Springer Series on Cultural Computing

Vasileios Mezaris
Claudia Niederée
Robert H. Logie *Editors*

Personal Multimedia Preservation

Remembering or Forgetting Images and Video

Springer

Springer Series on Cultural Computing

More information about this series at http://www.springer.com/series/10481

Vasileios Mezaris · Claudia Niederée
Robert H. Logie
Editors

Personal Multimedia Preservation

Remembering or Forgetting Images and Video

Springer

Editors
Vasileios Mezaris
Centre for Research and Technology Hellas
Thermi, Thessaloniki
Greece

Robert H. Logie
University of Edinburgh
Edinburgh
UK

Claudia Niederée
University of Hannover
Hannover
Germany

ISSN 2195-9056 ISSN 2195-9064 (electronic)
Springer Series on Cultural Computing
ISBN 978-3-319-73464-4 ISBN 978-3-319-73465-1 (eBook)
https://doi.org/10.1007/978-3-319-73465-1

Library of Congress Control Number: 2017962551

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

The ease of creating digital content, especially photos and videos in nearly all life situations, has created an explosion in the amount of personal digital content that is continuously generated. Although declining storage device prices as well as cloud and social media services provide short-term solutions for keeping this content, over the years what we once stored is subject to a random form of digital forgetting, or is just not revisited, because the mere numbers make this a tedious task, not to mention the impact of changing life situations.

In this book, we advocate a novel forgetful approach to dealing with personal multimedia content on the long run, which is inspired by the effectiveness of human forgetting as a mechanism for helping us humans to stay focused on the important things. We present the different theoretical foundations, technologies, as well as applications and results of studies that help the reader understand the problems and challenges associated with personal digital preservation, and the solutions that can be developed in response to these challenges.

The book is organized into three main parts. Part I presents the necessary Interdisciplinary Foundations, Part II covers Multimedia Preservation Theory, and Part III discusses Multimedia Preservation in Practice.

Part I: Interdisciplinary Foundations

The first question that is often raised when discussing the topic of digital preservation is whether this is something that is really necessary or important to the average individual. Chapter 1, "Multimedia preservation: why bother?," attempts to answer this very question, by explaining that just storing the content today does not mean that this content will remain accessible and meaningful in the long run. Based on the understanding of this fact, Chap. 1 proceeds to motivating a more intelligent and selective approach to personal multimedia preservation. This approach introduces and combines three key building blocks: (a) "managed forgetting," for focusing on important and useful content inspired by human forgetting and remembering; (b) “contextualized remembering,” for dealing with evolution and keeping content meaningful over time; and (c) “synergetic preservation,” for bridging the gap between active information use and long-term information management.

While Chap. 1 highlights problems and challenges associated with digital preservation, we need to acknowledge that the preservation of information, in general, is not a topic that first appeared in the digital age. To the contrary, it is something that was first dealt with as part of evolution: humans have evolved to be very efficient at preservation of what it is necessary to preserve, and forgetting trivial or irrelevant details when they are no longer needed. In order to allow us to draw inspiration from nature, Chap. 2 looks at remembering and forgetting in human memory. It discusses how the human superpowers of managed preservation and forgetting are achieved, and how a conceptual understanding of human memory function could be used to inspire the design of digital managed preservation and forgetting. Chapter 2 argues that human-inspired digital forgetting is key for achieving a truly synergetic relationship between human and digital memory, and uses a study for exploring and contrasting human management of photographic collections with managed preservation and forgetting of the same photo collection by an example digital system. Altogether, Chap. 2 highlights how understanding the human cognitive function can help us to inspire more useful digital storage systems.

Having discussed how human memory works, Chap. 3 takes us to the opposite side, discussing how computers can understand the digital multimedia content. While for humans understanding what, e.g., a photo depicts is something that comes naturally by just looking at the photo, it is not the same for computers: to them, in the absence of specialized understanding methods, a photo is nothing more than a huge array of bits. Thus, Chap. 3 discusses methods and algorithms that endow computers with capabilities to understand digital content, based on the premise that understanding the digital content is important for subsequently supporting intelligent preservation decisions. The methods discussed in this chapter include (a) photo/video annotation, which refers to the problem of assigning one or more semantic concepts to photos or video fragments; (b) photo/video quality assessment, which refers to the automatic prediction of the aesthetic value of a photo or a video; (c) near-duplicate detection, which aims to identify groups of very similar items in large media collections; and (d) event-based photo clustering and summarization, which concern the selection of the most characteristic photos of a photo collection so as to create a storyline that conveys the gist of this collection.

Part II: Multimedia Preservation Theory

Chapter 4 goes into more depth regarding the intelligent and selective approach to personal multimedia preservation that was sketched in Chap. 1, taking advantage of the insights provided in Chaps. 2 and 3. Specifically, Chap. 4 focuses on a core ingredient of managed forgetting: the assessment of the importance of information items. It introduces two key notions for describing this importance: "Memory Buoyancy," which, in the short-term, considers how information sinks away from the user, and "Preservation Value," which attempts to estimate the future importance of a digital resource in the long run. Chapter 4 then proceeds with outlining methods for Preservation Value computation for different exemplary settings. It also discusses managed forgetting beyond assessing the importance of information items, that is, methods that can be used to implement managed forgetting on top of the values for information importance. This includes methods such as information hiding, forgetful search, summarization and aggregation, as well as deletion.

