DIARY GENERATION FROM PERSONAL INFORMATION MODELS TO SUPPORT CONTEXTUAL REMEMBERING AND REMINISCENCE

Christian Jilek, Heiko Maus, Sven Schwarz, and Andreas Dengel

Knowledge Management Department,
German Research Center for Artificial Intelligence (DFKI) GmbH, Kaiserslautern, Germany
{firstname.surname}@dfki.de

ABSTRACT

Human remembering is a process vulnerable to producing errors in addition to unintentional forgetting which is often undesired. To support contextualised remembering and reminiscence we propose an application that is able to write diaries on demand. Without any considerable effort these diaries can be generated for every period of a person's life that is reflected in their personal information model. This model is a cornerstone of the Semantic Desktop approach which serves as the ecosystem that helps us realize our vision. In this paper, we describe our concept in more detail and also provide an overview of related diary and timeline applications. Additionally, we present a proof of concept implementation as well as results of an early user experience evaluation. Finally, we give an outlook on remaining challenges to be tackled as well as possible future work.

Index Terms— diary generation, personal information management, semantic desktop, data condensation and summarization, contextual remembering and reminiscence

1. MOTIVATION

According to findings in [1], preservation of information about specific events in human memory is heavily influenced by contextualising the information about those events within domains of knowledge or schema which group key features that are repeatedly experienced in specific scenarios. Moreover, recall from memory involves the reconstruction of details of events based on schema, rather than the retrieval of veridical records. On the one hand this can be considered a strength of human memory since overloading with redundant or irrelevant information is prevented, but on the other hand this process is vulnerable to generating errors. Some of these errors could be avoided with the use of external devices that can record events as they occur to support accuracy in subsequent recall [1]. Errors as well as unintentional forgetting are also fostered by the ever growing flood of information a person's memory has to deal with nowadays. People often have a hectic daily routine facing an overwhelming mass of information: (calendar) events, tasks, documents, e-mails, bookmarked websites, images, videos, etc. While forgetting irrelevant details of daily activity is actually healthy and important [1], humans also tend to forget important matters which they would like to remember later. For these reasons, reminiscing or especially remembering and reconstructing what happened in various periods of a person's life becomes an increasingly effortful and tedious endeavour that may even fail completely (for example, if no or too few pieces of contextual information are available anymore).

To support this process, we propose an application that is able to write diaries on demand for any desired period of a person's life based on the aforementioned information objects (which we hereafter call a user's *personal information model* or *PIMO* [2] for short). As a consequence, users are able to quickly get an overview of what actually happened in arbitrarily chosen periods of their lives - be it a week, a month, a year or even a decade. Additionally, a diary's emphases as well as the desired level of detail may be altered by the user. To provide this functionality, things contained in a user's PIMO are clustered, evaluated, summarized and filtered.

Conceptional details of our approach are provided in Section 2, followed by a review of related work (Section 3) and the description of its technical realization in a proof of concept implementation (Section 4). The results of a first user experience evaluation are presented in Section 5, whereas Section 6 contains a final conclusion and outlook on possible future work.

2. CONCEPT

To enable a meaningful retrospection on (parts of) one's private or professional life, possibly very large amounts of data are needed. Two important questions to be answered are *how this data can be acquired* (2.2) and *how it can be presented to the user* in a comfortable, not overwhelming way (2.1).

2.1. The diary as a user interface to browse personal data

Addressing the second question, we use the concept of a *diary*, especially since we wanted to take the aspects of a *timeline* and some kind of *editorial preparation of text* into ac-

count. Furthermore, a diary is an easily understood and well-known concept to most people (statistics resulting from long-term studies about adolescents writing diaries are, for example, mentioned in [3]). In contrast to providing a kind of log covering all of a user's activities as a large sequential collection of material, we would like to enable the user to easily get an *actual overview* of their past, even for large periods of time. To achieve this, it is necessary to identify relationships among possibly several thousands of individual information objects in a user's PIMO and create suitable abstractions from them. If a user, for example, looks back on the last decade, they should not be overwhelmed with a view showing plenty of individual events, but compact statements like project names, stages of life, life situations, etc. Examples for those are terms like school years, studies, marriage or the name of a place where a vacation or longer stay abroad has been spent. The user therefore *zooms out* of the overwhelming mass of details. If desired, these abstractions can easily be resolved by selecting a subperiod of time for concretization (*zooming in*), e.g. five years of a decade. Concretizations can be performed until the user reaches the actual, not further processed, basic material, i.e., concrete information objects like notes, photos, documents, etc.

