Advanced Memory Buoyancy
for Forgetful Information Systems

Christian Jilek[1,2], Jessica Chwalek[2], Sven Schwarz[1],
Markus Schröder[1,2], Heiko Maus[1], and Andreas Dengel[1,2]

[1] Smart Data & Knowledge Services Department, DFKI GmbH,
Kaiserslautern, Germany
{christian.jilek,sven.schwarz,markus.schroeder,
heiko.maus,andreas.dengel}@dfki.de
[2] Computer Science Department, TU Kaiserslautern, Kaiserslautern, Germany
j_chwalek14@cs.uni-kl.de

Abstract. Knowledge workers face an ever increasing flood of information in their daily lives. To counter this and provide better support for information management and knowledge work in general, we have been investigating solutions inspired by human forgetting since 2013. These solutions are based on *Semantic Desktop* (SD) and *Managed Forgetting* (MF) technology. A key concept of the latter is the so-called *Memory Buoyancy* (MB), which is intended to represent an information item's current value for the user and allows to employ forgetting mechanisms. The SD thus continuously performs information value assessment updating MB and triggering respective MF measures. We extended an SD-based organizational memory system, which we have been using in daily work for over seven years now, with MF mechanisms directly embedding them in daily activities, too, and enabling us to test and optimize them in real-world scenarios.

In this paper, we first present our initial version of MB and discuss success and failure stories we have been experiencing with it during three years of practical usage. We learned from cognitive psychology that our previous research on *context* can be beneficial for MF. Thus, we created an advanced MB version especially taking user context, and in particular context switches, into account. These enhancements as well as a first prototypical implementation are presented, too.

Keywords: Semantic desktop · (Intentional) Forgetting · Information value assessment · User context · Information management and knowledge work support

1 Introduction

Knowledge workers face an ever increasing flood of information in their daily lives, which is even intensified by technological trends like digital transformation. To counter this and provide better support for information management and knowledge work in general, we have been investigating solutions inspired by

human forgetting since 2013, starting with the EU-project *ForgetIT*[3] and continuing in the *Managed Forgetting* project[4], which is part of the recent priority program on "Intentional Forgetting in Organizations" by the German Research Foundation (DFG). On the one hand, these solutions are based on semantic technologies, especially semantic graphs/networks, in a knowledge work scenario consisting of a *Semantic Desktop* (SD) ecosystem for individual users embedded in an evolutionary organizational memory for groups. On the other hand, they are based on *Managed Forgetting* (MF) by which we understand an escalating set of measures ranging from temporal hiding to condensation to deletion, thus overcoming the binary keep-or-delete paradigm. A key concept of MF is the so-called *Memory Buoyancy* (MB), which is intended to represent an information item's current value for the user and allows to employ forgetting mechanisms. Following the eat-your-own-dog-food credo, we extended our SD-based organizational memory system, which we have been using in daily work for over seven years now, with MF mechanisms. These new forgetting capabilities were thus directly embedded in daily activities, too, enabling us to test and optimize them in real-world scenarios. We learned from cognitive psychology that our previous research on *context* can be beneficial for MF. Thus, a major enhancement we incorporated into a successor of the initial MB was to take user context into account.

In this paper, we first provide more details about SD and MF (Section 2). Next, we present the initial MB and discuss success and failure stories we have been experiencing with it during three years of practical usage (Section 3). In Section 4, we introduce our advanced MB version, which takes user context, and in particular context switches, into account. Last, we conclude this paper and give an outlook on possible future work in Section 5.

2   Semantic Desktop & Managed Forgetting

Semantic Desktop. The *Semantic Desktop* (SD) [18] is especially intended to capture knowledge that emerges from individuals and then spreads into groups like project teams. SD brings *Semantic Web*[5] technology to users' computing devices using a knowledge representation, i.e. giving resources unique identifiers (URIs) and allowing to make statements about them, e.g. using RDF[6], resulting in a semantic graph. Information items (files, mails, contacts, events, topics, . . .) that are separated on the computer (file system, mail client, web browser, . . .) but are related to each other in a person's mind, can thus be semantically represented and interlinked in a machine understandable way. As soon as such an item is semantically represented, it is called a "thing", which describes the item uniquely as an URI complemented by further statements like its type or a reference to the originating resource such as an URL or message-id of an

---

[3] 2013–2016, `www.forgetit-project.eu`

[4] 2016–2019, `www.spp1921.de/projekte/p4.html.de`

[5] `www.w3.org/standards/semanticweb`

[6] `www.w3.org/RDF`

e-mail. Capturing a user's mental model as accurate as possible is done in a *Personal Information Model* (PIMO) [19], which serves as the basis for knowledge representation in the SD. Shared parts of multiple PIMOs result in a *Group Information Model* (GIMO) forming the basis for an organizational memory.

