www.forgetit-project.eu

ForgetIT

Concise Preservation by Combining Managed Forgetting
and Contextualized Remembering

Grant Agreement No. 600826

Deliverable D3.3

Work-package	WP3: Managed Forgetting Methods
Deliverable	D3.3: Strategies and Components for Managed Forgetting – Second Release
Deliverable Leader	Nattiya Kanhabua (LUH)
Quality Assessor	LTU
Dissemination level	Public
Delivery date in Annex I	31 January 2015
Actual delivery date	28 February 2015
Revisions	2
Status	Final
Keywords	Digital Preservation; Dynamic Information Assessment; Time-aware Information Access; Managed Forgetting

Disclaimer

This document contains material, which is under copyright of individual or several ForgetIT consortium parties, and no copying or distributing, in any form or by any means, is allowed without the prior written agreement of the owner of the property rights.

The commercial use of any information contained in this document may require a license from the proprietor of that information.

Neither the ForgetIT consortium as a whole, nor individual parties of the ForgetIT consortium warrant that the information contained in this document is suitable for use, nor that the use of the information is free from risk, and accepts no liability for loss or damage suffered by any person using this information.

This document reflects only the authors' view. The European Community is not liable for any use that may be made of the information contained herein.

List of Authors

Partner Acronym	Authors
LUH	Claudia Niederée
LUH	Andrea Ceroni
LUH	Kaweh Djafari-Naini
LUH	Nattiya Kanhabua
LUH	Ricardo Kawase
LUH	Tuan Tran
DFKI	Heiko Maus
DFKI	Sven Schwarz

Table of Contents

Executive summary

In the previous deliverables D3.1 and D3.2, we have presented the foundations of managed forgetting, and our first ideas for computing *memory buoyancy*. In this deliverable D3.3, we present the extended components in support of the managed forgetting process including *preservation value* and *policy framework*. Memory buoyancy can be computed using a function of attention. Particularly, we substantially extend the computational method with several time-decay models and also leverage the access logs of a resource (e.g., the number of views/edits of a document), as well as report the performance of our proposed model evaluated using human assessment. Preservation value, on the other hand, is aimed to estimate the expected long-term benefit of a resource. We propose a conceptual model for preservation value, which is composed of its definition and purpose with respect to the two ForgetIT use case scenarios, i.e., personal preservation and organizational preservation. In addition, we present the case studies of preservation value as part of exploratory research, and discuss experimental results and our key findings. A policy framework can be seen as a main driver of the forgettor component and it is aimed at supporting different forgetting strategies by considering memory buoyancy, preservation value, and a set of user-specified rules. In this deliverable, we envision a conceptual model for the policy framework and discuss our activation options for the policies governing the preservation and forgetting process. We propose a computational model for further supporting activities in this task, which includes 1) definition and formats of policies, and 2) tools for defining the policies based on rule-based engines. In more detail, these formats and tools must allow a user to interact or provide a policy declaration (e.g. an option to ask the user if something should be really deleted). Furthermore, we present several system prototypes for advanced information value assessment methods and components in support of preservation value, and the integration of an extended set of managed forgetting options into the process. The proposed methods and system prototypes for managed forgetting have been integrated to the overall PoF Middleware. To this end, we report ongoing research activities and research plans for WP3.

1 Introduction

The goal of WP3 is to develop concepts and methods for managed forgetting and to integrate them into the Preserve-or-Forget framework. The methods aim to meet human expectations and to complement processes of human forgetting and remembering. The development of managed forgetting methods embraces the information value assessment of memory buoyancy (short to mid-term importance) and preservation value (long-term importance), as well as the implementation of forgetting options and strategies.

We have addressed the foundations of managed forgetting in the previous deliverable D3.1 by outlining the state-of-the-art in human and digital remembering and forgetting, as well as presenting several key research questions related to the managed forgetting concept. We explained how managed forgetting methods can complement human remembering and forgetting processes. In the deliverable D3.2, we focused on the computational model for memory buoyancy. We envisioned that managed forgetting can be regarded as functions of attention. We employed concepts in artificial intelligence and Semantic Web, and applied ontology from the PIMO semantic desktop system to define the information space. To this end, we presented our preliminary research results on complementing human memory.

This deliverable (D3.3) addresses the strategies and extended components in support of the managed forgetting process together with a report describing their functionality. Especially, this will include an extended policy framework, a richer set of forgetting strategies, advanced information value assessment methods and components in support of preservation value, and the integration of an extended set of managed forgetting options into the process. In more detail, we will present the extended components including *preservation value* and *policy framework*. Particularly, we substantially extend the computational method of memory buoyancy with several time-decay models and also leverage the access logs of a resource (e.g., the number of views/edits of a document), as well as report the performance of our proposed model evaluated using human assessment. Preservation value, on the other hand, is aimed to estimate the expected long-term benefit of a resource. We propose a conceptual model for preservation value, which is composed of its definition and purpose with respect to the two ForgetIT use case scenarios, i.e., personal preservation and organizational preservation. In addition, we present the case studies of preservation value as part of exploratory research, and discuss experimental results and our key findings. A policy framework can be seen as a main driver of the forgettor component and it is aimed at supporting different forgetting strategies by considering memory buoyancy, preservation value, and a set of user-specified rules. In this deliverable, we envision a conceptual model for the policy framework and discuss our activation options for the policies governing the preservation and forgetting process. We propose a computational model for further supporting activities in this task, which include 1) definition and formats of policies, and 2) tools for defining the policies based on rule-based engines. In more detail, these formats and tools have to enable an option for a user to the policy declaration (e.g. the option to ask the user if something should be really deleted). Furthermore, we present several system prototypes for advanced information value assessment methods, and the integration of an extended set of managed forgetting options into the process. To this end, we report ongoing research activities and research plans for WP3.

In the following, we discuss about the Success Indicators that have been defined in the Description of Work and listed below.

(1) Effectiveness of information assessment for memory buoyancy and preservation value

(2) Efficiency of information assessment

(3) Number of information assessment parameters considered

(4) Publication of assessment algorithms in peer-reviewed venue

The Success Indicator (2) Efficiency of information assessment has been reported in the previous deliverable (see Section 3.3, D2.2). The Success Indicator (1) Effectiveness of information assessment for memory buoyancy and preservation value will be explained in more detail in Section 2 and Section 3, respectively. The Success Indicator (3) Number of information assessment parameters considered will be discussed in Section 3.1.2. For (4) Publication in peer-reviewed venue, we have published recently publications in relevant conferences (i.e., [10, 23, 24, 34, 16, 31, 42, 43]), as well as under-submission (e.g., [7, 8, 32]).

1.1 Deliverable Organization

The detailed organization of the deliverable is outlined below.

- Section 2 presents the extended model for memory buoyancy including a conceptual level and a computational framework.
- Section 3 describes a definition and purpose of preservation value as well as a computation method. We report our exploratory research conducted by two use case studies.
- Section 4 explains our first idea for policy-based preservation including conceptual model and our proposed framework.
- Section 5 presents system prototypes in support of long-term preservation for two applications, i.e., photo summarization and timeline summarization in social media.
- Section 6 describes our ongoing exploratory research and outlines research plans for WP3.
- Section 7 summarizes and concludes the deliverable.

2 Memory Buoyancy

The previous deliverable D3.2 has focused on the study of the information value assessment framework of Memory Buoyancy (MB), one of the two primary concepts proposed within the ForgetIT project. A computational framework for MB values of digital objects has been proposed, which was inspired by two fundamental theories in human remembering, namely, decay and inference models. In the context of resource re-accessing, the model for MB assessment basically follows two flavors (Section 2.2.1, D3.2): Independently estimating the retention of digital objects in human brains by tracing the observed activity logs (Time-decay model), and associating them with their related objects, or with context information in the information space, in order to propagate the estimated MB scores along different connections to other digital objects unobserved by the activity logs (Propagation model). A case study of time-decay model has been conducted, which was based on the Weinbull distribution.

In this deliverable, we substantially extend the study with several new time-decay models, each of which was grounded by relevant research in decay effects of human remembering. We also incorporate the frequency information from activity logs of objects, and evaluate the performance of our proposed computational model using human judgment.

2.1 Conceptual Model

2.1.1 Time-Decay Memory Buoyancy

Time-decay based model estimates the memory buoyancy of a resource according to its temporal distance with the current context, i.e. how long time ago the resource was accessed and active in the information space. This approach considers all resources independently from each other, and uses activity logs as a primary sources of parameter estimation. In the deliverable D3.1, we mentioned Ebbinghaus, the earliest work in this line; the deliverable D3.2 extended with the Weinbull function. In fact, there have been a rich body of decay models proposed in philosophy at different scales and contexts. We first state the necessary condition of a decay function for it to be used to model the memory buoyancy.

Definition 1 *The memory buoyancy of a resource r at a given time t is the function $MB_{\mathbf{T}}(r,t)$ satisfying the following properties:*

1. $MB_{\mathbf{T}}(r,t) = 1$*, if there is a user interaction with r at t*
2. $MB_{\mathbf{T}}(r,t_1) \leq MB_{\mathbf{T}}(r,t_2)$*, if $t_1 > t_2$ (t_2 is closer to the creation time of r)*
3. $MB_{\mathbf{T}}(r,t) \rightarrow 0$*, if $t \rightarrow \infty$*
4. $MB_{\mathbf{T}}(r_1,t) \leq MB_{\mathbf{T}}(r_2,t)$*, if the last interaction of agent with r_1 is before the last interaction with r_2, or if the amount of interaction of r_2 is higher than that of r_1*

Next, we revisit from the literature the most promising candidates. To define a unified setting in which these functions can be used, let us denote $\delta_g(t)$ as the distance between the current time t and the time of the latest interaction with the resource in terms of the number of time unit g. Here g indicates the scale of the time unit, which can be hour, day, week or month depending on the nature of the information space. In our setting, as we work with the semantic desktop used in daily basis, we set g to a day granularity. In this work, we study the effectiveness of different time decay functions categories as follows.

Exponential decay function. Time-decay functions of this class have the form: $\lambda a^{-\beta\delta(t)}$, with λ, a, β are constants defined by specific configuration and domain. We leave the extensive analysis of different functions and parameters to future work and choose the most basic function in human memory and forgetting research, *Ebbinghaus forgetting curve*:

$$MB_{\mathbf{T}}(r,t) = e^{-\delta(t)/S} \tag{2.1}$$

where S is the decay rate, indicating how much the system should penalize old resources in decluttering.

Power law decay function. Similar to the exponential function but less aggressive in decluttering is a power law function: $a\delta t^{-b}$, with a, b are positive constants. In our study, we choose a equal to 1 to focus the analysis on the change in relative values of the memory buoyancy at different time points.

$$MB_{\mathbf{T}}(r,t) = \delta(t)^{-b} \tag{2.2}$$

The value of b depend on the type of interaction between the resource and the agent (for instance, a simple view access will leave the resource less recalled subsequently in the mental model than a thorough editing interaction). Since a learning model for posterior estimation of the parameter is not the focus of our work, we set up the parameter through empirical processes instead.

Weinbull distribution. Recent work in large-scale study of human memorysuggests that forgetting curve might well fit with the Weinbull distribution, with can be considered as the higher ordered variant of the exponential function:

$$MB_{\mathbf{T}}(r,t) = \mu e^{\sum_i - \frac{a\delta(t)^s}{s}} \tag{2.3}$$

where the parameter a measures the overall memory capacity of the system (how many data objects of the same type with d_i's that the system can store). Parameter s is one parameter of the Weinbull distribution and indicates the steepness of the forgetting function, i.e. how easily the system loses track of its member data objects. Parameter μ estimates the likelihood of initially storing the image of the resource in short- and long-term human memory.

Frequency-based computation. The decay functions above take into account the time of last interaction of the resource. In [3], Anderson et al. suggest that the frequency of interactions also play an important role in the memory buoyancy of a resource, as by the re-learning effect. Hence, for each of the above functions, we define a frequency-based extension as the aggregation of all decay function within the active window in the past:

$$MB_{\mathbf{F}}(r,t) = \frac{\sum_{i \in W} MB_{\mathbf{F}}(r,t)}{|W|} \tag{2.4}$$

where W is the set of interactions of resource r within the active window (in our setting, with day as time unit scale, we set the window to be one month past).

Each of the above time-decay models is used to calculate the memory buoyancy of items observed in the activity log, with two variants: Recency-based (the last access only), and frequency-based (averaging out all accesses in the past). After that, the propagation model is used (Deliverable D3.2, Section 2.2.1) to re-estimate the MB values, of both observed and unobserved items in the information space.

2.2 Computational Framework

2.2.1 Integration Workflow

The workflow of computation of the MB values has been described in Section 2.3.1 in D3.2. A background process, which is triggered by different strategies (e.g., periodically or on user's demand), computes the memory buoyancy using different models, and stores back the values in a repository. These values can be subsequently queries via RESTful services by other components through the middle-ware.

In this deliverable, we discuss how the workflow is used to evaluate the effectiveness of different conceptual models mentioned above. The challenge of such evaluation (and also the computation) flow is that input data of the framework is highly personal, very private data together with the detailed logs of user on daily basis. Exchange such kind of data over a network is undesirable, since it is both subject to potential intrusion from third-party and to privacy violation. To cope with this issue, we follow a pragmatic approach as follows. The components for background computation, together with the strategy configuration are bundled to client digest systems (PIMO-based desktop or TYPO3 web-based

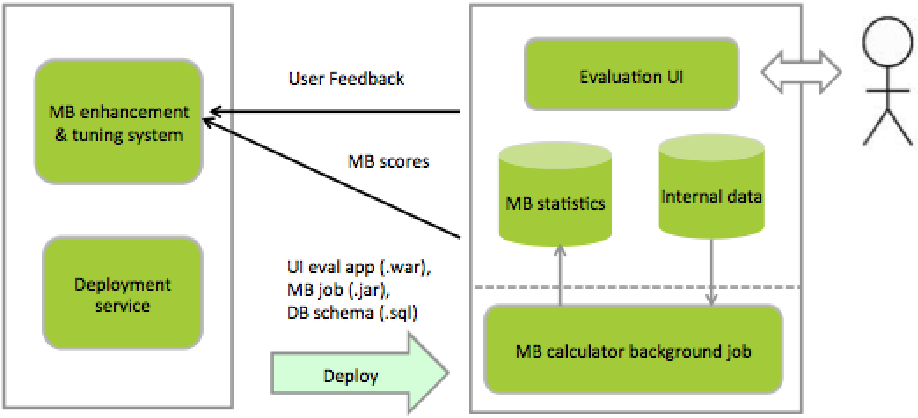

MB enhancement & tuning system
Deployment service
User Feedback
MB scores
UI eval app (.war), MB job (.jar), DB schema (.sql)
Deploy
Evaluation UI
MB statistics
Internal data
MB calculator background job

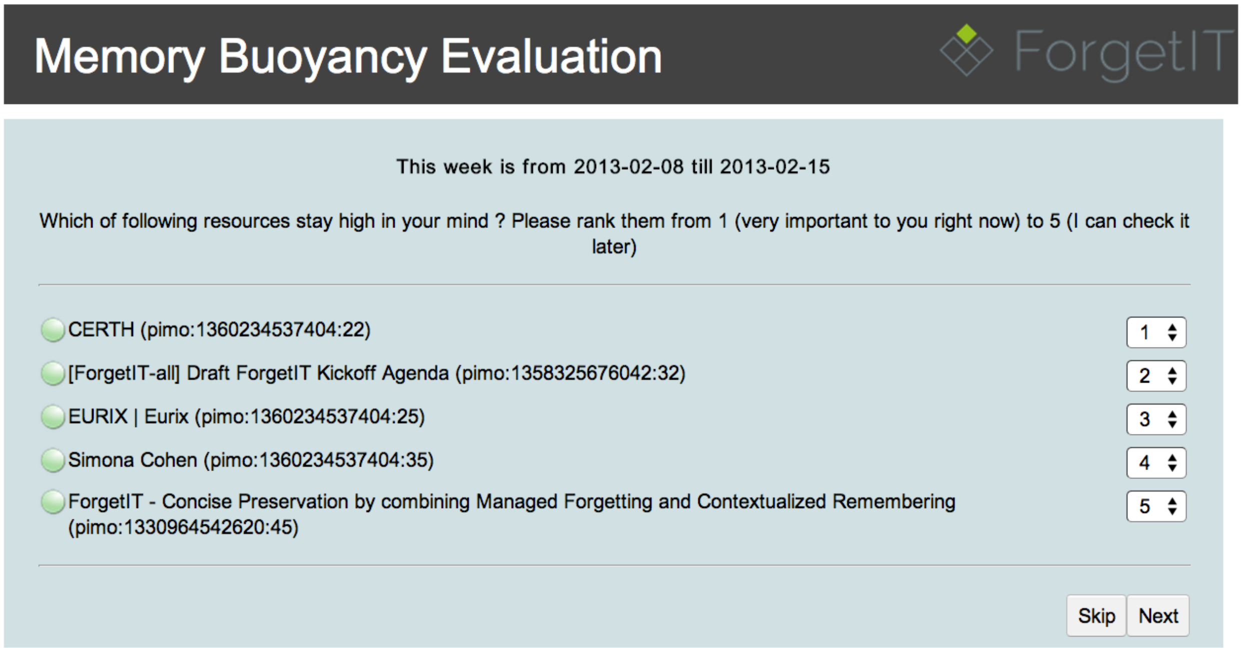
Figure 1: Workflow of Computation and Deployment of the MB Assessor Component

application). The components can then access to raw data locally and performs the necessary computation. In addition, a component of the evaluation user interface (EUI) is deployed in the client's environment as well. This component displays the information of user's resources, together with their estimated MB values, in a Web-based screen for users to evaluate. Finally, identifier of the resources (encoded), together with their computed MB scores and the feedback given by the users are sent back to the managed forgettor and are used in the quantitative evaluation, as well as in tuning and improvements of the models. The overall work flow for MB computation is illustrated in Figure 1.

2.2.2 Evaluation

In Section 6.1, we will discuss our approach of evaluation in the semantic desktop ranking case study. To accommodate such evaluation, we developed a lightweight interface that is to be integrated at the client side (PIMO / TYPO3). It queries the calculated MB values of items (stored in the managed forgetting repository), and matches with the meta-data and information of items (stored in the client), then displays the items in a Web-based form to ask for human feedback. This is illustrated in Figure 2. In the top-most label, it displays the time span within which the items are actively accessed (of course, for one item, there can be several active time spans) and based on which the MB values were calculated. Then it instructs the user with one question asking a user to rank the items based on how he/she would like to do with the documents based on their recalled context corresponding to the mentioned time span (for instance, the user was in tasks for a project meeting during then). A Likert scale style has been employed to suggest different actions to the user in 5 levels. They are:

1. **Pin - 1** The items is / was very important to the user within the mentioned time span. He would like to “pin” it as a shortcut in his spaces to easily access.
2. **Show - 2** The item was related to what user was doing, thus it should be displayed



Memory Buoyancy Evaluation

This week is from 2013-02-08 till 2013-02-15

Which of following resources stay high in your mind ? Please rank them from 1 (very important to you right now) to 5 (I can check it later)

- CERTH (pimo:1360234537404:22) — 1
- [ForgetIT-all] Draft ForgetIT Kickoff Agenda (pimo:1358325676042:32) — 2
- EURIX | Eurix (pimo:1360234537404:25) — 3
- Simona Cohen (pimo:1360234537404:35) — 4
- ForgetIT - Concise Preservation by combining Managed Forgetting and Contextualized Remembering (pimo:1330964542620:45) — 5

Figure 2: Screen shot of Evaluation of Documents' Memory Buoyancy.

in a relatively easy-to-access place (e.g. folders shown in front screen).

3. **Fade in - 3** The item was not related what user was doing within the time span, but it might be used in the time span immediately after, or might serve for other related purpose in a near future.

4. **Fade out - 4** The item was not related what user was doing within the time span, and it should be hidden to have spaces for other more important items shown up.

5. **Trash - 5** The item was not related what user was doing within the time span, and the user does not need it. It can be dispatched to the system to handle (e.g. archiving), with or without confirmation.

For each round of assessment, five items are shown to the user, and the user is asked to give a label based on one desired action mentioned above. If the user is uncertain about the most appropriate action, he or she can skip to the next round. The items are chosen as follows. For each round, one MB computation method is chosen randomly (e.g. only based on decay function, or decay function with frequency, or based on propagation, see Section 2.1.1 and Section 6.1.2 for the list of methods studied), proportional to how many evaluation feedback each method has. Then, a random time span of one week length is chosen, and the MB values of items are calculated based on the activity logs and meta-data available up to the last day of the time span. In this way, we can simulate the managed forgettor component as per different time periods when the data are continuously consumed and logs are produced. Based on these MB scores, 2 items are chosen at random in the top 10 highest scored item list, 2 from the 10 items with lowest scores, and 1 is chosen totally random to form the items shown to the user. We remark here that the information (labels of items, as in Figure 2, or activity logs, or the meta-data) are totally inside the client, only the MB scores (and the anonymized ids) of the items are sent back to the managed forgettor via the middleware.

