Learning and Visualizing Cultural Heritage Connections between Places on the Semantic Web

Tomi Kauppinen[1], Kimmo Puputti[1], Panu Paakkarinen[1], Heini Kuittinen[1], Jari Väätäinen[2], and Eero Hyvönen[1]

[1] Semantic Computing Research Group (SeCo)
Helsinki University of Technology
and University of Helsinki, Finland
http://www.seco.tkk.fi/
firstname.lastname@tkk.fi
[2] Geological Survey of Finland
http://www.gtk.fi
firstname.lastname@gtk.fi

Abstract. Semantic web techniques can be used to relate two things together. However, usually this relation is not accompanied with a measure that would tell how interesting the relation is. Data mining tradition provides interestingness measures; it is natural to try and fit semantic web and data mining traditions together. In this paper we use support and confidence values provided by association rule mining as interest measures for relations. The presented method is tailored to location ontologies in order to find out what interesting mutual relations two places have based on annotations in the cultural heritage domain. The method also uses ontology-based reasoning to group places together. We present tests of running the method against a set of over 60,000 annotations in order to find out cultural heritage connections between places.

1 Introduction

Cultural heritage collections contain rich semantic metadata about cultural objects such as museum objects, photographs, maps, paintings, poems, books, folk songs, and videos. This metadata contains explicit descriptions of objects, but in addition there is implicit hidden knowledge that has a potential to be discovered by data mining techniques. For example, in cultural heritage annotations of museum objects places co-occur in different roles such as *place of manufacture* or *place of usage*. This means that an item in museum might have originally been manufactured in Asia but with the help of Silk Road it has been used in Europe.

In essence, the research problems in here are:

- How to find out these implicit relations between places?
- How to visualize relations for a user?

In this paper we will present a method to create these semantic cultural heritage relations between locations to be used in e.g. visualization. The method examines first in which different roles places are in annotations of objects: the examined roles are *place of manufacture* and *place of usage*. Method then uses association rule mining [1] to relate locations together. The result can then be used with the help of visualization mechanisms to examine how different locations (or cultures) have exchanged e.g. goods and art over time.

However, one of the identified problems is that annotations in cultural heritage data make references to locations from different levels of granularity (e.g. countries vs. cities). For this reason we also present a revised version of our method that uses ontology-based inference to group locations by using topological relations [5] (*overlaps* and *touches*) and partonomy hierarchies. The method is then able to produce cultural relations also between grouped locations, for example, between “Sweden in different times” and “Finland in different times”.

The paper is organized as follows. Section 2 describes annotations of the semantic cultural heritage portal CULTURESAMPO used in the experiments. Section 3 describes a method that uses co-occurence techniques to create relations between locations taking into account their roles in the annotations and presents results of testing the method. Section 4 identifies problems of the first version of the method due to cultural heritage data, presents a revised method and shows the results of the revised method. Section 5 discusses the results and the related work in the field and Section 6 concludes the paper.

2 Materials

2.1 Cultural Objects of CULTURESAMPO

The material used in this research are annotations of cultural objects in the portal CULTURESAMPO Finnish Culture on the Semantic Web 2.0 [7]. The material consists of heterogeneous cultural content which comes from collections of 12 Finnish museums, libraries, archives, and other memory organizations is annotated using various ontologies. All of these annotations are made using the Resource Description Framework (RDF)[3] and a set of ontologies.

The dataset is metadata about over 60,000 objects, e.g. museum objects, photographs, maps, paintings, poems, books, folk songs, videos, et cetera coming from almost 100 different collections and over 1,5M other reference resources (concepts, places, times, etc.). There are many kinds of works of art, such as museum items, literary work, documents, illustrated works (paintings), etc. In our research the most important object type appears to be artifact, because artifacts commonly have different location roles such as place of manufacture, place of usage and place of collection. It is notable that around 9000 of 60,000 objects are museum items.

