PhotoMap: From Location and Time to Context-Aware Photo Annotations

Windson Viana[1*], José Bringel Filho[2], Jérôme Gensel, Marlène Villanova-Oliver, Hervé Martin

Laboratoire d'Informatique de Grenoble, équipe STEAMER
681, rue de la Passerelle, 38402 Saint Martin d'Hères, France

{Windson.Viana-de-Carvalho, Jose.De-Ribamar-Martins-Bringel-Filho, Jerome.Gensel, Marlene.Villanova-Oliver, Herve.Martin}@imag.fr

Abstract. Despite the growth of geotagged multimedia data on the Web, spatial and temporal metadata are still poorly exploited by Web search engines and multimedia systems. Interpretation and inference processes can enlarge these metadata towards high quality information that are useful for annotating multimedia automatically. In this paper, we propose a semi-automatic approach for annotating personal photos based on the use of OWL-DL ontologies together with the next generation of mobile devices. We present an ontology called ContextPhoto and a contextual photo annotation approach which improve the development of more efficient personal image management tools. ContextPhoto provides concepts for representing captured and inferred context information reusing Web standards that describes spatial, temporal and social networking data. In order to validate the context annotation process we propose, we have also designed and developed a mobile and Web location-based system. This new system, called PhotoMap, is an evolution of the related mobile annotation systems since it provides automatic annotation about the spatial, temporal and social contexts of a photo. We also present a demonstration of the PhotoMap application during a tourist tour in the city of Rome.

Keywords: Semantic Web, spatial ontologies, location-based services, mobile devices, and multimedia annotation.

1 Introduction

In the last years, the popularity of Web Geographical Information Systems (GIS), like Google Maps and Yahoo Maps, together with the possibility to reuse these systems through mashup techniques have increased the general public's interest in spatial information. Nowadays, people make available their personal multimedia content (i.e., photos, videos, blog posts, Web feeds, Wikipedia entries) on the Web, tagged with spatial information such as geographic coordinates and location description. Frequently, these spatial metadata describe where the multimedia content was produced, specially, for photos and videos. In other cases, metadata are used for georeferencing the main topics of some multimedia content. For example, metadata corresponding to an address or to geographic coordinates are linked to a set of news as Web feeds[3] (XML-based content) in order to indicate the geographic places that are related with each news item.

Despite this growth of *geotagged* multimedia content on the Web, spatial metadata are still poorly exploited by Web search engines and multimedia systems. Most of the systems, such as Yahoo! Flickr, Moving Blog[4], TripperMap[5] and WWMX (Toyama et al. 2003), only provide users with methods for

* Correspondent Author. Email: carvalho@imag.fr
[1] Supported by CAPES - Brazil
[2] Supported by the Program Alban, the European Union Program of High Level Scholarships for Latin America, scholarship no. E06D104158BR
[3] http://en.wikipedia.org/wiki/Web_feed
[4] http://www.movingblog.com/
[5] http://www.trippermap.com/

manually *geotagging* their content. In addition, one can navigate among the public-available multimedia documents by using a map-based interface. Although multimedia enriched annotation is a promising tool for outperforming the capabilities of multimedia systems, systems are still having little ability for deriving new useful information from users' geotagged spatial metadata. For instance, spatial metadata combined with date/time information associated with a photo shot provide significant information that can be applied for organizing photo collections and for increasing photo file descriptions (Matellanes et al. 2006, Naaman et al. 2004).

Web multimedia search engines could exploit these enriched annotations and the aggregated information from other spatial Web sources in order to offer innovated ways of searching multimedia documents instead of only querying filenames most of the time. For example, one can think of a new scenario in which a tourist, after arriving in a city, uses her mobile and location-based agent for searching public photos of the city on the Web. The agent could be configured for searching only photos taken by other tourists with very similar sightseeing preferences (e.g., people who like to see old churches). Moreover, one would like to select photos taken not so far by foot from her current position and taken during the same season. A location-based multimedia search engine will be able to answer this query if it is capable of integrating a wide variety of data from heterogeneous sources (i.e., user profiles, photo web sites, city streets information, and sensors data describing the current user situation) and if the content has metadata sufficiently enriched for the sources integration.

One of the main objectives of the Semantic Web is to make possible the development of such type of search engine that will help users to find the right data corresponding to a particular context of use (Shadbolt et al. 2006). Dey and Abowd (Dey and Abowd 2000) define context as "*any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant for the interaction between a user and an application, including the user and the applications themselves*". For instance, we can use contextual information for describing the photo shot situations (e.g., location, time, nearby persons).

In order to reach this goal, particularly for multimedia systems, *the first step* is the creation of standards (e.g., common vocabularies) for describing multimedia spatial metadata (and, if possible, to enlarge them with contextual metadata). These standards should also allow multimedia systems to share and to enrich multimedia annotations (Hollink et al. 2005). *The second step* is the design of methods for automating most of the multimedia annotation process. Spatial and temporal metadata (i.e., contextual information) describing the user's situation when she is creating a multimedia document should be exploited to infer contextual information for multimedia annotation. Furthermore, contextual metadata can be used together with content processing algorithms in order to increase the accuracy of automatic content annotation (Boutella 2003). In addition, enhanced contextual metadata will offer enough information to create innovative ways for organizing and navigating in personal multimedia collections. And finally, *the third step* consist in exploring the new and enriched multimedia annotation that will be made available in large-scale on the Web for delivering content to a user according to her context and her wishes.

In this paper, we propose an automatic approach for annotating a specific type of multimedia documents: personal photos. Our approach is based on the use of OWL-DL[6] ontologies together with the next generation mobile devices in order to accomplish the two first mentioned steps. Moreover, our approach set the fundamental bases of the *third step* towards the development of an intelligent location-based multimedia system. In this paper, we describe a vocabulary to represent captured and inferred context information reusing Web standards that describes spatial, temporal and social networking data. We present an ontology called ContextPhoto and a contextual photo annotation approach which allows the development of more efficient personal image management tools. ContextPhoto provides concepts for representing both spatial and temporal

[6] http://www.w3.org/TR/owl-guide/

contexts of a photo (i.e., where, when, what is near) and SWRL[7] rules for inferring the social context of a photo (i.e., who was nearby, was it a special event?). In order to validate the context annotation process we propose, we have also designed and developed a mobile and Web location-based system. This new system, called PhotoMap, is an evolution of the related mobile annotation systems since it provides automatic annotation about the spatial, temporal and social contexts of a photo (i.e., where, when, who and what was nearby). PhotoMap also offers a Web interface for spatial and temporal navigation in photo collections. In addition, users can look into the inferred and manual annotations of their photos and exploit them for retrieval purposes.

This paper is organized as follows: section 2 illustrates a generic annotation process that can be used for designing mobile photo annotation systems; section 3 presents briefly the technologies for representing image annotation and the ontology we have proposed for describing contextual information; section 4 describes the ContextPhoto ontology we propose for photo annotation; an annotation example that uses the ContextPhoto ontology is also depicted; section 5 gives an overview of the PhotoMap system describing its main components and processes; section 6 presents some related works in the field of image annotation in a situation of mobility; section 7 describes the first lessons learned during development and use of our annotation approach and, finally, we conclude in section 8 with some potential future works.

2 From Spatial/Temporal to Contextual and Content Annotation

Photo annotation consists in associating metadata with an image. Annotation highlights the significant role of photos in restoring forgotten memories of visited places, party events, and people. Moreover, the use of annotations allows the development of more efficient personal image tools. For instance, *geotagging* photos allows the visualization of personal photos in map-based interfaces such as Flickr Map[8]. In addition, people can find images taken near a given location by entering a city name or latitude and longitude coordinates into geotagging-enabled image search engines. Photo annotation works can be divided into two main categories: *context-based* and *content-based*. Furthermore, photo annotation approaches can be also classified according to the way they are performed: *manual*, *automatic* and *semi-automatic* approaches.

A *content-based* annotation characterises what the image depicts. For example, the objects and people that appear in a photo (e.g., Anne and her cat), spatial relations between the objects (e.g., the cat is on the left of Anne), or the activity illustrated on the photo (e.g., a birthday party) (Sarvas et al., 2004). Desktop image tools, such as Flickr[8] and ACDSee[9], support simple *manual* content annotation by using textual descriptions (i.e., tags). More elaborated systems, such as (Lux et al., 2004) and (Hollink et al., 2004), proposed *manual* content annotation assisted by image and spatial ontologies. For instance, one can describe in a formal way the image category (e.g., *onto:landscape*) and the spatial relations between image zones (e.g., the region A is *onto:on-the-left-of* region B). Despite of obvious advantages of image metadata, manual annotation requires a time-consuming effort from the user. In addition, annotation tools that automatically generate content annotations are still not convenient for photo organization since they do not fill the so-called semantic gap problem (Ames and Naaman 2007, Naaman et al. 2004). We claim, as the authors Naaman et al. (2004) and Sarvas et al. (2004) did, that metadata for describing the contextual information in the *automatic* annotation systems is particularly useful to photos organization and retrieval, since, for instance, knowing the location and the date/time of a photo shot often describes the photo itself a lot even before a single pixel is shown. Moreover, *automatic context-based* and *content-based* approaches can be combined to help systems to

[7] http://www.w3.org/Submission/SWRL/
[8] http://www.flickr.com/map/
[9] http://www.acdsee.com/

suggest for users other annotations and, mainly, structures for images organization. Hence, such generated information can reduce users' annotation effort, and transform this activity into a validation process.

Date and time of a photo shot are the most common contextual data (i.e., temporal data) acquired with traditional digital cameras. Image management tools as Adobe PhotoShop Album and (Pigeau and Gelgon 2004) use this timestamp in order to organize photos in a chronological order. They can also generate temporal clusters in order to facilitate the image consultation task. Another important source of image metadata is the EXIF file (i.e., *EXchangeable Image File*). When one takes a digital photo, most of the cameras insert contextual data into the header of the image file. This image metadata describes information about the photo (e.g., name, size, and timestamp), the camera (e.g., model, maximum resolution), and the camera settings used to take the photograph (e.g., aperture, shutter speed, and focal length). EXIF also allows the description of geographic coordinates using GPS tags. Several research projects, such as **PhotoCompas** (Naaman et al. 2004), **WWMX** (Toyama et al. 2003), and Web applications (e.g., Panoramio[10]) exploit spatial and temporal metadata together in order to organize photo collections. The major issues in these approaches are related to the lack of rich metadata associated with images. Most of the digital cameras only provide basic EXIF attributes. Additionally, as we mentioned above, manual association of location tags with an image file is also a time-consuming task. Furthermore, date/time and GPS coordinates are not the unique clues among all the context information that are useful to recall a photo (Matellanes et al. 2006, Sarvas et al. 2004).

