

# Ontology-Based Image Retrieval

Eero Hyvönen<sup>1,2</sup>, Avril Styrman<sup>1</sup>, and Samppa Saarela<sup>2,1</sup>

<sup>1</sup> University of Helsinki, Department of Computer Science  
firstname.lastname@cs.helsinki.fi,  
<http://www.cs.helsinki.fi/group/seco/>

<sup>2</sup> Helsinki Institute for Information Technology (HIIT)

**Abstract.** The binary form of an image does not tell what the image is about. It is possible to retrieve images from a database using pattern matching techniques, but usually textual descriptions attached to the images are used. Semantic web ontology and metadata languages provide a new way to annotating and retrieving images. This paper considers the situation when a user is faced with an image repository whose content is complicated and semantically unknown to some extent. We show how ontologies can then be of help to the user in formulating the information need, the query, and the answers. As a proof of the concept, we have implemented a demonstrational photo exhibition using the promotion image database of the Helsinki University Museum based on semantic web technologies. In this system, images are annotated according to ontologies and the same conceptualization is offered to the user to facilitate focused image retrieval using the right terminology. When generating answers to the queries, the ontology combined with the image data also facilitates, e.g., recommendation of semantically related images to the user.

## 1 The problem of semantic image retrieval

Images are a major source of content on the WWW. The amount of image information is rapidly raising due to digital cameras and mobile telephones equipped with such devices. This paper concerns the problem when an end-user is faced with a repository of images whose content is complicated and partly unknown to the user. Such situations are recurrent, e.g., when using public image databases on the web.

We approach this general problem through a case study. The Helsinki University Museum<sup>3</sup> will open a permanent exhibition in the autumn 2003 at Helsinki city centre on the Senates Square. A central goal of the museum is to spread knowledge about university traditions to the large audience. One living tradition of the faculties of the University of Helsinki is the promotion ceremonies through which the students get their master's and doctoral degrees and the faculties grant honorary doctoral degrees to distinguished scientists and persons outside the university. The ceremonies consist of many occasions and last for

---

<sup>3</sup> <http://www.helsinki.fi/museo/>

several days. The museum database contains over 600 photographs about the ceremony events and documents, ranging from the 17th to 21th century, and more images are acquired after every promotion. The contents of this image repository would provide the audience with an interesting view into the life of the university .

There are two basic approaches to image retrieval: 1) content-based image retrieval (CBIR) and 2) metadata based image retrieval. In CBIR [4] the images are retrieved without using external metadata describing their content. At the lowest level, features such as color, texture, shape, and spatial location are used. At a higher conceptual level, images with an object of a given type or a given individual are searched (e.g., images with a person in the front or images of the Eiffel tower). At the highest level, retrieval of named events or activities or pictures with emotional or symbolic significance are retrieved (e.g., pictures about "tennis" or pictures depicting "atonement"). An example of the CBIR approach on the web is the PicSOM system [10]. In the metadata-based approach image retrieval is based on textual descriptions about pictures. In practice, this approach is usually employed in image retrieval due to the great challenges of the CBIR approach when dealing with conceptually higher levels of content<sup>4</sup>.

A typical way to publish an image data repository online is to create a keyword-based query [1,2] interface to an image database. Here the user may select filtering values or apply keywords to the different database fields, such as the "creator", "time", or to the content descriptions including classifications and free text documentation. More complex queries can be formulated, e.g., by using Boolean logic. Examples of museum systems on the web using this approach include the Kyoto National Museum search facility<sup>5</sup>, Australian Museums Online<sup>6</sup>, and Artefacts Canada<sup>7</sup>.

Keyword-based search methods suffer from several general limitations [5, 8]: A keyword in a document does not necessarily mean that the document is relevant, and relevant documents may not contain the explicit word. Synonyms lower recall rate, homonyms lower precision rate, and semantic relations such as hyponymy, meronymy, antonymy [?] are not exploited.