Based on this setting, we compare the effectiveness of different components in our MB computation, which will be discussed in Section 6.

3 Preservation Value

Besides Memory Buoyancy, *Preservation value* is the second core information value considered in the ForgetIT project. It is crucial to preservation decisions, since it aims to estimate the expected long-term benefit of a resource. Figure 3 illustrates the difference between preservation value and memory buoyancy with respect to the time dimension. On one hand, preservation value is used for making a decision on whether the resource under consideration should be preserved for an expected long-term benefit. This is related to long-term information management considering time frames of decades. On the other hand, memory buoyancy is used to predict a short-term interest or importance. This refers to time frames between hours and days, or even short time frames depending on the application case.

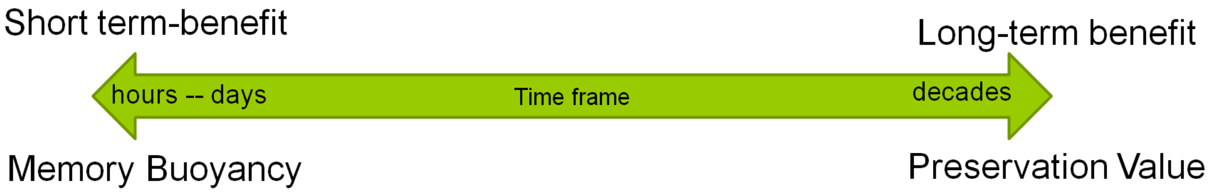


Figure 3: Preservation value vs. memory buoyancy with respect to the time dimension.

3.1 Conceptual Model

In this section, we present a working definition for preservation value and discuss a set of preservation value dimensions as a foundation for a systematic way of determining preservation value.

3.1.1 Definition and Purpose

In the ForgetIT project, we have agreed on the following working definition for preservation value, which clearly distinguishes it from other information values such as memory buoyancy:

Definition 2 ***Preservation value*** *is a value attached to a resource reflecting the benefit expected from the long-term survival of the resource.*

In particular, preservation value is used to determine the amount of investment to be made for ensuring the long-term survival and understandability of the resource under consideration. This can refer to preservation decisions such as how many copies to keep or to investment in semantic preservation, which also takes into account context evolution aspects. In the extreme case, the preservation value can also be used to make keep or delete decisions.

3.1.2 Dimensions for Value Assessment

The computation of preservation value, obviously, is a challenging task. It encompasses predicting the future value of a resource and is influenced by a variety of partially setting-specific factors. Therefore, it is not expected that there will be one single method, which can compute the preservation value for all possible settings, even if we just restrict to personal and organizational preservation. For example, other factors influence the decision, if I want to keep a photo or an email on the long run.

We still believe that there is enough common ground on what makes up a preservation value that a systematic approach can be defined on the conceptual level. For this purpose, we have defined a set of high level dimensions for drivers, which influence the preservation value. Within each of the dimensions different factors can be defined, which are relevant in the setting under consideration. From the consideration of the dimensions a better understanding of the preservation value itself as well as of the importance of those dimensions for the computation of preservation values is expected. They help in abstracting from the individual case.

The identification of those preservation value dimensions is based on intensive discussions with the application partners in the project and on the experiences gathered during the project. The following six dimensions have been identified:

Investment: In a wide sense, this dimension refers to the investment, which has been made into the resource and its improvement/change.

Gravity: This dimension refers to the relationship or closeness of a resource to important events, processes, and structures in the domain under consideration.

Social Graph: This dimension describes the relationship of the resource to the relevant social graph, i.e., the persons related to the resource, their roles and relationships. This might refer to the creators and editors of a resource as well as to persons related to the content of the resource.

Popularity: This dimension refers to the usage and perception of the resource.

Coverage: This dimension refers to factors, which considers the resource in relationship to other resources in the same collection. This includes factors such as diversity or coverage of sub-events, which are also used in making preservation decisions and, thus, influence preservation value.

Quality: This dimension refers to the quality of the resource.

The dimensions are also summarized in Table 1, together with example factors within each dimension for the case of personal preservation and organizational preservation, respectively.

For filling these dimensions with life we have investigated relevant factors within those dimensions for the case of determining preservation values for photos in a photo collection

Table 1: High-level dimensions for drivers of preservation value.

	Example – personal preservation	Example organizational preservation
Investment	Annotated photo	Highly revised or expensive press photo
Gravity	Related to important life situation	Importance for core business processes
Social Graph	Related to family members	Email from “the boss”
Popularity	Shared and liked photos	Frequently visited Web site
Coverage	Covering all sub-events of a trip	Representation of all departments/projects
Quality	Photo quality	Final version vs. draft

and for content in Social Networks. The results of those experiments are described in the sections 3.2 and 3.3, respectively.

It is noteworthy that this deliverable addresses preservation value in the personal preservation settings, whereas the preservation value study of organization will be presented in the next deliverable. The next section explains how the Semantic Desktop approach used in the personal preservation application scenario contributes to preservation value assessment along these dimensions.

3.1.3 Evidences for Preservation Value Assessment in the Semantic Desktop

This section explains the contributions of the Semantic Desktop approach for the assessment of the preservation value. It focusses on which evidences can be provided by the Semantic Desktop in addition to a solely investigation of a resource and its containment in a set of resources (such as a photo collection).

Considering the drivers for preservation value identified in the previous sections, we can identify several evidences for calculating the preservation value from the Semantic Desktop along the dimensions as explained in the following.

Investment: The Semantic Desktop provides various evidences for deriving an investment a user has spent on a resource. Activities on resources such as views are logged in the PIMO. This provides a log of user activities on a specific resources where a differentiation is available between such as access and modification (e.g., add/remove relation, modified text, added/removed literal,. . .). Whereas accesses shall be considered in the Popularity dimension, the modifications provide insights on activities of a user on a resource.

As an alternative to modification events, inspecting only the semantic representation of a resource – the so-called *thing* – also provides insight on investment: this can be done by counting the number of annotations (i.e., relations to other things), existence of a textual description (such as some notes about a picture), or the existence of a dedicated icon which differs from the class-specific icons (e.g., for a person, the image of a person could be set as an icon by the user). However, the latter does not reflect activates over time (i.e., adding ten annotations at once vs. one every week for 10 weeks), the constant change of the textual description or the removal of relations. Therefore, we consider harvesting the events for the preservation value is the most accurate choice.

A further contribution to investment can be derived from the semantics of the PIMO for the resources. For instance, there is a difference between, e.g., a resource of type *pimo:Document* and a resource of type *pimo:Contract*. Here, the preservation value of the contract resource can rise especially as the user made the effort of assigning this special type. However, most of such specific types would depend on dedicated applications such as a scanning application which allows to scan and import documents where the user is able to manually classify (or even supported by document analysis) with document types such as receipts, invoices, contracts, or certificates. The Semantic Desktop is capable to incorporate this.

Gravity: As the Gravity dimension refers to the relationship of resources, evidences for this dimension can be derived from the semantic graph of the PIMO. For instance, relations in the graph show which resources are connected to an event. Resources could be documents, photos, emails, notes; spread over different devices and/or available in different applications. Thus, the Semantic Desktop not only allows to express such relationships it also makes it possible to connect distributed resources.

Further, the semantic graph could also propagate importance of resources along specific relations. For instance, an event is connected to a photo collection where several photos have a large number of annotations or textual descriptions. Such importance indicators could be spread along certain properties indicating, e.g., containment (e.g., a part-of relation), contributing to the preservation value of the enclosing or closely connected resource. A similar spreading is already applied in the memory buoyancy calculation and could be adapted for the specific requirements of the preservation value.

Social Graph: The PIMO can contribute information about the social graph in various ways. First, relevant persons are expected to occur in the PIMO, e.g., as contacts, as attendees of events, as depicted and annotated in photo collections. Depending on the domain ontology extension included, a family tree could represent more detailed information about family relationships between persons in the PIMO as done in the PIMORE prototype (see deliverable D9.2 [17]) for close family members (parents, children, couples). Such information can also be collected from mentions in text which is analyzed by the PIMO either for annotation suggestions or by using semantic text composition with seed from WP4.

Furthermore, the PIMO contains information on who created a thing and who contributed. If it is private or shared in a group.

Popularity: As mentioned in the Investment dimension, evidences for popularity can be derived on resource accesses by the user logged by the PIMO. Extending this to a Group PIMO (e.g., a family in the Personal Preservation scenario or a organizational unit such as a project team in an organizational setting) – given that a resource is shared (which is another indicator for preservation value assessment) – other users' accesses can also be included, and considered differently, e.g., for sth. a user contributed to the group though not accessing it, but it is popular in the group.

Coverage: As done in the Gravity dimension, the semantic graph of the PIMO can be used to understand the relationships of collections of resources. First, to understand what a set of photos actually is which is handed over for preservation. The PIMO provides insights such as that it is, for instance, a photo collection of a holiday trip done five years ago. Further annotations could provide more information on locations, persons, topics, or additional material used for the trip.

Second, as the dedicated photo applications allow to add more details on a photo collection such as topics, this can contribute to decide for diversity or coverage, e.g., to give a higher preservation value to a photo annotated with a single topic (e.g., Cullen Skink (a Scottish soup)) vs. a topic used in many photos (e.g., like Edinburgh Castle), or by selecting photos for locations visited during a trip or by covering all workshop events of a project (such as photos taken by a ForgetIT project member at various ForgetIT workshops).

Quality: The Quality dimension is mainly fed by assessment of the resource by inspection, e.g., by image quality assessment. However, some indications can also be derived from the PIMO if such information is available, e.g., because a dedicated application contributed this. Such indicators could be a quality indication by the user by flagging the resource as favourite as done in the WP9 photo organization application (see deliverable [18]).

3.1.4 Leveraging Preservation Value

The most obvious way of using the preservation value is to make decision about preservation. A threshold can be set and only resources with a preservation value above this threshold will be considered for preservation (i.e., sent to the preservation system). Such a process can be automated or the resources above the threshold can be presented to the user for final decision.

As a time point for such preservation decisions, the time point, when the respective resource goes out of active use (decreased memory buoyancy for a specified time period) has been identified as a good candidate in the ForgetIT project. Before this point in time it might be expensive to keep track of the changes of the resource in the active system and to synchronize them with the resource copy in the preservation store. Waiting until long after this time point might put resources with high preservation value at risk. Other scheduling options would be a regular (e.g., monthly) computation of preservation values or a manually triggered preservation value computation (e.g. for a collection of photos or the resources of a project, which has been completed).

The preservation value cannot only be used for binary preservation decisions: preserve or not preserve[1]. On the one hand, the preservation value (or a category inferred from this value, see Figure 4 below) can also be provided to the preservation system as an indication of the effort to be invested for the respective resource. This can for example be used to select the adequate preservation aggregations as outlined in deliverable D7.1. On the other hand, the preservation value can also be used for deciding about other preservation options, e.g. just preserve a summary of a collection of images with low preservation value or store more context information for a resource with high preservation value.

Following this idea, as part of our approach a set of ”forgetting” options are defined in close collaboration with the activity on assessing human acceptance for managed forgetting. Clearly there will be no one-size-fits-all for managed forgetting, either. The challenge here is the definition of a flexible forgetting process that can implement adaptable forms of forgetting depending upon the needs and properties of the respective setting.

We are currently developing an adaptable framework for the managed forgetting process, which fixes the principle mechanisms of the process and can be customized along different dimensions: the parameters that are used for information assessment, the threshold used for memory buoyancy and preservation value for triggering forgetting actions and the options of forgetting considered. We are investigating the use of a policy framework that supports the definition of different forgetting policies. Policies have been shown to be an intuitive and powerful tool in the area of security management, e.g., for specification of access rights. In the preservation context, besides customizing the forgetting process, policies also can capture external constraints, such as legal preservation requirements or business requirements (e.g., to make sure that information pertinent to obsolete product versions is preserved).

The forgetting option are closely related and will build upon methods developed in WP4 such as methods for detecting redundancies as well as for condensation of textual and multimedia information objects.

3.2 Preservation Value for Personal Photos

Photo selection from personal photo collections is one of the scenarios that we considered to get insights for deriving a meaningful conceptual model for Preservation Value. Our aim is to better understand the human selection process for photo preservation and to investigate methods for supporting this process with no assumptions on prior user investment, such as, tagging, grouping, etc. For this purpose, we conducted a photo selection study with 35 users on real-world personal collections of some hundreds of photos each. In total, we obtained more than 8,000 photos for our experiment. Considering the insights that we got from the conducted survey, we designed and compared different approaches

[1] It has to be noted here again, that *not preserve* does not mean *delete* here. It just means the resource is not sent to the preservation system. It might still reside in the active system or any type of backup system.

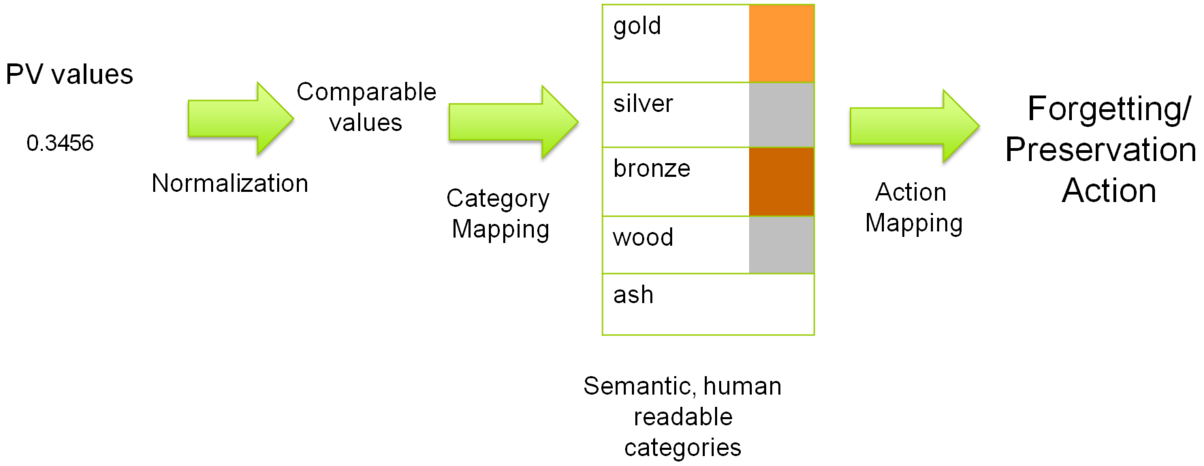


Figure 4: From PV to Managed Forgetting and Preservation Actions.

to make automatic selections from personal collections. We will explain in details both the user study and the automatic approaches in the rest of this section.

3.2.1 User Study

We performed a user study for a photo selection task. The goal of this user study is to gather insights on human photo selection for preservation and to collect ground-truth data for the automatic photo classification approach (see Section 3.2). In the user study, we asked participants to provide their personal photo collections and to select a subset of photos that they would want to preserve, i.e. to ensure that the selected photos stay accessible for a long period of time. For this purpose, we developed a photo selection application. The user study was complemented by a short survey about the task, which we asked the participants to fill after the photo selection task.

Participants. The experiment involved 35 users (30% females and 70% males) with 15 different nationalities: 24% of the participants came from Greece, 18% from Germany, 12% from Italy, 9% from China, 6% from Vietnam, and the rest from Ethiopia, Turkey, Kosovo, Iran, UK, Thailand, Sweden, Brazil, Albania, and Georgia. Regarding their ages, 61% of the participants are between 20 and 30 years, 24% between 30 and 40, 12% between 40 and 50, 3% between 50 and 60.

Task Definition. Since our task of selecting photos for preservation is not an everyday task for the users, it was important to find a good metaphor for supporting the task. After discussing a number of options, we decided to use the metaphor of a ”magic digital vault”, which incorporates the ideas of protection, durability and a sort of advanced technologies to keep things accessible in long-term. Therefore, before starting the photo selection, the task was explained with the following instructions:

Imagine you have an opportunity to protect some images from one of your

> *photo collections representing an event (a vacation, a business trip, a ceremony, etc.) by putting them into a magic digital vault. This vault protects the images against loss and ensures that they remain readable and accessible, and safely survive for the next decades (even in case of hardware crashes, new photo formats or software obsolete).*
>
> *Which photos would you put in the vault?*
>
> *When performing photo selection, please look through the entire collection before taking your final decision.*

Photo Collections. Previous works mostly consider either *public photo* collections (e.g., available on social media like Facebook and Flickr), e.g., [37], or photos from a *shared event* in which all the evaluators took part [45]. One difficulty we see with using public collections of photos from different people, even if they attended the same event, is that according to the different experiences of the individuals in the event they might also have a different level of appreciation for the same photo.

In contrast, we use personal photo collections. For instance, this can be photos from business trips, vacations, ceremonies, or other personal events the evaluator participated in. This means that each collection is not just a bunch of photos, which might exhibit different degrees of quality and aesthetics, but there are experiences, sub-events, and memories that might influence the selection behavior. We decided to focus on such personal collections because we wanted to observe the personal preservation photo selection decisions in a setting that is as realistic as possible.

In total, 39 collections were used in the experiment (four users evaluated two collections), resulting in 8,528 images. The size of collections ranges between 100 and 625 images, with an average size of 219 (with the standard deviation of 128.7). These collection sizes also emphasize the need for automated selection support, since manually browsing becomes time-consuming. We asked users for further information about their collections, such as, the main topic of the collection, whether they were previously pruned (e.g., by discarding low quality images), and when the photos were taken. Overall, 51% of the collections represent vacations, 30% business trips and conferences, and 18% other events (i.e., music festivals and graduation ceremonies). In addition, 23% of the collections were already pruned. The time when the collections were taken spans from 2007 to 2014 (64% in 2013-2014, 17% in 2012-2011, the rest in 2010-2007). Regarding user privacy, we made users aware that their images were temporally stored on remote servers (not publicly accessible) for feature extraction. The images will neither be inspected by humans nor given to third parties.

User Interface. We developed a desktop application that was used by the evaluators to import their collections and to select the photos they wanted to preserve. On the back-end, the photos were uploaded on our server, where they were temporally stored to extract features. Figure 5 shows the graphical user interface (GUI) of the application: the images contained in the imported collection are displayed in the bottom panel, while the ones selected by the users are shown in the top panel. Note that, faces appearing in

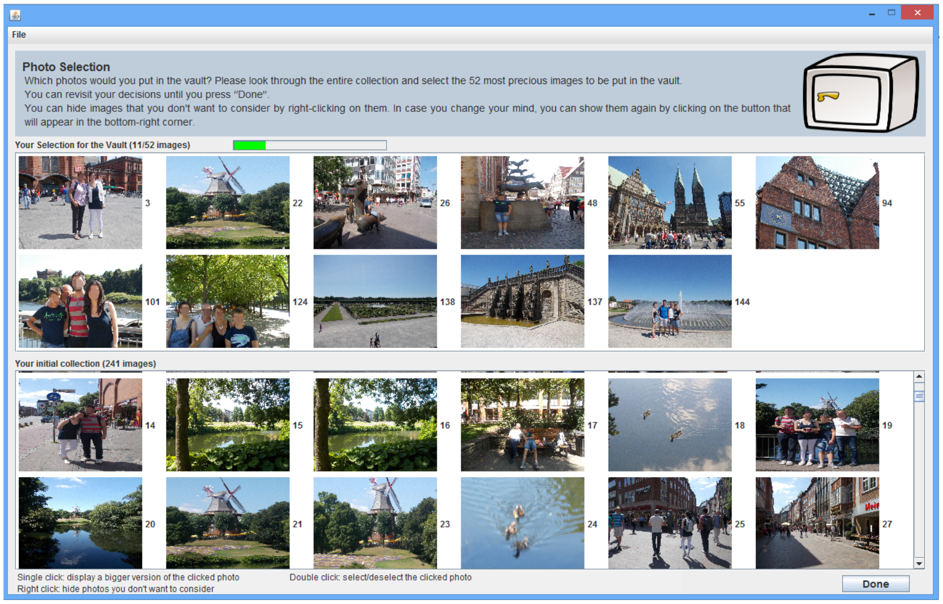


Photo Selection

Which photos would you put in the vault? Please look through the entire collection and select the 52 most precious images to be put in the vault.
You can revisit your decisions until you press "Done".
You can hide images that you don't want to consider by right-clicking on them. In case you change your mind, you can show them again by clicking on the button that will appear in the bottom-right corner.