Our approach addresses these issues by combining mobile devices as a primary source of context metadata with spatial, temporal, and social reasoning. The number of built-in sensors such as GPS, digital compass, and high resolution cameras for mobile devices increases constantly (Yamaba et al. 2005)(Viana et al. 2004). In addition, mobile devices are both pervasive and personal. Mobile phones track the person and have clues about her current situation allowing the acquisition of contextual information. Figure 1 shows an overview of a *semi-automatic content-based* and *context-based* annotation process designed for being used by mobile camera phones. Unlike traditional digital cameras, mobile devices can execute external applications (e.g, Brew, J2ME applications) allowing the development of mobile photo annotation systems (Ames and Naaman 2007, Naaman et al. 2004, Monaghan and O'Sullivan 2006). These applications can be designed for providing both manual content annotations and automatic context annotations. Manual annotation on the mobile device reduces the time lag problem that occurs in desktop annotation tools (Ames and Naaman 2007) since the user is still in the photo shot situation when she produces the photo metadata (Naaman et al. 2004).

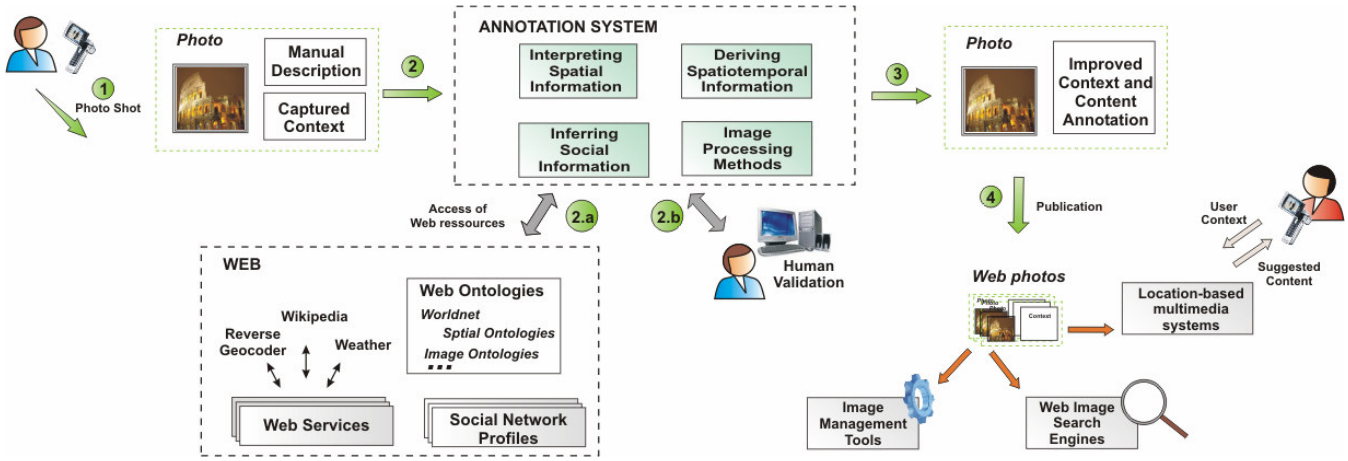

Figure 1 - Semantic image annotation process designed with mobile camera phones.

[10] http://www.panoramio.com/

Contextual metadata produced by the mobile device can be exploited for deriving high level annotations (see step 2 of Figure 1). For instance, annotation systems can transform GPS coordinates into a more suitable spatial representation (e.g., a city name). These system can combine location, orientation, and spatial reasoning in order to infer, for each photo, cardinal relationships (e.g., south of Rome) and 3D spatial relations (e.g., in front of the Coliseum) between the photo shot location and georeferenced content available on the Web (e.g., Wikipedia entries). The annotation system can also adopt an approach similar to (Naaman et al. 2004) in order to enrich the annotation of their photos using external context resources (e.g., weather conditions). In addition, annotation systems can use social network descriptions associated with contextual metadata for inferring the *social context* of a photo (e.g., the nearby user's friends) (Monaghan and O'Sullivan 2006). Furthermore, improved contextual annotations can be used together with algorithms of image process for increasing the accuracy of automatic content annotation. For example, both the location and camera settings information can be exploited in order to infer whether a photo is an outdoor or an indoor image (Boutella and Luob 2005).

Therefore, spatial, temporal and social reasoning can be applied to the initial contextual metadata in order to derive context and content annotations. This inferred information enables the development of smart image management tools that offer organization, retrieval and browsing functions outperforming the capabilities proposed by today's geotagging photo applications. In addition, location-based multimedia systems of the next generation will be able to exploit the availability of these new metadata for recommending multimedia content to a user whose current spatiotemporal context is *similar* to the context in which the multimedia have been created.

3 Image Annotation Representation

Existing approaches in multimedia annotation domain use several representation structures in order to associate metadata with images. These structures are attribute-value pairs inserted in the header of image files (e.g., EXIF, and IPTC formats), or more declarative representations such as the MPEG-7 standard and the RDF/OWL ontologies. In our work, we are especially interested in providing an interoperable annotation representation that is expressive enough in order to make it possible for next generation Web systems (e.g., Web search engines, location-based systems) to use, to share and to increase photo metadata. In order to satisfy these requirements, we have adopted an ontology representation for our image annotation. The use of ontologies for annotation is more suitable for making the content machine-understandable. Ontologies provide a common vocabulary to be shared between people and heterogeneous, distributed application systems. Moreover, annotations described in a formal and explicit way allow reasoning methods to be applied in order to infer about both the content and the context of a photo.

3.1. Content-based annotation and context ontologies

Ontology formally defines (specifies) concepts, relationships, and other distinctions that are relevant for modelling a domain (Gruber, 2008). Ontologies are typically used in two ways:

- *For defining a hierarchy of domain-specific concepts*. In this case, the ontology is created to describe categories (i.e., classes) of concepts and it uses the subsumption theory for expressing the hierarchy. In general, they are used to improve data retrieval. For example, spatial ontologies, such as Geonames[11],

[11] http://www.geonames.org

that represents the geographical administrative hierarchy of a Country (e.g., Continent> Country > Region > Department >City);

- *For provide representing a vocabulary and the relationships between terms in a formal way*. The ontology objective here is to reveal a common and standard domain vocabulary. The ontology content expresses semantic relations between terms (e.g., equivalency, disjunction). Typically, they are used to achieve data integration between heterogeneous systems. For instance, the RSS (Really Simple Syndication) small vocabulary defined in RDF allows the publication and aggregation of frequently updated Web content (e.g., news delivering).

In the field of multimedia annotation, the second ontology category is often encountered. For instance, image annotation ontologies, such as Dublin Core[12] and VRA Core Categories[13] specifies common terms for describing properties of the image itself (e.g., width, pixels quality) and some aspects of its creation (e.g., date, author name, license type). Other works, such as (Athanasiadis and Tzouvaras 2005, Schreiber et al. 2001, Hollink et al. 2004), take one step forward. They have proposed concepts for specifying subject matter of an image. For instance, these ontologies contain concepts for describing the visual features and their relationships (e.g., regions, person that appears in the image). However, these ontologies are well-suited for extensive *manual content-based* annotation and the elements for describing context concepts are limited.

The annotation process that we propose requires ontology for expressing a larger set of context properties and some important aspects (e.g., the subject type) for content-based annotation. In the domain of context-aware computing, the ontologies have been designed for characterising the user's context (Dey and Abowd 2000, Kirsch-Pinheiro et al. 2005, Petit et al. 2006). However, the challenges addressed by this community are different from those we face in the field of multimedia annotation. Approaches in context-aware computing mainly focus on the design and development of systems that can adapt their behaviour and content according to the current situation of the user and to her preferences. As a consequence, context ontologies in this field often limit the representation of the context to the concepts that are relevant for characterizing the interactions between the user and the system. However, some context information that is not directly useful for the interaction could reveal itself relevant for describing other system activities, such as data production. In our particular case, the observation of the circumstances in which a media has been created is more important than the contextual information related to the interaction between the user and the multimedia application. For example, even if the nearby people do not influence the interaction between the user and the digital camera application, this information can have a significant role for the annotation of the photo. Hence, we need to redefine context in order to fit with our specific needs and to create the corresponding ontology representation for supporting multimedia annotation.

3.2. Context definition for image annotation: Context Top ontology

We have modified the context definition given by (Dey and Abowd 2000) to get a more general definition. This new definition can be specialized for expressing the contextual concepts that we need for context-based photo annotation (see section 4). Thus, we have proposed the following definition:

"*Context is any information that can be used to describe the situation of entities and their relationships during an observation interval* Δt. *These observable entities are every abstract concepts and physical objects in a portion of space S that are relevant to a computer system for characterizing an entity or an action*".

[12] http://dublincore.org/

[13] http://www.vraweb.org/projects/vracore4/index.html

In our definition, an entity takes part of the context only if it is relevant and if its description can be captured by the context-aware system in its region of observation S. An entity in our definition, as for Dey, could be both abstract concepts (e.g., the user's activity, her humor, and her social relationships), and physical objects (e.g., nearby people, mobile phone). The region S represents the area of observation of the context-aware system. This spatial region S represents the intersection area between the system zones of interest and the area covered by the range of the system sensors. This region indicates the physical area that delimits the observation of the system. Unlike the Dey's definition, context for us is not only useful for describing the interaction between a user and a system, but for characterizing any entity and action that the system is interesting in know the situation. Each context-aware system determines the relevant entities (e.g., social and spatial-temporal information elements) to be acquired and to be represented as contextual information. For instance, in order to determine the context of a photo shot (i.e., the action of snap a photo) with a mobile phone, an annotation system can be interested in knowing who are the nearby people and what are the nearby objects (i.e., entities), and the social relations (i.e., abstract concepts) among the mobile phone's owner and these people in the "shot" time (i.e., the instant ***t***). In this case, the mobile display information is not a relevant entity since it does not increase the meaning of the photo shot.