Keyword-based search is useful especially to a user who knows what keywords are used to index the images and therefore can easily formulate queries. This approach is problematic, however, when the user does not have a clear goal in mind, does not know what there is in the database, and what kind of semantic concepts are involved in the domain. The university promotion ceremonies case discussed in the paper is an example of such a semantically complicated domain. Using the keyword-based approach would lead to the following problems:

---

<sup>4</sup> The term "content-based" in CBIR is unfortunate and confusing, since the textual metadata-based approach deals with explicit representations of content.

<sup>5</sup> <http://www.kyohaku.go.jp>

<sup>6</sup> <http://amonline.net.au>

<sup>7</sup> <http://www.chin.gc.ca>

**Formulating the information need** The user does not necessarily know what question to ask. One may only have general interest in the topic. How to help the user in focusing the interest within the database contents?

**Formulating query** The user cannot necessarily figure out what keywords to use in formulating the search corresponding to her information need. How to help the user in formulating queries?

**Formulating the answer** Generating image hit lists for keywords would probably miss a most interesting aspect of the repository: the images are related to each other in many interesting ways. In our case, the ceremonial occasions follow certain patterns in place and time and the people and surroundings depicted in the images reoccur in different events. These semantical structures should somehow be exposed from the data to the audience. The goal of an ordinary museum visitor is often something quite different from trying to find certain images. The user wants to learn about the past and experience it with the help the images.

We argue that semantic web technologies provide a promising new approach to these problems. In the following, semantic ontology-based annotation and retrieval of images is first discussed. After this the ontology used in our demonstrational system is presented, an annotation example is given, and the an ontology-based user interface to the image repository is illustrated. In conclusion, the contributions of this work are summarized.

## 2 Semantic image annotation and retrieval

The problem of creating metadata for images has been of vital importance to art and historical museums when cataloging collection items and storing them in a digital form. Following approaches are commonly used in annotating images:

**Keywords** Controlled vocabularies are used to describe the images in order to ease the retrieval. In Finland, for example, the Finnish web thesaurus YSA<sup>8</sup> is used for the task augmented with museum- and domain specific keyword lists.

**Classifications** There are large classification systems, such as the ICONCLASS<sup>9</sup> [20] and the Art and Architecture Thesaurus [14], that classify different aspects of life into hierarchical categories. An image is annotated by a set of categories that describe it. For example, if an image of a seal depicting a castle could be related to classes "seals" and "castles". The classes form a hierarchy and are associated with corresponding keywords. The hierarchy enriches the annotations. For example, since castles are a subclass of "buildings", keyword "building" is relevant when searching images with a castle.

**Free text descriptions** Free text descriptions of the objects in the images are used. The information retrieval system indexes the text for keyword-based search.

---

<sup>8</sup> <http://www.vesa.lib.helsinki.fi>

<sup>9</sup> <http://www.iconclass.nl>

Semantic web ontology techniques [5] and metadata languages [9] contribute to this tradition by providing means for defining class terminologies with well-defined semantics and a flexible data model for representing metadata descriptions. One possible step to take is to use RDF Schema [3] for defining hierarchical ontology classes and RDF [11] for annotating image metadata according to the ontology. The ontology together with the image metadata forms an RDF graph, a knowledge base, which can facilitate new semantic information retrieval services.

In our case study, we used this approach. Our idea is to first make ontological models of the concepts involved in the image repository. The ontologies form the core of the system and are used for three purposes:

**Annotation terminology** The ontological model provides the terminology and concepts by which metadata of the images is expressed.

**View-based search** The ontologies of the model, such as Events, Persons, and Places provide different views into the promotion concepts. They can hence be used by the user to focus the information need and to formulate the queries.

**Semantic browsing** After finding a focus of interest, an image, the semantic ontology model together with image instance data can be used in finding out relations between the selected image and other images in the repository. Such images are not necessarily included in the answer set of the query. For example, images where the same person occurs but in a different event of the same promotion may be of interest and be recommended to the user, even if such images do not match the query.

