

# A Linked-Data-driven and Semantically-enabled Journal Portal for Scientometrics

Yingjie Hu<sup>1</sup>, Krzysztof Janowicz<sup>1</sup>, Grant McKenzie<sup>1</sup>, Kunal Sengupta<sup>2</sup>, and Pascal Hitzler<sup>2</sup>

<sup>1</sup> University of California, Santa Barbara, CA, USA

<sup>2</sup> Wright State University, Dayton, OH, USA

**Abstract.** The Semantic Web journal by IOS Press follows a unique open and transparent process during which each submitted manuscript is available online together with the full history of its successive decision statuses, assigned editors, solicited and voluntary reviewers, their full text reviews, and in many cases also the authors' response letters. Combined with a highly-customized, Drupal-based journal management system, this provides the journal with semantically rich manuscript time lines and networked data about authors, reviewers, and editors. These data are now exposed using a SPARQL endpoint, an extended Bibo ontology, and a modular Linked Data portal that provides interactive scientometrics based on established and new analysis methods. The portal can be customized for other journals as well.

## 1 Introduction and Motivation

Linked Data is a paradigm for encoding, publishing, and interlinking structured data in a human and machine readable way to support the use and synthesis of these data outside of their original creation context. Almost all Linked Data available today come from two sources. Either they are converted from other forms of structured data such as databases and CSV files, or they are extracted from semi-structured and non-structured sources such as Web pages. While LinkedGeoData [1] is a typical example for the first case and is based on structured content from OpenStreetMap, DBpedia [2] represents the second case and is based on Wikipedia (originally its tabular content). Today most Web pages are dynamically created using content management systems that render content stored in a database using HTML and CSS. The Web presence of the Semantic Web journal (SWJ) is such a page<sup>1</sup>.

The journal started in 2010 and is published and printed by IOS Press. The journal's subtitle *Interoperability, Usability, Applicability* reflects its broad coverage of Semantic Web and Linked Data related research ranging from theoretical work on description logics and reasoning to applications in various domains such as the geo-sciences or digital humanities, and also includes topics from human-computer interaction, e.g., semantically-enabled user interfaces, as well as cognitive science research. More importantly, however, the journal has a unique

---

<sup>1</sup> <http://www.semantic-web-journal.net/>

feature in terms of the types of submissions it accepts and especially its review and editorial process. Besides classical research papers, surveys, application reports, as well as tools and systems papers, the journal also accepts descriptions of ontologies, and since 2012 also (Linked) Dataset descriptions. The SWJ review process is open and transparent, i.e., all submitted manuscripts are published on the journal's Web page together with the name of the responsible editor, the names of solicited reviewers, their full text reviews, as well as the decision status of the manuscript. Author's are also invited to post their reply letters publicly and any community member can contribute a volunteered review. Solicited reviewers can decide to have their name anonymized but only a minority makes use of this possibility. While we do not discuss the motivation for setting up this process and our experience running it here, details can be found in a recent publication [3].

More interesting for the Semantic Web community is the fact that this process creates a rich dataset that goes far beyond the bibliographic data provided by publishers. Essentially, it creates a detailed time line for each paper which contains structured and non-structured data. This provides us with new opportunities for scientometrics to evaluate and analyze scientific works, explore the network of authors, reviewers, and editors, and try to predict future research topic trends. Unfortunately, however, most of this information could not be exploited so far as it was stored in a closed-source journal management system (msTracker). Recently, the Semantic Web journal developed several extensions and modifications to the Drupal content management system, to merge its Web presence with the journal management process [4]. This new system stores all relevant data in a relational database with normalized database schema and thus allows to *triplify* these data; see figure 1.

In this paper, we publish this semantically rich dataset using the principles of Linked Data and serve those data through a SPARQL endpoint, a Pubby interface<sup>2</sup>, and a semantically-enabled Web portal. To capture the submitting-revising-and-resubmitting process typically experienced by most authors and their papers, we had to extend the bibliographic ontology BIBO [5], to be able to encode the different versions of a paper and the corresponding review comments for these versions in the paper's time line. We also created external links to *DBpedia* and the *Semantic Web Dog Food*<sup>3</sup> to enrich our local dataset, as well as to embed the SWJ dataset into the global Linked Open Data graph.

Based on the published SWJ dataset, a modular Linked Data portal has been designed and implemented to provide analysis functionality for scientometrics. A novel user interface has been designed as a middle layer to assists users who are unfamiliar with SPARQL and Linked Data but want to explore and query the dataset. Additionally, the SPARQL endpoint is also exposed to the public, and therefore data engineers can directly interact with the endpoint and integrate the data in their projects. To incorporate unstructured data such as reviews, abstracts, and the full paper content, we used Latent Dirichlet allocation (LDA)

---

<sup>2</sup> <http://wifo5-03.informatik.uni-mannheim.de/pubby/>

<sup>3</sup> <http://data.semanticweb.org/>

## Linked Brazilian Amazon Rainforest Data

Submitted by Tomi Kauppinen on 12/28/2012 - 06:19

Tracking #: 417-1539

**Paper Data**

**Authors:**  
Tomi Kauppinen, Giovana Mira de Espindola, Jim Jones, Alber Sánchez, Benedikt Gräler, Thomas Bartoschek

**Previous Version:**  
[Linked Brazilian Amazon Rainforest Data](#)

**Decision/Status:**  
Accept

**Responsible editor:**  
[Pascal Hitzler](#)

**Abstract:**  
The Linked Brazilian Amazon Rainforest Data contains observations about deforestation of rainforests and related things such as rivers, road networks, population, amount of cattle, and market prices of agricultural products. The Linked Data approach offers thus to combine ecological, economical and social dimensions together. Our aim has been to 1) dramatically shorten the time needed to collect information for a research setting concerning the Brazilian Amazon, and 2) via the linkage between datasets enable novel types of transdisciplinary research for the scientific community.