We have created a top ontology for representing our general definition of context. The Figure 2 shows the OWL-DL representation of our context definition, called *Context Top*. We have defined the *Context Top* ontology in order to allow modeling of the notion of context in various application domains. The *Context* concept is composed by a set of *Context_Element* instances (i.e., hasContextElement property). Context_Element concept defines the relevant entities observed by the context-aware system. We have introduced the *describesTheSituationOf* property for describing the relation among the *Context* concept and the concepts *Action/Context_Element*. The property *hasContext* is the inverse property of *describesTheSituationOf*. We have defined five types of context elements for representing the dimensions of context information: spatial, temporal, social, computational and spatiotemporal. The determination of their respective elements depends on the application that will use the context information.

For the spatial and temporal dimensions, instead of proposing new concepts for representing geometric spatial and temporal information, we have used and extended other Web ontologies. We have done this since the real advantage of ontologies on the Web comes from the mix of number of ontologies and vocabularies, and their respective interconnection (Shadbolt et al. 2006). Thanks to this approach, a semantic web system can *"understand"* some concepts and may discover new relationships even if it only knows the semantic of the reused ontologies.

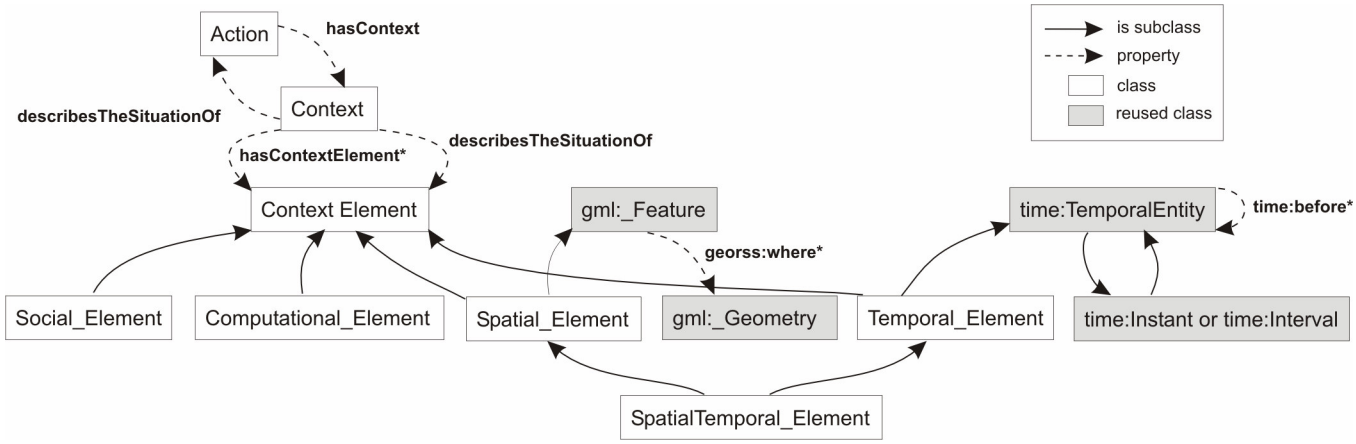

Figure 2 - Context Top ontology concepts and their relations with OWL-Time and GeoRSS ontologies

A similar effort has been done to represent spatial elements: we have used an OWL-DL representation of the GeoRSS ontology[14]. GeoRSS proposes a standardized way in which location is encoded with enough simplicity and descriptive power to satisfy most needs to describe the location of Web content. The GeoRSS GML version provides a variety of objects for describing geography features including coordinate reference systems, geometry, topology, time, units of measure and generalized values. GeoRSS GML provides a more simple description when compared to GML (Geographic Markup Language). However, its expressivity is sufficient to allow the aggregation and the share of geographic contents in the context of Semantic Web. We have adapted GeoRSS for describing location attributes in the Context Top ontology. The idea is to associate an ontology concept with a geographic feature for describing its geographic properties. Hence, spatial objects in our ontologies are expressed in a standard way. The Figure 2 shows the relation between the spatial element and the GeoRSS concepts (i.e., *gml:Geometry* and *gml:Feature*). A similar approach was chosen for temporal data representation. We have also used concepts of the OWL-Time[15] ontology. OWL-Time is a W3C ontology for describing the temporal content of Web documents and the temporal properties of Web services. The ontology provides a vocabulary for expressing facts about topological relations among instants and intervals, together with information about durations, and date/time data. OWL-Time was used for defining the temporal entities of Context Top ontology (see Figure 2).

4 Photo Annotation : ContextPhoto Ontology

4.1. What is useful for annotating a photo?

The main objectives of a photo annotation process are: *i)* facilitating the image retrieval from and navigation inside photo collections; and *ii)* highlighting people memories about the photo shot situation or the photo subject. Additionally, for professional use, the photo annotation is also important for revealing the configuration used of the photo camera (e.g., flash intensity, aperture). Consequently, the photo annotation concepts and relations should describe these characteristics.

Generally, people use contextual concepts for organizing and, consequently, retrieving their personal photos. For example, questions such as “Where and when has been taken the photo? Who was with me?” help people to categorize their photos since answers to these questions define the categories (e.g., photos taken in Paris, summer photos, and photos with Karol). Hence, majority of people apply spatial and temporal structures for categorizing, and subsequently, for researching their photo collections (Naaman et al. 2004). In order to permit the construction of these structures automatically, and to achieve the main objectives of the photo annotation process as mentioned above, we have extended Context Top ontology for representing the most important contextual attributes for photo annotation. Then, we have created an ontology, called ContextPhoto, for annotating photos and photo collections as well with contextual metadata and content descriptions. The Figure 3 shows the five context dimensions and their respective context elements in the ContextPhoto ontology.

[14] http://www.georss.org/
[15] http://www.w3.org/TR/owl-time/

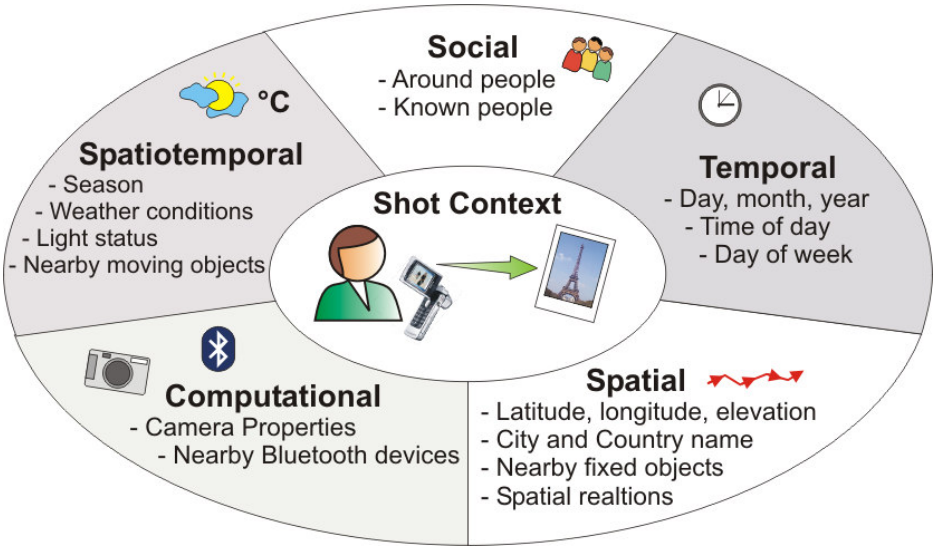

Figure 3 - Context elements of ContextPhoto ontology

The notion of place is one of the most useful information to recall personal photos (Naaman et al. 2004). However, one can treat this notion as having many senses (Christensen, 2001). The most common sense is the definition of a place as bounded space associated with a location or position within some order (e.g., a hierarchical order: Champs Elysées, Paris, France). The sense is derived from the political and geographical administrative organization of the space we are familiar with. A notion of place can also carry semantics that make sense for a few people or a community based on some shared experiences (e.g., 'our preferred beach'). Moreover, the concept of space often brings notions of relations among various geographical concepts. For instance, distance relations (e.g., near to the Eiffel Tower) and directional relations (on the centre of Champ de Mars). All these various senses for the notion of place should be used for annotating photos as an effort to facilitate the research and organisation of personal image collections. In order to achieve this objective, ContextPhoto support several forms for expressing spatial information associated with a photo: from geographic coordinates (latitude, longitude, elevation, coordination system) to derived spatial relations (e.g., near, in front of). It is important to note that the location recall clues can be related to both the camera location and the subject location. Hence, we have added to ContexPhoto the possibility of expressing these two kinds of location.

Time is another important aspect to be considered in the organization and the retrieval of personal photos. However, a precise date is not the temporal attribute the most used when a person tries to find a photo in a chronological way (Naaman et al. 2004). When a photo has not been taken at a date that can be easily remembered (e.g., birthday, Valentine's Day), people rather use the month, the day of week, the time of day, and/or the year information when formulating their query. Thus, the temporal elements of ContextPhoto allows the association of an instant (date and time) with a photo, and also of the different time interpretations and attributes listed above (e.g., night, Monday, July).

The combination of time and location can enhance the photo annotation. Usability tests show that the spatiotemporal attributes (e.g., season, the weather conditions, and the day light status) are useful photo clues that can be both used for search purposes and to increase the described information of a photo (Naaman et al. 2004, Matellanes et al. 2006). ContextPhoto contains concepts for expressing these spatiotemporal attributes whose values depend of both time and location information (e.g., the properties related to the weather conditions of the physical environment during the photo shot). Another significant context clue for remembering photo is the information about the presence of a person (e.g., who was with me? Who is in the

photo?) (Monaghan et al. 2006, Matellanes et al. 2006). The social dimension of ContextPhoto makes available concepts for expressing who were the nearby people and their social relations with the owner of the photo.

As we have mentioned above, camera configuration properties are also important concepts for annotating a photo. The computational context dimension describes the characteristics of the digital camera and the camera settings used to snap the picture (e.g., shutter speed, focal length). We have integrated the core attributes of the EXIF format in this context dimension. In addition, one can also represent the properties of the surrounding computation devices (e.g., RFID and bluetooth devices) captured in the zone of observation at the moment of the photo shot. The identification of these objects can be used for identifying the social context of a photo, or the nearby buildings or monuments.

Some concepts for annotating photos can not be easily categorized in only one context dimension of our ontology. However, the OWL W3C standard we used for representing our ontology support the multiple inheritance. With our ontology, one can represent a concept in more than a contextual dimension. For instance, a nearby georeferenced person can participate to the spatiotemporal context (i.e., subclass of Spatiotemporal Element) as well as she can be integrated to the social context (i.e., subclass of the Social element).