CULTURESAMPO annotations include the following location roles:

– place of discovery: a place from where an object was found
– place of manufacture: a place where an object was manufactured
– place of acquirement: a place from where an object was acquired
– place of creation: a place where an object was created
– place of photographing: a place where a photograph was taken at

[3] http://www.w3.org/RDF/

– place of subject: a place depicted in an object such as a painting
– place of usage: a place where an object was/is used
– place of context: a place relevant to an object in an unspecified way

See table 2.1 for overall statistics of different roles of location used in different types of objects.

	CW	MI	LW	IW	MP	D	N	WA
total number of items	50299	9005	7418	19655	8613	339	11161	116
place relations	43369	7656	6428	14927	8216	7	6021	114
place properties:								
place of manufacture	5449	4150	1299					
place of photographing	13001		13001					
place of acquirement	2427	2427						
place of usage	3203	3096						107
place of discovery	4040	4040						
place of context	768	661				7	100	
place of subject	2038			1924				114

Table 1. Number of different works of art related to different place properties. List of acronyms in the table: CW = cultural work , MI = museum item, LW = literary work, IW = illustrated work, MP = musical piece, D = document, N = narrative, WA = work of art

CULTURESAMPO (CS) makes use of a Place Ontology. CULTURESAMPO Place Ontology provides e.g. partonomy hierarchy of locations and coordinates of center points of locations. It is derived from the annotations and enriched and validated later on with coordinate and partonomy information from Place Name Registry (PNR)[4] gazetteer. PNR contains around 800000 place names and hence contains many times more Finnish place names compared to e.g. TGN (Getty Thesaurus of Geographic Names)[5], Geonet Names Server (GNS)[6], or Geonames-dataset[7]. However, descriptions of international places in CS Place Ontology have been enriched with coordinates from GNS.

3 Method

3.1 Relating Locations with Association Rules

As we saw in the previous section in many annotations multiple locations are mentioned in different roles, for example an artifact may have a place of manufacture and also a place of usage. This implies that two locations are related because of cultural activity

[4] PNR is provided by the National Land Survey of Finland (http://www.mml.fi)
[5] http:www.getty.eduresearchconducting_researchvocabulariestgn
[6] http://earth-info.nga.mil/gns/html/
[7] http://www.geonames.org/

between them. However, even if one co-occurrence already relates two locations it is not apparent how interesting this relation is.

In order to relate locations together and to measure their interestingness we apply a method for mining association rules [1] and define $support(A \Rightarrow B) = S_{A\Rightarrow B} (= S_{B\Rightarrow A})$ as follows

$$S_{A\Rightarrow B} = \frac{number\ \ of\ \ annotations\ \ containing\ \ both\ \ A\ \ and\ \ B}{total\ \ number\ \ of\ \ annotations} \tag{1}$$

Where A is a set of annotations of objects where a certain location have a certain role and similarly B is another set of annotations of objects.

Similarly, $confidence(A \Rightarrow B) = C_{A\Rightarrow B}$ is defined as

$$C_{A\Rightarrow B} = \frac{number\ \ of\ \ annotations\ \ containing\ \ both\ \ A\ \ and\ \ B}{number\ \ of\ \ annotations\ \ containing\ \ just\ \ A} \tag{2}$$

For example, figure 1 shows sets A and B where in all objects of set A Helsinki has the role of *place of usage*. These four objects are denoted with letters *k*, *l*, *m*, and *n*. Similarly Mumbai is defined in the role *place of manufacture* in all objects of set B, namely in the three objects denoted with letters *m*, *n* and *o*. The intersection of A and B contains those two objects *m* and *n* having Helsinki as the *place of usage* and Mumbai as the *place of manufacture*. We can then calculate e.g. $C_{A\Rightarrow B} = \frac{2}{4} = 0.5$ and $C_{B\Rightarrow A} = \frac{2}{3} \approx 0.67$.

The method has following steps:

1. Creation of sets $(A, B, \ldots)$ for each location in each role (e.g. "Helsinki" as a place of manufacture) based on annotations. One set A_i contains all the objects where e.g. "Helsinki" is mentioned in the role place of manufacture.
2. Association rule mining. Sets $(A, B, \ldots)$ are used for mining association rules. The minimum support and the minimum confidence can be used to prune out infrequent patterns.

