Nuno Miguel Soares Datia

Mestre em Eng. Informática

Using social semantic knowledge to improve annotations in personal photo collections

Dissertação para obtenção do Grau de Doutor em
Engenharia Informática

Orientador : Professor Doutor João Moura Pires, Prof. auxiliar,
Universidade Nova de Lisboa

FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

October, 2014

Using social semantic knowledge to improve annotations in personal photo collections

Às minhas flores . . .

Acknowledgements

Disclaimer: Although this document is written in English, I feel my acknowledgements should be done in my mother tongue. Sorry for those who can't read Portuguese.

Ao longo destes largos anos, encontrei pessoas fantásticas que me ajudaram no caminho até aqui. Cada um à sua maneira deu um contributo importante para o trabalho descrito neste documento. Nos altos e baixos, normais num processo de doutoramento, conjugados com as partidas e vicissitudes que a vida nos reserva, é com alguma emoção que cheguei ao fim de uma etapa. Espero não me esquecer de ninguém. Se o fizer, é o calor do momento e a adrenalina de entregar a dissertação que me toldaram a mente e baralharam as ideias.

Em primeiro lugar, gostaria de agradecer às instituições que tornaram este trabalho possível. À FCT, que me acolheu como aluno, e que ao abrigo de um protocolo com a minha instituição mãe, me deu as condições necessárias para terminar este doutoramento. Ao Instituto Politécnico de Lisboa (IPL) e Instituto Superior de Engenharia de Lisboa (ISEL) pelo apoio concedido pela bolsa SPRH/PROTEC/67580/2010, que apoiou parcialmente este trabalho. À Área Departamental de Engenharia de Electrónica e Telecomunicações e de Computadores, que fizeram uma cuidada gestão dos recursos humanos para que as dispensas de serviço docente fossem efectivamente praticáveis.

Embora seja da praxe deixar o agradecimento ao orientador, não escrevo palavras de cerimónia por obrigação, mas por sentir genuinamente uma enorme estima pelo Prof. Doutor João Moura Pires. Trabalhamos juntos à mais de uma década e neste tempo fui-me habituando ao rigor que ele impõem, principalmente na formalização dos problemas. Nos momentos menos bons, encontrei nele alguma da força necessário para continuar. É alguém com que se pode falar, debater, resultando sempre num frutuoso contributo. Por tudo, obrigado!

Aos restantes membros da minha Comissão de Avaliação de Tese (CAT), Prof. Doutor

Nuno Correia e Prof. Doutor Nuno Silva, agradeço todos os comentários, críticas, sugestões e caminhos apontados, que muito me ajudaram a melhorar este trabalho. Foi um previlégio privar convosco.

Ao Miguel, Zé, Rita, Luís, Sílvia, André, Bárbara, Filipa, Flávio, Maria, Pedro, Rui, Sérgio, Vítor, Manuel, João, Ricardo, David, Bruno, Sofia, Vítor, Luís, Sofia, Alexandre, Lara, Rui, Cátia, Nuno, Catarina, Sara, e Matilde o meu obrigado pelo tempo que despenderam a realizar os testes dos meus algoritmos. Foram testes que demoram a ser pensados e serem implementados para que tudo estivesse a 110% quando dessem o vosso importante contributo para validar o meu trabalho.

Deixo também aqui uma palavra de apreço aos meus colegas do ISEL, pelos bons tempos que passámos juntos. Embora estes anos tenham sido pautados por um maior afastamento, natural quando muitos estamos a desenvolver os trabalhos de doutoramento, é aí que se sente a importância da camaradagem que sempre existiu entre nós. Em particular, gostaria de agradecer ao Helder Pita o apoio que me deu durante estes anos, para que fosse possível chegar onde cheguei.

Por fim fica aqui o agradecimento à Matilde, embora quaisquer palavras que possa escrever fiquem aquém do seu real valor. Foi, sem dúvida, a pessoa que mais sentiu as minhas omissões durante este tempo, partilhando as minhas angústias e dúvidas, sobre este trabalho e não só. Obrigado, obrigado, obrigado, ...

Resta-me deixar um agradecimento final a todos vós, com um sincero e merecido obrigado.

Abstract

The characteristics of a personal photo collection set challenges in the archival and retrieval that are different from the challenges in general-purpose multimedia collections. The images in personal photo collections show large variability in the depicted items and have hidden semantics. Such features make it hard to find a fully automated solution to the archival and retrieval, that deals with sensory and semantic gaps. Since emotions and non-visual contextual information can be very important to address those problems, including the user in-the-loop is relevant. Thus, manual annotations are key, although their time-consuming nature may alienate users from doing them.

The approach followed in this dissertation uses social semantic knowledge, as a basis to build algorithms for supporting the archival and the retrieval of images from personal photo collections. It borrows from data warehousing the notion of a multidimensional space, capable of answering rare, personalised and previously unseen queries, based on a highly descriptive, social aware, hierarchical set of dimensions. Those dimensions are the "*when*", "*where*", "*who*" and "*what*". The user annotations are used to position photos in the multidimensional space, key to support the retrieval results, adapted to the user interacting with the system. To reduce the manual labour, the system relies on preprocessing the available information, gathered from the metadata and from previously inserted information, to suggest annotations that users will correct or accept. The suggestions are supported by a knowledge base of relevant concepts for a personal domain, stored as an ontology.

Two key algorithms are proposed, along with a prototype. The first algorithm, used during archival, does an automatic segmentation of a set of photos, keeping the spatio-temporal context coherent within segments. A second algorithm, used during retrieval, summarises a set of photos with clustering techniques and short descriptions, relying on hierarchies of textual terms, retrieved from the multidimensional space' dimensions.

The acceptance of the algorithms by the end users shows that using social semantic knowledge, supporting temporal regularities, and using textual human understandable terms to describe the context, are important to build reliable solutions for this domain.

Keywords: Personal Photo Collections; Context Separation; Annotations; Multimedia summarisation; Human factors; Empirical User Study

Resumo

As características de uma colecção de fotografias pessoais apresentam desafios no arquivo e recuperação que são diferentes dos existentes nas colecções multimédia de índole geral. As fotografias pessoais apresentam uma grande variabilidade de conteúdo, muitas vezes com semântica para além do que foi capturado. Tais características tornam difícil encontrar uma solução totalmente automatizada para o arquivo e recuperação, que lide com barreiras sensoriais e semânticas. As anotações manuais são fundamentais para lidar com essas características, embora o trabalho que exigem afaste os utilizadores desta tarefa.

A abordagem seguida nesta dissertação utiliza o conhecimento de carácter social, como base para construir algoritmos para o apoio do arquivo e recuperação de fotografias pessoais. Estes utilizam um espaço multidimensional assente num conjunto de dimensões fortemente descritivas, com informação de cariz social. Essas dimensões são o "quando", "onde", "quem" e "o quê". As anotações dos utilizadores são usadas para posicionar as fotografias nesse espaço multidimensional, elemento chave para suportar um resultado da recuperação adaptado ao utilizador a interagir com o sistema. Para reduzir o trabalho manual de anotação, o sistema pré-processa os metadados e outra informação inserida anteriormente, para sugerir anotações aos utilizadores. As sugestões são suportados por uma base de conhecimento de conceitos relevantes para um domínio pessoal, armazenados como uma ontologia.

O protótipo desenvolvido assenta sobre dois algoritmos chave. O primeiro, usado durante o arquivo, efectua uma segmentação automática de um conjunto de fotografias, mantendo um contexto espaço-temporal coerente dentro dos segmentos. Um segundo algoritmo, usado na recuperação, resume um conjunto de fotografias através de agrupamento e descrições concisas, utilizando hierarquias de termos textuais, retirados das dimensões do espaço multidimensional.

Os testes de utilizadores demonstraram a viabilidade dos algoritmos, indicando que o uso de conhecimento social semântico descrito em termos textuais simples e apoiado em regularidades temporais são importantes para construir soluções para este domínio.

Palavras-chave: Colecções de fotografias pessoais; Separação de contexto; Anotações; Sumarização multimedia; Factores humanos; Teste com utilizadores

Contents

List of Figures

List of Tables

List of algorithms

1

Introduction

> *The beginning is the most important part of the work.*
> ***Plato***

In this new *Digital Era* [Ros03], "capturing the past" it is easy, but making it useful in times to come seems more difficult to achieve. Taking a photo is now a common activity, driven by technological evolutions in digital devices. Recent innovations have transformed the common mobile phones into small computers with multimedia capture and production features, which are becoming ubiquitous in modern societies [KSFS05; Gan08]. Moreover, the technological evolutions in storage devices make larger storage devices available to ordinary people for less money [Whi11]. These factors contribute to the steady increase of digital photos for personal and social consumption. Millions of new photos are taken everyday, some of which are uploaded to online sites, like Flickr[1] [LKF10]. Recent research about digital photos for personal use [LKF10; WBC10], found that only few of them are deleted, even if they look similar or depict different angles for the some situation. This raises the question why some photos, of somewhat dubious quality, are kept? Researchers hypothesise that "*(...) there is an overlap between social and individual intentions, as people want to first share and then review the pictures later on themselves.*" [LKF10]. However, the main reason seems to be related with the important role photos have in our life. They are personal and collective memories of a community sharing a social bond [Zer96; Gye07; Hou09; KS10]. Maintaining those memories is a collective effort, done by several members of the different social groups we belong to. How often have you gone with some friends on holidays and each gather hundred of photos they want to share with the group? This simple action illustrates the problem of

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

managing a personal photo collection:

1. it includes photos from different photographs (*different social contexts*);
2. the photos came from different cameras belonging to different people (*many producers*);
3. many people are interested on those photos (*many consumers*);
4. the union of all the collections produce large collections (*dimensionality problems*).

Humans have limitations dealing with large volumes of information [Bux01], and their senses are also prone to saturate when over stimulated [Mil56]. These limitations contrast with the growing size of personal collections of photos [Gan08; HSW12].

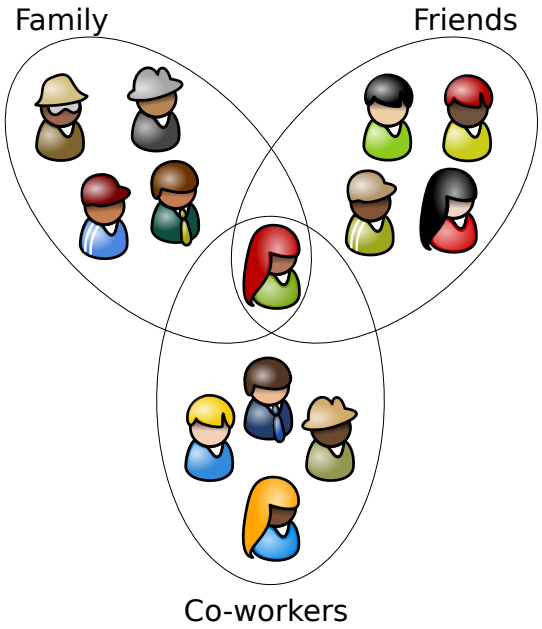


Figure 1.1: Example of social groups, from a self-centric point of view..

Figure 1.1 illustrates some of the social groups people can participate in. In each, one can be an active agent that contributes with photos. Within each our motivations can be different, as they depend on the context. When taking photos of a family meeting, the purpose is mainly to preserve a collective memory. Those moments are more private, and the photos will probably be shared only with the family members. On the other hand, when taking photos at a friends' party, the purpose of preserving a memory is also present, but the strongest motivation is probably self affirmation — to show that one is part of that group [Tom10]. As such, those photos will probably be shared publicly. Even in the same social group, there are photos that are more important than others, depending on the observer. For example, a photo with an old cousin may not have the same importance to young people, as to an older person. This is because photos are artefacts that tie each individual to the group [Gye07; Whi11], and this connection depends on

past memories and experiences. Therefore, it seems natural that our behaviour is conditioned by the group [Tri89], and thus, semantics depend on the self, on its relations with other members, and on the social context of the group itself.

The life cycle of photos, namely, *(i)* capture, *(ii)* archive, *(iii)* sharing, *(iv)* search, and *(v)* artefact generation [Har05; KSRW06], reflects their importance in maintaining a collective memory. The *capture* phase depends on the intention of taking a photo and the availability of a camera. The intentions may vary depending on the type of camera used. Taking photos using DSLRs [2] indicate some premeditation for taking photos, while the intentions are more diverse for pocket camera shots [LKF10], as they encourage frequent and spontaneous photos [VHDAFV05]. In the article "*The Ubiquitous Camera*", Kindeberg and colleagues organised the capture intentions along two dimensions. The first describe whether the subjects captured the photos for *affective* or *functional* reasons. The second outlines *social* versus *individual* intentions [KSFS05]. Among the intentions, affective reasons are predominant in both situations [LKF10]. This dominant intention relates to the need to create a chronology of photos for memory, identity, and narrative [VH07]. The *archive* of photos is essential to preserve those intentions and to better support later actions in the collection [KSRW06]. In this activity, users can enhance their value, by inserting personal annotations [KS05]. Recent evidence ([Hou09; WBC10]) confirms that as times passes by locating specific photos in the collection becomes harder. This just confirms what was already known — the passage of time affects the memory [Wag86; RW96]. From the many possible causes for this effect, we chose to present two: the problem locating events in time (*Forward and backwards telescoping*) [JS05] and retrospective memory problems (e.g. forgot someone' name) [ESB92]. The size of collections also difficult that task [WBC10]. Current solutions do not properly use the personal and social features to organise personal collections of photos [Whi11; HSW12]. As pointed out by Sarvas and Turpeinen [ST06], the characteristics of a personal photo collection and the type of end-user poses challenges in the archiving and retrieval that are different from the general-purpose collections. First, we must understand that the most valuable metadata is personal, and therefore full of semantics. Given its nature, it is also highly subjective and thus, its value depends on the user interacting with the collection [BBMN03]. The metadata of photos is becoming richer, but it is not personal or needs to be transformed to become valuable [DKGS04]. For example, the location where the photos were taken is represented by a pair of coordinates — their latitude and longitude. Those coordinates are valuable to the machine as is, but their value to the users is unveiled when they are represented as the name of the country, the name of the city, the name of the place, to name a few. The bottom line is that it is very difficult to automate semantic annotations in personal photo collections, so the users must take an active part in the process.

[2] Digital single-lens reflex camera

1.1 The problem

For photo collections in general, and for personal ones in particular, one of the biggest challenges is to address the semantic gap problem [SWSGJ00]. The semantic gap exists because the semantics of an object depends on the context that surrounds it. This means that any formal representation of the real world, in this case a photo, needs the translation of the contextual knowledge, which is high-level, into a series of low-level elementary grains that are at machine level. That translation is difficult to automate. First, there are some concepts with few, or no visual clues. For example, a honeymoon photo is in every aspects equal to other photo taken on vacations. The semantics depends on the specific time in the couple's life and not on the depicted items. Second, there is no direct connection between high-level concepts needed for the context reconstruction and the low-level features present in the photos [SCS01; KR08]. To some extent, the intention of the photo, and thus, its meaning, is only available to the photographer [LBB06]. In the last decade, several researchers stated how important the personal information is for the retrieval of photos, particularly for reducing the semantic gap [SWSGJ00; DJLW08].

Several studies have revealed the way the collections are structured is relevant to the users. If the organisation follows the users' labelling, most of the pictures can be found [KSRW06; DJLW08; WBC10]. For example, folder names using event names, date and location. However, if some of the information is absent, the users are unable to find a photo they know that exists somewhere [WBC10]. Nevertheless, users seem to resist to annotate their own collection, although they understand its importance. This is because the effort of keeping collections annotated does not compensate the gains [RW03]. As to the modern image organiser software (e.g. Picasa [3]), there seems to exist a direct relation between the size of the collections and the annotation effort. As the collections grow, the effort to keep photos annotated also grows. Due to this relation, it is difficult to convince users to annotate, even though they understand its value for long term retrieval [WBC10]. Nevertheless, recent studies [GH06; KBS07; CLP07; NY10] found the "social factor" can drive people to annotate. The description, characterisation and access to the photos' context in terms that both humans and machines can handle, enables the construction of a shared ground-truth that allows for a rich-full interaction between the two [Dey01]. Since the semantics emerges, to some extent, from the interaction between users and images [SGJ01], it seems wise to properly handle the context and the way it is described. The bottom line is that solutions that handle personal photo collections need to include user-driven methods [DJLW08; DMBDMJDM10].

Let $P = \{p_1, \ldots, p_n\}$ be a set of photos representing personal memories. Let $\mathcal{M} : p_i \rightarrow \{when, where, who, what\}$ map each photo to its context. In literature, we often see the terms "*who*", "*what*", "*when*" and "*where*" as desirable features to describe the context, named the 4*Ws*. The user annotations will help to map photos into the 4*Ws*. Such annotations should be semantically understandable by the machines, so they can act on behalf

[3] http://picasa.google.com/

of the users, enhancing the potential of the annotations. Namely, the solution should adapt to the users, based on the current state of knowledge about the context, the action performed, and the size of the set being manipulated. Despite the research made on the last decade, towards the mitigation of the semantic gap in personal collections, there are still problems that need to be addressed [DJLW08; JS10]. Specially the use of social context to better integrate personal factors into the solutions [SB11].

The research problem addressed in this dissertation, is stated as

> *How can we improve annotation in personal photo collections, guaranteeing an effective usage of that information on behalf of the users?*

In detail, this work seeks to address the following problems:

1. Annotating is time consuming. How can it be tackled?
2. How can personal annotations be used to adapt the results to the user interacting with the system?
3. How to deal with the (usually) large number of photos resulting from a query, to minimize the *searching+browsing* pattern?

1.2 Proposed approach

The work discussed in this dissertation uses a *social context* approach, triggered by a simple idea. Photographs are artefacts that, among others, are tied to past memories. Those memories exists in a given context, that we try to model by a multi-dimensional space, with social enhanced axis. It is named *multidimensional-context space*, denoted by $\mathcal{MCS}$. One of the most significant features of that space is the ability to consolidate and view the photos according to multiple dimensions. This type of model, used in *data warehouse* systems [CD97], enables straightforward and intuitive manipulation of data by users, if the description of each dimension is focused on their needs. To address them, we use attributes at different levels of detail to describe concepts like *when*, *where*, *what* and *who*. This approach tries to increase the information shared by humans and computers and thus, enables a rich-full interaction between them [Dey01]. To characterise the social context, we use descriptors for: *(i)* Family relations, *(ii)* Social relations, *(iii)* Temporal cycles, *(iv)* Spatial concepts, and *(v)* Cultural life scripts [BR04].

To delimit the scope of this approach, we will describe the target users, the assumptions made during the conception of the idea, and the way photos can be shared among users. We focused on personal photo collections, and in this domain, the number of users that archive and retrieve photos is small. They are family members and close friends, to name a few, with strong social bonds. Therefore, they share the semantics behind the context of the photos. Such constraints have some important impacts on the design of

the system, and set it apart from other photo sharing applications (e.g. Flickr). Thus, the term multi-user can be defined as *a set of users where each one has a direct social bond with, at least, another user in the set*.

In this work, we argue that *sharing* should embrace more than photos and include contexts and semantics. Events and shared knowledge can be fine-tuned and enhanced, using each user's perspective of the context. These actions enable a collaborative construction of a shared memory, as each member of a social group can contribute to a better context definition. However, this cannot be considered crowd-sourcing, as only few people participate in that sharing. Despite the fact that the annotation effort is more intensive during archiving, it is only natural that users wish to enhance the accuracy of the context, when they interact with the collection. This means the contexts stored in the $\mathcal{MCS}$ can change, as users introduce more information in subsequent actions (e.g. retrieval). Figure 1.2 is a representation of the approach, depicting its most important parts and

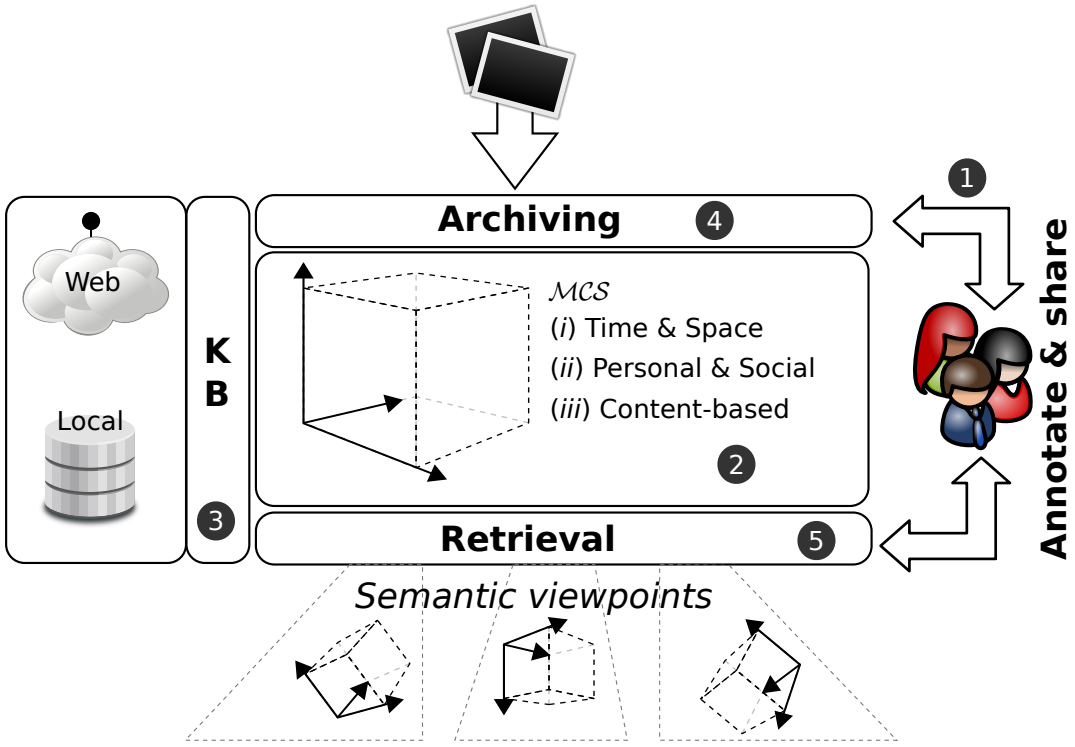


Figure 1.2: Overview of the approach. The dimensions of analysis contains descriptors for time, space, social features and content-based..

the way they are interconnected. It follows some of the processes described in [Har05], namely, the *archiving*, *annotation* and *querying*. The approach is named **MeMoT** [4], an acronym for **Me**mory, **Mo**ments and **T**eam work. The *human-in-the-loop* is a key part of the approach, identified by ❶. The annotations provided by users are key, and play an important role in the archival and retrieval processes. They are used to better position

[4] A prototype implements this approach is available at http://purl.org/mont/memot.

the photos into the $\mathcal{MCS}$, labelled as ❷. The interaction with the users is supported by a knowledge base, labelled as ❸, establishing a common ground of understanding between the users and the machine, in terms of semantics. The archival of photos is supported by an algorithm that enables the re-usage of annotations, by automatically separating the photos into groups of similar context (❹). The retrieval, labelled as ❺, benefits from a proper positioning of the photos in the $\mathcal{MCS}$. It uses its descriptive power and the information in the knowledge base to adapt the results to the user interacting with the system and to support short, self descriptive, summaries of sets of photos.

In conclusion, the key to address the research problem is to properly handle context, using terms that both people and machine can understand. Such terms should present different levels of detail, capable to adapt to different sets of photos, embedding the inherent human common impression. The separation of the context in the archival and the summaries during retrieval, sets the ground for a better manipulation of the personal photo collection. In the next subsections, each part of the approach is described in more detail.

1.2.1 Multidimensional context-space

The multidimensional nature of $\mathcal{MCS}$ is the key to a rich retrieval, because of the combinational power of the information in each dimensions. Each dimension in $\mathcal{MCS}$ is independent of the others. Thus, the retrieval performance is closely related to the expressiveness of the dimensions that will support the manipulation of such space. Populating each one with relevant information for a specific user, requires domain information and user intervention. We group the information available in the $\mathcal{MCS}$, as

1. time and space,
2. social/personal, and
3. content-based.

Each group is enhanced with common sense metadata that we use everyday to refer to events that happen in our lives. Whenever possible, that common sense is represented in different levels of granularity so as to deliver the most acceptable one in response to a query. There also exist hierarchic relations between levels of detail. A detailed description of the representational needs is presented in Section 3. For now, we present a brief overview.

1.2.1.1 Time and Space

Together, time and space give information about the "*where*" and "*when*" cues. Episodic memory needs both [Tul02; Fri04; Has09] to enable a recall of specific events. As Kellerman [Kel89] said, "*Time is a major dimension along which all events occur and around which human life cycles evolve*". We cannot dissociate time and location, as the correct perception

of time depends on the location. Terms like *Summer* or *Dawn* may derive from time information if it refers to a specific spatial location. People describe and understand time using a set of cycles that rule their lives: the natural cycles of days and years, but also the rhythm of the weekly cycle, with more social meaning than the formers [Zer85]. This view of time drives the terms that are relevant in the multidimensional context-space. For example, words like *afternoon*, *first weekend* and *April*, or just *Spring* and *sundays* are possible descriptors for the timestamp "2011–04–03T16:20:30.45+00:00". Table 1.1 shows the most relevant terms and the detail level of each one for time information. The location associated to a photo is expressed using latitude and longitude coordinates (spatial coordinates). Users, on the other hand, refer to a location using concrete names, like countries' names. $\mathcal{MCS}$ supports a set of spatial descriptors, namely:

Terms	**Cycle**	**Level of detail**
Day, Night, Afternoon, Morning, Dawn, Dusk, Twilight, Midnight, Noon	Day	Detailed
Weekend, Workday, Days of week	Week	Moderate
Months, Quarters, Seasons	Year	Summarized

Table 1.1: Example of the terms used to describe time in the $\mathcal{MCS}$ up to the year level.

1. the continent,
2. the country,
3. the administrative levels,
4. the city,
5. the area,
6. the point of interest.

They form a variable depth hierarchy that is necessary to deal with missing information and to accommodate different administrative levels that exist in each country.

1.2.1.2 Social information

The social and textual information comprises another dimension of the $\mathcal{MCS}$. It represents the high-level semantic metadata in the context, and contributes to the "*who*" and "*what*" cues. The most valuable information results from the interaction between users and the system. For example, identifying the people depicted in some images or entering the activity for a set of photos. Most of the time it is non-visual, user-dependant information that is important to complete the context. This information not only benefits the

retrieval but transforms an impersonal allocentric context into a user-centered one. The social descriptors can be enhanced automatically with calendar information containing holiday events (either civil, religious or school vacations), social events (like music performances or fireworks shows). The calendars can also include personal and work data. Gathering personal calendars, holiday events, and social events from web services was not addressed in this dissertation.

1.2.1.3 Content-based information

Content-based manipulation can be important if we want to increase the value of each photo, by automatic means. Although the high variability existing in personal collections defeats many of the content-based algorithms, if the goal is to get high level semantic information, there is relevant information that can be derived from content. In particular, if it is to be used in conjunction with other contextual information. For example, detecting faces [VJ01] or identifying the location[5] of the scene depicted [FPZ03; JAC11], can be used to improve the suggestion of annotations, supporting users on the contribution to the "*who*" and "*what*" cues. It also enhances the capability of $\mathcal{MCS}$ to provide photos sharing similar features, for example, depicting people. This may be of special interest, since users are interested in photos that have faces [SB09].

1.2.2 Knowledge base

The approach relies on a knowledge base (KB) for keeping the concepts of the domain, their relations, and users' assertions. It tries to cover most of the terms used in the $\mathcal{MCS}$, specially, the spatio-temporal terms, the social relations between people, and other relevant social terms. There is a lot of ground knowledge that can be captured to provide an extensive collection of concepts specific for a given culture. For example, in western societies there is a set of holidays, pre-defined vacation periods, and typical family arrangements, just to name a few. Although sometimes we are unaware of them, our social rhythms are dictated by those pre-established cultural "rules". The KB enables inference and reasoning, thus producing new knowledge, to be used either in archiving or in retrieval. For example, with the introduction of the *Son Of* assertion between John and Mary, we can infer that Bob is John's brother, if the KB has an assertion telling Bob is the son of Mary. Since the personal and social contexts change over time, the KB must be able to deal with them. Changes in context should not override previous assertions, as we want to maintain a personal and social memory. For example, our close friends may vary over time. But it should be possible to reason about photos containing the friends we had back in high school, even if we are not friends anymore. Nevertheless, the system is also autonomous to perform changes or updates based on new information coming from the outside world. Chapter 3 (Knowledge Representation) describes the knowledge

[5] By location we mean `Inside/Outside`.

representation in detail.

1.2.3 Archiving and metadata enhancement

When people take photos, they leave in the support medium a set of pictures that spans through different time periods, depicting several personal and social events, but also photos related to the ordinary day life. Sometimes that set of assorted photos is referred to as a *roll*. The archiving's main goal is to prepare rolls for later retrieval. This is achieved by positioning each photo in $\mathcal{MCS}$. Thus, the available information needs to be worked to meet the semantics we want to deliver to the users. Most of the time, the available information is the photo's metadata, which can vary with the camera and prior software processing tasks. It is assumed that temporal information is always present[6], as it is a common denominator [Gro10] for the available metadata standards [IPT10; JC10; Ado12]. It is assumed that spatial information is available, most of the time. Since the usage of smartphones is rising [DBGP11], and their location capabilities are becoming more precise [WM10], it is plausible to assume that this information will be ubiquitous in a near future. To help the insertion (or correction) of information, the set of photos are segmented using spatio-temporal regularities. We pay special attention to the daily cycle, using variable day limits to suit the photographer's shooting demands, producing the definition of *Logical Day*. Chapter 4 (Archival) describes the segmentation in detail.
After the segmentation, the system uses algorithms, web services and prior knowledge to improve the context of each photo (or group of photos), namely:

- uses known calendar algorithms [RD01] to settle time descriptors;
- uses web services, like the Geonames API[7], to settle human readable descriptors for spatial locations;
- uses web services, like Eventful API[8], to settle local public events information;
- uses content-based algorithms [VJ01] to derive features from photos.

Notice that a user can change the information and the segmentation proposed, whenever he thinks it is inappropriate to describe the context of certain photos.

1.2.4 Retrieval with multiple perspectives

The **MeMoT** retrieval power stems from the multidimensional nature of the context-space, as it enables the combinational power of the information present in the dimensions. It is specially tailored to deal with unpredictable queries, with requirements of

[6] Digitised photos may lack a correct temporal information, but they are out of the scope of this work.
[7] http://www.geonames.org/export/ws-overview.html
[8] http://api.evdb.com/

different grains. Despite its uncertain nature, we argue that users will only retrieve photos they know they have in their collections. For instance, a user won't search photos about the Brazilian Carnival if he knows he never went to Brazil.

Users want to query their collections of photos in the semantic level [HLMS08] and their interaction exhibits multiple semantics depending on the context [SBC07]. Namely, the semantics depend on the user who performs the retrieval and thus, that interprets the results returned from a query [SGJ01], constrained by his personal and social context. We call each user's own perspective its **semantic viewpoint**. These viewpoints originate from different references for some of the terms in the $\mathcal{MCS}$. So, it is necessary to perform the transformations from the multidimensional context-space towards the viewpoint, to present a relevant set of photos in response to a query. Many of the transformations occur at the spatio-temporal dimension. They are necessary whenever a reference to time (or time and space) in the query can be perceived differently depending on the photo's context. For example, the description of seasons for the same timestamp differs in the two hemispheres. The transformations are also performed in the social level. The occurrence in a query of relative references like *father* or *sister* can be mapped to several identities. Thus, the transformation towards the new viewpoint must deal with social and personal relationships to define, in query time, the proper terms to address the $\mathcal{MCS}$. Let us use an example. Assume that a user (Bob's brother) issues the query "*photos from Bob's last summer vacation, where my father is present*". In the spatio-temporal dimension, the query will include all the photos taken in the last summer vacation time-frame, according to the user's usual location. The term *father* can be transformed into multiple semantics terms, that depends on the different social relations. In this case it is transformed into the name of Bob's father.

To deal with the unpredictable size of the results set, photos are summarised in such a way they maintain the most important cues to the context comprehension. They are aggregated by context similarity, where each group has a proper short description, and there is a global selected detail for each dimension. Based on the summary, a user can judge the adequacy of the result without traversing it, as he has the notion of the underlying context of the set. This approach follows the principles stated by Shneiderman [Shn96] — *Overview first, zoom and filter, details on demand*. The overview is the aggregated data with a proper description for each group. The selected detail can be used to filter the groups for each available dimension, and is the starting point to drill up and down the context's details. In Chapter 5 there is a detailed explanation of the retrieval approach.

1.3 Contributions

The list of contributions to the current body of knowledge is the following.

One of the major contributions is the approach taken, focusing the problem in the context surrounding the photos, rather than the photos themselves. To handle the context, we propose a multidimensional modelling, with many dimensions of analysis, described simultaneously at different levels of detail. The terms that describe the dimensions are related to the people's activities, with a controlled vocabulary stored in ontology, named Memory Ontology [9]. The ontology reuses some concepts from other ontologies, but incorporates the imprecision people commonly use in their life, from naming places, to position events in time. Together, the multidimensional approach and the ontology provides a description of the context using a set of terms that both humans and machines can handle, supporting the algorithms and the processes needed in personal photo collections handling.

This work presents a theoretical formalisation of a segmentation problem. Such formalisation can be used in any field of knowledge where the objects to segment have a temporal order. Other contribution that follows this formalisation is the proposal of four binary relations between segmentations, addressing a lack in the literature for a theoretical support for qualitative comparisons.

A new segmentation algorithm is proposed, named Logical Day Event Segmentation algorithm, that is used to automatically detect events in photo sets. It is used to aid users during the annotation process in the archival. One of the most important aspects of this algorithm is the usage of the *logical day* concept, modelling the people's daily cycle of activities.

Another new algorithm, named Multimedia Short Summary, is presented. It is used during retrieval for summarising a set of photos, maintaining the most important context information with few summary groups. It also selects a proper level of detail for each dimension in the 4*Ws*, to describe each group in the summary.

The approach to the problem is demonstrated in a prototype [10] that implements the archival and retrieval of personal photos.

1.4 Document outline

The dissertation has been organised in the following way. The first chapter introduces the problem, overviews the solution and identifies the main contributions. Chapter 2 describes relevant related work, thematically organised, discussing their solutions and comparing them with the proposed approach. Chapter 3 describes what kind of information and knowledge needs to be in the system, to address the identified problems.

[9] Available at `http://purl.org/mont/mont.owl`.
[10] Available at `http://purl.org/mont/memot`.

It also presents the representation of some key concepts, establishing links to other representations in the knowledge representation field. Chapter 4 deals with the archival and annotation of rolls. It presents the *Logical Day Event Segmentation* (LDES) algorithm, that follows the problem formalisation and the segmentation comparison framework. Chapter 5 addresses the retrieval of photos, describing *Multimedia Short Summary* (MSS) algorithm for photo sets. It provides a theoretical formalisation of the problem and relevant use cases for the retrieval of photos in a multi-user scenario. Chapter 6 presents and analyses the results of the experiments made. It includes theoretical evaluations of the algorithms and results from the user's tests accomplished. Finally, Chapter 7 gives a summary of the findings, drawing some conclusions from the approach and pointing out topics for further research.

2

State-of-the-art

> *The measure of greatness in a scientific idea is the extent to which it stimulates thought and opens up new lines of research.*
>
> ***Paul Dirac***

This chapter describes some of the relevant research made in the last decade, related to photo collections. The organisation highlights some major issues, that are key in this field, giving context to our contributions.

2.1 Metadata

In the past decade, several organisations, mainly from the industry, set the standards on how to communicate metadata in digital photos. Metadata is stored inside the photo file (e.g. TIFF), using container formats specific to each standard. One of the most widely used in digital cameras is the EXIF [JC10]. The initial specification was made by JEITA[1], but the latest version was co-authored by CIPA[2]. The information format follows the TIFF specification, and contains a set of tags describing several features, like *DateTimeOriginal* and *Orientation*. The EXIF metadata is organised into different *Image File Directories* (IFD's), each one with a different purpose:

- *Exif IFD* — contains a set of tags for recording exif-specific attribute information;
- *GPS IFD* — records information about spatial localisation;

[1] http://www.jeita.or.jp/english/
[2] http://www.cipa.jp/english/

- *Interoperability IFD* — stores the information to ensure the interoperability between consumers.

The life cycle of digital media, especially photos, includes more players than just camera and end-users. The process flow starts with the *Creator actor* (generally a camera), goes through *Changer actors* (e.g. editing software) and ends at the *Consumer actor* (e.g. generally the end-user) [Gro10]. The EXIF standard was designed as an interchangeable format, and lacks the features to support the many perspectives that exist in the professional photo workflow. Those features are addressed by the International Press Telecommunications Council (IPTC), that created the Information Interchange Model (IIM) [IPT99], superseded by IPTC Core and Extension specifications. Together they define the IPTC Photo Metadata standard [IPT10]. The IPTC properties can be grouped as:

1. *Descriptive Metadata* — Describes the content of the photo (e.g. description/caption);
2. *Administrative Metadata* — Data about the content that cannot be retrieved or inferred from the picture (e.g. date created);
3. *Rights Metadata* — Information about ownership (e.g. creator).

Descriptive metadata, in particular, includes information that can be derived from other data, but shows the point of view of the creator. It includes a `location` property, that can be defined by hierarchical terms representing the world region, country, province/state, and city or any location outside a city. But also enables the insertion of a written description of "*who*", "*what*", "*when*", and "*why*", in the `description/caption` and `person` properties. IPTC uses namespaces to give semantics to their properties, namely using the Dublin Core [DCM09]. Dublin Core Metadata Initiative (DCMI) defines a set of metadata terms, that enables a controlled vocabulary, towards a better interoperability for metadata in the semantic web. Their *Metadata Element Set* includes, among others, the `date`, `creator`, and `subject`.

In 2001, Adobe Systems introduced the *Extensible Metadata Platform* (XMP), as a labelling technology especially designed to be used in digital assets. Later, in 2012, it become an ISO standard [Ado12]. XMP standardises the definition, creation and processing of extensible metadata [Gro10]. IPTC has adopted XMP as the successor of the IIM standard, resulting in the IPTC Photo Metadata standard (IPTC Core and Extension), where XMP is used to embed IPTC metadata inside the files. XMP is serialized in XML and uses a subset of the W3C Resource Description Framework (RDF) [W3C04]. Therefore, it is extensible, as each actor can define custom properties and namespaces to embed into the files.

Also in 2001, the Moving Picture Experts Group publish the *multimedia content description interface* standard, known as MPEG-7 [MKP02]. MPEG-7 is organised in descriptors,

that define the syntax and the semantics of a feature representation. There is also the *Description Scheme*, that defines the semantics of the relation between descriptors. The descriptor can be used to hold semantic information about concepts captured in media, for example, events and identities. However, MPEG-7 is not widely used in the industry, probably because of its complexity [Lux09]. On the contrary, EXIF, IPTC IIM and XMP formats are widely used in the industry [Gro10]. They can be found in TIFF, JPEG, and RAW[3] file formats, among others. These file formats cover most of the cameras available, including smartphones. Despite the fact that many recent camera models allow for the insertion of metadata describing the 4*Ws*, these features are exploited mainly by professional. Besides, the metadata varies according to the manufacturer and camera models. Building reliable solutions implies using common available data. As such, **MeMoT** built its reliability on top of common denominator metadata: creation time and spatial coordinates. Other metadata, if available, is used to increase the value of the collection.

2.2 Supporting the archival

Rodden and Wood [RW03] found that people organise their digital collections in a similar way as their physical counterparts — inside "shoeboxes". A digital representation of a shoebox is a folder labelled and dated. A latter study confirming this behaviour [WBC10], points out the importance of using a known structure to archive the collections. In a digital context, metadata can be used to automate the archival on behalf of the users, using different strategies. The *PhotoCompas* system [NSPGM04] provides an automatic sorting of personal photo collections, based on time and location. The temporal regularities, namely, the "burst" assumption [GGMPW02; Gar03], are used to settle event boundaries, tuned with the locations of two consecutive photos. Besides the event structure, they also settle location hierarchies, with a predefined fixed depth (2–3), where the lowest levels are created by clustering photos with similar spatial distribution. Both event and spatial hierarchies are used to support browsing, based on an organisation that mimics the way people think about their photos.

Elsweiler et al. [ERJ07] think browsing, as is, suffers from some problems. For example, it is not suitable to finding relatively small data sets, and is done over an organisation that forces a categorisation of photos based on some of their properties. Other approach is to change the focus from the organisation towards the role different types of memory play when retrieving an object from personal collections. The *PhotoMemory* [ERJ07], allows a multimodal interaction, to address three recurrent memory lapses: *(i) retrospective memory problems*, *(ii) prospective memory problems*, and *(iii) action slips*. The archival is more like a curation process where users can place photos in semantic groups and annotate images, when they are added or while browsing the collections. This process does not force any fixed organisation, something that users found inadequate when searching for photos [WBC10].

[3]For most RAW formats, since they follows TIFF' specification

Since semantics is important to users, Chai et al. use ontologies to support annotation, on the *OntoAlbum* system [CZJ08]. They support concepts needed for the management of personal photo collections, including social and event relations. Those concepts support the annotations made by users, but are also used to organise and browse the collections. A recent work from Figueirêdo et al. [FLPCSB12] presents the *PhotoGeo* system, that assists users on the annotation, storage and retrieval of their personal photos. The annotation has two targets: 1. events and 2. people. Events are detected using a combination of spatial and temporal clustering. The user influences the event generation changing the time and space granularities, based on predefined values. He can annotate the resulting event set, providing information for "*when*" and "*where*". Resolving the people's identities is a semi-automatic approach. The system uses contextual information to improve the current face detection recognition, and settles a list of candidates names. Deciding the identities provides information about the "*who*". Besides the event and people, the storage uses other information, namely, metadata available in the files.

The previous works illustrate the research carries out on the last decade. We can see trends changing from automatic [NSPGM04] to semi-automatic approaches [FLPCSB12], taking the user as a source of information [ERJ07; CZJ08; FLPCSB12]. It is also clear that information with more semantics [ERJ07; CZJ08] is more used, and thus, more valuable. Nevertheless, researchers do not emphasise the collaborative nature of constructing a collective memory. The approaches assume the archival as a *one-time-one-person* process. This assumption does not naturally fit the usage people give to their collections. The event structure, once settled does not change, even if new photos are added to the collection. However, an event can have photos from different users, archived in different moments. The event-detection algorithms use temporal and spatial information, but do not include support to time cycles that bound people's actions. The social information is restricted to people's relations, leaving out other common knowledge, namely, regularities that exits in most societies. Those in particular, are important to unveil the "*what*" in the context. **MeMoT** addresses these problems, and assumes a *multi-time-multi-user* approach, where the organisation, although highly structured, supports searching and browsing based on user defined restrictions.

2.3 Supporting the retrieval

The retrieval of photos is largely based on the information available to describe the context of photo collections. The richer the information is, the more efficient the retrieval will be. In [OGJLOS07], the authors index a set of textual descriptors, automatically generated from the EXIF and photo's content. Most of them describe a concept from different perspectives or levels of detail. For example, time-based fields include the terms "year", "month" and "day of week", to name a few. The textual descriptors cover the "*when*", "*where*" and "*who*" cues, enabling a text search retrieval without manual annotation. Using an opposite approach, in terms of flexibility, [CZJ08] allows users to retrieve their

photos guided by known concepts and properties, supported by an ontology. Figueirêdo et al. [FLPCSB12], uses in *PhotoGeo* system a half way solution. The user can define a set of filters, each for a specific type (e.g. Temporal). A filter has an operation that settles the restrictions you can make. For example, when the operation is `Distance`, it is possible to enter "*10Km from city (Lisbon)*". Free text queries are implemented using a conventional filter type, with a SQL `like` operation. [ZPFKV12] proposes a searching scheme for events on an annotated personal collection. Filtering can be based on: *(i)* free text, *(ii)* date, *(iii)* local, *(iv)* owner, *(v)* similarity, and *(vi)* event membership.
Other works like [ERJ07; DWC09; FTHSS10] focused on providing interactive retrieval modes, using a combination of multiple cues and free text. Elsweiler et al. [ERJ07], addresses three common memory lapses, by offering cues to help the user as he searches. The collection can be filtered using temporal, semantic and free-text cues. The collection is always visible on screen, and photos always have the same neighbouring photos, although they are empathised when matching the filters. Dörk et al. [DWC09] argue that visualisation widgets (*VisGets*) are a viable way to query and visualise web data, including image collections. The *temporal VisGet* implements an interactive bar charts. It shows the temporal distribution of photos and allows the formulation of temporal queries, for three fixed grains (year, month and day). The *location VisGet* shows the spatial distribution of photos. Zooming the map puts spatial constrains in the query. The last *VisGet* shows an interactive tag cloud, that filters the collection based on the textual information associated with photos. Fialho et al. [FTHSS10] uses interactive contextual browsing to search for events, although, it is not specific to photos. The user is visually guided through an interactive query refinement process, while visualising the results in different categories. Those categories are centred around the 4*Ws* and the representation events, using the LODE ontology [STH09].

The selected work illustrates two important key points when querying a personal photo collection: *(i)* multi-cue support, and *(ii)* unstructured vs. structured information. Supporting multi-cue enables users to refine their search in any of the 4*Ws*, providing a multitude of ways to express their needs. There are, however, different approaches for multi-cue support. One option is leaving the semantics of the terms in a query open and matching the terms using the textual descriptors attached to the photos [OGJLOS07]. This approach, although effective if the available textual descriptors are rich, defeats any attempt to interpret the query's cues using a-priori knowledge about the user querying the system. Thus, simple transformations from relative to absolute cues are impossible, and raise barriers to derive high-level semantics. Another approach, is restricting the filtering cues [DWC09; FTHSS10], based on the information available in the collection. This approach can provide effective cues to users. However, personal collections exhibit large variance in the 4*Ws*, leading to a large and complex cue sets. Since users have limited capacity to deal with large information sets [Mil56], restricting the filtering is not feasible, at least for the non-continuous information (e.g. location) in large variance sets. Another option is fixing the scope of search to one purpose [ZPFKV12]. The goal of the search is

clear to the users, but make impossible to include semantics in the queries, as the user interface forces predefined search actions.
Regarding the information structure, there are two possible approaches. The first is to use unstructured (or loosely structured) information [OGJLOS07; ERJ07; ZPFKV12; FLPCSB12]. This approach favours the adaptation to the users needs, as while annotating the collection each user can choose the words that will help a later retrieval. However, it limits the use of reasoning services to aid the users. The other approach is restricting the retrieval to include known concepts [CZJ08]. This approach clarifies the semantics, but as it forces a rigid structure makes it difficult for users to express the way they want, including using their own semantics.

MeMoT uses some of the techniques present in [OGJLOS07; ERJ07; DWC09] on top of a structured representation of concepts, similar to [CZJ08]. An ontology is used to suggest concepts to the users during annotation, and later during retrieval, based on free text input. This approach does not force the use of the suggested values, but once accepted by the users, enables the reasoning on their behalf. The query results set can be fine tuned by filtering the 4*Ws*. This filter input is selected among the most representative for the set, based on the semantic and relation between terms. The design follows the simplicity suggested by Hearst [Hea09] and Shneiderman's mantra [Shn96] "*Overview first, zoom and filter, details on demand*".

2.4 Motivating users

As the researchers move towards interactive annotation schemes, to complement photos with more semantic information, they face the difficulty of convincing users to annotate [KS05]. One of the most successful approaches is transforming the annotation into a game [AD04; SI09; JAC11]. The idea is to give a second purpose (annotate) to a joyful action (playing), where each user makes part of a massive computation architecture, in something called crowdsourcing. Another approach is to motivate annotation by showing alternative annotation-based presentation schemes [KS05]. The idea is to make the annotations effective, i.e., resulting on value to the users. For example, to automatically generate a pictographic family tree. The other approach is to make use of social incentives to motivate annotations [AN07]. Although the reasons that motivate annotations are many, they range from organisation to communication purposes. In particular, for public audiences, self-promotion plays a major role. The motivation seems to be related to self-satisfaction, either for self-promotion [AN07], incentive guided [KS05] or to kill time [AD04; SI09; JAC11].

Not all motivation techniques can be used to support annotation in personal photo collections. Crowdsourcing-based solutions [AD04; SI09; JAC11] are not suitable when the target users is restricted and the collection is known. However, raising the benefits of annotation [KS05; AN07], either in organisation or social communication is possible in collaborative scenarios, such as personal collection annotation. **MeMoT** empathise the

benefits of annotating, increasing their value through reasoning and re-usage of prior knowledge. Metaphorically speaking, it acts as an information amplifier, improving the retrieval.

2.5 Reducing the annotation effort

Raising the incentives for annotation is one way to motivate. But lowering the effort when inserting information is another approach, since users already know the importance of annotations [WBC10]. Researchers found they can leverage on past annotations to accelerate the process of manual annotation [NYGMP05; MO07; SSX07]. One [NYGMP05], determines patterns of re-occurrence and co-occurrence of different people in events, based on the context (time and location), using a ground truth of identity annotations. Those patterns enable suggestions for not yet known identities in photos, without using face-detection and face-recognition algorithms. Another [MO07], also suggests "*who*" and "*where*" annotations, based one time and location. However, they extend the source of information to online sources, like web services and social networks. The EXIF metadata is translated to RDF metadata and combined with other sources (e.g. GeoNames ontologies), providing support to identities and place annotations. The RDF is then combined with existing similar metadata to suggest annotations. The other [SSX07], uses recommendation algorithms to provide annotation support in the 4*Ws*. They exploit personal context as well as social network context, based on the assumption that the members of that network have highly correlated real-world activities. The context is modelled as a set of interconnected nodes, based on previous annotations for the 4*Ws*. Another solution for the reduction of the annotation effort, is bulk annotation [CWXTT07; SB07]. The main idea is to group photos that share some characteristics. Cui et al. [CWXTT07] developed an interactive photo annotation system, called EasyAlbum, focused on getting people's identity. The photos that share a similar scene/facial setup form clusters that the user can annotate. Also, the selection of photos forces the re-rank of photos or clusters to be labelled, so the annotation can be reused in more photos. In [SB07] the authors group photos with the same identities, based on the clothes colour, but also on similar activity (event). The other approach is to generated annotations based on predefined concepts [BSST07; VBFGVOM08; JAC11]. The idea is to map the low-level information into high-level concepts. Some mappings are done using binary classification models (support vector machines in [BSST07], and Regularized Least Squares Classifier in [JAC11]). They are based on context and content-based features, and classify photos using concepts like *In-/Outdoor* or *No-/People*. Other mappings are simple transformations of EXIF metadata, e.g. aperture and exposure time into *light condition* [BSST07]. Another approach is to transform time and location into human understandable descriptors, using known algorithms and online sources [VBFGVOM08].

The reduction of the annotation effort can be categorised in *(i)* Manual [CWXTT07;

SB07], *(ii)* Semi-automatic [NYGMP05; CWXTT07; MO07; SSX07; JAC11], and *(iii)* Automatic approaches [BSST07; VBFGVOM08; JAC11]. The first reduces the items to annotate. The second, relies on solutions that can exhibit error, and therefore, need human confirmation. The third is based on known (and correct) information, and its goal is to properly present descriptors for humans. **MeMoT** uses the set of three approaches. The archival uses segmentation techniques to group photos with similar context, and allow users to select many photos to annotate at the same time. The algorithms take into account the temporal cycles that govern our lives [Zer85]. During the insertion of annotations, **MeMoT** suggests contextual terms, derived from reasoning or ranking solutions, that need users approval. Finally, there are many descriptors generated on behalf of the user, covering the "*when*","*where*", "*who*" and "*what*" that enrich the context and do not require user intervention.

2.6 Handling large collections

The personal collections of photos are growing, making it more difficult to manage and find photos. A large and growing body of literature tries to cope with the size by summarising collections, highlighting relevant features for their specific goals. Platt et al. [PCF03] was one of the first to apply clustering for summarisation purposes, thus reducing the number of objects to browse. Clusters are settled based on temporal gaps between photos, although each can be further divided using color features, into small size clusters. Others [GGMPW02; NSPGM04; CFGW05; BPGP10], use the same clustering approach to simplify browsing. Another work [JNTD06], summarises a set of photos by ordering the set and selecting the top ranked photos, providing highly representative textual tags on relevant map locations. The ranking takes into account the photographer and tag information, and increases monotonically the relevance and image quality factors. It is applied to rank clusters found using an hierarchical algorithm based on location, and then to rank each photo. In another work [HDBW05], the authors developed a zoomable photo browser that combines time-ordering with a space-filling layout. Their goal is to support time-based visual searches over unstructured photo collections, using temporal clustering. Each cluster is visualised as a thumbnail, whose size depends on the zoom level and the dimension of the clusters. When thumbnails become too small, the clusters are merged to form larger ones. Columns represent years, and the clusters are placed from top to bottom, maintaining a temporal order from top-down, left-right. Sinha et al. [SPJ09] focused on three types of events (single day events, week events and year events) to provide a summarisation that preserves the information on the entire set and is semantically coherent. The photos are clustered using time, location, face and image features, guided by the event type. Each photo is ranked based on the presence of faces, the shoot rate and image features. The summary is settled using constraints gathered from the user's queries, weighting the "*where*", "*when*" and "*what*", and presenting the higher ranked photos. [OOO10] provides a summarisation procedure to assist the creation of

photo albums, inspired by principles of dramaturgy and cinematography. The authors divides the story into a three level hierarchy of acts, scenes and shots, using a temporal clustering algorithm. Then, based on information from the user' social network (other albums), the summary is settled guaranteeing the face ratio, time range, character relevance and aesthetics of the photos. A recent work by Sinha [Sin11] argues that a personal photo summarisation should have quality in the photos presented, covering important concepts of the collection and not containing redundant information. The photos are categorised in five concepts, *(i)* location, *(ii)* event type, *(iii)* visual, *(iv)* temporal, and *(v)* face. Using them, the authors model the summary as a multi-objective optimization problem, where they want to maximise diversity and coverage of the categories. They aggregate photos towards the optimal combination of values, and report an aesthetically attractive photo.

Previous research uses summaries as a common approach to cope with the size of photo collections. Such summaries are used to support browsing [PCF03; JNTD06; HDBW05; Sin11], searching [SPJ09] or to assist post-actions [OOO10]. Authors try to maintain the summary's coherence in some dimension, like time [PCF03; HDBW05; SPJ09] or space [JNTD06].**MeMoT** uses a clustering algorithm to produce summaries that support searching for a specific set. However, unlike [SPJ09], the goal is to maintain a global overview for the 4*Ws* of the set. The coherence can be one of the 4*Ws*. Most of the previous research, with the exception of [JNTD06], chooses a photo to represent a cluster, probably inspired by the adage "*A picture is worth a thousand words*". However, despite its value, there is no guarantee that a single photo is a valid representative of an event, or that it triggers in the user's memory the necessary cues for the 4*Ws*. Thus, **MeMoT** represents each cluster with a photo and a text description, complemented with a selected detail to give an overview of the context.

2.7 Adapting to the user's context

Santini et al. [SGJ01] argue that the meaning of a photo is contextual, depending on the query and the user querying the collection. This vision is shared by other researchers, that use context information to adapt the results to the user's needs. Evans et al. [EFVC06] claim that personalisation is a key facilitator in helping people find what they are looking for in large collections of photos. There are situations when the personalisation is group-based, by combining, comparing, or merging individual preferences. Those actions should occur at the semantic level, and thus, the knowledge necessary for the personalisation should be represented using solutions that enable conceptual reasoning. One way to include the personal preferences in retrieval is to store individual user profiles [VMCFCA05]. They include static and dynamic information, covering device profile, media preferences, name, birth date, nationality, residence, language, education and job. Since profiles are based on ontologies, the authors exploit domain concept semantics and the user interests in those concepts and their properties. They refer to this as *Semantic*

user preferences represented by pairs of concept-weight, where the weight $\in [-1..1]$ (negative to positive preference). Based on those weights, the results are filtered and ranked according to the user's preferences. Mylonas et al. [MVCFA08] take the ontology-driven approach further by using fuzzy representations to tackle the inherent uncertainty involved in the automatic interpretation of meanings during retrieval. User preferences are modelled with two distinct fuzzy sets, one for the positive preferences and other for the negative preferences. Jaffe et al. [JNTD06] incorporate a subjectivity factor in a photo collection summarisation algorithm, when ranking the clusters. The subjectivity is modelled by a relevance attribute that can take into account parameters such as recency, the time of day, the day of the week, the social network of the user and user attributes. Recent studies [OOO10; Sin11] build the adaptation on top of the social network of the user and the photos he shares. Sinha [Sin11] model the interest of a photo depending on image quality, but also on the number of likes, comments, and friends tagged. Obrador et al. [OOO10] analyse the user's online photo albums to learn some preferences (e.g. identities) used in the summarisation of photo collections for creating new photo albums to be shared online.