2.2. The Semantic Desktop and PIMO as means to provide semantically annotated & enriched data

Concerning the first question posed at the beginning of this section, we base our application on the concept of the *Semantic Desktop* (SD) [4], [5]. Its core idea is bringing *Semantic Web* technologies to the user's desktop[1]. Since these standards based on ontologies allow representing and organizing data across application borders, it is thus possible to explicitly express (major parts of) a user's personal mental model and make use of it in all their applications – or at least in those that integrate into their personal knowledge space. One of the SD cornerstones is the PIMO which serves as the basis for knowledge representation and provides a common vocabulary across different applications [6]. It consists of *concepts* (called "things" such as specific topics, projects, persons, tasks, ...), *associations* between them (persons are *member of* projects, a task *has topic* SD, ...), and finally, *associated resources* (documents, e-mails, web pages, pictures, ...) [6]. A more detailed motivation of the PIMO is given in [2]. The most recent SD implementation is in the form of a cloud-based service providing a service API based on JSON RPC that uses the PIMO schema with its classes and properties and intended semantics, relies on URIs to identify things and resources, and most importantly, defines a set of methods to access and manipulate the PIMO [6].

On the client side, we follow two strategies: first, ded-

---

[1]The term *Semantic Desktop* dates back to a time when desktop computers were the most prominent computing devices. Today, this term also comprises smartphones, tablet computers, etc.

icated PIMO-aware apps provide various beneficial PIMO-enabled services such as apps for photo organization, task management, note taking, or, a diary. Second, for standard programs (like Mozilla Firefox and Thunderbird, or Windows Explorer), we provide plug-ins connecting to the PIMO and extending them with PIMO vocabulary [6]. This allows for accessing, analysing, and introducing the respective information object such as web pages, e-mails, or files to the PIMO. As a consequence, the PIMO is integrated into computing environments users are familiar with. Annotating resources is done by tagging. For a given resource (a file, a website, an e-mail, ...) the system provides a list of recommended things from the user's PIMO by means of information extraction, which can then be accepted, complemented or corrected by them. The more users interact with their PIMO, the more complete and enriched with semantically annotated data it gets. This data serves as a rich basis for our diary app.

2.3. Joining these concepts in PIMO Diary

PIMO Diary, a diary application built on top of the SD, uses the things a user is concerned with (represented by their PIMO) and thus allows enriching stored memories with their proper context. The users decide for themselves on how much they would like to participate in shaping their diary. They may actively write entries, create photo collections that they additionally annotate and comment on, store a lot of supplementary material and so forth. Alternatively, they may just use their SD applications "the usual way" for personal information management (PIM) (i.e. take notes, set a calendar entry, tag an e-mail, etc.) without caring much about the outcome of a diary generation process in the near or distant future. Nevertheless, a diary generated from automatically collected data arising from daily SD usage without any considerable effort might still be an enjoyable added value for them.

Other usage scenarios are, for example, to utilize a person's episodic memory reflected by their diary to initiate or support a search process. Consider the case of trying to find a photo once taken on a trip to Paris for which you do not recall any other meta data. Browsing your diary for trips to Paris and descending towards the underlying information objects might help in being able to find the specific file quickly. Another feature we consider quite promising is sharing your knowledge and experiences with family, friends or colleagues. More details about having someone else's material be incorporated into one's own diary are presented in Section 4.

3. RELATED WORK

In addition to basing our application on the SD, we especially utilize Seed, a semantic editor proposed in [7], to identify things from a user's PIMO in a text (a note, a comment, etc.) they are currently writing. Furthermore, we focus on diary

and timeline related works, disregarding topics like life logging, trend and topic detection or augmented (personal) memories.

3.1. Diary-related works

Concerning diary-related works, relevant applications include *ComicDiary* [8], *AniDiary* [9], and *Smart Diary* [10]. These applications share the idea of acquiring the necessary user data by reading out the logs of mobile devices, in particular smartphones in most cases. Since many people carry a smartphone with them all day, it appears obvious to ask these "digital companions" about a user's experiences and events. Typical inputs are GPS data, call logs, contact and calendar data, message texts and app usage statistics (e.g. from photo viewers or music players). The first two apps noted above generate small comic stories from this data including only a few or no text lines, whereas the latter produces human-readable sentences using predefined templates filled during diary generation. Also a common idea is trying to achieve a certain diversity within the diary in order to make it more interesting and fun to view/read, for example by using different images for similar comic scenes. We, too, adopt this principle in our approach.