Your Selection for the Vault (11/52 images)

Your initial collection (241 images)

Single click: display a bigger version of the clicked photo
Right click: hide photos you don't want to consider

Double click: select/deselect the clicked photo

Figure 5: GUI used by participants to select their photos to preserve.

Figure 5 have been blurred for sake of privacy. The photos are selected and deselected by double-clicking on them, and can be enlarged and inspected in detail by a single mouse click.

The images in the collection are shown in the same order in which they were taken, since this makes the browsing, remembering, and selection easier and more realistic for the users. We also made a preliminary evaluation where images were shuffled before being presented. This resulted in higher evaluation time and a higher mental effort for the selection process, because it made picking from a set of related images very difficult. However, keeping the original order might introduce bias in the selection towards the early photos in the collection, since users might loose attention during the evaluation or even complete the selection without going through the entire collection. In Figure 6 we present the average distribution of selected photos with respect to their positions in the collections. The bar diagram shows that the selected photos follow a quite uniform distribution over the entire collections, with the only exception of the first 10% that tends to contain more selections (13.5% as compared to 10%), when assuming equal distribution. This confirms that there was no major bias in the user evaluation caused by the order of the photos.

User Evaluation. Before starting the evaluation, the users were personally introduced to the photo selection task as well as to the application that they were supposed to use. Further remarks and clarifications about both the task and the usage of the application were given, where needed. However, no guidelines were given about the detail criteria to use for selection, in order not to influence the selection process. After this, the users autonomously interacted with our application. In more detail, the application asked them to select 20% of photos from the collection for preservation. This selection percentage (20%) has been empirically identified as a reasonable amount of representative photos.

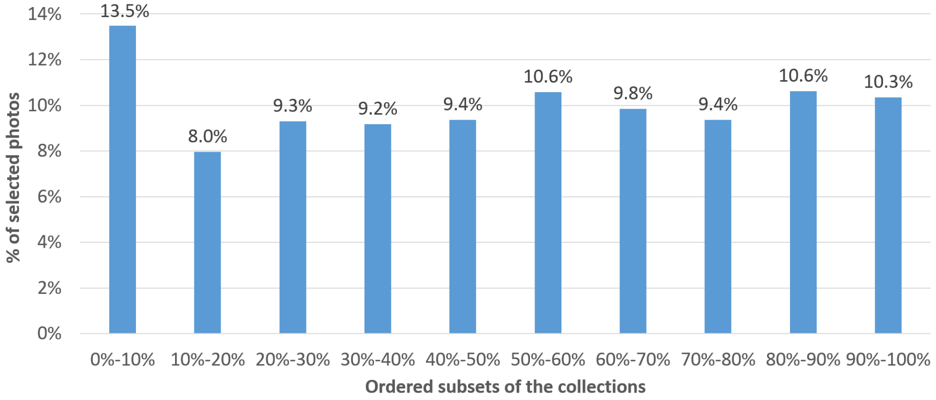


Figure 6: Average density of selected photos in adjacent and ordered subsets of the collection.

We also checked the selection of this number with the users in the survey.

User Survey and Results. The survey can be conceptually split into three groups. The first group includes questions about the collection used in the process: type of collection, year of photo taking and if the collection was cleaned. The results for those questions have already been reported above. The second group of questions refers to the scenario of photo selection process. The third group of questions look into the criteria that were taken into account during the selection.

Regarding the second group of questions, the users were asked to provide information about (1) which scenario they had in mind when selecting the images; (2) for whom they are preserving the images; (3) whether they would be ready to pay, and for how many years, if preservation was a paid service. The answers to each question were posed as multiple choices and are reported in Figure 7. Questions (1) and (2) reveal that the process of long-term preservation is centered around the owner of the photos: more than 70% of the evaluators said that they thought about own future reminiscence when they selected the photos, and almost 80% indicated themselves as a main recipient of the preservation. Looking at the preservation as a valuable service to be paid (question (3)), the evaluators were mostly split into two groups: either being ready to pay for many decades (39%) or needing flexibility to make new preservation decisions every 2-5 years (36%). In both cases, these answers highlight a clear need for preservation of personal photo collections.

In the third group of questions, we suggested different criteria and asked the users to rate how much each criterion was considered during the selection. The suggested criteria, based on the insights on *keep* and *delete* decisions in [46], link to the preservation value dimensions defined in Section 3.1.2 as follows. *Quality* is represented by “image quality”, *Gravity* by “memories evocation” and “important to me”, *Social Graph* by “somebody important”, *Coverage* by “event coverage”, and *Popularity* by “sharing on social media”.

The ratings were provided via star ratings on a scale between 1 and 5 (5 stars mean

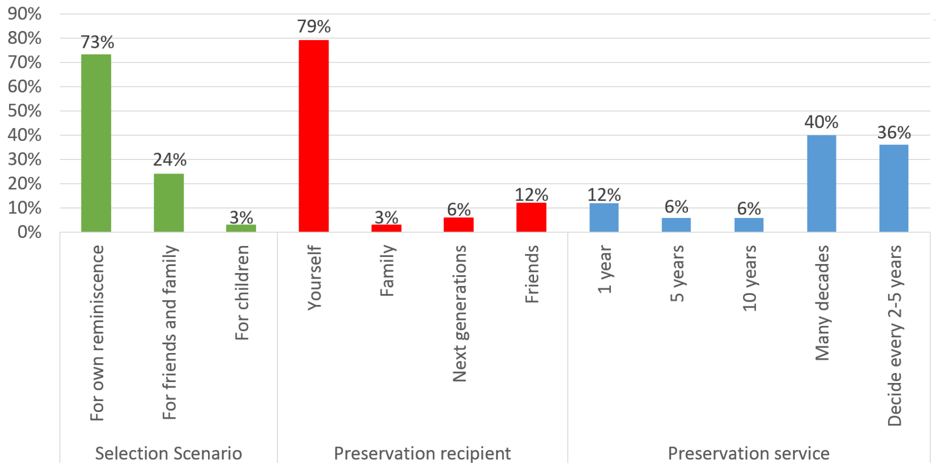


Figure 7: Survey results with respect to preservation scenario, preservation target group, and preservation as a service.

very important, 1 means not important at all). The criteria along with statistics about their ratings are reported as box plots in Figure 8. Note, that medians are represented as horizontal bold bars, while sample mean is indicated with a bold cross. An important finding of this evaluation is that the objective quality of photos is rated as the second least important selection criterion, after the sharing intent. This shows that quality and aesthetics, although being important and used for "general purpose" photo selection [25], is not considered very important in the case of selecting images for preservation. In contrast, criteria more related to reminiscence, such as event coverage, typical image, and "the picture evokes (positive) memories" are all rated high, with highest ratings for memory evocation. The remaining two criteria "picture is important to me" and picture "shows somebody important" refer to the personal relationship to the picture and are also both rated high.

These results anticipate that the task of predicting images to be selected for long-term preservation is likely to be difficult, since many of the criteria that are rated high, e.g. memory evocation, personal importance and "typical image", are difficult to assess for a machine, because they contain a high level of subjectivity. Another complicating fact is that there is no single dominating criteria, but a mix of highly rated results.

In the criteria ratings we can see clear differences to the ratings of the partially overlapping set of criteria reported in [45], where photos on a shared events were used and the selection was not directly related to preservation and reminiscence. Much higher ratings are given to criteria such as quality, whereas event coverage and important person are rated relatively low (although with high variance). Interestingly, photos that capture a memory are also rated high in this case.

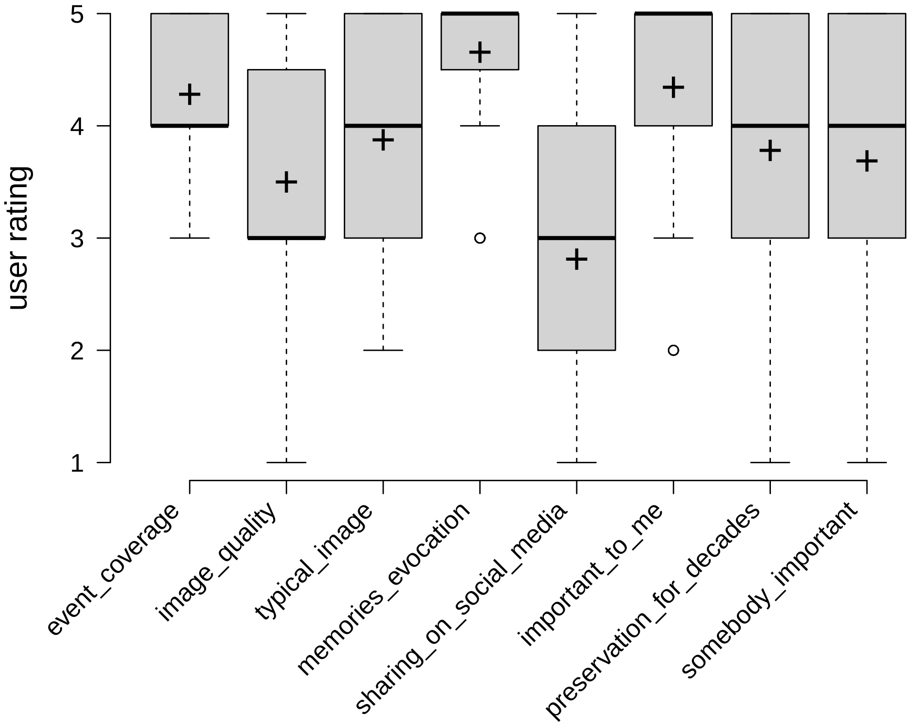


Figure 8: Boxplots of the different selection criteria.

3.2.2 Automatic Approaches to Photo Selection for Preservation

Due to the often large sizes of digital photo collections, automated photo selection, in more general, has already been studied in various other contexts, such as, photo summarization [26, 41, 40], the creation of collages [47], selection of representative photos [45, 11, 21], and the creation of photo books from social media content [37]. In the context of preservation of public images (available in the online photo sharing platform, e.g., Flickr), the previous work [12] conducted a qualitative study in order to assess the values of images for representing social history. To the best of our knowledge, this is the first work investigating automated photo selection in support of the long-term perspective for digital preservation and personal reminiscence.

Our approach is inspired by two insights got from how photo selection for preservation was performed in the user study (Section 3.2.1). Firstly, in post study discussions the participants to our user study reported a reduction strategy, taking out photos that are surely not considered for inclusion to ease the further selection. Secondly, our participants reported a coverage strategy when selecting photos, trying to cover all important sub-events or aspects of the collection under consideration. The importance of coverage is also implied by the results of the survey, which was part of our user study, and by other research work e.g., on social image search [40].

Therefore, we investigate and compare two models for the photo selection process for gaining deeper insights in it:

- *Reduction-oriented* photo selection considers photo collections as sets of two types of items, unique photos and groups of near duplicate photos. Photo selection is then modeled as a two phase selection process, first selecting valuable items from the collection and subsequently picking photos from the selected duplicate groups.
- *Coverage-oriented* photo selection considers photo collections as a sequence of depicted sub-events. Photo selection aims to reach coverage of the sub-events in addition to picking the most important photos.

For investigating the validity of those process models, we have implemented both of them resorting to a machine learning approach for automatic classification of valuable photos, based on a variety of features. In addition, we used a semantic event clustering method (for the coverage oriented process) and an advanced near–duplicate detection method (for the reduction-oriented method). These image processing methods have been developed in the context of WP4.

Other than [45, 40, 37], we do not rely on any additional user investment such as eye tracking information [45, 37] or photo annotation with text [41, 37, 40], because we believe it is exactly the reluctance of further investment that lets large photo collections unattended on our hard disks. In contrast to the photo selection works in [41, 26, 40], we evaluate our approach based on selection decisions taken by users on their own photo collections, i.e. we consider and aim at reflecting human expectations.

Overview

The problem that we tackle is emulating human behaviors when they select a subset of photos from a personal collection for long-term preservation. This means starting from a collection of personal photos $C = \{p_1, ..., p_n\}$, the system is supposed to suggest a subset $\hat{C}_p = \{p_{i_1}, ...p_{i_r}\}$, which is close to the set of photos that the user would have selected $C_p = \{p_{j_1}, ...p_{j_r}\}$ in the sense that there is a high overlap of selected photos. As anticipated, the task is to select the most valuable photos for preservation, i.e., $|C_p| = \alpha \cdot |C|$, where we set $\alpha = 0.2$ in our work.

We investigate two models, a *reduction–oriented* approach and a *coverage–oriented* approach for gaining deeper insights into the photo selection process for preservation. Figure 9 gives a high–level overview of our system framework. First, photos are processed to extract a set of features that are used as input to both the approaches to learn two models to predict the importance of each photo, i.e. the probability of a photo to be selected. The difference between the models learned in each method will be explained later in this section. The *reduction–oriented* method is inspired by the observation that users tend to first discard photos, especially redundant ones, to ease the selection process. It maps the collection C to a set of items $\{i_1, ..., i_n\}$, where each item i is a set of photos that can contain either one unique photo or different photos with near–duplicates.

In case of items i that are near–duplicate sets, photos are selected from those sets by using a second classifier to distinguish between important and not important near–

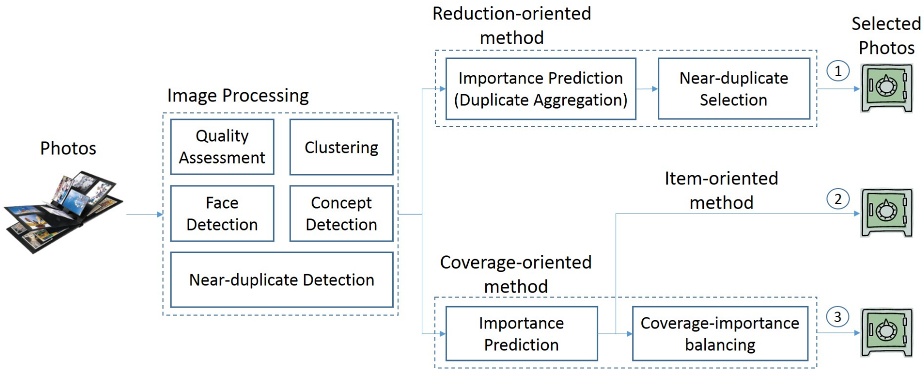


Figure 9: Overview of our approaches to personal photo selection.

duplicates. We split the selection process into two different steps under the assumption that users might consider different criteria in each one of them. When selecting a subset of photo from an entire collection, they might take joint decisions considering the overall collection, while the decision about which image to select among a set of near–duplicates might be taken based on individual characteristics of the images. The *coverage-oriented* approach is motivated by observations in the user study and by the results of our survey, which showed high ratings of event coverage as perceived selection criteria. Moreover, image clustering is widely used in other photo selection tasks (see, e.g., [37]). The coverage–oriented approach is thus a mapping from a collection C to a set of non-overlapping clusters c, where each cluster contains all the photos belonging to the same sub-event of the collection. Based on those clusters and on the predicted selection probabilities, photos are selected by balancing the coverage of the different clusters and the selection probabilities of the images.

Image Processing and Importance Prediction

The results of the survey suggest that the selection of personal photos for preservation is driven by a variety of factors, many of which are difficult to assess automatically. Therefore, in addition to assessing image quality, it is important to use advanced methods to extract semantics from photos. This is especially important, since we are considering scenarios where no user labeling or tagging is assumed, i.e. raw image content and metadata like time are the only available information that can be exploited. In general, rich feature sets will be used as a starting point for investigating which features are most important in this complex photo selection task. We also use methods for near–duplicate detection and image clustering, which are described in the context of the reduction-oriented and the coverage-oriented photo selection, respectively.

Image Processing. The images processing techniques that we employ are concept detection, quality assessment, face detection, clustering, and near–duplicate detection. The latter two techniques will be described in the following sections, as starting points of our *reduction–based* and *coverage–based* approaches.

Concept Detection consists in analyzing the visual content of an image and automatically assign concept labels to it. This moves the description of an image to a semantic level, where it is possible to identify abstract concept like *joy*, *cheering*, *entertainment*, as well as more concrete ones like *crowd*, *girl*, *stadium*. We used an extension of the 2–layer stacking architecture that was proposed in [29]. Our extension involves the use of a first–layer classifiers that exploits VLAD encoding [5] for the local features (e.g. SIFT) rather than Bag–of–Words encoding. Using this method, 346 concept detectors were trained for the 346 concepts defined as part of the TRECVID 2013 benchmarking activity [35]. As training corpus, the TRECVID 2013 dataset comprising 800 hours of video was used.

Regarding Image Quality Assessment, we computed four image quality measures, namely blur, contrast, darkness, and noise, along with their aggregated value (Minkowski sum), following the procedure presented in [30].

For Face Detection we have applied the Viola and Jones [44] approach using several detection classifiers. Each detected region is accepted as facial one by taking into account the number of classifiers that detected it, its color histogram (skin like color) and whether other facial features (eyes, mouth, nose) have been detected in it.

Features for Prediction. Five groups of features have been designed for being used in the photo selection task, based on the image processing results presented in the previous section as well as on more global characteristics. They will be used as input of the learning model that is used to predict the probability that an image is selected for preservation. In the following sections we will refer to the class of features using the names introduced hereafter, although the link between them and the preservation value dimensions defined in Section 3.1.2 is made explicit in their descriptions.

Quality-based features. They consists in the 5 quality measures described before: blur, contrast, darkness, noise, and their fused value. The assumption behind using this information is that users might tend to select good quality images, although their impact seems to be less important in subjective selections of humans [45]. This family of features corresponds to the *Quality* dimension defined in Section 3.1.2.

Face-based features. The presence and position of faces might be an indicator of importance and might influence the selection. We capture this by considering, for each image, the number of faces within it as well as their positions and relative sizes. In more details, we divided each image in nine quadrants, and computed the number of faces and their size in each quadrant. This results in 19 features: two for number and size of faces in each quadrant, plus an aggregated one representing the total number of faces in the image. These features are related to the *Social Graph* dimension defined in Section 3.1.2, although a more precise notion of who is in the photo, e.g. obtained via face clustering and tagging, would help in determining relations among people and the owner herself.

Concept-based features. The semantic content of images, which we model in terms of concepts appearing in them, is expected to be a better indicator than low-level image features, because it is closer to what a picture encapsulates. We associate to each image a vector of 346 elements, one for each concept, where the i-th value represents the probability for the i-concept to appear in the image. Although this class of features does not have a clear correspondence to any of the dimensions defined in Section 3.1.2, we believe that the semantics introduced by them can bring a great benefit to the predictive capabilities of the model.

Collection-based features. When users have to identify a subset of images to preserve, instead of just making decisions for each image separately, the characteristics of the collection an image belongs to might influence the overall selection of the subset. For the same reasons, but moving to a finer granularity, it might be worth considering information about the cluster an image belongs to, as well as whether there are near duplicates for an image. For each image, we consider the following collection-base features to describe the collection, cluster, near duplicate set the image belongs to: size of the collection, number of clusters in the collection, number of the clusters in the collection, number of near–duplicate sets, size of the near–duplicate sets (avg, std, min, max), quality of the collection (avg, std), faces in the collection (avg, std, max, min), size of the cluster (avg, std, max, min), whether the image is the centroid of the cluster, quality of the cluster (avg, std, max, min), faces in the cluster (avg), how many near–duplicates the image has. This family of features is a representative of the *Coverage* dimension defined in Section 3.1.2.

A general consideration about the comparison between the features described above and the user study presented in Section 3.2.1 has to be made. The concepts that resulted to be important from the user study, e.g. evocation of positive memories, typical images, overall and personal importance of images, are highly subjective and not directly recognizable by a machine. We started modeling event coverage through clustering and the coverage-based methods and introducing semantic features like concepts, but we plan to model the other important features as well. We also used image quality, although not perceived as important, for sake of comparison with the other features.

Importance Prediction. Once images in our collections have been described in terms of the features presented above, a prediction model represented by a Support Vector Machine (SVM) [13] is learned in both the *reduction–oriented* and *coverage–oriented* methods to predict the selection probabilities of new unseen images. Given a training set made of photos p_i, their corresponding feature vectors $\boldsymbol{f}_p$, and their selection labels l_p (i.e. *selected* or *not selected*), an SVM is trained and the learned model M can be used to predict the probability $P = M\left(\boldsymbol{f}_q\right)$ for a new unseen image q to be selected by the user. We use the selection information obtained in the user study as a training set for this classification task. The difference between the prediction models learned within the two methods consists in whether decisions made for near–duplicates are made coherent or not before the learning. As a matter of fact, different selection decisions for near–duplicate images would result in having patterns with similar features but different labels within the training set. Since in the *reduction–oriented* method handles near–duplicates in a dedicated way, which will be described in the corresponding section, we make the prediction model unaware of near–duplicates by making labels (i.e. human decisions) coherent within each near–duplicate set in the following way: (i) all the images within a set are marked as selected if at least one image in this set has been marked as selected; (ii) all the images within a set are marked as not selected if all the images in that set have not been selected. On the contrary, the prediction model used in the *coverage–oriented* approach is trained without modifying near–duplicate labels because the approach does not handle them in a dedicated way.