**Submission type:**  
Dataset Description

**Full PDF Version:**  
[swj417.pdf](#)

---

**Solicited Reviews:**  
▼ Click to Expand/Collapse

**Review #1**

By Dave Kolas submitted on 28/Jan/2013

**Suggestion:**  
Accept

**Review Comment:**  
This paper presents a dataset related to Amazon deforestation. This version of the paper improves clarity in the dataset descriptions. Particularly interesting are the descriptions of how the spatial relations between peices of data are used to merge

Fig. 1: A paper-page from the new Semantic Web journal management system.

[6] for topic modeling. It is used for simple word-cloud like paper topic visualizations as well as for more complex trending analysis. As all reviews, papers, and revisions have a time-stamp, we computed LDA topics per period (usually 3 months) to implement research topic trending modules. Descriptive statistics for each author, such as the number of papers published and the number of papers reviewed, are also included in this portal, as are information about the author's network. While the portal has been designed as a combination of a number of interactive analysis modules, all modules are implemented following the loose coupling principle, i.e., they can be easily separated, recombined, and integrated into other systems. This makes the Linked-Data-driven, semantically-enabled journal portal flexible and reusable. In fact, it will be adopted by a number of IOS Press journals in the near future.

The remainder of the paper is structured as follows. In section 2, we introduce the domain of scientometrics in a nutshell. Section 3 describes the SWJ dataset as well as the process of converting the data into triples and publishing them as Linked Data. In section 4, we discuss the design and implementation

of the scientometric analysis modules. Next, in section 5, we outline the overall system architecture and detail how these analysis modules are integrated and are functioning as a Web portal. Finally, in section 6, we conclude our work and point out limitations and directions for future work.

## 2 Scientometrics and Bibliographic Data in a Nutshell

Among the vast number of datasets available on the LOD cloud, several data hubs, such as DBLP<sup>4</sup> and CiteSeer<sup>5</sup>, have published bibliography data which provides the potential to reveal networks among researchers, their coauthors, papers, journals, and conferences. To semantically disambiguate the bibliographic vocabularies, researchers have developed a number of bibliographic ontologies to annotate and structure their data [7]. BIBO is one of these bibliographic ontologies. In fact, we used it before in *Spatial@LinkedScience* which serves as a Linked Data portal for the bibliography data in the field of Geographic Information Science (GIScience) [8].

The term *scientometrics* was coined by Vassily V. Nalimov in the 1960s, and refers to the science of measuring and analyzing of science, such as a discipline's structure, growth, change, and interrelations [9]. While scientometrics often studies the dynamics of a discipline, its methodologies and principles have also been applied to quantify an individual researcher's scientific output [10,11]. The large amount of bibliographic data available on the LOD Cloud allows new types of scientometrics. However, the important full text data about the articles are often missing, partially due to copyrights limitations. While there is an increasing number of open access journals, many of their articles are still not available on the LOD Cloud. As full texts often play an important role in analyzing research topics in a scientific discipline [12,13], a lack of such data limits the capacity of Linked Data to assist scientometric applications. Work related to ours has been made popular by the ArnetMiner portal [14] which also exports RDF. In contrast to our work, however, ArnetMiner, is restricted to direct bibliographic data enriched with some author related information. Moreover, while ArnetMiner is a centralized collection of data, our approach is decentralized and focuses on the semantic enablement of individual publishers and journals.

The unique review process of SWJ has collected not only bibliographic data and researcher networks, but also the content of manuscripts, review comments, as well as the authors' responses. Since such data are accessible to the general public under the open and transparent review process, their Linked Data version can also be contributed to the LOD Cloud and can be used to promote the understanding of different scientific research areas, such as the Semantic Web. With the full text data, text mining methods, such as LDA, can be applied to extract important topics in an article or a series of articles. When combined with the time periods, such key topics can also delineate the dynamics and trends of the research in a scientific area [15,16].

---

<sup>4</sup> <http://datahub.io/dataset/fu-berlin-dblp>

<sup>5</sup> <http://thedatahub.org/dataset/rkb-explorer-citeseer>

### 3 Data Publishing and Access

In this section we describe the process of converting the Semantic Web journal's data from the new Drupal system into Resource Description Framework (RDF) triples. We also discuss extensions to the BIBO ontology and give examples how to access our data via SPARQL and Pubby.

#### 3.1 Data source

While we converted most SWJ data to Linked Data, we focus on the data collected by our new journal management system. Data from the old system is less rich in terms of the available information, especially with respect to revisions and reviews. For instance, data about reviewers is only available starting from January 2013. Finally, while we also serve full reviews in RDF and they are also available on our Web page, for the moment, these Linked Data are access-restricted. We discussed the difficulties in *depublishing* Web data before [3]. This situation becomes even more difficult for Linked Data and we need to study and understand the implications before releasing this information. This is simply because the degree of availability of Linked Data is orders of magnitude above that of dynamically generated HTML pages.