Apart from the previously mentioned, rather academic examples, we could not find any application – neither free nor commercial – that actually generates diaries, i.e. texts based on users' experiences and events reflected by their information objects. All applications found are either diary apps in the sense that they help in organizing or writing daily notes as a kind of specialized text editor, or more or less PIM tools providing timelines reflecting the history of users' interactions with files, e-mails, etc.

3.2. Timeline-related works

During the last 20 years, several timeline applications have been developed: *LifeLines* [11], *Stuff I've Seen* (SIS) [12], *SIMILE Timeline* [13], *Continuum* [14], *YouPivot* [15] – just to name a few. We adopted – or at least were inspired by – some of their features, specifically details-on-demand, outlining, zooming and summarization (LifeLines), a "zoomable" and/or scalable histogram, always showing the whole dataset (SIS, Continuum), representation of hierarchical data (Continuum), clicking an information item opens it in the appropriate application (SIS) and "TimeMarks" (YouPivot). However, the application closest to our approach is *Microsoft Life Browser* [16] which itself is based on *Memory Lens* [17]. In the main part of its user interface, a user's information objects are aligned to so-called memory landmarks, i.e. events that happened during that time, e.g. a meeting, a conference, or a celebration. By using a slider ranging from *least* to *most memorable* the amount of data visible on the screen can be adjusted. Moving the slider to the *most memorable* position causes many (or most) of the data to disappear from screen.

Like all other tools mentioned before, no high- or higher-level abstractions are created causing them to be vulnerable to failing to provide the needed overview and clarity once the system contains lots of data. A possible solution could be to summarize several events (but also information objects) belonging to the same project by showing this project's name as a placeholder or landmark, as is undertaken in our approach. Following the principle of details-on-demand, clicking this landmark will re-display the formerly hidden details. LifeLines arranges some aspects in groups providing a kind of summarization, but since these groups seem to be predefined it also does not fit the needs of our approach.

Although they are similar to our approach, the *Facebook Timeline* and *Google+ Stories and Movies* are out of our scope for two reasons. First, they include too few sources: Google+, for example, only utilizes a user's photos or movies to create a story. We want to follow a more holistic approach by also incorporating files, bookmarked web sites, e-mails, etc. Additionally, as those platforms' main goal is social sharing, handing over very private or confidential data is at least a hindrance if not a major block. Although, we have to face the same trust issue when considering it as a cloud service, our ultimate goal is to provide a personal SD server at home.

In summary, to provide a meaningful and more holistic overview of a person's memories materialized in their personal information store, semantic relations between the different information objects have to be found and analyzed. Using the SD approach we already tackled the former aspect [5]. Given the semantic relations between things, we need to further analyze the data (clustering, condensation, etc.) to generate an attractive diary of a person's life. Last but not least, the look and feel of our application should be very appealing and closer to a classic diary or blog.

4. TECHNICAL REALIZATION

To prove our concept we implemented a diary component (realized as a JAVA servlet) into the SD infrastructure and created an *HTML5*-based app as a front-end.

4.1. User interface

A screenshot of our app's user interface is depicted in Figure 1. It shows the first two of ten entries covering the year 2014 (these are real-life examples taken from the PIMO running at our department). The default amount of entries to be generated is ten but may be altered by the user. This setting as well as several others and the *overall context* information are positioned on the right. The diary entries themselves form the main part of the app (left-hand side). Each of them consists of at least four parts: a date or time interval, a headline, a list of those underlying information objects that predominantly characterize the entry as well as the most prominent keywords originating from all objects. If these objects are photos them-
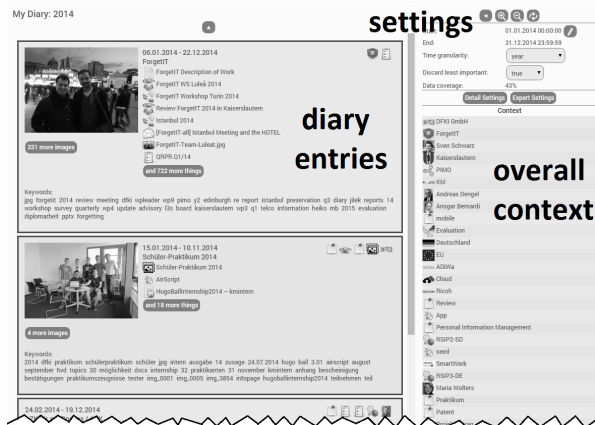
Fig. 1. User interface of the diary app: entries (left-hand side), settings and the overall context (right-hand side)

