Title	KonneXSALT: First Steps Towards a Semantic Claim Federation Infrastructure
Author(s)	Groza, Tudor; Handschuh, Siegfried; Möller, Knud; Decker, Stefan
Publication Date	2008
Publication Information	Tudor Groza, Siegfried Handschuh, Knud Möller, Stefan Decker "KonneXSALT: First Steps Towards a Semantic Claim Federation Infrastructure", Proceedings of the 5th European Semantic Web Conference (ESWC 2008), 2008.
Publisher	Springer
Link to publisher's version	10.1007/978-3-540-68234-9_9
Item record	http://hdl.handle.net/10379/420

KonneX$^{\mathbf{SALT}}$: First Steps Towards a Semantic Claim Federation Infrastructure

Tudor Groza, Siegfried Handschuh, Knud Möller, and Stefan Decker

DERI, National University of Ireland, Galway,
IDA Business Park, Lower Dangan, Galway, Ireland
{tudor.groza, siegfried.handschuh, knud.moeller, stefan.decker}@deri.org
http://www.deri.ie/

Abstract. Dissemination, an important phase of scientific research, can be seen as a communication process between scientists. They expose and support their findings, while discussing claims stated in related scientific publications. However, due to the increasing number of publications, finding a starting point for such a discussion represents a real challenge. At same time, browsing can also be difficult since the communication spans accross multiple publications on the open Web. In this paper we propose a semantic claim federation infrastructure, named KonneX$^{\mathrm{SALT}}$, as a solution for both issues mentioned above: (i) finding claims in scientific publications, and (ii) providing support for browsing by starting with a claim and then following the links in an argumentation discourse network (ADN) (in our case, by making use of transclusion). In addition, we join the web of linked open data, by linking the metadata contained in KonneX$^{\mathrm{SALT}}$with some of the known repositories of scientific publications.

1 Introduction

Dissemination, an important phase of scientific research, can be seen as a communication process between scientists. They expose and support their findings, while discussing claims stated in related scientific publications. This communication takes place over the course of several publications, where each paper itself contains a rhetorical discourse structure which lays out supportive evidence for the raised claims. Often this discourse structure is hidden in the semantics expressed by the publication's content and thus hard to discover by the reader.

Externalization, as defined by Nonaka and Takeuchi [1], represents *the process of articulating tacit knowledge into explicit concepts*. As such, it holds the key to knowledge creation. Consequently, the knowledge becomes crystallized, thus allowing it to be shared with and by others. Although made explicit, the externalized knowledge is dependent on the degree of formalization. In the case of the argumentation discourse based on claims, it can be a couple of keywords, or a weakly structured text, both possibly including direct references to the publications stating the actual claims.

In a previous paper [2], we have described SALT (Semantically Annotated LaTeX), an authoring framework for creating semantic documents and defined a Web identification scheme (named *claim identification tree*) for claims in scientific publications. The goal of the framework was to define a clear formalization for externalizing the knowledge captured in the argumentation discourses. We defined a special LaTeX markup syntax for annotating claims and arguments, and modeled the relationship between them by means of the SALT Rhetorical Ontology. The Web identification scheme allowed us to introduce a novel way for authoring and referencing publications, by giving authors the possibility of working at a fine-grained level, i.e., by citing claims within publications. The main goal was to support the creation of networks of claims, which span across multiple publications.

Having set the foundation, we have now reached the point where we can provide ways to use the externalized knowledge modelled as semantic metadata. In this paper, we introduce KonneX$^{\mathrm{SALT}}$, a semantic claim federation infrastructure, designed with the goal of finding claims in scientific publications and providing support for browsing argumentative discussions, starting from a particular claim. KonneX$^{\mathrm{SALT}}$was not conceived to be yet another search engine, but rather to represent a look-up service for externalized knowledge and realize efficiency through the minimization of the data to be indexed. By using latent semantic indexing, it also provides a means for discovering similarities among the managed claims. From a browsing point of view, KonneX$^{\mathrm{SALT}}$defines Argumentative Discourse Networks (ADN) as realizations of associative trails [3]. As a consequence, it provides a method for federating claims with the help of semantic technologies. In addition, for improving the readability of the ADNs, it makes use of transclusion (*the inclusion of part of a document into another document by reference* [4]). Finally, KonneX$^{\mathrm{SALT}}$contributes to the web of "linked open data" by linking the claims, and implicitly the publications hosting them, to publications referred to by social websites (e.g. Bibsonomy[1]) or managed by known repositories (e.g. DBLP[2]). Our ultimate goal is to transform KonneX$^{\mathrm{SALT}}$in an open hub for linking scientific publications based on externalized knowledge.