Making informed assessments of the importance of information items and decisions about their preservation is a big first step, but even this does not ensure that the preserved information will remain understandable and relevant in the long run. For this, Chap. 5 looks into contextualization methods. Fully understanding digital objects often requires knowing the wider context: for example, a family photo is practically useless if you do not know who are the people portrayed. This becomes even more important when considering the long-term preservation of documents, as not only is human memory fallible, but over long periods the people accessing the documents will change. Chapter 5 discusses methods for preserving the context associated with a digital item, and for assessing how this context evolves over time. It looks in detail at the relevant challenges and describes the development of a conceptual framework in which context information can be collected, preserved, evolved, and used to access and interpret documents. A number of techniques are presented showing real examples of context in action that fit within the framework, and applying to both text documents and image collections.

Chapter 6 takes the discussion of the Preserve-or-Forget (PoF) approach introduced in this book to the system level. It proposes a reference model—the PoF Reference Model which incorporates the techniques discussed in previous chapters for Preservation Value assessment, contextualization, etc., while at the same time paying special attention to the functionality which bridges between an Information Management System and a Digital Preservation System (DPS). The design of the PoF Reference Model was driven by the identification of five required characteristics: it has to be integrative, value-driven, brain-inspired, forgetful, and evolution-aware. The proposed PoF Reference Model consists of three layers; Chap. 6 goes on to discuss the main functional entities and the representative workflows of each of them, relating them to existing standards and practices in digital preservation. It also presents an architecture and an exemplary implementation for a system based on the PoF Reference Model.

Part III: Multimedia Preservation in Practice

Based on the foundations and methods presented in the two previous parts of this book, Chap. 7 presents the integration of preservation functionalities in a Personal Information Management (PIM) application. In this application, the "semantification" of the user's resources paves the way for more effective functionalities for automated preservation, forgetting, and remembering embedded in the daily activities of a user. The chapter also details a pilot based on this application, looking in depth into user activities such as photo organization, and the generation of diaries to remember past events. It investigates how forgetting functionalities can be embedded in applications and describes how different variants for forgetting are used in the pilot. Chapter 7 concludes with a discussion of experience of using the pilot in daily work.

Chapter 8 continues with investigating the application of the methods presented earlier in this book to the daily activities of users. In the first part of this chapter, a user study on a photo selection task is presented. Participants are asked to select subsets of the most important pictures from their own collections. Because evaluating the importance of photos to their owners is a complex process, which is often driven by personal attachment, memories behind the content, and personal tastes that are difficult to capture automatically, this study allows us to better understand the selection process. Then, based also on the findings of this study, the second part of this chapter presents methods for automatically selecting important photos from personal collections. Photo importance is modeled according to what photos users perceive as important and would have selected, and an expectation-oriented method for photo selection is presented, where information at both photo- and collection-level is considered to predict the importance of photos.

Thessaloniki, Greece
October 2017

Vasileios Mezaris
Claudia Niederée
Robert H. Logie

Acknowledgements

Most of the work reported throughout this book was supported by the EC's Seventh Framework Programme for Research, under contract FP7-600826 "ForgetIT: Concise Preservation by combining Managed Forgetting and Contextualized Remembering", 2013–2016.

Contents

Contributors

Walter Allasia EURIX Srl, Torino, Italy

Konstantinos Apostolidis Information Technologies Institute (ITI), Centre for Research and Technology Hellas (CERTH), Thermi, Thessaloniki, Greece

Andrea Ceroni L3S Research Center, Leibniz Universität Hannover, Hanover, Germany

Francesco Gallo EURIX Srl, Torino, Italy

Mark A. Greenwood The University of Sheffield, Sheffield, UK

Christian Jilek German Research Center for AI (DFKI), Kaiserslautern, Germany

Nattiya Kanhabua NTENT Inc., Barcelona, Spain

Robert H. Logie University of Edinburgh, Edinburgh, Scotland, UK

Foteini Markatopoulou Information Technologies Institute (ITI), Centre for Research and Technology Hellas (CERTH), Thermi-Thessaloniki, Greece; Queen Mary University of London, London, UK

Heiko Maus German Research Center for AI (DFKI), Kaiserslautern, Germany

Vasileios Mezaris Information Technologies Institute (ITI), Centre for Research and Technology Hellas (CERTH), Thermi, Thessaloniki, Greece

Kaweh Djafari Naini L3S Research Center, Hannover, Germany

Claudia Niederée L3S Research Center, Hannover, Germany

Elaine Niven University of Dundee, Dundee, UK

Ioannis Patras Queen Mary University of London, London, UK

Sven Schwarz German Research Center for AI (DFKI), Kaiserslautern, Germany

Nam Khanh Tran L3S Research Center, Leibniz Universität Hannover, Hannover, Germany

Tuan Tran L3S Research Center, Hannover, Germany

Christos Tzelepis Information Technologies Institute (ITI), Centre for Research and Technology Hellas (CERTH), Thermi-Thessaloniki, Greece; Queen Mary University of London, London, UK

Maria Wolters University of Edinburgh, Edinburgh, Scotland, UK

Part I
Interdisciplinary Foundations

Chapter 1
Multimedia Preservation: Why Bother?