selves or are associated with photos, they are displayed on the left-hand side of the entry. On the upper right corner there are the most prominent annotations summarized over all objects within the entry and illustrated as icons. A close-up of the first entry can be found in Figure 2. We see that it consists

Fig. 2. Example of a diary entry

of 559 information objects belonging to the ForgetIT project. The icons and headlines of the first eight ones are displayed and 551 additional information objects can be displayed on demand by clicking the respective button. Clicking any of these objects jumps to the specific resource's view, allowing for directly opening it (if the appropriate application is available). Among these most characterizing objects are calender events of project meetings in the cities of Kaiserslautern and Luleå, a photo collection belonging to another meeting in Istanbul, a document (the project's description of work), an e-mail (Istanbul meeting and hotel), tasks (quarterly reports) and a bookmarked website (a wiki page). In total, 332 photos were found, from which the one considered most important by the system (for evaluation criteria please see the next section) is shown on the entry's left-hand side. Again, the rest of

the photos may iteratively be shown by clicking the respective button below. The most prominent annotations summarized over all objects' annotations are the ForgetIT project itself (cubic symbol) and another relevant task, both displayed as icons in the upper right corner of the entry and providing the entry's most prominent context elements.

An overall context of the currently viewed period is displayed in the right part of the app. (A close-up excerpt as well as another example from 2012 are shown in Figure 3.) Such a context is generated from all things and annotations of
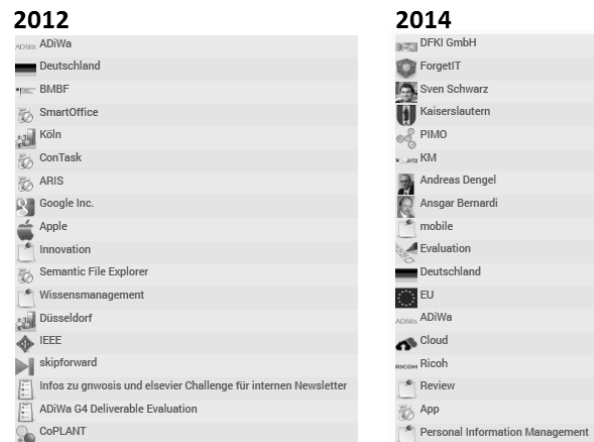

Fig. 3. Sample overall contexts of 2012 and 2014

a user's PIMO that are either associated with the selected time interval or have been created in it. It thus provides a compact overview of those things a user was concerned with the most during that particular time (based on their PIMO).

As stated in earlier sections, by clicking an entry the user zooms into the time period covered by it. In case of an entry covering a whole year, the next lower time granularity would be a half-year, a quarter, a month, etc. The time granularity can also be set directly in the upper right part of the app. Among these settings is also a button to zoom out of the current time interval. In case of a year, this would lead to a view covering five years and ten years after a subsequent click.

Another aspect we previously mentioned is incorporating shared information of others into one's own diary. This may either lead to completely new entries, e.g. representing the photo collection of a friend's holiday, or one's own entries are enriched. For example, consider the case of having a diary entry about a workshop consisting of your presentation, e-mails and notes. If your colleague also attended, took photos and shared them with you, they would appear as part of the same cluster previously only containing your material.

4.2. Diary Generation

Users are able to change the diary's default parameters like time period or number of desired entries to be generated. The

diary generation process itself completely runs on the server and works as follows: First, all objects of a user's PIMO either being created or having an associated date in the selected time interval (like an event of a meeting) are queried. Next, they are clustered according to their similarity. The similarity measure we used is a weighted sum of three components. For headlines and text bodies we create and evaluate term vectors like they are known from information retrieval [18]. Analogously, we create concept vectors for the objects' annotated concepts similar to the approach in [19]. These concept vectors are then extended by indirect relations found in a *spreading activation* process [20]. Two things that are both connected to similar resources, e.g. several persons and events of a project, will have a high similarity. Clustering similar things also fosters a higher diversity within the diary.

