Jul 1st, 12:00 AM

Filtering of Semantically Enriched Environmental Time Series

B. Bozica

J. Peters-Andersa

Gerald Schimak

Filtering of Semantically Enriched Environmental Time Series

B. Božić[a]**, J. Peters-Anders**[a]**, G. Schimak**[a]

[a]*Austrian Institute of Technology, Donau-City-Straße 1, 1220 Vienna, Austria*
(bojan.bozic@ait.ac.at, jan.peters-anders@ait.ac.at, gerald.schimak@ait.ac.at)

Abstract: In current time series processing there is a growing urge to add meaning to the information being processed in order to make it more readable for users as well as understandable/interpretable by machines. Especially in the environmental field of computer science, Semantic Web technologies have become more and more popular. Where we had simple time series data, consisting of time stamps an according values in the past, we are now dealing with complex time series slots consisting of not only time and information, but also a lot of additional meta-information.

Therefore, we developed a method for semantic filtering of time series meta-information and annotations, as well as methods to visualize the results to the user and show how this meta-information is composed. Our result is a Visualization and Filtering Component, which is able to process annotations similar to simple time series, but filters tags based on users, timing and geo-location. The benefits of this component are the ability to use semantic for filtering of annotations, the visualization/representation of annotations in a time series manner and furthermore the possibility to apply time series processing calculations on semantic annotations.

Within the TaToo project and as one of several show cases, we have implemented this method into an application called Climate Twins. Climate Twins demonstrates possible future changes of climate conditions at a regional level. This is done by indicating all regions, which have climate conditions today that are similar to future projected conditions in a certain region of interest. This similarity is based on the climate parameters 'temperature' and 'precipitation'. To be able to filter and visualize annotations on the Climate Twins application as time series, we integrated the Visualization and Filtering Component in the Climate Twins application.

Keywords: Time Series; Semantic Web; Climate Twins; Environment

1 INTRODUCTION

Time Series Processing has a long tradition in the fields of mathematics and computer science. Almost everything in terms of data can be represented as time series. Therefore, the logical consequence is to take Time Series Processing to the next level. This paper describes how Time Series Processing corresponds to the modern and popular Semantic Web technologies and how these technologies can be used to improve Time Series Processing. Our main focus is filtering with the help of semantically enriched time series and the preparation of the results for visualization in certain applications. Chapter 2 describes the Climate Twins application, chapter 3 the concepts and ideas behind semantically enriched time series, chapter 4 visualization and filtering, and the last chapter

gives an overview of the conclusions for the disciplines of time series processing and Semantic Web.

2 CLIMATE TWINS APPLICATION

To allow real world insights about future climate impact and appropriate adaptation, one can look at model regions, where the current climate appears to be similar to an expected future climate of a Point-Of-Interest (Ungar et al. [2011]). We call such region pairs with similar climate conditions (at different times) Climate Twins. From these (remote) current Climate Twin region parts we can learn how future climate impacts may be experienced in the POI (Point-of-Interest) and how to adapt there to the changing climate conditions, expected in the future. The idea of Climate Twins is to identify regions whose current climate conditions show high similarity to the expected future climate in the POI. The Climate Twins search tool is a web-based graphical user interface (GUI) allowing exploring climate change effects based on maps of current and future climate. To identify climatological similarity seems to be a simple exercise but the accuracy and validity of the result strongly depends on the indicators used and the similarity thresholds defined. A huge number of indicators in combination with narrow threshold ranges will reduce the number of matching regions significantly as well as few indicators combined with wide thresholds will show a big number of matching regions. The climate indicators used here are daily mean temperatures and daily precipitation because they are seen as the most important ones and provide sufficient input for proving the concepts applicability.

The most important part was to find a suitable matching method which strongly depends on the quantification of similarity between any two data vectors. This matching method now provides a unit-less similarity value able to be combined with similarity values of other indicators, information of the degree of similarity to derive statements like "more similar than" or "less similar than", and a consideration of many statistical properties because whole statistical distributions are being compared. Figure 1 shows a screenshot of the Climate Twins application. The left map shows the source of the calculation, which is a location in the Czech Republic. The red-marked region is the twin location for the calculation. The map on the right side shows all twin regions which have a similar climate compared to the source region in a different time. Finally, the lower frame enables the user to set all required parameters for the calculation.