SD Prototype. Our current SD research prototype [14] uses micro plug-ins [10] to capture in-app events and performs information extraction on these evidences in order to elicit the current user activity/context. Then, SD decides on and executes appropriate support measures. Users interact with the system using a sidebar and a web app called *PIMO5* (referring to *HTML5* as its base technology). Figure 1 shows a user browsing a website while the SD sidebar analyzes the visited website in real-time [9] showing concepts of the user's PIMO that are mentioned on that particular site. Users may thus easily tag this page
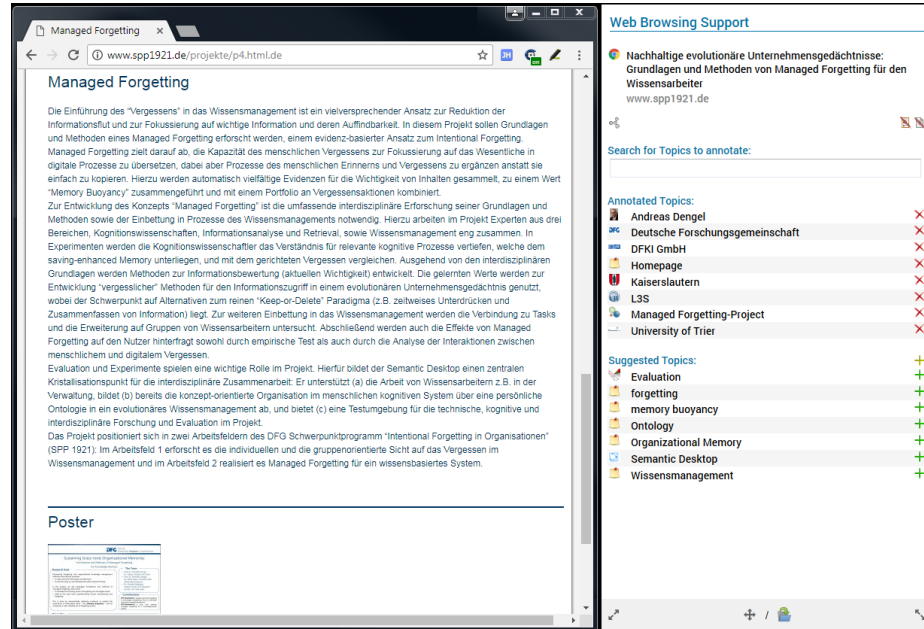

Fig. 1: SD sidebar (right) supporting user activities like web browsing (left)

with the respective topics. Similar plug-ins are available for email clients or the file system. SD's WebDAV[7]-based calendar *PimoDAV* is depicted in Figure 2: events entered into PimoDAV are automatically analyzed and added to the user's PIMO. Writing textual annotations (comments, notes, etc.) is possible using *Seed* [4], a semantic text editor (see Figure 3). While typing, Seed scans

---

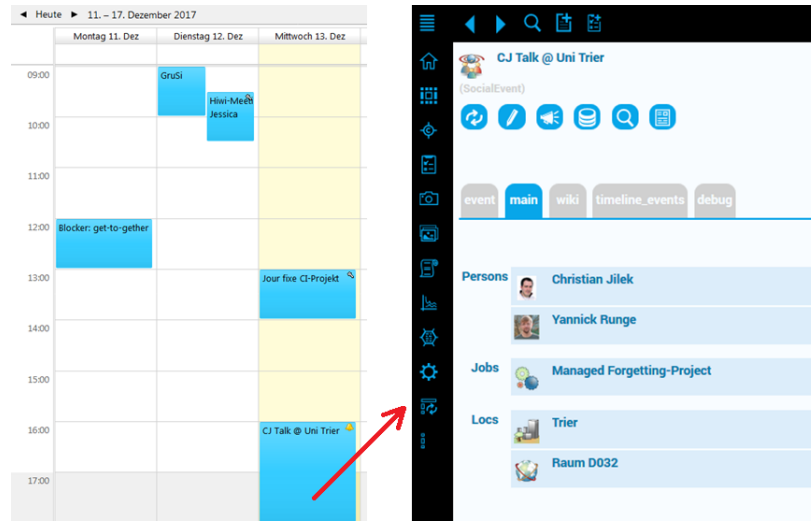[7] https://tools.ietf.org/html/rfc4918

Fig. 2: Browsing the representation of a particular calendar event in PIMO

Fig. 3: Semantic Text Editor *Seed*: data contextualization while typing

the text for entities found in the user's PIMO or other sources like DBpedia[8]. In the screenshot, those entities are highlighted in green. When hovering over such a concept further details appear (as it is the case for *PIMO (software)* in the screenshot). Content is thus contextualized while typing easing later retrieval. Speaking of retrieval: a user's PIMO can be searched using dynamically generated facets like illustrated in Figure 4. Here, the keyword *yannick* (first name of a colleague) is entered. The PIMO is then searched for all things having a literal (label, comment, etc.) matching that particular keyword. The resulting

