2014

Supporting Learning by Tracing Personal Knowledge Formation

Thaul, Witold

http://hdl.handle.net/10026.1/3030

COPYRIGHT STATEMENT

Supporting Learning by Tracing Personal Knowledge Formation

by

WITOLD THAUL

A thesis submitted to the Plymouth University in partial fulfilment for the degree of

DOCTOR OF PHILOSOPHY

School of Computing and Mathematics

Faculty of Science & Environment

In collaboration with

Darmstadt Node of the CSCAN Network

at University of Applied Sciences Darmstadt

May 2014

Abstract

Supporting Learning by Tracing Personal Knowledge Formation

Witold Thaul

MBA, Dipl.-Ing. (BA), B.Sc. (Hons)

Internet-based and mobile technologies enable new ways of learning. They offer us new possibilities to access an enormous amount of knowledge at any time and everywhere. Among many advantages, the adaptations require a rethinking of our previous learning behaviour patterns and processes. The challenge for students is no longer to get access to information and knowledge, but to select the right one and to deal with the information and knowledge overflow.

The aim of this research is to define, design and validate an advanced concept to support the contemporary learning processes. Therefore, the requirements for a new approach have been assessed, the available solutions from the related area of (personal) Knowledge Management have been investigated, and the weaknesses in the context of learning identified. The identified issues have been substantiated by university students via a quantitative survey. Besides several smaller aspects, knowledge fragmentation and the nescience of the knowledge formation process have been classified as the most critical ones. To overcome these problems, a methodological concept has been developed, and a corresponding technological design created.

The chosen approach is an intelligent, independent intermediate layer, which traces the different steps our knowledge entities are going through. Based on personal and individual configurations, the system provides a comprehensive and overall observation of nearly all our knowledge work activities. It supports the building and accessing of the knowledge formation paths for every important knowledge unit, later path combination and the access to automatically generated versions of our work. Moreover, it helps the users not only to remember what they did, but also gives them some strong indications why they did it. This is achieved by combining different knowledge actions and looking at the influences they have on each other. The suggested concept has been critically proved and confirmed via a qualitative expert analysis and backed up by a quantitative survey among university students.

List of Contents

List of Figures

List of Tables

List of Abbreviations & Acronyms

Apache	The Apache Software Foundation, operator of apache.org
API	Application Programming Interface
CEO	Chief Executive Officer
CSCAN	Centre for Security, Communications and Network Research
DFKI	Deutsches Forschungszentrum für Künstliche Intelligenz GmbH, German Research Center for Artificial Intelligence
DIKW	Data, Information, Knowledge and Wisdom
EPFL Lausanne	École polytechnique fédérale de Lausanne
ER	Entity–Relationship
flickr	Online photo management and sharing application from Yahoo!
FK	Foreign Key
FZI Karlsruhe	Forschungszentrum Informatik Karlsruhe
GB	Gigabytes / Gigabyte
H_DA	Hochschule Darmstadt, University of Applied Sciences
HP	Hewlett-Packard Development Company, L.P.
IBM	International Business Machines Corporation

ICCS-NTUA	Institute of Communications and Computer Systems at the National Technical University of Athens
IM	Information Management
KM	Knowledge Management
KMS	Knowledge Management System
LaTeX	Markup Language for the TeX Tpesetting Program
L3S Hannover	L3S Research Centre Hannover
Mandriva	Mandriva S.A.
MB	Megabytes
M.I.T.	Massachusetts Institute of Technology
Mozilla	Mozilla Foundation, a non-profit software organization
MS	Microsoft Corporation
NEPOMUK	Networked Environment for Personal Ontology-based Management of Unified Knowledge
NUI Galway	National University of Ireland, Galway
OM	Organisational Memory
OMS	Organisational Memory System
OU	The Open University, UK

PIM	Personal Information Management
PK	Primary Key
PKM	Personal Knowledge Management
SlideShare	SlideShare Inc., operator of slideshare.net
Thunderbird	Mozilla Thunderbird, a free, open source, cross-platform email, news and chat client
Tweet	Text-based messages in the social network Twitter
UCLA	University of California, Los Angeles
USQ	University of Southern Queensland, Australia
WebCT	Web Course Tools
XML	eXtensible Markup Language

Acknowledgements

The work, and indeed this PhD, would not have been possible without the help and support of my Director of Studies, Prof. Dr. Udo Bleimann. Thanks go to him for his tireless effort in steering me through the PhD process, from publishing papers to presenting at international conferences. His professionalism and experience has been invaluable, and I owe much of my success to his guidance.

Thanks must also go to Dr. Nathan Clarke, my other supervisor, who has spent a lot time proof reading papers and my thesis, in addition to providing very helpful experience and guidance throughout my studies.

Finally, I wish to thank my family, especially my wife Jessica for her support and patience over the past six years. She believed in me even when I was in doubt. Thank you!

Author's Declaration

At no time during the registration for the degree of Doctor of Philosophy has the author been registered for any other University award.

The postgraduate research project submitted for this research degree at the Plymouth University has not formed part of any other degree either at Plymouth University or at another establishment. It has been conducted and supervised under the agreement between the Plymouth University and the University of Applied Sciences Darmstadt concerning the establishment of the Darmstadt Node of the CSCAN Network at the Hochschule Darmstadt University of Applied Sciences (H_DA).

Relevant scientific seminars and conferences were regularly attended at which work was often presented, and several papers were prepared for publication.

Word count of main body of thesis: 47.310

Signed: _____

Date: _____

1   Introduction & Overview

This chapter introduces the area and content of this research. It provides an overview of the main issues associated with the subject of study. It then proceeds in establishing the aims and objectives of the research, followed by a brief summary of the thesis structure and the contents of each chapter.

1.1   Introduction

Internet-based and mobile technologies enable new ways of learning. They offer us new possibilities to access an enormous amount of information and knowledge at any time and everywhere. Among many advantages, the adaptations require a rethinking of our previous learning behaviour patterns and processes. The challenge for students is no longer to get access to information and knowledge, but to select the right one and to deal with the information and knowledge overflow.

The origin of this research can be found in the consideration of different approaches to support university level students in the changed learning environment. *Personal Knowledge Management* is a scientific discipline that concentrates on methodologies, processes and computer systems to support the individual in dealing with his personal information and knowledge base. The research presents a novel Personal Knowledge Management approach to support students and other knowledge workers to cope with these adaptations. The technology-supported application of Personal Knowledge Management methodologies in the academic learning environment is the underlying approach of this research.

1.2 Aims & Objectives

The changed learning environment requires students to deal with a different amount and quality of information and knowledge. Technology driven aspects require also an adaptation of the user behaviour. Current methodologies and systems do not fit entirely into the new working patterns and in many cases do not support the users sufficiently.

The aim of this research is to define, design and validate an advanced concept to support the contemporary learning processes. This has been achieved by assessing current approaches, identifying problematic issues, and defining and designing an applicable methodology and technological architecture to benefit students.

In this context, the following question has to be investigated:

How can students be better supported from a methodological and technological

point of view in the evolving learning environment?

In order to achieve this, the research can be divided into the following phases:

1. Assessment of the requirements for a new and advanced approach to support knowledge work within today's learning environments.

2. Investigation of existence and applicability of available approaches from the area of Knowledge Management in today's learning environments.

3. Design of a beneficial and applicable methodology to reduce existing problematic issues.

4. Design of a technological architecture to support the aims of the suggested methodology.

5. Confirmation of the feasibility and benefits of the theoretical concept and the technological system design.

The first phase provides a comprehensive review and discussion of the changed learning environment. It argues why there is a need to assist learners. This identified, the second phase proceeds with a literature review and evaluation of existing methodological and technological approaches. The outcome of this phase is to identify existing problems within the research area. In order to achieve the objective of a beneficial and applicable methodology, a new conceptual approach is being developed within the third phase. The fourth phase introduces a possible system design to implement the theoretical conceptual ideas. The final phase of the research is to evaluate the theoretical, as well as practical and technological concepts of the suggested approach.

1.3 Methodology

The previously introduced five phases of the research have been conducted according to the following research methodologies: the frame approach applied during the entire research is based on the participatory research (Lincoln 2001) methodology. With the goal to give university level students a better support within their learning processes, the idea of conducting research with and for research subjects (Truman & Pain 2013) has been met. According to Truman and Raine (2001), there are different modes of participation reflecting the different natures of user involvement. Table 1 outlines them and their impact on the user participation.

Table 1 : Modes of Participation in Participatory Research (Lincoln 2001)

Mode of Participation	Nature of User Involvement
Co-option	Token; representatives are chosen, but no real action
Compliance	Tasks are assigned, with incentives; researchers decide agenda and direct the process
Consultation	Users' opinions asked, researchers analyse and decide on a course of action
Co-operation	Users work together with researchers to determine priorities; responsibility remains with researchers for directing the process
Co-learning	Users and researchers share their knowledge to create new understanding and work together to form action plans with research facilitation
Collective Action	Users set their own agenda and mobilize to carry it out, in the absence of outside researchers or facilitators

For the present research, the mode *consultation* has been applied on different levels.

Furthermore, the identified results of phase one (assessment of the requirements), as well as of phase two (investigation of existence and applicability of available approaches) have been substantiated by university students via a quantitative survey. The scope of the survey is described in detail in section 4.4.

The conceptual outcome of the research resulting from phase three (design of a new beneficial and applicable methodology) and phase four (design of technological architecture) has been critically proved and confirmed via a qualitative expert analysis. Based on the design of a phenomenological study (Leedy & Ormrod 2005), with the main goal to understand the experience, ideas and thoughts, experts in the area of research have been identified. These experts have been interviewed and their opinions and beliefs on the theoretical and the practical aspects of the study analysed. The methodology and the scope of the concept confirmation, as well as the selection of the sampling are described in detail in chapter 7.

To ensure, that the main target group of the whole research, university level students, really benefit from the outcomes of the idea, the main thoughts have been confirmed via a quantitative student survey. The scope of the survey is described in detail in section 7.5 of the thesis.

1.4 Thesis Structure

The thesis addresses the outcomes of the aforementioned objectives in order and consists of the following chapters:

Chapter 2 provides a background to the changed learning environment. It discusses new technologies and internet driven ways of learning. The chapter continues describing problems resulting from these changes. It illustrates some challenges students, as well as universities and schools are facing. Putting the focus on the information and knowledge overflow, the chapter concludes that there is a need to antagonise the impacts by providing learners with better methodological and technological support.

Having firmly established the need for a methodological and technological support of learners, chapter 3 is dedicated to approaches that support learning. It begins with a justification why the research area of *Knowledge Management* is suitable to face the identified problems. Following, an introduction to the topic of Knowledge Management (KM) is given. The chapter continues by addressing the different areas of KM, putting the focus on the management of personal information and knowledge. Software systems, which support personal information and Knowledge Management activities, are addressed. Finally, an overview of semantic desktops and recommender systems methodologies is given.

Chapter 4 discusses the application of *Personal Knowledge Management* tools within the academic area. In this context, it identifies and reviews the existing problems. The chapter determines knowledge fragmentation and the nescience of the knowledge formation process as the main problematic issues and discusses these.

Based upon the findings, chapter 5 presents a new approach to overcome these problems in a sustainable manner. The suggested idea is based on the traceability of our knowledge formation. Designed as an independent layer, the idea allows a comprehensive and overall observation and consideration of nearly all our knowledge activities. It allows the building and accessing of the knowledge formation paths for every important knowledge unit, as well as the later path combination and the access to automatically generated (intermediate) versions of our work. Moreover, it helps the user not only to remember what he did, but also provides him with an indication of why he did this.

In chapter 6, the technological design of the new approach is described. The system consists of three main units and the database. The chapter introduces the *Configuration Unit*, the *Knowledge Storage Unit*, as well as the *Knowledge Access Unit*, defines their tasks and outlines their technological implementation design. Furthermore, the database design is introduced and explained.

Chapter 7 presents a confirmation of the suggested concept – both from a theoretical point of view concentrating on the idea, as well as from a technological point of view putting the focus on the implementation design described in chapter 6. The results from the confirmation process are being considered to enhance the actual system approach.

Finally, chapter 8 presents the main conclusions arising from the research, highlighting the principle achievements and limitations of the work. The chapter contains a discussion on suggestions for potential future research and development.

The thesis also includes a number of appendices containing additional information to support the main discussion.

2 The Evolving Learning Environment

The previous chapter introduced the context of this research and explained the aims and objectives. The identified research problem of this thesis is located in the educational area. For this reason, the following chapter introduces the area of learning and its environments. It discusses the challenges and problems students are facing nowadays, putting the focus on the knowledge and information overflow.

2.1 Introduction

The Internet changed the way we access and retrieve information and knowledge decisively. It also changed the way we learn. On the one hand, this offers many advantages and new possibilities; on the other hand, it poses many challenges. Existing methods, working patterns, as well as the supportive systems we are using need to be questioned and where necessary adapted.

2.2 Learning & its Environments

Oxford (Oxford 2013) defines learning as 'the acquisition of knowledge or skills through study, experience, or being taught'. According to Kolb (Kolb 1984), learning is 'the process whereby knowledge is created through the transformation of experience'. For Piaget (Simatwa 2010) learning is 'the mutual interaction of accommodation and assimilation'. He sees accommodation as the adaptation of our mental concepts based on the experi-

ences made by others. Assimilation on the other hand is the integration of our own experiences into existing mental concepts.

From a high-level perspective, learning can be reduced to the following two parts Know-How and Know-Why (Klein 2012). In the context of learning, Know-How is the acquisition of skills, and implies the physical ability to produce an or some actions. Know-Why on the other hand 'implies the ability to articulate a conceptual understanding of an experience' (Klein 2012). Although these approaches might still be valid on a conceptual high level, new technologies changes many concrete processes, workflows and steps linked to learning during the last two decades.

New ways of communication and presentation allow us to move from pure classroom based learning to concepts like e-Learning (Sendall et. all 2008) or the mixed form of blended learning (Strauss 2012). Even in the long-established area of distance education by mail the new web-based concepts have substituted former ways of communication, intermediation or for example assessment. This offers students new ways of studying and allows them to study from nearly any location worldwide in a highly time flexible and adaptable way. It allows students to learn according to the individual needs, affections or live situation (Lu et. al 2002). It is no longer a requirement to be on campus, or to have spare time during scheduled sessions. Nowadays, it is up to the student to decide when and where to learn which aspects, the institutions provide just a framework. The *Open University* in the UK (OU 2013) and the *University of Southern Queensland* in Australia (USQ 2013) are just two examples of established universities that focus on people worldwide who like to or need to study at a high educational level with approved standards, but in adaptable (to personal needs) and more flexible environments. WebCT (Web Course

Tools), discussion boards, video conferences, emails, instant messaging, collaboration tools, or video recorded on demand lectures are just some examples of how internet based technologies can be used nowadays to support learning flexibility (Lu et. al 2002).

However, this domain is still growing. Social networks like *Facebook* (Facebook 2013, Schock 2010) or *Twitter* (Twitter 2013, Academhack 2010, Young 2009) are gaining academic importance, and the integration of the different media and technology types is a challenge schools and universities need to meet. The access, acquisition, transmission, creation and development of knowledge, as a vital asset, affect methods and strategies of teaching, learning and research (Zuber-Skerritt 2005).

Easier communication is just one aspect resulting from the new technologies. Another and probably an even greater one is the easy access to information and knowledge. Ten years ago, university students usually had to spend a lot of time in libraries to search physically in books for information they needed. Often the required information source was not available, difficult to provide, or even strongly outdated. Services like *Wikipedia* (Wikipedia 2013a), *Google Scholar* (Google 2013a) and thousands of other websites offer contemporary instant access to information – mostly even for free and accessible even on the go by mobile devices.

As great as it sounds, there are several problems resulting from these new possibilities, from which at least two need to be considered here. Firstly, most of the information found is not validated. Usually it takes years before a book is published, it takes at least months before a journal paper is out, but it takes just a few seconds to publish a new post in a blog or to adjust a *Wikipedia* entry. The fact that it takes some months or even years for a book to be published has reasons, and probably the most important one is validation and correc-

tion of the contents. Most journal or conference papers are checked and double-checked by several reviewers who are typically experts in the related academic area. Entries within the Wikipedia encyclopaedia can be edited online live, are however assured by topic dedicated editors. A simple blog post on the other hand – even from an expert in his area – is typically published without any confirmation of correctness. The positive aspect is, that this way we do not only have access to much more information, but also to current information. However, the quality of the published information and knowledge might be questioned. Moreover, this behaviour enables a (partially) shifting of the duties regarding the validation and assessment of the correctness and value of the provided information and knowledge – from the publisher to the consumer.

The second and maybe even bigger problem is the enormous amount of available information and knowledge. This often results in an information and knowledge overflow. The next section introduces the area of information and knowledge overflow and discusses the associated issues.

2.3 Information & Knowledge Overflow

The second problem is the enormous amount of available information. To visualise the growth, it is interesting to have a look at the amount of information we create. Whilst 2000 every person worldwide produced 250 MB of recorded information (Lyman and Varian 2000), in 2003 that had grown to 800 MB (Lyman and Varian 2003), and in 2007 to around 43 GB (Gantz et al. 2011). More recently, the consumption of information rose extremely: in 2009, every US-American consumed 34 GB of information on an average day (Bohn and Short 2009). 2010 Eric Schmidt, the CEO of *Google*, stated at the *Techonomy Conference* in Lake Tahoe, that 'every two days we now create as much information as we did from the dawn of civilization up until 2003' (Siegler 2010).

The blue curve in Figure 1 demonstrates the volume and the evolution of the available hostnames (internet addresses). All of them could act as potential information and knowledge sources on the Internet. Even if we look at the currently active sites (red curve in Figure 1), we presently still have the possibility to access around 200 million websites.
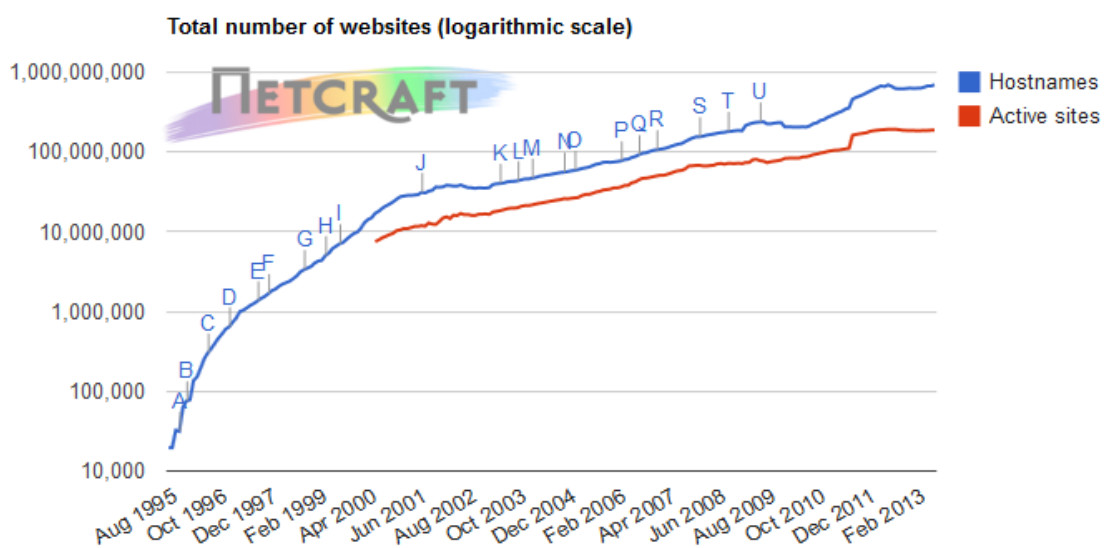
Figure 1 : Total Number of Websites (Netcraft 2013)

The enormous amount of information results not only from the growth in the number of information sources, but also in the increasing information that can be accessed from many of them. This become especially true, when the provided services on the available website started to offer not only the possibility to consume information, but also to participate and provide new information. Figure 2 illustrates the number of available English *Wikipedia* articles, which rose to around 4.2 million over the past twelve years. No other (traditional / printed) encyclopaedia has currently the possibility to react to incidents or changes as quickly as *Wikipedia* (Wikipedia 2013a). However, at the same time, there are many risks resulting from this. Manipulation attempts in the past have showed how big the impacts might be (Orlowski 2005, Singleton 2009).
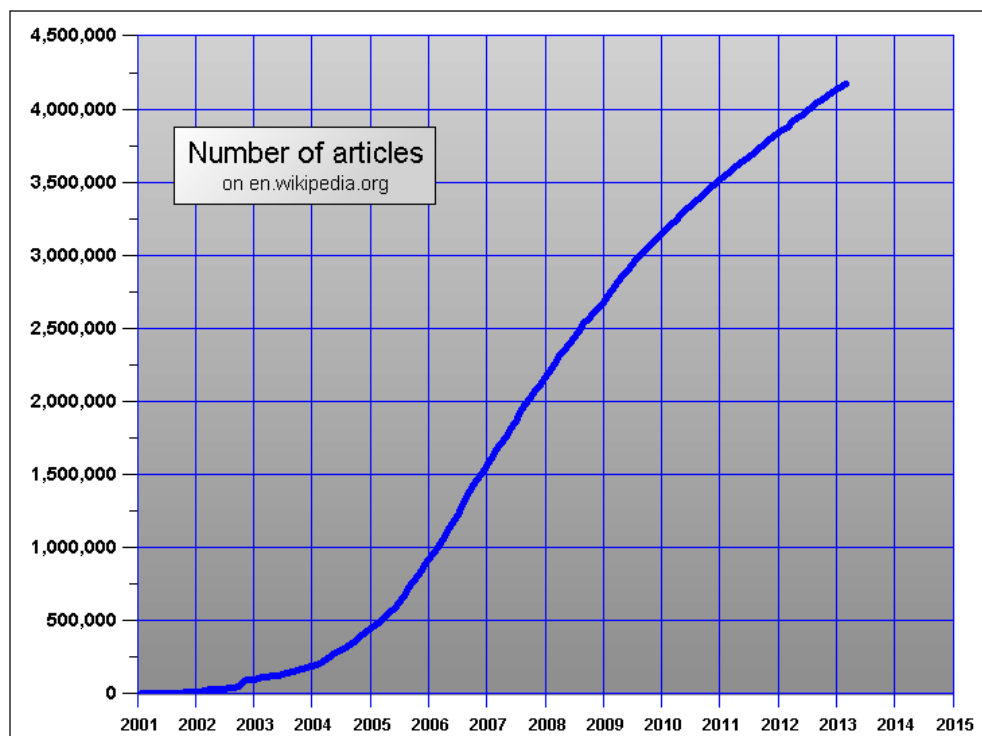
Figure 2 : History of Wikipedia (Wikipedia 2013b)

Whereas some years ago the trend was to have many different websites run by individuals, the area of social networks has shifted this approach towards the publication on cen-

tralised platforms. Social networks are also gaining also more and more importance in the academic area. Taking *Twitter* as an example demonstrates clearly, where the problems lay. On the one hand, you can find the latest information on a topic pretty quickly; on the other hand, you can find millions of irrelevant posts. To allow a more sufficient organisation, navigation and search of relevant tweet posts, hash tags have been introduced. They allow the comprehensive aggregation of related topics and discussions, and by doing this help the user to filter the for him relevant information.

Figure 3 illustrates the trend of tweets published on twitter every day. Beginning of 2010 the number almost reached the 50 million tweets per day. On the 6th of July 2012 Twitter's CEO *Dick Costolo* announced 400 million tweets per day number (Farber 2012), three months later this number had risen to 500 million (Terdiman 2012).
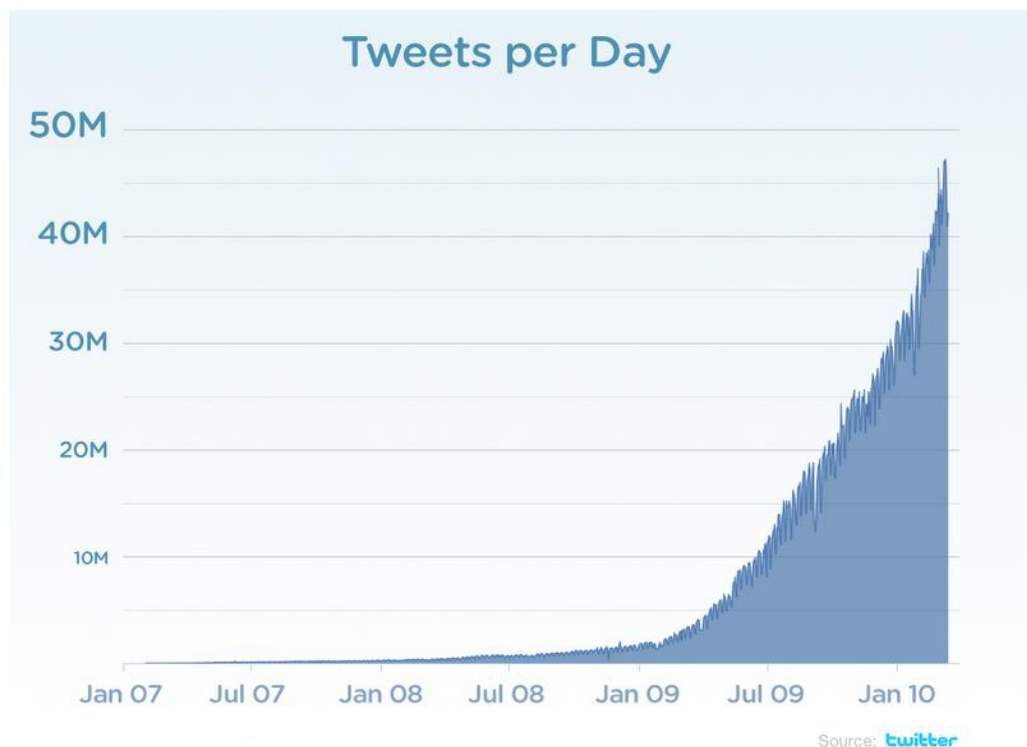
Figure 3 : Twitter - Amount of Tweets per Day (Weil 2010)

To visualise this, we can imagine a stack of books as high as two *Taipei 101* buildings published just on *Twitter* every single day (Figure 4). Extending this with entries on other platforms like *Facebook*, *LinkedIn*, *Evernote*, *CiteSeerX*, *ResearchGate*, or for example *SlideShare*, as well as on individual websites, blogs or discussion boards, the amount of available information is just enormous. In June 2013, *Facebook* for example had a total number of 1.11 billion users, who in average upload 350 million photos every day. Beside the personal profiles, there exist about 50 million pages, with an average of 36 monthly posts on every of them (Smith 2013). The challenge is no longer to get access to information and knowledge, but to find the right one. Depending on the individual behaviour patterns and methods, it is up to the individual to find ways to identify and manage his personal information and knowledge.
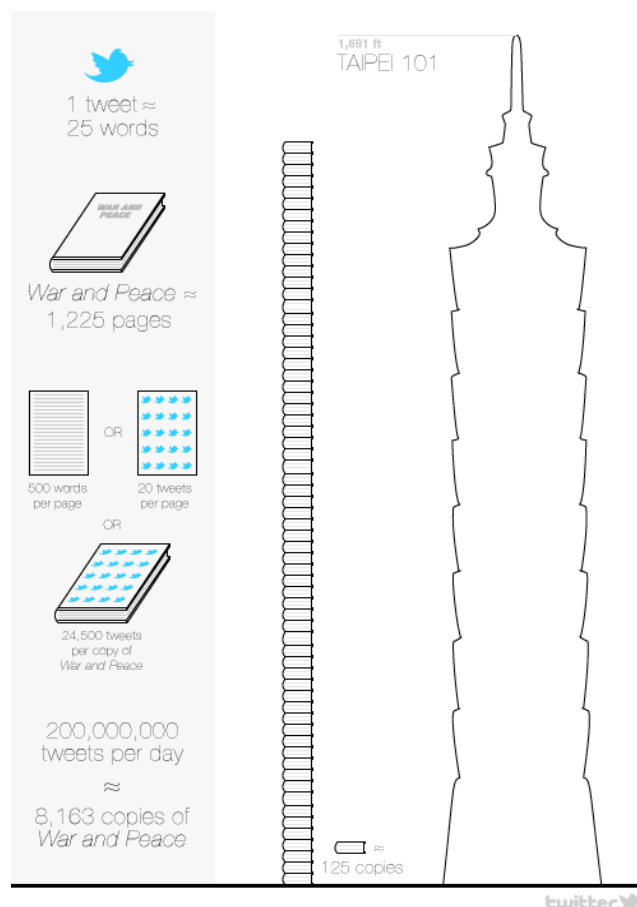
Figure 4 : 200 Million Tweets per Day (Twitter 2011)

As stated by *William Pollard* already decades before, 'information is a source of learning. But unless it is organized, processed, and available to the right people in a format for decision making, it is a burden, not a benefit' (PowerfulInformation 2013).

2.4 Summary

Internet based technologies have changed the way we learn. They allow us to study on a much more individual focused basis, and to adopt the way we learn to the life circumstance in a wide range. At the same time, the Internet allows us instant access to a huge amount of information and knowledge. However, exactly here lay the challenges and the risks. The paradox of choices which information to use, where to find it, and the question how trustable it is, are just some questions students are confronted with. Unfortunately, most of the information is not validated, and even if it seems to be, typically there does not exists any organisation that makes sure it is not being manipulated. Moreover, with the easy and fast access to the currently available information and knowledge sources students encounter an information and knowledge overflow. Currently there is very 'little knowledge about how people learn' in this new learning environment and 'how to facilitate their leaning to learn' (Zuber-Skerritt 2005), and at the same time it is wrong to expect traditional education systems 'to provide all the capacity and methods needed to support new-millennium knowledge workers' (Alley 1999).

3 Approaches to Support Learning

The previous chapter introduced the area of learning and learning environments. It reviews the changes on students' behaviours and processes resulting from the impact of the web-based and mobile technologies. The following chapter introduces approaches to support students in meeting the described challenges and problems.

3.1 Introduction

Meeting the resulting challenges can be faced differently. Learning and knowledge are linked very tightly to each other, and represent the same phenomenon observed from different perspectives (Reinmann & Eppler 2008). According to Steiner (Steiner 2006) learning represents an event or an action that cannot be directly observed, but leads to a relatively stable change in our knowledge. For that reason, the application of *Knowledge Management* (KM) concepts and methodologies to support learning processes and face the indicated problems and challenges has been chosen within that research. This chapter therefore introduces the area of Knowledge Management.

It begins with an introduction to the topic of *Knowledge Management*. It introduces the different areas of Knowledge Management, putting a focus on the management of personal knowledge. The different perspectives of *Personal Knowledge Management* (PKM) are presented, followed by an introduction into *Personal Information Management* (PIM). Moreover, the distinction between Personal Knowledge and Personal Information Management is discussed. Then, software systems that support Personal Information and

Knowledge Management activities are addressed. Finally, an introduction to the related research area of *Semantic Desktops* is given, and an overview of *recommender system* methodologies is given.

3.2 Knowledge

In 1985 Peter Drucker (Drucker 1985) for the first time mentioned the term *Knowledge Worker* while comparing the contribution of management in the 20th and the 21st century. He stated that as successful it was to increase manual worker productivity in the past, so important it will be to increase knowledge worker productivity in the future.

It is true, that by passing skills from generation to generation it can be argued that there have always been knowledge workers, but these skills mostly had to be learned once only. In general, they changed very little during the worker's lifetime. Today's knowledge workers on the other hand are not only required to have a much wider knowledge base, but their skills also require constant renewal (Barth 2004).

From talking about *Knowledge Economies* and *Societies*, the idea and necessity emerged of managing knowledge and *Intellectual Capital* (OECD 1999; Stewart 1997) as a vital asset and as a valuable resource (Zuber-Skerritt 2005). Not only work environment equipment, but also the everyday devices in our private lives become more complex (Wiig 2004). To face this, different approaches have been taken. They are based on diverse perspectives, but mainly focus upon the *technological*, *economic*, *sociological* and the *psychological* (Roehl 2000) aspects. These four perspectives build the fundament for a

holistic Knowledge Management perspective and are often referred to the *Pillars of Knowledge Management* as presented in Figure 5.

- The technological track was mainly found on the improvements of information- and telecommunication systems (Krcmer 1996); the referring pillar is the *technology* (Figure 5).

- The economic approach refers mainly to the question in which way knowledge and information can be organised and applied to maximise efficiency and increase the economic benefit (North 2005); the corresponding pillar is the *organisation* (Figure 5).

- The psychological perspective on the management of knowledge is putting the individual person and his phonemes like cognition, motivation, emotion, and social interaction in the focus (Reinmann & Eppler 2008). However, compared with the other approaches, the psychological perspective is only of rudimentary nature (Reinmann & Mandl 2004). The psychological approach refers to the pillar *human* (Figure 5).

- The sociological approach refers to a theoretical description of the management of knowledge within organisations. Furthermore, it models organisations as social systems (Wilke 2001); the referring image is the surrounding *culture* (Figure 5).
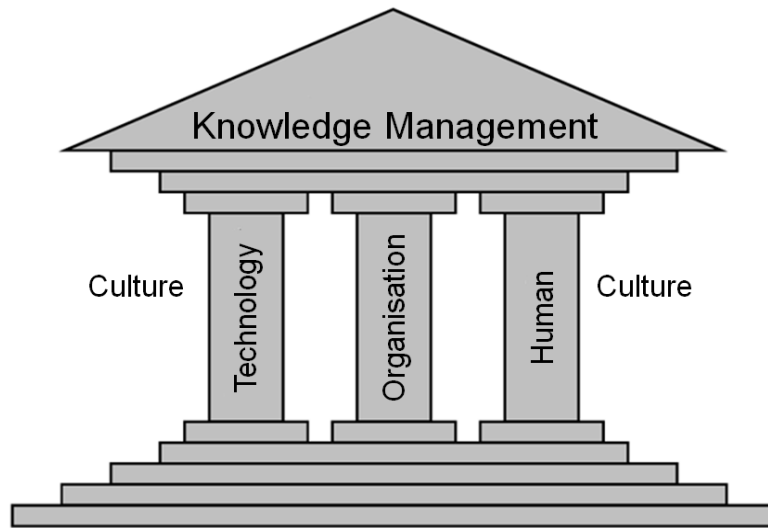
Figure 5 : Knowledge Management Pillars (adapted Baumann 2001)

This represents just one approach; however, there exist also others, like for example the four pillars categorisation (*Leadership*, *Organisation*, *Learning* and *Technology*) from the George Washington University (Stankosky 2005).

3.2.1 Data, Information, Knowledge

It is important to emphasize, that there is a need to distinguish between *data*, *information*, and *knowledge*. For Drucker (Drucker 1985) *data* is a set of raw facts. For Davenport (Davenport 2000) these facts are discrete and objective and are referring to events. *Data* itself, has 'little relevance or purpose'. In the organisational context, data is often associated with 'simply structured records of a transaction' (Tiwana 2002).

Endowed with 'relevance and purpose' (Davenport 2000) *data* becomes *information*. This happens by adding value or meaning to the data, by processing it (Tiwana 2002).

According to Tiwana, (Tiwana 2002) this can be done in different ways, for example by:

- *Contextualising* the data

  (explaining why the data was collected)

- *Categorising* the data

  (identifying key components or units identified)

- *Calculation* of the data

  (analysing mathematically or statistically)

- *Correcting* the data

  (removing errors)

- *Condensing* the data

  (summarising the data in a more useful form)

Actionable information can be seen as *knowledge* (Tiwana 2002). According to Wiig (Wiig 2004), the main idea and goal of information is its ability to describe, but the main idea and goal of knowledge is geared towards action. Davenport and Prusak define *knowledge* as 'a fluid mix of framed experience, values, contextual information, and expert insight that provides a framework for evaluating and incorporating new experiences and information' (Davenport and Prusak 2000). It is the key resource in intelligent decision making, and is much more expansive, deeper and richer than information (Tiwana 2002). Nonaka defines knowledge as a 'dynamic human process of justifying personal belief toward the truth' (Nonaka and Takeuchi 1995). Knowledge therefore can be seen as information in the context of human beings. According to Huber et al (Huber et. al 1998), knowledge is about context-specific meaning. Jennex goes even further by extending the concept of context by including associated culture that 'provides frameworks for under-

standing and using knowledge' (Jennex 2006). Transferring information to knowledge relates to (Davenport 2000):

- *Comparing*

  the information in the current situation to other experienced situations

- *Reflecting*

  the consequences the information might have on decisions or actions

- *Connecting*

  the information to others

- *Discussion*

  of the information with others (conversation)

Knowledge consists of several components like *experience*, *truth*, *judgment*, *intelligence*, and *rules of thumb*. *Experience* gives us a historical perspective to understand new situations. Analogue, *rules of thumb* represent answers to new problems based on previously solved problems. *Truth*, often also described as *ground truth*, reflects the fact of knowing what really works and what does not. Unlike information and data, knowledge is always based on *judgement*, and applied where needed and in a way that leads to a better performance, knowledge also qualifies as *intelligence* (Davenport 2000, Tiwana 2002, Davenport 2005).

Figure 6 represents the knowledge pyramid by Hey (Hey 2004). It demonstrates the dimensions *data*, *information* and *knowledge*, according to the maturity of their values.
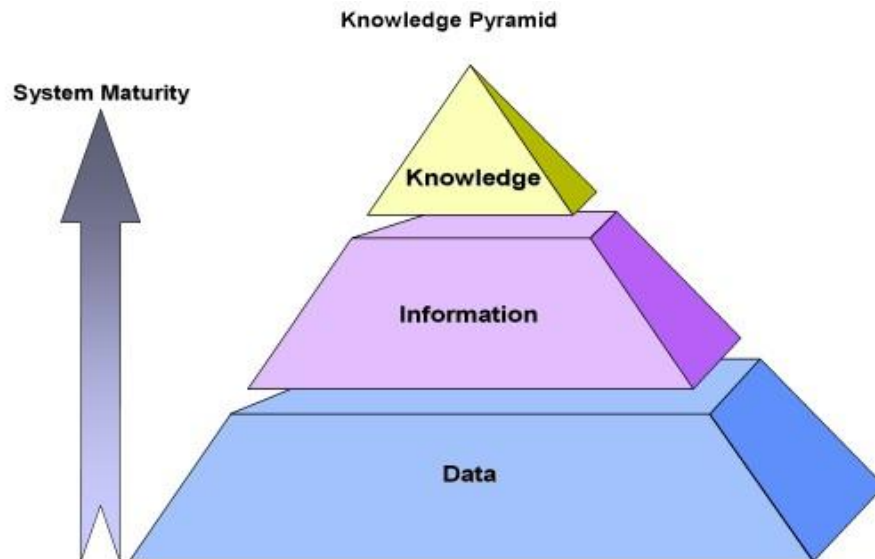
Figure 6 : Knowledge Pyramid (Hey 2004)

This hierarchy is often also related to the model of *DIKW* (Figure 7), which stands for *Data*, *Information*, *Knowledge* and *Wisdom*. It goes even a step further than Hey's Knowledge Pyramid, by extending the context with *wisdom*. *Wisdom* can be seen as the application of knowledge – focusing on novelty and the future (Clark 2004).
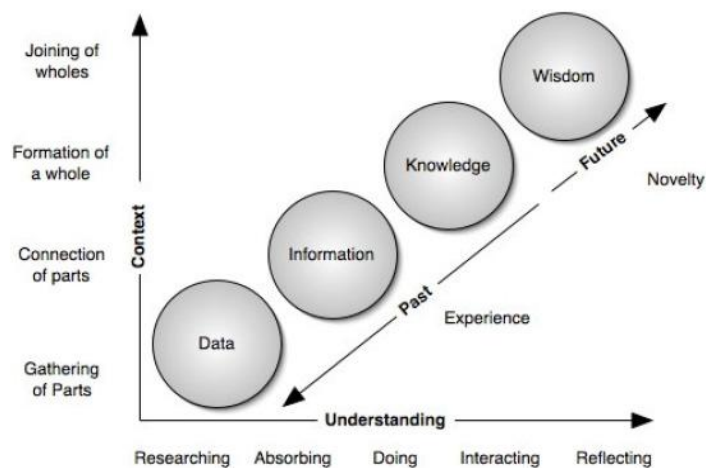
Figure 7 : DIKW Hierarchy (Clark 2004)

3.2.2 Tacit & Explicit Knowledge

According to Tiwana, there exist four key dimensions that allow the categorising of knowledge: *Type*, *Focus*, *Complexity* and *Perishability over Time* (Tiwana 2002). Table 2 characterises these knowledge dimensions:

Table 2 : Knowledge Dimensions & their Characteristics

Knowledge Dimensions	Characteristic
type	- technological - business - environmental
focus	- operational - strategic
complexity	- explicit - tacit
perishability over time	- low - high

The most known and applied (Alavi and Leidner 2001) classification is the one according to the *complexity* (Polanyi 1967, Nonaka 1994). *Explicit* and *tacit* knowledge are the possible values. As Polanyi (Polanyi 1996) stated, humans typically know more than they can tell. This part of knowledge that can be expressed in a systematic and formal language is called *explicit knowledge* (Nonaka and Takeuchi 1995). *Tacit knowledge* on the other hand is nearly impossible to reproduce in a document or database (Davenport and Prusak 2000). It is in the knower's head (Tiwana 2002) and is regarded as the most valuable and

actionable knowledge part (Nonaka and Takeuchi 1995). The knowledge iceberg meta-

phor in Figure 8 illustrates the relationship between explicit and tacit knowledge. While

around 95% of the entire knowledge can be seen as tacit knowledge based on experience,

thinking, competence, commitment and deed, only about 5% out our knowledge can be

seen as the explicit one (CDS 2010).

Figure 8 : Knowledge Iceberg Metaphor (CDS 2010)

3.2.3 Knowledge Management

Knowledge Management (KM) is a process of acquisition, sharing and utilization of knowledge. For Davenport Knowledge Management 'is the process of capturing, distributing, and effectively using knowledge'(Davenport and Prusak 1998). The Gartner Group (Duhon 1998) defines Knowledge Management 'as a discipline that promotes an integrated approach to identifying, capturing, evaluating, retrieving, and sharing all of an enterprise's information assets. These assets may include databases, documents, policies, procedures, and previously un-captured expertise and experience in individual workers'. Based on the soliciting definitions of the review board members of the *International Journal of Knowledge Management*, Jennex (Jennex 2005) defines Knowledge Management as 'the practice of selectively applying knowledge from previous experiences of decision making to current and future decision making activities with the express purpose of improving the organization's effectiveness'.

Knowledge management can be seen from an *individual* (*personal*), an *organisational*, and a *social* perspective. When talking about Knowledge Management nowadays, most people refer to the organisational perspective and think of 'corporate variety in which organisations try to get a handle on the vast amounts of knowledge locked inside the minds of individual employees across the organisation' (Miller 2005). The main focus lies on the organisation and the way organisations can gather a sustainable competitive advantage (Davenport and Prusak 1998) out of what they know. In this context, KM has reached an established status as a management discipline during the two last decades (Abecker and Völkel 2008). Within the domain of Knowledge Management many models have been developed, like for example the *Knowledge Building Blocks Model* by Probst

(Probst. et. al 1997), the *Knowledge Spiral Model* by Nonaka and Takeuchi (Nonaka and Takeuchi 1997), or the *Knowledge Market Model* by North (North 2005).

The practical *Building Blocks Model* has been developed by Gilbert Probst (Probst. et. al 1997) and his colleagues at the *University of Geneva*. The basis of the model builds the eight simple, and easy to follow building blocks: Knowledge *Goals*, *Identification*, *Acquisition*, *Development*, *Distribution*, *Preservation*, *Use* and *Measurement*. Each of them is associated to one or several Knowledge Management instruments as shown in Figure 9.
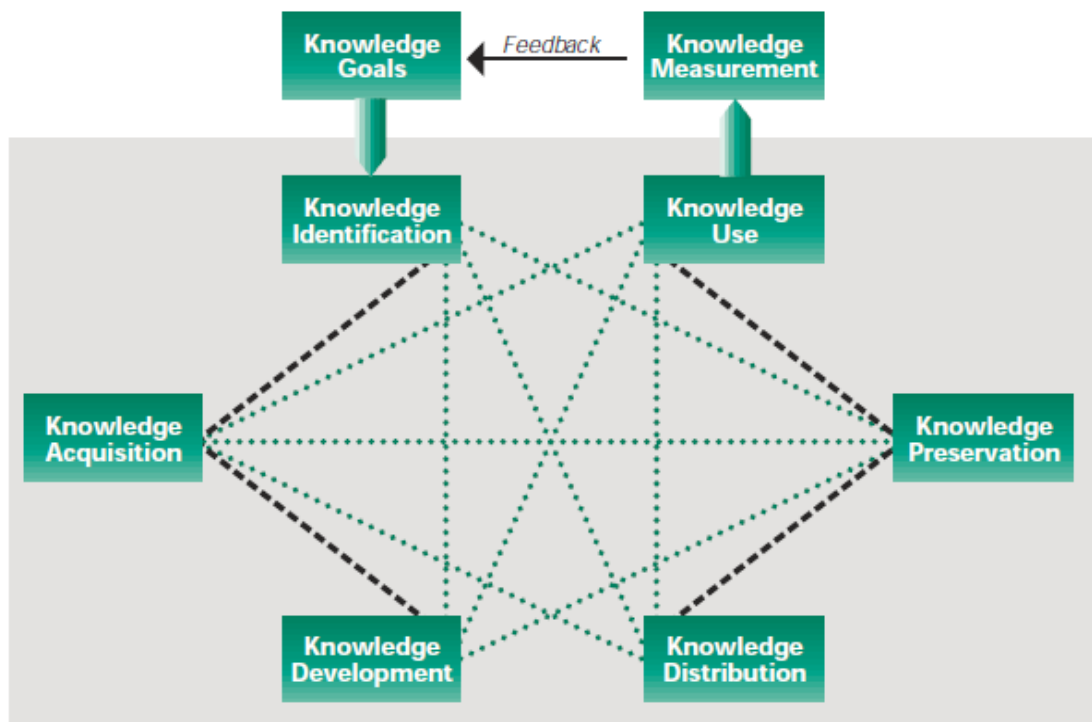
Figure 9 : The Building Blocks of Knowledge Management (Probst 1998)

Another commonly applied model is the *Knowledge Spiral Model* as described by Nonaka and Takeuchi (Nonaka and Takeuchi 1997). Based on empiric research within Japanese corporations, the aim of this theoretical model is to present how individual, per-

sonal knowledge can be integrated into the company's knowledge base. Figure 10 illus-

trates the several phases the spiral is passing.
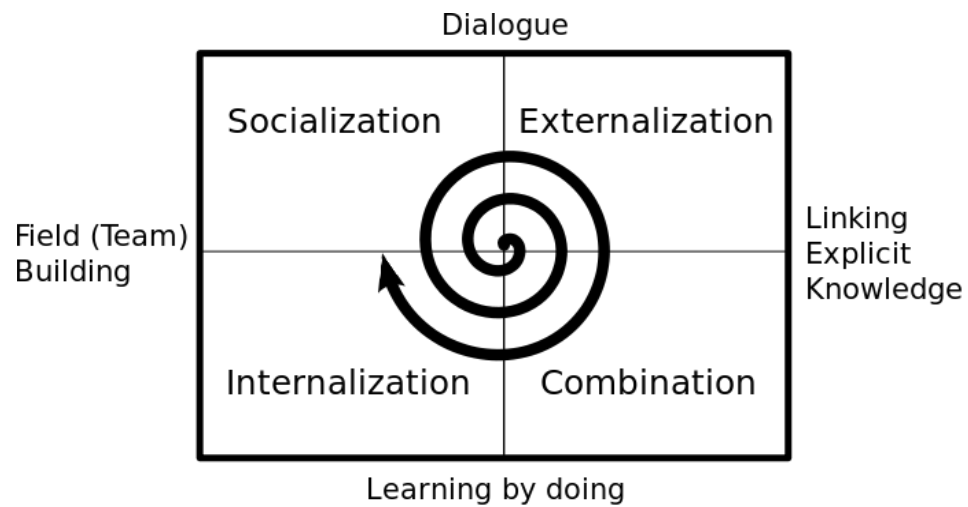
Figure 10 : Knowledge Spiral (Nonaka & Takeuchi 1997)

The *Knowledge Market Model* developed by North has been designed at the *University of Applied Sciences of Wiesbaden* within several practical projects (North 2005). The main idea of that model is the assumption, that knowledge is a limited resource; as such, it can and need to be treated and traded within organisations according to market mechanism like every other resource. From the companies' point of view, it is therefore necessary to develop and use the resource knowledge to improve their position in competition.

3.3 Personal Knowledge Management

As Davenport underlines, the knowledge-based society is 'no more effective than the sum of its knowledge workers' effectiveness' (Davenport 2005). This demonstrates the importance of not only concentrating on groups, organisations or societies, but to put the focus on the most important one (Abecker and Völkel 2008): the individual knowledge worker – in our case, the individual student or learner. The new generation needs to be more efficient in managing their own information and knowledge (Truch 2001) and at the same time take more responsibility for their own learning and growth (Smedley 2009). Personal Knowledge Management is essential to be responsible for what you know, whom you know, and what they know (Barth 2005).

In contrast to the traditional top-down organisational Knowledge Management, Personal Knowledge Management (PKM) focuses on the individual motivations and behaviour and describes a bottom-up approach to Knowledge Management (Pollard 2008). Unfortunately, compared with organisational KM, marginal research has been done on PKM more recently (Pauleen 2009).

Although the term *Personal Knowledge* has firstly been used in 1958 (Polanyi 1958), the origin of the discipline of Personal Knowledge Management can first be found in the University environment in 1999, in *UCLA*, Los Angeles (Frand and Hixon 1999) and later at the *Millikin University* in Decatur (Millikin University 2003).

One of the problems with PKM is that different people (Gurteen 2005) interpret it differently, which becomes obvious while looking at some of the following definitions:

- According to Frand PKM is a 'conceptual framework to organise and integrate information that we, as individuals, feel is important so that it becomes part of our personal knowledge base'. Moreover, it 'provides a strategy for transforming what might be random pieces of information into something that can be systematically applied and that expands our personal knowledge' (Frand and Hixon 1999).

- Paul Dorsey of Millikin University describes PKM as a set of seven skills (*retrieving information*, *evaluating information*, *organising information*, *collaborating around information*, *analysing information*, *presenting information* and *securing information*) that have a conceptual and a logical as well as a physical or a hands-on component (Avery et. al 2001).

- For Wright PKM is about 'connecting different types of problem-solving activities with specific cognitive, information, social and learning competencies, embedded within individual, social and organisational contexts' (Wright 2005, Wright 2007).

- Higgison defines PKM as 'managing and supporting personal knowledge and information so that it is accessible, meaningful and valuable to the individual; maintaining networks, contacts and communities; making life easier and more enjoyable; and exploiting personal capital' (Higgison 2005).

- Berkman says that PKM is about managing all the information that comes to your PC. He adds that it is also about contextual knowledge, which means making sense of that information and what it means to you; but that demonstrates how fuzzy the line between Knowledge and Information Management (section 3.4) is (Berkman 2005).

- Gurteen believes that PKM is more about how to use this information rather than how you organise it (Gurteen 2005). Martin understands PKM as 'knowing what knowledge we have and how we can organize it, mobilize it and use it to accomplish our goals – and how we can continue to create knowledge' (Martin 2006).

- For Tsui PKM is 'a collection of processes that an individual needs to carry out in order to gather, classify, store, search and retrieve knowledge in his/her daily activities. Activities are not confined to business / work-related tasks but also include personal interests, hobbies, home, family and leisure activities' (Tsui 2002).

All the found and here presented definitions point out, that Personal Knowledge Management is happening in a results-oriented, systematic and self-dependent way. A major role is taken by the planning activities, and enhanced by several processes regarding the handling and manipulation of information and knowledge.

Collaboration and cultural edge is part of some of the definitions. That points out, that PKM is not just focusing on individual knowledge workers as such, but also on collaborative, social and cultural aspects amongst them (Cakula and Osis 2010). Doing this, PKM assists as a basis for knowledge acquisition, sharing and usage (Martin 2006).

3.4 Personal Information Management

Following the previously introduced principals of the knowledge pyramid, Personal Knowledge Management is strongly related to the research area of Personal Information Management. Personal Information Management (PIM), first mentioned in the 1980s (Lansdale 1988), refers to activities people perform to manage their information. According to Jones Personal Information Management refers to 'both the practice and the study of the activities a person performs in order to acquire or create, store, organize, maintain, retrieve, use and distribute the information needed to meet life's many goals (everyday and long-term, work-related and not) and to fulfil life's many roles and responsibilities (as parent, spouse, friend, employee, member of community, etc.). Personal Information Management places special emphasis on the organization and maintenance of personal information collections in which information items, such as paper documents, electronic documents, email messages, web references, handwritten notes, etc., are stored for later use and repeated re-use' (Jones 2007).

Looking at the organisational part of Knowledge Management, the discussions are very similar. Nevertheless, within the personal area, the individual person himself defines where exactly knowledge starts and until when it is just information. What probably is information for most people, or from the point of view of an organisation, might already be knowledge for an individual. This becomes especially true when combining the value of the information with some other personal experiences (tacit knowledge), which others (individuals as well as organisations) might simply not have. Nevertheless, it is up to the individual knowledge worker to decide (consciously or otherwise) if it is knowledge for himself, or if this fragment is just a piece of information. In most cases, personal knowledge that for others is not more than information refers to explicit knowledge.

Until now, research regarding PKM can be divided into two categories. The first one looks at PKM as a part of whole organisation wide knowledge management systems and the question how it mutually interacts (Abecker and Völkel 2008). The second focuses on PKM tools and system, as well as their usage in the context on an individual (Li and Li 2009). The following section introduces and discusses the main aspect of the second approach – the technological systems behind PKM.

3.5 Personal Knowledge & Information Management Tools

For an effective management of personal knowledge it is necessary to use accordant technology (Barth 2004). Looking at information technology systems that support knowledge management activities, most of the research so far has been done from an organisational perspective. Stein and Zwass (1995) refer to *Organisational Memory Information Systems* (OMS), which can be used as a synonym to knowledge management systems (Alavi and Leidner 2001, Jennex and Olfman 2002), as the processes and IT components that are necessary to capture, store, and apply knowledge from the past on current and future decisions. Alavi and Leidner (Alavi and Leidner 2001) define *Knowledge Management Systems* (KMS) as 'Information Technology-based systems developed to support and enhance the organizational processes of knowledge creation, storage / retrieval, transfer, and application'. By taking Churchman's definition of a system, 'a set of parts coordinated to accomplish a set of gaols' (Churchman 1979), into consideration, Jennex and Olfman (Jennex and Olfman 2006) define KMS as 'a system that includes IT / ICT components, repositories, users, processes that use and / or generate knowledge, knowledge, knowledge use culture, and the KM initiative with its associated goals and measures'.

Personal Knowledge Management *tools* refer to software systems, which support Personal Knowledge Management activities. They should (Clemente and Pollara 2005) help the user to:

- Access what he knows

- Make clear what he needs to know

- Remember how he learned it

Most of the tools are mainly used to manage electronic documents, emails, internet bookmarks and other items, which the individual considers valuable (Frand and Hixon 1999). In order to be fully qualified as Personal Knowledge Management tools, the software systems should have at least some knowledge about the information and knowledge they manage, for example via semantic or ontology (Sauermann 2005). Therefore, according to a strong definition, most of them should be categorised as personal information management tools.

The Knowledge Management Magazine (KM Magazine 2000) identifies six categories of PKM tools, to which solutions could be classified to: *index, associative links, organise, meta search, collaboration services and web capture.*

Apshvalka (2005) describes PKM tools as complex systems that contain technological, social and psychological aspects. She divides PKM tools on three knowledge processes:

- Knowledge creation

- Knowledge codification

- Knowledge sharing

The operation of PKM tools is closely tied with 'knowledge workers perception, emotions, believes, surrounding society, environment, desires and goals' (Apshvalka 2005). Goal of PKM systems is to support as much as possible system's owner to make decisions and perform actions.

In 2004, Barth developed the Knowledge Management Magazine model further and identified the five categories: *indexing and searching* tools, *associative links and search* tools, *online meta-search* tools, *web capturing* tools and *organising and mapping* tools (Barth 2004). Table 3 briefly describes these categories.

Table 3 : PKM Tools (adapted from KM Magazine 2000, Barth 2004)

PKM Tool Category	Description
indexing & searching	Associated tools index the content of local hard drive, making complex and fast searches possible for the knowledge most likely to be immediately useful to the user.
associative links & search	Associated tools scan the content of the document the user is working on and capture relevant additional information from the Internet.
online meta-search	Associated tools send simultaneous queries to many individual search engines at the same time and deliver the results in a single list of hits.
web capturing	Associated tools simplify the clipping of web pages from the Internet for offline reading, archiving and reuse.
organising & mapping	Associated tools store, attach and annotate information in context by organising it into networks, hierarchies, and reports for later reuse.
collaboration services	Associated services offer independent knowledge workers and dispersed organizations online meeting space and other

PKM Tool Category	Description
	collaboration tools through secure, hosted web applications.
organise / share / map	Associated tools organise information in context into networks, hierarchies, and reports. Results can be shared with others by e-mail or published on the web.

Although the impression may occur, that there seems to be a tool for every need, this is not really true. Agnihotri and Troutt (Agnihotri and Troutt 2009) pointed out, that research about technology in PKM is only at its beginnings. Most tools associated with a category are not really fulfilling the category requirements. Moreover, PKM tools need to facilitate sharing information with others – directly or indirectly. It is essential to notice, that *personal* does not mean *isolated*. As Dorsey (Millikin University 2003) stated, at least two (*collaborating around information* and *presenting information*) of the seven PKM skills are directly linked to collaboration with others.

While building a knowledge management system, two different approaches can be applied: the *infrastructure / generic approach* and the *process / task approach* (Jennex and Olfman 2004). By identifying the information and knowledge needs of a process, the *process / task approach* focuses on the use of knowledge by 'participants in a process, task or project' (Jennex and Olfman 204) with the goal to improve effectiveness. Because the users typically understand the area of knowledge that is used, only a minimal context is captured. Focusing on 'network capacity, database structure and organization, and knowledge / information classification' (Jennex and Olfman 2004), systems based on the *infrastructure / generic approach* on the other hand are meant to 'capture and distribute knowledge for use throughout the organization' (Jennex and Olfman 204). It is also possible to combine both approaches as supported by Morrison and Weiser (1996).

3.6 Benchmarking of PKM Tool Usage

In many cases, the decision to use a dedicated Personal Information and / or Knowledge Management system is off-the-cuff. However, it can and should be considered from a cost-benefit perspective. As Abecker and Völkel stated, to make sense for a user, the benefit of using the system must be larger than the cost of managing it (Abecker and Völkel 2008).

The costs refer to the amount of time, effort and money the user spends on the following three components *Cost of Creation*, *Cost of Externalisation*, and *Cost of Retrieval*.

- *Cost of Creation* describes how much it costs the user to create the knowledge in such a way, that it is meaningful and useful to him.

- *Cost of Externalisation* refers to the question 'how much does it cost the user to integrate the knowledge into the Personal Knowledge Management system'.

- *Cost of Retrieval* specifies the costs the user might have to retrieve the needed information and knowledge from the system at a later point in time.

On the opposite of the costs stands the *benefit of use*. Here it is important to value the knowledge we are talking about, but as Iske and Boekhoff stated, 'the value of knowledge does not exist as such' and depends highly on the individual situation (Iske and Boekhoff 2002). Value of knowledge can be defined as 'increment in expected utility resulting from an improved choice' (Varian 1999) based on the available additional knowledge. It is not only important to contrast the actual benefit of having the needed knowledge restored from the system with not have this knowledge available at all, but also with the new acquisition of the same knowledge (again). Figure 21 illustrates a simplified model for the

cost / benefit analysis in the area of Personal Knowledge and Information Management according to Abecker and Völkel (Abecker and Völkel 2008).
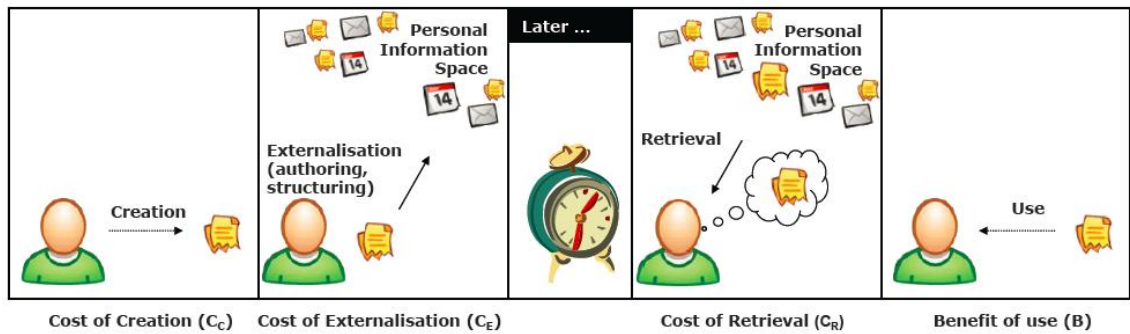
Figure 11 : Cost / Benefit Analysis in PKM (Abecker & Völkel 2008)

Another way to benchmark the usage of a Personal Knowledge Management system can be according to the *Technology Acceptance Model* by Davis (Davis 1989). The *Technology Acceptance Model* (TAM), as presented in Figure 12, suggests that the decision if a user will use a system or not, and if yes how, depends on several factors, which result in the *Perceived Usefulness* (PU / U) and the *Perceived Ease Of Use* (PEOU / E).
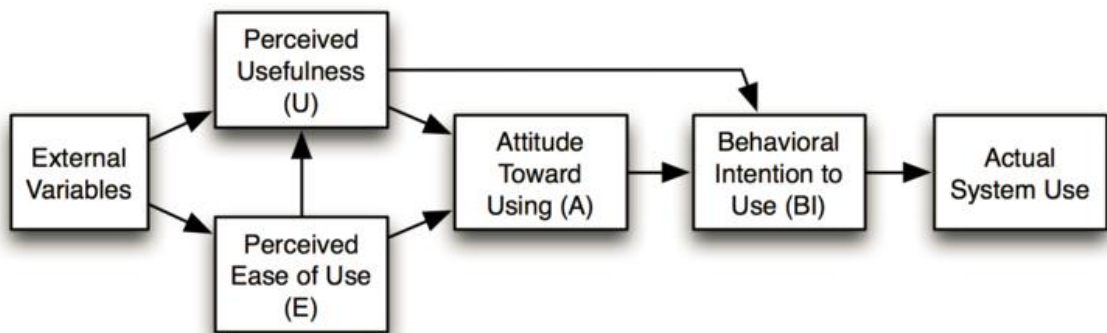
Figure 12 : Technology Acceptance Model (Davis 1989)

Davis (Davis 1989) defines *Perceived Usefulness* as 'the degree to which a person believes that using a particular system would enhance his or her job performance'. *Perceived Ease Of Use* (PEOU) is defined as 'the degree to which a person believes that us-

ing a particular system would be free from effort' (Davis 1989). Based on the TAM, there exist also other models like the *Unified Theory of Acceptance and Use of Technology* (UTAUT,) by Venkatesh (Venkatesh et al 2003) or the *TAM2* approach by Venkatesh and Davis (Venkatesh and Davis 2000).