MeMoT uses ontologies to represent domain knowledge, in particular, social concepts ranging from social relations to social regularities in time and activities. This approach is similar [VMCFCA05], but instead of modelling personal preferences, it enables a transformation between different perspectives in the collection, into one that is compatible with the user's context in the 4*Ws*. For example, a holiday in Argentina during last August, it is always a summer holiday for a person that lives in the north hemisphere. The adaptation is used on summaries, on archiving, and on retrieval.

2.8 Ontologies

Gruber [Gru+93] defines an ontology as "*A specification of a representational vocabulary for a shared domain of discourse — definitions of classes, relations, functions, and other objects*". In their work, the portability issues raised by sharing formally represented knowledge among systems, are clear. In the next decade, Tim Berners-Lee seminal paper [BLHL01], presents the vision of a World Wide Web (WWW) where content is designed to be understandable by humans, but can also be understandable and processed by machines, on behalf of humans. This automated reasoning is possible if machines have access to structured collections of information and a set of inference rules. The difference to "*old school*" knowledge-representation is that there is no centralised representation, and not all stakeholders share the same definition for common terms. This led to decentralised ontology creation, posing several problems for researches and developers on ontology matching [ES+07], for example. But, it also recentres the problem on the human user. Ontologies are used to specify the meaning of the terms in a vocabulary that is used within some domain. Since many domains shared concepts, like objects, space, time,

and people, researchers started to work on upper level ontologies, that are used to facilitate the semantic integration of domain ontologies. They define and axiomatise those general categories. Among the most important, the *Suggested Upper Merged Ontology* (SUMO) [NP01], DOLCE [MBGGOOSIcH02], and *Basic Formal Ontology* (BFO) [SG02]. Although those ontologies cover many concepts, they mostly cover *particulars*, described by their *endurant* or *perdurant* nature and their *qualities*. To support annotations of personal collections, it is necessary to have a broader vocabulary than the one offered by upper ontologies. There are general-purpose semantic knowledge bases, that try to capture common sense knowledge and expose it in a structured way. Among the most important, we have the WordNet [Mil95] that focus on formal taxonomies of words, the ConceptNet [LS04; HSA07] presented as a semantic network of common sense knowledge, and OpenCyc [Ope] available as a common sense knowledge logical framework. It is possible to join the upper level ontologies and common sense knowledge bases. SUMO is totally mapped to WordNet, and the OntoWordNet project [GNV03] aligned the top-level of WordNet to DOLCE.

During the last decade, researchers developed (or adapted) domain specific ontologies to deal with different aspects of a personal photo collection. One of the core concepts is the `event`, that can be modelled as an aggregative entity for the 4*Ws*. This is the approach taken by the Simple Event Model [VHMVSS12] and by the event model presented in the Music Ontology [RASG07]. They both define minimal events, relying on external vocabularies to refine the knowledge expressed. The first was designed to represent events derived from various sources, and supports incomplete and partial information. The later was developed to support events in the musical domain. However, it is domain agnostic and can be used as a general event ontology. In it, an event is an entity used by *agents* to classify relevant patterns of change. They follow Allen's ideas for interval temporal logic [AF94]. Other event ontology is the F event model [SFSS09; SFSS12], that supports many of the features identified in [WJ07], like participation, mereology (`part-of` relations), causal relations and correlations between events. F is built on top of DOLCE+DnS Ultralite (DUL). Another event model was proposed by Shaw et al. [STH09] providing an interoperable event model, by defining one class, the `Event`, with properties covering the 4*Ws*, in terms of agentivity, time, space, participation and causality. On those ontologies, the temporal information is usually expressed as instants or intervals, relying on Time ontology [HP06]. Spatial information is modelled as WGS84 coordinates[4] or uses more expressive ontologies. W3C Incubator Group [LSG07] started to work on a standard for multi-purpose spatial ontology, but it has not reached the draft or recommendation status. Nevertheless, Geonames [Geo12] provides an ontology that covers topological entities and links information with wikipedia and DBPedia. It is used by many researchers to integrate spatial knowledge (e.g. [VHMVSS12]).

[4]Basic geo vocabulary `http://www.w3.org/2003/01/geo/`

Identities and activities are more heterogeneous and domain-dependant than spatio-temporal information. One work developed for Kodak by Luoi et al. [LLCEJKLY07; YLLCEJKL08] defines a lexicon of terms to annotate photos, based on 7 categories: *(i) Subject Activity*, *(ii) Orientation*, *(iii) Location*, *(iv) Traditional subject matter*, *(v) Occasion*, *(vi) Audio*, *(vii) Camera motion*. The terms are specific to the personal domain, and their semantics is domain-dependant. For example, location is a categorisation that enables assertions like *kitchen* or *bedroom*. In another work [CZJ08], the authors developed an ontology to support annotations of photos, emphasising on family relations. Nevertheless, to represent identities, many of the event ontologies ([RASG07; STH09; VHMVSS12]), rely on the Friend-of-a-Friend [BM10] vocabulary. The representation of the activity is generally done representing terms that follow the simple knowledge organization system (SKOS) [MB09] guidelines. However, one possible taxonomy is available at [YLLCEJKL08].

MeMoT links to some of these ontologies [HP06; Geo12; RASG07; CZJ08; STH09; VHMVSS12] to support actions on collections of photos. It also uses vocabulary from OpenCyc, Dolce, Dublin core, BFO, FOAF, and SKOS. However, there are some sociological aspects that are relevant for the domain, but are not properly covered by those ontologies. Namely, the temporal cycles, the social relations, the activities's taxonomy and their temporal and social patterns of occurrence. We present a new ontology, called **MOnt**, an acronym for **M**emory **Ont**ology, that supports those important key features.

3

Knowledge Representation

"Common sense is not so common."
Voltaire

This chapter presents the knowledge representation requirements of **MeMoT** describing how they are addressed, in what concerns expressiveness, expandability and performance.

The knowledge base (KB) is the component responsible for keeping the information and knowledge used to describe the context surrounding photos in a personal photo collection. It consists of:

- a set of axioms describing conceptual entities and their relations, with a given semantics. They are common to a specific culture;
- a set of asserted facts that are specific instantiations for a collection of photos.

Those concepts are about events, time, space, content-based features, and social and personal information, covering the 4*Ws*. The core of the KB is the **MOnt** ontology, developed to support the operations (e.g. archiving) over a collection of photos. Ontologies are composed of taxonomic hierarchies of classes, class definitions, and the subsumption relation [Gru+93]. The advantage of ontologies over rigid taxonomies is that they allow for richer semantics, towards the adaptation of the users context. They can contain a broader scope of information, enabling semantic relationships between several taxonomies. But more important, they provide some properties that are important in this context, namely, *consistency checking*, *expandability* and *completion* [GP99; McC05]. The first two relate to the need to expand the facts in **MeMoT** towards the users' needs, keeping knowledge

consistent. The last one, enables small amounts of information obtained from the users to be expanded, adapted or complemented, towards the users and the processes needs. **MOnt** uses many concepts already defined in other ontologies, to represent spatial, temporal and event related concepts. It uses, among others, the *Time Ontology* [HP06], the *Geonames Ontology* [Geo12], and *WGS84 Geo Positioning* vocabulary [W3C03], the *Friend Of A Friend* (FOAF) vocabulary [BM10], and the *Linking open descriptions of events* [STH09]. This subject will be expand ahead in this chapter. For most of the concepts, the KB keeps knowledge at different levels of detail. These requirements arise from the need to:

(i) reason from partial cues referring to concepts that are present in a given context;

(ii) suggest information to complete the context based on assertions;

(iii) present concepts described with an adequate level of detail to the given context;

(iv) reason using different levels of detail;

(v) deal with imprecise information provided by the users.

While explaining the knowledge representation, to illustrate concepts, instances, properties and types, we will use the graphical representation showed in Figure 3.1. Other visual aids may be used for clarification purposes, but they will have no impact on the implementation. The concept and property's names follow the vocabulary best practices[1], using *CamelCase* and *mixedCase* notation respectively.

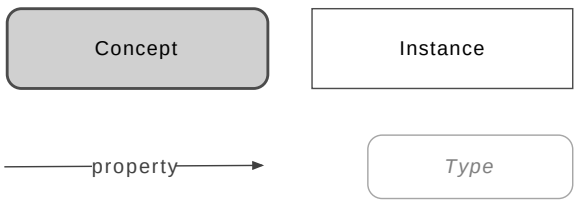


Figure 3.1: Graphical symbols used in diagrams when describing the knowledge representation needs for the KB. It follows the terminology used in concept languages, for example in description logics [BCMNPS03].

This chapter is organised as follows: The first four sections describe the proposed representation for the 4*Ws* — Section 3.1 describes the spatial related concepts, Section 3.2 characterises the time and date concepts, Section 3.3 details the social information and finally, Section 3.4 outlines the content-based concepts used in this approach. Next, the concept of event is introduced, discussing different types of events and their representational needs. In Section 3.6 the definition of Semantic Viewpoint is introduced, discussing

[1] http://wiki.opensemanticframework.org/index.php/Ontology_Best_Practices

how a transformation of perspective is possible using the KB. Some implementation details are discussed in Section 3.7. The chapter ends with an overview of the important information to retain.

3.1 Spatial concepts

The act of remembering relies on spatial cues to improve the reconstruction of past events. This is not strange, since space, along with time, are the constant dimensions in our life. However, spatial location information in the KB needs to address specific requirements:

1. To enable the representation of spatial locations at *different levels-of-detail (LoD)*, using a *human-readable* codification;

2. To provide support to handle *subjectivity* and *imprecision*.

3.1.1 Levels-of-Detail

The spatial information of photos has an inherent imprecision, arising from the multitude of sensors and services that provide localisation data. A spatial location is represented by latitude/longitude coordinates, usually using the World Geodetic System (WGS84) standard [Age04]. According to [WM10], the coordinate can have an error, that ranges from few meters up to 3 km. In **MeMoT**, the precision should be sufficient to identify a time zone, without ambiguity. This provides a spatial reference to time, and thus, enables the generation of temporal locations with different *LoD*. However, the way people refer to places is not by their latitude/longitude, but using a set of terms whose semantics is related to standard organisational units, like *country* or *city*. Thus, the relevant concepts in **MOnt** are the ones that describe what encloses a coordinate, following a set of standard organisational units. A place can be represented using instances of `SpatialLocation`

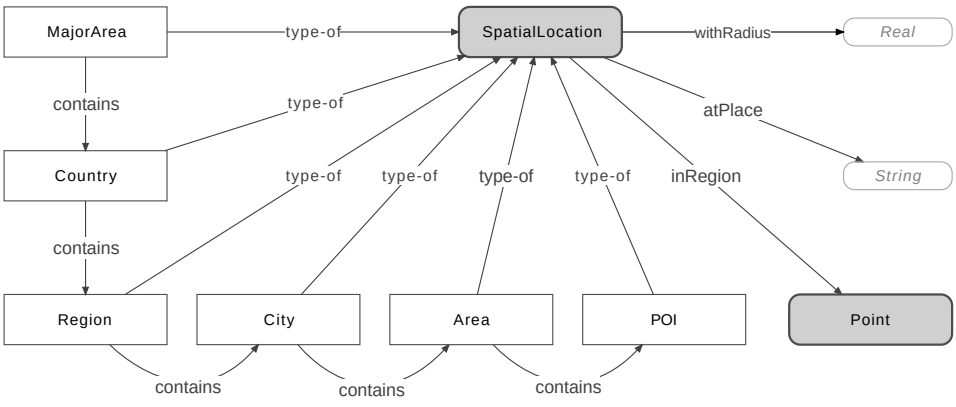


Figure 3.2: Overview of the spatial location concepts.

with different specificities, according to the semantics the user wants to use for a given

location, as illustrated in Figure 3.2. Some concepts are based on the Geonames ontology [Geo12], augmented with the standard organisational units people use to refer to places. For example, there is no concept of `Country` in that ontology, although there is a two letter country code[2], that can be used to determine the country. The suggested model is arranged hierarchically using the `contains` relation, providing an ordered path of *LoD*, with the following semantics:

(i) the `MajorArea` is the top level, lowest detail concept in the hierarchy. It can be an *hemisphere*, an *ocean* or a *continent*;

(ii) the `Country` concept is an administrative area that represents a *country*;

(iii) the `Region` concept is an area that contains the most adequate level inside a country;

(iv) the `City` concept is the most specific name of the closest populated place near the locale. It may be a *city*, or a *village*, to name a few. This is the default *LoD* for spatial locations;

(v) an `Area` inside a city, e.g., East Village in New York;

(vi) the `POI` concept is a *point of interest*. It may be a *monument*, a *street name*, etc.

Do note the above list are special cases of instances, that can be simultaneously, an `Instance` and a `Concept` depending on the context.
The mandatory `atPlace` data property is used to specify the name of the location. This property is functional, with a `SpatialLocation` as domain. The chain of `contains` relations, linking `SpatialLocation` instances, provides alternative descriptions of a place, that maps to standard spatial locations, whose semantics are important to people. The `contains` is a transitive object property, whose domain and range is `Thing`. It has an inverse property, named `isContainedIn`, that is not represented in Figure 3.2 to avoid clutter. Together they provide a mechanism to traverse the hierarchical relation of specificity between locations.

Sometimes a user wants to name an area around a coordinate. This is most useful when there is no knowledge about the geography or administrative boundaries — a common situation for remote, non populated places — but it is also necessary when the place name exists but with a non official status. For example, to name an estate where the family has a second house. The owners and the locals know the property by name, can say where it is, but officially the address says National Road nº 4. What is important as a memorable term is the estate's name not the road where it stands. In such situations, besides the use of the `atPlace` property to name the place, a user can also use two more properties: `inRegion` and `withRadius`. The first, is an object property that holds the coordinates of a place, with a given precision, depending on the source. For example, a

[2]ISO 3166, available at `http://www.iso.org/iso/country_codes.htm`

typical uncorrected GPS[3] unit provides four decimal places, enabling an accuracy up to 10 meters [WM10]. The `withRadius` data property expresses a circular area, in meters. The semantics of this two properties is the following. It is used to name a spatial region, centred in `inRegion`, up to `withRadius` apart.

There are two situations that are worth mentioning. One is that, for the same latitude/longitude we can have several `SpatialConcepts`, representing the different *LoD* we want to use. The other situation is related with the `POI` concept, as it is possible to have several points of interest in one spatial coordinate. For example, the building called *Casa dos Bicos*, in Lisbon, is a point of interest, for its history and peculiar architectural design. However, inside the building there is another point of interest, the *Fundação José Saramago*, that honours the life and work of the Nobel Prize writer.

3.1.2 Imprecision on named locations

The "imprecision" addressed in **MOnt** to support the spatial locations is different from the imprecision that is inherent to latitude/longitude reference, as this one can be inaccurate because of errors. For memory's sake, people used named locations to position events in space, and those have an inherent imprecision as many times they have ill-defined limits. What matters here is their names and the relations between them. Lets take the example of Broadway, in New York City (NYC). The name refers to a road in Manhattan, the name of a theatre, but it is mostly used to name the area where one can watch theatre shows. The same happens with the name East Village, an area in NYC. Its boundaries are not officially set, but people do refer to that area by its name, naturally handling the imprecision. Most of the time, name locations do have a single name for different levels of detail, some of which have ill-defined boundaries, as illustrated in Figure 3.3. Using the name *Marquês de Pombal* we may refer to an area, or to a statue inside that area. Although the statue is located in a square that follows *Avenida da Liberdade* — the avenue —, it is usually referred to as *Avenida da Liberdade* — the area. For annotation purposes, the name is what really matters. And the names used should be memorable so that they can later be used for retrieval purposes. Inserting the location of the statue in **MOnt** will use different *LoD*. Three new instances are created: *(i)* the `POI` instance, `MarquesPombal`, with property `atPlace` equal to "Marquês de Pombal"; *(ii)* the `Area` instance, `AvenidaDaLiberdade`, with property `atPlace` equal to "Avenida da Liberdade", and property `contains` equal to `MarquesPombal`; *(iii)* the `City` instance, `Lisbon`, with property `atPlace` equal to "Lisbon" and property `contains` equal to `AvenidaDaLiberdade`. If we query **MOnt** about points of interest in Lisbon, "Marquês de Pombal" will come up as being in Lisbon, since it is contained in Avenida da Liberdade, an area of Lisbon.

[3] An autonomous, or uncorrected, GPS unit has no correction applied. This is the typical scenario for smartphones and consumer level GPS units.

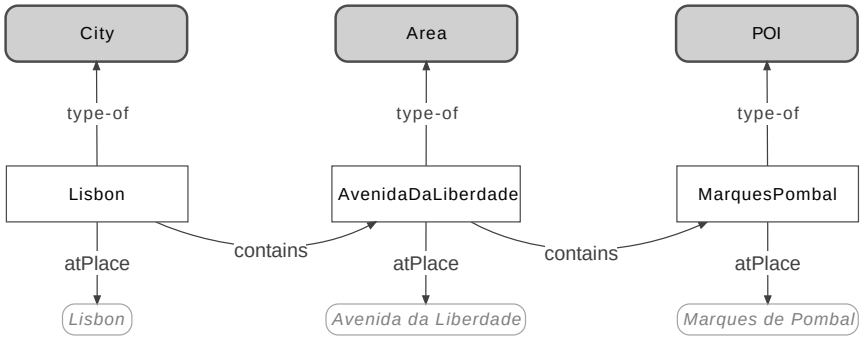


Figure 3.3: Example of an assertion of a spatial location.

3.1.3 Spatial hierarchies

The hierarchical structure of a spatial location can be determined traversing the chain of `contains` relations. However, this poses some constraints on the expansion of the knowledge base, since it locks the number and sequence of the named locations *LoD*. Since the KB supports different operations over a set of photos, is should provide mechanisms to settle different hierarchical organisations, as we should not assume that all operations have the same requirements. For example, during the archival of photos, all *LoD* must have named locations asserted to, in order to ensure that the information is complete. However, during retrieval, it is not necessary to use all *LoD*. We may want to remove some, e.g. the region, so the description of the place becomes more concise. As we know, for many populated places where people take vacations, the name of the region coincide with the name of a major city. As we will see in Chapter 5, the spatial hierarchies play an important role to settle the *LoD* for visualising a set of photos, during retrieval. To support such requirement, a set of helper concepts were created, as showed in Figure 3.4. The core concept is `Hierarchy`, modelled as an ordered list of spatial locations representing the hierarchy levels. A hierarchy has a name and a given depth, represented by the data properties `hasName` and `hasDepth` respectively. The `hasName` is a functional data property, that holds a common human understandable term. For example, the name "*Country-City-Area*". To model the ordered *LoD*, the `OrderedElement` concept is used. The `OrderedElement` is a "weak entity" of an `OrderedStructure`, meaning that its existence only makes sense in the context of an ordered list. Thus, it is mandatory that every `OrderedElement` belongs to an `OrderedStructure`, or a sub-concept, for example a `Hierarchy`. This association is made using the mandatory object property `isMemberOf`, whose domain is `OrderedElement` and range is `OrderedStructure`. An `OrderedElement` assigns a specific level to a given spatial location, using the `index` data property and the `refersTo` object property, respectively. For `SpatialHierarchy` the range is `SpatialLocation`. Do note the order of each level is hierarchy dependant, which means the same spatial location can be at different depths in distinct hierarchies. The navigation in the ordered list of levels is made using the object properties

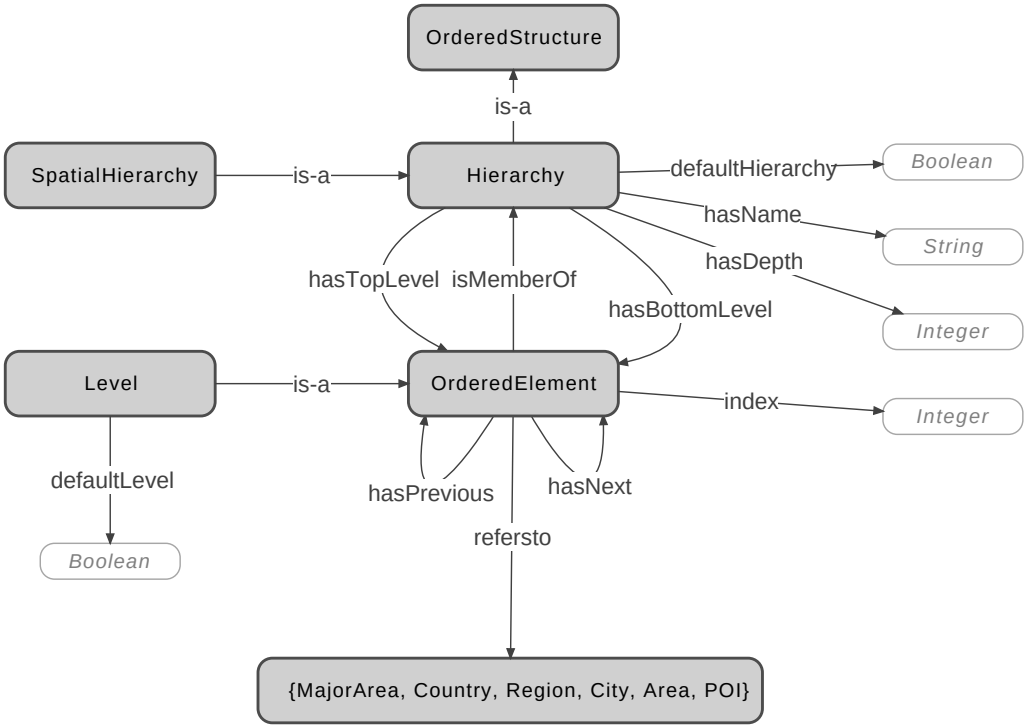


Figure 3.4: Hierarchy concept.

`hasPrevious` and `hasNext`. For presentation purposes (see Chapter 5 for further details), a level of a hierarchy can be marked as the most adequate level to describe a spatial location. This is achieved by the optional data property `defaultLevel`. A hierarchy has two object properties, `hasTopLevel` and `hasBottomLevel`, to establish direct connections with the extreme levels of the hierarchy. Those properties' domain is `Hierarchy` and their range is `OrderedElement`. Finally, in **MOnt**, there must exist a default spatial hierarchy, properly signalled using the object `defaultHierarchy` property. This simplifies the selection of the right hierarchy, if there are multiple ones.

Figure 3.5 illustrates an instantiation of a spatial hierarchy named *ShortLOD*, with 4 levels. The first is the Country, and the successive, more detailed, levels are the City, the Area, and the POI. When using this spatial hierarchy for a set of instances in **MOnt**, and if for some reason the places do not have information for any of this *LoD*, it is left to the implementation to provide a default case where the name associated to the `SpatialLocation` is an empty string. In no circumstance the order given by a spatial hierarchy can contradict the order specified in Figure 3.2, given by the `contains` relation.

3.1.4 Summary

In this section, we describe how a spatial location can be represented, when the focus is set on the perception users have of places. The goal is not to represent every position

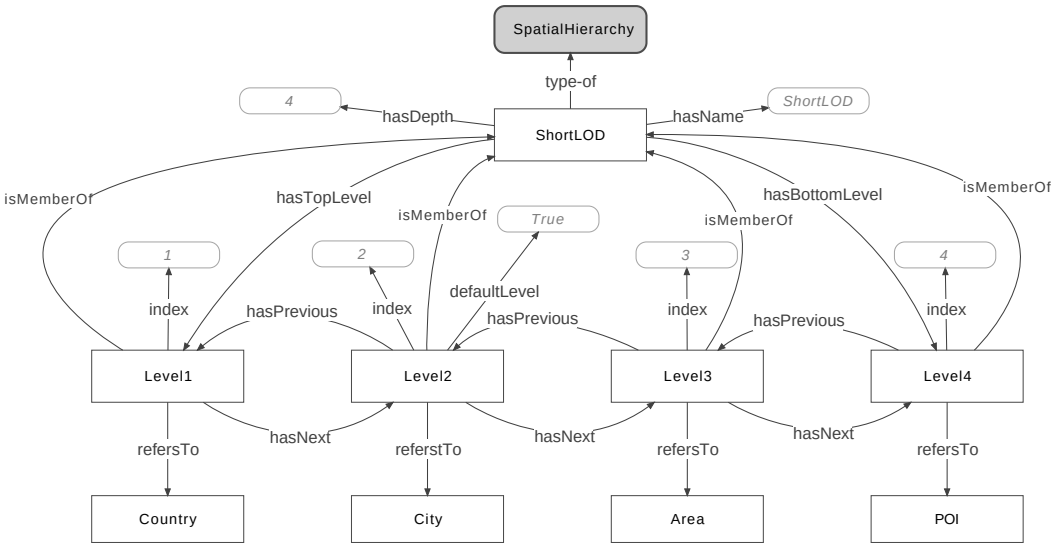


Figure 3.5: Examples of spatial hierarchies.

that exists in a photo collection, but to represent the different names of places that are relevant in the personal photo collections. Those places are arranged around a set of standard spatial location, for example, the City. Summing up:

- the description of a place is an understandable hierarchical set of spatial standard locations;
- the spatial standard locations are related by specificity, permitting to select the proper *LoD* to present a locale;
- it is possible to define spatial hierarchies of *LoD*s;
- the representation allows imprecision as it is not bound to a faithful representation of the reality, but rather to a representation of personal perceptions of that reality;
- a place can be represented in two ways: using only a name or defining a named area around a point.

3.2 Temporal concepts

We can see photos as memories that share the same characteristics as the autobiographical memory, especially time. Autobiographical memories can be dated, and the events are assumed to occur one after the other [Bur08]. Thus, time is inherent to recovering information from the past [Ten08]. We sense and reason about events using some elementary temporal notions, including: *(i)* duration, *(ii)* non-simultaneity, *(iii)* order, and *(iv)* past and present [LP09]. Despite time being a linear concept, our social organisation uses the calendar and specific temporal locations to accommodate events that embody

the seasonality present in our lives [Zer85]. Expressions like "last *summer*", "this *month*", "later, in the *afternoon*"are commonly used to position events in time. Such temporal terms are relative to some artefacts that are used to take account of time, like the calendar or Greenwich time. So it's not surprising that such an important dimension in our life reveals some interesting features, namely:

1. it is *spatial dependent*;
2. it has a *cyclic* representation in a *linear* space;
3. it can have *different granule* to express the temporal information;
4. it needs a context to be unambiguous;
5. it can be arranged *hierarchically*.

Spatial Dependence The perception we have of time, and thus, the words we use to describe it, depend on the context, including the spatial location. Temporal references, like *Day* or *Sunset*, depends on the location. For example, in a June day in Faro, south of Portugal, the sun sets at 8:51 p.m. whereas at Oporto, in the north of Portugal, it sets at 9:07 p.m. It turns out that one person in Oporto can describe the time 8:51 p.m. as being part of the day, while other person in Faro will describe the exact minute as sunset. Notice the time zone is the same in the two places. Time is unambiguous when relative to a referential, usually expressed as UTC[4].

Temporal Patterns Although time is thought to be linear, there is a set of temporal patterns that are repeated at regular intervals. Such patterns are the cycles that embody the seasonality of time, and they fall into three categories [Zer85; RWJM03]: *(i)* physio-temporal patterns, *(ii)* bio-temporal patterns, and *(iii)* socio-temporal patterns. The first are related to physical laws, and are dictated by natural phenomena, like sunrise, for example. The second regulates the livings things, for example the circadian rhythms. The last regulates the structure and dynamics of social life. It is the temporal patterns and cycles influencing the human activities, that we model. Artefacts, like calendars and cycles not only help to coordinate highly complex systems [SKK05] — a society for example — but also help to plan and schedule activities, and thus, set the grounds to take photos for memory purposes. Since we are intolerant to temporal anomalies [Zer85], the cycles' regular intervals of repetition are, in many cases, approximations to astronomical models, where the boundaries were settled by convention. However, other cycles are purely artificial, like the week. Nevertheless, the day, the week and the year are responsible for the rhythmic structure of social life [Zer85]. They repeat time after time, bending the continuous into a set of tangible, and for some, observable cycles that we use to position events.

[4]Coordinated Universal Time, defined by International Telecommunications Union Recommendation http://www.itu.int/dms_pubrec/itu-r/rec/tf/R-REC-TF.460-6-200202-I!!PDF-E.pdf

Imprecision and ambiguity Imprecision and ambiguity are inherent to temporal references, those used by us all the time. It is the context, either cultural, personal or social, that gives the semantics to those temporal references. For example, the *Day*, in the sense of day light presence, ranges from sunrise to sunset, but can have a different extension depending on a multitude of factors, from personal to cultural ones. For some people, the day stops at sunset, for others the day includes dusk[5]. Other temporal references, like *Afternoon*, lacks a precise definition. For example, the afternoon is *"between noon and evening"*, and so, the lower boundary is known and objective but the upper boundary is known but subjective, depending of social factors like culture and country, to name a few. For many time references, like the *Dusk*, there is a precise definition, but the correct placement depends on the accuracy of the clock. It is known that more vague descriptions of time seem to be much effective to memory recall [Bur08]. This notion of "vagueness" relates to the imprecision people have in dating and locating events, especially very distant ones [Fri04]. In this work we deal with temporal imprecision, by:

1. providing different levels of detail (granule), dealing with imprecision;

2. support terms with a ill-defined limits, for supporting ambiguity.

Time references with different grain Citing Bettini et al., "*The organization of human activities, as well as any communication related to these activities, must deal with an explicit or implicit temporal context, which is expressed in terms of an appropriate time granularity*" [BJW00]. The grain, or level-of-detail (*LoD*), at which time is used depends on the descriptive needs of the person and on the event to be retrieved. Many researchers, namely [CPP00; BKC03], argue that autobiographical memories contain knowledge at different *LoD*s, to help remember events at different locations in time [Bur08]. Granules like *minute* or *hour* may seem appropriate to recover close temporal events, residing in the short term memory, but they are not effective to recover memories with several years, because time information is held using a coarser grain in long term memory. In such cases, using time references like the *year* in conjunction with cycles, e.g. *Summer*, is key to provide the stepping stones to pass from almost oblivion to full recollection of episodes. Thus, mimicking the mechanics we use to remember and recover episodes from our memories seems appropriate to manage and recover photos.

Hierarchise time references The bends made on top of the linear nature of time enclose different *LoD*s, that are not independent from each other. Those bends take form of time references related to each other, mostly in a hierarchical partitioning scheme [Dea89]. The division of the Year in Quarters, and sub-sequentially in Months, Weeks and Days, specify an hierarchical structure of *LoD*s forming an ordered path from a coarse to a fine

[5] In this work we consider the civil dusk, although there are the nautical and astronomical dusk, settled at different angles.

grain. Such relation not only unveils the structured and ordered nature of time for people [Zer85], but also gives account of the need to use standard references to position events in time. Thus, although some references might be imprecise to locate an event, their hierarchic relations enables us to be more or less precise, by traversing a specific hierarchy. Temporal hierarchies are multiple and do not need to be disjoint. They can share some levels. For example, the hierarchies $Year \mapsto Quarter \mapsto Month \mapsto Day$ and $Year \mapsto Season \mapsto Week \mapsto Day$ share the top and bottom levels, but use different granules in the middle. The use of one instead of the other depends on the needs expressed by the user interacting with their memoirs.

3.2.1 Core concepts

For managing personal photo collections, we have three conceptual constructions to handle temporal information:

1. *Granules*,
2. *Cycles*, and
3. *Levels of Detail (LoD)*.

In this context, a granule is a temporal location. Examples of granules include `Day`, `Hour`, `Month`, `Week` and `Sunday`, to name a few. According to their needs, people use a granule of their choice to position events in time. This choice is often influenced by the proximity of the event. Granules can be arranged to form sequences of repetitions, forming cycles. For example, `Day` and `Night` are part of the Day cycle, as they form a cyclic repetition of temporal locations. A cycle tells the order of their members according to the passage of time. Do note granules and cycles can have the same name. The `Day` cycle includes the `Day` granule and the `Night` granule. The last conceptual construct, *level of detail*, is necessary to establish an order in granules from most specific to more general. Although people choose, and use, naturally the right level of detail when using temporal locations, machines need an ordered path which makes it possible to traverse the available granules.

Figure 3.6 illustrates how granules are represented in **MOnt**. A `TemporalLocation` is a super type granule, for `DateLocation` and `TimeLocation`. The `Week`, `Month`, and `Season` concepts are examples of first. `Afternoon`, `Night` and `Sunset` are examples of the second. The `hasName` is a functional data property that follows the semantics already discussed, and is used to name instances of `TemporalLocation`, with a common human understandable term. There are two types of granules worth mentioning:

1. *instants*, and
2. *intervals*.

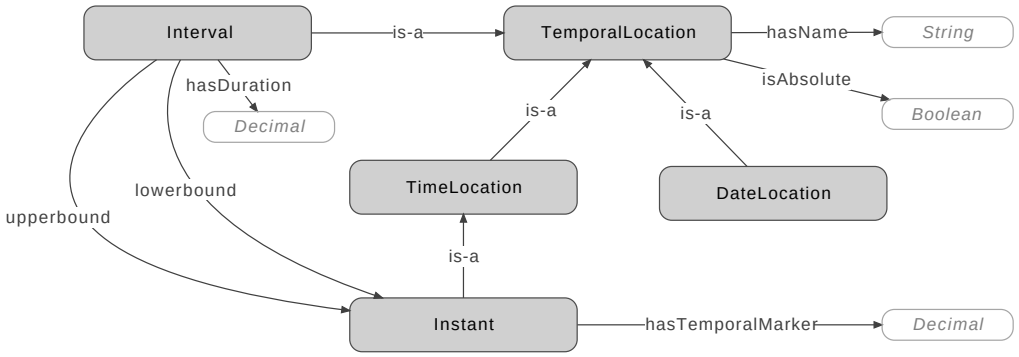


Figure 3.6: Definition of the `TemporalLocation` concept.

The word *instant* can be misunderstood, and needs a clarification for this domain. An instant is an indivisible granule, at human level. For example, `Noon` and `Sunset` are indivisible terms used to communicate temporal positions. In other words, an instant is a type of granule that is used to express the most specific detail about a temporal location. The `Instant` concept is a specialisation of `TimeLocation` that may have a temporal marker, settled using the data property `hasTemporalMarker`. The value of that property is expressed as a fraction of 24 hours. This way, an instant is positioned inside a day. This property only makes sense if the `isAbsolute` data property is set to `True`, indicating the temporal location does not depends on the context to be properly settled. The intervals have `lowerBound` and `upperBound` object properties, indicating the limits of the interval, represented by granules. Intervals can contain intervals and instants. Intervals can overlap each other but instants cannot. The `Interval` concept has a duration, expressed in terms of days, given by the data property `hasDuration`. The terminology is taken from Allen [All83].

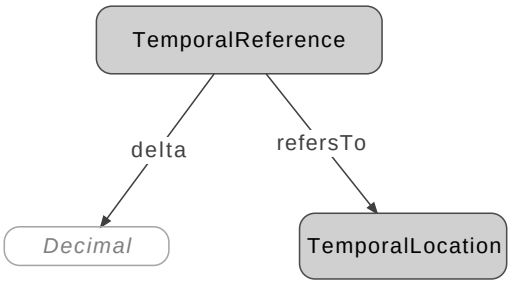


Figure 3.7: Definition of the `TemporalReference` concept, for supporting imprecise temporal references.

Supporting ambiguity Granules support the imprecision that exist in this domain. For example, saying that something happened last *Saturday* is correct, but imprecise since we do not know the exact time when it happened. However, granules do not support some ambiguity that is natural when people refer to temporal locations. The causes for this are related to ill-defined intervals of some granules, e.g the afternoon, or because the granule is precise but it may stretch beyond its duration, e.g. morning. To model such ambiguity **MOnt** defines the `TemporalReference` concept, as showed in Figure 3.7. The `delta` property indicates the value of the deviation around a temporal location, defining a range of $\pm$`delta`. The value is defined in terms of days, either multiples or fractions. Let's see an example. We want to define a reference that can deviate at most one hour from the conventional time. We define an instantiation called `HourReference`, whose `delta`=$\frac{1}{24}$. If we use this reference together with the concept `Morning`, that starts at dawn and ends at noon, we are expressing that we accept that mornings can start around dawn and end around noon, between 11 a.m. and 1 p.m.. A `TemporalReference` has a `TemporalLocation` indicating the temporal granule, as illustrated in Figure 3.6.

3.2.2 Cycles

The cyclic nature of time is an artefact to model the observable rhythmic repetitions in nature. Cyclic temporal locations are key to position past events, and thus, they should be used as cues to retrieve photos from **MeMoT**. There are three main cycles whose importance should be emphasised: *(i)* the daily cycle; *(ii)* the weekly cycle, and *(iii)* the annual cycle. A cycle settles an order over its set of elements. Moreover, a cycle is a repetition of

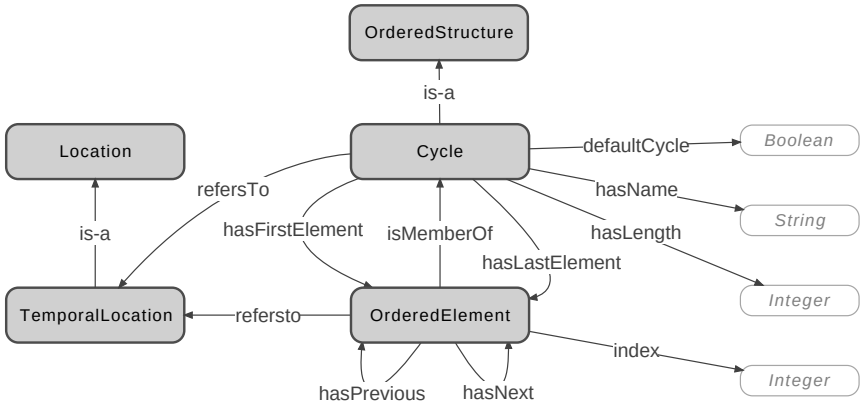


Figure 3.8: Cycle concept.

those elements, keeping their order. Figure 3.8 shows the definition of the cycle concept. A cycle is an ordered list of `OrderedElement`. Each one has a position in the cycle, given

by the integer data property `index`. As we see previously, an `OrderedElement` has two mandatory object properties, `hasPrevious` and `hasNext`. The first is used to settle the predecessor element in the cycle, and the second is used to indicate the successor element. The `hasFirstElement` and `hasLastElement` object properties, where domain equals `Cycle` and Range equals `OrderedElement`, are used to settle the extremes of the cycle. Do note those extremes are connected with `hasPrevious` and `hasNext` properties. This means the predecessor of the first element is the last element and the successor of the last element is the first element. As with hierarchies, an `OrderedElement` is associated to its order structured by a mandatory `isMemberOf` object property. A cycle has a name, a length, and a level-of-detail, settled using the mandatory data properties `hasName` and `hasLength`, and the object property `refersTo`, respectively. A cycle can be marked as the default cycle for presenting temporal concepts, related to social and cultural conventions. This is achieved using the optional boolean data property `defaultCycle`.

Daily cycle Figure 3.9 shows the important concepts associated with the day cycle. To avoid clutter, not all the information is depicted.

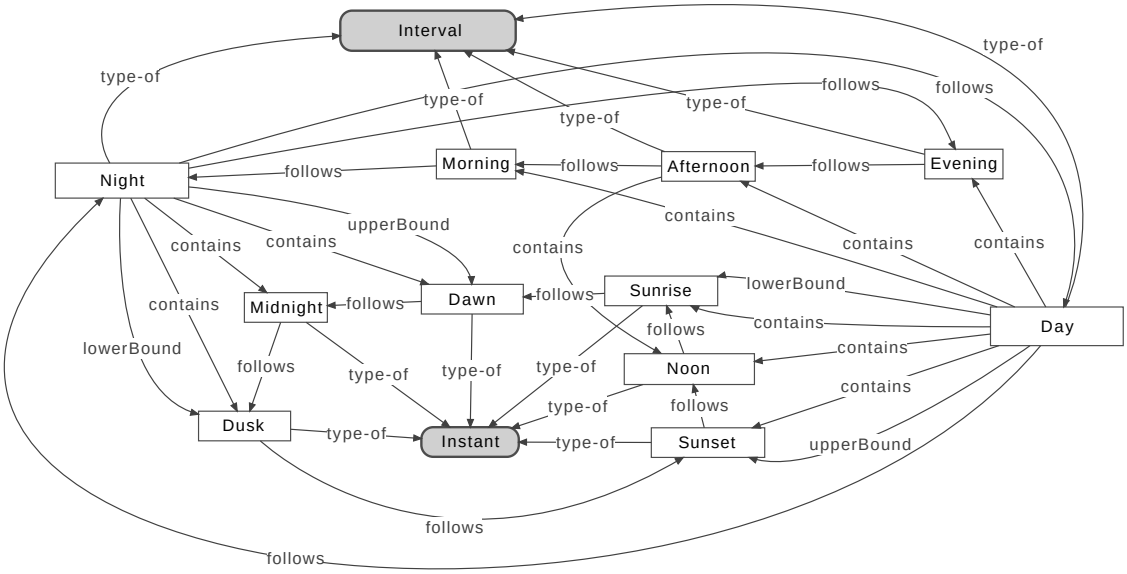


Figure 3.9: Overview of the day cycle.

We have main object properties to relate concepts: *(i)* `contains` and *(ii)* `follows`. The `contains` is a transitive object property, whose definition is equal to the one used in the spatial concepts. However, here it can be seen as a *partOf* relation [Arm97], expressing that besides the hierarchical relation between concepts, the temporal range of the *contained* concept is delimited by the container one. The `follows` data property is used to express a cycle of temporal locations, defining a traversal order in concepts. Besides, it also states that the temporal range of the concepts does not overlap. The `follows`

property, whose domain and range are `TemporalLocation`, is transitive. This characteristic will provide the support to determine that temporal location follows one after the other, the self included. Besides of such definitions, it is necessary to define named cycles using the concepts illustrated in Figure 3.8. Figure 3.10 shows the necessary properties to define a cycle in **MOnt**, using one of the many cycles inside the day. In this case, the example depicts the cycle for temporal locations inside a day.

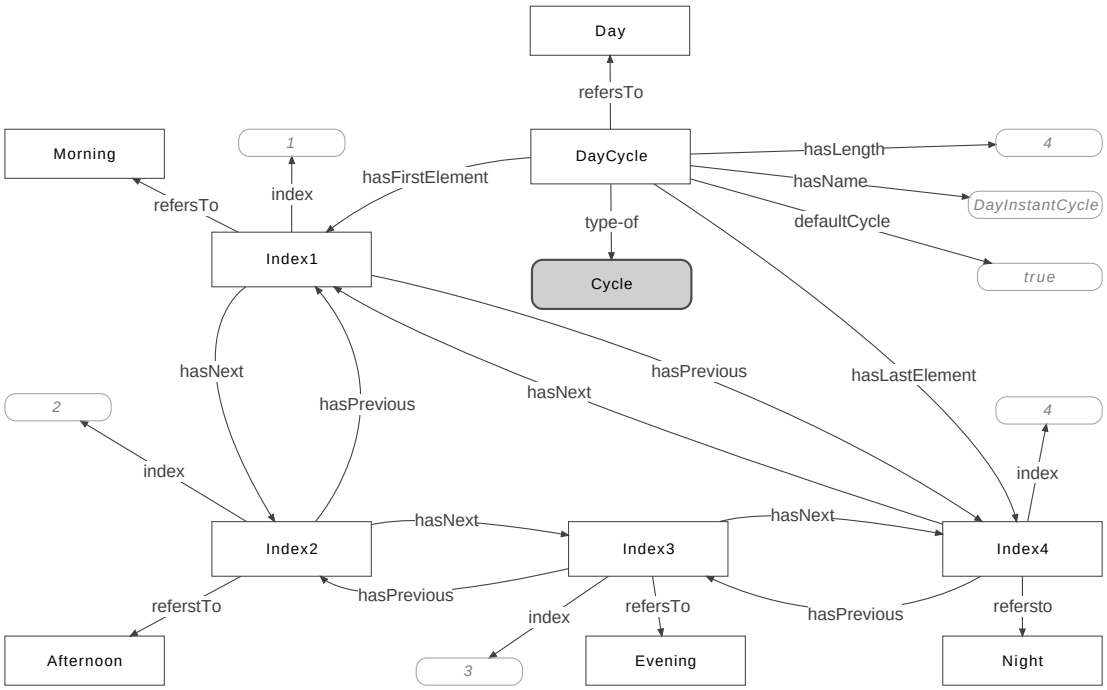


Figure 3.10: Example of a day cycle, that includes the memorable instants inside a day.

Weekly cycle Figure 3.11 shows the important concepts associated to the week cycle. The properties `follows` and `contains` are already discussed for the day cycle. Between the two cycles displayed, the one containing the week days is the default one. Although it is not depicted, each weekday links to the concepts showed in Figure 3.9. The default element in each cycle follows what is defined by the ISO-8601[6] standard, where the beginning of a week is on Monday. Using that reference, the `Weekend/Workweek` cycle starts at the `Workweek` and the weekday's cycle starts at `Monday`. There are other conventions, namely religious, that define other days as the starting of a new week. Those conventions are addressed by using the aforementioned concepts and properties.

Yearly cycle The last cycle is the *Yearly Cycle*, whose concepts are illustrated in Figure 3.12. The one depicted is the cycle from the north hemisphere, typical for the western culture. Although not depicted, here every concept is a specialisation of the `Interval`

[6] http://www.iso.org/iso/catalogue_detail?csnumber=40874

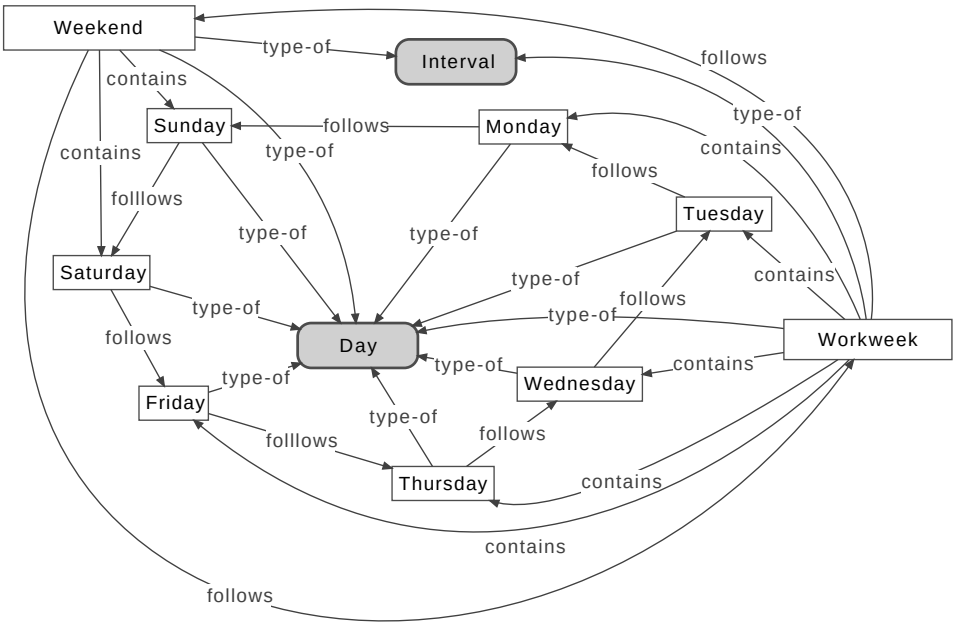


Figure 3.11: Overview of the week cycle.

concepts. The object property `includes` is similar to the `contains`, except it is not transitive [KA08]. Such need is necessary to model the relation of some months and seasons,

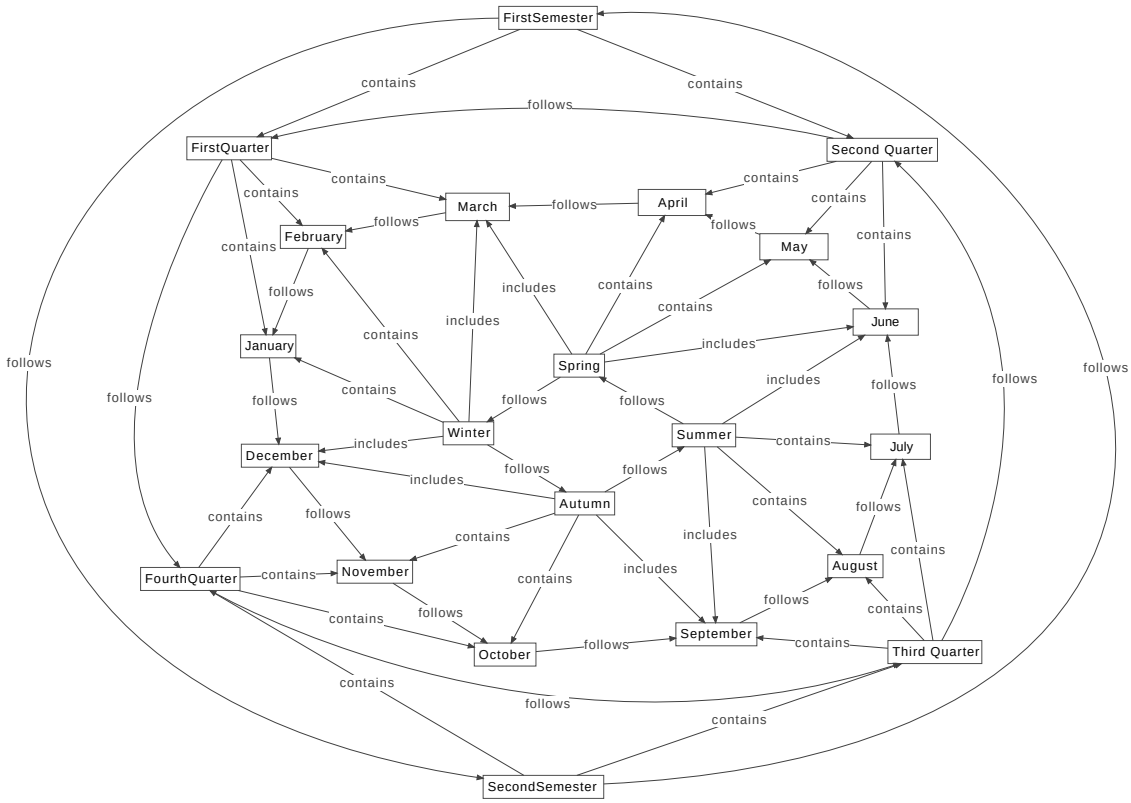


Figure 3.12: Overview of the year cycle in the northern hemisphere.

as the first belongs to two different seasons, and thus, they are partially contained in them. This odd situation exists because temporal locations are inspired in natural cycles, but are settled using conventions. While the seasons follow closely the earth revolution around the sun, the months are loosely inspired by the moon cycle. Since the references are different, the periods do not match. There are other types of seasons. For example, in tropical places there are only two seasons: the rainy and dry season. Nevertheless, it is possible to extend the ontology with the properties and concepts already defined, not only for providing those new cycles, but also to relate them with the existing ones.

3.2.3 Temporal hierarchies

As with space, we can organise temporal references hierarchically. But given its characteristics, the hierarchic organisation is richer, providing a set of references to multiple cycles. `Year`, `Week` and `Day` are important temporal references that can be used to settle levels of a hierarchy. They settle references to the three core cycles people use to reason about time. Each one represents complementary granules that can be used to refer to time. Unlike space, besides the order imposed by the hierarchic organisation of levels, there is an implicit order inside each one, dictated by temporal cycles. Let's use as an example, the hierarchy with levels $Year \mapsto Semester \mapsto Quarter \mapsto Month$, as illustrated in Figure 3.13. Except for the root level *Year*, that is not a cycle but has an implicit order,

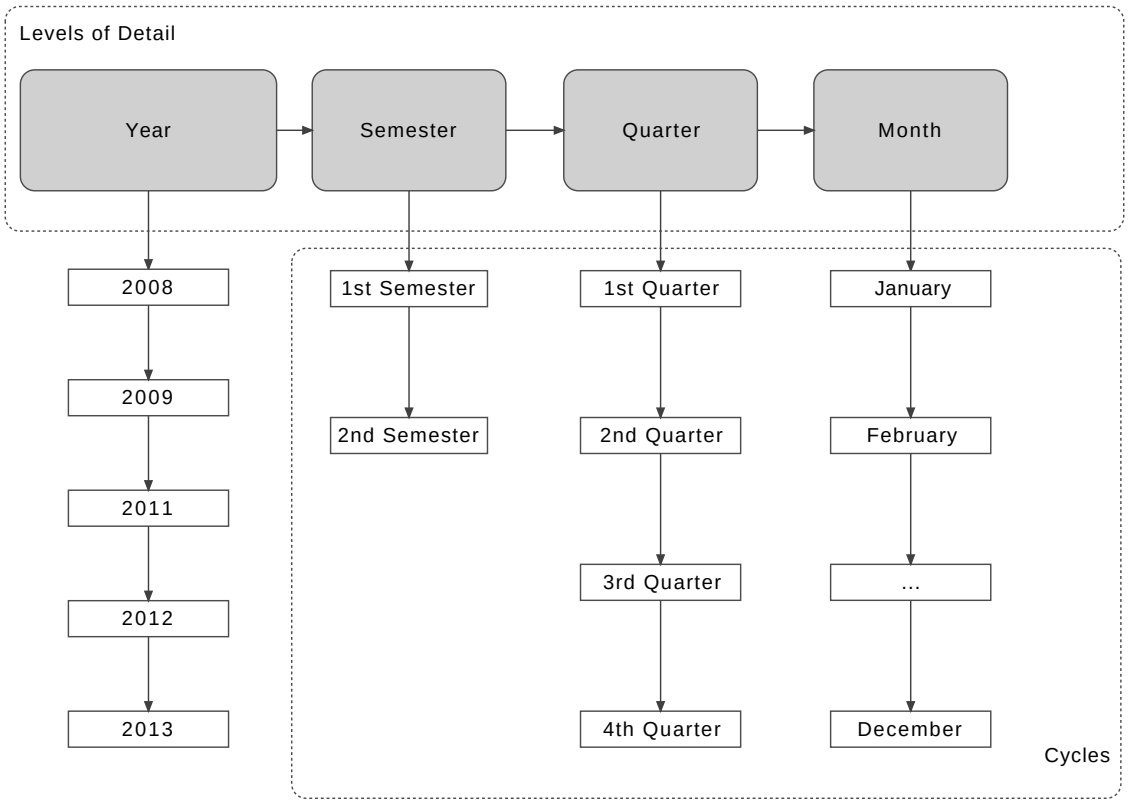


Figure 3.13: Illustration of a hierarchy of temporal references.

every other level is a named cycle. In general terms, all temporal references inside a year belong to a cycle. Thus, in temporal hierarchies it is possible to obtain an ordered list of members, using the mandatory object property `hasMember`, as illustrated in Figure 3.14. As with spatial hierarchies, using the correct *LoD* is important, since people are sensitive

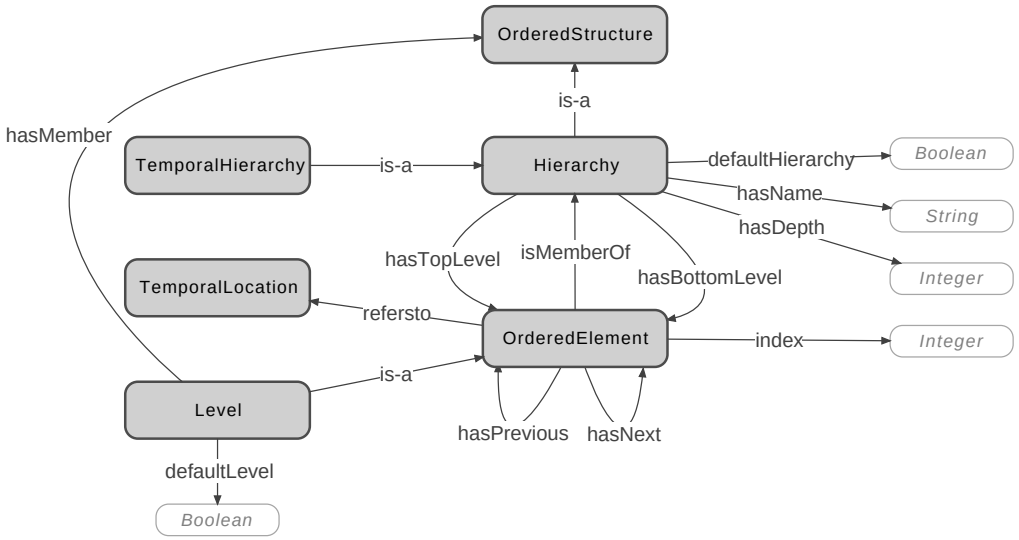


Figure 3.14: Temporal Hierarchy concept.

to it. As stated by Burt [BKC03], the level of detail presented to the user should be optimized, so the information is neither too specific nor too abstract, but still preserves the important information. This topic will be addressed in detail in Chapter 5, Retrieval.

