

## Chapter 11

# MODELING AND REASONING ABOUT CHANGES IN ONTOLOGY TIME SERIES

Tomi Kauppinen and Eero Hyvönen \*

*University of Helsinki, Department of Computer Science and  
Helsinki University of Technology, Media Technology and HIIT  
P.O. Box 5500, Otaniementie 17, FIN-02015 TTK*

**Abstract** Ontologies evolve when the underlying domain world changes at different points of time. The result then is a series of ontologies whose concepts are related with each other not only within one ontology valid at a moment but through the time, too. This chapter presents a model for representing ontology time series. The focus is on modeling partial overlap between concepts evolving over long periods of time, and the domain of application is historical geospatial reasoning. A framework is presented for representing and reasoning about conceptual overlap of concepts that evolve over an ontology time series. The idea is to provide the ontology developer with an intuitive change ontology for expressing local ontological changes in a declarative way. An algorithm is presented for reasoning about overlapping concepts globally over long periods of time. This algorithm can be applied, e.g., in concept-based information retrieval for ranking search results according to their relevance.

## ONTOLOGY CHANGE AND INFORMATION RETRIEVAL

Ontologies are a key technology underlying the Semantic Web (Fensel, 2004). They are used for defining vocabularies by which the metadata describing web contents is represented in a machine-interpretable way. Based on ontologies, intelligent content-based web services can be created and semantic interoperability of web systems enhanced.

\*Tomi Kauppinen and Eero Hyvönen. Modeling and Reasoning about Changes in Ontology Time Series. Published as a peer reviewed chapter (chapter 11) in Book: Ontologies: A Handbook of Principles, Concepts and Applications in Information Systems. Rajiv Kishore, Ram Ramesh, Raj Sharman (editors). Springer-Verlag, New York, 2007.

An important area of semantic web applications is information retrieval. In ontology-based search, content annotations and queries are based on concepts rather than on keywords. This leads not only to better precision and recall, but ontologies can be used as a navigational aid to help the end-user in formulating the queries and results. For example, in the semantic portal MuseumFinland<sup>1</sup> (Hyvönen et al., 2004) a location paronymy<sup>2</sup> is used for annotating museum artifacts with metadata about the place of manufacture and usage. The same ontological resources are exposed to the end-user as a hierarchical view<sup>3</sup> of categories to be selected when searching semantically related artifacts. A problem in applications like this is that the content in the underlying history-related databases is annotated using historical location concepts that have evolved as time has gone by. For example, an artifact may have been manufactured in East Germany—a country that does not exist any more in the location ontology used today. If the today's ontology is used for formulating a query concerning modern Germany, which may be natural from the end-user's viewpoint, then finding artifacts made or used in East Germany becomes problematic. To solve the mapping problem between query and annotation concepts, a spatio-temporal model of the ontological change from East Germany to current united Germany is needed.

More generally, the problem of areal change is quite common in the area of geospatial ontologies and reasoning, but is not discussed much in the literature (Visser, 2004; Stuckenschmidt and Harmelen, 2004). We investigated, for example, how Finnish counties and cities have been merged together, split into parts, changed name, and annexed to and from the neighboring countries. It turned out (Väättäinen, 2004) that after the year 1900 alone there have been nearly 900 changes in the borders and names of the counties and cities in Finland. The number of changes at similar level of granularity in Europe from the times of the Roman Empire until today would be extremely large. Think only how the map of Europe has changed during the last 100 years at the level of countries.

## OUTLINE OF THE CHAPTER

In this chapter we address the problem of ontological change from the information retrieval point of view. We investigate how to answer to a query based on concepts at a time  $t_q$  by using metadata annotated in terms of spatiotemporally related concepts from another point of time  $t_a$ . The time of the query concept is often after the metadata concept time, like in the query

“What modern British towns ( $t_q$ ) are former Viking settlements ( $t_a$ )?”,

but also the other way around, like in the query

“What Czechoslovakian cities ( $t_a$ ) are within the current area of EU ( $t_q$ )?”.

To deal with the problem, the concepts used at different times should somehow be mapped with each other.

In the following, we first formulate the information retrieval task addressed above. Then a model for representing a time series of ontologies is defined, a method for representing ontological changes in a partonomy time series is presented, and an algorithm for computing a partial overlap relation of concepts of different times is developed and illustrated with a realistic example. The overlap relation can be used for retrieving conceptually related objects and for ordering them according to their relevance. In conclusion, contributions of the work are summarized, related work discussed, and directions for further research outlined.

## **REASONING OVER ONTOLOGY TIME SERIES**

We focus on one aspect in the field of spatial reasoning (Visser, 2004; Stuckenschmidt and Harmelen, 2004): spatial overlap of regions. The other dimensions of spatial relevance, such as topology (of neighboring regions), directions (of related regions), and distances (between regions) are not considered here but could in principle be combined with partonomical relevance, as discussed in (Stuckenschmidt and Harmelen, 2004), chapter 8. Location partonomy will be used as the example domain, but the ideas presented could probably be extended to other domains dealing with other forms of conceptual overlap.

When dealing with historical data, the ontological vocabulary has to cover relevant location categories through different times of interest. There is a time series of location ontologies each of which is valid during a limited period of time. The next ontology in the series is needed whenever a set of simultaneous changes in the modeled domain occurs. This kind of evolution of ontology time series is due to changes in the underlying domain and should not be confused with ontology versioning (Klein and Fensel, 2001), database schema evolution, or ontology evolution (Noy and Klein, 2003) that deal with ontology refinements or other changes in the conceptualization (Klein, 2004; Stojanovic, 2004).

There has been an active philosophical discussion about how and whether things of the world endure or perdure as time goes by (Sider, 2001; Stell and West, 2004; West, 2003; Grenon and Smith, 2004). According to three-dimensionalism, things have only spatial parts, they endure, and are wholly present throughout the time interval of their existence. Four-dimensionalism challenges this view by asserting that things also have temporal parts in addition to their spatial parts. For example, the notion of a person has temporal parts such as childhood and death. According to this view, things can be seen as “space worms” that spread out in spacetime. In the SNAP/SPAN-approach

(Grenon and Smith, 2004), both views are supported by a combination of a three-dimensional SNAP-ontology and a four-dimensional SPAN-ontology.