3.7 Semantic Desktops

To manage personal working environments, the concept of *Semantic Desktops* emerged within the last ten to fifteen years. It has been seen as one of the main focuses of Personal Knowledge and Information Management. The term *Semantic Desktop* itself was firstly used by Decker and picked up by Sauermann in 2003 (Sauermann 2003). However, the related ideas were already touched in 1997 by Quan and his team from the *Computer Science and Artificial Intelligence Laboratory* of the M.I.T. – *the Massachusetts Institute of Technology* (Quan et al. 2003). Quan's project *Haystack* had the idea to build an integrated approach to let the individual knowledge worker manage his information in a most beneficial way from the personal point of view. Many project partners like *Nokia*, *HP* and *Microsoft* supported Haystack, and published a first prototype in 2004. Unfortunately, this prototype, called *Semantic Web Desktop*, was not further developed and is currently no longer available. In the meantime, the M.I.T. has given up the approach of developing an own semantic desktop, however, some of the resulting following projects still concentrate on parts of it mostly based on add-ons for the *Mozilla Firefox* web browser, or the *Thunderbird* email client. There were many other projects discussing the ideas, like the *MyLifeBits* project by *Microsoft Research* (Gemmell 2002), the *Ontooffice* project by *Semafora* (2013), as well as the open source projects *Chandler* (Chandler 2013) and *Fenfire* (Fenfire 2013).

Sauermann defines a semantic desktop as a 'device in which an individual stores all her digital information like documents, multimedia and messages. These are interpreted as Semantic Web resources, each is identified by a Uniform Resource Identifier (URI) and all data is accessible and query able as RDF graph. Resources from the web can be stored and authored content can be shared with others. Ontologies allow the user to express per-

sonal mental models and form the semantic glue interconnecting information and systems. Applications respect this and store, read and communicate via ontologies and Semantic Web protocols' (Sauermann et. al 2005). Sauermann sees the semantic desktop as 'an enlarged supplement to the user's memory' (Sauermann 2005).

According to Franz and Staab, the objective of 'the semantic desktop is to improve Personal Information Management (PIM) by combining all content available on the desktop and relevant to the user to easily manage that content, regardless of which type, and to simplify utilization of it' (Franz and Staab 2005). With the intention to create a central framework for the semantic desktop, the *European Union* founded the project *NEPOMUK- The* Networked Environment for Personal Ontology-based Management of Unified Knowledge (Nepomuk 2012, Nepomuk 2013). Beside the *German Research Center for Artificial Intelligence* (DFKI) in Kaiserslautern, many research partners like the *NUI Galway*, *EPFL Lausanne*, *FZI Karlsruhe*, *L3S Hannover*, *ICCS-NTUA Athens,* as well as companies like *HP*, *IBM*, *SAP*, *Mandriva*, *Thales*, and the *PRC Group* participated in that project. The NEPOMUK project was finished in 2008 and the results were two implementations, a KDE-based (KDE 2013) variant, and a Java-based (Java 2013) variant, from which some components are currently in use within other applications like the Linux distribution *Ubuntu* (Ubuntu 2013), as well as the Unix and Linux graphical user interfaces *GNOME* (Gnome 2013) and *KDE* (KDE 2013). The main focus here lies on a desktop search engine functionality, which was also extended within the project *Refinder* to search through documents stored in different cloud environments (Refinder 2013).

3.8 Recommender Systems

The previous sections referred to tools, systems and concepts that (might) support the users in managing the existing knowledge base. Another approach to link the individual knowledge fragments to each other might be the application of recommender systems concepts. Moreover, recommender system methodologies might be used to extend the existing knowledge base with relevant new aspects. The challenge to find relevant services and information easily and quickly is often seen as a great irony of the information revolution (Smyth et al. 2004). The enormous number of choices confuses the users and for many of them the best way to be successful and avoid mistakes is to use knowledge and experiences acquired by others (Velasquez and Palade 2007).

Recommender systems are, as first defined by Resnick and Varian, systems in which users share their preferences, and obtain, in a timely fashion, recommendations for unseen objects (Resnick and Varian 1997). Sarwar defines a recommender system as an electronic agent that helps users to find the most valuable products or services based on their historical preferences or tastes. They combine historical data on user preferences, information filtering and the application of patterns to suggest and predict items a user might like. By doing this, they offer users a more proactive and personalised information service (Sarwar et al. 2000).

There are two main different approaches, *Collaborative Filtering*, and *content-based Recommendations*. The *content-based recommender system* suggests the user services and products by analysing the items the user liked or used in the past (Balabanovic and Shoham 1997). Based on the information such as which books you have bought in the past, Amazon.com for example is able to recommend other books from the same authors or

about the same topics (Figure 13). However, there are several limitations of content-based recommender systems; by not knowing the user's preferences in advance. This means in the example of Amazon.com, that the predictions become better the more the customers buy, and nearly no recommendation are at all possible for new customers (Ehrlich and Malz 1995). To overcome these initial problems, very often additional information is used to create at least some useful recommendations. This can happen for example by using the user's browsing behaviour. So even without having bought any item, the visited products will be used as a basis for content-related recommendations.

Figure 13 : Content-Based Recommender System (Amazon.com 2013)

Another big limitation of the content-based recommender systems is the fact, that not all seen or bought items should be used for further recommendations. Just because a user bought one travel guide of Italy for example, does not mean that he is interested in buying also all the other available travel guides on Italy as well.

The *Collaborative Filtering* techniques approach on the other hand concentrates on recommending items other users, you have something in common with, have liked (Balabanovic and Shoham 1997). For example, when you have bought a book about travelling

Vietnam, and some other customers who bought the same book also selected a book about travelling Cambodia, the system would suggest you the Cambodia travel guide as well as presented in the Frequently Bought Together section of Figure 14.

Figure 14 : Collaborative Filtering Recommender System (Amazon.com 2013)

There also exist hybrid recommendations techniques (Smyth et al. 2005) where both approaches are combined. Depending on the first words you enter, search engines like Google (Figure 15) and Bing for example suggest to you relevant search string combinations based on the search string other users have been looking for in the past (Collaborative Filtering), as well as content related search strings (content-based Recommendation) you have been looking for in the past (Ansari et al. 2000).

Figure 15 : Hybrid Recommender System (Google.com 2013)

3.9 Summary

Based on the tight link between learning and knowledge the area of Knowledge Management has been identified as an adequate discipline to meet users' requirements in the changed learning environment and support their learning processes. This chapter starts with an introduction to the topic of data, information and knowledge. It presents the different available dimensions that allow the categorising of knowledge, like for example tacit and explicit knowledge. It discusses the scientific area of Knowledge Management, and classifies it into the social, organisational, and individual (personal) perspective. In this context some of the most common Knowledge Management models are introduced. Putting the focus on the individual approach to manage information and knowledge, the area of Personal Knowledge Management is introduced. The differentiation of Personal Knowledge and Personal Information Management is discussed, followed by an overview of Personal Knowledge and Information Management software systems. These systems are classified in categories, which are addressed and analysed. To outline in which cases the usage of Personal Knowledge and Information Management systems makes sense, the chapter continues with a discussion about benchmarking of the usage of such systems. The chapter finished with an overview of the related research area of semantic desktops, and a conceptual overview of the functionality and methodology of recommender systems.

4   Identified Problems in Supporting Learning

Chapter 3 introduced the area of Knowledge Management, focusing on Personal Knowledge and Information Management. Furthermore, it gave an overview over the current available categories of software systems to support the associated activities. The application of these tools in the changed learning environment and the resulting problems are identified and discussed in the following sections. In this context, *knowledge fragmentation* and the *nescience of the knowledge formation process* are determined as the main problematic issues within the current application of PKM tools in learning environments. Therefore, this chapter follows by addressing the issues of *knowledge fragmentation*. In this context, a deeper look at the components *content*, *descriptive metadata* and *relationships* is being taken. The next section is devoted to the topic of the *knowledge formation process*.

4.1   Introduction

During the research, the current usage of PKM application within learning environments has been examined. This has been done in by combining insights won from different occasions, like:

- Context related feedback discussions during and after several Knowledge Management lectures that were given at the *University of Applied Sciences in Darmstadt* (Germany) for Bachelor and Master level students.

- Informal discussions and expertise exchange at several national and international seminars and conferences on knowledge and information management and its systems.

- Informal discussions and expertise exchange at regular PhD seminars at the *University of Applied Sciences in Darmstadt* (Germany) in association with the *Plymouth University* (UK), and the *Cork Institute of Technology* (Ireland).

Based on these findings, four different application categories of PKM systems in the area of leaning have been identified:

1. Firstly, tools that are used by students as a matter of course, mostly even unconsciously.

   Desktop search applications like *Google Desktop* (Google 2011a) or *Copernic* (Copernic 2013) for example, have in the last years been part of the everyday work life, or are even integrated in newer operating systems like *Windows 7* (Microsoft 2010), *Windows 8* (Meyer 2011) or *OS X* (Apple 2013).
   One example in this category is the tool *Copernic Desktop Search* as presented in Figure 16. Beside data type comprehensive search functionality, it offers a direct preview of the results independently of the content type – even without having to start the designated viewer application. This way it allows a direct and instant allocation of the searched resources.
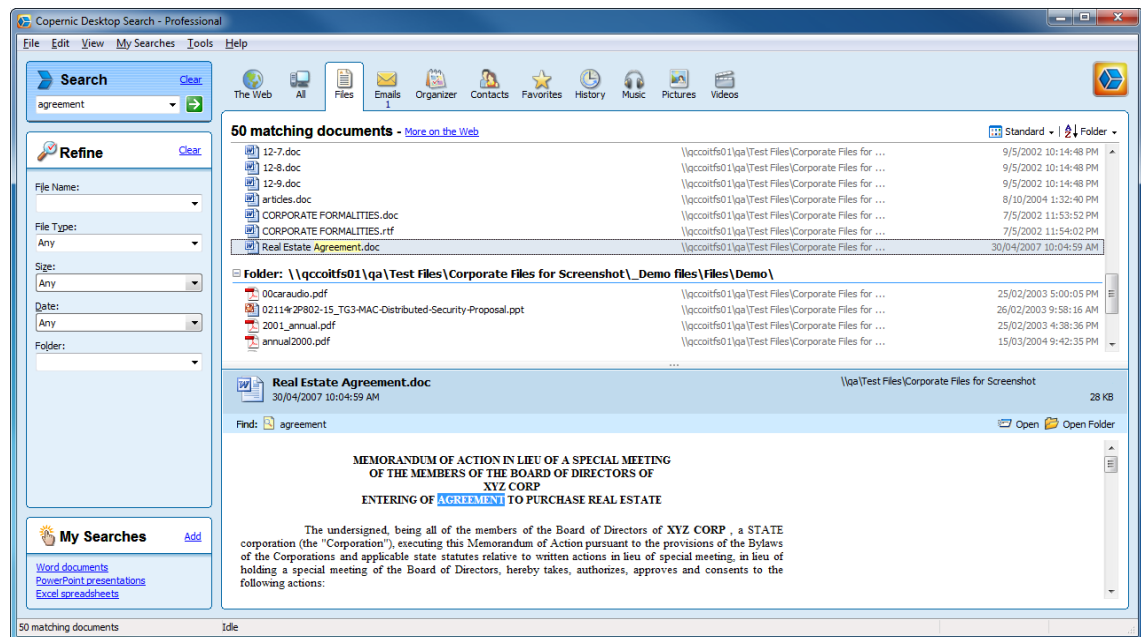
Figure 16 : Copernic Desktop Search with Preview (Copernic 2013)

2. The second type refers to tools that users are mostly not intentionally aware of.

Good examples are the tools *Knowledge Workshop* (LMS 2013a), *Personal Brain* (Personal Brain 2013) and *TheBrain* (The Brain 2013). Presented to other students during Knowledge Management lectures (Bachelor, Master) and PhD seminars at the *University of Applied Sciences Darmstadt* (Germany), the *Plymouth* University (UK), and the *Cork Institute of Technology* (Ireland) the astonishment and the enthusiasm has always been huge. Nevertheless, even an IT-related audience (in most cases) have just never heard of them – even though they are (mostly) available for students for free, typically come from an academic environment and have been available for years.

Figure 17 presents the graphical user interface of the tool *Knowledge Workshop*. It allows the storage and organisation of different document types, annotating, as well as extended indexed search functionality. Run as a Microsoft Windows ap-

plication, *Knowledge Workshop* does not support the simultaneous usage of several computers, mobile environments, as well as collaboration with other users.
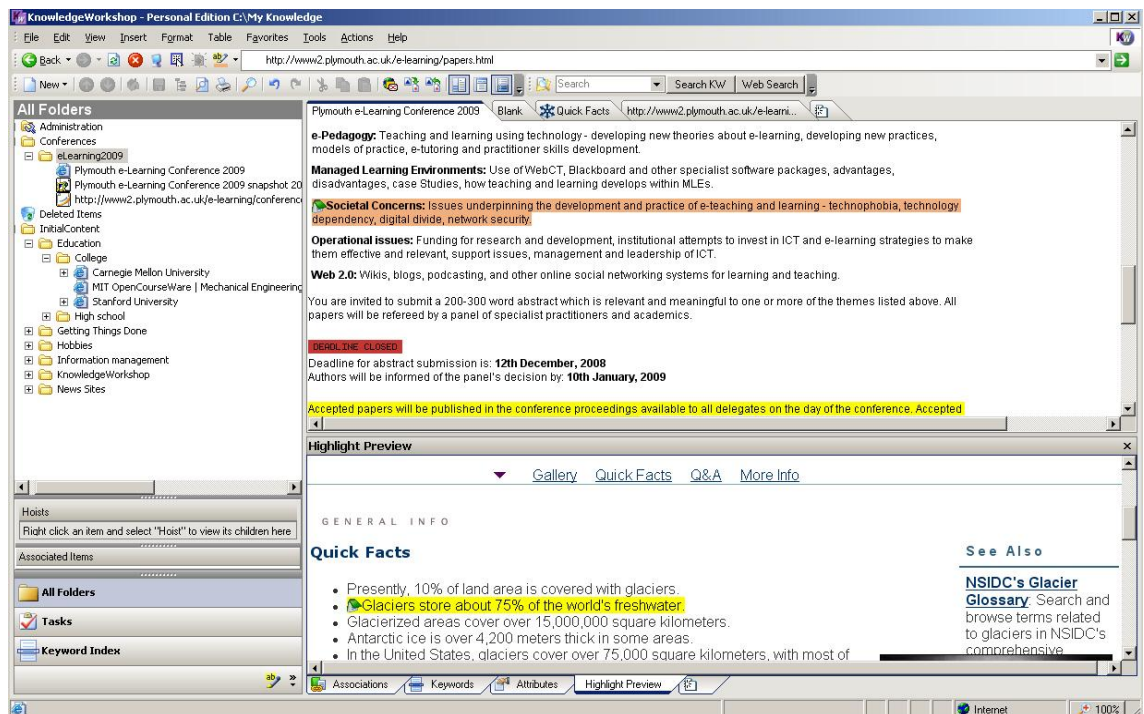
Figure 17 : Knowledge Workshop (LMS 2013a)

The mind-mapping tool *TheBrain* (TheBrain 2013), presented in Figure 18, offers functionality to structure the personal knowledge base in a completely different approach. Similar to the tool *Personal Brain* (Personal Brain 2013), the idea here is not a tabular and linear saving and organisation of the existing documents, but an interconnected mind mapping approach always starting from the point of view of an idea or a thought. *TheBrain* is offered as a Microsoft Windows desktop version. It does currently not support any mobile platforms or collaboration with other users, but it offers the possibility to store the personal knowledge base within a cloud environment. This allows at least the simultaneous usage of several Windows devices. Both solutions are based on normal mind mapping functional-

ity as offered also by many other commercial tools like MindGenius (MindGenius 2013), or open-source tools like FreeMind (FreeMind 2013). Principally, these tools might be used in the same manner as *TheBrain* or *Personal Brain*.

Figure 18 : The Brain (TheBrain 2013)

3. The third type refers to tools that the students are aware of, in theory the functionality even convinces them, but the expenditures to use them are just too high.

   This might be from a financial perspective, but it might also be from a workload perspective. If it is too complex and takes too long to integrate or import the student's knowledge base into the selected framework, usually the student will capitulate before getting it to a good starting point. The decision to use a tool or not refers to a cost-benefit benchmarking as described in section 3.6. However, the evaluation of the individual cost-benefit ratio is mostly done on instinct and is not based on founded data.

Figure 19 presents the application *Zotero* (Zotero 2013). Started as a *Firefox* web browser add-on, Zotero is currently also available as a *Microsoft Windows* operating system version, which allows running the application also under the *Google Chrome* and the *Apple Safari* (for Windows) web browsers. The tool is powerful in handballing and organising websites, and partially other document types like Adobe PDFs. Nevertheless, the mostly manual workload that is needed within the initial phase, to gain all the benefits out of the system is huge. Used in the upgraded version, the server (cloud-based) storage can be applied, which allows not only the usage of several devices, but also access via mobile platforms like Android (Android 2013) and iOS (iOS7 2013).
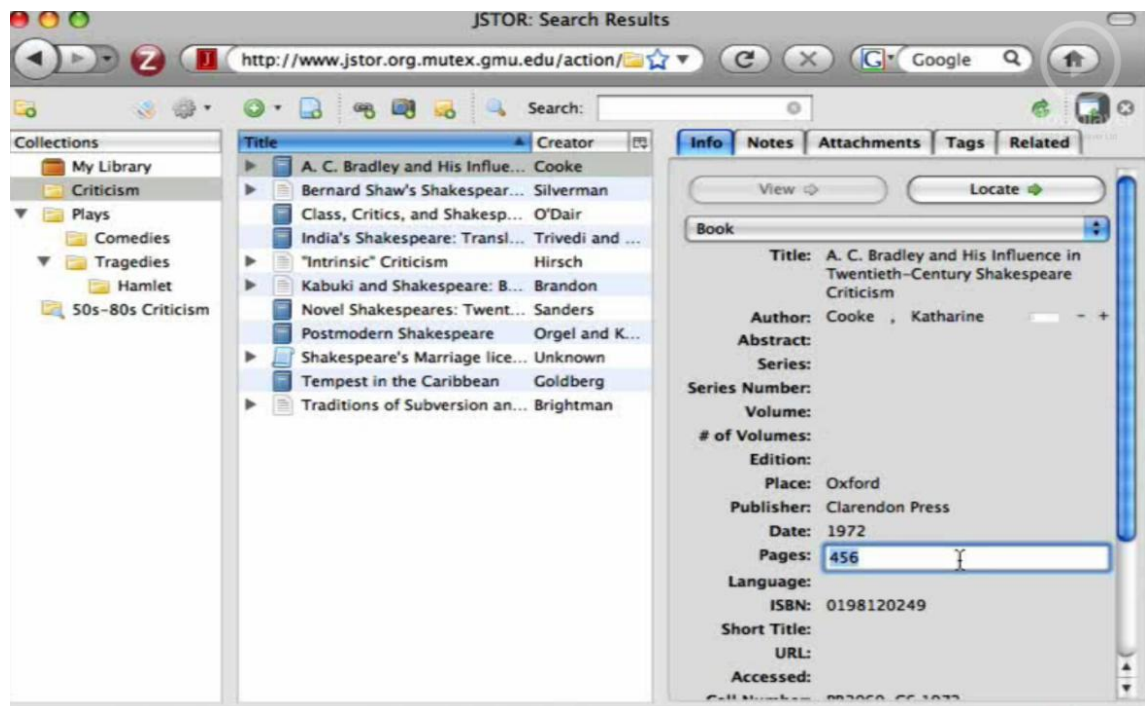
Figure 19 : Zotero (Zotero 2013)

4. The last type refers to solutions that are accepted and favoured, but have a short product life cycle.

The collaboration and communication tool *Google Wave* (Hölzle 2010) is a good example in this category. Google has stopped the service because of insufficient general acceptance (Google 2010). Nevertheless, especially in the academic area *Google Wave* gained a greater interest and acceptance. Another good example is the *Google Reader*, for which Google announced in March 2013 that on the 1st of July 2013 Google will retire the service (Green 2013). Due to a high dynamics within the software market of Personal Knowledge and Information Management tools, many approaches and prototypes are suspended after a short while (Reinmann & Eppler, 2008); however, some of them are built on very interesting ideas. Note: During the last months of this research, the tool *Knowledge Workshop* that was presented in section two, was also suspended due to 'business operational reasons' (LMS 2013b).

The application *Google Desktop,* introduced in section one, was also suspended on the 14th September 2011. The official reason is that the offered functionality became part of all modern operating systems (Google 2011b).
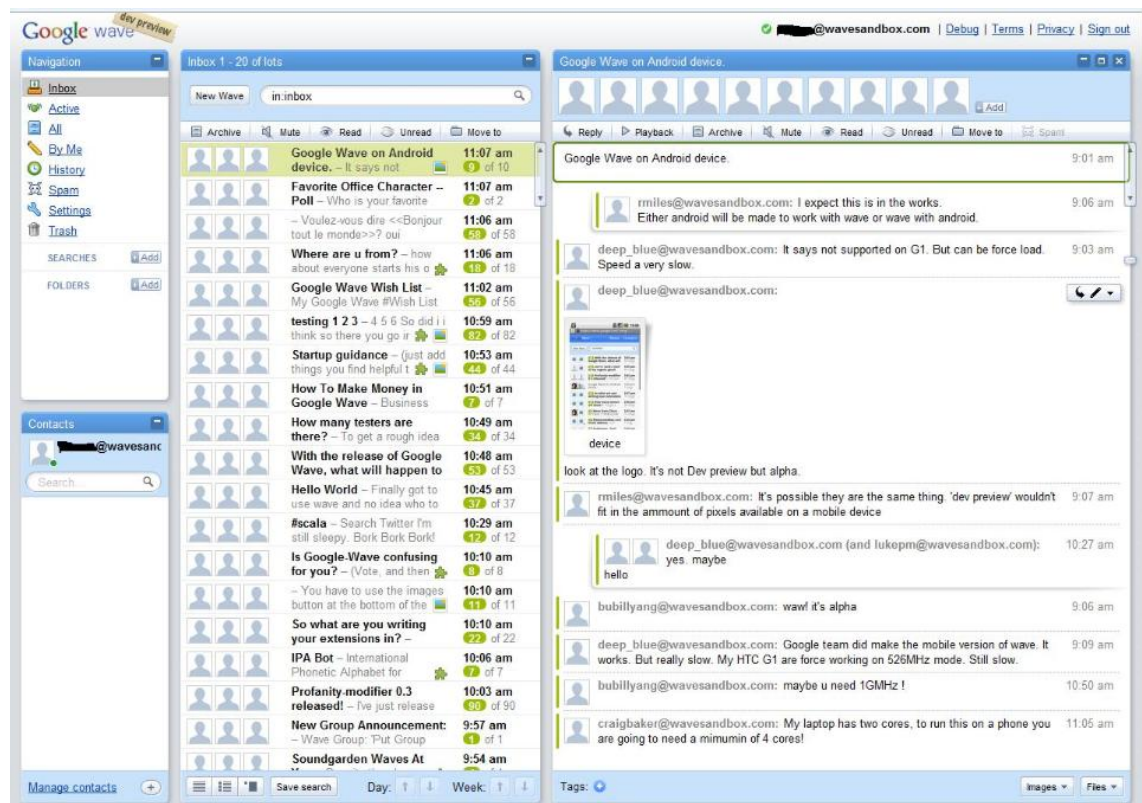
Figure 20 : Google Wave

These four types demonstrate just some aspects why the usage of PKM tools is currently not so widespread throughout students. Even where it is, or potentially could be, another problem becomes clear: the *knowledge fragmentation* (Karger 2006).

It is obvious, that from the user's perspective it would be great to have one persistent solution for all needs. This will enhance usability and user friendliness. Therefore, a natural step should be to integrate the in Table 3 introduced PKM functionalities into one tool. Due to the fact that knowledge work is multifaceted and versatile, the requirements on software systems to meet all of them are, or would be very high. Typically, vendors concentrate on individual tasks or groups of tasks, have their core competencies there and are, if at all, just providing average solutions in other fields. A good example is the search engines area; nearly every currently available Personal Knowledge and Information Man-

agement tool has an integrated search engine. However, whereas some of the tools are concentrating explicitly on *Indexing and Searching* (see Table 3; like for example *Copernic*), others, like *Knowledge Workshop* or *Zotero*, have their main tasks in the area of *Organising and Mapping*. The functional ranges, as well as the effectiveness of the implemented search engines vary a lot. Users therefore often prefer to search and find their documents in one tool and work with them within another tool.

At the same time, the dissatisfaction and the necessity of these auxiliary releases support a very low inhibition threshold to switch from one tool to another. Unfortunately, as mentioned previously, switching the personal knowledge base from one solution to another is usually associated to a high entry level that often requests an enormous start investment. Typically, this start investment, which can be directly linked to the cost of externalisation and retrieval as introduced in section 3.6, is not a monetary one, but based on intensive workload. In many cases, these entry (and exit) barriers are built on purpose. This way the solution suppliers want to tie their users in and not to offer them any easy possibilities to leave them – at least not without any disadvantages or problems.

To summarise, the following critical issues have been identified within this section:

- The simultaneous usage of different technical solutions for different knowledge work tasks and activities.
- The users' willingness to change the technical solution as soon as they think the new solution might benefit them.
- The often very short product lifecycles of the technical solutions.
- The problems associated with changing a technical solution.

Not only for an entire knowledge base movements (intended or unintended), but also to support the users within their task comprehensive work, a fundamental requirement must be to have access to the entire knowledge and information base. In this context, the distribution of our information and knowledge (knowledge fragmentation) needs to be further discussed. In those cases, where a complete access to our knowledge is not possible, a reproduction of the undertaken changes might help out. As long as we know what happened, we might be able to reproduce it in the new environments. This can happen on a larger scale to ensure a knowledge and information base switch from one solution to another, but it could also be applied on a smaller scale just to allow other systems the usage of insights won or saved within other environments.

The following subsections discuss the topics knowledge fragmentation, as well as the comprehension of the knowledge formation process.

4.2 Knowledge Fragmentation

Our entire personal knowledge base is distributed over several physical and logical locations. Beside our brain and memory as the largest and most important ones, we usually store our data, information, and knowledge in different ways and places, in an analogue or a digital way, like for example in notebooks, hard drives, databases, pads, memos, (PKM) applications. Where the single information and knowledge units are stored, depends on the kind of information and knowledge, as well as the personal preferences and behaviour patterns of the individual user. As described in chapter 3.2, tacit knowledge is archived within our brains, whereas explicit knowledge is usually saved externally. In many cases the entire picture consists of some tacit and some explicit parts of information and knowledge.

Section 3.5 introduced software systems (PKM tools) which should help the user in organising and managing the explicit parts of our knowledge base, as well as to support the relocation and retrieving of the tacit parts. As stated before, there are many reasons, inflicted by others, as well as self-inflicted ones, why we are using more than just one tool for everything. This leads to the situation that not only different knowledge pieces are distributed over different locations, but also the coherent parts of the same knowledge piece are saved in several places.

A single knowledge unit can be divided into the pure *content*, the associated *descriptive metadata*, and the *relationships* with other knowledge and information pieces. The following sections define these three categories and describe them based on an example of a digital photography.

Content

The statement of every knowledge unit can be seen as the *content* (FreeAdvice 2013, Definitions 2013). A digital photo itself is stored as a binary file, which represents the medium for the content. Opened in an editor it is not more than a set of raw facts (numbers and characters), the data behind the information (Figure 21).
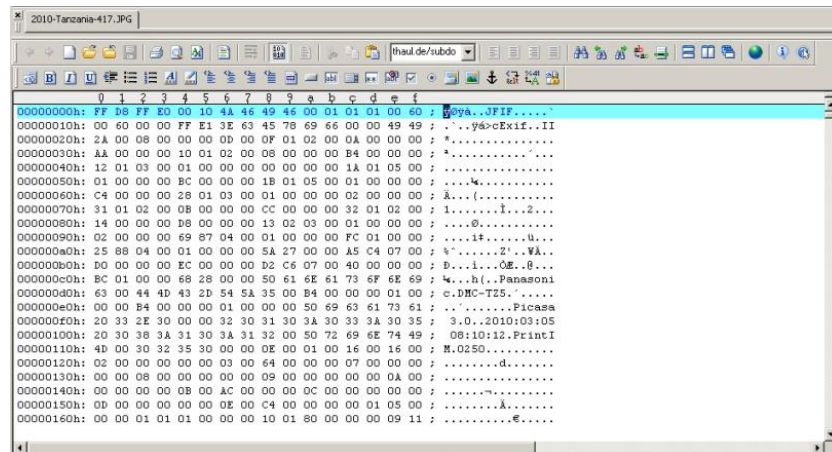
Figure 21 : Knowledge Fragmentation - Data

When we look at it using a picture viewer, we can see the taken picture – it offers us additional relevant information (Figure 22). In this form, the picture itself contains the content.

Figure 22 : Knowledge Fragmentation – Content

Descriptive Metadata

Descriptive Metadata, often also called metacontent, is defined as 'data that is used to describe other data' (Bargmeyer and Gillman 2011). In other words, it is content which describes the actual content. Depending on the information entity, the possible descriptive metacontent may vary fundamentally.

By working with the content, putting it into action (Tiwana 2002) and applying it we create personal knowledge. For example, we could rename a file, add some tags, a description, some notes, a title or some geo coordinates. These examples do not refer all to the content itself, but to the descriptive metadata. This metadata gives the content some (additional) meaning – personal meaning. This additional meaning is part of our personal knowledge. Maybe these facts and notes do not mean a lot to anyone else and are in fact just data, even maybe useless information; however, for us on the other hand they do – and that is where the difference between personal and general Knowledge Management takes place.

In our example, some of the descriptive metadata might be physically stored in the picture file itself, some might be stored in the application we use to work with it – but it is still metadata. Figure 23 demonstrates the same picture in the photo community *flickr* (2013). In the highlighted areas you can see an extract of the descriptive metadata, like for example the *title*, the *description of the photo*, information about *where it was taken*, *when it was taken*, with what *type of camera*, and which *tags* are associated to it.

Figure 23 : Knowledge Fragmentation - Metadata

Relationships

Beside the described content and the descriptive metadata our knowledge pieces usually have a context, or in other words, they are not isolated, but are linked to other information and knowledge units. In the case of the photo, we may for example store this photo in a specific album; this album on the other hand belongs to some categories. In the end, there is a whole structure of relationships and links between the categories, the albums and this picture, or between other pictures from the same album and this one. Another relationship might for example be found in the people presented on the picture. Other users' comments itself for example might be seen as metadata – but the environment in which the comments have been made is much more than that. The information by whom the comment was made, when it was made, who answered to this comment, and who is behind the user who made it, might be important for you. Again, these relationships are (in most

cases) of a personal importance, which means, that another person might probably put the picture in a completely different album-structure, or the persons who commented these photos might not be of any importance to him.

The highlighted area in Figure 24 illustrates the logical structure of relationships between our picture, its album, and the different collections (categories) it belongs to on flickr:
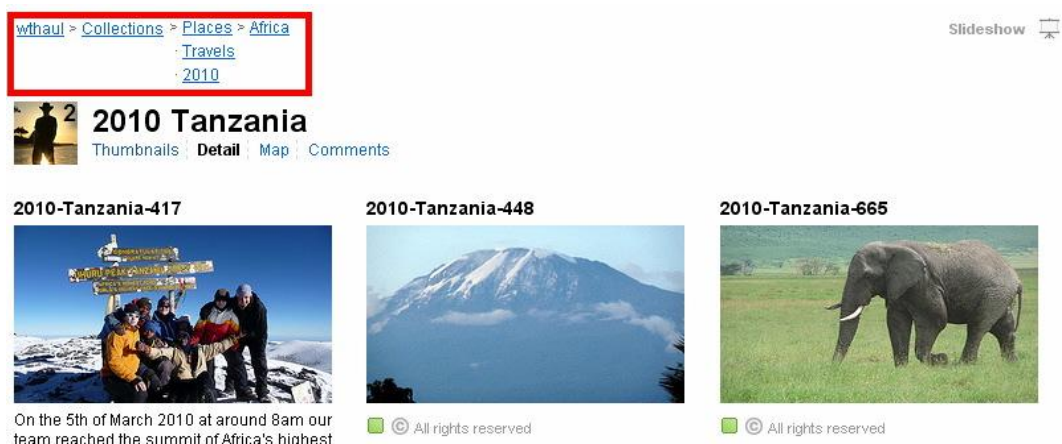
Figure 24 : Knowledge Fragmentation - Relationships

Challenges with Knowledge Fragmentation

The separation of *content*, *descriptive metadata* and the assigned *relationships* over different locations and systems results in several challenges. This becomes obvious when we for example decide to change the system we have used for our Personal Information and Knowledge Management (even partially). Here we need to find a possibility to transfer not only the content, but also the descriptive metadata and the relationships. Exporting the content is usually possible and in most cases even easy. Depending on the system in use, exporting the descriptive metadata may be (partly) possible as well. In many cases, this depends on the fact where the descriptive metadata is physically stored. Descriptive metadata, which is technically stored directly in the files, can usually be exported easily; meta-

data which is stored in the system but not in the file itself is usually difficult to capture. Extracting the relationships and connections is very often not possible at all.

In the case of *flickr* we could export the original file only by having a professional (paid) account (first exit barrier – note: this changed with the updated version announces on the 20th May 2013 (Spiering 2013). Even if the file had a meaningful file name during the upload (like "2010-Tanzania-417.JPG"), the downloaded file would be named differently like "7384278432984.JPG". Already here we have lost some of our descriptive metadata. Additional information we added using flickr (like the description and tags) would not be exported either. And the relationships, like the comments, the album structure, and the information about the people we marked on the pictures would also be lost. This illustrates how difficult it is to change a system – even if a competitor offers a new and better (more effective or e.g. just cheaper one) one.

Photos have been used just as an example to visualise, but the same behaviour can be found with all other media or document types as well – especially in the academic environment. Taking an assignment as an example might imply the usage of several software systems. We might use one tool for the word processing, another one for the design and creation of the needed graphics and pictures, the internet browser to assemble and validate the needed information, and a bibliography system for managing our references. A decision to switch now from one tool to another, for example from *Microsoft Word* to *LaTeX*, is directly linked to an enormous effort and in many cases not realistic. Therefore, one of the next steps in the design of technical PKM solutions must be a service, which can take care of the descriptive metadata, the relationships, as well as the content itself and allow accessing it completely and effectively at any time.

4.3 Knowledge Formation Process

To make sure the workload, users put into a system, is a long-term investment, there is either the need to have the possibility to export the holistic knowledge base from one environment and import it into another, or to reproduce the changes in a preferably automated way within the new environment. Section 4.2 discusses the issues resulting from the first approach (mainly knowledge fragmentation). The issues resulting from the second aspect refer directly to the traceability of the knowledge formation.

Before we can reproduce something, either manually or in an automated way, we need to know what exactly we need to reproduce – in other words what exactly happened. Applying information, putting them into action is, at least mostly, a multilevel activity. The way before the piece of information becomes our personal knowledge is an important one. We are doing things for a reason, and this reason is essential when we talk about knowledge. There might be several steps, several reasons, why in the end the knowledge is as it is, and why it is meaningful to us. In this context, it is fundamental to remember the reasons.

An academic assignment document for example is in most cases going through a non linear formation path. A student might start with a blank master document and by putting his different ideas together, by researching libraries and the internet or talking to others, his output (and personal knowledge) is growing step by step. It might happen, that the student decides to go some steps back to a previous version and to continue in a different direction from there on. A simplified knowledge formation path with different versions and inputs is presented in Figure 25:
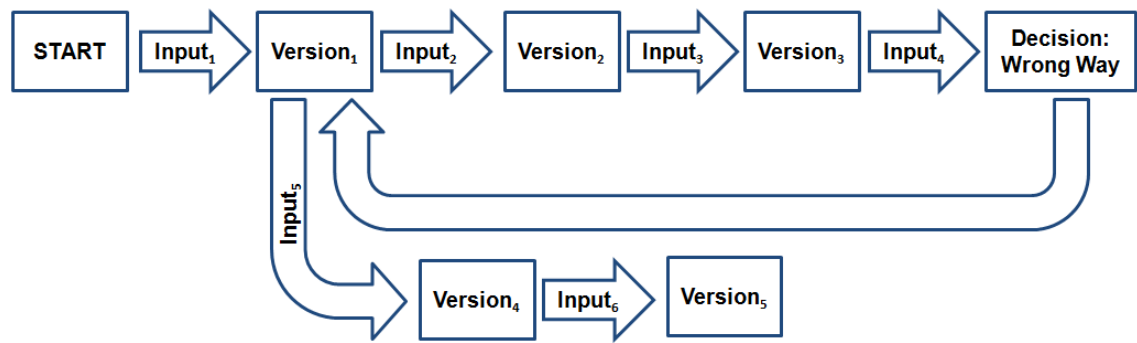
Figure 25 : Knowledge Formation Paths

Current PKM software systems do not consider this. They have the outcome, but they do not know, or at least remember the way the unit of knowledge was before it ended where and how it is now. Moreover, when we assume, that not all systems will ever allow us to access the metadata and the relationships introduced in the last section, the conclusion must be to create a possibility to reproduce the status. Therefore, we have to remember the intermediate steps as well.

You might for example prefer to use system *x* for the storage and working with your photos, but on the other hand, system *y* might have some functionality like *face recognition* which system *x* does not have. The normal handling will be to first use system *y* for the face recognition, and after this to use system *x* for the rest. However, what if system *y* does not want you to switch and therefore does not offer any possibilities to access the recognised faces meaningfully? We need a system which remembers the properties of the photo overall, that it was applied in the system *y*, what the outcomes (e.g. the names of the persons found on the photo) were, and what happened to it next. Having this information, it might be applied in a way, the target system *x* might handle it – in our case, the names of persons might be added to the photo, for example as tags.

The traceability of our actions and the resulting knowledge of the knowledge formation process are important for two reasons. Firstly, to support the replication of metadata and relationships in cases where we will not be able to export them; secondly to remember the way our knowledge went through and became what it is today – by, as far as possible, remembering why we decided to do things the way we did them.

4.4 Confirmation of Relevance

The confirmation of the identified problems and challenges within the evolving learning environment (as described in chapter 2), as well as the identified problems students encounter currently (as described in chapters 3 and 4) has been on the one hand executed during the already mentioned formal and informal discussions with bachelor, master and Ph.D. level students. Furthermore, to confirm the issues from a broader student perspective, a qualitative student survey has been conducted among students of the Plymouth University in the UK.

4.4.1 Scope of the Confirmation

The student survey has been conducted via the online platform SurveyMonkey (Survey-Monkey 2014) and has been distributed via email with the corresponding web link to 308 undergraduate and graduate students (based on availability) at the Plymouth University.

Confirmation Questions

The questions asked have been divided into three parts referring to the following areas:

- Statistical Information

- Information Overload

- Current Issues

Statistical Information (section 1)

Goal: To allow a more detailed interpretation of the provided answers, the first section refers to statistical and demographic data.

Sex: [ ] Male   [ ] Female

Age: _____

Faculty of   [ ] Arts & Humanities

                    [ ] Business

                    [ ] Health and Human Sciences

                    [ ] Science and Environment

What is the highest level of school you have completed?

High school degree or equivalent

    [ ] Some college but no degree

    [ ] Associate degree

    [ ] Bachelor degree

    [ ] Graduate degree

Information Overload (section 2)

Goal: To confirm that the in chapter 2 described evolving learning environment and the identified challenges (especially the information overload) are really relevant issues within the contemporary learning environments.

1) How much do you rely upon the internet for learning?
    [ ] Not at all
    [ ] Slightly

[ ] Moderately

[ ] Very

[ ] Extremely

2) Compared with traditional academic information sources like books and journals, how often do you use the internet as a primary information and knowledge source?

[ ] Never

[ ] Rarely

[ ] Sometimes

[ ] Often

[ ] Always

3) Compared with a book or a journal article, most of the published information on the internet is not validated. How concerned are you regarding the correctness of the used information when you use it within your studies?

[ ] Not at all

[ ] Slightly

[ ] Moderately

[ ] Very

[ ] Extremely

4) Search engines like Google or Bing help us find the information we are looking for. However, mostly a search query results in thousands of results. From all these results, how easy is it for you (in the academic context) to identify the ones which offer you the answers you are really looking for?

[ ] Very easy

[ ] Easy

[ ] Moderate

[ ] Hard

[ ] Very Hard

5) Besides all the benefits of the information technology there are also critical aspects like information overload. For your studies, do you consider information overload as a issue / problem?

[ ] Strongly disagree

[ ] Disagree

[ ] Undecided

[ ] Agree

[ ] Strongly Agree

6) Once you have found interesting information (like an interesting website), do you sometimes encounter a problem of locating it again later (e.g. 2 months later)?

[ ] Strongly disagree

[ ] Disagree

[ ] Undecided

[ ] Agree

[ ] Strongly Agree

Current Issues (section 3)

Goal: To confirm that the identified problems of *knowledge fragmentation* and *the nescience of the knowledge formation process* are really relevant issues within the contemporary learning environments.

1) Do you think remembering why you (or your lecturer, your study group) decided to do things a certain way is important?

[ ] Strongly disagree

[ ] Disagree

[ ] Undecided

[ ] Agree

[ ] Strongly Agree

2) Do you sometimes encounter the problem of no longer remembering why you (or your lecturer, your study group) decided to do things a certain way?

[ ] Strongly disagree

[ ] Disagree

[ ] Undecided

[ ] Agree

[ ] Strongly Agree

3) Typically students are using different software systems (like MS Word, Excel, PowerPoint, Dropbox, Google Drive, Evernote, Email-Accounts) for different tasks. Do you sometimes encounter problems with this because your information and knowledge is separated over several locations (also different systems, hard drives, online-portals)?

[ ] Strongly disagree

[ ] Disagree

[ ] Undecided

[ ] Agree

[ ] Strongly Agree

4) Have you ever tried to switch a system you are using (e.g. e-mail provider, word processing, any kind of online systems) due to new and better features, but gave up due to problems with compatibility, data migration, or just the needed work-load?

[ ] Strongly disagree

[ ] Disagree

[ ] Undecided

[ ] Agree

[ ] Strongly Agree

4.4.2 Findings of the Confirmation

A total of 308 students were contacted of which 58 started the survey and 56 completed it. The full results can be found in Appendix E. Eight of the participants were female, 48 male. 53 were currently enrolled at the *Faculty of Science and Environment*, whereas three were students of the *Business Faculty*. The majority of the participants were bachelor level students with some college experience but none had finished their degree (Figure 26).
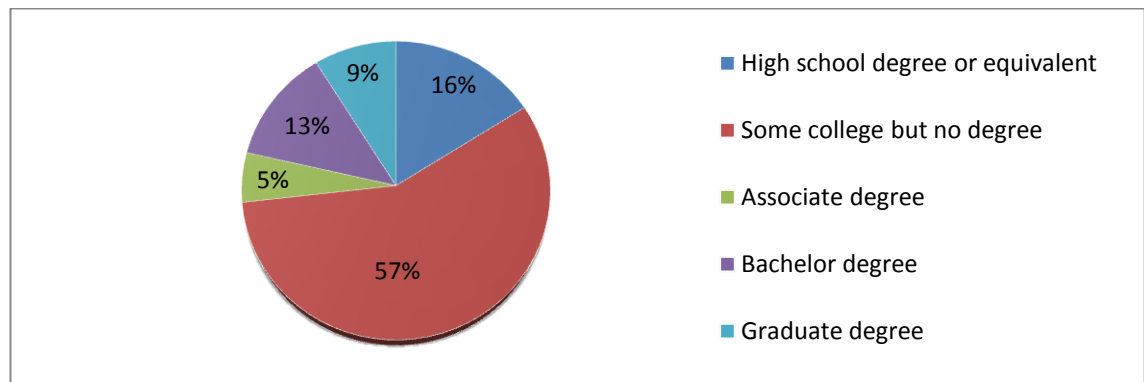
Figure 26 : Student Survey – Highest Level of School

The survey confirmed the general assumptions and findings, as well as the identified challenges and problems of the research. 100 % of the students rely on the Internet for learning purposes and for 91% of them, the Internet is very important (selecting either 'very' (43%) or 'extremely' (48%) categories). 23% of the students always use the Internet as the primary information and knowledge source, with a further 63% using it on an frequent basis.

A large majority (89%) of the students were concerned over the validity of the information published on the Internet (Figure 27).
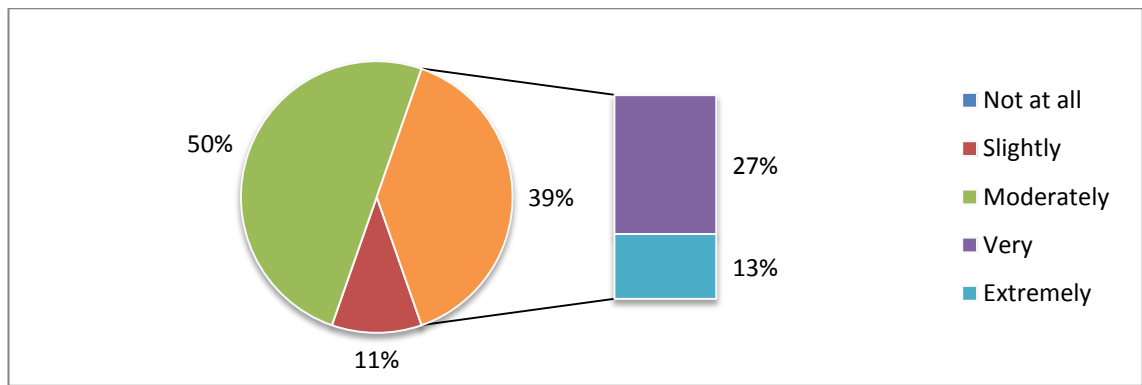
Figure 27 : Student Survey – Concern regarding Correctness of Information

When participants were asked how easy they found it to locate information using search engines (i.e. Google or Bing), only 25% considered in either easy or very easy to do so. More than half (52%) found it moderately difficult with a further 23% experiencing real problems (Figure 28).
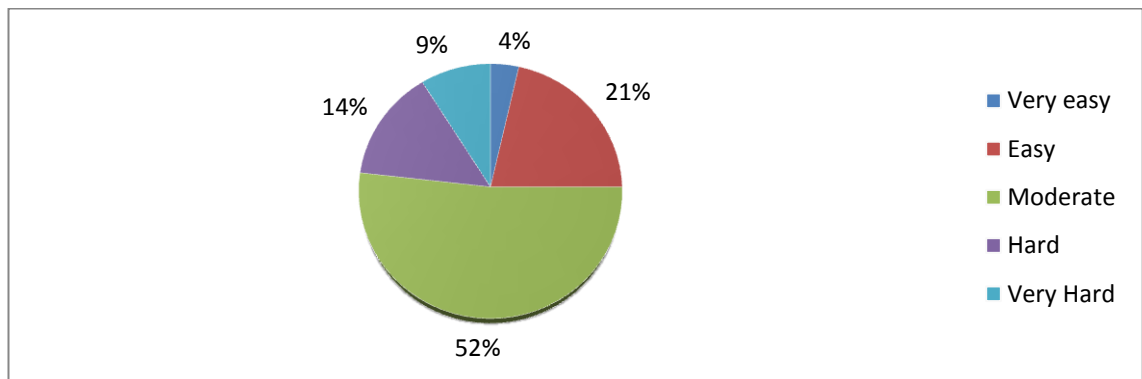
Figure 28 : Student Survey – Identification of the Correct Search Results

From all the participants, 63% consider information overload as a problem (selecting either 'agree' (52%) or 'strongly agree' (11%) categories). 14% are undecided whereas 23% do not see it as a critical issue (Figure 29).
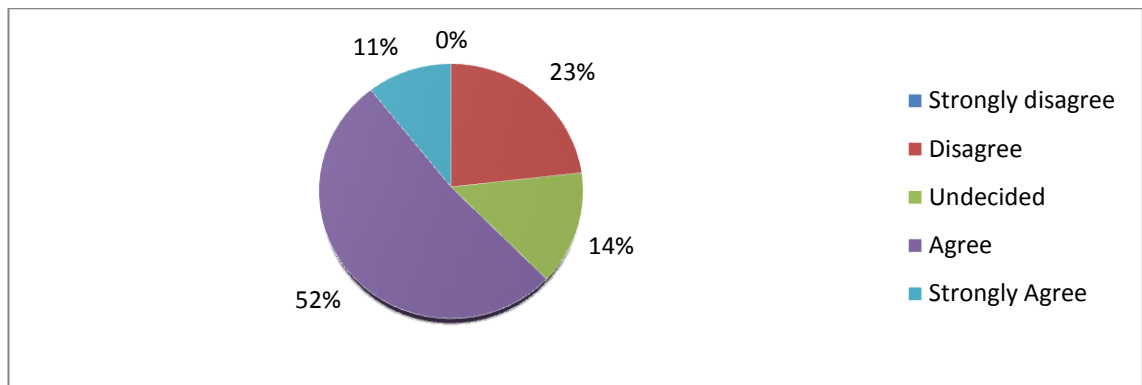
Figure 29 : Student Survey – Problems with Information Overload

While working with their information and knowledge base, 48% of the participants have problems in locating and finding interesting information (e.g. websites) again after a certain time (selection either 'agree' (36%) or 'strongly agree' (12%) categories as presented in Figure 30).
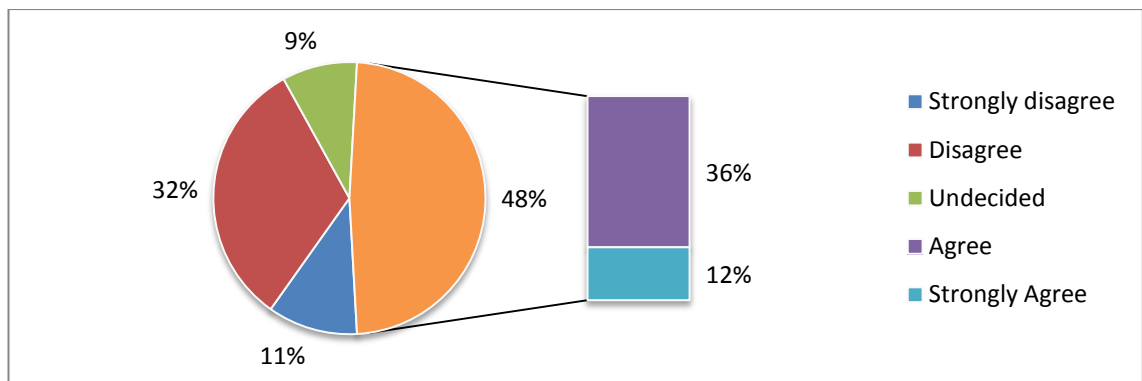
Figure 30 : Student Survey – Problems with Locating Again Information

82% of the students agree (selecting either 'agree' (59%) or 'strongly agree' (23%) categories), that remembering why we, or our environment (like for example the lecturer or the study group) decided to do things a certain way is important. At the same time, 56% encounter problems of no longer knowing why things are as they are (selecting 'agree' (43%) or 'strongly agree' (13%) categories as presented in Figure 31).
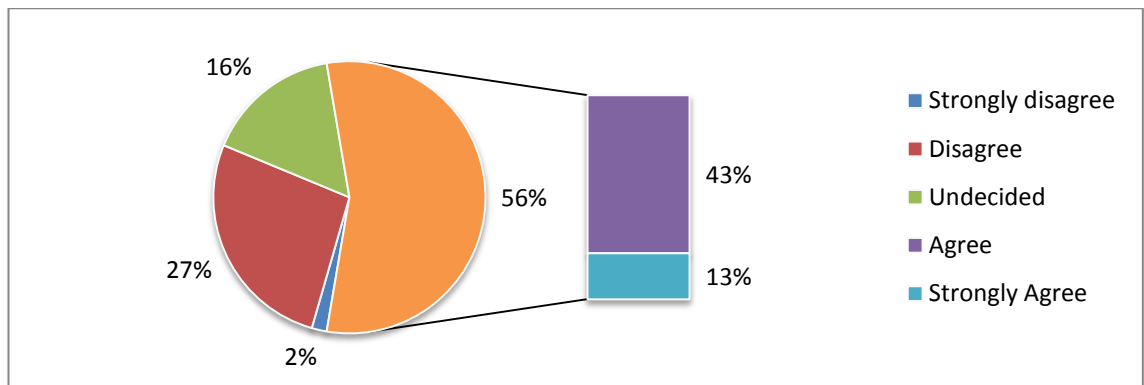
Figure 31 : Student Survey – Problems with Remembering Why

The identified problems of knowledge fragmentations have also been confirmed by 57% of the students (selecting 'agree' (43%) or 'strongly agree' (14%) categories as presented in Figure 32). Nearly half of the students (48%) have already encountered problems while switching a system in such a manner that they gave up before succeeding (selecting 'agree' (37%) or 'strongly agree' (11%) categories).
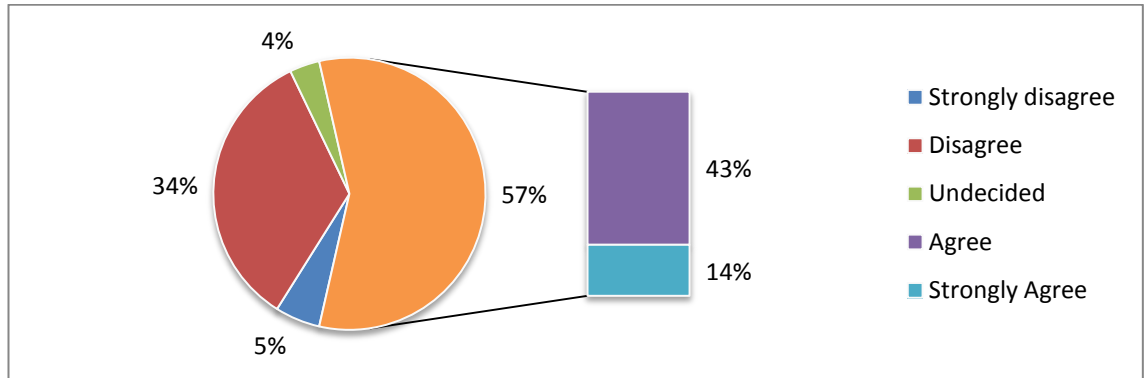
Figure 32 : Student Survey – Problems with Knowledge Fragmentation

Within the answers, no patterns according to age, subject of study or highest level of completed school could been observed.

4.5 Summary

Chapter 4 describes the current application of PKM software systems within learning environments. It introduces the four identified types of applications and addresses the problematic issues resulting from knowledge fragmentation and the nescience of the knowledge formation process. Taking a deeper look into knowledge fragmentation the three knowledge piece categories *content*, *descriptive metadata* and *relationships* are introduced and described. Their impact on students' behaviour with PKM tools is analysed. This is followed by assessing the problems of current systems resulting from not remembering how the knowledge piece became what it currently is. In the last section of chapter 4, a student survey that has been conducted to confirm the identified findings, is introduced and its results outlined and discussed.

In a fast changing and diversified system-landscape a practicable and beneficial solution must allow easy and sufficient change of the currently used systems. The tool vendors put intentionally high entry and exit barriers to stop their clients from changing to a competitor's solution. To overcome this, there is a need to have access to the content, the descriptive metadata and the relationships with other content pieces, export it from one environment and import it into another. In cases where this information is either not available or accessible, the possibility to reproduce the work within a different environment is needed. Moreover, the possibility to reproduce our work is based on the historisation of our actions, and could so help us to remember why we decided to do things the way we did them. These insights might be of high importance for our next actions, and could at the same time help us with other, similar decisions – even by developing behaviour patterns to improve our performance and efficiency.

5   New Approach: The Holistic Knowledge Formation System

The last chapter identified from both, a technological and user perspective the requirement for a different and new approach to allow a holistic view on the individual knowledge and information base. It also stated why it would make no sense to integrate all functionalities into one software solution; the market is changing fast and the specialised vendors provide better solutions in their specific areas than vendors with all-rounder approaches do. Based on the identified issues, this chapter 5 presents a new approach to overcome the problems in a sustainable manner.

5.1   Objectives

Having the previously described problems and challenges in mind, the main objectives of the suggested approach are as follows:

1. Comprehensive and automated versioning of the user's knowledge work

    Irrespective of the kind of the knowledge work the user is undertaking, and irrespective of the used application(s) or system(s), as well as the associated file formats or versions, the solution should offer the possibility to historicise the user's knowledge work in a comprehensive and sustainable way. Although this requirement has already been (partially) implemented within large organisational document and knowledge management systems, and can at the same time be achieved in a mostly manual way also on a smaller (even personal) basis, it is also a fundamental requirement for the new approach.

2. Tracing of the personal holistic knowledge formation process

   While historicising the users' knowledge work, the solution should trace the knowledge formation process of every single knowledge item. Moreover, it should allow the graphical presentation of the knowledge formation process, as well as the navigation between the different versions of the certain knowledge item. Navigation between the different linked knowledge items should also be offered. Based on context-sensitive search functionality, an effective finding of previously applied knowledge items should be offered. The search functionality should have the possibility to be limited to special kind of knowledge work, to technical file formats, to certain applications or systems, as well as to certain time points or periods in which the searched event appeared.

3. Support the users within their learning processes

   The suggested solution should help the users to recover or find again already known knowledge or information items. It should support the users in remembering which inputs led to the intermediate statuses (versions) of individual knowledge items. Furthermore, it should support the users with the reproduction of already undertaken actions, and support building on analogies and learning from patters based on previous works.

4. Respect of user privacy and data security

   The suggested solution should respect the users' privacy. It should allow the users to decide individually which knowledge items should be captured, and which should not. Furthermore, it should give the users the possibility to retroactively

remove captured items. The data should be stored safely and be secured of unwanted access by others.

5. Based on process / task and infrastructure / generic approach

By combining the in section 3.5 introduced *process / task approach* and the *infrastructure / generic approach*, and applying them from a personal knowledge management perspective, the solution would support specific activities (based on the process / task approach), and at the same time integrate the information and knowledge into a single system. This dual approach would ensure the user a complete and entire view of and access to his personal knowledge base.

5.2 Use Cases

Having described the main objectives of the suggested approach, the following section illustrates examples of typical use cases from students' everyday academic life, in which the solution should support students.

1. Finding and accessing of existing information and knowledge

   Being sure that a certain information is stored somewhere on his PC; a student tries to find it applying different search approaches: manual browsing though his folder structure, the search from his word processing software, the search from his PDF application, and finally also a operation system based desktop search. However, the student cannot identify the desired information. The reasons might be different, like for example:

   - The searched information is not structured enough

   - The searched information is not assigned to the right context

   - The student does not know the essential search terms

   - The search terms which are being used are offering to many results

   - The information is stored somewhere where the searches were not performed (like for example on an external hard drive or a cloud storage)

   - The searched information is saved in a file or format which cannot be contextually searched with the applied methods and applications

2. Finding again and accessing of already identified information and knowledge

    While working on an assignment, the student searches the internet for interesting papers and websites related to his current topic of interests. He performs several web searches, and visits several websites, from which some were more interesting and relevant than others. Some of the found information sources had further links to other sources (like other websites, linked documents etc.), whereas some were accessed directly based on search engine results. At this time, the student does not take any explicit notes and does not save any of the skimmed material.

    A few days later, the student works again on the same assignment and remembers that some days ago he was on a website that was really relevant for his work and which information (like a graphic or a table) he would like to use right now. Unfortunately, although he spends a lot of time for searching again, the student is not able to find again the desired information. This might have different reasons like for example:

    - Not having taken any notes

    - A too long and unclear web browser history

    - Not remembering when exactly it was

    - In the meantime deleted browser history

3. Content comprehensive linking of information and knowledge

    While working on an assignment for subject A, the student comes across an information which is not relevant at all for his current work – however, it might be very useful for an assignment the student needs to o in a few weeks for another

subject B. Focusing just on the current work, no notes of this special information is being taken.

A few weeks later, the student works on the other assignment for subject B and remembers that some weeks ago there was something really relevant and useful. Unfortunately, although he spends a lot of time for searching again, the student is not able to find again the desired information. This might have different reasons like for example:

- Not having taken any notes

- A too long and unclear web browser history

- Not remembering when exactly it was

- In the meantime deleted browser history

4. Sustainable understanding and traceability of previous decisions

Working on a group work over the whole semester, the team makes a decision to do things a certain way after evaluating different possibilities. A few months later, the chosen decision is being questioned, but no team member can really remember why the decision has being met – everyone just knows, that 'we have talked about it and we had good arguments to do it that way'. The team just has the result, but not the arguments why they decided to go that way. The selection needs to be questions again and the long evaluation processes fulfilled for a second time.

5. Sustainable traceability of knowledge formation

After a productive meeting, a student inserts a paragraph into his thesis document. A few months later while working on a thesis, he does not understand the content

of this particular paragraph any longer. He tries to recover where the content came from, and why it is as it is – unfortunately, he does not remember and cannot even link it to the meeting so he could ask the other participant(s) for help.

These exemplary use cases illustrate typical challenges and problems in students' everyday life. After having introduced and described the suggested new approach in the following sections and chapters, section 6.8 illustrates how the system would help with these use cases.

5.3 Intermediate Layer

The last two sections concentrated on the objectives the suggested approach should meet, as well as on exemplary use cases in which it should support students within their learning processes. The following sections of the chapter describe the suggested approach from a theoretical and methodological point of view.

As mentioned before, current Personal Knowledge Management (PKM) tools do not cover the different steps our knowledge went through. Usually they just look at the latest outcomes and results. Some of them have an extended *undo* functionality. Working with *undo* functionality usually replaces the old path through the new one and makes a later path combination impossible. The intermediate results (like $version_2$, $version_3$ and $version_4$ in Figure 33) are typically deleted, and some potentially valuable insights removed and no longer available for future work.
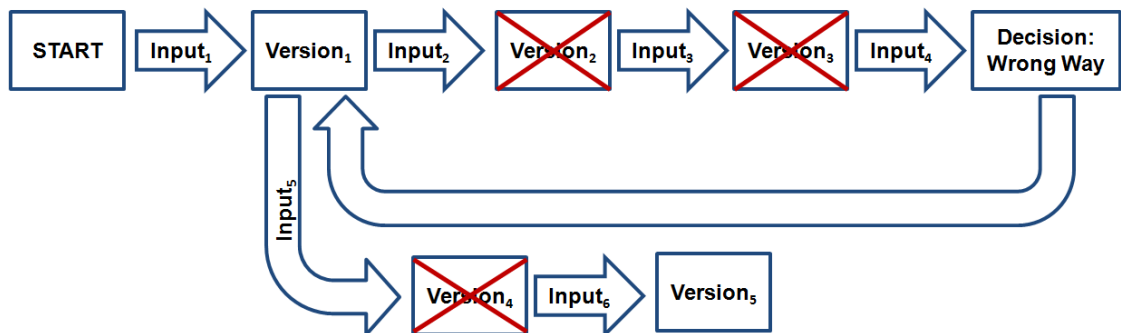
Figure 33 : Knowledge Formation Paths

Moreover, the undo functionality in current systems looks just on its own environment – but fragmented knowledge requires an overall reflection, a reflection that also considers the inputs of events and detractions that happed within other tools or environments.