The resulting clusters are then evaluated according to their derived importance. All things of a cluster are associated with a type-based value representing their potential of being a memory landmark. For example, vacations or projects get a higher basic value than e-mails or notes. If a cluster contains things with a high memory landmark potential, its importance value is raised accordingly. Additionally, we consider aspects like cluster size, annotation intensity, association with rich media like photos, etc. Following the idea that rare contexts might be more memorable than common ones [17], we also add a rarity bonus to a cluster's importance value if a rare person, location, topic etc. is among its objects.

If a user has specified to exactly get as many entries as desired, the least important ones are discarded. For the remaining ones, headlines and keyword lists are generated. Additionally, each entry's associated photos are ordered according to their importance with the most important one being displayed next to the entry by default. Relevant aspects considered are whether the user has set a photo to be a *favorite*, a user-set preference score, added comments or annotations and an image's quality score as described in [21]. To generate an appropriate headline, the importance of every object within a cluster as well as its annotations are evaluated similar to the inter-cluster importance evaluation. Depending on these results the headline of the most prominent object or annotation (or a concatenation of both) is used. The keyword list is generated from all headlines and text bodies collected from the cluster's objects, whereas terms appearing in headlines are associated with a slightly higher weight. The idea behind this is that headlines are a kind of human-made summarization. Last, the overall context is generated from all available objects regardless whether they were sorted out in the importance evaluation or not. Then, the final diary entries are sent to the client.

5. EVALUATION

Our system has been evaluated by four industrial engineering students who were in their final year before completing their master's studies. They used the SD and PIMO for about three months before getting access to our application for about three weeks. During the evaluation they were asked 20 questions. Ten of them were taken from the USE questionnaire [22] covering the aspects of usefulness, ease of use, ease of learning and satisfaction. The others were specifically phrased to cover the different features of our app, e.g. meaningful clusters, headlines or text summaries, and especially whether the app succeeds in providing a satisfactory retrospection on the user's life (reflected by their PIMO). All items were measured on a seven-point Likert scale and were phrased such that a score of seven is best in all cases. The average of all 80 items resulted in 6.4, which we consider a success. But since this has only been an early evaluation, the group of participants was small and there might be a bias resulting from the general innovative character of the app, further evaluation is necessary.

6. CONCLUSION AND OUTLOOK

With PIMO Diary we presented an application that supports users in contextual remembering and reminiscence while being integrated into their daily interaction with their Semantic Desktop and connected PIM tools. Without any or very limited extra effort diaries for any arbitrarily chosen period of their lives can be generated. Still, there are further interesting research questions. Please consider the case of writing a document for several periods of one's life, for example a report in which the last chapter is added many months later. At which position in the diary should this document appear? At the date of its last modification? During the time interval when most of the work has been done? Additionally, more results form image analysis such as concept detection could be used and the performance in clustering and condensation could be further improved. By the latter we especially refer to textual condensation: keywords could be replaced by natural language summaries. We also thought of so-called *topic lanes*, in which a cluster is not depicted as a diary entry but as a lane with different spots covering all events and information objects during that time. Using such lanes zooming into "hot spots" of a certain topic would be much easier. Nevertheless, since our early user evaluation appears very promising we consider putting more effort into this project to be a rewarding investment. For more details on our app and the current SD research prototype please visit our web documentation [23].

7. ACKNOWLEDGEMENTS

8. REFERENCES

[1] R. Logie, E. Niven, M. Wolters, and V. Mayer-Schönberger, "D2.2: Foundations of forgetting and remembering - preliminary report," Deliverable D2.2, ForgetIT consortium, March 2014.

[2] L. Sauermann, L. van Elst, and A. Dengel, "PIMO – a framework for representing personal information models," *Proc. of I-Semantics*, vol. 7, pp. 270–277, 2007.

[3] I. Seiffge-Krenke, "'Dear Kitty, you asked me...': imaginary companions and real friends in adolescence," *Praxis der Kinderpsychologie und Kinderpsychiatrie*, vol. 50, no. 1, pp. 1–15, 2001.

[4] L. Sauermann, A. Bernardi, and A. Dengel, "Overview and outlook on the semantic desktop.," in *Proc. of the 1st Workshop on The Semantic Desktop at ISWC*, 2005.