Our approach will use a sequence of three-dimensional models (actually, only two-dimensional areas are considered). Each member ontology in an ontology time series defines a snapshot, where the objects representing two-dimensional regions stay the same. Every ontology may be used quite rightly for both annotations and for querying. We will represent areal changes between regions in successive ontologies, but the identity of a region is changed after each change. We therefore do not have the four-dimensional notion of the same concept, say, “Finland” changing — or merely having different temporal parts — through its history as time goes by, although the change chains between areas in successive ontologies can be seen as a kind of space worms in space-time.

As in (Holi and Hyvönen, 2004) the reasoning problem of this chapter is formulated as follows: Given is a set of regions whose extensions are geospatial areas that may overlap with each other. If a resource has been annotated using a concept  $A$ , and the user makes a query by using another concept  $Q$ , then our task is to determine how *relevant* is  $A$  with respect to  $Q$ .

It is argued that the notion of relevance can be expressed in a natural way as the *proportional overlap*:

$$p = |A \cap Q|/|A| \quad (1.1)$$

Here the notation  $|X|$  denotes the extensional size of the set  $X$ . For example, if the annotation concept  $A$  denotes the area of the Nato countries in 1960, and the query concept  $Q$  denotes the area of EU represented as a geospatial area (e.g., a polygon of points or pixels), then  $p$  tells how much EU covers Nato (1960) and gives a measure of how likely it is that an object within the range of the Nato (1960) is also within the range of EU.

The practical result of this chapter will be a method and an algorithm for deriving the *global proportional overlap relation*

$$o : Q, A \rightarrow p, p \in [0, 1] \quad (1.2)$$

between any query concept  $Q$  and annotation concept  $A$  in an ontology time series. By using the relevance relation  $o$ , all overlapping concepts for a query concept  $Q$  can be found both in the past and in the future. Moreover, the values  $p$  can be applied as a measure of relevance to sort the search results and to visualize the mutual relevance of evolved concepts. For example, Figure 1.1 depicts the merger of East and West Germany that constitutes modern Germany. If we query with the area of East Germany, and the metadata is annotated with the newer concept of Germany, we get relevance  $o(E.Germany, Germany) = 0.3$ , because East and West Germany cover Germany exhaustively and exclusi-

vely in proportion 30%/70%. Since East and West Germany are disjoint we get  $o(E.Germany, W.Germany) = 0$ , and so on. This kind of reasoning based on one *local* change is straightforward, but when the change chains become longer and are intermingled, determining *global* overlaps becomes a challenge.

An obvious approach to determine  $o$  would be to map the concepts to physical areas in the real world represented as polygons (Visser, 2004; Stuckenschmidt and Harmelen, 2004). In this way, overlaps between concepts can be determined by geometrical intersection algorithms as customary in Geographical Information Systems (GIS). A problem of this approach is, however, that the physical areas corresponding to the concepts have to be known exactly, which is not always the case when dealing with historical areas. For example, in our case study for modeling counties and cities in Finland, such information is not available. We therefore decided to take another avenue, where the idea is to model the changes, that were known and easier to represent, and then compute the global overlap relation based on local proportional overlaps related to individual changes. In this way one can determine not only the overlap table but also explicate the series of changes and ontology versions that evolves through time in a concise, semantically interpretable ontology. This representation can be used in other reasoning and visualization tasks, too, like for selecting query concepts from an ontology valid at a particular historical time.

## A MODEL OF ONTOLOGY TIME SERIES

This section presents a model of ontology time series. The model will be used as the basis for deriving the global proportional overlap relation between concepts.

### ONTOLOGY TIME SERIES

A *temporal ontology*  $O = \langle R, T \rangle$  is a set of ontology resources  $R$  that persist over a time span interval  $T$ . An ontology resource  $r \in R$  is a tuple  $\langle name, T_r, P \rangle$  where  $name$  is the name,  $T_r \supset T$  is the time span, and  $P$  is the set of additional *properties* of the resource. An ontology resource  $r \in R$  represents, e.g., a two dimensional geospatial region and its characteristics in the world. We will consider semantic web ontologies represented in terms of RDF triples (Brickley and Guha, 2004), where a resource is characterized by an identity (URI) and related property triples of form  $\langle uri, property, value \rangle$ . Intuitively, each ontology resource can be identified by a URI associated with a name, an interval  $T_r$  representing the time span during which the resource persists, and other properties. All resources persist over the whole time span of the ontologies in which they belong.

At certain *change points*  $t_i$  of time, one or more resources  $r$  become obsolete or new resources emerge due to a change in the underlying domain conceptua-

lized by the ontology. For example, East and West Germany were reunited into Germany in 1991. This means that the old notions of East and West Germany became obsolete after 1991, and the new concept of Germany was introduced as depicted in Figure 1.1. New concepts are typically related to old ones in various ways. Here, for example, the geospatial area of Germany is the union of that of the two disjoint merged countries.

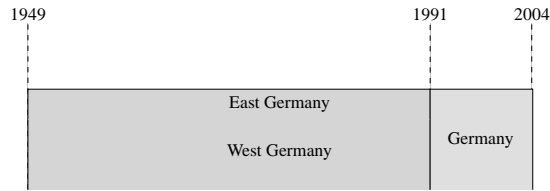


Figure 1.1 Individuals *West Germany*, *East Germany* and *Germany* of an ontology. X-axis depicts time and y-axis the relative areas of the countries.