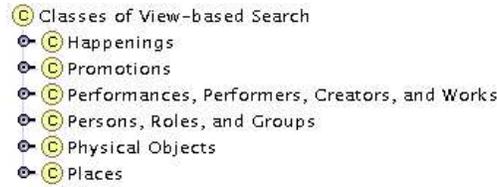
In the following, the construction of a comprehensive ontology for promotion concepts is discussed. It is then shown, how such an ontology facilitates semantic-based information retrieval of images as envisioned above.

### 3 The promotion ontology

The promotion ontology describes the promotional events of the University of Helsinki and it's predecessors, the Empirical Alexander's University and the Royal Academy of Turku. The top-level ontological categories are depicted in figure 1.<sup>10</sup> The classes of the ontology represent people of different roles (**Persons, roles, and groups**), events and happenings (**Happenings**) that take place in different locations (**Places**), physical objects (**Physical Objects**), speeches, dances, and other performances (**Performances, Performers, Creators, and Works**), and a list of all promotions from 17th century until 2001 (**Promotions**). The main goal of the ontologization process was to create an ontology suitable for the photograph exhibition and to offer the programmers the basis to implement the exhibition, either on the web or as an internal information kiosk application<sup>11</sup>.

<sup>10</sup> In this paper we use English names for classes etc., but the actual implementation is in Finnish.

<sup>11</sup> The publication of the photographs representing, e.g., living people on the public WWW has legal consequences that have to be considered first.



**Fig. 1.** Top-level classes of Promotion-ontology.

The stable and unchanging things of the subject domain, i.e., continuants [19], are presented with classes in the ontology. The changing things, occurrents, are presented with instances. For example, in figure 2 the Cathedral of Helsinki has its own class. On the other hand, buildings that are not regularly used in promotions do not have a subclass of their own, but are instances of the general class **Buildings**.

The instances of the ontology have literal-valued properties, such as *name of the person*. These properties are typically used to provide a human-readable presentation of the instance to the user. Each instance, e.g., a particular person, is related to the set of promotions in which the instance occurs. In this way, for example, the persons performing in a particular promotion are easily found.

The top-level classes of the ontology are briefly discussed in the following.

**Promotions** Class **Promotions** has several properties to describe the features of every single promotion, such as the central person roles. All the instances of type **Promotions** hold at least 1) the date of the promotion ceremony, 2) the university under which the promotion was held, and 3) the faculty that arranged the promotion.

**Persons, Roles, and Groups** The same person may appear in many roles in several promotions. For example, a person, i.e., an instance of some (sub)class of **Persons**, can be a master-promovend in one promotion and a doctor-promovend in another one. To represent this, a person's unique person-instance is related to (possibly) many instances of the different **Roles**. The fundamental properties of instances of **Person** classes are the name, other name(s), and the promotion(s) in which the person participated to in different roles. The properties of the sub-classes of class **Roles** include the person(s) in the role, and the promotion(s) in which the person participated in the role. The **Groups** classes represent collections of persons and roles in a group. Groups have a string-valued label and instance-valued property "person-roles of this group". For example, a group of master-promovends has a label "Master promovends of promotion  $x$ ", and instances of the particular masters as the values of the property "person-roles of this group".

**Places** Class **Places** is divided into four subcategories (cf figure 2), **Squares and Parks**, **Buildings**, **Islands and Harbours**, and **Streets and Roads**. These classes define two literal-valued properties: name and address of the place. The name of the place describes the place and is used as the label of the instances.



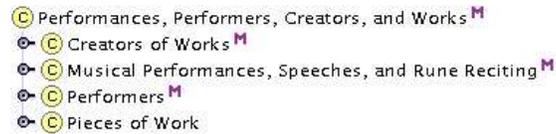
**Fig. 2.** Places and subclasses.

There is a regularly used set of **Squares and Parks** with direct relations to promotional ceremonies as well as a few **Buildings** that have become traditional sites for promotional happenings. These continuants have their own classes. **Islands and harbours** and **Streets and Roads** don't have further subclasses.