3.2 Results of The First Version of The Method

We applied the described method to produce relationships between locations using the roles *place of manufacture* and *place of usage* in annotations of CULTURESAMPO . This method was implemented using Weka[8].

When using the association rule mining without ontological inference the method found 803 relations between locations in roles *place of manufacture* and *place of usage*. In practise, the implementation produced following kind of relations expressed in Turtle[9], shown partially in the following example (relation stores also number of objects in question, links to objects, support value etc.). In this example Lahti is in a role of a place of usage and Hollola of a place of manufacture.

[8] www.cs.waikato.ac.nz/ml/weka/
[9] http://www.w3.org/2007/02/turtle/primer/

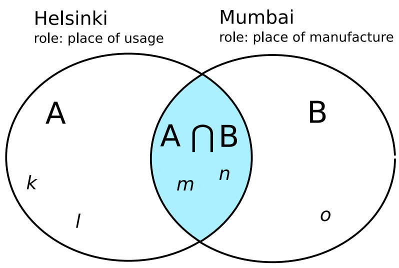

Fig. 1. This picture depicts two sets A and B where A is a set of objects where Helsinki is in the role *place of usage*. B is a set of objects where Mumbai is in the role *place of manufacture*.

```
relation:mininginstance17457
      rdf:type
              relation:culturalheritage ;
      relation:relationType
              relation:manufacture2usage ;
      relation:argument1
              csplace:hollola_manufacturingplace ;
      relation:argument2
              csplace:lahti_usageplace ;
      relation:confidence1to2
              "0.009" .
      relation:confidence2to1
              "0.611" ;
```

3.3 Visualization of Results

In order for a user to be able to examine how different locations (or cultures) have exchanged e.g. goods and art over time we implemented a visualization system that uses the cultural relations produced by the method.

Figure 2 shows a visualization of results on a map [10]. A user is interested in seeing where items used in a city called Lahti have been manufactured. As a first step, the user has selected Lahti in the role *place of usage*. As a result, all those locations that have the role *place of manufacture* in same item annotations are shown and again the thickness of the arrow depicts how large a portion of objects used in Lahti are from a certain location such as Japan, Italy, France, Brasil, etc.

The problem, however, is that locations are at different levels of granularity (countries versus cities), or represent close-by regions (neighbouring cities) or even over-

[10] We make use of the openly available Google Maps API to provide the map visualization.

lapping historical regions. In the next section we examine these problems further and propose a revised method to tackle them.

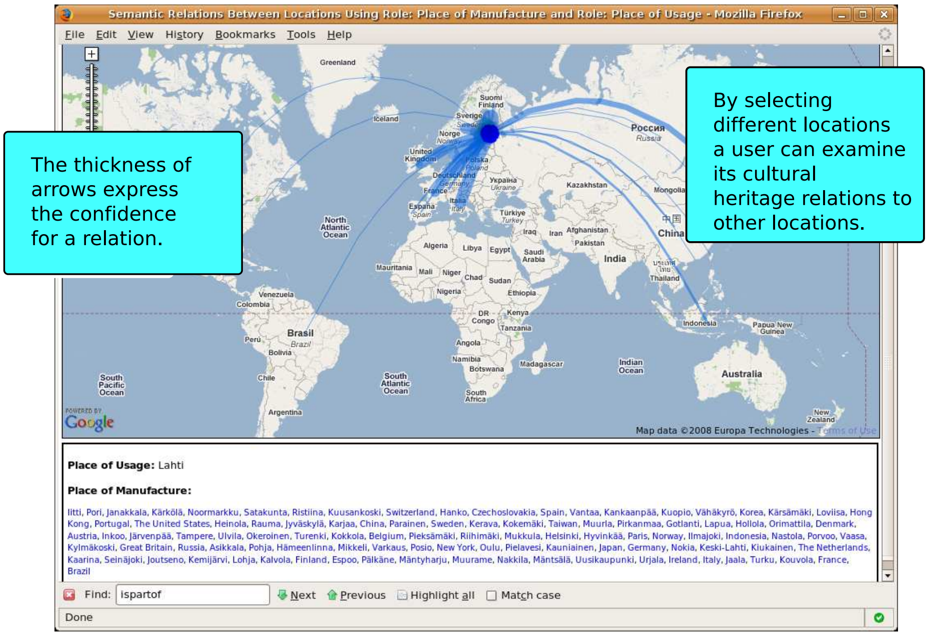

Fig. 2. This time a user has selected Lahti as a role *place of usage* and locations having a role *place of manufacture* in same item annotations are shown.