An *ontology time series* is a tuple  $\langle O, S \rangle$ , where  $O = \langle R, T \rangle$  is a temporal ontology and  $S = [t_0, \dots, t_n]$  is a sequence of change points  $t_i < t_{i+1}$ ,  $i = 0 \dots n - 1$ , such that  $T = [t_0, t_n]$ . It is assumed that the time span limits of the ontology resources  $r \in R$  define exhaustively and exclusively the set of change points, which means that an ontology time series with change points  $t_i, i = 0 \dots n$ , define a series of  $n$  successive *period ontologies*. Figure 1.2 illustrates the idea: there are three ontology resources  $r_j$  whose end-points de-

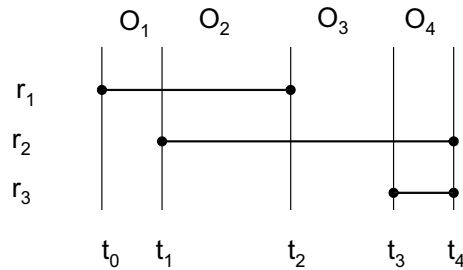


Figure 1.2 An ontology time series of four ontologies  $O_i$  and three resources  $r_j$ , whose time span is depicted by the horizontal lines over the five successive change points of time  $t_k$ .

fine five distinct change points  $t_k$  and four period ontologies  $O_i$ , such as the ontology

$$O_1 = \langle \{r_1\}, [t_0, t_1] \rangle \text{ and} \\ O_2 = \langle \{r_1, r_2\}, [t_1, t_2] \rangle.$$

In an ontology time series, each resource belongs to, and its persistence time spans over at least one period ontology. An individual ontology in the series corresponding to a period  $T$  can be constructed by collecting all ontology resources that span over  $T$ .

## MODELING SPATIAL RELATIONSHIPS BY CHANGE BRIDGES

We introduce the notion of the *change bridge* for representing overlap changes in an ontology time series. A change bridge is associated with a change point and tells, what current concepts become obsolete (if any), what new concepts are created (if any), and how the new concepts overlap with older ones. A bridge is defined by the tuple

$$\langle t, OLD, NEW, covers, coveredBy \rangle, \quad (1.3)$$

where  $t$  is a change point of time,  $OLD$  is the set of resources that become obsolete at  $t$ , and  $NEW$  is the set of new resources introduced at  $t$ . The function

$$covers : n, r \rightarrow p \in [0, 1] \quad (1.4)$$

tells how much each new resource  $n \in NEW$  covers the other resources  $r \in R \cup NEW$  in the ontology time series  $O = \langle R, T \rangle$  at the moment  $t$ . It is represented in terms of proportional overlap  $p = covers(n, r) = |r \cap n|/|r|$ . In the same vein, *coveredBy* is a function

$$coveredBy : n, r \rightarrow p \in [0, 1] \quad (1.5)$$

that tells how much each new resource is covered by the other older ones. Its value is the proportional overlap  $p = coveredBy(n, r) = |r \cap n|/|n|$ . For brevity of descriptions, we make the assumption that new resources do not overlap with other resources unless otherwise stated.

For example, the areal merger involved in the reunion of the two Germanies (Figure 1.1) can be represented by the bridge below:

$$t = 1991 \quad (1.6)$$

$$OLD = \{E.Germ., W.Germ.\} \quad (1.7)$$

$$NEW = \{Germ.\} \quad (1.8)$$

$$covers = \{\langle Germ., E.Germ. \rangle = 1, \langle Germ., W.Germ. \rangle = 1\} \quad (1.9)$$

$$coveredBy = \{\langle Germ., E.Germ. \rangle = 0.3, \langle Germ., W.Germ. \rangle = 0.7\} \quad (1.10)$$

The *covers*-values tell how much the new concept Germany covers the old resources, and the *coveredBy*-values tell how much the old concepts East and West Germany cover the new one. It follows from the absence of other *covers* and *coveredBy* values that East and West Germany exhaustively and exclusively cover the area of Germany. Since East and West Germany are in the list *OLD* their persistence is terminated at  $t$ , and since Germany is in the list *NEW*, a new resource whose persistence starts at  $t$  is added into the ontology. Notice that resources are never removed from the ontology, only their time span is updated.

## GENERATING A TIME SERIES BY BRIDGES

A set of bridges at a moment  $t$  can be used to introduce a new period ontology in the following way.

1. For each resource in *OLD*, the upper limit of the time interval is opened and set to  $t$ .
2. For each resource in *NEW*, the time interval  $[t, +\infty)$  is set and the resource is added into the ontology of the time series.
3. A new change point  $t$  is added into the time series ontology.

An initial time series ontology can be created by a set of simple bridges that introduce new concepts and their partonomy. After this, bridges for renaming, merging, and splitting areas can be used. After each change point  $t$ , a new period ontology is implicitly defined as the set of resources that persist after  $t$ .

During this evolutionary process, the global overlap relation between concepts could in principle be automatically constructed as a table in the following way. The rows  $i$  and columns  $j$  represent the resources in the order  $X_0, X_1, \dots, X_n$  in which these have been introduced. When a new  $n$ :th resource is introduced, the cells  $[i, n], i = 0 \dots n - 1$ , of the  $n$ :th row can be filled based on the *cover*-function of the corresponding bridge. The cell  $[n, n]$  has value 1. In the same vein, the cells  $[n, j], j = 0 \dots n - 1$ , of  $n$ :th column can be filled with the values of the *coveredBy*-function. This means that cells  $[i, j], j \leq i$ , will always



contain the global mapping  $coveredBy(X_i, X_j)$  and, in a symmetrical way, cells  $[i, j]$ ,  $i \leq j$ , will contain the values  $covers(X_i, X_j)$  for all resource pairs  $(X_i, X_j)$ .

By using such a table, the problem of determining proportional overlap of a resource  $X_i$  with respect to any other resource  $X_j$  can be solved by a simple table lookup at position  $[i, j]$  of the table. A major problem, however, remains: a set of bridges has to be defined which is easy to use from the human viewpoint and is yet sufficient for modeling the changes in the domain. Obviously, it is not usually feasible in practice to consider the global  $covers$ - and  $coveredBy$ -functions explicitly when creating bridges. Otherwise, for example, modeling a modern county border change in Italy would involve considering overlaps with areal concepts used during the times of the Roman Empire. What is needed is a mechanism that can infer such global overlaps over long periods of time based on local change descriptions that can be modeled easily by the humans. In the following we show how this can be accomplished.