Claudia Niederée, Vasileios Mezaris, Heiko Maus and Robert H. Logie

Abstract Multimedia content and especially personal multimedia content is created in abundance today. Short- to mid-term storage of this content is typically no problem due to decreased storage prices and the availability of storage services. However, for the long-term perspective, i.e., preservation, adequate technologies and best practices for keeping the content accessible and meaningful are still missing. Instead, the breakdown of devices and changes in technologies lead to some form of random survival and random forgetting for digital content. In this chapter, we motivate a more intelligent and selective approach to personal multimedia preservation. This approach introduces and combines three key building blocks: (a) "managed forgetting" for focusing on important and useful content inspired by human forgetting and remembering; (b) "contextualized remembering" for dealing with evolution and keeping content meaningful over time; and (c) "synergetic preservation" for bridging the gap between active information use and long-term information management.

C. Niederée
L3S Research Center, Hannover, Germany
e-mail: niederee@L3S.de

V. Mezaris (✉)
Information Technologies Institute (ITI)/Centre for Research and Technology Hellas (CERTH), 6th Km. Charilaou-Thermi Road, 57001 Thermi-Thessaloniki, Greece
e-mail: bmezaris@iti.gr

H. Maus
German Research Center for AI (DFKI), Kaiserslautern, Germany
e-mail: heiko.maus@dfki.de

R. H. Logie
University of Edinburgh, Edinburgh, Scotland, UK
e-mail: rlogie@ed.ac.uk

V. Mezaris et al. (eds.), *Personal Multimedia Preservation*, Springer Series on Cultural Computing, https://doi.org/10.1007/978-3-319-73465-1_1

1.1 Abundant Creation and Random Survival

With the advent of digital photography, taking photos and videos is nearly effortless and requires few resources, and the widespread use of smartphones with continuously improving camera features and storage capacities has accelerated this trend. A further trigger for growing multimedia content creation is social media with its multifaceted opportunities and incentives for content creation and sharing. In addition, nowadays, taking photos and videos is tolerated nearly everywhere. Hundreds of photos are easily taken by an individual participating in public events, such as concerts, as well as in private events such as a holiday trip. Furthermore, photos and videos are also very often taken of more mundane aspects of life, such as food or in support of everyday activities such as shopping, further increasing the amount of personal multimedia content to be dealt with. Thus, personal multimedia content is created in abundance and with a tendency for further growth.

This raises the question of what happens to all of this personal multimedia content, which has to be considered in the short-term, mid-term, and the long-term perspective. With the decreasing price of storage media, it is not really a problem to store all of this content; either in traditional digital storage devices such as hard disks or also on the cloud, since storage is offered as a service by several cloud storage providers. This situation fosters the adoption of a keep-it-all strategy for the short- to mid-term perspective, where the majority of content created is kept "somewhere" (more or less systematically organized).

However for the mid- to long-term perspective, just storing all of this content often ends up as a kind of "dark archive" of photo, video, and other content collections, which are rarely accessed (and enjoyed) again. The mere size of the collections makes going through them as well as sorting or annotating them a tedious task.

Furthermore, for the long-term perspective, there is the risk of losing personal content by a random form of "digital forgetting" [191]: Over decades, storage devices such as hard disks may break down, and employed storage media become subject to decay, loss, and accidental destruction or even theft. Moreover, even cloud storage is vulnerable if the companies providing the service go out of business. Furthermore, with the development and adoption of new technologies, existing formats and storage media quickly become obsolete. These developments make random parts of personal collections inaccessible. Just consider, for instance, how difficult it would be today to access photos stored years ago in .mos format in a floppy disk, or that even accessing photos stored much more recently in a DVD is not that straightforward any more, given that your new ultrabook laptop does not cater for reading this type of external media. This leads to a random form of survival for personal multimedia content. This weakness of digital content of being seemingly easy to keep but also easy to "lose" has raised discussions about "digital dark ages" already in the late 90s [51, 212]. It is possible to read a 200-year-old book, or look at 100-year-old printed photos, but in a hundred years from now, will the technology exist to read the file and media formats being used today? And, even if pure format readability is ensured by timely transformation to more up-to-date formats and copying in contemporary storage media, more investment is still required for ensuring long-term interpretability of content. Over the years, we might simply forget what or whom a photo shows or why it was taken.

What is actually missing are best practices and supporting technologies for dealing with personal multimedia content in the long run. Both the risk of dark archives and of digital forgetting highlight the need to select, supported by automated methods, the most important content and to invest some effort into keeping them enjoyable and accessible not only over a lifetime but possibly also for future generations.

1.2 State of Affairs in Personal Preservation

"Digital Preservation"—i.e., secure long-term storage of content, considering time frames of decades and longer—is a systematic approach for avoiding random digital forgetting. It embraces tools, technologies, as well as organizational aspects.

However, while preservation of digital content is now well established in memory institutions, such as national libraries and archives, it is still in its infancy in most other organizations, and even more so for personal content. There are several obstacles for the wider adoption of preservation technology in organizational and personal information management: There is a considerable gap between active information use and preservation activities. Active information use refers to dealing with information objects for everyday private or professional activities, typically supported by some information management environment, such as a content management system in an organization or a desktop environment in the context of personal information management. In addition, especially in personal information management, there is typically little awareness for preservation. Although the need for personal preservation has been recognized in theory (e.g., [271]), this has not yet propagated to more practical settings and solutions. As a consequence, readiness for investing considerable resources in terms of time and money for preservation is low. Finally, establishing effective preservation and concise and usable archives still requires a lot of manual work for selecting content that is relevant for preservation and for keeping the archives accessible and meaningful long term, thus entailing expenses much larger than just the storage costs. This is further aggravated by the fact that no benefits are seen for moving from more or less systematic backup to systematic preservation.