**Fig. 3.** Physical Objects and subclasses.

**Physical Objects** Class **Physical Objects** (cf. figure 3) has several subclasses, namely **Headgear**, **Sculptures**, **Vehicles**, **Flowers**, **Flags, Marks, and Badges**, **Printed Matters**, and **Other Things**. These classes have more subclasses that specify the physical objects involved in promotions in more detail. The properties of all physical objects are the name of the object, that is also used as the label of the instances, and the manufacturers of the object. In addition, the subclasses have their own additional properties, such as the literal-valued "language of a document" and the instance-valued "physical situatedness". **Printed Matters**, such as rune books may consists of several instances of **Pieces of Work**.



**Fig. 4.** Classes about audial performances.

**Performances** Class **Performances, Performers, Creators, and Works** (cf. figure 4) has subclasses **Musical Performances, Speeches, and Rune Reciting**, **Performers**, **Pieces of Work**, and **Creators of Works**.



**Fig. 5.** Happenings and subclasses.

**Happenings** Class **Happenings** (cf. figure 5) ties all the other classes semantically together. Every happening has properties that describe

- the place of the happening (an instance of **Places**).
- the people who participated in the happening in different roles (instances of **Persons and Roles**),
- the performances of the happening (instances of **Performances, Performers, Creators, and Works**),
- the physical objects used in the happening (instances of **Physical Objects**),
- the **name of happening** (a literal value), and
- the **date of happening** (a literal value).

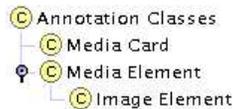
The class **Happenings** is divided into subclasses that collect happenings of similar nature together. The class **Sequence of Happenings** represents all the happenings in a hierarchy based on the chronological ordering of the happenings. The sequence of happenings is specified with a class-valued property **previous**. The sequence of two successive happenings, say, that **A** is followed by **B**, is specified by giving class **B** the property **previous** with the value **A**.

**Ontology construction** There are several partly conflicting goals to keep in mind when designing the ontology. The ontology not only should be semantically motivated, but also easy to construct and maintain to the ontologist. At the same time, the annotation work based on it should be simple to the annotator. Furthermore, the ontology and the annotated instance data should be in a form that is easy to use by the application programmer and efficient to run by the exhibition software.

In our work, two major difficulties were encountered during the annotation and implementation process:

1. Annotation process posed new demands to the ontology, which lead to changes in the ontology after many annotations were already done. How to manage such changes so that the annotator wouldn't have to redo the annotation work?
2. Application programmers pose new demands to the ontology and to the annotations in order to satisfy the demands of the end-user interface. As a result, changes in both the ontology and the annotations were needed.

## 4 Annotation of the images



**Fig. 6.** Classes used in annotating the photographs.

We used the following annotation scheme: every image is associated with a set of instances of the promotion ontology. They occur in the image and hence characterize its content. This is basically how annotations are done when using, e.g., the ICONCLASS system [20]. The linkage between the image and its content is based on the classes of figure 6. Class **Image Element** defines information about the photographs: the name of the photographer, textual description about the subject of the photograph, a reference to the actual image file, and a reference to an instance of the class **Media Card** that has as one of its properties, the list of instances of the promotion ontology.

The ontology was created with the Protégé-2000 ontology editor<sup>12</sup>[7], using RDF Schema as the output language. Protégé was also used for annotating the photograph.

For example, the image of 7 is annotated in the following way:

---

<sup>12</sup> <http://protege.stanford.edu>



**Fig. 7.** Honorary doctor Linus Torvalds on a procession to divine service at the entrance of the Cathedral of Helsinki June 2, 2000.