## A METHOD FOR DETERMINING GLOBAL COVERINGS CHANGE BRIDGES FOR EXPRESSING CHANGED SITUATIONS

Our method, called ONTOFLUX, uses a set of change bridges to form mappings between concepts from different period ontologies. Each bridge type specifies a typical territorial change type. Our initial analysis of a database (Väättäinen, 2004) suggests that at least the following types are needed in practice: *addition* (a new region is formed), *usedtobe* (the name of a region is changed), *removal* (a region ceases to exist), *merged* (several distinct regions are merged into a new region), and *split* (a region is divided exhaustively into several distinct regions). We consider here the *merged* and *split* bridges whose interplay causes problems from the modeling and reasoning points of view.

The *merged* bridge is defined by

$$\langle t, \{old_1, \dots, old_n\}, \{new\}, covers, coveredBy \rangle \quad (1.11)$$

where  $\cap\{old_i\} = \emptyset$  and  $new = \cup\{old_i\}$ . The functions  $covers(new, r)$  and  $coveredBy(new, r)$  have a non-zero value for each older resource  $r$  that intersects with  $new$ .

The *split* bridge is defined symmetrically by

$$\langle t, \{old\}, \{new_1, \dots, new_n\}, covers, coveredBy \rangle \quad (1.12)$$

where  $\cap\{new_i\} = \emptyset$  and  $old = \cup\{new_i\}$ . The functions  $covers(new_i, r)$  and  $coveredBy(new_i, r)$  have a non-zero value for each older resource  $r$  that intersects with  $new_i$ .

In our implementation, these bridges are represented as instances of the *change bridge classes* in RDF(S) (Brickley and Guha, 2004) and are created using the Protégé-2000-editor<sup>4</sup>. For example, an instance of the *merged* bridge is depicted in Figure 1.3. The property *before* refers to the concepts before the change time point *hasTime*, and property *after* refers to the merged new concept after the change. The values for *covers*- and *coveredBy*-functions are determined by considering the areas involved.

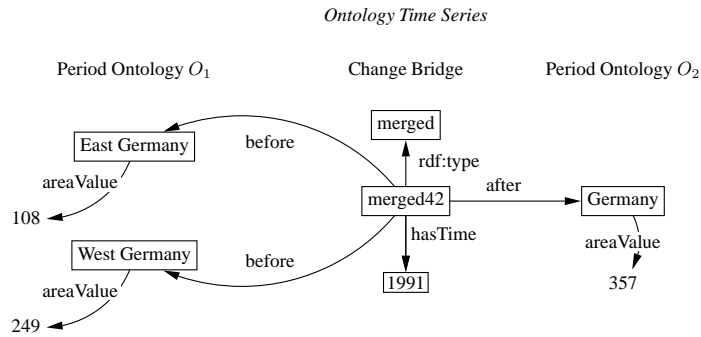


Figure 1.3 An RDF instance of the *merged*-bridge (Equation 1.11) from the change bridge ontology.

## BRIDGE CHAINS

Local bridges relating concepts of successive period ontologies form chains that span over wider time intervals. For example, Figure 1.4 represents some geographical regions at the eastern border of Finland during 1906-1989. These regions have been split and merged over the years, and in 1944 a new country border between Finland and the Soviet Union was established, which makes the ontological modeling of the regions even more challenging.

Figure 1.5 depicts the changes in the same region during the 20th century as a set of chained bridges in more detail. In the first split on the left, a part was separated from Viipuri in 1906 into a new county Nuijamaa that was later in 1944 divided into two halves: one for Finland and one (annexed) for the Soviet Union. The Finnish half was finally merged into Lappeenranta in 1989. The property *areaValue* tells the geographical size of the attached region in square kilometers. Notice that the concepts Viipuri (-1906) (upper left corner) and

Lappeenranta (1989-) (upper right corner) are related with each other through two different chains of bridges.

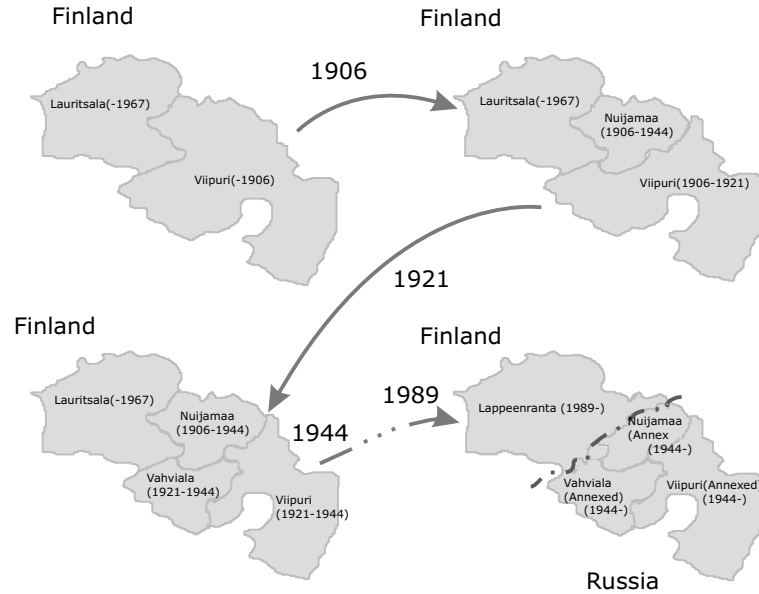


Figure 1.4 An example map that represents geographical regions of Lauritsala, Vahviala, Viipuri, Nuijamaa, and Lappeenranta in different periods of time. The small area inside Lauritsala (-1967) is Lappeenranta (-1967). The gray thicker dotted line represents the border between Finland and Russia (and the former Soviet Union).

