

# ScienceWISE: A Web-based Interactive Semantic Platform for Scientific Collaboration

Karl Aberer<sup>1</sup>, Alexey Boyarsky<sup>123</sup>, Philippe Cudré-Mauroux<sup>4</sup>,  
Gianluca Demartini<sup>4</sup>, and Oleg Ruchayskiy<sup>5\*</sup>

<sup>1</sup> Ecole Polytechnique Fdrale de Lausanne, Switzerland  
`{firstname.lastname}@epfl.ch`

<sup>2</sup> Instituut-Lorentz for Theoretical Physics, Universiteit Leiden, The Netherlands

<sup>3</sup> Bogolyubov Institute for Theoretical Physics, Kiev, Ukraine

<sup>4</sup> eXascale Infolab, University of Fribourg, Switzerland

`{firstname.lastname}@unifr.ch`

<sup>5</sup> CERN TH-Division, PH-TH, Geneva 23, Switzerland

`oleg.ruchayskiy@cern.ch`

## 1 Introduction

Organizing scientific knowledge in systematic ways becomes increasingly important. However, the creation of intra- and inter-disciplinary knowledge bases is hindered by the heterogeneity and the scale of the information to consider. This calls for *scientific community-run systems* (replacing classical publishers of encyclopedias) allowing to combine presentation of new results, in-depth discussions, “user-friendly” introductions for young scientists, and meta-data to relate semantically similar concepts or pieces of content. Today, there are no standard tools to insert, store and query such meta-data online, which mostly remains “in the heads of the experts”. *The goal of the ScienceWISE (SW) project is to improve this situation.*

ScienceWISE<sup>6</sup> allows a community of scientists, working in a specific domain, to generate dynamically as part of their daily work an *interactive semantic environment*, i.e., a field-specific ontology with direct connections to the text of research papers. The SW system includes a number of crucial elements: (a) an expanding collection of field-specific community-ranked encyclopedia articles (mostly about physics at this point of time); (b) an ontological structure (concepts and logical relations between them) encompassing this encyclopedia; (c) established connections from ontology entries to a vast collection of research papers ([ArXiv.org](http://ArXiv.org))<sup>7</sup>; (d) an operational platform, allowing scientists to annotate and conceptually index (bookmark) research papers, link them against the ontology, and validate and dynamically update the ontology through annotations.

In this demonstration we present the SW system with a particular focus on its semantic-based expert finding functionality. By leveraging on the SW ontology, the system is able to rank authors of scientific papers by exploiting the concepts contained in their papers as well as the relations among concepts. In the following we briefly describe the SW platform together with its ontology and main functionalities. We then present the expert finding module that ranks authors when answering keyword queries.

## 2 The ScienceWISE Platform

The central use-cases of the SW system are *annotations* (expanding the semantic space of a paper with “Supplementary material”) and *scientific bookmarking*. To *annotate* a paper, authors are

---

\* Authors are listed in alphabetical order.

<sup>6</sup> Accessible at <http://ScienceWISE.info>

<sup>7</sup> Since early 90s <http://ArXiv.org> provides free full-text access to preprints of research papers in physics, mathematics, some branches of CS, biology, etc. In certain areas of physics and mathematics it contains close to 100% of all manuscripts.

Fig. 1. A concept in the ScienceWISE ontology and its use for bookmarking.

presented with a list of concepts and definitions that are automatically identified in the paper and ordered by their relevance. As the user select concept, our system dynamically produces a hyperlinked version of the manuscript, inserting hyperlinks to relevant definitions/resources, thus expanding the paper with additional details, comments or pedagogical material. Competing scientific viewpoints are represented as alternative resources and definitions about the same concept.

User can *bookmark* any ArXiv.org paper using the SW ontology (*conceptual indexing*). The system automatically selects the most relevant concepts for characterizing a paper, which can be further fine-tuned by the user. A concept navigation panel (see Figure 1) allows to classify bookmarked papers, create collections and easily navigate to any bookmarked paper with a few clicks.

Both for annotation and bookmarking users can add concepts, definitions, resources, and relations that deem necessary. This occurs exclusively through the SW ontology, thus creating a mechanism to expand it manually and validate the results of automated expansion.