4.2. ContextPhoto main concepts

Besides the capacity of annotating photos with contextual metadata, ContextPhoto contains concepts for expressing photo collections metadata. Moreover, a photo collection can be associated with spatial and temporal data for describing a *spatiotemporal event*. The authors in (Naaman et al. 2004) and (Matellanes et al. 2006) have pointed out that grouping photos in events (e.g., a vacation, a tourist visit) is one of the most common way for people to recall and to organize their photos. Hence, we have introduced the *EventCollection* element in ContextPhoto. It represents a collection of photos associated with a time interval (inherited from the OWL-Time ontology) and an ordered list of track points (i.e., timestamp and geographic coordinates). The Figure 4 shows the main concepts of ContextPhoto and their relationships with those of Context Top ontology.

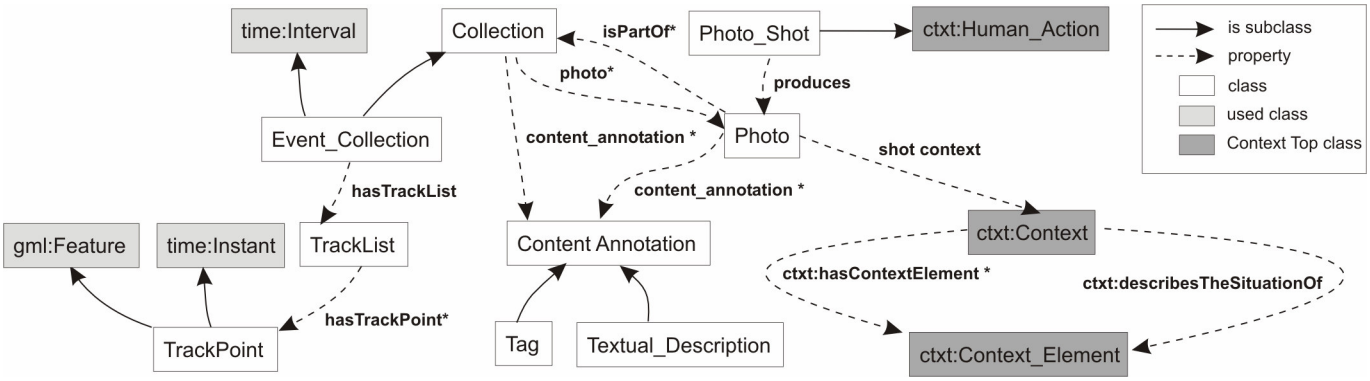

Figure 4 - ContextPhoto ontology

The *Photo* concept contains the basic image properties (e.g., name, width and height) described using the Dublin Core vocabulary. Besides those properties, each photo has two annotation types: content annotation (*Content Annotation*) and contextual annotations (*ctx:Context*). An annotation system can use the *Content Annotation* in order to describe the image content through a manual textual annotation (e.g., keyword tags) or by integrating ContextPhoto with other image annotation ontologies such as the Visual Descriptor Ontology

(VDO) (Athanasiadis and Tzouvaras 2005). In an effort to represent the contextual metadata, we have extended, in accord to the requirements we have presented in section 4.1, the five context dimensions of Context Top Ontology. These concepts correspond to the major elements for describing a photo (i.e., where, when, who, with, what) (Naaman et al. 2004, Matellanes et al. 2006, Sarvas et al. 2004).

4.3. Spatial concepts and spatial relations

With regard to the spatial context dimension, we have extended the *Spatial Element* concept for incorporating different semantic representation levels of a location description. We have introduced two spatial concepts: *Place* and *Georeferenced_Object*. The *Place* concept represents the location where the camera or the photo subject was positioned at the photo shot. *Place* has an address description and, as a *Spatial Element*, it can express its coordinates by reusing ontology geoRSS concepts and theirs relations. The address is semantically divided in street, city, and country data. The *GeoReferencedObject* represents contextual points of interest for annotating a photo. An annotation system may be concerned in representing spatial relations between the photo location and some points of interest (e.g., city monuments). Hence, we have used the spatial ontology proposed in (Reitsma and Hiramatsu 2006) in order to represent topological, distance, and directional relations between a photo location (i.e., the *Place* concept) and georeferenced objects. The table 1 shows the DL Syntax and the respective OWL description of the spatial concepts used in ContextPhoto. The table 1 also presents the geometric and spatial relations among those concepts.

Table 1 - Spatial concepts and relations of ContextPhoto (note that only a small part is presented)

	DL Syntax	OWL description
(1)	$\top$	*owl:Thing*
(2)	gml:Feature $\sqsubseteq \top$ gml:Geometry $\sqsubseteq \top$	*gml:Feature* and *gml:Geometry* are *rdfs:subClassOf owl:Thing*
(3)	ctxt:Spatial_Element $\sqsubseteq$ gml:Feature	*ctxt:Spatial_Element* is *rdfs:subClassOf gml:Feature*
(4)	geoRSS:where $\exists$geoRSS:where. $\top \sqsubseteq$ gml:Feature $\top \sqsubseteq \forall$geoRSS:where.gml:Geometry	*geoRSS:where* is a *rdfs: ObjectProperty* *rdfs:domain = gml:Feature* *rdfs:range = gml:Geometry*
(5)	Place $\sqsubseteq$ ctxt:Spatial_Element	*Place subClassOf ctxt:Spatial_Element*
(6)	Georeferenced_Object $\sqsubseteq$ ctxt:Spatial_Element Wikipedia_Entry $\sqsubseteq$ Georeferenced_Object	*Georeferenced_Object* is *subClassOf ctxt:Spatial_Element* *Wikipedia_Entry* is *subClassOf Georeferenced_Object*
(7)	hasSpatialRelation $\exists$ hasSpatialRelation. $\top \sqsubseteq$ Place $\top \sqsubseteq \forall$ hasSpatialRelation.Georeferenced_Object	*hasSpatialRelation* is a *rdfs: ObjectProperty* *rdfs:domain = Place* *rdfs:range = Georeferenced_Object*
(8)	distanceRelation $\subseteq$ hasSpatialRelation near $\sqsubseteq$ distanceRelation	*distanceRelation* is *rdfs:subPropertyOf hasSpatialRelation* *near* is *rdfs:subPropertyOf distanceRelation*
(9)	3Drelation $\sqsubseteq$ hasSpatialRelation inFrontOf $\sqsubseteq$ distanceRelation	*3DRelation* is *rdfs:subPropertyOf hasSpatialRelation* *infFrontOf* is *rdfs:subPropertyOf 3DRelation*

According to (Gruetter et al., 2007), the spatial relations derived from RCC-8 calculus can not be calculated using the OWL reasoning methods available nowadays (e.g., Classification reasoning). However, the authors show that a hybrid spatial knowledge representation approach can be developed. The spatial relations can be calculated using traditional reasoning approaches and the inferred relations should be represented in OWL. For instance, an annotation system can explore GPS and Compass data, and a GIS database for calculating spatial relations between a photo location and tourist attractions (e.g., photo was taken besides the Coliseum). Afterwards, the annotation system represents the results with the ContextPhoto spatial concepts. The representation of those spatial concepts and relations in the annotation ontology

highlights the photo description. Moreover, this information can be exploited by an image search engine for expanding photo queries semantically.

4.4. ContextPhoto Annotation Example

ContextPhoto contain concepts for expressing both manual content annotation information and automatic (captured and inferred) contextual annotation. A mobile and location-based annotation system can exploit ContextPhoto as the common vocabulary for representing content and context metadata of mobile photos. The Figure 5 illustrates a visual representation of an annotation example.

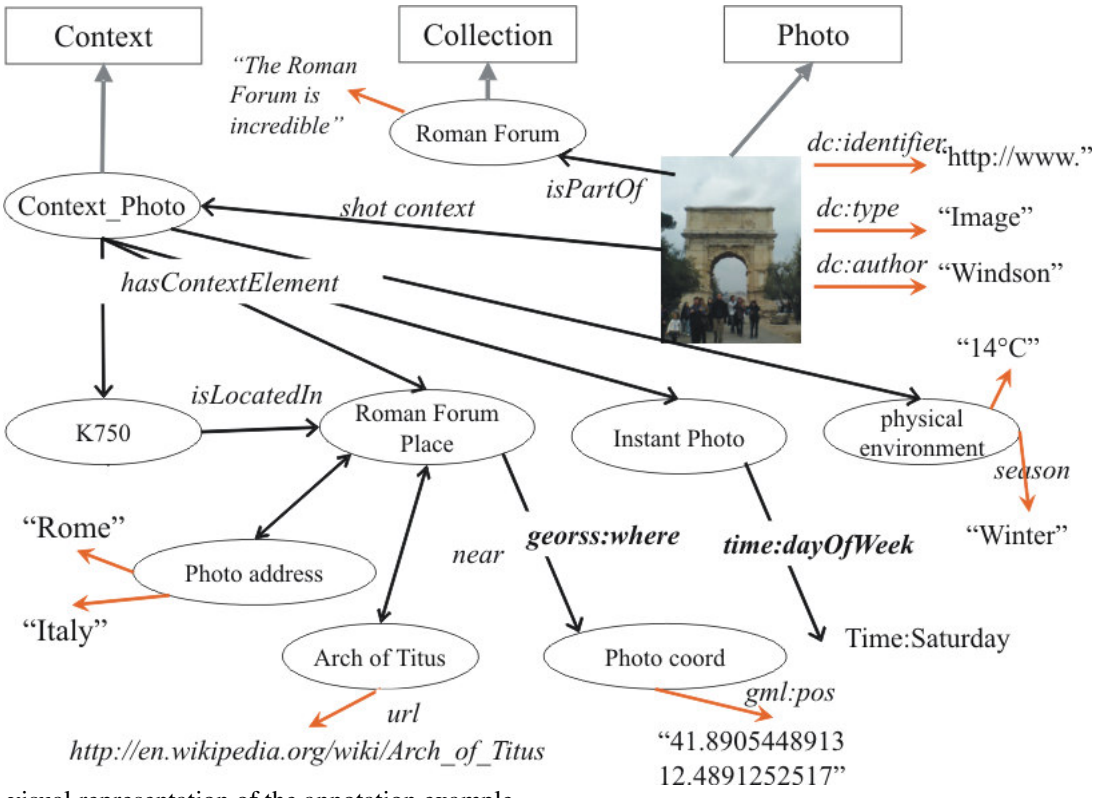

Figure 5 - A visual representation of the annotation example

The example contains a reduced and simplified representation of the spatial, computational, temporal and spatiotemporal context dimensions of the image annotation. The manual content annotation of the photo collection is also presented in the example. The example can be easily traduced in natural language as: "*Windson* has taken a photo with a *K750* camera. The photo is accessible at *http://www-lsr.imag.fr/users/Windson.Viana/images/romanforum3.jpg*. This photo is part of the *Roman Forum* collection. *Windson* has written about this collection: *"The Roman Forum is incredible"*. *Windson* has snapped the picture a *Saturday* (*12/01/2008*). The camera was located at the geographical coordinates *41.89054489135742 12.48912525177002*, what is close to the address *Via Sacra, Rome, Italy*. This place is near to the *Arch of Titus* (*http://en.wikipedia.org/wiki/Arch_of_Titus*). In this *winter* day, the temperature was *14°C*".