Each local bridge can be defined easily but the global areal relations are not necessarily so obvious. For example, if a museum artifact  $x$  has been manufactured in the region of Viipuri (-1906), and the end-user is looking for material that has been manufactured within the area of modern Lappeenranta (1989-), how likely is it that the artifact  $x$  indeed is what the end-user is looking for? In the following we show, how the local bridges created at different instants of time can be used for deducing the global covering between arbitrary two concepts over a complete ontology time series.

The method has the following phases:

1. Local Bridges. Changes are modeled as instances of the bridge ontology.
2. Local Coverings. The bridges of the ontology time series, represented in RDF, are transformed into a form where the local covers- and coveredBy-functions are made explicit.
3. Global Coverings. Global overlaps are calculated by chaining local coverings and by considering different change paths between concepts.

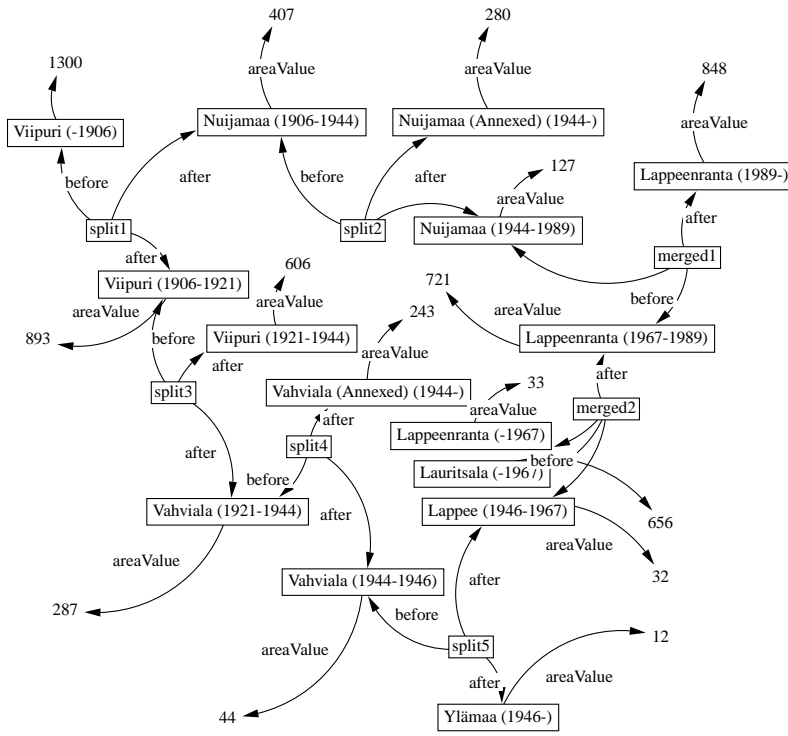


Figure 1.5 An example of chained change bridges. Each region is associated with a literal value for its area in square kilometers.

## CALCULATING LOCAL COVERINGS

The meaning of an ontological bridge, such as the one in Figure 1.3, is essentially defined in terms of the *covers*- and *coveredBy*-mappings. We therefore first transform the bridges into such mappings. A local mapping  $\text{covers}(a, b) = p$  can be represented in RDF by an instance of the class *Covers* with property values  $\text{hasCoverValue} = p$ ,  $\text{coverer} = a$ , and  $\text{coveree} = b$ ; a mapping  $\text{coveredBy}(a, b) = p$  is an instance of the class *CoveredBy* with property values  $\text{hasCoverValue} = p$ ,  $\text{coverer} = b$ , and  $\text{coveree} = a$ . Functions *covers* and *coveredBy* are inverse relations of each other.

For example, Figure 1.6 depicts the covering that can be generated from the upper part of the bridges of Figure 1.5 by using the semantic definitions of the

merged and split bridges. Dotted arcs *covers p* and *coveredBy p* are a shorthand notation for corresponding instances of classes *Covers* and *CoveredBy* with overlap value *p*. The value *p* is calculated based on the areal sizes of the concepts given in Figure 1.5. For example, since the size of Nuijamaa (1906-1944) is 407 and the size of Viipuri (-1906) is 1300, Nuijamaa covers Viipuri by value  $407/1300 = 0,3131$  and is coveredBy by Viipuri by value  $407/407 = 1$ .

The coverage graph, based on split and merged bridges, is always a directed acyclic graph (DAG), because these bridges always introduce new concepts in time by definition.

By traversing the *covers*-arcs, the coverage of a newer concept with respect to an older one (in terms of creation time) can be determined. By traversing *coveredBy*-arcs, the coverage of an older concept with respect to a newer one is determined. These coverage chains will be used as the basis for calculating global coverages.

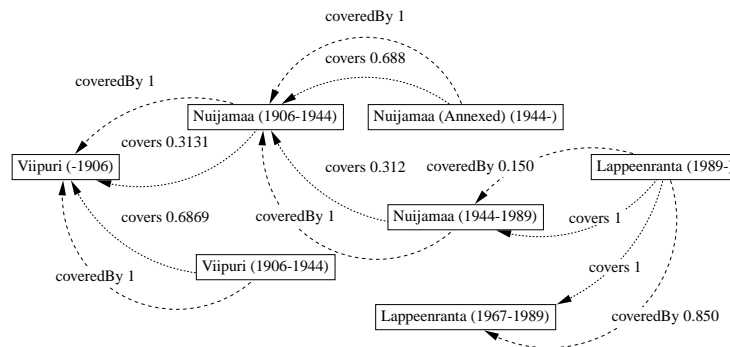


Figure 1.6 Local covers-mappings created from the upper part of the bridges of Figure 1.5. Dotted arcs *cover p* and *coveredBy p* tell quantitatively how much related concepts overlap, i.e., cover each other in terms of geographical area.