**The SW Ontology** The SW Ontology<sup>8</sup> is the result of a community effort made by domain experts. To create the initial version of the ontology, we have performed a semi-automated import from many science-oriented ontologies and online encyclopedias. After this initial step, SW users (who are domain experts) are now allowed to edit elements of the ontology (e.g., adding new definitions or new relations) in order to improve its quality. Presently, the SW ontology counts more than 15 000 unique entries, each with its own definitions, alternative forms, and semantic relations to other entries. The semantic relations are both of general (e.g., *is a part of*) and field-specific (*is a model of*, *is observed in*) nature.

**The SW Document Management system.** In addition to the ScienceWISE ontology, SW has a back-end document management system, based on the CERN's *Invenio system*<sup>9</sup>, allowing for documents and metadata storage, indexing, and effective search of research documents (see Figure 2). The SW collection of scientific papers is currently growing, aiming to reach half a million papers in various domains of physics, mathematics, computer science, etc. within the coming months.

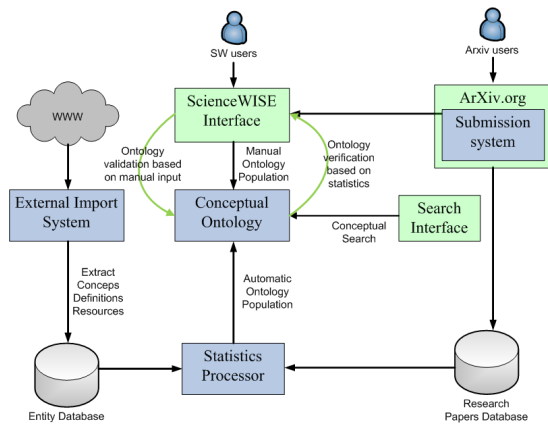
**Users.** Since April 2011 the SW system is public and accessible by scientists via ArXiv.org as well as via the CERN Document Server<sup>10</sup> and Inspire<sup>11</sup>. The system currently counts above

<sup>8</sup> Accessible for browsing via <http://ScienceWISE.info/ontology>

<sup>9</sup> <http://invenio-software.org>

<sup>10</sup> <http://cds.cern.ch>

<sup>11</sup> <http://inspirebeta.net>: comprehensive bibliographic database in high-energy



**Fig. 2.** The architecture of the ScienceWISE system; Black arrows represent the interactions between the various modules of our system and external entities such as users, the Web, and the ArXiv.org document server.

200 active users, thousands of conceptually indexed and annotated papers, and is now receiving several new registrations *daily*.

**Related Systems.** There exists a number of systems (such as BibSonomy.org [2], Connotea, Citeulike, or Mendeley) that offer to tag and share research papers. Also, many bibliography managing systems (including the CERN Document Server) allow to add private or public comments to any record they keep. MathSciNet<sup>12</sup> has a long-standing tradition of open peer reviews for mathematical papers. Additionally, there is a number of attempts to create “scientific blogs” and/or “discussion forums” to supplement scientific publications with expanding and clarifying notes.<sup>13</sup>

SW is conceptually very different from all those systems since it is built on a highly-structured, collaborative and expressive ontology, rather on static taxonomies or keywords. The SW ontology is used to automatically establish links between keywords and scientific concepts, to semantically annotate resources, etc. The SW ontology allows our system to perform those tasks automatically, but lets the users fine-tune the suggested annotations. As a result, users are empowered with very powerful semantic functionalities to search for papers or experts (see below), and to browse collections of papers using their “ontological neighborhoods”.

### 3 Ranking Authors by their Expertise

One of the many interesting functionalities that can be built on top of the SW ontology and the collection of annotated research papers is *semantic search*. SW users can search for papers, definitions, and concepts. In addition, they can see who are the experts in the areas related to their queries. Specifically, we rank experts by considering all possible relations available in our platform by taking advantage of concepts identified in the document, their semantic neighborhood, and the authors who contributed to the document.