Our customized Drupal system treats each submitted manuscript as a *node*, and manages all publication data, including abstracts, authors, assigned editors, reviewers and review comments, the editor's decisions, authors' responses, paper categories, as well as PDF URLs, in a MySQL database. In total, this database contains information about more than 300 unique manuscript submissions since Fall 2010, multiple revisions for most of these manuscripts, about 1000 researchers with author, editor, and reviewer roles assigned to them, full bibliographic information for all papers, decision statuses, information about the submission type, assigned editor, PDF documents, public comments, and so forth. The dataset is constantly growing.

#### 3.2 Converting Data from Relational Database to RDF

The first step to creating a Linked Data version for the SWJ data is to convert the data from the relational database to RDF triples. Following Linked Data principles, we employ URIs as identifiers for papers, researchers, and other entities in the data. These URIs are developed using the namespace of the server (<http://semantic-web-journal.com/sejp/>) appended by the name of the entity (e.g., a paper's title or an author's name) with the words connected by hyphens. This URI convention makes it easy for human users to understand the meaning of the entities. To reduce the length of URIs for paper entities, stop words, such as *of* and *on*, contained in the paper title have been removed. Paper revisions are encoded using their underlying Drupal node ID. Meanwhile, we also remove the diacritical marks (which are contained in some authors' names) from the URIs to avoid potential issues in URI decoding. For reviewers who requested to remain anonymous during the review process, we use *salted MD5* to protect their

privacy while maintaining *identity*, i.e., there are multiple anonymous reviewer URIs. Several URI examples are listed below and will lead to a Pubby interface when opened in a Web browser:

- Author: a researcher whose name is *Matthew Rowe*  
<http://semantic-web-journal.com/sejp/page/matthew-rowe>
- Article: A version of *Approaches to visualising linked data: A survey*  
<http://semantic-web-journal.com/sejp/page/node/118>
- Submission types: List of all survey articles so far  
<http://semantic-web-journal.com/sejp/page/surveyArticle>
- Decision status: List of all accepted articles so far  
<http://semantic-web-journal.com/sejp/page/accept>

With the designed URI convention to identify each entity in the SWJ database, a set of formal vocabularies is still required to express the entities' relations. The BIBO ontology mentioned before fits many of our requirements. However, the existing BIBO ontology is unable to express the history of a paper during which the paper was first submitted, then reviewed, sent back for revision, and then resubmitted, since BIBO is originally designed to capture bibliographic data about papers that have already been published. To the best of our knowledge there is no ontology that captures the (internal) workflows of journals, conferences, and so forth. While a lack of such vocabularies may not cause much issue when used to express a bibliographic dataset, it does affect the results for scientometric analysis. For example, almost all papers are revised at least one time. Thus, a reviewer may need to review the same paper multiple times. Without tracking the history of the paper, the time and effort of the reviewer could not be distinguished from reviewing the paper just once.

In order to express the submission history of a paper, we extended the BIBO ontology with an *AcademicArticleVersion* class which is defined as a subclass of the *Article* class in the BIBO ontology. We reuse the *hasVersion* and *isVersionOf* relations from the Dublin Core Metadata Initiative (DCMI) Metadata Terms<sup>6</sup>. Two object relations *hasPreviousVersion* and *hasNextVersion*, as well as a datatype property *isLatestVersion* are further created to capture the relations between a paper's different versions. Figure 2 shows how these classes and relations are configured to encode multiple paper versions.

To capture the author order of a paper, we employ the RDF list properties recommended by the W3C RDF/XML Syntax Specification<sup>7</sup>, in which the orders are expressed as *rdf:\_1*, *rdf:\_2*, *rdf:\_3*, and so forth. Meanwhile, to simplify queries for the papers of an author (instead of having to know whether he/she is the first or second author beforehand), we use the *creator* relation from *DCMI terms* to connect each paper with all its authors. Figure 3 shows how these relations are used to capture the author order information.

Finally, to give a more comprehensive view of the extended BIBO ontology, figure 4 shows how information about a particular paper is encoded and stored in our triple store. It shows reviewers, authors, submission types, and so forth.

<sup>6</sup> <http://dublincore.org/documents/dcmi-terms/>

<sup>7</sup> <http://www.w3.org/TR/REC-rdf-syntax/#example17>

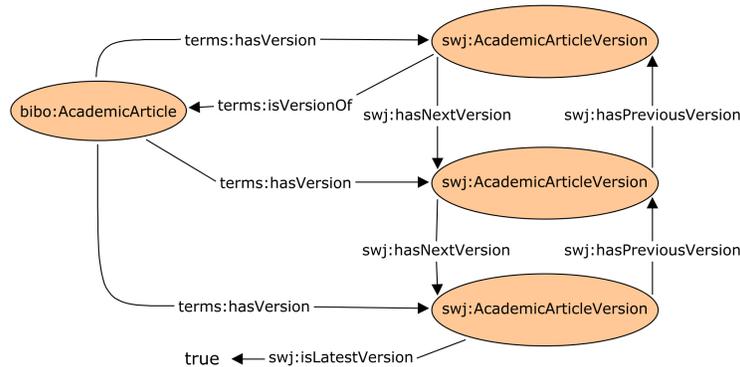


Fig. 2: Classes and relations used to express a paper's history.