<p><b>Description</b></p> <p>Textual description about image.jpg</p> <hr/> <p><b>ImageFile</b></p> <p>image.jpg</p> <hr/> <p><b>Photographer</b></p> <p>Mr. Photographer</p> <hr/> <p><b>Media Card</b>    V C + -</p> <p>mediacard</p>	<p><b>MediaCard ID</b></p> <p>mediacard</p> <hr/> <p><b>IncludedInstances</b>    V C + -</p> <ul style="list-style-type: none"> <li>⊕ Doctor's sword</li> <li>⊕ Group of conferred doctors</li> <li>⊕ Procession to promotion service</li> <li>⊕ Doctor's hat</li> <li>⊕ Audience</li> <li>⊕ Torvalds Linus</li> <li>⊕ Entrance of Cathedral of Helsinki</li> </ul>
---	---

**Fig. 8.** Instances of annotation classes of figure 6. The left side depicts filled fields of an instance of **Image Element**, and the right side depicts an annotated instance of **Media Card**, that was used to annotate figure 7.

- Step 1:** The annotator takes the photograph and creates an empty instance of the class **Image Element**.
- Step 2:** The annotator fills in the empty fields of the **Image Element** instance in figure 8, and creates an instance of class **Media Card**.
- Step 3:** To include the needed metadata within the new **Media Card** instance, the annotator browses the ontology, starting from the top-level classes of figure 1. If a new instance is needed, e.g., if this is the first photo to be annotated where the **Person** Linus Torvalds is present, the annotator has to create one. Once the instance is created, the annotator can use it again to annotate other images about Linus Torvalds. The image of figure 7 is well-annotated with several instances as seen in figure 8. However, one could still add a few instances to it, such as the decoration in the upper left corner, and the diploma of the person carrying the decoration. The choice depends on how detailed semantics are needed and on the annotators choices.

The photos come from the image database of the Helsinki University Museum. This database was transformed into a repository of images and RDF-instance data and annotated further according to the ontology. At this phase the instance RDF metadata was checked and edited; the original information of the database was not sufficient alone and much of the data content was written in the free text description fields.

The ontologization annotation process lasted several months and included examining a wide range of historical materials. Several people took part of the process at different stages, including the domain experts and clients of the museum and programmers of the exhibition software. The full version of the ontology has 575 classes and 279 properties (slots). Currently, the knowledge base contains annotations of 619 photographs and has 3565 instances of ontology classes. The first demonstrational version of the exhibition discussed in this paper uses a much simpler ontology and annotation data of about 100 photos.

## 5 Semantic image retrieval

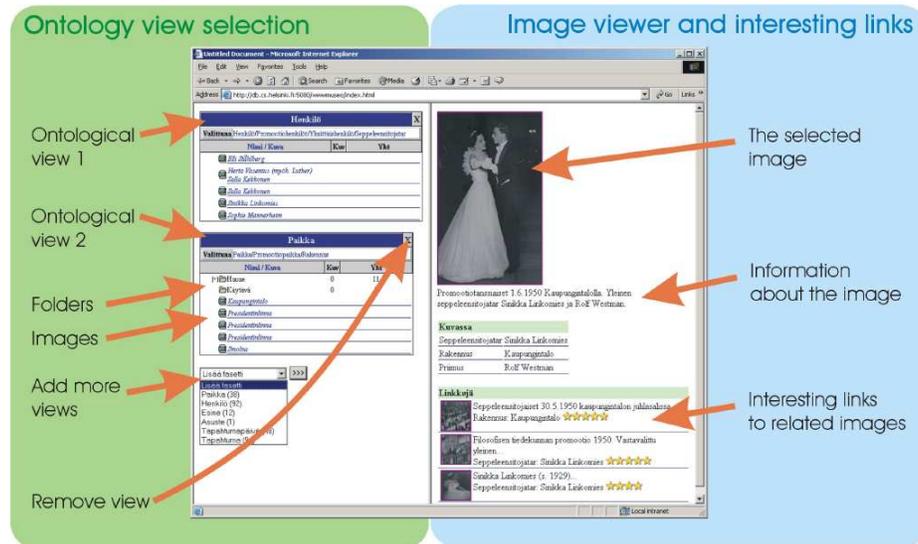


Fig. 9. User interface to the image server.

Based on the ontology, a web server was implemented to support semantic image retrieval. Figure 9 illustrates its appearance on an ordinary web browser. The system provides the user with the following semantics-based facilities.