In the last chapter problems like short product-lifecycles of PKM systems, the users' willingness to switch to better and newer solutions, as well as the enormous knowledge fragmentation has been discussed. To overcome these issues, and to make sure that the solution will work independently of the used systems and in a sustainable manner, it should be implemented in a logically independent layer. As presented in Figure 34, this layer should handle the knowledge work that is being done within any environment, bringing results from tools of the different PKM categories (as introduced in section 3.5) together. To ensure user friendliness and usability, the solution should be running as a background process, and would connect the different content portions and functionalities from the separate technical environments in use. This would offer a flexible and durable additional value to the market of PKM tools and could make their usage really interesting and effective for a broader audience. This layer needs to remember and to trace the different statuses our knowledge has gone through. This way it will support the replication of content, descriptive metadata and relationships in cases where it is not possible to export them due to high exit barriers. Mainly, it will remember the way our knowledge went through and became what it is today.
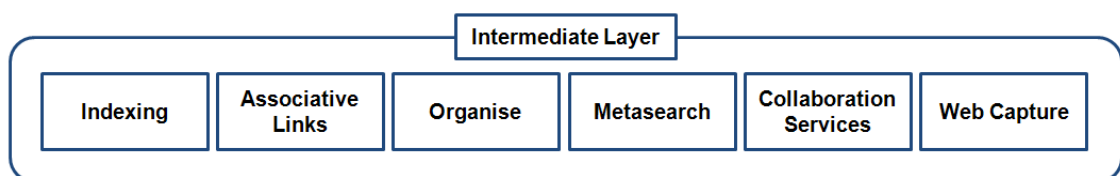
Figure 34 : Intermediate Layer

5.4 Knowledge Formation Process

We can divide our handling into several events we are undertaking. For example, we *read a book*, *receive an e-mail*, *write an e-mail* in response, *visit a website*, *meet other people*, talk to them or *write some notes*. In this context, *event types* represent categories of actions we might undertake during our knowledge work. The concrete events assigned to these types (categories) are then *instances* of the several event types. They inherit the behaviour and characteristics of the assigned event type categories.

Figure 35 represents a simplified illustration of a timeline with several instances of different event types: after *visiting a website* we *edited a presentation document*, *received a message*, *worked on a text document*, *visited several websites*, *read a PDF document* and so on.

Figure 35 : Knowledge Relevant Events

For simplification, we do not consider the duration of the several events at this stage; instead, we define for every kind of event a relevant point of time – called *observation time*. For the event type *receiving a phone call* for example the observation time might be the start of the phone call, whereas for the event type *editing a text document* the observation time might be the moment the document is saved. Table 4 introduces the three possible observation times:

Table 4 : Observation Times

Observation Time	Description
start	The point in time the event begins will be considered as relevant.
stop	The point in time the event stops will be considered as relevant.
save	The point in time the event is being interrupted to save or buffer an (intermediate) version or result will be considered as relevant.

The events are related to one or several entities like for example *books*, *persons*, *websites*, *documents*, *conversations*. Some of the entities refer to the *content*, some to the *descriptive metadata*, and some to the *relationships* as introduced in section 4.2. At the same time some of them are of a basic (single) nature; some consist of a composite of several basic entities. The event type *reading a book* for example is related to the entity *book*, which refers to a specific book and is therefore of a basic entity nature. The event type *sending an email* on the other hand refers to the *email* itself as an entity; this (or the conversation behind this email) consists of a composition of the *email subject*, the *email sender* and the *email receivers*.

Every event has (at least) one *central entity*. Which entity this is, can in some cases depend on the personal point of view of the individual user. Whereas the event type *visiting a website* might have several entities (e.g. *author, publisher, title, abstract, keywords, content, URL ...*) associated, only the *URL* describes the specific website and acts therefore as the central entity. All the other associated entities act just as additional content related information or additional descriptive meta information. Table 5 represents some examples of events and their central entities.

Table 5 : Examples of Event Types & their Central Entities

Event Types	Central Entity
reading a book	an explicit book
editing a text document	an explicit text document
receiving a message	sender
visiting a website	an explicit website
receiving an e-mail	conversation which can be a composite of the sender and the subject or a composition of the sender, the subject and the receivers

Many of the events are related to each other in different ways. One of the strongest relations is, if the *central entity* is the same one. For example, during a week, we can have several events, which are dealing with the same document we are using for our academic assignment. Assuming that the events include writeable actions (actions that change the *content*, the *descriptive metadata*, or the *connections*), in many cases, we will talk about new versions of the same entity; in our case new versions of the same document (Figure 36).

next version of the same document

Figure 36 : Next Version of the Same Document

5.5 Holistic Knowledge Formation Process

Based on the central entity, the suggested system can recognise different versions of the same instance and assign them to each other. Giving the system the possibility to do it not only for text documents but also for many different entity types, it will be already a huge benefit compared to the currently available solutions. This will allow a comprehensive versioning of the users' knowledge work, irrespective of the kind of knowledge work and the system in use. As described so far, this idea is partially being handled by some semantic desktop approaches. Therefore, this alone would unfortunately not meet the formulated objectives. Moreover, the added value compared to a repository of different versions of the specific knowledge part, as well as to the theoretical semantic desktop approaches would not be big enough. For this reason, and to meet objectives b) and c) at least the same attention should be put on the inputs that influence the different steps. From a chronological perspective, the question is:

What happened between version$_x$ and version$_{x+1}$

that had an impact on the new insights in version$_{x+1}$?

On the timeline in Figure 37 we refer to *Input$_1$*, *Input$_2$* and *Input$_3$*.
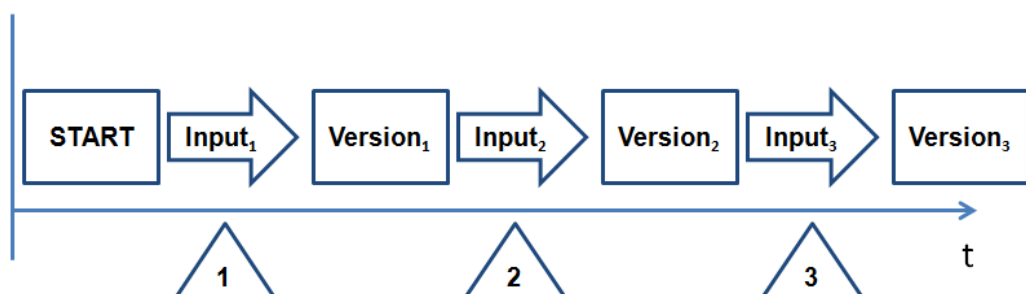
Figure 37 : Knowledge Formation Process on the Timeline

In other words: we assume, that what happened chronologically between *version₁* and

version₂ of the document within our knowledge formation process is (also) responsible for

the new insights and findings that lead to version₂. Large organisational document and

knowledge management systems partially capture the reasons for decisions already,

mostly in a manual way asking the user to describe why this is being done. In parts of

some industries, like the power industry (Jennex 2008) or the banking industry, this is

even part of a regulatory requirement. However, this approach here is different focusing

on the personal view and handling it automatically for the user.

Figure 38 shows an example of an assignment text document. Between *versionₓ* and *ver-*

sionₓ₊₁ of the assignment document the user fulfilled the following events:

- Visited three websites

- Read a PDF document

- Received a phone call

- Opened another text document

- Received an e-mail

- Got another phone call

- Visited another two websites

Figure 38 : Possible Influences on Next Version

All of these events might have an impact on the inputs for the new version$_{x+1}$. However, in practise, usually not every event, which happened chronologically between the two versions of the document, would have an impact on the new version. In many cases, the events did not have anything to do with it at all.

From the point of view of this document, these irrelevant events might be considered as noise (Figure 39). This does not mean that events classified as noise from the point of view of our assignment document are unimportant (noise) in general. They might have a huge influence on other events (e.g. other documents we are working on), but from the point of view of versio$_{x+1}$ of the specific considered assignment document, they might be irrelevant noise.
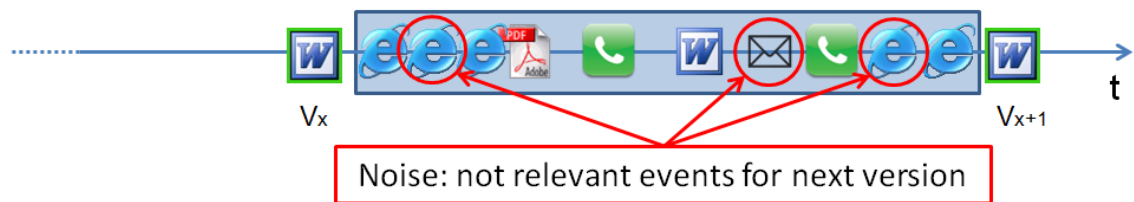
Figure 39 : Irrelevant Noise Events

Within the implementation of such an intermediate layer, it is therefore important to divide the events in relevant and irrelevant events (Figure 40) – always from the point of view of each and every event (instance of an event type).

Figure 40 : Irrelevant & Relevant Events for V_{x+1}

5.5.1 Definition of Event Types

As described before, the actions we are undertaking within our knowledge work are classified as instances of event types. Event types themselves represent predefined categories of actions. They store different data and information, and can have different behaviours and characteristics. The instances inherit the behaviours and characteristics of the assigned event type categories.

The definition of the considered event types is a two-step task. First, the user needs to define which event types are in general important within his knowledge work (from his personal point of view), and how these are characterised.

In the second step, the user needs to define which of these event types are only giving input or influencing others, and which ones should be *followed*. For all instances of event types that *should be followed*, a chronological holistic knowledge formation process will be saved and a timeline built.

The suggested approach offers a list of predefined *standard event types*. Nevertheless, it is flexible enough to allow the definition of new event types and their characteristics. Within

the provided predefined standard event types, a suggestion of standard *follow event types*, like for example *editing text documents* and *editing presentation documents* is given – visualised with the arrows in Figure 41.

Figure 41 : Standard Follow Event Types

At the same time, the system allows the user to define which other event types should be followed and how this should be done (like for example PDF documents and phone calls in Figure 42).

Figure 42 : Additional Personal Followed Types

As stated by Jefferson (Jefferson 2006), there is a need for systems with ability to adjust to individual style and working methods. Therefore, it might for example make sense to $user_1$ to follow the formation of a text document, whereas it might make no sense to follow the history of phone calls. Nevertheless, the event type *receiving phone calls* is still of important value as an input (Figure 41) to other events.

User$_2$ might like to add *receiving phone calls* as a follow event type, which will instruct the system to create a holistic knowledge formation process for all central entities of the event type phone call – basically a history timeline of his calls with the concrete person.

5.5.2 Storage of Event Types

A flexible database structure (described in section 6.3), which allows configuring which event types are in general relevant to the user, and how they should be handled, has been designed. Figure 43 shows an example of a possible configuration for the event type *Editing a text document.*

First, some basic aspects like the event type name, the visual icon, the fact if instances of this type should be followed, as well as the in the previous section introduced *observation time* need to be assigned. Moreover, the user has the possibility to define, if external entries of instances of this event type are allowed and possible. This makes especially then sense, when the user is working in different (technological) environments.

Second, the user has the possibility to configure which physical file extensions should be connected to this event type. The system can use the predefined extensions for automated detection and assignment of instances to certain event types.

Finally, for each type a definition of information (entities) which can or has to be stored is predefined. In the example of the text document, it might be the logical *name* as the central entity (underlined), *title, content, description, author, size, location* and the associated *keywords*. Some of the listed entities are mandatory, some are optional only. Due to the fact, that the recognition of versions of the same instance is based on the central entity, the

central entity (in this case the logical name) is always required. Some of the entity can

have a wider range, meaning that more than just one entry is possible. In our example the

text documents we are editing can have several key words.

Event Type		
Icon	Name: To Be Followed: External Entries Possible: Observation Time:	Editing a text document yes yes save

Entities				Extensions
info	range	required		.doc, .docx, .rtf, .sxw, …
Name	1	yes		
Title	1	no		
Content	1	no		
Description	1	no		
Author	1	no		
Keyword	n	no		
Size	1	yes		
Location	1	yes		
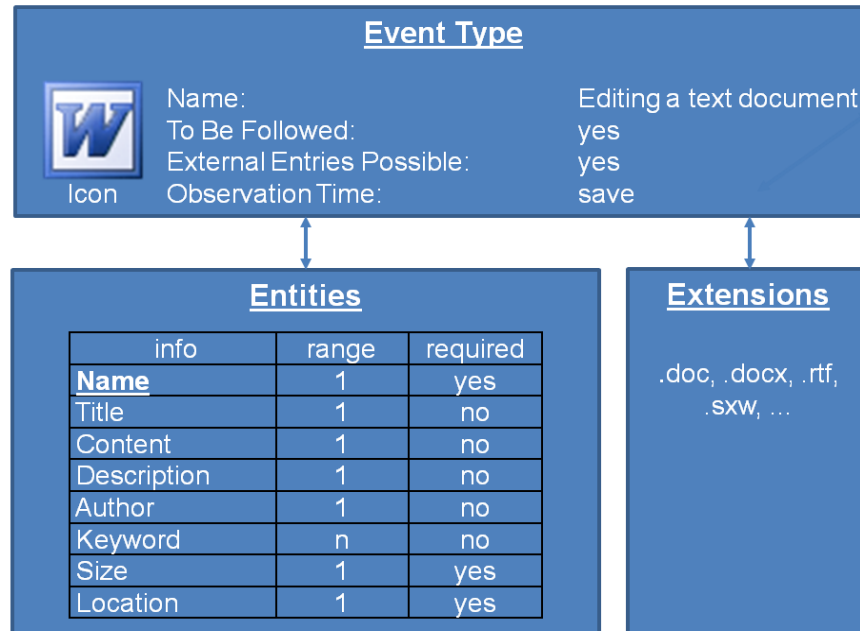
Figure 43 : Event Type - Editing a Text Document

Figure 44 represents another example, the event type *Having a phone call*. Concerning a

phone call, the relevant entities might be just the *caller*, the *callees* and the *duration of the*

call, whereas the caller and the callees together form the composition of the central entity.

Moreover, some additional *notes* might be made for every phone call. There might be

several of them, but they are not required.

Event Type

Icon	Name: To Be Followed: External Entries Possible: Observation Time:	Having a phone call no yes start

Entities

info	range	required
Caller	1	yes
Callee	n	yes
Duration in seconds	1	yes
Notes	n	no

Extensions

no extension
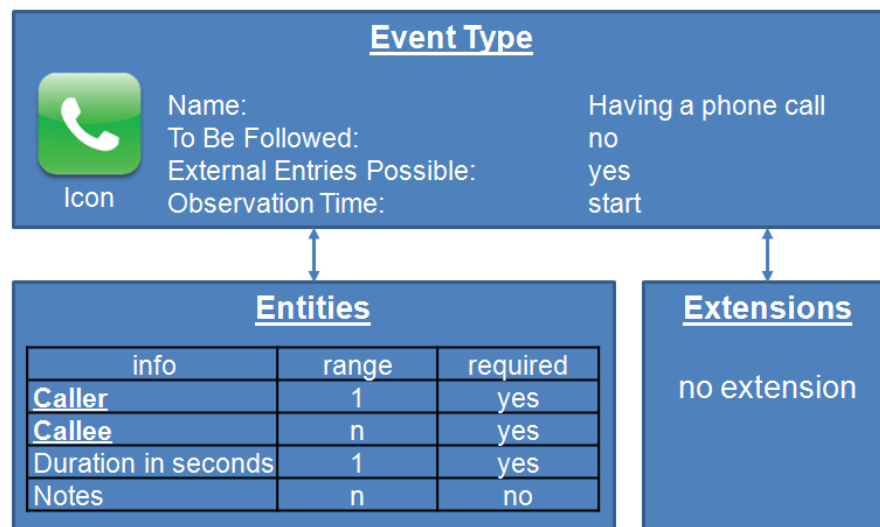
Figure 44 : Event Type - Receiving a Phone Call

Not all relevant events happen within the desktop environment. It is therefore necessary to open the system to external entries. An application programming interface (API) allows the import and synchronisation of events that happened for example on the Smartphone (phone calls, text messages, visited websites, ...). The API can also be used to consider information that are stored within external systems like electronic calendars. In this case, it might be difficult to capture the content of a face-to-face meeting, but the information that there was a meeting, where it took place, and who attended it, might be helpful as well – and this information might be accessed electronically. Besides this, manual entries, like for example the fact that the student reads a paper book, are allowed as well. Section 6.4.4 describes the API and the manual import functionality from a technological perspective.

5.5.3 Influences of Events on other Events

As mentioned before, not every event has an impact on other events and we need to consider only the relevant ones (Figure 45).

Figure 45 : Events with Relevant Influences

However, even within the relevant events the influences can vary a lot. In the example of our academic assignment, the user *visited two websites*, *read a PDF document*, *received a phone call*, *worked on another text document*, *received another phone call* and finally *visited another website*. All these events had an impact, but not to the same extent. We should therefore assign these events according to their *level of influence* (Figure 46). The higher the *level of influence* is, the more important the linked event was for the outcome of the current version$_{x+1}$. It is important to understand, that this is happening from the point of view of version$_{x+1}$ of our assignment document – the same events might also influenced other actions, but to a completely different level. The assignment of the influences always happens according to the observation time of an event. In the case of instances of the event type *Editing a Text Document* this means, that as soon as the user saves a text document, all the relevant events that might have an impact on the just saved version of the document will be identified and assigned.
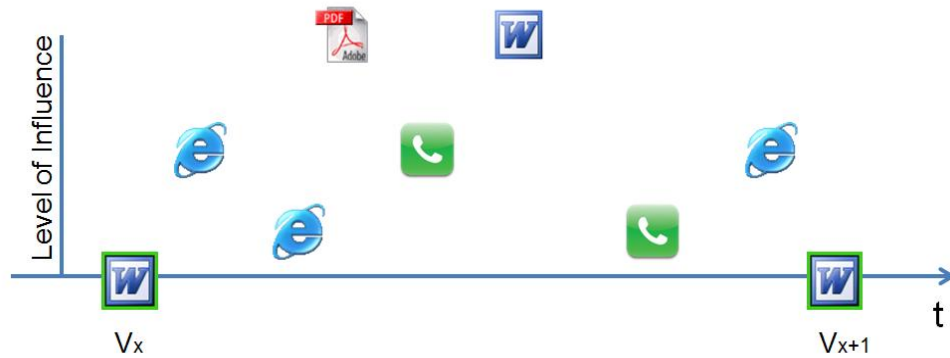
Figure 46 : Levels of Influences

There are several possibilities to automate the assignment of levels of influences to an event. From the Personal Knowledge Management perspective, it makes sense to offer some algorithms and let the user decide which one to use and to which extent. This is especially true in the field of academic learning, where different learner types need to be served. According to Brückner around 70 different models to characterise learners exist (Brückner 2005). Röll developed a model that merges seven preference determination models out of the existing 70, and defines six different learner-types: *analyst*, *creator*, *communicator*, *constructor*, *organiser* and *preceptor* (Röll 2005). A student does typically not correspond to a single learner-type, but his learning preferences are distributed through all six preferences – typically, two preferences are developed strong, two medium, and two weak as presented in the example in Figure 47.
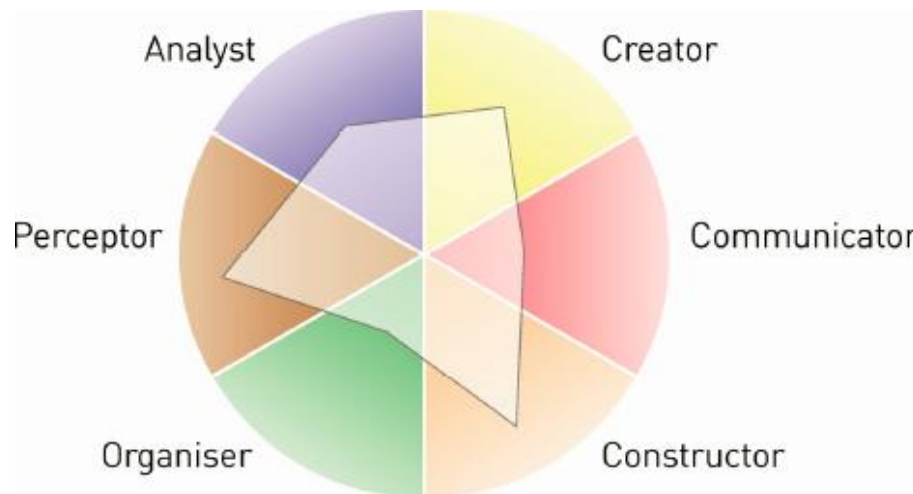
Figure 47 : Learner Types (adapted from Lerno 2013)

This underlines the need to allow the single user to decide individually how the system should work, which event types to consider in which way, but also which weighting algorithms to apply. Subsection 6.4.3 introduces the possible implementation of the automated calculations.

5.5.4 Knowledge Access

The last subsections describe how the relevant information is being identified, extracted, assessed and finally saved. The reason for all this is to allow a later fast and easy access to the information and knowledge base. Depending on the actual task, as well as on the user's individual preferences, access to the information and knowledge base should be offered in different ways like table-based and graph-based, and allow the user to find what he is looking via search functionality or via navigation. These knowledge access possibilities are described in Table 6:

Table 6 : Knowledge Access

Knowledge Access	Description
table	All knowledge items should be presented in a table form. The different versions of the same items should be aggregated presenting the user the number of available versions, as well as the timestamp of the last saved version. The table should present (and mark) the central entity, as well as dynamically the other entities. In cases where the content of the entities does not fit into the column, an enlargeable abstract / thumbnail should be presented. The predefined icon of the certain event type should be assigned to the items to allow an easy visual recognition. The table should be sortable by the different columns. By selecting an entry within the table, the application should present the user the knowledge formation process of the selected item in a visual graph form as described in the next row.
graph	Having selected the knowledge item of interest (either via table, search functionality or navigation), its full knowledge formation process should be presented on the timeline. The visualisation should ensure a clear arrangement and view even if many versions exist. Those events that had an impact on a certain version should be also graphically presented in the chart according to their importance and point in time. By selecting an item of interest on the graph, the associated entities (descriptive meta information) should be presented. In cases where the contents do not fit into the available area, an enlargeable abstract / thumbnail should be presented.
search	The user should have a possibility to search for the required knowledge item or items. This should be allowed according to the following criteria:

Knowledge Access	Description
	Search • Over the full knowledge and information base • Within instances of specific event types only • Within certain entities only • Within a certain timeframe only The search functionality should also allow a combination of any of the mentioned ways. Beside the generic search of the relevant knowledge items, the search functionality should also allow an internal content search within the knowledge items. Depending on the context, the results of the search should be displayed in the table form or the graph form.
navigation	On the graph view, the user needs an easy and fast possibility to navigate between the certain knowledge items. Being on the presentation of the knowledge formation process of one knowledge item, the link to any other related or relevant (influencing) knowledge item should be given.

5.6 Key Aspects

To outline the main characteristics of the suggested approach, the following section summarises the key aspects from a theoretical point of view. The practical concrete implementation design is described in the next chapter 6.

The suggested holistic knowledge formation system would offer a new Personal Knowledge Management approach. It could help to face the earlier discussed challenges and requirements within the evolving learning environment in a sustainable manner.

The implementation of the functionality as a background process would ensure user friendliness and usability. As an independent layer, the solution would be autonomous from and not dependent on the currently and in future available (specific) solutions; it would however be able to consider the knowledge work being done within other environments.

The flexible configuration would consider different personal working patterns, behaviours, processes, learner types and other personal preferences. The idea is based on the observation of our actions. These actions would be categorised within different event types. For every preconfigured event type, the system would log the corresponding activities according to the steps one to four in Table 7.

To support the users in recovering already known or used knowledge items, in reproduction of already undertaken action, as well as in remembering why he decided to do things a certain way, the suggested approach would trace the inputs our knowledge item experi-

enced during its formation. Steps six to eight of Table 7 refer to the assessment part of the linked events.

Table 7 : Approach Steps

Step		Description
1	observe	The system observes every process that is happening within the current environment.
2	identify	According to the user's configuration, the system identifies the relevant processes and links them to the preconfigured event types.
3	capture	The identified processes are captured and made available for the system's next steps.
4	extract	According to the user's configuration of the event types, the system extracts the relevant data and information.
5	save	The system saves the extracted information within the user's database.
6	weight	For all other events, that might have an impact on the current event, the level of influence is determined according to the user's configuration.
7	adjust	To ensure a correct weighting, the user might adjust the automatically assigned levels of influence.
8	save	The levels of influences are saved within the user's database.

Figure 48 visually represents the different steps the system is undergoing within the creation phase of the information and knowledge base. The blue activity refers to step number one in Table 7 and represents a loop waiting for a relevant event to happen. The green activities represent the historicising part (steps two to four) of the suggested approach, whereas the grey activities stand for the assessment part (steps five to seven) of other events on this one.

Figure 48 : Traceability of Knowledge Formation

According to the user's configuration, events that happen not within the main technological environment can be imported belatedly. Depending on the event type, this can happen manually, but where possible also in an automated way via electronic interface.

The suggested system captures the user's behaviour, the visited websites, the meetings he was attending, the messages he received and wrote, the documents he read but also the ones he generated. Bringing all the captured information together could be used to generate a whole profile of the user, which could theoretically be used for different purposes by different parties (in a positive and a negative way). It is therefore essential, that in the end the system would respect the users' privacy by allowing them to decide which items to capture, and which not to capture. It would also cover the relevant aspects of data security. How this is being handled in practice, is described in section 6.7.

The approach would offer an easy and fast access to the stored information and knowledge in different ways: a searchable, sortable and filterable table presentation, as well as a graph presentation with navigation functionality.

As stated before, to qualify as a full Personal Knowledge Management tool, the system should also help the users remembering how they learned things (Clemente and Pollara 2005). This suggested approach will not only offer a strong support of students' learning processes, but at the same time assure a fully qualification as a complemented Personal Knowledge Management tool.

5.7 Summary

To overcome the identified problems of current PKM solutions within learning environments, chapter 5 introduces the suggested new approach that is based on the traceability of our knowledge formation. This chapter starts with the main objectives of the proposed approach. It continues by illustrating typical use cases from students' life, which currently often result in challenges and problems. In such cases, students should be supported by the proposed approach. In the following sections, the suggested approach is introduced and described from a theoretical and methodological point of view.

Designed as an independent layer, the suggested system allows a comprehensive and overall observation and consideration of nearly all our knowledge work activities. Based on an individual configuration, the system supports the building and accessing of the knowledge formation paths for every important knowledge item, later path combination and the access to automatically generated (intermediate) versions of our work. Moreover, it helps the users not only to remember what they did, but gives them also some indications why they did it. This is being achieved by combining different knowledge actions and looking at the influences the different actions may have had on each other.

The concrete system design of the suggested approach is described in the next chapter 6.

6   System Design of the Holistic Approach

The last chapter introduced the theoretical aspects of the suggested new approach of a holistic knowledge formation system. How the concrete system design of the suggested holistic knowledge traceability system could look like, is described in the following chapter.

6.1 Introduction

The described theoretical approach consists of the two main tasks: collecting the information and knowledge of interest, and allowing the users to access the stored information and knowledge in an adequate way. Moreover, the importance and need of an individual configuration has been discussed and argued.

To meet the tasks, the suggested system design consists of three main units and the database (Figure 49). The first one, the *Configuration Unit*, is responsible for the personal configuration of the system, especially the different event types, their attributes and entities. The theoretical aspects of the *Configuration Unit* have been introduced in subsection 5.5.1 and the possible implementation is described and discussed in the following section 6.2.

Based on the selected configuration, the *Knowledge Storage Unit* is responsible for identifying, extracting, and saving the relevant data and information. Furthermore, the *Knowledge Storage Unit* assigns the levels of influences as introduces in the last chapter. The *Knowledge Storage Unit* is described and discussed in section 6.4.

The third part is the *Knowledge Access Unit*. It allows an easy and fast access on the knowledge formation processes of the saved instances of the event types. Section 6.5 covers the design and implementation of this unit based on the theoretical aspects and requirements listed in subsection 5.5.4.

All three units are connected to the central database. Section 6.3 presents possible implementation example of the database. It describes the entire entity relationship model followed by an explanation of every table and the several columns.

Figure 49 illustrates the high-level system architecture concept of the entire system.
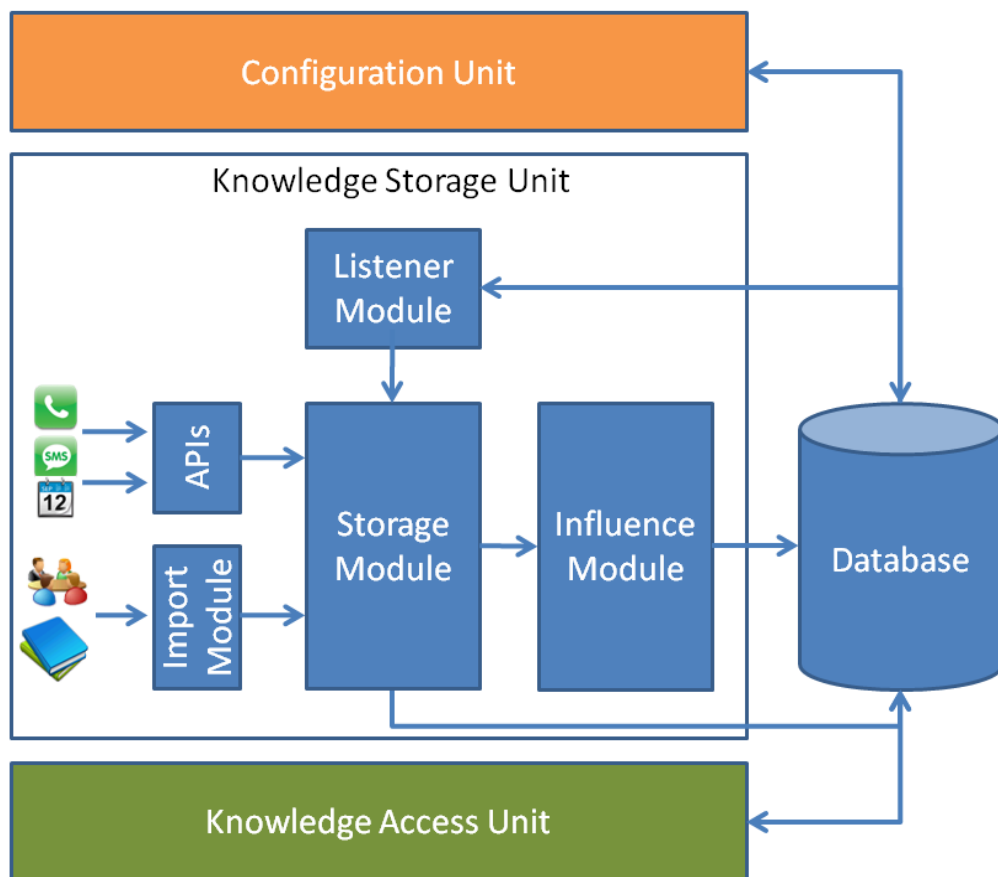
Figure 49 : System Architecture View

6.2 Configuration Unit

The *Configuration Unit* takes care of the entire personal system adjustments and configurations. One of the most critical aspects within software design is usability and simpleness (Microsoft 2000). This ensures that the users will really make use of the system. The full system design follows therefore the requirement of a well-structured, simple and clear application. To ensure this also for the *Configuration Unit*, the suggested approach therefore consists of only four tabs: *Event Types*, *Entities*, *Extensions* and *Add-Ons*.

6.2.1 Tab Event Types

On the tab *Event Types* the user can configure the different types of events. One row in the table represents one event type (Figure 50):

Figure 50 : Configuration Unit – Event Types

Every event type has a *name* and a short *description*. Furthermore, as theoretically introduced in subsection 5.5.1, the user has the possibility to decide:

- Should instances of these event type be *followed*?

    (meaning that a holistic knowledge formation process will be created for them)

- Should *external entries* be possible and allowed?

    (electronically via APIs or manually via a graphical user interface)

- Which *observation time* according to the entries introduced in Table 4 should be considered for instances of this event type?

Moreover, there might be several event types that are *linked to* each other and refer to the same central entity, like for example *Reading an E-Mail* and *Writing an E-Mail*. Both events might be handled differently (e.g. due to the observation time), but in general together they give the overall picture. To make sure to connect instances of these two event types, the user should have the possibility to enter the ID of the linked event type in the column *Linked To*. Usually it will be a bidirectional connection, but to allow also a one-directional connection for some event types, the user should have the possibility to consider every connection type separately.

There is a need to edit and delete existing entries, as well as to add new entries to the event type table. To follow again a simple and user-friendly approach, a context-sensitive menu with functions to manage existing entries is offered. With an *Add New Event Type* button, the user gets the possibility to extend the list. In addition, every event type has an associated graphical icon. This can also be managed via the context menu and will be used later for graphical representations on the timeline.

6.2.2 Tab Entities

The tab *Entities* allows the user to create and manage the different entries, in which the information of the specific events will be saved (Figure 51). These entries are meant to capture the real *content*, as well as the *descriptive metadata* and the *relationships*. Examples might be a name, the content, a description, keywords, hash tags and many others – depending on the event type.

Figure 51 : Configuration Unit – Event Entities

Every entity has a meaningful entity *name* and is of a certain *type*. The *name* should help the user to recognise the relevant kind of knowledge work. In respect of the *types*, the user can select between *text*, *number, time/date* and *media*. This differentiation should on the one hand allow the assignment of the right storage type within the database. In this context, it can also be used for validation reasons – a non-numeric value for example cannot be saved within an entity of the number type. On the other hand, it should make later

searches within the system faster and more efficient. Moreover, depending on the type, different applications are possible, like for example the use of different numeric figures for calculations, determination of the duration of timeframes based on several time/date entries, or combining several text values to a longer, coherent text.

Beside the configuration of entity types, there is a need to assign them to the different event types. Within the last column *Associated Event Types*, the user can link the entities to one or several event types as preconfigured on the first tab (subsection 6.2.1). By opening the pull down menu and entering a range of *one* or *n* for an event type, the link is established. If an entry within this entity is required for an event type, the user needs to mark this within the column *required*. Those entities that should act as central entities (as described within subsection 5.5.1), need to be marked in addition. In all other cases, it depends on the user's preferences which information he would like to store mandatory.

The entity sender within Figure 51 for example acts as a central entity for events of the types *email income*, as well as *email outcome*. Entries marked as central are automatically required – otherwise no automated software-based assignment might be possible.

6.2.3 Tab Extensions

As mentioned in chapter 5, the idea of the system must be to recognise the event type that is currently happening as far as possible in an automated way. To make sure this works in practise, the different possible extensions of physical files can be associated to the several event types. As soon as a file with one of the listed extensions is opened or executed by the user or the operating system, the application can assign the associated event type and the linked functionality automatically. Here it is always a (maximum) one to one link –

meaning that one extension can always be linked to maximum one event type only. Otherwise, the assignment would not be unique and an automated selection of the event type would not be possible. In the unlikely case where one extension can be theoretically linked to several event types, the extension has to be listed without a direct association. In this case, a request prompt might ask the user to select between the assigned options.

The configuration and assignment of physical file extensions can be done on the tab *Extensions* (Figure 52). As in the previously introduces tabs, the table represents an overview of all existing extensions. New entries can be added via the *Add New Extension* button, and the association to a certain preconfigured event type is done via the pull down menu in which all existing event types are listed.
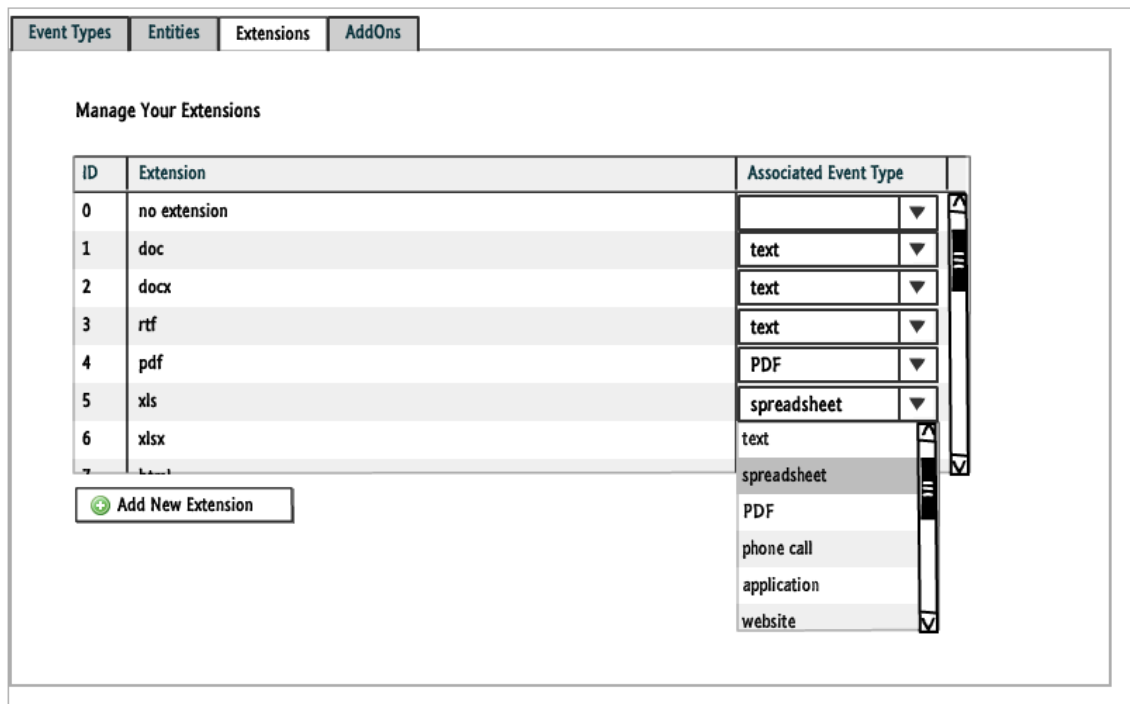
Figure 52 : Configuration Unit – Extensions

6.2.4 Tab Event Add-Ons

To configure the *event types*, *entities* and *extensions* the way the user wants them to have and behave, is only one part of the solution. These settings need to be backed up with functionality. To make sure that for every event type the right algorithms is being applied while extracting and weighting the (right) data and information, it is necessary either to offer the needed functionality in front, or to allow the users to design simply their own solutions. The idea is to offer this functionality via optional add-ons. Such a modular concept has been successful within a wide range of contemporary applications. Implementing the suggested approach under the open source (The Open Source Initiative 2013) regulations, would for example allow a community driven, flexible, adaptable and growing functionality. However, this means, that if the data extraction and later import into the system's knowledge base should happen automatically, there is a need to have a specific piece of software, which can be simply integrated into the system, and will take over the required functionality for instances of exactly this certain event type. On the tab *Add-Ons* the user is meant to import new, as well as manage the already imported add-ons and assign them to his event types.
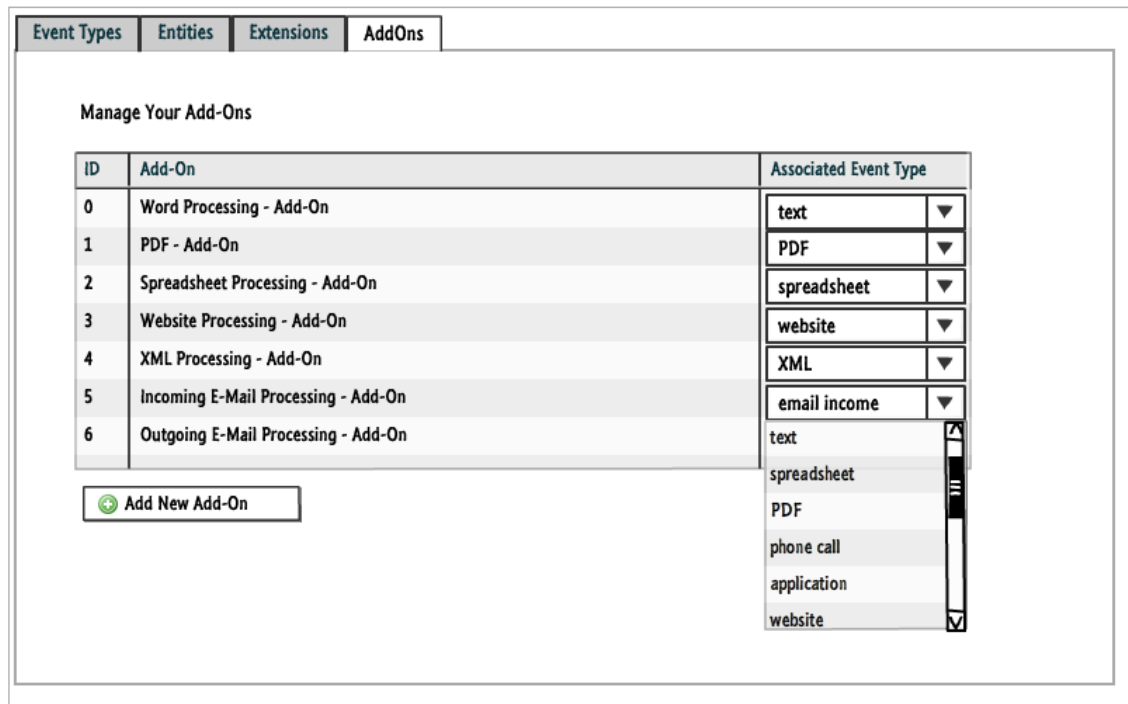
Figure 53 : Configuration Unit – Add-Ons

Depending on the file format, the functionality to extract the right data and information may vary a lot. For many existing standard file types such functionality already exists and is in many cases available under open source license agreements for free; some example are *Apache POI* (Apache 2013a) for Microsoft Office documents, or *Apache PDFbox* (Apache 2013b) for PDF documents. However, once the relevant data and information is extracted, it needs to be inserted into the system – and this can be done in the same way for all event types. It is therefore possible to define an interface how certain information could be integrated into the system. This interface is described in subsection 6.4.4 and could also be published and used by external add-on developers to map the right fields with the corresponding extracted information.

6.3 Database Structure

To consider the different personal configurations, a flexible database structure needs to support the application. This could be implemented in different ways, like for example within an object-oriented database (Dittrich et. al. 2011), as a multidimensional data warehouse (Chaudhuri and Dayal 1997), or within a relational database. During the research, the last approach, the relational database one, has been designed. It allows the storage of the configuration settings, the actual information (content), the descriptive metadata, the relationships, as well as the influences between the several entries.

Figure 54 represents the entity–relationship diagram of the database structure. The orange framed area marks the tables responsible for the storage of the configuration data. The other tables (green–blue framed) store the content, descriptive metadata, as well as the influences. Once stored, this data and information is also used within the *Knowledge Access Unit* as described in section 6.5.
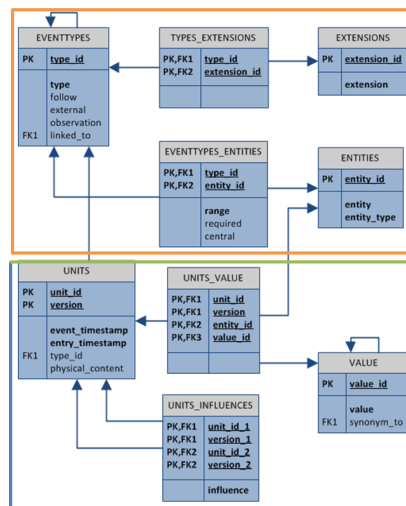
Figure 54 : Database Structure

The detailed database design is described in Appendix A.

6.4 Knowledge Storage Unit

The last two sections described the *Configuration Unit* and the *database design* of the suggested application. To fill the database according to the individually configured values, the *Knowledge Storage Unit* is used. Technologically the *Knowledge Storage Unit* consists of the *ListenerModule*, the *Storage Module*, the *Influence Module*, as well as the *API*s and Interfaces for automated and manual data import (Figure 55).
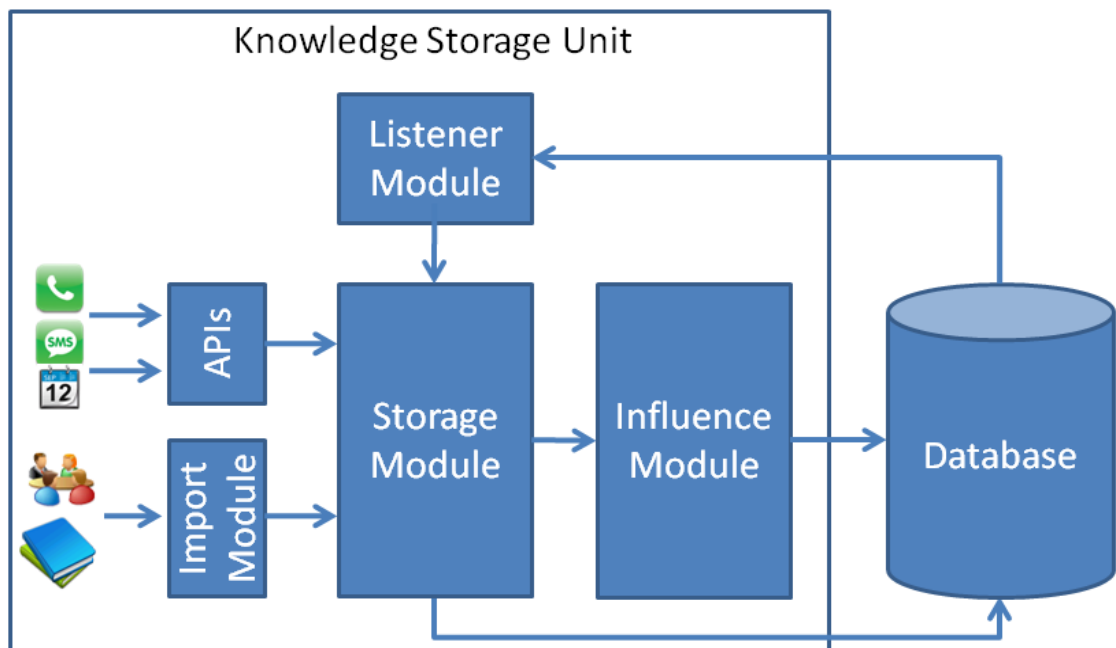
Figure 55 : Knowledge Storage Unit

6.4.1 Listener Module

According to the personal configuration as described in subsection 5.5.2 the *Listener-Module* is running as a background process on the desktop environment, observing the actions being performed and comparing them to the configuration. As soon as something relevant happens, meaning an action that fits to a preconfigured event type is being performed, the *Listener Module* informs the *Storage Module*.

6.4.2 Storage Module

Being informed by the *Listener Module*, the *Storage Module* captures what is happening. It extracts the predefined information, and assigns it to the configured entities. The extraction of the relevant data depends on the format of the central entity the current event type is referring to. For Microsoft Office documents for example it would be possible to apply the *Apache POI* (Apache 2013a) or the *catdoc* (Freecode 2013) functionality, for RTF documents the *unrtf* (FSF 2013) one, and for PDF documents the *Apache PDFbox* (Apache 2013b), *pdftotext* (Glyph & Cog 2013) or *pdfinfo* (Haas 2013) one. All of these modules are proven algorithms, and are available under open source licenses. Here the Storage Module is taking advantage of the provided functionalities by the different Add-Ons (see subsection 6.2.4). Once the relevant information is extracted correctly, it is assigned to the associated entities. These entities are then saved within the database as a new unit version of the instance.

6.4.2.1 Major Versions

Some users tend to save intermediate version very often. To avoid complexity and allow a better overview, the system should summarise these smaller versions and create *major versions*. Meaning, that while saving the data, a comparison with the old version should be done. Depending on the event type, this could happen either via comparing the *content* the *file size*, or by just looking at a *timeframe since the last version*. In many cases, a combination of these parameters might make sense; especially by combining the *timeframe since the last version* with the other parameters. Comparing the content might be, depending on the actual event type, a difficult and time-consuming task. Looking just at the file size might be too dangerous by assuming that if the file size did not change, the

file itself did not change as well. In most cases, this will be true, but by combining this with the time factor, the results might be much better.

In addition, the fact that no other events happened in the meantime could also be used to identify a version as a *security saving* only without being a real informative value added version. If such a saving is being identified, it should temporary be stored in the database. This will make sure that in the unplanned event of a system breakdown the information will not be lost. As soon as the next major version is being created, the temporary saved version or versions can then be removed again. This temporary saving and clearance should happen as often as it is needed to identify a major version according to the criteria suggested above. However, a general approach could be based on the fact that a version becomes only then a major version, if enough other knowledge tasks happened in the meanwhile.

Figure 56 illustrates a major version V_x of a certain document, followed by five further versions that classified themselves as security savings only ($V_{xs1} - V_{xs5}$).
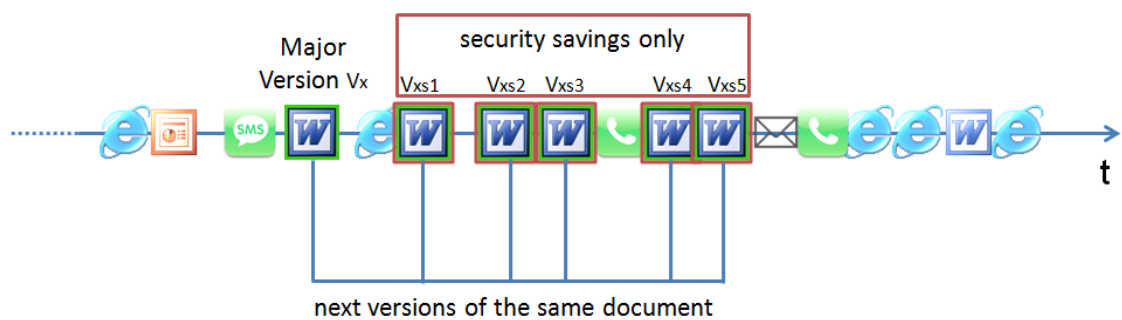
Figure 56 : Security Savings

Once a new saving is being identified as a major version (V_{x+1}), the corresponding information is being saved and the influences calculated starting with the time of the last major

version. In our example, we should therefore not consider only the timeframe between the security saving v_{x5} and the major version V_{x+1}, but the entire time frame since the last full version V_x. Once this action is finished, the temporary security savings might be removed from the system as presented in Figure 57.
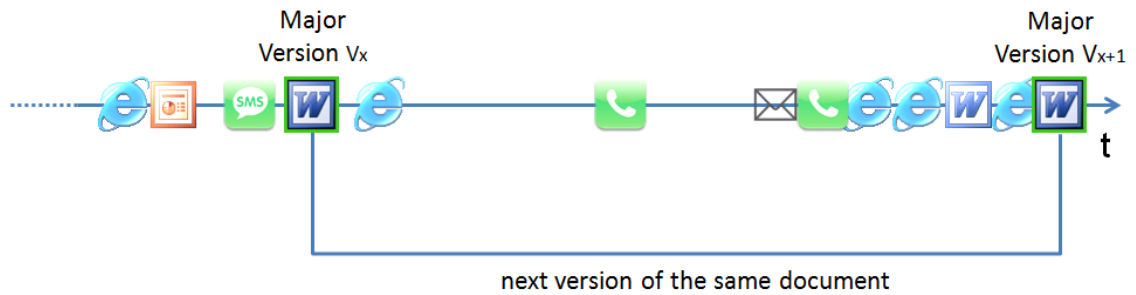
Figure 57 : New Major Version replacing Security Savings

In the case, that no further major version will be created, the last security saving represents the last available (final) version of the document. In such situations, the last security saving should also become the last official version of the item. This could either happen on the fly continuously, meaning, that the last saving would always be regarded as a full major version, or as a subject of a preconfigured timeframe as presented in Figure 58. The advantage of doing it continuously would be the fact that at any point in time the system would have the latest version available (as a major version with all extracted information). The disadvantage on the other hand might be, that even this version was not really a new version, the whole extractions, savings and calculation of the influence levels need to be done – without having the upfront guarantee, that a few seconds later this version might be replaces by a new major version.
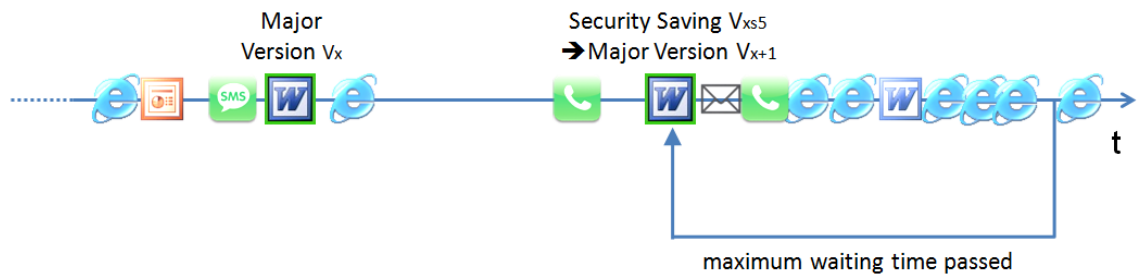
Figure 58 : Existing Security Saving becomes Major Version

6.4.2.2 Virtual Versions

As some users tend to save their work every few moments, others might tend to have very long timeframes between their versions. In this case, too many changes may happen, based on too many influencing events. It would become difficult to offer the user an easy, understandable and clearly laid out overview of his knowledge work. Therefore, the system might build virtual versions of our knowledge work, if the period since the last version is larger than a preconfigured timeframe and the specific document (central entity according to Table 5) is still in use as illustrated in Figure 59.
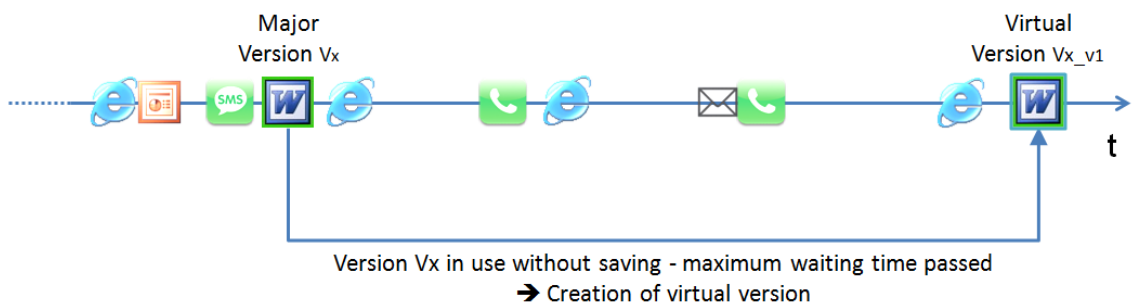
Figure 59 : Creation of Virtual Version after Predefined Time Period

Another approach might be not to look at a preconfigured timeframe, but at the amount and quality of other associated events. At the point, when more than a preconfigured number of other relevant events happen, a virtual version might be created (Figure 60).

This approach however would require a live comparison of the influences, and not as originally meant only at the saving phase.
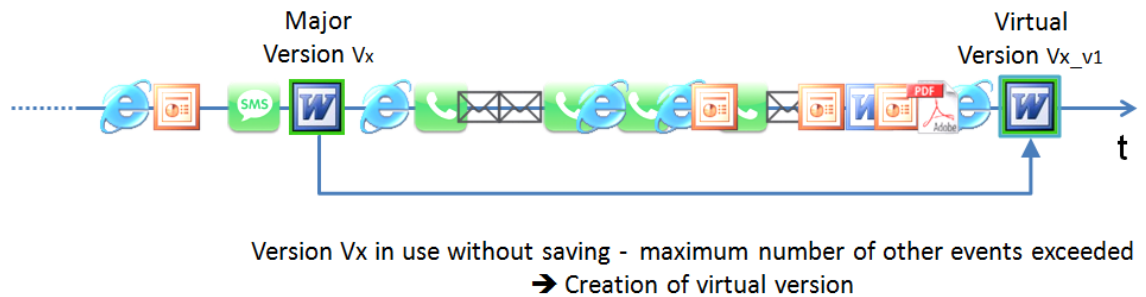
Figure 60 : Creation of Virtual Version after Maximum of Events Happened

How big the timeframe, as well as the number of the other relevant events should be, should be left to the user. However, the fact that the user is not making too many intermediate savings of his work, could let us assume, that he does not really need them. Therefore, a saving every 12 or even 24 hours should be sufficient, if the number of other related relevant events does not exceed a number of around 20. Based on the chosen display presentation, presenting more than 20 relevant events would make the application user-unfriendly and confusing. A combination of both approaches could also here make sense; however, if based on the timeframe approach, the virtual versions should be removed if real major version would follow. In the case of the maximum number of related other events, the virtual versions should be kept as such for clearness reasons within the presentation.

The described functionality so far represents the versioning part within the suggested system. The functionality described in the next subsections is referring to the questions why the user decided to do things a certain way.

6.4.3 Influence Module

Once these steps are fulfilled, the *Influence Module* identifies which other events might have had an impact on this particular event. For instances with a version larger one, meaning events referring to central entity that exists already in the system, the timeframe between the last and this current version will be considered. For events with new central entities, a fixed timeframe backwards will be taken into consideration. All events that happened in the identified timeframe will be examined and the levels of influence determined. How big the timeframe should be, is again up to the individual user and its user behaviours and patterns and should therefore be individually configurable. Some users tend to create a document prior starting any other related knowledge work. In such case, a short time period might be sufficient. Others tend to collect their inputs first, and once they have won enough insights, they start to include the findings in their work (document). For such users a much longer period should be chosen.

6.4.3.1 Content-Based Influence Assignments

To calculate the actual level of influence, a mixture of several algorithms is possible. The proportion depends mainly on the event type. For electronic content-based events, a comparison of the contents is one possibility. The implementation itself could be build up on the statistical method of keyword analysis (Heyer et al. 2002). Here an abstract of each and every event is made. Therefore, a list of stop words is filtered of the content, leaving behind the meaningful content. This meaningful content is then used to build a keyword vector. The system can now compare the abstracts of the current event and all the chronologically relevant other events, and assigns the level of influences.

Taking the latest working version of a conference paper document as an example, all the events that chronologically happened after the last version (Version$_{73}$ in Figure 61) need to be considered.

Knowledge Formation Process of
I-Know Conference Paper - IKnow2012.docx [55010]

Type	ID	Version	Influence on new V 74
PDF	52003	1	2
PDF	52001	2	5
E-Mail	52007	1	3
Text	52021	6	2
E-Mail	52011	1	1
Phone Call	52020	1	4
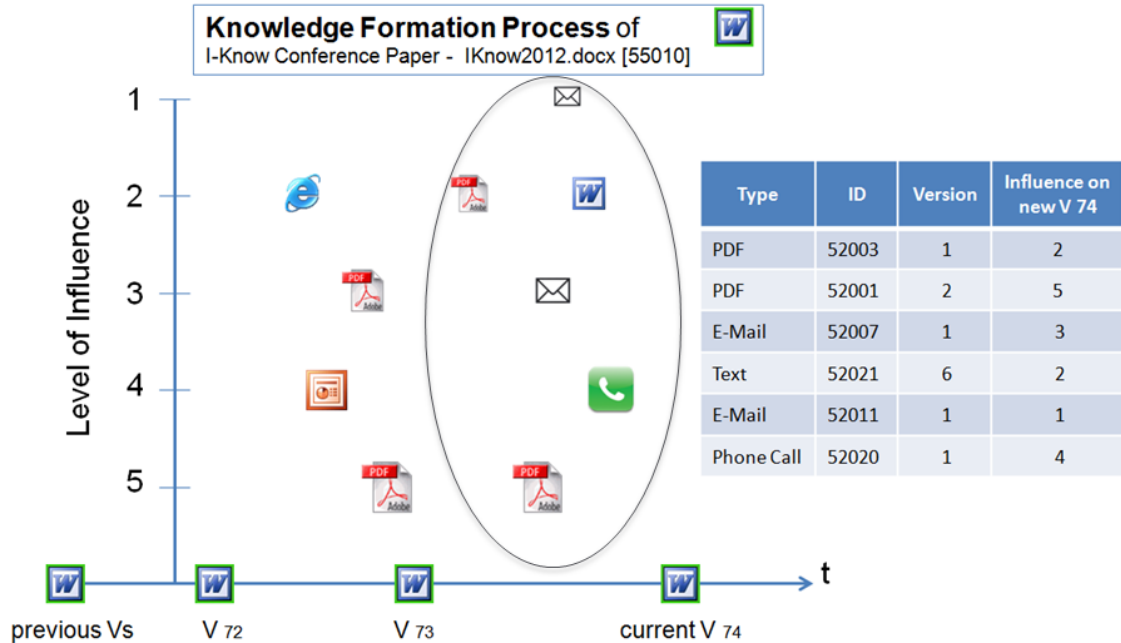
Figure 61 : Assigning Influences to Relevant Events

For all those events (marked with the oval) the system should create content abstracts as they happen. This abstracts should then be saved within the database, and as soon as their corresponding events have been identified as relevant, these abstracts will now be compared with the newly created abstract of the latest conference document version$_{74}$. For each event, a level of influence is calculated and assigned (Figure 61).

In respect of the content-based comparisons, another issue needs to be considered: the language of the compared input sources. Assuming that the input PDF document for example is in English, the language of the paper document we are working on is German, we need to use a dictionary and a thesaurus to allow a high quality content comparison

results. This however could be a challenging task including several steps like identifying the right language, translating the contents to the reference language, and finally compare the contents. Considering topics like synonymous, multiple meanings of the several words, as well as quality of context translation (in contrast to single word translation), this might be in practise a challenging issue. Currently there are several internet services available (including Google Translate (Google 2013b)) offering such possibilities. However, although the results are contentiously improving, the real satisfaction with such functionality would also depend on the affected languages, as well as the areas of topics the users are using them for.

The algorithms applied within the content-based recommendation methodology as described in section 3.8 can also be applied here to detect related items. The research, which is being done in this field, identifies many possible applications and as we can see from the user perspective, the results become constantly better. As mentioned before, the biggest problem within the content-based recommender systems is that to make sufficient recommendations, there is a need for a good knowledge base as a starting point. Within the suggested system this is given.

6.4.3.2 Manual Influence Assignment

Knowing that the automatic assignment will never be as good as the user's knowledge at this time, there is a need to allow the user to adapt the automatically calculated influences. The application should therefore ask the user to confirm or if needed to adjust the levels of influences as presented in Figure 62. This could happen via a pup-up window on which the user would have the possibility to tune the values of the bars that represent the individual events.
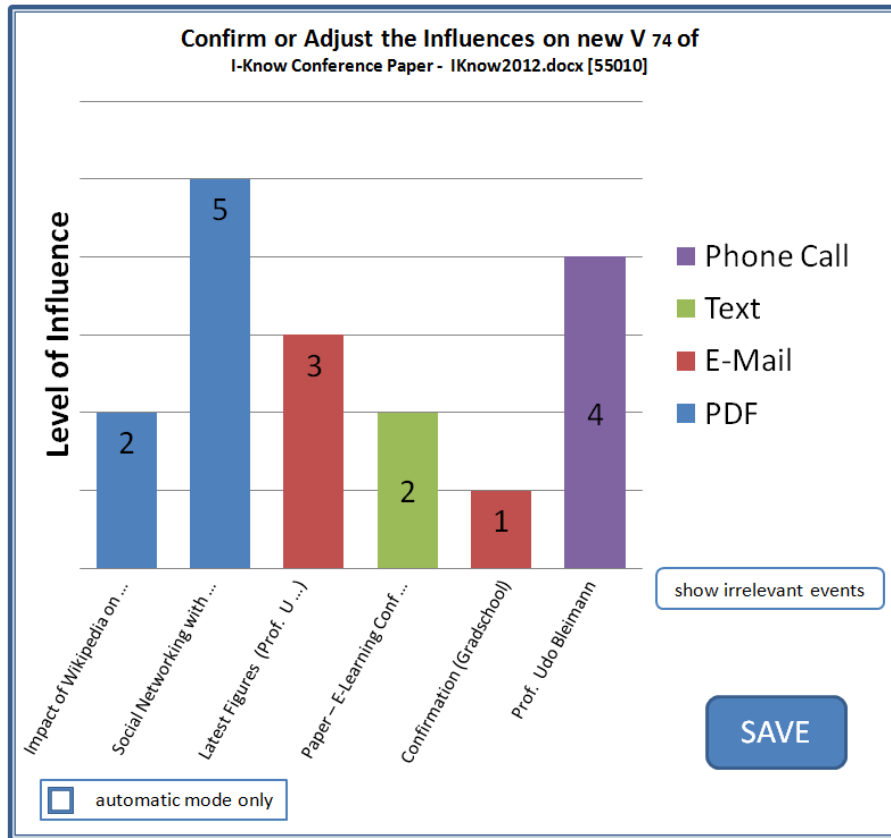
Figure 62 : Result of Weighting – Confirmation / Adjustment

If the automatic algorithm is good enough for the user, he could skip this step in future by marking the *automatic mode only* button. The result will be a small notice during the saving process as presented in Figure 63.
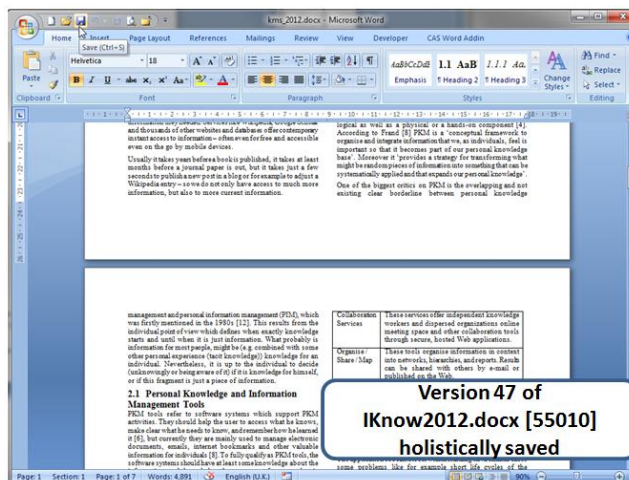
Figure 63 : Automatic Mode

In normal mode, the application will show only the relevant events, meaning events, which had a positive level of influence on the current event. This allows an easier overview and if necessary fast adjustment. Nevertheless, also the as irrelevant marked events should be made assessable and manually assignable, for example by pressing the 'show irrelevant events' button (Figure 62).