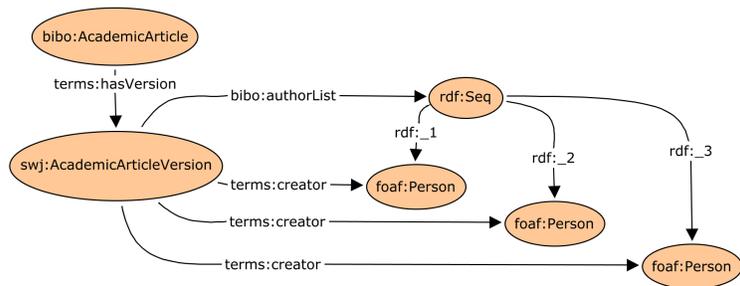


Fig. 3: The order of the authors for a paper.

Based on the extended BIBO ontology, we developed a customized Java converter using the *OWL API*<sup>8</sup>. The Java converter reads records from the SWJ MySQL database, generates RDF triples, and publishes them using Apache Jena's SPARQL server *Fuseki*<sup>9</sup>. This process has to be repeated regularly to keep the triple store synchronized with the journal's data. Once a paper gets published, it receives a page number, DOI, volume, and so forth. This data is imported from the IOS Press server, converted to RDF, and merged with the existing entities. The SPARQL endpoint is exposed at <http://semantic-web-journal.com:3030/sparql.tpl>. A Linked Data interface to the SPARQL endpoint has been created using Pubby<sup>10</sup>, and can be accessed at <http://semantic-web-journal.com/sejp>. Together this makes the data readable and accessible by humans and machines. Consequently, the SWJ data can be queried to arrive at results not directly represented by the portal described in

<sup>8</sup> <http://owlapi.sourceforge.net/>

<sup>9</sup> [http://jena.apache.org/documentation/serving\\_data/](http://jena.apache.org/documentation/serving_data/)

<sup>10</sup> Pubby is a Java web application that facilitates the development of a frontend for Linked Data served through a SPARQL endpoint. Details about Pubby can be found at <http://wifo5-03.informatik.uni-mannheim.de/pubby/>

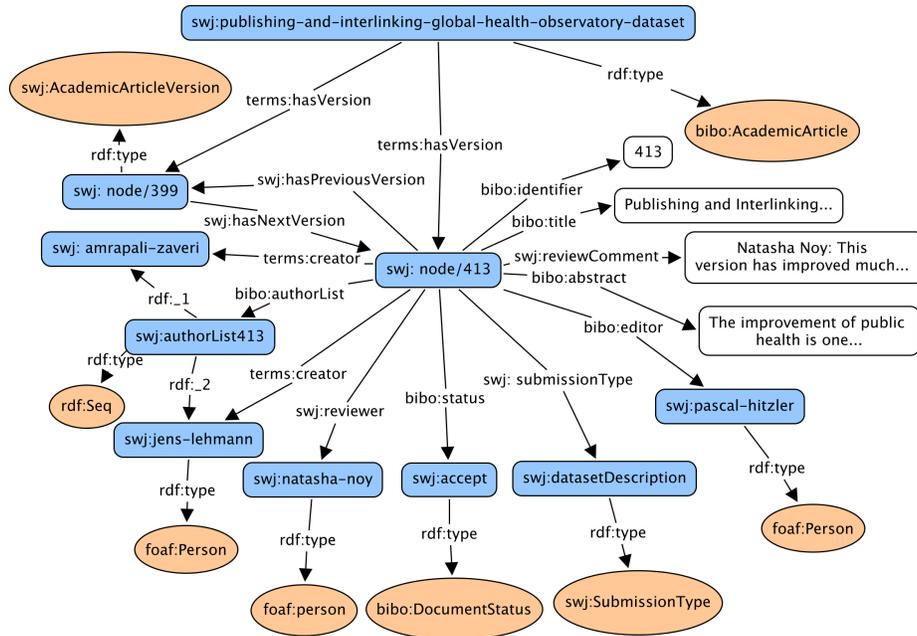


Fig. 4: A comprehensive view of the extended BIBO using a paper as an example (orange ellipses represent classes, blue rectangles represent entities, and white rectangles represent literals.)

section 4 below. The following SPARQL query, for example, will show authors of accepted papers that at the same time also edited a paper. This includes some members of the SWJ editorial boards as well as guest editors from special issues and their editorials (or regular articles that they published at SWJ before).