The Figure 6 shows the RDF serialization of the annotation example. One can observe that the collection description is not in the same annotation file of the photo. Instead of this, a link is created between the two Owl documents (see in the RDF serialization the *isPartOf* property).

```
.....
<ctxtphoto:Photo rdf:ID="romanforumphoto3">
   <dc:identifier rdf:datatype="http://www.w3.org/2001/XMLSchema#string">http://www-
      lsr.imag.fr/users/Windson.Viana/images/romanforum3.jpg</dc:identifier>
   <dc:author rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Windson</dc:author>
   <ctxtphoto:isPartOf>
      <rdf:Description rdf:about="http://www-
      lsr.imag.fr/users/Windson.Viana/ontologies/RomanForumCollection.owl#Roman_Forum">
      <ctxtphoto:photo rdf:resource="#romanforumphoto3" />
   </rdf:Description></ctxtphoto:isPartOf>
   <ctxtphoto:shot_context>
      <cxt:Context rdf:ID="context_photo">
      <cxt:hasContextElement>
         <ctxtphoto:Photo_Shot_Instant rdf:ID="instant_photo">
         <time:inXSDDateTime rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2008-01-12
      15:20:30</time:inXSDDateTime></ctxtphoto:Photo_Shot_Instant>
      </cxt:hasContextElement>
      <cxt:hasContextElement>
         <ctxtphoto:Place rdf:ID="romanforum3Place">
         <georss:where>gml:Point rdf:ID="photo-coord">
            <gml:pos rdf:datatype="http://www.w3.org/2001/XMLSchema#string">41.89054489135742 12.48912525177002</gml:pos>
            </gml:Point> </georss:where>
         <ctxtphoto:address>
      <ctxtphoto:Address rdf:ID="photoAddress">
         <ctxtphoto:street_address rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Via Sacra</ctxtphoto:street_address>
         <ctxtphoto:country rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Italy</ctxtphoto:country>
         <ctxtphoto:city rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Rome</ctxtphoto:city>
         <ctxtphoto:countryabrev xml:lang="en">IT</ctxtphoto:countryabrev>
      </ctxtphoto:Address> </ctxtphoto:address>
      <ctxtphoto:near>
         <ctxtphoto:WikipediaEntry rdf:ID="archoftitus">
         <georss:where>
            <gml:Point rdf:ID="wikipoint1">
               <gml:pos rdf:datatype="http://www.w3.org/2001/XMLSchema#string">41.8907 12.4886</gml:pos>
            </gml:Point></georss:where>
          <ctxtphoto:near rdf:resource="#romanforum3Place" />
       <ctxtphoto:url
       rdf:datatype="http://www.w3.org/2001/XMLSchema#string">http://en.wikipedia.org/wiki/Arch_of_Titus</ctxtphoto:url>
       </ctxtphoto:WikipediaEntry> </ctxtphoto:near>
    </ctxtphoto:Place>
  </cxt:hasContextElement>
 <cxt:hasContextElement>
    <ctxtphoto:Physical_Environment rdf:ID="physical_enviroment">
        <ctxtphoto:temperature_value rdf:datatype="http://www.w3.org/2001/XMLSchema#float">14.0</ctxtphoto:temperature_value>
        <georss:where>
             <gml:Point rdf:resource="photo-coord"/>
             <ctxtphoto:temperature_unit
      rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Celsius</ctxtphoto:temperature_unit>
           <ctxtphoto:season rdf:datatype="http://www.w3.org/2001/XMLSchema#string">winter</ctxtphoto:season>
     </ctxtphoto:Physical_Environment>
 </cxt:hasContextElement>
.....
```

Figure 6 - RDF Serialization of the annotation example

4.5. Social Context Inference: ContextPhoto and SWRL rules

One of the innovating features of the ContextPhoto ontology is the ability for describing the social context of a photo. The main idea is to represent who were the people present close to the location where the photo has been taken (see Figure 7). The description of social context is inspired from the proposal of Monaghan and O'Sullivan (2006) and FoafMobile[16], and uses the Bluetooth address of personal devices for detecting a person's presence. The method that we propose consists in associating the Bluetooth address of a personal device with a personal profile. In addition, the personal profiles can be interconnected for representing social networks. Hence, when someone takes a photo with her mobile camera phone, the mobile device searches for

[16] http://jibbering.com/discussion/Bluetooth-presence.1

Bluetooth devices nearby and the camera application annotates the photo file with the nearby Bluetooth addresses. Later, a system reads the photo annotation and searches the photographer profile and the profiles of people who belong to her social network. With both data, the system can infer if the nearby devices annotated in the photo metadata correspond to the devices of some photographer's acquaintances.

We have introduced concepts for expressing nearby devices in ContextPhoto. Hence, an annotation system can use the computational dimension of our ontology for annotating a photo with the detected bluetooth devices. The Figure 7 shows an example of device detection and its corresponding annotation. Each identified Bluetooth address is described by using the BTDevice class. In the example, three addresses of the nearby Bluetooth devices have been detected (i.e., 00:0A:D9:EB:66:C7, 00:0A:D9:EB:50:B3, and 00:0A:D9:EF:33:Q2). At this stage, the names of the nearby persons and their relations with the photographer are still unknown (i.e., the social context was not yet inferred).

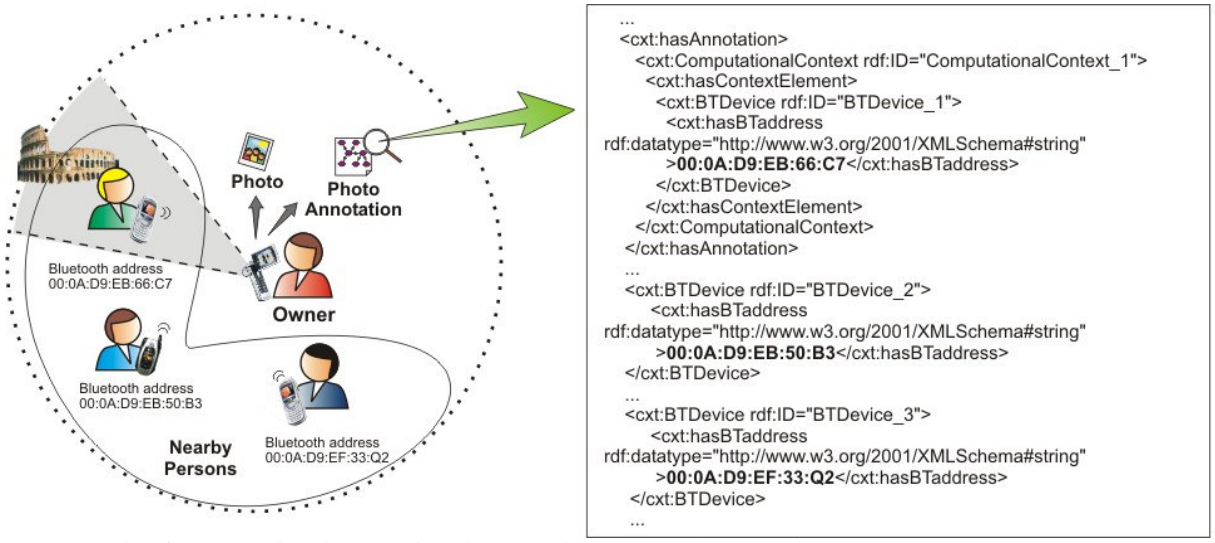

Figure 7 - Example of computational annotation that uses the ContextPhoto ontology.

In order to represent people and their social relations, we have used Friend-Of-A-Friend[17] (FOAF) profiles. FOAF is a RDF ontology that allows the description of a person (name, personal image, and birthday) and of her social networks (the person's acquaintances). FOAF does not support natively the association of the Bluetooth address with a profile. Hence, we have defined in ContextPhoto the concept *Person* that permits this association. The *Person* is defined as a subclass of *foaf:Person* that has a Bluetooth device identification. We have also created a subclass of Person, the *Owner* concept, for describing the owner of the photo collection annotated with ContextPhoto (e.g., the photographer or the annotation system user).

We suppose that a Web image management system (e.g., Flickr) can describe its users and social networks using FOAF profiles (see Figure 8). These Web tools can provide interfaces which allow users to link their FOAF profile with other profiles on the Web. The system can permit a user to associate Bluetooth addresses for identifying her FOAF profile by using the vocabulary available in ContextPhoto. Moreover, this association can also be set up automatically. One can download a mobile application for acquiring the Bluetooth address of one's mobile device and, later, the mobile application sends this information to the image management system server. Subsequently, the Web server executes the association of the acquired Bluetooth address with the user's FOAF profile.

[17] http:// www.foaf-project.org/

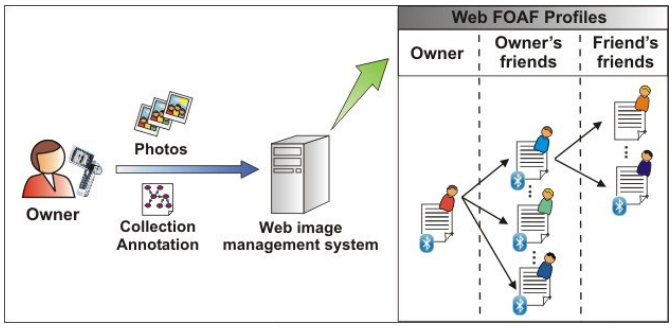

Figure 8 - Example of the web image management system that uses the ContextPhoto ontology and FOAF Profiles

With the availability of photo annotated with computational contexts and the existence of Web profiles identified with Bluetooth address, an annotation system can consider the execution of an inference process responsible for deriving the social context of a photo. An annotation system can, then, determine who was present at the moment of a photo shot. We propose an inference process that combines ContextPhoto and SWRL rules for inferring this information. This process reads the photo annotation and gets the *Owner* identification. Then, a system can use the FOAF profile of the photo owner as a start point of a search. The tool navigates through the properties "*foaf:Knows*" and "*rdf:seeAlso*" of the owner FOAF profile to get the FOAF profiles of people that the owner knows. All the profiles that have been found are used to instantiate *Person* individuals. After the instantiation of the *Person* and *Owner*, the system can use a rule-based engine in order to infer which acquaintances were present when the photo was shot. Table 2 shows the SWRL rule we have used to infer the presence of an owner's friend. After the end of the inference process, individuals representing the nearby acquaintances are included in the social dimension of ContextPhoto annotation.