3.2.4 Summary

In this section, we describe how a granule is represented when the focus is set on the perception users have of time. This includes the usage of temporal references with ambiguity. Since there is a cyclic nature in time, **MOnt** covers most of the important concepts and instances related to time, in particular, using a set of standard temporal granules. About time related aspects, we have:

- "Granules" is a core concept in time, represented in **MOnt** by the `TemporalLocation` concept;
- Granules can be perceived as *instants* or *intervals*. Instants are indivisible granules, at human level;
- Granules are imprecise by definition, as they refer to time at a given level-of-detail. Granules can be associated with an interval of variation around a temporal marker, allowing ambiguity;
- Each `TemporalLocation` has information about its relationship with an observer, either absolute or relative;

- The `contains` relation defines a hierarchy of temporal concepts. The upper level of the hierarchy delimits the time frame where the other levels occurs. This means each one has a `lowerBound` and `upperBound`;
- The `follows` property establishes an order between temporal concepts;
- The concepts linked by a `contains` relation can be perceived simultaneously;
- The concepts related by a `follows` relation have a non-simultaneity property [LP09];
- It is possible to define cycles of `TemporalLocations`;
- It is possible to define temporal hierarchies of *LoD*s, where each level is a cycle.

3.3 Social concepts

The spatio-temporal information can locate an event in space and time, since it is not tied to any personal or social context. However, the semantics in a set of photos goes beyond that. When a collection of photos is about our memories, the context is always self-referent [Ten08]. Quoting Fivush, "*Cultures define canonical forms of social interactions and activity, such that individuals within a culture develop a shared representation of reality that guides what are considered appropriate and inappropriate behaviours and interactions*" [Fiv11]. Since we are intended to deal with personal collections of photos, it is vital to define social and personal concepts in the KB.
In this section, the placeholder name `self` will be used as a reference to someone that is interacting with **MeMoT**. The placeholder name *who* refers to someone that exists in the memory of the *self* and is depicted in some photo. Figure 3.15 shows the central

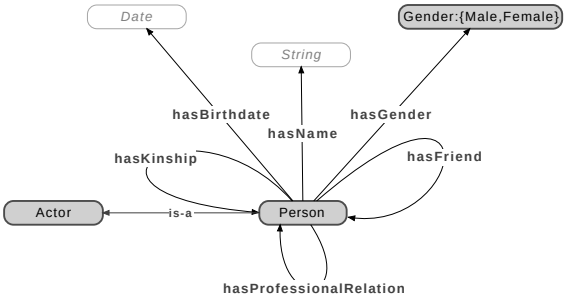


Figure 3.15: Overview of the personal and social concepts.

concept in any social modelling: the `Person` concept, with a specific type of `Actor`. In **MOnt**, a person is characterised by three properties. The functional data properties `hasBirthdate` and `hasName` define the date of birth and the name of a person, respectively. The functional object property `hasGender` defines the Gender of the person. A person can be connected with other persons by three bonds:

1. *Familiar*, using the `hasKinship` transitive object property;
2. *Social*, using the `hasFriend` symmetric object property;
3. *Professional*, using the `hasProfessionalRelation` object property.

Those object properties are a specialisation of a more general one, named `hasRelanshionship`.

Familiar The `hasKinship` denotes a relation between people due to genealogical or marriage relations. For each concept in Figure 3.16 there is specialisation of the `hasKinship` property. For example, `hasMother`[7]. All the relations in a direct path

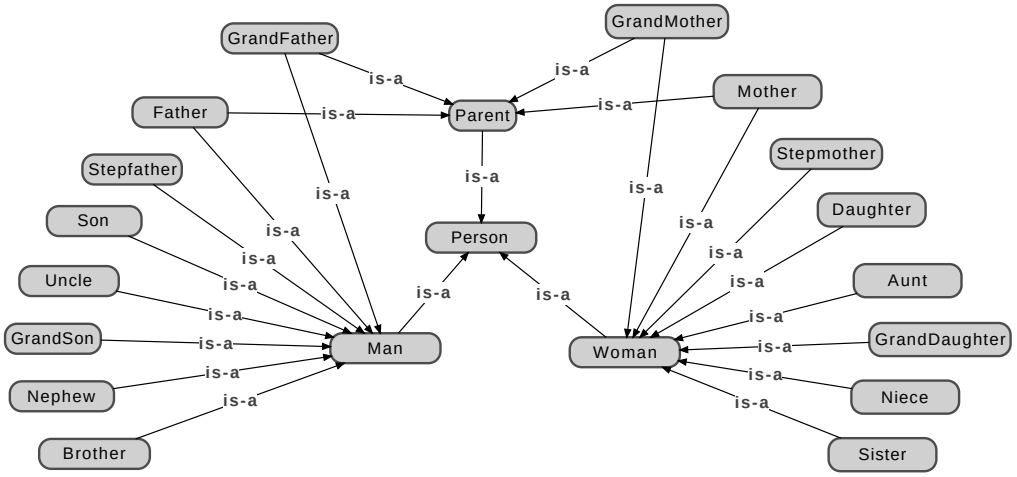


Figure 3.16: Characterisation of genealogical concepts.

of the tree formed by the concepts `Mother`, `Father`, `Son`, `Daughter`, `Brother`, and `Sister`, are permanent. This means they are *endurant* concepts. The `Ascendant` and `Sibling` are concepts derived from the ones represented in Figure 3.16, following their common semantics.

Social Friendship is a social relation that is not permanent. The life course changes the network of relationships [All08]. For example, the friends from our childhood may not be our friends today. However, in an old group photo, the concept *Friend* is appropriate to describe, partly, the social context, because it states the relation we have with those people in that particular time frame. Nevertheless, the term *Friends* can be misleading. The increasing use of social network applications makes it easier to keep "latent ties" with people who share some offline bond. The reasons people connect are varied, and

[7]For the complete list of properties, please consult `http://purl.org/mont/doc`

the word *Friend* should not be assumed as a synonymous of friendship in the everyday sense [BE08]. In **MOnt**, the object `hasFriend` is used to model all the situations where one can say he is friend with another. Although recent evidences found asymmetric friendship relations [BN13], it seems more reasonable that, in a personal photo collection, we follow the Aristotle's Ethics[8] for friendship. Thus, it is modelled as a symmetric property, a common assumption in the literature (e.g. [Bru06]). There are two sub-properties of `hasFriend`, useful to model stronger relations. One is the symmetric, irreflexive object property `hasPartner`, that models love situations between two persons. The other is the `hasSpecialFriend`, an object property that is used to model stronger relationships, stating that some person is more important that others.

Professional The professional relations are also non permanent, as they represent relations that rely on circumstantial professional factors, that may change over time. **MOnt** has two specialisations of the object property `hasProfessionalRelation`, the `hasCoworker` and `hasSuperior` properties. All the three are irreflexive relations.

3.3.1 Groups

Recent evidences [NDR10] have demonstrated that self-categorisations regulate our organisation in groups. Simply, because we feel more comfortable in coming together with fellow group members. However, one can be part of many groups, each one used to carry related activities. This is uncorrelated with the bonds we share with the in-group members. For example, we can have a group to play football and another group to watch opera. As such, social organisation goes beyond a rigid conceptualisation of relations, as showed in Figure 3.17. The `Group` is a collections of `Person`, that can be named, using

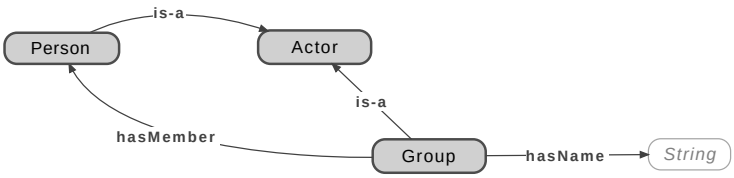


Figure 3.17: Group of persons.

the functional data property `hasName`. Each group is related to its participants using the object property `hasMember`.

3.3.2 Summary

This section has described the concepts in **MOnt** to provide the means to express a relationships between people, useful to describe the "*who*" in personal photo collections. Summing up, to deal with the social aspect of the context, we have:

[8] http://plato.stanford.edu/entries/aristotle-ethics/

- Concepts to represent personal, social and professional relationships;
- the `hasKinship` specific concepts, if blood related (e.g., `Father`), are endurant, making the assertions valid throughout any time frame;
- the `Kinship` concept non-blood related (e.g., `Stepmother`), are perdurant, making the assertions valid in a specific time frame;
- the `hasFriend` and `hasProfessionalRelation` properties that express perdurant natures. The assertions relying on those concepts are valid in a specific time frame;
- `Person` and `Group` are subtypes of `Actor`. The first represents single individual and the later represents a named collection of individuals.

3.4 Content-based concepts

Several studies from the past decade (e.g. [SWSGJ00; DJLW08]) reiterate the difficulty to derive high semantics from low-level features. The content of a photo can be described using colour descriptors [SGS10]), edge histogram [PJW00], or invariant features [Low04]. Those features provide efficient ways for computers to deal with larger sets of images, and are appropriate for specific tasks, like detecting objects or scenes. But their use is not user friendly. The terms to use should be textual, and easily understandable for the regular user. For consumer oriented personal multimedia retrieval, the solutions should be dependable, and less prone to errors [HLMS08]. Thus, the content-based concepts yield by the KB should be:

1. Absolute concepts, independent of a personal point of view;
2. Able to be generated automatically, with a very low rate of false positive;
3. Common in personal photo collections;
4. Relevant for usage in the collections.

Considering the requirements mentioned, two concepts were used, namely, **People Detection** and **Scene**. They provide information for the "*what*" and "*who*" cues used in $\mathcal{MCS}$.

3.4.1 People detection

`Person` is an important concept in a personal photo collection, not only because there are many photos depicting portraits, but also because of its value for retrieval [NH-WGMP04]. For identities, the social part of the context, discussed earlier, suffices to support the identities and the relations between people. There isn't yet a dependable solution for an algorithm to correctly identify a depicted person without a large ground

truth [CWXTT07]. The algorithms have improved, but their performance relies on users' feedback. On the other hand, torso [AMB07] and face [JLM10; LWZ11] detection are possible with low false positive rates, although sacrificing sensitivity. With that information, despite being lower level, one can use it to improve the generation of high level semantics, along with a recognition algorithm. Thus, despite the quality or existence of such algorithms, what matters is the knowledge representation in **MOnt**. As such, we model `Torso` and `Face` as concepts related to the `People` concept presented earlier. They are related with `hasPart` object properties, as showed in Figure 3.18. Capturing those

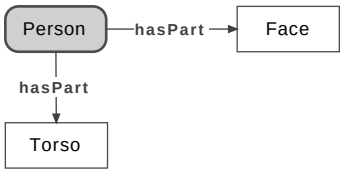


Figure 3.18: Relation between `Person` concept and some content-based concepts.

concepts provides the support needed to infer more knowledge about the photos. The detection of either a torso or a face indicates the presence of people. But it also gives hints about the type of shot. For example, if there is only one face, without a torso, the photo is probably a portrait.

3.4.2 Scene

Scene classification in photos has been extensively addressed by researchers, for example in [SP98; PS05; KPK10; JAC11]. There are common types of scenes in a personal photo collection, namely, *sunset*, a *beach*, or a *snow* scene. The scene information can help during the suggestion of activities and leverage other contextual information, e.g. spatial and temporal locations. For example, a set of photos with many scenes depicting snow, taken in the winter, could indicate a winter holiday. Figure 3.19 illustrates how the scene is modelled in **MOnt**. The concept `Scene` is described by its name, using the functional property `hasName`. A scene can happen in a `SceneLocation`, that can be of two types, represented by the individuals `Indoor` and `Outdoor`. An activity can have

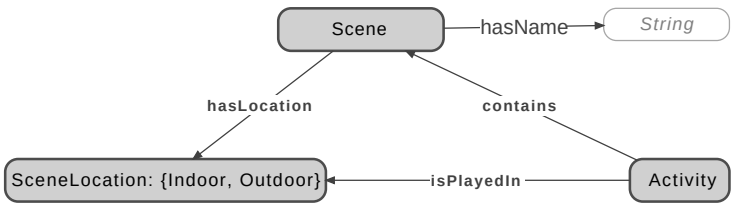


Figure 3.19: Relation between `Activity` and `SceneLocation` concepts.

scenes of a certain type, modelled using the `contains` object property. The object property `isPlayedIn` indicates where an activity can happen. However, we do not model uncertainty or preference for either sub-concepts of `Scenelocation`, when an activity can be played in both.

3.4.3 Summary

The content of the photos is not the focus of this work, since content is hard to describe at a semantics level. Therefore, the **MeMoT** implementation does not integrate content-based algorithms that contribute to refine this part of the context. Nevertheless, **MOnt** supports three concepts, `People`, `Scene` and `SceneLocation`, that complement the context of personal photo collections.

3.5 Event concepts

Several studies, as [Wag86; Con05], suggest the autobiographical memory is made of a course of actions based on some, often incomplete, remembered cues. It involves linking, contextualising, and interpreting memories trying to recreate the temporal, spatial, and social context of the remembered episode [Hab11]. Quoting Burt [BKC03],

> *"In the study of human memory, the terms episode and event, and more recently theme, are frequently used to refer to knowledge corresponding to autobiographical memory."*

The representation of events is widely discussed in literature, e.g. in [WJ07]. In **MOnt**, the concept `Event` describes something ("*what*") that happens at a given time ("*when*") and place ("*where*"), generally with people involved ("*who*"). It addresses most of the aspects described in [WJ07], leaving out the causal and experiential aspects of an event, as those are less important for people in a personal domain. As such, an event joins the four cues used to describe the context of a photo, as showed in Figure 3.20. An event also

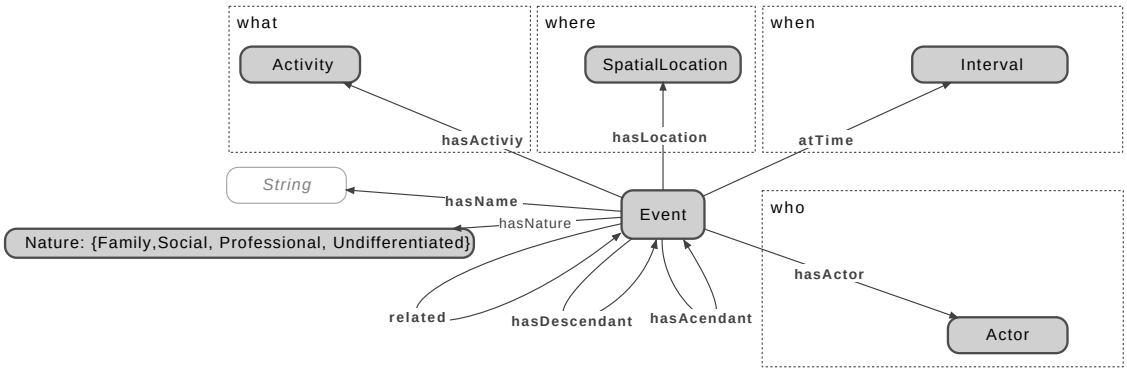


Figure 3.20: Characterisation of an event.

aggregates an *Activity*, an *Actor*, and has a *nature*.
The `hasNature` property follows some relations depicted in Figure 3.15, marking the

event as familiar, social, professional, or any combination of the three. It is not intended to be an alternative characterisation of the social aspect of the context. Instead, it intents to characterise the social groups involved in the event, complementing the "*who*" clue, much in the thematic sense that Burt refers to in [BKC03]. The `related` object property is used to connect two events, that are related. For example, to say that one event is a consequence of another. Events can be hierarchically organised. The event *Summer Holiday*, for example, may be subdivided in fine grain events like *Balloon Ride* and *Sand Sculpture*. The hierarchy is constructed using two transitive object properties. The `hasAscendant` property defines the main event, that is unique. The `hasDescendant` property defines an event that follows.

The `Activity` concept has the information about the reason of the event, addressing the

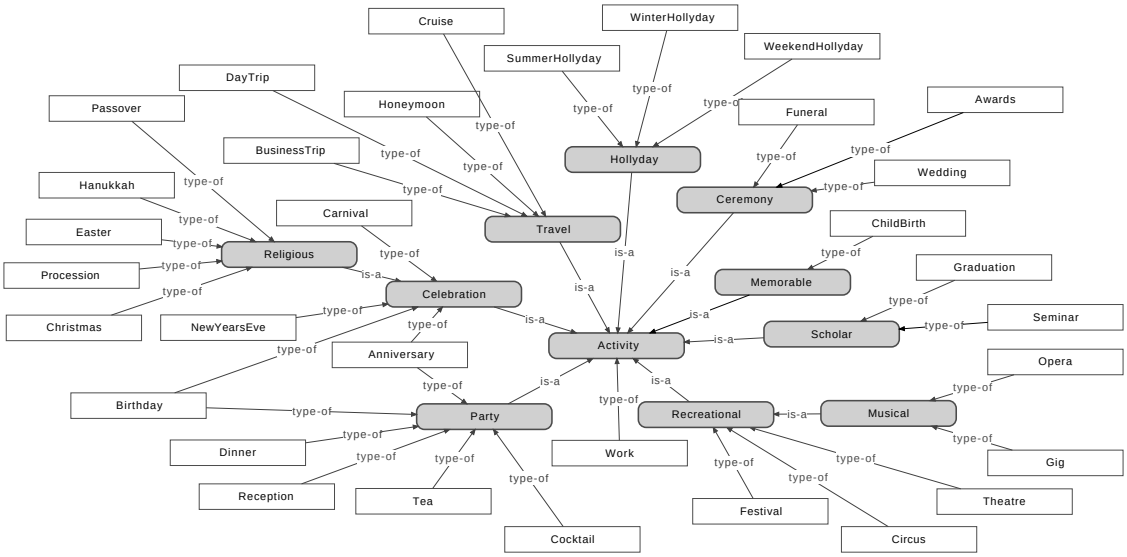


Figure 3.21: Characterisation of a situation describing the motif of an event.

"*what*" cue. It comprises a set of typical activities carried during the life course, arranged around base concepts, as showed in Figure 3.21[9]. The KB keeps several hierarchies of specialisation, providing different levels of abstraction about the description of an activity. For example, an event may be described as an `Opera` concert, but also as a `Musical` event or, more generic, a `Recreational` event. In the former example, `Recreational` act as coarse-grained categorisation of the activity, while `Opera` is the fine-grained specification of the purpose of the event. An `Activity` has some mandatory properties, as showed in Figure 3.22. It must provide a typical duration, an expected regularity and the repetition cycle. At first, it may seem strange to provide a typical duration for an activity if the event has the `atTime` object property. However, each holds different knowledge. The `when` property provides the actual duration of the event, specified in the assertions. The data property `typicalDuration` settles the usual duration of the activity for a

[9]The list of concepts is not exhaustive. It illustrates the type and scope of the activities that can be represented in the KB.

given culture, in terms of days. The `recurrent` data property indicates whether the activity has a regular schedule. For example, summer vacation is a regular activity, since it happens, most of the time, once a year. The `typicalRegularity` property tells the usual recurrence pattern of the activity, for a given culture. If the activity is regular, it can express the pattern of recurrence using a temporal location. For example, to express a daily activity, the `Day` concept is used.

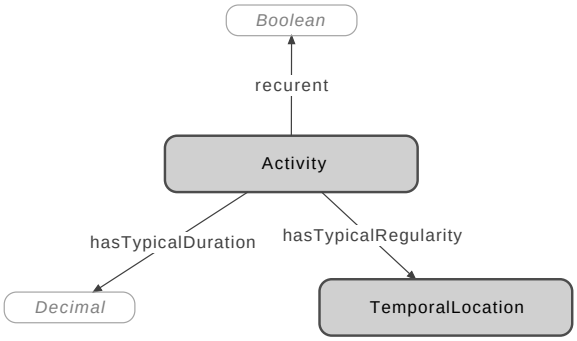


Figure 3.22: Properties of the `Activity` concept.

3.5.1 Events taxonomy

Despite the fact the representation of the `Event` concept showed in Figure 3.20 does not draw any distinction between events, not all are equally important to everyone's life [SM97]. From a knowledge representation point of view, events are classified according to their cultural and social value, following the ideas present in [BR04; Hab07; Bur08]:

1. *Life events*
2. *Global events*

Life events are those considered to be normative by a specific culture and thus, expected to be experienced by everyone. They are believed to be social and culturally important [BR04]. They also have an expected time frame, although within imprecise boundaries. For example, the event "*Primary education*" is a normative event that is compulsory to all children, in most of the western world. It covers a period that starts at the age of 5–6 years and goes on to the age of 10–11 years. Besides, it is expected to be experienced once in a lifetime. Figure 3.23 shows the properties of life events. The `atTime` object property refines the inherited property from `Event`. `LifeStage` are intervals that model observable changes in individuals during life, essentially based on biology. As expected, they go from birth to death. The life stages' instantiations are an adaptation of the classical Greek 7 stages of life to today's reality [Arm08]. Besides their temporal location, life events are expected to follow each other in a given order, forming what is often called a *live script*. Life Scripts are representations of a course of actions, denoting the expectation in the order and timing of life events, according to an ideal life course for a specific culture. These

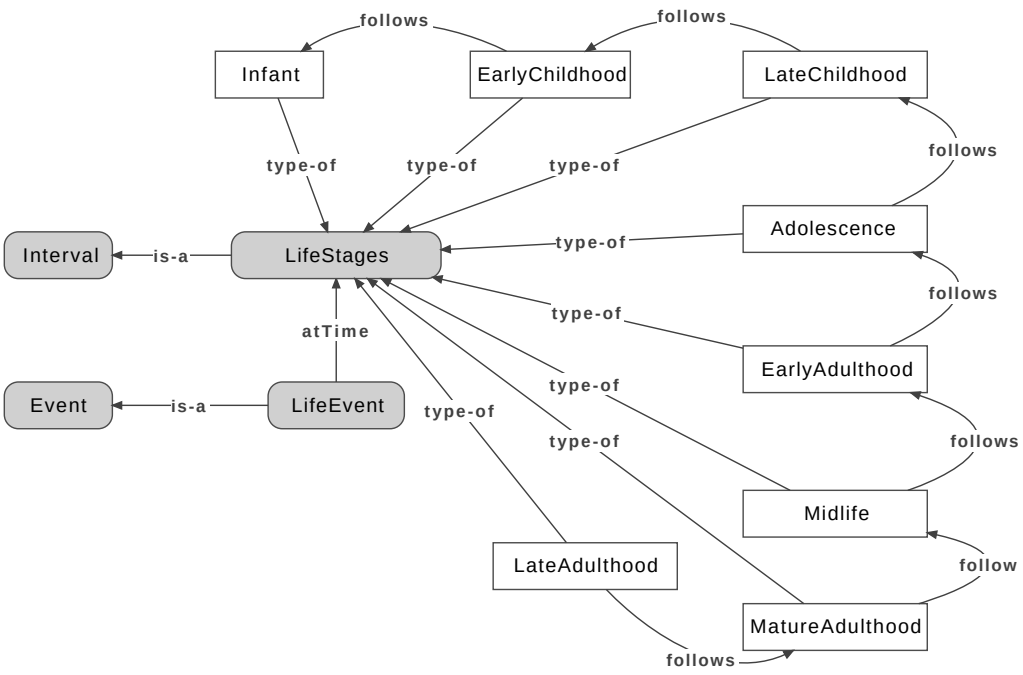


Figure 3.23: Properties of life events.

pre-defined scripts enable us to reason about events in a cultural level [BR04] and allows people to position an event in time [Bur08]. The life script is represented in the KB by a set of life events connected by a `follows` property. As an example, Figure 3.24 shows a possible life script for the western civilisation. The `Primary school` and `Retirement` events are the limits of the sequence. Do notice that life scripts are an ideal course of ac-

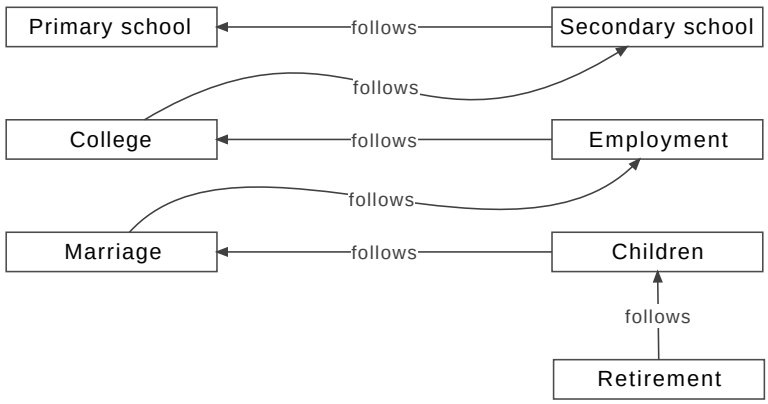


Figure 3.24: Example of a life script.

tions, and that the life events depicted in a photo collection may have a different temporal location and order. Such events, if stored as assertions, produce what is called a *life story*, that is unique to everyone. Nevertheless, life scripts are useful, not only because they settle landmark events that aid the remembrance of other past events, but also because they are a culturally shared part of our semantic knowledge [BR04]. Therefore, they are

a property of cultures themselves. This can be used to enhance the suggestion of annotations, but also to improve algorithms accuracy, by removing unlikely events given the current time-frame and assertions.

Global events are important non-normative events in each one's life story. They are much more personal, with less cultural weight. Examples of such events are "*Death of a relative*", "*Leaving home*", or "*Participation in Olympic Games*". They are usually non-recurrent events with an expected time frame to occur. This means that most of them have an age norm, meaning that a global event has the properties depicted in Figure 3.23.

3.5.2 Summary

In this section, we presented the *Event* concept as an aggregation of the 4*Ws*. Since the KB role is to provide the meta concepts from which we reason about the annotations in the photos, there are some key features that are collection wide and need to be represented. Summing up, the KB guarantees:

- a representation of the `Activity` concept, delivering a subsumption relation for several types of activities;
- that each activity instantiation must have an indication of its expected regularity and repetition cycle, or is otherwise marked as non-regular;
- that each activity instantiation must have its typical duration;
- that each `Event` can be related to an activity instantiation, using the `hasActivity` property (the "*what*");
- that each `Event` can be related to a spatial location, using the `hasLocation` property (the "*where*");
- that each `Event` can be related to a time interval, using the `atTime` property (the "*when*");
- that each `Event` can be related to an individual or groups of individuals, using the `hasActor` property (the "*who*");
- that each `Event` can be marked as familiar, social, professional or undifferentiated event, using the `hasNature` property;
- a categorisation of different events into different types, representing landmark events;
- the `LifeEvent` concept to support life events as normative events, providing expected temporal locations for important cultural happenings;
- the support for life scripts, as cultural dependent sequences of cultural happenings, modelled as `LifEvent` individuals related by `follows` properties;

- the `GlobalEvent` concept that represents non-normative personal important events;
- the `LifeStage` concepts, for supporting intervals to locate the life and global events.

3.6 Semantic Viewpoints

The assertions in the KB represent a perspective of the state of the world, that although correct, is not unique. This is because we conceive what surrounds us, objects, properties, to name a few, as reachable from other points of view, different for our own [Gru00]. In this work, the state of the world are assertions about the context surrounding sets of personal photos. Lets assume one wants to query about photos where the context includes the terms Summer and father. The query can be "photos from last summer with my father". The set of photos satisfying such query can also be reached using a different formulation, like "photos from last summer with my *uncle*". In this case, the person issuing the query is my cousin. What changes is the representation of the same person. The terms father and uncle refer to the same person, depending on the observer. But the same set of photos can also be reached using the query "photos from last winter with my *father*". In this case, the observer is my brother who lives in a different hemisphere. Thus, the temporal reference is different, so the same interval in time — summer — is referred to using different words, that represent those alternative representations. In the example above, it is necessary to build a query context — the year, the person issuing the query — to be able to determine what last summer means, and the identity of the person called father (or uncle). We call this semantic viewpoints. In other words, a semantic viewpoint is a user vision of the world and so, the interpretation of the facts are user dependant. A user vision of the world can be stated as an *egocentric* one that sees facts from a particular point of view, exists in a *allocentric* context, and is independent of that point of view [Gru00]. The terms *egocentric* and *allocentric*, follow the definition proposed in [Kla98], for representations *self-to-object* and *object-to-object* respectively. The context is settled by several contributions from a network of people connected by several bonds. Admitting that we do not have any reference, every assertion is independent from its creator and is part of the whole. However, to settle the proper semantics, we need to choose a point of view, turning our vision of the context egocentric (or *self-to-object*).

3.6.1 Transformations

There are three transformations that can be used to settle the proper viewpoint:

1. egocentric → allocentric
2. allocentric → egocentric
3. egocentric → egocentric

All possibilities are necessary, as they complement each other, but their use depends on the available assertions.

Transforming an egocentric cue is possible when the semantics of the cue can be determined using the KB, and there is an alternative allocentric cue. This is not always possible, since the knowledge in KB is incomplete, and represents only a part of the possible assertions of the collection. Similarly, transforming an allocentric cue is possible if the semantics is known and if there is an alternative representation, specific to the user's point of view. To support such operations, the concepts stored in the KB need to be marked as *relative* or *absolute*. For many domain relations, e.g. social relations, the KB has the necessary predicates to support transformations. However, to fully cover the viewpoint adjustments, it is also necessary to express the equivalence between concepts, and under what circumstances it occurs. This is done using `DualConcept`, as showed in Figure 3.25. The `hasObject` relation indicates the left side of the equivalence, and the `hasSubject` relation expresses the other side. The `hasPreCondition` and `hasPostCondition` relations express the coherence conditions of the duality, for the left and right side, respectively. If there is more than one condition in each side, the semantics is the conjunction of the conditions. A `Condition` can be generic or it can be narrowed to be interpreted in the user context (`SelfCondition`), the owner of the photos context (`OwnerCondition`) or to the annotator of a certain assertion (`AnnotatorCondition`). In the cases where the condition refers to a specific entity, the `refersTo` relation can be used. A condition has a `DomainRule` that encodes a specific rule, that is implementation dependant. For example, it can express the value of a certain property of something outside the KB, but known to the system. Do notice that `DualConcept` is not self-dual,

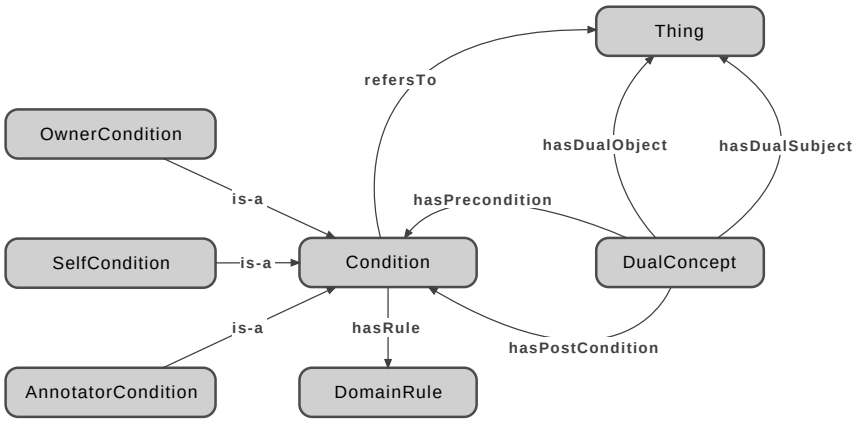


Figure 3.25: Dual concept representational needs.

meaning that *A dualOf B* $\not\equiv$ *B dualOf A*. However, a `DualConcept` with no conditions is

self-dual and it is the same as to express that *A sameAs B*. The *DualConcept* is a mapping concept. It addresses the transformation needs in **MeMoT**, in a simple manner. However, if more complex transformations are required, the concept can be extended in ways that are discussed in [MMSV02; Euz04].

3.6.2 Spatio-temporal concepts

Most of the temporal concepts are absolute. For example, the concepts shown in Figure 3.9 related to the day cycle retain their semantics independently of the user or its location. A photo taken at *Midnight* is described as a *Midnight* everywhere. However, there are some concepts that depend on the observer, namely, `Season`. For example, the term *Summer* should match the term *Winter* if the spatio-temporal references for two different users from different hemispheres. They are relative concepts, and the transformation is egocentric $\rightarrow$ egocentric. Using the `DualConcept`, this duality is expressed assigning `hasDualObject` to `Summer`, `hasDualSubject` to `Winter`, `hasPreCondition` to a `SelfCondition` containing the rule "North", and `hasPostCondition` to a `Condition` containing the rule "South". This expresses that *user is from North and Summer*$\equiv$ *Winter if South*.
The spatial concepts are absolute, but the same place can be named differently by two people, denoting preferences or different visions of the world. Thus, it may be necessary to use the `sameAs` relation to denote the preferable naming depending on the user who made the assertion. An example is the `Yugoslavia` and `Serbia`, both individuals of `Country` concept, where the former was, in the 20^{th} century, a country that includes the area occupied by Serbia. Since it no longer exists, but the inhabitants may use the old reference, it represents an egocentric $\rightarrow$ allocentric transformation.

3.6.3 Social concepts

Most of the social concepts are relative, since they express relations between people. The exceptions are the `Woman`, `Man` and `Person` absolute concepts. The relative references are transformed from egocentric $\rightarrow$ allocentric, generating an absolute representation of the object according to a specific viewpoint. For instance, the term *brother* becomes `Carlos`, when Bob is querying for his brother' photos. While it is not guaranteed that absolute references are unique in the KB, the personal nature of the photo collections minimizes clashes. Name collisions are already dealt within the family and social relations. The way we refer to each person we know, whose importance is such that we keep photos of them, tends to be unique. The multiple names that a person usually has in some circles, can be represented with the `sameAs` relation. Thus, the terms in the KB present the same characteristics, as they mimic life.

3.6.4 Content-based

The content-based dimension needs no work in the viewpoint transformation. The terms used are absolute, by design, and thus independent of any semantic viewpoint.

3.7 Implementation details

The knowledge base consists of two major parts:

1. an ontology, named **MOnt**

2. a relational database (RDB)

The first is used as a metadata repository, holding assertions that are collection wide, independent of a specific collection and can be reused in multiple contexts, for example, the terms `Day`, `New York`, and `Father`, to name a few. The second stores the data about photos, following the semantics determined by the first. Figure 3.26 illustrates those parts and how they are connected. This architecture separates certain types of assertions.

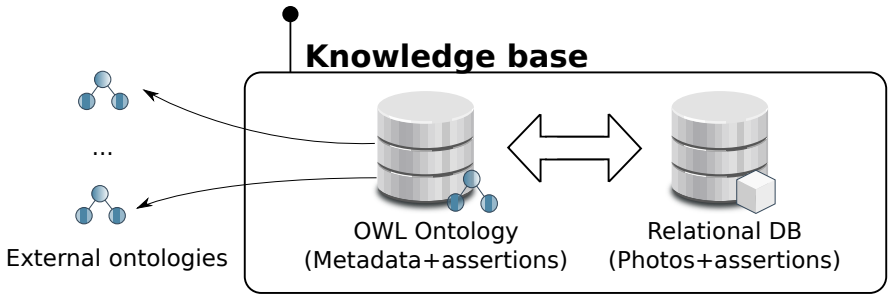


Figure 3.26: The KB architecture.

The assertions in the ontology are only for concepts and individuals that can be reused collection wide. This means the ontology, for example, will not hold assertions of the latitude/longitude of every photo in the collection, nor any assertion referring to a specific photo. On the other hand, the RDB will hold the photos and their specific metadata, following the semantics stored in **MOnt**. All assertions, either in the ontology or in the relational database, reference the identity of the person who made it. Whenever a user complements the metadata of a photo, the terms used are taken from the ontology concepts, creating new assertions attached to the photo, that are included in the RDB and in **MOnt** as instantiations of those concepts. If the terms are unknown, they are inserted as textual tags in the RDB only.

3.7.1 Ontology

The ontology that is part of the knowledge base, called **MOnt**[10], is developed in OWL 2.0 [W3C09], encoded using RDF/XML. More precisely, **MOnt** uses the OWL DL subset, of the $\mathcal{SHOIN}$ family, which are decidable fragments of first order logic [HPS03; KSZ07]. Since OWL has a well-known limitation of allowing only binary relations between classes [SFE11], some concepts (e.g. `Conditions`, `DualConcept`, and `OrderedElement`) were implemented using reification, explicitly expanded has already shown in Figures 3.25, 3.8, and 3.4, to implement n-ary relations [NR06]. For example, `OrderedElement` represents an n-ary relation between `Cycle` and a `TemporalLocation`, where the position (index) is associated to the relation. The concepts in **MOnt**, whenever possible, where defined using sufficient and necessary conditions, making them fully defined classes. For example, the concept `Woman` is defined as Woman $\equiv$ Person $\sqcap \exists$ hasGender {Female}. However, the major concerns during the development of **MOnt** was to achieve an ontology *(i)* decidable, *(ii)* with performance, and *(iii)* interoperable.

The current implementation of **MOnt** holds a representation of concepts for the Western culture, using a set of terms that are common for a wide range of people. The design is agnostic in terms of sex, race, religious or other creeds. This means that, although it covers a wide range of situations, it will only provide basic support for a a suitable representation of the social environment in many of the cultures in the world. Nevertheless, key extension points are available. As a final remark, no internationalisation issues were addressed in the implementation of the ontology, as this was considered out of the scope for building a proof-of-concept prototype. For the complete concepts and properties, please consult the online documentation, available at `http://purl.org/mont/doc`.

Decidability Although OWL-DL is decidable, some authors (e.g. [HST00; MSS05]) have reported that using certain types of constructions can lead to undecidability. One of the cases is the usage of number restrictions with inverse roles. Since **MOnt** contains no number restrictions, it is therefore decidable from that point of view. Another is the usage of a combination of OWL-DL and rules. **MOnt** uses this type of combination, especially to implement people relations, discussed on page 45. Thus, to maintain the decidability of the ontology, DL-safe rules were used [MSS05], expressed in terms of SWRL [HPS-BTGD+04].

Performance OWL-DL complexity is known to be NEXPTIME-complete in both concept satisfiability and ABox consistency [HPS03]. However, in [BCMNPS03] the authors stated that reasoner's implementations have demonstrated they can perform well in realistic applications with high worst-case complexity. In fact, such situations rarely occur in real world. In **MOnt** the number of concepts is relatively small, less than five hundred. The only unbound number of assertions concerns the individuals in the ABox.

[10]The ontology is available at `http://purl.org/mont/mont.owl`

However, they are related with places, people and activities present in the collections, which won't grow large enough to cause performance problems. Nevertheless, there are some important performance issues, because of the way **MOnt** is used. In Figure 3.26, we see the ontology is used in conjunction with the relational base, and acts as a metadata repository that drives the insertion of new knowledge about the photos. This is a typical *read-write* scenario. Since we want to maintain the ontology consistent, all assertions need to be verified one by one, retracting the faulty ones[11]. For example, inserting a new assertion telling that John is his own father will lead to inconsistency since `hasFather` object property is irreflexive, as thus is retracted. This forced us to compute consistency for each new assertion. Even though the reasoner used, the Hermit OWL Reasoner[12], is based on the hypertableau calculus [MSH09], it leads to poor performance dealing with new assertions for thousand of photos during archiving. So we transform the insertion of assertions into an asynchronous operation, decoupling the producer (the application) from the consumer (the data store), using a message queue. This solution provides **MeMoT** with faster response times when archiving large photo sets. The overview of this setup in shown in Figure 3.27. Do notice the KB needs to be online to serve queries, so **MeMoT** maintains a cache of a stable version of the ontology, that gets updated given a predetermined, configurable, criteria. The current implementation refreshes the cache at the beginning of a process (archival or retrieval).

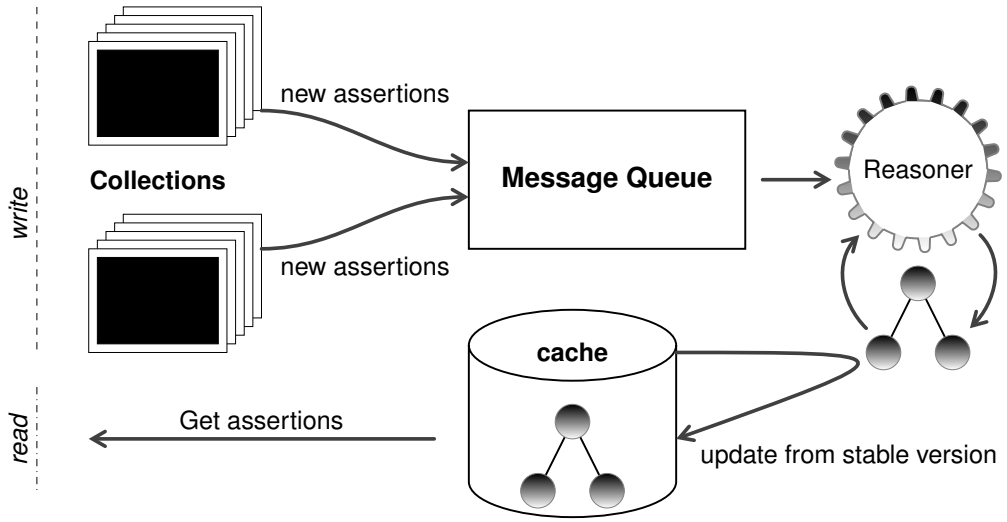


Figure 3.27: Decoupling insertion and consumption of assertions.

Connections to other ontologies In **MOnt** we reuse concepts from different ontologies, to improve the ontology matching. We use the SKOS [MB09] concepts `broadMatch`, `closeMatch` and `exactMatch` to link to external concepts. Although the following description is not exhaustive, it shows the main bridges between this work and the research done in the area, for some of the core components on **MOnt**. The notion of event, as

[11] Dealing with updates to the KB using contradictory facts were out of scope of this work
[12] `http://hermit-reasoner.com/`

described in page 50, is a central concept in a personal photo collections. In **MOnt**, it builds its foundations on the class `mont:Event`, that is linked with the same concept present in many ontologies and vocabularies. It is a closed match to the event concept exposed on LODE [STH09] and on Dublin core [DCM09]. It is also related to SEM [VH-MVSS12] and the event ontology [RASG07] as a broad match. To express its perdurant nature, it is linked with the `dolce:perdurant` class available on DOLCE lite [Gan06], and with `bfo:occurent` of BFO [SG02], using `skos:closeMatch`. We also reuse the notion of spatio-temporal entities, that plays an important role in this domain. As such, we related the `mont:event` with the concept `bfo:spatiotemporal_region`, using `skos:broadMatch`.

Since an *event* is modelled as an aggregation of the 4*Ws*, we also reuse concepts from other ontologies. The "*what*" is modelled as a taxonomy of activities that starts at the class `mont:Activity`. Such class is an exact match of the OpenCyc [Ope] concept `cyc:HumanActivity`.

For the social part of the ontology, that describes the relations between people, we use FOAF [BM10] vocabulary. The class `mont:Person` and the data property `mont:hasName` are exact matches of `foaf:Person` and `foaf:name`. The object properties `hasRelationship`, and its sub-properties, has a broad match with `foaf:knows`. The personal connectors that we used are similar to those found in [CZJ08] and in family ontology[13]. We also include a broad match between `mont:Person` and `mpeg:PersonType` of MPEG7. The `mont:Group` is a broad match with the `dul:Actor` concept from DOLCE+DnS Ultralite ontology [Gan07].

The location of an event, illustrated in Figure 3.2, uses concepts from the geonames ontology [Geo12]. The class `mont:Place` is an exact match with `geonames:Place` and a close match with `geonames:SpatialThing`. Geonames lacks typification for `City`, `Countries`, and all other concepts that **MOnt** uses to describe a location. Nevertheless, the `skos:broadMatch` is used to match them with the `geonames:SpatialThing` class. In particular, the `mont:POI` is a close match with the `geonames:SpatialThing`. All spatial classes are connect with the `dolce:spatial-region` and `bfo:two-dimensional-region` using `skos:broadMatch`.

The temporal classes are related with the time ontology [HP06]. The `mont:TemporalLocation` is an exact match with the `time:TemporalEntity`. The `mont:Instant` and `mont:Interval` are also exact matches with their counterparts in the time ontology. The helper structure `Mont:OrderedElement` follows the *Ordered List Ontology* (olo) [AF10]. The properties `Mont:index`, `Mont:hasNext`, and `Mont:hasPrevious` are an exact match with `olo:index`, `olo:next`, and `olo:previous`, respectively. The properties `Mont:hasLength` and `Mont:hasDepth` are an exact match with `olo:length`. The concepts `Mont:OrderedElement` and `Mont:OrderedStructure` are an exact match with `olo:Slot` and `olo:OrderedList`, respectively.

[13] Available at TONES repository, http://rpc295.cs.man.ac.uk:8080/repository/download?ontology=http://www.mindswap.org/ontologies/family.owl&format=RDF/XML

3.7.2 Relational database

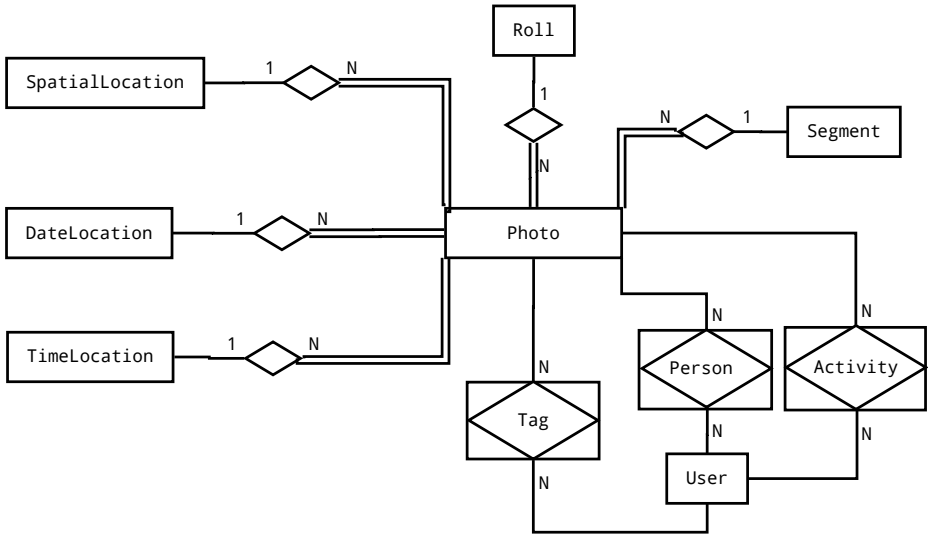


Figure 3.28: Data model of the relational database, responsible for supporting the $\mathcal{MCS}$.

The relational database implements the $\mathcal{MCS}$ and so it contains the information for all photos present in **MeMoT**. It includes the url of each photo and all the available descriptors for the 4*Ws*. The implementation was carefully crafted, to support two important requisites in this domain: *(i)* capability to handle large collections of photos, *(ii)* and performance during retrieval. The solution was inspired by the methodology used in data warehouse systems, that has proven over the years its adequacy and efficiency in dealing with high volumes of data [KR02]. The data model follows the star schema approach, as illustrated in Figure 3.28. Each photo is represented as factual data (dark rectangle), that is described by several dimensions containing human understandable descriptors (white rectangles). The arrows indicate a foreign key integrity constraint. The richer the descriptors are, the more expressive the description of the photos is. Generally, such expressiveness is desirable, as it accommodates changes in the analytic requisites, that, in this domain, we can transpose to changes in the way users retrieve photos. The descriptors for `Roll`, `Segment`, `Spatial`, `Date` and `Time` are unique for a photo. For example, a change in the spatial location will change the location of the photo in the $\mathcal{MCS}$, overriding the previous value. Although this is true for the knowledge base data, we do not change the original photo's metadata. This allows us to reconstruct the descriptors for the original data, if necessary. The spatio-temporal information of a photo is unique and independent of any viewpoint. Although this is true for the coordinates, a certain person may prefer different textual descriptors for some locations. To maintain coherence between the metadata, the spatio-temporal descriptors and the real world, we cover this necessity by relying on a tagging approach.
The descriptors for `Tag`, `Person` and `Activity` can be associated to many photos, and

are viewpoint dependant. The things you see in photos and the way we interpret them are asserted this way in the $\mathcal{MCS}$, providing information about the users' viewpoint. Given the nature of the domain, and supported by the literature (e.g. [WBC10]), after the organisation of a set of photos, new interactions for annotations are rare. Thus, we derive non structured, denormalised information from the assertions of tags, persons and activities, for indexing purposes. For example, the photos will have a field containing the concatenation of all the tags. This improves the performance during retrieval.

3.8 Summary

In this chapter, we introduce the information considered necessary to describe the context surrounding a personal photo collection. Besides the concepts related with the 4*Ws*, that are common in this research field, we also include social and cultural information, towards a better coverage of the events and activities that exist in those collections. To our knowledge, it is the first time that *life events* and *life scripts* are used in this domain. Since personal photo sets, being private, exhibit a narrower variance in terms of information needs than public collections, it is possible to use domain knowledge to support a better manipulation of the user assertions. Using a knowledge representation that supports reasoning, we provide mechanisms that offer controlled vocabulary, that can be used to suggest annotations using prior-knowledge and inference. Besides the annotation purpose, domain knowledge enables the manipulation of vocabulary towards the vision of the world of each user that we referred to as *semantic viewpoint*. We also separate the domain knowledge, common to all photo sets, from individual assertions, specific to each photo. The first are stored in an ontology that acts as a meta-repository for the second. Individual assertions are stored in a star-schema like data store, called $\mathcal{MCS}$, tailored to support large photo sets. The design of the data repository follows the solutions developed in the data warehouse field. In particular, the use of different levels of detail for the information, that enables the choice of the proper level that matches the semantic needs of the user.

In the next chapters we will use the information and the concepts discussed here to support the archival and retrieval of photos in the **MeMoT** system.

4

Archival

> *If you wish to forget anything on the spot, make a note that this thing is to be remembered.*
>
> ***Edgar Allan Poe***

The primary function of photography is to capture our family life, and to attach us to our own past and to the past of our social groups [Gye07]. So, we can see photos as out-of-self-memories that share characteristics with autobiographical memory, especially in what concerns time. When people take photos, they leave in the support medium a set that spans through different time periods, depicting several personal and social events, but also photos related to the ordinary day life. For archival purposes, the photos are manipulated in sets, sometimes referred to as *rolls*. Even in the absence of an explicit transfer of the photos to repository, as in a cloud services like MyShoeBox[1], we can think of a roll as a set of photos uploaded in one session.

Definition 4.1 (Roll). *A* roll *is a set of photos that are manipulated together for archival purposes.*

Rolls can be divided into several subsets, each representing a different *event*. The division is intended to be done automatically based on the metadata of each photo. As in autobiographical memories, events are assumed to occur one after the other [Bur08].

Definition 4.2 (Event). *An* Event *is a happening that spans thorough time and space. Events form a sequence of happenings that do not overlap.*

[1] http://shoeboxapp.com

Events embody the seasonality that is present in our life. They are controlled by the natural cycles of days and years, but also by the rhythm of the weekly cycle, with more social meaning than the formers [Zer85]. Notice that time is one of the most important dimensions of the social world [ZS93]. However, photos are difficult to interpret outside the context where they were taken. In fact, the meaning of each photo is closely related to the purpose of the shot. Besides, content semantics are complex to be formally described in a complete (or unambiguous manner), and need information from context to settle a precise interpretation. That is where the user plays an important role, inserting metadata and increasing the value of the collections. Those handmade contributions occur in different moments, namely:

- in the catalogue process
- during image retrieval

The first time users have the opportunity to insert new metadata is during an explicit catalogue process[2]. However, annotating each photo is not an option, because of its time-consuming nature. This poses an interesting challenge to researchers on how to motivate users to do so [Han08; VA06; KS05], but because they are based on crowdsourcing, their solutions are not adequate to personal collections of photos. Knowing beforehand that some photos share a context allow us to create groups that enable the insertion of metadata in batches, reducing the manual labour [CWXTT07; YNC09].

These rolls will be subject to the archival process, showed in Figure 4.1. It consists of the following steps:

1. Segment the roll, so that the photos inside each segment share a similar context;
2. Enhance the information associated to each photo or event, for each axis of the multidimensional context-space, using the KB terms;
3. Insert the photos in multidimensional context-space and update the KB with the new assertions.

This chapter describes the archival and annotation of rolls. It presents the *Logical Day Event Segmentation* (LDES) algorithm that segments a collection of photos using their spatio-temporal information:

1. An automatic segmentation algorithm that uses the spatio-temporal information present in the photos, without user intervention;
2. An approach that considers the temporal cycles that govern our lives, especially the day cycle;
3. A framework for comparing different segmentations of the same set of photos.

[2]Recent cameras support annotations when the photos are still in the medium. However, this is an exception, and not the rule.

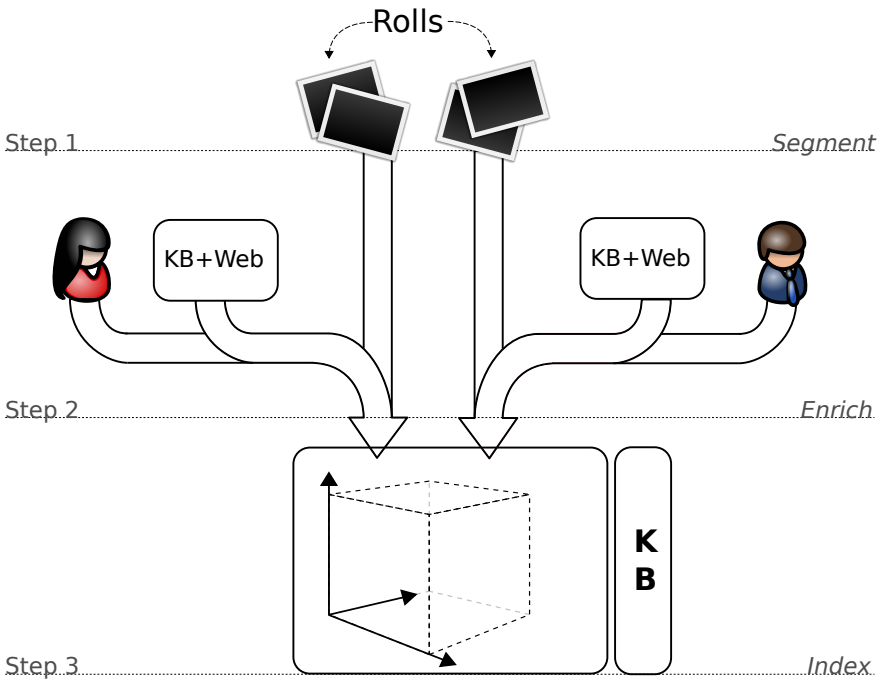


Figure 4.1: Overview of the archival process..

The segmentation of a collection of photos is, at its core, a time segmentation problem. The minimum requirement to divide a roll is that each photo is timestamped with its creation time. We assume the timestamps are in local time and the error is less than one hour. Later, other types of metadata, namely, spatial coordinates, can be used to fine-tune the boundaries of each segment. In this chapter, the problem is formalised and some tests are presented, showing the performance of the algorithm and its capacity to tackle the problem.

4.1 An interface for archival

One of the archival goals is to increase the available semantic information for the collection, an important tool to deal with sensory and semantic gaps. However, the way the photos are presented to the user, the simplicity of adding or changing suggested annotations plays an important step towards the user acceptance of the solution. The interface should "disappear", concentrating the user's attention on *what to do* and not on *how to do it*. Although the focus is not on the interface, the developed prototype tries to accomplish that goal, using a minimalist design[3]. Today, and probably in the near future, cloud based services will hold many of our digital assets, photos included [OSHT12]. The explicit transference of photos will probably be replaced by a background uploading procedure, dismissing the need for an interface like the one in Figure 4.2. However, the importance of an archival step is such, that it will hold for the benefits it provides.

[3]Information about the the prototype is available at `http://goo.gl/1Bs9cc`

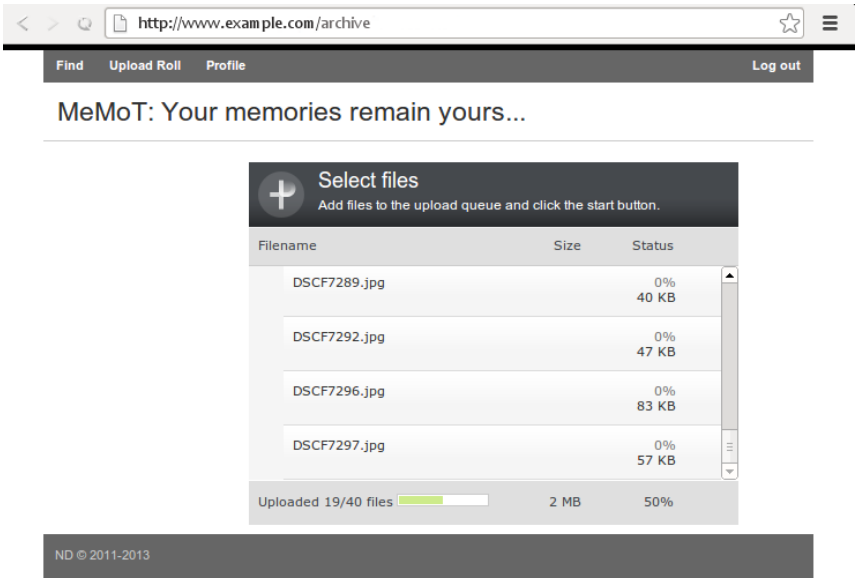


Figure 4.2: Uploading interface implemented in the prototype.