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX dc: <http://purl.org/dc/terms/>
PREFIX bibo: <http://purl.org/ontology/bibo/>
PREFIX swj: <http://semantic-web-journal.com/sejp/>

SELECT distinct ?author
{
  ?article dc:creator ?author;
          bibo:status swj:accept.
  ?_a     bibo:editor ?author.
}

```

## 4 Analysis Modules

In this section, we provide a brief overview of some of the interactive modules developed for the Linked Data-driven, semantically-enabled journal portal (SEJP) so far. Currently 20 modules have been deployed and new modules are under development. These modules will also be customized for other IOS Press journals.

### 4.1 Statistical Summary Modules

The exposed Linked Data allows us to summarize the work of authors, reviewers, editors, and the state of the journal as such. Examples include the number of paper that a researcher has reviewed, a list of her/his of community services (e.g., conference committee membership), or the acceptance rate of the journal.

To acquire a more comprehensive view about a researcher, as well as to link out to other data sources on the LOD Cloud, we have created external links to the Semantic Web Dog Food. SWDF contains information about many researchers in the Semantic Web community, their papers, and their community services. With the combination of the two data sources, a researcher's information may look like follows:

```
swj:researcher swrc:affiliation http://data.semanticweb.org/organization/uc-santa-barbara
swc:holdsRole http://data.semanticweb.org/conference/pcm2004
terms:isCreatorOf http://semantic-web-journal.com/sejp/node/204
bibo:isEditorOf http://semantic-web-journal.com/sejp/node/108
swj:isReviewerOf http://semantic-web-journal.com/sejp/node/122
```

Using the combined information, we developed several statistical summary module. Two of them are shown in Figure 5. In the first case the percentage of papers per type is displayed. In the second case a radar charts shows the normalized percentage of a researcher's activity with respect to the Semantic Web journal.

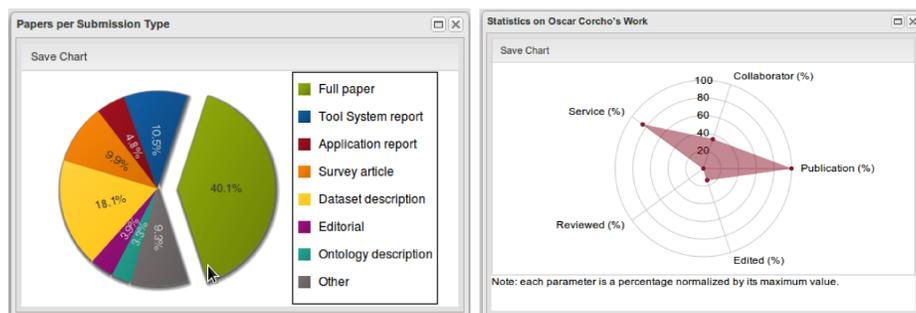


Fig. 5: The submission type and radar chart modules.

## 4.2 TreeMap of Co-Authors Module

Most SEJP modules are highly interactive and many of them enrich the SWJ data with external information. For example, Figure 6 shows a TreeMap visualization of the co-authors of Frank van Harmelen. Clicking on the universities will show the individual authors with pictures. The size of the colored fields indicates the number of co-authors from the particular institution. The module does not only contain SWJ data but collects co-author information from Microsoft's Academic Search API<sup>11</sup> and caches them. While the overall data quality is good, errors in Microsoft's identity resolution will also appear in the TreeMap. The figure also shows a fragment of the general people module which gives access to modules about individual authors, reviewers, and editors. For instance, the buttons activate a citation map and the radar chart of activities.

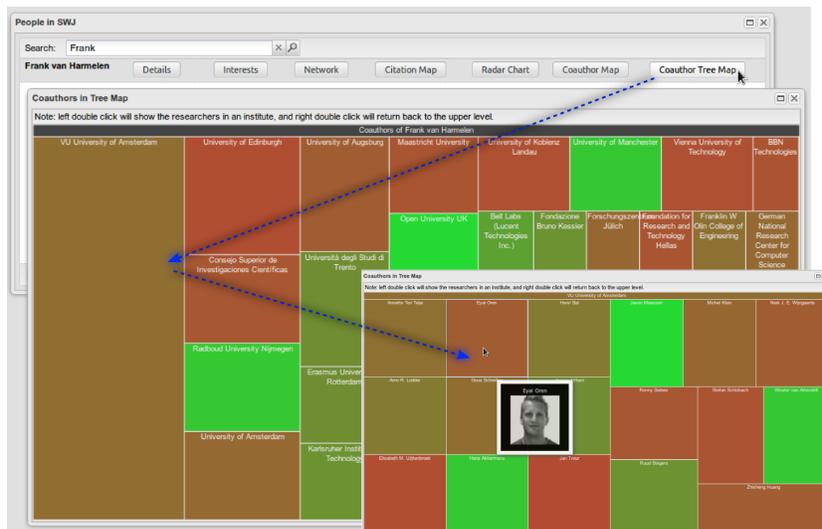


Fig. 6: The SEJP TreeMap module.

## 4.3 Collaborator Network Visualization

The network module is another example. It aggregates data from multiple sources to visualize relations between researchers and their work. Figure 7 shows a partially expanded and interactive example. It depicts jointly authored papers by Hitzler and Janowicz, their common co-authors, as well as the papers they edited for the Semantic Web journal. Clicking on a node will expand the graph and reveal additional information – in this example all titles of papers edited by Hitzler.

<sup>11</sup> <http://academic.research.microsoft.com/>

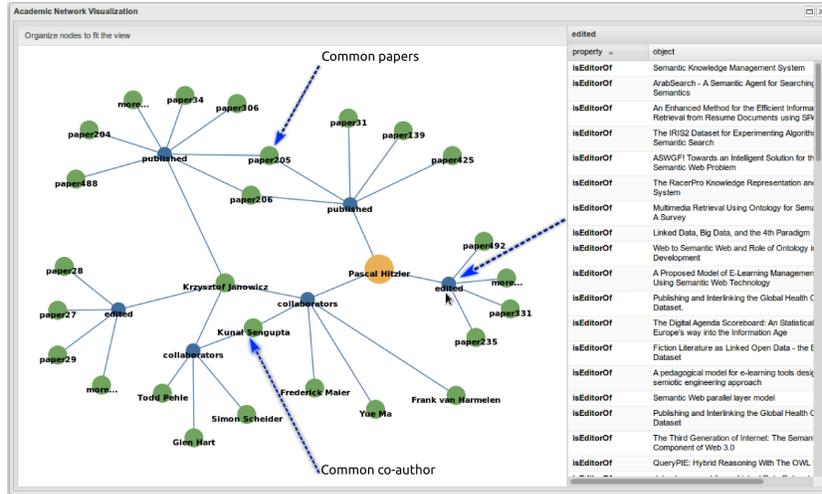


Fig. 7: Partially expanded network showing joint publications and co-authors.

The module can, for instance, support editors in detecting conflicts of interest. The data can also be exploited for sociability analysis as part of scientometrics as done by ArnetMiner before. Our data, however, provides additional relations and could be linked to ArnetMiner in the future.