2.1 The Climate Twins TaToo Use Cases

The Climate Twins application is fully functional and can be freely accessed. Nevertheless, there was room for improving its functionality and accessibility. For this purpose TaToo Public Services have been integrated regarding the following four Use Cases (UC 1 - 4):

- UC 1: Make it easier for interested people from different fields of science, from politics, public authorities and business to find the Climate Twins application.

- UC 2: Give the users a possibility to tag the Climate Twins application and its community-added resources for other users.

- UC 3: Give the users the opportunity to find additional information beyond the Climate Twins data on their specific regions of interest.

- UC 4: Give them the possibility to add such additional information by themselves for the sake of other users.
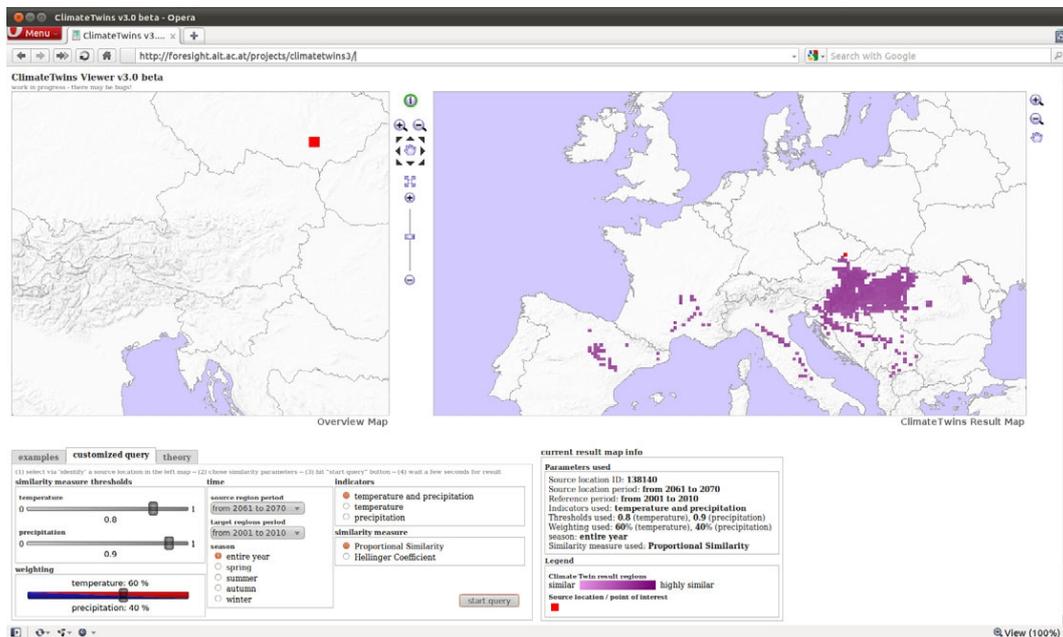
Figure 1: Screenshot of the Climate Twins Application (Ungar et al. [2011]).

Figure 2: Tagging Service Integration.

In this paper we will have a look at Use Case 4.

Use Case 4: Add additional information on a particular Climate Twin region. The benefit of TaToo lies in the support of the extension of the basic information from the Climate Twins model (data on temperature and precipitation) by the users of the Climate Twins application themselves for the sake of other users. TaToo enables the upload of geo-located additional climate-related information. The primary beneficiaries are the users of the Climate Twins application. But also the Climate Twins resource provider benefits from the additional geo-located resources added to his temperature and precipitation database.

Before being able to find additional climate-related information for a particular region other users have to add such resources. Accordingly, TaToo enables also the upload of geo-located resources. Figure 2 shows the integration of the TaToo Tagging Service in the Climate Twins application.