Although the task of photo selection is potentially subject to the preferences of each user [39, 48], we learn a common model for all the users as a first step in the investigation of this new complex photo selection task. Moreover, this can be the starting point for personalization, i.e. adapting the initial general model to better meet the preferences of single users.

Reduction–oriented Photo Selection

Near–duplicates are in the core of the reduction–oriented selection process. Therefore, we take a closer look into the occurrence of near–duplicates in our data set, before discussing our approach and the way we handle them in more detail.

Near-Duplicates. Within our dataset, we observed that more than a half of the photos (53.4%) are near–duplicates (or just duplicates for short in what follows), which suggests that adequately dealing with this redundancy plays an important role in the photo selection process. Moreover, when also considering the photo selections made in the user experiment, we noted that most of the duplicate sets (i.e. sets containing photos that are near–duplicates of each other) contain both selected images (usually one) and not selected ones. As we already discussed, this aspect is taken into account when uniforming the labels of near–duplicates before training the prediction model.

Near-Duplicate Detection. Near–duplicate detection identifies photos that are quite similar. In many cases, the photographers shoot a scene multiple times which results in creating near–duplicate images that are important to be detected in a preservation application. We used the method described in [4] to identify near duplicates. Initially, it forms a vocabulary by applying k–means on image SIFT features and then encodes the image using Locally Aggregated Descriptors (VLAD) encoding. The images with very similar VLAD vectors are efficiently retrieved using a KD forest index and Geometric Coding [49] is further used on these candidate near–duplicates in order to finally accept or reject the hypothesis that they are near–duplicates.

In contrast to more strict near–duplicate detection methods, this approach results in a more relaxed notion of near–duplicates, which we believe is a better fit to the targeted rather selective photo selection task. Figure 10 illustrates the different types of duplicates the approach detects. The images in the upper row represent a more traditional near–duplicate situation, where the images can be hardly recognized as different. The images in the bottom row represent what we call *semantic duplicates*: they are different in terms of colors and background, but the concepts in the images are the same as in both images there is a person singing and playing the guitar on a stage.

Relaxed Near-Duplicate Aggregation. Our assumption is that people might take decisions based on different criteria when (1) selecting a subset of photos from a collection for

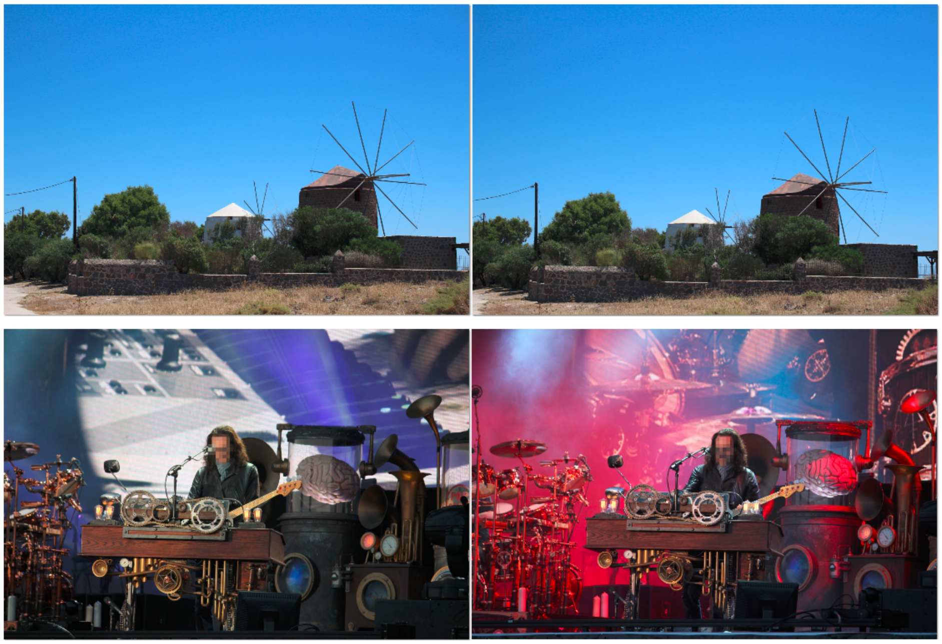


Figure 10: Two different examples of near–duplicates.

preservation, and (2) choosing which photo to select from a set of near–duplicates. Therefore, we split the reduction–oriented approach into two steps: *relaxed near–duplicate aggregation* and *near–duplicate selection*. This also splits the entire problem of photo selection into two distinct problems: (1) select a subset of images from a collection, without considering near–duplicates (i.e. applying the aggregation described below); (2) for each selected image that had originally near–duplicates, pick the one that was actually selected by the user.

For *relaxed near–duplicate aggregation*, we logically consider near–duplicates as a single image, and our system makes a single decision for them. This is in line with other works on photo album summarization [11], where usually the near–duplicates are identified and only the representative (based e.g. on centroid / similarity information) is selected. For selecting the image that represents the duplicate set, we rely on our importance prediction method: we select the image with the highest importance prediction within the duplicate set as the representative of it. After this step, all the images in this revised collection are considered as singles, either because they actually are or because they are a representative of a duplicate set.

Subsequently, the images are ranked based on the predictions, and the top 20% is finally selected (20% because this is the original amount of photos that was selected by the users). For assessing this selection step, we make human decisions coherent within each near–duplicate set in the same way as it was done before the training of the model (see Section 3.2.2): all the images within a set are marked as selected if at least one image in the set has been marked as selected; all the images within a set are marked as not selected if all the images in the set have not been selected.

Near-Duplicate Selection. For the duplicate sets selected in the previous step, we still have to decide which of the photos in the set to select. Since we consider this as a distinct type of selection problem, we decided to learn a second classifier to predict the importance of each duplicate image in a set and to pick the one with highest importance as the best candidate. Similarly to the first prediction model, an SVM has been used as prediction model. The training set used in this case is restricted only to duplicate images.

As the discussion of the results will show, this kind of decision is actually driven by another set of features than the original image importance prediction.

Coverage-oriented Photo Selection

The approach for coverage–oriented photo selection also consists of two steps. It first uses a clustering approach in order to cluster the photos that belong to the same sub-event in a photo collection. In a second step, photos are selected from each cluster by balancing predicted photo importance with coverage of the different clusters. Although kept into account with the *Collection-based* features in Section 3.2.2, the *Coverage* dimension (Section 3.1.2) is dominant and explicitly considered in this photo selection method.

Clustering Approach. We identify sub-events within a collection by clustering images based on time and the extracted concepts. Based on the outcomes of [36], where image clustering has been tested using a variety of clustering methods and several image representations, we adopted in our experiments the K–means - model vectors (concept detection confidence score vectors) combination that achieved the best performance. Furthermore, the clustering method in [36] has been slightly modified in order to take into account the near–duplicate results and make sure that all images of a near duplicate group will be assigned to the same cluster. In contrast to the common strategy of selecting centroids from each clusters as candidates for the selection, we don't take such hard decisions but we take information about centroids into account as feature in the importance prediction model.

For capturing sub-events we merge the results of a more semantic clustering based on concept features with the results of a temporal clustering of the images, giving priority to the temporal clustering. Figure 11 shows an example of a cluster identified by our approach.

Photo Selection from Clusters. As mentioned above, the photo selection step has to balance predicted photo importance and sub–event coverage with each other. In our approach we experiment two methods to achieve this. The first one is based on the idea of "round-robin", i.e. visiting each cluster in a round–robin fashion and picking the photo with the highest predicted importance from it. While also considering predicted photo importance, because the photo having the highest predicted importance is always selected from each cluster, this method gives priority to coverage. As an example, in case the collection contains 10 clusters, the method picks the most important photo from each cluster, getting in total 10 photos. Subsequently, it repeats this process with the second most important image from each cluster, continuing until the targeted 20% of the

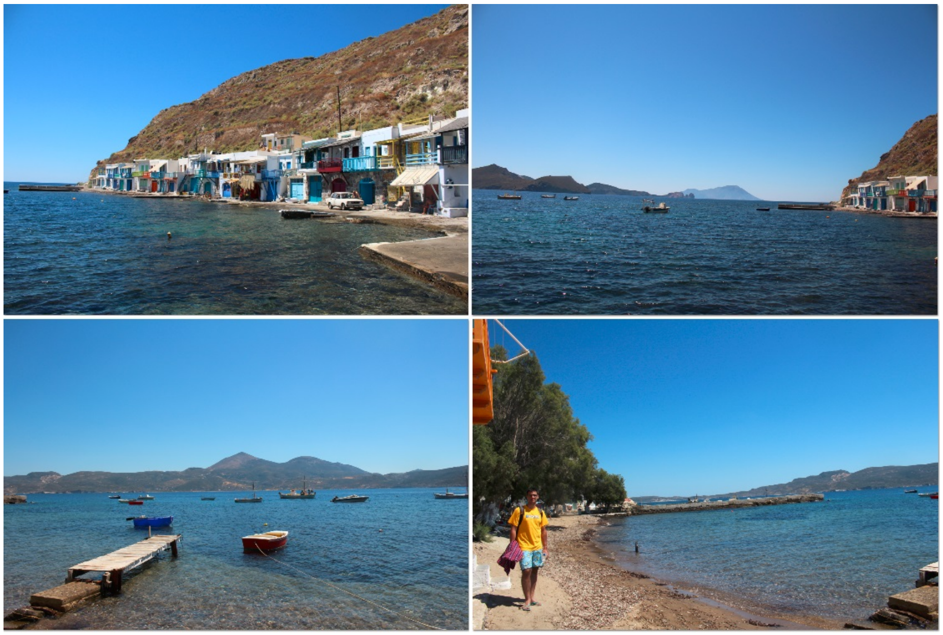


Figure 11: Example of a cluster containing four images.

collection is reached.

The second method, which we call "mixed method" is similar to the previous one, but it takes more photos from bigger clusters. Each cluster is given a number of "placeholders" in the subsets to be filled, based on its relative size in the collection. For instance, let us assume that a collection is split into two cluster: cluster A has a relative size of 0.8 and cluster B of 0.2. Assuming that 10 images have to be selected, then 8 will be taken from cluster A and 2 from cluster B. Selections within each cluster are always done according to the predicted importance of the photos (i.e. most important photos first).

Experiments And Results

In this section, we evaluate our proposed approaches to photo selection, and present the analysis results achieved when using the previously described approach to photo selection. Besides providing the numeric performances, we extensively discuss and compare the dominant criteria behind each method, and we map such analysis to the pattern exhibited by humans when making selections.

Experimental Setup. Since the overall goal of our work consists in emulating the human behaviors in selecting the subsets of photos from a personal collection, we compare the automatic selections generated by our methods with the ones done by the users. This is done by measuring the precision of the selections performed by the different methods, i.e., the fraction of the number of suggested photos that were selected by the users with respect to the total number of photos in the selection. Since the users were asked to select the 20% of their collection, we let the methods select the 20% of each collection as

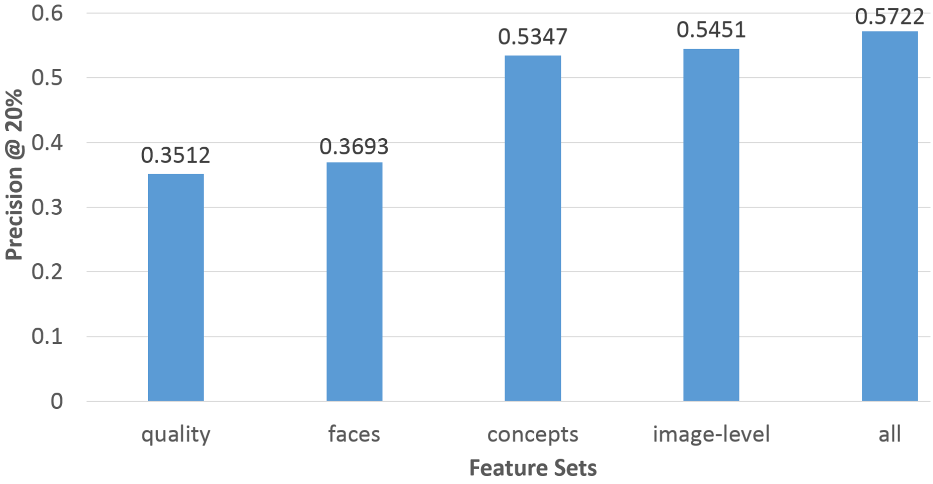


Figure 12: Average precision in the selected 20% of photos, for different feature sets, when aggregating near–duplicates.

well. The results presented in the rest of this section are averaged over the 39 collections that we considered in our work. Note that our baseline method that randomly selects the 20% of each collection would have an average precision equal to 0.2.

The prediction models employed in our methods were built by using the Support Vector Machine implementation of LibSVM[2]. In particular, we trained an SVM model with Gaussian Kernel via 10–fold cross validation over the entire set of photos. The open parameters were tuned via grid search to $C = 1.0$, $\gamma = 1.0$.

Reduction–oriented Method. The first part of the evaluation focuses on the role of near–duplicates in personal collections. We analyze how their presence affects both the performances and decision criteria of the selection. Figure 17 presents the performance of the first step of the reduction–oriented photo selection. In this case, we aggregate near–duplicates unifying selection decisions among them. Recall that, during the aggregation, any non-selected photo having a near–duplicate that has been selected is marked as selected as well stressing the similarity of near–duplicates. This modified collection is used for training and for testing for the first set of results. Different prediction models have been trained by using the subsets of the features described before, so that the impact of different groups of features can be analyzed. *Quality* and *faces* features are the ones that perform worst, individually. For the quality features, this is expected from the results of the survey and has also already been observed for other photo selection tasks [45]. In addition, *faces* features alone also do not seem a very good indicator.

The performances achieved when only using *concepts* features are clearly better than the ones of *quality* and *faces* features: they are able to capture the semantic content of the photos, going beyond their superficial aesthetic and quality. Examples of concepts with a high importance in the model are *person*, *joy*, *cheering*, *entertainment*, and *crowd*. The

[2] http://www.csie.ntu.edu.tw/~cjlin/libsvm/

Rank	Selection of subsets	Selection of near-duplicates
1	Near-duplicates of image	Entertainment concept in image
2	Avg blur in cluster	Contrast in image
3	Similar images	Face frequency in near–duplicate set
4	Face frequency in cluster	Faces in image
5	Minimum darkness in cluster	Noise in image
6	Face frequency in collection	Joy concept in image
7	Number of clusters	Min contrast in near–duplicate set
8	Person concept in image	Near–duplicates of image
9	Faces in image	Faces in the center of image
10	Entertainment concept in image	Cheering concept in image

Table 2: Most important features for the two different models: importance prediction and near–duplicate selection.

model trained with the combination of all aforementioned features, denoted *image-level* because the feature values are extracted from image level, improves (precision of 0.545) the performance of using concept features alone. This indicates that leveraging quality and faces features, and semantic measures, such as, concepts, can better the overall performance.

If we include global features for each photo representing information about the collection, the cluster, and the near–duplicate set the image belongs to, we get a comprehensive set of features, which we call *all*. The precision of the selection for this model further increases up to 0.57: this reveals that the decision is also driven by considering general characteristics of the collection the image belongs to: e.g. number of images, clusters, average quality of images in the collection and in the same cluster, how many duplicates for the image there are. Although we use a slightly simplified setting by unifying the duplicates, this are rather promising results given set we work on the pure image information.

Near–duplicate Selection The second and even more challenging problem is making selection decisions among the images belonging to the same near–duplicate set. That is, for each set of near-duplicated images, we want to predict which are the ones that the user would select, which even for human users are sometimes difficult. As we already discussed during the description of our approaches, decisions within each near-duplicate set are taken via a second classifier. To get deeper insights, we compare the two problems in terms of the differences in the features that are most important in the respective prediction models. Table 2 contains the list of top ranked features for the two models, where the rank is determined by the correlation of the feature with respect to the class (user selections), computed via χ^2 test.

The left-hand column contains the top ranked features for the model used to predict the selection probability of images after near–duplicate aggregation (see above). It is possible to observe that most of them represent global information about size, face frequency, quality of the collection, cluster, and near–duplicate set the image belong to. Interestingly, the first ranked represents the number of near–duplicates for a given image: this could represent the behavior of people that take many and similar pictures to what they perceive as important. In the set there are also a few features at the level of the single image, representing concepts (person, entertainment) and faces appearing in it. Note that there

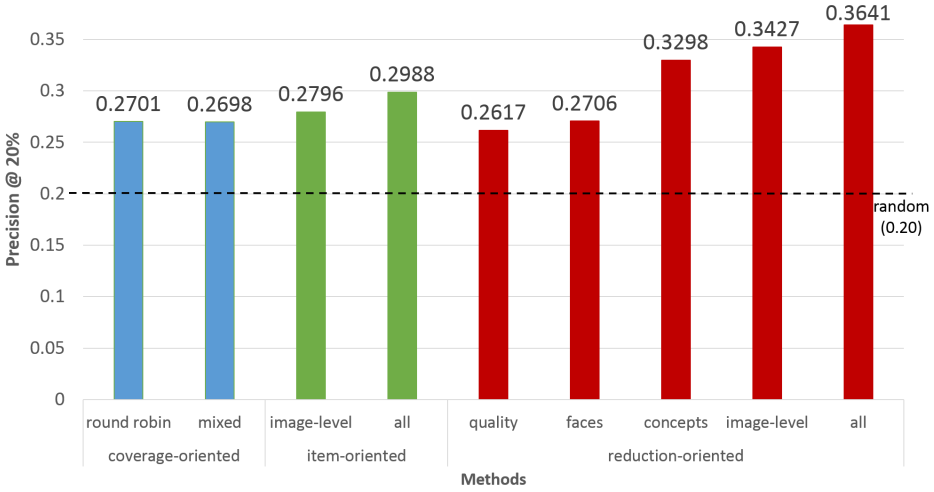


Figure 13: Average precision of the selections done by different methods, making selections within near–duplicate sets.

are no features representing the quality of single images: the quality–related features in the list are all global. Considering the feature list on the right-hand side, representing the most important features for selecting among near–duplicate images, we can note a different composition. Most of the features only refer to single images, e.g. to the concepts and faces appearing in them. In particular, the quality aspect is more important in this case. There are also a couple of features about the near–duplicate set of the image, which makes sense because when deciding whether to select an image or not the user might take its near–duplicates into account.

To summarize, this comparison shows that users exhibit two different patterns when selecting photos: (1) they consider all the images of a collection when selecting a subset of it without pay much attention to their quality, and (2) in case of duplicated images they make decisions based on the content and quality of the single images. Mapping the results to the user surveys. They said that they pay attention to the coverage aspect, the semantic ("the photo contains something important"), and the presence of people when selecting the photos. This is confirmed by the analysis of the first selection. However, when they have to make decisions between similar photos (second selection) they consider other criteria.

Overall performance Figure 13 shows the performance of the reduction-oriented methods, when we combine the two previously described steps, prediction with near–duplicate aggregation and near–duplicate selection. Actually, Figure 13 shows the average precision of the selection done by all the differen methods considered. The *reduction–oriented* method is the best performing one in the comparison, confirming the intuition that the process of selecting a subset of photos from a collection and the one of identifying which near–duplicate to select are driven by different criteria and should be addressed as separated challenges. In line with the results reported in Figure 17, *concepts* and *all* features lead to improvements with respect to *quality* and *faces* features. However, further research effort has to be spent in order to bridge the gap between these results and the ones achieved when aggregating near–duplicates.

Finally, we also analyzed the performance of the reduction-oriented method for the photo collections of individual users. We observed a considerable amount of fluctuations in the results obtained for individual collections. For instance, the average results for the *reduction–oriented* method (with global features) of 0.36 had a standard deviation of 0.144, with individual results above 0.6 for 10% of the collection and below 0.2 for another 10%. This suggests that there is room for improvements by adapting the general model to the user preferences.

Item-oriented Method. Figure 13 also shows the results for the item-oriented method. The *item-oriented* method represents the situation where the selection process does not consider aggregation of near–duplicates: the first prediction model is used to assign a selection probability to each single image, and the 20% with highest probability is eventually selected. Two separate experiments have been conducted by including only *image-level* features in the model, as well as *all* features. The reason of the split is that *all* features already consider near–duplicate information to some extent.