#### 4.4 Geospatial Influence Visualization

Traditional scientometric analysis for individual researchers often focuses on the number of their publications and citations, and uses numeric values to evaluate a researcher's work, such as the H-index [11]. However, a numeric value may not be enough. Consider two researchers both having 100 citations for a given paper. The citations of one researcher are limited within a single country, e.g., the U.S., while those for the other researcher are spread throughout the world. By just looking at the numbers, we may conclude that both papers have similar academic influence. This, however, hides many interesting spatial aspects, e.g., the detection of *hot regions* in the geography of science [17], and thus our understanding of how ideas spread and why scientific communities are often local.

Based on this motivation, we linked the authors in the SWJ dataset with more general information from Web-based academic search engines. In the current version the module links authors with the information on Microsoft Academic Search since it provides more complete profile information. In particular, we are interested in affiliation of researchers which can be used to geolocate them. However, in later versions, other academic search engines, such as Google Scholar and ArnetMiner may also be integrated. Figure 8 shows the spatial (and temporal) distribution of the citations for a paper of a US-based first-author. The green symbols indicate places where the paper was cited and the color fades

out the older the citation is. The top citing authors (and their affiliations) are connected with links and also shown on the right side. For computational reasons only the first author is considered.

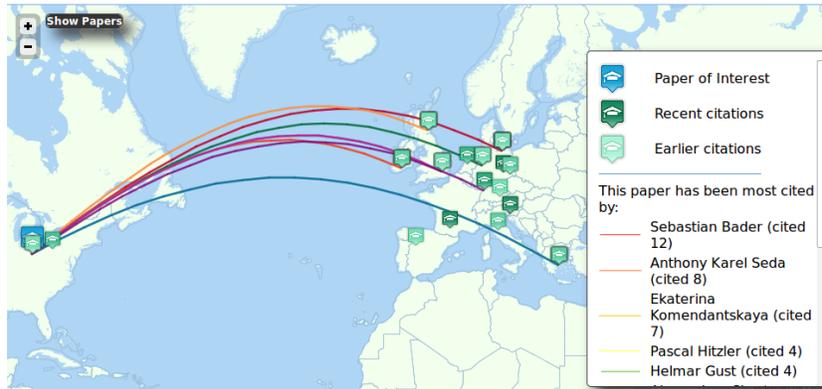


Fig. 8: Visualization of an author's citations

#### 4.5 Hive

The Hive graph module is another example, it visualizes the relationship between authors, papers, and keywords. Depending on which axis is given priority, the module will show the papers and keywords associated with authors, the authors and keywords associated with papers (see figure 9), or the authors and papers associated with a certain keyword (see figure 10). This powerful exploration tool can be used to understand what topics are common to most authors, the areas of expertise of particular authors, the position a certain paper takes within the research field, and so forth. In the case depicted in figure 10, the mouse is moved over the keyword axis (the keywords are mined from the fulltexts of all papers) and the hive chart shows all authors (and papers) that have a strong association to the (stemmed) keyword *ontology*.

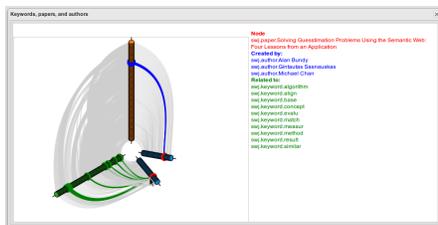


Fig. 9: The Hive module showing representative keywords for a paper and its

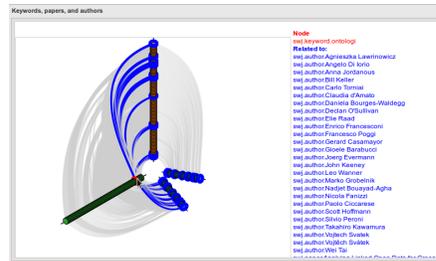


Fig. 10: The module showing all authors (and papers) for a given keyword.

## 5 Architecture and User Interface

In this section, we describe the software framework developed based on the linked SWJ data and the analysis modules. We discuss the architecture of the system, and show how the modules designed in section 4 can be flexibly assembled to form a Web portal. The developed Web portal can be accessed at <http://sejp.semantic-web-journal.net/> and will be constantly extended and updated adding new data and functionality.

### 5.1 Architecture

The Web portal is constructed based on a client-server architecture which uses asynchronous JavaScript requests (AJAX) for communication. Figure 11 shows the overall architecture. The server hosts the published Linked Data and the analysis modules. Each module is a combination of a JavaScript file, a Java Servlet, and a set of SPARQL queries (including queries to external data). In some cases additional data is loaded, e.g., state boundaries in the new TopoJSON format. These modules can be easily assembled to form a Web portal like the one shown in this paper. They can also be separated and integrated with other systems, i.e., they are loosely-coupled. Almost all modules are interactive and are dynamically updated. More complex queries involving external sources are cached or pre-computed.