4 Revised Method

4.1 Why the Method Needs Revising?

As we saw in the previous section association rule mining was able to relate places together. However, there are problems related to the nature of annotations. Namely, annotations are characterized by the following features:

Semantic granularity: Annotations contain references to locations on different levels of granularity. For example, some museum items may refer to Paris while an other item refer to France.

Use of nearby locations: Nearby locations are used in annotations. For example, two neighboring villages may be referenced in different annotations.

Mismatches due to historical changes: Locations have changed their names, or they have merged and split, which causes semantic mismatch between locations from different times. For example, in one annotation a reference is made to (historical) Bombay while in other annotation a reference is made to (contemporary) Mumbai.

To provide a revised method that is able to take into account these problems we will make use of SAPO (The Finnish Time-Location Ontology) [11]. SAPO contains historical locations (mainly municipalities), changes between them, and temporal properties (like when a location has existed) and spatial properties (like polygonal boundaries). SAPO defines also 784 *overlaps*- and 2645 *touches*-relations between historical municipalities. The *overlaps*-relations between municipalities were generated using knowledge about changes e.g. merges and splits between locations [9]. The *touches*-relations between neighbouring municipalities were generated to SAPO automatically by examining polygons of historical municipalities.

4.2 Grouping Places as a Solution

As said, annotations contain references to locations that are topologically and mereologically close i.e. they overlap, touch or are in a partonomy hierarchy. A practical example of this is depicted in Figure 3 showing locations near the current border between Finland and Russia. A municipality called Imatra *overlaps* many historical municipalities, namely Ruokolahti, Jääski and Joutseno. On the other hand all these three historical municipalities were neighbors of each other i.e. they *touch*. Different municipalities near the current border have also been in different partonomy hierarchies i.e. as *part of* Finland or as *part of* Russia.

For these reasons we will revise our method and also group places together. The following example illustrates the idea of these groupings.

All places of Figure 3 have been used in annotations. For example, a hat in the bottom left corner is annotated with Joutseno in the role *place of usage* while Jääski is in the same role in an annotation of the horn shown in the bottom right corner. Hence a group called G_1 = "Joutseno and its neighbouring municipalities as places of manufacture" contains these two municipalities. Another group is G_2 = "Imatra and its historical overlapping municipalities as places of usage". Further on, as Figure 3 shows, two different museum items were manufactured either in Imatra (a chair in 1957) or in an overlapping historical Ruokolahti (a shepherd's whistle in 1920). Hence these two items belong to the group place G_2.

Next, we will describe a revised method that takes into account these characteristics to produce relations between locations by using following principles:

1. Interestingness of relations are measured based on amount of co-occurrences as was done in the first version of the method
2. Locations are grouped and new "group places" created
   (a) if places are in the same partonomy,
   (b) if places are neighbouring locations (i.e. they touch) or
   (c) if places overlap for historical reasons.

4.3 Revising the Method

We will next present different types of grouping criteria of locations in sections 4.4, 4.5, and 4.6. These types include grouping by partonomy (by using *partof*-relations), grouping by neighborhood (by using *touches*-relations), and finally grouping by diachronic

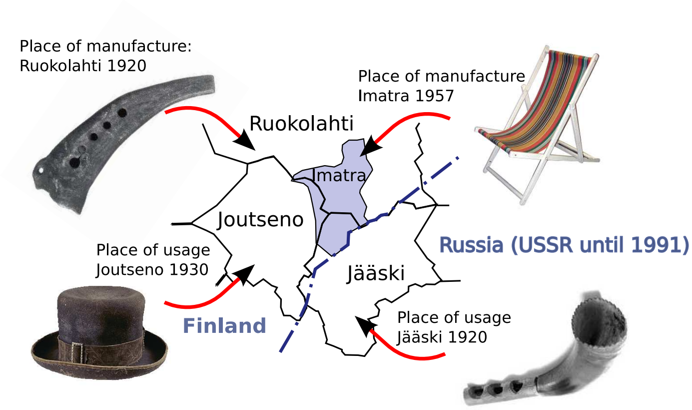

Fig. 3. An example of complexity: locations can be from different partonomical levels, they might overlap due historical reasons or they are neighbors.

regions[11] by using *overlaps*-relations. Groupings will be used by the method to infer relations between groups of locations rather than just between locations.