6.4.4 Import Module & Interface

The *Import Module* represents a simple manual data entry possibility. The masks are designed and created dynamically according to the entity configuration the user has made before. For every event type, which has been assigned with *external entry possible = yes,* an import functionally is offered based on the corresponding Add-Ons. This makes sure, that events which happen outside the desktop environment and are not able to be imported in an automated way, can still be inserted and considered.

Figure 64 represents the graphical user interface of the manual *Import Module* for the two exemplary event types *Text* and *Phone Call*. For the *Text* event the relevant document has to be selected, the time and date on which this event happened entered, and the new event can be imported. From here on, the described *Storage Module* will extract the relevant information and save them according to the described manner. To ensure better data quality, the user has the possibility to explicitly mark the entry as a new one (not a further version of an existing one). This will tell the system not to try and find a possible accordance to already existing entries. For instances of new event types, this functionality need also to be integrated into the certain add-on as presented in subsection 6.2.4.
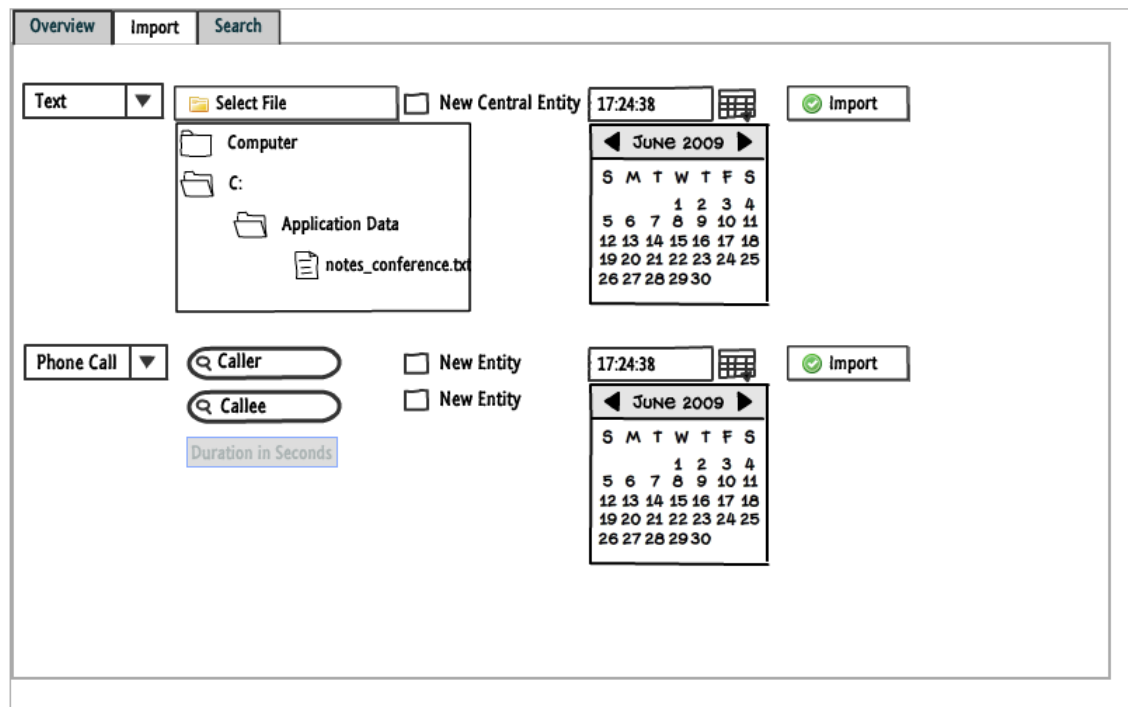
Figure 64 : Knowledge Storage Unit - Manual Import GUI

The second example refers to the event type *Phone Call*. The user has the possibility to search the database for already existing callers and callees, or to enter new ones. In addition, here, the event time and data needs to be entered, and the event can finally be inserted into the system. Depending on the personal configuration of the event type, other criteria like for example the duration of the call might also be considered.

To import the electronic captured events, which happened outside the system, an XML (Extensible Markup Language) based *application programming interface* (API) has been designed. This API can be used to implement interfaces for the common systems and services like iOS, Android, Windows Phone, Google Calendar and others. It is also the same interface, which needs to be accessed while designing and developing add-ons for new event types as mentioned in subsection 6.2.4. Figure 65 presents the schema of the XML import interface. XML files based on this schema can import one or several events (units)

of different event types into the system. Beside the standard (static) information like the *unit_id*, the associated *version number*, the assigned *event type id*, and the *timestamp* describing when the event happened, each entry can consist of one or several entities. Each entity entry is built up of the *entity id* that is directly referring to the entity type (as presented in Table 19 in Appendix A), and the value of the explicit entry.

```
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">
<xs:element name="unit">
 <xs:complexType>
  <xs:sequence>
   <xs:element name="UNIT_ID" type="xs:int" />
   <xs:element name="VERSION" type="xs:int" />
   <xs:element name="EVENT_TYPE_ID" type="xs:int" />
   <xs:element name="EVENT_TIMESTAMP" type="xs:dateTime" />
   <xs:element name="ENTITIES">
    <xs:complexType>
     <xs:sequence>
      <xs:element name="ENTITY_ID" type="xs:int" />
      <xs:element name="VALUE" type="xs:string" />
     </xs:sequence>
    </xs:complexType>
   </xs:element>
  </xs:sequence>
 </xs:complexType>
</xs:element>
</xs:schema>
```

Figure 65 : XML-Import Schema

Once external events are added (either manually or via the interface) the data needs to be stored within the database, and the levels of influences assigned. Due to the fact, that the external import of an early event might happen belatedly chronologically after some

other, automated captured events, the system needs to calculate not only the level of influences of other events on the imported one, but also of other on the imported one. Figure 66 presents an example of an external inserted phone call event. All the events, which happened before the phone call on the timeline, could have an impact on the inserted phone call event. On the other hand, all the already stored events that happened chronologically after the phone call need to be checked regarding the question if the phone call had an impact on them. If it had, then not only the influences of the other events on our phone call need to determined, but also the influences our phone call had on other events supplemented.
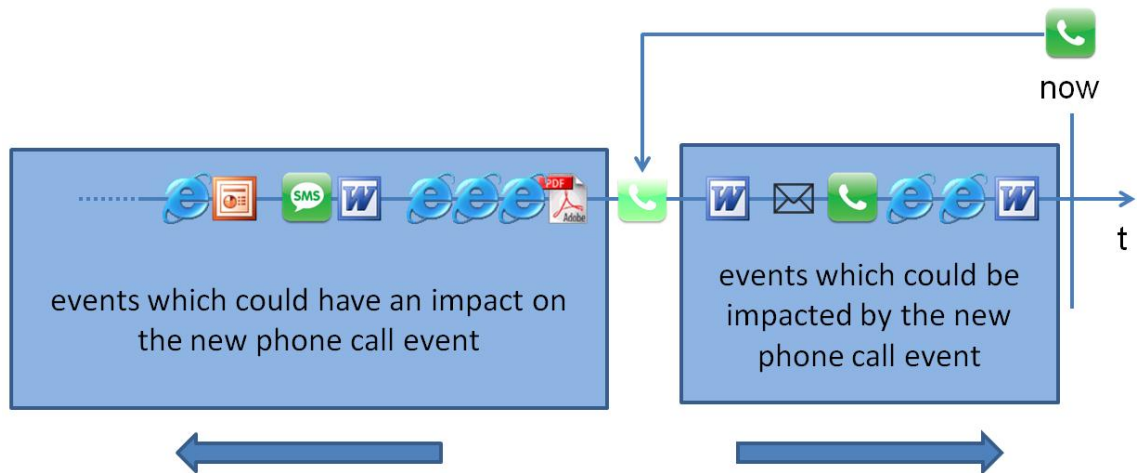
Figure 66 : Levels of Influences within External Inputs

6.5 Knowledge Access Unit

The approach is to offer an overview and a timeline of the knowledge formation process for the learning processes the student is going through. This will not only allow an intelligent versioning, but also go beyond this by remembering the whole picture and allowing the user to go back, reproduce, and combine different paths of his past work. It will not always answer, but in many cases at least help essentially to remember why the user has chosen a certain way.

There are many possible ways to access this collected knowledge and information afterwards. One is a table of the central entities like for example the explicit documents we are working on. Depending on the configuration of the associated event type, different meta-information (entities) can be displayed. Here all versions of the same entity are aggregated into one entry (Figure 67).

	Unit	ID	Versions	Last Version	Type –Dependent Information 1	Type –Dependent Information 2	...
...							
	Title Impact of Wikipedia on ...	55002	2	2012-01-15 12:33:01	*Author* Mike Mayer	*Keywords* Social, Learning, e-Learning	
	Sender Max Mustermann	55003	-	2012-01-15 12:58:23	*Content* Found what you were ...		
	Title Paper Risk Management ...	55004	4	2012-01-15 13:34:54	*Author* Witold Thaul	*Keywords* I-Know Conference 2012	
	Title How many tweets?	55008	1	2012-01-15 15:20:11	*Author* John G., twitter.com	*Keywords* Twitter, Facebook, LinkedIn	
	Caller / Callee Witold Thaul – Udo Bleimann	55009	43	2012-01-15 16:01:43	*Caller* Witold Thaul	*Callee* Prof. Udo Bleimann	
	Title I-Know Conference Paper	55010	75	2012-06-23 16:57:12	*Author* Witold Thaul	*File* IKnow2012.docx	
	Subject New Figures	55011	3	2012-01-15 17:17:15	*Sender* Prof. Udo Bleimann	*Receivers* Witold Thaul, Max...	
...							

Figure 67 : Database of Personal Knowledge Parts

The first column visually represents the event type with the assigned event type icon. In the second column, the central entity is presented, followed by the unique unit ID, the number of stored versions and the timestamp indicating when the last version was created. The following columns refer to the entities assigned to the event type and can differ from event type to event type. For a better understanding and overview, the referred entity name is always displayed above the actual value. Whenever the user needs more information on the entire knowledge formation processes a certain entity has gone through, he can click on the entry and open the associated timeline as presented in Figure 68.
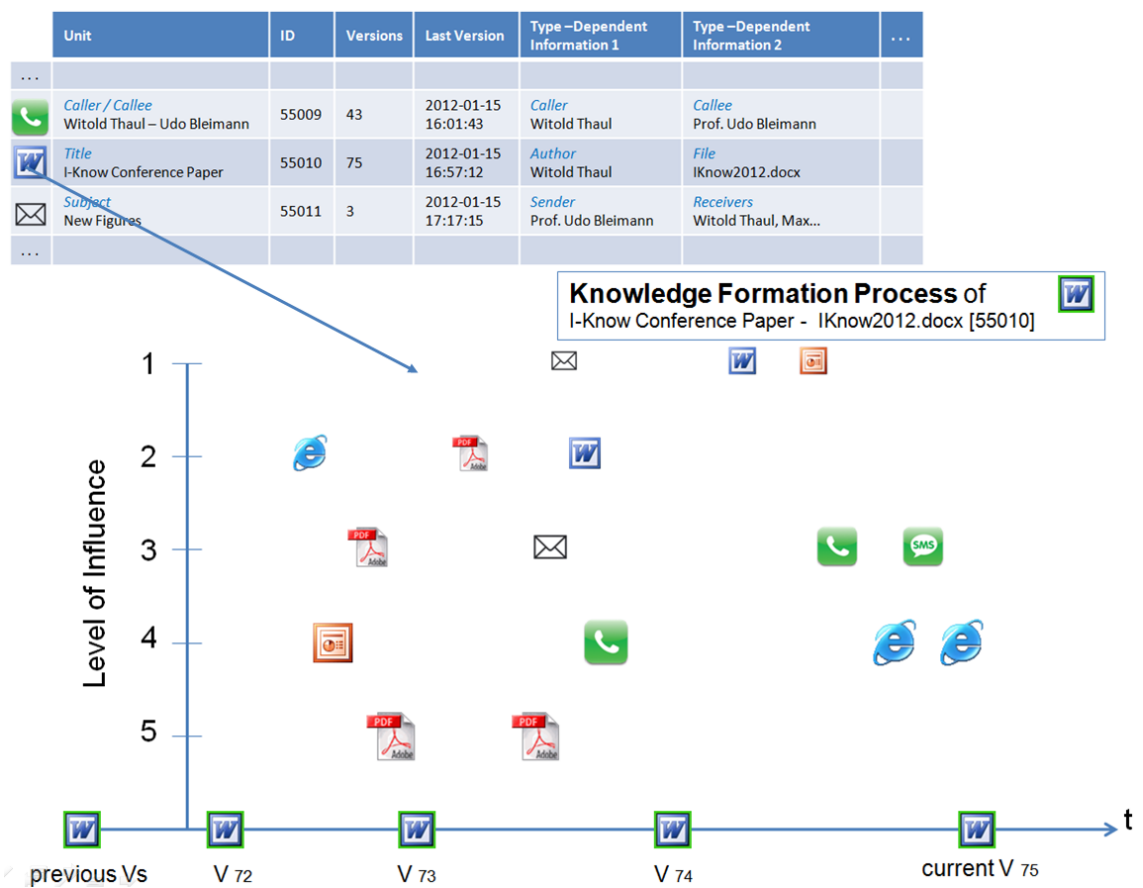
Figure 68 : Holistic Knowledge Formation Process

Here not only older versions of the same instance can be found, but also the assigned other events, which influenced the certain versions are displayed. By clicking on them,

the user gets all the associated information (entities) in a summery view added by a short context sensitive menu. The menu gives the user the possibility to readjust the level of influence, to remove the marked event from the point of view of the corresponding event (meaning setting the level of influence to 0), or to remove the marked event completely from the database. Furthermore, the user can open the knowledge formation process of the linked event, which will then be opened in the same manner as the one described above and represented in the lower part of Figure 68. This way, navigation through the entire knowledge base is possible.

6.5.1 Unit Search

To support the user in finding the entries he is looking for, the system should contain extended search functionality. Beside a context sensitive search term entry field, the possibility to restrict the search by defining the *event type* should be given. Based on the event type selection an additional entity selection should be possible. Furthermore, a limitation to a certain timeframe, at which both limits (from / till) are optional would be offered. These optional entries would offer the possibility to find the needed information faster and in a more structured way. Figure 69 represents a possible graphical user interface for the search functionality.
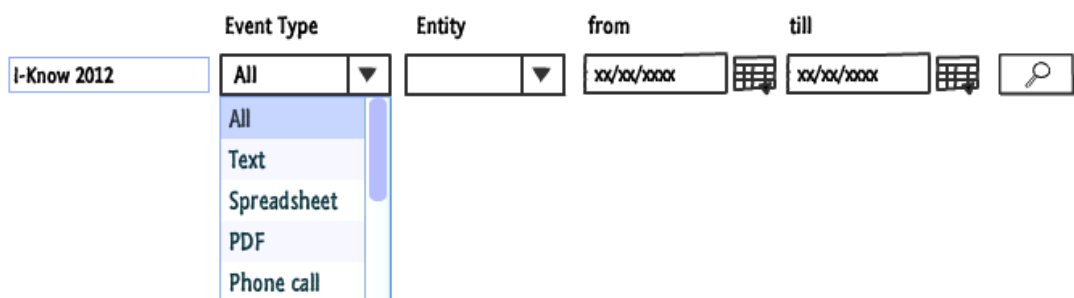
Figure 69 : Knowledge Access Unit - Search GUI

The results will be represented in the already introduced table form as presented in Figure 67. Beside the search functionality, the table should be able to sort the entries by clicking on the column label.

6.5.2 Context Search

Event types associated to a central entity of a text-based nature (like for example *texts*, *spreadsheets*, *presentations*, or *PDFs*) offer the possibility to not only look at the creation process of the whole central document, but allow also a deeper look into the single contents. These documents may contain text paragraphs, sentences, single words, and tables. A typical question in this context might be:

Since when is this single paragraph in that document and what was responsible for it?

To answer this question the solution offers a simple functionality. The user could for example mark the relevant content block, clicks on the *history search icon* in the context sensitive menu, and the software would be looking for the first version of that concrete document in which the marked content block appeared.

Figure 70 shows a Microsoft Word text document in which the user marked a paragraph, and selected the *History Search* option within the context sensitive menu.
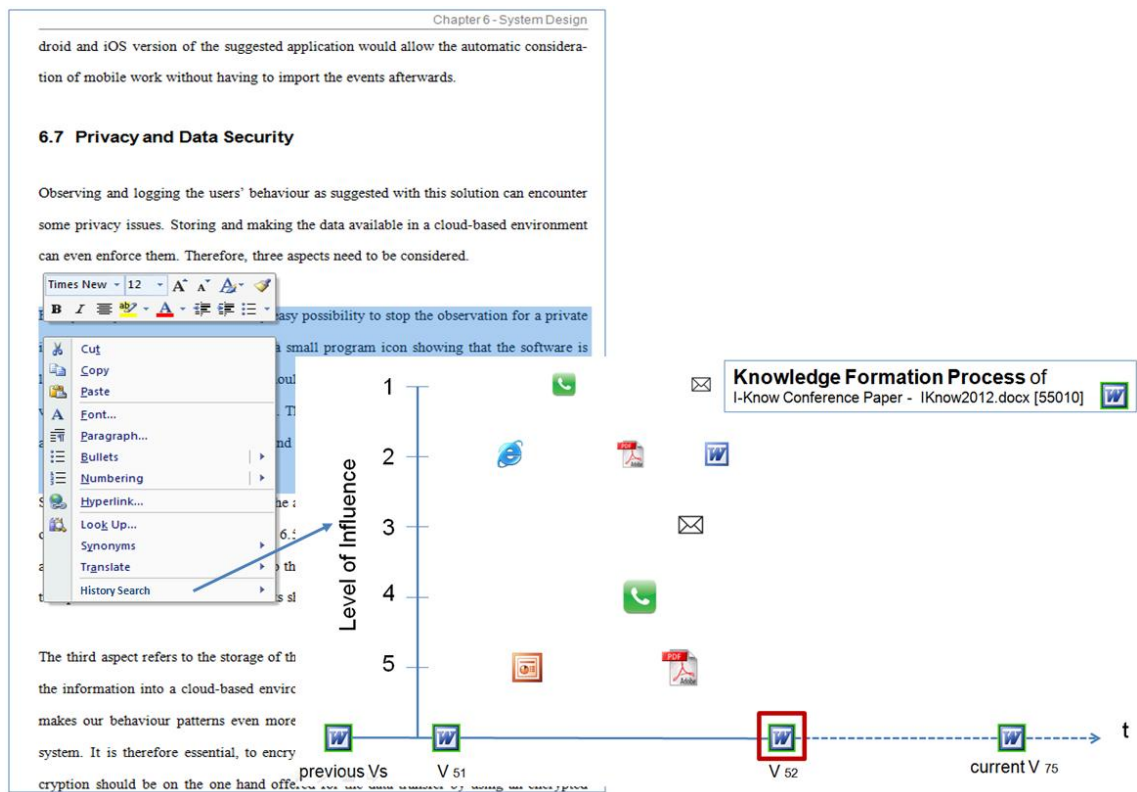
Figure 70 : Context History Search

The system opens the knowledge formation process overview of that document, representing version 52 as the first version in which the marked contents appeared. Furthermore, the previous version 51 is displayed, as well as all the relevant events, which happened in between. The user gets the information when (date, time and version number) the relevant paragraph was inserted into the document, and where the inputs most probably came from – in the example of Figure 70 the marked text was most probably based on some finding in the presented PDF document, the PowerPoint document, or based on the phone call.

Depending on the applied comparison algorithm (as described in subsection 6.4.3), the functionality can also be extended to images, and other included multimedia objects.

6.6 Data Access

To ensure a consistent, and at the same time distributed access to the knowledge base, the approach suggests to operate the database in a web-based cloud environment. This will on the one hand allow the usage of several devices at the same time, and gives them so also the possibility to access the same data basis. Looking especially at the increased amount of mobile internet usage via smart phones or tablet PCs, this data storage approach allows the access to the personal knowledge base from nearly everywhere. To gain a full advantage, not only the database should be stored centrally in a cloud-based service, but also the application should be made available for the different operating systems. Providing a version for Microsoft Windows versions, as well as for Linux OS and Apple iOS will allow connecting the (more or less) stationary workplaces to the cloud, offering a mobile Android and iOS version of the suggested application would allow the automatic consideration of mobile work without having to import the events afterwards.

6.7 Privacy & Data Security

Observing and logging the users' behaviour as suggested with this solution can encounter some privacy issues. Storing and making the data available in a cloud-based environment can even enforce them. Therefore, three aspects need to be considered.

Firstly, the system should have a very easy possibility to stop the observation for a private incognito mode. The idea is to have a small program icon showing that the software is loaded. In the active mode, the icon could be green, by clicking on the icon, the logging would stop and the icon will become for example red. This very simple way would allow the user to interact with the application in a very fast and user friendly way.

Second, if a user forgets to switch off the application, there is a need for an easy clearance of the unintended records. In section 6.5 the possibility to delete individual events has already been introduced. In contrast to this, the functionality here requires the possibility to input a timeframe in which all events should be removed. A confirmation of the removing wish should be offered to minimise user mistakes with potentially high consequences.

The third aspect refers to the storage of the data. As mentioned in the last section, putting the information into a cloud-based environment offers many advantages. Nevertheless, it makes our behaviour patterns even more accessible than it would be in a locally stored system. It is therefore essential, to encrypt the information in an adequate manner. Encryption should be on the one hand offered for the data transfer by using an encrypted connection (for example 128 bit https connection), and should also be applied for the storage of the data. Due to the fact, that local devices can also get lost, stolen or externally accessed, the suggestion would be to encrypt it anyway, if stored locally or in a cloud.

6.8 Student Support with Described Use Cases

In section 5.2 a few exemplary use cases from students' everyday life have been outlined. With the suggested approach, in every of this cases a strong student support might be offered:

1. Referring to the described problem of *finding and accessing of existing information and knowledge* (use case 1 on page 79):

   The system supports the user by storing all information and knowledge in one central place. Furthermore, it extracts relevant descriptive metadata and information, and makes it so available and searchable for the user. Irrespective of the kind of technical file or application behind the information, the suggested system allows to use one central search, based on the content, the descriptive metadata, as well as in respect of the related context.

2. Referring to the described problem of *finding again already identified information and knowledge* (use case 2 on page 80):

   On the one hand, the system supports the user by keeping a history of relevant items even if the user is not manually and explicitly doing it. On the other hand, it brings the information and knowledge items into a context and links them with the relevant other knowledge sources and events, which makes it much easier for the user to find again the needed sources.

3. Referring to the described problem of *content comprehensive linking of information and knowledge* (use case 3 on page 80):

   The system supports the user by not only looking at the current working area, but on all areas of interest. Doing so, even accidently found knowledge and information pieces will be identified, captured, extracted, saved and assigned correctly to the referred knowledge work area. That way, they can be found and identified later within the right context.

4. Referring to the described problem of *sustainable understanding and traceability of previous decisions* (use case 4 on page 81)

   By building a whole knowledge formation process from the point of view of every knowledge item, the user gets the possibility to look back and determine which decisions were met when and by what and /or whom they have been influences. For example, it might be a result of a paper, some notes, a meeting or a phone call. This will not in every case be sufficient to answer every detailed question, but it will help to remember, understand and comprehend the circumstances.

5. Referring to the described problem of *sustainable traceability of knowledge formation* (use case 5 on page 81)

   The system offers the user the possibility to determine when a certain knowledge part has been inserted into our work. For example by marking a paragraph, the student can see when the paragraph has been inserted for the first time. Bringing that in a context and presenting which other events had influenced this, the system will help the students to trace back and to understand what happened. Referring to

a meeting or a phone call, the system might not have for example the content of the discussions. However, the fact that the student will know that there was meeting, with whom it was, where it was, how long it was, which documents have been opened during this meeting etc, will again help to remember, understand and comprehend the circumstances.

6.9 Summary

The research suggests the traceability of our personal knowledge formation to support learning processes. Chapter 5 introduced the theoretical concepts, followed by the system design in this chapter 6. The system consists of three main units and the database. The *Configuration Unit* is responsible for the personal and individual configuration of the system.

The *Knowledge Storage Unit* is build of three modules. The identification of relevant data and information is being handled within the *Listener Module* in accordance to the undertaken configuration. The *Storage Module* extracts and stores the identified information, whereas the *Influence Module* assesses the individual levels of influences of other events on the current event and saves them.

The later usage of the stored data is being handled by the *Knowledge Access Unit*. It allows a sortable and searchable table-based access to the entire personal knowledge base. Every entry can be viewed on the timeline giving access to the intermediate versions and presenting the other events that had an impact on the concrete outcomes. By clicking on

the linked icons, an easy and seamless navigation over the stored contents, metadata, relations, and their own knowledge passes is offered.

Beside the functionality and the graphical user interface design of the three units and their modules, the underlying flexible database structure is being introduced within this chapter. The external possibility to import data into the system in a manual, as well as in an automated way via a XML-based interface has been presented. The chapter continues with a review of the different possibilities to store and run the system in a local or a cloud-based environment. This is followed by a discussion about the identified security and privacy issues. How the system might support students within the previously outlined examples of possible use cases concludes the chapter.

7   Confirmation of the Concept

The last two chapters have been describing the suggested approach to support learning by tracing the personal knowledge formation from both, a theoretical and a practical implementation perspective.

Chapter 7 concentrates on the confirmation of the concept of that new approach. It starts with a description and justification of the chosen confirmation method. Having this in mind, the chapter continues by describing the confirmation scope, and illustrating the applied methodologies. The sample of the confirmation is characterised, followed by a detailed discussion of the findings. The chapter continues by substantiating the finding on a broader basis, and finally summarises the findings and results of the confirmation phase.

7.1 Method of Confirmation

Having suggested a new concept for personal knowledge management system, its effectiveness needs to be determined. From an organisational point of view, Turban and Aronson (Turban and Aronson 2001) provide the following three arguments for measuring the success of a Knowledge Management System:

- To provide a basis for company valuation

- To stimulate management to focus on what is important

- To justify investment in KM activities

However, also from an academic, and a personal point of view, measuring KMS effectiveness is crucial to understand 'how these systems should be built and implemented' (Jennex and Olfman 2004).

From the organisational knowledge management point of view, there exist several KMS success models that could be applied to determine the effectiveness of the suggested approach, like the *KM Value Chain* (Bots and de Bruijn 2002), the *Massey, Massey, Montoya-Weiss, and Driscoll KM Success Model* (Massey et. al 2002), the *Lindsey KM Effectiveness Model* (Lindsey 2002), and the *KMS Success Model* (Jennex and Olfman 2006). Although all models might be applied to both KMS building approaches, the *infrastructure / generic approach* and the *process / task approach* (see section 3.5), the *Massey at al. Model*, as well as the *KMS Success Model* do not have any limitations (Jennex and Olfman 2004). Based on the broadly accepted *IS Success Model* (DeLone and McLean 2002), the *KMS Success Model* has been chosen as the mostly suitable one for this research.

Figure 71 illustrated the KMS Success Model:
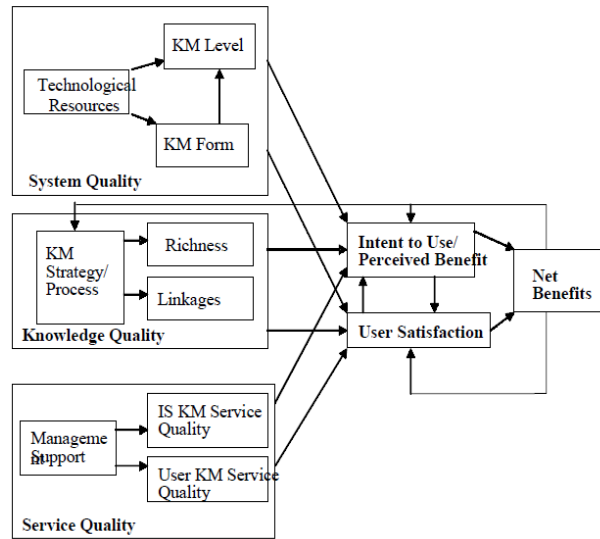
Figure 71 : KMS Success Model (Jennex and Olfman 2006)

The model consists of the five dimensions *System Quality*, *Knowledge Quality*, *User Satisfaction*, *Intent To Use / Perceived Benefit* and *Net Impact* which are briefly described in Table 8:

Table 8 : Dimensions of the KMS Success Model

Dimension	Description
System Quality	How well are knowledge creation, storage, retrieval, transfer, and application performed and how much of the knowledge is represented by the system?
Knowledge Quality	Is the right knowledge with sufficient context captured and made available for the users at the right time?
User Satisfaction	How satisfied are users with the system?
Intent To Use / Perceived Benefit	How is the perception of the benefits of the system for the user?
Net Impact	Which impact has the system on the user's performance?

Although the model, as well as the others, have been intended to assess the effectiveness of organisational knowledge management systems, the principals can also be applied here.

As with most of the currently available Personal Knowledge and Information Management tools, the main benefit of the suggested approach lies also in the medium-term and the long-term usage. On the one hand, the knowledge base needs a certain minimum basis to become useful and helpful to the user. On the other hand, the idea of helping the user remembering previous decisions only makes sense if enough time passed since the decision. Here, the real assets and benefits for the users can be achieved best. To evaluate the idea on a quantitative basis, not only a running prototype would be needed, but also a sample of dedicated, and representative long-term users. Such a long-term end user trial would be the most appropriated evaluation of the suggested system. However, due to the maximum length and the economic constrains of the research project this approach has been classified as not suitable.

To ensure a fundamental and deep confirmation of the presented ideas, a qualitative expert analysis has been chosen therefore. This approach was seen as the most significant opportunity to obtain useful research outcomes.

As Seymour stated, innovation always starts with people, and if we 'forget this even for a moment', we 'run the risk of delivering feature-rich rubbish into already overcrowded lives' (Seymour 2002). To ensure this, as well as a fundamental and deep confirmation of the presented ideas, a qualitative expert analysis has therefore been chosen. This approach has been considered as the most significant opportunity to obtain useful research outcomes.

Based on the qualitative design of a phenomenological study (Leedy & Ormrod 2005), with the main goal to understand the experience, ideas and thoughts, experts in the area of research have been identified. These experts have been interviewed and their opinions and beliefs on the theoretical and the practical aspects of the study analysed.

7.2 Scope of Confirmation

The confirmation process for each individual participant has been divided into two parts. In the first part the methodology and the suggested implementation of the solution has been introduced to the participant. Based on a PowerPoint presentation containing the theory and the prototypical graphical user interfaces as explained and presented in chapters 5 and 6, the functionality and usability of the proposed system have been explained to the participant. Open questions regarding the functionality have been clarified.

In the second part, an interview with the participants has been held. The format has been open and unstructured. Nevertheless, a framework of questions has been designed, to guide the discussion and cover all relevant issues. All interviews have been voice recorded and later on typewritten. They can be found in Appendix B.

The participants have been informed that any information that they provide might be used within the PhD work and possibly within other publications or presentations. They gave their consent to this upon accepting the participation. All consent forms are available in Appendix C.

Moreover, the participants have also had to provide their name, job title, and their experience level / duration within the area of Knowledge Management. The goal here has been to interview real experts in the area of Knowledge Management. Their experiences guarantee an objective and sufficient confirmation of the whole PhD research. All participants, as well as their classification as appropriate experts are introduced in section 7.3.

Confirmation Questions

The expert interviews were of an in-depth, open and mostly unstructured format. This approach has given the needed flexibility to adjust the interview dynamically according to interesting issues. Nevertheless, to ensure the coverage of every interesting and relevant topic, a framework of questions were designed. They focus upon the following four categories:

- Comprehensive historiography of work

- Remembering why we did something

- Reproduction of work

- System design and implementation

Each of the categories will be introduced in the following paragraphs. For each category, a brief statement regarding the objectives (goal) has been worked out:

Comprehensive Historiography of Work

Goal: To find out, if versioning / historisation of work from the personal point of view is still an important (partially) unsolved issue - especially looking at a system compre-

hensive solution (like the suggested independent layer). Where do the problems lie, and what are the reasons for them?

Do you historicise your work?

YES

    How?

    Do you use a software system for that?

    YES

      Which one?

      What is good, what is not good / missing?

      Does it save the versions automatically?

      YES

            Did you ever have problems with it?

            YES

                What kind of problems?

      Did you also try other systems?

      YES

            Which ones?

            Why did you decide to switch?

      NO

            Why not?

    NO

      Why not?

      Have you tried some of the available ones?

      YES

            Why did it not work out for you?

      NO

            Why not?

    Do you have your own file logic system?

    YES

      How does it look like?

      Do you use the date within filenames?

> Do you use the version within filenames?
>
> Do you delete the working versions after the final version is ready / later?
>
> YES
>
> > When usually?
> >
> > Did you ever wish you had not? What happened?
>
> NO
>
> > Why not?
> >
> > Did you ever wish you had?
> >
> > Which type of work are you versioning?
>
> In practise, how easy is it for you to work with your versions / history?
>
> What do you think of the idea to track versions of our knowledge within an independent layer on a comprehensive basis like in the suggested approach?

Remembering why we did Something

Goal: The presented solution suggests not only the versioning of our (intermediate) work results, but also to observe and remember what inputs led to these results. The answers to the following questions should help to find out, how new, important, and helpful this is or might be for our daily knowledge work.

> How important is it to remember why you decided to do things a certain way / what the inputs were?
>
> What do you do to make sure you 'remember'?
>
> Which kind of information should be considered?
>
> Do you use a software system for that?
>
> YES
>
> > Which one?
> >
> > How does it work?
> >
> > What exactly does it save?

What do you do with the 'missing' information?

Is it saved permanently? If not, for how long?

What is good, what is not good / missing?

Beside this, any additional functions of operating systems / browsers, ...?

NO

Why not?

Have you tried one?

YES

Which one?

Why did you decide not to use it?

Do you use a paper notebook (in addition)?

YES

What is your "system"?

What is good, what is not good / missing?

In practise, how easy is it for you to recover why you decided to do things a certain
way / relocate your input sources?

What do you think of the idea to track not only versions of our knowledge, but also
the events which led to them like in the suggested approach?

Reproduction of Work

Goal: Within a fast changing technological environment, the ability to change or ex-
change a certain system is a necessity for an easy and fast adoption of the user. Often
this is blocked by the vendors due to technological or business purposes. Remember-
ing what we did to get the results we have, would allow a (partially) reproduction of
the status quo in a different environment. The answers to the following questions
should help to find out, how innovative and practicable this approach is.

Software vendors often set high entry / exit barriers. Could reproduction be an ade-

quate alternative for exporting your data and information from 'such' systems?

What should be considered here?

Which problems do you see here?

Is the centralised and comprehensive versioning of our knowledge and saving of the relevant inputs on these versions (like in the suggested approach) an applicable way to face these problems?

System Design & Implementation

Goal: To find out, how practicable, useful and sophisticated the suggested implementation approach is.

The presented solution suggests the tool independent and comprehensive saving of versions of our work. While doing this, the different possible influences will be considered and assigned to the saved versions. In general, how do you like the presented implementation approach?

The suggested implementation defines a version of our work according to a point of time in which we either save it (e.g. a document, sending of emails), or receive it (e.g. a phone call, e-mail). Is this reasonable?

Some knowledge workers use to save their work every few minutes, which could lead to a very high number of versions. The idea here is to skip some intermediate versions, if no other events happened in the meantime. What do you think of this approach?

Any other ideas how to handle this?

The implementation suggests saving all versions, and in the standard mode displaying just the latest five ones – is this appropriate or how many versions should be saved?

Why? Should this be individually configurable?

How many should be displayed?

Considering a following version of a knowledge piece, every event which happened

between the last version and the current one will be considered as a possible influence. Is this reasonable?

Considering a first version of a knowledge piece, the system will consider all events which happened in a predefined (configurable) time slot.

Do you see any other possibility to solve this?

On the Y-axis the levels of influences will be displayed. The more values available, the more detailed the classification is. At the same time it makes it more difficult and time-intensive for manual assignment. How many values should the scale regarding the level of influence have?

Why?

Should this be individually configurable at the beginning of the usage (later adjustment might be difficult)?

There are different possibilities to automatically determine the levels of influence.

The implementation suggests allowing the integration of various algorithms by providing adequate interfaces. These methods can be assigned to the different event types and mixed individually. Do you think this approach can ensure a good automatic assignment?

Initially the implementation will offer the Statistical Method of Keyword Analysis? Here the adjusted abstracts of the events will be compared and their relationship determined. Is this an adequate method?

To ensure an easy and fast manual correction possibility of the automatic assignment, an optional pop-up will be shown to the user and the possibility of manual adjustment offered. To make it easier to recognise the several events, some of the descriptive metadata will be displayed and the certain event types will be displayed in preconfigured colours. Is this an adequate method?

Due to the fact, that not every event happens within the desktop system environment, some of the events need to be assigned belatedly (e.g. after the synchronisations with the smart phone). The impact here is that in the meantime, the knowledge formation process is not complete – even without knowing that something is missing. Do you think that this is a problem regarding the fact that the time-slot should not be too long?

Would it make sense to let a part of the system run in an internet based environment, to support a faster and more flexible information exchange?

In this case, would you have any privacy concerning having your data stored partially in the web?

In general, if there is a possibility to deactivate the system at any time, would there be (still) any privacy / security issues concerning you?

YES

Which ones?

What would be the best way to face them?

All knowledge entities will be displayed to the user in the form of a sortable and searchable table. Here the main descriptive meta information will be presented upfront. Which meta information this is, can be configured for the different event types (like word documents, phone calls, ...). How do you like this way of overview?

Would you prefer another access possibility?

YES

Which one?

The graphical presentation of the holistic knowledge formation process intends to give a fast and clear overview. What do you think of this approach?

For clarification the different weightings will also be reflected within different sizes of the icons. What do you think of this approach?

By sliding the mouse over the different icons you get a direct overview of the most important meta information, like the title, or the sender of the mail. This can be individually configured. By clicking on the icon all assigned information will be displayed, and if available the linked documents opened. What do you think of this approach?

Is the suggested system implementation useful?

What in particular?

What is missing?

Is the suggested approach practicable?

What in particular?

What could be better?

In which areas would you use it?

Where will be the largest benefits of such a system?

7.3 The Participants

The following section is devoted to the sampling of the selected participants for the confirmation of the research. The goal is to have experts in the area of Knowledge Management to estimate the value of the suggested approach. The expertise should in total be linked to the following three characteristics:

- Academic Backgrounds

- Organisational Backgrounds

- Personal Backgrounds

The focus should be clearly put on learning processes and learning environments. Based on the general conditions of a phenomenological study (Creswell 1998, Leedy & Ormrod 2005), a total number of seven to ten participants has been selected prior the start of the assessment. During the study, the seven firstly selected participants were already able to discuss and evaluate the theory and system design of the suggested approach in such a manner, that no additional findings could be expected. Therefore, the sample remains at the seven participants. The final selection was based on the candidates' experience and knowledge, but also on availability and willingness to participate. They will be introduced shortly in the following, and a motivation for their selection will be given.

Dr. Ian O'Keeffe

Dr. Ian O'Keeffe, Research Fellow, *Knowledge and Data Engineering Group* (KDEG), *School of Computer Science and Statistics*, *Trinity College Dublin*, *The University of Dublin*, Ian.OKeeffe@scss.tcd.ie, https://www.scss.tcd.ie/ian.okeeffe/ − interviewed 06.09.2012, I-Know 2012 Conference, Graz, Germany.

Dr. O'Keeffe is an expert in the area of Knowledge Management, with a focus on learning methodology and processes. He has contributed to the knowledge in this area with several ACM, IEEE and other high standard publications, like:

- AMASE: A framework for composing adaptive and Personalised Learning Activities on the Web (Proceedings of 11th International Conference on Advances in Web-Based Learning, 2012).

- Personalized Activity Based eLearning (Proceedings of the 12th International Conference on Knowledge Management and Knowledge Technologies, 2012).

- Personalizing Educational Activities (IEEE 12th International Conference on Advanced Learning Technologies (ICALT), 2012).

- Personalised Web Experiences: Seamless Adaptivity across Web Service Composition and Web Content (User Modelling, Adaptation and Personalization, 2009).

- Principles for Designing Activity-based Personalized eLearning (Seventh IEEE International Conference on Advanced Learning Technologies (ICALT), 2007).

- Combining Adaptive Hypermedia Techniques and Ontology Reasoning to Produce Dynamic Personalized News Services (Adaptive Hypermedia and Adaptive Web Based Systems, 2006).

- Just-in-time Generation of Pedagogically Sound, Context Sensitive Personalized Learning Experiences (International Journal on eLearning (IJeL), vol. 5, no. 1, 2006).

He has a broad theoretical background in the areas of Knowledge Management, Personal Knowledge Management, as well as learning processes. At the same time, he has many years of practical experiences in these subjects. His recent postgraduate level studies enhance this and allows him also to adopt the point of view of a learner. This qualifies Dr. Ian O'Keeffe as a comprehensive and critical candidate for the confirmation.

Prof. Dr. Andreas Peter Schmidt

Andreas Peter Schmidt, Professor at the *Faculty of Computer Science and Business Information Systems* at the *University of Applied Sciences in Karlsruhe* (Germany) – HS_Karlsruhe, andreas_peter.schmidt@hs-karlsruhe.de, http://andreas.schmidt.name/ - interviewed 28.03.2013, HS_Karlsruhe, Karlsruhe, Germany.

Prof. Dr. Andreas Peter Schmidt is an international well-known expert in the area of Knowledge and Information Management. With many years of professional experience from an academic, organisational and personal perspective, he can not only reflect the suggested approach against existing solutions, but also project the idea into the academic environment. Prof. Dr. Andreas Peter Schmidt gained his expertise among others in the following positions:

- Researcher at the FZI Research Centre for Information Technologies in Karlsruhe, Germany; engaged in projects like:

- o CoastBase (www.coastbase.org) - a European integration project for building a warehouse for coastal data and information.

  o Learning in Process (www.professional-learning.eu), leading the scientific activities, developed a competency-oriented methodology for supporting work-integrated learning on demand.

  o Im Wissensnetz - (In the Knowledge Web, www.im-wissensnetz.de) – development of the methodology of more informal learning processes in emerging domains and incorporation of social awareness.

- Scientific coordinator of the FP7 IP MATURE (www.mature-ip.eu), which builds upon the knowledge maturing model, building a bridge between e-learning and Knowledge Management. This is now further developed in the Learning Layers project (hwww.learning-layers.eu).

- Department manager within the research division Information Process Engineering at the FZI Research Centre for Information Technologies in Karlsruhe responsible for the competence area Knowledge and Learning.

- Assistant Lecturer at the University of Karlsruhe with a lecture on Information Integration and Web Portals.

Prof. Dr. Cosima Schmauch

Cosima Schmauch, Professor at the *Faculty of Computer Science and Business Information Systems* at the *University of Applied Sciences* in Karlsruhe (Germany) – HS_Karlsruhe. cosima.schmauch@hs-karlsruhe.de, http://www.home.hs-karlsruhe.de /~scco0001/ – interviewed 27.02.2013, HS_Karlsruhe, Karlsruhe, Germany.

Prof. Dr. Cosima Schmauch is the Associate Dean of the *Faculty of Computer Science and Business Information Systems* at the *University of Applied Sciences in Karlsruhe*, Germany. She is a professor for 22 years, and an expert in the area of Knowledge Management, knowledge-based systems, as well as in distributed computing. Here her main focus lies on semantic technologies, rule based systems, fuzzy logic and intelligent software agents. Her professional experience, the direct and close link to her students and their requirements, as well as her strong cooperation with industry partners in the mentioned areas give her a perfect basis for an objective and critical confirmation of the suggested approach.

Monica Morrison

Monica Morrison, Director *Moedi wa Kitso* (Flow of Knowledge), Maun, Botswana, monicamorrison@gmail.com, http://www.linkedin.com/profile/view?authType=NAM &id=39948468 – interviewed 07.09.2012, I-Know 2012 Conference, Graz, Austria.

Monica Morrison is a very experienced knowledge worker. She looks back at more than 30 years of professional experience in the area data and information management. During this time, she had the chance to learn different approaches, for different use-cases, from different perspectives. She concentrated also on the personal perspective of Knowledge Management and is always in search of new methods and technologies to support Information and Knowledge Management. With her practical experience and knowledge, Monica Morrison gives the insights of a real power user, and offers the possibility to question the sensuousness and usability of the suggested approach.

Former Knowledge Management related working experience:

- Communications and Information Specialist, OKACOM (The Secretariat of the Permanent Okavango River Basin Water Commission), Maun, Botswana (www.okacom.org).

- Senior Librarian, University of Botswana HOORC, Botswana.

- Knowledge Management Advisor, focus on the development of a Personal Knowledge Management programme, United Nations Dag Hammarskjold Library (icty.org).

- Knowledge Management Advisor, International Labour Organization Library (INFOR, www.ilo.org).

- Content Management Advisor, UNAIDS Secretariat (www.unaids.org).

- Information Resource Manager, Royal Commission on the Toronto Waterfront, Canada Ministry of Indian and Northern Affairs.

- Librarian / Editor, Canadian Arctic Resources Committee.

- Indexer, Canadian Museum of Civilization Canadian Centre for Folk Culture.

Prof. Dr. Andreas Schmidt

Andreas Schmidt, Senior Research Fellow at the Institute for Applied Computer Sciences of the Karlsruhe Institute of Technology (KIT) and professor at the Faculty of Computer Science and Business Information Systems at the University of Applied Sciences in Karlsruhe (Germany). andreas.schmidt@kit.edu – interviewed 27.02.2013, KIT, Karlsruhe, Germany.

Andreas Schmidt is a professor at the *Department of Computer Science and Business Information Systems* of the *Karlsruhe University of Applied Sciences* (Germany). He is lecturing in the fields of database information systems and model-driven software development. Additionally, he is a senior research fellow in computer science at the *Institute for Applied Computer Science* of the *Karlsruhe Institute of Technology* (KIT). His research focuses on web-based technologies, with emphasis on databases, Knowledge Management systems, column stores and generative programming. In the field of Knowledge Management he developed different systems for Knowledge Management in distributed, intercompany environments. He is one of the main architects of the *Provida* and *MINA-Base* systems, which focus on the representation of knowledge in microsystem technology. He is regularly presenting his work at international conferences. His outstanding technical background within the area of Knowledge Management systems, as well as his long experience with and within learning environments make him an adequate interview partner regarding the technical issues, implementation, as well as differentiation against other technical solutions.

Nikolaos Pappas

Nikolaos Pappas, PhD student related to semantic indexing and retrieval of network media, *Ecole Polytechnique Fédérale de Lausanne* (EPFL), *Idiap Research Institute*, Switzerland, nikolaos.pappas@idiap.ch, http://people.idiap.ch/npappas – interviewed 06.09.2012, I-Know 2012 Conference, Graz, Austria.

Having finished his Dipl.Eng in Information and Communication Systems Engineering, his M.Sc. in Technologies and Management of Information and Communication Systems – Information Management, Nikolaos Pappas is currently doing his PhD in the area of

Knowledge Management, related to the semantic indexing and retrieval of information. His broad theoretical background and his own situation as a recent Master student, and a current PhD student make Nikolaos Pappas an appropriate candidate to look at the suggested approach and its environment from a critical user perspective, who is able to assess and evaluate it against the available methodologies and solutions.

Andi Rexha

Andi Rexha, PhD student, at the Graz University of Technology (TU Graz) and the Know Center, Kompetenzzentrum für wissensbasierte Anwendungen und Systeme Forschungs- und Entwicklungs GmbH, Austria, arexha@know-center.at − interviewed 06.09.2012, I-Know 2012 Conference, Graz, Austria.

Andi Rexha represents the typical targeting group of the suggested approach. He has in the past, and is also currently within his PhD studies still encountering the typical challenges and problems university students are facing every day. Moreover, his area of interests and research, Knowledge Management, as well as his working environment within the *Know Center* in Graz give him the perfect possibility to assess the value of such a solution not only from a user perspective, but to put it also in relation to current and past research that is and was going on in this area.

7.4   Findings of the Confirmation

7.4.1 Keeping a History of our Knowledge Work

Historicising our knowledge work is seen as 'essential' (O'Keeffe 2012, l. 3). For O'Keeffe it is based on a mix of build in versioning, and manual version points (O'Keeffe 2012, l. 7-9) based on dates within filenames (O'Keeffe 2012, l. 15). He definitely keeps all the versions (O'Keeffe 2012, l. 19, 24), not being really sure if he uses them often (O'Keeffe 2012, l. 24), but because he 'lost some information' (O'Keeffe 2012, l. 26) in the past, he keeps the versions (O'Keeffe 2012, l. 26-29).

Prof. Schmidt prefers to use version numbers within the filenames. Regarding his lecture preparation for example he tends to sit down and write one or two points that he would like to change in the next version of the slides – 'new ideas, spelling, problems, more expressive examples or ideas' (Schmidt 2013, l. 5-56) he got from feedback from the students. The next time he gives the lecture he takes a further look at the slides and then at his notes. If he has time, he will do it instantly shortly after he has made the notes. The latest point for the adaption is when he gives this presentation again (Schmidt 2013, l. 61). All in all the current methods work for him, 'but probably it could be a lot better if' (Schmidt 2013, l. 18) he 'would change to a more sophisticated system' (Schmidt 2013, l. 18-19).

Prof. AP Schmidt uses versioning mainly for 'backup, really rarely for some other things' (Schmidt, AP 2013, l. 14). He also applies 'versioning with documents by adding your own initials in the document names' (Schmidt, AP 2013, l. 6-7), or the 'the date prefix at the beginning of the filename' (Schmidt, AP 2013, l. 18). In a collaborative context 'usu-

ally one is the master, and the others are contributing, so then you have the date, then what it is about, then everybody adds his initial to characterise who contributed to what. The other possibility is working with version numbers – version 1, version 2, version 3' (Schmidt, AP 2013, l. 19-22). Beside these standard patterns, some users are applying specific tools for historization. Prof. AP Schmidt for example is using Evernote (Evernote 2013) or Google Docs (Schmidt, AP 2013, l. 28). Within personal use, he embeds his 'notes within Evernote, and works with them within Evernote' (Schmidt, AP 2013, l. 90-91). In collaborative work, he 'can just observe, that a lot of these collaborative parts moved to Google Docs, and only for the very final version move back to Word to make it nice looking. Just to avoid this sending around, and messing up with different versions' (Schmidt, AP 2013, l. 23-26). He and his colleagues create a folder within Google Docs where they 'keep that notes and the material. So you will then have the paper, and then some diagrams, and the other separate items within the same folder' (Schmidt, AP 2013, l. 93-95). Depending on the task (e.g. software development), others (Pappas 2012, l. 5, Schmidt 2013, l. 8) are using specialised automatic versioning system like Subversion (Apache 2013c).

Prof. Schmauch uses typically 'several versions of documents' (Schmauch 2013, l. 7-8). This results mostly from the fact that she is 'not quite sure whether' (Schmauch 2013, l. 10) she wants 'to go back to this version' (Schmauch 2013, l. 10). She also does apply the typical patterns of 'date in the file name' (Schmauch 2013, l. 13). The main reason why she is not using any system for that so far is lack of time 'to find out which systems are available' (Schmauch 2013, l. 97).

Morrison also tries to keep a history of her knowledge work. She tries 'all these tricky thinking ways to be the records management guru' (Morrison 2012, l. 5) and relies 'a lot on the traditional records management model, so a file classification system, and folders, and categorisation of documents' (Morrison 2012, l. 6-7) and is this way 'really document oriented' (Morrison 2012, l. 8).

The findings of section 4.1 have been confirmed here once again. Some users just do not know about the availability of appropriated tools. Prof. Schmidt for example just 'missed the idea' so far (Schmidt 2013, l. 10), but he thinks that it might be 'a good idea' (Schmidt 2013, l. 11). Rexha is not versioning much because of laziness concerning manual work, and he does 'not have a tool' (Rexha 2012, l. 4) for this. In addition, the problem that many 'tools disappear' (Morrison 2012, l. 23) has been stated. 'They come and then they go, they just disappear. There are just not enough people using them' (Morrison 2012, l. 23-24). Even if the tools remain the functionality sometimes disappear. 'Way back, one of the earlier versions of Microsoft Word had versioning build it, and it was fantastic' (Morrison 2012, l. 24-26) but 'it just gone away, so then you had to go back to the old fashion system, if you wanted to track a version, you had to basically keep it as an additional filename' (Morrison 2012, l. 28-40). Currently 'there are tools for everything, but to get tools that you can use consistently, and change and train people to also use them, is difficult' (Morrison 2012, l. 44-46). Once you have found a system, 'you have to try to find out, does this make sense in my context. How does it make sense? And then you see somebody else, trying it that way, and you think good idea, try it out yourself and you change again. Simple things like tagging or putting into notebooks or whatever, so very simple things. But still, you have not found really a stable way of doing it' (Schmidt, AP 2013, l. 241-246).

To find out the need for improvement in this area, the success of the applied methodologies among the interviewees has been discussed. Especially the loss of information and data has been mentioned several times. Rexha stated 'in the history it [comment: loss] happened. [...] And then it is really difficult to return and find [comment: the information sources] back' (Rexha 2012, l. 13-15). In such cases he 'tried to research again the document and try to find which was the webpage' (Rexha 2012, l. 18) he 'was looking at for that work and that terminology' (Rexha 2012, l. 18-19), but 'it is a lot of work' (Rexha 2012, l. 21). For Prof. Schmidt the application of the mentioned methodologies 'works [...], but probably it could be a lot better if' (Schmidt 2013, l. 18) he 'would change to a more sophisticated system (Schmidt 2013, l. 19). He also encountered the loss of information; 'it happened. Rarely, but it happen from time to time' (Schmidt 2013, l. 28). Prof. Schmauch thinks that her methodologies do not work for her, at least 'not in any case' (Schmauch 2013, l. 34). Even from the dedicated point of view of for example software development, the success is just partially given. Pappas is just 'more or less' (Pappas 2012, l. 7) happy with it; if he has a mistake, he can go back, roll back, but he 'cannot track what else happened, what systems were used, if a browser was used, or if there was a phone call' (Pappas 2012, l. 7-9). He added 'all this I do not know. I only know the versions of the certain text. And maybe the code also, but that is all' (Pappas 2012, l. 9-10).

For Morrison this is a 'very sensitive issue' (Morrison 2012, l. 79) that happens a lot (Morrison 2012, l. 80). One reason for this is, that where we 'do have IT departments, there is often no clear definition of responsibility for backups and for preservation of information' (Morrison 2012, l. 80-82). The result is that 'we lose a lot of stuff, and [...] it is a huge, a huge cost (Morrison 2012, l. 85-86). She wished 'there would be more studies

done on return on investment here, and on cost of lost of memory, because it is a hard sell, records management is a hard sell' (Morrison 2012, l. 85-88).

For Prof. AP Schmidt the challenge lies in the integrated usage of different systems and methodologies. For him 'each of it as such is easy, what sometimes is complicated is when you switch from one to another, and try to find out how was it in that tool, so that in Google Docs it looks differently, in Evernote it looks differently, in your own file system you have your own style, then you have different styles depend on to whom you are collaborating with – so the differences in that are more challenging' (Schmidt, AP 2013, l. 46-51). Due to this 'multiple channels' (Schmidt, AP 2013, l. 97) he is 'not really happy – but' (Schmidt, AP 2013, l. 101) does not 'know how to do it better (Schmidt, AP 2013, l. 102).

7.4.2 Remembering Why We Did Things a Certain Way

Besides the historization and versioning of our knowledge, the link between the separate knowledge items has been discussed. In this context the question regarding the importance of remembering why we decided to do things a certain way was raised. Prof. Schmauch thinks that 'this is important and is a good idea to have this information in addition to the' (Schmauch 2013, l. 61-62) actual content. Within her teaching environment 'this is also the topic, which' (Schmauch 2013, l. 57) she discusses with her students, for example when they write their thesis. She asks 'them to write down every little influence, the chapter of a book they had read or any conversation with the supervisors, something like this which had/has influence on their writing or the chapter or something like this' (Schmauch 2013, l. 58-60). To make sure she remembers she does currently 'not do very much' (Schmauch 2013, l. 65). In a second document, she collects 'information on web-

sites and additional information' (Schmauch 2013, l. 66-67). Unfortunately, this approach 'does not work' (Schmauch 2013, l. 73). Especially because she does not 'not have the direct relation between the part of a chapter and the notes, [...] just the collection of the notes; and [...] not [...] the direct link' (Schmauch 2013, l. 73-75).

Rexha agrees (Rexha 2012, l. 25) on the importance. In practice he tries 'to put it in a text file, notes, and try to recall every document, and try to recall which is the base, the line' (Rexha 2012, l. 27-28) he is 'trying to go through. But still it is kind of difficult, because one day you use a different vocabulary, and another do not remember what you were thinking' (Rexha 2012, l. 28-30). Usually he keeps notes, 'like *there was an error there*, or *that was problematic*, but usually like with other notes – you lose them' (Rexha 2012, l. 41-42). 'It is not easy, if you for example after one year want to find the right notes. (Rexha 2012, l. 44-45), so the biggest problem is that he gets lost (Rexha 2012, l. 7).

According to Prof. Schmidt, 'it depends on the work you are doing. It could help you in various situations, from time to time, or very often you decide with your stomach feeling and you have probably additional information' (Schmidt 2013, l. 107-110). Normally he can remember or there is a red line he can follow. Probably if he 'had additional information, which was the base of my decision one or two years ago, it would be easier' (Schmidt 2013, l. 83-85).

From Prof. AP Schmidt's perspective, especially the reasons behind our decision are of high importance (Schmidt, AP 2013, l. 110). 'If we think about paper writing, *why we have it in that order*, *why have we included that*. so it is more the why part, that rarely gets in documents. Because you discuss that, and think *yes, that makes sense*, then you move it around, and then later on you could get the impression, that it could be better the other

way around, and you forget about why you did it that way in the first place, you move it around again, and later discover *ah we discussed that problem before* – and move it back. So that you have to repeat the same decision process again and again, because you never document it. And that happens actually quite often, that you forget about the reasons. Because this is also hard to write down. And you are happy that you found it, and in that moment you do not think that you will need the justification later on, because it is not so much about references that you need, but it is more about *ah ok – that makes sense, because when we do it otherwise that would be the problem*' (Schmidt, AP 2013, l. 110-124). Within his personal work, he tries to 'keep notes in Evernote' (Schmidt, AP 2013, l. 80). 'If it is collaborative, usually we try to keep some minutes document or something like that of the process, where just we also note down action items or things like that. Sometimes the problem is that everything happens under time pressure and then you just quickly discuss, run to the next meeting and forget about it' (Schmidt, AP 2013, l. 80-85). Based on the applied methodologies it is 'hard' (Schmidt, AP 2013, l. 127) to recover why he decided to go that certain way. Typically it is that he and his colleagues 'forget about *why did we do that,* and then' (Schmidt, AP 2013, l. 129) try 'to collaboratively reconstruct the argument. Which takes some time, and is not necessary. It is just like that – collaboratively reconstruction. Sometimes you at the end discover something like *ah, I think we have talked about this before*' (Schmidt, AP 2013, l. 129-133).

O'Keeffe tries to put 'everything into one folder' (O'Keeffe 2012, l. 47) and this way makes sure that the connected items can be found again. In practice, 'it works to an extent; you know you have it somewhere, and you can search again for it. But it does not track the influences for the thing you are producing, the artefact you are developing. You do not have this kind of source history, you only have – at best – your bibliography in that

scientific context, but especially when you are not working in the scientific context, if you are 'just' a technical writer, having those sources would be great. At the moment I would not, if I would write a quick report, I would not write a bibliography, and then I would not be able to track them, I would only be purely working on my own memory, I would not have a set of files related specifically to this concrete document' (O'Keeffe 2012, l. 50-60). That gives him a big selection, and he never knows which was the one he used for that document (O'Keeffe 2012, l. 37-38), so in the end it 'is definitely not' working for him, 'you have a feeling, I know I referenced something, but usually you can never find it again' (O'Keeffe 2012, l. 63-64).

7.4.3 The Suggested Approach

Having discussed the current situation, methodologies, as well as the benefits, challenges and problems the users are having within historization and defragmentation of their knowledge work, the following subsections looks at the general feedback regarding the suggested methodology and system.

Prof. Schmauch thinks that 'it would be a big help' (Schmauch 2013, l. 40). Currently she does not use a lot of support when writing documents, but she knows that she could need this. She stated 'I need this but I do not have it' (Schmauch 2013, l. 41-42). If there was a system, 'I think I would use it' (Schmauch 2013, l. 43). She continues that as a computer scientist she likes 'the idea of not controlling but monitoring actions of users to help them, to support them' (Schmauch 2013, l. 45-46). She teaches 'monitor based systems and the software agent is a part of this area and' (Schmauch 2013, l. 46-47) she likes 'the idea to have software agents who help the users' (Schmauch 2013, l. 49), like in the proposed system (Schmauch 2013, l. 49). From her point of view, conversations and discussions are

a really important part in a decision making process (Schmauch 2013, l. 81-82), therefore she likes the idea of a comprehensive system which does not look only on documents, but also on other event types. From her perspective, the suggested approach is 'a very nice idea' (Schmauch 2013, l. 102).

Morrison feels very 'conscious about that whole issue' (Morrison 2012, l. 67-68) due to the fact, that she is person who 'thought about it a lot, and is constantly struggling to think of ways to that we can do this better' (Morrison 2012, l. 69-70). She thinks that 'this is very useful, because you see here the major different, substantive content of version one and version two is based on this founded source, and this source is mainly responsible for the production of version two, and that tracks the thinking process, the establishing process' (Morrison 2012, l. 207-211). 'It will be even practically if you do it on a daily basis. Because, you work on a paper, or prepare a PowerPoint for a meeting and you do it over the whole week, so if you can see the major influences from each day, that would be fine' (Morrison 2012, l. 219-222). In general she thinks that 'it is very good' (Morrison 2012, l. 211) and 'would help - definitely' (Morrison 2012, l. 118).