A personal information space consists of a substantial number of information objects connected to the person's life such as wedding videos, travel pictures, or graduation keepsakes. It requires serious dedication and cognitive effort to organize all these data and keep them accessible as time passes. Moreover, these digital artifacts often represent past moments but are not associated with a physical memento. Therefore, they form valuable resources for the user and future generations. If the material is lost or corrupted due to improper conservation, it will be useless and memories might be lost. Most users still use backups as their main form of preservation. John et al. in [186] surveyed 2600 academics and members of the digital public about their preservation strategies. 60% of the respondents relied on backups. If data were lost, which happened to 30% of the participants, the most common reason for the loss (70% of all cases) was inability to find the files again.

Furthermore, many people follow the keep-it-all strategy. Marshall [271] points out five main reasons: (a) assessing value of resources in advance is difficult, (b) keep-it-all is currently the best practice, (c) deletion and (d) sorting out resources are cognitively demanding exercises, and (e) archived information resources play an important role as memory prosthesis.

There are preservation guidelines aimed at the general public that show how to go beyond backups. For example, the Library of Congress raises awareness of personal archiving solutions on their website and provides practical information [232, 234]. The recommended steps[1] are as follows:

1. Identify what you want to save.
2. Decide what is most important to you.
3. Organize the content (descriptive file names and folders).
4. Save copies in different places.
5. Manage your archive (including migration plans).

These recommendations are very helpful but leave all steps to the user, i.e., what to save, how to organize, where to store (hard disk and online storage), and when to migrate. This puts a high burden on the user: various decisions need to be made and it requires discipline in, e.g., maintaining and updating the archive.

This cognitive up-front effort is one of the reasons why the cloud storage offered by DropBox, Microsoft OneDrive, or Google Drive is not a preservation system in itself, but only a tool in a larger preservation strategy. Started as syncing, file sharing, and backup solutions, those services offer structuring methods such as file folders or (keyword-) tags, but do not comply with preservation best practices such as the Open Archival Information System (OAIS) standard [68]. Other services, such as Amazon Cloud, comply with OAIS, but do not support users before ingesting data into the store. Either way, users are left on their own for large parts of the preservation process. An overview of the preservation functionality of major cloud services can be found in [331].

Another issue is preserving social media content. While service providers carefully store users' data for in-depth analysis, they often do not provide any support for dedicated archiving and preservation. Even when archivists are called in at a later stage, preservation is not optimal, since it is not part of the business model [308]. As shown, for example, in the survey reported in [189], users often post information in social media, which might be worth preserving in a personal archive such as information about lifestyle and data about travel, festivities/parties, and funny events. Many services have emerged around curating social media content in a form that is easily accessible, and would lend itself well to further preservation. At the most basic level, Twitter allows users to download a compressed file that includes all of their tweets, which can then be viewed in a browser. The Storify service allows users to curate

[1]There are some specializations depending on the media, e.g., video or social media.

conversations on Twitter—tweets that are replies to each other are collated into a single web page and can be annotated further [81].

Having a look at services like Flickr where people collect and share digital moments, or services such as Twitter that focus more on real-time short messaging, long-term preservation is not the first thing which comes into one's mind. Nevertheless, Twitter also treats concerns on archiving: they introduced a service to download a zip archive of the user's tweets. Furthermore, Google and Facebook have introduced formal ways of handling an account in case a user is deceased, such as handing it over to a dedicated contact person with full or partial access [57].

1.3 Preserve or Forget: Managed Forgetting and Digital Preservation

To ease the adoption of more robust and beneficial preservation practices for personal multimedia content, we propose the introduction of a radically more adoptable and sustainable approach, the "Preserve-or-Forget" approach to intelligent preservation management, which combines three novel concepts:

- Inspired by the role of forgetting in the human brain, we envision a concept of **managed forgetting** for systematically dealing with information that progressively ceases in importance and finally becomes obsolete, as well as for redundant information. This concept is expected to help in preservation decisions and to create direct benefits for active information use. At first glance, forgetting seems to contradict the idea of preservation: Preservation 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 in the digital world today. As discussed above, this is triggered, for example, by changing hardware, hard disk crashes, technology evolution, and changes in life circumstances. We aim to replace such random forgetting processes with managed forgetting, where users are optimally supported in their explicit decisions about what to keep, and how what is kept to be organized and preserved. In particular, we envision an idea of gradual forgetting, where complete digital forgetting is just the extreme, and a wide range of different forgetting actions such as summarizing are foreseen. This draws on the principles of a highly efficient process of forgetting in the human brain for information that is trivial, redundant, or only required on a single occasion for a short time.
- For bridging the chasm that still separates active information use from content preservation activities, we envision the concept of **synergetic preservation**, which couples information management and preservation management, making the intelligent preservation integral to the content life cycle in information management. This clearly supports easier adoption, and by enabling a rich information flow from the information context to the preservation context more intelligent and informed preservation decisions, e.g., for preservation selection and contextualization.

- To bring preserved information back into active use in a meaningful way, even if a long time has passed since their transition into the archive, we envision the concept of **contextualized remembering**, again inspired by processes in the human brain. The idea here is to already equip resources with rich context information when packaging them for preservation (thus preparing them for long-term interpretation) and to gradually evolve this context information, reflecting the evolution in terminology, semantics, and interpretation context, thus reaching a semantic level of preservation.