## CALCULATING GLOBAL COVERINGS

Calculating the coverings of the global overlap relation table can be done in two steps. First, the coverings of newer concepts with respect to older ones are calculated by traversing the local *covers* arcs and by accumulating coverings. Second, the covered-by relations from an older concept to a newer ones are computed in the same way by traversing *coveredBy*-arcs. Due to the similarity in the computation, we describe below only how to deal with the first case of traversing *covers* arcs.

The global cover value  $covers(A, B)$  between concepts  $A$  and  $B$  can be determined by first enumerating all possible paths (chains)  $CoveringChain_k$  of local coverings from  $A$  to  $B$ . The value for each chain is the product of the local *covers*-values of each edge:

$$CoveringChain(A, B)_i = \prod_{k=1}^n covers_k, i = 1 \dots n, \quad (1.13)$$

where  $i$  is an index for the covering paths between  $A$  and  $B$  and  $n$  is the number of *covers*-edges on the  $i$ :th path between  $A$  and  $B$ . Multiplication is possible because at each node the outgoing *covers*-arc tells how much the node covers the next one.

It is possible that the covering is accumulated through different paths. Then the global covering between  $A$  and  $B$  is accumulated by the different covering chains between  $A$  and  $B$ . Depending on the bridges used on the path, four different cases arise. First, if there are only *split* bridges on the path, then the global accumulated covering is simply the sum of path coverings:

$$GlobalCovering(A, B) = \sum_{i=1}^n CoveringChain(A, B)_i, \quad (1.14)$$

The formula is based on the observation that when a split is made, the parts are mutually exclusive, i.e., the parts do not overlap. Therefore, all concepts on alternative *covers*-paths between two nodes are mutually exclusive and their effect on global covering is purely additive.

Second, if there are only *merged* bridges on the path, then the same formula 1.14 is applicable by analogous reasons.

Third, it is easy to see that the formula is also applicable for paths in which a *split* is followed later by a *merged* bridge: alternative coverings remain purely additive.

The fourth remaining case, where a *merged* is followed later by a *split* is a bit more complicated. Figure 1.7 illustrates the situation. Here East Germany (EG) and West Germany (WG) are merged into Germany. Let us assume a hypothetical future, where Germany is split again into two parts, North Germany (NG) and South Germany (SG), whose sizes are the same as those of East and

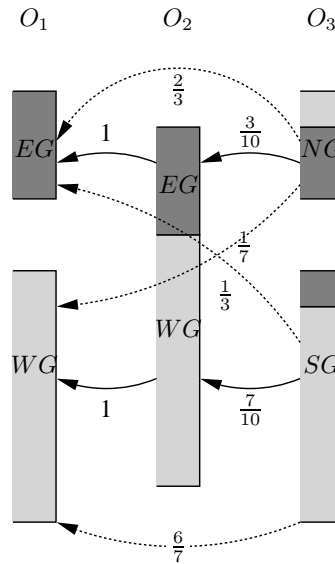


Figure 1.7 Resources  $EG$ ,  $WG$ ,  $G$ ,  $NG$  and  $SG$  of an ontology time series.

West Germany, respectively. If we multiply the local arcs using Formula 1.14, then we would get, e.g., the following overlap value:

$$\text{covers}(NG, EG) = 3/10 * 1 = 3/10$$

However, it is clear that this result is wrong: we cannot automatically say anything about the global coverings between South/North Germany and East/West Germany because there is no information telling us where the north-south boundary lays.

However, this information should have been provided directly by the user when (s)he created the *merged* bridge, because its definition demands that the functions *covers* and *coveredBy* are given. This information is illustrated in the figure by darker and lighter grey that indicate the areas of East and West Germany, how they are merged into Germany, and how they are redistributed further between North and South Germany. For example, roughly two thirds of East Germany will be within North Germany. This information is given by the following global *covers* functions:

$$\begin{aligned} \text{covers}(NG, EG) &= 2/3, \text{covers}(NG, WG) = 1/7 \\ \text{covers}(SG, EG) &= 1/3, \text{covers}(SG, WG) = 6/7 \end{aligned}$$

Therefore, the formula 1.14 for accumulating the global covering of paths in situations where a *merged* is followed later by a *split* must be modified by

the following additional condition: if there is a direct arc  $covers(A_1, A_2)$  on a path (from  $A$  to  $B$ ) given by a bridge (i.e., by the user), then any (indirect) path  $A_1, \dots, A_2$  should be omitted from the enumeration.

The same kind of situation arises, when a new areal concept  $A$  is introduced by an *addition* bridge, and other concepts intersect with  $A$ . This corresponds to a merger. If  $A$  or a part of  $A$  is later split, then the *covers* relation must be specified between the split parts and the original areas overlapping with  $A$ .

To sum up, the global overlap table can be created easily in most cases by enumerating covering paths and adding the effects. The problematic case is the *split* bridge, where local covering definitions are not always sufficient for determining global coverings. The user is obliged to specify some additional global covering values. This situation may occur if 1) the split area *old* intersects with an original region created by the *addition* bridge or 2) *old* is a part of a region created formerly by a *merged* bridge. Luckily, such regions can be found by inspecting the topology of the bridge DAG, and the potentially overlapping concepts can be pointed out to the bridge modeler. The user then has to consider the mappings *covers* and *coveredBy* only with respect to these problematic areal concepts.

## AN APPLICATION CASE

We are applying ONTOFLUX to build a Finnish Temporal Region Ontology (Suomen Ajallinen PaikkaOntologia, SAPO) based on a real dataset from (Väätainen, 2004). The dataset defines different areas of Finland and the changes that have occurred from the beginning of the 20th century. The descriptions of the dataset are not fully machine-understandable, and the idea is to change them into an ontology time series. A Perl script was written that created an initial RDF(S) ontology of the dataset and this ontology is being edited further by hand as a Protégé-2000 project. The bridges of Figure 1.5 are one part of the ontology being constructed.

Currently SAPO consists of 667 different regions in time, that is, Finnish counties that have existed during a period from the beginning of the 20th century until today. We have created the change bridge knowledge base of Figure 1.5 to test our method in determining global coverings of regions. An initial analysis of the dataset suggests that there will be in total 887 different change bridges (Table 1.7) between the regions, excluding the *addition* bridges that introduce new concepts.