Also the other participants like the idea; from Pappas point of view 'it will help a lot' (Pappas 2012, l. 51), for Prof. Schmidt it 'sounds very good' (Schmidt 2013, l. 36) and also Rexha confirms that 'the main idea is really nice' (Rexha 2012, l. 35). O'Keeffe thinks that having the events linked together 'would be really powerful. Because it is always the case, when you write a report or a paper: *I saw it, it was really interesting, but where was it?* It is not always captured within the bibliography, it is sometimes more a thing or a feeling, and you import it, it influences your writing, it is not technically a reference, [...] you lose it' (O'Keeffe 2012, l. 39-44).

Prof. AP Schmidt thinks that the proposed solution 'seems interesting, because the system is looking what you are doing. You do not have to do it actively and write down what happened. The system is doing that for you. After you finish your work or make a new version and it pops up, you remember and you have the possibility to enrich the information with additional text and so on' (Schmidt, AP 2013, l. 136-141)? 'You sometimes forget, you do not take proper notes, and then you search again, and that happens quite often, that you have to search again, because you do not have a full reference, because you did it quickly, and in between (Schmidt, AP 2013, l. 59-62). In his opinion, this 'is even more important than the versioning part, because this really causes many problems, especially if you have many of these content switches, which is a real pain. You tend to forget, sometimes you are also in the situation asking yourself *well why have I decided to do it that way? have we decided it?,* or *is it just by accident that it is here*? – and you try to remember, because you did not keep any minutes on the last online call you had, or whatever – so that happens quit often. Tracing back why things are as they are [...] would be great – because that is something that helps you just in case you trace some aspects back to where they come from – I think that this would be very useful (Schmidt, AP 2013, l. 66-76).

Reproduction of Work within Other Environments

As mentioned before in section 4.2, the transfer of data, information and knowledge from one technological environment to another is very often a difficult task, if possible at all. An alternative to transferring the results is the reproduction of the undertaken work. The application of the suggested system as a basis for such a reproduction has been discussed with and assessed by the interviewed experts.

For O'Keeffe it sounds very reasonable, because 'you have a log, you can replay it essentially' (O'Keeffe 2012, l. 76). And this could work 'not only in a different environment, but even in the same context. It is something that we do enough. If you think back to a Linux command line, you have the facility that lets you record your sessions, and then you get an automated script, which you can run in the future. As you worked through the problem, you were trying to solve the problem; you set the commands, and generated this script. And what you are talking about is that on a much larger scale. [...] Not only the reallocation where you might want to replay it to import it into different application or run it on a different system, but also just for processes which you repeat over, and over again. Yes, so if you have it, it is basically a log of the process, you can reply it in another context, or even use it as a learning tool. There is a process you went through doing your searching and finding your documents and doing these things and so on, if you can replay it, you can see how the influences went, which things were good approaches, and which things were dead-ends and so on' (O'Keeffe 2012, l. 76-91).

Prof. Schmauch agrees (Schmauch 2013, l. 136). She thinks 'that many people want to have a feature like *I take my data and put it in another system to use them there*. [...] If you see that you have all the emails which were important for this paper development and you have them all together in your system and you can take them out' (Schmauch 2013, l. 132-136).

From Morrison's perspective, it 'is desirable' (Morrison 2012, l. 191). For her it is about 'the whole issue of creation, [...] the issue of building stores of information, [...] the staff that you have worked with, and the staff that you think is good, [...] and what value does that have to your activity, is it better to abandon it, or anyway, and start again' (Morrison

2012, l. 191-196). The suggested approach in her opinion would allow that 'you just need to look at your thinking process, that gathering process' (Morrison 2012, l. 197-198), and that would make it 'much simpler' (Morrison 2012, l. 196).

For Prof. AP Schmidt the idea of reproduction, based on the knowledge what happened, could for sure (Schmidt, AP 2013, l. 160) be an adequate alternative for exporting from one and importing into another system. He knows the problems of 'losing some details' (Schmidt, AP 2013, l. 168) when 'moving to another system' (Schmidt, AP 2013, l. 167), but according to the suggested way the user 'will not lose it, it will just be accessible via some other place or way' (Schmidt, AP 2013, l. 168-169). However, it is more about that 'emotional thing, that you could if you need to, if you would like to, but [...] most of the time you will not be using the history, it will always be something for rather rare occasions, but [...] in some cases not having it will cost really a lot of time to reconstruct it' (Schmidt, AP 2013, l. 178-182). Therefore, he 'would guess that this will give you a feeling of assurance, it is there, you do not have to worry about reconstructing it, you do not have to worry about the history (Schmidt, AP 2013, l. 183-185).

7.4.4 System Design

This subsection refers to the interview feedback regarding the system design and implementation suggestions.

Major & Virtual Versions

Resulting from the varying user behaviour, the approach suggests the creation of virtual major version. This makes especially sense, when the user saves his work very often. This

idea was welcomed by the experts. Rexha thinks that it 'would be good to make a major version once a day. Otherwise, you get lost' (Rexha 2012, l. 48- 49). Prof. Schmauch suggests making the creation of versions depending on the changes in the central entity (Schmauch 2013, l. 181), especially, because 'some users do not change anything and then save again' (Schmauch 2013, l. 183-184). Prof. Schmidt absolutely agrees. He suggests even 'an explicit check in or check out of major versions' (Schmidt 2013, l. 201), from his point of view 'the intermediate savings are not interesting' (Schmidt 2013, l. 202). It could even be 'a user triggered action, when such a new version should be saved and not' (Schmidt 2013, l. 207-208). Prof. AP Schmidt adds that 'usually you will observe patterns of high level of activity and then stability' (Schmidt, AP 2013, l. 257-258). 'You are not working continuously like never. You have then your periods of activity and then the stability of somebody else. I think you can find there easily patterns, that are reasonably occurred. There are still some probably pathological cases, where it will not really reflect reality, but I think in most cases it is like that, because you then spend some time on it, to these things and you can consider that as close activity which leads to a new version' (Schmidt, AP 2013, l. 260-267). Therefore, he is confident that the idea of virtual major version will work (Schmidt, AP 2013, l. 257-258). Prof. Schmidt adds that it might be wise to define major version according to 'the stand from opening to the closing of the document' (Schmidt 2013, l. 165-166) and not looking so much on the actual user savings at all.

Relevant Timeframes

Regarding the question, which might be the relevant timeframe to look at the start of the knowledge formation process from the point of view of a certain document, the experts suggested a pragmatic method. 'The second version could be seen as the first base, the

first version. The [real] first version is just a start up, and usually there will be not so many important documents as later on during the work in every other version. Therefore showing the second version as the ground of the document should be reasonable' (Rexha 2012, l. 63-68). Prof. Schmauch agrees on that: 'the user might have developed some ideas before, but I think if he or she decides to write the paper now, and then open or start a document' (Schmauch 2013, l. 223-225) would be the right starting point. So it could even end up in a mix of taking the creation time of the document as the starting point for observation, and the beginning of the editing as the time for the first version (Schmauch 2013, l. 229). Prof. Schmidt's idea went in the same direction saying that when 'I create a document [...] this is my version zero' (Schmidt 2013, l. 231-232). Alternatively, the system might ask the user to manually assign related events to the first version by showing 'what happened before, one day, ten days' (Schmidt 2013, l. 230).

According to Prof. AP Schmidt, the right approach might vary a lot on the individual situation, the event type, as well as on the user behaviour. 'In some cases you start with some concrete documents already, very early, you collect things. And in other cases you just start discussing and then in some later point then you create some artefacts' (Schmidt, AP 2013, l. 291-294). He adds, that 'it also depends on the individuals. Some of them tend to create always a document or something like that to put things in. Others tend to discuss first to get a clear picture and then start' (Schmidt, AP 2013, l. 294-297). For such a system to be successful, it does not require only a type specific relevant implementation, because it is 'not only will be a technical solution, but it will be related to your own personal practice, which has to be modified properly' (Schmidt, AP 2013, l. 215-217). Nevertheless, we are 'constantly changing it. So it is not really a stable practice' (Schmidt, AP 2013, l. 225-226). 'I can just observe, how you write papers has changed classically in the

last year. Just because of the tools that are available, collaboration, it is much easier now to collaborate than it was before. So you are also going for more frequent iterations then. So that changes' (Schmidt, AP 2013, l. 217-221).

Talking about the observation time of the several event types (as defined in section 5.4), the idea to let the user decide has proved reasonable. For Prof. A. Schmidt and Prof. Schmauch it should be linked to the start of an action, like for example the opening of a document (Schmidt 2013, l. 174, Schmauch 2013, l. 149). 'It seems that the opening of a document would be the most interesting point, because here you start something, which you think could help you (Schmidt 2013, l. 190-192). However, Prof. A. Schmidt thinks that it is probably not so important to differentiate here anyway, especially when talking about short term-events like writing an email for example (Schmidt 2013, l. 184-185).

Event Types

The general feedback regarding the structure and logic concerning the different event types and their entities was very positive. Only Prof. Schmidt and Prof. AP Schmidt brought up the point that there should be the possibility to link different event types and their central entities to an overall event. 'The work I am doing here, is writing a document. But in real life the work could be *organise a conference*. So I have a lot of things to do. I have calls, discussions with my colleagues. This concerns all the documents I have. So all the emails I have to write, it is not only concentrated on one – I mean the target is not only on one document. It is a bunch of documents for example or emails or phone calls' (Schmidt 2013, l. 393-399).

Prof. AP Schmidt agrees on that. In his point of view considering a simple document as the point of reference might be in some cases 'a bit too simplistic' (Schmidt, AP 2013, l. 298-299). 'In many cases I have observed that you start with one thing where you collect your thoughts. Then later on start to create the paper and then you find out we need a main figure for it and create a main figure. All of that is related to each other. And the calls are not related to the diagram or the paper, but to all of them. So that you need the notion of collection for it' (Schmidt, AP 2013, l. 300-305).

Influence Weighting

The automated weighting of the influence levels is from a methodological and a techno-logical point of view one of the biggest challenges in the suggested approach. For the text-based event types this is in general seen as possible and well 'understood' (O'Keeffe 2012, l 149.). Nevertheless, the pure comparison of contents is being seen critical. From Pappas perspective 'you cannot rely on keyword matrix techniques, because those key-words are not humanly annotated, and do not have any semantic meaning behind them' (Pappas 2012, l. 59-61). The suggested approach of comparing vectors cleaned of any stop words could not be sufficient enough. Pappas stated, that there is a need to 'use something more sophisticated, like the semantic meaning behind it. So instead of using vector space models, apply semantic space model' (Pappas 2012, l. 63-64). The idea would be to 'project all your documents in a semantic space and compare the similarity there' (Pappas 2012, l. 65-66). Based on human feedback, Pappas suggest also a method of training 'an algorithm, like *learning to rank* – so if a user said that this is more relevant than the other, then you can train the software specifically for that individual person what the ranking for that element is' (Pappas 2012, l. 71-74). This would result in training and learning the software what is important for the user, instead of using an ad hock measure

for similarity (Pappas 2012, l. 74-76). O'Keeffe welcomes the idea of semantic approaches as well, especially for 'trying to find synonyms' (O'Keeffe 2012, l. 150).

Prof. AP Schmidt goes even a step further, and thinks that 'comparing these vectors, [...] will not capture reality adequately. Because in many things, this really important influence is not really connected to something similar, but is just a good idea from something which might not be related at all' (Schmidt, AP 2013, l. 335-338). He thinks that often it is even 'more dissimilar than similar' (Schmidt, AP 2013, l. 339-340). 'The easy ones are with similarity, and there are definitely cases where you can apply this, but then there' (Schmidt, AP 2013, l. 344-346) are other, more complicated events, and in these cases he would 'guess that similarity is not influence' (Schmidt, AP 2013, l. 340). From his point of view, 'influence is really related to how much has actually changed in terms of content not really volume but quality. [...] Because sometimes it is a small change, but it is a big step forward (Schmidt, AP 2013, l. 341-344). However, 'that is hard to capture' (Schmidt, AP 2013, l. 342-343) and at the same time 'hard to come up with an algorithms actually for the influence in these cases' (Schmidt, AP 2013, l. 352). He backs up his argument with the example of figures; assuming that during some work on a paper we have 'concentrated a lot on creating figures to get the whole story' (Schmidt, AP 2013, l. 347-348), 'it is hard to see the similarity, it is somewhere hidden' (Schmidt, AP 2013, l. 348-349).

Morrison added another point to this discussion. Due to the fact, that it might be so obvious what the participants are working on or talking about, sometimes the content related key words might not even be mentioned. For example 'I can also have a meeting with my boss, because I am preparing a PowerPoint for the upcoming meeting, and I might not mention that words presentation, PowerPoint, anything – I just ask him a question that I

will use then to clarify the presentation' (Morrison 2012, l. 278-281). For her the whole automated influence assignment part is a 'tough point' (Morrison 2012, l. 278) and a messy thing [...] trying to deal with' (Morrison 2012, l. 282).

Prof. A. Schmidt liked the idea of content comparison in general; however, he suggested the comparison of the content vectors as described above, but referring not to the whole document, but based only at the changes between the current, and the last version of the document. This way the software would build the delta between the two versions, and 'compare the delta with' (Schmidt 2013, l. 259) the possibly linked other units to see how they influenced this delta. In the above example, this will mean to compare in the first step version 74 with version 73 of the conference paper document. The changes are then being used to create a vector the same way as described above. In a second part, this vector can then be compared with the vectors of the other events. This way we are only referring to the relevant changes asking the question *did the other events have an impact on these changes, and if yes how*.

Prof. Schmauch added the idea of referring to the intensity the other event has been dealt with. In a desktop or smart phone environment, this can be evaluated by checking the duration of time in which the event has been used. It could also be possible to advice the 'user that if you have open a document very long, then the relevance is higher than if you close it very fast' (Schmauch 2013, l. 168-170). Furthermore, it would also be possible 'to monitor which application is in front of the screen, so which application is active' (Schmauch 2013, l. 265-266), in which document the user is currently 'scrolling or inside' (Schmidt 2013, l. 168). Here, taking the event *Reading of a PDF document* as an example, it would not be sufficient to look at the timeframe the document was open, but the

time in which the document was in use, meaning, that the PDF application was active / not hidden on the screen. This approach might lead to an 'adaption of the user's behaviour' (Schmauch 2013, l. 154-155), but this way the user would get the possibility to mechanically adjust the influence levels without having to adjust it manually. This approach would also support the problem mentioned by Prof. A. Schmidt that users might have some documents open for several days (Schmidt 2013, l. 186-188). When we would just take the opening of the document as an indication for the influence of other events, we might end up assigning the levels to only a few, or even the wrong events. Observing the actual real usage of the conference paper, and not only the fact when it was opened, would allow a much more accurate association to the real user's behaviour.

As an alternative and supplement to the partially problematic automated assignment of influence levels, the manual adjustments have been discussed with the experts. The general functionality was seen as 'good' (Pappas 2012, l. 105), and using it would for some users be not a problem; Morrison for example said 'for me no problem. because I do it a lot of time, when I get an email, and I process that action, and I drag it into a folder, and categorise, because I do that, that is what I do, I classify' (Morrison 2012, l. 289-291). Having applied it, she 'will be glad later on, because' (Morrison 2012, l. 294) she finds it again. However, she thinks that 'most people [...] might struggle with that, they are really engaged in the process itself, and they are not really thinking if I do this and I spend one second to do this' (Morrison 2012, l. 291-294). Prof. AP Schmidt adds that 'if you do not have time to keep actually some minutes and some action items, because you are running from one meeting to the next, this will not help either. Before you do that, you would write down the action items' (Schmidt, AP 2013, l. 375-378). In Morrison's opinion it would be about proving 'the usefulness of a system' (Morrison 2012, l. 295), for example

with a 'good demo, and if you can solve just one of their problems, just one, and they can see how it helps, how it works, than it will work' (Morrison 2012, l. 297-298). Here it might be extremely important 'to find then the right moment' (Schmidt, AP 2013, l. 379) when to display the adjustment popup window. 'Because otherwise you just click it away, or the extreme case, you uninstall it, because it pops up too frequently' (Schmidt, AP 2013, l. 380-381). O'Keeffe also thinks that it should be implemented as a 'background process, that is automatically doing it, because if it is manual – people are lazy, so they will not do it' (O'Keeffe 2012, l. 106). Pappas agrees on the requirement to run in the background and not being too annoying (Pappas 2012, l. 35).

In respect of the scale of the influence levels the suggested up to seven distinguishing possibilities have been assessed as too high. Prof. AP Schmidt thinks that 'you cannot really distinguish, so that is a bit related to these fluid discussions that if seven levels make sense or not, which we had quite long in one of the projects where we came to the conclusion, probably for statistical reason it is a good thing to have seven, but for the respondent it doesn't make sense, which is a kind of paradox result of discussion. So I would guess a maximum of 5 would actually make sense. And I could also imagine that it is just 3' (Schmidt, AP 2013, l. 321-328). Knowing 'that 7 is a number which is stated if you talk about what people can differentiate in their short term memory' (Schmauch 2013, l. 247-249), it at the same time offers you 'a lot of possibilities to differentiate, but' (Schmauch 2013, l. 246) she is 'not quite sure, whether you need this' (Schmauch 2013, l. 247). Prof. Schmidt agrees; for him 'it should be easy' (Schmidt 2013, l. 268), so it would be better to 'have only three or five' (Schmidt 2013, l. 268) possibilities. 'If you have seven or eight...if it is more rough to categorise them' (Schmidt 2013, l. 268-269).

Visualisation

Prof. AP Schmidt likes the visualisation of the stored knowledge items. In his point of view that is the way we will 'sometimes [...] do it [...]. You are looking for the version of that diagram, you are looking for a certain call, when was it, what was it related to. If you can then from there go further to something that is related to that. Because sometimes your memory works like that. Ah it was for that meeting and maybe it is not the meeting as such, but it is a resource connected with the meeting. Or a change that you have made at that meeting or whatever. So I think, that going according to those categories would be one way, the other would be the timeline definitely and then this chaining. So from there you go to something related. Yes, I think these views make sense' (Schmidt, AP 2013, l. 397-406) and 'might help to reconstruct – definitely' (Schmidt, AP 2013, l. 139).

Prof. Schmauch agrees; she likes 'this graph figure with these icons' (Schmauch 2013, l. 299), and for Prof. A. Schmidt it is 'a nice graphical visualisation, when you also see the relevant other events' (Schmidt 2013, l. 365-366). For Morrison it is very easy to understand, she stated that it took her 'just three seconds to understand what' (Morrison 2012, l. 332-333) is happening there, nevertheless, 'as a visualisation – a little empty, not very rich on information' (Morrison 2012, l. 331), and 'could be prettier, it could be more compact' (Morrison 2012, l. 327). She would like to have some additional information, like 'some indications of date, time' (Morrison 2012, l. 323). As she stated, 'that would be interesting for' (Morrison 2012, l. 3243 her 'to be able to see, that this here took [...] three weeks to do, but over the course of three weeks' she 'discovered that' (Morrison 2012, l. 324).

Referring to the amount of versions presented on the timeline, Prof. A.P. Schmidt thinks that 'usually it is the recent things that you want to look at, it is not the early ones. However, in some cases you might want to go back' (Schmidt, AP 2013, l. 276-277). Beside the possibility to display a fixed number of versions, it might also make sense to display 'rather about a period of time' (Schmidt, AP 2013, l. 279-280). If needed, the user can then still 'explicitly go back' (Schmidt, AP 2013, l. 283). Prof. Schmauch agrees and thinks 'that the latest events will be much more important than the earlier events' (Schmauch 2013, l. 198-199). So for example 'just show the latest five version' (Schmauch 2013, l. 199), and still apply the major version functionality.

Beside the graphical presentation form, the suggested table form has been broadly accepted. For Rexha 'an aggregation of the different events' (Rexha 2012, l. 88) is definitely important, and the suggested approach allows him to 'find even a single element' (Rexha 2012, l. 96). For him 'this way it is perfect' (Rexha 2012, l. 96-97). Prof. A. Schmidt would in addition like to have a table-based presentation not from the perspective of the central entity of an event, but from a chronological point of view (Schmidt 2013, l. 363-364).

Manual Imports & API

Due to the fact, that not every event happens within an environment in which the system can capture and consider it, the belayed import takes place via the manual interface, as well as via the described electronic API interface might lead to a time difference concerning the entire knowledge formation process.

Prof. Schmauch does not see any problems regarding the delayed integration of manual imported events. 'The user is responsible for importing the information, so if he or she does not have had imported information and looks at the last version or looks at the events and does not see events, which are not imported, it is his or her responsibility' (Schmauch 2013, l. 281-282). O'Keeffe agrees and sees 'the power of this being used in the medium to long term; it is not about immediately what you are doing, so the fact that there is an intermediate inconsistency between these two the user might not even notice' (O'Keeffe 2012, l. 161-164). As long as it is possible to give the manually inserted event a 'time-stamp in the past' (Schmidt 2013, l. 282), it is fine, and the suggested solution would definitely be nice (Rexha 2012, l. 72).

When discussing the area of via API included information, like for example calendar information, the question occurs, if such information, without having the real content, is helpful at all. Looking at an appointment, my calendar might give me the information that there was a meeting, when it was, with whom it was, and where it was – but it will most probably not be able to give me the discussed contents of our meeting. The assumption of the suggested approach is that this still is worth a lot, and helps to remember. This was also confirmed by the participant, mainly 'because it is context' (Morrison 2012, l. 251). 'Most of our knowledge is contextualise in events, so [...] just the fact that you are re-minded that you spoke to a certain person, will probably jurg our memory to remember what you have spoke about in that context of this certain piece of work' (Morrison 2012, l. 256-259). Prof. Schmidt agrees, 'at least this meta information helps you to remember' (Schmidt 2013, l. 130).

Privacy & Data Security

As O'Keeffe stated, the suggested system might 'have the problem, that people are feeling this kind of big brother thing of the system tracking them, but if there is a real benefit to them, they will quickly get used to it and if they can see an concrete outcome that is meaningful to them and helps them (O'Keeffe 2012, l. 107-110). Regarding privacy and data security two main topics have been an issue; on the one hand the possibility to switch on and off the logging, as well as deleting unwillingly saved events, on the other hand the storage of the information in a cloud-based environment.

The possibility to stop the system very easily is an essential requirement on the solution. The suggested approach of a simple start / stop button has been widely accepted. In Rexha's opinion, that is the 'best solution and enough' (Rexha 2012, l. 79). 'Usually users are lazy, and putting too many options makes such a system impracticable. Therefore, a simple start / stop button would be great'(Rexha 2012, l. 79-81). O'Keeffe agrees and sees the definite need for 'a flag to switch in on and off' (O'Keeffe 2012, l. 126). Especially if you want to do something 'personnel, I do not want to be observed' added Prof. Schmidt who sees it as an 'important issue' (Schmidt 2013, l. 302). Morrison agrees, although she thinks that 'there will be times' she 'will forget to switch on and times' she will forget 'to switch off. Because that is how it is' (Morrison 2012, l. 303-304).

Once the user forgets to switch off the system, there is a need for a 'possibility to assess the history and manually delete things' (O'Keeffe 2012, l. 127). Morrison thinks that 'you have to give an individual the right to do that, [...] to give people that functionality' (Morrison 2012, l. 310-311) to delete entries. Beside the privacy aspect, this functionality might be used to mark entries as 'spam for later classification of the relevant things that

will truly happen' (Schmidt 2013, l. 303). It is 'similar to Amazons service where you can go in and delete some things you have been searching for, do not use them again, they are not longer relevant' (O'Keeffe 2012, l. 127-130).

An interesting aspect was brought in by Morrison. Such a system could also be used for 'soft performance measurement' (Morrison 2012, l. 156), and here, she 'would be afraid to see how inefficient' she is, looking at self-reflecting questions like *how many times did I visited that website?* (Morrison 2012, l. 156-158). Beside the point that in the short-term it could 'really tiny your self esteem' (Morrison 2012, l. 163), it could in the long-term also improve her way of working (Morrison 2012, l. 165).

In respect of the storage of the information within a cloud-based environment, Prof. AP Schmidt is 'a bit relaxed' (Schmidt, AP 2013, l. 389). In his opinion there are already 'too many things store in clouds or any, to be worrying about it' (Schmidt, AP 2013, l. 390). Prof. Schmauch agrees on this, and thinks that the information can be encrypted, the transport can also be encrypted, and in the end 'there is a lot of data in the cloud and also data which shouldn't be seen by other people' (Schmauch 2013, l. 290 - 292). For O'Keeffe it is more about the possibility to have it under your own control. 'As long as there is trust in the system that it is not pushing data to some anonymous service some-where, and your data is being shared or used in some other way, then your data is totally under your own control' (O'Keeffe 2012, l. 123-125). In this case, it does not matter 'if it is in the cloud, it is your space, even if you use services like Google for hosting, it is still your space. It will be fine' (O'Keeffe 2012, l.136-138). Important is to have it under con-trol meaning that it is not part of some bigger service. As long as it is encrypted, Prof. A. Schmidt does not 'see any problem' (Schmidt 2013, l. 293). The fact of working 'with

three or four computer' makes it 'important, that the information is linked' (Schmidt 2013, l. 293-294) and distributed. Even the participants with more scepticism concerning clouds do see the advantages and would be happy with a cloud solution if it was 'an option, as long as it is by the user's decision initially, that is fine, but not forcing it to have it on the cloud in order to use it. Maybe the user could decide if he wants to have it stored locally, or in the cloud or both' (Pappas 2012, l. 98-99).

However, if saved locally or within a cloud-based environment, due to the fact, that 'people are moving a lot these days' (Morrison 2012, l. 362) they 'would like of course to carry this with them' (Morrison 2012, l. 364) – so it needs to be portable.

7.4.5 Collaboration

Collaboration with others has been a constant and important topic during the confirmation phase of the research. Not only are two of the seven Personal Knowledge Management skills directly linked to collaboration with others (Millikin University 2009), but also the interviewed experts mentioned collaboration several times. For Professor AP Schmidt the success of the suggested system was even dependent upon the 'collaborative situation' (Schmidt, AP 2013, l. 56). Only if this is being supported 'well, then this (the system) will be definitely helpful' (Schmidt, AP 2013, l. 56-57). The main reasons for that are that most of his papers are collaborative. Morrison agrees on the personal view of the collaborative contents. 'As soon as you enter an institution or an organisation where you work with other people, capturing what you do is really important for the whole' (Morrison 2012, l. 92-94). Prof. AP Schmidt continues that 'because we collaborate more easily, we do it more often, so that any solution that we find will just mean that you collaborate more

which will just increase the problems that you having – just in different areas (Schmidt, AP 2013, l. 102-106).

7.4.6 Usefulness of the Suggested Approach

The experts agreed on the usefulness of the idea and the suggested implementation approach. For Morrison it is 'absolutely' (Morrison 2012, l. 340) useful. Especially 'for people working in offices, [...] and academics as well, because the tracking of development of their ideas, this is reflected in documents' (Morrison 2012, l. 340-342). 'For those two kinds of people' she thinks 'it is fantastic' (Morrison 2012, l. 343). Furthermore, she could image that 'you could use it anywhere' (Morrison 2012, l. 344).

For O'Keeffe the solution is definitely useful (O'Keeffe 2012, l. 167), and he 'would use it for work mostly' (O'Keeffe 2012, l. 195). 'Being able to assign information to the sources, being able to look at a document and see the phone call' he 'had with some person, the documents' he 'read, and link those to the changes that' he 'made in the documents, whatever it was' he 'was doing [...] would just be extremely powerful' (O'Keeffe 2012, l. 195-199). It is like 'linking everything together, and having the understanding of where the information came from' (O'Keeffe 2012, l. 199). Especially in writing a similar document, people tend to go back to the original, and kind of repurpose bits of it. Bring in all the source material together would be really helpful (O'Keeffe 2012, l. 201-204).

Looking at it from a more technical point of view, and referring to the usability and the possibility to collaboration, Prof. AP Schmidt states that 'if you can solve the technical difficulties, yes, it could be definitely useful' (Schmidt, AP 2013, l. 409-410). For him 'paper writing, proposal writing would be definitely an area where this would make sense.

[...] Maybe lecture preparations. Maybe also thesis supervision. That's [...] a different use case, which has different characteristics of interactions. [...] Going back, what were my comments to the last version, what were the responses there? [...] Which I now keep manually, but if it would be automatically, it would be great (Schmidt, AP 2013, l. 422-429)'. From the point of view of his students he could imagine that it might be applied anywhere 'where they have to prepare some documents', like writing thesis, or project work (Schmidt, AP 2013, l. 438). He sees the 'traceability of how things came about actually' (Schmidt, AP 2013, l. 442) as the largest benefit of the approach. 'It is not the versioning itself, [...] but making traceable how the changes came about, help you to remember' (Schmidt, AP 2013, l. 442-444). He 'would definitely try it out'' (Schmidt, AP 2013, l. 422).

To the question of such a system is useful, Prof. Schmauch answered 'I would buy it' (Schmauch 2013, l. 321). The largest benefit from her point of view 'would be the information' (Schmauch 2013, l. 332). Having a look 'on a text document or a paper' (Schmauch 2013, l. 332-333) at a later point in time, wanting to rewrite it, 'then it would be very helpful to see all the additional information which was relevant when I produced the paper. It is something like a memory. I forget a lot of things and a lot of paper I had read in a different environment. So if I see, there is a paper I had looked at this time or for this purpose, I could use it for another paper too' (Schmauch 2013, l. 334-338).

For Rexha the solution is definitely useful, and 'looks really nice' (Rexha 2012, l. 110). It would be a practicable solution that 'traces everything what' he is 'doing and has some filters on it' (Rexha 2012, l. 112), which 'would be perfect' (Rexha 2012, l. 114). He would use it mainly at work. 'Here the largest benefit would be to arrange the work, to

arrange what' (Rexha 2012, l. 115-116) he is 'doing, saving time, of course. Finding documents, which have already been used before better' (Rexha 2012, l. 116-117). Pappas also would like to use it, 'mostly in a private modus in work' (Pappas 2012, l. 111).

Furthermore, the participants see further application possibilities. Having 'some analytics available, so that at the end of the year you can look how much time did you spend on a project – because that again would be very useful if you are trying to evaluate your work and trying to decide how to enhance it in future' (Morrison 2012, l. 349-352). 'For example, you can say that if you are only using websites, so maybe consider reading a little more documents or collaborating with others' (Pappas 2012, l. 29-31). Inserting more intelligence, by allowing not only to observe which applications have been used, but also how they have been used, would make the solution a learning log (O'Keeffe 2012, l. 190). It would allow recording 'your experience, and your thought process as well' (O'Keeffe 2012, l. 190-191).

7.5 Substantiation of the Confirmation

The confirmation of the concept has been achieved with the help of the experts as presented in the last sections. However, to ensure, that the main target group of the whole research, university level students, really benefit from the outcomes of the idea, the main thoughts have been substantiated via a student survey among students of the Plymouth University (UK).

7.5.1 Scope of Substantiation

As the first parts of the student survey (described in section 4.4) also this part has been conducted via the online platform SurveyMonkey (SurveyMonkey 2014) and has been distributed via a email with the corresponding web link to 308 undergraduate and graduate students (based on availability) at the Plymouth University.

Substantiation Questions

Beside the question referring to the concept of the system, statistical information (as described in subsection 4.4.1) has been surveyed. Before the questions, some short explanations of the methodological concepts of the proposed approach have been given. The full survey can be found in Appendix D.

System Concept (section 4)

Goal: To allow a more detailed interpretation of the provided answers, the first section refers to statistical and demographic data.

1) At any point in time the system would offer you the possibility to go back to previous versions or statuses of your work. Would this be useful for you?

[ ] Strongly disagree

[ ] Disagree

[ ] Undecided

[ ] Agree

[ ] Strongly Agree

2) Imagine that this way you would also at any point in time be able to find again already used resources independently of their kind or format. Would this be useful for you?

[ ] Strongly disagree

[ ] Disagree

[ ] Undecided

[ ] Agree

[ ] Strongly Agree

3) At the same time, the system would store everything in one central place, irrespective of the environment in which the actual action happened. Would this be useful for you - also in respect of changing the actual system or tool you would like to use for a task?

[ ] Strongly disagree

[ ] Disagree

[ ] Undecided

[ ] Agree

[ ] Strongly Agree

4) The system would automatically link your different actions to each other and determine which dependencies and influences they had on each other. For example, a relevant website or a journal article could be linked to your assignment. This way it might help you to remember why you (your lecturer, your study group) decided to do things a certain way. Would this be useful for you?

[ ] Strongly disagree

[ ] Disagree

[ ] Undecided

[ ] Agree

[ ] Strongly Agree

5) In general, would such a system support you with your studies?

[ ] Strongly disagree

[ ] Disagree

[ ] Undecided

[ ] Agree

[ ] Strongly Agree

7.5.2 Findings of the Substantiation

As described in section 4.4, a total of 308 students were contacted of which 58 started the survey and 56 completed it. The full results can be found in Appendix E. Eight of the participants were female, 48 male. 53 were currently enrolled at the *Faculty of Science and Environment*, whereas three were students of the *Business Faculty*. The majority of the participants were bachelor level students with some college experience but no finished degree (Figure 26).

As presented in Figure 72, the usefulness of the proposed aspects of versioning and historization has been agreed by 88% of the participants (from which 36% strongly agreed). 10 more percent were undecided, and only 2% though that it might not be helpful.
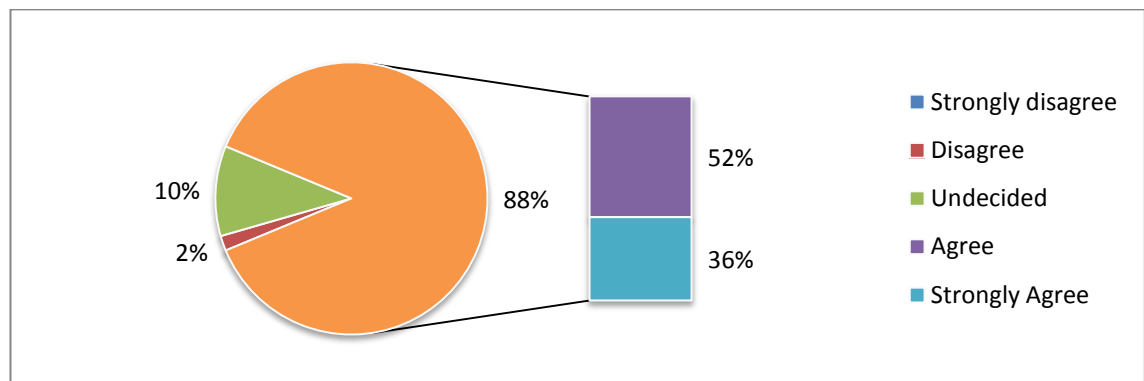
Figure 72 : Student Survey – Usefulness of Presented Historization

The possibility to navigate through the history and browse through the different events and their associated entities (like for example documents), and with this the possibility to find again already used resources independently of their kind or format has been seen as useful by 95% of the participants (selecting 'agree' (54%) or 'strongly agree' (41%) categories as presented in Figure 73).
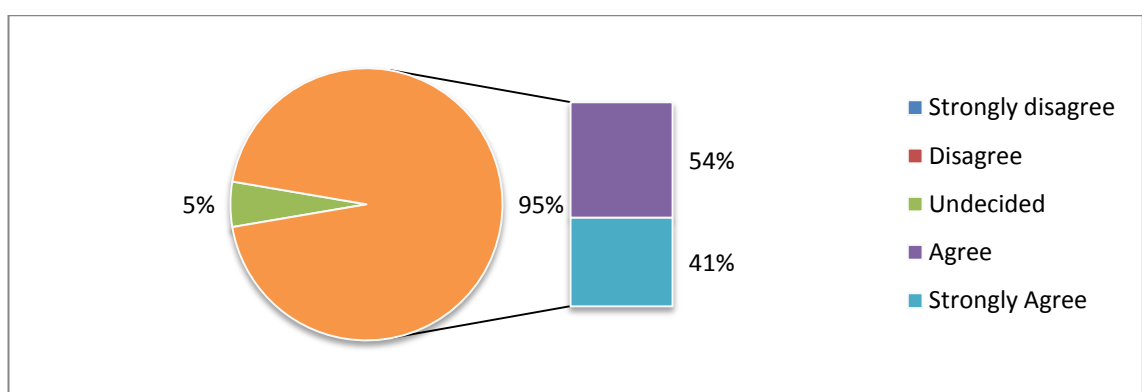
Figure 73 : Student Survey – Usefulness of Finding Again

Although similar features might be available in other systems, there still seems to be a strong need for a centralised storage of our personal knowledge work. 84% are convinced that the proposed centralised architecture, that is supposed to store everything in once central place irrespective of the actual environment in which the actual action happened, would be useful (selecting 'agree' (55%) or 'strongly agree' (29%) categories as presented in Figure 74).
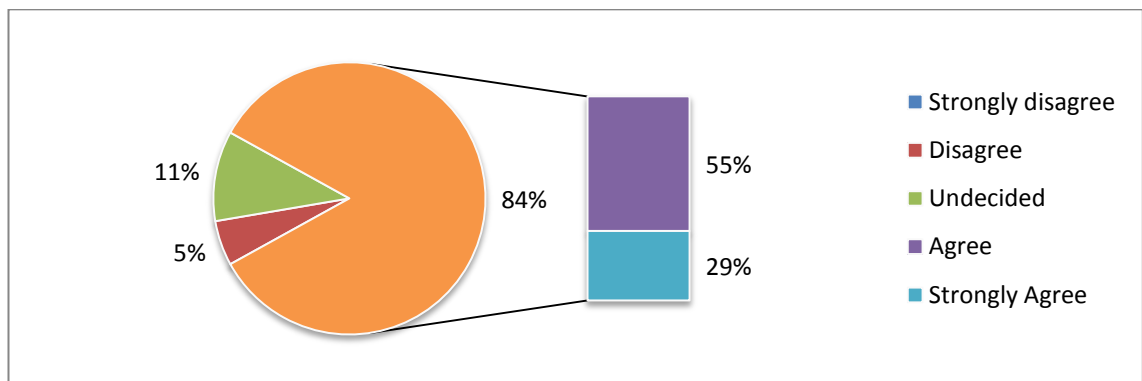
Figure 74 : Student Survey – Usefulness of Central Architecture

The usefulness of linking the different knowledge work events to each other and determining the dependencies and influences they had on each other has also been confirmed by over 80% of the students (selecting 'agree' (59%) or 'strongly agree' (23%) as presented in Figure 75).
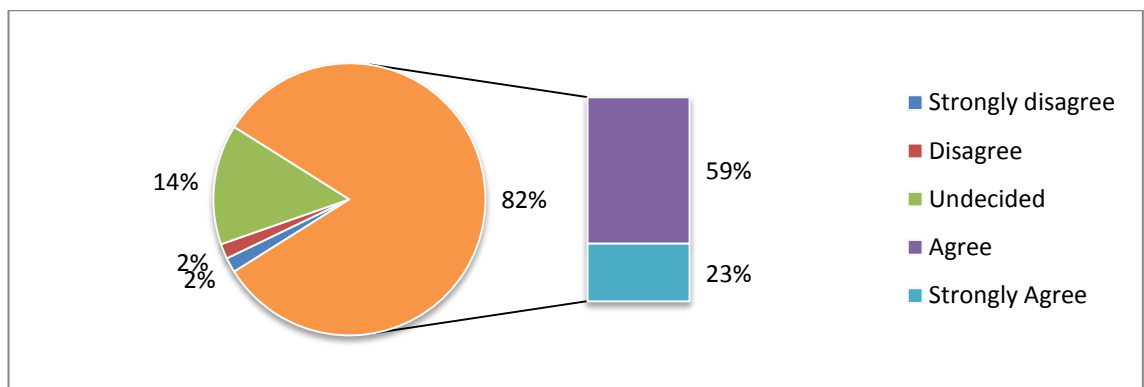
Figure 75 : Student Survey – Usefulness of Event Linking

Questioned if a system as the proposed one would support the students with their studies, has been agreed by 86% of the participants (selecting 'agree' (48%) or 'strongly agree' (38%) categories as presented in Figure 76).
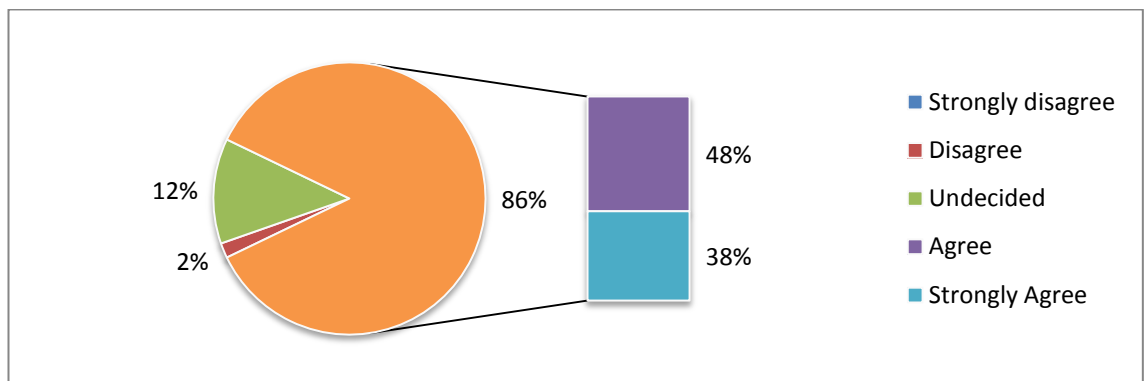
Figure 76 : Student Survey – Usefulness of Proposed System

7.6 Enhancements of the Suggested Approach

As described in the last section, the confirmation phase brought up several requirements to enhance the suggested approach. The following subsections discuss these requirements and if possible give some ideas how to meet them.

7.6.1 Major & Virtual Versions

The worked out characteristics of major versions as described in the subsubsection 6.4.2.1, and of the virtual versions as described in subsubsection 6.4.2.2, have been welcomed and assessed as well qualified. Prof. AP Schmidt brought up the idea of enhancing the assessment criteria by not only looking at a passed time-frame and the number of related other events, but also to take the activity level into consideration. He stated that 'usually you will observe patterns of high level of activity and then stability' (Schmidt, AP 2013, l. 257-258). Information like how often the keyboard or the mouse have been used by the user while the application was active, or how long the corresponding application was active, could for example be used to determine the activity level. If a certain activity level is exceeded, a virtual version might be created independently of the time passed since the last saving as well as the number of other related events.

7.6.2 Relevant Timeframes

In subsection 6.4.3 the question which timeframe to consider when dealing with the very first versions of our knowledge work was discussed. Beside the offered possibility of a configurable timeframe, the participants of the confirmation phase suggested a pragmatic approach: the first version of the corresponding document is version zero, and the system

should not consider anything that happens before this. As soon as the users create an empty document first, before starting to do any other related work, this approach would work fine. For those users who tend to collect their inputs first, and once they have won enough insights, start to include them in their documents, this would require a change of their user patterns and behaviours. However, as Prof. AP Schmidt stated, users are 'constantly changing' (Schmidt, AP 2013, l. 225).

7.6.3 Event Types

To implement the suggested idea of linking different events together would be possible. It would be possible to make collections to which several units might be linked. It would then still be possible to observe the single events and their entities, but at the same time, it would be possible to aggregate them to one central entity, which would be 'a mixture of several items' (Schmidt 2013, l. 409). Probably, 'it would not change [...] anything. [...] But to reduce it to one document, would be a limitation' (Schmidt 2013, l. 410-411). Within the database design, it would mean the enhancement of the UNITS table (Table 9) with an additional column *collection* that refers to the collections that are saved as entries within the new table COLLECTIONS.

Table 9 : Enhanced Table UNITS

unit _id	ver- sion	type _id	FK_type	event _timestamp	entry _timestamp	physical _content	colle- ction
1	1	1	Text Doc	2013-01-18- 17-24-16	2013-01-18- 17-24-16	1.DOCX	47
2	1	4	Phone Call	2013-01-18- 18-12-16	2013-01-18- 18-12-16		47
3	1	6	Website	2013-01-19- 12-13-17	2013-01-19- 12-13-17	3.JPG	
1	2	3	Text Doc	2013-01-19- 12-13-21	2013-01-19- 12-13-21	1.DOCX	47
4	1	5	Application	2013-01-19- 14-12-23	2013-01-19- 14-12-23		

Table 10 represents an example of the *COLLECTIONS* table. The first column *id* repre-

sents the id the entries in the *UNITS* table are referring to, the second column *collection*

represents the collection name which stands behind the id. This value can be used within

the graphical user interfaces whenever needed. From the representation point of view, the

corresponding icons could be mixed together and presented as a collage, whereas the ac-

tual relevant event is highlighted. On the timeline, the referring events of all the separate

main events could then be presented.

Table 10 : Table COLLECTIONS

id	collection
45	course A
46	course B
47	PhD Work
48	project X

7.6.4 Influence Weighting

Due to simplicity and usability reasons of the whole application, the sophisticated implementation of the automated weighting algorithms is one of the most essential and critical parts of the suggested approach. At the same time, the largest concerns have been seen here. The suggested compression of the changed context only instead of the whole document can be easily implemented. The main idea is the same, as while comparing the whole document; it just needs to be done in several steps. Referring to the example in Figure 68, the following steps need to be undertaken at the time t_{x+1}:

1) Creation of the content vector of the new version V_{x+1}

2) Removing of the V_x contents from the V_{x+1} content vector

3) Comparing the cleared content vector with the other chronological relevant events

Figure 77 : Comparison of Changed Content

The second mentioned idea was to include the timeframe duration in which the related document was active on the screen. The implementation of this approach seams also possible within the context of document-based events. Such functionality has been partially already implemented within semantic desktops and could be applied here as well.

However, the other identified topics, especially regarding semantics as well as referring to non document-based events, require a further research and should be handled within a successive research project.

7.6.5 Visualisation

The visualisation of the knowledge formation path was in general seen very positively. However, the following requirements were formulated by the interviewees:

- display of maximum 3 to 5 influence levels
- display a maximum of the last 5 versions of the document with the possibility to go further back in the history (if needed)
- enhance the versions with a creation timestamp

Figure 78 represents the adjusted graphical visualisation of the knowledge formation process. Beside the reduction of the weighting scale to three values, the reduction of the displayed versions to four, also the timestamps of the versions have been included.
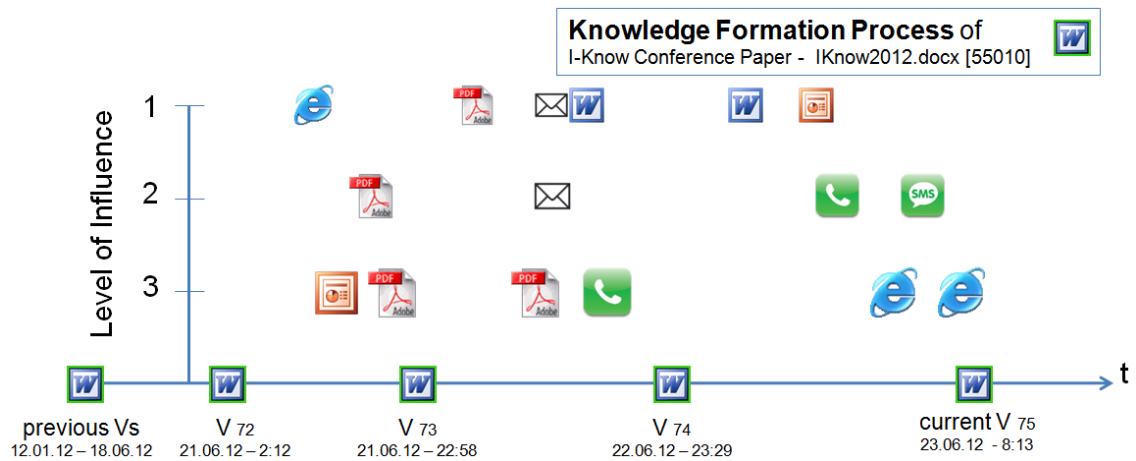
Figure 78 : Holistic Knowledge Formation Process with Dates

The adjusted visualisation offers an easier and faster overview of the relevant information, and allows at the same time a better chronological classification.

7.6.6 Collaboration

Collaboration with others has been seen as an essential and necessary requirement for the suggested system. The technical questions regarding this topic are outlined below, however the challenge here lies on the ethical questions regarding user privacy, data security, as well as the possible knowledge of the users' behaviour patterns. Assuming that the system will run in a central, cloud-based environment, and that the privacy and security issues are really clear to any user, the technological implementation would not be a huge problem.

To extend the suggested functionality to a multi-user system, the entity relationship model of the database design (Appendix A) needs to be adjusted, for example according to the one presented in Figure 79. The main difference is the existence of the four tables *GROUPS*, *USER_GROUPS*, *USERS*, and *UNITS_GROUPS* on the right side of the dia-

gram. The idea behind it is that every user has its unique user id. His entries are always assigned to this unique *user-id*. On the other hand, every user can create own *user groups*. Once a *user group* is created, the user can assign individual users to this group. Based on a timeframe, the available user groups might get access to the individual knowledge units.
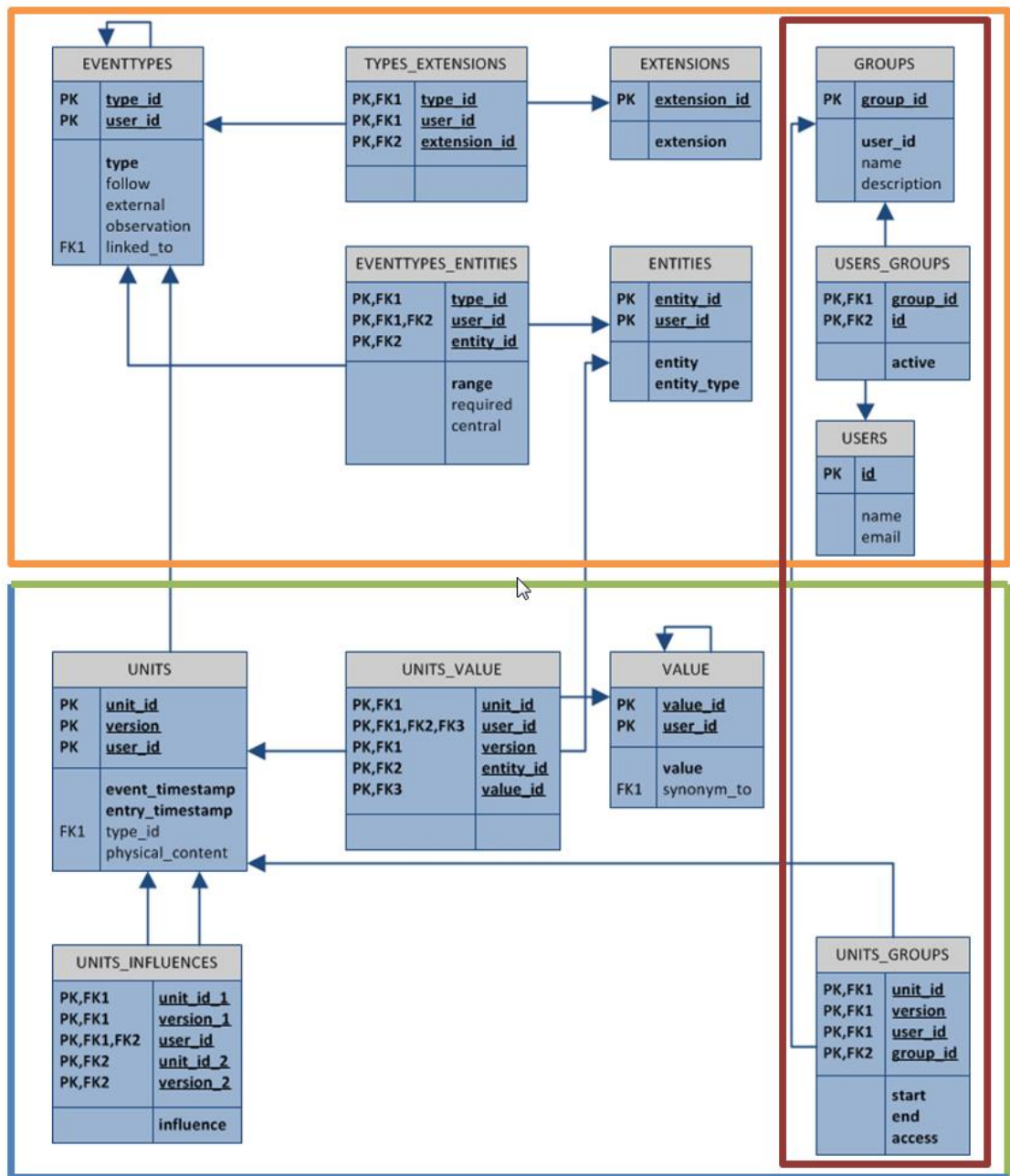
Figure 79 : Database Structure for Collaborative Environment

Table GROUPS

The table *GROUPS* stores the different users groups. Every *user group* has its unique group id (column: *group_id*), is owned by one single user (column: *user_id*), and has a short name (column: *name*) and optionally also a description (column: *description*).

An example of user groups is shown in Table 11. For a better understanding, the column *FK_user_name* has been inserted to visualise the actual value behind the element *user_id*.

Table 11 : Table GROUPS

group_id	user_id	FK_user_name	name	description
1	1	Witold Thaul	PhD	PhD supervisors and me
2	1	Witold Thaul	PhD Students	PhD fellow students
3	2	Udo Bleimann	KM Class 13	Master Class of Knowledge Management 2013

Table USERS

The table *USERS* contains the individual users of the system. Beside the unique user id (column: *id*), every entry contains a name (column: *name*) and an email-address (column: *e-mail*). This allows other users a clear recognition of the wanted collaboration partner. Additional information might be used as well, are however not really necessary. Table 12 represents a sample user table.

Table 12 : Table USERS

id	name	e-mail
1	Witold Thaul	witold.thaul@plymouth.ac.uk
2	Udo Bleimann	udo.bleimann@h-da.de
3	Nathan Clarke	N.Clarke@plymouth.ac.uk
4	Christoph Wentzel	christoph.wentzel@h-da.de

Table USER_GROUPS

The table *USER_GROUPS* links the different individual users with the preconfigured user groups. The entry in the column active describes if the user is still active within the system. Possible values are *yes* (1) or *no* (blank).

The example in Table 13 presents the members of the group *PhD* (1) owned by user *Witold Thaul* with the id 1.

Table 13 : Table USER_GROUPS

group_id	FK_group_name	id	name	active
1	PhD	1	Witold Thaul	1
1	PhD	2	Udo Bleimann	1
1	PhD	3	Nathan Clarke	1
1	PhD	4	Christoph Wentzel	1

Table UNITS_GROUPS

The table *UNITS_GROUPS* assigns the different user groups to the knowledge units. By doing this, it gives the members the access to the in general private resources. The entry in Table 14 assigns the *PhD group* (column: *group_id*) to the unit *55010* (column: *unit_id*) – the *I-Know Conference Paper*. The access starts on the 11.02.12 (column: start) and has an open end (column: *end*). It allows the participants a writeable access (column: *access*). Possible Values: *writable* (1), read-only (blank).

Table 14 : Table UNITS_GROUPS

unit _id	FK_unit_name	version	user _id	group _id	FK _name	start	end	access
55010	I-Know Conference Paper	47	1	1	PhD	11.02.2012		1

7.7 Summary

The interviewed experts confirmed from different perspectives the need for an additional support within their knowledge work. The suggested idea of tracing and creating versions of our knowledge work, irrespective of the knowledge work type, as well as of the environment in which it takes place, is a most welcome idea. However, the biggest advantage of the suggested approach is seen in the idea to support the user in remembering why he or she has chosen to go a certain direction. The link of certain knowledge work tasks, enhanced with the statement about the kind of influence these tasks had on each other is a new, and valuable improvement of the current situation.

Regarding the presented implementation design, the general echo was quite positive. There has been a homogeneous wish to simplify the scaling of the weighting down to three to maximum five values. Any higher in-depth classification is seen as a challenge in the day-to-day usage. Moreover, the suggested presentation could be enhanced with some additional information, like for example the date and time of the single versions.

For most interview partners the collaboration with other is an important key topic. The upgrade of the approach to a multi-user solution is from a technological perspective principally no problem and has been explained in the previous section. The challenge here lies more on the ethical questions regarding user privacy, data security, behaviour patterns, as well as possible conclusions about the users' productivity.

The largest concern regarding the suggested approach has been seen in the algorithms for automatic detection of the influence levels between the certain knowledge work parts. To be user-friendly and practicable, and at the same time effective, most of the assignments

need to be done in an automatic mode. The suggested method of comparing contents with each other has in general been seen as an adequate approach. The idea of not comparing the whole content, but just the incremental part since the last version with the external document was one idea how this might be improved. Another idea was to include the timeframe duration in which the related document was active on the screen. Both ideas, as well as the original one, can be used within the context of text-based documents. However, even in a text-based context not every constellation can be handled. Sometimes it is just a short moment in which we get an idea, sometimes it is a completely different content, which brings us to an idea. Furthermore, an adequate solution that fits all the events that are not text-based need to be found. The application of semantics in this part might be an approach, but this area is worth being further research. For now, the suggested methods, enhanced with manual adjustment possibility would need to be applied.

To summarise the interview partners really see a new and sophisticated idea here. 'You have got a fan here' (Morrison 2012, l. 366) said Morrison, Prof. Schmidt 'really like the idea' (Schmidt 2013, l. 423), and Prof. Schmauch 'would buy it' (Schmauch 2013, l. 321) if it was available. By extending the functionality by not only looking at the high level of applications, but for example how they have been used, which search has been performed, it can even become 'more than just the tracking software, it becomes a learning log, where you can record your experience, and your thought process as well, and things like why was that document important in contrast just to the fact that it was important' (O'Keeffe 2012, l. 189-192).

Assuming a running version of the system, including the already described enhancements, the five dimensions of the *KMS Success Model* (as introduced in section 7.1), can be determined as follows:

1. System Quality

   To assess the dimension *System Quality*, the question 'how well knowledge creation, storage, retrieval, transfer and application are performed' (Table 8) needs to be answered. The suggested system supports the observation of all knowledge relevant events irrespectively of the actual format or environment in which they occur. This ensures a high ability to 'bring knowledge to bear upon current activities' (Jennex and Olfman 2006)'. The centralised approach enables the storage of all captured knowledge items in one central place in a connected form, which also supports the later transfer and retrieval of the needed knowledge to a high extend. Based on the identified system enhancements, a high amount of the created and managed knowledge could be integrated and computerised. In total, these characteristics ensure a high System Quality value.

2. Knowledge Quality

   Referring to the dimension *Knowledge Quality*, the question if 'the right knowledge with sufficient context is captured and made available for the user at the right time' (Table 8) needs to be determined. The system is meant to observe every knowledge work relevant event. To identify the (contextually) really relevant information and knowledge, the events are weighted and assessed according to their contextual belonging and importance. The extracted descriptive metadata and rela-

tionships also ensure a sufficient context of the captured items. This has also been confirmed by the interview partners. According to them, the right knowledge with sufficient context is captured and made available for the user whenever he needs it. This is especially true, if we assume that the current technological problems with the automated assessing of the levels of influences between the several events can be solved. The dimension *Knowledge Quality* can therefore be assigned to a high value.

3. User Satisfaction

A final assessment of the dimension *User Satisfaction* is difficult at this stage and would required the system to be implemented. Based on the concepts, mock-ups, and screenshots there has been a (partially very) positive feedback by the interviewed experts, as well as by the students during the survey. This however would need to be confirmed during a real usage phase of the system.

4. Intent To Use / Perceived Benefit

Asked if the users would use such as system if it would be available, the majority answered with a clear 'yes' and encouraged the author to keep on with the development of the software. Therefore, the dimension *Intent To Use / Perceived Benefit* is announced to be of a high value. It would however in the end depend on overcoming of the identified problems and issues, and also on the final implementation of the system (as mentioned by several interview partners).

5. Net Impact

For 86% of the surveyed students, as well as for all the interviewed experts the concepts of the system seem useful; they would like to use it with the goal to improve their effectiveness and performance within learning processes. Many of them could also imagine the application of the system in other areas of their lives. The dimension *Net Impact* can therefore be assigned to a high value.

Applying the results from the confirmation of the concept on Davis's *Technology Acceptance Model* (subsection 3.6) would also offer a quite positive view. The *Perceived Usefulness* has been committed by the experts, and also rated relatively high by most of the students. At the same time, the *Perceived Ease Of Use* has been rated high, especially when assuming that the current technological problems with the automated association of influence levels would be settled.

8 Conclusions & Future Work

To conclude the thesis, this chapter summarises the achievements and limitations of the research. It also identifies new research areas within which the work presented could be enhanced.

This research set out to explore how learners can be supported from a methodological and a technological point of view in the evolving learning environment. To achieve this, the research was divided into the following phases:

1. Assessment of the requirements for an effective and sustainable approach to support knowledge work within today's learning environments.

2. Investigation and understanding of the currently existing approaches and identifying their problematic issues.

3. Design of an applicable and beneficial methodology to reduce and overcome the existing weaknesses.

4. Design of a technological architecture to support the aims of the suggested methodologies.

5. Confirmation of the feasibility and benefits of the theoretical concept and the technological system design.

8.1 Achievements of the Research

The aim of this research was to define, design and confirm an advanced concept to support students with their contemporary learning processes. This has been achieved by assessing current approaches, identifying problematic issues, and defining and designing an applicable methodology and technological architecture to benefit students.

Concrete the following results have been achieved:

1. The problems and challenges resulting from the evolving learning environment have been identified and confirmed by university level students via a survey. These include issues like the paradox of choices which information and knowledge to use, where to find it, how to organise it, how to recover it, as well as questions like how trustable the identified knowledge is. Resulting from these subjects, the requirement for a more sophisticated methodological and technological support of the students within their daily knowledge work has been formulated (chapter 2).

2. Learning and knowledge are linked very tight to each other. They represent the same phenomenon from different perspectives. The acquisition, distribution, and the effective usage of information and knowledge (Davenport 1994) are covered by the research discipline of Knowledge Management (KM). Therefore, the area of Knowledge Management has been identified as an adequate field to better support the learners.

   To cover the personal, individual aspects of knowledge work, different working patterns and behaviours, as well as different learner types, the subarea of Personal

Knowledge Management (PKM) has been assessed from a methodological, technological, and cost perspective. Furthermore, the related research area of semantic desktops has been determined, as well as the different approaches within the area of recommender systems discussed (chapter 3).

3. The current application of Personal Knowledge Management Systems within learning environments has been critically discussed. In this context, four types of applications within the academic environment have been identified, and their main problems analysed.

   Knowledge and information fragmentation has been identified as one of the central problematic issues. Furthermore, the *nescience of the knowledge formation process* has been recognized as another controversial topic. These two aspects have been examined in detail (chapter 4) and confirmed by university level students via a survey. Based on the findings the overall objectives on a new methodological and technological solution have been formulated (chapter 5):

   - Comprehensive and automated versioning of the user's knowledge work

   - Tracing of the personal holistic knowledge formation process

   - Support the users within their learning processes

   - Respect of user privacy and data security

   - Based on the process / task and the infrastructure / generic approach

4. A novel Personal Knowledge Management approach to meet the formulated objectives and support learners has been methodologically described (chapter 5). The main idea is to trace the user's knowledge work, independently of the kind of work he is performing, as well as of the technological environment in which it oc-

curs. By tracing the entire knowledge work, the knowledge formation process of every single information and knowledge item is created and stored within the system. The events, during which the knowledge work results are created and developed, are used as the central instance of the approach. All other information like the actual content, the descriptive metadata, as well as the relationships to other knowledge items is then linked to those corresponding events. Individual configuration and customisation of the system reflect the personal user preferences and behaviours (chapter 5).