---

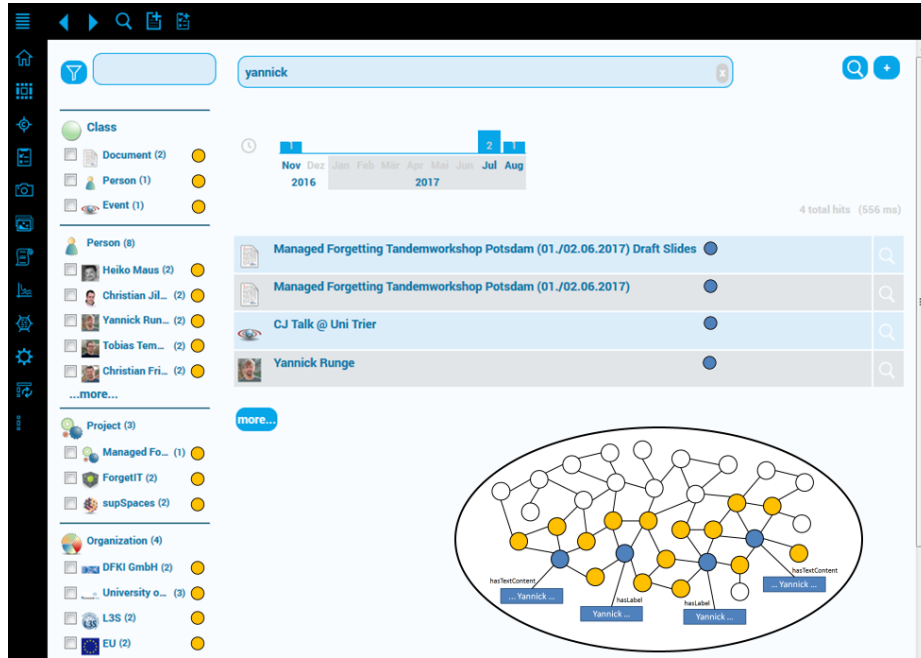[8] a semantic representation of Wikipedia, `https://wiki.dbpedia.org/`

Fig. 4: Using dynamic facets (yellow) to search things in PIMO (blue)

list (highlighted using blue dots) can be further filtered using facets dynamically calculated from the results' relations in the semantic graph, e.g. types, associated persons, projects, locations, etc. (highlighted with yellow dots).

So far, the system would sooner or later run into scaling problems typically found in SD systems [3]. That is another reason why we started investigating solutions based on human forgetting, which are discussed in the following.

Managed Forgetting. We started our investigations on information management and knowledge work support measures based on human forgetting and SD technology in the previously mentioned *ForgetIT* project. It was an EU-funded project consisting of a large international team. In its course, the *Memory Buoyancy* (MB) representing an information item's short-term value for the user as well as the *Preservation Value* (PV) as its long-term perspective counterpart were conceived [16,17]. A second aspect of the project was contextualized remembering: here, the idea is to store contextual metadata in addition to the actual material to ensure future comprehension of the text, e.g. who were the mentioned persons, terms, topics, etc. An application in this regard is the aforementioned semantic text editor *Seed* [4] that allows for data contextualization while typing. Last not least, the project introduced the concept of *Managed Forgetting* (MF), by which we understand an escalating set of measures overcoming the binary keep-or-delete paradigm. They range from temporal hiding,

to condensation, to adaptive synchronization, archiving and deletion. As stated before, MB is a key concept of MF and will be presented more thoroughly in the next section. Additionally, we show examples of how we incorporated MF into the SD. Further details going beyond this short overview can be found in [13], [11] and [7]. Focusing more on self-(re)organization aspects realized by means of MF, the more recent and still ongoing *Managed Forgetting* project refines several outcomes of ForgetIT like temporal hiding or MB. The extensions of the latter are especially addressed in Section 4.

3   Memory Buoyancy

Memory Buoyancy. Memory Buoyancy (MB) [16,17] is intended to represent an information item's current (short-term) value for the user. It follows the metaphor that items starting to lose relevance "sink away" from the user's sight, but are brought back up (by their higher buoyancy) as soon as they regain relevance.

In this section, we present our initial version of MB as well as success and failure stories we experienced with it during three years of daily usage in our organizational memory system. Last, we discuss problems leading to these failures and how we solved or intend to solve them.

Related Work. MB is a form of information value assessment (IVA), to our best knowledge, the only one designed and implemented according to findings of cognitive psychology about human memory and cognition. However, there are also other approaches mentioned in literature, which we will shortly mention in the following without going into details: Wijnhoven, Amrit et al. investigated identifying and dealing with *information waste*, for example in the file system [23] or on the internet [1]. Hasan & Burns speak of *waste data*: In [6], they analyze the impact of such data on computing environments and propose solutions on how to cope with it. Turczyk et al. [22] investigated *file valuation* in the area of information lifecycle management, which seeks to store files on different storage systems according to their (business) value. Gyllstrom & Pedersen propose *LostRank* [5], an approach to estimate which documents are most likely to be lost for the user (e.g. important but not recently used) and are thus harder to re-find.

Design Principles of MB in SD. The calculation of MB in our SD evolved from the insights of the approach in [21] and finally follows design principles shortly presented in the following. The basic ones are inspired by human brain activity applied to the user's mental model as represented in a semantic graph and discussions with the team of Prof. Logie (Psychology, University of Edinburgh) who presented their insights in [12].

- MB is updated every time a thing is stimulated (which in worst case could be after each click of a user).
- MB depends on:
  - user interaction (e.g. viewing, modifying, annotating),

- the thing itself (e.g. mails are assumed to be more ephemeral than a presentation), and
- its connections in the semantic network (PIMO).

– MB values are normalized ($0 \leq$ MB ≤ 1).
– Single access of an item should not directly lead to an MB of 1.0 (Fig. 5a).
– Multiple accesses in quick succession (e.g. every minute) are treated reluctantly (Fig. 5b).
– Multiple accesses every day, for example, will saturate against 1.0 (Fig. 5c).
– MB drops for things that are not stimulated:
  - first, there is a steep decline, then
  - a long-tail of slow decline.
– We added rules and heuristics to deal with requirements of various domains:
  - upcoming events should stimulate connected things,
  - finished items (tasks, events) shall decrease faster unless other indicators speak against this,
  - times with low user interaction should not lead to massive decay in MB
– MB utilizes the semantic network (PIMO): a form of *spreading activation* [2] is performed using additional heuristics for
  - types (e.g. emails are faster forgotten than persons), or
  - the number of relations that connect things.