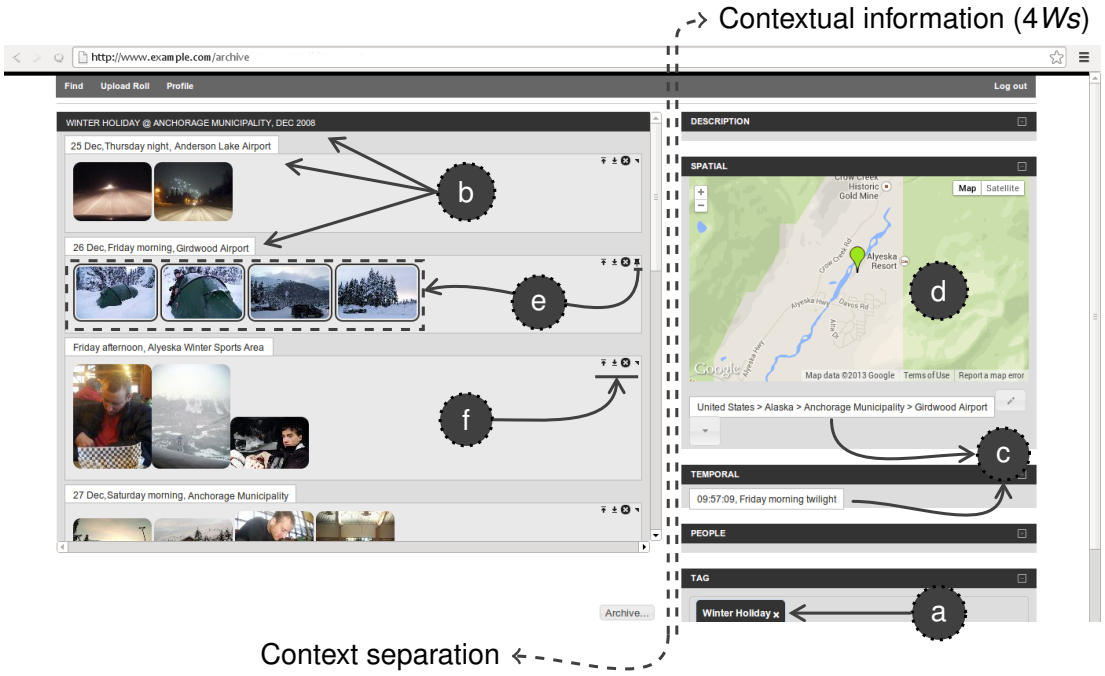


Figure 4.3: Interface for archiving a set of photos.

The user interface (UI) depicted in Figure 4.3 is used to archive photos in **MeMoT**. It has two distinct zones, both supporting some automation on behalf of the user. On the right hand-side, users see their photos, properly segmented, based on temporal and spatial context similarity. The algorithm used to perform the context separation is the LDES, that will be discussed in the next sections. On the right hand-side, the interface shows contextual information for the 4*Ws*. Whenever it's possible, such information reflects the state of what is selected on the left hand-side. However, what is really important is the

ability to automate and suggests information based on the **MeMoT**'s current knowledge and on the context of the photos being archived. In Figure 4.3, labelled with ⓐ, we can see an example of an automatic suggestion for an activity description. In this particular case, the source of information was the information present in **MOnt**. For example, for the first two photos, besides the `Winter Holiday` tag, the interface also suggests `Christmas`. Another example of automation is the annotation for the roll and for each group of photos (labelled as ⓑ). These descriptors are generated using the same principles as discussed in Chapter 5 for the MSS summarisation algorithm. The contextual information, namely, the spatial and temporal information, is presented to maximize the information for the user. Either using visual feedback about the location of the photos (labelled as ⓓ) or, displaying qualitative descriptors with the support of the **MOnt** (labelled as ⓒ). Once again, the right hand-side is supposed to summarise the information of whatever is selected on the left hand-side.

Users have the opportunity to change the information that is displayed. This includes segmentation, suggestions, and spatial information. The only exception is temporal information, that we consider as correct [4]. The user can select one or many photos and change the annotations, as illustrated in Figure 4.3, ⓔ. Changing the proposed segmentation is done by dragging the photos from one segment to another, or by splitting or joining segments. The user can an also add or remove segments from the segmentation. For those actions, there are specific buttons (labelled as ⓕ).

The archival interface follows some of the desired features from Human Computer Interaction (HCI), namely: *(i)* has affordances (e.g. Figure 4.3, ⓕ); *(ii)* provides feedback on the user's action (e.g. dragging the marker changes the spatial summary); *(iii)* prevents users from making errors (e.g. preventing changes in the temporal order of photos). For example, the drag and drop facility is available for modifying the segmentation. However, the temporal order of the photos is held, i.e., if the user drags a photo to the next segment, all the following photos are also moved.

4.1.1 Setting the context

Citing Dey, "Context is all about the whole situation relevant to an application and its set of users." [Dey01]. In this work, the relevant information is the terms that people use to refer their memories, organised along the 4*Ws*. As such, context is the set of most specific terms that are used to describe a happening. This happening can be, for example, a photo, where the context is the description of that particular moment that was captured. But it can be a segment, where the context has the level of detail that is suitable to describe all the photos it contains. The set of terms used to characterise the context is used to position a photo in the $\mathcal{MCS}$. For example, a photo will occupy different positions in $\mathcal{MCS}$, if the the spatial part of the context of a photo is described using different *LoD*, like `USA/New York/Manhattan/Broadway` or `USA/New York/Manhattan`.

[4] The support for changes in temporal information is outside the scope of this work.

The left-hand side of the interface, illustrated in Figure 4.3, presents a context separation at the segment level, and on the right-hand side, there is the context description, organized along the 4*Ws*. When a user archives a set of photos, the segments' structure and the photos specific information are inserted in the relational database. Besides, the terms on the right-hand side are store in **MOnt** as assertions of granules of information to describe the context. Let us use as an example, the roll illustrated in Figure 4.3. Considering the selected photos, identified by , the following assertions are made to the ABox:

1. `2008`, of the `Year` concept;
2. `United States`, of the `Country` concept;
3. `Alaska`, of the `Region` concept;
4. `Anchorage Municipally`, of the `City` concept;
5. `Girdwood Airport`, of the `POI` concept.

The terms `Friday`, `Morning`, `Twilight`, and `Winter Holiday` are already known to **MOnt**, as they are included in the domain knowledge. Once those terms are known, they can be suggested in posterior interaction with the user. In the relational database, additionally to the segment information, each photo is stored along the spatial coordinate, and the terms that describes the context of the photo. Those terms are the ones that exist in **MOnt**.

The main support of the archival is the segmentation algorithm, that enables the batch insertion of annotations. In the next sections, we address this problem, introducing the LDES algorithm.

4.2 The segmentation problem

In this section, we formalise the notions of segments, segmentations, and a set of relations between segmentations. Those formal definitions will support the comparison of different segmentations for the same set of photos.
Let P be a sequence of N photos, ordered non-descendingly by their creation timestamps. P is represented by the pairs (t_n, g_n), where t_n is the timestamp for photo n, and g_n is the spatial location for photo n, or, if the location is unavailable, $g_n = null$. Without loss of generality, for segmentation purposes, we represent all photos taken at the same second (the time grain available in EXIF) as a single instant, thus

$$\forall n \in [1..N-1] : t_n < t_{n+1} \tag{4.1}$$

Let $T = [t_1, \ldots, t_n, \ldots, t_N]$ be the sequence of t_n in P, satisfying (4.1).

Definition 4.3 (successor in T). *An element $t_j \in T$ is the successor of $t_i \in T$, denoted by $(t_i)^{\succ} = t_j$, when there is no element in T between t_i and t_j, and thus $j = i + 1$.*

The elements in T can be arranged to form non-empty, temporal contiguous sub-sequences of T, leading to the notion of segment.

Definition 4.4 (a segment). *A segment, denoted by $s = [t^-, t^+]$, is a non-empty, sub-sequence of T. t^- is the lower limit of the segment and t^+ is the upper limit, where $t^- \leq t^+$, holding*

$$\forall t_n \in T\ ,\ t_n \in s \Leftrightarrow t^- \leq t_n \leq t^+$$

Definition 4.5 (element of a segment). *An element $t \in T$, is an element of a segment $s = [t^-, t^+]$, denoted $t \in s$, if and only if $t^- \leq t \leq t^+$*

Notice that a segment can be singular, when $t^- = t^+$, and can be equal to T, when $t^- = t_1$ and $t^+ = t_N$.

Definition 4.6 (a segmentation). *Given $T = [t_1, \ldots, t_n, \ldots, t_N]$, a segmentation S, represented by $S = [s_1, \ldots, s_k, \ldots, s_K]$, is a non-empty ordered set of segments from T, where:*

i) $\forall k \in [1..K-1], (t_k)^{\succ} = t_{k+1}$

ii) $s_1 = [t_1, t_n]$

iii) $s_K = [t_i, t_N]$

The cardinality of a segmentation S ranges from 1 to N. In the first case, $t_1^- = t_1$ and $t_1^+ = t_N$. In the later, S contains N singular segments, where $t_n^- = t_n^+$.

4.2.1 Relations between segments

From Definition 4.4, it is easy to see that two segments of T, s_1 and s_2, are equal when both of their limits are equal, $t_1^- = t_2^-$ and $t_1^+ = t_2^+$. We will denote the `equal` relation by the symbol $=$. For simplicity, to represent $\neg(s_1 = s_2)$, we will use the symbol $\neq$.
To ease the understanding of the upcoming definitions, we will use three segments of T, defined as $s_1 = [t_1^-, t_1^+]$, $s_2 = [t_2^-, t_2^+]$ and $s_3 = [t_3^-, t_3^+]$.

Proposition 4.1 (properties for the equal relation). *The equal relation is reflexive, symmetric, and transitive.*

Definition 4.7 (precede relation). *Given $s_1 = [t_1^-, t_1^+]$ and $s_2 = [t_2^-, t_2^+]$, s_1 precedes s_2, denoted as $s_1 \prec s_2$, iif $t_1^+ < t_2^-$*

Proposition 4.2 (properties for the precede relation). *The precede relation is irreflexive, asymmetric and transitive.*

Proof. Let $s_1 \prec s_2 \Rightarrow \neg(s_2 \prec s_1)$. Now assume that $s_1 \prec s_2 \wedge s_2 \prec s_1$ is true. From Definition 4.7 we know that $t_1^+ < t_2^- \wedge t_2^+ < t_1^-$. This means there are elements between t_2^+ and t_1^+ that are shared between both segments, which contradicts Definition 4.7, proving the asymmetric nature of the relation.
Finally, if $s_1 \prec s_2 \wedge s_2 \prec s_3$ then, from Definition 4.7, we get that $t_1^+ < t_2^- \wedge t_2^+ < t_3^-$. From Definition 4.4, if $t_2^- < t_2^+$ thus $t_1^+ < t_3^-$, which is the same as $s_1 \prec s_3$. This proves the transitivity for the relation. □

Following Definition 4.7, it is possible to define a more strict precedence order, the *contiguous precedence*, that maintains the properties in Proposition 4.2, except the transitivity.

Definition 4.8 (precede contiguous relation). *Let $\prec_c$ represent the* `precede contiguous` *relation, where*

$$s_1 \prec_c s_2 \Leftrightarrow t_n = t_1^+ \wedge (t_2^-)^{\succ} = t_{n+1}$$

Lemma 4.1 (precedence relation). $\forall s_i, s_k \in S_1 \ : \ s_i \prec s_k \vee s_k \prec s_i$

Proof. Let's choose a segment s_i from S. Now, every other segment we may choose, let's say s_k, contains elements that are greater or lower than the ones from s_i, because segments from a given segmentation do not share elements. From the definition 4.4, we know that if they are greater, then $t_i^+ < t_k^-$. From Definition 4.7 this is the same as $s_i \prec s_k$. Otherwise, $t_k^+ < t_i^-$ or $s_k \prec s_i$. □

Definition 4.9 (contained relation). *A segment s_2 is **contained** in s_1, denoted by $\sqsubset$, when all elements of s_2 are also elements of s_1,*

$$s_1 \neq s_2 \wedge \forall t \in s_2 \Rightarrow t \in s_1$$

Proposition 4.3 (properties for the contained relation). *The contained relation is irreflexive, asymmetric and transitive.*

Proof. The irreflexive property is a direct consequence of Definition 4.9, since the contained relation does not contemplate the equal case. Therefore, $\neg\,(s_1 \sqsubset s_1)$ is true, for any segment s_1.
Let us demonstrate the transitive property. If $s_2 \sqsubset s_1 \wedge s_1 \sqsubset s_3$, by Definition 4.9 we can rewrite as $(t_1^- \leq t_2^- \wedge t_1^+ \geq t_2^+) \wedge (t_3^- \leq t_1^- \wedge t_3^+ \geq t_1^+) \wedge (s_1 \neq s_2) \wedge (s_1 \neq s_3)$. Since the relation $\leq$ is transitive, $t_1^- \leq t_2^- \wedge t_3^- \leq t_1^-$, means $t_3^- \leq t_2^-$. Using the same principle,

$t_1^+ \geq t_2^+ \wedge t_3^+ \geq t_1^+$ led to $t_3^+ \geq t_2^+$. So, if $t_3^- \leq t_2^- \wedge t_3^+ \geq t_2^+$ holds, we say that $s_2 \sqsubset s_3$, proving the transitivity.
The asymmetry is proved, since a relation that is irreflexive and transitive, is also asymmetric. □

Two segments overlapped, if they have some common elements and they are not equal, neither one contains the other.

Definition 4.10 (overlap relation). *Two segments overlapped each other, denoted as $s_1 \ O \ s_2$ iff*

$$\begin{aligned} \exists t \in T \ : \ (t \in s_1 \ \wedge \ \ t \in s_2) \wedge \\ \exists t \in s_1 \ : \ t \notin s_2 \wedge \\ \exists t \in s_2 \ : \ t \notin s_1 \end{aligned}$$

Proposition 4.4 (properties for the overlapped relation). *The `overlap` relation is symmetric.*

Proof. From Definition 4.10 if $s_1 O s_2$ then there is at least one t that belongs to both segments. This means that $t_1^- \leq t \leq t_2^+ \vee t_1^+ \leq t \leq t_2^-$. Assuming that $s_2 O s_1$ is false, than $\neg(t_1^- \leq t \leq t_2^+ \vee t_1^+ \leq t \leq t_2^-)$, producing $t_1^- > t > t_2^+ \wedge t_1^+ > t > t_2^-$. Rearranging the expression, we can see that $(t_1^- > t \wedge t_1^+ > t) \wedge (t > t_2^+ \wedge t > t_2^-)$, which is a contradiction since t will not belong to s_1 nor s_2. □

Figure 4.4: Illustration of the non-transitive nature of the overlapped relation.

Figure 4.4 depicts an example that demonstrates the non-transitive nature of the relation. Notice that $s_1 \ O \ s_2$, $s_2 \ O \ s_3$, but $s_1 \ O \ s_3$ is false, since they don't have any t in common.

Figure 4.5 illustrates the relations between segments, and Table 4.1 summarises them, presenting their symbols and properties. The rectangles represent the timestamps range of a segment.

4.2.2 Relations between segmentations

We will present some binary relations supporting the comparison between two segmentations. The only requisite for this comparison is that they are defined over the same P. Since there are many ways to segment a set of personal photos, the binary relations allow us to understand how segmentations differ. Since segmenting a set of personal photos is an important step towards annotation, knowing that two segmentations present different

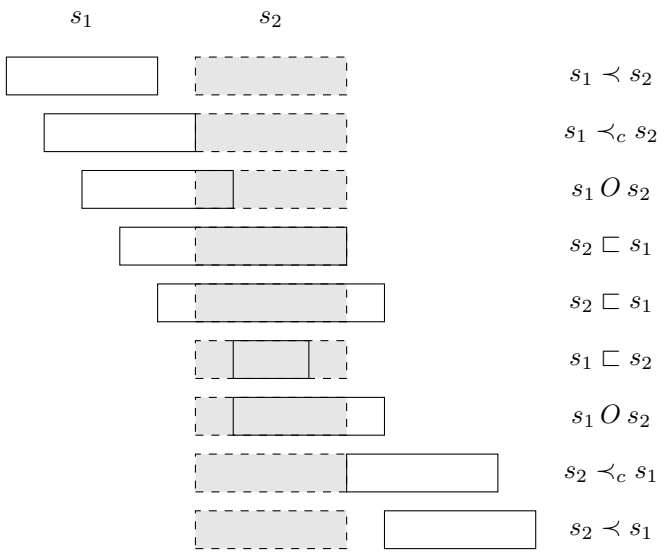


Figure 4.5: Illustration of the relations between segments.

	Symbol	Symmetric	Asymmetric	Reflexive	Irreflexive	Transitive
`equal`	$=$	•		•		•
`precede`	$\prec$		•		•	•
`precede contiguous`	$\prec_c$		•		•	
`contained`	$\sqsubset$		•		•	•
`overlap`	O	•				

Table 4.1: Summary of the binary relations between segments.

grains of segmentations (although compatible), reveals the annotations of one segmentation are more detailed than the annotations of the other.
There are four scenarios of comparison between segmentations: the *equality* (4.6), and other four, divided in *compatible* (Figures 4.7 and 4.9) and *incompatible* (Figure 4.10).

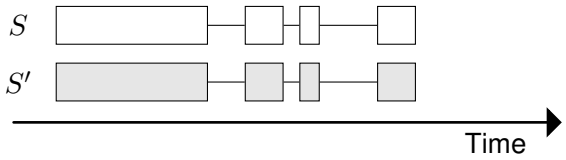


Figure 4.6: Equal segmentations.

Figure 4.6 shows two equal segmentations S and S'. They segment T in the same segments, so their cardinality is equal.

Definition 4.11 (equal segmentations). *Two segmentations S and S' are* `equal`*, denoted by $S \equiv S'$, when all the segments at the same index, are equal. Thus*

$$\forall s_k \in S, \forall s_l \in S' \; : \; k = l \; \Rightarrow s_k = s_l$$

For simplicity and clarity, we will denote the $\neg(S \equiv S')$ as $S \not\equiv S'$.

Proposition 4.5 (properties for the equal relation). *The* `equal` *relation, defined between segmentations, is reflexive, symmetric and transitive.*

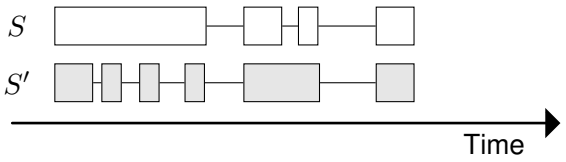


Figure 4.7: Compatible segmentations.

Figure 4.7 represents the case of two **compatible** segmentations.

Definition 4.12 (compatible segmentations). *Let S and S' be two segmentations of T. S and S' are* `compatible`*, denoted by $\lesseqgtr$, when they are not equal and*

$$\forall s_k \in S, \exists\, s_l \in S' \;:\; s_k = s_l \vee s_k \sqsubset s_l \vee s_l \sqsubset s_k$$

In two compatible segmentations, defined over the same T, there is no overlapping between segments of the two segmentations.

Proposition 4.6 (properties for the compatible relation). *The compatible relation is irreflexive and symmetric.*

Proof. Let us assume that $S \lesseqgtr S$ is true. From Definition 4.12 we known that they cannot be equal, and thus the statement is false, proving the reflexivity.
We also know that each segmentation shares a common restriction: that some segments are a refinement of the other. This duality and the fact that all other segments are equal, make the compatible relation symmetric. □

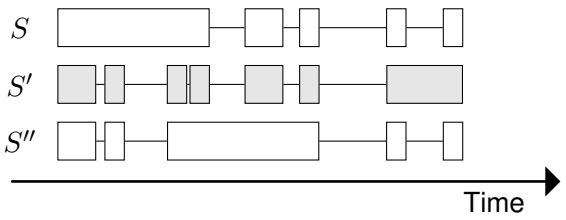


Figure 4.8: Illustration of the non-transitive nature of the compatible relation.

The compatible relation is not transitive. Figure 4.8 depicts an example that demonstrates the non-transitive nature of the relation. Notice that $S \lesseqgtr S'$ and $S' \lesseqgtr S''$, but when we compare S and S'', it is clear that $S \lesseqgtr S''$ is not true. For instance, the first segment of S overlaps the third segment of S''.

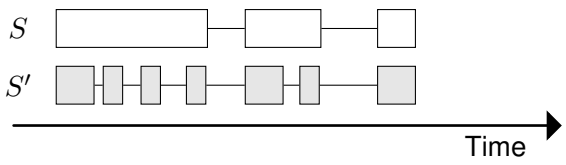


Figure 4.9: S' is a refinement of S.

Figure 4.9 illustrates a special case of compatibility, called *refinement*.

Definition 4.13 (refinement relation). *Let S and S' be two segmentations of T. S' is said to be a* `refinement` *of S, denoted by $S' \lhd S$, when $S \not\equiv S'$ and each segment of S' is equal or contained in one segment of S, holding*

$$\forall s_k \in S', \exists^1 s_l \in S \ : \ s_k = s_l \vee s_k \sqsubset s_l$$

For simplicity and clarity, we will denote the $\neg(S' \lhd S)$ as $S' \not\lhd S$.

Lemma 4.2. *Since the refinement relation means a fine grain division of T, if $S^{'} \lhd S$, then $|S^{'}| > |S|$*

Proof. If $S^{'} \lhd S$ then from Definition 4.13, $S^{'} \not\equiv S$, then at least one segment of $S^{'}$ is contained in one segment of S. This means that one or more segments of $S^{'}$ have at least one less element than a segment of S. Otherwise, Definition 4.13 will not hold. Thus, this element must be in another segment of $S^{'}$. This means the elements from s_k are split in at least two elements of $S^{'}$, making $|S^{'}| \geq 1 + |S|$, proving the lemma 4.2. □

Proposition 4.7 (properties for the refinement relation). *The* `refinement` *relation is irreflexive, asymmetric and transitive.*

Proof. The demonstration of the irreflexive is a consequence of the definition, since the segmentations must be different.
We will demonstrate the asymmetric property by contradiction. Let $S^{'} \lhd S \wedge S \lhd S^{'}$ be true. Using lemma 4.2, we can rewrite $|S^{'}| > |S| \wedge |S| > |S^{'}|$, which is a contradiction, because a value cannot be simultaneously greater and lesser than some value. So it proves the asymmetry of the relation, $S^{'} \lhd S \Rightarrow \neg(S \lhd S^{'})$.
Finally, the demonstration of the transitive property. Let $S^{'} \lhd S \wedge S \lhd S^{''} \wedge S^{'} \not\lhd S^{''}$ be true. Let s_j be a segment from S, s_k be a segment of $S^{'}$, and s_l a segment from $S^{''}$. From Definition 4.13, we know each segment of $S^{'}$ is equal or have its limits within the limits of one segment of S. The same happens between S and $S^{''}$. Thus, we can rewrite $(t_j^- \leq t_k^- \wedge t_k^+ \leq t_j^+) \wedge (t_l^- \leq t_j^- \wedge t_j^+ \leq t_l^+) \wedge (t_l^- \geq t_k^- \vee t_k^+ \geq t_l^+)$. Since the relations $\leq$ and $\geq$ are transitive, $t_l^- \leq t_k^- \wedge t_k^+ \leq t_l^+ \wedge t_l^- \geq t_k^- \vee t_k^+ \geq t_l^+$. The expression is true if and only if $t_l^- = t_l^- \wedge t_c^+ = t_k^+$. But this is in contradiction with Definition 4.13, because $S^{'}$ and $S^{''}$ cannot be equal. So, it is proved that the refinement relation is transitive. □

The refinement relation is a special case of compatibility between two segmentations, where the segments on the refined segmentation are equal or contained in the segments of the other segmentation.

The relation shown in Figure 4.10, is the `incompatible` relation, denoted by $\|$.

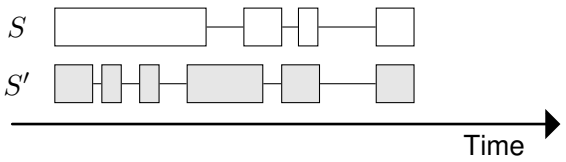


Figure 4.10: Incompatible segmentations.

Definition 4.14 (incompatible segmentations). *Let S and S' be two segmentations from the same T. They are `incompatible`, if some segments of S overlaps some segments of S'*

$$\exists\, s_j \in S, \exists\, s_k \in S' \; : s_j \; O \; s_k$$

For the example shown in figure 4.10, there are many cases of of overlapping between segments: $s_1 \, O s'_4$, $s_2 \, O s'_4$, and $s_2 \, O s'_5$.

Proposition 4.8 (properties for the incompatible relation). *The `incompatible` relation is irreflexive and symmetric.*

Proof. The irreflexive property is a consequence of Definitions 4.14 and 4.6, since no segment of a segmentation overlaps to any other segment of the same segmentation.
We will prove the symmetric property by contradiction. Let's assume that S is incompatible with S' but the symmetric is not, that is $S \| S' \wedge \neg(S' \| S)$ is true. Expanding those using Definition 4.14, we get $[\exists\, s_j \in S, \exists\, s_k \in S' \; : s_j \; O \; s_k) \wedge (\forall\, s_j \in S, \forall\, s_k \in S' \; : \neg(s_j \; O \; s_k)]$. This is an obvious contradiction, because no two segmentations can simultaneously have no overlapping segments and at least two that overlap, proving that $S' \| S \wedge S \| S'$ is true.□

The `incompatible` relation is not transitive. We will use an example to illustrate why, depicted in Figure 4.11. S is incompatible with S' due $s_1 \, O \, s'_4$ and S' is incomptible with S'' because $s'_4 \, O \, s''_2$. However, when we compare S and S'', instead of incompatible, we notice that $S'' \lhd S$. This simple example shows the non-transitive nature of the relation.

The binary relations between segmentations are summarised in Table 4.2. It resumes their symbols and properties. The presentation order, from top to bottom, indicates their descending level of compatibility, where `equal` is more compatible than `incompatible`.

4.2.3 Distance function between segmentations

As stated before, the comparison between segmentations can only take place if they result from the same set of timestamps T. Sometimes, besides the relations between them,

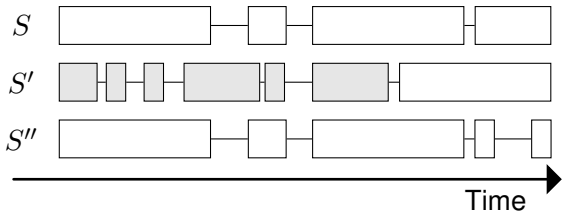


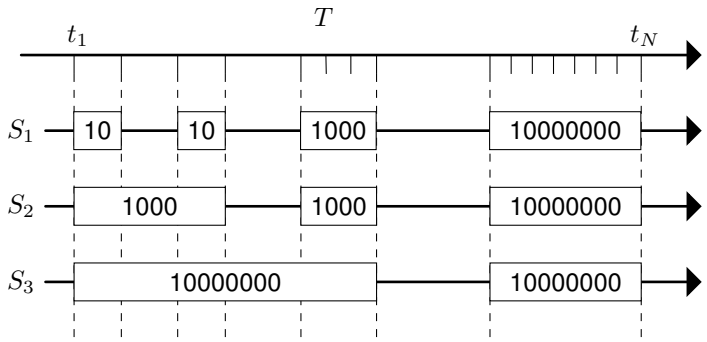
Figure 4.11: Illustration of the non-transitive nature of the `incompatible` relation.

	Symbol	Symmetric	Asymmetric	Reflexive	Irreflexive	Transitive
`equal`	$\equiv$	•		•		•
`refinement`	$\lhd$		•		•	•
`compatible`	$\leqslant$	•			•	
`incompatible`	$\parallel$	•			•	

Table 4.2: Summary of the binary relations between segmentations.

it is convenient to quantify how distant two segmentations are. For example, when comparing segmentations produced by different algorithms.

To calculate the similarity between a pair of segmentations more easily, we need to represent the segmentations in a different way. Using T as the reference, we can represent each segmentation as a binary vector, with an equal size of T. Each position of the vector marks the absence or existence of a segment start, using 0 and 1 respectively.



$$S_1 = (1,0,1,0,1,0,0,0,1,0,0,0,0,0,0,0)$$
$$S_2 = (1,0,0,0,1,0,0,0,1,0,0,0,0,0,0,0)$$
$$S_3 = (1,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0)$$

Figure 4.12: Vectorial representation of segmentations, illustrated for the refinement relation.

Figure 4.12 illustrates three segmentations represented this way. T has 16 timestamps, that are segmented differently, as shown in the depicted segmentations S_1, S_2, S_3. On top, the segmentations are graphically represented using rectangles. They represent the timestamps range of a segment. As we can see, there is a refinement relation between

segmentations, namely, $S1 \lhd S2$ and $S2 \lhd S3$. In Figure 4.12, on the bottom is the corresponding vector representation. The same relation is possible to observe. Take S_1 and S_2, as an example. The first four positions represent the first segment of S_2, containing 4 timestamps. In the same positions, S_1 has an extra 1. This indicates the beginning of a new segment. From that point on, all the 1s are aligned, meaning that only one segment of S_2 was refined. The nature of the segmentation and the selected representation, led to the following characteristics:

(i) It is impossible to have a vector filled with zeros, since a segmentation is a non-empty set of segments;

(ii) The first position of the vectors represents the early segment of both segmentations, and thus, is always equal to one. This also means they are aligned with T;

(iii) The number of ones ranges from 1 to $|T|$. The minimum happens when a segmentation has a single segment and the maximum is achieved when the segmentation has only singular segments;

(iv) The number of zeros is comprised between 0 and $|T| - 1$, achieved when the segmentation has only singular segments and a single segment, respectively.

With this representation, it is possible to use several distance functions defined in the literature, namely the Hamming distance [Mya07]. However, there are metrics for comparing segmentations, in the literature. Although they were developed for text segmentations, we can apply them to compare segmentations of sets of photos (e.g. [NSPGM04]). The most common metrics are the P_k [BBL99], WindowDiff [PH02] and PR_{error} [GCA06]. These are relative measures, whose range is normalised between 0 and 1, enabling a comparison outside the same T. They use a sliding window and take into account the mismatches between 1s inside that window. Using the segmentations in the Figure 4.12, and other three special cases,

$$S_4 = (1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1)$$
$$S_5 = (1,0,1,0,1,0,1,0,1,0,1,0,1,0,1,0)$$
$$S_6 = (1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)$$

we can calculate the values for WindowDiff (WD), using a window size of 3. The results are presented in Table 4.3. The last line illustrates an important difference between metrics. Since ones represent the beginning of segments, the hamming distance indicates the number of mismatches between segment starts. On the other hand, WD (and the others metrics) differentiate between false positives, false negatives and near misses. This means that a shift of a segment start of 1 position will not be penalised as much as a missing segment start.

Segmentations	WindowDiff	Hamming
S_1,S_2	0.143	1
S_2,S_3	0.143	1
S_1,S_3	0.286	2
S_1,S_4	0.928	12
S_1,S_5	0.429	4
S_1,S_6	0.429	3

Table 4.3: Illustration of distance metrics for segmentations.

4.3 The segmentation algorithm

In this work, an event is a happening that spans through time and space, forming sequences of happenings that do not overlap. The separation can be done either by clustering or segmenting a set of photos. Since the temporal order of happenings is key, segmenting is a way to separate events in a set of photos. The segmentation algorithm is supported in assumptions:

(i) Timestamps are ubiquitous in today's digital photos;

(ii) PPCs exhibits a bursty nature of the shots taken by the photographer [GGMPW02; Gar03];

(iii) Personal and social activities are scheduled, performed, and dictated by temporal rhythms, among them the natural cycles of days and years [Zer85];

(iv) Temporal order is important, as people tend to recall events using their order of occurrence [Fri04; SJRLC08; KSM12];

(v) Personal memories are about happenings in time and space [Tul02].

The Logical Day Event Segmentation (LDES) algorithm uses the creation time of photos to fulfil a base segmentation. The spatial information is used to fine-tune the limits of each segment. We take advantage of the bursty nature of the shots made by the photographer to detect gaps in the collection, both in time and space. Figure 4.13 shows the diagram of the algorithm, representing its three steps. Parameters w, f_t and f_g are used in different phases of the algorithm and will be described next.

4.3.1 Step 1 — Day finding

The first step of the algorithm has two goals:

1. produce a segmentation S_{days} where each segment contains the photos for a single *logical day*;

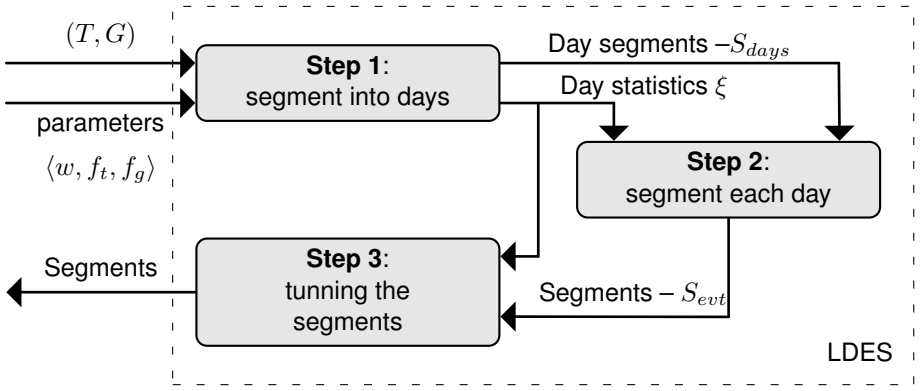


Figure 4.13: Overview of the LDES algorithm.

2. produce a set of statistics, denoted by ξ, for each segment, that will drive the later steps.

A *logical day* is the calendar day assigned to all the photos taken in one day of activity. It gets its name from the fact that, for us humans, daily activities may span two calendar day. For example, if someone gets out of bed at 10 a.m. in the 1^{st} of May and goes to bed at 2 a.m. in the next day. For recalling purposes, the day that matters is the 1^{st} of May, because our notion of *day* only ends when we rest. Thus, 1^{st} of May is the logical day assigned to all photos taken in that period. Since the notion of *logical day* follows closely the daily cycle and activities people do, the assignment changes on a daily basis, depending on the photos we have in the collection.

To settle a logical day, it is necessary to have a sequence of timestamps that spans from the last hours of one "standard" day into the early hours of the next one. If the timestamps fall within a window w, a parameter of the algorithm, they are considered to belong to the same logical day — the first day. If not, the logical and "standard" day are considered to be the same. The rational behind this window is that people need to rest a few hours, and this usually happens at night. Thus, after a "day" in people's activity documented by a set of photos, comes a larger gap. The assignment of one logical day ends when such gap is reached. LDES does not address exceptional cases, like New Year's eve; those cases should be treated at the application level, for example, setting w accordingly, for special days. If the gaps between photos are regular in two consecutive days, thus falling inside the window w, the logical day is the same as the standard day. We also assume the temporal information can be wrong, with errors reflecting a constant delta from the correct temporal reference. Since the gap between photos is unchanged by those errors, the LDES is not affected.

Besides the segmentation into logical days, this step produces the daily statistics ξ. These are important for the next steps of the algorithm, namely, to detect segments within the day. For each segment, representing a logical day, the statistics include:

1. the number of photos;

2. the maximum time gap;

3. the minimum time gap;

4. the average time gap between two consecutive photos.

With such information, the algorithm can adjust the segmentation to the shot behaviour the photographer had each day.
Listing A.1 (in Appendix A) shows the pseudo-algorithm for the first step of LDES.

4.3.2 Step 2 — Event finding

The second phase of the algorithm takes the segmentation S_{days} and the statistics ξ to produce a segmentation S_{evt}, where $S_{evt} \lhd S_{days}$. This refinement divides each segment of S_{days} into fine grained segments that are close to events (e.g. to visiting a museum). Since each day has its own statistics, the algorithm adapts the cut points to each daily set of photos. The decision to create a new segment is based on a reference value, denoted by Δ_t. Whenever the time gap between two consecutive photos is bigger than Δ_t, a new segment is created. Δ_t is calculated as

$$\Delta_t = f_t \times \Big[\text{average time gap } + (\text{maximum time gap} - \text{minimum time gap})\Big] \quad (4.2)$$

where $0.1 \leq f_t \leq 0.9$. Setting $f_t = 0.5$ is a recommended design value, that stands in the middle of the scale, providing a good separation of bursts in several scenarios. A lower value of f_t will produce more segments, and on the contrary, an higher value tends to join low density bursts, producing few segments. The rational behind the formula of Δ_t is the following. In the personal domain, each day is usually dedicated to few specific events,

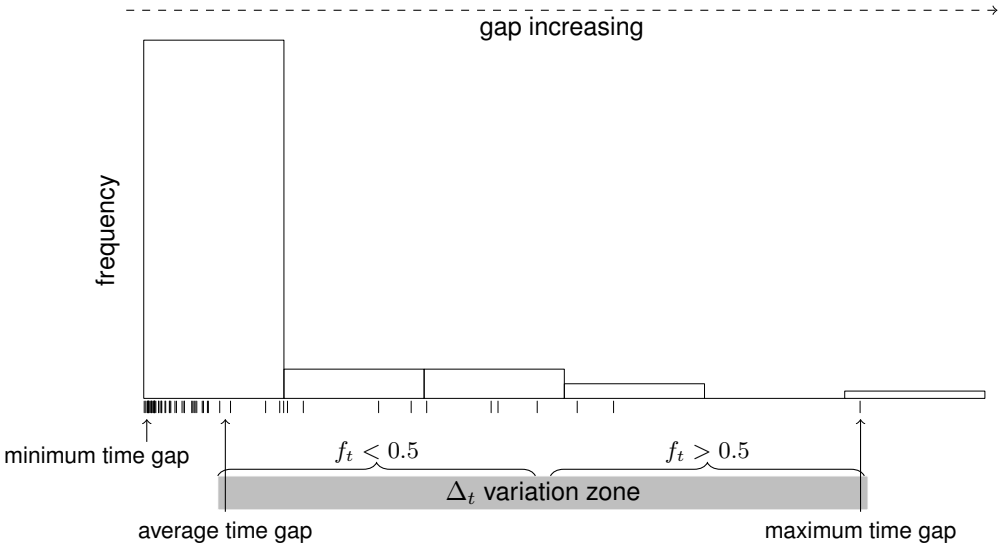


Figure 4.14: Representation of a typical gap spread, in a personal photo collection.

that are worth being photographed. It seems natural that time gaps between consecutive

photos change in different periods of the day, even for the same event. Fatigue, new subjects, and pauses, among others, influence the shooting behaviour. Although time gaps follow an exponential distribution, with low values more frequent than higher ones, there are different shot patterns that must be attained. Those are becoming more frequent as the usage of camera phones increase, and the photo taking habits change [CH09]. We consider the following:

A. Burst of photos;

B. Just two photos;

C. A single, uniformly separated group of photos;

D. Low density (sparse) groups of photos.

The maximum, the minimum and the average time gaps are used to handle these situations, contributing differently in each case, as encoded in (4.2). Shot pattern A is the most common. The typical gap distribution in this situation is presented in Figure 4.14. There are many small gaps, representing the time separation within the bursts of photos, and few large gaps, the potential separations between events [Gar03]. In this scenario, the minimum time gap tends to zero, the average time gap tends to be small, and the maximum time gap tends to be much larger than the others, leading to $\Delta_t \approx f_t \times$ maximum time gap. Since maximum time gap $\gg$ average time gap, the most frequent situation is

$$\text{average time gap} < \Delta_t < \text{maximum time gap}.$$

For $f_t = 0.5$, the close dense bursts are separated from the ones that are far apart.
If there are just two photos (shot pattern B), there is only one time gap and thus, the maximum, the minimum and the average time gap are all equal. In such cases, $\Delta_t = f_t \times$ time gap. Knowing that $0.1 \leq f_t \leq 0.9$, Δ_t is always lower than the gap between the two photos, separating them into different segments. If the gap is small, turning out they belong to the same event, the next step of the algorithm will use the spatial location of photos to correct such situation, joining the two photos.
In shot pattern C, the temporal information could be insufficient to identify the segments. Since there is a steady shooting pattern, all the time gaps are very similar. As such, the minimum and the maximum time gaps tend to cancel out in (4.2), making $\Delta_t \approx f_t \times$ average time gap. This means the $\Delta_t <$ average time gap, increasing the cardinality of the segmentations. The next step of the algorithm will use the spatial location to fine tune the segmentation, joining segments with similar spatial locations.
Shot pattern D happens when there are few photos documenting each event. In such cases, the gap between photos of the same event is larger but still smaller than the gap photos separating events. Assuming the minimum time gap does not tend to zero in such

situations, the difference (maximum time gap − minimum time gap) tends to approximate the average time gap, leading to $\Delta_t \approx f_t \times 2 \times$ average time gap. With $f_t \geq 0.5$, only gaps greater than average produce new segments. This behaviour is confirmed on the tests performed in the datasets described in Table B.1.

Despite (4.2) uses different temporal statistics to adapt Δ_t to different shooting patterns, it is possible to generate under- or over-segmented collections. In such situations, the next step uses spatial information to guarantee spatio-temporal coherent segmentations. Listing A.2 shows the pseudo-algorithm for the second step of LDES.

4.3.3 Step 3 — Event tunning

The temporal information may not be sufficient to separate activities. Knowing that it takes time to move in space, there are two situations that need further analysis:

- there is a time gap between two consecutive segments, but their spatial location is similar;
- time gaps inside a segment exhibit a regularity, but the spatial locations indicates two or more places.

This last step takes care of these situations, analysing segmentation S_{evt} and producing the final segmentation, S_{final}.

The first thing that is necessary is to validate the spatial coordinates. Although the location based services, available in many digital devices (e.g. smartphones), are becoming more accurate, such accuracy depends on the hardware characteristics and on environmental conditions. The location precision can vary from a few meters up to 3 km [WM10]. The spatial information is used to compute the distance of each g_n to the predecessor, g_{n-1}, denoted by δ_n. By definition, the distance for the first photo, δ_0, is 0, and so is the distance for all photos without spatial coordinates. Then, the outliers in the spatial data are detected. Three statistical methods were tried: one using Tukey's rule [MTL78], other the Local Outliers Factors [BKNS00], and another using a speed based solution. The development test shows the first is too sensitive, given the location's range of variation, and the second has an impact on the LDES performance, producing results similar to the latest. The proposed solution, speed based, employs a simple moving average over the spatial and temporal distances, considering a small window of size $2k + 1$. The central point of the window is the photo whose coordinates are being assessed. Do note that, for the initial and last photos, the window is only $k + 1$. This is because there is no previous or successive photos, respectively. The speed based method assumes that most of the photos in a personal photo collection are taken on foot. As for travelling long distances, it takes times. The average spatial gap between photos inside the $2k + 1$ window

is compared to a spatial reference value S_{ref},

$$S_{ref} > \frac{\delta_{n-k} + \ldots + \delta_n + \ldots + \delta_{n+k}}{2k+1} \tag{4.3}$$

and the average speed between photos inside the same window is compared to a temporal reference value V_{ref}

$$V_{ref} > 120 \times \left(\frac{\delta_{n-k}}{t_{n-k} - t_{n-k-1}} + \frac{\cdots}{\cdots} + \frac{\delta_n}{t_n - t_{n-1}} + \frac{\cdots}{\cdots} + \frac{\delta_{n+k}}{t_{n+k} - t_{n+k-1}} \right) / (2k+1) \tag{4.4}$$

The constant 120 is an approximation to the kilometres in one degree. A photo p_n is marked as an outlier when (4.3) and (4.4) are both `True`. The reference values were set as $S_{ref} = 5km$ and $V_{ref} = 100km/h$, making all photos taken at high speed and with large[5] gaps between them marked as outliers. In all the experiments we use $k = 1$. With this procedure at most 1% of spatial coordinates are marked as outliers for each collection used in the experiments. When the coordinates are missing, or are considered outliers, the photos are marked accordingly. This means the only information used from them is the temporal information. After the outlier detection, the spatial information is used to fine-tune the limits of each segment, using the $split$ operation in first place, and then using the $join$ operation.

The $split$ operation evaluates each segment s_k from S_{evt} to check if it can be refined, producing a set of segments denoted by $split_s = \{s_1, s_2, \ldots, s_n\}$, such $t_k^- = t_1^- \wedge t_k^+ = t_n^+$. If the distance between two consecutive photos is greater than a reference value Δ_s, a new division is set. The reference value is calculated as

$$\Delta_s = (1.5 - f_g) \times \sigma_k \tag{4.5}$$

where σ_k is the standard deviation of δ_n for the valid photos[6] in s_k, and $0 \leq f_g \leq 1$. This means the split reference f_g, a parameter of the algorithm, is between $\frac{1}{2}$ and $\frac{3}{2}$ standard deviations. $split_s$ is kept if it lessens the distance's variance between photos inside each segment, producing more compact segments than the original one, verifying the strict inequality

$$\frac{\left(\sum_{s \in split_s} \sigma_s \right) \times \frac{1}{|split_s|}}{\sigma_k} < f_g \tag{4.6}$$

The numerator gives the mean of the standard deviation of δ_n in each $s \in split_s$ and the denominator gives the standard deviation of δ_n in the original s_k. The first spatial gap of each segment is left out of the calculus, as it represents a cutting point, thus it is not

[5] Considering a person on foot

[6] Photos with spatial coordinates neither null nor outliers

representative of the gaps inside a segment. Do notice that when $f_g = 0$, no split is done. On the other hand, with $f_g = 1$ small reductions in the standard deviation will produce more splits.

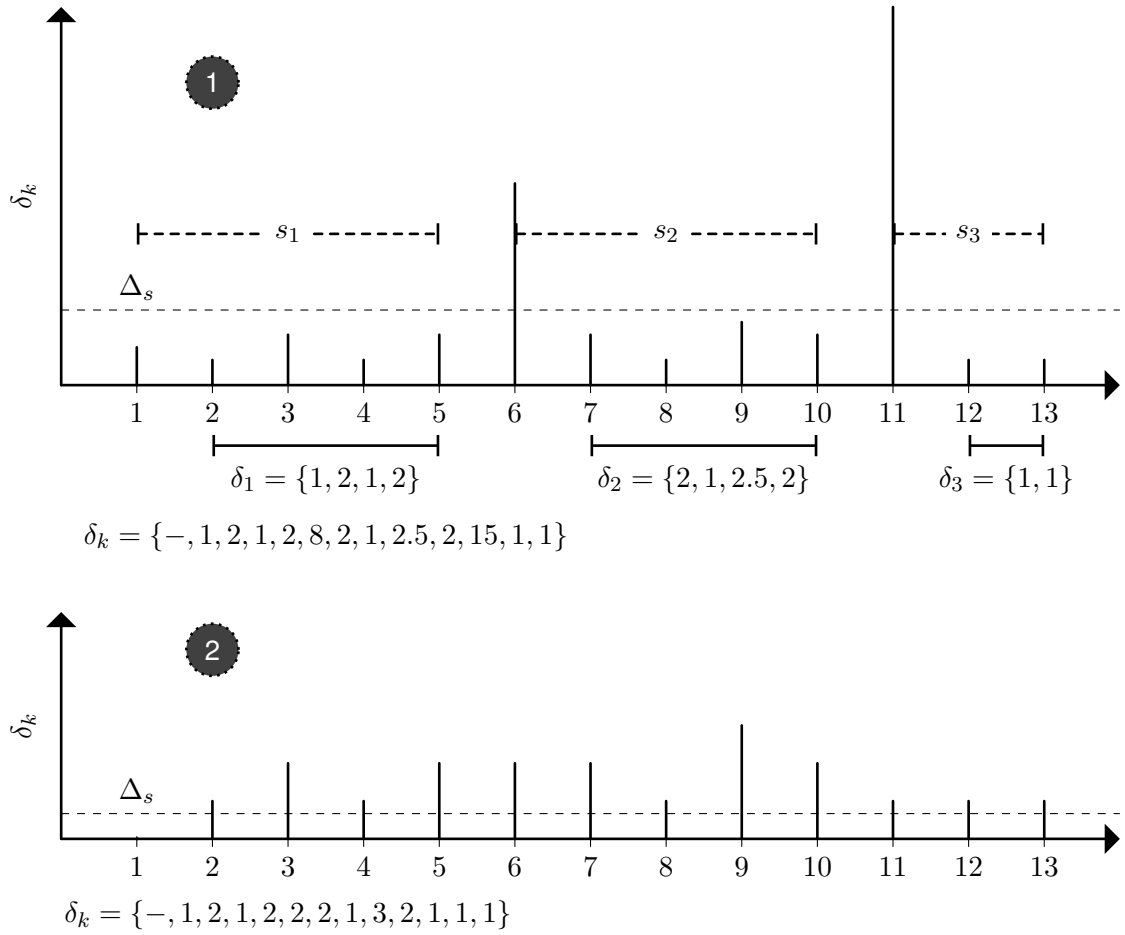


Figure 4.15: Examples illustrating the *split* operation.

Figure 4.15 shows two examples of a segment *split*, using $f_g = 0.5$. The first case, denoted by ❶, illustrates a segment where there are large spatial gaps. In this example, positions 6 and 11 verify (4.5). Thus, $split_s$ gets three segments, s_1, s_2 and s_3 starting at positions 1, 6 and 11 respectively, as illustrated. To make the split, it is necessary that new segments lessen the distance's variance between photos. The depicted spatial gaps δ_1, δ_2, δ_3 and δ_k are used in (4.6). Do note they omit the first gap of each segment. Since $split_s$ verifies (4.6), the split is done.

The second case, denoted by ❷, illustrates a scenario where there are no spatial gaps that stand out for difference. All positions, except the first, verifies (4.5). In this case, $split_s$ gets thirteen singular segments, whose δ_s is undefined. Therefore it does not verify (4.6), and no split is done.

The $join$ operation is performed, after the $split$ operation is completed for every segment $s_k \in S_{evt}$. It takes two contiguous segments, s_1 and s_2, and produces a new segment s_3 where $t_3^- = t_1^- \wedge t_3^+ = t_2^+$. The rational is as follows. It doesn't make sense to re-join something that was split using a similar criteria — to enhance the spatial cohesion inside each segment. However, the extremes segments can be spatially close to the contiguous segments, indicating they belong to the same event. Considering the original segments in S_{evt}, joining segments can benefit the context coherence, in situations where the spatial location is very similar, thus removing temporal gaps inside an event. The advantage of combining two segments s_1 and s_2 are verified using

$$\frac{\sigma_{s_3}}{(\sigma_{s_1} + \sigma_{s_2}) \times \frac{1}{2}} < 1 - f_g \tag{4.7}$$

where s_3 is the segment resulting from the join of s_1 and s_2. The numerator represents the standard deviation of δ_n after the join and the denominator represents the mean of the standard deviation of δ_n in the segments before the join. By definition, whenever the numerator is zero, the left hand side of the equation in (4.7) is also zero, independently of the value for the denominator. With $f_g = 1$ no join is performed. On the other hand, when $f_g = 0$, just a minor decrement in the standard deviation results in the combination of two segments. Figure 4.16 shows the conditions where segments can be joined. If they result from a $split$, represented as dashed rectangles, only the first and last segments in $\{s_1, s_2, \ldots, s_n\}$ can be joined with others. For the original segments in S_{evt}, no restrictions are imposed.

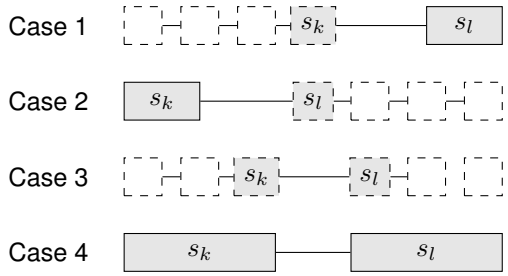


Figure 4.16: Cases where the join operation is possible. The segments resulting from a $split$ operation are dashed.

Figure 4.17 shows two examples of the $join$ operation, using $f_g = 0.5$. The first case, denoted by ❶, illustrates the two segments with no obvious advantage to join them, so the division settled using the temporal information is kept. As illustrated, the standard deviation of the spatial gaps in the original segments, δ_{s_1} and δ_{s_2} is very similar to the standard deviation of the joined segments, denoted by δ_{s_3}. Thus, Equation (4.7) is not satisfied, and the two segments are left untouched.
The second case, denoted by ❷, shows a situation where there is an advantage to join the two segments, since the spatial information represents a similar pattern for both. The join is done, since (4.7) is satisfied as the standard deviation of the spatial gaps in the

resulting segment in zero.

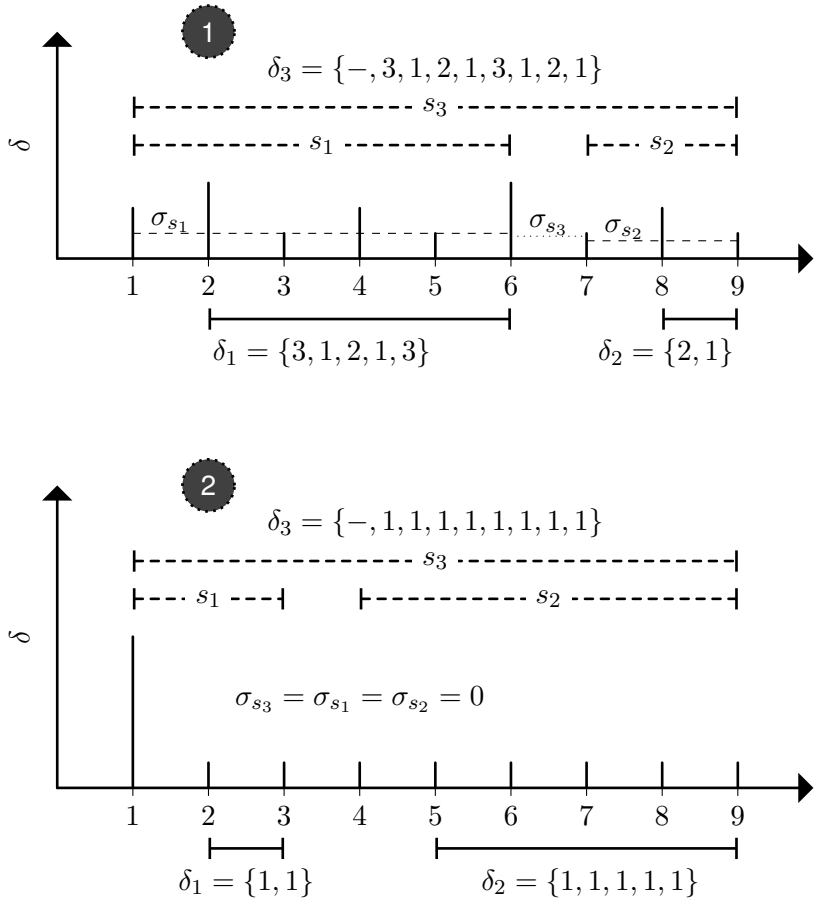


Figure 4.17: Examples illustrating the $join$ operation.

From the above explanations, we can see that lower values of f_g will produce less segments, and higher values will produce more segments. Setting $f_g = 0.5$ will produce a balance between the splits and joins. An important remark is that no $split$ or $join$ operation can break the notion of *logical day*.

4.4 Summary

In this chapter we describe the archival of a personal photo collection, emphasising the importance of a proper pre-archival processing step. Such step consists of the automatic segmentation of a roll, a way to reduce the labour needed to insert manual annotations. Three of the contributions of this thesis were presented and discussed, namely:

1. The theoretical framework supporting the segmentation, including a set of binary relations for comparing segmentations of the same set of photos;

2. The definition of the *logical day* concept;

3. An automatic segmentation algorithm, LDES, that uses the spatio-temporal information available in the photos, without user intervention.

The binary relations, summarised in Table 4.2, cover an omission in the literature. As far as we know, until this date there is no framework that enables a qualitative comparison between segmentations. Therefore, the five relations presented here are a contribution to the body of knowledge in this field. This is also the first work using the notion of logical day applied to personal photo collections. Knowing the importance of time in our daily live, the notion of logical day adapts the natural day cycle to the social day cycle, where the boundaries of each one not always coincide. Finally, the LDES algorithm implements a segmentation where the notion of logical day is fundamental. The usage of spatio-temporal information only guarantees that all segments are temporally and spatially delimited, despite the content of the photos. Since content varies dramatically in the same activity, our approach delivers a less detailed segmentation, contributing to our goal of supporting manual annotations. Given a set P with N photos, the LDES produces a segmentation S. The operations performed in P to get S are $\mathcal{O}(n)$, assuming an ordered set P satisfying (4.1). If we consider this requirement as part LDES complexity, then it changes to $\mathcal{O}(n \log(n))$. In our experiments we use LDES with a stable sort algorithm, as part of a web application, and found no performance issues.

The relevance of the contributions is validated by experiments involving users. Section 6 reports and discusses the results obtained.