The remainder of the paper is structured as follows: in Sect. 2 we present the relevant research performed in this field. Following, we introduce our use-cases (Sect. 3) and then, in Sect. 4, we provide background information to motivate the decisions taken for achieving our goals. In Sect. 5 we present the design and implementation of KonneX$^{\mathrm{SALT}}$, and before concluding in Sect. 7, we discuss some ethical and technical challenges discovered during our research (Sect. 6).

2 Related Work

The relevant literature for our work can be split in two main categories: search engines focused on scientific publications and hypertext systems close to our browsing goals.

---

[1] http://www.bibsonomy.org/
[2] http://dblp.uni-trier.de/

Table 1. Scientific publication search engines overview

	Google Scholar	CiteSeer	Science-Direct	DBLP	KonneX^{SALT}
Focus	Full publications	References	Full publications	Shallow metadata	Shallow metadata, references, claims
Population	Crawling	Crawling	Manual	Manual	Author-driven
Full search	+	−	+	−	−
SW oriented	−	−	−	±	+
Linking	−	−	−	−	+
Openness	+	+	−	±	+

Table 1 shows a brief comparative overview of some of the known search engines for scientific publications and shows how KonneX^{SALT} compares to them in a number of features. Traditional search engines focus on indexing the full content of the publications, inserted manually or via automatic crawling, but provide almost no means for creating/re-using semantic data (the only exception is DBLP, which offers an RDF dump). In contrast, our approach minimizes the quantity of data to be indexed and provides only semantically linked data.

On the browsing side, the roots of KonneX^{SALT} heavily reside on the work elaborated by visionaries like Vannevar Bush and Ted Nelson. In 1945, in his famous article "As we may think" [3], Bush described the Memex as a the first proto-hypertext system. The Memex was envisioned as an electronic device, which linked to a library, was able to display films or books from the library and when necessary, to follow cross-references from one piece to another. In addition to following links, Memex was also able to create links, based on a technology which combined electromechanical controls and microfilm cameras and readers. This is generally regarded as a main inspiration for and the first step towards today's hypertext.

In the Memex, Bush also introduced the concept of *associative trails*, which are realized in our approach by means of the argumentative discourse networks. Bush defines associative trails as *"a new linear sequence of microfilm frames across any arbitrary sequence of microfilm frames by creating a chained sequence of links, along with personal comments and 'side trails' "*. In a similar manner we (re)construct the argumentation discourse structure (the *new linear sequence of microfilm frames*), based on the claims stated in publications (*the arbitrary sequence of microfilm frames*), and following the positions and arguments expressed in papers referring to these claims (*chained sequence of links ...*).

In 1960, Ted Nelson founded Project Xanadu. It represented the first attempt to design and implement a hypertext system. Xanadu was describing the ideal approach of realizing a word processor which would allow versioning, visualization of difference between versions, and especially non-sequential writing and reading. Consequently, a reader was able to choose his own path through an electronic document, based on *transclusion*, or *zippered lists* [5], i.e., creating a document by embedding parts of other documents inside it. In KonneX^{SALT} we use transclusion in conjunction with the argumentative discourse networks to improve readability and the browsing process.

The Compendium methodology, developed as part of the ScholOnto project[3] was a major inspiration for our work. In [6], the authors describe the modeling foundation for capturing claims inside scientific publications, while [7] details the set of sensemaking tools which can be used for visualizing or searching claims in scholarly documents. As part of the work, transclusion is mentioned and handled in [8]. While our goals are very similar, there is a major difference in the modeling approach that we took and in our ultimate goal, i.e., to establish a new way of citing scientific publications not via usual references, but via claim identifiers.