In addition to these conceptional design principles, there are also technical requirements:

– MB is time-dependent which poses a challenge for calculation as MB values change permanently just as time passes.
– MB values must always be up-to-date (simply doing an overnight update is not satisfactory).
– Especially since it is a very interactive system, its high overall performance must be kept despite the additional MB calculation.

We thus split MB into two parts: a static, time-independent part and a dynamic, time-dependent one as shown in Figure 5d. This allows us to only update MB when applications or services actually query the resources, which reduces the additional calculation effort.

As mentioned before, this is only a short overview of the design principles: more explanations and details can be found in [13].

Success Stories. We incorporated MB in our productively used SD-based organizational memory system. Thus, we experienced several success and failure stories during three years of practical usage. In this section, we present one prominent example of each category.

The major success story we experienced is that rising and fading out of information items works quite well and is already helpful. Figure 6 shows the example of a workshop: Viewing this calendar entry during the meeting is depicted on the left-hand side. All involved persons (top), emails (middle) and tasks (bottom) associated with the event are shown. After eight months (middle part of

(a) single access

(b) multiple accesses in quick succession

(c) multiple accesses every day
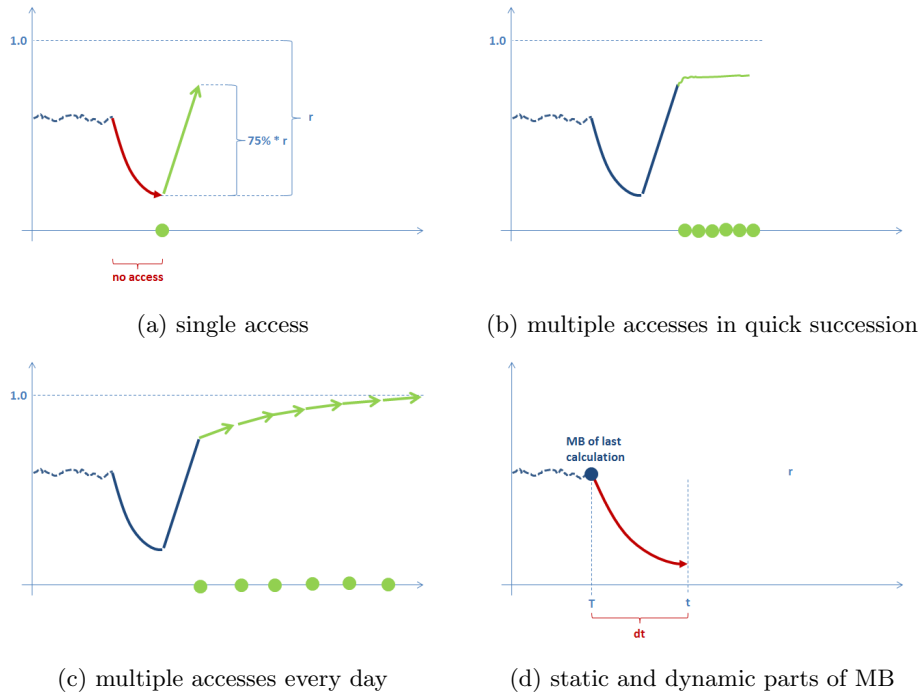
(d) static and dynamic parts of MB

Fig. 5: Selected design principles for MB in a SD

the figure), tasks and emails have already dropped in relevance and are thus not shown anymore. What remains are primarily persons involved in the project as well as the presentation slides and the project thing itself. After two years (right-hand side), the presentation is not visible anymore (possibly since it was just one presentation among many in the project), whereas photos taken at the event (probably often referred to and shown to colleagues) as well as the project itself remain. The same is true for several of the involved persons. A second example especially also showing MB rising is illustrated in Figure 7. It shows the MB graph of a task and two subtasks. The old task T9.4 (green) is revisited, thus rising in relevance (left-hand side), when writing a deliverable in task T9.5 (pink) and during the final project review (right). The parent work package task (WP9) is plotted in blue.

Failure Stories & Problems. A failure story we experienced was pathological behavior with our search user interface. The interface is equipped with a slider to manually lower the MB threshold, which was intended to be used if a user explicitly wants to remember something that has already been forgotten by the system. Such things have a low buoyancy and would thus not appear in the list of search results otherwise. What actually happened is depicted in Figure 8 (images left to right): users started to habitually drag the slider in the low MB