5

Retrieval

> *One could go even further and say that one's own memories form the backbone of one's existence. Much of human behaviour can be understood only when placed in the perspective of the past, as it is remembered.*
>
> ***Willem Wagenaar***

The nature of a personal photo collection pose challenges in the archiving and retrieval, that are different from general-purpose databases [ST06]. The actions taken on the collections include querying, sharing and creating derivative artefacts (e.g. albums) [Har05]. Each action needs a proper support to answer information needs in an effective way, meeting user expectations. The tags, activities, or people's names, to name a few, have a specific meaning according the personal and social context of the user interacting with the system — their **semantic viewpoint**. This has a major impact for retrieval. When the user who retrieves photos is the same who archive them, then a match exists between the terms issued in the query, and the terms stored in the contexts. However, in a multi-user scenario, there are several users, each one with his own semantic viewpoint. This means the same set of photos can be retrieved using different queries, but similar in the semantic level. For example, "*Photos with my father*" and "*Photos with my uncle*" should generate the same set photos when the queries are issued by two cousins. Thus, the retrieval is much more than a simple query issued to the multidimensional context-space. It requires some reasoning about the equivalence of different terms, in different semantic viewpoints.
For the particular case of retrieving a set of photos from an event, the users need to remember the context of a specific event, so they can formulate a query. This includes

spatio-temporal cues, but also tags or other information he inserted during archival. Remembering is the result of linking, contextualising, and interpreting temporal, spatial, and social cues. Since the context is usually expressed by "*when*", "*where*", "*who*" and "*what*", a query is a specification of cues for the 4*Ws*. When a user writes a query, he wants a specific set of photos, let say G. Performing the query produces a set A, that can be different from expected. The intersection between G and A can be empty, A can be a subset of G, G can be a subset of A or A can partially overlap G. There are several reasons contributing to this. The user may enter wrong temporal cues. It is known [JS05] that a person may say that an event happened earlier than it did (backward telescoping), or that it happened more recently (forward telescoping). The user may enter wrong or ambiguous spatial descriptors. For example, Paris can be a city in France or in the USA. The cardinality of A tends to be large and judging the relationship between G and A can be difficult, as it collides with our capacity to handle large volumes of information [WH99; Bux01]. Understanding the context of A is important to reformulate the query, so the next query execution produces a set closer to the expected one. Thus, it is important how A is shown to users. In cases of complex information retrieval, which a photo retrieval is an

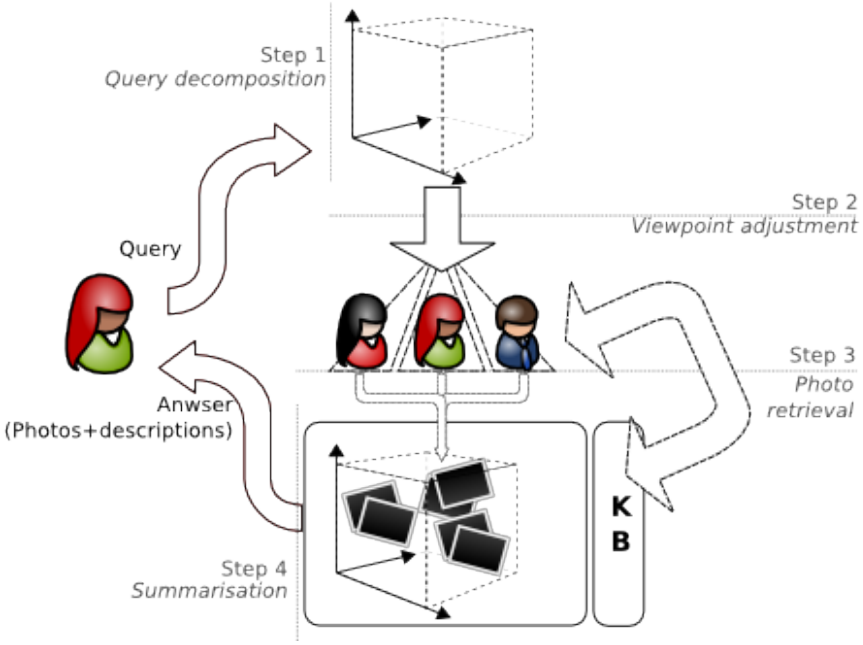


Figure 5.1: Overview of the retrieval process..

example, users may want (or need) to modify the queries and view intermediate results until their information needs are satisfied [MVCFA08]. The retrieval process is divided into four main steps, as showed in Figure 5.1:

1. The query is *decomposed* into terms, according the dimensions of the multidimensional context-space;

2. The query is *translated* to the viewpoints of the relevant users;

3. The photos whose contexts *intersects* the query specification are selected;

4. The set is summarised and returned to the user.

If the user needs to reformulate the query, the retrieval is an interactive process of several instantiations of the four mains steps described above.

In this chapter we address the retrieval of a set of photos. First, we present relevant use cases for the retrieval of photos in a multi-user scenario. Then we discussed the retrieval from the $\mathcal{MCS}$, containing both photos and the context description. We detailed the retrieval steps, including the viewpoint adjustment and focused on the last step, summarisation of geotagged photos in a compact way. We describe a novel approach called *Multimedia Short Summary* (MSS). It comprises: *(i)* a partition over a set of photos, *(ii)* a small textual description for each element of the partition, and *(iii)* a selection of an auxiliary grain of detail.

5.1 Use Cases

This section provides a series of use cases, illustrating the retrieval. For each use case, we define the initial conditions, many of them relating the way the photos were archived. Figure 5.2 shows the names and relations of a sample social context, used to simplify the comprehension of the use cases. We explore the duality of viewpoints between the user who archived the photos and the user who is performing the query.

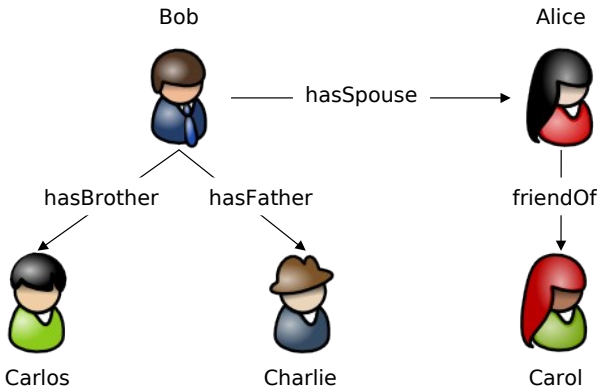


Figure 5.2: Sample social context for the retrieval use cases.

5.1.1 Retrieval of owned photos with temporal cues

This use case illustrates how enhanced information, and the KB can be used to deliver a set of photos with a given temporal location. The temporal cues can be relative, absolute or a mix. They also can use any on the cycles illustrated on Section 3.2. Bob issues the following queries "*Photos from **last summer***", "*Photos from **last August***", and "*Photos from ***summer** of **2011***". They lead to the same result in terms of photos.

Conditions:

(i) The KB has common domain knowledge, but no user assertions;

(ii) Bob is the user that queries the system;

(iii) The collection has only 12 photos from a day in August, 2014, taken at Lisbon;

(iv) The photos are tagged with *summer holiday*;

(v) Bob was the user who archived the photos.

Expected results:
In the query "*Photos from last **summer***", summer defines an implicit time range and *last* is a relative qualifier used to restrict the set of summers. Thus, only the last period ranging from 22^{nd} June to 20^{th} September is selected. The query "*Photos from last **August***" works similar to the previous one, except the period range is replaced by August, confining the search to only one month. In the query "*Photos from **summer** of **2011***", all the restrictions are explicit. Apart from that, the result is the same. Notice the summary of the set is the most specific descriptor, that in this case is a "*day in August at Lisbon, 2011*".

5.1.2 Retrieval of owned photos with spatio-temporal cues

This use case illustrates how the multidimensional context-space and the KB can be used to deliver a set of photos with a given spatio-temporal location. The cues in the query are matched against the spatial location information discussed in Section 3.1. Bob issue the query "*Photos of **last summer** in **Lisbon***".

Conditions: The same as in use case 5.1.1.

Expected results:
The word *Lisbon* is a spatial cue that can be matched against different levels of detail of the `name` property of the `Place` concept. Concerning the contexts in the multidimensional context-space, they belong to their named location hierarchy, representing different administrative place's descriptions. The term *summer* defines an implicit time range and *last* is a relative qualifier used to restrict the set of summers. The response to the query is a set of 12 photos. Their description is "*one day in August at Lisbon, 2014*".

5.1.3 Retrieval of photos with spatio-temporal cues, archived by others

This use case demonstrates what happens when the user querying the system is different from the one who archived the photos, for queries using spatio-temporal cues. Here, Alice issues the query "*one day in August, 2014 at Lisbon*".

Conditions: The same as in use case 5.1.1, except that Alice is the user who queries the system.

Expected results:
The spatial cue is matched against the information available in each context, since it is part of the *LoD* in which Lisbon is part of. The system returns a set of 12 photos, whose description is "*one day in August, 2014 at Lisbon*".

5.1.4 Retrieval of photos with inference

In this use case, Alice issues the "*last* ***summer trip***". However, the KB has user assertions, namely it has information about Alice and Bob.

Conditions:

(i) The same as in use case 5.1.1, except that Alice is the user who querying the system;

(ii) The KB has common domain knowledge and assertions for some persons of Bob' social circle, as illustrated in Figure 5.2;

(iii) Alice is known to live in Cambridge.

Expected results:
The temporal cue *summer* is matched against the information available in each context. The term *Trip* is absent from any context. However, there is a KB rule stating that a spatial location different from a user base location is a trip. Since Alice is known to live in Cambridge, having photos from Lisbon is consider a to be a trip. Therefore, the set of 12 photos are retrieved described as "*one day in August at Lisbon, 2014*".

5.1.5 Retrieval of photos from different hemisphere

Carlos lives in Australia, in a different hemisphere than his brother Bob. They went together to a Lisbon trip in August. Bob and Alice were the photographers who later share the photos with Carlos. The photos have Lisbon local time. Carlos issued the query "*Photos from winter holiday*".

Conditions: The same as in use case 5.1.4, except that Carlos is the user who querying the system.

Expected results:
The query "*Photos from winter holiday*" has the relative temporal reference *winter*. From the KB knowleged, `Winter` and `Summer` are equivalent in different hemispheres. Thus the term *winter* is translated into the equivalent term, from Bob's perspective — *Summer*. A set with 12 photos is returned, described as "*winter holiday at Lisbon, 2014*".

5.1.6 Implicit temporal cues

Sometimes the query defines a temporal interval, based on a pre-conception about when the activity should occur. For example, in the query "*brunch at Lisbon*", the term *brunch* is used to define the activity, but it can also be used to settle the time frame to search. Usually, a brunch is taken between late morning and early afternoon.

Conditions: The same as in use case 5.1.4.

Expected results:
Since the KB has the description of the concept `Brunch`, that includes the typical time frame, the temporal restriction is settle. Thus, all the photos taken in Lisbon between late morning and early afternoon are retrieved. In fact all the segments in $\mathcal{MCS}$ that intersects with this time frame are return with the description "*late morning in August at Lisbon, 2014*".

5.1.7 Summary without information

This use case illustrates the case where a set is not properly summarised, because the common denominator is too abstract and thus, contains no relevant information. The example is supported by a set of three photos, each taken in a different year, namely 2011, 2012 and 2013. Their context share nothing in common in terms of temporal location below the year, spatial location, or social information. Each one is marked to depict a person, namely *Carol*, *Dave* and *Trent*. Those names are not asserted in the KB or are not social related to Carlos.

Conditions:

(i) The KB has common domain knowledge and assertions for some persons of Bob' social circle, as illustrated in Figure 5.2;

(ii) Carlos is the user that queries the system.

Expected results:
For the three photos, the only common denominator is the *decade*. Thus, the description of the photos are "*Photos from several years*" that is too abstract to summarize a set of consecutive years.

5.1.8 Social denominator as a summary for a set

This use case shares the same set of three photos with the previous one, and Carlos is the user interacting with the system. However, now there is more information about

the social relation between Carlos and the persons depicted on the photos, as showed in Figure 5.3.

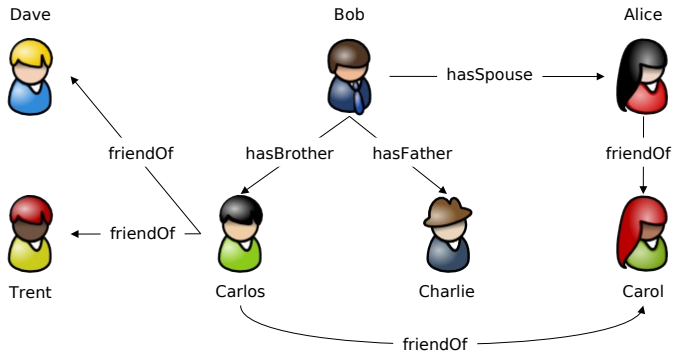


Figure 5.3: Second social context sample for the retrieval use cases.

Conditions:

(i) The same as in use case 5.1.7;

(ii) The KB has domain assertions plus other for some persons of Bob' social circle, as illustrated in Figure 5.3.

Expected results:
Since the three photos depict persons that have the same relationship with Carlos — they are their friends, the set is summarised as "*Friends*".

5.2 Retrieving a set of photos

Retrieving a set of photos from **MeMoT** is done using the minimalist interface showed in Figure 5.4. It boroughs its simplicity from a known web search engine. The design try to make things simple and productive [Gal07], hiding the details and complexity of the system underneath. The user enters the query in the text box. After a few characters, **MeMoT** starts suggestion terms representing concepts, or individuals, that are asserted in the KB. We limit the number of hints, not only for aesthetic reasons, but primarily because choosing between few alternatives is more efficient [Wei09]. The reason for suggesting cues is twofold. First, it lowers the writing effort, and second, it drives the user to select the semantics of the terms. This situation is illustrated in Figure 5.5. The example shows a choice between `July` being an instance of `Month` or a instance of `Person`. Even in the situations where the cue, given the known facts, is not ambiguous, this approach unveils the meaning of the terms issued in the query. Thus, it provides information about the user' intentions, that can be used later. The suggestions are made to every cue inserted in the query interface, independently.

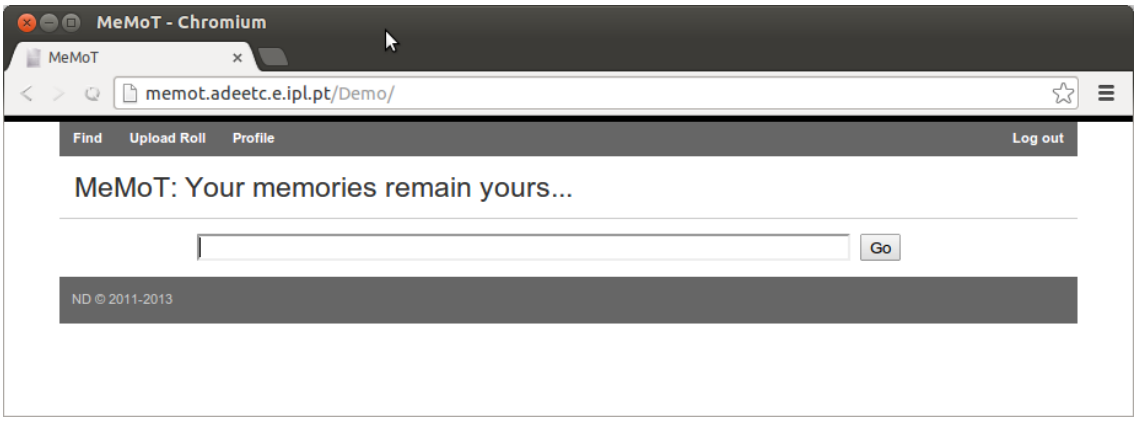


Figure 5.4: Query interface of the **MeMoT** system.

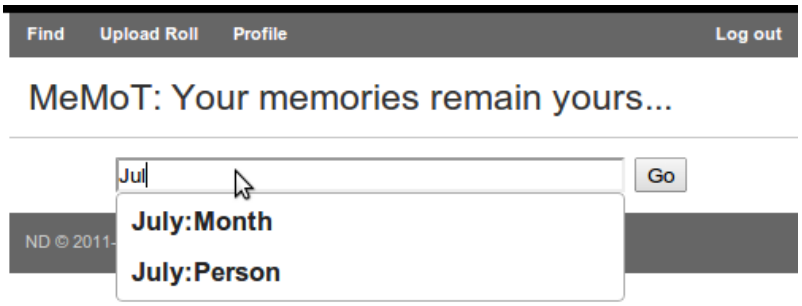


Figure 5.5: Cue suggestion example. Users set the proper semantics aided by the KB assertions.

5.2.1 Query decomposition

Figure 5.6: Query example.

The query is not analysed using natural language processing, since it is out of the scope of this work. Instead, the query is interpreted as a conjunction of terms that put restrictions on specific parts of the context, defining (partially) a sub-space of the $\mathcal{MCS}$, containing the relevant photos. The semantics of the terms is settled by the users, as explained earlier, allowing the assignment to a dimension of the $\mathcal{MCS}$ where the restriction

is intended. The terms unknown to **MOnt** have no explicit semantics, and are used for restrictions purposes only. They cannot be used in the viewpoint transformations. Since each dimension of the multidimensional context-space is independent of the others, it can be manipulated separately, for transformations and comparisons purposes. Figure 5.6 shows an example of a query that retrieves a set of photos from a snow trip. This work do not focused on developing (or using) an expressive query language. It uses a simple, yet effective, conjunction of terms separated by commas. They are assigned as restrictions to the $\mathcal{MCS}$ dimensions using the semantics settled by the user, using the hints **MeMoT** provides. In this example, the cues put restrictions on 3 dimensions, "*where*","*what*" and "*when*", on three levels of detail, `Country`, `Activity` and `Year`, respectively. The annotations were provided automatically by the system, as described in chapter 4. After the decomposition, we are able to get:

- The identification of the user issuing the query;
- A set of spatio-temporal restrictions (possible empty). For example, *Paris* and *2011*, as illustrated in Figure 5.6;
- A set of social restrictions, both on the "*who*" and "*what*" cues (possible empty). For example, *Winter Holiday*, as illustrated in Figure 5.6;
- A set of content-based restrictions, confined to the available features in the system (possible empty). For example *Face*, telling **MeMoT** to search for photos that are known to have people's faces;
- A bag of words with unknown semantics (usually empty). For example, using the the tag *canon*, restricting the result to photos for that brands of cameras.

5.2.2 Viewpoint adjustment

The previous step prepare the manipulation of the query and its terms. The viewpoint adjustment is tried, to those terms whose semantics is known. The adjustment use the the `DualConcept`, the `sameAs`, and other assertions in **MOnt** to expand the query with others representations of the same terms. Figure 5.7 illustrates the viewpoint adjustment. Lets us suppose that a query is composed by four terms, where the first three has a know semantics. Thus, a match is only tried on the A, B and C terms. If B is matched against a `DualConcept` assertion, the query is transformed, replacing B for a disjunction of terms. Such disjunction is made of the original term, B, and a conjunction of a new term and a condition for its applicability, $(E \wedge condition)$. The conjunction is derived from the definition of the `DualConcept`. Using the example in Figure 5.6, `WinterHoliday` can be transformed into `WinterHoliday` $\vee$ (`SummerHoliday` $\wedge$ `South Hemisphere`). Now, let us assume that C is matched against a `sameAs` assertion. The query is transformed in a similar way, replacing the original term C by a disjunction of terms, composed by the C and F, an equivalent term but with a different representation. For example, in families it

is usual that members have nick names. If a query use the term `Bob`, it can be expanded to `Bob` $\wedge$ `Bobby`. Finally, let us suppose that C represents a relative term, like a family relation, e.g. the term `Brother`. If Bob is issuing the query, an absolute representation can be derived using **MOnt**, if the asserted facts contain the family relation illustrated in Figure 5.3. Thus, `Brother` is transformed into `Brother` $\vee$ `Bob`.

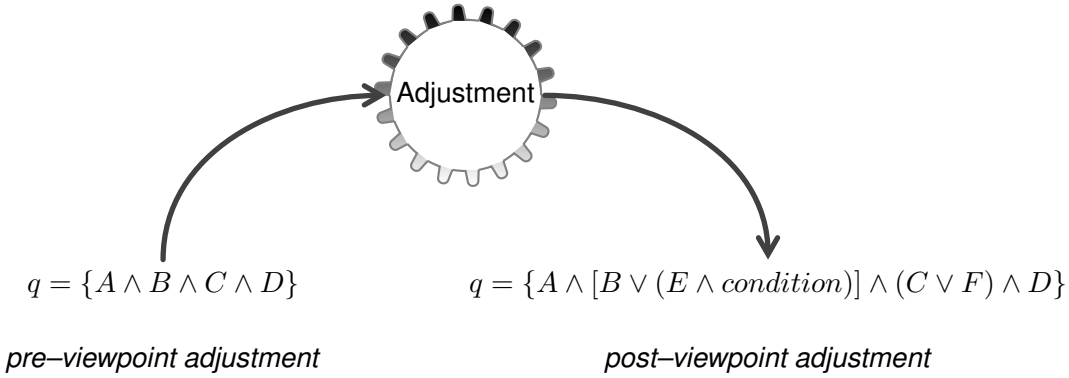


Figure 5.7: Illustration of the viewpoint adjustment.

As a final remark, do notice the response time of **MeMoT** is enough to reasoning while the user is typing the query. Because of this, the viewpoint adjustment is done in background, preparing the query with the necessary terms to make the retrieval faster.

5.2.3 Photo selection

The previous steps produce the terms, and respective semantic, that are used to retrieve objects from $\mathcal{MCS}$. Since it is implemented as a multidimensional repository, the selection of photos is a matter of direct the restrictions to the correct dimensions. Figure 5.8 illustrates this process. We use the cube metaphor (although $\mathcal{MCS}$ is a 4 dimensional space) to exemplify the selection. Each dimension is described by a set of attributes, with different levels of details level-of-detail, as already discussed in Chapter 3. Each term is used to restrict the values of the attributes in the dimensions. The conjunction of all restrictions in the cube produce a slice, representing all the photos that share the pairs $< attribute, value >$. Thus, the selection of photos is a matter of getting the positions in $\mathcal{MCS}$ that satisfy the query, i.e., the context being described by that set of terms in that query. Since $\mathcal{MCS}$ is implemented using a relational database, the terms are combined into a SQL query. The semantic of each term allow us to select the proper attribute in the tables, making the selection more efficiently. For example, if we know that $July$ is a month, then the restriction can be made in the `Date` table, in the fields `month`. Since **MOnt** is used as a metadata repository for $\mathcal{MCS}$ there are some assertions to aid this mapping. **MOnt** has two data properties, `tableId` and `fieldId`, that allow each term to be properly mapped to the a specific field of a specific table. If the semantics of the term is unknown, it is matched against all fields in all dimension tables. The result set contains not only the url of the photos, but also the description of the context of each one,

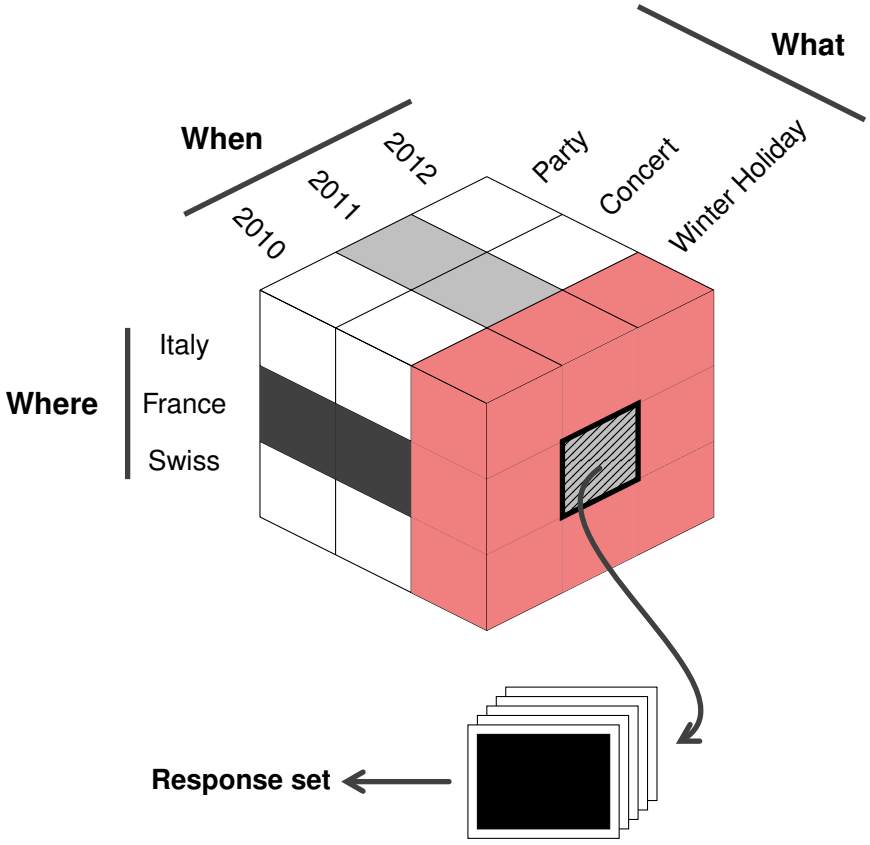


Figure 5.8: Photo selection example.

using the levels of the default hierarchy for each non-empty dimension, populated it the values retrieved from the $\mathcal{MCS}$.

Describing a set of photos With the richness of the multidimensional context-space, we can push the retrieval further, making possible to get contextual information. Photos are the visual part of the context, as they document an instant from a given perspective. This description is a summary of the set, containing both visual information and textual cues about the 4*Ws* enunciated earlier. The more information is available, the better the summary is. Some works in the literature [Pig10; NSPGM04], use features like spatial location and time, to cluster similar photos. They use such cluster to derive meaningful text descriptors. However, the relation between the photographer and the viewer is left behind. Even in the work of Jaffe [JNTD06], the photographer is treated as a part of the context, which is inherently truth, but the different semantics of the *producer* and *consumer* are not addressed. The summary of a set should be:

Human readable The summary should use common references a user would use to describe a set.

Concise It should not be a full length sentence, but a set of keywords around the 4*Ws*.

Meaningful to the user constructing the set The summary should be created using the context of the user requiring the summary, transforming the relative references towards the user semantic viewpoint.

Self contained The meaning of the summary should be understandable by the user from which it was generated, without the necessity to access more information.

The resulted set is summarized as described in the next section.

5.3 The summarisation problem

Let $P = \{p_1, \ldots, p_m\}$ be a set of photos representing personal memories, described by a set of attributes. The attributes are organised in dimensions, denoted by W_i, where W_1 = "*when*", W_2 = "*where*", W_3 = "*who*" and W_4 = "*what*". The set of attributes of a dimension W_i is $A_i = \{A_{i1}, \ldots, A_{iN_i}\}$, where N_i is the number of attributes of W_i and A_{ij} represents the attribute j of dimension W_i.
Let $D(A_{ij})$ denote the domain of A_{ij} and $\#D(A_{ij})$ its cardinality, and $D(A_{ij}|P)$ denote the domain of A_{ij} observed in P and $\#D(A_{ij}|P)$ the number of distinct values of A_{ij} in P. The attribute values are textual, human-understandable descriptions of W_i at different conceptual levels. Examples of attributes are *Time-of-day* and *Season-in-year* for W_1, *Country* and *City* for W_2, *Name* and *Kinship* for W_3, and *Activity-type* for W_4. The values for each attribute are derived from the photos' metadata or are introduced by the users, or both of them, as explained in Chapter 4, Archival. We assume the attributes for W_1 and W_2 are always present. This is a viable assumption as the photo's timestamp is ubiquitous in digital cameras, and the localization data is becoming more reliable in recent smartphones [WM10]. The attributes for W_1 and W_2 were set during the archival of roll, as they are the mandatory part of the photo's context. As already discussed (see Chapter 3 for more details), the attributes are presented in different levels-of-detail. It is important to know their order of specificity, for summarisation purposes.

Definition 5.1 (less specific relation). *The less specific relation, denoted by $<_s$, states the semantic of one attribute A_{ij} is less specific than another A_{ik}.*

For example, *Month* is less specific describing an instant in time than *Day*. Thus, we can represent that relation by *Month* $<_s$ *Day*. A relation $<_s$ is transitive and defines a partial order over attributes of the same dimension. The information about$<_s$ is stored in the knowledge base (KB). When two attributes have the same specificity, we denote the relation by $=_s$. This happens, for example, when the two are a translation of the same concept in two different languages, like *Year* and *Année*. It is possible that two attributes are incomparable in terms of their specificity. For instance, *Quarter* inc_s *Season*, because they refer to a similar division of an year, and thus no one is less or more specific than the other, but their are not equal also. We denote that relation by inc_s, where

$$A_{ij}\ inc_s\ A_{ik} \Rightarrow\ \neg(A_{ij} <_s A_{ik})\ \wedge \neg(A_{ik} <_s A_{ij})\ \wedge \neg(A_{ij} =_s A_{ik})$$

Given a partial order $(<_s, =_s, inc_s)$ based on the attribute specificity, we can define a compatible total order $(<_s^r, =_s^r)$, named *relaxed specificity*, where X and Y are attributes of A_i.

$$\begin{cases} X <_s^r Y \quad \text{if} \quad X <_s Y \ \vee \\ \qquad\qquad\quad (X\ inc_s\ Y \wedge\ \#D(X) < \#D(Y)) \\ X =_s^r Y \quad \text{if} \quad X =_s Y \ \vee \\ \qquad\qquad\quad (X\ inc_s\ Y \wedge\ \#D(X) = \#D(Y)) \\ X \leq_s^r Y \quad \text{if} \quad X <_s^r Y \ \vee X =_s^r Y \end{cases} \tag{5.1}$$

When the attributes are incomparable, their cardinality is used to decide which is more specific.
For summarisation purposes, we use (5.1) over a subset of A_i, to produce a non-decreasing ordered sequence governed by the attributes' relaxed specificity, denoted by A_i^r, where

$$\forall A_{ij}^r, A_{ik}^r \in A_i^r : A_{ij}^r \leq_s^r A_{ik}^r \Leftrightarrow j < k, i \in [1..4]$$

We define a matrix M_i for each dimension W_i, where the rows represent the photos in P, and the columns are given by A_i^r. Thus, the first column of M_i is a less specific attribute and the last column is one most specific attribute. We assume that all the elements of M_i have values.

The summarisation problem is stated as follows. Given a set of photos P, represented by the set $M = \{M_1, \ldots, M_4\}$, we want to reduce the data presented to the user, keeping the necessary information so they understand the context. The summary comprises:

1. a partition over the set of photos,
2. a small textual description for each element of the partition, and
3. a selection of an auxiliary grain of detail.

Figure 5.9 sketches the output of the algorithm. In this example, P contains photos from 1 week visit to the Dublin, summarised by 4 clusters. The description of each group is the most specific that is suitable to all photos that it contains. For example, the second group contains photos taken in several days in January in Dublin, and so, the description that is more specific if *Jan 2014,Dublin*. The selected detail for "*when*" is the *month*, showing the temporal range at that level and the distribution of photos in each. Notice that January 2014 has more photos than December 2013. For "*where*", the *POI* was selected to show the spatial distribution within the group. The "*what*" shows the annotated activities in the set, made during the archival of the photo. The "*who*" was left empty on purpose, to demonstrate that if the information for one dimension is not completely filled it cannot

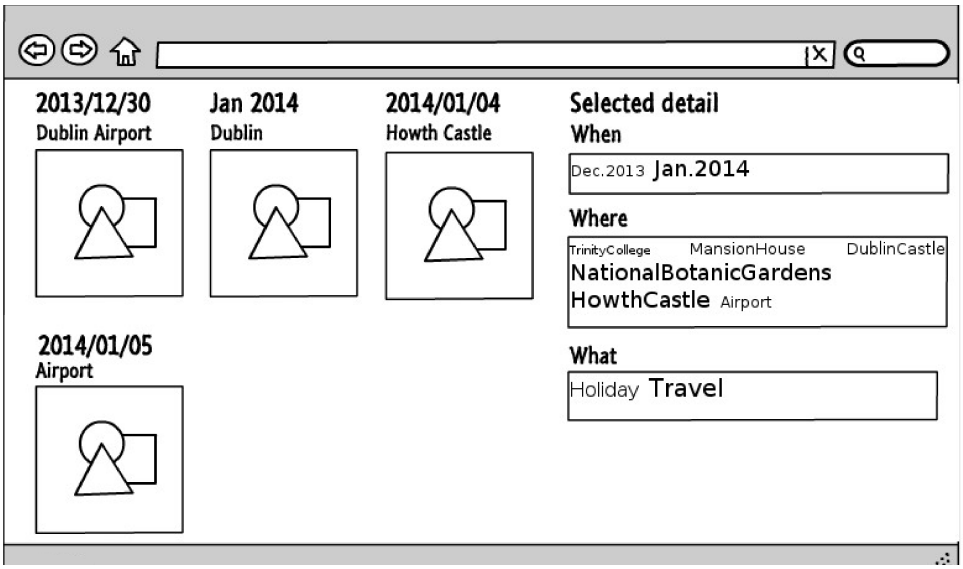


Figure 5.9: Example of MSS in a user interface.

be used in the summarisation algorithm. Using a tag cloud approach, it is possible to represent, in a concise manner, the distribution of photos for each term. Besides, each term is selectable, enabling the interface to give a visual hint about the groups where that term, or conjunction of terms, is present. For example, selecting `Airport` should highlight the first and fourth group.

5.4 Multimedia Short Summary algorithm

In this section we will describe the three components of the MSS algorithm:

1. a partition $Q = \{Q_1, \ldots, Q_k\}$ over P;
2. a description of each Q_i using one value for a selected column of each M_i;
3. a concise detail for P using, for each M_i, all values of a chosen column.

From this point forward, and without loss of generality, we consider only the temporal and spatial dimensions. Nevertheless, MSS extends to the four dimensions, as long the information is complete for every photo. This means that every step of the algorithm, described next, can be made for any of the dimensions, as illustrated in Figure 5.9.

5.4.1 Clustering multimedia objects

The clustering algorithm is used to present a summary of P. The summary is a partition Q, limited in size, as the capacity of working memory is a well known bottleneck in human information processing [BH74]. We consider 16 clusters as the largest accepted value, that enables a succinct presentation of P at most in a 4×4 matrix. When the

temporal order of the photos is kept intra- and inter-cluster, the algorithm produces *temporal dominated* clusters. If the places with equal descriptions are kept together, then Q has *spatial dominance*. The clustering algorithm has three inputs, namely, the limit for the number of clusters (κ), the set M and the type of dominance. For a better explanation of the algorithm we introduce the concept of *less specific discriminant attribute*.

Definition 5.2 (less specific discriminant attribute). *The* less specific discriminant attribute, *denoted by* $A_i^{LSD(\kappa)}$*, with at most* κ *distinct values, is the first attribute that guarantees*

$$A_i^{LSD(\kappa)} = \begin{cases} A_{ij} & \text{the first left most attribute where } 1 < \#D_s(A_{ij}^r|P) \leq \kappa \\ A_{i1} & \text{if } \forall j, \#D_s(A_{ij}^r|P) > \kappa \ \vee \ \#D_s(A_{ij}^r|P) = 1, j \leq N_i \end{cases} \tag{5.2}$$

Year	Season	Month	WeekDayName	Day	TimeOfDay	Hour
2014	Winter	01	Thursday	02	night	07 p.m.
2014	Winter	01	Friday	03	night	08 p.m.
2014	Winter	01	Saturday	04	morning	08 a.m.
2014	Winter	01	Saturday	04	morning	08 a.m.
2014	Winter	01	Saturday	04	morning	09 a.m.
2014	Winter	01	Saturday	04	morning	09 a.m.
2014	Winter	01	Saturday	04	morning	09 a.m.
2014	Winter	01	Saturday	04	evening	05 p.m.
2014	Winter	01	Sunday	05	afternoon	01 p.m.

Table 5.1: Example of a matrix M_1, for illustrating the behaviour of the MSS.

Table 5.1 illustrate the M_1 matrix and the attributes that were used to describe the temporal part of the context, namely, the `Year`, `Season`, `Month`, `WeekDayName`, `Day`, `PeriodOfDay`, and `Hour`. For M_2 the attributes used were the `Continent`, `Country`, `City`, and `Place`. The M_1 values for each attribute are uniquely represented using the complete date, but for presentation purposes, their values were simplified. The rational for $A_i^{LSD(\kappa)}$ is straightforward. As the groups should be a concise summary for the photo set, we should start looking from the most general attribute until we reach one whose number of values are able to generate at most, the number of groups needed. Using the example in Table 5.1, if $\kappa = 4$, by applying (5.2), we stop at *WeekDayName*, since the number of distinct values are four: *Thursday*, *Friday*, *Saturday* and *Sunday*. However, if $\kappa = 2$, we can see the number of values in the attributes are either 1 (*Year*, *Season*, and *Month*), or greater than κ. In that situation, the less specific discriminant attribute will be the *Year*.

The clustering algorithm starts selecting the $A_i^{LSD(\kappa)}$ for each matrix. This method was inspired in the attribute-oriented induction [HF96], although we do a specialisation instead of a generalisation.

The clustering is done by grouping elements from P that share the value for the $A_i^{LSD(\kappa)}$

attribute in the dominant matrix. As mentioned earlier, we can choose the type of dominance in MSS. That choice has impact on the order or evaluation of each matrix. If the type of dominance is temporal, the order of evaluation are M_1, followed by the M_2. Otherwise, M_2 is the first. For each group in the dominant matrix, we perform recursively the same action in the other matrix, using its own $A_i^{LSD(\kappa)}$. This creates a two level tree, where each cluster in the top level has one or more clusters in the bottom level. The cluster solution is found in the lowest level where the number of clusters do not exceed the κ value. Listing A.4 shows the pseudo-algorithm for the clustering procedure.

5.4.2 Descriptors for a cluster

The spatio-temporal context is distinct among clusters, so we can briefly describe each one using a textual value from a spatial attribute and another value from a temporal attribute[1]. To be effective, the a cluster description should be:

1. textual,
2. short — to be effective communicating the context of the photos in the cluster, and
3. useful — it should add value to describe the context of the photos in the cluster.

The description is based in attribute values of M_i. To keep a short description, each group is described with just one value from one attribute of each matrix. The chosen value (and the chosen attribute) has to be as representative as possible and simultaneously, as specific as possible. As an example, consider a set of 200 photos, taken on two consecutive days in 2012. 20 were taken in the Summer and 180 in the Spring. Describing the set as "*Photos from 2012*" is correct. But since it is to vague, it lacks informative power. If we use the description "*Photos from Spring, 2012*", we correctly describe 90% of the photos, and produce a more specific cue to show the temporal location of the set. We decided to sacrifice the precision, keeping the information useful for the users.
For each matrix, the attribute used to describe the cluster is picked among the most specific one, as long as:

- the most common value covers a minimum percentage of the photos in the group ($coverage$);
- the most common value is $ratio$ times greater than the second most common value.

If no attribute satisfies both conditions, we choose the A_{i1}^r.
Using the examples in Tables 5.1 and 5.2, we will exemplify the selection of one value from M_1 and another from M_2 to describe each cluster. Let us assume the if we use $\kappa = 4$, and the clusters found were $\{1\}$, $\{2\}$, $\{3, 4, 5, 6, 7, 8\}$, $\{9\}$. Using $coverage = 80\%$ and $ratio = 3$, the values from M_1 for describing each cluster are:

[1] If other dimensions were included, they are also included in the textual description.

- *7 p.m.* — `Hour` — for the first cluster. Since it contains only one case, the description is always set to the most specific value.
- *8 p.m.* — `Hour` — for the second cluster, for the same reason.
- *morning* — `TimeOfDay` — for the third cluster, since that attribute its the most specific one whose $coverage \geq 80\%$ (83%) and the $ratio \geq 3$ (5:1).
- *1 p.m.* — `Hour` — for the last cluster.

The values from M_2 for describing each cluster are:

- *Abbey Theatre* — `Place` — for the first cluster. Since it contains only one case, the description is always set to the most specific value.
- *Dublin Castle* — `Place` — for the second cluster, for the reasons already mentioned.
- *Dublin* — `City` — for the third cluster.
- *Airport* — `Place` — for the last cluster.

Continent	Country	City	Place
Europe	Ireland	Dublin	Abbey Theatre
Europe	Ireland	Dublin	Dublin castle
Europe	Ireland	Dublin	Trinity College
Europe	Ireland	Dublin	Ha' Penny Bridge
Europe	Ireland	Dublin	Temple Bar
Europe	Ireland	Dublin	City Hall
Europe	Ireland	Dublin	Christ Church Cathedral
Europe	Ireland	Dublin	Christ Church Cathedral
Europe	Ireland	Dublin	Dublinia
Europe	Ireland	Dublin	Airport

Table 5.2: Example of a matrix M_2, for illustrating the behaviour of MSS.

5.4.3 Selection of a proper level-of-detail

As described in the two previous sections, we partition the original set into a reasonably small number of subsets ($\leq$ than 16). The previous step settles a local description of each cluster. In this section we will describe how we find a set of textual descriptions for each dimension, allowing a globally description of P at a proper level of detail. Using a global selected detail has two major advantages:

1. shows a brief summary, starting at a given conceptual level, that complements the description of the clusters, and

2. provides a way to filter on demand each dimension separately.

Thus, we want a level where the information is detailed and limited, but sufficient to disclose the underlying context. For each matrix M_i, we select the most specific attribute that verifies $\#D_s(A^r_{ij}|P) \leq L_i$. The value for L_i considers the nature of each dimension, enabling MSS to choose any attribute as a level of detail. It selects only less specific attributes when the most specific ones present higher cardinalities. The idea is to become more imprecise describing the set being summarise, if it reduces the number of information presented to the user.
For M_1, the attributes we use have values related to a specific date. For example, *08 a.m.* is represented differently in two different days. This means that it is very likely the cardinality of the `Hour` attribute will rise when the number of day increase. But then, probably the indication of which days are included in the set is more useful than the number of hours. During development, we found that using $L_1 = 12$ adapts MSS to the temporal dispersion of the sets, showing a good balance between relevance and compactness. For example, for one year sets, the proper level of detail can be set to the `month` attribute, it there are photos from every month. Using the example in Table 5.1, we see that `Hour` has 6 different values. Thus, for that example, the selected level of detail will be the hour.
For M_2, the attributes used show great cardinality differences between the two most specific attributes. The analysis of several geotagged personal set of photos confirmed that they include many places. If we use a small value, e.g. $L_2 = L_1$, MSS commonly choose the *City* as a level of detail. This lowers detail relevance, since many sets have just one city. We found that, for the spatial information, a good compromise between relevance and compactness is setting $L_2 = 40$. Using the example in Table 5.2, we see that `Place` has 9 different values. Thus, for that example, the selected level of detail will be the place.

5.4.4 Final remark

Given a set P with n photos, the MSS output is:

1. a partition for P;

2. a short textual description for each group in the partition; and

3. a set of textual descriptions used to globally describe the context of P.

This output depends, mainly, on the spatio-temporal context of the photos, not on their number. The MSS time complexity is, in worst case, $\mathcal{O}(n \log(n))$, where n is the number of photos, provided the number of attributes used to describe the photos is much lower than the number of photos of P. Setting the partition is the major contributor to this complexity. MSS is computed fast, even for sets with thousands of photos, making it usable to take part in a user interface.

Using the example described inTables 5.1 and 5.2, the overall result of the MSS is the following. The clusters and respective descriptions are:

Cluster 1 : 7 p.m., 02/01/2014 @ Abbey Theatre;

Cluster 2 : 8 p.m., 03/01/2014 @ Dublin Castle;

Cluster 3 : morning, 04/01/2014 @ Dublin;

Cluster 4 : 1 p.m., 05/01/2014 @ Airport.

The proper level-of-detail for the temporal information is:

$$07p.m. \mid 08p.m. \mid 08a.m. \mid 09a.m. \mid 05p.m. \mid 01p.m.$$

The proper level-of-detail for the spatial information is:

$$AbbeyTheatre \mid Dublincastle \mid TrinityCollege \mid Ha'PennyBridge \mid$$
$$TempleBar \mid CityHall \mid ChristChurchCathedral \mid Dublinia \mid Airport$$

5.5 Summary

In this chapter we described the retrieval of a set of photos, using **MeMoT**. The usage of the KB enable the usage of hints during the retrieval, that provide the semantic of the terms used in the query. This is helpful to increase the retrieval performance, but more important, it is essential to do the viewpoint adjustment in behalf of the users. This adjustment increase the change of getting the wanted photos using user dependent query' cues, even if the photos were archived and annotated by other user. It also enables the manipulation of the context' description in behalf of the user, particularity for setting concise description of the sets. We concentrate on the summarisation of the result set, as a mechanism to deal with large collections and yet, retaining the necessary information to construct the context in therms of space and time. Although we developed the algorithm to be used in photo sets, it can be used with any set objects, with spatio-temporal descriptors. The summarisation is performed using the *Multimedia Short Summary* algorithm, that uses concept generalisation to summarise a set of temporal and spatial referenced photos. The current version of the algorithm implies the information for each dimension is complete, i.e., not allowing missing information for any value in the attributes. Since the information for M_3 and M_4, representing the "*who*" and "*what*" respectively, requires user intervention during the archival, this is more difficult to achieve. Nevertheless, all the step are extended to any dimension, as long as the information is complete for them. This means the clustering, the description of each cluster and the proper, global, level-of-detail can be set using information from M_1, M_2, M_3 and M_4.

The evaluation of the algorithm is presented at Section 6.2, considering only the spatio-temporal information.

6

Experiments and Results

> *What most experimenters take for granted before they begin their experiments is infinitely more interesting than any results to which their experiments lead.*
>
> ***Norbert Wiener***

This chapter is concerned with the evaluation of the algorithms presented in Chapters 4 and 5. Both algorithms share the same method of assessment. In a first phase, the algorithms were evaluated using systematic tests, towards the characterization of their response to a parametrization change. In a second phase, both algorithms were assessed in user testing, so we can understand their performance in 'the wild'. Next, we first describe the evaluation of LDES, followed by the evaluation of MSS. In the end we draw some conclusions based on the evidences resulting from the tests.

6.1 LDES algorithm evaluation

The LDES algorithm, discussed in Chapter- 4, produces a segmentation of a set of photos based on their temporal and spatial information. We control LDES using a set of parameters, namely: *(i)* w, affects the logical day assignment; *(ii)* f_t, affects the way temporal information is segmented; and *(iii)* f_g, affects how the the spatial information is segmented. The algorithm was evaluated against several personal collections of photos to determine:

1. the impact of the parameters on the output of the algorithm;

2. the qualitative value of segmentation.

Stats	Stats.	No. Photos	Photos w/ Geo. (%)	Day range	Km range
Min.	2	16	72	3	3
1st Qu.	4	78	100	7	71
Median	9	156	100	10	579
Mean	10	212	98.7	144.6	1690.9
3rd Qu.	13	248	100	20.5	2353
Max.	41	1395	100	1738	9459

Table 6.1: Descriptive statistics for the dataset used in the experiments.

The first item was assessed using a set of experimental tests, whose descriptions and results are reported in Section 6.1.1. The second item was assessed using experimental user tests, whose description and results are reported in Section 6.1.2.

6.1.1 Sensitivity and compatibility tests

For this tests, we have used 39 photo sets gathered from personal collections of holiday photos, most of them available at Picasa Web Albums. They average 221 photos, ranging from 16 to 1395. The temporal range varies from three days to four years, while the diagonal of the bounding box of photo set ranges from 3 km to 9459 km. We have included a photo set with photos from two years apart, to see how the LDES handle this cases. Table 6.1 summarises the photo sets used. For a more comprehensive information, please consult Table B.1. It resumes the distribution of the numbers of photos in the datasets, the percentage of photos with spatial information, the temporal range of the datasets, expressed in days, and the range of the bounding box, expressed in km.

In this experiment, we decided not to refer to the sets of photos as rolls. The reason relates to the way the sets were collected. Since they are posted online, by their owners, they may have already went through a selection and filter processing, which is not quite the idea behind the notion of roll. Nevertheless, the photo sets maintain other characteristics of the rolls, namely, their heterogeneity in terms of temporal and spatial dispersion and their diversity of depicted items. Figure 6.1 shows the histogram for the number of days in the photo sets. It resembles a reverse J-shaped distribution, and exhibits a power law.

6.1.1.1 Exploring LDES sensitivity to parametrization

The first set of analyses examines the impact in the segmentation of changes in the parameters f_t and f_g. The first, f_t, controls the threshold that governs the detection of the temporal gaps for creating new segments, ranging between 0.1 and 0.9. This parameter affects the cardinality of segmentation, i.e., when it is small it forces the algorithm to segment more; when is large it produces larger and fewer segments. The second, f_g, governs the sensitivity to spatial gaps betweens photos, raging in the interval $[0..1]$. Smaller values make LDES less sensitive to changes in location, producing less segments. We settled

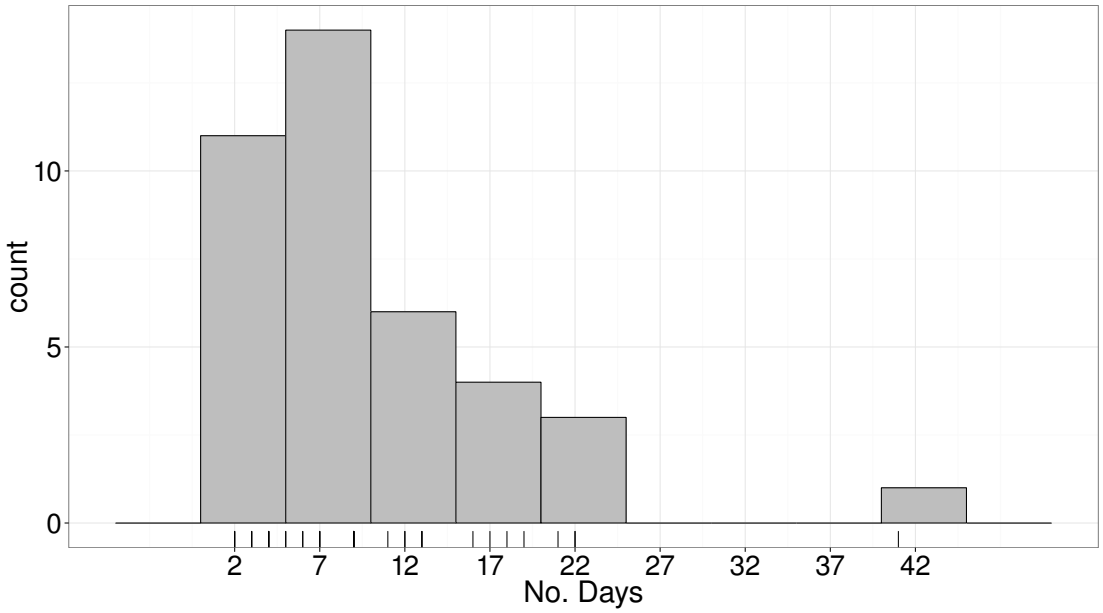


Figure 6.1: Characterisation of the distribution for the no. of days in the photo sets.

$w = 4$, because it is assumed that most of the people sleep more than 4 hours a night. Thus, gaps larger that this value will stop the assignment of the logical day to a regular day.

We evaluate the impact that a change in f_t has in the two first steps of the algorithm. For each photo set, we make 5 runs. In each one the parameter f_t is changed using the values $\{0.1, 0.25, 0.5, 0.75, 0.9\}$. Since the number of photos in each set is different, for comparison purposes, we use the *relative number of segments*, denoted by $\#S_r$, given by

$$\#S_r = \frac{|S|}{|T|}$$

where $|S|$ is the number of segments the segmentations, and $|T|$ gives the numbers of photos in the dataset. Do notice that $0 < \#S_r \leq 1$.

Figure 6.2 illustrates the runs. They are represented in the x-axis, and the *relative number of segments* is represented in the y-axis. Besides the representation of the five used in box-plots [MTL78], the mean is also represented by a black diamond shape. As expected, the number of segments decreases when f_t grows. We can also see a decrease in the variance, as the value of f_t increases. The changes in the parameter f_t produces segmentation with different number of segments. However, what is needed to answer is what are the relations those segmentation have with each other. For example, does the type of relations vary with the increase of the f_t or it stands still? This analysis was performed using the relations proposed in Chapter Archival, section 4.2.2. Figure 6.3 is and illustrative example on how the relations change with different values for f_t. It shows a square matrix where each cell represents the relation between two segmentations, produced with the f_t depicted in the axis. For example, a segmentation produce with $f_t = 0.1$ is a refinement

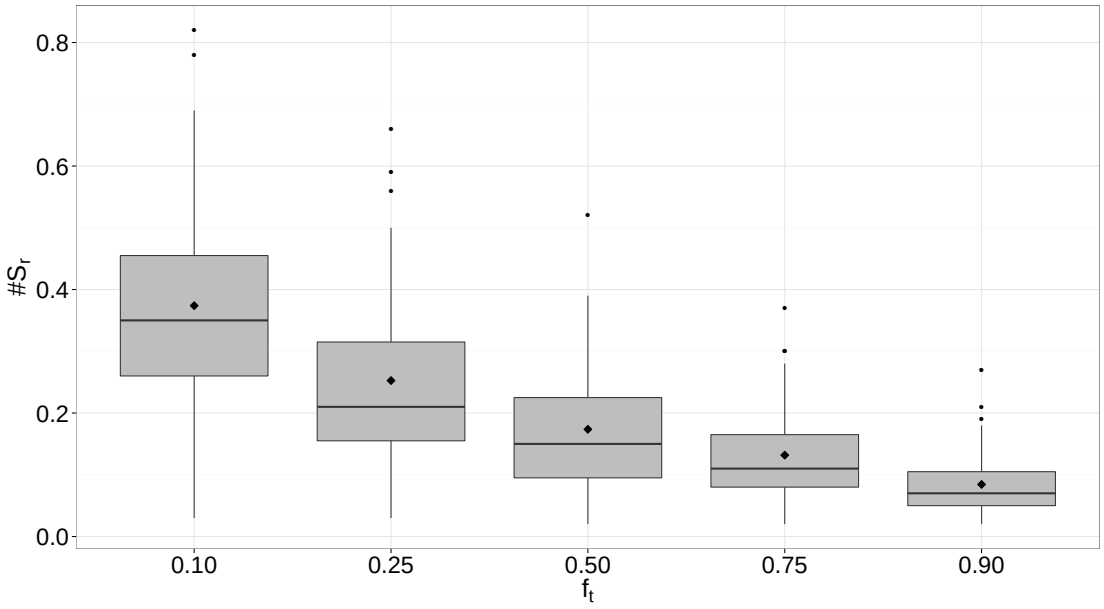


Figure 6.2: Variation of the relative number of segments when f_t changes, considering only a temporal segmentation.

of a a segmentation produced with $f_t = 0.75$ in the axis of the cell. Analysing the matrices for the 39 datasets, we found only two relation — `equal` and `refinement`. This result is important, since it demonstrates that changing the f_t parameter produces compatible segmentations, at most, a refinement of the other. We refer to this behaviour as a change in the zoom at which the algorithm "set the events". We count the occurrence of each

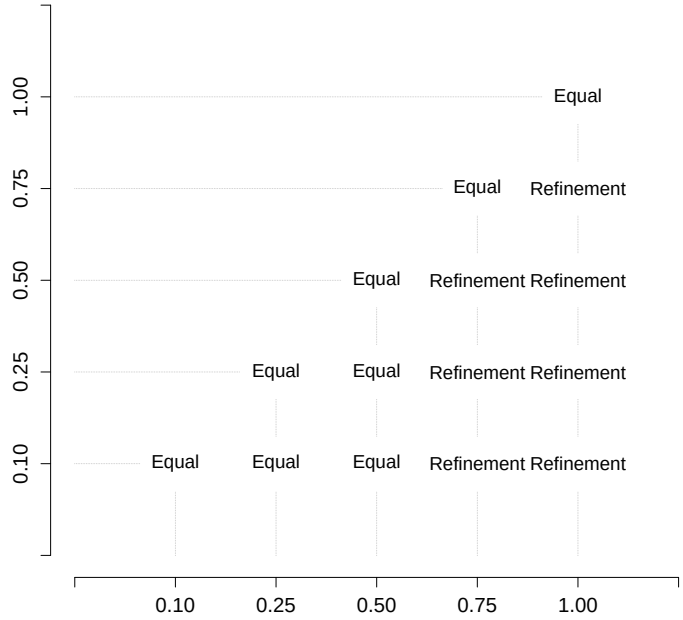


Figure 6.3: Illustration of variation in the the relations between segmentations, when f_t changes, for one dataset.

relation, comparing one parametrization with the other four. It was observed that the number of refinement relation varies between $[97\%, 99\%]$. This means that, changing the f_t will produce, more than 97% of the time a refinement of a segmentation.

There are no "golden rule" to decide what a good segmentation is, since it depends on the preferences of the user. Setting $f_t = 0.5$ is a recommended design value, that stands in the middle of the scale, providing a good separation of bursts in several scenarios. It allows, for example, users to visually explores the segmentations of a photo set, by changing the value of f_t online and see the result that suits them best.