The vision of our Preserve-or-Forget approach is a transition from pure archives to managing and preserving concise knowledge ecosystems, coupling information management and preservation. Realizing such an ecosystem requires a concise, diversity-aware, and evolution-aware preservation approach which includes a careful selection of what to preserve taking into account coverage, diversity, importance and overlap/redundancy of information, the explicit contextualization of preserved resources into self-contained objects to ensure long-term interpretability, as well as adequately dealing with evolution and with information becoming obsolete.

In this book, we discuss the conceptual foundations, architectural aspects, as well as effective methods for implementing the Preserve-or-Forget approach. Many of the discussed methods and technologies have been developed in the European project ForgetIT, in which the editors and authors of this book were involved.

Chapter 2
Preserving and Forgetting in the Human Brain

Robert H. Logie, Maria Wolters and Elaine Niven

Abstract Humans have evolved to be very efficient at managed preservation of what is necessary to preserve. Humans are also extremely efficient at forgetting trivial or irrelevant details when they are no longer needed. Indeed, managed preservation and forgetting could be viewed as a set of human 'superpowers' achieved through use of a lifetime of accumulated knowledge, highly effective contextualisation, aggregation, organisation, summarisation and reconstruction of key features of experiences. But humans are poor at preservation of large amounts of detail. Typically, memories are partially reconstructed during the retrieval process, and this reconstruction process can sometimes lead to false memories. Many of these strengths and limitations of human memory are well understood by human memory researchers, although important questions and uncertainties remain. In complete contrast, digital systems excel in preserving large amounts of detail, and are getting better at contextualisation. But they remain rather poor at systematic forgetting of irrelevant detail. Often, digital forgetting occurs by accident through disk crashes, incompatible upgrades of software and hardware, lost or stolen storage devices. Even if the data are still present and safely stored, insufficient indexing and poor information retrieval may result in those data effectively being forgotten. This chapter will provide a detailed overview of the state of the science of human memory, based on empirical studies and conceptual modelling. It will discuss how the human superpowers of managed preservation and forgetting are achieved, and show how a conceptual understanding of human memory function could be used to inspire the design of digital managed preservation and forgetting. It will argue that human-inspired digital forgetting is key for achieving a truly synergetic relationship between human and digital memory, and explore how such a synergetic relationship can address aspects of the paradox that massive investment in technology has not necessarily led to the expected increase in

R. H. Logie (✉) · M. Wolters
University of Edinburgh, Edinburgh, Scotland, UK
e-mail: rlogie@ed.ac.uk

M. Wolters
e-mail: maria.wolters@ed.ac.uk

E. Niven
University of Dundee, Dundee, UK
e-mail: e.niven@dundee.ac.uk

V. Mezaris et al. (eds.), *Personal Multimedia Preservation*, Springer Series on Cultural Computing, https://doi.org/10.1007/978-3-319-73465-1_2

productivity (IT/productivity paradox). Next, we will describe an in-depth study of personal, digital photograph collections that were contributed by volunteer research participants. This study explored human management of photographic collections and contrasted it with managed preservation and forgetting of the same photo collection by an example digital system that incorporates automated conceptualisation and forgetting. The chapter will conclude with a summary of how understanding human cognitive function can help to inspire more useful digital storage systems that offer reliable and usable tools to complement and support human memory rather than attempt to replace it.

2.1 Human Memory and Forgetting

2.1.1 What Is Human Memory?

Human memory takes many forms and serves a wide range of purposes that are essential for humans to function in everyday personal and working life. Among the lay public, it is most widely associated with preservation and retrieval of information about public and personal events. However, scientific study takes a much broader view of human memory to include the acquisition, preservation and retrieval of knowledge and skills (semantic memory and procedural memory), events and experiences across a person's lifetime (episodic memory), and remembering to carry out intended actions (prospective memory). It also applies to the temporary storage and moment-to-moment updating of information required for a focus on the current task, an ability known as 'working memory'. Finally, it applies to a range of control functions that can suppress or inhibit information that is irrelevant or redundant and that can detect or recognise whether information has been encountered previously or is linked with previously preserved information.

2.1.2 Research Methods for Human Memory

Each individual has different experiences and different memories of those experiences, as well as different knowledge and skills accumulated over their lifetime. Also, people vary in the efficiency with which they can learn and retrieve new knowledge, and encode, forget, preserve or retrieve details of events that they have experienced. However, the study of human memory is based on the assumption that the general principles that govern human memory organisation and function are the same across all healthy human adults, and are the result of the effect of evolution on the human brain. So, there are common principles across all humans for learning, forgetting, encoding, preserving, retrieval and so on. Many of these principles also apply to many animals. An analogy would be that the principles governing the function of other aspects of physiology, such as the heart, the liver, the kidneys, the lungs, the immune system or the endocrine system, are the same across all healthy human adults, even if these systems differ in their efficiency between individuals. On this principle, if we are collecting objective memory performance data, we need only study the principles of functioning of human memory in a single healthy adult and these principles should generalise to all healthy human memory.

However, experimental data are inevitably noisy and an individual who is studied might have some underlying anomaly such as subtle undetected brain damage, or be simply uncooperative. Therefore, in practice, multiple healthy participants are recruited and allocated randomly to different experimental groups for the purposes of comparison. Using this approach, the numbers in each group need not be large, although there is usually a matching in the age range and educational level to mitigate the variability from differences in memory efficiency. Data are then averaged and analysed statistically across participants within each group to reduce the impact of possible idiosyncrasies of any one participant. In contrast to studies that involve subjective opinions, self-report or survey data, in experimental studies of this kind, there is less emphasis on sampling from cross sections of the population. Therefore, most research on human memory involves designing and running experiments with human volunteers who are presented with material and are subsequently tested on their memory for that material.