Table 2 - SWRL rules for inferring friend's presence

Owner(?owner) ^ Person(?person) ^ SocialContext(?scctxt) ComputationalContext(?compctxt) ^ BTDevice(?btDv) foaf:knows(?person, ?owner)^ hasBTDevice(?person, ?btDv) hasContextElement(?scctxt, ?owner) ^ hasContextElement(?compctxt, ?btDv) → hasContextElement(?scctxt, ?person)

5 PhotoMap

PhotoMap is a mobile and Web location-based system for photo annotation. We use the image annotation process proposed in the Section 2 (see Figure 1) for guiding the PhotoMap design. The three main goals that we aim with PhotoMap are: (1) to offer a mobile application that enables users to take pictures and to group theirs photos in spatiotemporal collections (see Figure 9).; (2) to propose a Web system that organizes the user's photos using some automatically acquired spatial and temporal data; and (3) to improve the users recall of their photos showing inferred spatial, temporal, and social information. PhotoMap is structured according to a client-server model.

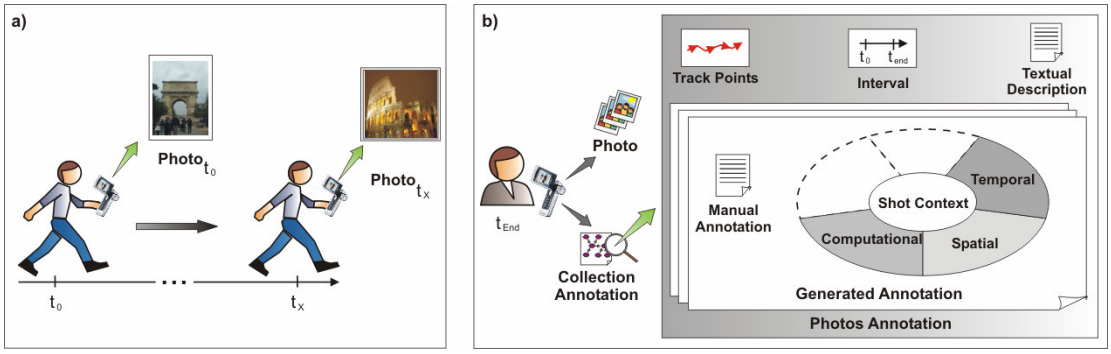

Figure 9 - The way Photomap is used and the generated annotation.

The client application runs on J2ME-enabled devices. We used the XMobile tool (Viana et al. 2008) in order to generate client interfaces adapted to mobile device. The PhotoMap client application allows the users to create photo collections representing events (e.g., tourist visits). The user can give a name and a textual description when she starts a collection. The client application runs a background process that monitors the current physical location of the mobile device (see Figure 9). It accesses the device sensors (e.g., built-in GPS, Bluetooth-enabled GPS receiver) via the Location API (i.e., JSR 179) or via the Bluetooth API (i.e., JSR 82).

The gathered coordinates (latitude, longitude, and elevation) are stored in order to build a list of track points. This list represents the itinerary followed by the user to take the pictures. The acquired track list is associated with the metadata of the current photo collection. Moreover, the PhotoMap client application captures the photo shot context when a user takes a picture with her camera phone. The mobile client gets the geographic position of the device, the date and time information, the bluetooth addresses of the nearby devices, and the configuration properties of the digital camera.

All these metadata are stored for each photo of the collection. After a photo shot, the user can also add manual annotations. The textual description of the collection, its start and end instants, and the track points list are added to the annotation. For each photo, the gathered context and the possible manual annotation are stored in the ontology instantiation. The taken photos and the generated annotation (i.e., ContextPhoto instances) are stored in the device file system. Afterwards, the user uploads her photos and the annotation metadata of the collection to the server application. The user can execute this upload process from her mobile phone directly (e.g., via HTTP) or via a USB connection.

5.1.Interpretation, Inference and Indexation Processes

The PhotoMap server is a J2EE Web-based application. Besides acting as an upload gateway, the server application is in charge of the photo indexation, inference, and interpretation processes. After the transmission of a collection and its metadata, the PhotoMap server reads the annotations associated with each photo and then executes some internal processes which enhance the contextual annotation as described in this section. When a user sends a photo collection to PhotoMap, the annotation contains only information about the computational, spatial and temporal contexts of a photo. Thus, PhotoMap accesses off-the-shelf Web Services in order to augment the context annotation.

First, PhotoMap executes an **interpretation** of the gathered spatial metadata. The interpretation process consists in translating spatial and temporal metadata in a more useful representation. First, the PhotoMap server uses a Web Service to transform GPS data of each photo into physical addresses. The *AddressFinder*[18]

[18] http://ashburnarcweb.esri.com/

Web Service offers a hierarchical description of an address at different levels of precision (i.e. only country and city name, a complete address). The responses of the Web Service are stored in the Datatype properties of *Address* instances. PhotoMap server also uses the GPS data of a photo for acquiring the near Wikipedia documents, since this information is useful for enriching the photo annotation and because, in some cases, the information is directed related to the photo subject. In a concrete way, PhotoMap uses the photo location and the service *Find Nearby Wikipedia Entries* available in GeoNames[11] site for acquiring the three nearest objects. The results (i.e.,Wikipedia entries) are included in the ContextPhoto instance.

The second interpretation phase performs the separation of temporal attributes of the date/time property. The PhotoMap server calculates day of week, month, time of day, and year properties using the instant value. Later, the PhotoMap server gets information about the physical environment at the photo shot time. PhotoMap derives temperature, season, light status, and weather conditions using the GPS data and the date/time annotated by the PhotoMap client. We use the *Weather Underground*[19] Web Service to get weather conditions and temperature information. In addition, PhotoMap uses the *Sunrise and Sunset Times*[20] Web Service to get the light status. The season property is calculated using the date and GPS data.

After the interpretation process, the server application executes the **inference** process in order to derive the social context of the photo (see section 4.5). At the end of the inference process, individuals representing the present acquaintances are associated with the *Shot Context* of the current photo. Hence, the final OWL annotation of a photo contains spatial, temporal, spatial-temporal, computational and social information.

Then, the PhotoMap server executes an **indexation** process in order to optimize browsing and interaction methods. The number of photo collection annotations will quickly increase in our system. To avoid problems of performance linked to sequential searches in these OWL annotations, spatial and temporal indexes are generated for each collection using the PostgreSQL database extended with the PostGIS module. Using the spatial information of the collection track points list, PhotoMap calculates a minimum bounding box that includes the entire itinerary. The generated polygon is stored as a GiST index and it is associated with a string indicating the file path of the OWL annotation. The collection start and end times are used to create temporal indexes which are also associated with the collection file path.

5.2 Browsing and Querying Photos

PhotoMap offers graphical interfaces for navigation and query over the users' photo collections. We have designed the PhotoMap portal using map-based interfaces for browsing photos purposes. Usability studies (Toyama et al. 2003) show that map interfaces present more interactivity advantages than browsing photos only with location hierarchical links (e.g.; Europe > France > Grenoble). Furthermore, with a map-based interface, we can easily represent the itineraries followed by users when taking their photos. Besides the map-based visualization, users can look into the inferred and manual annotations of their photos and exploit them for retrieval purposes.

[19] http://www.weatherunderground.com/
[20] http://www.earthtools.org/

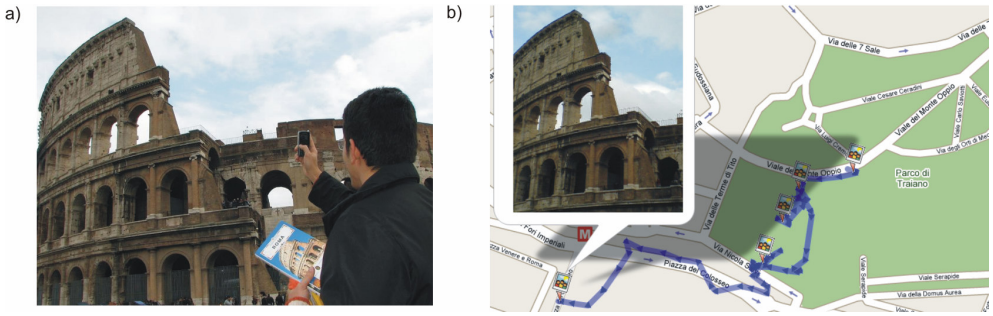

Figure 10 - PhotoMap in a real situation

Figure 10 illustrates an example of a photo collection taken during a tourist tour in Rome (77 GPS data and 5 photos). We have used a Sony Ericsson K750i phone to snap the photos near the Coliseum. The mobile phone had our annotation application installed and was connected to a Bluetooth-enabled GPS. At the beginning of the tour, the user has activated the client application for taking his photos and for tracking his followed path. Figure 10a shows the user taking a photo with the K750i phone. In the Figure 10b, the itinerary followed by the user and the placemarks for each taken photo can be seen on a map view of Rome.

Figure 11 shows a screen shot of the PhotoMap Web site. The rectangular region on the left side of the Web Site is the *menu-search view*. This window shows PhotoMap main functionalities, the social network of the user, and a keyword search engine for her photo collections. On the right side, from top to bottom, are shown the *event-collection view*, the *spatial view*, and the *temporal query window*. When a user enters the Web site, PhotoMap uses the temporal index to extract the ten latest collections. PhotoMap then shows, in the *event-collection view,* thumbnails of a symbol photo of each collection (i.e., informed by the user when she synchronizes the collection with PhotoMap) and the collection names annotated by the user. The latest collection is selected and the itinerary followed by the user is displayed in the *spatial view* (i.e., in the figure, the roman forum collection is selected). Placemarks are inserted in the map for each photo, and the user can click to view the photo and the generated annotation. Figure 11 illustrates the photo visualization when the user clicks on a placemark.