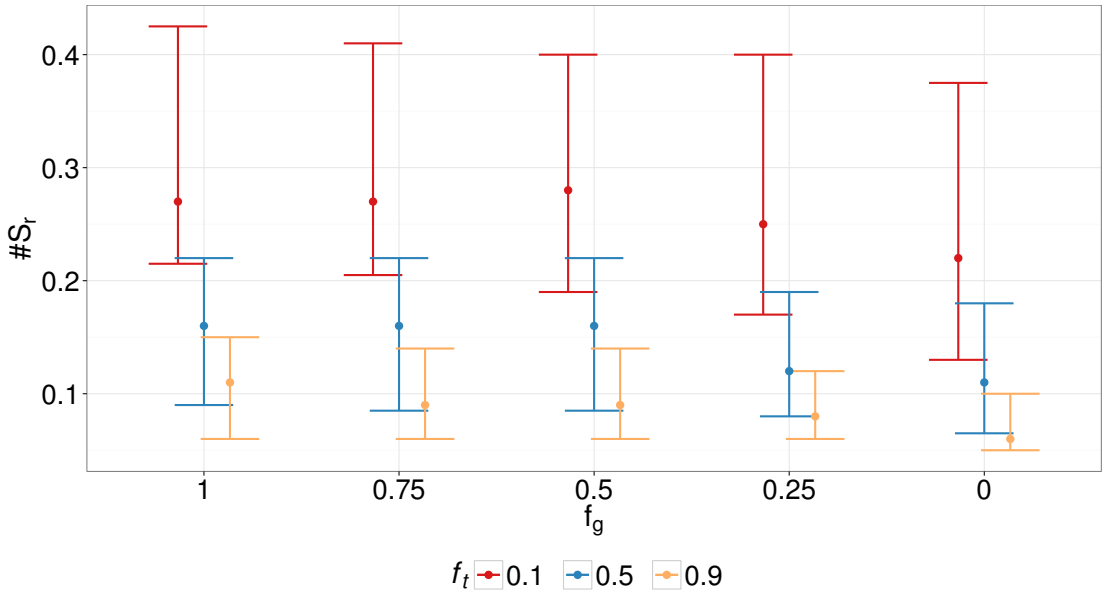


Figure 6.4: Variation of the relative number of segments when f_g changes, considering multiple values for f_t.

The analysis of the effect on changes in parameter f_g were done in conjunction with changes in f_t. This is a direct consequence of the algorithm design. Do note the spatial information is used to tune a temporal segmentation. Thus, the choice a value for f_t affects the result of whatever f_g is chosen. For a given value of f_t, the parameter f_g was changed using the values $\{1, 0.75, 0.5, 0.25, 0\}$ respectively. This procedure was repeated for three values of f_t: $\{0.1, 0.5, 0.9\}$. The results are showed in Figure 6.4. Each bar depicts the median (middle), the first quantile (bottom) and the third quantile (top) for the relative number of segments. As we can see, using a lower value for f_t produces an broad inter-quantile than the one achieved using higher values for the same parameter. The median is also less stable for lower f_t values. Figure 6.5 shows how the *relative number of segments* changes with the value of f_g. As we can see, higher values for f_g tends to produce similar results, with the most significant changes occurring with lower values for f_g. Nevertheless, the segmentations produced are different, depending on the parametrization.

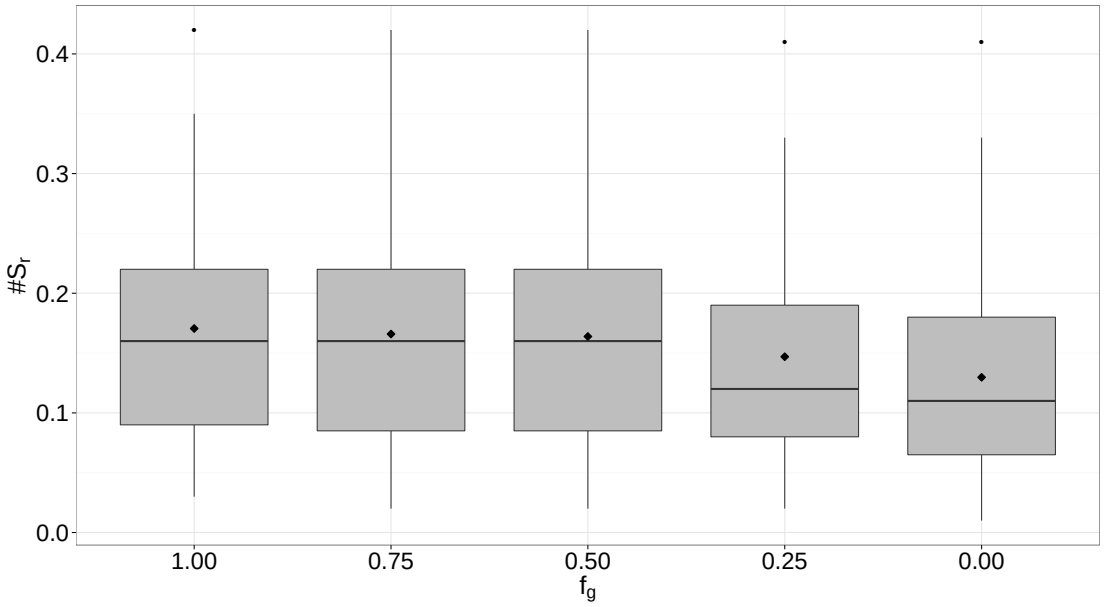


Figure 6.5: Variation of the relative number of segments when f_g changes, for $f_t = 0.5$.

Proceeding in a similar way as to the f_t parameter, the relation between segmentations were analysed, for different f_g values. The process is equal to the one described earlier, where the 39 square matrices were used, representing the relation of each segmentation with the other four. In this case, we see a richer of relations, as illustrated in Figure 6.6. The y-axis represents the frequency count. The results are consistent in the ones found earlier. The `refinement` relation is still the most common, with the `equal` one in second place. However, there are some `compatilbe` relations, specially with the default value in f_t. However, they are absent for $f_g \leq 0.25$. The `incompatible` relation, although it appears for $f_t = 0.1$ and $f_t = 0.5$, its expression is insignificant. This is one unanticipated finding, but important, since LDES uses the `join` operation to tune the segmentation using spatial data. Such operation is, in theory, prone to produce incompatible relations. However, the experiments show that, in practice, the algorithm tends to produce compatible segmentations despite the values of f_t and f_g. The results also show a more rich, less flat, response in terms of relations in contrast more stable result in terms of the number of segments. The extreme combination $f_t = 0.1$ and $f_g = 1$ produces all the four relations, while the other extreme combination $f_t = 0.9$ and $f_g = 0$ give us only `refinement` and `equal` relations. Other interesting finding is how changes in parameter f_g affects the curves for relations `equal` and `refinement`. It is visible in Figure 6.6 that they are negatively correlated; that is, the rise of one relation' count is followed by a similar decrease in the other' count.

As for f_T, deciding a "proper" default value for f_g follows the same assumptions:

- The notion of a proper segmentation depends on the analysis of the user, and so,

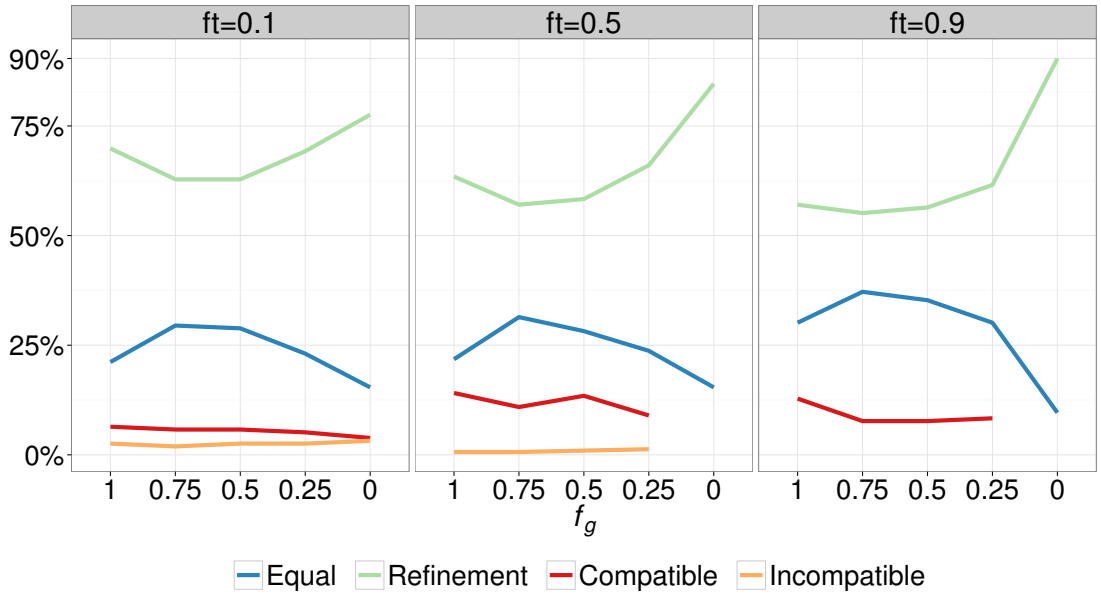


Figure 6.6: Relations between segmentations when f_g changes, considering multiple values for f_t.

is subjective. This means that we cannot say, for example, that a more refined segmentations is better;

- There is no obvious criteria to optimise.

Thus the default value was settle to $f_g = 0.5$, a value that stands in the middle of the interval variation of the parameter. This way, users can increase or decrease the parameter values, observing the changes in the segmentation and deciding which is best, given their preference.

The default values for f_g and f_t are used to settle the baseline parametrization of LDES that was used in an experimental user test, described in Section 6.1.2.

6.1.1.2 Evaluation of Context Segmentation

In this section, we evaluate the LDES segmentation against four baseline segmentations:

1. **LDES-T** — LDES using only temporal information, without spatial refinement;
2. **SO-T** — A segmentation using fixed social temporal markers, representing the periods of day. New segments are created for the hours $\{7, 12, 14, 17, 19, 21, 24\}$;
3. **SP-2** — A fixed spatial threshold segmentation. A new segment is created for every gap of 2 km between photos;
4. **SP-5** — Similar to the previous segmentation, but using a 5 km threshold.

The hours used to settle **SO-T** is an attempt to put bounds for known temporal periods, like *Morning* or *Evening*. Do notice, however, that those periods lack a formal and universal accepted range, and such hours are an interpretation of those periods modelled by a South Western point of view. For **SO-T** and **SP-5**, we try model a bounding box of a normal visit to a city or a place. In our vision, most of the photos are used to document events that happened in populated places. A part of those photos are taken during trips to cities. We think that 2 km and 5 km radius are a realistic upper bound for a context boundary. Is someone travels such distance without taking a picture, it is likely that the context has changed.

We use the PR_{error} [GCA06] and the WindowDiff [PH02] metrics to compare the segmentations. Some of them were used in others work for photo's context separation (e.g. [NSPGM04; GKS12]). Those measures compare a reference segmentation (the reference data) with an hypothesised segmentation, taking into account how far away is a hypothesise count point from a reference one. Those measures range from $[0, 1]$, representing *all cut points are equal* to *no cut points in common*, respectively. For PR_{error} we settle three different scenarios:

(i) equal costs for miss and false alarms (new segments starts);

(ii) false alarms (FA) costs three times higher than miss costs;

(iii) miss costs three times higher than FA costs.

Those scenarios help us see the variation of FA and miss in each baseline.

For each of the 39 datasets, there were created five segmentations: four are the baseline ones and the fifth is the LDES segmentation, that acts as the reference segmentation. Each of the four baseline segmentations were compared with the LDES using the four metrics described above. The results are shown in Figure 6.7. We also compare the segmentations using the relations proposed in Chapter Archival, section 4.2.2. The results are shown in Figure 6.8.

The evaluation demonstrates that **LDES-T** gets the most similar results, in all measures (Figure 6.7). This is an expected outcome, since the spatial information is used to fine tune the temporal segmentation. Few segmentations are incompatible to any produced by the temporal part of the algorithm, as showed in Figure 6.8. This result is explained by the fact that LDES uses the spatial information most of the time to split the temporal segmentation, producing compatible and refinement relations.

Another important finding was the results considering **SO-T**. If we look at Figure 6.8, it seems that **SO-T** is the most diverging of the four, since most of the relations are incompatible. However, in Figure 6.7 we can see the comparison results are low. The high number of incompatible relations indicates a disagreement in the cut points. However, they are close misses, that are less penalised by WindowDiff and PR_{error}, than false alarms or real misses. This means the temporal behaviour has a shift from the typical settlement

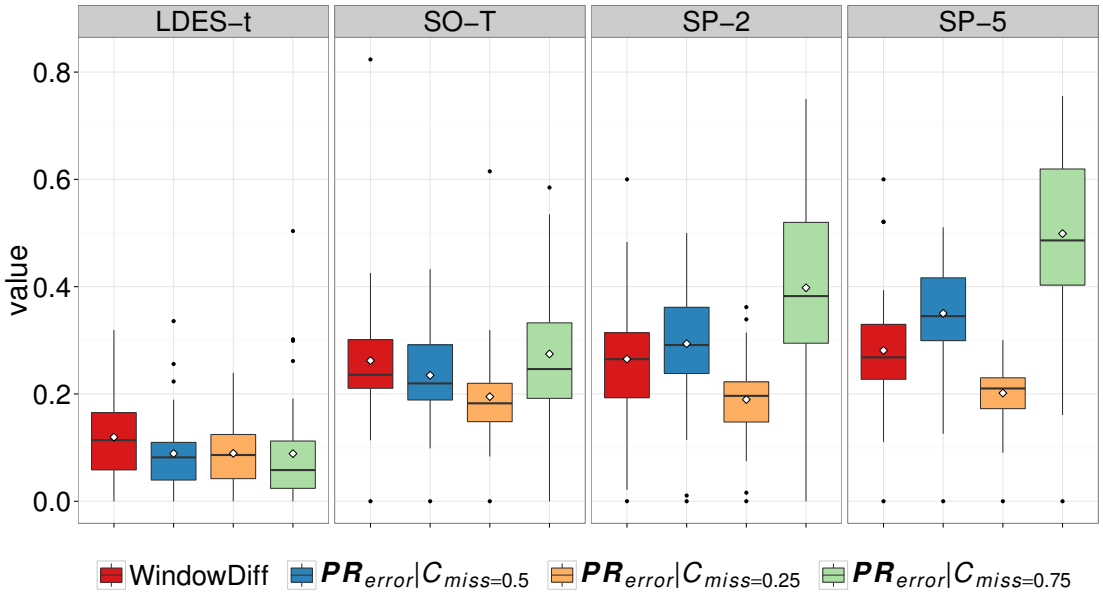


Figure 6.7: PR_{error} and WindowDiff results for different baseline segmentations.

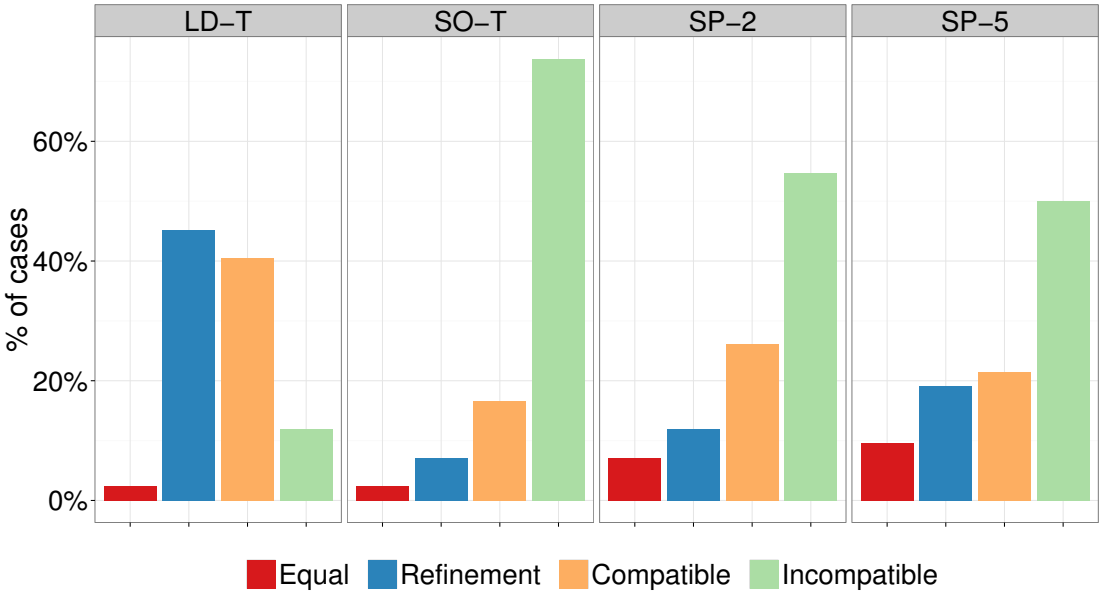


Figure 6.8: Relations for different segmentations.

for the day periods. This fact reinforces the need to consider the temporal cycles when handling personal photo collections.

The results, as shown in Figure 6.7, also indicate that larger spatial distances increase the dissimilarity between segmentations. A possible explanation for this might be that users confine their activities to small areas, e.g., part of a city, in each time period. When we

increase the spatial range, we force the removal of context divisions, increasing the artificiality of the context separation. Other evidence is the higher difference from LDES segmentation and **SP-2** and **SP-5**, comparing to **SO-T**. This reinforces the idea that temporal information is important to define the major cut points. The relation between segmentations in **SP-2** and **SP-5**, presents few modifications, as shown in Figure 6.8. There is, however, a larger difference between the `refinement` and `compatible` relation in the two cases.

6.1.2 Users test

The LDES was tested by a set of volunteers. The research statement of the study is that users accept the segmentations suggested by LDES, with minor changes.

The empirical study involved 14 participants who have provided us with some of their rolls (35 in total). There was a balance in terms of gender, with eight males and six females. Their age ranged from 21 to 55 years, covering a wide range of ages (see Figure 6.9). Most of them are computer science students. We provided them with some guidelines to help the roll selection, namely:

1. the photos should have location information;

2. the temporal range of each roll should have more than 1 day;

3. the rolls should reflect real sequences of photos, without a pre-selection.

The participants were responsible for selecting the rolls. For the record, we used all of the provided sets, even though some do not fulfil all three guidelines.

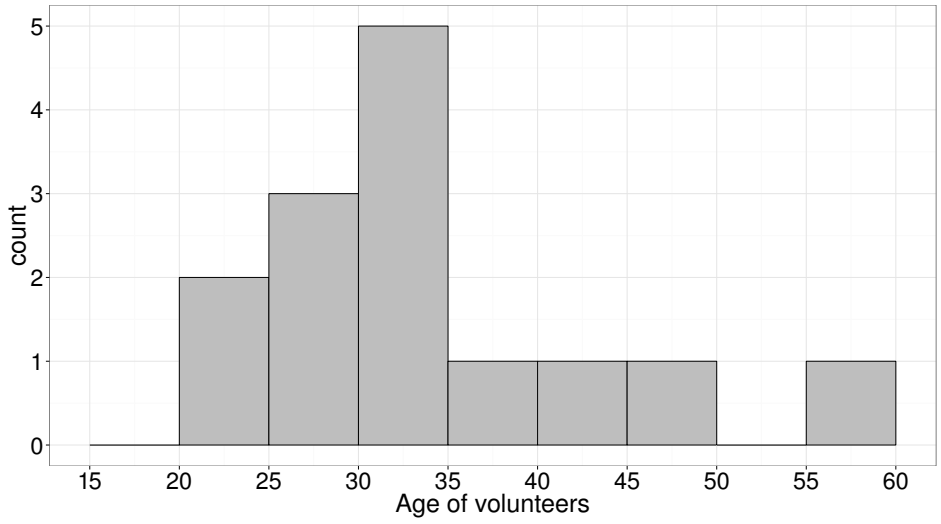


Figure 6.9: Age distribution of the participants in the LDES study.

6.1.2.1 Characterisation of the rolls

The 35 rolls are summarised in Table 6.2. Most of them have localisation information, with 68% of georeferenced photos. However, five rolls contain only photos with temporal information. In terms of temporal range, about 75% of the rolls range between 1 to 4 days approximately.

Stats.	No. Days	No. Photos
Min.	1	5
1st Qu.	1.5	28.5
Median	3	48
Mean	4.2	101.3
3rd Qu.	4	185
Max.	23	351

Table 6.2: Descriptive statistics for rolls used in the study.

From the distribution of the number of days, depicted in Figure 6.10, along with an observation of the rolls, we can tell that participants select the rolls to be, mostly, from weekend trips and one-week holidays.

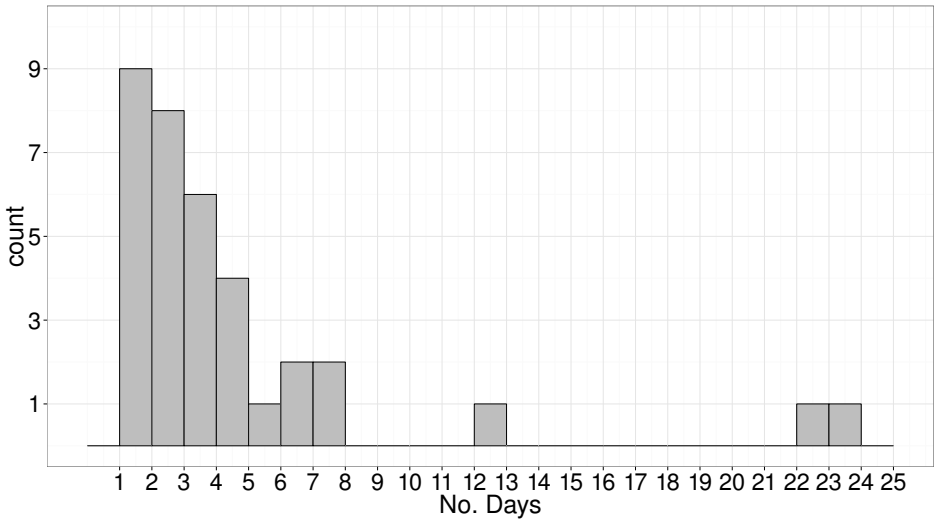


Figure 6.10: Characterisation of the distribution for the no. of days in the rolls.

The camera types used to capture the photos are shown in Figure 6.11. As we can see, more than 60% of the participants choose rolls taken from their smartphones. Thus, the 68% of geo tagged photos can be explained, since the smartphone models used are equipped with localisation services and have a built-in GPS tracker.

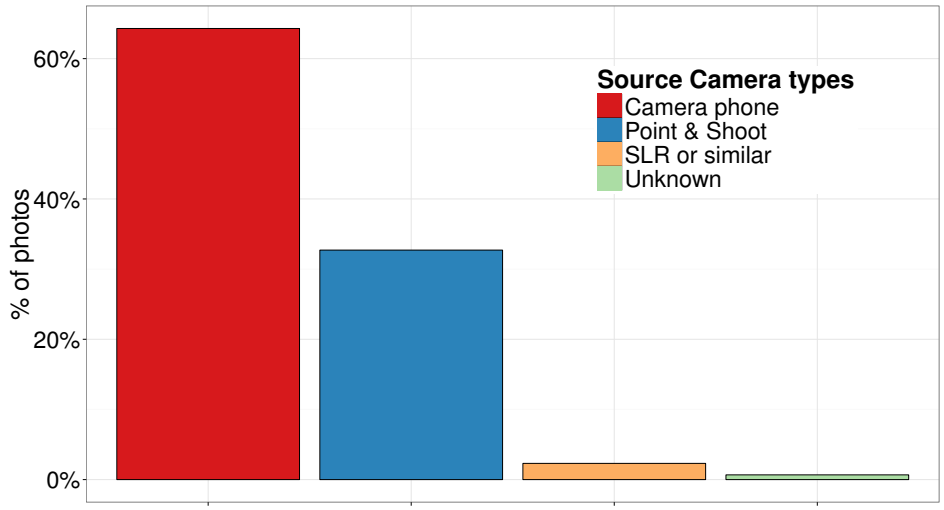


Figure 6.11: Type of cameras used to take the photos in the rolls.

6.1.2.2 Test Design

The experimental unit in this empirical user test is the pair $< participant, roll >$. There are some participants that share some rolls. Two rolls are shared between two participants each, and four rolls are shared between two participants. Those rolls represent situations where the participants are simultaneously the photographer and the subject. The segmentations are presented independently to each user. Thus, the 14 participants, interacting with 35 different rolls, makes the 43 experimental units in the test.

Figure 6.12 describes the flow used in the experimental user tests, made using a web application. It includes:

1. a *questionnaire*;
2. a *training phase* and;
3. the *test* itself.

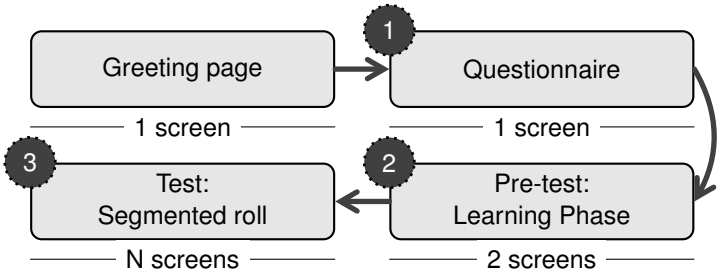


Figure 6.12: Representation of the test flow.

Step 1: The questionnaire In a first phase of the test, users need to complete a questionnaire with general questions about how they manage their photo collections. Most of

them are closed-ended questions, presented as follows:

1. *Sex*: ○F ○M;
2. *Age*: _____([18..120]);
3. *What type of camera you use most?*:
   ☐Cameraphones (iPhone, Android, ...), ☐Point and Shoot Cameras, ☐Reflex or similar;
4. *When you geotag your photos...*:
   ☐I have no idea what is geotagging, ☐the camera does it for me, ☐with a GPS tracker, ☐by hand, with less detail, ☐by hand, with as much detail as possible, ☐I do not geotag photos;
5. *What program do you use to manage your collection of photos?*:
   ☐None, ☐Aperture, ☐Digikam, ☐F-Spot, ☐iPhoto, ☐Picasa, ☐Lightroom, ☐Shotwell, ☐Windows Live Photo Gallery, ☐Other: _____;
6. *What online library do you use to publish your photos to?*:
   ☐ None, ☐ Facebook, ☐ Flickr, ☐ Instagram, ☐ PicasaWeb, ☐ SmugMug, ☐ Other: _____;
7. *To which online storage service do you save your photos?*:
   ☐None, ☐Dropbox, ☐Google Drive, ☐iCloud, ☐Microsoft OneDrive, ☐MyShoebox, ☐Other: _____;

With the exception of the first two questions, all the others allow multiple answers.

Step 2: The training phase After the questionnaire, the participant passes to a *pre-test learning phase*. The goal is to let the participant learn how photos are presented in the segments and how to change the segmentation, permitting an exploratory interaction. The learning phase has two steps, that share a similar layout:

1. a first one, introduces some simple terms and calls the participant attention to the way the segmentation is presented, showing the basic interaction, and
2. a second one, where the participant can freely interact with the user interface (UI), changing the segmentation at will. This includes the creation of new segments, and changing photos from one segment to another.

A participant can repeat the learning steps, as it is possible to go back and forth thorough them. Figure 6.13 depicts the first step of the learning phase. The left-hand side, marked as ⓐ, displays a representation of a segmentation. The tooltips indicate the locations of the labels, what a segment is, and identifies the contents of a segments — photos. The photos are labelled with letters, so the order can be checked, after the participant changes

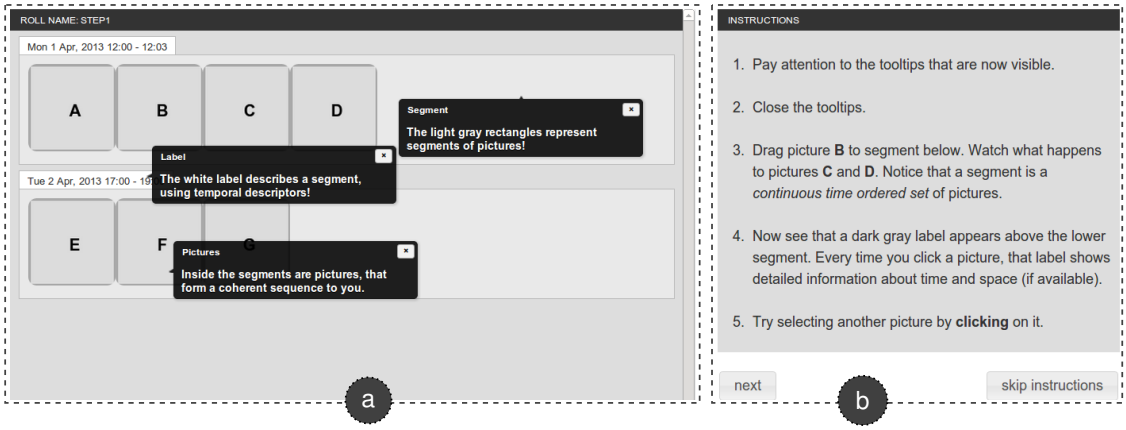


Figure 6.13: Interface for the learning phase.

the segmentation. On the right-hand side, marked as (b), the instructions are presented. They are just a simple script of actions, pointing out the artefacts depicted and the way we can interact with them. The left-hand side layout is, in every way, a representation of the test interface. The participant can, at any time skip the instructions. However, before the test begins, a pop-up asks to confirm that action.

Step 3: The test We take great care on how the participants "see" the segmented collections during the test. We have developed a minimalist interface, yet, powerful enough to allow changes in the segmentation, in a simple way. Citing [May99], "We want powerful functionality, but a simple, clear interface. We want ease of use but also ease of learning". To achieve these goals, the development cycle included several empirical tests [Nie94] using real users. Those tests allow us to tune the interface and the flow of the dialogues. Figure 6.14 shows the test interface, depicting a sample roll. We use some of the Gestalt principles [Joh+10] for representing a segment. The *proximity* law was applied to the photos of one segment, reinforcing the idea of group. The *figure/ground* law was also applied, making contrast between the background of the test and the foreground of the segments. Both principles reinforce the perception of a group. Each segment is annotated with a short description of the temporal information of the photos in a segment. The annotation consists of the day and hour range, representing the timestamps, to the minute, of the first and last photos in the segment (see Figure 6.14, (a1)). If there are photos from more than one day in a single segment, the annotation depicts a range of two dates (see Figure 6.14, (a2)). The photos are chronologically presented from left to right, and the segments are chronologically presented from top to bottom. When a photo is selected, it is possible to see more of its spatio-temporal information, as illustrated in Figure 6.14, (b). The interface was also designed to take several guidelines into account [RLLOBBC-CKMMRSSW04]. Namely, the UI:

1. provides *feedback* on the user's action. For example, selecting a photo shows more spatio-temporal information about it (see Figure 6.14, (b));

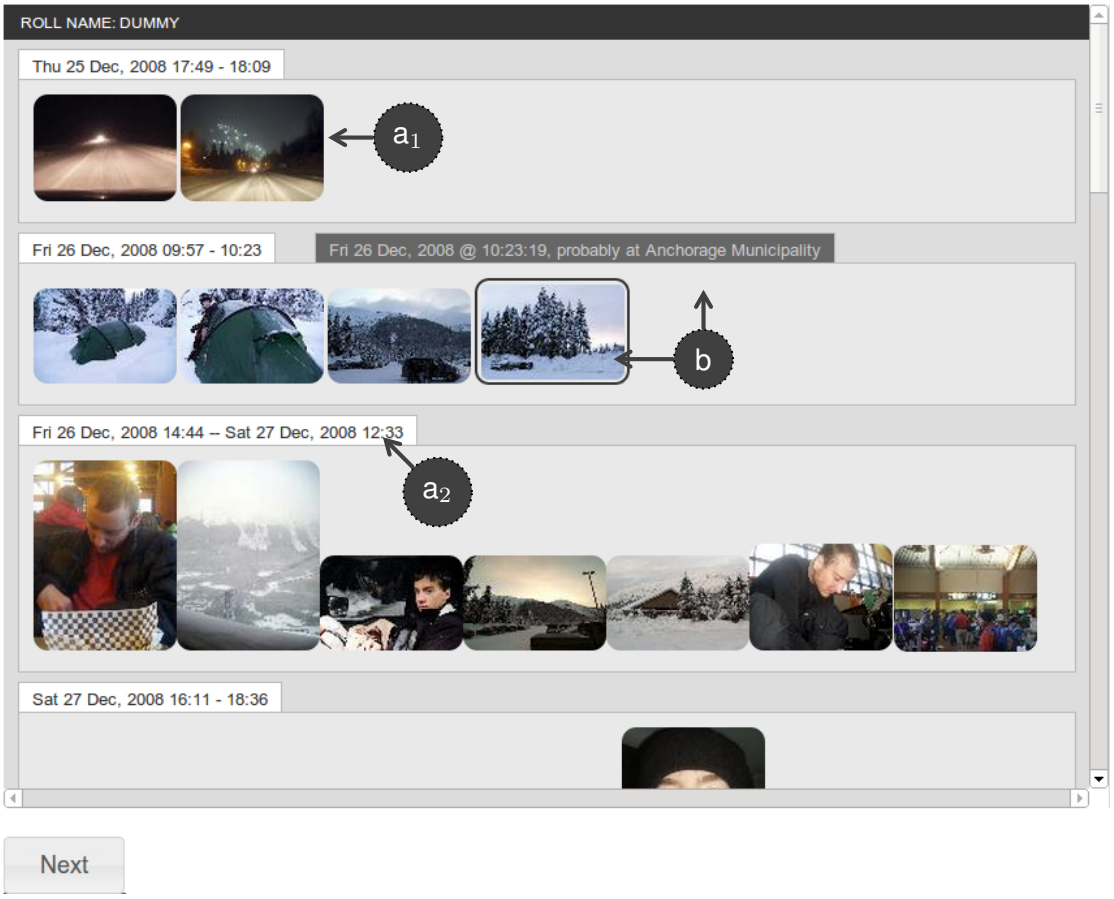


Figure 6.14: Example of the test interface.

2. it is *consistent*, as a change in a segment produces changes in annotations;

3. *prevents users from making errors*. The drag and drop facility is available to modify the segmentation. However, the temporal order of the photos is held, i.e., if the user drags a photo to the next segment, all the following photos are also moved.

A participant can take two types of actions in the proposed segmentation:

1. `Split`

2. `Join`

   (a) `Move-One`

   (b) `Move-Many`

The `split` action is triggered by selecting a photo and hitting the 's' key. The segment with the selected photo is split in two. The photos until the selected photo are kept in the existing segment. The selected photo and the ones that follows in the same segment are moved to a new segment.
The `join` action is done by drag and dropping photos from one segment to another. The

`Move-One` action is a special case of a join, where only one photo is moved. Notice this action consists of moving the first photo from one segment into the previous one, or moving the last photo in a segment into the next one. The `Move-Many` action is another special case of a join. When the selected photo is dragged from one segment and dropped into another (that must be contiguous), other photos are also moved, guaranteeing the temporal order intra- and inter-segment. If the photo is dragged to the predecessor segment, all the preceding photos of the selected one are moved. If the selected photo is moved to the successor segment, then all the photos following the selected one are also moved. If all the photos in one segment are moved into another segment, a complete `Join` is performed and the empty segment is removed from the interface. Otherwise, a partial join is done (`Move-One` or `Move-Many`). All the actions are presented and explained to the participant in the learning phase, that precedes the test itself, where the participant in invited to explored them.

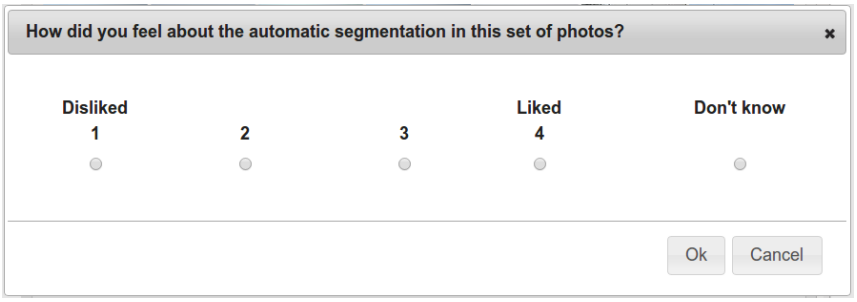


Figure 6.15: Question to assess the perceived quality of the automatic segmentation.

During the test, when a participant finishes a test screen, he must rate the LDES proposed segmentation using a 4+1 Likert item, as depicted in Figure 6.15. The extra option is the *Don't know* choice. Researchers found significant differences when this option is omitted [Lie10], namely, an higher increase on the weak agree/disagree than for the agree/disagree. Nevertheless, we want the participants to take a position, either positive or negative, and since they own the photos, we believed that none will go undecided. Our assumption was confirmed by the results. We use an unipolar scale (1..4) as, according to [OGW95], a bipolar scale would shift the responses towards the positive side of the scale. Thus, the design of the scale is conservative and, if biased, is into the lower part of the scale. The question used is short, in a neutral tone, to improve comprehension and to avoid bias, respectively [Lie10].

6.1.3 Power analysis

We have performed a power analysis for the empirical test, using $\alpha = 0.05$, varying the power from 70% to 90% and the effect size from small (0.2) to large (0.8). As Figure 6.16 shows, for the current number of experimental units, we can detect medium to large

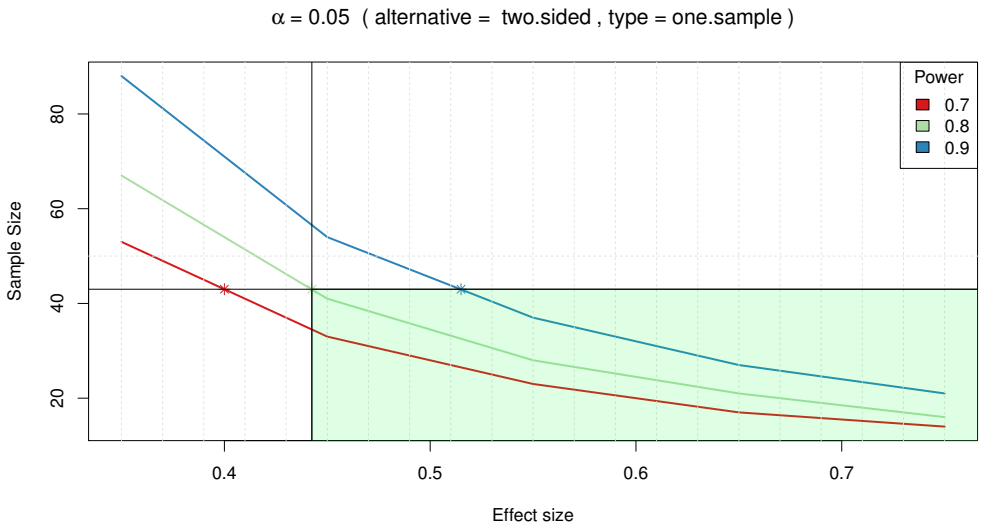


Figure 6.16: Power analysis for the given study.

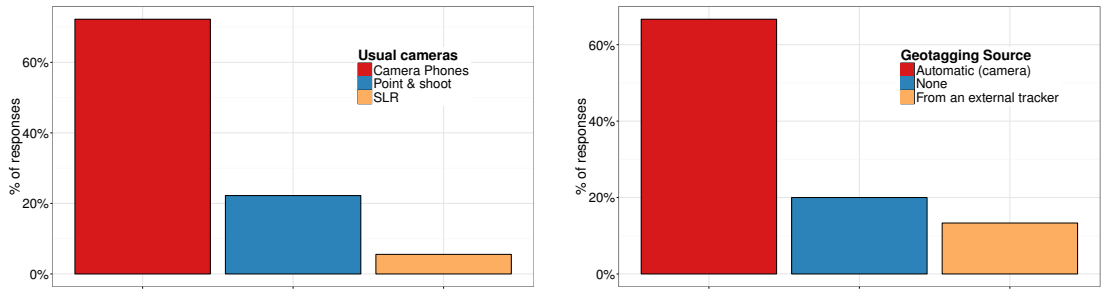
effects sizes[1] with a power greater than 80%, $\beta = 20\%$, [Coh92]. Typically, researchers agree this value is the accepted value for a good power [Coh88; Sel12]. The shaded area on the bottom right, represents the reachable zone, considering the number of experimental units we tested. Depending on the effect size, we can achieve up to 99% power, for a high effect size (equal to 0.8). Since the research statement is that users do accept the segmentation proposed, something that was confirmed by the experimental test, we have an effect size from medium to high. In fact, for high effect sizes, a sample size of 15 experimental units would suffice.

6.1.4 Survey analysis

The analysis of the survey enables us to better describe the group of participants in the study, in terms of their habits of capturing/archiving photos. Figures 6.17(a) and 6.17(b) show the results of the questions "*What type of camera do you use most?*" and "*When you geotag your photos...*" respectively. As we can see, the *cameraphone* is the participant's primary camera, followed by the point and shoot ones. In fact, almost 93% select *cameraphone*, as a single response or in conjunction with another option. Only one participant does not use a camera phone. Thus, it is not surprising the geottaging source is mostly done by the camera itself, as depicted in Figure 6.17(b). From the EXIF data, we confirmed that all camera phones are smartphones equipped with a built-in GPS tracker.

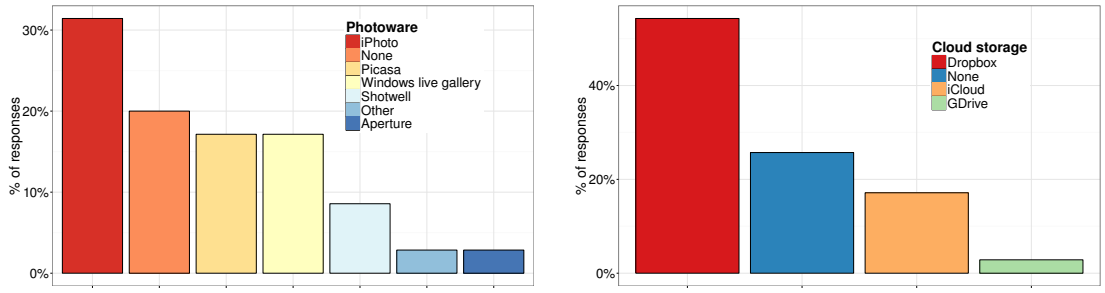
Figures 6.18(a) and 6.18(b) shows the results for the questions "*What program do you use to manage your collection of photos?*" and "*To which online storage service do you save your photos to?*", respectively. The results show that, despite the fact that most participants manage their photos using specifics programs, 20% do not manage their photos using "photoware". A closer look on their responses reveal that 25% do not use any online

[1]The effect size values are for a student's t-distribution.



(a) Usual cameras (b) Geotagging sources

Figure 6.17: Characterisation of the usual cameras used by the participants and the geotagging source.



(a) Software used to manage photos (b) Cloud storage used to backup photos

Figure 6.18: Storage and organisation.

storage service, but 85% use social networking sites, specially Instagram and Facebook. This means they still select the photos to post them online. Figure 6.19 resumes the responses to question "*What online library do you use to publish your photos?*". As we can see, about 15% of the participants do not publish their photos online.

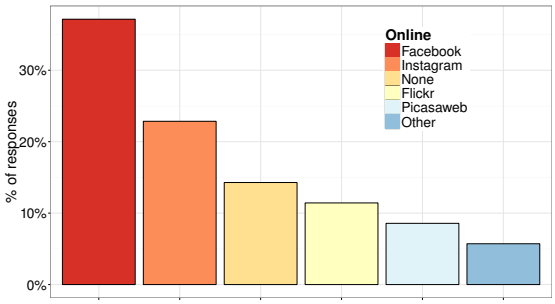


Figure 6.19: Social networking sites used by the participants.

From the above results, we can describe the participants as people who use their smartphones to take photos, they later share online. Most of them also use an online storage service to backup their photos, besides the usage of a desktop program to locally store and manage their photos.

6.1.5 Experiment analysis

As stated earlier, the research statement that we want to assess in this empirical study is that users will accept the temporal coherent segmentation provided by LDES, with none or minor changes. During the empirical study, we collected several data, namely:

1. the participant's modified segmentation;
2. the stream of actions made on the segmentation, by the participant;
3. the perceived quality of the segmentation.

The LDES parametrisation is displayed in Table 6.3. Those were the default values that were set, as described earlier. Do note that in a real situation the users can be able to modify the thresholds, providing more or less refined segmentations depending on their needs.

LDES	
Parameter	Value
f_t	0.5
f_g	0.5
w	4

Table 6.3: Parametrisation of LDES used in the empirical user test.

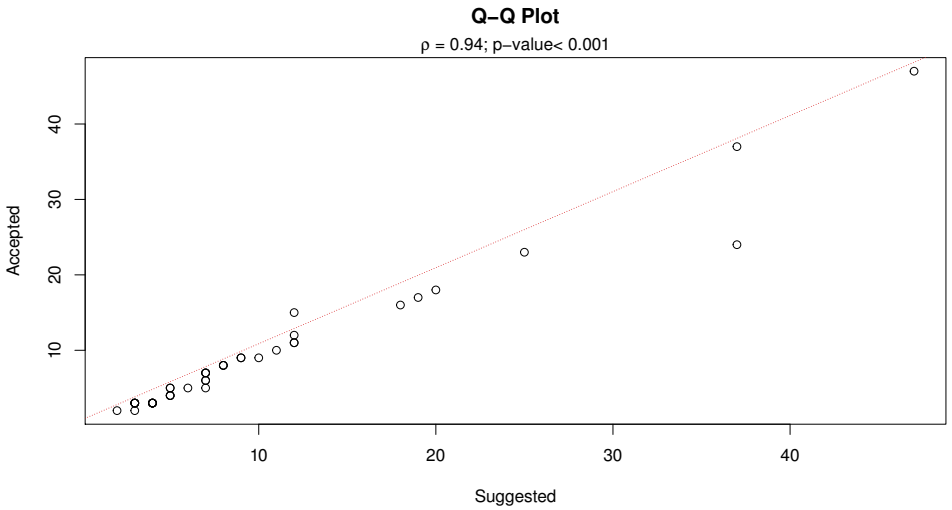


Figure 6.20: Quantile-Quantile plot for the no. of segments for the suggested segmentations and user-modified segmentations.

Acceptance of LDES segmentation From the collected data we are able to derived other indicators, some of which will be used next. One of the indicators is the *number of segments* in the segmentation. The distribution for the number of segments, in the suggested segmentations and in the modified ones, is very similar. This can be observed in the Quantile-Quantile (Q-Q) plot, showed in Figure 6.20. The values are positively correlated, with $\rho = 0.94$, with a p-value < 0.001. Besides the number of segments, we analysed the changes made by the participants to the segmentations using 4 actions:

1. `Split`: one segment is divided in two. A new segment is created;
2. `Join`: two segments are merged into a single segment. One segment is removed;
3. `Move-One`: One photo is moved from one segment to another. No segment is created. One segment may be removed;
4. `Move-Many`: Many photos are moved from one segment to another. No segment is created. One segment may be removed.

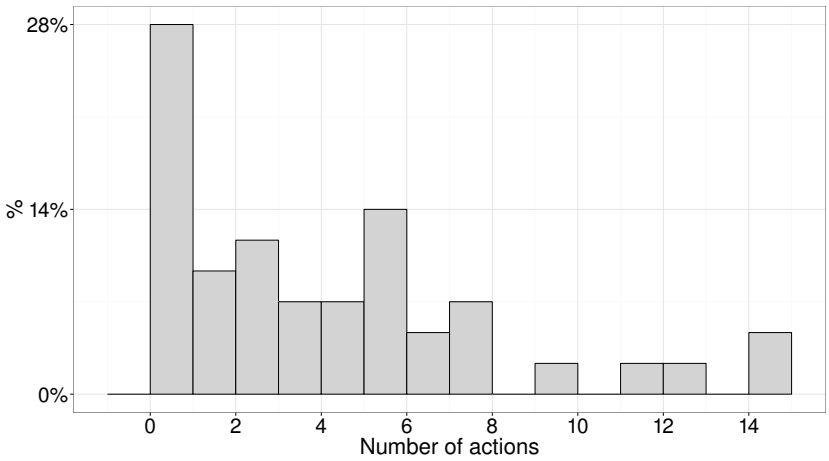


Figure 6.21: Distribution of the number of user actions made in the experimental units.

Figure 6.21 shows the distribution of the number of actions made by the users to the proposed segmentations. One result that stands out is that the most common behaviour is not to change the segmentation, representing about 30% of the cases. However, in a few segmentations we can see an higher number of actions. Looking at the data, those cases represent a misalignment between the level of detail (*LoD*) the users want in certain parts of their rolls, and the *LoD* given by LDES. One of the participants that make more actions said about it

> *Participant 8*: "If the roll contains photos from a longer vacation (more than a week), I do not want to see it divided lower than the day. However, for a weekend holiday it seems OK."

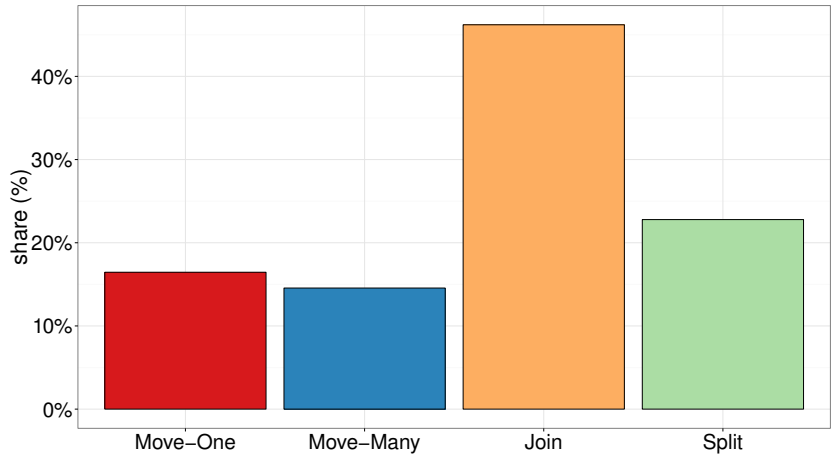


Figure 6.22: Share of action types made by the users to the proposed sgmentations.

This result is confirmed by the share each type of action has in the changes made by the user, as depicted in Figure 6.22. The `join` action was the most common, used in 46% of the changes, followed by the `Split` action (23%). Those two are responsible for the higher number of actions made to some of the segmentations. Nevertheless, the over-segmentation (corrected by the `Join`) and the under-segmentation (corrected by the `Split`) are made by the users in small portions of their rolls, generally confined to one logical day. An important finding is the `Move-One` and `Move-Many` are used occasionally. This shows that, most of the time, important cut points are well identified by LDES. Together, Figures 6.20, 6.21 and 6.22 show a strong evidence that segmentations are accepted by the participants. The number of segments proposed by the algorithm are highly correlated with the ones that are accepted by the users. The actions made by the participants are small, with median < 2.

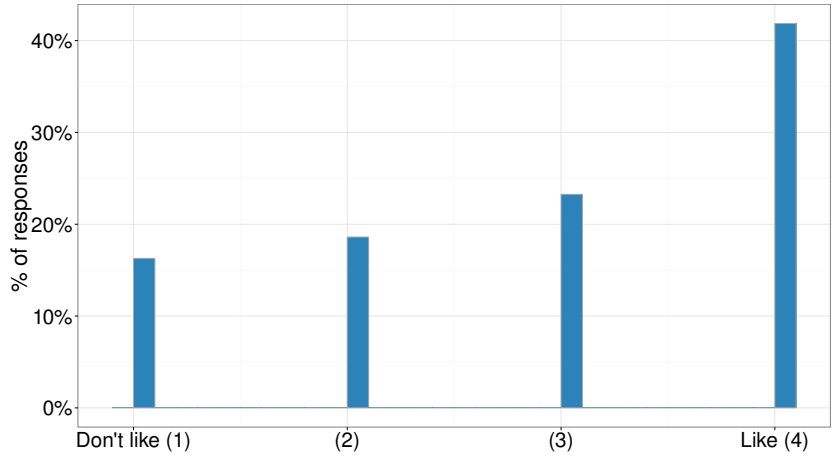


Figure 6.23: Responses to the quality of the proposed segmentation.

From the data in Figure 6.23, it is apparent that participants like the segmentations produced by LDES, since more than $\frac{2}{3}$ of the responses are positive (one sample t-test,

$p < 0.001$). Only in 17% of the responses, the participants select *don't like*. The responses for the quality of the segmentation are more diverse, towards the lower part of the scale, when the spatial information is absent. The perceived quality is lower when the geographic information is absent (one sample t-test, $p < 0.001$).

Logical day LDES introduces the notion of *logical day*, as an important mechanism to match people's behaviour to the notion of the day cycle. We analysed the rolls to see whether the logical day are different from the standard day. It was found that on 7% of the cases they differ. The analysis of the segmentations modified by the participants shows that 100% of the logical days were kept. These results suggest the concept is important to maintain a temporal coherence in the segments, going beyond the strict boundaries of temporal cycles. Further analysis showed that some participants join, sparsely, photos from two days, producing *multi-day* segments. The statistical tests revealed that multi-day segments appear specially in segmentations having higher cardinality[2] (one sample t-test, p-value < 0.01). However, this was done sparsely, without an apparent criteria. This may indicate the size of the segmentation (which may or not be directly related with the temporal and spacial range) may influence the perception users have of the context.

Singular segments Other important finding is that singular segments are kept by the participants. Figure 6.24 depicts the Q-Q plot for the number of singular segments in each experimental unit, considering the suggested segmentation and the ones that exists in the final segmentation, after the participants made their changes. As we can see, they have a similar distribution, and they are positively correlated, with $\rho = 0.97$, with a p-value < 0.001. These are interesting results, as it seems that LDES is capable to isolate photos that have their own context. This situation is becoming more frequent in personal photo collections, as there are more photos taken from smartphones. Further statistical tests revealed that singular segments in the participants segmentations are more frequent in the segmentations with higher cardinality(one sample t-test, p-value < 0.01).

Relations and measures We have compared the segmentations before and after the participants intervention, using qualitative and quantitative analysis. In the first case, we used the binary relations introduced in Section 4.2. In the second case, we used the PR_{error} [GCA06] and the WindowDiff [PH02] measures to compare the segmentations. The PR_{error} was set in three different scenarios:

1. *equal costs* for miss and false positive (FP);
2. *FP costs are three times higher* than miss costs;
3. *miss costs are three times higher* than FP costs.

[2]No. of segments greater than the median

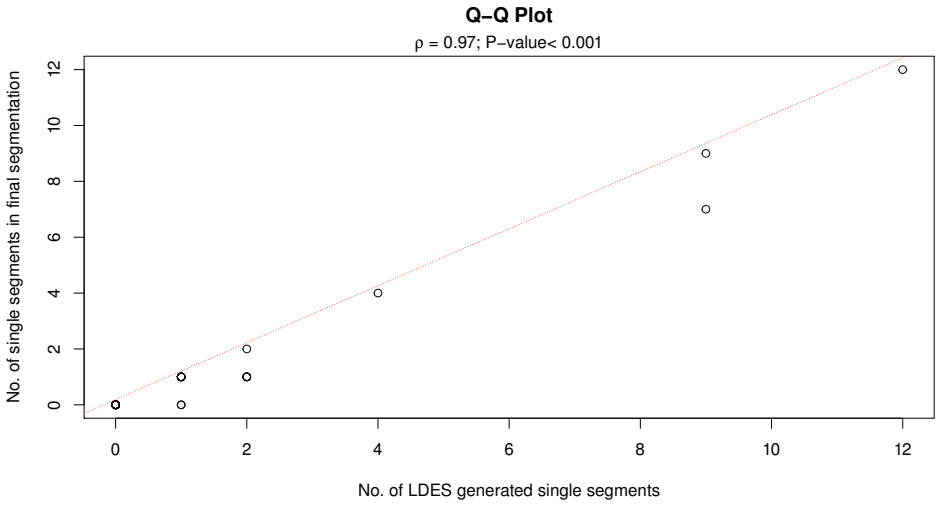
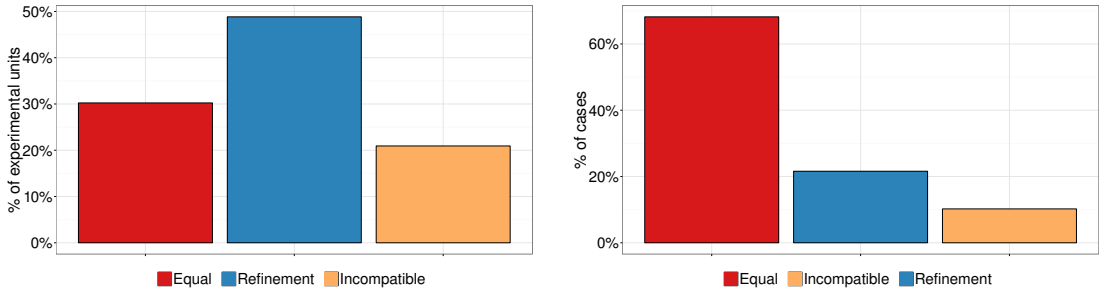


Figure 6.24: Quantile-Quantile plot for the no. of single segments in the LDES segmentations and participants' segmentations.