Experiments are designed according to theoretical, conceptual models of how a specific aspect of human memory might function, and conclusions are drawn from detailed analyses of the pattern of memory errors that result from different experimental manipulations or different kinds of material. Some of the experiments involve assessing the processes and accuracy of retrieval of real-life events, whereas others involve relatively artificial materials. There is also a large amount of research of this kind with volunteers who have suffered specific forms of brain damage, and these studies can reveal some of the characteristics of healthy human memory as well as the nature of memory impairments from which the patients suffer. Other experiments involve exploring the patterns of brain activation while human volunteers are completing memory tasks. The development of the theoretical, conceptual models is driven by the patterns of results from these experiments, and converging evidence across studies.

2.1.3 General Principles of Human Memory

The research approach described in the previous section has generated a large volume of evidence for some general principles of memory function. Most human memory researchers agree about these principles (for an overview see [27]), but there are ongoing debates about the details of the conceptual models of memory and the interpretation of patterns of results. As a result, there is currently no universally accepted conceptual model of human memory. An example of one conceptual model of human memory is illustrated in Fig. 2.1. Some of the details of this model have

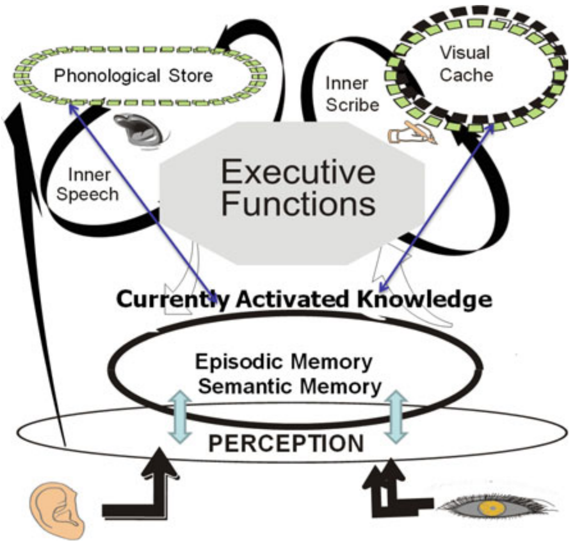

Fig. 2.1 A Conceptual Model of Human Memory

been developed by the first author of this chapter (e.g. [251–253]), based on a simpler model originally proposed in [29]. However, it has characteristics that are similar to other contemporary conceptual models.

In summary, this conceptual model indicates that information from auditory, visual and other forms of perception (e.g. tactile) activates stored knowledge accumulated over a lifetime regarding knowledge about the world and about the self ('semantic memory') and preserved information about individual events ('episodic memory') related to the perceived stimuli. Some of the activated knowledge is held on a temporary basis in a collection of interacting, domain-specific temporary memory systems or components of working memory, and processed by a range of executive functions. For example, combinations of meaning, shape and sound may be held together as Currently Activated Knowledge. Details of recently perceived stimuli that have been seen or heard may be held as sound-based codes in the phonological store component or as visually based codes in the Visual Cache component. Both types of code decay within around 2 s, but the inner speech component can allow the sound-based codes to be held for longer by mentally repeating the sounds. The Inner Scribe component holds and can mentally rehearse sequences of movements and can allow visual codes to be held longer by mentally rehearsing the codes held in the Visual Cache.

It should be noted that the theoretical, conceptual models of human memory such as the one shown in Fig. 2.1 are used as frameworks to generate hypotheses and to guide the design of memory experiments. They are not formal computational models that have clearly defined characteristics for each component, or that describe precisely what information flows along the arrows between components. There are some formal computational models of specific functions of human memory (for example of the phonological store and inner speech shown in Fig. 2.1), and these are used to run simulations of the behavioural data patterns obtained from memory experiments with human volunteers. However, these formal models are beyond the scope of the current chapter. Figure 2.1 is included here to set a context for the reader who is unfamiliar with the approaches and style of research summarised later in this chapter.

2.1.4 Semantic and Episodic Memory

A key distinction is between semantic and episodic memory. Semantic memory includes knowledge acquired through life experience (e.g. language, facts about the world, people and the self) and learned skills, sometimes referred to as procedural memory (e.g. swimming, riding a bicycle, mathematics). Episodic memory refers to memory for specific experiences of events that took place at a particular place and at a particular time (e.g. a particular holiday, meeting, lecture or social event or what you had for dinner yesterday). In [394], the authors introduced the concept of episodic memory as a system that underlies the 'what-when-where' specifics of an event, and as such is distinct from factual knowledge in semantic memory [391]).

Forgetting from episodic memory is rapid and substantial. Forgetting from semantic memory is much less rapid and information is well preserved over long periods or never lost. Semantic memory is thought to develop across the lifetime by extracting features that are common across similar events, and building what are known as schema for specific types of events, a concept first proposed by Bartlett [34]. Details of the occasion on which the information was first encountered are forgotten.