The method of determining global coverings using the RDF(S) ontology has been implemented in Java with the help of the Jena library<sup>5</sup>. Table 1.1 depicting the global overlap table corresponding to Figure 1.5 was computed using this software. The x- and y-axis list the concepts in the order of their creation. Some concepts are not shown in the table in order to save space.



<i>change bridge</i>	<i>count</i>
Merged	302
Split	421
Usedtobe	164
all bridges	887

*Table 1.1* It is expected initially that 887 change bridges are needed to define the changes of 667 different temporal regions of Finland from the beginning of the 20th century until 2004.

The value in a cell  $[X, Y]$  tells the global value  $covers(X, Y)$  or, conversely, the global value  $coveredBy(Y, X)$ . For example, the current city of Lappeenranta (1989-) covers the area of historical Viipuri (-1906) by 0.12, i.e. 12%, and Lappeenranta (1989-) is covered by Viipuri (-1906) by 19%.

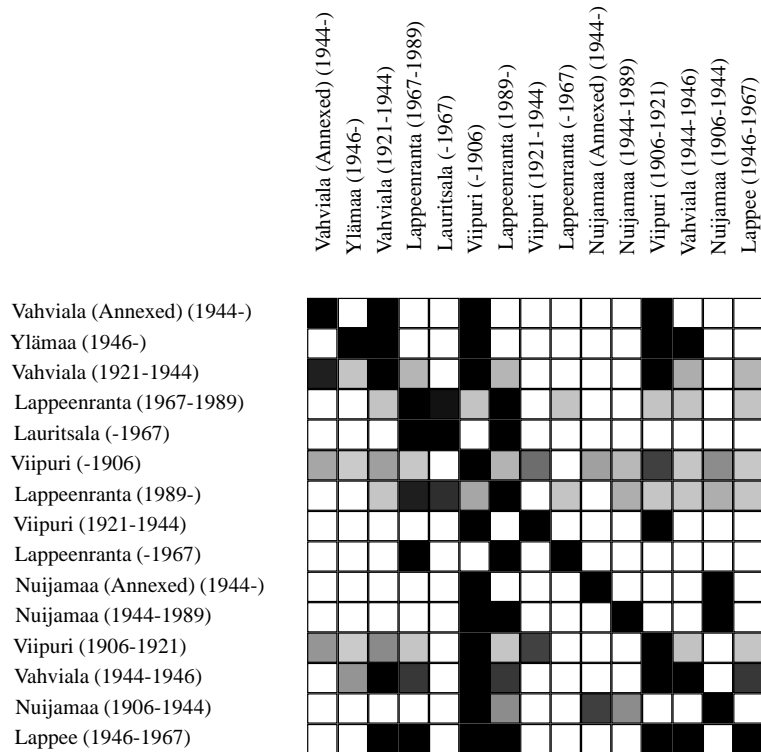
When querying a database with Lappeenranta (1989-), an object annotated with Viipuri (-1906) would match with this value with relevance value 12%—a result that many users could find a bit surprising due to the turbulent changes on the Finnish eastern border. A more obvious result in the table is that Lappeenranta (1989-) does not overlap with Viipuri (1921-144) at all (0%). In general, the table  $covers(X, Y)$  tells, what annotations  $Y$  match the query  $X$  and the  $covers(X, Y)$  values tell their order of relevance.

$covers(X, Y)$	Viipuri (-1906)	Nuijamaa (1906-1944)	Viipuri (1906-1921)	Vahviala (1921-1944)	Viipuri (1921-1944)	Lappeenranta (1967-1989)	Lappeenranta (1989-)
Viipuri (-1906)	1	0.31	0.69	0.22	0.47	0.025	0.12
Nuijamaa (1906-1944)	1	1	0	0	0	0	0.31
Viipuri (1906-1921)	1	0	1	0.32	0.68	0.036	0.036
Vahviala (1921-1944)	1	0	1	1	0	0.11	0.11
Viipuri (1921-1944)	1	0	1	0	1	0	0
Lappeenranta (1967-1989)	0.044	0	0.044	0.044	0	1	1
Lappeenranta (1989-)	0.19	0.15	0.038	0.038	0	0.85	1

*Table 1.2* Table describing some of the global coverages between ontology concepts illustrated in Figures 1.5 and 1.6.

The complete global overlap table of the concepts in Figure 1.5 is visualized in Figure 1.8. Here the black color indicates a full 100% coverage between the temporal regions and the white color a 0% coverage, accordingly. Different shades of grey indicate the level of coverage: the darker the box, the higher is the coverage. From this illustration it is easy to see the mutual asymmetric

coverages between the regions, and that the overlapping relation in this case is fairly complicated.



*Figure 1.8* All the coverages visualized using colored boxes. The black color indicates a full 100% coverage between the temporal regions and the white colour a 0% coverage, accordingly. Different shades of grey indicate the level of coverage between regions: darker the box, the higher is the coverage between the regions.

## DISCUSSION

This chapter formulated the problem of reasoning over a time series of evolving ontologies. The focus was on modeling partial overlap between concepts that change over long periods of time. The domain of application was information retrieval using spatiotemporal reasoning.

A method, ONTOFLUX, was presented for computing a consistent set of global covering relations between overlapping resources in an ontology time series. The method was based on a bridge ontology that was used for expressing local ontological changes in a declarative way. In most cases global coverings could be computed automatically based on local coverings that are easy to determine for the human modeler. When using the *split* bridge, some global coverings may need to be specified by the human user, but the concepts involved can be detected based on the topology of the covering bridge graph and be pointed out for the user.

We think that the idea of storing all ontological resources in one time series ontology that is evolved by adding local change bridges is economical from the ontology management point of view and helps the modeler's work. The period ontology corresponding to a particular moment can be explicated from the implicit ontology series description easily.