5. The described versioning part of the system allows access to the entire digitised knowledge and information base of the user, including different intermediate versions. From a process point of view, this allows a later path combination, as well as an unrestricted access to the required information. It helps overcoming the knowledge and information fragmentation problems.

6. By relating events to each other, and by assessing the dependencies and influences they have on each other, the system enhances the versioning part by looking not only on what happened, but also by helping the user to remember why it happened. It supports the user to find again already used resources, even if they are no longer available within the original location. By combining the corresponding information of several independent knowledge items, and the independencies they have on each other, a holistic knowledge formation process for single knowledge items, as well as for the entire knowledge base is be created. This could also be used to reproduce undertaken tasks within other technological environments, as well as within other contexts. To reduce the knowledge and information overflow,

it by default presents context-related events to the user, giving him at the same time an indication of the effect they had on the outcome.

7. Beside the methodological description, a technological design of the system has been created (chapter 6). The service has been divided into the functional units *Configuration Unit*, *Knowledge Storage Unit*, *Knowledge Access Unit* and the *database*. The units have been conceptually developed and discussed from a technological and a presentation point of view. In this context, the database schema has been defined, the graphical user interfaces drafted, and the underplayed functionality described.

8. For the representation of and the access to the stored knowledge and information base several ways have been described. A sortable and searchable table-based solution gives an overview of all existing knowledge items. By aggregating the different versions and intermediate steps, a better overview is offered. Moreover, a time-related graph representation of the knowledge formation path has been designed. A seamless navigation between the separate knowledge graphs, as well as an extended search functionality allow the user an easy and fast finding, relocating and reuse of the for him in this specific context relevant information. The search functionality cannot only be used on a document base, but look even deeper into certain documents answering questions like 'when did I include this paragraph into this document' and also 'why did I include this paragraph into this document'.

9. Exemplary use cases from students' everyday life have been outlined, and typical challenges and problems illustrated (section 5.2):

   - Finding and accessing of existing information and knowledge

   - Finding again and accessing of already identified information and knowledge

   - Content comprehensive linking of information and knowledge

   - Sustainable understanding and traceability of previous decisions

   - Sustainable traceability of knowledge formation

   Based on the suggested theoretical and practical approach, the support of the students for every of these use cases has been pointed out (section 6.8).

10. Keeping the *KMS Success Model* and the *Technology Acceptance Model* in mind, the suggested solution has been controversially assessed by a qualitative approach applying expert interviews. The participants recognised the suggested idea as a new and useful approach to enhance the current status. The new ideas of the confirmation have been partially directly converted into enhancements of the suggested system; the other parts have been used as ideas for future research. To ensure that the main target group of the research, university level students, benefit from the outcomes of the idea, the main thoughts have been substantiated via a quantitative student survey among university level students.

The research has met all of the objectives originally outlined in chapter 1 and has resulted in the design of a new approach to support learners. The suggested concept is a software system implemented as an intelligent intermediate layer running as a background process, and observing and versioning the students' knowledge work. The idea of capturing the

knowledge creation events independently of their original type and the environment in which their occur, overcomes the identified problems resulting from knowledge and information fragmentation. Linking the events to each other (according to the time in which they happen) help the users to remember and reconstruct their thinking process. The assessments of the influences the several events had on each other is a novel contribution and assist the users with the problems of the nescience of the knowledge formation process. From the point of view of every single knowledge item, the suggested approach allows a holistic view on the knowledge formation process. In many cases this can help to remember why certain decision have been met, to reproduce earlier statues, to recover and find again previously found knowledge items, as well as to reapply previous process patterns within other problems, questions and technological environments. The suggested idea is originally meant to be applied within learning environments, but can be also usefully applied by every other kind of knowledge worker.

8.2 Limitations of the Research

Despite having met the overall objectives of the research programme, the work inevitably contains a number of limitations. The principal points are presented below:

1. The two main advantages of the suggested approach are the system comprehensive versioning of our knowledge work, as well as the information, which other events lead to the decisions we made. However, for this to work, the implemented algorithms need to meet the working patterns of the single user. The system will work relatively well for users which tend to digitalise their intermediate results promptly. Users who tend to collect their ideas first, and start writing them down with a certain time shift will dismiss the idea of the system. To gain the most advantages of the software, users with such behaviour would need to adopt their personal working patterns.

2. The automated measurement of the influence levels between the several events is the most critical part within the implementation. Until now, only methods which are connected to content based events have been discussed at all; even here the applicable algorithms are being questioned by the confirmation participants in some cases. However, to make sure the application is user friendly and not annoying, there is definite need for an automated assignment – therefore further research is required here.

   Moreover, events that are not linked to content based entities, like for example meetings or phone calls, need the application of algorithms from a completely different direction. In both cases, semantic methods might be applicable here.

3. The collaborative mode has not been part of the original implementation design; however, a high-level concept has been offered during the confirmation phase as described in subsection 7.4.5.

4. Up to now, the suggested solution brings improvements in the management of the existing knowledge and information base. Regarding the acquisition of new information and knowledge, the approach supports the users only by helping them remembering how they did it before in other (similar) situations. Nonetheless, it does not give the users any indications what other inputs might be useful, where to find them, and how valuable they are.

5. So far, the concept has been described from a methodological point of view, as well as designed from a system implementation point of view. However, to date no prototype has been implemented. The confirmation of the idea was based on a qualitative research, but could possibly lead to different outcomes if based on a long-term prototype usage.

6. One of the main reasons for the requirement of additional support for learners was the extremely increased amount of available and accessible information and knowledge. The suggested system itself produces and stores even more additional data and information. This paradox might be seen as a limitation, is on the other hand necessary for the concept to be applicable.

Despite these limitations, the research programme has been able to made valid contributions to knowledge and provided sufficient proof of concept for the ideas proposed.

8.3 Directions of Further Research

This research has only touched the surface of the optimal Personal Knowledge and Information Management system to support learners in the enhanced and fast changing learning environment. The overall advantage of the suggested approach have been described in the previous sections and chapters; the limitations have also been outlined. These areas, together with new ones, define fields where future work is possible or preferable, and could conduct to build upon and enhance that undertaken project. They are summarised in the paragraphs below.

Additional Weighting Algorithms

As mentioned before, the sophisticated implementation of the weighting algorithms is one of the most essential and critical parts of the suggested approach. It is therefore necessary to have a deeper look into the possible techniques.

On the one hand, the during the confirmation phase discussed (subsection 7.4.4) concerns within the area of document-base events need to assessed, and possible solutions determined. On the other hand, event types based on central entities for which the comparison of the content is not easily possible (if at all) need to be assessed and methods for an automated influence assignment defined. In both cases, the application of semantic ontology functionality might offer some useful procedures.

Furthermore, in this context the inverse approach could also be applied, meaning that not only the system is trying to identify and weight the relevant events, but in addition, the

user also has the possibility to manually mark important events and assign their levels of influences on others as their happen.

Comprehensive Management

Having the holistic picture of the knowledge formation processes of the entire personal information and knowledge base, a clever comprehensive management of the several knowledge items might be offered within the system. An example might be an internal information system based on cross references; if for example one document which had an high impact on other documents changed, so maybe the other (influenced) documents need to be adjusted as well. Such dependencies are worth further consideration and should be analysed.

Integration of Recommender Systems Methodology

For learners and other users the proposed solution offers a huge benefit looking backwards. It supports them within their daily work, assists them with the system overall versioning. However, looking forward only the deduction of certain behaviour patterns is supporting them so far.

The integration of the in section 3.6 introduced recommender systems methodologies would extend the backwards perspective of the suggested approach. It could give the users indications what other inputs might be useful, where to find them, and how valuable they are. Having the entire knowledge and information base of a user in a central database, is a condition for a substantial starting point for good content-based recommendations (Davies et al. 2006). There are different approaches possible; for example to rate the

information stored in the systems by matching the content to the user's behaviour: if a user has 100 documents within his knowledge base, and he uses ten of them frequently, their content and their keywords should be weighted more heavily than others'. Moreover, learners may highlight some sentences in their documents; they may add key words, descriptions, tags or some other additional information. A document with several manually annotated passages might be much more valuable (to the certain user) than a document which was just stored in the system and never opened again.

Applied in the academic area, recommender systems integrated in the suggested approach could gain an additional advantage from the nature of many academic documents. Keywords, abstracts and author's name, e-Mail addresses, university names, journal names – these are all highly qualified input parameters for successful recommendations. Moreover, references included in the documents may also be used to show the users possible interesting directions.

After the reasoned integration of the collaboration functionalities into the suggested system, the integration of the Collaborative Filtering techniques might be as well possible. For example, if a user is referring to a scientific paper stored in his own system, and another user has the same paper in his associated with several other papers and hyperlinks, this information might be also valuable to the first user. It is important to mention the privacy aspect here, but there are ways to guarantee the users privacy and make other users the knowledge retrieving processes easier: for example according to the suggested user rights and groups as indicated in the previous sections. Another possibility would be not to show the user the recommended items, but only inform him, that another user has

something relevant to this context. In this case, the user might contact the other one and

ask for advice.

References

1. Abecker, A. and Völkel, M. (2008) "Cost-Benefit analysis for the design of personal knowledge management systems". Proceedings of *ICEIS 2008 - International Conference on Enterprise Information Systems*.

2. Academhack (2010) *"Twitter for Academia"*. http://academhack.outsidethetext.com/home/2008/twitter-for-academia, last accessed: 2013-07-23.

3. Agnihotri, R. and Troutt, M. (2009) "The effective use of technology in personal knowledge management - a framework of skills, tools and user context". *Online Information Review*, vol. 33, no. 2, pp. 329 – 342.

4. Alavi, M. and Leidner, D. (2001) "Review: Knowledge management and knowledge management systems: Conceptual foundations and research issues". *MIS Quarterly*, vol. 25, no. 1, pp. 107 – 136.

5. Alley, L. (1999) "Diverting a Crisis in Global Human and Economic Development: a New Transnational Model for Lifelong Continuous Learning and Personal Knowledge Management". *Higher Education in Europe*, vol. 24, no. 2.

6. Amazon.com (2013) *"Amazon.com: Online Shopping for Electronics, Apparel, Computers, Books, DVDs & more"*. http://www.amazon.com, last accessed: 2013-07-23.

7. Android (2013) *"Android"*. http://www.android.com, last accessed: 2013-07-23.

8. Ansari, A., Essegaier, S. and Kohli, R. (2000) "Internet recommendation systems". *Journal of Marketing Research*, vol. 37, no. 3, pp. 363 – 375.

9. Apache (2013a) *"Apache Software Foundation - Apache POI - the Java API for Microsoft Documents"*. http://poi.apache.org, last accessed: 2013-07-23.

10. Apache (2013b) *"Apache Software Foundation - Apache PDFBox - Java PDF Library"*. http://pdfbox.apache.org, last accessed: 2013-07-23.

11. Apache (2013c) *"Apache Software Foundation - Apache Subversion"*. http://subversion.apache.org/, last accessed: 2013-07-23.

12. Apple (2013) *"iOS: Understanding Spotlight Search"*. http://support.apple.com/kb/ht3636, last accessed: 2013-07-23.

13. Apshvalka, D. and Grundspenkis, J. (2005) "Personal and organizational knowledge management as a driving force for business process effectiveness" in Proceedings of *Int. Conf. on Information Technologies for Business*, Vilnius University, Kaunas, pp. 9 – 14.

14. Avery, S., Brooks, R., Brown, J., Dorsey, P. and O'Conner, M. (2001) "Personal knowledge management: Framework for integration and partnerships". Proceedings of *ASCUE Conference.*

15. Balabanovic, M. and Shoham, Y. (1997) "Fab: Content-based, collaborative recommendation". *Communication of the ACM*, vol. 40, no. 3, pp. 66 – 72.

16. Bargmeyer, B., Gillman, D. (2011) "Metadata Standards And Metadata Registries: An Overview". *Environmental Protection Agency* and *Bureau of Labor Statistics*, Washington, DC. http://www.bls.gov/ore/pdf/st000010.pdf, last accessed: 2013-07-23.

17. Barth, S. (2004) "Self-Organization: Taking a Personal Approach to KM. Knowledge Management Tools and Techniques: Practitioners and Experts Evaluate KM Solutions". *Madanmohan Rao,* Butterworth-Heinemann.

18. Baumann, S. (2001) "Theoretische und praktische Einführung von softwareunterstütztem Wissensmanagement in einem IT-Beratungsunternehmen". University of Applied Sciences (Fachhochschule) Schmalkalden.

19. Berkman, R. (2005) "Information Advisor", citied in Miller, R. (2005) "The Evolution of Knowledge: This Time It is Personal". *EContent*, vol. 28, no. 11, pp. 38 – 41.

20. Bohn, R. and Short, J. (2009) "*How Much Information? 2009 - Report on American Consumers"*. http://hmi.ucsd.edu/pdf/HMI_2009_ConsumerReport_Dec9_2009.pdf, last accessed: 2013-07-23.

21. Bots, P. and de Bruiin, H. (2002). "Effective Knowledge Management in Professional Organizations: Going by the rules". Proceedings of *35th Hawaii International Conference on System Sciences*, IEEE Computer Society Press.

22. Brückner, S. (2005) "Structure patterns depending on various learner types in an e-learning context". Diploma Thesis. University of Applied Sciences, Darmstadt, Germany.

23. Cakula, S. and Osis, K. (2010) "Knowledge Management in Course Content Development and Teaching". Proceedings of *EduLEARN'10 International Conference on Education and New Learning Technologies*. Barcelona, Spain, pp. 87 – 96.

24. CDS (2010) "*Cognitive Design Solutions - Explicit & Tacit Knowledge*". http://www.cognitivedesignsolutions.com/KM/ExplicitTacit.htm, last accessed: 2013-07-23.

25. Chandler (2013) "*Chandler Project – The Note-to-Self Organizer 1.0!*". http://chandlerproject.org, last accessed: 2013-07-23.

26. Chaudhuri, S. and Dayal, U. (1997) "An overview of data warehousing and OLAP technology". *ACM SIGMOD*, vol. 26, no. 1, pp. 65 – 74.

27. Churchman, C. (1979) "The system approach". *Dell Publishing*, New York in Jennex, M. and Olfman, L. (2006) "A Model of Knowledge Management Success". *International Journal of Knowledge Management*, vol. 2, no. 3, pp. 51 – 68.

28. Clark, D. (2004) "*Understanding and Performance*". http://www.nwlink.com/~donclark/performance/understanding.html, last accessed: 2013-07-23.

29. Clemente, B. and Pollara, V. (2005) "PKM: Mapping the Course, Marking the Trail". *IT Pro November*, vol. 12.

30. Copernic (*2013*) "*Copernic Desktop Search - The best desktop search tool*". http://www.copernic.com/en/products/desktop-search, last accessed: 2013-07-23.

31. Creswell, J. (1998) "Qualitative inquiry and research design: Choosing among five traditions". *Thousand Oaks*, CA, Sage.

32. Davenport, T. and Prusak, L. (1998) "Working Knowledge: How organizations manage what they know". *Harvard Business School Press*, Boston, Massachusetts.

33. Davenport, T. (2005) "Thinking for a Living: How to Get Better Performances And Results from Knowledge Workers". *Harvard Business School Press*, Boston, Massachusetts.

34. Davis, F. (1989) "Perceived usefulness, perceived ease of use, and user acceptance of information technology". *MIS Quarterly*, vol. 13, no. 3, pp. 319 – 340.

35. Davies, S., Allen, S., Raphaelson, J., Meng, E., Engleman, J., King, R. and Lewis, C. (2006) "Popcorn: The Personal Knowledge Base". Proceedings of *Conference on Designing Interactive Systems*, University Park, PA, USA, June 26 – 28.

36. DeLone, W. and McLean, E. (2002) "Information Systems Success Revisited". Proceedings of *35th Annual Hawaii International Conference on System Sciences*, IEEE Computer Society.

37. Definitions (2013) "*What does content mean?*". http://www.definitions.net/definition/content, last accessed: 2013-07-23.

38. Dittrich, K., Dayal, U. and Buchmann, A. (2011) "On Object-Oriented Database Systems". *Springer Publishing Company*.

39. Drucker, P. (1985) "Management: Tasks, responsibilities, practices". *Harper & Row*.

40. Duhon, B. (1998) "It is All in our Heads". Inform, vol. 9, no. 12, pp. 9 – 13.

41. Ehrlich, K. and Maltz, D. (1995) "Pointing the way: Active collaborative filtering". *CHI Proceedings Papers*.

42. Evernote (2013) "*Remember everything with Evernote, Skitch and our other great apps*". https://evernote.com, last accessed: 2013-07-23.

43. Facebook (2013) "*Facebook helps you connect and share with the people in your life*". http://www.facebook.com, last accessed: 2013-07-23.

44. Farber, D. (2012) "*Twitter hits 400 million tweets per day, mostly mobile*". http://news.cnet.com/8301-1023_3-57448388-93/twitter-hits-400-million-tweets-per-day-mostly-mobile, last accessed: 2013-07-23.

45. Fenfire (2013) "*The Fenfire Project*". http://fenfire.org, last accessed: 2013-07-23.

46. Flickr (2013) "*Welcome to Flickr - Photo Sharing*". http://www.flickr.com, last accessed: 2013-07-23.

47. Frand, J. and Hixon, C. (1999) "Personal knowledge management: Who, what, why, when, where, how?". Speech working paper.

48. Franz, T. and Staab, S. (2005) "SAM: Semantics Aware Instant Messaging For the Networked Semantic Desktop". Proceedings of *1st Workshop On The Semantic Desktop. 4th International Semantic Web Conference*, Galway, Ireland.

49.   FreeAdvice (2013) *"What is content in terms of intellectual property?"*. http://law.freeadvice.com/intellectual_property/communications_law/content_definition.htm, last accessed: 2013-07-23.

50.   Freecode (2013) *"catdoc - Freecode"*. http://freecode.com/projects/catdoc, last accessed: 2013-07-23.

51.   FreeMind (2013) *"Free mind mapping software"*. http://freemind.sourceforge.net, last accessed: 2013-12-23.

52.   FSF (2013) *"UnRTF - GNU Project - Free Software Foundation (FSF)"*. http://www.gnu.org/software/unrtf, last accessed: 2013-07-23.

53.   Gantz, J., Chute, C., Manfrediz, A., Minton, S., Reinsel, D., Schlichting, W. and Toncheva, A. (2011). "The Diverse and Exploding Digital Universe: An Updated Forecast of Worldwide Information Growth Through 2011". *IDC*, March.

54.   Gemmell, J., Bell, G., Lueder, R., Drucker, S. and Wong, C. (2002) "Mylifebits: Fulfilling the memex vision". In: *ACM Multimedia*, December 1-6, Juan-les-Pins, France, pp. 235 – 238.

55.   Glyph & Cog (2013) *"Xpdf & Open Source"*. http://www.glyphandcog.com/Xpdf.html, last accessed: 2013-07-23.

56.   Gnome (2013) *"GNOME 3: Ease, comfort and control"*. http://www.gnome.org, last accessed: 2013-07-23.

57.   Google (2011a) *"Google Desktop Download"*. http://desktop.google.com/en, last accessed: 2011-09-10.

58.   Google (2011b) *"Google Desktop Update"*. http://googledesktop.blogspot.de/2011/09/google-desktop-update.html, last accessed: 2013-07-23.

59.   Google (2013a) *"About Google Scholar"*. http://scholar.google.com/intl/en/scholar/about.html, last accessed: 2013-07-23.

60.   Google (2013b) *"Google Translate"*. http://translate.google.com, last accessed: 2013-07-23.

61.   Green, A. (2013) *"Power Down Google Reader"*. http://googlereader.blogspot.de/2013/03/powering-down-google-reader.html, last accessed: 2013-07-23.

62. Gurteen, D. (2005) citied in Miller, R. (2005) "The Evolution of Knowledge: This Time It is Personal". *EContent*, vol. 28, no. 11, pp. 38 – 41.

63. Haas, J. (2013) "*Linux / Unix Command: pdfinfo*". http://linux.about.com/library/ cmd/blcmdl1_pdfinfo.htm, last accessed: 2013-07-23.

64. Hey, J. (2004) "*The Data, Information, Knowledge, Wisdom Chain: The Metaphorical link*". http://www.dataschemata.com/uploads/7/4/8/7/7487334/dikwchain.pdf, last accessed: 2013-07-23.

65. Heyer, G., Quasthoff, U. and Wolff, C. (2002) "*Möglichkeiten und Verfahren zur automatischen Gewinnung von Fachbegriffen aus Texten*". http://www.researchgate.net/ publication/220003645_Mglichkeiten_und_Verfahren_zur_automatischen_Gewinnung_ von_Fachbegriffen_aus_Texten, last accessed: 2013-07-23.

66. Higgison, S. (2005) "Your say: Personal knowledge management". *Insight Knowledge*, vol. 7.

67. Hölzle, U. (2010) "*Update on Google Wave*". http://googleblog.blogspot.com/2010/08/ update-on-google-wave.html, last accessed: 2013-07-23.

68. Huber, G., Davenport, T., and King, D. (1998). "Some perspectives on organizational memory". Working paper for the task force on organizational memory. Proceedings of *31st Annual Hawaii International Conference on System Sciences*.

69. iOS7 (2013) "*iOS7 – the mobile OS from a whole new perspective*". http://www.apple. com/uk/ios/ios7, last accessed: 2013-07-23.

70. Iske, P. and Boeckhoff, T. (2002) "The value of knowledge doesn't exist" in Karagiannis, D. and Reimer, U. (2002). PAKM, vol. 2569 of *Lecture Notes in Computer Science*, pp. 632 – 638. Springer.

71. Java (2013) "*java.com: Java + You*". http://www.java.com/en, last accessed: 2013-07-23.

72. Jefferson, T. (2006) "Taking it personally: personal knowledge management". *The journal of information and knowledge management systems*, vol. 36, no. 1, pp. 35 – 37.

73. Jennex, M. (2005) "What is knowledge management?". *International Journal of Knowledge Management*, vol. 1, no. 4, pp. i – iv.

74. Jennex, M. (2006) "Culture, context, and knowledge management". *International Journal of Knowledge Management*, vol. 2, no. 2, pp. i – iv.

75. Jennex, M. (2008) "The Impact of Culture and Context on Knowledge Management" in *Current Issues in Knowledge Management*, edited by Jennex, M., IGI, pp. 6 – 11.

76. Jennex, M. and Olfman, L. (2002) "Organizational memory/knowledge effects on productivity: A longitudinal study". Proceedings of 35th *Annual Hawaii International Conference on System Sciences*, IEEE Computer Society Press.

77. Jennex, M. and Olfman, L. (2004) "Assessing Knowledge Management Success/Effectiveness Models!". Proceedings of 37th *Hawaii International Conference on System Sciences*, IEEE Computer Society Press.

78. Jennex, M. and Olfman, L. (2006) "A Model of Knowledge Management Success". *International Journal of Knowledge Management*, vol. 2, no. 3, pp. 51 – 68.

79. Jones, W. (2007) "Keeping Found Things Found: The Study and Practice of Personal Information Management". *Academic Press, Morgan Kaufmann Publishers*, Burlington.

80. Karger, D. (2006) "It is All the Same to Me: Data Unification in Personal Information Management". *Communications of the ACM*, vol. 49, no. 1.

81. KDE (2013) "*KDE - Experience Freedom!*". http://kde.org, last accessed: 2013-07-23.

82. Klein, D. (2012) "The Strategic Management of Intellectual Capital". *Routledge*, Woburn, MA.

83. Kolb, D. (1984) "Experiential learning: Experience as the source of learning and development". *Prentice-Hall*, New Jersey.

84. KM Magazine (2000) "Personal Knowledge Management Tools". *Knowledge Management Magazine*, vol. 12.

85. Krcmar, H. (1996) "Informationsmanagement". *Springer*, Heidelberg.

86. Lansdale, M. (1988) "The psychology of personal information management". *Applied Ergonomics*, vol. 19, no. 1, pp. 55 – 66.

87. Leedy, P. and Ormrod, J. (2005) "Practical Research – Planning and design". *Pearson Merrill Prentice Hall*.

88.   Lerno (2013) *"leno.de – Der Lernwegweiser im Internet"*. http://www.lerno.de/lerno/ index.html, last accessed: 2013-07-23.

89.   Li, G. and Li, Y. (2009) "A Study on Blog Based Personal Knowledge Management". Proceedings of *2nd International Workshop on Knowledge Discovery and Data Mining*. Moscow, Russia, pp. 248 – 251.

90.   Lincoln, Y. (2001) in Reason, P. and Bradbury, H. "Handbook of Action Research: Participative inquiry and practice", *Sage Pubn Inc*.

91.   Lindsey, K. (2002) "Measuring Knowledge Management Effectiveness: A Task-Contingent Organizational Capabilities Perspective". Proceedings of *Eighth Americas Conference on Information Systems*, pp. 2085 – 2090.

92.   LMS (2013a) *"Learning Management Solutions - Information management, learning enhancement and knowledge management software"*. http://www.kworkshop.com, last accessed: 2013-02-20.

93.   LMS (2013b) *"Knowledge Workshop - no longer available"*. http://www.kworkshop. com, last accessed: 2013-07-23.

94.   Lu, J., Yo, C. and Liu, C. (2002) "Learning style, learning patterns, and learning performance in a WebCT-based MIS course". *Information & Management*, vol. 40, pp. 497 – 507.

95.   Lyman, P. and Varian, H. (2000) "*How Much Information, 2000*". http://www2.sims. berkeley.edu/research/projects/how-much-info, last accessed: 2013-07-23.

96.   Lyman, P. and Varian, H. (2003) "*How Much Information, 2003*". http://www2.sims. berkeley.edu/ research/projects/how-much-info-2003, last accessed: 2013-07-23.

97.   Martin, J. (2006) "*PKM: the basis of corporate and institutional knowledge management*" in Martin, J. and Wright, K. (2006) "Managing Knowledge: Case Studies in Innovation". *Spotted Cow Press*, Edmonton.

98.   Massey, A., Montoya-Weiss, M. and O'Driscoll, T. (2002) "Knowledge Management in Pursuit of Performance: Insights from Nortel Networks". *MIS Quarterly*, vol. 26, no. 3, pp. 269 – 289.

99.   Meyer, D. (2011) "*Microsoft explains Windows 8 desktop search*". http://www.zdnet. com/microsoft-explains-windows-8-desktop-search-4010024605, last accessed: 2013-04-09.

100.   Microsoft (2000) "*Usability in Software Design*". http://msdn.microsoft.com/en-us/library/ms997577.aspx, last accessed: 2013-07-23.

101.   Microsoft (2010) *"Windows Search"*. http://www.microsoft.com/windows/products/ winfamily/desktopsearch/default.mspx, last accessed: 2010-08-23.

102.   Miller, R. (2005) "The Evolution of Knowledge: This Time It is Personal". *EContent*, vol. 28, no. 11, pp. 38 – 41.

103.   Millikin University (2003) "*Personal Knowledge Management at Millkin University*". http://www.millikin.edu/pkm, last accessed: 2010-08-23.

104.   MindGenius (2013) "Mind Mapping Software". http://www.mindgenius.com, last accessed: 2013-12-23.

105.   Morrison, J. and Weiser, M. (1996) "A Research Framework for Empirical Studies in Organizational Memory". Proceedings of *29th Annual Hawaii International Conference on System Sciences*, IEEE Computer Society Press.

106.   Morrison, M. (2012) Thesis Confirmation Interview. Appendix B, Part 4. Interviewed 07.09.2012, I-Know 2012 Conference, Graz, Germany.

107.   Nepomuk (2012) "*The Social Semantic Desktop*". http://nepomuk.semanticdesktop.org/ nepomuk, last accessed: 2012-07-23.

108.   Nepomuk (2013) "*Nepomuk - KDE UserBase Wiki*". http://userbase.kde.org/Nepomuk, last accessed: 2013-07-23.

109.   Netcraft (2013) "*Internet Research, Anti-Phishing and PCI Security Services*". http://news.netcraft.com, last accessed: 2013-07-23.

110.   Nonaka, I. (1994) "A dynamic theory of organizational knowledge creation". *Organization Science*, vol. 5, no. 1, pp. 14 – 37.

111.   Nonaka, I. and Takeuchi, H. (1995) "The knowledge creating company". Oxford University Press, New York.

112. North, K. (2005) "Wissensorientierte Unternehmensführung. Wertschöpfiung durch Wissen". *Gabler*, Wiesbaden.

113. O'Keeffe, I. (2012) Thesis Confirmation Interview. Appendix B, Part 1. Interviewed 06.09.2012, I-Know 2012 Conference, Graz, Germany.

114. OECD (1999) "Measuring and Reporting Intellectual Capital: Experience, Issues and Prospects". *Programme Notes and Background to Technical Meeting and Policy and Strategy Forum*, Paris.

115. Orlowski, A. (2005) "*Wikipedia founder admits to serious quality problems - Yes it is garbage, but it is delivered so much faster!*". http://www.theregister.co.uk/2005/10/18/wikipedia_quality_problem, last accessed: 2013-07-23.

116. OU (2013) "*Distance Learning Courses and Adult Education - The Open University*". http://www.open.ac.uk, last accessed: 2013-07-23.

117. Oxford (2012) "*Definition of learning in Oxford Dictionaries*". http://oxforddictionaries.com/definition/english/learning, last accessed: 2013-07-23.

118. Pappas, N. (2012) Thesis Confirmation Interview. Appendix B, Part 6. Interviewed 06.09.2012, I-Know 2012 Conference, Graz, Germany.

119. Pauleen, D. (2009) "Personal knowledge management: putting the 'person' back into the knowledge equation". *Online Information Review*, vol. 33, no. 2, pp. 221 – 224.

120. Personal Brain (2013) "*PersonalBrain – Ihr digitales Gedächtnis*". http://www.pbrain.de, last accessed: 2013-07-23.

121. Polanyi, M. (1958) "Personal Knowledge: Towards a Post-Critical Philosophy". *Routledge & Kegan Paul Ltd*, London.

122. Polanyi, M. (1967). "The tacit dimension". *Routledge and Keoan Paul*, London.

123. Polanyi, M. (1996) "The tacit dimension". *University of Chicago Press*, Chicago.

124. Pollard, D. (2008) "PKM: A bottom-up approach to knowledge management". *Knowledge Management in Practice: Connections and Context*, ed. Srikantaiah, T. and Koenig, M. *Information Today*, pp. 95 – 114.

125. PowerfullInformation (2013) *"The Nature of Information"*. http://www.powerfulinformation.org/page.cfm?pageid=pi-infoquotes, last accessed: 2013-07-23.

126. Probst, G. (1998) "Practical Knowledge Management: A Model That Works". Arthur D. Little, issue 2.

127. Probst, G., Raub, S. and Romhardt, K. (1997) "Wissen managen. Wie Unternehmen ihre wertvollste Ressource optimal nutzen". *FAZ Gabler*, Frankfurt am Main.

128. Quan, D., Huynh, D. and Karger, D. (2003) "Haystack: A platform for authoring end user semantic web applications". In Proceedings of *International Semantic Web Conference*. 2003, pp. 738–753.

129. Refinder (2013) *"Refinder – Search above the Cloud"*. *https://www.getrefinder.com/about*, last accessed: 2013-07-23.

130. Reinmann, G. and Eppler, M. (2008) "Wissenswege – Methoden für das persönliche Wissensmanagement" *Verlag Hans Huber*, Bern.

131. Reinmann, G. and Mandl, H. (2004) "Psychologie des Wissensmanagements. Perspektiven, Theorien und Methoden". *Hogrefe*, Göttingen.

132. Resnick, P. and Varian, H. (1997) "Recommender Systems". *Special Section in Communications of the ACM*, vol. 40, no. 3.

133. Rexha, A. (2012) Thesis Confirmation Interview. Appendix B, Part 7. Interviewed 06.09.2012, I-Know 2012 Conference, Graz.

134. Roehl, H. (2000) "Instrumente der Wissensorganisation. Perspektiven für die differenzierende Interventionspraxis". *Gabler*, Wiesbaden.

135. Sarwar, B., Karypis, G., Konstan, J. and Riedl, J. (2000) "Analysis of Recommendation Algorithms for E-commerce". Proceedings of *2nd ACM Conference on Electronic Commerce*, pp. 158 – 167.

136. Sauermann, L. (2003) "The gnowsis-using semantic web technologies to build a semantic desktop". *Diploma thesis*, Technical University of Vienna.

137. Sauermann, L. (2005) "The Gnowsis Semantic Desktop for Information Integration". Proceedings of *Intelligent Office Appliances Workshop of the 3rd Conference on Professional Knowledge Management in Kaiserslautern*, Germany, WM2005.

138. Sauermann, L., Bernardi, A. and Dengel, A. (2005) "Overview and Outlook on the Semantic Desktop". *ISWC*.

139. Schmauch, C. (2013) Thesis Confirmation Interview. Appendix B, Part 3. Interviewed 27.02.2013, HS_Karlsruhe, Karlsruhe, Germany.

140. Schmidt, A. (2013) Thesis Confirmation Interview. Appendix B, Part 5. Interviewed 27.02.2013, KIT, Karlsruhe, Germany.

141. Schmidt, A.P. (2013) Thesis Confirmation Interview. Appendix B, Part 2. Interviewed 28.03.2013, HS_Karlsruhe, Karlsruhe, Germany.

142. Schock, M. (2010) "*Facebook and Academic Performance | Dr Shock MD PhD*". http://www.shockmd.com/2010/04/19/facebook-and-academic-performance, last accessed: 2013-07-23.

143. Semafora (2013) "*Semafora – True Symantics*". http://www.semafora-systems.com/en, last accessed: 2013-07-23.

144. Sendall, P., Ceccucci, W. and Peslak, A. (2008) "Web 2.0 Matters: An Analysis of Implementing Web 2.0 in the Classroom". *Information Systems Education Journal*, vol. 64, no. 6.

145. Seymour, R. (2002) "Surprise Me". *Design Council*, London.

146. Siegler, M. (2010) "*Eric Schmidt: Every 2 Days We Create As Much Information As We Did Up To 2003*". http://techcrunch.com/2010/08/04/schmidt-data/, last accessed: 2013-07-23.

147. Simatwa, E. (2010) "Piaget's theory of intellectual development and its implication for instructional management at presecondary school level". *Education Research and Reviews*, vol 5(7), pp. 366-371, July 2010, ISSN 1990-3839.

148. Singleton, J. (2009) "*Wikipedia open to manipulation*". http://www.one-reality.net/wikipedia.htm, last accessed: 2013-07-23.

149.   Smedley, J. (2009) "Modelling personal knowledge management". *OR Insight*, vol. 22, no. 4, pp. 221 – 233.

150.   Smith C. (2013) "*June 2013 Update – By The Numbers: 32 Amazing Facebook Stats*". http://expandedramblings.com/index.php/by-the-numbers-17-amazing-facebook-stats, last accessed: 2013-07-23.

151.   Smyth, B., Sulliva, D. and Wilson, D. (2004) "Preserving Recommender Accuracy and Diversity in Sparse Datasets". *International Journal on Artificial Intelligence Tools*, vol. 13, no. 1, pp. 219 – 235.

152.   Smyth, B., Sulliva, D. and Wilson, D. (2005) "Understanding Case Based Recommendation: A similarity Knowledge Perspective". *International Journal on Artificial Intelligence Tools*, vol. 14, no. 1 and 2, pp. 215 – 232.

153.   Spiering, M. (2013) "*A better, brighter Flickr*". http://blog.flickr.net/en/2013/05/20/a-better-brighter-flickr/, last accessed: 2013-08-04.

154.   Stankosky, M. (2005) "Advances in knowledge management: university research toward an academic discipline" in "Creating the discipline of knowledge management: the latest in university research". *Elsevier*, Butterworth-Heinemann, Amsterdam.

155.   Stein, E. and Zwass, V. (1995). "Actualizing organizational memory with information systems". *Information Systems Research*, vol. 6, no. 2, pp. 85 – 117.

156.   Steiner, G. (2006) "*Lernen und Wissenserwerb*" in Krapp, A. and Weidenmann, B. "Pädagogische psychologie". *Beltz*, Weinheim, pp. 137 – 2002.

157.   Stewart, T. (1997) "Intellectual Capital: The New Wealth of Organisations". *Doubleday*, New York.

158.   Strauss, V. (2012) "Three fears about blended learning". *The Washington Post*, 22. September 2012.

159.   SurveyMonkey (2014) "*Free online survey software & questionnaire tool*". http://www.surveymonkey.com, last accessed: 2014-12-23.

160.   Terdiman, D. (2012) "*Report: Twitter hits half a billion tweets a day*". http://news.cnet.com/8301-1023_3-57541566-93/report-twitter-hits-half-a-billion-tweets-a-da/, last accessed: 2013-07-23.

161. The Open Source Initiative (2013) "*The Open Source Initiative*". http://opensource.org, last accessed: 2013-07-23.

162. TheBrain (2013) "*TheBrain : Mind Mapping Software, Brainstorming, GTD and Knowledgebase Software*". http://www.thebrain.com, last accessed: 2013-07-23.

163. Tiwana, A. (2002) "The Knowledge Management Toolkit". 2nd edn, *Prentice Hall*, London.

164. Truch, E. (2001) "Managing personal knowledge: The key to tomorrow's employability". *Journal of Change Management*, vol. 2, no. 2, pp. 102 – 105.

165. Truman, C. & Pain, R. (2013) "*Participatory Research Background and outlook*". http://www.lancaster.ac.uk/researchethics/5-2-outlook.html, last accessed: 2013-11-23.

166. Truman, C. and Raine, P. (2001) "Involving users in evaluation : the social relations of user involvement in health research*". Critical Public Health*, vol 11, no. 3, pp. 215 – 232.

167. Tsui, E. (2002) "*Technologies for Personal and Peer-to-Peer (P2P) Knowledge Management*".   http://go.webassistant.com/wa/upload/users/u1000062/1000246/p2p_km.pdf, last accessed: 2013-07-23.

168. Twitter (2011) "*Twitter is a real-time information network powered by people all around the world that lets you share and discover what's happening now*". http://blog.twitter.com/2011/06/200-million-tweets-per-day.html, last accessed: 2013-07-23.

169. Turban, E. and Aronson, J. (2001) "Decision Support Systems and Intelligent Systems". 6th edn, *Prentice Hall*, London.

170. Twitter (2013) "*Twitter*". http://twitter.com, last accessed: 2013-07-23.

171. Ubuntu (2013) "*The world's most popular free OS | Ubuntu*". http://www.ubuntu.com, last accessed: 2013-07-23.

172. USQ (2013) "*Studying by distance education*". http://www.usq.edu.au/future-students/study-options/online, last accessed: 2013-04-09.

173. Varian, H. (1999) "The economics of search". In SIGIR '99: Proceedings of *22nd annual international ACM SIGIR conference on Research and development in information retrieval*, p. 1, New York, USA. ACM.

174. Velasquez, J. and Palade, V. (2007) "Building a knowledge base for implementing a web-based computerized recommendation system". *International Journal on Artificial Intelligence Tools*, vol. 16, no. 5, pp. 793 – 828.

175. Venkatesh, V., Davis, F. (2000) "A theoretical extension of the technology acceptance model: Four longitudinal field studies". *Management Science*, vol. 46, no. 2, pp. 186 – 204.

176. Venkatesh, V. and Morris, M., Davis, G. and Davis, F. (2003) "User Acceptance of Information Technology: Toward a Unified View". *MIS Quarterly*, vol. 27, no. 3, pp. 425 – 478.

177. Weil, K. (2010) "*Twitter Blog: Measuring Tweets*". http://blog.twitter.com/2010/02/measuring-tweets.html, last accessed: 2013-07-23.

178. Wiig, K., (2004) "People-focused knowledge management: how effective decision making leads to corporate success". *Butterworth-Heinemann*, Oxford, pp. 6 – 19.

179. Wikipedia (2013a) "Wikipedia, the free encyclopedia". http://en.wikipedia.org/wiki/Main_Page, last accessed: 2013-05-01.

180. Wikipedia (2013b) "History of Wikipedia". http://en.wikipedia.org/wiki/History_of_Wikipedia, last accessed: 2013-04-09.

181. Willke, H. (2001) "Systematisches Wissensmanagement". *Lucius und Lucius*, Stuttgart.

182. Wright, K. (2005) "Personal knowledge management: supporting individual knowledge worker performance". *Knowledge Management Research and Practice*, vol. 3, pp. 156 – 165.

183. Wright, K. (2007) "Rethinking Knowledge Work: Supporting Work and Learning through Personal Knowledge Management". *KRW Knowledge Resources*.

184. Young, J. (2009) "*10 High Fliers on Twitter*". The Chronicle of Higher Education. http://chronicle.com/article/10-High-Fliers-on-Twitter/16488, last accessed: 2013-04-09.

185. Zotero (2013) "*Zotero*". http://www.zotero.org, last accessed: 2013-04-09.

186. Zuber-Skerritt, O. (2005) "A model of values and actions for personal knowledge management". *The Journal of Workplace Learning*, vol. 17, no. 1, pp. 49 – 64.

Appendix A – Database Design
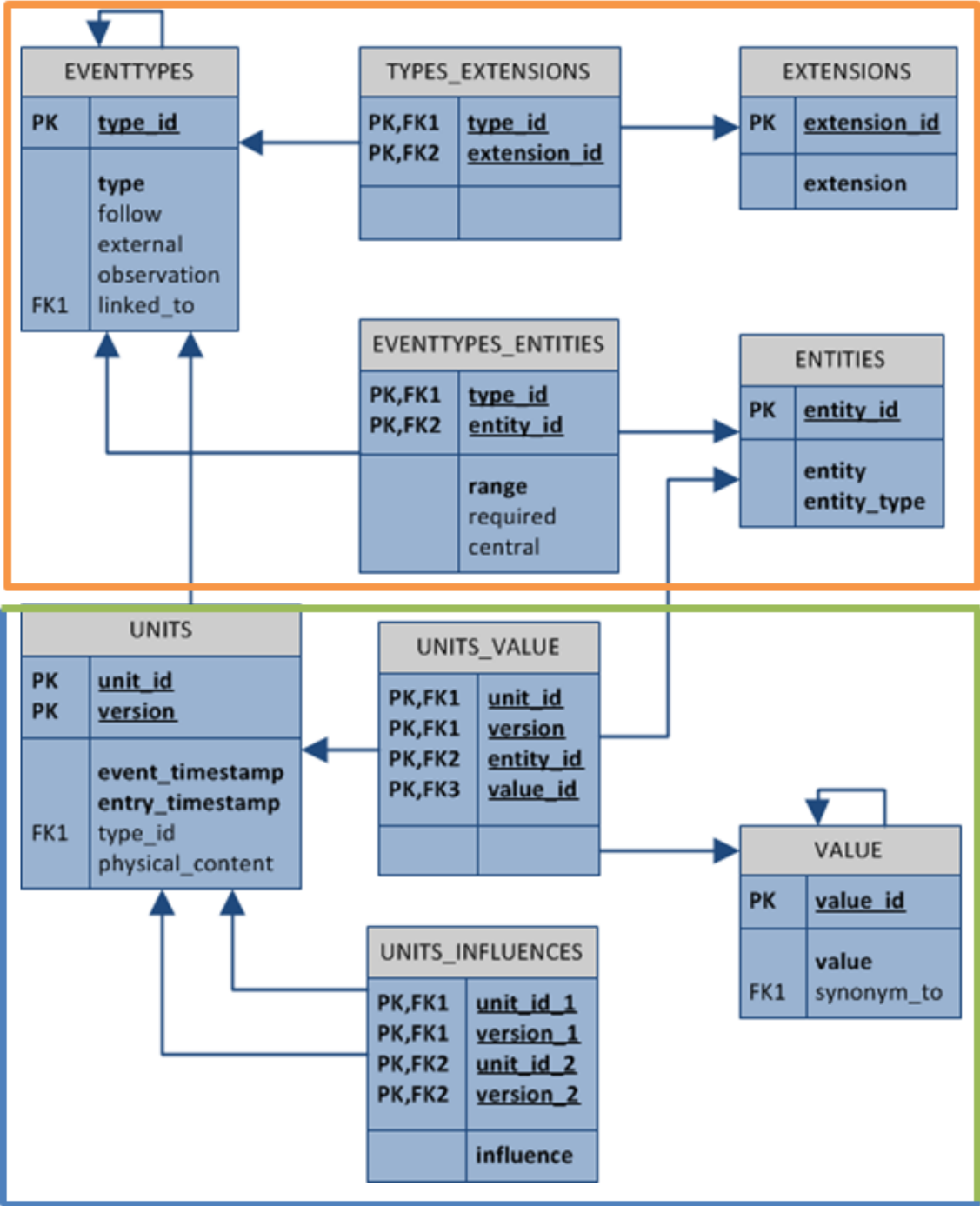
Figure 80 : Database Structure

Part 1- Table EVENTTYPES

The table *EVENTTYPES* stores the different event types. Beside the event type name and the description, the following classification is being made for every event type:

- Shall it be able to save versions of the instances of this type?

    Column: *versions*

    Possible Values: yes (1), no (blank)

- Shall instances of this type be followed, meaning that a holistic knowledge process will be recorded for every instance of this type (or should they just be used as input parameters for other events)?

    Column: *follow*

    Possible Values: yes (1), no (blank)

- Is it possible to make manual external entries for instances of this type?

    Column: *external*

    Possible Values: yes (1), no (blank)

- To which point in time shall this events be considered and stored?

    Column: *observation*

    Possible Values: start, stop, save

- Are instances of this event type directly linked to instances of other event types?

    Column: *linked_to*

    Possible Values: type_id of the linked Event Type (from the same table)

An example set of event types is shown in Table 15:

Table 15 : Table EVENTTYPES

type _id	type	description	ver- sions	fol- low	exter- nal	obser- vation	linked _to
0	unknown						
1	Text	Editing a text doc...	1	1	1	save	
2	Spreadsheet	Editing a spread...	1	1	1	save	
3	PDF	Reading a PDF ...	1	1	1	start	
4	Phone Call	Having a phone call			1	start	
6	Website	Visiting a website	1		1	start	
7	E-Mail in	Reading an E-Mail	1	1	1	start	8
8	E-Mail out	Writing an E-Mail	1	1	1	stop	7

Part 2 - Table EXTENSIONS

The tables *EXTENSIONS* stores the possible physical file extensions as shown exemplary in Table 16. Depending on the used environment these extensions may vary or be adjusted during the usage.

Table 16 : Table EXTENSIONS

extension_id	extension	extension_id	extension
1	doc	6	xlsx
2	docx	7	html
3	rtf	8	htm
4	pdf	9	php
5	xls	10	msg

Part 3 - Table EVENTTYPES_EXTENSIONS

The table *EVENTTYPES_EXTENSIONS* links the different physical file extensions from the table *EXTENSIONS* with the different event types from the table *EVENTTYPES*. This link helps the system to detect the event type according to the file, which is currently being used by the user or the system. For a better understanding of Table 17, the two columns *FK_type* (foreign key of the type) and *FK_entension* (foreign key of the extension) have been inserted to visualise the actual value behind the columns *type_id* and *extension_id*.

Table 17 : Table EVENTTYPES_EXTENSIONS

type_id	FK_type	extension_id	FK_extension
0	unknown	0	no extension
1	Text	1	doc
1	Text	2	docx
1	Text	3	rtf
3	PDF	4	pdf
4	Phone Call	0	no extension
6	Website	7	html
6	Website	8	htm
6	Website	9	php
7	E-Mail in	0	no extension
7	E-Mail in	10	msg

Part 4 - Table ENTITIES

The different entity types (Table 18) are defined within the table *ENTITIES*. They describe the different possible parameters of the content, the descriptive metadata, and some of the relationships introduces in section 4.2. Here, independently of the multiple usage of the same entity type within several event types, every type is defined only once. Moreover, the type of the entity (text, number, time/date, media) is stored in column *type*.

Table 18 : Table ENTITIES

entity_id	entity	type
1	Name	text
2	Title	text
3	Content	text
4	Description	text
5	Author	text
6	Keyword	text
7	Caller	text
8	Callee	text
9	Duration in Seconds	number
10	Size	number
11	Location	text
12	Screenshot	media
13	Sender	text
14	Receiver	text
15	Time	time/sate

Part 5 - Table EVENTTYPES_ENTITIES

Once defined in the table *ENTITIES*, the different entity types might be associated to several event types according to the table *EVENTTYPES*. This links are stored within the table *EVENTTYPES_ENTITIES* as showed exemplary in Table 19. For a better understanding, the two columns *FK_type* (foreign key of the type) and *FK_entity* (foreign key of the entity) have been inserted to visualise the actual value behind the columns *type_id* and *entity_id*.

Beside the link itself, the following classification is being made for every link:

- Every link contains the range, which defines, how many entries of these information type might be associated to the one event type. Each website for example has just one title, but could have several (n) keywords.

  Column: *range*

  Possible Values: one (1), many (n)

- Some of the entities are required to fully characterise and describe an event. Whereas they might be required for one event type, they still might be used optional only for other event types. They need therefore be marked within this relation table.

  Column: *required*

  Possible Values: yes (1), no (blank)

- The meaning and importance of central entities have been issued in the last chapter. They are mainly used to identify different versions of the same instances (e.g. documents). It is therefore required to mark the central entity for every event type.

If an entity should act as a central entity (as described in subsection 5.5.1) for the linked event type, a correspondent flag is required.

Column: *central*

Possible Values: yes (1), no (null)

Table 19 : Table TYPES_ENTITIES

type_id	FK_type	entity_id	FK_entity	range	required	central
1	Text	1	Name	1	1	1
1	Text	2	Title	1		
1	Text	3	Content	1		
1	Text	4	Description	1		
1	Text	5	Author	1		
1	Text	6	Keyword	n		
1	Text	10	Size	1	1	
1	Text	11	Location	1	1	
4	Phone Call	7	Caller	1	1	1
4	Phone Call	8	Callee	n	1	1
4	Phone Call	9	Duration in sec	1	1	
4	Phone Call	4	Description	1		
6	Website	2	Title	1		
6	Website	3	Content	1		
6	Website	4	Description	1		
6	Website	5	Author	n		
6	Website	6	Keyword	n		
6	Website	10	Size	1	1	

type_id	FK_type	entity_id	FK_entity	range	required	central
6	Website	11	Location	1	1	1
7	E-Mail in	2	Title	1	1	
7	E-Mail in	3	Content	1		
7	E-Mail in	5	Author	n	1	
7	E-Mail in	13	Sender	1	1	1
7	E-Mail in	14	Receiver	n	1	1
7	E-Mail in	10	Size	1		

Part 6 - Table UNITS

The actual instances of the event entities are stored within the table *UNITS*. Every unit is assigned to an event type (column: *type_id*) according to the entries in Table 19. For better understanding, the column *FK_type* represents the value behind the id. Moreover, every entry is characterised with a version (column: *version*), and the information when the event happened (column: *event_timestamp*), as well as when the even have been entered into the system (column: *entry_timestamp*). If available (depending on the event type), the physical files are stored in the column *physical_content*. Depending on the event type functionality (add-on as described in subsection 6.2.4), the actual file format may vary from the original one. For example, instead of saving the HTML code of a website, it might make more sense to save just a screenshot of it. Due to the fact that the content and metadata information is extracted and stored in a readable and searchable way anyway (within the table VALUES), the pure screenshot of the original website might be of higher importance to the user by having a higher recognition value of his previous knowledge work.

Table 20 represents an example of possible entries:

Table 20 : Table UNITS

unit_id	version	type_id	FK_type	event_timestamp	entry_timestamp	physical_content
1	1	1	Text Doc	2013-01-18-17-24-16	2013-01-18-17-24-16	1.DOCX
2	1	4	Phone Call	2013-01-18-18-12-16	2013-01-18-18-12-16	
3	1	6	Website	2013-01-19-12-13-17	2013-01-19-12-13-17	3.JPG
1	2	3	Text Doc	2013-01-19-12-13-21	2013-01-19-12-13-21	1.DOCX
4	1	5	Application	2013-01-19-14-12-23	2013-01-19-14-12-23	

Part 7 - Table VALUE

Within the table *VALUE* the concrete information of the type instances are stored. There may be several synonym values for the same information; therefore the structure allows linking these information together by setting the s*ynonym_to* column with the id of the linked entry. In the example presented in Table 21 the entry number six stands for the name *Witold Thaul*, the entries 12 and 13 on the other hand are *email address* and the *phone number* of the same person – therefore there is a direct link. For the entry WM, which stands for *Wissensmanagement*, the German term for Knowledge Management – there is a link set back to entry number 2 – KM.

Table 21 : Table VALUE

value_id	value	synonym_to
1	New approach on semantic desktop	
2	KM	
3	PKM	
4	semantic desktop	
5	WM	2
6	Witold Thaul	
7	Udo Bleimann	
8	520	
9	iknow.docx	
10	http://www.golem.de	
11	Golem News	
12	witold.thaul@plymouth.ac.uk	6
13	+491765507677	6
14	Nach nur wenigen Monaten bekommt ...	

Part 8 - Table UNITS_VALUE

Table UNITS_VALUE links the several units (Table 22) with the concrete information stored in the table *INFOVALUES*. In the columns *FK_entity* and *FK_value* also the values behind the columns *entity_id* and *value_id* are presented for a better understanding.

Table 22 : Table UNITS_VALUE

unit_id	v	entity_id	FK_entity	value_id	FK_value
1	1	1	Name	9	iknow.docx
1	1	5	Author	6	Witold Thaul
1	1	6	Keyword	2	KM
1	1	6	Keyword	3	PKM
2	1	7	Caller	6	Witold Thaul
2	1	8	Callee	7	Udo Bleimann
2	1	9	Duration in Seconds	8	520
3	1	1	Name	10	http://www.golem.de
3	1	2	Title	11	Golem News

Part 9 - Table UNITS_INFLUENCES

After assigning the level of influence, the value will be stored within the table *UNITS_INFLUENCES*. It links the two corresponding events, in the correct version and describes their coherency. The example in Table 23 represents the influence level 3 a phone call (the 12th one) between Witold Thaul and Udo Bleimann had on version two of the text document iknow.docx.

Table 23 : Table UNITS_INFLUENCES

unit_id_1	unit_version_1	unit_id_2	unit_version_2	influence
1	2	2	12	3

Appendix B – Interviews

Part 1 - Dr. Ian O'Keeffe

1	Thaul, W.:	Do you historicise your work when you for example write a paper for a
2		conference?
3	O'Keeffe, I.:	I would historicise it, yeah. That would be essentially.
4	Thaul, W.:	How do you do it? Do you use a system for this or do you do it manu-
5		ally by for example using a for example logical file system by entering
6		the date or a version?
7	O'Keeffe, I.:	Like for a paper it will generally be a mixed of both. It will be like us-
8		ing the build in versioning, for example in Word, it will track your
9		changes, but I tend then to have version points. Saving files with differ-
10		ent version names, as well as the internal of the software, giving snap-
11		shots of how you progress.
12	Thaul, W.:	And when we now think about not only of Word files or text files in
13		general, but also thinking of for example PDFs...
14	O'Keeffe, I.:	Yes, definitely then I will be using filenames, a system of versions and
15		dates within filenames.
16	Thaul, W.:	You have just said that beside the functionally of Microsoft Word you
17		create your own manual versions. When do you usually delete the old
18		versions? Do you delete them at all?
19	O'Keeffe, I.:	No, I do not tend to. I keep them all.
20	Thaul, W.:	And do you really use them? I mean you are now at the I-Know 2012,
21		maybe next year you think there was something important you would
22		like to catch up. In practice – do you really use your old versions, or is
23		it more just there.
24	O'Keeffe, I.:	I do not know if I use them, I definitely keep them. I am not sure if I
25		ever really got back to a previous iteration of something.
26	Thaul, W.:	Have you ever lost some information by for example overwriting, or
27		deleting old version. You have just mentioned you do not delete old
28		versions, but….
29	O'Keeffe, I.:	Yeah, it happened to me – that is why I keep versions.

30	Thaul, W.:	So you learned of it?
31	O'Keeffe, I.:	Yes.
32	Thaul, W.:	And what do you think of a independent layer which tracks your ver-
33		sions, and not just for word and text functionality, but in general.
34	O'Keeffe, I.:	I guess it is a little bit of a difference between tracking in the document,
35		and tracking the sources you are using. Like at the moment I use some-
36		thing like Mandelane, so keep all my bibliography in all my documents
37		and so on. But that gives me a big selection, and I never know which
38		was the one I used for that document. So having them tight to docu-
39		ments, to version, would be really powerful. Because it is always the
40		case, when you write a report or a paper: I saw it, it was really interest-
41		ing, but where was it? It is not always captured within the bibliography,
42		it is sometimes more a thing or a feeling, and you import it, it influences
43		your writing, it is not technically a reference. Therefore, it is not kept in
44		your Mandelane, you lose it.
45	Thaul, W.:	Interesting. Let us come to the topic: remembering why you did thing
46		the way you did them. How do you try to make sure you remember?
47		You have just said: putting everything into one folder.
48	O'Keeffe, I.:	Yes…
49	Thaul, W.:	So is it not working really well for you?
50	O'Keeffe, I.:	It works to an extent, you know you have it somewhere, and you can
51		search again for it. But it does not track the influences for the thing you
52		are producing, the artefact you are developing. You do not have this
53		kind of source history, you only have – at best – your bibliography in
54		that scientific context, but especially when you are not working in the
55		scientific context, if you are *just* a technical writer, having those sources
56		would be great. At the moment I would not, if I would write a quick re-
57		port, I would not write a bibliography, and then I would not be able to
58		track them, I would only be purely working on my own memory, I
59		would not have a set of files related specifically to this concrete docu-
60		ment.
61	Thaul, W.:	And do you think your memory will still be able to handle this looking
62		for example half a year back, or a full year back?

63	O'Keeffe, I.:	No, it is definitely not – you have a feeling, I know I referenced some-
64		thing, but usually you can never find it again.
65	Thaul, W.:	Great, another topic I would like to discuss is reproduction of work.
66		When we know what and why we have done something, we are able to
67		reproduce it again. We when now think of users using certain systems
68		which tight them and do not allow to export. Applications, which rec-
69		ognised what happened, would allow reproducing it in another system.
70		As an example, maybe you know flickr the photo sharing community, if
71		you use this and want to switch to a competitor like Google Picasa, all
72		the changes you have done within flickr (like Geotagging) cannot really
73		be exported and would be lost. Knowing what happened, you could re-
74		produce it within a different environment. Would this be helpful from
75		your point of view?
76	O'Keeffe, I.:	Yeah, you have a log, you can replay it essentially. Yes, not only in a
77		different environment, but even in the same context. It is something that
78		we do enough. If you think back to a Linux command line, you have the
79		facility that lets you record your sessions, and then you get an auto-
80		mated script, which you can run in the future. As you worked through
81		the problem, you were trying to solve the problem; you set the com-
82		mands, and generated this script. And what you are talking about is that
83		on a much larger scale. Even just for not only the reallocation where
84		you might want to replay it to import it into different application or run
85		it on a different system, but also just for processes which you repeat
86		over, and over again. Yes, so if you have it, it is basically a log of the
87		process, you can reply it in another context, or even use it as a learning
88		tool. There is a process you went through doing your searching and
89		finding your documents and doing these things and so on, if you can re-
90		play it, you can see how the influences went, which things were good
91		approaches, and which things were dead-ends and so on.
92	Thaul, W.:	Very interesting thoughts! Basically you covered all my questions
93		within this area. When we now come back to the implementation, I
94		have presented you before. In general, how much do you like the idea?
95		A system running in the background, maybe as an add-on for a seman-

96		tic desktop, maybe as a standalone application, and every time you save
97		your work, does not matter if it is a Word file, a LaTeX, PowerPoint,
98		PDF or whatever, there is the explained pop-up, or even assuming that
99		the automated assignment is good enough that there is even no need for
100		manual adjustments and popups, and in the end you will have some-
101		thing like this, basically an overview of all your knowledge parts pieces
102		and you can select one and for each of them you have an entire forma-
103		tion process as a result. Would this be interesting from your point of
104		view?
105	O'Keeffe, I.:	Yes, I think it has to be a background process that is automatically do-
106		ing it, because if it is manual – people are lazy, so they will not do it. I
107		guess you always have the problem, that people are feeling this kind of
108		big brother thing of the system tracking them, but if there is a real bene-
109		fit to them, they will quickly get used to it and if they can see an con-
110		crete outcome that is meaningful to them and helps them, and it has to
111		be in the background, and has to be looking at not only applications
112		opening and context switching, but then also at more detailed internal
113		activity, like in the browser example when you search for some things,
114		the system needs to be able to see the things you searched for, and the
115		things you collect.
116	Thaul, W.:	You have just mentioned *big brother*. This is an important issue. If
117		there was a very simple switch on / off bottom, which you could just
118		press while for example searching for a birthday present where you
119		would not like to be tracked, would that be fine with you, or do you see
120		any other privacy and security issues here?
121	O'Keeffe, I.:	I guess you are talking about a desktop application anyway, so it is
122		something which is under your control, you could uninstall it anyway.
123		As long as there is trust in the system that it is not pushing data to some
124		anonymous service somewhere, and your data is being shared or used in
125		some other way, then your data is totally under your own control. I
126		think then it is ok –but definitely a flag to switch in on and off or even
127		the possibility to assess the history and manually delete things, similar
128		to Amazons service where you can go in and delete some things you

129		have been searching for, do not use them again, they are not longer
130		relevant. Maybe that might be a useful feature.
131	Thaul, W.:	You have just said *as long as you have it under your control*. Could you
132		also imagine to use such a system and save the data in a cloud environ-
133		ment, to make mobile access and synchronisation easier – like the
134		phone calls or the emails you have written using your mobile device.
135	O'Keeffe, I.:	I guess the thing about having it under control is more about that the
136		tool is not part of some bigger service, essentially if it is in the cloud,
137		and it is your space, even if you use like Google for hosting, it is still
138		your space. It will be fine. It will definitely be powerful to move be-
139		tween applications, because there is always the idea, which goes back to
140		Google Wave thing, where you start your document on your laptop, but
141		then you move to your phone, and so on. So processes are diverse now,
142		mobile.
143	Thaul, W.:	Talking about the levels of influences, I just mentioned one algorithm
144		as an example, the statistical method of keyword analysis, where you
145		make vectors out of the contents, clear them by getting rid of all the ir-
146		relevant words, and then comparing the vectors. Do you have any other
147		ideas for appropriate algorithms, which could be used here?
148	O'Keeffe, I.:	I'm not an expert in this area, but anytime I had to do anything with that
149		I just resorted to frequency, it is an easy thing to do, well understood. I
150		guess you could go with some semantics, trying to find synonyms and
151		things like that. But you have to benefit this, because benefit on the cost
152		of additional complexity...
153	Thaul, W.:	Exactly... We have been just talking about the point that not everything
154		happens within the desktop environment, some things happen within a
155		mobile environment, on the go. There will be an API to connect the
156		events back together. Nevertheless, when I save something on my desk-
157		top now, there will appear a popup with the assigned events, which
158		happened in the meantime, but maybe without having the information
159		what happened on my mobile in the meantime. Do you think this could
160		be a problem?