In terms of navigation of argumentative discourse using hypertext systems, the main direction was given by Horst Rittel [9] when he introduced the IBIS (Issue-Based Information Systems) method. IBIS focuses on modeling *Issues* together with the *Positions* and *Arguments* in a decisional process. In 1987, Conklin et. al introduced a hypertext version of IBIS, called gIBIS, first by modeling team design deliberation [10] and then by exploring policy discussions [11]. Other approaches that make use of the IBIS method are for example the DILIGENT Argumentation Ontology [12] designed to capture the argumentative support for building discussions in DILIGENT processes, or Aquanet [13] designed to model the personal knowledge via claims and arguments.

3 Use-cases

Due to the increasing number of scientific publications, finding relevant literature for a particular field represents a real challenge. At a more fine-grained level, if we look for specific claims, we are faced with an even more cumbersome task. One of the main reasons for this problem is the lack of a central access point to all publications. Information about publications is scattered over the Web. In theory, this would not really present an issue, if there was a uniform way of representing this information. In practice, however, each of the currently existing publishing sources (e.g. journals, conferences, publication repositories) handles the representation of the scientific publications in its own particular manner. Therefore, the same publications exist in different environments, maybe modelled from different aspects, but without any explicit links between the instances of the same publication.

For example, let us consider a publication X: (i) X might be listed in DBLP, together with the publication environment (e.g., the conference where it was presented) and some extra information about the authors of the publication (e.g. other papers they published), (ii) its citation record can be found using CiteSeer, (iii) independent reviews about the publication might be found on Revyu.com, (iv) while even more information about the authors might be found in OntoWorld.org. All these information sources provide different (partially overlapping) facets of the same publication, some in a traditional fashion, while others

[3] http://kmi.open.ac.uk/projects/scholonto/

in the “Semantic Web” way, i.e., by describing the information in RDF (the last two). In the following, we detail the approach KonneX$^{\mathrm{SALT}}$takes in providing solutions for both finding publications and linking them.

3.1 Search and Browse

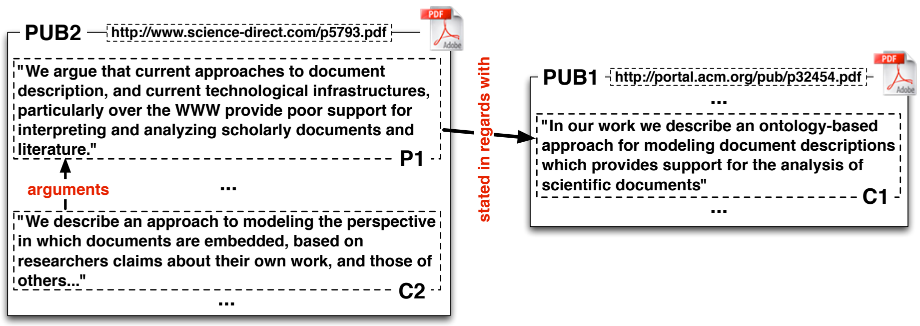

Fig. 1. Example of argumentative discussion in different publications

Current solutions for finding scientific literature, such as Google Scholar or CiteSeer, follow a traditional direction, by crawling the open Web for publications, storing them locally and performing full-text indexing. Therefore the search operation includes in terms of cost all the previous three operations, while in terms of time consumption is expensive because the search space comprises the full text of all publications.

Contrary to this approach, KonneX$^{\mathrm{SALT}}$tries to make the search operation as in-expensive and efficient as possible. In order to achieve this, we rely on the authors' incentive for externalizing their knowledge and on minimizing the search space only to this knowledge. As we show in Sect. 4, the benefit an author receives by annotating their publications with SALT is proportional to the effort they put in the annotation process. For the lowest possible level of annotation (i.e., claims and positions), the benefit is already clear: one can easily find the claims stated in one's publications and browse the argumentation line in the related work starting from such a claim. At the same time, the author themselves might find interesting connections between their claims and others, about which they might have not been aware. From an efficiency point of view, by indexing only the claims in publications, we restrain the search space to the minimum. This, combined with latent semantic indexing, provides us with a high level of accuracy, as well as the option of clustering similar claims.