## RELATED WORK

The problem of modeling change in partonomy time series has not been discussed much in the literature, although there is lots of research going on related to ontology versioning (Klein, 2004; Stojanovic, 2004) and spatiotemporal ontologies (Sider, 2001; Stell and West, 2004; West, 2003; Grenon and Smith, 2004). In GIS systems, overlap of physical areas is usually determined by representing the real world in terms of intersecting polygons (Visser, 2004; Stuckenschmidt and Harmelen, 2004). However, in application cases like ours, such geometrical modeling may not be feasible because precise geometrical information is not available or it could be difficult to create and computationally difficult to use. Local change bridges could be expressed more easily, be used for deriving the global covering information needed in the application, and for presenting the ontologies at different times. An additional benefit of dealing with change mappings is that this notion is more general than that of areal two-dimensional overlap. This suggest that the same approach could perhaps also be used in other more complex application domains dealing with other forms of conceptual overlap, e.g., by using more than two dimensions.

Spatial reasoning is a research field of its own (Cohn et al., 1997). Approaches such as RCC-8 (Randell et al., 1992), however, typically deal with qualitative reasoning and provide relations like  $PO(x,y)$ , i.e.,  $x$  partially overlaps  $y$ , without quantitative values. Our approach, in contrast, is quantitative. We are interested in the amount of overlap between  $x$  and  $y$ . These values are provided by the functions  $covers(x,y)$  and  $coveredBy(x,y)$ .

## FURTHER RESEARCH

The work of modeling the Finnish SAPO ontology has just started and is continued based on the initial promising results presented in this chapter. We plan to further examine the problematic combinations of merge and split bridges, and how the human modeler can be supported in specifying the covering relations of the bridges as easily as possible.

We believe that there are many useful ways to use partonomical coverage data in, e.g., information retrieval and in visualizing partonomical ontologies and their changes. At the moment, an ontology browser for ontology time series is being implemented in order to visualize the changes through time, and for using the ontologies in actual annotation work.

## Acknowledgments

Discussions with Mirva Salminen, Jari Väättäin, Arttu Valo, Eetu Mäkelä, Markus Holi, Miikka Junnila, and Eero Carlson are acknowledged. Our research was funded mainly by the National Technology Agency Tekes.

## Notes

1. <http://museosuomi.cs.helsinki.fi>
2. This partonomy is a part-of hierarchy of individuals of the classes Continent, Country, County, City, Village, Farm etc.
3. The idea of view-based search in general is discussed e.g. in (Hearst et al., 2002).
4. <http://protege.stanford.edu>
5. <http://www.hpl.hp.com/news/2004/jan-mar/jena2.1.html>

## References

- Brickley, D. and Guha, R. V. (2004). RDF Vocabulary Description Language 1.0: RDF Schema W3C Recommendation 10 February 2004. Recommendation, World Wide Web Consortium.
- Cohn, A., Bennett, B., Gooday, J., and Gotts, N. (1997). Representing and reasoning with qualitative spatial relations about regions. In Stock, O., editor, *Temporal and spatial reasoning*. Kluwer.
- Fensel, D. (2004). *Ontologies: Silver bullet for knowledge management and electronic commerce (2nd Edition)*. Springer-Verlag.
- Grenon, P. and Smith, B. (2004). SNAP and SPAN: Prolegomenon to geodynamic ontology. *Spatial Cognition and Computation*, (1):69–105.
- Hearst, M., Elliott, A., English, J., Sinha, R., Swearingen, K., and Lee, K.-P. (2002). Finding the flow in web site search. *CACM*, 45(9):42–49.

- Holi, M. and Hyvönen, E. (2004). A method for modeling uncertainty in semantic web taxonomies. In *Proceedings of WWW2004, Alternate Track Papers and Posters*, New York, USA.
- Hyvönen, E., Junnila, M., Kettula, S., Mäkelä, E., Saarela, S., Salminen, M., Syreeni, A., Valo, A., and Viljanen, K. (2004). Finnish Museums on the Semantic Web. User's perspective on museumfinland. In *Selected Papers from an International Conference Museums and the Web 2004 (MW2004)*, Arlington, Virginia, USA.
- Klein, M. (2004). *Change Management for Distributed Ontologies*. PhD thesis, Vrije Universiteit Amsterdam.
- Klein, M. and Fensel, D. (2001). Ontology versioning on the Semantic Web. In *Proceedings of the International Semantic Web Working Symposium (SWWS)*, pages 75–91, Stanford University, California, USA.
- Noy, N. and Klein, M. (2003). Ontology evolution: Not the same as schema evolution. *Knowledge and Information Systems 5*.
- Randell, D. A., Cui, Z., and Cohn, A. (1992). A spatial logic based on regions and connection. In Nebel, B., Rich, C., and Swartout, W., editors, *KR'92. Principles of Knowledge Representation and Reasoning: Proceedings of the Third International Conference*, pages 165–176. Morgan Kaufmann, San Mateo, California.
- Sider, T. (2001). *Four-Dimensionalism. An Ontology of Persistence and Time*. Clarendon Press, Oxford.
- Stell, J. G. and West, M. (2004). A 4-dimensionalist mereotopology. In Varzi, A. and Vieu, L., editors, *Formal Ontology in Information Systems*, pages 261–272. IOS Press.
- Stojanovic, L. (2004). *Methods and Tools for Ontology Evolution*. PhD thesis, University of Karlsruhe, Germany.
- Stuckenschmidt, H. and Harmelen, F. V. (2004). *Information Sharing on the Semantic Web*. Springer-Verlag, Berlin Heidelberg, New York.
- Visser, U. (2004). *Intelligent information integration for the Semantic Web*. Springer-Verlag, Berlin Heidelberg, New York.
- Väätäinen, J. (2004). A database containing descriptions of changes of counties in Finland. The Geological Survey of Finland (GSF), Espoo, Finland.
- West, M. (2003). Replaceable parts: A four dimensional analysis. In *COSIT-03 - Workshop on fundamental issues in spatial and geographic ontologies*, Ittingen, Switzerland.