direction instead of rephrasing their query if sought things were not found in the first place. We address this problem in our improved *Forgetful Search* interface, for example by showing users how much of their information is currently covered by the search results. Additionally, we work on thematically clustering potential search results in the forgotten part so that users get a brief overview where to continue the search if sought things are not yet appearing in the results. This is addressed in a future paper.
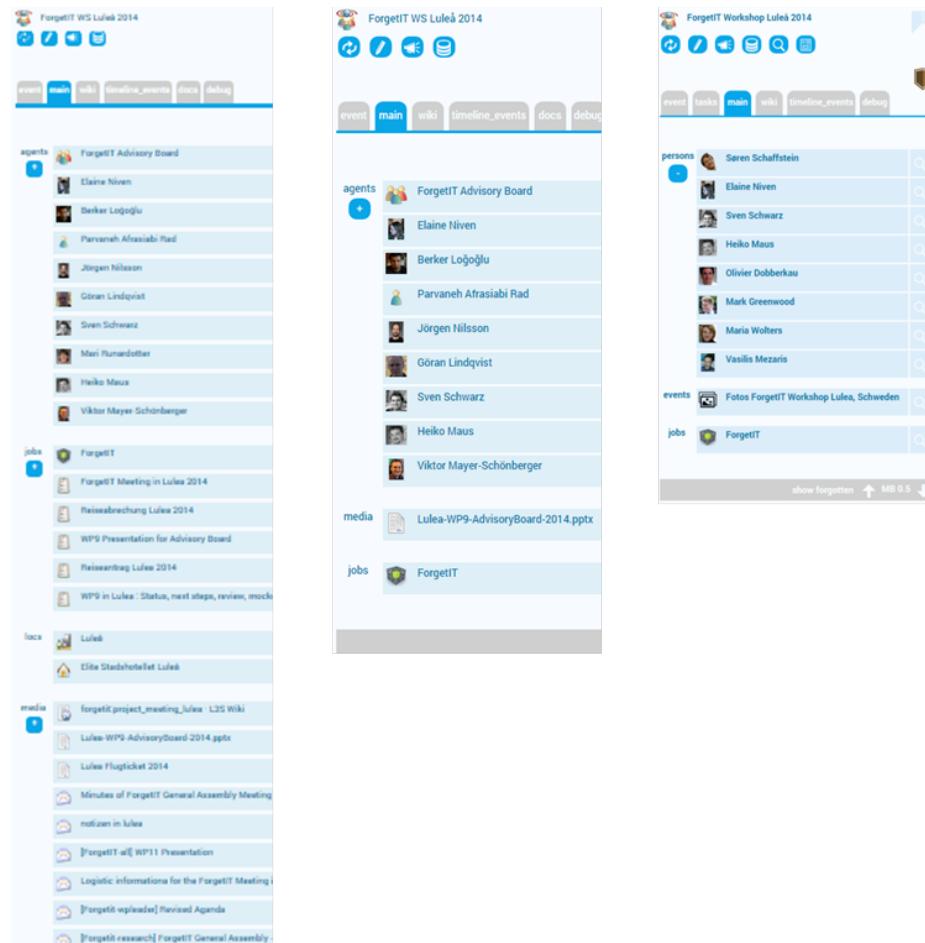

Fig. 6: Details of a calendar event gradually fading out: a comparison of the state during the meeting (left), after 8 months (middle), and after 2 years (right)

The MB slider failure is not the only problem we experienced: In [8], we name more problems and challenges also addressing forgetful information systems in general. Several issues regarding MB can be traced back to two major problems:
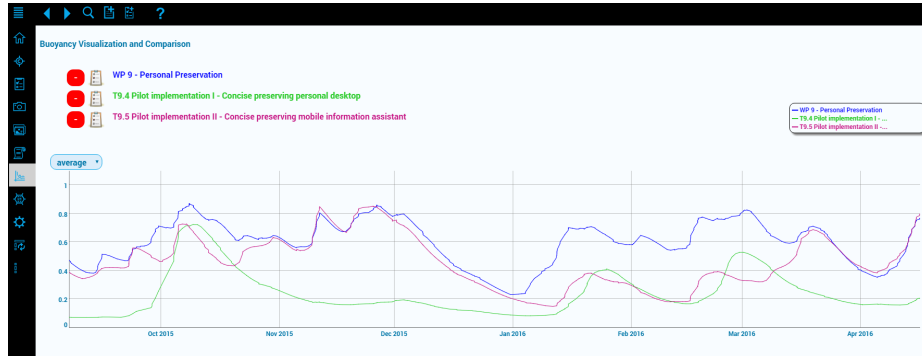
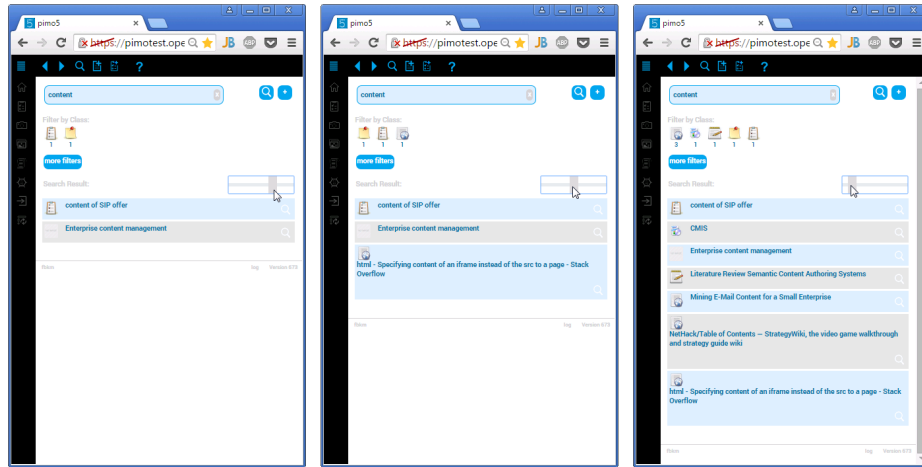Fig. 7: MB graph for a task (blue) and two subtasks (green, pink)