For example, a restaurant schema includes tables, menus, food, conversation, waiting staff and paying a bill, but we probably cannot recall when we first learned these features of restaurants. In the same way, we know that the capital of France is Paris, but are unlikely to remember when that fact was first encountered. This means that the schema sets the general context for each restaurant visit, and provides a 'framework' on which to build the memory for key details of specific visits to restaurants. The features that are common to each restaurant visit need not be stored on every occasion. The same is true of any common experience, such as a working day, a visit to a swimming pool, a train or aeroplane journey. The framework or context can then be used to aid retrieval of information about specific events. When recalling a restaurant visit, we can assume that there was a menu, food, a table, a bill, etc., and so only have to store and retrieve the key information such as who else was at the table, and what was important about the conversation or about the food. In summary, human memory tends to preserve generic information that is repeated across similar experiences and events without 'tagging' that information with a time and place. Human memory tends to forget details that are unique to individual experiences or events, except when the unique features of a particular event are particularly important for the individual.

A further role for a schema or context is in the understanding or interpretation of presented information or events. Take, for example, the following paragraph.

> The procedure is actually quite simple. First, you arrange the items into different groups. Of course, one pile might be sufficient depending on how much there is to do. If you have to go somewhere else due to lack of facilities, that is the next step; otherwise, you are pretty well set. It is important not to overdo things. That is, it is better to do too few things at once than too many. In the short run, this may not seem important but complications can easily arise. A mistake can be expensive as well. After the procedure is completed, one arranges the material into different groups again. They then can be put into their appropriate places. Eventually, they will be used once more and the whole cycle will then have to be repeated [52, 134-5].

This passage is difficult to understand and also is difficult to remember because there is no context. Most of the material from the passage will already have been forgotten as you are reading this sentence. However, after the context for the passage is given as ‘washing clothes’, then the interpretation of the text is trivial and memory for the sequence of procedures can be generated from existing knowledge in the schema without having to remember the exact wording. The use of the schema can be repeated every time this kind of activity is required, and precise details of each occasion do not have to be preserved in memory, unless, for example, an error on one occasion is to be avoided on future occasions, or a small change to the procedure results in a benefit that should be remembered for future repetition.

However, even when a context is available, substantial forgetting of detail occurs within minutes. For example, although the general meaning of the washing clothes paragraph could be regenerated, the precise wording is unlikely to be remembered accurately. Likewise, within a few seconds of reading the text of the current paragraph, any reader will have forgotten the exact wording used but will remember the meaning of the text. For readers who are not already familiar with the topic of how human memory functions (i.e. have no accurate or detailed existing schema or context), many of the detailed facts presented in this chapter will be forgotten within an hour after it has been read, unless this material is relearned before this forgetting occurs (e.g. [195]).

Therefore, a great deal of information concerning an event is never stored in memory. Because there is a large number of schemata and a large amount of information accumulated in semantic memory over the lifetime of each individual, the human memory system can select what information is necessary to set the context for the current environment or information presented, and can inhibit or ignore information that is irrelevant or can be assumed from the context. This aspect of human memory is a major strength in that it avoids the distraction of information that is irrelevant or redundant for the current task, and avoids the storage of large amounts of irrelevant or redundant information, making it very efficient for storing and retrieving key details about an event, or retrieving key facts and skills that are required for the current task.

2.1.5 Forgetting from Episodic and Semantic Memory

The process of forgetting from semantic or episodic human memory typically refers to the inability to retrieve information that has previously been stored, and this is often

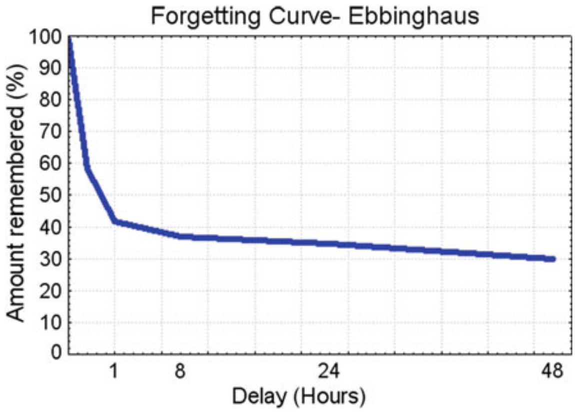

Fig. 2.2 Human forgetting over time

viewed as an unwelcome limitation. However, detailed analysis shows forgetting to be more complex, and to be a benefit to humans most of the time. As should be clear from the previous section, a substantial amount of detail is never encoded in memory. It is equally well established that of the details that are encoded, a substantial amount is forgotten within a short time after the initial experience. This prevents the memory system from being filled with information for which there is no clear context, or that is largely irrelevant, or which is required only on a temporary basis, and preservation is not normally required. Consequently, only information that is important for understanding and functioning in the world tends to be preserved.

The forgetting of information that lacks context was first subject to systematic study by the German researcher Ebbinghaus [122] who experimented with learning and remembering 'nonsense material', specifically three-letter syllables (e.g. BAZ FUB YOX DAX LEQ VUM . . .) for which he had no established schema. Therefore, this kind of material was selected to assess 'pure' episodic memory without the support of semantic memory. In his experiments, he would spend several minutes trying to learn sequences of these nonsense syllables, and then tested his memory by attempting to relearn the material at different time periods after the initial learning. Typical results from his experiments are shown in Fig. 2.2. It is clear from the figure that most of the forgetting had occurred within one hour of the learning, but the small amount of material that was retained after one hour was retained for at least 48 h.