The user locates her POI in the Climate Twins GUI. For this POI she knows interesting
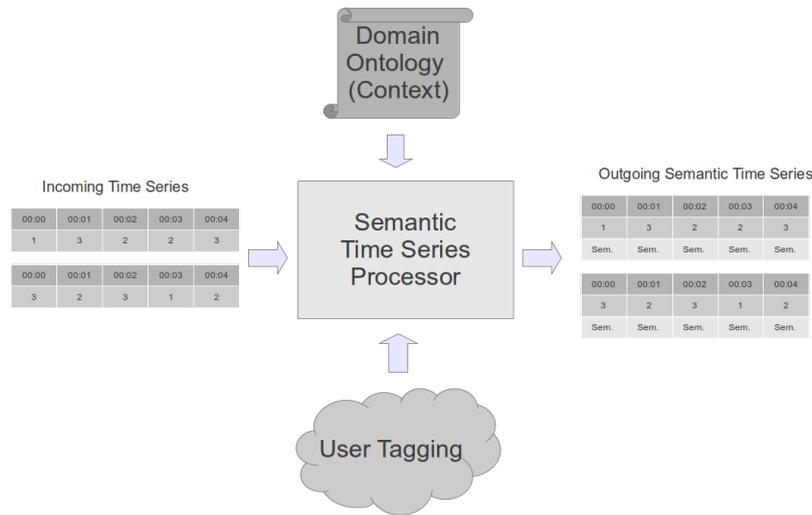
Figure 3: The process of semantic-enrichment of time series data.

additional material which she wants to share with other users. Now the user uploads the URL to this additional resource and adds meta-information to it (e.g. intended audience). The tag is stored as a URI in the TaToo Knowledge Base and will be presented to any other user who will access this respective POI in the future.

3 SEMANTICALLY ENRICHED TIME SERIES

To improve the capabilities of visualization and filtering in Time Series Processing, we introduce the semantic time series approach (Božić [2011]). Our idea is to enrich time series with meta-information, integrate ontologies into processing, and hence make connections between data and meta-data. This process makes it possible to dynamically change the process of Time Series Processing from general to domain-specific by adding an ontology, or to switch the domain by replacing an ontology, during run-time.

The core component for this task is a semantic-enabled Time Series Processor, which uses domain-specific ontologies as input and enriches data of time series with its meaning. As a consequence, the Time Series Processor provides domain-specific calculations and therefore prepares time series for visualization in a semantic context, but also filters time series in a form that is more meaningful to a user.

Figure 3 demonstrates how a semantic Time Series Processor contributes to the process of meta-information enrichment of time series and improves transformation and processing, while preparing time series to be visualized and enabling semantic filtering.

The central component in semantically-enriched Time Series Processing is the semantic Time Series Processor, as shown in figure 3. Much like traditional Time Series Processors it has a certain number of time series as input and provides the result of its processing in form of output (or result) time series. The big difference to traditional time series processing is the transformation of incoming time series to outgoing time series.

To be precise, the semantic Time Series Processor not only provides means of calculation, aggregation, and transformation of time series regardless of the environment and the context they are processed in. Moreover, it is able to take into account for the tagging of users, which are interested in time series and are willing to add addional information to them (e.g. to validate and rank them), but also the domain which has a specific ontology, and for which time series are being processed. This allows us to have not only transformed, but meaningful time series, which can be interpreted in a more confortable way and without further transformation in the specific domain, as an output.

The additional Semantics (or meta-data) in time series can be:

- Units of slot values (e.g. °C for temperature),

- Regulations for slots (e.g. threshold of temperature for heat warnings),

- Meaning (°C signalizes temperature).

Listing 1 shows a simple example of an ontology, which could be used by a semantic Time Series Processor to add a context to the processing.

```
@prefix owl:    <http://www.w3.org/2002/07/owl#> .
@prefix tatoo:  <http://www.tatoo-fp7.eu/environmental-context.owl#> .
@prefix rdf:    <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .

tatoo:isInterestedIn    rdf:type        owl:ObjectProperty .
tatoo:InterestedParty   rdf:type        owl:Class .
tatoo:TimeSeries        rdf:type        owl:Class .

tatoo:EnvironmentalAgency   rdf:type             tatoo:InterestedParty ;
                    tatoo:isInterestedIn tatoo:AirQualityTimeSeries .

tatoo:AirQualityTimeSeries      rdf:type            tatoo:TimeSeries .
tatoo:WaterQualityTimeSeries    rdf:type            tatoo:TimeSeries .

tatoo:EnvironmentalActivist     rdf:type            tatoo:InterestedParty ;
                tatoo:isInterestedIn tatoo:WaterQualityTimeSeries ,
                                     tatoo:AirQualityTimeSeries .

tatoo:Company       rdf:type                tatoo:InterestedParty ;
                tatoo:isInterestedIn        tatoo:WaterQualityTimeSeries .

tatoo:EnvironmentalResearch rdf:type            tatoo:InterestedParty ;
                    tatoo:isInterestedIn tatoo:AirQualityTimeSeries ,
                                         tatoo:FinancialTimeSeries .
```

Listing 1: Example ontology for semantic processing in Turtle code.