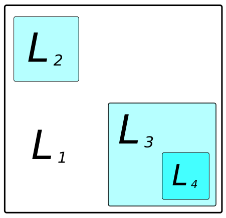

Fig. 4. Locations L_1, L_2, L_3 and L_4 in a partonomy.

In essence, the revised method goes as follows. The step one creates initial sets of items for each location in each role. Step two uses different methods to produce groups of locations — these will be explained in the following subsections. And step three produces weighted relations between locations (or groups of locations).

4.4 Grouping by Partonomy

In this variation of the inference, mereological relations i.e. containment relations (part-of-relations) are used to group locations that are in the same partonomical hierarchy.

[11] The term “diachronic” comes from Greek where “dia” means through and “chronos” means time. By diachronic region we mean here a location together with all historical locations that overlap it because of e.g. merges, splits or name changes.

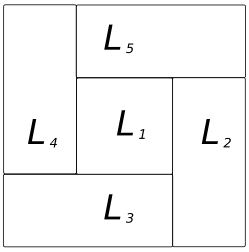

Fig. 5. Locations L_2, L_3, L_4 and L_5 are neighboring locations with location L_1 i.e. they all have a $touches$ relation with L_1.

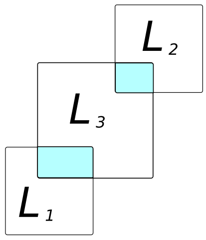

Fig. 6. Locations L_1 and L_2 overlap the location L_3.

See Figure 4 where locations L_2, L_3 and L_4 are parts of location L_1. Note that L_4 is a part of L_3 but because L_3 is a part of L_1, accordingly, also L_4 is a part of L_1 due the transitive nature of part-of relationship.

Grouping means that all roles (like place of manufacture) locations have in annotations are propagated up in the partonomy hierarchy. As a result new location groups are generated such as “India and all its municipalities in role of manufacturing”. If e.g. “Mumbai” is mentioned in the role *place of manufacture* in one annotation, “India and all its municipalities in the role of manufacturing” will also get the role *place of manufacture* in the same annotation. The partonomy of a location is a union $\bigcup_P$ of location itself and all locations that are *part of* it.

Grouping by partonomy enables to answer to questions like “In which continents items manufactured in Asia are used in?”

4.5 Grouping by Neighboring Locations

Neighboring locations of the location L_1 form together “neighborhoods” of the location L_1. This idea is used when locations are grouped by neighboring locations to be used by the method. In essence the neighborhood of a location L_1 is formed by using *touches*-relations location L_1 has with other locations.

In Figure 5 locations L_2, L_3, L_4 and L_5 are neighboring locations with location L_1 i.e. they all $touch$ L_1. For example, if “Helsinki” *touches* “Vantaa”, “Espoo” and “Sipoo”, then the neighborhood of “Helsinki” contains all these neighbors and “Helsinki” itself.

Grouping by neighborhood means in practice that location roles are propagated to “neighborhoods”. For example, if an item is manufactured in “Espoo” then the “neighborhood of Helsinki in the role of manufacturing” will get the role *place of manufacture* in this annotation in the inferred model as well. The neighborhood of a location is a union $\bigcup_N$ of region itself and all regions it *touches*.

Grouping by neighborhood enables to answer questions like “In which approximate locations items manufactured in Helsinki (or nearby) are used in?”

4.6 Grouping by Diachronic Regions

Annotations in cultural heritage domain typically make references to historical locations that may overlap even if these locations have different names. This phenomenon is used as a motivation when historical locations are grouped to form *diachronic regions*. Figure 6 illustrates the idea: both locations L_1 and L_2 overlap the location L_3. These overlaps are due to historical changes. For example, Mumbai changed its name from Bombay in year 1996 [12] which means their spatial extensions heavily overlap.