Regarding the *item-oriented* method, it is possible to observe that the exploitation of global information in the *all* set leads to an improvement because such features contain information about the presence of near–duplicates in the collections, although they are not separately handled like in the case of the *reduction–oriented* approach. Overall, the performance of the *item-oriented* is better than the performance of the *coverage-oriented* approach, but worse than the *reduction-oriented* approach.

Coverage-oriented Method. The last part of the discussion consists in an analysis of the role of coverage in the selection process. The *coverage-oriented* approach handles grouping of related photos via the identification of image clusters, each one potentially representing a sub-event in the collection. Although coverage emerged from the surveys as one of the most important criteria for the users, the *coverage–oriented* methods does not perform especially well in our experiments. Actually it is the worst performing of the three analyzed methods. Both methods applied for selecting images from the clusters, mixed and round–robin, expose a similar performance

There are different possible reasons for this discrepancy and the not so strong performance of the coverage based approach. It is interesting to analyze this in some more detail, because there are several methods that exploit clustering as part of photo selection tasks [40, 37]. Firstly, there might be a discrepancy between the perceived and the applied selection criteria. Secondly, there might also be a difference between the clusters generated automatically and the users' perceptions of sub-events. A third aspect to consider is that not all the clusters might be important for the users. Within a collection, there might be photos from a sub-event that the user either simply does not like or considers less important than others. As to be expected, only for a few clusters (7.3%) all images of the cluster are selected.However, for a considerable part of the clusters (43%) no images were selected at all. Given these statistics, the selection done by any pure *coverage–based* method that picks an equal number of images from each cluster will contain at least 43% of not selected images. One possible solution of this problem might be a ranking of clusters.

3.2.3 Insights from the Study

This section summarizes the main insights obtained in this work, both from the user study (Section 3.2.1) and the automatic methods for photo selection (Section 3.2.2). We also compare them with respect to the preservation value dimension defined in Section 3.1.2.

From the results of the automatic methods for photo selection, *quality* and *faces* features are the ones that perform worst. For the quality features, this was already expected from the results of the survey, revealing that the *Quality* dimension is not of primary importance for preservation in personal scenarios. The *faces* class of features alone also was not a very good indicator, although the presence of importance people was rated as highly important in the survey. The introduction of more powerful processing techniques like face clustering, to know who is actually appearing in the photos, might help in making this class of features more discriminative.

Since a wide part of the state of the art methods for photo selection and summarization [41, 40, 11, 37] considers clustering and/or coverage as primary concept for generating selections and summaries, we clustered photos and compared clustering results with human selections. Moreover, the *event coverage* criterion considered in the user survey, representable through clustering, has been identified as important during our study (Section 3.2.1). These high expectations on the *Coverage* dimension were not confirmed by the experiments, which showed that emphasizing coverage by selecting images fairly from each cluster did not achieve better results than predicting the importance of single images alone. In our opinion, one of the main pitfalls of applying clustering to emulate human selections for long-term preservation is that not all the clusters might be important for the users. There might be photos from a sub-event that the user either simply does not like or considers less important than others. Given the highest expectations in the *Coverage* dimension, we plan to further investigate how to incorporate clustering and coverage in our photo selection methods.

3.3 Preservation Value in Social Media

Preservation of human generated content in social media is another scenario in which we investigate a conceptual model for Preservation Value. In our study we take into account the different dimensions of the preservation value by using not only the content itself but also different types of features for the popularity of the content as well as the social graph and the relationship of the persons relevant to the content. A preliminary study of our analysis was presented in D3.2 together with the first data analysis. In this deliverable we present a larger user study and deeper analysis of the new collected data, which especially focused more on the temporal aspect of the ratings collected from the users for Facebook content of different types. More crucially our new work doesn't stop at the analysis of the data. Rather we use a training set for learning to rank memorable posts based on an extensive set of features. Building models for ranking users is not covered in the D3.2. Finally, by applying feature selection method and evaluating their performance for the ranking models, we also identify a compact yet effective set of core features that are most helpful in ranking memorable posts. This is also a new aspect which is not covered in D3.2.

User Study

In order to better understand human expectations and build a ground truth of memorable social network posts, we set up a user study on top of the Facebook platform. The main goal of this evaluation was to collect participants' opinions regarding retention preferences for their own Facebook posts, shares and updates for Facebook posts from different time periods.

Evaluation

For facilitating the participation, we prepared an intuitive evaluation in a form of Facebook apps. The setup and first data analysis for a preliminary evaluation have already been described in D3.2. In order to participate, users have to log in with their Facebook credentials and grant the app the permissions to access some of the participant's Facebook information, such as, profiles, timeline, and friendship connections. After that, participants were presented with a running list of their posts.

In the study, each participant had to judge their own posts on a 5-point Likert scale answering the following question: *Which posts do you think are relevant and worth keeping for the future needs?* Once a post is rated using 5 points starting from 0 (irrelevant) to 4 (extremely relevant), it fades out providing more space for posts to scroll up. The evaluation interface of a single post contains information about its author, creation date, description, image, etc.

We asked participants to judge about 100 to 200 of their posts. It is important to note that we are not judging participant's memory skills, but instead we are collecting their personal opinions. Due to that, we presented participants' posts in a chronological order starting from the latest. In this case, for more active users, 100 posts may date back to just a few days, reaching up to months for less active ones.

In order to not only get ratings for the most recent posts shown to users (as it can be biased in displaying posts), we picked starting time points differently among participants in our user study. We defined three distinct groups of participants with respect to the recency of the posts they evaluated. Each participant was randomly assigned to one of the groups at the beginning of evaluation (according to his/her Facebook id). Participants in Group 1 (*recent*) were assigned to evaluate posts starting from the most recent ones (February 2014) in their timelines. Participants in Group 2 (*mid-term*) were assigned to evaluate posts starting from January 2011 and back to this date. Finally, Group 3 (*long-term*) received posts from January 2009 and before (if available). This results in rated posts spanning from 2007 to 2014.

We took extra care regarding the participants' privacy and to comply with Facebook's Platform Policies[3]. First of all, the data collected will not be disclosed to third parties. The data cached represent the minimal amount of required information for the experiments and we avoided analyzing the content itself. If a participant deletes the evaluation app, her cached data is removed from our servers.

Evaluation Results & Data Analysis

In the following, we describe our dataset collected from the user study and give an overview analysis. The study was performed between the second week of November 2013 and the third week of February 2014. We had 41 participants, 24 males and 17 females, with age ranging from 23 to 39 years old. In total, there are 8494 evaluated posts. Additionally, once the users provided us authorization to access their data, we were able to collect general numbers that help us to depict the general use of Facebook social network. From 2008 to 2014, on the average, there are about 30 to 100 posts annotated per year and user. For 2007, only 3 participants evaluated posts, thus we discarded posts from this year for most of the analysis.

Facebook defines seven types of posts, namely, link, checkin, offer, photo, question, swf and video. This basically describes the type of content that is attached to a post. In the dataset, we found the following distribution among these categories: 42.5% of evaluated posts consists of status updates, followed by shared links (33%), photos (19%) and videos (4%). We disregarded *swf*, *offer* and *checkin* types, which do not have sufficient occurrences to be significant. We also investigated the distribution of different content types over years. Our observation is that there is a clear increase in the use of videos and photos over time. Several factors help us to explain this change in behavior. First, the catch up of broadband connection allowed users to quickly upload large amounts of data (photos and videos). Second, the dissemination of smart phones with embedded cameras played an important role. Nowadays, anyone can quickly take a snapshot and upload it on the Web. Statistics from photo sharing website Flickr[4] show that the most used cameras are, by far, embedded smart phone cameras[5]. The rate of links and status information changes over years, however, there is no clear trend seen.

In addition, the second characterizing field (so-called *status_type*) defines predefined actions taken by the user. For example, the type 'photo' can have *status_* type 'added_photos', 'tagged_in_photo', or 'shared_story'. Due to space limitation, we left out the analysis of these types, nevertheless, we use these features in our experiments. The distribution of ratings in the full dataset shows a clear dominance (57%) of irrelevant posts (with rating 0). This is followed by 16% of posts with rating 1; 13% with rating 2; 8% with rating 3 and 6% of top-rated posts (with rating 4). This results in an average rating of 0.92 with standard deviation 1.58 and variance 1.26.

When looking into the individual content types, we found out that *photos* have the highest average rating of 1.94 followed by *videos* with an average rating of 1.27. The average

[3] https://developers.facebook.com/policy/

[4] http://www.flickr.com

[5] http://www.flickr.com/cameras

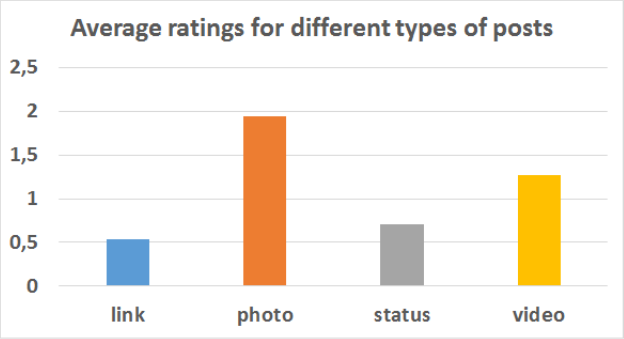


Figure 14: Average Ratings by Content Type.

ratings of *status* updates (0.71) and *links* (0.54) are much lower suggesting a dominating role of photos and videos for memorable posts.

Besides other features relevant for identifying memorable posts, we are interested in the role of time in deciding about content retention, i.e., whether older content on the average is rated lower than more recent content. For this purpose, we analyzed the dependency between content ratings and the age of content. In Figure 15, the dash line shows the average rating for the different years of content creation.

The image shows a clear trend where participants in the evaluation assigned higher ratings to more recent posts. This is in line with the idea of a decay function underlying the content retention model. The decrease in the average rating with growing age of content is especially steep in the first year (2013/2014). However, we also see an increase in the rating values in 2008, where we leave this for further investigation. In Figure 15 (solid lines), we see the development of the average ratings for individual content types (videos have only been included started in 2010 because the number of videos in the dataset before 2010 is very small). Once more, we observe an increase of ratings for the most recent content. However, we also see very high average ratings for older photos (older than 5 years). Thus, we assume that seeing these older (already forgotten) photos again caused some positive surprise, which resulted in higher ratings. This perception would support the idea of creating Facebook summaries for reminiscence. However, this still would require further investigation, since unfortunately only a rather small number of photos was available for rating in 2008 and thus the observation is only based on a rather small data set.

For analysing the temporal behavior in more detail, Figure 16 shows the distribution (ratio) of ratings 0 to 4 among the content categories over years (content age). Besides the dominance of content rated as irrelevant (blue line, rating 0), we can see a clear decrease of content rated as irrelevant for the more recent content. Accordingly there is an increase of content rated *relevant*, which is shown in the doted green line as an aggregation of content rated from 1 to 4. An increase of positively rated can especially be seen for content rated as very relevant (red line, rating 4), which increases its rating by a factor of 4 from 5 percent to about 20 percent between 2010 and 2014. Figure 16 also supports

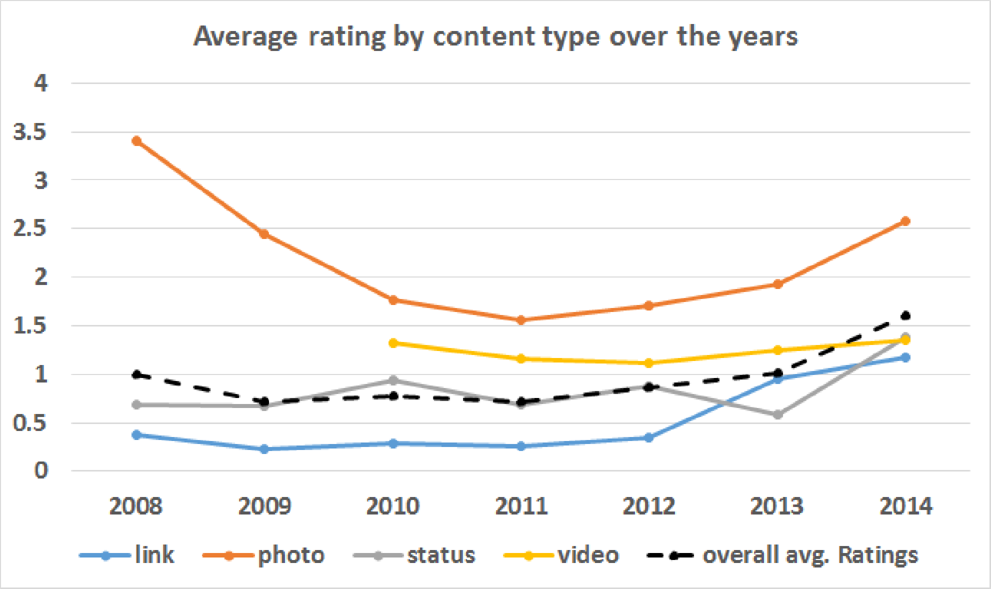


Figure 15: Observed average ratings for increasing content age (the dash line), and the average ratings for individual content types over increasing content age (solid lines).

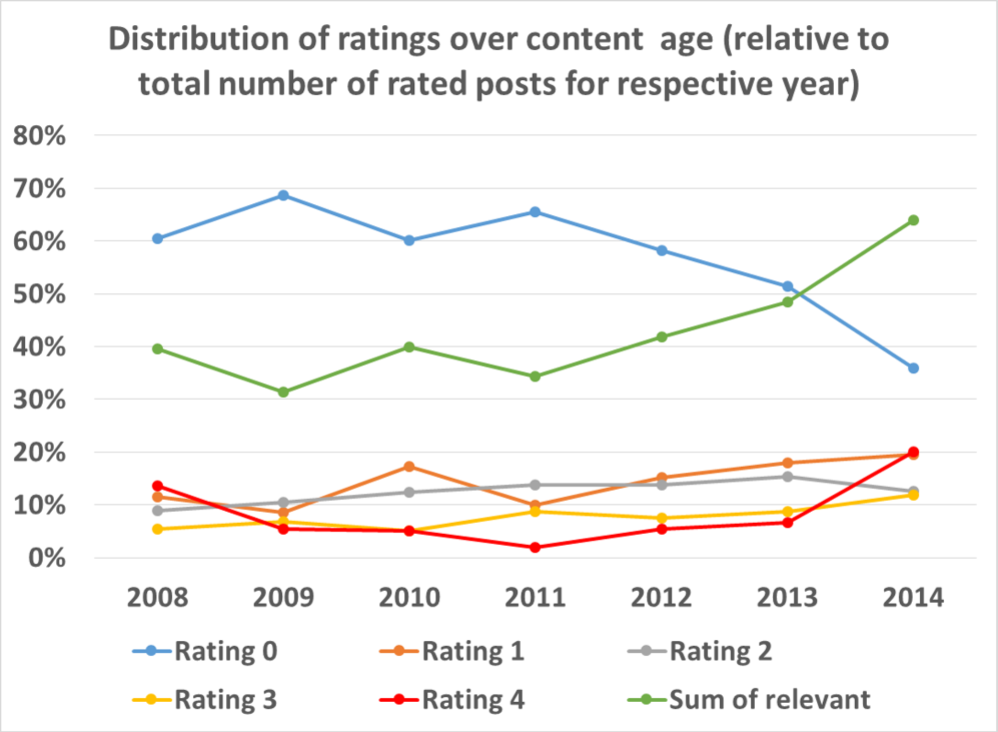


Figure 16: Distribution of the ratings per content types over years.

the idea of a decay function (forgetting) when assessing the memorability of older posts.

On Facebook, users are able to *comment* or *like* a particular post. *Comments* are a very common feature on the Web. Any user (with a required permission) is able to express her opinion on the subject. *Like* is a particular feature of Facebook where users vote to show their appreciation on a given post. Both actions have a limited audience that is imposed by the author of the original post. In the most common case (default setup), a user's post is visible for her network of friends, and those are the ones that are able to comment and like the post. Our first analysis of results already suggests that the average number of comments and average number of likes for the evaluated posts have an impact on the rating of the post: the higher the number of likes and number of comments, the higher the rating.

From these statistics, one can deduce first ideas for determining features that have a high impact in the identification of memorable posts. Roughly speaking, recent photos with high number of likes and high number of comments, seems to be the best evidence.

Qualitative Feedback

To further complement our user studies, we asked the participants to fill a short survey in order to give us qualitative feedback to understand people's Facebook habits. All participants stated that the main use of Facebook is for communication, 83% claimed using Facebook in private life matters and 50% for business. Additionally, 58% of the participants claimed to use Facebook as a personal archive.

Regarding this preservation aspect, half of the participants claimed to have deleted (at least once) older posts. Almost 74% of the participants also stated that they have removed their tag (once or more) associated to a particular picture. Interestingly, when asked if they 'care' about their profiles and timelines, only 50% answered yes, contradicting with the fact that they sometimes 'clean' their profiles.

To summarize this section, we performed a user study in order to collect a ground truth for memorable posts. The primary data analyzes show that most of the posts (60%) are considered expendable and, temporal aspects have significant impact on memorability perception. In this light, in the following sections we present a series of experiments in order to empirically uncover the best features that identify memorable items.

Features for Retention

The annotated dataset provides the basis for building models for ranking memorable blocks, as well as the feature selection experiments, which allow identifying a compact core feature set for the task at hand.

For capturing factors that might influence retention decisions of users, we compiled a broad set of 139 features. They can be categorized in 5 groups: temporal, social interactions, content-based, privacy, and network. The inclusion of temporal features is inspired by the idea that retention preferences are influenced by a decay function as it was also confirmed by the data analysis in the previous section.

For temporal features, we consider the temporal aspect of the post in terms of creation date, age, and lifetime. While *age* is the time between the evaluation and creation date, i.e., the time the post was created, *lifetime* is measuring the active time of a post starting at the time it was created to the last update. We also use variants of the age feature, which use the time of the last update and the time of the last commenting, respectively, instead of the creation time.

The social features capture core signals of social interaction in a Social Web application, covering the features that are typically used in Facebook analysis: number of likes (No.Likes), number of comments (No.Comments), and number of shares (No.Shares).

The next group of features are content-based features. We use the type of post and some specific meta data about the post. This is based on the content-based features offered by Facebook and includes, for example, features such as status type, type, hasLink, hasIcon, and app type etc. We map each categorical feature (like status type) to multiple binary features. To respect user privacy and the privacy policies, the only text-based feature in our set is the length of text included in posts and comments. In other words, we do not utilize the textual content of posts.

The privacy features are based on the privacy options offered by Facebook, which are used to restrict the access to a post to a particular set of user.

Furthermore, for each post we compute standard network measures, such as, cluster coefficient, density, diameter, path-length and modularity. The network features are extracted from the network of all users involved in a post, and also on the networks of likes and comments.

We also use a personalized normalization (Pers.) for the social and network features based on the average values of the collections of individual users. This better adapts the features to the individual characteristics and behavior of individual users. Including the normalized version of the features, we use 5 temporal, 6 social features, 47 content-based, 13 privacy, and 68 network features. The number of network features is relatively high since we analyze the typical network features for the different types of social graphs that are implied by Facebook interactions separately (e.g., likes, messages, and comments), and also include a normalized version of each of these features.

Learning Models for Ranking with Feature Selection

Based on the candidate features, our goal here is ranking a user's posts to identify the most memorable ones as it is a crucial stage for various interesting applications, such as constructing targeted summaries. To this end, we adopt strategies from web search domain, where machine-learned rankers are heavily investigated and, as far as we know, incorporated into commercial search engines [28]. If we make an analogy, a user in our case corresponds to a query in the search setup, and user's posts correspond to the documents retrieved for the query. During the training stage, for a given user u, we construct an $m-$dimensional feature vector F for each post of this user, and augment the vector with the label r assigned to this post (obviously, labels are the ratings collected in the user study, after a mapping to 0-4 range as the posts rated as 1 are not at all considered memorable). For the testing, we feed vectors in the form of $< u, F >$ to the learnt model for each user in the test set; and the model outputs a ranked list of posts for each user. We evaluate the success of the ranking using a typical metric from the literature, namely, Normalized Discounted Cumulative Gain (NDCG), which is a rank sensitive metric that take into account graded labels. We report NDCG scores at the cut-off values of {5, 10, 15, 20}.

In the experiments, we employ a well-known algorithm, namely RankSVM, from learning-to-rank literature [22]. Instead of single data instances, RankSVM considers the pairs of instances (posts of a user, in our case) while building a model. We apply leave-one-out cross validation due to our relatively small dataset including 41 users participated in the user study described above.