In other words, SEJP provides a growing library of self-contained and interactive JavaScript analysis modules that can be recombined and styled (using CSS) to develop portals for other journals and conferences. These modules can directly communicate with the SPARQL endpoint, the stored LDA topics, external geo-data, and so forth. In the case shown below we use Java Servlets as additional *facades*. These Servlets can play multiple roles, e.g., cache data, restrict access, render more complex visualizations on the server (e.g., for mobile devices), and so forth.

### 5.2 User interface

We designed a flexible user interface to demonstrate how the modules jointly form a portal. The JavaScript framework ExtJS 4.1<sup>12</sup> has been used to facilitate the UI development process. More complex modules also make use of the D3 library.<sup>13</sup> In addition to the analysis modules, the user interface also provides data exploration and key word query functionality. Users unfamiliar with SPARQL queries can easily use those key word queries to access the data served by the Web portal.

The user interface shown in figure 12 consists of four major parts. The top panel shows a grouped drop-down menu which allows the user to load different modules. Beneath it, the workbench canvas displays the modules and allows the

<sup>12</sup> <http://www.sencha.com/products/extjs>

<sup>13</sup> <http://d3js.org/>

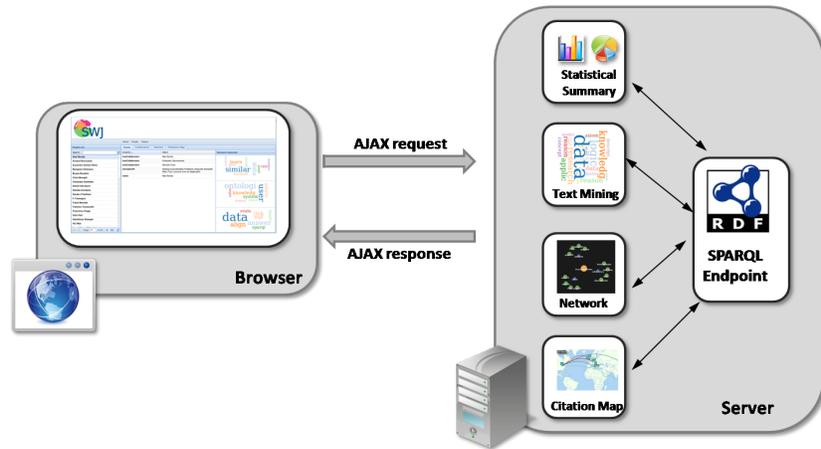


Fig. 11: Basic architecture overview.

user to rearrange them in different ways. One of these modules, for instance, shows the percentage of submitted manuscripts per type. Finally, some modules group access to other modules. For example, the people module gives access to modules showing data about authors, reviewers, and editors. It also contains a search box to find people by name.

## 6 Conclusion and Future Work

In this work, we converted data collected from the Semantic Web journal to RDF and published them as Linked Data. In addition to the content offered by traditional bibliographic datasets, our data also contain an entire time line for each paper together with metadata from SWJ's unique open and transparent review process. This enables novel scientometric applications, insights into scientific networks, allows to study the spread of ideas, discover new trends, and so forth.

To capture the information about a paper's time line, we extended the BIBO ontology by allowing a paper to have several versions and by establishing sequential links among these versions. We also added new classes and roles to model reviewers and their reviews, decision statuses (such as whether a paper is under major revision), submission types, and so forth. The data are published via a SPARQL endpoint along with a Pubby Linked Data interface for data exploration. External links to the Semantic Web Dog Food and DBpedia are also established, and we make use of the Microsoft Academic Search API as an additional source of information.

Based on the published data, we exemplarily showcase some of the 20 analysis modules such TreeMap of collaborators or a Hive plot visualizing the relation between authors, papers, and areas of expertise. The SEJP modules can used

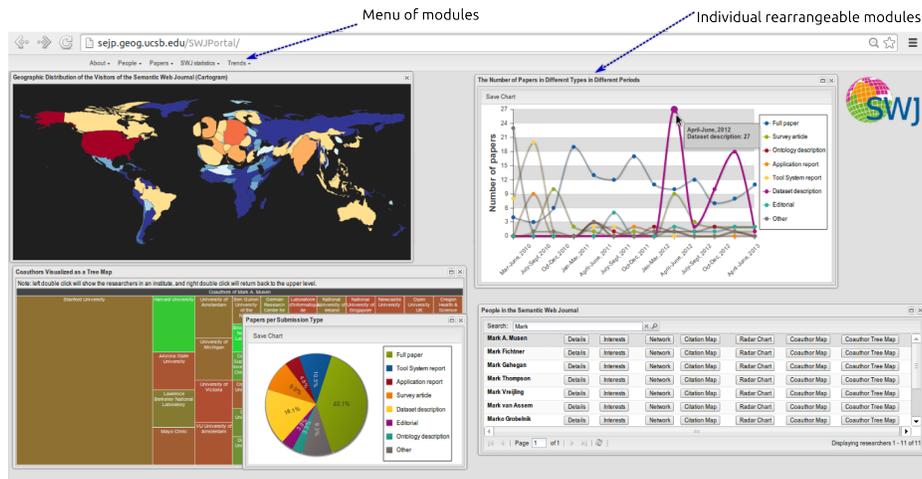


Fig. 12: Example of an user interface showing the menu bar and multiple overlapping modules of the Semantic Web journal's SEJP installation.

for multiple tasks such as finding suitable reviewers, browsing for potential co-authors and papers of interest, or exploring research trends by mining for latent topics in recently submitted manuscripts. Several of the developed modules also take a spatial perspective and provide users with insights about spatial citations patterns and the distribution of research interests. Finally, those interactive modules were made accessible via a Web-based user interface. The interface is developed in a workbench-style, i.e., different modules can be loaded at the same time and arranged to support multiple perspectives on the same data.