**View-based filtering** On the left, the user can open ontologies for filtering photographs of interest. In the figure, the ontologies for Persons (Henkilö) and Places (Paikka) have been opened. Additional views could be opened with the tool below. The ontologies are the same that were used when annotating the images. They tell the user the relevant concepts related to the promotions and underlying images. In this way the ontologies help the user in formulating and focusing the information need.

Queries can be formulated by opening ontological views and by selecting their classes. A query is the conjunction of the selections made in the open ontologies. In the figure, the selection was **Person=GarlandBinder** and **Place=Building**. The metaphor of opening directory folders with images is used here. This view-based idea to information filtering along different indexing dimensions ("facets") is an adaptation of the HiBrowse system developed for a bibliographical information retrieval system [15].

**Image recommendations** The answers to filtering queries are hit lists as customary in search engines, such as Google<sup>13</sup> and AltaVista<sup>14</sup> on the web. However, in contrast to such systems each hit is semantically linked with other images based on the ontological definitions and the annotations. In figure 9 the thumbnail photos beneath the dancing image are links to recommended images. They do not necessarily match the filtering query but that are likely to be of interest. They point, for example, to other images where the same garland binder occurs during the same promotion but not in a **Building**, or photos taken within the same **Place** but depicting only persons in other roles.

By clicking on a recommended thumbnail photo, the large image in view is switched and a new set of recommended images is dynamically generated beneath it. This idea is vaguely related to topic-based navigation used in Topic Maps [13] and the book recommendation facility in use at Amazon.com.

Our image server keeps a log of the session in order not to recommend same images over and over again. Furthermore, a persistent counter for visited images is maintained in a log file. In this way, more popular images can be ranked higher in the recommendations than less popular ones.

The system was implemented in Java with the help of Java Server Page technologies [6], JSP Tag libraries [18], the Apache Tomcat servlet engine<sup>15</sup>, and HP Lab's Jena toolkit (version 1.4.0)<sup>16</sup> [12].

## 6 Discussion

This paper showed that ontologies can be used not only for annotation and precise information retrieval [16, 17], but also for helping the user in formulating

---

<sup>13</sup> <http://www.google.com>

<sup>14</sup> <http://www.altavista.com>

<sup>15</sup> <http://www.apache.org/>

<sup>16</sup> <http://www.hpl.hp.com/semweb/>

the information need and the corresponding query. This is important in applications such as the promotion exhibition, where the domain semantics are complicated and not necessarily known to the user. Furthermore, the ontology-enriched knowledge base of image metadata can be applied to constructing more meaningful answers to queries than just hit-lists. For example, in our demonstration implementation, the underlying knowledge base provided the user with a semantic browsing facility between related recommended images.

The major difficulty in the ontology based approach is the extra work needed in creating the ontology and the detailed annotations. We believe, however, that in many applications – such as in our case problem – this price is justified due to the better accuracy obtained in information retrieval and to the new semantic browsing facilities offered to the end-user. The trade-off between annotation work and quality of information retrieval can be balanced by using less detailed ontologies and annotations, if needed.

## Acknowledgements

Kati Heinämies and Jaana Tegelberg of the Helsinki University Museum provided us with the actual case database and helped in annotating the images. Tero Halonen and Tiina Metso provided expert knowhow about promotions and university traditions. Robert Holmberg, Kim Josefsson, Pasi Lehtimäki, Eetu Mäkelä, Matti Nykänen, Taneli Rantala ja Kim Viljanen participated in the team that implemented the demonstrational system. Our work was partly funded by the National Technology Agency Tekes, Nokia, TietoEnator, the Espoo City Museum, and the Foundation of the Helsinki University Museum, and was supported by the National Board of Antiquities.