Grouping by diachronic regions propagates roles locations have in annotation also to "diachronic" regions they overlap. For example, if an item is manufactured in "Bombay" this knowledge is propagated to "diachronic Mumbai". Hence a diachronic region is a union $\bigcup_D$ of region itself and all overlapping regions.

Grouping by diachronic regions enables to answer questions like "In which modern (or historical) locations items manufactured in Mumbai (or any of its historical predecessors) are used in?"

4.7 Results of the Revised Version of the Method

The revised method, including the association rule mining (apriori), was implemented in Java using Jena library [3] in order to provide reasoning for groupings. The efficiency was not the main goal. By using the ontological inference i.e. when grouping locations together in different ways the result was a lot more relations. By grouping locations using partonomy hierarchy the method was able to create 5017 relations between locations in these two roles. By adding grouping by diachronic regions the result was 5799 relations. Grouping both by partonomy and neighborhood resulted in 13128 relations. By combining all different groupings the result was 13806 relations. All these relations are ranked by support and confidence.

Among these results are more generalized ones than what we were able to produce with the first version of the method. For example, a user can examine relations on a country level, or between two countries that both have had different borders and perhaps even different names in different times. Furthermore, instead of examining relations of a single city like Helsinki a user is able to examine the relations from whole neighborhood of Helsinki to other parts of the world.

5 Related Work and Discussion

Association rule mining alone was able to produce cultural heritage relations between places in different roles. In our visualization system all these relations are shown and the width of the arrow indicates the confidence value for that relation.

However, as we identified, the granularity of references to places in annotations creates problems. As a solution we used topological relations of geospatial ontologies in a revised method. Grouping by partonomy, neighborhood or diachronic regions produced substantially more relations. By combining these different grouping types the method was able to produce even more relations.

[12] http://en.wikipedia.org/wiki/Mumbai

Data mining techniques have been applied in the area of the semantic web for example for mining frequent characteristics [8] from knowledge bases. These characteristics can be e.g. places of living or ages of clients of some bank. Furthermore, association rule mining have been applied to analyze and to structure folksonomies [13] and for extending an existing ontology [10]. The grouping technique we proposed is somewhat related to the tradition of mining multilevel association rules (see e.g. [6], page 244) where concept hierarchies are used for providing these "groups". Multirelation association rule mining is another closely related field (see e.g. [4]). Spatial association rules have also been proposed [12], they are rules like "most big cities in Canada are close to the Canada-U.S. border". Wiki content has been used as a source to extract content to allow for querying relations between places [2] and e.g. for revealing that Innsbruck and Leipzig both share the same twin town, Krakow. The interestingness measures we used were limited to support and confidence but the presented framework could easily utilize also other interestingness measures (such as lift and conviction) as well (see e.g. [14] for discussion and overview of measures). However, in that case the visualization part should also be altered accordingly.

Our approach is different from existing ones in that it uniquely combines 1) spatial tradition (to be able to explicate spatial relations), 2) ontology-based reasoning (to be able to group locations by spatial relations), and 3) co-occurrence techniques (to explicate the confidence and support of a relation) to produce semantic relations between locations using roles and 4) application of the methods in the cultural heritage domain.

6 Conclusions

In this paper we introduced a method for creating cultural heritage relations between places based on activities between them. In order to create them we used a set of annotations of a semantic cultural heritage portal. We concentrated in two roles locations have in annotations: *place of manufacture* and *place of usage* and examined whether two locations co-occur in annotations in these roles. By employing co-occurrence techniques the method examined how strong this pattern is compared to other co-occurrences. However, as was shown, annotations use locations from different levels of a partonomy, from neighboring locations or historically overlapping regions. For this reason we presented a revised method that is able to group locations (by e.g. partonomy) with the help of ontology-based reasoning.

Acknowledgments

Our research is done in the National Semantic Web Ontology Project in Finland [13] (FinnONTO) 2003–2007, 2008-2010 funded mainly by the Finnish Funding Agency for Technology and Innovation (Tekes) and in an EU project SmartMuseum[14] supported within the IST priority of the Seventh Framework Programme for Research and Technological Development.