Figures 6.25(a) and 6.25(b) present the results of the qualitative comparison between LDES segmentations and the ones the participants submitted during the test. In case (a), the segmentations are compared as a whole. However, to better understand the behaviour of participants, we decided to analyse separately each logical day inside the segmentations (b). In this case we do not consider the situations of multi-day segments. It can be seen from Figure 6.25(a) that almost 80% of the segmentations are compatible,



(a) Qualitative comparison between LDES segmentations and modified segmentations.

(b) Qualitative comparison between LDES segmentations and modified segmentations, considering the division in logical days.

Figure 6.25: Qualitative comparison between segmentations.

30% of which are `equal`. This means that most of the time, the important cut points are well identified by LDES. The `refinement` relation, which represents almost 50% of the cases, tells us the participants need to insert or remove cut points. In either way, such behaviour represents a difference in the level of detail the participants want to see, and the level of detail LDES provides. However, as shown in Figure 6.25(b), these changes in cut points happen in few segments, since almost 70% of the logical days were perfectly segmented by LDES. These results are in line with the previous evidences found when

analysing the number and type of actions made to the segmentations by the users.

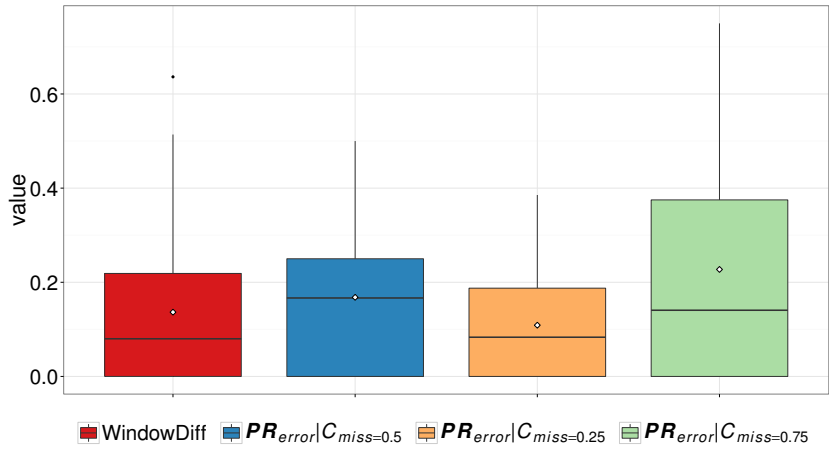


Figure 6.26: Distance measures between LDES segmentations and participant's modified segmentations.

The quantitative analysis confirms the results shown above. From the data in Figure 6.26, it is apparent that the segmentations are similar. The medians, in all scenarios, are above 0.2. The upper quartiles for the two measures, considering PR_{error} with equal costs, are under 0.3. The sizes of the first two boxplots are similar, indicating that there is not such great difference between measures.

Another interesting observation is there are more misses than false positives, given the values for different parametrisations of PR_{error}. This is in line with the characteristics of the measures. WindowDiff penalises all pure false positive the same way, regardless of how close they are to an actual boundary [PH02] and misses are less penalised than false positives [GCA06]. Those characteristics explain the lower values in WindowDiff and the behaviour of PR_{error} when the cost of a miss is changed. This also reinforces the results depicted in Figure 6.25(a), where almost 50% of the segmentations have a `Refinement` relation between them, as a result of the `Join` being the action most used by the participants (see Figure 6.22). Thus, when participants changed the segmentations, they mainly lowered the level of detail for some days in their rolls. This was done removing cut points, but maintaining the ones that are key to separate the context.

6.2 MSS algorithm evaluation

The MSS algorithm was evaluated against several personal collections of photos to determine:

(i) the impact of the parameters in the clustering solution;

(ii) the quantitative difference between our algorithm and other clustering procedures;

(iii) the quality of MSS as a summarisation procedure.

The first two items were determined in experimental tests, whose descriptions and results are reported in Section 6.2.1. The last item was assessed using experimental user tests, whose description and results are reported in Section 6.2.2.

6.2.1 Evaluating the cluster step

Stats.	No. Days	No. Photos	No. Cities
Min.	1	10	1
1st Qu.	4	72.5	3
Median	7	157	5
Mean	9.2	188	7.8
3rd Qu.	12	248	8.5
Max.	41	607	59

Table 6.4: Descriptive statistics for the dataset used in the MSS experiments.

We used 39 photo sets collected from personal collections of holiday photos, most of them available at Picasa Web Albums. They differ from the ones described in Table B.1 for three reasons. First, because MSS needs all the photos with location information. Second, because we want to reduce the day range, to accommodate more specific contexts. Finally, because we want to increase the number of photo sets with 1 day. The last two reasons are for summarisation purposes, related to the motivation for the existence of the MSS. Table B.2 (in Appendix B) details the photo sets used, and Table 6.4 shows a summary. As we can see, they average 157 photos, ranging from 10 to 682. The temporal

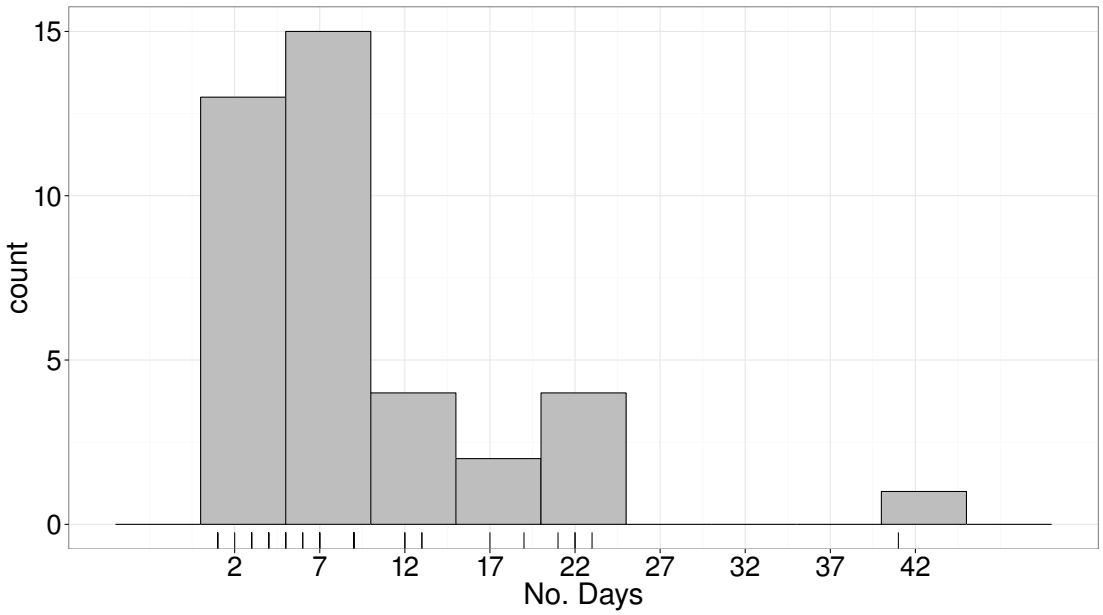


Figure 6.27: Characterisation of the distribution for the no. of days in the photo sets for MSS experimental test.

context varies from 1 to 41 days (Figure 6.27), while the spatial context ranges from 1 to 65 different cities (Figure 6.28).

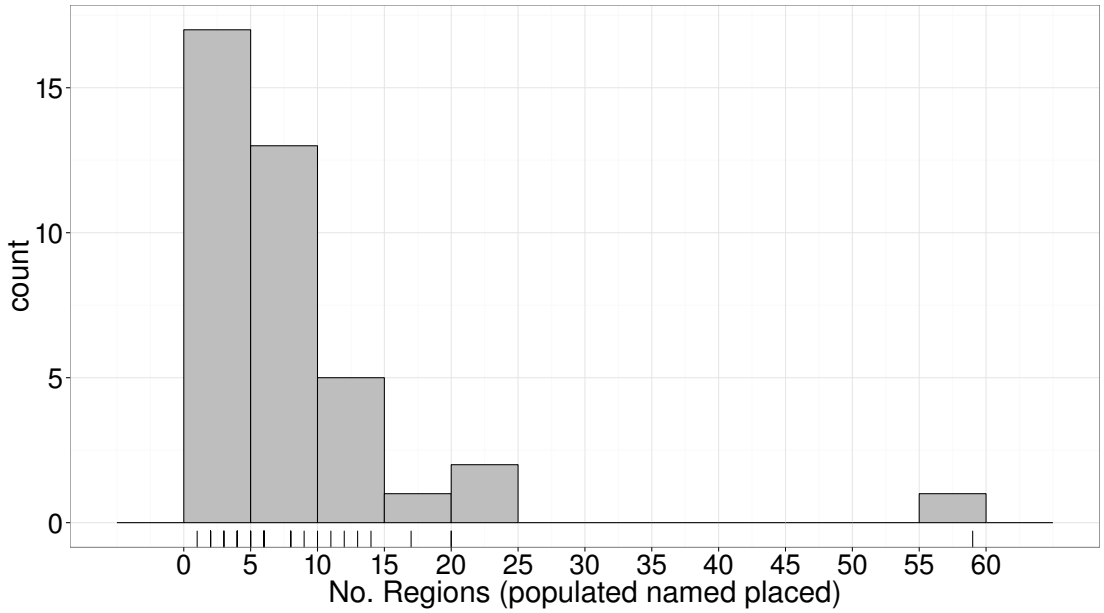


Figure 6.28: Characterisation of the distribution for the no. of cities (populated places) in the photo sets for MSS experimental test.

6.2.1.1 Sensitivity to κ and dominance

The MSS summarisation algorithm has two parameters that govern its behaviour. The first is κ that specifies the maximum number of groups that MSS can produce. The second is dominance that indicates which information has a greater influence on the clusters — spatial or temporal. We study the sensitivity of MSS to changes in the dominant dimension and in the value of κ between $[2..16]$. For each of the 39 datasets, we make 14 runs with a different value of κ, repeating them for each dominance. Figure 6.29 shows the aggregated results for the 39 datasets. The y-axis indicates the number of groups produced for each MSS run, and the x-axis shows the value of κ. The results for the two dominances are depicted side by side, to ease the comparison between the two. The number of clusters increases in steps, as κ increases. This is an expected behaviour, as κ limits the number of groups. The number of clusters increases when a specific attribute has a cardinality that exceeds κ. Until then, it stays at the same attribute and thus, it produces the same number of clusters. The median of the number of segments found seems to stabilise for $\kappa \geq 12$, independently of the dominance used. We found that MSS with spatial dominance produces less segments. For higher values of κ, the division is less affected by changes in the parameters, as the distributions of the result are more similar.

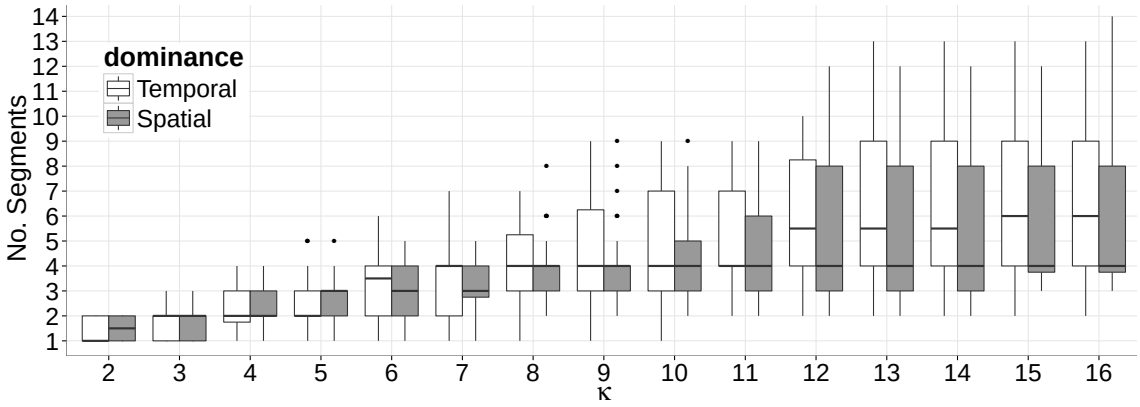


Figure 6.29: Number of segments found varying κ.

6.2.1.2 Comparison with other clustering algorithms

The MSS was compared with other two clustering approaches:

1. one that uses attribute induction, identified by AOI [HF96], and
2. hierarchical clustering algorithm, named AGNES [KR+90].

We chose them because MSS is both hierarchical and attribute induction based. For comparison purposes, we settle $\kappa = 16$, the maximum value admitted in the work.

AGNES produces 16 clusters for every set[3]. This results in similar descriptions for different clusters, which reduces their value as a summary. AOI creates clusters that change with the spatio-temporal context of P. On average, it produces 10 clusters for single day photo sets and 6 cluster for multiples day data sets. For single day photo sets, the clusters are over-detailed for a summary, as it often uses the most specific attribute to discriminate the set. It is important to notice that AOI and AGNES do not guarantee temporal order in the partitions. This causes discontinuities in time between clusters, forcing more verbose descriptions.
We compare the inter- and intra-cluster distances for the clustering produced by each algorithm. The dissimilarity measure used was based in the *Gower's General Similarity Coefficient* [KR+90]. For the inter-cluster distance, the dissimilarity is calculated using the medoid of the cluster. Table 6.5 shows the average dissimilarity for the partitions produced for each algorithm. The inter-cluster dissimilarity is very similar in both algorithms, with a slight advantage for MSS. The higher values for intra-cluster dissimilarity in MSS is a consequence of design, as we choose the partition cuts between less specific attributes. We prefer to have better separated groups than homogeneous ones, as they need to be described differently. The results shows the MSS found segmentations with similar performance as the clusters found by the other algorithms, but better for producing distinct descriptions for each group.

[3]The only exception was one set which has only 10 photos.

	Intra-cluster	Inter-cluster
MSS	0.1379	0.576
AOI	0.1044	0.572
AGNES	0.0855	0.522

Table 6.5: Intra- and inter-cluster average dissimilarity.

6.2.2 User test

The MSS was tested by a set of volunteers. We wanted to know how long they take to answer a questionnaire about the spatio-temporal context of a set and how good they perform on the answers. We recruited 20 volunteers, mainly computer science students, both graduate and non-graduate. Their age ranges from 21 to 38, with 40% of women. We choose 12 comparable photo sets based on their context, as described in Tables 6.6 and 6.7.

	No. Photos	No. Days	No. Regions	No. Place	No. Country
DS-U01	40	4	2	5	1
DS-U02	126	6	9	70	1
DS-U03	17	2	2	9	2
DS-U04	280	4	7	58	4
DS-U05	179	13	5	32	4
DS-U06	356	22	15	79	1
DS-U07	183	23	31	57	4
DS-U08	230	41	23	57	4
DS-U09	513	21	65	212	10
DS-U10	566	7	31	165	9
DS-U11	44	3	6	15	2
DS-U12	239	9	14	47	5

Table 6.6: Characterisation of the photo set used in MSS users test.

Stats.	No. Days	No. Photos	No. Countries	No. Regions	No. Places
Min.	2	17	1	2	5
1st Qu.	4	105.5	1.8	5.8	27.8
Median	8	206.5	4	11.5	57
Mean	12.9	231.1	3.9	17.5	67.2
3rd Qu.	21.2	299	4.2	25	72.2
Max.	41	566	10	65	212

Table 6.7: Descriptive statistics for the data in Table 6.6.

The test used the following summarisation strategies:

1. ***G-Day*** — group by day;

2. ***G-City*** — group by city;

3. ***MSS-T*** — MSS with *temporal* dominance;

4. ***MSS-S*** — MSS with *spatial* dominance.

For *G-Day*, the spatial description is taken from the first photo of the group. The temporal description in *G-City* is either a day or a range of days, if there is more than one day in the group. *G-Day* and *G-City* are relevant to establish a baseline because they use important concepts to people. Currently, people make most of their activities in cities. According to [Pil02], there is a strong relation between memory and city, where the last produces a dense network of encounters, that helps to map memories to space and time. The notion of day, despite being a natural separator of activities, is also the common denominator "*social cycle*" for events [Zer85]. Also, users are aware of similar grouping procedures, since they are widely used in commercial programs like Apple's iPhoto.

6.2.2.1 Test Design

Each test had 12 steps. In each one, the volunteers saw a photo[4] set summarised by one of the four strategies described above, guaranteeing a different one in each group of 4 screens, as illustrated in Table 6.8. The photo sets in each wave[5] share similarities, namely:

(i) 4 sets have less than 10 days and 10 cities — [DS-U01..DS-U04];

(ii) 4 sets have multiple months and more than 10 days — [DS-U05..DS-U08];

(iii) 4 sets have more than 6 cities in multiple continents — [DS-U09..DS-U12].

There is no repetition of sets with different summarisation strategies. A strategy was never repeated twice in a row. The volunteers were divided in two groups. The first group watched the sets in a sequence, that was inverted for the second group. This would help us check if the order changes the results. The descriptions of the sequences are showed in Table 6.8.

The interface used was similar to the one depicted in Figure 6.30. On the left hand-side a collection is summarised. On top, the summarisation strategy is presented (ⓐ). In the middle there is the selected detail for temporal information (ⓑ), and finally, in the bottom, there is the location's selected detail (ⓒ). With *G-Day* and *G-City* each group was labelled with the temporal (mm/dd/yyyy) range and the city (the most common), and the selected detail (ⓑ and ⓒ) is not presented. For *MSS-T* and *MSS-S* the descriptions are generated using MSS. All the strategies have the no. of photos for each group. The right-hand side, labelled as ⓓ, has the questionnaire, with 4 open questions about time and 5 about space:

[4]The first of the cluster

[5]A sequence of four summarisations strategies.

	Group 1		Group 1	
	summarisation	photo set	summarisation	photo set
screen 1	*MSS-T*	1	*G-Day*	12
screen 2	*G-Day*	2	*MSS-T*	11
screen 3	*MSS-S*	3	*G-City*	10
screen 4	*G-City*	4	*MSS-S*	9
screen 5	*MSS-S*	5	*G-City*	8
screen 6	*G-Day*	6	*MSS-T*	7
screen 7	*MSS-T*	7	*G-Day*	6
screen 8	*G-City*	8	*MSS-S*	5
screen 9	*MSS-S*	9	*G-City*	4
screen 10	*G-City*	10	*MSS-S*	3
screen 11	*MSS-T*	11	*G-Day*	2
screen 12	*G-Day*	12	*MSS-T*	1

Table 6.8: Sequence of photo sets and the summarisation used during the MSS users' test.

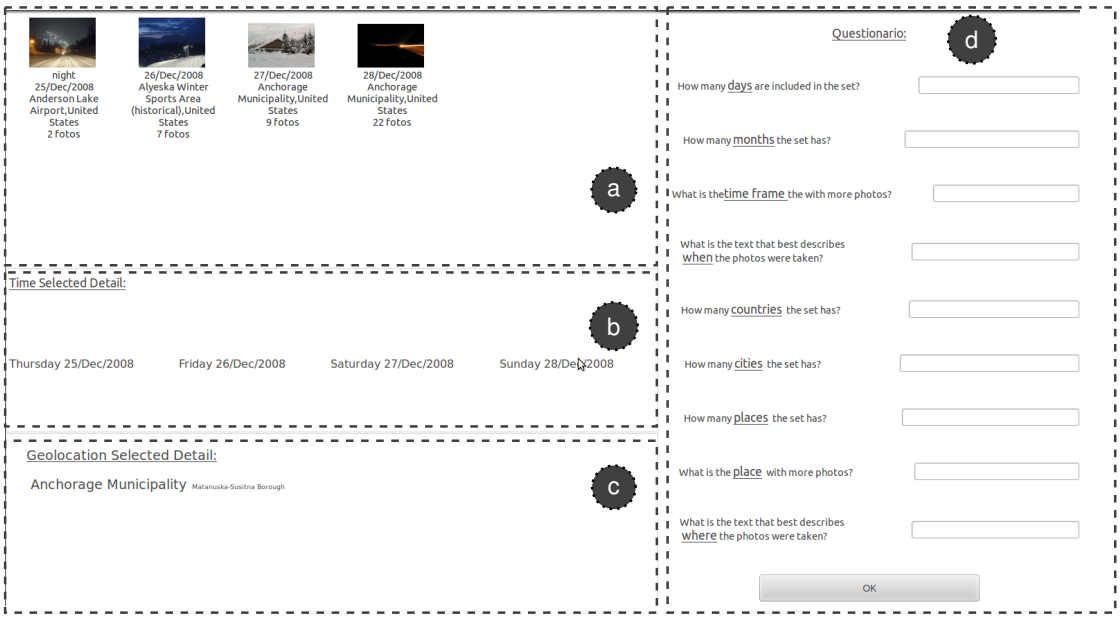


Figure 6.30: Example of the MSS user' test interface, for a *MSS-T* screen.

Time context

$Q1$ — "how many days are included in the set?"

$Q2$ — "how many months the set has?"

$Q3$ — "what is the time frame with more photos?"

$Q4$ — "what is the text that best describes "*when*" the photos were taken?"

Spatial context

$Q5$ — "how many countries the set has?"

$Q6$ — "how many cities the set has?"

$Q7$ — "how many places the set has?"

$Q8$ — "what is the place with more photos?"

$Q9$ — "what is the text that best describes "*where*" the photos were taken?"

With this questions we can detect if the context is properly captured by the users. In the beginning of the test, each volunteer saw a small video explaining what we expected them to do.

6.2.2.2 Response time analysis

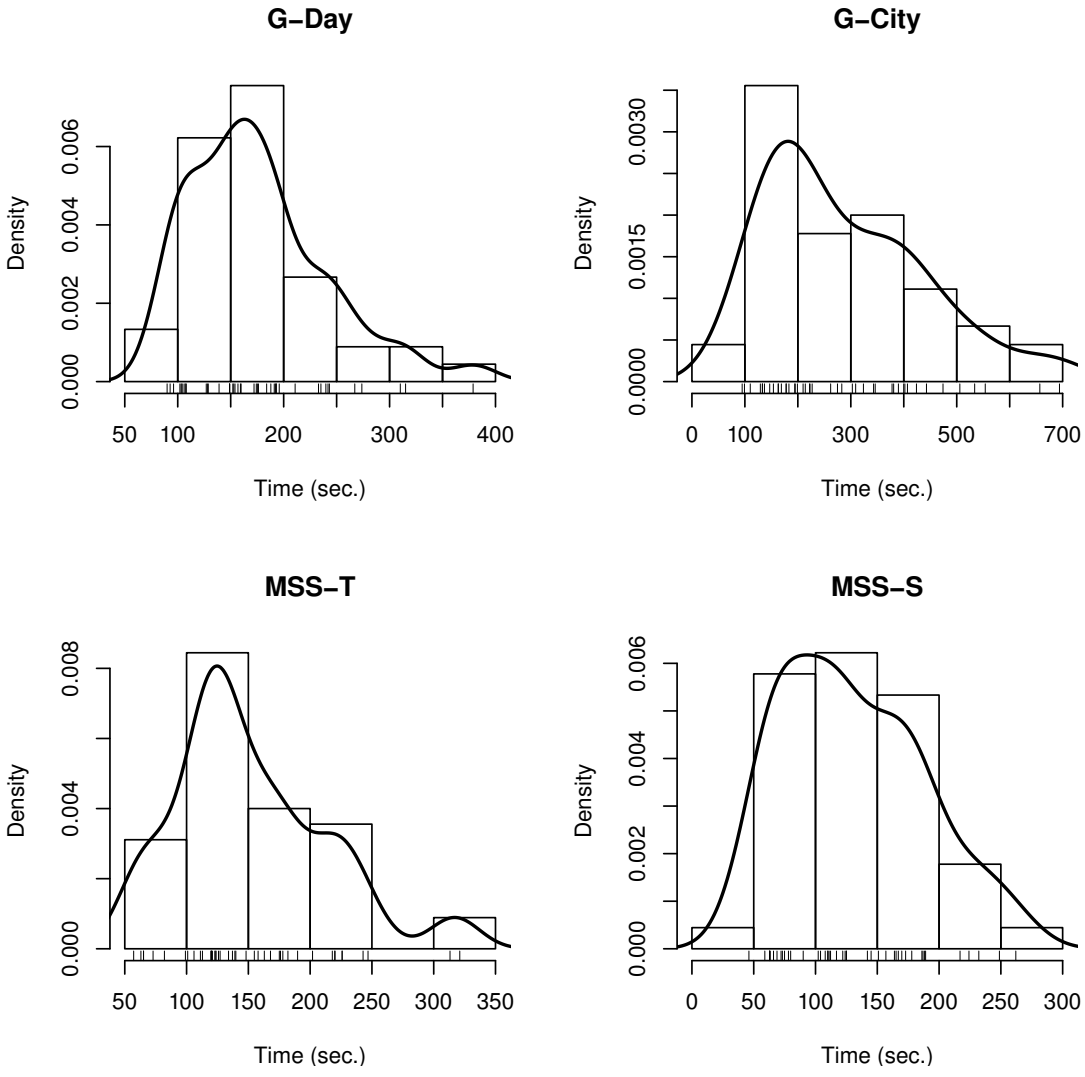


Figure 6.31: Distribution of response times for each screen, grouped by type of summary.

The division of the volunteers in two groups help us to detect any ordering or recency effects. The results show the order of screen presentation does not change the average response time in each screen (p-value < 0.01). Response times have a large variability,

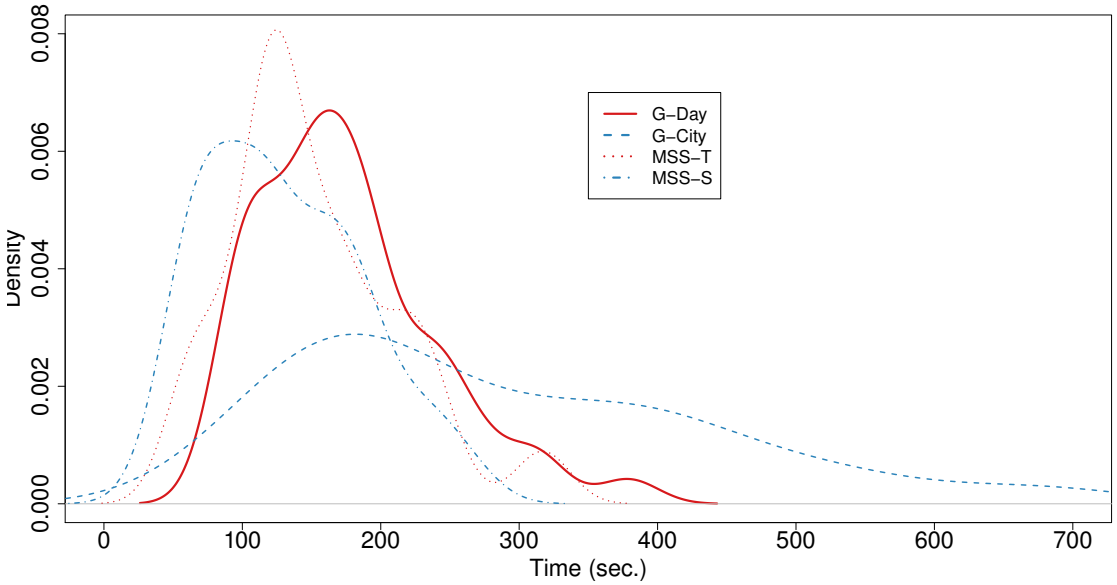


Figure 6.32: Kernel density estimation of response times, grouped by type of summary.

depending on the user and the type of summary. The values range from 46 seconds to 705 seconds, almost 12 minutes. The distribution of response time is shown in Figure 6.31, grouped by type of summary. To ease the comparison between them, we estimate the density distribution for each type of summary, shown in Figure 6.32, using a density estimator with a Gaussian kernel. We can see that MSS, in both dominances, presents lower response times than the other strategies, with peak density around 100 seconds. The baseline strategies response peaks have an offset of 50 seconds. The difference is smaller when comparing *G-Day* and *MSS-T*, and larger between *G-City* and *MSS-S*. It is possible to see that *G-City* exhibits the largest variance in the group, while *MSS-T* has the smallest variance.

Figure 6.33 shows the response time variation for each screen, grouped by their type.

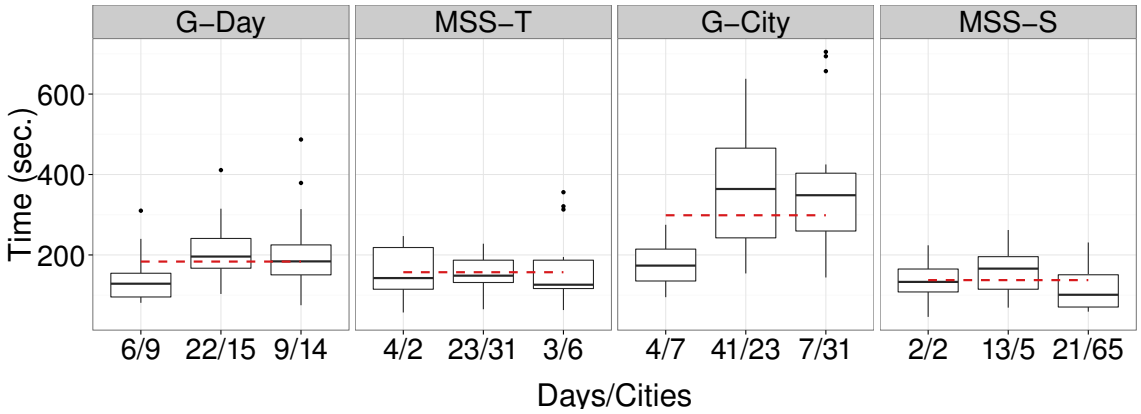


Figure 6.33: Response times for each step, grouped by type of summary.

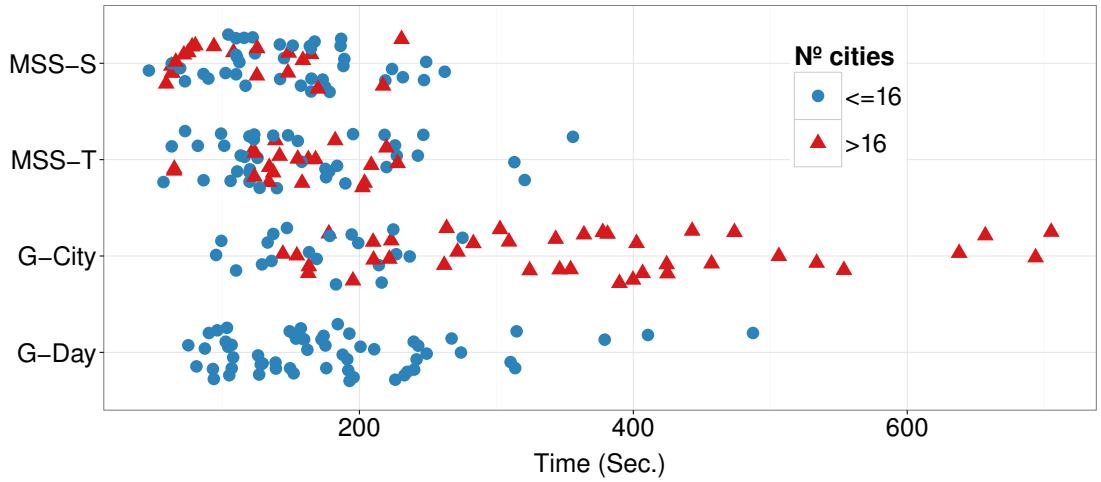


Figure 6.34: Response time comparison for each type, separating the screens with more than 16 cities from the others.

The y-axis shows the response time in seconds. The x-axis tick labels are pairs. The dashed lines represent the mean response time in the summarisation strategy. Each one has the number of days and the number of cities in the dataset summarised in the screen. The results show the volunteers answer quicker when the summary is MSS. On average, they spent 54% more time answering at type *G-City* than at type *MSS-S*, and 14% more time for type *G-Day* comparing with type *MSS-T*. It is also noticeable that two tests stand out with higher response times, both of type *G-City*. They share a higher number of cities, which causes an increase in the response times. Since the number of days in those sets is 41 and 7, we exclude the higher number of days as the cause of this increase. The screen with the highest number of cities, 65, uses type *MSS-S*. As we can see, the response time in that case is no different when compared with others with less cities.
The user tests confirm that limiting the number of groups presented to the user has advantages. Figure 6.34 shows the response time for each summary, but now separating the sets with the number of cities above 16 from the others. We can see the response time with MSS is similar in both dominances. However, type *G-City* shows more often higher response times with higher number of cities. The statistical significance test supports our findings (p-value < 0.05).

6.2.2.3 Response accuracy analysis

The users were able to answer the questionnaire using MSS and, most of the time, the accuracy is the best among the four types. Notice the number of clusters with MSS is, on average, $\frac{1}{3}$ and $\frac{1}{5}$ less than using *G-Day* and *G-City* respectively. *MSS-S* is always better than *G-City*. *MSS-T* is better than *G-Day* except in two cases. One is question $Q1$ (*how many days the set has*). We foresee this result, because the information available to users in screens using MSS was, most of the times, at a higher conceptual level (e.g. the Month). The other exception is question $Q7$ (*how many places the set has*). What should be reported

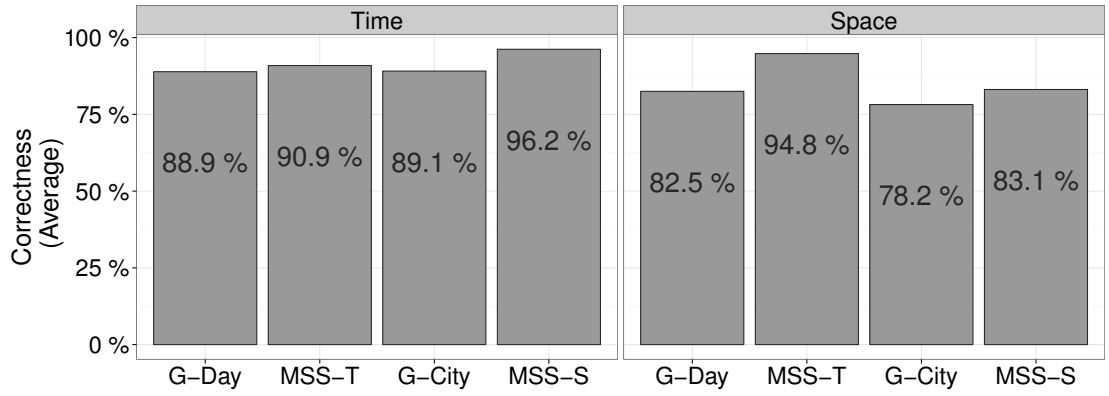


Figure 6.35: Proper context description.

as a "place" was left for the user to decide. In average, 50% of the volunteers choose the *City* as the conceptual level for "place", except for *MSS-T* screens, where only 25% chose the same level. Figure 6.35 shows the adequacy of descriptions for the temporal and spatial context. As we can see, the descriptions are more often correct when the set of photos is displayed using MSS. We assume as acceptable, a term or phrase that covers most of the photos, and is a proper label for the holiday. For example, *"United States"* is a proper label for the spatial context of a holiday in the USA, even though there is one day spent in Canada, during a visit to the Niagara Falls.

The user testing reveals some differences between male and female volunteers. This was an unexpected finding. One difference was found in the response time. It is statistical relevant the evidence that females take more time answering the questions than males, when observing MSS summaries (two sample t-test, p-value < 0.005). The variation, in average, is 65 seconds. The other difference is related with responses containing generalised concepts. A concept is generalised, when it is broader than the ones available to the user. For example, if 3 groups are described as "2013 @ Lisbon", "2013 @ Sintra" and "2013 @ Oporto", the description "2013 @ Portugal" contains a generalisation of the location, Portugal. Questions 3 and 4 (those about temporal information) and questions 8 and 9 (those about spacial information) are the ones where users can enter general terms, as they need to describe the spatio-temporal context of the set being summarised. Table 6.9 shows the percentage of responses that include a generalisation. We

Question	%
Q3	10.2
Q4	61.3
Q8	24.3
Q9	74.9

Table 6.9: Percentage of responses that include generalisations.

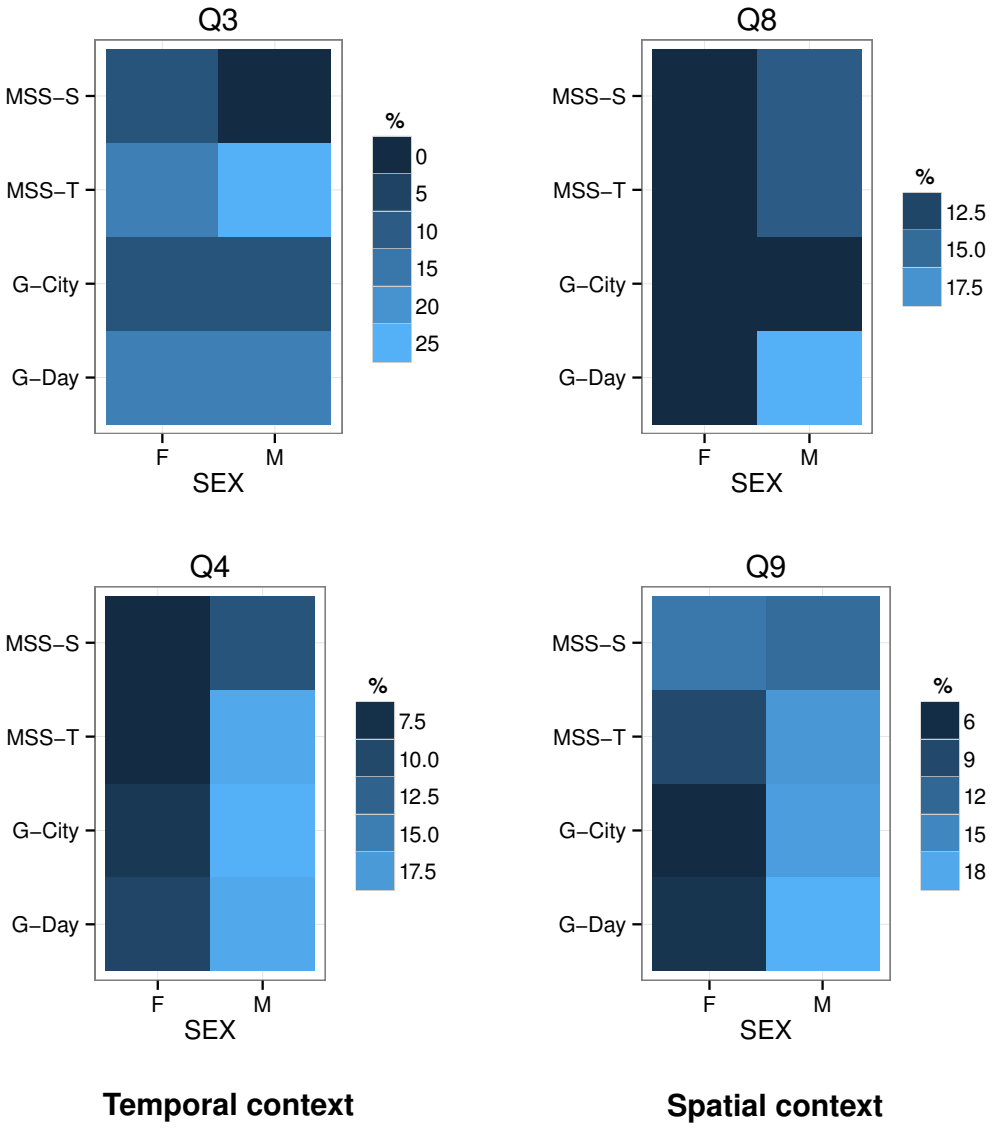


Figure 6.36: Concept generalisation in survey. Each graphic represents the share of generalisation, for a combination of *Sex* and *Type of summary*.

can draw two conclusions. The first is users generalise more when they write about the context (two sample t-test, p-value < 0.001). The second, is the generalisation is more often used when describing the spatial location of the sets. Figure 6.36 shows the share of generalisations, in percentage, in each of the 4 questions. The values are discriminated by *sex* and *type of summary*. Lighter colours represent a higher percentage value. The results show that males generalised more than females in their responses (two sample t-test, p-value < 0.01). This is most evident when describing the set that is summarised (questions 4 and 9).

6.2.3 Discussion

The analysis of MSS response to changes in the number of groups reveals a stabilisation when $\kappa \geq 12$, never reaching the 16 limit for any set. These results provide further support for the hypothesis that MSS is suitable to be used in devices with smaller screens, where space is limited. Nevertheless, it is important to bear in mind that MSS was not designed to support browsing, but to aid the retrieval of a specific set of photos. Besides the reduced number of objects, MSS also includes a selected level of detail (*LoD*). From the analysis of the responses to questions 3, 4, 8 and 9, it seems it was important to settle the temporal and spatial description of the context. However, the test was not designed to assess each component of MSS, and thus, we cannot measure the importance of each component alone. The evaluation shows the MSS has achieved its goals, namely: *(i)* it is capable to set a summary of a set of photos, using a concise number of groups, and, most important, *(ii)* it does it in a way that users are able to understand the spatio-temporal context of the photo set. We show that limiting the groups has advantages, at least for the spatial part of the context. The threshold of 16 groups proved its adequacy in the user test, for producing faster response times. Despite the maximum number of groups were fixed, independently of the size of the set being summarised, the experiments indicate the inter-cluster context separation is suitable to address MSS' goals. The number of objects displayed to the user are, on average, reduced 48 times. Despite such reduction, MSS maintains enough information to give the users cues they need to identify correctly the spatio-temporal context of a set. With less objects to look at, the context is reconstructed in less time, comparing to other summarisation algorithms. We support our conclusions, not only in the evidences gathered from the results, but also on how the test was designed. Using datasets that not belong to the participants in the test, enables us to state the description of the context was solely made using the available information in the summaries.

The user tests reveal statistical relevant genre influence in the responses. The first is that male participants often use generalised terms to describe the context. The second is that females take more time answering the questionnaire, when MSS is the summarisation used. Since these *genre* were not taken into account when designing the test, we lack the information to determine if this happens because of MSS or because of other external factors.

6.3 Conclusion

This chapter describes the experiments done to assess the quality of the two core algorithms proposed in this work: 1. the *Logical Day Event Segmentation* algorithm, used as a pre-processing step for archival and 2. the *Multimedia Short Summary*, used to summarise the context of a set of photos during retrieval. The process of experimentation is equal for both algorithms, and is divided in two phases. In the first phase the algorithms are tested

with different parametrisations, so they can be characterise under several conditions. The second phase of experiments is the user testing. The user testing was carefully crafted, using the best practices gathered from two areas of knowledge, namely, psychology and computer science. From the first, we used their experience in user testing to design the questions and to set the number of participants. This last issue is very important, to assure the correct α and β needed for hypothesis testing. From the second, we take their expertise on developing test interfaces that are not intrusive to users. The careful design of the user tests makes possible to draw important conclusions about the adequacy of the algorithms proposed fulfilling their purpose.

6.3.1 LDES

In the LDES experiments, the aim was to assess if users accepted an automatic segmentation of their own rolls, that is temporal dominant, and incorporates spatial information. The segmentation algorithm, LDES, incorporates the notion of logical day, that follows the day cycle, adapting it to the perception people have about where a day ends. Such features are the result of the importance of temporal and spatial information to settle the context of personal photo collections, taking into account the social semantics knowledge.

During the test, if the users are not satisfied with the segmentation of their rolls, they can change it, dragging photos from one segment to another. The most obvious finding to emerge from this test is that the segmentations pre- and post-user action are almost equal. This is supported not only by the similar number of segments in each, but also in the small amount of changes the users made. It was also verified that, among the four actions available to the users to change the segmentations, `Join` was the most common. Another major finding was that users accept all the logical days LDES suggested. The relevance of LDES is clearly supported by the current findings, where users have a positive reaction to the segmentation of their rolls. Together, these results suggest the users accept the LDES segmentation. They made few, or no modifications to the segmentations, and when they did, it was mainly to correct the *Level Of Detail* (*LoD*) in some parts of the roll. Changing the *LoD* is made by reducing the number of segments, maintaining the key cut points. Nevertheless, this was done sparsely. This is confirmed by the response to the perceived quality of the segmentation, that was positive, with a tendency towards the highest value.

The results also support the relevance of the *Logical Day*, as a key piece to settle context boundaries. This indicates that incorporating temporal cycles is important, if they are modelled to the way users perceive them. The test has demonstrated the acceptance of singular segments by the users. Such result is important to assess the LDES capability to detect isolated photos, a common feature in rolls gathered from smartphones.

To our best knowledge, this is the first time that binary relations has been used to explore the relation between segmentations in personal photo collections, exploring a qualitative approach complementary to a quantitative one. During the analysis, it became clear they are an important tool to understand or confirm the results of the tests.

The quantitative metrics, despite their importance, are insufficient for a complete analysis of the data.

6.3.2 MSS

The Multimedia Short Summary algorithm uses concept generalisation to summarise a set of temporal and spatial referenced photos. It is most useful to assist users during retrieval, providing an overview of the context using a limited number of information. The goal of the user test is to see if a reduction on the available information maintains or enhances the capability of the users to reconstruct the context of a photo set. The tests show MSS has a proper inter-cluster context separation, providing descriptions with enough information to give users the cues they need to identify correctly the spatio-temporal context of a set.

One major finding is that users take less time understanding the context using MSS, comparing to other summarisation strategies. This is particularly evident when the strategies, MSS an the other, use spatial dominance. The test also demonstrates that limiting the number of groups has advantages, at least for the spatial part of the context. The threshold of 16 groups proved its adequacy in the user test, showing that users are quicker in response times with few groups.

Other major finding is the perception of the context is not affected by the reduction of the available information. With MSS, users were better at describing the context than with others summarisation strategies.

7

Conclusions and Future Work

> *We must not confuse the present with the past. With regard to the past, no further action is possible.*
>
> ***Simone de Beauvoir***

This dissertation addresses the archival and retrieval of photos in personal collections, from a context manipulation point of view. It is known that those photos show large variability in the depicted items and may have hidden semantics. Such features lead to a set of difficulties in the analysis of photos, defeating a full automation of solutions on behalf of the users. Thus, the human factor is key in this domain. The so called *human-in-the-loop* approach sets the constraints needed to deal with the double role users play. On one hand, users are *clients* of the solution. On the other hand, they are an important source of information that will contribute to strengthen the archival, and thus, enhance the retrieval of photos. Thus, by means of annotations, the solution outsources the knowledge necessary to transform a generic photo management solution into a personal one, tailored to the user interacting with the system.

The research question under analysis is about supporting annotations, during the archival, and to explore that information to augment the knowledge about the context, both in archival and retrieval. Such approach lowers the annotation effort and improves the retrieval of a set of photos. We must remember this research is not about developing a full-fledge annotation system, that tries to seek as many tags as possible to cover most of the depicted items. There is no 'guessing' what annotations should be issued based on content analysis. Instead, the solution proposed is built on top of regularities and habits, providing the user with a directed aid to certain actions, namely, annotation and query.

The process used to tackle the problem consists of three steps:

1. narrow the domain to address;
2. gather and systematise knowledge;
3. develop specific algorithms for context manipulation, during archival and retrieval.

The first step is essential to concentrate the effort on solving a couple of problems with feasible and dependable solutions. We restrict the photo collections to those of personal use. Such restriction does not diminish the relevance of the problem, but allows us to make plausible assumptions about the domain. For example, the set of users interested on certain photos share a common knowledge about the context where they were taken. Those assumptions were important to tackle the problems identified. In step 2, the goal was to find a set of entities, their interconnections, their properties, and their relevance for describing the context in a personal domain. Once they are identified, the knowledge is systematised and stored in a way that it can be inserted, replaced and inferred. The last step concerns the developing of new algorithms based on the assumptions made, using the knowledge gathered in the second step. They comprise a segmentation algorithm — that separates the context in photo sets, preparing them for archival — and a summarisation algorithm that is used to improve the retrieval of photo sets, offering to the users a concise description of the context, at a proper level-of-detail. It resulted on a set of artefacts, namely:

1. a proof of concept prototype, named **MeMoT**, that is available at `http://purl.org/mont/memot`;
2. an ontology for describing the contexts that exists in personal photo collections, available at `http://purl.org/mont`.

In the next section, we summarise the contributions and the research findings. In Section 7.1, the conclusions for the foundational components are presented. They cover the ontology and the multidimensional context space. Next, Section 7.2 overviews the findings and conclusions for the archiving process, specially concerning the logical day event segmentation (LDES) algorithm. Finally, in Section 7.3, the findings and conclusions for the retrieval of photos are presented, in particular for the multimedia short summary (MSS) algorithm. We finalise the chapter pointing out future lines of work.

7.1 Foundational components

The **MeMoT** system was built around a knowledge base responsible for storing the personal photo collections (PPC), their annotations, the users and, most important, the metadata of all those entities. The metadata represents the common grounds from which the algorithms and the annotations are built, and thus, it is independent of a specific collection. The knowledge base core components are the **MOnt** ontology — the metadata, and the $\mathcal{MCS}$ multidimensional context space, as described in Chapter 3. Their development

was inspired by data warehouse (DW) solutions, where the goal is to provide query performance of unpredictable analysis, that works well against very large sets of data. The DW approach uses specific repository designs, based on a rich set of descriptive dimensions and metadata. There are some similarities between the DW domain and the PPC domain, but also some important differences. Among the similarities, we can count the following. The PPCs are actually large and their size continues to grow. The querying (or retrieval) is unpredictable, as finding photos is done differently by each person. Finally, the set of dimensions used to describe the context is quite standard in the literature, confined to the 4*Ws* — "*when*", "*where*", "*what*" and "*who*". However, an important difference exists in the characterisation of each dimension. The descriptive terms that enrich some of them depend not only on the domain, but also on the user's perspective. This is true for the 4*Ws*. The design of the knowledge base, with separated repositories for the metadata and the multidimensional context space is in part justified by the need to provide such adaptation to the users. The domain knowledge, common to all photo sets, is stored in the ontology — the meta-repository. Individual assertions are stored in the $\mathcal{MCS}$ — the data-repository. To our knowledge, this is the first time this setup is used for managing personal photo collections. On the past, works like [SJ08; LFBS08] used multiple dimensions to describe the content and context of the photos, but there is no metadata unifying the whole solution. The metadata, stored using the **MOnt** ontology, supports the lexicon of the annotations, but also enables new knowledge to be inferred from the assertions made to the collections. For example, from an incomplete set of family relations, other relations can be inferred.

7.1.1 Ontology

The metadata plays an important role, as it describes the domain of discourse shared by the intervening — the **MeMoT** system and the users. This enables a better comprehension of the semantics, benefiting some automations and suggestions on behalf of the users. Using an ontology to store the metadata is a natural solution, as we want to provide consistency checking, expandability and completion. The reasoning capabilities that can be built on top of ontologies allow their evolution with the dynamics of the system. In particular, new assertions can drive the inference on new knowledge. This was important for social relations, especially of the `Kinship` types. For example, the introduction of a fact stating the relation between two users, can be used to derive other relations, depending on the assertions already in the ontology. This simple feature reduces the effort demanded to users during annotation. The ontology is used almost in every algorithm and action implemented in **MeMoT**, including:

1. suggestions of annotations, during the archival;
2. suggestions of cues, used to perform a retrieval of photos;
3. settling the hierarchies that should be used in the MSS algorithm, for summarising

the context of a set of photos;

4. support for the viewpoint adjustments.

MOnt reuses many ontologies available in the literature, ranging from upper ontologies to domain specific ones. To our knowledge, it is the first time that life events and life scripts, two concepts borrowed from the sociology field, are used in this domain.

7.1.2 Multidimensional context space

The $\mathcal{MCS}$ enables the combinational power of the information present in the dimensions. The spatial, date and time information are built using common denominator terms, that are independent of the users. They represent the common sense knowledge we use everyday when we refer to those important dimensions in our lives. Dependent terms, including tags, activities and persons, are stored using a fixed structure with variable lexicon. The semantics of the terms are given by the ontology. In the worst case, they fall into the *tag* description, where the semantics is user dependent. Each photo is simultaneously represented in this space using different levels of detail, enabling users to pick the level that suits them better, without any additional processing. Another important aspect of $\mathcal{MCS}$ is that it works with incomplete information for most of the dimensions, excluding the temporal and spatial ones.

This strategy proved to be a proper solution for supporting the archival and retrieval activities, and the algorithms developed to support them. Although performance tests were outside the scope of this dissertation, the development tests show no bottlenecks writing or reading to the $\mathcal{MCS}$. Despite the absence of more robust stress tests to assure the quality of the solution, particularly regarding scale issues, the design seems to fit the personal photo collection domain, providing the same benefits as the ones found in DW solutions: *(i)* simplicity of the data model; *(ii)* performance.

7.2 Archival

The main contribution to the archival of personal photo collections is a segmentation algorithm, named LDES.

The claim is that regularities and cycles can be used to settle boundaries between events of similar context, helping to group photos that share annotations. This will help to reuse annotations and thus, it simplifies the manual labour. LDES sets its grounds on temporal regularities, mainly regarding the notion of day, extending it to the notion of logical day. This new notion of logical day is aligned with the daily activities of people, taken between rest periods. Such rest periods are important marks for people, but are often absent in the photoware available. LDES uses spatial information to fine tune the temporal segments, generating a segmentation that is temporal dominant. This is based on evidences from the literature, where the temporal order is important for people.

The user test supports the claim, providing evidences about the importance of the LDES algorithm. The first evidence is that users accept the automatic segmentation of their collection, making few or no changes. Most of the changes made consist of joining two segments. The overall perceived quality of the segmentations is positive. Another conclusion is the importance of the logical day to settle the boundaries for activities that continues after midnight. All the logical days were kept by the users. The singular segments, containing just one photo are also kept by the users, demonstrating the design goals of LDES were correct to address the needs of a proper segmentation algorithm for personal photo collections.

The experimental tests have demonstrated the importance of the theoretical comparison framework, developed along the LDES algorithm formalisation. To our best knowledge, this is the first time that binary relations have been used to explore the qualitative relation between segmentations in personal photo collections, as a complement to a quantitative approach. The analysis of the results from these two perspectives gives more insight about the causes and sets the grounds to a better explanation of the results.

The LDES algorithm can be used to regain more control over the archival of photos, either offline or online. Such control includes the structure, but also key annotations, that will help future action on the collections. In particular, the division in rolls and segments can help settle events whose activity is derived using the available annotations.

7.3 Retrieval

The main contribution to the retrieval of personal photo collections is the summarisation of a set of photos using the MSS algorithm. The claim, confirmed in the user tests, is that users can gain if they are assisted during the retrieval using a controlled amount of information, resulting from a summarisation process. The MSS algorithms consist of three parts:

1. a partition for a photo set;
2. a short text description for each group in the partition; and
3. a concise detail for the photo set, using an automatic level of detail selected for each dimension.

MSS can be set to operate in *temporal dominance*, where the temporal order of the photos is kept intra- and inter-group. Otherwise, the places with equal descriptions are kept inside each group, maintaining only the temporal order intra-group. This is called *spatial dominance*.

The advantages of MSS are *performance* and *focus*. Performance, because as the user test reveals, users take less time understanding the context using MSS, comparing to other summarisation strategies. This is particularly clear when using spatial dominant

summarisations strategies. Focus, because users were still capable to describe the spatio-temporal context of the photo sets, with better accuracy than with other summarisation strategies, with less groups. Thus, focusing on the right information, with the proper highlights, can produce accurate context comprehension.

There is also evidence of the relevance of the algorithm. Tests show an increasing stability on the groups created when the maximum numbers of allowed groups increases, reaching a steady position when this number if greater the 12. Besides, the limits design to be 16 groups at most, were also confirmed to be a proper solution in the user test, for producing fast results.

This type of summarisation strategy can be used in several scenarios where the information to search is vast and spatio-temporal. Although the MSS algorithm was developed and tested to assist the retrieval of personal photos, it is not bound to any specific type of object. This means that any spatio-temporal object can be used, as long as it is described by a set of hierarchical dimensions representing different levels of detail.

7.4 Future work

This research has raised many questions that need further investigation. In the following lines the most important ones are highlighted. This section is of particular importance, as it points out how the results achieved in this dissertation can be used to push even further the body of knowledge in this area.

LDES A natural progression of this work is to analyse how LDES can be extended to support hierarchical segmentation, incorporating large cycles. In particular the weekly cycle. There should be more research to understand if the logical day concept can be extended to the upper cycle or, if at that level of detail, the use of standard boundaries is enough for users. This is an important issue if the number of photos to archive is high, and spans through a large time frame.
Other area of investigation is the alignment of multi-roll archiving situations, that presents a lag between the temporal information. This would allow the generation of incremental segmentations, including new rolls into existing segmentations, to integrate the new information in existing segments and generating new segments whenever necessary.
Other research direction is to allow different segmentations to co-exist over the same roll, for different users. This allows a more personal structure for the storage of the photos.
Finally, since the usage of cloud storage is increasing, and users are using them to upload photos, LDES should evolve to support stream segmentation. This can be done from one camera stream or, with temporal alignment automatic solutions, from multi-cameras source.