161	O'Keeffe, I.:	I do not think so. I see the power of this being used in the medium to
162		long term; it is not about immediately what you are doing, so the fact
163		that there is an intermediate inconsistency between these two the user
164		might not even notice...
165	Thaul, W.:	Ok, great, coming to the end, do you think a system like that one would
166		be useful?
167	O'Keeffe, I.:	Yes, definitely.
168	Thaul, W.:	Something in particular? Do you miss something? You have already
169		mentioned the point of looking deeper into the applications. Any other
170		ideas / thoughts here?
171	O'Keeffe, I.:	If you think of the system of being completely automated, you will be
172		missing some inputs, and it would be like in e-learning, I was talking
173		about the idea of blended learning, a mixture of digital and class-room,
174		I think the same idea would apply here, because when you are working
175		on that sort of tasks, you talk to people in your office, you got an phone
176		call, and there might be something in there that you would like to cap-
177		ture, so a way of reflecting of the process, and capturing the information
178		that are coming from that side, from the real world - might be a benefit.
179		Maybe not for everybody, but it could be an interesting facility. Be-
180		cause it would allow you to capture manually only the important, sig-
181		nificant things.
182	Thaul, W.:	I was thinking a lot about this topic, and the idea was basically to allow
183		synchronisation of our calendars which contain many of these informa-
184		tion, nevertheless even if in my calendar for example is the I-Know
185		conference marked, this meeting both of us are just having is not men-
186		tioned and this information might be lost. And therefore the idea was to
187		allow manual inputs, still being not sure about how many will really use
188		it, because as you said, people are lazy.
189	O'Keeffe, I.:	Yes, I think if you provide this facility, it becomes more than just the
190		tracking software, it becomes a learning log, where you can record your
191		experience, and your thought process as well, and things like why was
192		that document important in contrast just to the fact that it was important.

193	Thaul, W.:	Yes, that is exactly what it is about: why is it important. If it was avail-
194		able, would you use it?
195	O'Keeffe, I.:	Yes, I would use it for work mostly. I think being able to assign infor-
196		mation to the sources, being able to look at a document and see the
197		phone call I had with some person, the documents I read, and link those
198		to the changes that I made in the documents, whatever it was I was do-
199		ing, it will just be extremely powerful – linking everything together, and
200		having the understanding of where the information came from. When
201		you come back and you write a similar document, you tend to go back
202		to the original, and you kind of repurpose bits of it, and if you can then,
203		as you repurpose those bits, then bring in all the source material you
204		had, you will be able to go back and read into them. Yes.

Part 2 - Prof. Dr. Andreas Peter Schmidt

1 2	Thaul, W.:	I have four areas I would like to talk to you about. The first one is the comprehensive historisation of work. Do you historicise your work?
3 4 5 6 7 8 9 10 11	Schmidt, A.P.:	On the one hand I use backups with histories; on the other I use Evernote, which has some implicitly history functions. But I have to admit that I do not go back to history version that often − if at all. And the other thing is that you usually doing this versioning with documents by adding your own initials in the document names, these typical patterns. And then we using quite a lot services like Google Docs, which has a build in versioning, but even here using it rather rarely, if there are some problems technically, if something disappears suddenly, because you have pasted something in and that did not work.
12 13	Thaul, W.:	So you are using it more for backup and not so much for the content itself.
14	Schmidt, A.P.:	More backup, really rarely for some other things.
15 16	Thaul, W.:	You have mentioned you are using these standard system / patterns? Do you mean a file system, and adding the date for example?
17 18 19 20 21 22 23 24 25 26	Schmidt, A.P.:	Well it depends. Most of my papers are collaborative, and we have different systems, one is with the date prefix at the beginning of the filename. Usually one is the master, and the others are contributing, so then you have the date, then what it is about, then everybody adds his initial to characterise who contributed to what. The other possibility is working with version numbers − version 1, version 2, version 3 − so that is the typical staff. Recently I can just observe, that a lot of these collaborative parts moved to Google Docs, and only for the very final version move back to Word to make it nice looking. Just to avoid this sending around, and messing up with different versions.
27 28	Thaul, W.:	Do you encounter any problems with the tools you use? You have mentioned Google Docs, or Evernote.
29 30 31	Schmidt, A.P.:	I think the main issue with Google Docs is the area of references, and the second problem refers to the styles that are for example required by some conferences. So that you always have that extra step to format it.

32	Thaul, W.:	And this is the reason why at the latest stage you again export it to Microsoft Word.
33		
34	Schmidt, A.P.:	Yes. Also also – LaTeX it is a good thing, but the problem is that I work a lot in an interdisciplinary area, and LaTeX is only for technical people. And that is a real problem – I did my PhD in LaTex and was happy with it, but now it is just Word. Everyone knows how to use Word... just this collaborative part...
35		
36		
37		
38		
39	Thaul, W.:	And do you delete your old version?
40	Schmidt, A.P.:	In most cases I keep them – I have a subfolder *old* – in some cases I try to keep the history in Evernote, send them around but keep them in Evernote – I tend to keep it, but not really using it too much – just in case, having it.
41		
42		
43		
44	Thaul, W.:	To summarise, how easy is it in practise for you to work with these version within the software tools which are available and you are using?
45		
46	Schmidt, A.P.:	I think each of it as such is easy, what sometimes is complicated is when you switch from one to another, and try to find out how was it in that tool, so that in Google Docs it looks differently, in Evernote it looks differently, in your own file system you have your own style, then you have different styles depend on to whom you are collaborating with – so the differences in that are more challenging. So that is a major problem.
47		
48		
49		
50		
51		
52		
53	Thaul, W.:	What do you think in general, without going too much into detail, of an additional system, running as a background service, and taking care of your versioning?
54		
55		
56	Schmidt, A.P.:	If it deals with the collaborative situation well, then this will be definitely helpful. I think also the other part that you have explained, I think that even would be more useful, tracing where things came from, because you sometimes forget, you do not take proper notes, and then you search again, and that happens quite often, that you have to search again, because you do not have a full reference, because you did it quickly, and in between....
57		
58		
59		
60		
61		
62		

63	Thaul, W.:	This brings me to my second main topic, remembering why we did
64		things we did them. How important is it from your perspective to re-
65		member why we decided to do it a special way.
66	Schmidt, A.P.:	I think that this is even more important than the versioning part, because
67		this really causes many problems, especially if you have many of these
68		content switches, which is a real pain. You tend to forget, sometimes
69		you are also in the situation asking yourself *well why have I decided to*
70		*do it that way?*, *have we decided it?* or *is it just by accident that it is*
71		*here* – and you try to remember, because you did not keep any minutes
72		on the last online call you had, or whatever – so that happens quit often.
73		Tracing back why things are as they are, any help, I think would be
74		great – because that is something that helps you just in case you trace
75		some aspects back to where they come from – I think that this would be
76		very useful. I do not know how frequently you would use it, but when
77		you need it, any alternative would be much, much more effort – and in
78		many cases it would be just like....I do not know.
79	Thaul, W.:	And what do you currently do to make sure you remember?
80	Schmidt, A.P.:	I try to keep notes in Evernote, for myself, if it is personal. If it is col-
81		laborative, usually we try to keep some minutes document or something
82		like that of the process, where just we also note down action items or
83		things like that. Sometimes the problem is, that everything happens un-
84		der time pressure and then you just quickly discuss, run to the next
85		meeting and forget about it. So these are the two main strategies, one is
86		the personal, the other is the collaborative, where we that then in
87		Google Docs, sometimes in a Wiki, if it is part of a project for example,
88		but usually it is more Google Docs at least in the last two years.
89	Thaul, W.:	Ok, and the link between the real content and these notes?
90	Schmidt, A.P.:	Depends – if it is personal a personal one, I embed these notes within
91		Evernote, and work with them within Evernote. So I keep them insight
92		the same note for example. Or I do also use Notelinks, If it is collabora-
93		tive, and it is Google Docs we create a folder where we keep that notes
94		and the material. So you will then have the paper, and then some dia-
95		grams, and the other separate items within the same folder.

96	Thaul, W.:	And does that work well for you – especially the link?
97	Schmidt, A.P.:	The problem is again multiple channels. Some things come via email,
98		some things are online, some things are just for yourself – and bringing
99		together each of them is a bit tricky. I try to put as much as possible into
100		Evernote, because that is a single place then, but that does not work to-
101		gether with Google Docs – of course. So I am not really happy – but I
102		do not know how to do it better. And I have also the impression that be-
103		cause we collaborate not more easily, we do it more often, so that any
104		solution that we find will just mean that you collaborate more which
105		will just increase the problems that you having – just in different areas –
106		but that is just a personal impression.
107	Thaul, W.:	I agree. Looking again at the reasons why we have done things the way
108		we have done them – remembering the history – what kind of informa-
109		tion should be considered from your perspective?
110	Schmidt, A.P.:	Especially this reasoning behind it. I mean if we think about paper writ-
111		ing, *why we have it in that order*, *why have we included that*. So it is
112		more the why part, that rarely gets in documents. Because you discuss
113		that, and think *yes, that makes sense*, then you move it around, and then
114		later on you could get the impression, that it could be better the other
115		way around, and you forget about why you did it that way in the first
116		place, you move it around again, and later discover *ah we discussed that*
117		*problem before* – and move it back. So that you have to repeat the same
118		decision process again and again, because you never document it. And
119		that happens actually quite often, that you forget about the reasons. Be-
120		cause this is also hard to write down. And you are happy that you found
121		it, and in that moment you do not think that you will need the justifica-
122		tion later on, because it is not so much about references that you need,
123		but it is more about *ah ok – that makes sense, because when we do it*
124		*otherwise that would be the problem.*
125	Thaul, W.:	Looking at the way you work currently– how easy is it for you recover
126		in such situations why you decided to go that certain way?
127	Schmidt, A.P.:	Hard. Usually we are having any kind of regular meetings, because
128		nearly all papers for example are collaborative ones, and then it is just

129		that I forget about *why did we do that*, and then we trying to collabora-
130		tively reconstruct the argument. Which takes some time, and is not nec-
131		essary. It is just like that – collaboratively reconstruction. Sometimes
132		you at the end discover something like *ah, I think we have talked about*
133		*this before*.
134	Thaul, W.:	Having said all this, what do you think in general of not only making
135		versions of your work, independently of the content, but maybe not
136		looking at the whys, because this will be difficult to capture, but at least
137		say that there was an event, there was a meeting, like this one here, that
138		let me come to that conclusion, that...
139	Schmidt, A.P.:	It might help to reconstruct – definitely. If it is an easy visualisation
140		how it went, then usually you think along that – *ah ok*, *ah that was then*,
141		*that came after that* – it may help to facilitate your memory. Because
142		this reconstruction is usually hard, because you simply have sometimes
143		the feeling, that there was a discussion about something, but maybe not,
144		thoughts like *have we tried it the other way around or not* – for things
145		like that it might be easier to reconstruct your memory, maybe – I could
146		guess, yes. But I would have to try out, so I am not sure, but I could see
147		a potential to aid your memory actually – if it is closed to how you
148		work, so if you consider paper writing as a story line for example, so it
149		should be like a story line, then it might be closer to how your brain
150		works – worth finding out!
151	Thaul, W.:	You have mentioned already that when you are using one tool, you usu-
152		ally have difficulties to use another tool at the same time. One of the
153		finding during my research was also, that it is often very difficult to
154		switch from one tool to another. Once you have your knowledge base
155		within system A, it is really difficult to switch to system B – even if the
156		new system B might be better. Do you think that the idea of reproduc-
157		tion, based on the knowledge what happened, could be an adequate al-
158		ternative for exporting from one and importing into another system –
159		where it is not possible?
160	Schmidt, A.P.:	Sure.

161	Thaul, W.:	Just to give you an example. When you are using one system, and
162		within this system you edit your stuff, make some comments, some
163		comments, add some meta information, which will not be included
164		while trying to export the knowledge base. But as long as you remem-
165		ber that there was an additional information I edit, and you can repro-
166		duce it in an automated way, would this be...
167	Schmidt, A.P.:	I think if the issue of moving to another system is mainly that you are
168		losing some details, yes. Because this way you will not lose it, it will
169		just be accessible via some other place or way. But I think usually mov-
170		ing means a lot more, because of the way you organise things might be
171		very different. So for example if you are moving from a tagging based
172		environment to a folder based environment, or the layout is broken, or
173		with graphics – this is really a pain, when you are moving from one sys-
174		tem to another and the result looks really different, and you can do dif-
175		ferent things – that will not be solved by that. That is the hard part of
176		moving, but I think it could help if it is about the history that you are
177		moving, so if it is about moving and then the other you start as if it was
178		the first version, then this could help, definitely. That is more this emo-
179		tional thing, that you could if you need to, if you would like to, but I
180		think most of the time you will not be using the history, it will always
181		be something for rather rare occasions, but I think in some cases not
182		having it will cost really a lot of time to reconstruct it, and therefore I
183		would guess that this will give you a feeling of assurance, it is there,
184		you do not have to worry about reconstructing it, you do not have to
185		worry about the history.
186	Thaul, W.:	When we now look at the implementation on a high level, the idea is to
187		present it on a timeline. Depending on the event type, we are talking
188		about, for example visiting a website and writing a document might
189		have a completely different time structure, whereas here the relevant
190		point in time might be when I visited the website. Here the relevant
191		point in time might be when I have saved the last version or something
192		like this. The idea is basically for each time to define a point of time, to
193		make it possible at least to present it somehow. Otherwise, I will always

194		have timeframes and this will make it even more complicated. Do you
195		think this is a sufficient solution? Or do you have any thoughts, any
196		ideas in this area?
197	Schmidt, A.P.:	It depends a bit what the notion of versioning there is. There are very
198		different practices, when it comes to Google Docs for example, it is a
199		constant update interval, when we are moving to Microsoft Word, can
200		be a daily or longer periods, when you create a new version. I think that
201		depends a bit on that. And it is also then the question what the activities
202		would actually be. I mean keep the note and then add something to the
203		note or the new version. So when is the new version actually?
204	Thaul, W.:	Exactly, do you have any ideas in this area?
205	Schmidt, A.P.:	Actually that is quite hard, because it depends a bit on the type of the
206		thing you are doing, also on the way you are doing it. So what I have
207		found in code for example, there is also this different strategy of having
208		frequent commits to SVN for example. Word does having rather these
209		release oriented commits. And both have their advantages and disad-
210		vantages, but create a very different history and if you navigate a his-
211		tory of frequent committing, then sometimes can be quite hard to find
212		the right thing. On the other side quite usually you have these big
213		changes and it is hard to track down why something was made. I think,
214		in general, it is very hard to say anything about that. It is rather that you
215		have to look at specific cases and the problem is that it will not only be
216		a technical solution, but it will be related to your own personal practice,
217		which has to be modified properly. And I can just observe how you
218		write papers has changed classically in the last year. Just because of the
219		tools that are available, collaboration, it is much easier now to collabo-
220		rate than it was before. So you are also going for more frequent itera-
221		tions then. So that changes. And that you also have on the fly live things
222		that you can change in real time. Like Google Docs which we did not
223		have before. So I think they are changed quite a lot and your practice
224		changed. And you are still struggling actually with it, to make it better,
225		because you think it could be better. And you are constantly changing
226		it. So it is not really a stable practice. So I think, there you have then a

227		mutual dependency between the technology and your own practice.
228		And that makes it hard to say what is the right thing, but it rather,
229		probably has to be something that can be adapted to your practice. So
230		you can adapt your practice, but you also can adapt how the tool reacts
231		to your practice. That you have this mutual dependency between the
232		two, that makes it hard. Appropriation would be the key word here, that
233		how you appropriate such a possibility, that that changes according to
234		your practice and changes your practice. And then you change back to
235		your appropriation. And I think that is the hard thing. If you have a look
236		for example how Evernote and its possibilities are used by different
237		people, radically different. Everybody has started like that, then do it
238		now like that.
239	Thaul, W.:	I agree, yes. And it is changing a lot, too. Every time there is a new fea-
240		ture coming...
241	Schmidt, A.P.:	You have to try to find out, does this make sense in my context. How
242		does it make sense? And then you see somebody else, trying it that way,
243		and you think good idea, try it out yourself and you change again. Sim-
244		ple things like tagging or putting into notebooks or whatever, so very
245		simple things. But still, you have not found really a stable way of doing
246		it.
247	Thaul, W.:	Ok, you have mentioned also the personnel behaviour, also talking
248		about major versions and so on. Some people – when we stay at the ex-
249		ample of text documents – are saving every 20 seconds just for security
250		reasons and behaviour. This will in the end let do...
251	Schmidt, A.P.:	It is like a Google Doc.
252	Thaul, W.:	And the idea to make it still reasonable and still applicable also on the
253		one hand for the user to visualize it, but also to apply this algorithms
254		here, the idea is to make like major versions...
255	Schmidt, A.P.:	Automatically?
256	Thaul, W.:	Automatically.
257	Schmidt, A.P.:	I think you can come up with major versions, because usually you will
258		observe patterns of high level of activity and then stability. So I think
259		that works. So I think that is a part I would be post confident about, be-

260		cause there is a clear activity caching usually. You are not working con-
261		tinuously like never. You have then your periods of activity and then
262		the stability of somebody else. I think you can find there easiliy pat-
263		terns, that are reasonably occurred. There are still some probably patho-
264		logical cases, where it will not really reflect reality, but I think in most
265		cases it is like that, because you then spend some time on it, to these
266		things and you can consider that as close activity which leads to a new
267		version.
268	Thaul, W.:	This was from the working perspective, from the saving perspective.
269		When we now look at the usage perspective. So later on, I access my
270		knowledge base. Does it make sense from your point of view to visual-
271		ize all these versions, even the major versions, I mean when we are
272		talking about the paper where you have been working a few months on
273		it, still can be 20 or 50 versions? Or, from your personnel point of view,
274		shall we concentrate just on the last few? The idea of the system was
275		just to represent the last up to five and then...
276	Schmidt, A.P.:	I think usually it is the recent things that you want to look at, it is not
277		the early ones. However, in some cases you might want to go back, be-
278		cause that was...So if it is in default a system of the most recent things
279		that were happened the last...I would not talk about versions, but rather
280		about a period of times. For example weeks or something like that. I
281		would guess, probably two weeks, most recent things and then the other
282		things would be going back, *we have that – I do not remember*. And
283		then you would explicitly go back and say, oh when was it, and trying
284		to find out, ok it could be like that. And that would be the tool actually
285		to try to find out that. So I would guess like that.
286	Thaul, W.:	The time period in which we consider potential inputs is the time period
287		between the version x and version x+1. This can be difficult, if this is
288		the start, if this is the very first version. Do you have any ideas, any
289		thoughts, how this can be handled? From which point of view should I
290		start observing regarding this version?
291	Schmidt, A.P.:	It also varies a lot, I think. In some cases you start with some concrete
292		documents already, very early, you collect there things. And in other

293		cases you just start discussing and then in some later point then you cre-
294		ate some artefacts. So I think it depends a bit. And I think it also de-
295		pends on the individuals. Some of them tend to create always a docu-
296		ment or something like that to put things in. Others tend to discuss first
297		to get a clear picture and then start. I think it is hard. It is also the ques-
298		tion, probably it is not a single document. Maybe that is a bit too sim-
299		plistic to consider a simple document as the point of reference, maybe it
300		is a collection. Because in many cases I have observed that you start
301		with one thing where you collect your thoughts. Then later on start to
302		create the paper and then you find out we need a main figure for it and
303		create a main figure. All of that is related to each other. And the calls
304		are not related to the diagram or the paper, but to all of them. So that
305		you need the notion of collection for it. Because the other thing would
306		be then too easy, or too much into *how do you text write*, so more some-
307		thing like a multiple thing. If you are moving beyond paper and are for
308		example considering proposal writing it is usually quite a number of
309		things. And quite complex patterns between them. Maybe it is also not a
310		good idea to concentrate on these individual pieces, but rather concen-
311		trate on the collection. Maybe then it is easier, because you also do not
312		think of not in...ah it was just a proposal...and then you had some..ah
313		there was this input coming in. And then maybe zoom in and say, ah ok,
314		that has changed, as a consequence out of that. Maybe that is more real-
315		istic, so how you work. It is not that you work in simple and single arte-
316		fact.
317	Thaul, W.:	When we think of the level of influences. I have presented on these
318		pictures seven different levels. How many do you think make sense?
319		What is important, very important, less important and so on. Any
320		thoughts on this?
321	Schmidt, A.P.:	No real thoughts. My good feeling is that seven is too much. You can-
322		not really distinguish, so that is a bit related to these fluid discussions
323		that if seven levels make sense or not, which we had quite long in one
324		of the projects where we came to the conclusion, probably for statistical
325		reason it is a good thing to have seven, but for the respondent it does

326		not make sense, which is a kind of paradox result of discussion. So I
327		would guess a maximum of 5 would actually make sense. And I could
328		also imagine that it is just 3. But 7, I would say...I do not know what the
329		difference between 6 and 7 would be.
330	Thaul, W.:	I completely agree.
331	Schmidt, A.P.:	But that is just a feeling. I do not know.
332	Thaul, W.:	OK then this determination of the level of influence. I have mentioned
333		this comparison of the content vectors. Do you have any other ideas
334		how this could be done?
335	Schmidt, A.P.:	The negative thing, I would say, comparing these vectors, I think it will
336		not capture reality adequately. Because in many things, this really im-
337		portant influence is not really connected to something similar, but is just
338		a good idea from something which might not be related at all, some-
339		how. And you just thought, ah that is it. And I think this is more dis-
340		similar than similar. So I would guess that similarity is not influence. I
341		would say influence is really related to how much has actually changed
342		in terms of content not really volume but quality. And that is hard to
343		capture actually. Because sometimes it is a small change, but it is a big
344		step forward. I think it is hard to tell what influence actually. I mean the
345		easy ones are with similarity, and there are definitely cases where you
346		can apply this, but then there are those other things. I specially think
347		know about figures. Well we have concentrated a lot on creating figures
348		to get the whole story. That is quite hard, there it is hard to see the simi-
349		larity, it is somewhere hidden. So it is more than really maybe the col-
350		location time might be closer.
351	Thaul, W.:	The time is considered anyway.
352	Schmidt, A.P.:	But it is hard to come up with an algorithms actually for the influence in
353		these cases.
354	Thaul, W.:	In these difficult cases do you think these manual adjustments or the
355		possibility of manual adjustments are sufficant?
356	Schmidt, A.P.:	Maybe if you are showing different categories, maybe if there is a con-
357		crete meeting, concrete phone call. So the call with X is not phone call?
358	Thaul, W.:	It is.

359	Schmidt, A.P.:	So for example not a category phone call, but it is an individual phone
360		call.
361	Thaul, W.:	It is a meeting between Professor....
362	Schmidt, A.P.:	So it is taken from the calendar, more or less. This meeting.
363	Thaul, W.:	Exactly.
364	Schmidt, A.P.:	I think that would be much more useful. Because that has given me the
365		idea and that has brought us a huge step forward.
366	Thaul, W.:	Everything is concrete.
367	Schmidt, A.P.:	Ok, I just thought it was a category.
368	Thaul, W.:	I can click on it and I can see the website.
369	Schmidt, A.P.:	I thought this bar char was about categorizing it into phone calls, meet-
370		ings, documents, because each of them was mentioned once.
371	Thaul, W.:	So we have talked about these manual adjustments – is this sufficient
372		from your point of view? Or do we need it at all or adjust this auto-
373		matic?
374	Schmidt, A.P.:	I mean the only thing is that manual would be useful to express some-
375		thing. It is just the point that, if you do not have time to keep actually
376		some minutes and some action items, because you are running from one
377		meeting to the next, this will not help either. Before you do that, you
378		would write down the action items, and not to reconstruct them. So that
379		is a bit the point. You need to find then the right moment, when does
380		this plays. Because otherwise you just click it away. Or the extreme
381		case, you uninstall it, because it pops up too frequently. So that would
382		be critical actually.
383	Thaul, W.:	And would you have any doubts having this data also stored in a cloud
384		environment which will allow basically also events which had more
385		used smartphone or maybe other computers or within Google Docs or
386		whatever, if they are encrypted?
387	Schmidt, A.P.:	I am a bit relaxed belong that. Too many things store in clouds or any,
388		to be worrying about it.
389	Thaul, W.:	We have talked a lot about the saving. When we now look at the pres-
390		entation: the aggregation of one and the same event to different ver-
391		sions, so in the end we have one entity of this and this paper. I can click

392		on it and I see all the versioning. Bringing it to the user in a table form,
393		in a sortable table form, in a searchable table form, and after selecting
394		which event I am looking at, present it in such a time chart. Is this a
395		presentation with which you could be happy with or do you have any
396		other ideas?
397	Schmidt, A.P.:	Sometimes you will do it like that. You are looking for the version of
398		that diagram, you are looking for a certain call, when was it, what was it
399		related to. If you can then from there go further to something that is re-
400		lated to that. Because sometimes your memory works like that. Ah it
401		was for that meeting and maybe it is not the meeting as such, but it is a
402		resource connected with the meeting. Or a change that you have made
403		at that meeting or whatever. So I think, that going according to those
404		categories would be one way, the other would be the timeline definitely
405		and then this chaining. So from there you go to something related. Yes,
406		I think these views make sense.
407	Thaul, W.:	To summarize, the suggested system implementation, as you have seen
408		it here on a high level – do you think this is useful or could be useful?
409	Schmidt, A.P.:	I mean, if you can solve the technical difficulties, yes, it could be defi-
410		nitely useful. And if it is really also about this usability part, so really
411		waiting for the right moment for that, because that is critical. I mean if
412		you have something popping up, then it will be just like get rid of it. But
413		if it is in the right moment, and if you then find out, that is useful actu-
414		ally to do so. If you can solve these two key aspects, then it could be
415		definitely useful.
416	Thaul, W.:	And do you miss from your point of view anything else, besides the
417		points you have just mentioned?
418	Schmidt, A.P.:	That would probably come when using it.
419	Thaul, W.:	Let us assume that all these problems you mentioned could be solved
420		and it would be available. Would you use it? And if yes, in which ar-
421		eas?
422	Schmidt, A.P.:	So, I would definitely try it out. And I think this paper writing, proposal
423		writing would be definitely an area where this would make sense.
424		Maybe lecture preparations, which are different, times cape. Maybe also

425		thesis supervision. That is also again different use case, which has different characteristics of interactions. So going back, what were my comments to the last version, what were the responses there. It is more like that. Which I now keep manually, but if it would be automatically, it would be great.
426		
427		
428		
429		
430	Thaul, W.:	And from the perspective of your students, could you imagine such a system as useful and practical, when you think of their tasks and their duties?
431		
432		
433	Schmidt, A.P.:	I mean they probably have similar things, like writing thesis – I think this is the much closest. In the project work, they have to be doing, that might make sense, if the collaborative perspective is represented. I do not know exactly about how they prepare for exams or something like that. So I am not sure how useful it would be, how relevant the artefacts are. But anything where they have to prepare some documents.
434		
435		
436		
437		
438		
439	Thaul, W.:	Then let us come to the last question. What would be from your point of view the largest benefit of such a system, which is currently just not available?
440		
441		
442	Schmidt, A.P.:	Traceability of how things came about actually. So it is not the version itself, I would say, but making traceable how the changes came about, help you to remember.
443		
444		
445	Thaul, W.:	Thank you very much.

Part 3 - Prof. Dr. Cosima Schmauch

1	Thaul, W.:	Do you yourself historicise your work?
2	Schmauch, C.:	No, I do not.
3	Thaul, W.:	Why not?
4	Schmauch, C.:	I do not know. I think I do not have the time or I am not organized that
5		way.
6	Thaul, W.:	Ok.
7	Schmauch, C.:	When I work with a text document, then I have several versions of
8		documents. Every time, if I think, this is a version, that I should store or
9		save this version, because I am not quite sure whether I want to go back
10		to this version.
11	Thaul, W.:	Ok and how do you typically do this? Some people for example are
12		using the date within the file name.
13	Schmauch, C.:	Yes, normally I use the date in the file name. I do not use something
14		like SVN (Comment: Apache Subversion tool). Our students have to
15		use this, if they implement applications. I would use it, if I implement
16		an application, but just for documents, I do not use it.
17	Thaul, W.:	And have you ever tried one of these tools? I mean what you have just
18		mentioned is a repository tool for software development, but there are
19		also tools, just addicted for example for text documents or when you are
20		doing your slides for your lectures for example. Have you tried ever us-
21		ing such assistance?
22	Schmauch, C.:	No.
23	Thaul, W.:	And you have just mentioned that you are saving different versions of a
24		document, if you think it is necessary or it is important. Is it just text
25		documents or in which area do you see the benefit of saving different
26		intermediate steps?
27	Schmauch, C.:	Yes, I think so. I think it is just different versions of documents. If I
28		know, I have, let us say... there are some chapter I have to work on, and
29		I have an idea and I bring this in, I write this down, and I am not quite
30		sure whether this is a good idea, then I save the version before and then
31		I work with the document and then I save this version.

32	Thaul, W.:	And does this really work for you? Do you always find what you are
33		looking for?
34	Schmauch, C.:	No, I do not think so, not in any case.
35	Thaul, W.:	And what do you think – very basically maybe, not going too much into
36		detail from their presented idea of the system but - of an idea of an in-
37		termediate layer, not Microsoft Word or PowerPoint, just a background
38		layer taking care of all this what you have just mentioned? What do you
39		think of this, in general?
40	Schmauch, C.:	I think that would be a big help, because I do not use a lot of support
41		when I write documents. But I know that I could need this. I need this
42		but I do not have it. I do not have the time to choose a system. So if
43		there were a system, I think I would use it.
44	Thaul, W.:	Ok, sounds good.
45	Schmauch, C.:	And I am a computer scientist and I like the idea of not controlling but
46		monitoring actions of users to help them, to support them. I teach moni-
47		tor based systems and the software agent is a part of this area and I like
48		the idea to have software agents who help the users and this could fit. It
49		is similar to your system.
50	Thaul, W.:	Sounds good. These were more or less the questions about the version-
51		ing itself, but what I showed you also on the other slides was the idea of
52		not just versioning itself but also having a look at why we did things the
53		way we did them or why we decided to do them this way. How impor-
54		tant is it from your point of view, from your experience also in your
55		work here, to remember why we decided to go this way and this way?
56	Schmauch, C.:	Yes, I think it is important. If I write a paper and develop some ideas
57		and this is also the topic, which I discuss with my students when they
58		write their thesis. So I ask them to write down every little influence, the
59		chapter of a book they had read or any conversation with the supervi-
60		sors, something like this which had/has influence on their writing or the
61		chapter or something like this. Yes and I think this is important and is a
62		good idea to have this information in addition to the text.
63	Thaul, W.:	How do you handle this with your own work? What do you do to make
64		sure you remember?

65	Schmauch, C.:	I do not do very much, I think. I have some kind of second document
66		where I collect such things. Where I collect information on websites
67		and additional information what I just had read there. I do not think that
68		I write...or some ideas coming of a conversation or something like that.
69		Just in a -not formally- in a separate document, I have just some notes.
70	Thaul, W.:	Does this always work? I mean, you are working now maybe on one
71		lecture document, maybe on a paper, maybe on some other stuff. One
72		month later, two month later...
73	Schmauch, C.:	No, it does not work. Especially I do not have the direct relation be-
74		tween the part of a chapter and the notes. I just have the collection of
75		the notes; I do not have the direct link.
76	Thaul, W.:	When we see this whole academic area in which you are, which kind of
77		information do you think are worth being considered? We have men-
78		tioned already text documents or maybe some presentations or meetings
79		– is there anything else where you think this is really important as well
80		and which is maybe not mentioned so far?
81	Schmauch, C.:	When we make a decision, I think the conversation or discussion is
82		really important too. So it is about discussions, then we have email con-
83		versations, documents and, I do not know, our thoughts.
84	Thaul, W.:	It is the most difficult one?
85	Schmauch, C.:	Yes, also when we are at home and we think about decisions we have to
86		make.
87	Thaul, W.:	This is definitely difficult to capture.
88	Schmauch, C.:	Yes, I think you just have to speak in a recorder.
89	Thaul, W.:	Yes, exactly. You have mentioned that if you are doing something like
90		this, you also prepare a document with some additional notes. So in the
91		end you have the whole work and you have some additional notes
92		documents. There are also systems available for this purpose – you
93		never tried one?
94	Schmauch, C.:	No.
95	Thaul, W.:	Why not? You have already mentioned a lack of time – is this the rea-
96		son?

97	Schmauch, C.:	Yes, I do not have time to find out which systems are available. I think
98		this is the main reason.
99	Thaul, W.:	In general, what do you think of the idea not just to track the versions,
100		independently of the system we are currently in, but also track the in-
101		puts / the influences, the main idea?
102	Schmauch, C.:	I think it is a very nice idea.
103	Thaul, W.:	One of the problems which I discovered during my research why this
104		came up was also the problem of reproduction of work. When we are
105		using a system, it is very time consuming to start using it. You need to
106		fill it with your knowledge, with information and that is work. At the
107		same time most of the software vendors are very egoistic, they do not
108		want the users let go. Therefore, it is very often, that it is not possible to
109		export the way it should be. When I am working on a system, include
110		maybe some additional value, some additional knowledge, some meta
111		information for example, I cannot exit very easily. The idea behind this
112		is also, that when I know what I have a different version and when I
113		also know what I did basically to come to this conclusion, I will be able
114		maybe not to export my knowledge base from one solution and import
115		it into another, but I will be able to reproduce it within another envi-
116		ronment, in an idealistic way at least. Do you think that this is a good
117		idea and this could work? Do you see it maybe as a problem, in gen-
118		eral?
119	Schmauch, C.:	So the different environment would be like? I do not understand.
120	Thaul, W.:	Let us maybe take maybe an abstract example of photo community.
121		When I am having my photos at flickr, I can add some tags, some geo-
122		location tags and so on. And then I decide maybe half a year later I
123		would like to have them within Picasa, because they have now new fea-
124		tures like face recognition and so on. So flickr does allow me to export.
125		But when I export, I lose many information like geo information, like
126		the name, like the description, many of the meta information. They are
127		doing this in purpose and they are doing this to let me stay there. And
128		the idea is, when I know what happened, I can reproduce it in the new
129		system. Do you think in general, maybe not just looking at this photo

130		example, this could be an interesting approach to become more vendor
131		independent?
132	Schmauch, C.:	Sure, I think that many people want to have a feature like *I take my data*
133		*and put it in another system to use them there*. It could be an advantage
134		of the system. If you see that you have all the emails which were impor-
135		tant for this paper development and you have them all together in your
136		system and you can take them out. Yes, sure.
137	Thaul, W.:	Now more specific to the suggested implementation. Here you see the
138		time line and basically we always save at a certain point of time but
139		some events like for example a phone call has a beginning and a start.
140		And there I occur some difficulties, so basically what is the relevant
141		point. If a phone call for example starts here and ends there – is it still
142		important, is it not important. Do you think that taking one special mo-
143		ment like for example the beginning of the phone call or -when we
144		think of text documents – the saving time of the text document is the
145		right approach to do it? Or do you have any thoughts, any ideas how
146		this could be handled in a different way?
147	Schmauch, C.:	So it would not be the saving of a document, but if I open a document
148		and read the document, which could be relevant for my own paper I am
149		writing...I think it is the beginning of the event that could be of interest
150		and there could also be something like monitoring how long the docu-
151		ment is open or how long the call is going on. I think calls are not going
152		very long, but if I open a document and do not read the document, so
153		the user has to know, that if a document is open, it is relevant, as soon
154		as I close it, it becomes irrelevant....there might be some adaption of the
155		user's behaviour.
156	Thaul, W.:	These are exactly the problems I do occur here. I see basically three
157		different times – the start, the end as well as maybe in some cases the
158		saving, which could be still in the middle. When I look for example at a
159		PDF document or website, for me it is pretty obvious, that it is the be-
160		ginning, because when I am surfing I am going to the website and in
161		most cases at least this will be the time when I open it, I will also read
162		it. But when I think of a text document it is more complicated, because I

163		can open it and it can stay open for the whole day. So which time frame
164		to take? I also agree with you that the user behaviour is very important
165		here and it can vary from one user to another. The further idea behind
166		the implementation is to let the user decide for every event time what
167		time to take.
168	Schmauch, C.:	But you could also think about the length. You could tell the user that if
169		you have open a document very long, then the relevance is higher than
170		if you close it very fast, because if this document is not relevant, then I
171		close it.
172	Thaul, W.:	That is interesting. There is the possibility to check which application is
173		currently active on the screen. So this could maybe also go together
174		with your suggestion.
175		Another problem I was also facing was that some users are saving every
176		few minutes. So we have hundreds and thousands of versions. How I
177		would like to face this problem would be making major versions, which
178		still will save every single version, so once you click on it, you will see
179		the individual smaller version, but on the first view you see the major
180		versions only. Do you think this will be ok?
181	Schmauch, C.:	Yes, some users have fear that the system crashes and they do not have
182		the version. Yes, I think that there should be major versions with the
183		major changes and small versions...some users do not change anything
184		and then save again. But this is not a question of space; I think, it is just
185		a question of the number of versions and the importance.
186	Thaul, W.:	Exactly, and from what you have mentioned, when I have 200 versions
187		and I present all of them here, it makes no sense any more. Also the
188		other events, which take influence on the version - if you have a version
189		every 2 minutes, there will be basically no possibility to assign other
190		events as influences, but when I take at least a timestamp of half an
191		hour between two versions, then everything which happened between,
192		can be seen as a real input, as something with a real influence.
193		This lets me come to the question. Even if there are more than 20 ver-
194		sions of it, I believe it makes sense to present just the latest 5 or 7 ver-
195		sions and still give the user the possibility to go back if he wants but in

196		the end it is all about usability and about making it simple and giving
197		the user a good overview. Do you think that make sense?
198	Schmauch, C.:	Yes, I think so, I think that the latest events will be much more impor-
199		tant than the earlier events. Yes, just show the latest five version or so.
200		And if you have many versions without events in between than you
201		could just overleap them.
202	Thaul, W.:	Now a more technical question. Maybe you have an idea. When we are
203		having something like this, so I basically have a version 2, and I am
204		working on this and save it, it becomes version 3. The timeframe for the
205		other events I need to look at is pretty simple, because these are all the
206		events which happened between version 2 and version 2. But where it is
207		difficult, is at the beginning. When I have the first version, so when ba-
208		sically do I start my observation? I was investigation a lot in this area,
209		but did not really find any solution. My pretty pragmatic approach will
210		be to make a fix time frame, maybe two hours, 3 hours, maybe a time
211		frame which can the user adjust, because I do not really know when it
212		was beginning. What do you think of this idea? Or do you maybe have
213		another one?
214	Schmauch, C.:	So, lets say the version 1 is saved. But do you know the time when the
215		document is created?
216	Thaul, W.:	Yes, I have this information as well. So basically when I have this and I
217		am working on the new version and I save the new version, all these
218		events are relevant. This is pretty easy. But when we.just imagine this is
219		version 1 this becomes a bit tricky. This time frame here is difficult to
220		define. I know when it was opened, so this could also be for example
221		when this document was started or when it was saved. So these were
222		maybe the two parameters to take.
223	Schmauch, C.:	I think I would take this time. The user might have developed some
224		ideas before, but I think if he or she decides to write the paper now, and
225		then open or start a document ...But if there are several other documents
226		with some ideas, then he or she could read them again, so these were
227		events then.
228	Thaul, W.:	Exactly.

229	Schmauch, C.:	Yes, it could be the beginning of writing.
230 231 232 233 234 235	Thaul, W.:	Yes, ok. When we look at the presentation of the saved results, how many levels of influences should be displayed. From the point of view of the algorism, I can make as many as I want, but from your user perspective or your personal perspective, what do you think would be an appropriate number to make it reasonable, because when I make 10, it will be a very big picture and...
236 237	Schmauch, C.:	Yes, it would be very difficult to find out whether there is some event having more influence than another or why.
238	Thaul, W.:	Exactly.
239	Schmauch, C.:	So, why did you choose seven?
240 241 242 243 244 245	Thaul, W.:	I have talked to several people and the answer her was always something between 3, so basically very important, important, still important and ten. And in the end I had a very short...I asked some students to make some questions and seven was the median number, so in the end I decided for 7. Do you think 7 is too much and maybe 3 is enough? Do you have an idea?
246 247 248 249 250 251 252 253	Schmauch, C.:	I think 7 is a lot. Well you have a lot of possibilities to differentiate, but I am not quite sure, whether you need this. I know that 7 is a number which is stated if you talk about what people can differentiate in their short term memory. In many cases you take the maximum of 7, because it is too difficult to differentiate between more than 7 items. But I think you could take 5. Because if you see this, then you have to think about why this document 1 is more relevant than the PowerPoint. So not more than 7.
254 255 256 257 258 259	Thaul, W.:	Interesting notice with the short term memory. Maybe a more specific question from your backgrounds: I have mentioned that I see the need for a mixture of different algorithms for the assignment of the level of influences. The one I have mentioned was this statistical method of keyword analysis – which is basically comparing these two content vectors. Do you have any other idea how this could be done?
260	Schmauch, C.:	So on the base of text comparison?

261	Thaul, W.:	Not necessary. In the end I have one event and another one, and I want
262		to compare them in any way to make sure how important for example
263		this email was. Until now, I came up with solutions, which are content
264		comparing. Do you have maybe other ideas how this could be done?
265	Schmauch, C.:	As you mentioned before that if I can monitor which application is in
266		front of the screen, so which application is active. Then I follow from
267		this the person is reading a document and the length of the activation of
268		an application or of a document.
269	Thaul, W.:	Not everything is happening within our desktop environment, some-
270		thing is happening here (comment: pointing to the timeline picture),
271		something is happening face to face or on the phone on the way home...
272		Even if I import these events from the smartphone, from the phone,
273		from my online calendar whatsoever, there is a time lack between it.
274		When at this point I import a phone call which was here (comment:
275		pointing to the timeline picture), the system will still correct all these
276		other events, so this will be done. But nevertheless for this time frame
277		while it was not imported, there is still something missing. Do you think
278		that is a problem?
279	Schmauch, C.:	The User is responsible for importing the information, so if he or she
280		does not have had imported information and looks at the last version or
281		looks at the events and does not see events, which are not imported, it is
282		his or her responsibility.
283	Thaul, W.:	More or less the last area regarding the implementation. I think it would
284		make sense to store the database online in a cloud environment, so you
285		also have access from different devices. Talking to many people about
286		this, I have always heard these policy issues, privacy issues. What is
287		your opinion about this? Would this be a problem?
288	Schmauch, C.:	I do not think so. You can encrypt the information; you can transport it
289		in encrypted way. I think there is a lot of data in the cloud and also data
290		which should not be seen by other people.
291	Thaul, W.:	As I really see the benefits, connecting different device types and hav-
292		ing the information on the go like for example here. From the graphical
293		presentation point of view the idea would be to have a sortable and

294		searchable table. Every event will be presented aggregated and there is
295		also information about the number of versions which exist. When I
296		click on the event type, something like this or similar will open. Do you
297		think this is a wise presentation of this or do you have any other ideas or
298		suggestions how this could be done better or more sufficient?
299	Schmauch, C.:	I like this graph figure with these icons. This is very nice. So the table
300		contains these events?
301	Thaul, W.:	Exactly, basically the table is the overview of all my knowledge base. I
302		have for example here my PHD theses, 75 versions of it, and when I
303		click on this, I see the latest version and so on. I can also click on this,
304		then I will see all the other ones. So for every event within the table, I
305		can click and get this. But the first search and the first access always
306		happens here (note: referring to the table). From here on I can click for
307		example on this document and then it will open exactly the same graph
308		for this document. This one will be done there and so on. Lets imagine
309		this PDF has a six influence on this version, but maybe it has a 7 or 8 on
310		this or maybe it is not important at all for this one. It always depends on
311		the point of view.
312	Schmauch, C.:	And the starting point is out of this table.
313	Thaul, W.:	Exactly. Is this ok or any other suggestions?
314	Schmauch, C.:	I think I would rearrange the columns. I am not sure if it is important,
315		whether this is a text. So if I have events and I see.the PHD thesis. This
316		information is the most important, because I do not look for text, I look
317		for special events.
318	Thaul, W.:	Ok, good point. But in general? Sortable and searchable?
319	Schmauch, C.:	Yes, ok. Great.
320	Thaul, W.:	Do you think such a system could be useful?
321	Schmauch, C.:	I would buy it.
322	Thaul, W.:	Sounds fantastic. Do you miss anything?
323	Schmauch, C.:	I do not. I have to think about it.
324	Thaul, W.:	If you have any ideas, I am more than happy, if you send me them by
325		email. You have already mentioned that you would buy it. And would

326		you use it only in your professional environment or could you imagine
327		also using it in your private environment as well?
328	Schmauch, C.:	I do not think for the private environment. Most times, when I sit in
329		front of the computer it is for professional environment.
330	Thaul, W.:	What would be the biggest benefit for you, if you would have such a
331		system observing you and helping you?
332	Schmauch, C.:	I think it would be the information, if I have a look later on a text
333		document or a paper, that I wanted to rewrite or something like this,
334		then it would be very helpful to see all the additional information which
335		was relevant when I produced the paper. It is something like a memory.
336		I forget a lot of things and a lot of paper I had read in a different envi-
337		ronment. So if I see, there is a paper I had looked at this time or for this
338		purpose, I could use it for another paper too.
339	Thaul, W.:	This was basically all from my questions. Thank you very much for
340		your time and I hope it was also interesting for you.

Part 4 - Monica Morrison

1	Thaul, W.:	I would like to start with some questions about your way of historisa-
2		tion of your work. Do you historicise your work?
3	Morrison, M.:	I try. I am a records manager ever. I am a librarian by training, but I am
4		a records manager by default. So I work a lot with records systems, and
5		I try all these tricky thinking ways to be the records management guru,
6		so I rely a lot on the traditional records management model, so a file
7		classification system, and folders, and categorisation of documents, so I
8		am really document oriented. I have never ever considered capturing
9		phone calls, except to take notes when I make a phone call. I started us-
10		ing Skype for the last type, last year, and I was very happy to see, at one
11		point that I could save, that even without a good bandwidth, so it is
12		sometimes easier to type to each other, rather than to talk – so I was
13		happy to see that I could save a substantive exchange. Then I lost the
14		ability to do it... I am struggling with it a lot, yes, I am experimenting a
15		lot, with different ways to do it, but my basic approach is the records
16		management approach, which is to document something that happened
17		in the course of developing a product or a project, and make sure that it
18		all goes to a file, either electronic, or paper.
19	Thaul, W.:	And do you use systems which support you with this? Or do you prefer
20		to work with a logical system by using for example the date or the ver-
21		sion number within the filename, or they store some meta information
22		within the documents. How do you do it?
23	Morrison, M.:	What I find is that a lot of tools disappear. They come and then they go,
24		they just disappear. There are just not enough people using them. Way-
25		back, one of the earlier versions of Microsoft Word had versioning
26		build it, and it was fantastic! And I used it most to teach to people, be-
27		cause it is a huge, one of the biggest problems, is tracking the versions
28		of documents, it just gone away, so then you had to go back to the old
29		fashion system, if you wanted to track a version, you had to basically
30		keep it as an additional filename, and in fact my last job I was in, that
31		was what my boss was doing very carefully, every time we worked to-

32		gether, we tracked changes, and then saved the document with a new
33		extension in the filename, saying that you see version 1, you see version
34		2 – that drives me crazy, but at least we can do it. But that is a big prob-
35		lem. Maybe I better answer in relation to a specific example. Maybe
36		you can phrase the question in a different way.
37	Thaul, W.:	From what I have understood, you used some system for this, but they
38		dissapeard, so they are not constantly available, you are also using – or
39		let us call it are forced to use by your boss for example – to use these
40		logical systems, within filenames, extensions, versions – but do you
41		think that there are currently good enough systems available on the
42		market that really support you, and not just for word processing, but
43		also for other tasks like presentations, calculations, or other tasks?
44	Morrison, M.:	No, there are not. There are tools for everything, but to get tools that
45		you can use consistently, and change and train people to also use them,
46		is difficult. I worked about eight, nine years ago with a SharePoint in-
47		stallation, and investigated the collaborative facilities of that tool, and at
48		every time I was in the business to get other people to use it, and it was
49		just so difficult, it was too complicated for people, they just wanted to
50		work with their email and attachments, and track changes... I keep com-
51		ing back to the whole issue of creating documents, because that is a lot
52		of what I do, and when we look at telephone calls, who calls an office,
53		who contacts us, it is very important for an organisation, even within an
54		organisation, and in the old days we had these wonderful telephone
55		message pads, that had a carbine, so the secretary answers the phone,
56		she takes the message, she writes it on a little paper, rips the paper off,
57		and puts it on your desk. But there is that carbine, so suddenly you have
58		this wonderful record keeping system, of who was contacting the or-
59		ganisation, and the timeline as well, so that was a very simple thing, but
60		it was very useful, and now that we do not have secretaries, sometimes
61		no receptions, and telephone lines are going directly to peoples' desks,
62		we lose a lot of that corporate memory – this is when you were talking
63		to me, I had to realise, that my business is trying to get corporate mem-
64		ory into the organisational memory, to a point where people can reuse

65		any purpose information. A lot of people are quite happy with more and
66		more independent working, they are lone ranges, but my work is trying
67		to get those lone ranges to share what they have got. I am very con-
68		scious about that whole issue, I am probably not a typical person to in-
69		terview, because I thought about it a lot, and am constantly struggling to
70		think of ways to that we can do this better.
71	Thaul, W.:	That is really interesting; I like your view. You have answered most of
72		my question of this area of historisation. I would like to come to the
73		topic of overwriting or loosing information. If might happen from time
74		to time, that you just loose something, overwrite it, maybe not you, but
75		your colleagues. How big do you think, from your personal perspective
76		and from your experience in this field, can the impact be? Do you think
77		it is possible to measure that impact – for example in the time need to
78		recover the information, or in money?
79	Morrison, M.:	This is a very sensitive issue, because where I come from, it happens a
80		lot. Because even when people do have IT departments, there is often
81		no clear definition of responsibility for backups and for preservation of
82		information – people still work on their own PCs, save stuff on their
83		own hard drives, this is much more in the part of the world where I am
84		working, because networks do not work well, and there is no one to
85		maintain a server or whatever, so we lose a lot of stuff, and I think it is a
86		huge, a huge cost, and I wish there would be more studies done on re-
87		turn on investment here, and on cost of lost of memory, because it is a
88		hard sell, records management is a hard, hard sell – and again, you are
89		interested in a tool that supports an individual worker, but you know I
90		see that if you really are an individual worker, then your information is
91		your own, then it is your problem, and you can find ways to deal with it,
92		but as soon as you enter an institution, or an organisation where you
93		work with other people, capturing what you do is really important for
94		the whole organisation – the last job I had, within the first three months
95		I set up a common document storage on a server, and for emails, so that
96		we can store it centrally and perhaps one day use it again – and one day
97		IT support people came in and took everything of the servers, it took me

98		literary three months to get everything organised, and it just disap-
99		peared, and they said *oh, but we have asked the officer administrator –*
100		*and he said that there was nothing on in –* so a lot has to do with man-
101		agement and good management, so I know when it comes to organisa-
102		tions it is often not really a tool problem, it is a management problem,
103		and it is the same with individuals, it is a management problem, a sys-
104		tem like this can capture automatically a lot of the stuff you do, just by
105		tracking it, that is great... You know, I move around, and I change jobs
106		a lot, the average time for a job is two years, the last one was three, that
107		before was four – very long-term jobs for me, and I move from country
108		to country, and I am an Canadian working at Africa, being living in
109		Europe and North America – I have boxes of stuff! What I am talking
110		about, thousands and millions of documents, and you will never go and
111		look at them again, but you can not give them up, because when I go
112		back once, and look at them, I think *that was a really good idea...*
113	Thaul, W.:	Very interesting... You have already mentioned, that in general a system
114		that logs what you are doing, and not doing it just for documents, but
115		also observing your phone calls, observing your other activities, putting
116		them all this into a relation, adding some sense, putting it on a time-
117		line... would this help you?
118	Morrison, M.:	Oh, it would help- definitely. It is just scary. It is scary because then
119		there is the potential for someone to abuse this information, and looking
120		back to the keynote speaker on the first day, it is being done in a way,
121		statistics, how am I doing, am I exercising enough, am I achieving
122		enough, am I productive – I can see this as a very good learning tool,
123		and a self improvement tool, but as soon as you digitise anything, then
124		it is available, even if you protect it with security measures, nothing is
125		perfect, so... I would try it, because as I said, the problems are there,
126		very much in my easy all the time, but I would worry about it
127	Thaul, W.:	Worry about privacy and security issues mainly?
128	Morrison, M.:	Yes.
129	Thaul, W.:	You have mentioned that it would be not only interesting to look at it
130		not just from a personal perspective, but also from an organisational

131		perspective. But having a system observing and logging what you are
132		doing from a company perspective could be very dangerous, defi-
133		nitely...
134	Morrison, M.:	Yes.
135	Thaul, W.:	But as long as you are doing it for yourself, as long as you have the
136		possibility to switch it on and off, and configure for yourself which kind
137		of events you like to be observed – because everyone is different, what
138		is important for me might be not important for you, maybe for security
139		issues, maybe for some other reasons. And that is also a reason not to
140		design it as a top down solution, but a bottom up one, and let the user
141		decide what should be looked at, maybe on a encrypted USB stick..
142	Morrison, M.:	The last job I was on, was a diplomatic world. And the principal in ths
143		world is that you are behind the scenes, and it never gets recorded. TO
144		achieve some of the things you need to achieve in a diplomatic envi-
145		ronment, they just cannot be recorded, they cannot be even references.
146		That does not point to my work very much because I am a technical
147		person, administrative, but for my boss for example, although he
148		worked a lot via e-mail, a lot of the negotiation and giving things done,
149		was done through conversations that probably will never ever been cap-
150		tured. It is a pity, because that is how things really happen, how they
151		work, and if you knew how they really worked, then maybe everybody
152		can become better – but you know, in those kind of environments face
153		saving is a huge thing, so I just think of beyond my personal use, and I
154		am kind of a technical person, I have very little to hide, but the one
155		thing that worries me, it that if I see that happening, I would be actually
156		using it as a performance measurement, soft performance measurement,
157		I would be afraid to see how inefficient I was – *how many times did I*
158		*visited that website?*
159	Thaul, W.:	Interesting approach – self performance measurement. Many people I
160		talked to had some doubts that their boss might use this as a perform-
161		ance measurement. But self performance measurement is another ap-
162		proach...
163	Morrison, M.:	Can really tiny your self esteem...

164	Thaul, W.:	But could also improve your way of working?
165	Morrison, M.:	Yes definitely. So you go back to these boxes. and find out things like
166		*did I say that? – that was really smart.*
167	Thaul, W.:	One aspect I identified, when looking at all these issues was that when
168		we are using systems, as you have also mentioned, they are disappear-
169		ing for example or changing, so new systems are coming. The way of
170		switching from one system to another is not always easy. The reason for
171		that is mostly, that the vendors do not really want the users to switch,
172		they want them to keep there. Just maybe an easy example. I am using
173		flickr for photo sharing. All of my photos are online. Most of them pri-
174		vately. Now if I would like to switch to Picasa for example, it is nearly
175		not possible switching everything, because all the stuff I have done
176		within flickr, it will be lost. The idea behind something like this would
177		be to track the changes, so reproduction of work would be easier and
178		maybe even possible in an automatic way. When we think maybe a step
179		further. Not just this, but a step further.
180	Morrison, M.:	With your step, you should be able to move it, right?
181	Thaul, W.:	Exactly.
182	Morrison, M.:	The problem is this: if you get something for free, so is it really yours?
183	Thaul, W.:	Yes, definitely. But you know, we will head away by recording what I
184		have done for example added some tags or added some geo location
185		tags or changed the description or whatsoever. Just staying with this
186		picture example. Would make it possible for me to do the same work
187		somewhere else again. Maybe even in an automated way. So it will al-
188		low switching of applications of new systems, if I would like to, or if
189		they would disappear, easier. Do you think this is true what I am saying,
190		could you image this could work?
191	Morrison, M.:	That is desirable, it is desirable. It means that the whole issue of crea-
192		tion, you know, the issue of building stores of information, stores that
193		are collections – the staff that you have worked with,. and the staff that
194		you think is good, and also looking on networking, and what value does
195		that have to your activity, is it better to abandon it, or anyway, and start
196		again – but the premise of your work probably is much simpler, this

197		way you are saying that and what is valuable, you just need to look at
198		your thinking process, that gathering process, so the big answer is yes, I
199		think it is important.
200	Thaul, W.:	When we now come to this concrete implementation. The idea is that
201		we always look at the time frame between two versions of our work.
202		And when I am saving this version, I am looking back to the time frame
203		which happened between these two versions observing everything
204		which happened and weighting their levels of influences of all these
205		events. Do you think that the way of observing this time frame is the
206		right one or maybe we should go another way, for example extend it?
207	Morrison, M.:	Make a bigger gap. No no, I think this is very useful, because you see
208		here the major different, substantive content of version one and version
209		two is based on this founded source, and this source is mainly responsi-
210		ble for the production of version two, and that tracks the thinking proc-
211		ess, the establishing process, I think it is very good. So it is a lot of de-
212		tails, a lot of staff to capture, but if you tell me the machine can do this,
213		it is ok...
214	Thaul, W.:	OK, sounds very good. And you have just mentioned, it is a lot of cap-
215		ture - it depends also how people work. Some people save every two
216		minutes. There might be a problem that we will have too many ver-
217		sions. The idea behind this is to let the software aggregate some ver-
218		sions to make a major version.
219	Morrison, M.:	It will be even practically if you do it on a daily basis. Because, you
220		work on a paper, or prepare a PowerPoint for a meeting and you do it
221		over the whole week, so if you can see the major influences from each
222		day, that would be fine. But I am not an academic. I mean I am office
223		work.
224	Thaul, W.:	The idea behind the system how the prototype is implemented. Really
225		every version is kept, but be aggregated to major versions and making it
226		configurable for the users. So that you can say every day or every half
227		day or whatever. And then you see these versions and when you click
228		on it, you see the real versions, if you really need them. So you really
229		do not lose any changes you have made, but exactly what we have said,

230		we put them together in the overview, because otherwise you will have
231		200 versions and this is not practical.
232	Morrison, M.:	Yes, and it is very demanding what you are tracking, phone calls, meet-
233		ings, text docs, websites, email, PDF – yeah – so a meeting could be
234		anything from a conversation in a hallway to a planned meeting with a
235		fixed agenda – yes....
236	Thaul, W.:	Exactly. And you have just mentioned meetings. Meetings and phone
237		calls are difficult to capture. So therefore the idea is to offer APIs which
238		allow us to synchronize for example our calendars or the phone history
239		from my phone, so just use this information. You already mentioned
240		your suspects regarding security. So this is how this will be done. It
241		might happen that I will just have a information, there was a meeting,
242		from my calendar, it was the I-Know, I know it was in Graz, but that is
243		all. There will be no information that we really met, that the meeting
244		was with you. Depending on which information I gave the system, I still
245		have the possibility to manipulate it manually, but in an automatic
246		mode, I can just get, what I have. But the idea behind it was, that the in-
247		formation that there was a phone call with you for example might be
248		helpful to me, even if I do not really know what we have discussed. But
249		just the information, there was this very important phone call we had,
250		might be helpful. Do you agree with this?
251	Morrison, M.:	Oh yes, because it is context. You know our memories are is jugged by
252		a whole variety of things and the key note speaker this morning was
253		talking of events, and how complicated they are, and first I was thinking
254		ah come on, you are making a big thing out of nothing, but actually
255		when I started thinking about it, you know, this brad interpretation of
256		events, it is true, and most of our knowledge is contextualise in events,
257		so yeah, just the fact that you are reminded that you spoke to a certain
258		person, will probably jurg our memory to remember what you have
259		spoke about in that context of this certain piece of work. What I find a
260		little difficult to understand, that in any given day in my last jobs I will
261		be doing maybe 10 major tasks during a day, so how does the system
262		know which task that phone call is related to?

263	Thaul, W.:	I agree. Phone calls are really the most difficult ones, I think. I would
264		like to answer the questions maybe first let uss say with documents, be-
265		cause there we have emails or PDFs or whatever. We can very easily
266		compare what it is about, the content. But currently the prototype is
267		really implemented just regarding the content. I think that this is not
268		enough. Especially regarding meetings and phone calls, we need also
269		the social connection algorism which allows us…you know, as long as I
270		know for example this phone call was with my professor, even if I did
271		not track this phone call and then look what we have talked about, just
272		the information, that was with my professor, and was maybe followed
273		ten minutes later by an email I have sent him regarding a conference.
274		And I am writing here paper for this conference which is academic, and
275		also maybe stands in the connection. This is the algorism, which is
276		missing, I am still working on it. Currently I am just comparing the con-
277		tent.
278	Morrison, M.:	Yes, that is a tough point, because I can also have a meeting with my
279		boss, because I am preparing a PowerPoint for the upcoming meeting,
280		and I might not mention that words presentation, PowerPoint, anything
281		– I just ask him a question that I will use then to clarify the presentation,
282		so it is messy, it is a messy thing you are trying to deal with.
283	Thaul, W.:	And therefore we have the possibility to manually adjust. Do you think
284		this is practicable? I mean the users are usually lazy and when I will
285		save this new version of the PowerPoint there will be a pop-up with all
286		these events. There will be for example this meeting, that will remind
287		you. Would it be ok for you, usable for you to grab this mouse and put
288		this meeting which is assumed that it is very important and put it up?
289	Morrison, M.:	For me no problem. because I do it a lot of time, when I get an email,
290		and I process that action, and I drag it into a folder, and categorise, be-
291		cause I do that, that is what I do, I classify, but for most people I think,
292		they might struggle with that, they are really engaged in the process it-
293		self, and they are not really thinking if I do this and I spend one second
294		to do this, I will be glad later on, because I find it. Either way, you have
295		just to prove the usefulness of a system so much first to get people to

296		start to do that, to add items manually, this kind of information... But it
297		could be good demo, and if you can solve just one of their problems,
298		just one, and they can see how it helps, how it works, than it will work...
299	Thaul, W.:	We already talked about this security issue. So assume it is really personal, it is not from a company point of view. And if you just had a button *switch on the job recording* and it will work. Do you think this would be for you personally ok?
300		
301		
302		
303	Morrison, M.:	Yes, and there will be times I will forget to switch on and times I will forget to switch off. Because that is how it is...
304		
305	Thaul, W.:	And later on the possibility just to click on an event and delete it, if you like. I mean I know sometimes you are surfing for a present for someone who is using the same computer or whatever. You really need to get rid of this event. If there is a possibility to switch on switch off and later on delete events completely of the system.
306		
307		
308		
309		
310	Morrison, M.:	You have to give an individual the right to do that, so yes, you have to have this possibilities....to give people that functionality.
311		
312	Thaul, W.:	The main idea was to present it in a table form, so you will all have in a sortable and searchable table form. So you will have all your text documents, all your emails, all your phone calls. So you can just…these text documents…we have three versions of this text document. So not every version is an extra text document, but just one of it. I click on it and then I get this overview of what happened with this evaluation of this document. Do you like it, do you have any other ideas how this could be presented?
313		
314		
315		
316		
317		
318		
319		
320	Morrison, M.:	The table is nice and need. And it is understandable for someone who likes to work with text. I think that the visualisation is important and I think you can probably improve on this. For example I would like to have some indications of date, time, that would be interesting for me to be able to see, that this here took me three weeks to do, but over the course of three weeks I discovered that – yes, I think that would be another piece of information I would like to see on that visualisation. It could prettier, it could be more compact, I think somehow, but I do not know enough about the options of visualisations to say, but I assume if
321		
322		
323		
324		
325		
326		
327		
328		

329		you click on here, it will take you to that document, and if you click
330		here it will show you a pop-up with all the details. Yes – it just looks –
331		as a visualisation – a little empty, not very rich on information. You
332		know I understand it perfectly, it took me just three seconds to under-
333		stand what you are doing there.
334	Thaul, W.:	OK, sounds very good. I like this with the time frame. I mean the time
335		is there just to give an indication, but really you would really suggest
336		dates for example?
337	Morrison, M.:	A date stamp maybe, yes.
338	Thaul, W.:	Then the last general question. Do you think such an implementation is
339		useful in general?
340	Morrison, M.:	Oh yes, absolutely. For people working in offices, especially, and aca-
341		demics as well, because the tracking of development of their ideas, this
342		is reflected in documents, so that I think absolutely, for those two kinds
343		of people I think it is fantastic. I have not thought beyond that, I mean
344		you could use it anywhere, but those are the environments that I under-
345		stand.
346	Thaul, W.:	Very good. I mean this came from the academic world. It started really
347		with the idea to support bachelor and master level students working on
348		their assignments, master thesis or whatever.
349	Morrison, M.:	And then there is the idea to have some analytics available, so that at
350		the end of the year you can look how much time did you spend on a
351		project – because that again would be very useful if you are trying to
352		evaluate your work and trying to decide how to enhance it in future.
353	Thaul, W.:	That is really an idea. I do not really think that this will be part of my
354		PHD thesis any longer, but of the product definitely. Not just looking
355		back, but allow also a forecast, an improvement of your own learning.
356		So statistics are definitely a part of it. Do you think anything is missing
357		here? We have already talked about this company point of you, but
358		when we stay at this personal point of view. For you tracking different
359		events which happening, you as the user, can configure yourself which
360		events you want to track and how do you want them to track. Do you
361		miss anything?