[13] http://www.seco.tkk.fi/projects/finnonto/

[14] http://smartmuseum.eu/

References

1. R. Agrawal, T. Imielinski, and A. N. Swami. Mining association rules between sets of items in large databases. In P. Buneman and S. Jajodia, editors, *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*, pages 207–216, Washington, D.C., FebruaryJune–FebruaryAugust˜ 1993.
2. S. Auer and J. Lehmann. What have Innsbruck and Leipzig in common? Extracting semantic from wiki content. In E. Franconi, M. Kifer, and W. May, editors, *Proceedings of the 4th European Semantic Web Conference (ESWC)*, volume 4519 of *Lecture Notes in Computer Science*, pages 503–517. Springer, 2007.
3. J. J. Carroll, I. Dickinson, C. Dollin, D. Reynolds, A. Seaborne, and K. Wilkinson. Jena: Implementing the semantic web recommendations. Technical Report HPL-2003-146, HP Labs, December 24, 2003.
4. L. Dehaspe. *Frequent pattern discovery in first-order logic*. Ph.D. Dissertation, K.U.Leuven, December 1998.
5. M. Egenhofer. A formal definition of binary topological relationships. In *Foundations of Data Organization and Algorithms, 3rd International Conference, FODO 1989*, volume 367 of *Lecture Notes in Computer Science*, pages 457–472, 1989.
6. J. Han and M. Kamber. *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers, 2000.
7. E. Hyvönen, E. Mäkelä, T. Kauppinen, O. Alm, J. Kurki, T. Ruotsalo, K. Seppälä, J. Takala, K. Puputti, H. Kuittinen, K. Viljanen, J. Tuominen, T. Palonen, M. Frosterus, R. Sinkkilä, P. Paakkarinen, J. Laitio, and K. Nyberg. Culturesampo – finnish culture on the semantic web 2.0. thematic perspectives for the end-user. In *Proceedings, Museums and the Web 2009, Indianapolis, USA*, April 15-18 2009.
8. J. Józefowska, A. Ławrynowicz, and T. Łukaszewski. Frequent pattern discovery from owl dlp knowledge bases. In *Managing Knowledge in a World of Networks, 15th International Conference, EKAW 2006*, pages 287–302, Podebrady, Czech Republic, October 2-6, 2006,.
9. T. Kauppinen and E. Hyvönen. *Modeling and Reasoning about Changes in Ontology Time Series*, pages 319–338. Integrated Series in Information Systems. Springer-Verlag, New York (NY), New York, NY, January 15 2007.
10. T. Kauppinen, H. Kuittinen, J. Tuominen, K. Seppälä, and E. Hyvönen. Extending an ontology by analyzing annotation co-occurrences in a semantic cultural heritage portal. In *Proceedings of the ASWC 2008 Workshop on Collective Intelligence (ASWC-CI 2008), 3rd Asian Semantic Web Conference (ASWC 2008), Bangkok, Thailand*, December 8-11 (postponed to early 2009) 2008.
11. T. Kauppinen, J. Väätäinen, and E. Hyvönen. Creating and using geospatial ontology time series in a semantic cultural heritage portal. In *S. Bechhofer et al.(Eds.): Proceedings of the 5th European Semantic Web Conference 2008 ESWC 2008, LNCS 5021, Tenerife, Spain*, pages 110–123, June 1-5 2008.
12. K. Koperski and J. Han. Discovery of spatial association rules in geographic information databases. In *SSD '95: Proceedings of the 4th International Symposium on Advances in Spatial Databases*, pages 47–66, London, UK, 1995. Springer-Verlag.
13. C. Schmitz, A. Hotho, R. Jäschke, and G. Stumme. Mining association rules in folksonomies. In *Proceedings of the 10th International Federation of Classification Societies Conference*, pages 261–270. Springer, 2006.
14. P.-N. Tan, V. Kumar, and J. Srivastava. Selecting the right objective measure for association analysis. *Information Systems*, 29(4):293 – 313, 2004.