

Preservation and Forgetting: Friends or Foes?

Nattiya Kanhabua and Claudia Niederée

L3S Research Center / Leibniz Universität Hannover
Appelstrasse 9a, Hannover 30167, Germany
{kanhabua,niederee}@L3S.de

Abstract. Humans are very effective in remembering by abstraction, pattern exploitation, or contextualization. On the other hand, humans are also capable of forgetting irrelevant details, an important role in the human brain helping us to focus on relevant things instead of drowning in details by remembering everything. The research question that we address in this paper is: *Can we learn from human remembering and forgetting in order to develop more advanced preservation technology?* In particular, we aim at studying how a managed or controlled form of forgetting can play a role in digital preservation, including personal and organizational archives as well as collective memories. Our research goal is twofold: 1) to establish effective preservation for more concise and accessible digital memories, and 2) to enable the easier and wider adoption of preservation technology. The concept of managed forgetting is discussed in more detail in the research work of the European project ForgetIT, which investigates the proposed concept by mean of an integrated information and preservation management approach.

Keywords: Digital Preservation; Dynamic Information Value Assessment; Time-aware Information Access; Managed Forgetting

1 Preservation and Forgetting

At first glance, forgetting seems to contradict the idea of preservation, which is about keeping things, not about throwing them away. However, if no special actions are taken for long-term preservation, we already face a rather random digital forgetting process in the digital world today, including the Web, the Social Web, as well as private and professional information management context. Things get lost for various reasons, namely, hardware crashes, change of employment, or by deletion of referenced material as has been shown in [5] for the Social Web. In addition, the pure and growing volume of digital content created nowadays makes it necessary to take decisions on what to preserve, just consider the amount of content shared every day on platforms, such as, Twitter or YouTube.

In this paper, we envision the concept of *managed forgetting* in digital memories, which will actively support users and systems in their decisions about what to keep, and what can and/or should be forgotten. We argue that this concept could bring a variety of benefits for preservation technology. The concept

of *managed forgetting* is inspired by the important role of forgetting in the human brain, where forgetting enables us to focus on the things that are relevant instead of drowning in details by remembering everything. This concept has a twofold purpose: 1) to create immediate benefit by focusing on relevant content in a time-aware fashion, and forgetting (unfocusing) things that are no longer important, and 2) to replace random forgetting of information, e.g., by hardware crash, with managed forgetting, where users are optimally supported in their explicit decisions about what to keep, and how, what is kept, should be organized and preserved. We will systematically deal with information in an information space that progressively ceases in importance and finally becomes obsolete, as well as with redundant information. Furthermore, *managed forgetting does not mean deletion*. With managed forgetting the system is able to detect such information, and to trigger *forgetting actions*, which can be taken from a wide variety of possible forgetting actions including elimination of redundancies, aggregation, modification of ranking, and, finally, also deletion.

To some extent, managed forgetting is closely related to work on appraisal and selection of content to be preserved. However, we view the information value assessment underlying the managed forgetting as a *dynamic, repeated process*. This implies the revisiting of preservation decisions based on information value re-assessment leading to more dynamic archival approaches as they are, for example, also propagated by preservation models as the record continuum model [6]. Beyond these concrete ideas, it is important to foresee managed forgetting in our digital world today, which is dominated by a *keep it all* paradigm, in a more fundamental way: As already observed nowadays, there is a problem of information persisting much longer than intended in the Web and causing, e.g. problems in future life situations, such as, seeking employment [4]. When Web and Social Web content is archived and made available by more advanced and effective technology, which by itself is desirable, this engraves the problem even further and adds another facet to forgetting in preservation. More precisely, there should be some control and regulations on what and how long things about individuals, and maybe also organizations, are remembered. Therefore, forgetting is of core importance for projects that aim at archiving Web and especially Social Web content, such as, ARCOMEM.

2 Managed Forgetting in Context

In this section, we give a brief overview of our proposed concept of managed forgetting in the ForgetIT¹ project, which focuses on preservation in the personal and organizational context. Finally, we present key research challenges that need to be carefully investigated within the project. The more detailed description and a case study for further strengthening the need for systematic forgetting support can be found in [2].