Figure 17 reveals the performance of RankSVM for ranking posts using all the proposed features. As a baseline, we also train a model using three basic social features, namely, the number of likes, comments and shares. We choose the latter features for the baseline ranker as they are the most intuitive popularity signals in social web and very likely to be involved in practical applications, such as the Facebook's post ranking applications discussed before. The results show that the proposed features are actually very useful, and using these features for training a ranker yields relative effectiveness improvement of up to 26 percent in comparison to the baseline ranker trained with the basic social features (i.e., compare the first and last bars for each metric in Figure 17).

The next question we address is: Can we identify a subset of the proposed features that has the highest impact in ranking memorable posts? While feature selection methods are widely applied for various classification tasks, only a few works have investigated their performance in a learning-to-rank framework [20, 15, 16]. Here, we adopt two filtering-type approaches: In the first approach, so-called TOP, we use each feature on its own to rank the posts of the users, and then choose top-N features with highest effectiveness in terms of the NDCG@10 scores [15]. Secondly, we adopt the so-called GAS (Greedy search Algorithm of Feature Selection) introduced by Geng et al. [20]. In GAS, we again compute each feature's isolated effectiveness, but additionally, we also compute pairwise feature similarity, i.e., to what extent the top-10 rankings generated by two different features correlate. To compute the similarity of two ranked lists, we use Kendall's Tau metric. Then, the feature selection proceeds in a greedy manner as follows: In each iteration, first the feature with the highest effectiveness score is selected. Next, all other features' effectiveness scores are discounted with their similarity to the already selected feature. The algorithm stops when it reaches the required number of features, N.We experiment for all possible values of N, from 1 to 138 (as $N = 139$ is the case with all features), and evaluate the performance. Figure 17 also demonstrates that feature selection with TOP strategy (for the best-performing value of N, which is found to be 41) cannot outperform models trained with all available features. However, using GAS strategy and with only 30 features we can achieve the same effectiveness as using all 139 features (i.e., compare the third and last bars for each metric in Figure 17).

For this latter case, we analyze the features selected by GAS in each fold to identify the

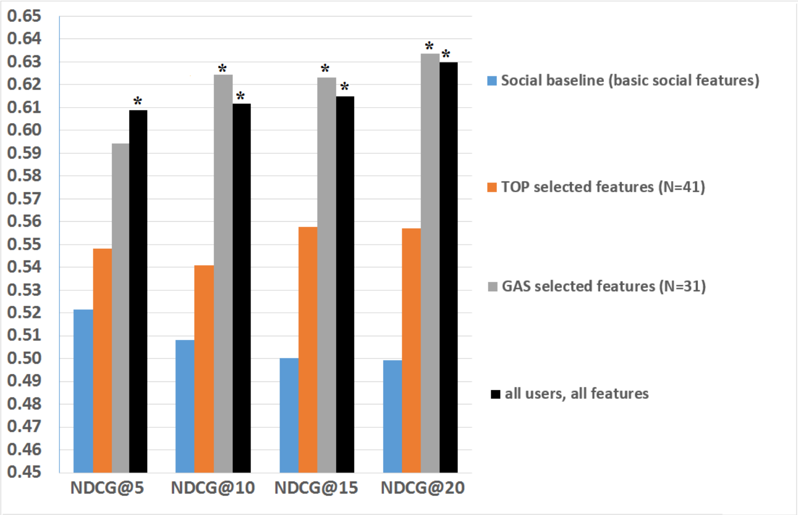


Figure 17: Effectiveness of ranking general

most promising features for the task of ranking posts. As the absolute value of the weights assigned to features by the RankSVM model (built using a linear kernel) can reflect the importance of features [9], we average these absolute values for the features appearing in the learnt model in each fold. Table 3 shows the features with the highest average scores in the models built after feature selection with GAS.

We see that the most discriminative features cover all four feature categories except the privacy. Regarding top-10 features in the list, content-based and social categories, with 4 and 3 features respectively, are the most influential ones. We also observe that features that are adapted to the behavior pattern of individual users such as pers.no.likes play an important role for ranking memorable posts. Interestingly, the feature content type is outperformed by the feature status type. Our explanation is that status type can be considered as a refined form of content type, which captures what has been done with the content (e.g. 'added photo'). This more fine grained information has proven to be more important for the ranking task than the information on the content type alone. Network features computed from the graphs in the social network have less impact for the characterization of memorable posts. And finally, the temporal feature age is also among the top-10 features, a finding in line with our earlier claims on the relationship of the time and human retention preferences.

In Figure 17, we present the results of the ranking posts. Our results are encouraging in that, even in a relatively small setup with 41 users in total, the personalized approach improves the NDCG@5 and NDCG@15 scores in comparison to using a single ranking model (i.e., compare the first and last bars in the figure for each metric). Finally, we also experimented with feature selection in this setup, where we applied Top-N and GAS strategies. In Figure 17, we only report the results for the better performing GAS strategy. It turns out that, feature selection using GAS strategy is outperforming the basic social features significantly, denoted with $\star$ in Figure 17, as well as the set of all features for NDCG@10, NDCG@15, and NDCG@20.

Table 3: Best features selected by GAS. The core social features (baseline) denoted by *.

Category	Feature	Weight
content-base	Pers.Length Description	1.676
social	No.Shares*	1.190
social	Pers.No.Shares	0.471
social	No.Likes*	0.453
network	Overlap.No.Friends (all)	0.392
content-base	Status_Type	0.329
temporal	age (created time)	0.327
content-base	Pers.Length.Story	0.326
network	Cluser Coefficient	0.320
content-base	Length Message	0.269
social	Pers.No.Likes	0.260
content-base	APP_Type	0.259
content-base	hasMessage	0.200
content-base	Length Story	0.199
content-base	hasDescription	0.167
content-base	hasStory	0.151
content-base	Type	0.140
content-base	Length Comments	0.140
content-base	Pers.Length Message	0.115
temporal	CreatedTime	0.112
network	Density (all)	0.091
social	Pers.No.Comments	0.065
content-base	Length Description	0.063
network	Overlap.No.Friends (likes)	0.059
network	Pathlength (like)	0.023
network	Overlap.No.Friends (tagged)	0.009
social	No.Comments*	0.006

4 Policy-based Preservation

As already discussed in the previous sections, high flexibility is required, when dealing with memory buoyancy and preservation value. Especially, when exploiting those values and when basing decisions upon them, a variety of preservation settings and preferences have to be supported. Therefore, in the ForgetIT project, we complement the machine learning-based approach for deriving memory buoyancy and preservation value from observed evidences with a policy-based solution, which enables the user (or organization) to customize preservation strategies according to their needs.

Policies (implemented in terms of rules) can be used for different purposes in our preservation framework: (a) they can be used for mapping the actual computed numeric values for memory buoyancy and preservation value to semantic categories (see figure 4 for an example mapping for the preservation values) and (b) in a second step they can be used for deciding upon actual forgetting actions such as archival, summarization, level of contextualization, etc.

In this section, we focus on the foundations for the policy framework, its computational framework and on option (a) i.e. mapping the raw values to semantic labels in a customized fashion leaving the actual decision about the action upon the semantic label to th respective application. Work on option (b) for the mapping is left to the next version of the policy framework.

4.1 Conceptual Model

The starting point of our conceptual model of the policy framework are **preservation scenarios**, i.e., the general setting, in which the preservation decisions are taken. Example preservation scenarios are ”management of private photo collections”, ”management of personal professional document collections in a project context” and ”management of press releases in a content management system”. A preservation scenario is defined by the types of documents considered, the purpose and tasks related to the documents and the more general context such as organizational vs. personal preservation. Applications using the PoF framework will typically focus on one (or a small set of related) preservation scenario. There are, however, also applications such as the Semantic Desktop, which target different types of preservation scenarios such ”as management of private photo collections” and ”management of personal professional document collections in a project context”

Different preservation scenarios require different ways of dealing with forgetting and preservation, in order to reflect the special characteristics and requirements of the respective preservation scenario. Therefore, for each preservation scenario supported by an application **forgetting policies** are pre-defined that can be customized by the user. (Obviously, it is not desirable that such rule sets are built from scratch by the user.) As a further step of abstraction it is envisioned to organize possible policy options into more fundamental

forgetting strategies, which enables the user to chose from a small set of fundamental approaches for preservation such as a conservative approach vs. a more deliberate and focused approach. The strategy selection, would then be translated into a set of policies that can be further customized by the user.

Hence, the preservation policy framework is organized into four areas: the high-level layer of preservation scenarios and the orthogonal forgetting strategies, which are together mapped to a customizable set of forgetting policies, which are in turn mapped to one or more sets of rules, which implement the respective policies.

As a foundation for the definition of policies in terms of rules a **policy vocabulary** is required, which enables the expression of rules in terms of relevant concepts. In more detail, for the ForgetIT policy framework, the policy vocabulary consists of the following parts:

Raw forgetting values These are the values for memory buoyancy and preservation value as they are computed using the evidences learned from the application contexts. In the rules they are uses as the starting points for the mappings to preservation categories;

Preservation categories These are semantic categories expressed in terms of category labels, which the forgetting values are to be mapped to. Example labels are ”gold”. ”silver”and bronze for the preservation value. Their main purpose is to give meaning to the raw forgetting value, thus easing decision taking upon the value in the respective application.

Domain concepts A core building block of the formulation of forgetting policies are relevant domain concepts typically encapsulated into a data model of the domain. This might include relevant roles of persons, document types (including semantic document types such as ’agenda”, domain events (such as ”closing of a project”) and in more general, all relevant kinds of resource properties and relationships. These domain concepts are used to express rules, which do not treat all resources in the same way. It, for example, enables to express a rule like ”never delete emails from the boss”.

Forgetting actions These are actions, which can be performed based on preservation categories. As already mentioned above the definition of forgetting actions is currently left to the respective application.

The rules defined based on the policy vocabulary are organized into policy-specific rule sets. In addition to those rule sets the system will also contain sets of more generic meta-rules, which help in the implementation of the policies. This, for example, includes rules for conflict-resolution.

4.2 Computational Framework

From the computational perspective, there are many ways to implement a policy framework and integrate it into an information space (personal or organizational). One widely accepted solution in enterprises is *Business Rule Management System* (BRMS)[6], a software system that enables specifying, analyzing, executing and deploying logical rules used within different units of an organization. In ForgetIT, we inherit the BRM models partially, and adapt them for our preservation scenarios. The re-use and adaptation of BRM model has many benefits. BRMS is a well-established approach in information systems management and databases, especially in enterprise resource planning (ERP) area. This makes the policy component in ForgetIT easier to adopt for organizations. In addition, there are several open-source BRMS solutions widely adopted in industry (as well as in academics) such as Drools, OpenRules, BaseVISOR, etc. In ForgetIT, our policy framework is developed using Drools[7], one of the most popular BRMS open-source framework developed in Java. While Drools targets a full-fledged, end-to-end business rule management with several software components, including an integrated development environment (IDE) for software developers, in ForgetIT, we only make use (with adaptations) of the following components, which are relevant to our policy framework:

- **Drools Expert**: This is a rule engine based on the RETE [19], a common pattern-matching algorithm for implementing logical rules in production systems. In ForgetIT, we run the RETE algorithm via Drools Expert API to execute rules and identify conflicts, and develop a simple conflict resolution based on some meta rules (see Rules Repository, 4.2).
- **Drools Workbench**: This is a web-based application to enable humans to specify and manage rules using graphical editors (called Drools Guided Rule Editor). In ForgetIT, we re-use the Drools Workbench and simplify its interfaces for organizational preservation scenarios.

Figure 18 sketches the essential components of a typical business rule management system, and how it is supposed to be integrated with other components, as is advised through an agile development process [6, 2]. Although this architecture is heavily tailored towards enterprise settings, we will adapt each component to match the requirements of our policy framework as discussed below.

Rules Repository

This is the core of a BRMS, and also of our policy framework. It keeps all rules in a centralized place for the rule engine to be able to identify conflicts, to execute and derive information or new facts. In our policy framework, rules can be human-specified or inferred from user's preferences, and are grouped in different *rule sets*, which implement forgetting policies. One rule set usually corresponds to the forgetting policies within one

[6] `http://en.wikipedia.org/wiki/Business_rule_management_system`

[7] `http://www.drools.org/`

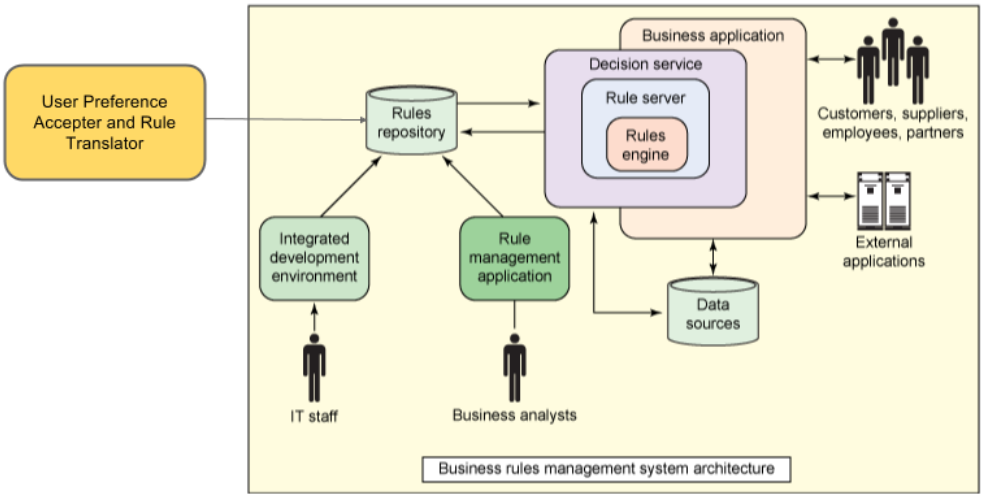

User Preference Accepter and Rule Translator
Rules repository
Business application
Decision service
Rule server
Rules engine
Customers, suppliers, employees, partners
External applications
Integrated development environment
Rule management application
Data sources
IT staff
Business analysts
Business rules management system architecture

Figure 18: Architecture of a desired Policy-based Framework[8]from BRMS perspective

preservation scenario (for instance, rules for preserving personal documents at work), but a preservation scenario can have more than one rule set (for instance, each type of user preference has one set of rules, see Authorship below). The grouping of rules into different sets allows the managed forgettor to load only relevant rule sets to execute in a specific preservation scenario, hence keeping the reasoning complexity and responding time of the web services low.

Besides the domain-specific rule sets, we also store common rules that are shared and can be applied in all preservation scenarios (personal or organizational). These are called *meta rules*. For instance, the rules used to resolve the conflicts between different rule sets will be stored as the meta rules.

Drools stores its rules in a centralized repository built by the JBoss Guvnor [9] - an SOA repository management tool written in Java. Rules that are edited via Drools Workbench are output as text files with `.drl` extension. In ForgetIT, we use a relational database (MySQL) to store references to these rule files (i.e., location of the corresponding artifacts in the Guvnor repository). The database also stores the meta rules, the mappings between rules and the rule sets, and the mappings between the rule sets and forgetting policies in support of preservation scenarios.

Authorship: Within a preservation scenario, rules are associated with users or user groups. In personal scenarios, the user can specify her preferences, each linked with a separate rule set. She can customize or add her own rule set. In organization scenarios, each unit

[9] https://developer.jboss.org/wiki/Guvnor

in an enterprise can have its own policies. For instance, upon the closure of a project, an administration group can have different policy towards preserving documents from a technical group. Hence, our rules repository organizes preservation scenarios into different author groups (users or organizational units), and assigns each group with a rule set.

Business Application

The business application consumes business rules to perform its functions and features, and delivers the results either directly to users or to other applications / components. In our case, the business applications correspond to the two components in the managed forgettor, which compute the memory buoyancy and preservation values. After using machine learning algorithms to assess the values, these two systems will call the decision service (policy framework) to map and adjust the values based on information about user preferences, organization policies, contexts, etc.

Data Source

The data source stores data items used and generated within the business applications. In our case, the data source can be personal documents of a semantic desktop using the PIMO model, or resources in a TYPO3-based system. The data model of items (i.e. structures of the data schema and their relationships) is defined using POJO standard and imported into the Rules Repository, so that the rule management system (Drools Workbench) can understand the rules and load the rule set properly to the editors. Each item has two label fields to store the MB and PV values as results of the policy framework. For each label, the data source stores the timestamp of when the label was assigned to the item, so that it can identify out-dated labels and trigger the re-computation if necessary.

Decision Service

The decision service component provides service for applying rules in the managed forgetting components (memory buoyancy and preservation value assessors). This forms the core computation unit of our policy framework. Upon requests from the business applications, the decision service queries the data store and returns the labels of items (preservation categories for MB and PV). It will send the computation / re-computation requests to the Rule Server in some of the following cases:

- The labels are not available, i.e. the information assessment (MB or PV) is not performed yet on the item.
- The labels are available only through the machine learning process, not yet checked against human-defined policies.
- The labels were assigned to the item too long (i.e. after the assignment items metadata gets updated) and subject to be out-dated.

The Rule Server is responsible for applying rules to items on demand and output the corresponding preservation categories. It first finds the applicable rule sets for the items, based on the current preservation scenario (and customized set of policies) and on the context / meta-data of the items, and create a session for applying the rules (in Drools, it is called *knowledge session*). In one knowledge session, the Rule Server creates a *rule engine* instance via Drools Expert API. The rule engine fires all the rules in the rule sets against the data of the items and generate the labels. If it detects any conflict in applying the rules, resulting in more than one labels for an item, the rule engine informs the Rule Server about the need for conflict resolution. The Rule Server then loads the meta rules and uses them to resolve the conflicts. If a conflict cannot be resolved, it sends the error messages back to the decision service, which can then decide to halt the computation or to forward the messages to the business application.

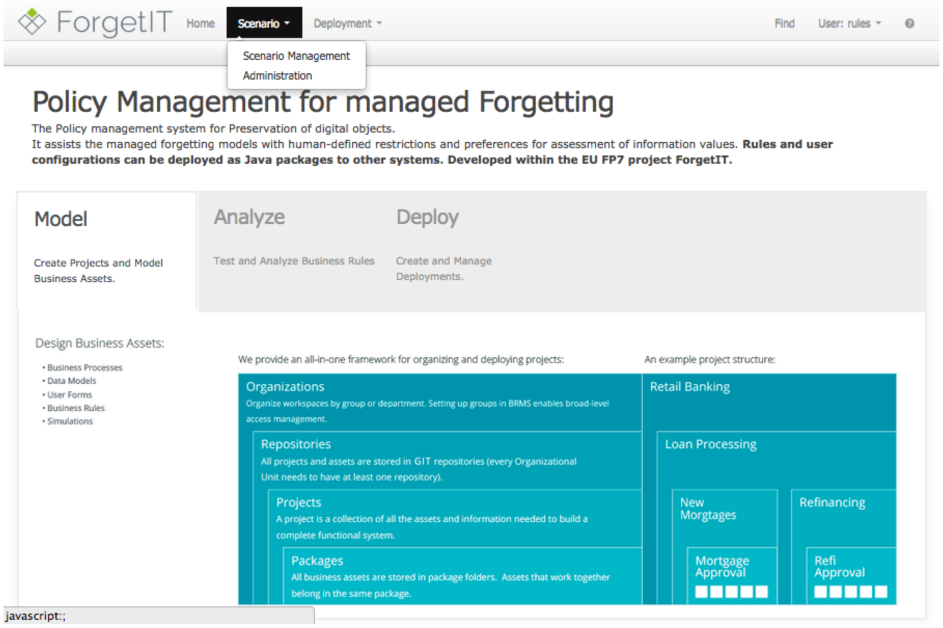
Integrated Development Environment

The integrated development environment (IDE) is a component in a BRMS that allows software developers to query, handle rules directly from the rules repository. It also supports debugging the rules in the presence of conflicts, via some visualization tools (such as Drools Eclipse Plugin). In our policy framework, this is not the focus, thus we skip this component.

Rule Management Application

The rule management application allows rule experts in an organization to analyze and manage all rules from the rules repository. An expert, once granted permissions to the repository, can use the management application to access all authorized preservation scenarios and rule sets, which implement forgetting policies. Figure 19 illustrates the main features of a web-based rule management application using Drools Workbench. The expert manages the model in two levels: Administration and Scenario Management. The Administration allows defining and managing the common configurations such as organization units (or author groups), URI of the rules repository, or monitoring activities log of the other experts accessing the same repository. The Scenario Management enables defining and configuring a new preservation scenario, specify the relevant data models for the scenario, as well as creating plocies in terms of rule sets for different organization units and editing rules using graphical or plain-text interfaces.