Fig. 8: Pathological behaviour in search: habitually lowering the MB threshold instead of rephrasing the query if things are not found in the first place

Our initial MB version is too much dependent on the connectivity of the semantic graph (PIMO) and it does not take user context into account. Let us consider an example to illustrate the latter: Figure 9 (left-hand side) shows the context of a *Trip to Rome in July 2018*. Its elements are ordered by descending buoyancy. Looking at these elements from low to high MB, the context's "golden thread" could be as follows: The owner of the context and a colleague, Peter Stainer, planned to meet in Mannheim and then travel on to Rome. An email was written asking in which hotel to stay. This email was answered a bit later. Then, the owner of the context booked a ticket with Deutsche Bahn, checked two different hotels and presumably booked one of them. At some point, the user switches to another context to plan a different meeting (middle section of Fig. 9). By chance, Peter Stainer is also an attendee of that meeting, the location is Mannheim and
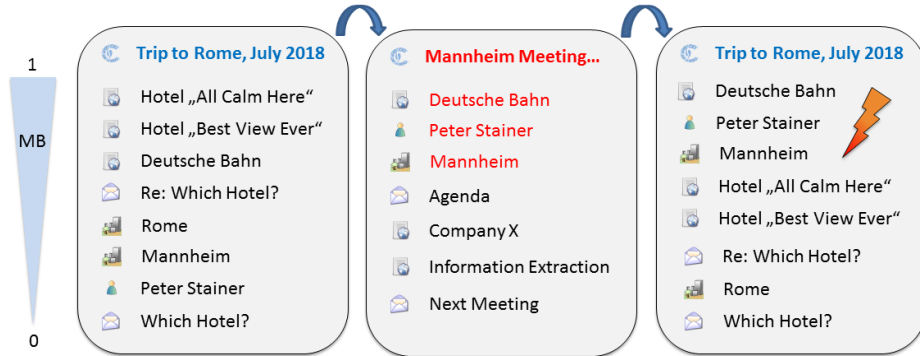
Fig. 9: Problem of context-free MB when switching contexts

they again travel by train (Deutsche Bahn). Thus, elements also present in the first context are now stimulated. When later revisiting the *Trip to Rome* context (Fig. 9, right-hand side), a context-free MB (like our initial version) would destroy the context's "golden thread" by re-ordering the elements although there was no action in that particular context which would justify this.

By "injecting" contexts into the semantic graph, we are able to kill two birds with one stone: Working with(in) context spaces, like proposed and realized in [10], ensures that information items are usually associated with at least one context leading to an increase of PIMO's connectivity. Second, we have a better handling of context switches avoiding cases like the example described before. Thus, one of the major enhancements of a successor version of MB was to take user context into account, which is discussed in the next section.

4   Advanced Memory Buoyancy

We learned from cognitive psychology as well as experience in real-world scenarios with MB that our previous research on *context* (e.g. [20,15,7]) can be beneficial for MF. Thus, the major enhancement we had to incorporate into a successor of our initial MB version was to take user context into account. This especially also included context switches: As an addition to the aforementioned gradually changing MB, we needed another value applicable in scenarios of sudden changes.

Local MB. We therefore introduced a *local MB*, representing a resource's current value for a specific user in a specific context. Stated more formally, this results in:

$$f_{\text{MB}_{local}}\colon (\text{ resource, user, context }) \to \text{MB}_{local}$$

Figure 10 (top section, gray shapes) shows different contexts of two users (green: User1, red: User2). As an example, we see that the website of Deutsche Bahn is
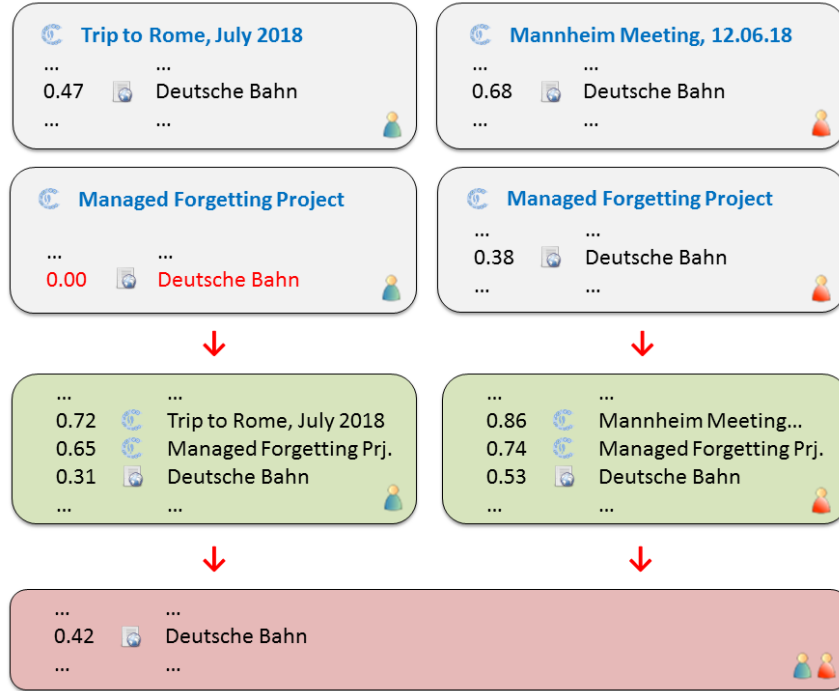
Fig. 10: Advanced MB versions: local (top), global (middle), group (bottom)

relevant for User2 in the context of a Mannheim meeting as well as the Managed Forgetting project. Its MB varies in both contexts: 0.68 in the first and 0.38 in the second. The same website is also relevant for User1 in the context of a Trip to Rome. User1 is also involved in the Managed Forgetting context, a shared context, but for them the Deutsche Bahn website is not relevant there, indicated by an MB value of zero. This is an example to illustrate that the very same thing can be very important in one context while being totally irrelevant in another. Second, a certain thing can belong to a certain context (like the Managed Forgetting project) whereas its relevancy in that context varies from user to user (here: 0.00 for User1 vs. 0.38 for User2).