Although the work of Ebbinghaus was important for understanding memory and forgetting, it was unclear as to why the material was being forgotten. The research was criticised because of the reliance on memory for nonsense material and because only one individual was tested, namely Ebbinghaus himself. It is rare that adult humans are required to remember material for which they have no schema or context, and as mentioned above, most experimental studies of memory involve an investigation of the aggregate results from groups of individuals. The issue of what might be the main causes of forgetting has been the subject of scientific debate ever since the time of Ebbinghaus, with the major possibilities being decay of the material over time, or other material causing interference with the memory representation.

In the case of decay, information is lost over time through gradual deletion from memory of material that is rarely or never accessed and retrieved. In the case of interference, the forgetting may arise from an inability to retrieve key details of an event because of interference from previously stored details about similar events (‘proactive interference’—e.g. [398], or because of interference from stored details of similar subsequent events (‘retroactive interference’, e.g. [287]. More recent studies have demonstrated interference-based forgetting of a first language when trying to learn a second language [177]), an example of retroactive interference: learning of the new language interferes with memory for the previously learned language. Other studies have shown that parking a car in different spaces in the same car park multiple times (e.g. at work or near a retail centre) can make it difficult to remember where the car was parked today [97]. This is a common experience and often suggests to people that their car has been stolen until they realise that they are looking in the space that they used for their car yesterday or last week. This is an example of proactive interference: multiple similar previous experiences interfere with the ability to remember details of the most recent instance of this experience.

The Ebbinghaus forgetting function applies beyond the forgetting of context-less material. It also applies to the forgetting of material that can be supported by context from semantic memory. Take, for example, the results from a study published 100 years after the Ebbinghaus studies. McKenna and Glendon [288] tested memory in people who had undertaken and successfully completed a first aid course. At intervals varying from 3 to 36 months, they were tested on their memory for their ability to diagnose the health problem associated with particular symptoms, their resuscitation technique and performance as well as on a total score for the knowledge they had retained from the course. Despite spending several days on the first aid course, and passing the test at the end of the course, within 3 months they had forgotten 70% of their knowledge about diagnosis, and after 6 months they had forgotten 60% of even their best preserved ability, namely their technique for cardiopulmonary resuscitation (CPR). However, over the following 30 months, the rate of forgetting was very much slower than it was during the first 6 months. In the case of Ebbinghaus, learning took place over a few minutes with no schema or context, and forgetting was over periods of minutes and hours. In the McKenna and Glendon study, learning took place over several days and involved information and skills within the schema or context of first aid care. In this latter case, the forgetting was over periods of months rather than hours. So, context as well as amount of initial learning greatly slows down the speed of forgetting. However, the shapes of the forgetting functions were remarkably similar, even if over different time periods. Equivalent results were found in [31] for English native speakers remembering Spanish learned at school over delays of up to 50 years, and retention of information learned at University up to 30 years later [88]. In both studies, there was substantial forgetting within the first few years after leaving the formal learning environment, but then a much slower rate of forgetting thereafter.

These well-established studies show that when memory is supported by context or schema, then the material can be retained for periods of months or years, but even with this support, most of the forgetting of details still occurs within a relatively

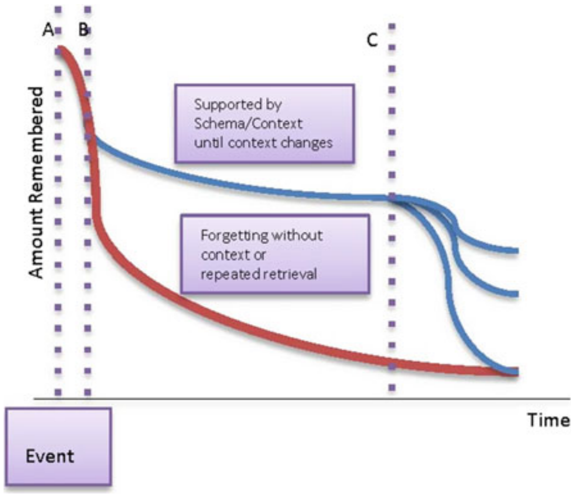

Fig. 2.3 Forgetting with context support versus forgetting without context support

short period and the material that remains after that initial period is forgotten much more slowly. If never 'relearned' from time to time, all of the information may be forgotten. For example, the people in the study in [31] who used Spanish in their daily lives after leaving school retained their knowledge of Spanish very much more successfully than those who had few subsequent opportunities to practice using the Spanish they had learned. The same was true of the material learned at university in the study in [88]. The difference in forgetting supported by context and without such support is illustrated in Fig. 2.3.

A further everyday example of the support from context (also illustrated in Fig. 2.3) is remembering the number of a hotel room for the period of staying in the hotel, or remembering a flight number. The context of being in a particular hotel is supported by the repeated requirement to retrieve the room number when asking for the key, going into breakfast, or returning to the room, but after leaving the hotel there is no requirement to retrieve the room number and so it is forgotten. The same is true for the flight number which need only be retrieved while at the airport but is not retrieved again after the travelling is complete, and so is forgotten. Paper and electronic aids are of course extremely useful in these circumstances, and are used widely to avoid the need to retain this kind of information in memory even on a temporary basis.

Context and schema work well in supporting memory most of the time, but because much of memory retrieval involves reconstruction of details based on schema rather than actual memory for details, the reconstruction process can result in major errors and false memories that the individual is convinced are genuine. For example, a witness to a crime or accident can have a false memory for details of the people present or of the incident. These false memories can arise because people assume 'what must have happened' based on their schema for such events rather than what actually happened. False memories also arise because of subsequent experiences (retroactive interference) and can result in accusing an innocent bystander of being