Figure 20 shows an example graphical interface for one rule using Drools Guided Rule Editor. Once the data model, which encapsulates the relevant domain concepts, has been specified and the rules repository has been connected, the editor can read rules from `.drl` files in the repository, and use the data model to visualize the rules in the editor, with all available fields of the related schema. The expert can use the interactive dialog to modify the rules, or create a new rule from the available schema (policy vocabulary). He or she can also write the rules directly in Drools Rule Language format [1]. After editing the rule, the expert can validate the rules using “Validate” button on the top right of the



Policy Management for managed Forgetting

The Policy management system for Preservation of digital objects.
It assists the managed forgetting models with human-defined restrictions and preferences for assessment of information values. **Rules and user configurations can be deployed as Java packages to other systems. Developed within the EU FP7 project ForgetIT.**

Model

Create Projects and Model Business Assets.

Analyze

Test and Analyze Business Rules

Deploy

Create and Manage Deployments.

Design Business Assets:

- Business Processes
- Data Models
- User Forms
- Business Rules
- Simulations

We provide an all-in-one framework for organizing and deploying projects:

Organizations
Organize workspaces by group or department. Setting up groups in BRMS enables broad-level access management.

Repositories
All projects and assets are stored in GIT repositories (every Organizational Unit needs to have at least one repository).

Projects
A project is a collection of all the assets and information needed to build a complete functional system.

Packages
All business assets are stored in package folders. Assets that work together belong in the same package.

An example project structure:

Retail Banking
Loan Processing
New Morgtages
Mortgage Approval
Refinancing
Refi Approval

Figure 19: Screenshot of the Rule Management Application - Home Screen

editor screen. Finally, the expert invokes the “Deployment” feature on the top banner of the management screen (Figure 19) to deploy the rule set into the rules repository.

User Preference Acceptor and Translator

The BRMS architecture addresses the traditional scenarios in enterprise business. To use them in personal preservation scenarios, where the users typically have little knowledge about a business rule work flow, we add a new component named “user preference acceptor and translator” as in Figure 18. The main purpose of this component is to enable users to specify - in a user-friendly way - their preferences for preserving their personal documents in some restricted preservation scenarios (such as digital photo preservation), and to map these preferences into a predefined rule set.

4.3 Case Study: Personal Professional Preservation Scenario

As a case study, we investigated the component features in the *professional personal asset* scenario, where the goal is to decide which of the personal assets (documents) used at work to be preserved. This includes various types of assets such as minutes and presentations created and used for professional tasks and for professional events (such as business meetings, project closure, etc.). Figure 21 shows the domain concepts relevant for policy definition in terms of a data model used in this scenario. The most important concepts are:

- **Person** This concept defines the persons involved in the professional events in dif-

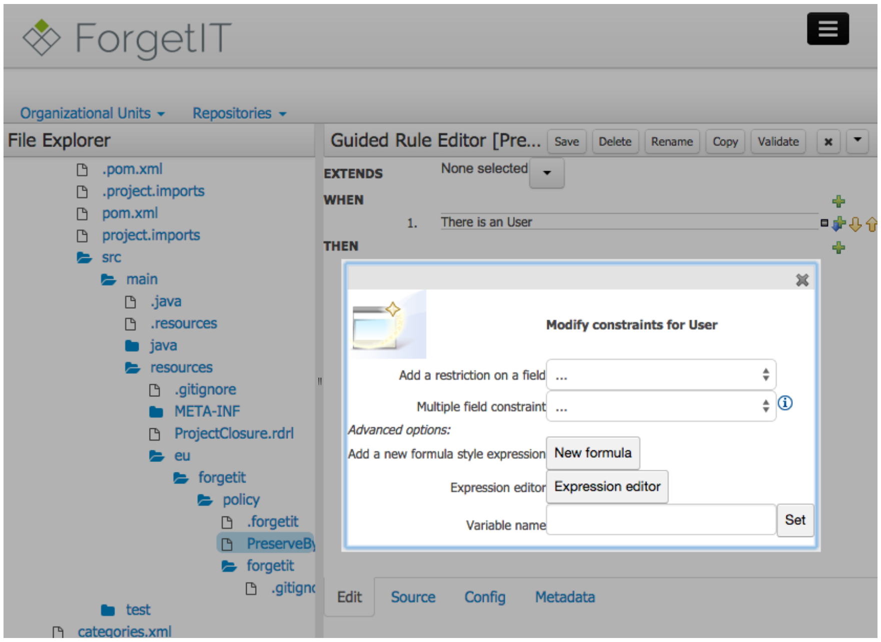


Figure 20: Screenshot of the Rule Management Application - Rule Editor

ferent roles (project manager, administrator, secretaries, etc.) and who are authors or related to some digital resources

- **Document** This is the main asset to be preserved. There are three types of documents targeted in the scenario: Images, Files (text documents, Office files, etc.) and Emails. Each document has one person as the author and can be accessed by or shared with other persons.

- **Collection** Documents are grouped into collections. The preservation rules can be applied to individual documents, or to the entire collection.

- **Task** Each document at work is associated with some business or personal tasks (for example, for a meeting, or for bookmarking a trip information, etc.).

- **Project** In this scenario, persons work in different projects (finished or ongoing).

- **Topic** Besides Tasks, documents are also linked to some high-level topics as another organizing dimension. For instance, an email can have topics such as administration, business trip, and so on.

- **Event** This concept defines the business events in which persons participate in. For example, a project meeting is a business event. Event takes place in some **Locations**, and can be linked to some collections, such as a photo collection capturing the social event of the project meeting, etc.

The data model hast to be imported into the rules repository for the rule engine to be able to execute the rules. On the other hand, based on this data model, we can define beforehand a number of fixed rule sets encoding different preservation policies. These

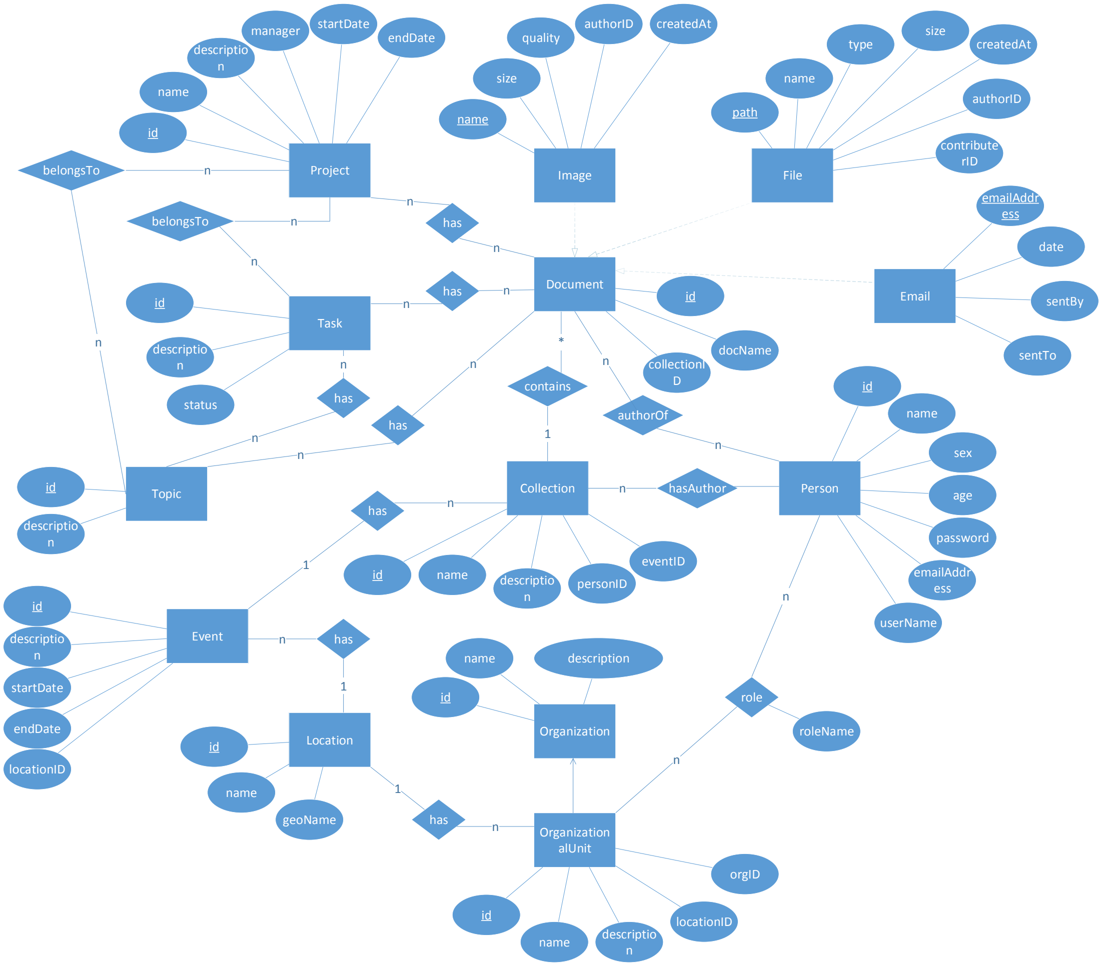


Figure 21: Data Model capturing relevant domain concepts for personal professional preservation scenario

sets are derived from on the surveys conducted in Work package 2 and 10 for photo preservation and TYPO3 preservation scenarios, respectively. Then, from the user profile (specified in the registration step), the system maps to the most appropriate rule set using fixed mapping strategies. The rule engine is used to fire these rule sets against the user's data to label the items, as mentioned above. These labels are displayed in a graphical interface to the user for correction. The user can also edit the rules directly if she needs to adapt the rules for her purpose.

5 System Prototypes

This section presents two system prototypes in support of long-term preservation for two applications, i.e., photo summarization and timeline summarization in social media.

5.1 Photo Summarization and Selection for Preservation

A desktop application, described in D9.3, has been developed to assist users in browsing their collections and selecting sub sets of them for long term preservation. Since it is also endowed with the selection approach described in Section 3.2, we summarize the behavior of the application in the followings with the aim of showing how our photo selection component has been plugged in it.

The application allows the user to import her own collections, obtain automatic selections, revise them, and finally store the selected photos for long term preservation. Two back-end components are available to the user: a photo summarisation component, developed within the context of WP4, and our photo selection component whose behavior has been described in Section 3.2. These components are all available as alternative options, which the user can compare and choose for the final selections. The selection automatically generated for the current collection can be revised by the user, who can either remove selected photos or add not selected photos. Once the user is done with editing, the finally selected images are stored for long term preservation. Moreover, the feedback that the user gave by revising the automatically generated selection is sent back to our components, and used to update the selection model. Assuming a sufficient number of selections iteratively done by the same user, the model would be able to adapt to the particular preferences of the user and, in turn, to improve the goodness of the selections. Given the availability of user feedback, different strategies can be investigated to update the model. Currently, the update of the model is achieved by (i) adding the new labeled data (i.e. the features extracted by the new collection along with the selections of the user) to the dataset, (ii) training the model with the new dataset off-line, (iii) make the new learned selection model available for the next selection. Other updating strategies are under investigation, and they are highlighted as a future research direction in Section 6.2

5.2 RememberMe App

In our RememberMe! app we implemented several summarization strategies of the users' timeline. The reason behind that is two folded. First, we want to provide the users the option to choose the summary they are more pleased with. We believe that, by giving the users this option we can please a higher number of users who will eventually share the application, consequently reach a broader audience and attracting more users. Second, we are able to collect data to analyze and compare which are the best and most preferred strategies. In total we provide 4 distinct template strategies and an additional customizable one described as follows (Figure 22):

Number of likes: This strategies solely relies on the number of likes that each item in the users' timeline has. The top 20 most liked items are presented in the summary.

Number likes over time: This strategy relies on the number of likes that each item has in different time periods of the users' timeline. First, the timeline of the user is divided in 20 'buckets' sorted by time. Each bucket contains one twentieth of the users' posts. On each bucket the strategy picks the top most liked item.

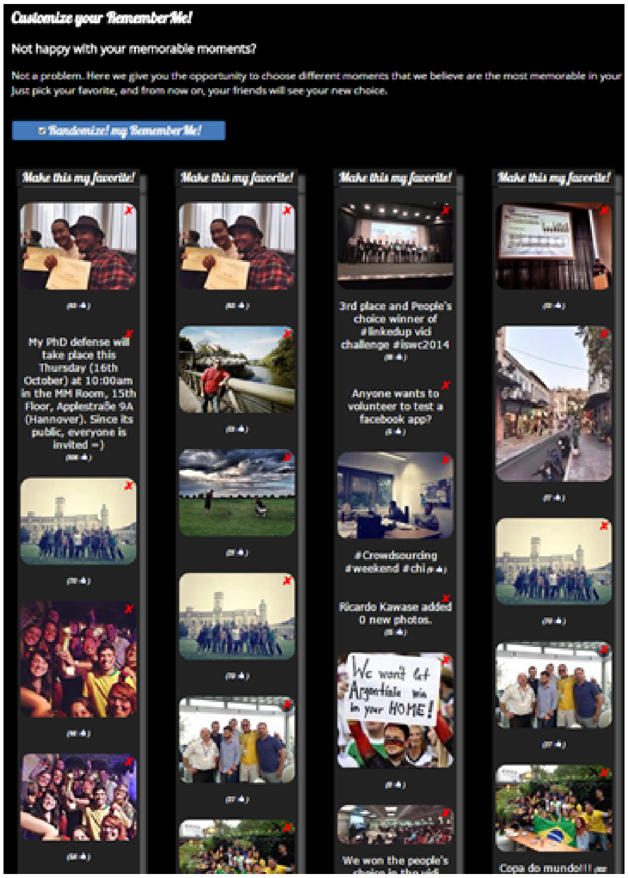
Top active week: This strategy first detects the top 20 most active weeks of the users. From the items belonging to these weeks, the strategy chooses the most liked post.

Random: This strategy randomly selects 20 posts from the users' timeline.

Customizable: With this strategy, the users have the option to manually add items to their summary. To do so, the user must go to Facebook and add the hashtag #RememberMe to the posts they want to appear in their summary. There's no limit for the number of posts an user can add to her summary.

In addition to that, all the strategies allow the users to randomize the exhibition of the posts, i.e. the 'slide show' summary present items in a random fashion other than the conditions imposed by the strategies.

By providing these different strategies we are able to monitor how users choose and switch from one strategy to the other. Additionally, the customizable strategy give us solid ground truth of the items that the users really want to see in their summaries. This ground truth is important for us to perform further research in order to identify the most prominent features of the memorable posts. Figure 22 illustrates the interface where the users are presented with the different summarization strategies.



Customize your RememberMe!

Not happy with your memorable moments?

Not a problem. Here we give you the opportunity to choose different moments that we believe are the most memorable in your
Just pick your favorite, and from now on, your friends will see your new choice.

My PhD defense will take place this Thursday (16th October) at 10:00am in the MM Room, 15th Floor, Applestraße 9A (Hannover). Since its public, everyone is invited =)

3rd place and People's choice winner of #linkedup vidi challenge #iswc2014

Anyone wants to volunteer to test a facebook app?

#Crowdsourcing #weekend #chi

Ricardo Kawase added 0 new photos.

We won the people's

Copa do mundo!!!

Figure 22: The customized view of the user for selecting one of the 4 strategies.

6 Ongoing Research and Research Plans

This section outlines our ongoing research and plans for the next steps in WP3, which aims at providing reusable building blocks for managed forgetting. In more detail, we present two main ongoing research topics, namely, (1) using the Semantic Desktop as the information assessment framework of Memory Buoyancy, and (2) leveraging an adaptive method for personalized photo selection. In addition, we plan to conduct exploratory research in the last year of the ForgetIT project, which includes the policy-based framework, forgetting actions and strategies, investigating preservation value in the organizational use case, as well as the architecture work package (in collaboration with WP8).

6.1 Semantic Desktop: Re-visited

As one case study for evaluating the effectiveness of the Memory Buoyancy (MB) calculation methods, we investigate the use of MB computation in decluttering resources in personal and organizational information spaces. After some time of use, information spaces such as personal desktops and organizational Wikis would profit from any kind of automated spring-cleaning. More precisely, they tend to clutter with information items, which have been collected or created for some purpose or task, but are no longer relevant. Going a step further, it would be even better to declutter the information space regularly, thus keeping it more organized and focused, as well as reducing the cognitive effort for accessing required information items. For the personal information space, this idea of decluttering has gained further importance with the wide use of mobile devices with their restricted storage capacities for all types of personal information management tasks.

In this case study, our approach of decluttering in semantic information spaces is to provide a ranking of documents (photos, text files, etc.) in accordance to their MB values, so as to assist users easily find and organize the information based on their short-term and long-term need. Here in information retrieval (IR) terminology, a document is ranked higher if it is relevant to the user ad-hoc information need (for instance, looking up files to prepare for a meeting), as well as to the user recent activities (for instance, files of recent meetings are more relevant). The latter requirement in fact correspond to the temporal prior for a document [27]. In this case study, we study how different MB values can be integrated into the document temporal priors estimation to improve the ranking performance of documents in personal and organization settings.

6.1.1 Experiment Setup

Datasets Here we introduce the data used in our experiment. Table 4 summarizes the statistics of the two data and topics used. We evaluated our methods using two real datasets. The first dataset Person is PIMO desktop collection, which was collected via installed PIMO clients in individual's personal computers and at work in DFKI (see D9.2), and was used on daily basis in the course of three years. The collection is equipped with a database of access logs from 17 users (7 active). To preserve the privacy, all data were anatomized, a running prototype and evaluation user interface were deployed at each computer of the evaluator, and only statistics and human feedback (detail below) were sent back and could be accessed from outside.

Table 4: Statistics of the Datasets

Data	Person	Collaboration
No. of documents	20363	1437
Time span covered	23.09.2011 - 17.09.2014	09.10.2008 - 10.09.2014
No. of users	17	268
No. of relationships	155539	126326
No. of entries in activity log	337528	217588

The second dataset Collaboration was used to test the generality of the method beyond the personal collection setting. It is L3S wikis dump, the archive of website used by L3S members size as an administration, collaborative and knowledge sharing platform. Beside that, the wikis has also been used as a platform for exchange information, discussions and collection of resources for a wide variety of national and international (EU) projects including ForgetIT. As such, the data consist of information not only about L3S internal activities, but also external undertaking with outside partners. Data are web pages mentioning wiki information, discussion, regulation and policy, technical documents, meeting notes, etc. and are stored in Dokuwiki[10] format. Multiple access and edits to one page is allowed. A proxy server was installed to log HTTP request to the pages. In addition, backup plugin was used to record all changes made to a page, including timestamp, contributor account (within and outside L3S), type of action and the modified content.

Graph Construction To evaluate the propagation method, we need the graph for both collections. For PIMO dataset (Person), we used explicit semantic relationships between documents as annotated by users. This semantic annotations, including concepts, relationships between resources, facts and constant values are developed by dedicated Ontology (PIMO)[38] in RDF style, and form the semantic graphs with documents and concepts in graph nodes and relationships between them form the graph edges. Here we re-used the graph to perform our propagation model, described in D3.2. Weights of one edge from one document to another is estimated by the number of connecting semantic relationships. For the L3S wiki dataset, to have relationships between pages, we created three kinds of relationships and aggregated them into a unified graph, taking average of edges' weight.

1. **isSimilarTo** relationships are established between two pages when their Cosine Similarity measure exceeds a threshold θ (we observed $\theta = 0.$ gives a reasonable

[10] `https://www.dokuwiki.org/dokuwiki`

graph density). For both datasets, we preprocessed the activity log, with timestamps are rounded up to the day scale. Using this relationship, propagation can be triggered in both ways.

2. **HyperLinks** Two pages are connected when there is a hyper links referring to one from the other. Weights were equally set to 1.
3. **sameTopics** Pages are grouped in different namespaces, each encoding some topics or events created by users. We connect two pages if they share the same namespace. The number of namespaces shared is used as the edge weight.