Global MB. We saw that the MB of a certain resource may vary from one context to another. For several applications (e.g. synchronization, archival) it is helpful to be able to give a summarizing statement about a resource's relevance. Thus, we also introduced the *global MB* providing a context-free relevancy information of a resource for a certain user:

$$f_{\mathrm{MB}_{global}}: (\text{ resource, user }) \rightarrow \mathrm{MB}_{global}$$

It is very similar to our initial MB version and addresses less abrupt changes. In Figure 10, the green shapes in the middle section contain global MB values.

Note that the individual contexts have their global MB value. So, when leaving a certain context, the local MB values are frozen and will not change anymore until the context is revisited and therefore active again. Nevertheless, the MB of the context as a whole, i.e. its global MB, may rise or drop. In our current prototype, the global and local MB values are calculated independently of each other. We though plan to investigate the possibility of deriving the global MB from all local values instead of assessing all user evidences multiple times (once for each MB type). The same is true for the group MB.

Group MB. To especially address our organizational memory scenario, we additionally introduced a *group MB* value representing a resource's relevance for all users of the GIMO (red shape at the bottom of Figure 10):

$$f_{\mathrm{MB}_{group}}\colon \text{resource} \to \mathrm{MB}_{group}$$

Experimental Prototype. All these advanced MB values are still experimental, especially the group ones. Since further experiments are necessary, we developed a prototypical implementation depicted in Figure 11. We have modeled
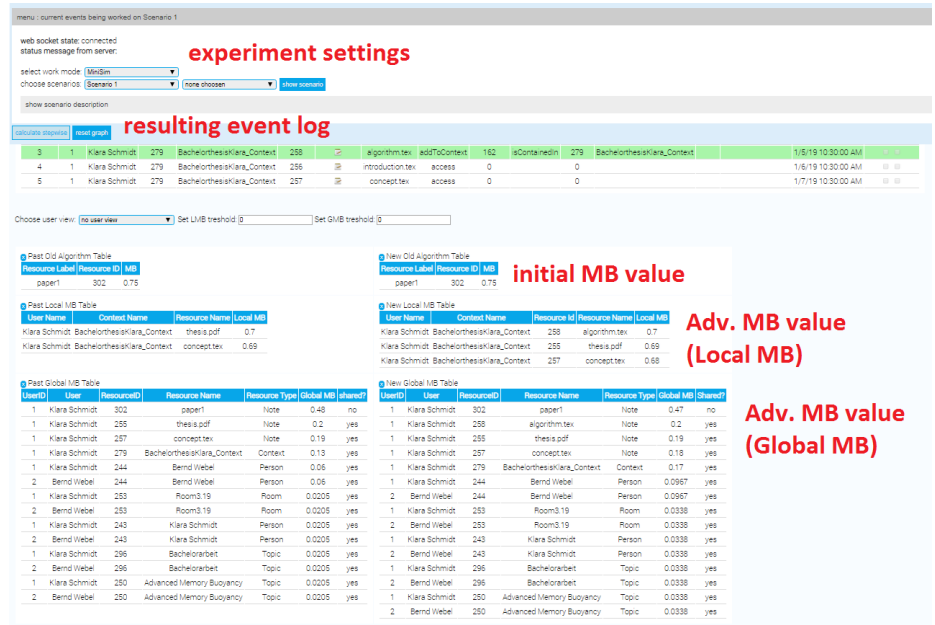
Fig. 11: Prototype implementation of Advanced MB for experiments

typical knowledge work scenarios: a single person working on a task, a group of people collaboratively working on a task, whereas in some scenarios only some members of the group are active while others are just readers, a situation shortly

before or after an event, etc. In the first section of the experiment app, one can choose a specific scenario. Each scenario is represented by a specific list of events, i.e. a generated list of user and system events, serving as input to the MB calculation (together with the semantic graph, PIMO, as the second input source). The lower half of the figure shows the different MB values: the initial MB as well as the new advanced MB values (group MB not visible on the screenshot). The left tables contain the MB values before, the right ones the values after the processing of each event.

5   Conclusion & Outlook

In summary, we incorporated MB into our productively used SD-based organizational memory system. We thus gained three years of practical experience with MB in daily work. Lessons learned from this as well as further cooperation with cognitive psychologists ultimately led to an advanced MB in the form of local, global and group values. They were presented in this paper as well as a first prototype to conduct experiments with them.

In the future, we plan to do more user experiments for in-detail evaluation. We are also interested in scaling experiments using "big personal data", i.e. all files of a person's or department's computer(s) instead of a selection stored in an organizational memory. In [8] we presented more challenges, especially also addressing forgetful information systems in general. One of the them is gaining and keeping user's trust in such a system, which we also mentioned in this paper when presenting our experience with the misused MB slider (see Section 3). We will especially address this aspect in our future paper on *Forgetful Search*.