362	Morrison, M.:	People are moving a lot these days, they are not getting long-term con-
363		tracts, they are not staying a long time in a permanent environment, so
364		people would like of course to carry this with them – and I would be as-
365		suming you have thought of that.
366		But yeah – you have got a fan here!

Part 5 - Prof. Dr. Andreas Schmidt

1 2 3	Thaul, W.:	When we look at the historisation of work – do you historicise your work, when you work on papers. On publications, on documents and so on?
4 5	Schmidt, A.:	I make different versions, I make new names of the files, like version 1, version 3, version 5 and so on.
6 7	Thaul, W.:	Ok, have you ever used a software system for this? As there are many software systems available.
8	Schmidt, A.:	Only for software versioning, but not for writing papers.
9	Thaul, W.:	And why not?
10 11 12 13	Schmidt, A.:	I never tried it. Why not? I missed the idea. I just reached some new ideas from you some minutes ago. Probably it would be a good idea. And it could be either some file called futureidea.txt where I make some marks and so on.
14 15	Thaul, W.:	But currently you mainly use your own logical file system with version or the date of the document or whatever?
16	Schmidt, A.:	Right.
17	Thaul, W.:	Does this work for you?
18 19	Schmidt, A.:	It works for me, but probably it could be a lot better if I would change to a more sophisticated system.
20 21	Thaul, W.:	Imagine you have like ten versions of the document – do you delete your old versions from time to time?
22 23	Schmidt, A.:	After conference is finished and one year ago and I clean all my system, because I am looking for some more space.
24 25	Thaul, W.:	So it may happen that you delete some versions and maybe you lose some inputs also. Some input you could maybe use in another paper?
26	Schmidt, A.:	Right.
27	Thaul, W.:	In practice, did you ever wish you had not deleted something?
28 29	Schmidt, A.:	It happened. Rarely, but it happen from time to time. Or as I mentioned before – what the hell was I doing then? Why?
30 31	Thaul, W.:	Why – I will come to this in the next section. In general, what do you think of the idea to track versions of our knowledge independently, so

32		not within a Microsoft Word, or LaTeX or PowerPoint or whatsoever,
33		but as an independent layer, which is observing what we are doing?
34		Like a semantic desktop maybe observing what you are doing, making
35		these versions for you, so saving you time, in general?
36	Schmidt, A.:	Sounds very good.
37	Thaul, W.:	You have just mentioned this question *Why have I done this*. Basically
38		we are talking about remembering why we did things the way we did
39		them. How important is it from your point of view to remember why we
40		decided to go this way or this way or to do things we did them? Espe-
41		cially looking at your job and also at the academic environment.
42	Schmidt, A.:	It is a point of writing and that is so important compared to software
43		engineering. Here when I decide, why we go this way, are stronger rea-
44		sons. I think, if you write a paper at a later point, you have already an
45		overview, a stronger concept of what you want to write. So I think it is
46		more important in software development.
47	Thaul, W.:	What do you do to remember? I have heard that in software develop-
48		ment you have like a versioning tool?
49	Schmidt, A.:	Yes, that is right.
50	Thaul, W.:	Besides software development, when we think of the slides for your
51		lectures, of tutorials you are going to give at the university – do you
52		make any additional notes to make sure you know where this is coming
53		from?
54	Schmidt, A.:	After 50 % of my lessons I sit down and write one or two points that I
55		want to change in the next version of the slides – new ideas, spelling,
56		problems, more expressive examples or ideas I got from feedback from
57		the students or I mentioned myself while I was giving the lesson.
58	Thaul, W.:	And what are you doing with these notes?
59	Schmidt, A.:	The next time I give the lesson I take a further look at the slides and
60		then at my notes. If I have time, I will do it instantly shortly after I have
61		made the notes. But the latest point is when I give this slideshow again.
62	Thaul, W.:	I could image that you are having this same lesson maybe once a year
63		or every half a year. So does it really work? Meaning that half a year
64		later, do you find these notes?

65	Schmidt, A.:	It is easier, if I make the changes immediately, that is right. So it is an-
66		other strong point for your concept.
67	Thaul, W.:	Which kind of information should be considered there? You have men-
68		tioned already your own remarks, the comments from the students, any
69		additional ideas?
70	Schmidt, A.:	I have no other feedback – just what my own brain says to me or what
71		the students are saying to me.
72	Thaul, W.:	And you are not using any system for this. *Why not* is basically the
73		same answer as previously, I think? As there are also systems which
74		can handle this in a certain way.
75	Schmidt, A.:	Simple answer – I do not know these systems. I use Frame Maker for
76		making my slides. It is a program that is probably old than you, not in
77		the current version, but in general...
78	Thaul, W.:	How easy is it for you to recover why you decided to do things a certain
79		way and maybe relocate the resources? When you look at a assignment
80		or a homework – how easy is it, how good does it work?
81	Schmidt, A.:	I think, it works, but as I said before it good work better. I do not have a
82		problem sitting there why I did this in my slide. Normally I could re-
83		member or there is a red line I could follow. Probably if I had additional
84		information, which was the base of my decision one or two years ago, it
85		would be easier.
86	Thaul, W.:	How it came to this approach was that currently there are several sys-
87		tems, which can support the user with his knowledge work. But these
88		systems very often have the problem that they tie the users, they do not
89		want the users to let go. This advantage that you have to put a lot of in-
90		vestments mostly not in form of money but in form of time, and if there
91		is a new system coming up the market - and there is happening quite a
92		lot in this area – it is very, very difficult to change the system. And this
93		is not only true for document systems, but even we look at easier exam-
94		ples like photo community systems. Just imagine you have your photos
95		hosted with flickr, you add some geo information, some descriptions
96		and so on. Then for example Picase comes with another solution,
97		maybe a better solution with some new functionalities and you would

98		like to switch. They do not really allow you to export everything. One
99		idea behind such a system was – even if I am not able to export every-
100		thing – as soon as I know what happened and what I have done, I can
101		reproduce it in the new system. This is maybe not exactly what I am
102		showing you here, but from the logical point of view. I have a status
103		quo and I have the next status, and I know between those statuses I have
104		a run an application for face recognition, for text tagging, geotagging.
105		Do you think that such approach could be a good solution to overcome
106		this vendors problematic.
107	Schmidt, A.:	Difficult question. I would say, it depends on the work you are doing. It
108		could help you from in various situations, from time to time, or very of-
109		ten you decide with your stomach feeling and you have probably addi-
110		tional information, but I am not sure if they are helpful one year later,
111		because its more intuitive than fact based. I think it depends on what
112		you are really doing.
113	Thaul, W.:	To comment what you have said. I also have events not just like text
114		documents but also like a meeting, a phone call or an email. And from a
115		meeting like ours, I do not have the content in this knowledge base, but
116		knowing that I met you here, that I met you regarding talking about my
117		thesis. Maybe this additional information, which helps me and gives me
118		some more input. and also I basically have this version and I am look-
119		ing how did I come to this solution. I still do not really remember what
120		we have talked about, but I know there was something, maybe he is the
121		one I should ask.
122	Schmidt, A.:	You probably remember that we met, but you do not have the idea what
123		is here, here, or here. Even if you remember, there was the meeting. I
124		remember there was this point, this point and that point. And if you do
125		not have the information, you probably are losing information com-
126		plete.
127	Thaul, W.:	Do you think this information can be helpful?
128	Schmidt, A.:	Yes.
129	Thaul, W.:	So even if it is not just the real content, like text documents, PDS or
130		whatever, just some meta information, who met whom?

131	Schmidt, A.:	Yes, but at least this meta information helps you to remember.
132	Thaul, W.:	Now maybe talking about the implementation itself. In general, how do
133		you think the idea is, having it implemented this way that we make
134		these versions in an automatic way, but at the same time look at what
135		else happened and try to weight it. Your general impression of this
136		idea?
137	Schmidt, A.:	The solution you propose seems interesting, because the system is look-
138		ing what you are doing. You do not have to do it actively and write
139		down what happened. The system is doing that for you. After you finish
140		your work or make a new version and it popes up, you remember and
141		you have the possibility to enrich the information with additional text
142		and so on?
143		Yes, seems good. Probably you have to add something, like if you have
144		a discussion, the system cannot...
145	Thaul, W.:	Exactly. I can show it maybe on another slide. This is a very high level
146		architecture of this. You basically have a configuration unit where you
147		from a personal point of view configure which events should be looked
148		at and so on. Then there is this knowledge storage unit, which basically
149		consists of your listener which is observing what is happening as a
150		background process on your computer. As soon as according to this
151		configured events something relevant is happening, then this informa-
152		tion will be extracted in the way configured above. It will be saved in a
153		database and then all the other relevant events will be identified and the
154		influence measured. And as soon as you have, what you have just
155		asked, manual imports, you can either import it really manually, or you
156		can also use APIs for standard devices like smartphones, tablets or
157		whatever. So this is the main idea here. There is another point, we are
158		talking about a timeline. Every event happens somewhere in the time-
159		line and I had some problems with assigning the right point in time for
160		each event. For a document, it could be the opening of a document, the
161		closing of the document, the saving. There are completely different time
162		lines, when... I can open a document in the morning and I can close it in
163		the afternoon. Then it will be difficult for me.

164	Schmidt, A.:	It is more complicated with a browser.
165	Thaul, W.:	Exactly. Do you have any ideas, how to overcome this?
166	Schmidt, A.:	Probably it is enough if you have the stand from opening to the closing
167		of the document and more important it is the interval between two ver-
168		sions that occurs. Probably if you have the possibility to register if you
169		are scrolling or inside a document. I do not know what you can look for,
170		but I do not think that the exact point is so important. I would say it is
171		more important. Let us start from the opening, but this would be a point
172		where I would really be interested in the document. You have the first
173		idea that could be a solution for your problem. And then probably last
174		point. The sequence would be for you probably important, but in this
175		point I would say the start or the opening of the document.
176	Thaul, W.:	The idea I had was to make it flexible for every event type. For a word
177		document for example, as soon as I save it, this will be the relevant...
178	Schmidt, A.:	Save or open it? If it is an external source, you probably do not save it.
179	Thaul, W.:	That is true. But I mean here, in this document, when I save it. When I
180		open a PDF, it is read only anyway. When I receive an email – but via
181		email I would for example differentiate between an incoming email and
182		an outgoing email. Outgoing email, when I start writing it, because then
183		I had this idea. Whereas incoming email, when I open it and read it and
184		so on.
185	Schmidt, A.:	But if I write an email, I normally do not write it longer than 5 minutes.
186		I do not think that this is a big difference. The more problematic point is
187		when I read a conference paper, probably I read it three days. I start
188		here, the next day I read the next section and then I finish on the third
189		day.
190	Thaul, W.:	Ok, I see the point.
191	Schmidt, A.:	It seems that the opening of a document would be the most interesting
192		point, because here you start something, which you think could help
193		you. And if it is now finished in ten minutes or even shortly before you
194		save the next version. Probably... this is my opinion.
195	Thaul, W.:	Some users are saving their work every two minutes. This can result in
196		hundreds or thousands of versions. It will not only result in hundred of

197		versions, but it will also give the impression that between two versions
198		nothing really happened. And this would make this whole system not
199		really working. Therefore, the idea was to put these individual versions
200		together to so-called major versions. Would this be ok from your point
201		of view?
202	Schmidt, A.:	Yes, absolutely. An explicit checking or checkout of major versions, the
203		intermediate savings are not interesting I would say. .If I save a docu-
204		ment thirty seconds or hours, it would not be so interesting. I would say,
205		now is a point, where I make another major version, but I think here is a
206		good time, because when you show up, what would have been an influ-
207		ence and I give additional information or I rerank some points, I do not
208		want to do it every two minutes. So I think it would be an user triggered
209		action, when such a new version should be saved and not all the inter-
210		mediates.
211	Thaul, W.:	And I would usually display just the latest versions and if the user likes
212		to allow going further in the past, just to make it easier and give the user
213		an easier and better overview. Do you think that this is ok or do you
214		think it is really important to have a full view the whole time?
215	Schmidt, A.:	No, probably it would be interesting to see the last two versions, be-
216		cause I know what has happened between the previous and the last ver-
217		sion.
218	Thaul, W.:	When we look at the time between this version and this version, it is
219		pretty obvious which events should be taken in consideration by calcu-
220		lating these levels of influence. These is basically every event which
221		happened between these two points. The problem is when we have the
222		very first version. So I do not have a version number zero. Do you have
223		any ideas which period I could take to make sure I already look at the
224		right other events, which could influenced me?
225	Schmidt, A.:	Up to the point, when you save for the first time, yes.
226	Thaul, W.:	Maybe on this picture it is better. When we talk about these two, it is
227		obvious. But when we are talking about this one, which I save for the
228		very first time- how can I identify from your point of view the right
229		events which could have influenced me here?

230	Schmidt, A.:	Is the system observing my computer all the time, he could give me
231		what happened before, one day, ten days, so I could select some of
232		them. Or an easier solution would be, I create a document and this is my
233		version zero.
234	Thaul, W.:	So, just start from the first one, ok. On the Y-achsis we do have the lev-
235		els of influences. I currently present seven. The more you have the more
236		detailed it can be, but the less overview you have. Do you have a sug-
237		gestion how many would make sense?
238	Schmidt, A.:	From the feeling I would say three to five.
239	Thaul, W.:	It was interesting. The number seven – Prof. Schmauch just mentioned
240		this before – is the number which our brain can save directly without
241		having any problems and recognize directly. So seven is this magic
242		number. It was just a coincidence. The determination of the level of in-
243		fluence can be done in different ways. I have mentioned this statistical
244		method of analysis. Do you have any other ideas how I could realise
245		this? Any other algorism, which would be reasonable in this scenario?
246	Schmidt, A.:	Firstly, it depends on the source. If you have text, the number of algor-
247		ism to analyse what is the context of the text... How does the system
248		currently measure the influence?
249	Thaul, W.:	Currently there is just this one algorism with this statistical method of
250		analysis. It builds a vector, cleans up the vector throwing away all fill
251		words...
252	Schmidt, A.:	and then compares it with the vector in the document?
253	Thaul, W.:	Exactly. So it compares the vector of this version, with the vector of this
254		version, with this and so on. And if there is any comparison being done
255		it is highly rated, and if it is not, it is low rated. But in the end, I think, it
256		should be a mixture of different algorism. Especially when we look at
257		meetings, when we look at phone calls, emails and so on. It is not al-
258		ways about the content. There is also a very high social part.
259	Schmidt, A.:	Probably it would be interesting to see the delta between these two
260		documents and compare the delta with this document to see the influ-
261		ence.

262	Thaul, W.:	Yes, that is good. And some other points like discussions or so, you
263		only have the meta information you are doing or you add manually. To
264		ensure a fast and easy correction of the automatically assigned levels, a
265		user interface like this has been designed. Do you think this is reason-
266		able? You can just grab it with the mouse and adjust the values...
267	Schmidt, A.:	Yes, and plus or minus button.
268	Thaul, W.:	Exactly. Is it ok?
269	Schmidt, A.:	Yes. It should be easy. Although I think, it is better to have only 3 or 5.
270		If you have 7 or 8...if it is more rough to categorise them.
271	Thaul, W.:	Ok, so you would even go with 3 to 5.
272	Schmidt, A.:	Yes, 3 to 5. I would say 5 should be the maximum.
273	Thaul, W.:	There is another point. Not all the events happen within the desktop
274		environment like our meeting here or on the smartphone or the tablet.
275		To overcome this, I have showed you on the other slide already, there
276		has been designed these APIs to be able to integrate meta information
277		from calls, text messages, electronic calendars and so on, and as well a
278		manual import module where I just can input meta information about
279		our meeting and so on. Never the less this input of this additional in-
280		formation can be in the time much later than these automated generated
281		inputs. We do have a time lack there. Do you think this time lack might
282		be a problem for your personal work?
283	Schmidt, A.:	No, but if I insert it manually I can give it a timestamp in the past.
284		However, I think that the most important is, between which versions it
285		occurs. The sequence could be important. As in a discussion, when I
286		mention a paper and you go and have a look at the paper and so on. And
287		in this paper you could find another link to an interesting paper.
288	Thaul, W.:	Another technological question. The idea would be to run this in a web-
289		based environment in a cloud, just to make it available from different
290		sources, accessible from different devices. The system is observing
291		many the things you are doing. Concerning this and having stored this
292		information in an external environment. Would you have any concerns
293		regarding the privacy and security?

294	Schmidt, A.:	It could be encrypted, so I do not see any problem. I work with three or
295		four computer, so it would be important, that the information is linked.
296		But I think it should be really encrypted. I do not like that anyone's eyes
297		are looking how I am working or how I am not working.
298	Thaul, W.:	Maybe even more important. In general the system, as you just men-
299		tioned, is observing how you are working or how you are not working.
300		If there would be a simple button *Start / Stop the observation*, if you
301		want to do something privately like looking for some private presents...
302	Schmidt, A.:	Yes, important issue, I see here two points. Firstly, if I want to do some-
303		thing personnel, I do not want to be observed. And secondly, this is
304		spam for later classification of the relevant things that will truly happen.
305	Thaul, W.:	But this start / stop button would be enough in your opinion?
306	Schmidt, A.:	Yes, but another idea is, that you register web access ...no...well I think
307		such a button would be enough.
308	Thaul, W.:	Then once again to the other slide. When we now talk about the access
309		of the results, so on the one hand we save what happened, but when we
310		later want to access it, the main idea is, that we have basically a table
311		based application which I can sort, which I can also search and were all
312		my events are saved like for example a paper and the versions of the
313		paper. And as soon as I want to know what happened with this paper, I
314		just click on it and I will see this overview. And again on this overview
315		I can click on any other icon and this will then change to the overview
316		from the point of view of the other event.
317	Schmidt, A.:	This is a paper you read, but not the paper you are writing?
318	Thaul, W.:	This is the one I am writing.
319	Schmidt, A.:	Ok, this is the one you are writing.
320	Thaul, W.:	I can click on it, and then it is open. And this is my current version of
321		the paper, this is version 2 and version 1. and so I see, between version
322		3 and 2, this and this...
323	Schmidt, A.:	In the first moment I understood it differently. There was a phone call,
324		which influenced me, there was an external text which influenced me.
325	Thaul, W.:	This is not what influenced you. These are the several instances. And
326		there is a phone call conference between me and my supervisor and we

327		already have 43 of them. If I want to have a look at them, when they
328		were, what they were about and after maybe what documents or what-
329		ever...
330	Schmidt, A.:	Ok you are right. This is already aggregated information.
331	Thaul, W.:	But I see from your reaction that on the first point you were not too
332		satisfied with it. Any suggestions how this could be presented to the
333		user?
334	Schmidt, A.:	Probably I did not understand it right, but my first impression was that
335		the history of what happened. But this is not the version number, there
336		are 43 versions meaning for these real calls between you and your Pro-
337		fessor, and your three versions of your paper...
338	Thaul, W.:	All of them are saved behind this entry.
339	Schmidt, A.:	Yes, but you also have a possibility to have a table view in chronologi-
340		cal order.
341	Thaul, W.:	Yes.
342	Schmidt, A.:	Ol. No, I mean, then I missed the point. I think this is ok. When I open
343		it and see the single events and some information about these. And I
344		think the second point is something special, because this is what itself
345		the paper you...
346	Thaul, W.:	This is my instance. My instance of the event type text document. I am
347		working on this text document named *personnel knowledge manage-*
348		*ment system configuration*. I am the author. So these are the meta in-
349		formation basically coming. Once I open it, I can see, ok, this is the lat-
350		est version and these are the previous versions and all the events that
351		took impact on my work. The idea is, that for every type, I can define, if
352		there is such a history being made or not. From what I have understood,
353		in general this would be ok, but you would also wish to have a table on
354		a chronological order to see what happened.
355	Schmidt, A.:	Yes.
356	Thaul, W.:	And this graphical presentation. Do you think this is ok, to present it in
357		such a way or do you have any other ideas?

358	Schmidt, A.:	I mean, I see the PDF, but I do not have any idea what the PDF is about.
359		Probably it would be helpful, if you go on with the cursor and you get a
360		pop-up with the title or the URL of a webpage.
361	Thaul, W.:	This is the idea how it should work exactly. When you just make a
362		mouse-over you will see what PDF is it about, so you basically see the
363		main entity, which would be in this example the name of the document.
364		When you click it, then you will access it.
365	Schmidt, A.:	But probably in a tabular version, where you have one one line for each
366		of the external events, but then you can say what the title of the docu-
367		ment is... but then only between two versions for example. This is a
368		nice graphical visualisation, when you also see the relevant other
369		events... but probably from time to time I would prefer to see more de-
370		tailed information in table form.
371	Thaul, W.:	Now some last questions in general. Is the suggested system design
372		useful from your point of view?
373	Schmidt, A.:	Yes, I like the idea. The only point I can see here is, the work is reduced
374		to one document.
375	Thaul, W.:	This is always from the point of view of the document. When I want to
376		have a look at another one, this one for example, I just click on it and
377		then I get again a knowledge formation process of this document.
378	Schmidt, A.:	When I work on a project like with Markus for example, I do not work
379		only on one document, I work on several documents. And even tele-
380		phone calls are work items. Did you consider this or do you really re-
381		duce it to one document you are working on?
382	Thaul, W.:	I did consider it this way that all of them are considered. So basically
383		this phone call you have just mentioned is here, and once you click on
384		it...and there is a relation between this document we are talking about
385		and this phone call. Otherwise, it would not be here, it would be some-
386		where down there, meaning not available or relevant here. It will still be
387		available in the system, because this phone call might be important for
388		another work, but for this it would not. Therefore it is considered. But
389		the presentation view is always from the point of view of one item. Not
390		necessarily one document, I have chosen the document, because I think

391		it is the simplest one to understand, but you can also look at it from the
392		point of view of a website for example. So how often did you visit this
393		website? And how other websites gave you the link to this website and
394		so on? It is all interconnected and once I click on this for example...
395	Schmidt, A.:	But what I mean is, the work I am doing here, is writing a document.
396		But in real life the work could be '*organise a conference*'. So I have a
397		lot of things to do. I have calls, discussions with my colleagues. This
398		concerns all the documents I have. So all the emails I have to write, it is
399		not only concentrated on one – I mean the target is not only on one
400		document. It is a bunch of documents for example or emails or phone
401		calls.
402	Thaul, W.:	I am not sure if I see the point, because all these events which you have
403		mentioned – meetings with your colleagues, some emails, some other
404		documents. They are here.
405	Schmidt, A.:	Yes, but could I reduce a result to one document for example or would
406		not be the result 10 documents? Fifty filled in forms or...
407	Thaul, W.:	I see, multidimensional. I will think about this.
408	Schmidt, A.:	Develop some software for example.
409	Thaul, W.:	Yes, sure. That is a point.
410	Schmidt, A.:	Probably, it would not be a big problem. You only have to broaden it
411		and it is not a complete document, it is a mixture of several items. It
412		would not change probably anything. It could be the same. But to re-
413		duce it to one document, would be a limitation, I think.
414	Thaul, W.:	Ok, interesting, nobody mentioned this before. But I will think about
415		this. I was always thinking I am considering the other events.
416	Schmidt, A.:	Yes, but from my point this point here influences the result here. But
417		this point here could influence probably not only one document but a lot
418		of.
419	Thaul, W.:	And it is – so that phone call for example is influencing this document
420		to an extend of 3 for example, and at the same time is influencing the
421		other event to an extend of 1.
422	Schmidt, A.:	Ah, ok, I see.
423	Thaul, W.:	Any other ideas what is missing, could be better?

424	Schmidt, A.:	No, besides this point. If you give me a prototype and I play around
425		with it, I could give you some more hints. But I really like the idea.

Part 6 - Nikolaos Pappas

1	Thaul, W.:	Do you historicise your work?
2	Pappas, N.:	I am using an automatic versioning system.
3	Thaul, W.:	So you are versioning, and using a system for this. Great. Which one
4		are you using?
5	Pappas, N.:	Subversion.
6	Thaul, W.:	Are you happy with it?
7	Pappas, N.:	More or less. If I have an mistakes I can go back, roll back, but I cannot
8		track what else happened, what systems were used, if a browser was
9		used, or if there was a phone call. All this I do not know. I only know
10		the versions of the certain text. And maybe the code also, but that is all.
11	Thaul, W.:	Thinking of the files, do you also have your own logical system like the
12		date or version number within the filename?
13	Pappas, N.:	No, the system is taking care of this.
14	Thaul, W.:	And talking about other types of work, so for example not text docu-
15		ments but presentations?
16	Pappas, N.:	I do not histories presentations, no.
17	Thaul, W.:	Do you also delete your old version after some time? If yes, when?
18	Pappas, N.:	No, the system transparently keeps all the versions, so I do not really
19		need to delete it.
20	Thaul, W.:	Ok, let us come to the area of remembering why we did something.
21		Coming from here to here was maybe also because I visited a Website,
22		which was important, maybe I read a paper, which was important, or I
23		got an E-Mail from my professor that was important. How important do
24		you think is it, within the knowledge work you are doing, to remember
25		to let you come the conclusion where you are now?
26	Pappas, N.:	Well, I think that it is interesting to know some statistics about the evo-
27		lution of a version, but I do not know if I can come up with an exact
28		thing that will help me to remember. You want to track the knowledge
29		process – yeah. For example, you can say that if you are only using
30		Websites, so maybe consider reading a little more documents or col-

31		laborating with others. In case, we can get some additional information
32		out of that.
33	Thaul, W.:	Would that be useful for you?
34	Pappas, N.:	For me personally – it depends how the implementation will be. If it is
35		not annoying and running in the background, I will be using such sys-
36		tem and the statistics afterwards.
37	Thaul, W.:	What do you currently do to make sure you remember? Do you for ex-
38		ample use some notes. You have mentioned that you are using your sys-
39		tem for the versioning, but do you do something to remember why, or
40		what happened?
41	Pappas, N.:	I usually keep notes. Like *there was an error there*, or *that was prob-*
42		*lematic*, but usually like with other notes – you lose them.
43	Thaul, W.:	So you do some notes but sometimes lose them?
44	Pappas, N.:	Yes, because it is not easy, if you for example after one year want to
45		find the right notes.
46	Thaul, W.:	Would it be helpful for you if you had a system that takes these notes
47		you are currently taking manually and saving them automatically.
48		Maybe in a different form, as if showing you that there was an impor-
49		tant meeting, or a phone call, or an email from a participate which you
50		met here that was really important for you?
51	Pappas, N.:	It will help a lot, but with the restriction, that it is not covering all the
52		details. But I can imagine an application that works in the background
53		and you just add some comments. So the system could store these in-
54		formation and also the mentioned statistics what you used also.
55	Thaul, W.:	You have been asking regarding different algorithms to weight the in-
56		fluence. Do you know any algorithms that from your point of view or
57		based on your knowledge could be interesting for such a implementa-
58		tion?
59	Pappas, N.:	When you want to compare contents, you cannot rely on keyword ma-
60		trix techniques, because those keywords are not humanly annotated, and
61		do not have any semantic meaning behind them. The most simple ap-
62		proach is to vectorise the documents, but this will still rely on the exist-
63		ing of the word, so you have to use something more sophisticated, like

64		the semantic meaning behind it. So instead of using vector space mod-
65		els, apply semantic space models. Try to project all your documents in a
66		semantic space and compare the similarity there. I would think about
67		that, because most of the time when you want to compare the similarity,
68		and want to get related result, and not only based on the keywords, you
69		definitely have to think about such methods also. But it of course de-
70		pends on the question which task you want to evaluate. If you have for
71		example some human feedback, you could train an algorithm, like
72		*learning to rank* – so if a user said that this is more relevant than the
73		other, then you can train the software specifically for that individual
74		person what the ranking for that element is. So instead of using an ad
75		hock measure for similarity, you train and you learn the software what
76		is the weight for the user.
77	Thaul, W.:	Individually for every user differently?
78	Pappas, N.:	Yes, but of course you can the finding for others as well, by collaborate
79		these information maybe.
80	Thaul, W.:	Sounds good, I like that idea. When showing this system to some other
81		people, they mentioned issues regarding security and privacy. If there
82		would be a simple button switch on / off would that it will be fine with
83		you?
84	Pappas, N.:	That should be fine. And also, if the information is stored locally – that
85		is important. Because when you say we store it in the cloud, then even
86		people know that it is secure, nevertheless, most of the time I will not
87		trust it! I know that many do, and I know that many companies save
88		their data, but I am not using it, and even if some others might, as long
89		as you say it is stored locally, you just do not have to think about that.
90		The user is aware of it, he can delete his files, can backup them, he can
91		do everything.
92	Thaul, W.:	It is interesting. The last person I was talking to he was very much into
93		this could idea, everything should be centralised, and accessed via mo-
94		bile.
95	Pappas, N.:	Yeas, I know, that is the current trend and we do it all the time. So
96		maybe having it as an option, as long as it is by your decision initially,

97		that is fine, but not forcing it to have it on the cloud in order to use it.
98		Maybe the user could decide if he wants to have it stored locally, or in
99		the cloud or both.
100	Thaul, W.:	The graphical presentation. Would it be useful for you? So for example
101		you are just right here, here are the different versions, from the 40
102		available once just the last three displayed. If you like, you can go fur-
103		ther back here. Is that a visualisation that would be helpful for you?
104		Any suggestions or ideas?
105	Pappas, N.:	It is good, but it could be much easier than that. Like a list to give the
106		ranks from the most relevant influences, maybe with some bars, would
107		be especially easier to use on mobile devices with smaller displays.
108	Thaul, W.:	If the system would run in the background and the algorithms we have
109		talked about would satisfy you, would you use it? And if yes, more in
110		private or business modus – or even both?
111	Pappas, N.:	Oh yes, mostly in a private modus in work.

Part 7 - Andi Rexha

1	Thaul, W.:	Do you historicise your work?
2	Rexha, A.:	No.
3	Thaul, W.:	Why not?
4	Rexha, A.:	Because I am lazy. Well I try, but I do not have a tool.
5	Thaul, W.:	Have you tried some tools – there are some available?
6	Rexha, A.:	No, not a tool. Just text files. I try to log by myself which are the docu-
7		ments which are relevant for me. But then I get lost.
8	Thaul, W.:	And do you use a logical system? Some people for example use the
9		version number or the date within the filename.
10	Rexha, A.:	Not in the filename, probably – if any – in the Metadata.
11	Thaul, W.:	Did you ever deleted something you have been working on and later on
12		thought it would have been great having it?
13	Rexha, A.:	Yes, in the history it happened. Especially not on work, but more re-
14		lated to techniques and methodology. And then it is really difficult to
15		return and find back the techniques. The name of the technique for ex-
16		ample.
17	Thaul, W.:	And what did you do in such cases where you lost something?
18	Rexha, A.:	Tried to research again the document and try to find which was the
19		Webpage I was looking at for that work and that terminology.
20	Thaul, W.:	Is it usually a lot of work?
21	Rexha, A.:	Yes, it is a lot of work.
22	Thaul, W.:	So let us look at the fact of remembering why you did something,
23		which means not just versioning your work, but to look at what hap-
24		pened in between. Is it import for you to have this kind of information?
25	Rexha, A.:	Yes sure!
26	Thaul, W.:	What do you do to make sure you remember what happened?
27	Rexha, A.:	I try to put it in a text file, notes, and try to recall every document, and
28		try to recall which is the base, the line I am trying to go through. But
29		still it is kind of difficult, because one day you use a different vocabu-
30		lary, and another do not remember what you were thinking.

31	Thaul, W.:	Let us come back to the implementation I have just presented to you. In
32		general: how do you like the base idea of a system running in the back-
33		ground, observing what you are doing, and helping you to remember
34		what happened?
35	Rexha, A.:	The main idea is really nice, but I am not comfortable with the evalua-
36		tion how the system makes suggestions what is really relevant and what
37		is irrelevant. It can be relevant for one kind of document and irrelevant
38		for another kind of document regarding different topics.
39	Thaul, W.:	Definitely – so we are talking about the weighing algorithms, which are
40		behind it. There is not just one algorithm, but a mixture of different al-
41		gorithms. And from my perspective in the area of personal knowledge
42		management it means that it can be adjusted by the users.
43	Rexha, A.:	Yes, that is right.
44	Thaul, W.:	Some knowledge workers save their work pretty often like every five
45		minutes. This could lead to a very large number of versions. In this
46		case, the idea behind the system is to aggregate, to make automated ma-
47		jor versions. What do you think of this idea?
48	Rexha, A.:	Yes, I think it would be good to make a major version once a day. Oth-
49		erwise, you get lost if you do that. You would have notes for every sav-
50		ing you make – this way you would get lost.
51	Thaul, W.:	Ok, now looking backwards at what happened. The system will in gen-
52		eral display not more than five versions back, still offering you the pos-
53		sibility to go further back, but for overview reasons.
54	Rexha, A.:	In my opinion, having a system, which saves five, or even two major
55		versions would be enough. Talking about major versions and real ver-
56		sion it might be interesting to click on the space between version 1 and
57		version 2 and the system would probably display a popup or zoom in to
58		display all the real versions in between.
59	Thaul, W.:	Considering a first version of a knowledge piece, the system will con-
60		sider all events, which happened in a predefined (configurable) time
61		slot. How do like the idea? Do you see any other possibility to solve
62		this?

63	Rexha, A.:	That is kind of tricky. In my opinion, the second version could be seen
64		as the first base, the first version. The first version (from your point of
65		view) is just a start up, and usually there will be not so many important
66		documents as later on during the work in every other version. Therefore
67		showing the second version as the ground of the document should be
68		reasonable.
69	Thaul, W.:	Due to the fact that not every fact happens within a desktop environ-
70		ment, we will have some APIs to connect other applications, like for
71		example mobile devices, or online calendars.
72	Rexha, A.:	That would definitely be nice, yes.
73	Thaul, W.:	Would you have any security issues or concerns here?
74	Rexha, A.:	No, I do not think so, that would be fine.
75	Thaul, W.:	And do you see any other, general security or privacy issues. Because
76		basically the system will observe what you are doing, but I there was a
77		button which allows very easy to switch it on and off – would this be
78		ok, or are there any points you see?
79	Rexha, A.:	I think that is the best solution and enough. Usually users are lazy, and
80		putting too many options makes such system impracticable. Therefore,
81		a simple start / stop button would be great.
82	Thaul, W.:	What do you think about the presentation form in the table form? Be-
83		cause of the fact that we are not talking just about one document, or one
84		phone call, we have a mixture of different entities. The table form al-
85		lows you to select what you are looking for, or to search for it in practi-
86		cable way, sort the table, and so on. How do you like the idea? Any
87		other suggestions?
88	Rexha, A.:	Important is definitely an aggregation of the different events. So for
89		example an information on how many phone calls, how long were they,
90		and so on.
91	Thaul, W.:	Yes, this is exactly what we thought. This here is an aggregation for
92		example on the phone calls between me and my professor, and when I
93		click on it, I see all the single phone calls, as well as the information
94		what major events happened between them. Basically what had an im-
95		pact on the phone call, or what led to the phone call.

96	Rexha, A.:	Oh yes, this way you can find even a single element, this way it is per-
97		fect. This is really a very interesting system, but do you have different
98		relevant for the same document. So for example when I see on version
99		one that this document is not too relevant for my, then I proceed with
100		my work and later on the same document will again have an impact, but
101		this time it is relevant.
102	Thaul, W.:	That is exactly what we are doing here, because we are always looking
103		at the events on the timeline. So one and the same document may have
104		a huge impact on event x but no impact on event y, even if event x and
105		y are versions of the same work. But to make it clear, currently we are
106		only looking at the timeframe between the versions, so you need to
107		open the document again to make it relevant.
108	Rexha, A.:	Oh, then it is ok, this is what I meant! Then it is fine.
109	Thaul, W.:	In general, is the suggested system useful?
110	Rexha, A.:	Yes! Yes, to me it looks really nice. Really nice.
111	Thaul, W.:	Is it practicable? So would you use it?
112	Rexha, A.:	Sure. If it traces everything what I am doing and has some filters on it,
113		it would be perfect.
114	Thaul, W.:	And would you mainly use it in business or private area?
115	Rexha, A.:	I would mainly use it at work. Here the largest benefit would be to ar-
116		range the work, to arrange what I am doing, saving time, of course.
117		Finding documents, which already used before better. Oh yes...

Appendix C – Consent Forms

Part 1 - Dr. Ian O'Keeffe

Faculty of Science and Technology Ethical Application Form ps 2011/12 Final

UNIVERSITY OF PLYMOUTH

FACULTY OF SCIENCE AND TECHNOLOGY

Human Ethics Committee Sample Consent Form

CONSENT TO PARICIPATE IN RESEARCH PROJECT / PRACTICAL STUDY

Name of Principal Investigator: **Witold Thaul**

Title of Research: **Assessment and critical evaluation of the methodology and implementation approach of an extended personal knowledge management system**

Brief statement of purpose of work

With the aim of supporting students and other learners the PhD research project has identified methodologies and designed an implementation approach within the area of personal knowledge management.

The methodology and the implementation approach need to be assessed and critically validated. This shall be achieved on a qualitative basis within expert interviews. The suggested solution will be presented, followed by a critical, open discussion – based on a framework of questions.

The interview will be voice recorded and the results will be used within the PhD work and possibly also within some other future presentations or publications.

The entire interview should not take longer than 45 minutes.

The objectives of this research have been explained to me.

I understand that I am free to withdraw from the research at any stage, and ask for my data to be destroyed if I wish.

I understand that my anonymity is guaranteed, unless I expressly state otherwise.

I understand that the Principal Investigator of this work will have attempted, as far as possible, to avoid any risks, and that safety and health risks will have been separately assessed by appropriate authorities (e.g. under COSHH regulations)

Under these circumstances, I agree to participate in the research.

Name: ...... *IAN O' KEEFFE* ......

Signature: ...... *Ian O' Keeffe* ......   Date: ...... *06/09/12* ......

22

315

Part 2 - Prof. Dr. Andreas Peter Schmidt

Faculty of Science and Technology Ethical Application Form ps 2011/12 Final

UNIVERSITY OF PLYMOUTH

FACULTY OF SCIENCE AND TECHNOLOGY

Human Ethics Committee Sample Consent Form

CONSENT TO PARICIPATE IN RESEARCH PROJECT / PRACTICAL STUDY

Name of Principal Investigator: **Witold Thaul**

Title of Research: **Assessment and critical evaluation of the methodology and implementation approach of an extended personal knowledge management system**

Brief statement of purpose of work

With the aim of supporting students and other learners the PhD research project has identified methodologies and designed an implementation approach within the area of personal knowledge management.

The methodology and the implementation approach need to be assessed and critically validated. This shall be achieved on a qualitative basis within expert interviews. The suggested solution will be presented, followed by a critical, open discussion – based on a framework of questions.

The interview will be voice recorded and the results will be used within the PhD work and possibly also within some other future presentations or publications.

The entire interview should not take longer than 45 minutes.

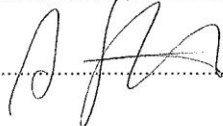The objectives of this research have been explained to me.

I understand that I am free to withdraw from the research at any stage, and ask for my data to be destroyed if I wish.

I understand that my anonymity is guaranteed, unless I expressly state otherwise.

I understand that the Principal Investigator of this work will have attempted, as far as possible, to avoid any risks, and that safety and health risks will have been separately assessed by appropriate authorities (e.g. under COSHH regulations)

Under these circumstances, I agree to participate in the research.

Name: *Andreas Schmidt*

Signature: .... Date: 28.3.13

22

Part 3 - Prof. Dr. Cosima Schmauch

Faculty of Science and Technology Ethical Application Form ps 2011/12 Final

UNIVERSITY OF PLYMOUTH

FACULTY OF SCIENCE AND TECHNOLOGY

Human Ethics Committee Sample Consent Form

CONSENT TO PARICIPATE IN RESEARCH PROJECT / PRACTICAL STUDY

Name of Principal Investigator: **Witold Thaul**

Title of Research: **Assessment and critical evaluation of the methodology and implementation approach of an extended personal knowledge management system**

Brief statement of purpose of work

With the aim of supporting students and other learners the PhD research project has identified methodologies and designed an implementation approach within the area of personal knowledge management.

The methodology and the implementation approach need to be assessed and critically validated. This shall be achieved on a qualitative basis within expert interviews. The suggested solution will be presented, followed by a critical, open discussion – based on a framework of questions.

The interview will be voice recorded and the results will be used within the PhD work and possibly also within some other future presentations or publications.

The entire interview should not take longer than 45 minutes.

The objectives of this research have been explained to me.

I understand that I am free to withdraw from the research at any stage, and ask for my data to be destroyed if I wish.

I understand that my anonymity is guaranteed, unless I expressly state otherwise.

I understand that the Principal Investigator of this work will have attempted, as far as possible, to avoid any risks, and that safety and health risks will have been separately assessed by appropriate authorities (e.g. under COSHH regulations)

Under these circumstances, I agree to participate in the research.

Name: .....Cosima Schmauch.....

Signature: .... Date: .....27.2.2013.....

22

Part 4 - Monica Morrison

UNIVERSITY OF PLYMOUTH

FACULTY OF SCIENCE AND TECHNOLOGY

Human Ethics Committee Sample Consent Form

CONSENT TO PARICIPATE IN RESEARCH PROJECT / PRACTICAL STUDY

Name of Principal Investigator: **Witold Thaul**

Title of Research: **Assessment and critical evaluation of the methodology and implementation approach of an extended personal knowledge management system**

Brief statement of purpose of work

With the aim of supporting students and other learners the PhD research project has identified methodologies and designed an implementation approach within the area of personal knowledge management.

The methodology and the implementation approach need to be assessed and critically validated. This shall be achieved on a qualitative basis within expert interviews. The suggested solution will be presented, followed by a critical, open discussion – based on a framework of questions.

The interview will be voice recorded and the results will be used within the PhD work and possibly also within some other future presentations or publications.

The entire interview should not take longer than 45 minutes.

The objectives of this research have been explained to me.

I understand that I am free to withdraw from the research at any stage, and ask for my data to be destroyed if I wish.

I understand that my anonymity is guaranteed, unless I expressly state otherwise.

I understand that the Principal Investigator of this work will have attempted, as far as possible, to avoid any risks, and that safety and health risks will have been separately assessed by appropriate authorities (e.g. under COSHH regulations)

Under these circumstances, I agree to participate in the research.

Name: *Monica Morrison*

Signature: *[signature]* Date: *7/09/2012*

22

Part 5 - Prof. Dr. Andreas Schmidt

Faculty of Science and Technology Ethical Application Form ps 2011/12 Final

UNIVERSITY OF PLYMOUTH

FACULTY OF SCIENCE AND TECHNOLOGY

Human Ethics Committee Sample Consent Form

CONSENT TO PARICIPATE IN RESEARCH PROJECT / PRACTICAL STUDY

Name of Principal Investigator: **Witold Thaul**

Title of Research: **Assessment and critical evaluation of the methodology and implementation approach of an extended personal knowledge management system**

Brief statement of purpose of work

With the aim of supporting students and other learners the PhD research project has identified methodologies and designed an implementation approach within the area of personal knowledge management.

The methodology and the implementation approach need to be assessed and critically validated. This shall be achieved on a qualitative basis within expert interviews. The suggested solution will be presented, followed by a critical, open discussion – based on a framework of questions.

The interview will be voice recorded and the results will be used within the PhD work and possibly also within some other future presentations or publications.

The entire interview should not take longer than 45 minutes.

The objectives of this research have been explained to me.

I understand that I am free to withdraw from the research at any stage, and ask for my data to be destroyed if I wish.

I understand that my anonymity is guaranteed, unless I expressly state otherwise.

I understand that the Principal Investigator of this work will have attempted, as far as possible, to avoid any risks, and that safety and health risks will have been separately assessed by appropriate authorities (e.g. under COSHH regulations)

Under these circumstances, I agree to participate in the research.

Name: *Andreas Schmidt*

Signature: *[signature]*   Date: 28/2/13

22

Part 6 - Nikolaos Pappas

UNIVERSITY OF PLYMOUTH

FACULTY OF SCIENCE AND TECHNOLOGY

Human Ethics Committee Sample Consent Form

CONSENT TO PARICIPATE IN RESEARCH PROJECT / PRACTICAL STUDY

Name of Principal Investigator: **Witold Thaul**

Title of Research: **Assessment and critical evaluation of the methodology and implementation approach of an extended personal knowledge management system**

Brief statement of purpose of work

With the aim of supporting students and other learners the PhD research project has identified methodologies and designed an implementation approach within the area of personal knowledge management.

The methodology and the implementation approach need to be assessed and critically validated. This shall be achieved on a qualitative basis within expert interviews. The suggested solution will be presented, followed by a critical, open discussion – based on a framework of questions.

The interview will be voice recorded and the results will be used within the PhD work and possibly also within some other future presentations or publications.

The entire interview should not take longer than 45 minutes.

The objectives of this research have been explained to me.

I understand that I am free to withdraw from the research at any stage, and ask for my data to be destroyed if I wish.

I understand that my anonymity is guaranteed, unless I expressly state otherwise.

I understand that the Principal Investigator of this work will have attempted, as far as possible, to avoid any risks, and that safety and health risks will have been separately assessed by appropriate authorities (e.g. under COSHH regulations)

Under these circumstances, I agree to participate in the research.

Name: Pappas Nikolaos

Signature: .... Date: 06/09/2012

22

Part 7 - Andi Rexha

UNIVERSITY OF PLYMOUTH

FACULTY OF SCIENCE AND TECHNOLOGY

Human Ethics Committee Sample Consent Form

CONSENT TO PARICIPATE IN RESEARCH PROJECT / PRACTICAL STUDY

Name of Principal Investigator: **Witold Thaul**

Title of Research: **Assessment and critical evaluation of the methodology and implementation approach of an extended personal knowledge management system**

Brief statement of purpose of work

With the aim of supporting students and other learners the PhD research project has identified methodologies and designed an implementation approach within the area of personal knowledge management.

The methodology and the implementation approach need to be assessed and critically validated. This shall be achieved on a qualitative basis within expert interviews. The suggested solution will be presented, followed by a critical, open discussion – based on a framework of questions.

The interview will be voice recorded and the results will be used within the PhD work and possibly also within some other future presentations or publications.

The entire interview should not take longer than 45 minutes.

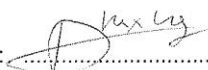The objectives of this research have been explained to me.

I understand that I am free to withdraw from the research at any stage, and ask for my data to be destroyed if I wish.

I understand that my anonymity is guaranteed, unless I expressly state otherwise.

I understand that the Principal Investigator of this work will have attempted, as far as possible, to avoid any risks, and that safety and health risks will have been separately assessed by appropriate authorities (e.g. under COSHH regulations)

Under these circumstances, I agree to participate in the research.

Name: ...Andi Rexhe...............

Signature: .... Date: ..6.03.2012..

22

Appendix D – Survey Form

Information and Knowledge Management in Learning Environments

1. Introduction & Statistical Information

Thank you for taking the time to complete this survey on **Information and Knowledge Management** within your studies.

The survey is part of a PhD research project at the *School of Computing and Mathematics* in the area of Knowledge Management.
The aim of the research is to an approach to better support students in the evolving learning environment.

This survey should only take about 5 minutes of your time. Your answers will be completely anonymous and will be used within a PhD research.
All questions require an answer in order to progress through the survey.

If you have any questions about the survey, please contact us at witold.thaul@plymouth.ac.uk.

1. Sex

◯ Female   ◯ Male

2. Age

[_____]

3. Faculty

◯ Arts & Humanities

◯ Business

◯ Health and Human Sciences

◯ Science and Environment

4. What is the highest level of school you have completed?

◯ High school degree or equivalent

◯ Some college but no degree

◯ Associate degree

◯ Bachelor degree

◯ Graduate degree

Information and Knowledge Management in Learning Environments

2. Information Overload

The following section refers to the topic of information and knowledge overload within your studies - mainly resulting from the information technology and the internet.

1. How much do you rely upon the internet for learning?

◯ Not at all   ◯ Slightly   ◯ Moderately   ◯ Very   ◯ Extremely

2. Compared with traditional academic information sources like books and journals, how often do you use the internet as a primary information and knowledge source?

◯ Never   ◯ Rarely   ◯ Sometimes   ◯ Often   ◯ Always

3. Compared with a book or a journal article, most of the published information on the internet is not validated.
How concerned are you regarding the correctness of the used information when you use it within your studies?

◯ Not at all   ◯ Slightly   ◯ Moderately   ◯ Very   ◯ Extremely

4. Search engines like Google or Bing help us find the information we are looking for. However, mostly a search query results in thousands of results. From all these results, how easy is it for you (in the academic context) to identify the ones which offer you the answers you are really looking for?

◯ Very easy   ◯ Easy   ◯ Moderate   ◯ Hard   ◯ Very Hard

5. Besides all the benefits of the information technology there are also critical aspects like information overload. For your studies, do you consider information overload as a issue / problem?

◯ Strongly disagree   ◯ Disagree   ◯ Undecided   ◯ Agree   ◯ Strongly Agree

6. Once you have found interesting information (like an interesting website), do you sometimes encounter a problem of locating it again later (e.g. 2 months later)?

◯ Strongly disagree   ◯ Disagree   ◯ Undecided   ◯ Agree   ◯ Strongly Agree

Page 2

Information and Knowledge Management in Learning Environments

3. Current Issues

The following section refers to problems and issues students might encounter during their studies.

1. Do you think remembering why you (or your lecturer, your study group) decided to do things a certain way is important?

◯ Strongly disagree ◯ Disagree ◯ Undecided ◯ Agree ◯ Strongly Agree

2. Do you sometimes encounter the problem of no longer remembering why you (or your lecturer, your study group) decided to do things a certain way?

◯ Strongly disagree ◯ Disagree ◯ Undecided ◯ Agree ◯ Strongly Agree

3. Typically students are using different software systems (like MS Word, Excel, PowerPoint, Dropbox, Google Drive, Evernote, Email-Accounts) for different tasks. Do you sometimes encounter problems with this because your information and knowledge is separated over several locations (also different systems, hard drives, online-portals)?

◯ Strongly disagree ◯ Disagree ◯ Undecided ◯ Agree ◯ Strongly Agree

4. Have you ever tried to switch a system you are using (e.g. e-mail provider, word processing, any kind of online systems) due to new and better features, but gave up due to problems with compatibility, data migration, or just the needed workload?

◯ Strongly disagree ◯ Disagree ◯ Undecided ◯ Agree ◯ Strongly Agree

Information and Knowledge Management in Learning Environments\
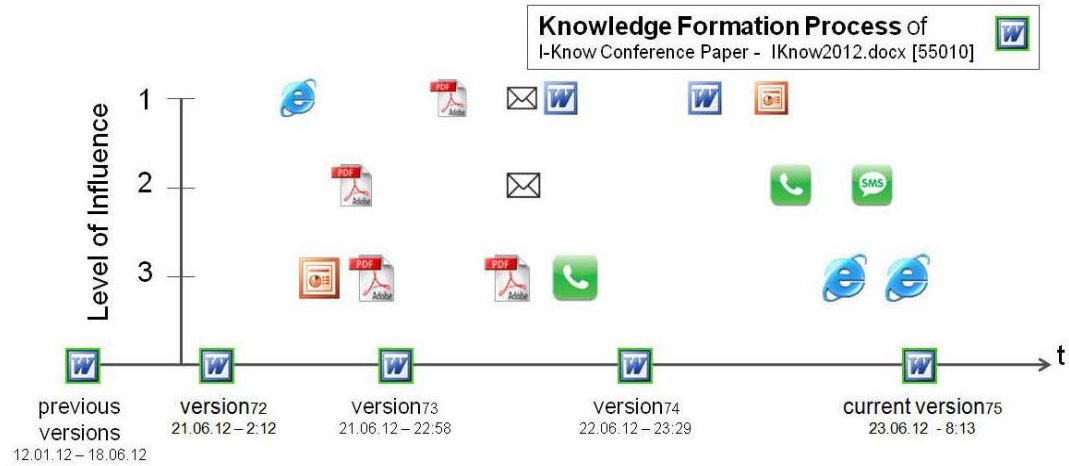
4. System Design

The following section refers to the system design of the proposed solution.

Imagine a system (on your PC and your Smartphone) which observes what you are doing and *automatically* captures the (for your studies) relevant actions – *independently of where (in which environment) they happen*. It automatically creates *versions of your work* (also from your visited (relevant) websites), stores them in *one central place* and offers you the possibility to access them at any time - also in a graphical manner on a timeline.

To *help you remember* why you decided to do things a certain way, the system *links the different actions* to each other - depending on their influence on each other. So you will not only see the different versions of your work (e.g. your assignment or thesis), but also all the other actions (books, journals, websites, emails) *which were important for your work*.

A *table based database* of your whole knowledge (left picture), and a *timeline based graphical presentation* of the single items (right picture) would allow you an easy access and seamless navigation though your information and knowledge.

	Unit	ID	Versions	Last Version	Type – Dependent Information 1	Type – Dependent Information 2	...
...							
	Title Impact of Wikipedia on ...	55002	2	2012-01-15 12:33:01	*Author* Mike Mayer	*Keywords* Social, Learning, e-Learning	
	Sender Max Mustermann	55003	-	2012-01-15 12:58:23	*Content* Found what you were ...		
	Title Paper Risk Management ...	55004	4	2012-01-15 13:34:54	*Author* Witold Thaul	*Keywords* I-Know Conference 2012	
	Title How many tweets?	55008	1	2012-01-15 15:20:11	*Author* John G., twitter.com	*Keywords* Twitter, Facebook, LinkedIn	
	Caller / Callee Witold Thaul – Udo Bleimann	55009	43	2012-01-15 16:01:43	*Caller* Witold Thaul	*Callee* Prof. Udo Bleimann	
	Title I-Know Conference Paper	55010	75	2012-06-23 16:57:12	*Author* Witold Thaul	*File* IKnow2012.docx	
	Subject New Figures	55011	3	2012-01-15 17:17:15	*Sender* Prof. Udo Bleimann	*Receivers* Witold Thaul, Max...	
...							

Knowledge Formation Process of
I-Know Conference Paper - IKnow2012.docx [55010]

1. At any point in time the system would offer you the possibility to go back to previous versions or statuses of your work. Would this be useful for you?

◯ Strongly disagree   ◯ Disagree   ◯ Undecided   ◯ Agree   ◯ Strongly Agree

Information and Knowledge Management in Learning Environments

2. Imagine that this way you would also at any point in time be able to find again already used resources independently of their kind or format. Would this be useful for you?

◯ Strongly disagree ◯ Disagree ◯ Undecided ◯ Agree ◯ Strongly Agree

3. At the same time, the system would store everything in one central place, irrespective of the environment in which the actual action happened.
Would this be useful for you - also in respect of changing the actual system or tool you would like to use for a task?

◯ Strongly disagree ◯ Disagree ◯ Undecided ◯ Agree ◯ Strongly Agree

4. The system would automatically link your different actions to each other and determine which dependencies and influences they had on each other. For example, a relevant website or a journal article could be linked to your assignment. This way it might help you to remember why you (your lecturer, your study group) decided to do things a certain way. Would this be useful for you?

◯ Strongly disagree ◯ Disagree ◯ Undecided ◯ Agree ◯ Strongly Agree

5. In general, would such a system support you with your studies?

◯ Strongly disagree ◯ Disagree ◯ Undecided ◯ Agree ◯ Strongly Agree

Appendix E – Survey Results

Sex	Age	Faculty	S1.Q4	S2.Q1	S2.Q2	S2.Q3	S2.Q4	S2.Q5	S2.Q6	S3.Q1	S3.Q2	S3.Q3	S3.Q4	S4.Q1	S4.Q2	S4.Q3	S4.Q4	S4.Q5
Male	26	Science	Graduate degree	4	3	4	3	2	2	4	4	4	5	3	4	3	4	3
Female	28	Science	Bachelor degree	4	3	2	4	2	2	4	5	5	4	4	4	5	3	4
Female	23	Business	Bachelor degree	5	4	3	3	2	4	2	3	2	5	4	4	4	3	4
Female	30	Business	Graduate degree	5	4	5	5	5	5	4	4	5	5	4	5	4	5	5
Male	47	Science	Some college but no degree	5	4	3	3	4	2	2	4	5	5	5	5	5	5	5
Male	26	Science	High school	5	5	4	3	4	5	4	4	2	4	5	5	4	4	4
Male	19	Science	Some college but no degree	3	4	2	2	4	2	4	4	4	4	5	5	4	4	4
Male	24	Science	Some college but no degree	5	5	3	3	4	4	3	3	4	4	4	4	4	4	5
Male	39	Science	Graduate degree	5	5	4	5	4	4	3	4	4	3	4	4	4	4	4
Male	19	Science	High school	3	4	4	3	4	2	4	2	2	2	4	4	4	4	4
Male	22	Science	Some college but no degree	3	5	3	4	5	5	4	4	4	4	5	5	5	5	5
Male	20	Science	Some college but no degree	4	4	3	2	4	2	5	2	4	4	5	4	4	5	4
Male	21	Science	Some college but no degree	4	4	3	3	4	2	4	4	4	3	4	3	3	3	3
Male	19	Science	Some college but no degree	4	4	3	4	4	2	4	4	4	3	4	4	4	5	5
Male	18	Science	High school	4	4	3	2	4	3	2	3	4	3	4	4	4	4	4
Male	18	Science	Some college but no degree	4	5	3	3	4	2	4	4	2	2	4	4	2	4	4
Male	27	Science	Some college but no degree	4	4	3	2	2	4	4	2	4	4	5	5	5	4	5
Male	36	Science	High school	3	3	5	2	4	2	4	2	1	2	4	4	4	4	4
Male	20	Science	Some college but no degree	4	5	2	3	2	3	3	2	2	4	5	4	5	2	2
Male	18	Science	Some college but no degree	4	4	5	4	2	4	4	3	4	4	4	4	2	4	5
Female	20	Science	Associate degree	5	5	2	4	2	4	5	5	2	3	5	5	4	1	4
Male	33	Science	Some college but no degree	5	4	4	2	2	4	4	2	4	2	5	5	5	5	5
Female	34	Science	Some college but no degree	3	4	3	3	3	3	4	4	2	3	4	4	4	4	4
Male	21	Science	High school	5	4	3	3	2	2	4	4	2	4	5	5	5	4	5
Male	23	Science	Associate degree	4	4	4	4	4	5	4	2	3	4	5	4	4	4	5
Male	25	Science	Bachelor degree	5	4	3	2	2	1	5	2	1	2	3	4	4	5	5
Male	20	Science	Graduate degree	4	4	3	3	4	2	2	2	2	2	4	4	4	4	4
Male	27	Science	Some college but no degree	5	5	2	2	4	1	3	4	2	1	5	5	4	4	4
Male	32	Science	Some college but no degree	4	4	5	3	4	2	5	5	5	3	4	5	5	4	4
Male	20	Science	High school	4	5	3	3	4	2	5	1	2	1	4	5	3	4	4
Male	23	Science	Some college but no degree	5	4	3	3	4	4	5	5	4	3	4	4	4	5	3
Male	21	Science	Bachelor degree	4	4	3	1	2	4	3	2	2	5	4	4	3	3	3
Male	20	Science	Bachelor degree	5	4	3	3	2	1	5	4	2	1	3	4	4	4	4
Male	20	Science	Some college but no degree	5	3	3	2	3	4	5	2	1	4	5	5	4	4	5
Male	39	Science	Some college but no degree	5	4	5	4	3	4	5	3	4	4	3	3	3	3	3
Male	21	Science	Associate degree	5	4	4	3	4	4	4	4	4	4	3	5	4	5	4

Sex	Age	Faculty	S1.Q4	S2.Q1	S2.Q2	S2.Q3	S2.Q4	S2.Q5	S2.Q6	S3.Q1	S3.Q2	S3.Q3	S3.Q4	S4.Q1	S4.Q2	S4.Q3	S4.Q4	S4.Q5
Male	24	Science	Some college but no degree	5	4	3	3	4	3	4	5	4	3	5	5	4	4	5
Male	22	Science	Some college but no degree	4	4	4	3	5	2	4	4	2	2	4	4	4	4	4
Male	23	Science	High school	5	5	4	2	2	1	5	4	4	2	5	4	4	3	5
Male	21	Science	Bachelor degree	5	4	3	2	3	3	4	4	2	2	4	4	4	4	4
Male	21	Science	High school	4	3	3	3	4	4	4	3	4	4	4	5	5	4	5
Male	18	Science	Some college but no degree	4	4	4	3	3	2	4	4	4	4	4	4	4	4	3
Male	23	Science	Some college but no degree	5	5	4	1	4	1	4	2	5	4	4	4	4	4	5
Male	18	Science	Some college but no degree	4	4	3	3	4	4	5	3	2	4	4	5	3	3	3
Male	28	Science	Some college but no degree	5	4	3	3	4	5	4	3	2	3	4	4	5	4	5
Female	23	Science	Some college but no degree	5	2	4	3	5	5	4	5	4	4	3	5	4	4	4
Male	22	Science	Bachelor degree	4	4	5	3	4	4	4	4	4	2	5	5	5	5	5
Male	18	Science	Some college but no degree	5	4	3	3	4	4	4	4	3	2	5	4	4	4	4
Male	24	Science	Graduate degree	4	3	3	4	3	4	4	2	2	4	4	4	4	3	4
Male	19	Science	Some college but no degree	4	4	4	5	4	4	4	4	2	1	4	5	5	4	4
Female	19	Science	Some college but no degree	5	5	4	3	5	2	4	4	4	2	5	4	5	4	5
Male	20	Science	Some college but no degree	4	4	2	2	3	4	4	2	4	2	4	4	2	4	4
Male	20	Science	Some college but no degree	5	4	3	3	3	1	5	4	4	3	5	5	4	5	5
Male	21	Science	Some college but no degree	5	3	4	3	4	2	4	2	5	1	4	5	5	5	5
Female	21	Science	Some college but no degree	4	4	3	5	4	4	3	3	5	4	2	3	5	4	4
Male	23	Business	High school	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5