MSS The summarisation algorithm relies on cluster' dominance to guarantee coherence in the summary. Selecting the proper dominance needs further research on how to find if it can be set automatically, based on the context (user and set included). Further work needs to be done to allow MSS to handle the 4*Ws*. Especially, to handle noisy data in "*where*" and incomplete information to support "*who*" and "*what*". Since their source is, mostly, user annotations, they can be missing for some photos. The approach should follow the current state of MSS, using attribute induction on social relation hierarchies (for identities), and rely on domain specific ontologies to hierarchically organise activities.

MOnt The ontology supports most of the important concepts in a personal domain. However, it lacks internationalisation support, not allowing users to use different languages to retrieve photos from $\mathcal{MCS}$ using their idiom.
More research is needed to support the evolution of the assertion, mimic the user's life changes. The current implementation does not allow, a user to change is wife, for example, without removing the old assertion. Future implementations should support multiple assertions, even if in contradiction, if the time frames they occur in are separate.

MeMoT We need to investigate which machine learning algorithms can be used to learn from previous interactions to better support and adapt to the users preferences. We can, for example, increase the value of annotations. The results of such algorithms should be trustworthy and reliable, two important features when dealing with users and their memories.
The current implementation selects the first photo of the group, when MSS is used to summarise a set of photos. More research is needed to better understand how the context can be misunderstood by the users, if a wrong choice of a photo is made. Since people pay much attention to visual artefacts, it is important the selected a photo that satisfies some quality criteria, but also, that can be seen as a visual translation of the textual description assigned to the group.
Considerably more work will be needed to determine how rolls and segments can be used to settle an hierarchical structure of events, browsable and searchable at different levels of detail. The current research focused on retrieving a set of photos. However, a user might want to retrieve a set of events, that share a common context for some given dimensions.
Regarding retrieval, further research is needed to assess if the dominance in MSS can be used to change the user interface accordingly, given different highlights and perspectives of the information.
More testing is needed to assess two situations: 1. to study the user's reaction to the suggestions made during the archival, and 2. to analyse the enhancements **MeMoT** brings to the retrieval. A further study could assess how **MeMoT** performs in long-term retrieval, after a period of continuous usage.

The current state of **MeMoT** could also be extended to the usage of algorithms for people detection and recognition, and scene detection. This would allow a better annotation support.

Bibliography

[AF10] S. A. Abdallah and B. Ferris. *The Ordered List Ontology*. online: http://smiy.sourceforge.net/olo/spec/orderedlistontology.html. 2010.

[Ado12] Adobe. *Extensible Metadata Platform*. ISO 16684-1:2012. 2012.

[Age04] N. G.-I. Agency. *World Geodetic System*. online: http://earth-info.nga.mil/GandG/publi-cations/tr8350.2/tr8350_2.html. 2004.

[AD04] L. von Ahn and L. Dabbish. "Labeling images with a computer game". In: *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems*. Vienna, Austria: ACM, 2004, pp. 319–326. ISBN: 1-58113-702-8. DOI: http://doi.acm.org/10.1145/985692.985733.

[All08] G. Allan. "Flexibility, friendship, and family". In: *Personal Relationships* 15.1 (2008), pp. 1–16.

[AF94] J. F. Allen and G. Ferguson. "Actions and events in interval temporal logic". In: *Journal of logic and computation* 4.5 (1994), pp. 531–579.

[All83] J. Allen. "Maintaining knowledge about temporal intervals". In: *Communications of the ACM* 26.11 (1983), pp. 832–843. ISSN: 0001-0782.

[AN07] M. Ames and M. Naaman. "Why we tag: motivations for annotation in mobile and online media". In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM. 2007, pp. 971–980.

[Arm97] D. M. Armstrong. *A world of states of affairs*. Cambridge Univ Press, 1997.

[Arm08] T. Armstrong. *The Human Odyssey: Navigating the Twelve Stages of Life*. Sterling, 2008.

[AMB07] K. Arras, O. Mozos, and W. Burgard. "Using Boosted Features for the Detection of People in 2D Range Data". In: *Robotics and Automation, 2007 IEEE International Conference on*. 2007, pp. 3402–3407. DOI: 10.1109/ROBOT.2007.363998.

[BCMNPS03] F. Baader, D. Calvanese, D. McGuiness, D. Nardi, and P. Patel-Schneider. *The description logic handbook: theory, implementation, and applications*. Cambridge University Press, 2003. ISBN: 0521781760.

[BH74] A. D. Baddeley and G. Hitch. "Working Memory". In: ed. by G. H. Bower. Vol. 8. Psychology of Learning and Motivation. Academic Press, 1974, pp. 47 –89. DOI: 10.1016/S0079-7421(08)60452-1.

[BN13] B. Ball and M. Newman. "Friendship networks and social status". In: *Network Science* 1 (01 Apr. 2013), pp. 16–30. ISSN: 2050-1250. DOI: 10.1017/nws.2012.4.

[BBL99] D. Beeferman, A. Berger, and J. Lafferty. "Statistical models for text segmentation". In: *Machine learning* 34.1-3 (1999), pp. 177–210.

[BBMN03] O. Bergman, R. Beyth-Marom, and R. Nachmias. "The user-subjective approach to personal information management systems". In: *Journal of the American Society for Information Science and Technology* 54.9 (2003), pp. 872–878. ISSN: 1532-2890. DOI: 10.1002/asi.10283.

[BLHL01] T. Berners-Lee, J. Hendler, and O. Lassila. "The semantic web". In: *Scientific american* 284.5 (2001), pp. 28–37.

[BR04] D. Berntsen and D. Rubin. "Cultural life scripts structure recall from autobiographical memory". In: *Memory & Cognition* 32.3 (2004), pp. 427–442.

[BJW00] C. Bettini, S. Jajodia, and S. Wang. *Time granularities in databases, data mining, and temporal reasoning*. Springer, 2000.

[BSST07] S. Boll, P. Sandhaus, A. Scherp, and S. Thieme. *MetaXa—Context-and content-driven metadata enhancement for personal photo books*. 2007.

[BE08] D. Boyd and N. Ellison. "Social network sites: Definition, history, and scholarship". In: *Journal of Computer-Mediated Communication* 13.1 (2008), pp. 210–230.

[BKNS00] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander. "LOF: identifying density-based local outliers". In: *ACM Sigmod Record*. Vol. 29. 2. ACM. 2000, pp. 93–104.

[BM10] D. Brickley and L. Miller. *FOAF Vocabulary Specification*. http://xmlns.com/foaf/spec/. W3C, 2010.

[Bru06] J. K. Brueckner. "FRIENDSHIP NETWORKS*". In: *Journal of Regional Science* 46.5 (2006), pp. 847–865. ISSN: 1467-9787. DOI: 10.1111/j.1467-9787.2006.00486.x.

[BPGP10] P. Bruneau, A. Pigeau, M. Gelgon, and F. Picarougne. "Geo-temporal structuring of a personal image database with two-level variational-Bayes mixture estimation". In: *Adaptive Multimedia Retrieval. Identifying, Summarizing, and Recommending Image and Music*. Springer, 2010, pp. 127–139.

[Bur08] C. Burt. "Time, language, and autobiographical memory". In: *Language Learning* 58 (2008), pp. 123–141.

[BKC03] C. Burt, S. Kemp, and M. Conway. "Themes, events, and episodes in autobiographical memory". In: *Memory & cognition* 31.2 (2003), p. 317.

[Bux01] W. Buxton. "Less is more (More or less)". In: *Buxton Design, Toronto Ontario* (2001), p. 17.

[CLP07] C. Cattuto, V. Loreto, and L. Pietronero. "Semiotic dynamics and collaborative tagging". In: *Proceedings of the National Academy of Sciences* 104.5 (2007), pp. 1461–1464.

[CZJ08] Y. Chai, X. Zhu, and J. Jia. "OntoAlbum: An Ontology Based Digital Photo Management System". In: *Proceedings of the 5th international conference on Image Analysis and Recognition*. Springer. 2008, pp. 263–270.

[CD97] S. Chaudhuri and U. Dayal. "An overview of data warehousing and OLAP technology". In: *ACM Sigmod record* 26.1 (1997), pp. 65–74.

[CH09] P. Cobley and N. Haeffner. "Digital cameras and domestic photography: communication, agency and structure". In: *Visual Communication* 8.2 (2009), pp. 123–146.

[Coh88] J. Cohen. *Statistical power analysis for the behavioral sciences*. Psychology Press, 1988.

[Coh92] J. Cohen. "A power primer." In: *Psychological bulletin* 112.1 (1992), p. 155.

[Con05] M. A. Conway. "Memory and the self". In: *Journal of Memory and Language* 53.4 (2005), pp. 594 –628. ISSN: 0749-596X. DOI: DOI:10.1016/j.jml.2005.08.005.

[CPP00] M. A. Conway and C. W. Pleydell-Pearce. "The construction of autobiographical memories in the self-memory system." In: *Psychological review* 107.2 (2000), p. 261.

[CFGW05] M. Cooper, J. Foote, A. Girgensohn, and L. Wilcox. "Temporal event clustering for digital photo collections". In: *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP)* 1.3 (2005), pp. 269–288. DOI: http://doi.acm.org/10.1145/1083314.1083317.

[CWXTT07] J. Cui, F. Wen, R. Xiao, Y. Tian, and X. Tang. "EasyAlbum: an interactive photo annotation system based on face clustering and re-ranking". In: *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*. San Jose, California, USA: ACM, 2007, pp. 367–376. ISBN: 978-1-59593-593-9. DOI: http://doi.acm.org/10.1145/1240624.1240684.

[DJLW08] R. Datta, D. Joshi, J. Li, and J. Z. Wang. "Image retrieval: Ideas, influences, and trends of the new age". In: *ACM Comput. Surv.* 40.2 (2008), pp. 1–60. ISSN: 0360-0300. DOI: http://doi.acm.org/10.1145/1348246.1348248.

[DKGS04] M. Davis, S. King, N. Good, and R. Sarvas. "From context to content: leveraging context to infer media metadata". In: *MULTIMEDIA '04: Proceedings of the 12th annual ACM international conference on Multimedia*. New York, NY, USA: ACM, 2004, pp. 188–195. ISBN: 1-58113-893-8. DOI: http://doi.acm.org/10.1145/1027527.1027572.

[DCM09] DCMI. *The Dublin Core metadata element set, ISO 15836:2009*. online: http://dublincore.org/specifications/. 2009.

[DMBDMJDM10] K. De Moor, K. Berte, L. De Marez, W. Joseph, T. Deryckere, and L. Martens. "User-driven innovation? Challenges of user involvement in future technology analysis". In: *Science and Public Policy* 37.1 (2010), pp. 51–61. DOI: 10.3152/030234210X484775.

[Dea89] T. Dean. "Using temporal hierarchies to efficiently maintain large temporal databases". In: *Journal of the ACM (JACM)* 36.4 (1989), pp. 687–718.

[Dey01] A. Dey. "Understanding and using context". In: *Personal and ubiquitous computing* 5.1 (2001), pp. 4–7. DOI: http://dx.doi.org/10.1007/s007790170019.

[DBGP11] T. M. T. Do, J. Blom, and D. Gatica-Perez. "Smartphone usage in the wild: a large-scale analysis of applications and context". In: *Proceedings of the 13th international conference on multimodal interfaces*. ICMI '11. Alicante, Spain: ACM, 2011, pp. 353–360. ISBN: 978-1-4503-0641-6. DOI: 10.1145/2070481.2070550.

[DWC09] M. Dork, C. Williamson, and S. Carpendale. "Towards visual web search: Interactive query formulation and search result visualization". In: *WSSP. Madrid, Spain* (2009).

[ESB92] M. Eldridge, A. Sellen, and D. Bedkerian. *Memory Problems at Work: Their Range, Frequency and Severity*. Rank Xerox, EuroPARC, 1992.

[ERJ07] D. Elsweiler, I. Ruthven, and C. Jones. "Towards memory supporting personal information management tools". In: *Journal of the American Society for Information Science and Technology* 58.7 (2007), pp. 924–946. ISSN: 1532-2890. DOI: 10.1002/asi.20570.

[Euz04] J. Euzenat. "An API for ontology alignment". In: *The Semantic Web–ISWC 2004*. Springer, 2004, pp. 698–712.

[ES+07] J. Euzenat, P. Shvaiko, et al. *Ontology matching*. Vol. 18. Springer, 2007.

[EFVC06] A. Evans, M. Fernández, D. Vallet, and P. Castells. "Adaptive multimedia access: from user needs to semantic personalisation". In: *Circuits and Systems, 2006. ISCAS 2006. Proceedings. 2006 IEEE International Symposium on*. IEEE. 2006, 4–pp.

[FPZ03] R. Fergus, P. Perona, and A. Zisserman. "Object Class Recognition by Unsupervised Scale-Invariant Learning". In: *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on* 2 (2003), p. 264. ISSN: 1063-6919. DOI: http://doi.ieeecomputersociety.org/10.1109/CVPR.2003.1211479.

[FTHSS10] A. Fialho, R. Troncy, L. Hardman, C. Saathoff, and A. Scherp. "What's on this evening? Designing User Support for Event-based Annotation and Exploration of Media". In: *1st International Workshop on EVENTS-Recognising and tracking events on the Web and in real life*. 2010, pp. 40–54.

[FLPCSB12] H. F. de Figueirêdo, Y. A. Lacerda, A. C. de Paiva, M. A. Casanova, and C. de Souza Baptista. "PhotoGeo: a photo digital library with spatial-temporal support and self-annotation". In: *Multimedia Tools and Applications* 59.1 (2012), pp. 279–305.

[Fiv11] R. Fivush. "The development of autobiographical memory". In: *Annual review of psychology* 62 (2011), pp. 559–582.

[Fri04] W. Friedman. "Time in autobiographical memory". In: *Social Cognition* 22.5: Special issue (2004), pp. 591–605.

[Gal07] W. O. Galitz. *The essential guide to user interface design: an introduction to GUI design principles and techniques*. Wiley. com, 2007.

[Gan06] A. Gangemi. *DOLCE-Lite*. Available at http://www.w3.org/TR/owl2-overview/. WonderWeb Foundational Ontologies Library, 2006.

[Gan07] A. Gangemi. *DOLCE+DnS Ultralite ontology*. online: http://www.loa.istc.cnr.it/ontologies/DUL.owl. 2007.

[GNV03] A. Gangemi, R. Navigli, and P. Velardi. "The OntoWordNet Project: extension and axiomatization of conceptual relations in WordNet". In: *On the move to meaningful internet systems 2003: CoopIS, DOA, and ODBASE*. Springer, 2003, pp. 820–838.

[Gan08] J. F. Gantz. "The Diverse and Exploding Digital Universe". In: 2008.

[Gar03] U. Gargi. *Consumer media capture: Time-based analysis and event clustering*. Tech. rep. Technical Report HPL-2003-165, HP Laboratories, 2003.

[Geo12] GeoNames.org. *GeoNames Ontology*. online: http://www.geonames.org/ontology/docu-mentation.html. 2012.

[GCA06] M. Georgescul, A. Clark, and S. Armstrong. "An analysis of quantitative aspects in the evaluation of thematic segmentation algorithms". In: *Proceedings of the 7th SIGdial Workshop on Discourse and Dialogue*. Association for Computational Linguistics. 2006, pp. 144–151.

[GH06] S. A. Golder and B. A. Huberman. "Usage patterns of collaborative tagging systems". In: *Journal of information science* 32.2 (2006), pp. 198–208.

[GP99] A. Gómez-Pérez. "Evaluation of taxonomic knowledge in ontologies and knowledge bases". In: (1999).

[GKS12] J. Gozali, M. Kan, and H. Sundaram. "Hidden Markov Model for event photo stream segmentation". In: *Multimedia and Expo Workshops (ICMEW), 2012 IEEE International Conference on*. IEEE. 2012, pp. 25–30.

[GGMPW02] A. Graham, H. Garcia-Molina, A. Paepcke, and T. Winograd. "Time as essence for photo browsing through personal digital libraries". In: *Proceedings of the second ACM/IEEE-CS joint conference on Digital libraries* (2002), pp. 326–335.

[Gro10] M. W. Group. *Guidelines For Handling Image Metadata*. online: http://www.metadataworkinggroup.org/pdf/mwg_guidance.pdf. 2010.

[Gru+93] T. R. Gruber et al. "A translation approach to portable ontology specifications". In: *Knowledge acquisition* 5.2 (1993), pp. 199–220.

[Gru00] R. Grush. "Self, World and Space: The Meaning and Mechanismsof Ego-and Allocentric Spatial Representation". In: *Brain and Mind* 1.1 (2000), pp. 59–92.

[Gye07] L. Gye. "Picture this: the impact of mobile camera phones on personal photographic practices". In: *Continuum* 21.2 (2007), pp. 279–288.

[Hab07] T. Habermas. "How to tell a life: The development of the cultural concept of biography". In: *Journal of Cognition and Development* 8.1 (2007), pp. 1–31.

[Hab11] T. Habermas. "Autobiographical reasoning: Arguing and narrating from a biographical perspective". In: *New Directions for Child and Adolescent Development* 2011.131 (2011), pp. 1–17.

[HF96] J. Han and Y. Fu. "Attribute-oriented induction in data mining". In: *Advances in knowledge discovery and data mining*. American Association for Artificial Intelligence. 1996, pp. 399–421.

[Han08] A. Hanbury. "A survey of methods for image annotation". In: *Journal of Visual Languages & Computing* 19.5 (2008), pp. 617 –627. ISSN: 1045-926X. DOI: DOI:10.1016/j.jvlc.2008.01.002.

[HLMS08] A. Hanjalic, R. Lienhart, W.-Y. Ma, and J. R. Smith. "The Holy Grail of Multimedia Information Retrieval: So Close or Yet So Far Away?" In: *Proceedings of the IEEE* 96.4 (2008), pp. 541 –547. ISSN: 0018-9219. DOI: 10.1109/JPROC.2008.916338.

[Har05] L. Hardman. "Canonical processes of media production". In: *Proceedings of the ACM workshop on Multimedia for human communication: from capture to convey*. ACM. 2005, pp. 1–6. ISBN: 159593247X.

[Has09] M. Hasselmo. "A model of episodic memory: mental time travel along encoded trajectories using grid cells". In: *Neurobiology of learning and memory* 92.4 (2009), pp. 559–573. ISSN: 1074-7427.

[HSA07] C. Havasi, R. Speer, and J. Alonso. "ConceptNet 3: a flexible, multilingual semantic network for common sense knowledge". In: *Recent Advances in Natural Language Processing*. 2007, pp. 27–29.

[Hea09] M. Hearst. *Search user interfaces*. Cambridge University Press, 2009.

[HP06] J. R. Hobbs and F. Pan. *Time Ontology in OWL*. http://www.w3.org/TR/owl-time/. W3C, 2006.

[HPS03] I. Horrocks and P. F. Patel-Schneider. "Reducing OWL entailment to description logic satisfiability". In: *The Semantic Web-ISWC 2003*. Springer, 2003, pp. 17–29.

[HPSBTGD+04] I. Horrocks, P. F. Patel-Schneider, H. Boley, S. Tabet, B. Grosof, M. Dean, et al. *SWRL: A Semantic Web Rule Language Combining OWL and RuleML*. Available at http://www.w3.org/TR/owl2-overview/. W3C Member Submission, 2004.

[HST00] I. Horrocks, U. Sattler, and S. Tobies. "Practical reasoning for very expressive description logics". In: *Logic Journal of IGPL* 8.3 (2000), pp. 239–263.

[Hou09] N. A. V. House. "Collocated photo sharing, story-telling, and the performance of self". In: *International Journal of Human-Computer Studies* 67.12 (2009), pp. 1073 –1086. ISSN: 1071-5819. DOI: 10.1016/j.ijhcs.2009.09.003.

[HSW12] E. van den Hoven, C. Sas, and S. Whittaker. "Introduction to this special issue on designing for personal memories: past, present, and future". In: *Human–Computer Interaction* 27.1-2 (2012), pp. 1–12.

[HDBW05] D. F. Huynh, S. M. Drucker, P. Baudisch, and C. Wong. "Time quilt: scaling up zoomable photo browsers for large, unstructured photo collections". In: *CHI'05 Extended Abstracts on Human Factors in Computing Systems*. ACM. 2005, pp. 1937–1940.

[IPT99] I. P. T. C. IPTC. *Information Interchange Model*. http://www.iptc.org/std/IIM/4.1/specification/IIMV4.1.pdf. 1999.

[IPT10] I. P. T. C. IPTC. *IPTC PhotoMetadata*. http://www.iptc.org/std/photometadata/specification/IPTC-PhotoMetadata-201007.pdf. 2010.

[JNTD06] A. Jaffe, M. Naaman, T. Tassa, and M. Davis. "Generating summaries and visualization for large collections of geo-referenced photographs". In: *MIR '06: Proceedings of the 8th ACM international workshop on Multimedia information retrieval*. Santa Barbara, California, USA: ACM, 2006, pp. 89–98. ISBN: 1-59593-495-2. DOI: http://doi.acm.org/10.1145/1178677.1178692.

[JS10] R. Jain and P. Sinha. "Content without context is meaningless". In: *Proceedings of the international conference on Multimedia*. MM '10. Firenze, Italy: ACM, 2010, pp. 1259–1268. ISBN: 978-1-60558-933-6. DOI: 10.1145/1873951.1874199.

[JLM10] V. Jain and E. G. Learned-Miller. "Fddb: A benchmark for face detection in unconstrained settings". In: *UMass Amherst Technical Report* (2010).

[JC10] JEITA and CIPA. *Exchangeable image file format 2.3*. http://www.cipa.jp/english/hyoujunka/kikaku/pdf/DC-008-2010_E.pdf. 2010.

[JAC11] R. Jesus, A. J. Abrantes, and N. Correia. "Methods for automatic and assisted image annotation". In: *Multimedia Tools and Applications* 55.1 (2011), pp. 7–26.

[JS05] E. Johnson and L. Schultz. "Forward telescoping bias in reported age of onset: an example from cigarette smoking". In: *International journal of methods in psychiatric research* 14.3 (2005), pp. 119–129.

[Joh+10] J. Johnson et al. *Designing with the mind in mind: Simple guide to understanding user interface design rules*. Morgan Kaufmann, 2010.

[KBS07] H. Kang, B. B. Bederson, and B. Suh. "Capture, annotate, browse, find, share: novel interfaces for personal photo management". In: *International Journal of Human-] Computer Interaction* 23.3 (2007), pp. 315–337.

[KR08] M. Kankanhalli and Y. Rui. "Application Potential of Multimedia Information Retrieval". In: *Proceedings of the IEEE* 96.4 (2008), pp. 712–720.

[KR+90] L. Kaufman, P. Rousseeuw, et al. *Finding groups in data: an introduction to cluster analysis*. Vol. 39. Wiley Online Library, 1990.

[KSZ07] Y. Kazakov, U. Sattler, and E. Zolin. "How many legs do I have? Non-simple roles in number restrictions revisited". In: *Logic for Programming, Artificial Intelligence, and Reasoning*. Springer. 2007, pp. 303–317.

[KA08] C. M. Keet and A. Artale. "Representing and reasoning over a taxonomy of part–whole relations". In: *Applied Ontology* 3.1 (2008), pp. 91–110.

[Kel89] A. Kellerman. *Time, space, and society: geographical societal perspectives*. Kluwer Academic Pub, 1989. ISBN: 0792301234.

[KPK10] W. Kim, J. Park, and C. Kim. "A novel method for efficient indoor–outdoor image classification". In: *Journal of Signal Processing Systems* 61.3 (2010), pp. 251–258.

[KR02] R. Kimball and M. Ross. *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling*. Wiley, 2002.

[KSFS05] T. Kindberg, M. Spasojevic, R. Fleck, and A. Sellen. "The ubiquitous camera: An in-depth study of camera phone use". In: *IEEE Pervasive Computing* 4.2 (2005), pp. 42–50.

[KSRW06] D. Kirk, A. Sellen, C. Rother, and K. Wood. "Understanding photowork". In: *Proceedings of the SIGCHI conference on Human Factors in computing systems*. ACM. 2006, pp. 761–770.

[KS10] D. S. Kirk and A. Sellen. "On human remains: Values and practice in the home archiving of cherished objects". In: *ACM Transactions on Computer-Human Interaction (TOCHI)* 17.3 (2010), p. 10. DOI: 10.1145/1806923.1806924.

[Kla98] R. L. Klatzky. "Allocentric and egocentric spatial representations: Definitions, distinctions, and interconnections". In: *Spatial cognition*. Springer. 1998, pp. 1–17.

[KS05] J. Kustanowitz and B. Shneiderman. "Motivating Annotation for Personal Digital Photo Libraries: Lowering Barriers While Raising Incentives". In: *Univ. of Maryland Technical Report HCIL-2004* 18 (2005).

[KSM12] S. C. Kwok, T. Shallice, and E. Macaluso. "Functional anatomy of temporal organisation and domain-specificity of episodic memory retrieval". In: *Neuropsychologia* 50.12 (2012), pp. 2943 –2955. ISSN: 0028-3932. DOI: http://dx.doi.org/10.1016/j.neuropsychologia.2012.07.025.

[LFBS08] Y. Lacerda, H. de Figueir, C. Baptista, and M. Sampaio. "PhotoGeo: a self-organizing system for personal photo collections". In: *Tenth IEEE international symposium on multimedia*. IEEE. 2008, pp. 258–265.

[LP09] R. Le Poidevin. "The Experience and Perception of Time". In: *The Stanford Encyclopedia of Philosophy*. Ed. by E. N. Zalta. Winter 2009. 2009.

[LWZ11] J. Li, T. Wang, and Y. Zhang. "Face detection using SURF cascade". In: *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*. 2011, pp. 2183–2190. DOI: 10.1109/ICCVW.2011.6130518.

[LSG07] J. Lieberman, R. Singh, and C. Goad. *W3C Geospatial Ontologies*. http://www.w3.org/2005/Incubator/geo/XGR-geo-ont/. W3C, 2007.

[Lie10] P. Lietz. "Research into questionnaire design". In: *International Journal of Market Research* 52.2 (2010), pp. 249–272.

[LS04] H. Liu and P. Singh. "ConceptNet—a practical commonsense reasoning tool-kit". In: *BT technology journal* 22.4 (2004), pp. 211–226.

[LLCEJKLY07] A. Loui, J. Luo, S.-F. Chang, D. Ellis, W. Jiang, L. Kennedy, K. Lee, and A. Yanagawa. "Kodak's consumer video benchmark data set: concept definition and annotation". In: *Proceedings of the international workshop on Workshop on multimedia information retrieval*. ACM. 2007, pp. 245–254.

[Low04] D. G. Lowe. "Distinctive image features from scale-invariant keypoints". In: *International journal of computer vision* 60.2 (2004), pp. 91–110.

[LBB06] J. Luo, M. Boutell, and C. Brown. "Pictures are not taken in a vacuum-an overview of exploiting context for semantic scene content understanding". In: *IEEE Signal Processing Magazine* 23.2 (2006), pp. 101–114. DOI: 10.1109/MSP.2006.1598086.

[Lux09] M. Lux. "Caliph & Emir: MPEG-7 photo annotation and retrieval". In: *Proceedings of the 17th ACM international conference on Multimedia*. ACM. 2009, pp. 925–926.

[LKF10] M. Lux, M. Kogler, and M. del Fabro. "Why did you take this photo: a study on user intentions in digital photo productions". In: *Proceedings of the 2010 ACM workshop on Social, adaptive and personalized multimedia interaction and access*. SAPMIA '10. Firenze, Italy: ACM, 2010, pp. 41–44. ISBN: 978-1-4503-0171-8. DOI: 10.1145/1878061.1878075.

[MMSV02] A. Maedche, B. Motik, N. Silva, and R. Volz. "Mafra—a mapping framework for distributed ontologies". In: *Knowledge engineering and knowledge management: ontologies and the semantic web*. Springer, 2002, pp. 235–250.

[MKP02] J. M. Martínez, R. Koenen, and F. Pereira. "MPEG-7: the generic multimedia content description standard, part 1". In: *Multimedia, IEEE* 9.2 (2002), pp. 78–87.

[MBGGOOSIcH02] C. Masolo, S. Borgo, A. Gangemi, N. Guarino, A. Oltramari, R. Oltramari, L. Schneider, L. P. Istc-cnr, and I. Horrocks. *WonderWeb Deliverable D17. The WonderWeb Library of Foundational Ontologies and the DOLCE ontology*. 2002.

[May99] D. J. Mayhew. "The Usability Engineering Lifecycle". In: *CHI '99 Extended Abstracts on Human Factors in Computing Systems*. CHI EA '99. Pittsburgh, Pennsylvania: ACM, 1999, pp. 147–148. ISBN: 1-58113-158-5.

[McC05] D. L. McCuinness. "Ontologies come of age". In: *Spinning the semantic web: bringing the World Wide Web to its full potential* (2005), p. 171.

[MTL78] R. McGill, J. W. Tukey, and W. A. Larsen. "Variations of box plots". In: *The American Statistician* 32.1 (1978), pp. 12–16.

[MB09] A. Miles and S. Bechhofer. *SKOS simple knowledge organization system reference*. Tech. rep. W3C, 2009.

[Mil56] G. Miller. "The magical number seven, plus or minus two: some limits on our capacity for processing information." In: *Psychological review* 63.2 (1956), p. 81.

[Mil95] G. A. Miller. "WordNet: A Lexical Database for English". In: *Communications of the ACM* 38 (1995), pp. 39–41.

[MO07] F. Monaghan and D. O'Sullivan. "Leveraging ontologies, context and social networks to automate photo annotation". In: *Semantic Multimedia*. Springer, 2007, pp. 252–255.

[MSS05] B. Motik, U. Sattler, and R. Studer. "Query answering for OWL-DL with rules". In: *Web Semantics: Science, Services and Agents on the World Wide Web* 3.1 (2005), pp. 41–60.

[MSH09] B. Motik, R. Shearer, and I. Horrocks. "Hypertableau reasoning for description logics". In: *Journal of Artificial Intelligence Research* 36.1 (2009), pp. 165–228.

[Mya07] G. Myatt. *Making sense of data: a practical guide to exploratory data analysis and data mining*. Wiley-Blackwell, 2007.

[MVCFA08] P. Mylonas, D. Vallet, P. Castells, M. Fernández, and Y. Avrithis. "Personalized information retrieval based on context and ontological knowledge". In: *The Knowledge Engineering Review* 23.01 (2008), pp. 73–100.

[NHWGMP04] M. Naaman, S. Harada, Q. Wang, H. Garcia-Molina, and A. Paepcke. "Context data in geo-referenced digital photo collections". In: *Proceedings of the 12th annual ACM international conference on Multimedia* (2004), pp. 196–203.

[NSPGM04] M. Naaman, Y. J. Song, A. Paepcke, and H. Garcia-Molina. "Automatic organization for digital photographs with geographic coordinates". In: *JCDL '04: Proceedings of the 4th ACM/IEEE-CS joint conference on Digital libraries*. Tuscon, AZ, USA: ACM Press, 2004, pp. 53–62. ISBN: 1-58113-832-6. DOI: http://doi.acm.org/10.1145/996350.996366.

[NYGMP05] M. Naaman, R. B. Yeh, H. Garcia-Molina, and A. Paepcke. "Leveraging context to resolve identity in photo albums". In: *JCDL '05: Proceedings of the 5th ACM/IEEE-CS joint conference on Digital libraries*. Denver, CO, USA: ACM Press, 2005, pp. 178–187. ISBN: 1-58113-876-8. DOI: http://doi.acm.org/10.1145/1065385.1065430.

[Nie94] J. Nielsen. "Usability inspection methods". In: *Conference companion on Human factors in computing systems*. ACM. 1994, pp. 413–414.

[NP01] I. Niles and A. Pease. "Towards a standard upper ontology". In: *Proceedings of the international conference on Formal Ontology in Information Systems - Volume 2001*. FOIS '01. Ogunquit, Maine, USA: ACM, 2001, pp. 2–9. ISBN: 1-58113-377-4. DOI: 10.1145/505168.505170.

[NY10] O. Nov and C. Ye. "Why do people tag?: motivations for photo tagging". In: *Commun. ACM* 53.7 (July 2010), pp. 128–131. ISSN: 0001-0782. DOI: 10.1145/1785414.1785450.

[NDR10] D. Novelli, J. Drury, and S. Reicher. "Come together: Two studies concerning the impact of group relations on personal space". In: *British Journal of Social Psychology* 49.2 (2010), pp. 223–236. ISSN: 2044-8309. DOI: 10.1348/014466609X449377.

[NR06] N. Noy and A. Rector. *Time Ontology in OWL*. http://www.w3.org/TR/swbp-n-aryRelations/. W3C, 2006.

[OOO10] P. Obrador, R. de Oliveira, and N. Oliver. "Supporting personal photo storytelling for social albums". In: *Proceedings of the international conference on Multimedia*. ACM. 2010, pp. 561–570.

[OSHT12] W. Odom, A. Sellen, R. Harper, and E. Thereska. "Lost in Translation: Understanding the Possession of Digital Things in the Cloud". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '12. Austin, Texas, USA: ACM, 2012, pp. 781–790. ISBN: 978-1-4503-1015-4. DOI: 10.1145/2207676.2207789.

[OGJLOS07] N. O'Hare, C. Gurrin, G. Jones, H. Lee, N. O'Connor, and A. Smeaton. "Using text search for personal photo collections with the MediAssist system". In: *Proceedings of the 2007 ACM symposium on Applied computing*. ACM. 2007, pp. 880–881. ISBN: 1595934804.

[OGW95] C. O'Muircheartaigh, G. Gaskell, and D. B. Wright. "Weighing anchors: Verbal and numeric labels for response scales". In: *JOURNAL OF OFFICIAL STATISTICS-STOCKHOLM-* 11 (1995), pp. 295–308.

[Ope] *OpenCyc*.
online: http://www.opencyc.org/. 2012.

[PJW00] D. K. Park, Y. S. Jeon, and C. S. Won. "Efficient use of local edge histogram descriptor". In: *Proceedings of the 2000 ACM workshops on Multimedia*. ACM. 2000, pp. 51–54.

[PS05] A. Payne and S. Singh. "Indoor vs. outdoor scene classification in digital photographs". In: *Pattern Recognition* 38.10 (2005), pp. 1533–1545.

[PH02] L. Pevzner and M. A. Hearst. "A critique and improvement of an evaluation metric for text segmentation". In: *Computational Linguistics* 28.1 (2002), pp. 19–36.

[Pig10] A. Pigeau. "MyOwnLife: incremental and hierarchical classification of a personal image collection on mobile devices". In: *Multimedia Tools and Applications* 46.2 (2010), pp. 289–306.

[Pil02] S. Pile. "Memory and the city". In: *Temporalities, autobiography and everyday life* (2002), pp. 111–127.

[PCF03] J. Platt, M. Czerwinski, and B. Field. "PhotoTOC: automatic clustering for browsing personal photographs". In: *Information, Communications and Signal Processing, 2003 and the Fourth Pacific Rim Conference on Multimedia. Proceedings of the 2003 Joint Conference of the Fourth International Conference on* 1 (2003), 6–10 Vol.1. DOI: 10.1109/ICICS.2003.1292402.

[RASG07] Y. Raimond, S. Abdallah, M. Sandler, and F. Giasson. "The music ontology". In: *Proceedings of the International Conference on Music Information Retrieval*. Citeseer. 2007, pp. 417–422.

[RLLOBBCCKMMRSSW04] L. M. Reeves, J. Lai, J. A. Larson, S. Oviatt, T. S. Balaji, S. Buisine, P. Collings, P. Cohen, B. Kraal, J.-C. Martin, M. McTear, T. Raman, K. M. Stanney, H. Su, and Q. Y. Wang. "Guidelines for Multimodal User Interface Design". In: *Commun. ACM* 47.1 (Jan. 2004), pp. 57–59. ISSN: 0001-0782. DOI: 10.1145/962081.962106.

[RD01] E. Reingold and N. Dershowitz. *Calendrical calculations*. Cambridge Univ Pr, 2001.

[RW03] K. Rodden and K. Wood. "How do people manage their digital photographs?" In: *Proceedings of the SIGCHI conference on Human factors in computing systems* (2003), pp. 409–416.

[RWJM03] T. Roenneberg, A. Wirz-Justice, and M. Merrow. "Life between clocks: daily temporal patterns of human chronotypes". In: *Journal of biological rhythms* 18.1 (2003), pp. 80–90.

[Ros03] R. Rosenzweig. "Scarcity or abundance? Preserving the past in a digital era". In: *The American Historical Review* 108.3 (2003), pp. 735–762.

[RW96] D. C. Rubin and A. E. Wenzel. "One hundred years of forgetting: A quantitative description of retention". In: *Psychological review* 103.4 (1996), pp. 734–760.

[SGS10] K. E. A. Van de Sande, T. Gevers, and C. G. M. Snoek. "Evaluating Color Descriptors for Object and Scene Recognition". In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 32.9 (2010), pp. 1582–1596. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2009.154.

[SB09] P. Sandhaus and S. Boll. "From usage to annotation: analysis of personal photo albums for semantic photo understanding". In: *Proceedings of the first SIGMM workshop on Social media*. ACM. 2009, pp. 27–34.

[SB11] P. Sandhaus and S. Boll. "Semantic analysis and retrieval in personal and social photo collections". In: *Multimedia Tools and Applications* (2011), pp. 1–29. ISSN: 1380-7501.

[SGJ01] S. Santini, A. Gupta, and R. Jain. "Emergent semantics through interaction in image databases". In: *Knowledge and Data Engineering, IEEE Transactions on* 13.3 (2001), pp. 337–351. ISSN: 1041-4347.

[ST06] R. Sarvas and M. Turpeinen. *Social Metadata for Personal Photography*. 2006.

[SBC07] A. Scherp, S. Boll, and H. Cremer. "Emergent semantics in personalized multimedia content". In: *Journal of Digital Information Management* 5.2 (2007), p. 55. ISSN: 0972-7272.

[SFSS09] A. Scherp, T. Franz, C. Saathoff, and S. Staab. "F–a model of events based on the foundational ontology dolce+ DnS ultralight". In: *Proceedings of the fifth international conference on Knowledge capture*. ACM. 2009, pp. 137–144.

[SFSS12] A. Scherp, T. Franz, C. Saathoff, and S. Staab. "A core ontology on events for representing occurrences in the real world". In: *Multimedia Tools and Applications* 58.2 (2012), pp. 293–331.

[SKK05] L. Seebeck, R. M. Kim, and S. Kaplan. "Emergent temporal behaviour and collaborative work". In: *ECSCW 2005*. Springer. 2005, pp. 123–142.

[Sel12] H. J. Seltman. "Experimental design and analysis". In: *Online at: http://www. stat. cmu. edu/, hseltman/309/Book/Book. pdf* (2012).

[SI09] L. Seneviratne and E. Izquierdo. "Image annotation through gaming (TAG4FUN)". In: *Proceedings of the 16th international conference on Digital Signal Processing*. Institute of Electrical and Electronics Engineers Inc., The. 2009, pp. 940–945.

[SCS01] I. Sethi, I. Coman, and D. Stan. "Mining association rules between low-level image features and high-level concepts". In: *Proceedings of the SPIE Data Mining and Knowledge Discovery* 3 (2001), pp. 279–290.

[SM97] R. Settersten and K. Mayer. "The measurement of age, age structuring, and the life course". In: *Annual Review of Sociology* 23 (1997), pp. 233–261.

[SFE11] P. Severi, J. Fiadeiro, and D. Ekserdjian. "Guiding the representation of n-ary relations in ontologies through aggregation, generalisation and participation". In: *Web Semantics: Science, Services and Agents on the World Wide Web* 9.2 (2011), pp. 83–98.

[STH09] R. Shaw, R. Troncy, and L. Hardman. "Lode: Linking open descriptions of events". In: *The Semantic Web* (2009), pp. 153–167.

[SSX07] B. Shevade, H. Sundaram, and L. Xie. "Modeling personal and social network context for event annotation in images". In: *Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries*. ACM. 2007, p. 134.

[Shn96] B. Shneiderman. "The eyes have it: A task by data type taxonomy for information visualizations". In: *Visual Languages, 1996. Proceedings., IEEE Symposium on*. IEEE. 1996, pp. 336–343. DOI: 10.1109/VL.1996.545307.

[Sin11] P. Sinha. "Summarization of archived and shared personal photo collections". In: *Proceedings of the 20th international conference companion on World wide web*. ACM. 2011, pp. 421–426.

[SJ08] P. Sinha and R. Jain. "Semantics In Digital Photos: A Contenxtual Analysis". In: *Semantic Computing, 2008 IEEE International Conference on*. 2008, pp. 58 –65. DOI: 10.1109/ICSC.2008.87.

[SPJ09] P. Sinha, H. Pirsiavash, and R. Jain. "Personal photo album summarization". In: *Proceedings of the 17th ACM international conference on Multimedia*. ACM. 2009, pp. 1131–1132.

[SWSGJ00] A. W. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. "Content-Based Image Retrieval at the End of the Early Years". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22.12 (2000), pp. 1349–1380. ISSN: 0162-8828. DOI: http://doi.ieeecomputersociety.org/10.1109/34.895972.

[SG02] B. Smith and P Grenon. *Basic formal ontology*. Draft. Downloadable at www.ifomis.org/bfo/. IFOMIS, 2002.

[SJRLC08] P. St. Jacques, D. Rubin, K. LaBar, and R. Cabeza. "The short and long of it: Neural correlates of temporal-order memory for autobiographical events". In: *Journal of Cognitive Neuroscience* 20.7 (2008). cited By (since 1996)29, pp. 1327–1341.

[SB07] B. Suh and B. B. Bederson. "Semi-automatic photo annotation strategies using event based clustering and clothing based person recognition". In: *Interacting with Computers* 19.4 (2007), pp. 524 –544. ISSN: 0953-5438. DOI: http://dx.doi.org/10.1016/j.intcom.2007.02.002.

[SP98] M. Szummer and R. Picard. "Indoor-outdoor image classification". In: *Content-Based Access of Image and Video Database, 1998. Proceedings., 1998 IEEE International Workshop on*. 1998, pp. 42 –51. DOI: 10.1109/CAIVD.1998.646032.

[Ten08] I. Tendolkar. "How Semantic and Episodic Memory Contribute to Autobiographical Memory. Commentary on Burt". In: *Language Learning* 58.s1 (2008), pp. 143–147.

[Tom10] C. L. Toma. "Affirming the self through online profiles: beneficial effects of social networking sites". In: *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM. 2010, pp. 1749–1752.

[Tri89] H. C. Triandis. "The self and social behavior in differing cultural contexts". In: *Psychological review* 96.3 (1989), pp. 506–520.

[Tul02] E. Tulving. "Episodic memory: From mind to brain". In: *Annual review of psychology* 53.1 (2002), pp. 1–25.

[VMCFCA05] D. Vallet, P. Mylonas, M. A. Corella, J. M. Fuentes, P. Castells, and Y. Avrithis. "A semantically-enhanced personalization framework for knowledge-driven media services". In: *Proceedings of IADIS International Conference on WWW/Internet (ICWI 2005)*. 2005.

[VHMVSS12] W. R. Van Hage, V. Malaisé, G. K. de Vries, G. Schreiber, and M. W. van Someren. "Abstracting and reasoning over ship trajectories and web data with the simple event model (SEM)". In: *Multimedia Tools and Applications* 57.1 (2012), pp. 175–197.

[VHDAFV05] N. Van House, M. Davis, M. Ames, M. Finn, and V. Viswanathan. "The uses of personal networked digital imaging: an empirical study of cameraphone photos and sharing". In: *CHI'05 extended abstracts on Human factors in computing systems*. ACM. 2005, pp. 1853–1856.

[VH07] N. A. Van House. "Flickr and public image-sharing: distant closeness and photo exhibition". In: *CHI '07 Extended Abstracts on Human Factors in Computing Systems*. CHI EA '07. San Jose, CA, USA: ACM, 2007, pp. 2717–2722. ISBN: 978-1-59593-642-4. DOI: 10.1145/1240866.1241068.

[VBFGVOM08] W. Viana, J. Bringel Filho, J. Gensel, M. Villanova-Oliver, and H. Martin. "PhotoMap: from location and time to context-aware photo annotations". In: *Journal of Location Based Services* 2.3 (2008), pp. 211–235.

[VJ01] P. Viola and M. Jones. "Rapid Object Detection using a Boosted Cascade of Simple Features". In: *cvpr* 01 (2001), p. 511. ISSN: 1063-6919. DOI: http://doi.ieeecomputersociety.org/10.1109/CVPR.2001.990517.

[VA06] L. Von Ahn. "Games with a purpose". In: *Computer* 39.6 (2006), pp. 92–94.

[W3C03] W3C. *WGS84 Geo Positioning: an RDF vocabulary*. http://www.w3.org/2003/01/geo/wgs84_pos. W3C, 2003.

[W3C04] W3C. *Resource Description Framework*. Available at http://www.w3.org/standards/techs/rdf. W3C Standard, 2004.

[W3C09] W3C. *OWL 2 Web Ontology Language: Document Overview*. Available at http://www.w3.org/TR/owl2-overview/. W3C Recommendation, 2009.

[Wag86] W. A. Wagenaar. "My memory: A study of autobiographical memory over six years". In: *Cognitive Psychology* 18.2 (1986), pp. 225 –252. ISSN: 0010-0285. DOI: DOI:10.1016/0010-0285(86)90013-7.

[WM10] S. von Watzdorf and F. Michahelles. "Accuracy of positioning data on smartphones". In: *Proceedings of the 3rd International Workshop on Location and the Web*. ACM. 2010, p. 2.

[Wei09] S. M. Weinschenk. *Neuro web design: what makes them click?* New Riders Publishing, 2009.

[WJ07] U. Westermann and R. Jain. "Toward a common event model for multimedia applications". In: *IEEE MULTIMEDIA* 14.1 (2007), p. 19. DOI: http://dx.doi.org/10.1109/MMUL.2007.23.

[Whi11] S. Whittaker. "Personal information management: From information consumption to curation". In: *Annual Review of Information Science and Technology* 45.1 (2011), pp. 1–62. ISSN: 1550-8382. DOI: 10.1002/aris.2011.1440450108.

[WBC10] S. Whittaker, O. Bergman, and P. Clough. "Easy on that trigger dad: a study of long term family photo retrieval". In: *Personal and Ubiquitous Computing* 14.1 (2010), pp. 31–43.

[WH99] C. Wickens and J. Hollands. *Engineering psychology and human performance.* Prentice Hall New Jersey, 1999.

[YNC09] R. Yan, A. Natsev, and M. Campbell. "Hybrid Tagging and Browsing Approaches for Efficient Manual Image Annotation". In: *IEEE MultiMedia* (2009), pp. 26–41.

[YLLCEJKL08] A. Yanagawa, A. C. Loui, J. Luo, S.-F. Chang, D. Ellis, W. Jiang, L. Kennedy, and K. Lee. *Kodak consumer video benckmark data set: concept definition and annotation*. Tech. rep. Columbia University, 2008.

[Zer96] E. Zerubavel. "Social memories: Steps to a sociology of the past". In: *Qualitative Sociology* 19.3 (1996), pp. 283–299. ISSN: 0162-0436.

[Zer85] E. Zerubavel. *Hidden Rhythms: Schedules and Calendars in Social Life*. University of California Press, 1985.

[ZPFKV12] C. Zigkolis, S. Papadopoulos, G. Filippou, Y. Kompatsiaris, and A. Vakali. "Collaborative event annotation in tagged photo collections". In: *Multimedia Tools and Applications* (2012), pp. 1–30.

[ZS93] J. Zuzanek and J. Smale. "Life-cycle variations in across-the-week allocation of time to selected daily activities". In: *SOCIETY AND LEISURE-MONTREAL-* 15 (1993), pp. 559–559. DOI: http://dx.doi.org/10.1007/0-306-47155-8_6.

Algorithms

A.1 LDES

Algorithm A.1 Day finding

Input: T - An ordered set of timestamps
G - A list of locations for each element of T
w - A threshold value used to detect logical days
Output: A segmentation S
A list of time statistics ξ for each segment

STEP_1(T, G, w, S)

```
1  S = ∅
2  ξ = ∅
3  // Obtain the first timestamp
4  lastTS = T(0)
5  for i = 0 to T.length
6      if DATE(lastTS) ≠ DATE(T(i)) ∨ T(i) − lastTS < w
7          // Add the new segment s = [t⁻, t⁺]
8          S.append(segment(lastTS, T(i − 1)))
9          ξ.append(currStat)
10     else
11         currStat = UPDATESTATISTICS(currStat, T(i))
12     lastTs = T(i)
13 S.append(segment(lastTS, T(T.length − 1)))
14 ξ.append(currStat)
15 return ξ
```

Algorithm A.2 Event Finding

Input: T - An ordered set of timestamps T
S - A segmentation with day segments
ξ - the statistics for the segments
f_t - a real value
Output: A segmentation S_{evt}, that is a refinement of S

STPE_2(T, S, ξ, f_t)

```
1  // The new segments
2  S_evt = ∅
3  for i = 0 to S.length
4       currStats = ξ(i)
5       lastTS = S(i).t⁻
6       for t_j = S(i).t⁻ to S(i).t⁺
7            // computeThreshold implements Equation 4.2
8            if lastTS − t_j) > ComputeThreshold(currStats, f_t)
9                 S_evt.append(segment(lastTS, t_j − 1))
10                lastTS = t_j
11      if lastTS = S(i).t⁻
12           // No refinement was done. Maintain segment S(i)
13           S_evt.append(S(i))
14      else
15           if lastTS! = S(i).t⁺
16                // Assure no timestamps were left behind)
17                S_evt.append(segment(lastTS, S(i).t⁺)
18 return S_evt
```

Algorithm A.3 Event tunning

Input: T - An ordered set of timestamps
G - A list of locations for each element of T
S - a segmentation containing the activities inside each day
w - A threshold value used to detect logical days
Δ_g - A list of spatial distances
fg - a real value
Output: A fine-tuned segmentation S_{final}

STEP_3(S, T, G, w, f_g)

$S_{split} = \varnothing$
$\delta_n = \varnothing$
// Compute the distance between photo i and its predecessor
// If the distance is an outlier, gets δ_nNULL
$\delta_n.\,append($COMPUTEDISTANCEANDVALIDATE$(G, T))$
for $i = 0$ **to** $S.\,length$
    $S_{tmp} =$ SPLIT$(S[i], T, \delta_n, f_g)$
    // S_{split} gets appended with the split result
    $S_{split} = S_{split} + S_{tmp}$
return JOIN$(S_{split}, \delta_n, f_g, w, T)$

SPLIT(s, T, δ_n, f_g)

// *ComputeSplitReference* implements Equation 4.5
$\Delta_s = ComputeSplitReference(s, f_g, \Delta_g)$
$S_{split} = \varnothing$
$lastTS = s.t^-$
for $t_i = s.t^-$ **to** $s.t^+$
    if Δ_g for photo with $t_i > \Delta_s$
        $S_{split}.\,append(segment(lastTS, t_i)))$
        $lastTS = t_i$
if $lastTS = s.t^-$
    // No split was done. Maintain segment s
    $S_{split}.\,append(s)$
else
    if $lastTS! = s).\,t^+$
        // Assure no timestamps were left behind)
        $S_{split}.\,append(segment(lastTS, s).\,t^+)$

// *VerifySplitGain* implements Equation 4.6
$toSplit = VerifySplitGain(S_{split}, s, f_g, \delta_n)$
if $toSplit =$ TRUE
    return S_{split}
else
    return $\{s\}$

JOIN$(S_{split}, \delta_n, f_g, w, T)$

$S_{final} = \varnothing$
// A valid segment verifies the cases in Figure 4.16
for each $i \in$ valid segment for joining in S_{split}
    if $DifferentLogicalDay(w, S_{split}(i), S_{split}(i+1))$
        CONTINUE
    $s = segment(S_{split}(i).\,t^-, S_{split}(i+1).\,t^+)$
    // VerifyJoinGain implements Equation 4.7
    $toJoin = VerifyJoinGain(s, S_{split}(i), S_{split}(i+1), f_g, \Delta_g)$
    if $toJoin =$ TRUE
        $S_{final}.\,append(s)$
        $i = i + 1$
    else
        $S_{final}.\,append(S_{split}(i))$
return S_{final}

A.2 MSS

Algorithm A.4 MSS Clustering algorithm

Input: M - an array containing M_1 and M_2;
$dominant$ - the index indicating the dominance;
κ - the maximum number of groups.
Output: A partition Q

MSSCLUSTER($M, dominant, \kappa$)

```
1  Q = [ ][ ]
2  lsAttrs = A_i^{LSD(κ)} for each matrix in M
3  STABLESORT(M[dominant], lsAttrs)
4  INNERCLUSTER(M, lsAttrs, 0, M.length, 1, Q)
5  return Q
```

INNERCLUSTER($data, lsAttrs, lb, hb, level, cluster$)

```
1  if data.Length == 0
2      return
3  level = 0, idx = 0
4  R = REMOVEFIRST(data)
5  d = REMOVEFIRST(lsAttrs)
6  last = R.first[d]
7  for each case in R
8      if case[d] ≠ last
9          INNERCLUSTER(data, lsAttrs, last, case, level + 1, cluster)
10         last = case, idx+ = 1
11     ADDCLUSTER(cluster[level], idx)
12 INNERCLUSTER(data, lsAttrs, last, data.last, level + 1, cluster)
```

Algorithm A.5 Description of a MSS Cluster

Input: $matrix$ - A matrix with the objects for one group;
$coverage$ - A value $\in]0, 1]$ that limits the selection of the attribute;
$racio$ - The value that specifies the ratio between the first and second most common values
Output: The most common value and the index of the attribute

PROPERDESCRIPTION($matrix, coverage, racio$)

```
1  for each column c in matrix in reverse order
2      n = c.length
3      get the count for the two most common values
4      if there is only one count or
           (most common count > n × coverage and
               most common count ≥ racio× second common count)
5          return most common value, c
6  return most common value, 0
```

B

Datasets

Dataset	No. Photos	No. Days	Day range	Photos w/ Geo. (%)	Km range
DS-1	179	13	14	100	1,098
DS-2	47	3	9	100	128
DS-3	97	5	7	100	77
DS-4	90	18	1,738	81	593
DS-5	171	9	10	100	1,161
DS-6	72	7	8	100	166
DS-7	40	4	5	100	80
DS-8	513	21	21	100	9,459
DS-9	306	6	20	100	8,242
DS-10	607	9	29	100	8,465
DS-11	164	9	9	100	15
DS-12	88	4	5	100	44
DS-13	126	6	7	100	172
DS-14	156	13	15	100	1,228
DS-15	17	2	3	100	4
DS-16	81	7	8	100	2,729
DS-17	448	12	13	100	2,555
DS-18	16	2	3	100	3
DS-19	280	4	5	100	2,151
DS-20	103	11	11	100	50
DS-21	44	3	3	100	20
DS-22	75	12	12	100	593
DS-23	263	16	720	100	1,366
DS-24	356	22	27	100	162
DS-25	238	9	9	100	3,260
DS-26	53	3	1,019	96	37
DS-27	194	2	3	100	8
DS-28	582	17	19	100	3,826
DS-29	97	5	7	100	77
DS-30	89	4	6	100	31
DS-31	186	9	13	100	78
DS-32	64	4	5	100	65
DS-33	157	22	26	100	686
DS-34	73	19	20	100	3,981
DS-35	230	41	44	100	4,435
DS-36	201	9	40	100	579
DS-37	111	9	10	100	421
DS-38	258	7	8	100	1,543
DS-39	1,395	13	1,708	72	6,357

Table B.1: Characterisation of the photo sets used in the experimental test of LDES.

Dataset	No. Photos	No. Days	Day range	No. Cities
DS-1	179	13	14 days	5
DS-2	47	3	9 days	2
DS-3	43	5	331 days	1
DS-4	171	9	10 days	8
DS-5	72	7	8 days	4
DS-6	40	4	5 days	2
DS-7	513	21	21 days	59
DS-8	306	6	20 days	10
DS-9	607	9	29 days	20
DS-10	291	12	15 days	4
DS-11	164	9	9 days	5
DS-12	126	6	7 days	9
DS-13	156	13	15 days	5
DS-14	17	2	3 days	2
DS-15	81	7	8 days	6
DS-16	448	12	13 days	14
DS-17	280	4	5 days	6
DS-18	44	3	3 days	4
DS-19	356	22	27 days	6
DS-20	238	9	9 days	13
DS-21	392	5	15 days	8
DS-22	26	1	11 hours	3
DS-23	182	1	15 hours	3
DS-24	36	1	17 hours	4
DS-25	138	1	12 hours	2
DS-26	10	1	8 hours	2
DS-27	137	3	316 days	2
DS-28	183	23	25 days	20
DS-29	582	17	19 days	17
DS-30	97	5	7 days	4
DS-31	89	4	6 days	3
DS-32	186	9	13 days	6
DS-33	64	4	5 days	2
DS-34	157	22	26 days	6
DS-35	73	19	20 days	8
DS-36	230	41	44 days	11
DS-37	201	9	40 days	2
DS-38	111	9	10 days	12
DS-39	258	7	8 days	6

Table B.2: Characterisation of the photo sets used in the experimental test of MSS.