References

1. Amrit, C., Wijnhoven, A.B.J.M., Beckers, D.: Information waste on the world wide web and combating the clutter. In: Proc. of the European Conf. on Information Systems 2015 completed research papers (2015)
2. Crestani, F.: Application of spreading activation techniques in information retrieval. Artificial Intelligence Review **11**(6), 453–482 (1997)
3. Dragan, L., Decker, S.: Knowledge management on the desktop. In: Knowledge Engineering and Knowledge Management. pp. 373–382. Springer (2012)
4. Eldesouky, B., Maus, H., Schwarz, S., Dengel, A.: Seed, a natural language interface to knowledge bases. In: Human Interface and the Management of Information. Information and Knowledge Design. pp. 280–290. Springer (2015)
5. Gyllstrom, K., Pedersen, E.R.: Not gone, but forgotten: Helping users re-find web pages by identifying those which are most likely to be lost. Proc. of the Workshop on Desktop Search at SIGIR (2010)
6. Hasan, R., Burns, R.: The life and death of unwanted bits: Towards proactive waste data management in digital ecosystems. In: Third Int'l Conf. on Innovative Computing Technology (INTECH 2013). pp. 144–148 (2013)

7. Jilek, C., Maus, H., Schwarz, S., Dengel, A.: Diary generation from personal information models to support contextual remembering and reminiscence. In: 2015 IEEE Int'l Conf. on Multimedia & Expo Workshops, ICMEW 2015. pp. 1–6 (2015)
8. Jilek, C., Runge, Y., Niederée, C., Maus, H., Tempel, T., Dengel, A., Frings, C.: Managed forgetting to support information management and knowledge work. KI – Künstliche Intelligenz, Special Issue: Intentional Forgetting (2018), in print
9. Jilek, C., Schröder, M., Novik, R., Schwarz, S., Maus, H., Dengel, A.: Inflection-tolerant ontology-based named entity recognition for real-time applications (2018), submitted for publication
10. Jilek, C., Schröder, M., Schwarz, S., Maus, H., Dengel, A.: Context spaces as the cornerstone of a near-transparent and self-reorganizing semantic desktop. In: The Semantic Web: ESWC 2018 Satellite Events. pp. 89–94. Springer (2018)
11. Jilek, C., Schwarz, S., Maus, H., Dengel, A.: Managed forgetting, data condensation & preservation in application. In: 2016 ACM Int'l Joint Conf. on Pervasive and Ubiquitous Computing: Adjunct. pp. 1046–1053. UbiComp '16, ACM (2016)
12. Logie, R.H., Wolters, M., Niven, E.: Preserving and forgetting in the human brain. In: Personal Multimedia Preservation: Remembering or Forgetting Images and Video, pp. 9–45. Springer (2018)
13. Maus, H., Jilek, C., Schwarz, S.: Remembering and forgetting for personal preservation. In: Personal Multimedia Preservation: Remembering or Forgetting Images and Video, pp. 233–277. Springer (2018)
14. Maus, H., Schwarz, S., Dengel, A.: Weaving personal knowledge spaces into office applications. In: Integration of Practice-Oriented Knowledge Technology: Trends and Prospectives, pp. 71–82. Springer (2013)
15. Maus, H., Schwarz, S., Haas, J., Dengel, A.: CONTASK: Context-sensitive task assistance in the semantic desktop. In: Enterprise Information Systems. pp. 177–192. Springer (2011)
16. Niederée, C., Kanhabua, N., Gallo, F., Logie, R.H.: Forgetful digital memory: Towards brain-inspired long-term data and information management. SIGMOD Record **44**(2), 41–46 (2015)
17. Niederée, C., Kanhabua, N., Tran, T., Naini, K.D.: Preservation value and managed forgetting. In: Personal Multimedia Preservation: Remembering or Forgetting Images and Video, pp. 101–129. Springer (2018)
18. Sauermann, L., Bernardi, A., Dengel, A.: Overview and outlook on the semantic desktop. In: Proc. of the 1st Workshop on The Semantic Desktop at ISWC (2005)
19. Sauermann, L., van Elst, L., Dengel, A.: PIMO – a framework for representing personal information models. In: Proc. of I-Media '07 and I-Semantics '07. pp. 270–277. Know-Center, Austria (2007)
20. Schwarz, S.: Context-Awareness and Context-Sensitive Interfaces for Knowledge Work Support. Ph.D. thesis, TU Kaiserslautern (2010)
21. Tran, T.A., Schwarz, S., Niederée, C., Maus, H., Kanhabua, N.: The forgotten needle in my collections: Task-aware ranking of documents in semantic information space. In: Proc. of the 2016 ACM Conf. on Human Information Interaction and Retrieval (CHIIR'16). pp. 13–22. ACM (2016)
22. Turczyk, L., Groepl, M., Liebau, N., Steinmetz, R.: A method for file valuation in information lifecycle management. AMCIS 2007 Proceedings p. 38 (2007)
23. Wijnhoven, F., Amrit, C., Dietz, P.: Value-based file retention: File attributes as file value and information waste indicators. Journal of Data and Information Quality **4**(4), 15:1–15:17 (2014)