In the literature there are two main approaches to the expert finding problem: 1) build author profiles and search over those or 2) retrieve papers first and then rank the authors of those papers [1]. SW adopts the latter approach together with the following scoring function to compute the

<sup>12</sup> from the American Mathematical Society, see <http://www.ams.org/mathscinet>

<sup>13</sup> See e.g. [CosmoCoffee.info](http://cosmocoffee.info) – researchers’ forum, related to cosmology; [PhysicsForums.com](http://physicsforums.com) – general question and answer forum, student oriented; scientific blogs of individual scientists or groups (e.g. <http://blogs.discovermagazine.com/cosmicvariance>).

ranking of an author  $a$  given a user query  $q$ :

$$score(q, a) = \sum_{d_i \in D_q} \left( \frac{BM25(q, d_i)}{\sum_{d_i \in D_q} BM25(q, d_i)} \frac{\sum_{c_j \in C_q} f(d_i, c_j) rel(c_j, q)}{|C_q|} imp(a, d_i) auth(d_i) \right) \quad (1)$$

where  $D_q$  is the set of retrieved papers for query  $q$ ,  $BM25(q, d_i)$  is a standard scoring function for document retrieval [3],  $C_q$  is the set of concepts described in the query  $q$ ,  $f(d_i, c_j)$  is a function that represents the relation between a document  $d_i$  and a concept  $c_j$ <sup>14</sup>,  $rel(c_j, q)$  is a weighting function that represents the importance of a concept with respect to a query,  $imp(a, d_i)$  is a function that represents the relation between an author  $a$  and a document  $d_i$ <sup>15</sup>, and  $auth(d_i)$  is an authority score for document  $d_i$ <sup>16</sup>.

An extension of the above method exploits *query extension* techniques: the set of query concepts  $C_q$  is extended by considering (possibly with a lower weight) related concepts in the SW ontology. For example, the query “dark matter direct detection experiments” directly relates to two elements in the ontology (*dark matter* and *direct detection experiments*). In addition, semantically related concepts (‘Cold dark matter’, ‘weakly interacting massive particles’, etc.) can be added to  $C_q$  in order to improve recall and the quality of the top-k results.

## 4 Demonstration

In the demonstration we propose, users are expected to upload papers on `www.scienceWISE.info` dynamically, to semi-automatically bookmark papers, and to manually modify the underlying SW ontology using a visual interface. Users can also search for experts using semantic concepts, and compare the results of various expert finding algorithms based on i) a simple tf/idf scheme ii) the semantic search functionality described above and iii) semantic search with query expansion. In the latter case, they have to possibility to adjust the results by varying the semantic scope considered in the ontology. The expert search interface also allows to fine-tune the results and to disambiguate experts based on additional ontological information<sup>17</sup>.

## 5 Acknowledgments

This project is carried out as part of the “AAA/SWITCH – e-infrastructure for e-science program” under the leadership of SWITCH, the Swiss National Research and Education Network, and has been supported (in part) by funds from the ETH Board and by the Swiss National Science Foundation under grant number PP00P2.128459.

## References

1. Krisztian Balog, Leif Azzopardi, and Maarten de Rijke. Formal models for expert finding in enterprise corpora. In *SIGIR*, pages 43–50, 2006.
2. A. Hotho, R. Jäschke, C. Schmitz, and G. Stumme. Bibsonomy: A social bookmark and publication sharing system. In *Proceedings of the Conceptual Structures Tool Interoperability Workshop*, pages 87–102, 2006.
3. Stephen E. Robertson and Hugo Zaragoza. The Probabilistic Relevance Framework: BM25 and Beyond. *Foundations and Trends in Information Retrieval*, 3(4):333–389, 2009.

<sup>14</sup> Currently we compute it as the number of times a concept is mentioned in  $d_i$ .

<sup>15</sup> Currently we compute it as  $(\# \text{ of authors for } d_i.)^{-1}$

<sup>16</sup> Currently we compute it as the number of citations  $d_i$  has obtained so far.

<sup>17</sup> For example, experts on *Dark Matter* can work on astrophysical evidences (astronomers), on laboratory searches (experimental particle physics) and on models of dark matter (theoretical high-energy physicists). Our algorithm and interface allow to distinguish between these different fields of expertise.