6.1.2 Evaluation Methods

Human Assessment. We seek to evaluate the impact of the frequency features and the propagation in retrieving existing documents. For each dataset, we asked the contributors (3 for dataset Person, 3 for dataset Collaboration) to pick up some dates in the past when they all experienced the similar sets of events in their work (e.g. a project meeting), and was reflected in the data and the activity log. To specify the temporal interest, the contributors were asked to provide one simple hint for each date as a query. (For dataset Collaboration, hints were key words, while for the Person collection, hints were the working documents of the contributor). To adjust the different scenario of human need in the two datasets. We set up the human assessment as follows. For each result, human judge with respect to the query and hit time different values:

1. Person: Human are asked for desired actions on the document: Pin as shortcut (4), display for current use (3), display for future use (2), hide from current screen (1), don't need anymore or irrelevant (0) (see Section 2)
2. Collaboration: Relevant to current context (2), relevant but not to current context (1), irrelevant (1)

To ensure fair comparison in the two datasets, we set up different sets of baselines for each dataset. For semantic desktop, since the text availability is very rare and noisy (e.g. title of photos are always machine-generated), we used personalized PageRank as the baseline. For Collaboration, we use Query likelihood model (QLM), structured entity model using predicate folding (SEM) [11],Temporal Query Model (TQM), and timebased LM (TLM) [27] as baselines. Each baseline was incorporated with two kinds of document priors: The decay functions, and the decay functions with propagation. We evaluated thee two decay functions from Section 2.1.1: Weinbull and MCMC-1 functions, with parameter empirically chosen. Also, for both datasets, we compared against T-Fresh [14], a state of the art in document freshness-aware propagation ranking Web search results. We chose the parameter reported in the paper [14]. Finally, we compared methods against the traditional MAP, Precision @ 10 and NDCG@10 scores.

[11] with 4 groups of attributes as [33]. We set equal weight as it had been shown to perform the best

6.1.3 Experimental Results

Influence of Frequency

In the first experiment, we studied the impact of frequency in decay functions. We incorporated different decay functions and their frequency-recency variants into the baseline, and compare against TLM (Collaboration). The results are shown in Tables 5 and 6. We can see that using frequency-recency decay obviously improves the retrieval performance, as it has more evidences about which resources are more active at the time. The MCMC-1 function gives higher performance in the Collaboration collection, while the Weinbull function has the best performance in the Person collection.

Table 5: Impact of Frequency-Recency Decay in Collaborationcollection

Method	MAP	Prec@10	NDCG@10
QLM	0.147	0.21	0.361
TLM	0.391	0.422	0.49
QLM+Weinbull	0.356	0.253	0.412
QLM+freqWeinbull	0.561	0.526	0.538
QLM+MCMC-1	0.429	0.316	0.460
QLM+freqMCMC-1	0.575	0.537	0.549

Table 6: Impact of Frequency-Recency Decay in Person collection

Method	MAP	Prec@10	NDCG@10
PersonalizedPR	0.11	0.3	0.175
PersonalizedPR+Weinbull	0.201	0.25	0.198
PersonalizedPR+freqWeinbull	0.423	0.511	0.4
PersonalizedPR+MCMC-1	0.132	0.243	0.2
PersonalizedPR+freqMCMC-1	0.26	0.391	0.42

Influence of Propagation

In this experiment, we compare the document prior with and without the propagation step. From the first experiment, we choose the decay functions with frequency-recency versions, as they are shown to perform better in both cases. For Collaboration collection, we chose MCMC-1, and for Person collection, we chose Weinbull.

The results for the Collaboration is shown in Table 7. To our surprise, allowing propagation in estimating document prior actually harm the overall retrieval performance. Analyzing deeper, we think one cause is that the propagation of temporal document assumes the

Table 7: Impact of Propagation in Collaboration collection

Method	MAP	Prec@10	NDCG@10
SEM	0.155	0.225	0.376
TQM	0.319	0.231	0.439
T-Fresh	0.220	0.218	0.357
QLM+freqMCMC-1	0.575	0.537	0.549
QLM+freqMCMC-1+Propagation	0.515	0.21	0.411
SEM+freqMCMC-1	0.443	0.51	0.576
SEM+freqMCMC-1+Propagation	0.397	0.521	0.576
TQM+freqMCMC-1	0.445	0.382	0.667
TQM+freqMCMC-1+Propagation	0.445	0.42	0.59

co-trigger of related documents into the human memory, and therefore give a better estimation of the “relative importance” of a document with regard to the user interest. In the Collaboration dataset, however, one documents are accessed and edited by many people, with different intent and purposes in mind. Therefore, the relationships in document indeed have a wrong signal. What is more, when we remove or change the heuristics used to construct the relationships graph, the results also changed (detailed information is omitted due to space limit), suggesting more advanced mechanism to model document relationship in an organization is needed.

Table 8: Impact of Propagation in PIMO collection

Method	MAP	Prec@10	NDCG@10
PersonalizedPR+freqWeinbull	0.423	0.511	0.4
PersonalizedPR+freqWeinbull+Propagation	0.681	0.79	0.667
T-Fresh	0.173	0.23	0.157

Finally, when looking into the Person collection, propagation significantly improves the retrieval performance (Table 8). This is attributed to the rich availability of semantic relationships in our datasets, which are maintained manually by human over a long time. What is interesting to note is that, T-Fresh is slightly worse than the baseline (Table 6). This can be explained by the fact the T-Fresh algorithm is designed to optimize the impact of hyperlinks between web pages, using temporal random surfer model. In the desktop search scenario, user rarely explore her own computer (unless for reminiscence purpose, which is not frequent), but usually has a clear (but difficult to express) information need in mind. This again emphasizes the importance in designing good algorithms for searching desktop collection, as contrast to normal Web search setting.

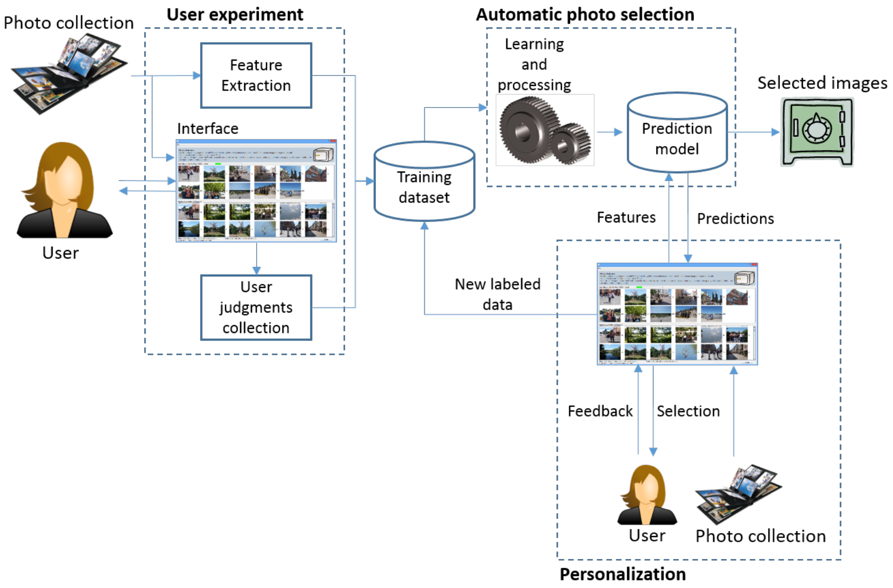


Figure 23: GUI used by participants to select their photos to preserve.

6.2 Adaptive Photo Selection

As anticipated in Section 5.1, different strategies can be investigated for exploiting user feedback and adapting the photo selection model to the user preferences. The overall scenario of photo selection, including personalization, is sketched in Figure 23. The boxes regarding the *User experiment* and the *Automatic photo selection* have been discussed already in Section 3. The online scenario providing user adaptation is plugged in by acquiring a new collection of the user, presenting the selection done according to the current selection model, allowing the user to revise the selection, exploiting the user feedback to update the selection model. The last point is currently subject to further studies. In particular, we aim at investigating (i) how to update the selection model given the new data of the user and all the previously available data, possibly given by other users, and (ii) different granularities of the human-machine interaction. We list issues and envisioned ideas in the rest of this section.

6.2.1 Interaction Granularity

We envision two alternatives to regulate the interaction between the photo selection system and the user. The first is asking the user to revise the selection for a collection only once, and then updating the model for the next collection. The second is allowing the user to do a small number of modifications, updating the selection model, and proposing a new selection for the same collection, until the user is satisfied. These two cases differ from each other in how the human effort in revisiting is distributed. In the former case the user concentrates his/her effort only once, which might help him/her in handling all the modifications that lead to the overall final selection. In the latter case, the user is asked to give feedback more frequently, after a new selection for the same collection is shown. This might represent a difficulty for the user in keeping trace of the changes applied over the selections, but might also bring to an overall lower amount of modifications for the same collections, since the model have the chance of refine the selection different times according to the user feedback.

6.2.2 Personalization Strategies

In an ideal case, where a *sufficient* amount of annotated collections are available for the same user, the selection system could be trained only with them and already achieve personalization with respect to that user. Besides the fact that the sufficient number of collections for being in such situation is not known a priori, a more critical and challenging situation is when there are only few collection being annotated for the same user, and many other collections from other users. We propose different ideas to dial with this *cold-start* scenario.

- **User clustering.** Some users might exhibit similar patterns when performing selections, then a kind of personalization could be achieved by identifying clusters of similar users and learning a selection model for each cluster. The notion of similarity between users is based on the characteristics of their selections, along with those of their original collections: the more users perform similar selections, e.g. they prefer images with faces than high-quality images, the more they are considered as similar. Since such clustering would be based on user selection, it can be done only for users who has already performed at least one selection. All the users available in the dataset are clustered based on their selections, and when a known user imports a new collection then the selection model related to the cluster the user belongs to is used to make the selection.

- **Collection clustering.** Another dimension for the clustering, besides user selections, could be collections themselves: it might come out that some similar collections are handled similarly by users who are overall different each other. This would result in learning a selection model for each cluster of collections, whose similarity is based on the collection features described in Section 3.2, and for any new collection using only the model representing the cluster the collection falls in.

- **Balancing general and personal models.** Considering a given user and all the other user available in the dataset at a given point in time, it would be possible to train a personalized model and a general mode. The former would be trained only on the collection belonging to the given user, while the latter would exploit all the collections of the other users. The question that arises is: how to merge the decisions of these models? And also: how does their mutual influence changes while new collections are introduced in the dataset? Recall that, in a limit case where there are sufficient collections for the given user, it might be sufficient to train the selection model only on them. But in a situation where only few collections (e.g. 2-5 over a total of

50 collections) are available for the current user, one might want to consider also collections and evaluation data coming from other users, since training only on the personal collections might overfit the model. At a given time point, where a total of N evaluated collections are available and k of them belongs to a given user, two distinct selection models would be created: a personalized model, trained on the k collections of the user, and a generalized model, trained on the remaining $N - k$ collections. A first strategy to merge the decisions of the models might depend on the proportion between N and k: if very few collections are available for the user, one might rely more on the generalized model; however, with more and more new evaluations done by the given user, the personalized model might become more and more influential. This and other strategies will explored in the next future.

7 Conclusions

This deliverable describes the extended components in support of the managed forgetting process including *preservation value* and *policy framework*. Particularly, we substantially extend the computational method with several time-decay models and also leverage the access logs of a resource (e.g., the number of views/edits of a document), as well as report the performance of our proposed model evaluated using human assessment. We propose a conceptual model for preservation value, which is composed of its definition and purpose with respect to the two ForgetIT use case scenarios, i.e., personal preservation and organizational preservation. In addition, we present the case studies of preservation value as part of exploratory research, and discuss experimental results and our key findings. In addition, we envision a conceptual model for the policy framework and discuss our activation options for the policies governing the preservation and forgetting process. We propose a computational model for further supporting activities in this task, which include 1) definition and formats of policies 2) tools for defining the policies based on rule-based engines. Furthermore, we present several system prototypes for advanced information value assessment methods and components in support of preservation value, and the integration of an extended set of managed forgetting options into the process. To this end, we outlined planned research activities towards realizing the concept of managed forgetting in the coming months of the project.

References

[1] Drools documentation. chapter 7: Rule language reference. http://docs.jboss.org/drools/release/6.0.1.Final/drools-docs/html/DroolsLanguageReferenceChapter.html. Accessed: 2015-01-02.

[2] Increase business agility through brm systems and soa. https://www.ibm.com/developerworks/library/ar-brmssoa. Published: 2008-05-27, Accessed: 2010-09-30.

[3] J. R. Anderson and L. J. Schooler. Reflections of the environment in memory. *Psychological Science*, 6(2):396–408, 1991.

[4] K. Apostolidis, C. Papagiannopoulou, and V. Mezaris. CERTH at MediaEval 2014 synchronization of multi-user event media task. In *Proc. MediaEval 2014 Workshop*, 2014.

[5] R. Arandjelovic and A. Zisserman. All about vlad. In *CVPR '13*, 2013.

[6] M. J. Boyer and H. Mili. *Agile business rule development: Process, Architecture and JRules examples*. Springer, Berlin; Heidelberg; New York, 2011.

[7] A. Ceroni, M. Fu, V. Solachidis, N. Kanhabua, V. Mezaris, and C. Niederee. Investigating human behaviors in selecting personal photos to preserve memories. In *Under-submission*, 2015.

[8] A. Ceroniand, O. Papadopoulos, V. Solachidis, N. Kanhabua, V. Mezaris, and C. Niederee. To keep or not to keep: An expectation-oriented photo selection method for personal photo collections. In *Under-submission*, 2015.

[9] Y.-W. Chang and C.-J. Lin. Feature ranking using linear svm. In *Proc. of WCCI Causation and Prediction Challenge*, pages 53–64, 2008.

[10] S. Chelaru, E. Herder, K. D. Naini, and P. Siehndel. Recognizing skill networks and their specific communication and connection practices. In *Proceedings of the 25th ACM Conference on Hypertext and Social Media*, HT '14, pages 13–23, 2014.

[11] W.-T. Chu and C.-H. Lin. Automatic selection of representative photo and smart thumbnailing using near-duplicate detection. In *Proc. of MM '08*, 2008.

[12] A. Copeland. The use of personal value estimations to select images for preservation in public library digital community collections. *Future Internet*, 6(2), 2014.

[13] C. Cortes and V. Vapnik. Support-vector networks. *Mach. Learn.*, 20(3):273–297, 1995.

[14] N. Dai and B. D. Davison. Freshness matters: in flowers, food, and web authority. In *SIGIR*, pages 114–121, 2010.

[15] V. Dang and W. B. Croft. Feature selection for document ranking using best first search and coordinate ascent. In *Proc. of SIGIR'10 Workshop on Feature Generation and Selection for Information Retrieval*, 2010.

[16] K. Djafari Naini, R. Kawase, N. Kanhabua, and C. Niederee. Characterizing high-impact features for content retention in social web applications. In *Proceedings of the Companion Publication of the 23rd International Conference on World Wide Web Companion*, WWW Companion '14, pages 559–560, 2014.

[17] ForgetIT. D9.2: Use cases & mock-up development. Deliverable, ForgetIT consortium, M12 2014.

[18] ForgetIT. D9.3: Personal preservation pilot i: Concise preserving personal desktop. Deliverable, ForgetIT consortium, M24 2015.

[19] C. L. Forgy. Expert systems. chapter Rete: A Fast Algorithm for the Many Pattern/Many Object Pattern Match Problem, pages 324–341. Los Alamitos, CA, USA, 1990.

[20] X. Geng, T.-Y. Liu, T. Qin, and H. Li. Feature selection for ranking. In *Proc. of SIGIR'07*, pages 407–414, 2007.

[21] E. Guldogan, J. Kangas, and M. Gabbouj. Personalized representative image selection for shared photo albums. In *Proc. of International Conference on Computer Applications Technology*, 2013.

[22] T. Joachims. Optimizing search engines using clickthrough data. In *Proc. of KDD'02*, pages 133–142, 2002.

[23] N. Kanhabua and W. Nejdl. On the value of temporal anchor texts in wikipedia. In *SIGIR 2014 Workshop on Temporal, Social and Spatially-aware Information Access (TAIA'2014)*, 2014.

[24] N. Kanhabua, T. N. Nguyen, and C. Niederée. What triggers human remembering of events? a large-scale analysis of catalysts for collective memory in wikipedia. In *Proceedings of the Digital Libraries Conference 2014*, DL '14, 2014.

[25] C. Li, A. C. Loui, and T. Chen. Towards aesthetics: A photo quality assessment and photo selection system. In *Proc. of MM '10*, 2010.

[26] J. Li, J. H. Lim, and Q. Tian. Automatic summarization for personal digital photos. In *ICICS-PCM '03*, 2003.

[27] X. Li and W. B. Croft. Time-based language models. In *CIKM*, pages 469–475, 2003.

[28] T.-Y. Liu. Learning to rank for information retrieval. *Foundations and Trends in Information Retrieval*, 3(3):225–331, 2009.

[29] F. Markatopoulou, V. Mezaris, and I. Kompatsiaris. A comparative study on the use of multi-label classification techniques for concept-based video indexing and annotation. In *MultiMedia Modeling*, 2014.

[30] E. Mavridaki and V. Mezaris. No-reference blur assessment in natural images using fourier transform and spatial pyramids. In *Image Processing, 2014. ICIP 2014. Proceedings. 2014 International Conference on*. IEEE, October 2014.

[31] K. D. Naini and I. S. Altingövde. Exploiting result diversification methods for feature selection in learning to rank. In *Proceedings of the 36th European Conference on IR Research*, ECIR '14, pages 455–461, 2014.

[32] K. D. Naini, I. S. Altingövde, Kaweh, R. Kawase, E. Herder, and C. Niederee. Analyzing and predicting privacy settings in the social web. In *Under-submission*, 2015.

[33] R. Neumayer, K. Balog, and K. Nørvåg. On the modeling of entities for ad-hoc entity search in the web of data. In *Advances in Information Retrieval*, pages 133–145. Springer, 2012.

[34] T. N. Nguyen and N. Kanhabua. Leveraging dynamic query subtopics for time-aware search result diversification. In *Proceedings of the 36th European Conference on IR Research*, ECIR '14, pages 222–234, 2014.

[35] P. Over, G. Awad, M. Michel, J. Fiscus, G. Sanders, W. Kraaij, A. F. Smeaton, and G. Quéenot. Trecvid 2013 – an overview of the goals, tasks, data, evaluation mechanisms and metrics. In *TRECVID 2013*, 2013.

[36] C. Papagiannopoulou and V. Mezaris. Concept-based image clustering and summarization of event-related image collections. In *Proc. 1st ACM Workshop on Human Centered Event Understanding from Multimedia (HuEvent'14) at ACM Multimedia (MM'14)*, November 2014.

[37] M. Rabbath, P. Sandhaus, and S. Boll. Automatic creation of photo books from stories in social media. *ACM Trans. Multimedia Comput. Commun. Appl.*, 2011.

[38] L. Sauermann, L. van Elst, and A. Dengel. PIMO – A Framework for Representing Personal Information Models. In T. Pellegrini and S. Schaffert, editors, *I-SEMANTICS Conference 5-7 September 2007, Graz, Austria*, J.UCS, pages 270–277. Know-Center, Austria, 2007.

[39] A. E. Savakis, S. P. Etz, and A. C. P. Loui. Evaluation of image appeal in consumer photography. 2000.

[40] B.-S. Seah, S. S. Bhowmick, and A. Sun. Prism: Concept-preserving social image search results summarization. In *Proc. of SIGIR '14*, 2014.

[41] P. Sinha, S. Mehrotra, and R. Jain. Summarization of personal photologs using multidimensional content and context. In *Proc. of ICMR '11*, 2011.

[42] T. Tran, A. Ceroni, M. Georgescu, K. D. Naini, and M. Fisichella. Wikipevent: Leveraging wikipedia edit history for event detection. In *Proceedings of the 15th International Conference on Web Information Systems Engineering, Part II*, WISE ’14, pages 90–108, 2014.

[43] T. Tran, M. Georgescu, X. Zhu, and N. Kanhabua. Analysing the duration of trending topics in twitter using wikipedia. In *Proceedings of the 2014 ACM Conference on Web Science*, WebSci ’14, pages 251–252, New York, NY, USA, 2014. ACM.

[44] P. Viola and M. J. Jones. Robust real-time face detection. *International journal of computer vision*, 57(2):137–154, 2004.

[45] T. C. Walber, A. Scherp, and S. Staab. Smart photo selection: Interpret gaze as personal interest. In *CHI ’14*, 2014.

[46] M. K. Wolters, E. Niven, and R. H. Logie. The art of deleting snapshots. In *Proc. of CHI EA’14*, 2014.

[47] J. Xiao, X. Zhang, P. Cheatle, Y. Gao, and C. B. Atkins. Mixed-initiative photo collage authoring. In *Proc. of MM ’08*, 2008.

[48] C.-H. Yeh, Y.-C. Ho, B. A. Barsky, and M. Ouhyoung. Personalized photograph ranking and selection system. In *Proc. of MM ’10*, 2010.

[49] W. Zhou, H. Li, Y. Lu, and Q. Tian. Large scale image search with geometric coding. In *Proc. of MM ’11*, 2011.