In the future, we plan to add additional modules, provide more complex interaction techniques, links between the different modules, and integrate more data sources. In terms of analysis, we will especially focus on research trending, e.g., to understand the dynamics of a discipline or an individual researcher. The presented deployment of SEJP to the Semantic Web journal is just the beginning. In the near future the portal will be used by other IOS Press journals (based on the data they provide) as well. Finally, by contributing the SWJ dataset to the Linked Data cloud, we hope to further open up scientific review processes, make them more transparent, and document the research field.

## References

1. Stadler, C., Lehmann, J., Höffner, K., Auer, S.: LinkedGeoData: A core for a web of spatial open data. *Semantic Web* **3**(4) (2012) 333–354
2. Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., Ives, Z.: DBpedia: A nucleus for a web of open data. In Aberer, K., Choi, K.S., Noy, N.F., Allemang, D., Lee, K.I., Nixon, L.J.B., Golbeck, J., Mika, P., Maynard, D., Mizoguchi, R., Schreiber, G., Cudré-Mauroux, P., eds.: *The Semantic Web, 6th International*

- Semantic Web Conference, 2nd Asian Semantic Web Conference, ISWC 2007 + ASWC 2007, Busan, Korea, November 11-15, 2007. Volume 4825 of Lecture Notes in Computer Science. Springer (2007) 722–735
3. Janowicz, K., Hitzler, P.: Open and transparent: the review process of the Semantic Web journal. *Learned Publishing* **25**(1) (2012) 48–55
  4. Hitzler, P., Janowicz, K., Sengupta, K.: The new manuscript review system for the Semantic Web journal. *Semantic Web* **4**(2) (2013) 117
  5. D’Arcus, B., Giasson, F.: Bibliographic Ontology Specification. Online: <http://bibliontology.com/specification> (November 2009) Last accessed 2013-05-12.
  6. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. *the Journal of machine Learning research* **3** (2003) 993–1022
  7. Shotton, D., Portwin, K., Klyne, G., Miles, A.: Adventures in semantic publishing: exemplar semantic enhancements of a research article. *PLoS computational biology* **5**(4) (2009) e1000361
  8. Keler, C., Janowicz, K., Kauppinen, T.: spatial@linkedsience Exploring the Research Field of GIScience with Linked Data. In Xiao, N., Kwan, M.P., Goodchild, M., Shekhar, S., eds.: *Geographic Information Science*. Volume 7478 of Lecture Notes in Computer Science. Springer Berlin Heidelberg (2012) 102–115
  9. Hood, W.W., Wilson, C.S.: The literature of bibliometrics, scientometrics, and informetrics. *Scientometrics* **52**(2) (2001) 291–314
  10. Braun, T., Glänzel, W., Schubert, A.: A Hirsch-type index for journals. *Scientometrics* **69**(1) (2006) 169–173
  11. Hirsch, J.E.: An index to quantify an individual’s scientific research output. *Proceedings of the National academy of Sciences of the United States of America* **102**(46) (2005) 16569
  12. Glenisson, P., Glänzel, W., Janssens, F., De Moor, B.: Combining full text and bibliometric information in mapping scientific disciplines. *Information Processing & Management* **41**(6) (2005) 1548–1572
  13. Brody, T., Carr, L., Gingras, Y., Hajjem, C., Harnad, S., Swan, A.: Incentivizing the open access research web: publication-archiving, data-archiving and scientometrics. *CTWatch quarterly* **3**(3) (2007)
  14. Tang, J., Zhang, J., Yao, L., Li, J., Zhang, L., Su, Z.: Arnetminer: extraction and mining of academic social networks. In: *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM (2008) 990–998
  15. Wang, X., McCallum, A.: Topics over time: a non-markov continuous-time model of topical trends. In: *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM (2006) 424–433
  16. Zhou, D., Ji, X., Zha, H., Giles, C.L.: Topic evolution and social interactions: how authors effect research. In: *Proceedings of the 15th ACM international conference on Information and knowledge management*, ACM (2006) 248–257
  17. Bornmann, L., Waltman, L.: The detection of ”hot regions” in the geography of science — A visualization approach by using density maps. *Journal of Informetrics* **5**(4) (2011) 547–553