## References

1. M. Agosti and A. Smeaton, editors. *Information retrieval and hypertext*. Kluwer, New York, 1996.
2. R. Baeza-Yates and B. Ribeiro-Neto. *Modern Information Retrieval*. Addison-Wesley, New York, 1999.
3. D. Brickley and R. V. Guha. *Resource Description Framework (RDF) Schema Specification 1.0, W3C Candidate Recommendation 2000-03-27*, February 2000. <http://www.w3.org/TR/2000/CR-rdf-schema-20000327/>.
4. J. P. Eakins. Automatic image content retrieval — are we getting anywhere? pages 123–135. De Montfort University, May 1996.
5. D. Fensel (ed.). The semantic web and its languages. *IEEE Intelligence Systems*, Nov/Dec 2000.
6. D. K. Fields, M. A. Kolb, and S. Bayern. *Java Server Pages*. Manning Publications Co., 2002.
7. W. Grosso, H. Eriksson, R. Ferguson, J. Gennari, S. Tu, and M. Musen. Knowledge modelling at the millenium (the design and evolution of protege-2000. In *Proceedings of 12th Workshop of on Knowledge Acquisition, Modeling and Management (KAW-1999)*, Banff, Alberta, Canada, 1999.

8. E. Hyvönen, K. Viljanen, and A. Häätinen. Yellow pages on the semantic web. Number 2002-03 in HIIT Publications, pages 3–15. Helsinki Institute for Information Technology (HIIT), Helsinki, Finland, 2002. <http://www.hiit.fi>.
9. Eero Hyvönen, Petteri Harjula, and Kim Viljanen. Representing metadata about web resources. In E. Hyvönen, editor, *Semantic Web Kick-Off in Finland*, number 2002-01 in HIIT Publications. Helsinki Institute for Information Technology (HIIT), May 2002. <http://www.cs.helsinki.fi/u/eahyvone/stes/semanticweb/>.
10. M. Koskela, J. Laaksonen, S. Laakso, and E. Oja. The PicSOM retrieval system: description and evaluations. In *The challenge of image retrieval, Brighton, UK*, May 2000. <http://www.cis.hut.fi/picsom/publications.html>.
11. O. Lassila and R. R. Swick (editors). Resource description framework (RDF): Model and syntax specification. Technical report, W3C, February 1999. W3C Recommendation 1999-02-22, <http://www.w3.org/TR/REC-rdf-syntax/>.
12. B. McBride, A. Seaborne, and J. Carroll. Jena tutorial for release 1.4.0. Technical report, Hewlett-Packard Laboratories, Bristol, UK, April 2002. <http://www.hpl.hp.com/semweb/doc/tutorial/index.html>.
13. Steve Pepper. The TAO of Topic Maps. In *Proceedings of XML Europe 2000, Paris, France*, 2000. <http://www.ontopia.net/topicmaps/materials/rdf.html>.
14. T. Peterson. Introduction to the Art and Architecture thesaurus, 1994. <http://shiva.pub.getty.edu>.
15. A. S. Pollitt. The key role of classification and indexing in view-based searching. Technical report, University of Huddersfield, UK, 1998. <http://www.ifla.org/IV/ifla63/63polst.pdf>.
16. A. T. Schreiber, B. Dubbeldam, J. Wielemaker, and B. J. Wielinga. Ontology-based photo annotation. *IEEE Intelligent Systems*, 16:66–74, May/June 2001.
17. G. Schreiber, I. Blok, D. Carlier, W. van Gent, J. Hokstam, and U. Roos. A mini-experiment in semantic annotation. In I. Horrocks and J. Hendler, editors, *The Semantic Web – ISWC 2002. First international semantic web conference*, number LNCS 2342, pages 404–408. Springer-Verlag, Berlin, 2002.
18. G. Shachor, A. Chase, and Magnus Rydin. *JSP Tag Libraries*. Manning Publications Co., 2001.
19. J. Sowa. *Knowledge Representation. Logical, Philosophical, and Computational Foundations*. Brooks/Cole, 2000.
20. J. van den Berg. Subject retrieval in pictorial information systems. In *Proceedings of the 18th international congress of historical sciences, Montreal, Canada*, pages 21–29, 1995. <http://www.iconclass.nl/texts/history05.html>.