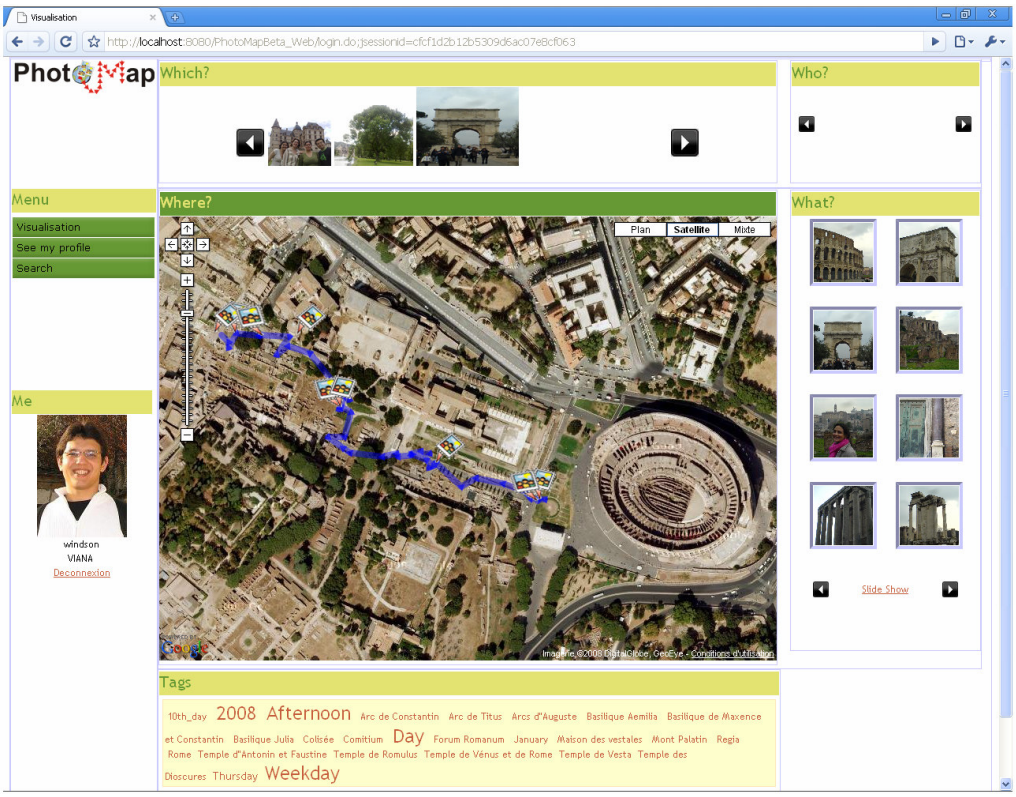

Figure 11 - PhotoMap Web site.

Using this web application, the user can read the captured and inferred annotations. Figure 12a illustrates the visualization of a photo collection over the satellite image of Rome. In the "Captured" panel (see Figure 12b) are described the initial context data annotation: GPS data, date/time, and a physical address of a nearby Bluetooth device. The "Inferred" panel (see Figure 12c) shows the information derived by the server application: address, temperature, and season. The panel shows the FOAF information about the user's friends present in the photo shot (i.e., name and personal photo). Moreover, the system shows the three nearby Wikipedia entries. In this case, coincidently, the first Wikipedia entry corresponds to the photo subject.

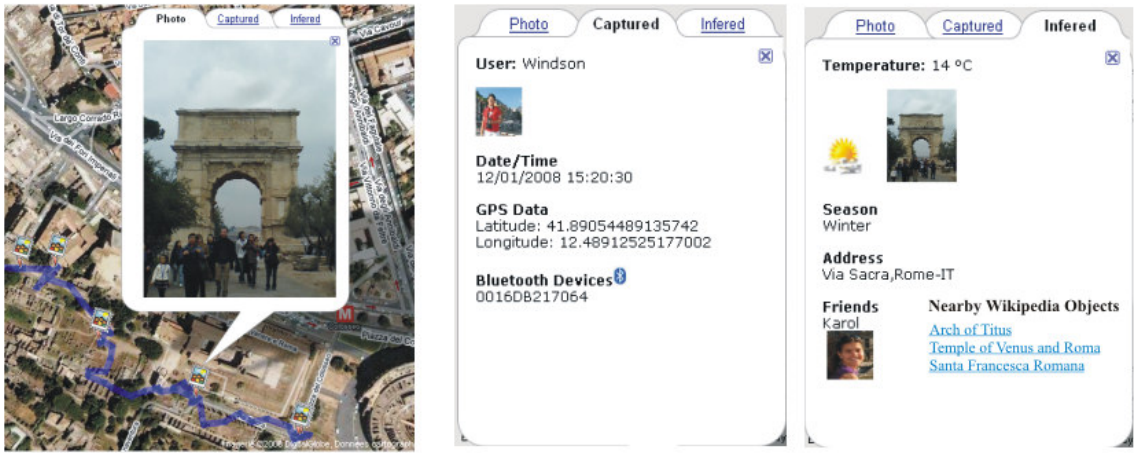

Figure 12 - An illustration of PhotoMap captured and inferred annotations.

6 Related Works

Some research works and off-the-shelf tools have explored both spatial and temporal metadata of an image in order to organise photo collections and to generate annotation metadata. For example, **PhotoCompas** (Naaman et al. 2004) use Web Services to derive information about the location and physical environment (e.g., weather conditions, temperature). PhotoCompas calculates image clusters using time and spatial data in order to represent event recall clues. All the generated information is used for browsing the photos of the collection. **WWMX** (Toyama et al. 2003) is a Microsoft project with the goal of browsing large databases of images using maps. WWMX offers different possibilities of displaying images in a map and uses GIS databases in order to index the location information of the photos.

We can also mention recent works, such as **PhotoCopain** (Tuffield et al., 2006). PhotoCopain is a desktop system for annotating images with a RDF representation of the EXIF standard. The proposition combines EXIF metadata (and if it is available, the GPS photo coordinates), user calendar data, and a geographical gazetteer service for inferring context annotation information (i.e., the city where the photo was taken). PhotoCopain integrates the context information with imaging processing methods in order to classify the images in categories (i.e., portrait, landscape). The generated annotation is represented in the ICMP format that allows user to use the AKtiveMedia tool (Chakravarthy et al., 2006) for visualising the image metadata and for validating them.

The approaches mentioned above are based on the investigation of the EXIF data available in image headers. It represents a primary advantage since there are a huge number of images pre-annotated with this metadata format on the Web. However, these approaches limit the set of useful contextual metadata. For example, most of the Web images do not have spatial metadata. Some works, like ours, addresses this issue by using mobile devices as source of context metadata in order to increase the number and the quality of contextual information. For instance, the **MMM Image Gallery** (Sarvas et al. 2004) proposes a mobile and location-based application for semi-automatic photos annotation using the Nokia 3650 devices. At the photo shot time, the application captures the GSM cell ID, the user identity and date/time of the photo. The image is sent to the server associated with its annotation. The server combines the low level properties of the image with location information in order to derive information about its content (e.g., is the Eiffel Tower the photo subject?). After that, the user uses a mobile XHTML browser in order to read and validate the generated annotation. The imprecision of the GSM cell ID and the slow response time of the server during the image upload and validation processes were identified as the major usability problems of MMM. **Zonetag** (Ames 2001) is a Yahoo! research prototype allowing users to upload their photos to Flickr from their mobile phones directly. ZoneTag leverages location information from camera phones (GSM cell ID and GPS coordinates) and suggests tags to be added to the photo. ZoneTag derives its suggestions from the tags used by the community for the same location, and from the tags assigned to the user photos.

In (Monaghan et al**,** 2006), the authors show the design of a mobile and Web system that combines Bluetooth addresses acquisition, FOAF profiling and face detection tools in order to identify people in a photo. The authors suggest that the generated annotation is represented in a RDF file and embedded in the header of the image. They propose to use the Bluetooth addresses of the nearby devices as key inputs to a FOAF Query Service that is able to deliver the requested FOAF profiles. The work in (Monaghan et al**,** 2006), has inspired the inference process we use in order to derive the social context of a photo shot. However, part of the original approach seems to us not easily feasible. First, this proposition does not take into account the distributed characteristics of FOAF profiles. Moreover, designing a service that indexes all the FOAF profiles of the Web is not realistic.

The work we present in this paper is an evolution of the location-based multimedia applications mentioned above. We have proposed not only a location-based annotation tool, but also a semi-automatic annotation

process, and photo annotation ontology. We have integrated and extended some of the contextual data proposed in these approaches. In addition, we have offered a formal way for representing them and inferring from them. For instance, we have used the approach of (Monaghan et al**,** 2006) for deriving friends' presence by combining bluetooth addresses and FOAF profiles. Nevertheless, we have modified the proposed method in order to allow its feasibility. In spite of use a questionable FOAF index service, we have used the FOAF profile of the photo owner as the starting point of a search from which we access to profiles that are useful to annotate the photo (i.e., the profiles of friends, of friends of friends etc.). Furthermore, the system we have developed to validate our approach is both a mobile application for semi-automatic image annotation and a Web system for the management of personal image collections. Besides the ability for acquiring and deriving context automatically, the PhotoMap mobile client allows user to create their event collections and to insert textual annotation to them. One can use the PhotoMap Web site to view her photo collections (and, also the followed itinerary) and to exploit the inferred annotation for organization and retrieval purposes.

7 Usability Tests and Lessons learned

An enriched image annotation derived from few spatial and temporal data appears as a promising tool for organizing personal photo collections and for improving the quality of the image searching engines. Our initial experiments showed the users' surprise when they see their mobile collections structured in the Web tool. The rich annotation describing friend's presence information, itinerary followed and near objects offers to users new forms to share, to organise and to browsing their photos. In addition, users have enjoyed the possibility to see information about objects, places and people that they did not necessarily know at the shot time of the picture.

Users have created collections in some Europe cities (e.g., Rome, Lyon, Grenoble …). They have used two configurations: one version of Photomap installed in a Sony Ericsson K750i connected by bluetooth with a GPS receiver and another version installed in a Nokia N95 which contains a built-in GPS. The users have taken their collections using a **non-network based version** of PhotoMap. In this version, the mobile application accesses only the context information available in the device (e.g., GPS coordinates) and the inference and interpretation processes are executed in a postponed situation, when the users synchronize their mobile devices with our server application using a desktop Java application.