This ontology tells the semantic Time Series Processor which type of time series it deals with and who is interested in the results. Therefore, the processor is able to interpret the time series in a way they are important for the domain (e.g. if the time series is of type *tatoo:AirQualityTimeSeries* the processor knows that the values describe the quality of air). Of course, there have to be more precise definitions of these classes *TimeSeries*, *InterestedGroup*, etc. to define in what kind of values a company is interested in or what kind of values a water quality time series is able to describe and so on, but the main point is that such information is defined in the ontology and can be loaded and changed dynamically, so that the processor does not have to change its behavior when processing time series for different domains.

Apart from ontologies, tagging is the other important functionality in semantic Time Series Processing. Our approach is that users are able to tag time series and thus add new

instances to the domain specific ontologies. This means that users are able to add new RDF (Manola and Miller [2004]) triples and enhance meta-information of a time series. For example a user knows an additional interested party for a time series and she decides to add this information through a web interface. So, she selects "interested party" as a class and types the name of it (e.g. "University"). As a next step, the system would create the RDF triple *tatoo:University rdf:type tatoo:InterestedParty* and add it to our semantic repository. Thereby, we have two different kinds of meta-information enrichment, one by a specific domain through the ontology and one by a user through tagging.

The most important part of tagging is not always the additional meta-information from a user, but the possibility for users to evaluate time series. This is, in most of the cases, much more valuable to a system. Evaluation or ranking can be done much the same way, just by adding another kind of triples, for example something like the following: *tatoo:AirQualityTimeSeries tatoo:hasRank "5"*, which can indicate that the quality of this particular time series is ranked with 5 stars by the user who generated this triple.

4 VISUALIZATION AND FILTERING

The Visualization and Filtering Component is a special kind of a Tagging Processor and is responsible for the visualization of tags to users and filtering by users (Dihé [2011]). The Visualization and Filtering Component can be subdivided into two subcomponents:

- The Tag Visualizer: Visualizes tags associated with a resource in a way selected by the user e.g. tag cloud, tree, etc. grouping tags, highlighting certain tags, etc.

- The Tag Filter: Filters tags by specific tag values.

The specification process in service-oriented architectures often stops at the service interface level. The TaToo Architecture goes one step further by specifying the functionalities of client components and underlying business logic components in addition to the publicly visible functionality of the system exposed through the interfaces and operations of services. Therefore, the TaToo Framework Architecture is designed as an n-Tier architecture which comprises the following 4 Tiers:

1. The Presentation Tier which is concerned with user interaction and the presentation and aggregation of information.

2. The Service Tier decoupling Business Tier and Presentation Tier as well as serving as a layer to enforce interoperability.

3. The Business Tier which is responsible for the core functionality (business logic) of the TaToo System.

4. The Data Tier which is concerned with the storage of semantically enriched information, and other data (registered resources to be harvested, user information, etc.).

Figure 4 shows the Core Components Building Block, which provides the business logic of the TaToo system.
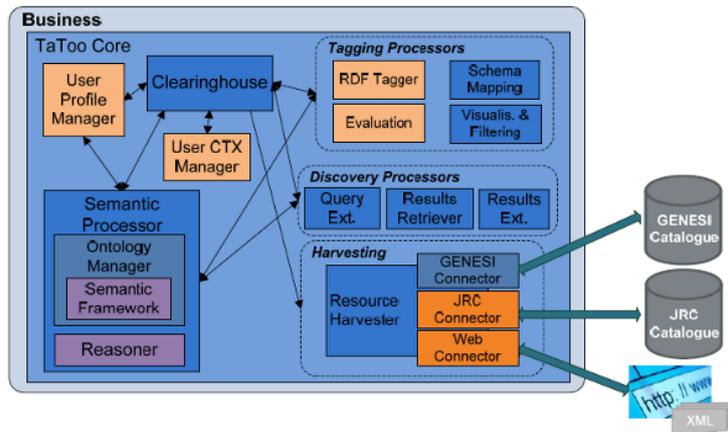
The main components of this Building Block are:

Figure 4: TaToo Core Components Building Block (Dihé [2011]).