[5] A. Dengel, "Knowledge technologies for the social semantic desktop," in *Knowledge Science, Engineering and Management*, pp. 2–9. Springer, 2007.

[6] H. Maus, S. Schwarz, and A. Dengel, "Weaving personal knowledge spaces into office applications," in *Integration of Practice-Oriented Knowledge Technology: Trends and Prospectives*, M. Fathi, Ed., pp. 71–82. Springer, 2013.

[7] B. Eldesouky, H. Maus, S. Schwarz, and A. Dengel, "Seed, a natural language interface to knowledge bases," in *Proc. 17th Intl. Conf. on Human-Computer Interaction, Los Angeles, USA*, 2015.

[8] Y. Sumi, R. Sakamoto, K. Nakao, and K. Mase, "Comic-Diary: Representing individual experiences in a comics style," in *UbiComp 2002: Ubiquitous Computing*, vol. 2498 of *Lecture Notes in Computer Science*, pp. 16–32. Springer, 2002.

[9] S.-B. Cho, K.-J. Kim, K.S. Hwang, and I.-J. Song, "Ani-Diary: Daily cartoon-style diary exploits bayesian networks," *Pervasive Computing, IEEE*, vol. 6, no. 3, pp. 66–75, 2007.

[10] J. Liao, Z. Wang, L. Wan, Q. Cao, and H. Qi, "Smart Diary: A smartphone-based framework for sensing, inferring and logging users' daily life," *IEEE Sensors Journal*, vol. 15, no. 5, 2014.

[11] C. Plaisant, R. Mushlin, A. Snyder, J. Li, D. Heller, and B. Shneiderman, "LifeLines: using visualization to enhance navigation and analysis of patient records.," in *Proc. AMIA Symposium*. American Medical Informatics Association, 1998, pp. 76–80.

[12] M. Ringel, E. Cutrell, S. Dumais, and E. Horvitz, "Milestones in time: The value of landmarks in retrieving information from personal stores," in *Proc. Interact*, 2003, vol. 2003, pp. 184–191.

[13] Massachusetts Institute of Technology, "SIMILE Widgets: Timeline," Application software, 2009, http://www.simile-widgets.org/timeline/.

[14] P. André, M. L. Wilson, A. Russell, D. A. Smith, A. Owens, and M. C. Schraefel, "Continuum: Designing timelines for hierarchies, relationships and scale," in *Proc. 20th Annual ACM Symposium on User Interface Software and Technology*. 2007, UIST '07, pp. 101–110, ACM.

[15] J. Hailpern, N. Jitkoff, A. Warr, K. Karahalios, R. Sesek, and N. Shkrob, "YouPivot: Improving recall with contextual search," in *SIGCHI Conf. on Human Factors in Computing Systems*. 2011, CHI'96, pp. 1521–1530, ACM.

[16] Microsoft Corporation, "Microsoft Research: Lifebrowser," Video, 2012, http://research.microsoft.com/apps/video/default.aspx?id=159531.

[17] E. Horvitz, S. Dumais, and P. Koch, "Learning predictive models of memory landmarks," in *Proc. CogSci 2004: 26th Annual Meeting of the Cognitive Science Society*, 2004.

[18] G. Salton, A. Wong, and C.-S. Yang, "A vector space model for automatic indexing," *Communications of the ACM*, vol. 18, no. 11, pp. 613–620, 1975.

[19] H. Liu, J. Wang, and D. Xu, "A simple and effective Concept Vector for WordNet semantic measurement," in *Advanced Computer Theory and Engineering (ICACTE)*, 2010, vol. 2, pp. 342–345.

[20] F. Crestani, "Application of spreading activation techniques in information retrieval," *Artificial Intelligence Review*, vol. 11, no. 6, pp. 453–482, 1997.

[21] E. Mavridaki and V. Mezaris, "No-reference blur assessment in natural images using fourier transform and spatial pyramids," in *IEEE Intl. Conf. on Image Processing (ICIP), France*, 2014, pp. 566–570.

[22] A.M. Lund, "Measuring usability with the USE questionnaire," *Usability interface*, vol. 8(2), pp. 3–6, 2001.

[23] C. Jilek, H. Maus, S. Schwarz, and A. Dengel, "ForgetIT Personal Preservation Pilot, esp. Ch. 8: Contextual Remembering with PIMO Diary," Web documentation, 2015, https://pimo.opendfki.de/wp9-pilot/.