The main usability issue pointed out by users is linked to the time of GPS initiation. The application take 2-4 minutes with the K750i and 1-2 minutes with the N95 to get useful coordinates in a Europe city. Users have reported that this delay interferes their current activities (e.g., walk), since one should stay static and look frequently at the mobile application until it achieves the acquisition of the first geographical coordinates. This boring process can be repeated if users unintentionally quit the PhotoMap application during a photo collection. This initiation time will probably be reduced in few years with new approaches of location acquisition (e.g., A-GPS, Galileo) and, their integration in the mobile devices.

Another usability problem is the workflow we have proposed for photo shooting process (i.e., take a photo, view initial annotation, complete annotation, save photo). Most of the users preferred a workflow closer to the user interaction available in tradition digital camera applications. As it was observed in others mobile annotation experiences (Matellanes et al. 2006) (Sarvas et al. 2004), users want to see or to complete image annotation in mobile situation only for few and favourite photos. We will modify the PhotoMap interface in the next mobile application version for better fits users' preferences.

Tests concerning captured data also have produced some drawbacks. Some collections have presented unrealistic users' itineraries. For instance, one collection generated in the downtown of Lyon city (France) has traced a line over the river that was very distant of the bridge the user has walked through. In this same collection, some photos were placed 100-200 meters distant from their original point. These problems concern

the GPS precision errors and they are also related to the captured model we have used to keep trace of the users' itinerary. In order to avoid store and execution memory surcharge, we have decided to store part of the GPS coordinates by using a pre-defined acquisition frequency. However, some noise is generated when one stops (e.g., for taking a photo) and some "empty" zones appear when users move quickly (i.e., take a bus for example). Methods for refining the trajectory data and decrease the noise will be studied and adapted to mobile devices capabilities. In addition, the acquisition frequency will be automatically calculated using the speed attribute available in GPS data. These modifications will allow a more realistic and useful track of the user's itinerary.

Although users understand and appreciate that enriched annotation can be used for multimedia content recommendation (e.g., the example we have mentioned in the introduction section), they argue that the historic of the itinerary and the acquaintance's presence information could be both dangerous and embarrassing in some situations. Hence, a privacy-aware model for the contextual information (e.g., to provide privacy for the user location information) needs to be designed and integrated in our annotation process. The security policy enforcement should respect user's privacy preferences (i.e., the user's security policies) and try to avoid the shrinkage of the quality of enhanced contextual metadata.

8 Conclusion and Future Work

The availability of multimedia content and context metadata represents the first steps of the Web transformation from a Web of documents to a Web of knowledge. In this paper, we have proposed a semi-automatic image annotation process based on the integration and extension of existing and well-understood technologies and services. The information represented in our ontology combined with the inference rules and interpretation process presented in the section 5.1 illustrates how the initial set of spatial data can be increased towards a complete description of the photo shot situation, describing this way the photo context. Additionally, PhotoMap validates the proposed annotation approach offering interfaces for navigation and visualisation over the user's photo collections and annotations.

In the future, we will refine the PhotoMap system and our context annotation process to avoid the problem of divulgate user private information. We will use an extension of the FRAMESEC (Bringel Filho et al., 2005) in order to provide user's privacy protection. Also, we will investigate ways to recommend PhotoMap multimedia content and tourist trajectories to a mobile user by using similarity measures between her current context and the contexts associated with these multimedia documents.

References

Ames, M. and Naaman, M., 2007. Why We Tag: Motivations for Annotation in Mobile and Online Media. *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems.* CHI '07. ACM, New York, NY, 971-980. DOI= http://doi.acm.org/10.1145/1240624.1240772 ACM Press, New York, NY, USA, 971-980.

Athanasiadis, Th., Tzouvaras, V., Petridis, K., Precioso, F., Avrithis, Y. and Kompatsiaris, Y., 2005. Using a Multimedia Ontology Infrastructure for Semantic Annotation of Multimedia Content. *In Proc. of 5th International Workshop on Knowledge Markup and Semantic Annotation (SemAnnot '05).* Springer, Galway, Ireland, November 2005.

Boutella, M. and Luob, J., 2005. Beyond pixels: Exploiting camera metadata for photo classification. *Pattern Recognition*, (38), No. 6, June 2005, pp. 935-946.

Bringel Filho, J., Viana, W., Braga, R. and Andrade, R. M. C., 2005. FRAMESEC: A Framework for the Application Development with End-to-End Security Provision in the Mobile Computing Environment. *AICT-SAPIR-ELETE '05: Proceedings of the Advanced Industrial Conference on Telecommunications/Service Assurance with Partial and Intermittent Resources Conference/E-Learning on Telecommunications Workshop*, IEEE Computer Society, Washington, DC, USA, 72-77.

Christensen, C.B., 2001. Place and Experience: a Philosophical Topography. *Mind,* Oxford University Press, Volume 110, Number 439, 1 July 2001, pp. 789-792(4).

Chakravarthy, A., Ciravegna, F., Lanfranchi.,V. 2006. AKTiveMedia: Cross-media document annotation and enrichment. *In Proceedings of the Fifteenth International Semantic Web Conference (ISWC2006).*

Dey, A.K. and Abowd, G.D., 2000. Towards a Better Understanding of Context and Context-Awareness. *HUC '99: Proceedings of the 1st international symposium on Handheld and Ubiquitous Computing*, Springer-Verlag, London, UK, 304-307.

Gruber, T.,2008. Collective knowledge systems: Where the Social Web meets the Semantic Web. Web Semantics: Science, Services and Agents on the World Wide Web . Volume 6, Issue 1, , Semantic Web and Web 2.0, February 2008, Pages 4-13.Doi:10.1016/j.websem.2007.11.011

Grütter, R., Bauer-Messmer, B. 2007. Towards Spatial Reasoning in the Semantic Web: A Hybrid Knowledge Representation System Architecture. *The European Information Society. Leading the Way with Geo-Information. Part* 8, LNGC, 349-364, December 2007.

Hollink, L., Nguyen, G., Schreiber, G., Wielemaker, J., Wielinga, B. and Worring, M., 2004. Adding Spatial Semantics to Image Annotations. *Proc. of 4th International Workshop on Knowledge Markup and Semantic Annotation*.

Kirsch-Pinheiro, M., Villanova-Oliver, M., Gensel, J., and Martin, H., 2005. Context-aware filtering for collaborative web systems: adapting the awareness information to the user's context. *Proceedings of the ACM Symposium on Applied Computing (SAC),* Santa Fe, New Mexico, USA, March 13-17, 1668-1673.

Lux M., Klieber W., Granitzer M., 2004. Caliph & Emir: Semantics in Multimedia Retrieval and Annotation, *Proceedings of 19th International CODATA Conference*: The Information Society: New Horizons for Science, Berlin, Allemane, 2004

Matellanes, A., Evans, A. and Erdal, B., 2006. Creating an application for automatic annotations of images and video. *Proc. of 1st International Workshop on Semantic Web Annotations for Multimedia (SWAMM)*, Edinburgh, Scotland, May. Available at: http://www.image.ntua.gr/swamm2006/resources/paper20.pdf

Monaghan, F. and O'Sullivan, D., 2006. Automating Photo Annotation using Services and Ontologies, *MDM'06: Proceedings of the 7th International Conference on Mobile Data Management*, IEEE Computer Society, Washington, DC, USA, 79-83 .10-12 May 2006. Doi: 10.1109/MDM.2006.39

Naaman, M., Harada, S., Wang, Q., Garcia-Molina, H. and Paepcke, A., 2004. Context data in geo-referenced digital photo collections. *MULTIMEDIA '04: Proceedings of the 12th annual ACM international conference on Multimedia*, New York, NY, USA, 196-203.

Pigeau, A. and Gelgon, M., 2004. Organizing a personal image collection with statistical model-based ICL clustering on spatio-temporal camera phone meta-data. *Journal of Visual Communication and Image Representation*, 15(3), 425-445.

Petit,M., Ray C. and Claramunt C., 2006. A Contextual Approach for the Development of GIS: Application to Maritime Navigation. *Proc. of Web and Wireless Geographical Information Systems, 6th International Symposium, W2GIS 2006,* Hong Kong, China, December *4-5,* Springer, Lecture Notes in Computer Science, 4295, 158-169.

Reitsma, F. and Hiramatsu, K., 2006. Exploring GeoMarkup on the Semantic Web. *9th AGILE International Conference on Geographic Information Science: shaping the future of Geographic Information Science in Europe*, 20-22.

Schreiber, A. Th., Dubbeldam, B., Wielemaker, J. and Wielinga, B. J., 2001. Ontology-based photo annotation. *IEEE Intelligent Systems*, 16(3), Piscataway, NJ, USA, 66-74.

Sarvas, R., Herrarte, E., Wilhelm, A. and Davis, M., 2004. Metadata creation system for mobile images. *MobiSys '04: Proceedings of the 2nd international conference on Mobile systems, applications, and services*, ACM, Boston, MA, USA, 36-48.

Shadbolt, N., Berners-Lee, T. and Hall, W., 2006. The Semantic Web Revisited. *IEEE Intelligent Systems, IEEE Educational Activities Department,* Piscataway, NJ, USA, 21 (3), 96-101.

Toyama, K., Logan, R. and Roseway, A, 2003. Geographic location tags on digital images. *MULTIMEDIA '03: Proceedings of the eleventh ACM international conference on Multimedia*, Berkeley, CA, USA,156-166.

Yamaba, T., Takagi, A. and Nakajima, T., 2005. Citron: A context information acquisition framework for personal devices, *Proc. of 11th International Conference on Embedded and real-Time Computing Systems and Applications*, 489-495.

Tuffield, M., Harris, S., Dupplaw, D. P., Chakravarthy, A., Brewster, C., Gibbins, N., O'Hara, K., Ciravegna, F., Sleeman, D., Wilks, Y. and Shadbolt, N. R., 2006. Image annotation with Photocopain. In: First International Workshop on Semantic Web Annotations for Multimedia (SWAMM 2006) at WWW2006, May 2006, Edinburgh, United Kingdom.

Viana, W. ; Castro, R. M. C. ; Machado, J.; Filho, B. ; Magalhaes, K. ; Giovano, C. . Mobis: A Solution for the development of secure Applications for mobile device. In: ICT: International Conference on Telecommunications, 11th, Fortaleza-CE, Brazil, 2004. Lecture Notes in Computer Science, v. 3124, p. 1015-1022, 2004. DOI:10.1007/b99377

Viana, W., Andrade, R. M. C., XMobile: A MB-UID environment for semi-automatic generation of adaptive applications for mobile devices. Journal of Systems and Software 81(3): 382-394. 2008. DOI:10.1016/j.jss.2007.04.045