- Clearinghouse: Is the central entry point to the business tier from the service tier. It is responsible for the coordination and communication of all the other components in the business tier.

- Semantic Processor: Retrieves ontologies on the basis of user-defined properties like domain or context, stores and retrieves semantic annotations for resources, and searches for semantic annotations by providing a SPARQL (Prud'hommeaux and Seaborne [2008]) query interface.

- Ontology Manager: Manipulates RDF, supports SPARQL queries, and supports other functionalities to access and manage semantic data.

- Reasoner: Infers new knowledge from available RDF triples and ontologies, and checks for incongruent information while managing ontologies.

- Tagging Processor: Provides tagging functionality to users.

- Discovery Processor: Provides discovery of semantic resources.

- Resource Harvester: Retrieves external meta-information on resources (i.e. SOAP services (Box et al. [2000]), REST services (Booth et al. [2004]), etc.) and web sites.

The Visualization and Filtering Component can be invoked by the Tagging Service through the Clearinghouse to perform visualization and filtering. It depends on the TaToo Semantic Processor to retrieve the meta-information (tags in RDF-Format) from the Knowledge Base.

As a first step, visualization and filtering functionality is needed to extend time series processing in a semantic way. The Visualization and Filtering Component can be seen as a kind of Tagging Processor. Visualization can be understood as the visualization of tags (e.g. a tag cloud) of resources selected by a user, but also grouping tags, highlighting certain tags, etc. The filtering functionality filters tags by specific tag values, such as user name, tagged resource, time, location, etc.

The objective of visualization and filtering is to visualize tags to the user and perform filtering operations requested by the user. The Time Series Processor with visualization and filtering functionality takes metainformation (e.g. an RDF document) in XML format and user input to perform certain operations.

5 CONCLUSIONS

The conclusions of our work on visualization and filtering of semantically enriched time series can be subdivided into three different categories. The first category are the improvements our work can contribute to Time Series Processing in general, the second category are contributions to the field of Semantic Web, and the third category is the contribution to the TaToo project and thus improvements of TaToo's Validation Scenarios (especially of the Climate Twins application, which is discussed in this paper).

The improvements for time series processing are quite obvious, as shown in chapter 2. The idea is to enrich time series through ontologies by experts on the one hand, and tagging from general users on the other. Therefore, the improvement is that a lot of processing time can be saved, because of information structuring, adding context for a domain, and user tagging.

The contributions to the field of Semantic Web are a new approach for data representation, namely as time series. The advantage is that data can be processed sequentially and that a lot of methods from Time Series Processing can be transfered to and used by Semantic Web technologies.

In the case of the Validation Scenario Climate Twins, having the possibility to discover additional information on a particular POI increases the use-value of Climate Twins significantly, because regional climate conditions are not only characterized by many more parameters than temperature and precipitation alone, but also due to the almost openended number of topics which are related to climate issues.

REFERENCES

Booth, D., H. Haas, F. McCabe, E. Newcomer, M. Champion, C. Ferris, and D. Orchard. *Web Services Architecture*. W3C, http://www.w3.org/TR/ws-arch/, 2004.

Box, D., D. Ehnebuske, G. Kakivaya, A. Layman, N. Mendelsohn, H. F. Nielsen, S. Thatte, and D. Winer. *Simple Object Access Protocol (SOAP) 1.1*. W3C, http://www.w3.org/TR/2000/NOTE-SOAP-20000508/, 2000.

Božić, B. Simulation and modelling of semantically enriched time series. In *MODSIM 2011*, Perth (AU), 12/2011 2011. TaToo Project, TaToo Project.

Dihé, P. D3.1.2 semantic service environment and framework architecture - v2. page 155, 08/2011 2011.

Manola, F. and E. Miller. *RDF Primer*. W3C, http://www.w3.org/TR/2004/REC-rdf-primer-20040210/, 2004.

Prud'hommeaux, E. and A. Seaborne. *SPARQL Query Language for RDF*. W3C, http://www.w3.org/TR/rdf-sparql-query/, 2008.

Ungar, J., J. Peters-Anders, and W. Loibl. Climate twins - an attempt to quantify climatological similarities. In *ISESS 2011*, Brno (CZ), 06/2011 2011. TaToo Project, TaToo Project.