¹ <http://www.ForgetIT-project.eu>

2.1 Personal Preservation

The creation, handling, and sharing of electronic information within the personal sphere, e.g., via mobile devices in their daily lives, has seen a unprecedented growth and change in recent years. Cornerstones for such development are new technical devices and corresponding changes in our everyday behaviors. Digital photo- and videography create large data volumes and numerous artifacts that are considered a valuable part of personal remembrance [3]. Participative content generation and sharing in Web 2.0 solutions and social interaction via networks and platforms have gained wide acceptance, ranging from media-specific sharing, e.g., Flickr, over text and video distribution channels, e.g., Twitter and Youtube, up to web-based documentation and sharing of nearly complete life histories as encouraged by Facebook. From the user's point of view, the personal information space of an individual thus evolves quickly into a large, distributed network of interlinked information elements, which represent very important aspects of the individual's life.

Keeping this information space not only well-structured and accessible, but also sustainably available over time in spite of changes in hardware, software, and data formats, is a crucial challenge for which satisfying solutions are not yet available. Successful personal information management [1] needs to combine the aspects of long-term technical availability of information (with "long-term" meaning at least a lifetime, if not across generations) with suitable accessibility of the information, which asks for sustainable structures and situation-specific information retrieval and delivery.

2.2 Organizational Preservation

In organization or work-related settings, there are already promising organizational information and knowledge management systems, e.g., CMSs, which include sophisticated workflow approaches to support the process from creation of content to its publication on the Intranet or Web. However, this support does not go beyond the publication step (in OAIS² terminology, the producer sphere). Long-term preservation and information access is not part of the content workflow, and no archiving features, except the versioning of individual resources, are available.

As a result, preservation processes, if in place at all, are organized independently from the production process, cutting the content lifecycle into two unconnected pieces. With this separation, access to organizational content removed from the publication system is difficult and costly. Consequently, a lot of obsolete or even outdated information stays online, because no one dares to delete it. Instead of a binary model (online vs. archived, keeping vs. deleting), an institutional information management system must enable seamless access to recently phased-out content, including changing its status back to active. The seamless access in turn fosters a bolder attitude to removing information from active content, leading to a significant reduction of content debris.

² <http://nost.gsfc.nasa.gov/isoas/>

2.3 Research Challenges

Although inspired by human memory, managed forgetting is meant to complement rather than copying the process of human remembering and forgetting. In particular, we envision an idea of gradual forgetting, where complete digital forgetting is just the extreme and a wide range of different levels of condensation for preservation is foreseen. This concept is expected to both help in preservation decisions and to create direct benefits for active information use. The aim is to strike a balance between preservation and managed forgetting, taking into account constraints for digital forgetting (e.g. legal regulations) and help the user to decide what to preserve. Managed forgetting also offers an immediate benefit from adopting preservation by helping to keep the active information spaces more focused. This is a challenging task implying three key challenges:

- An interdisciplinary concept for flexible and gradual **managed forgetting** that meets **human expectations** and is driven by the goal of the digital memory complementing human memory;
- Development of flexible and multifaceted **information value assessment** methods in support of managed forgetting;
- Development of adequate **forgetting actions** for textual and multimedia content, e.g., summarization, aggregation, detection of redundancy, and consideration of diversity;
- Understanding of **constraints** for forgetting, e.g. legal regulations, as well as **societal and individual implications** of forgetting (and not forgetting) information.

Acknowledgments This work was partially funded by the European Commission Seventh Framework Program under grant agreement No.600826 for the ForgetIT project (FP7 / 2013-2016).

References

1. E. Cutrell, S. T. Dumais, and J. Teevan. Searching to eliminate personal information management. *Commun. ACM*, 49(1):58–64, 2006.
2. N. Kanhabua, C. Niederée, and W. Siberski. Towards concise preservation by managed forgetting: Research issues and case study. In *Proceedings of the 10th International Conference on Preservation of Digital Objects*, iPres '2013, 2013.
3. S. E. Lindley, C. C. Marshall, R. Banks, A. Sellen, and T. Regan. Rethinking the web as a personal archive. In *Proceedings of the 22nd international conference on World Wide Web*, WWW '13, pages 749–760, 2013.
4. V. Mayer-Schönberger. *Delete - The Virtue of Forgetting in the Digital Age*. Morgan Kaufmann Publishers, 2009.
5. H. M. SalahEldeen and M. L. Nelson. Losing my revolution: how many resources shared on social media have been lost? In *Proceedings of the Second international conference on Theory and Practice of Digital Libraries*, TPD L '12, pages 125–137, 2012.
6. F. Upward. Modelling the continuum as paradigm shift in recordkeeping and archiving processes, and beyond - a personal reflection. *Records Management Journal*, 10(3):115–139, 2000.