As mentioned, in addition to the search functionality, the externalized knowledge gives us the possibility of enhancing the readability and browsing of the argumentative discourse spanning across multiple publications. Fig. 1 depicts an example of such a discourse: `PUB1` claims `C1`, while `PUB2` provides a position `P1`

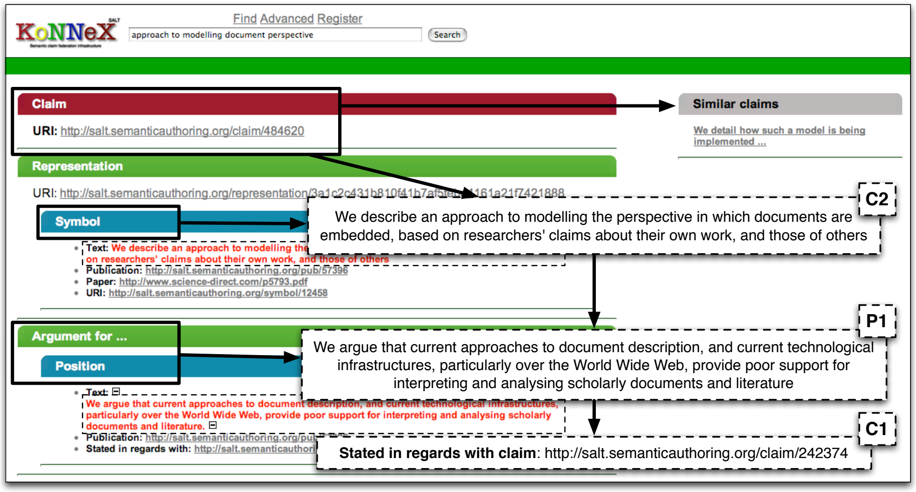

Fig. 2. Browsing a discussion in KonneX$^{\mathrm{SALT}}$

in regards to the claim in `PUB1` and arguments it with its own claim `C2`. Finding either of the claims in KonneX$^{\mathrm{SALT}}$ will result in the opportunity of browsing the entire discussion, as shown in Fig. 2.

3.2 Linking Open Data

As the Semantic Web grows, the focus in R&D moves more and more from producing semantic data towards the reuse of data. In this category we can also fit approaches for solving one of the collateral aspects of producing semantic data, i.e. replication. Different or overlapping aspects of the same data now exist on the Semantic Web. The Linking Open Data[4] initiative emerged from the necessity of creating bridges between those data islands. We join this effort by providing the means for linking publications in different environments. Fig. 3 depicts such an example, where the reader is advised to follow the links to find more information on this publication. The same linking information is also available as triples for reasoning purposes.

4 Background

As an intermediary step towards implementing the technical solutions to solve the use-cases we have just outlined, we will now provide some background information on the frameworks we use to achieve our goals and motivate our decisions. We start with a short description of SALT, then emphasize the evolution of the

[4] `http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData`

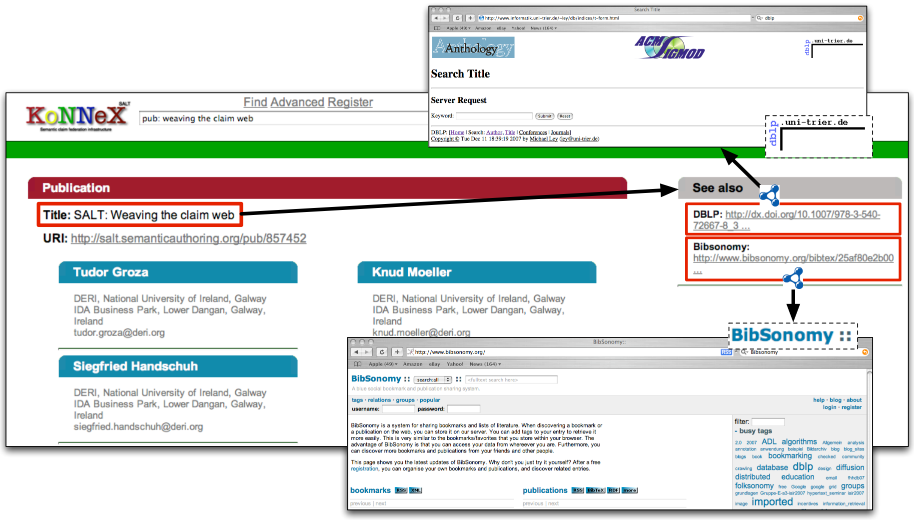

Fig. 3. Linked data in KonneX$^{\text{SALT}}$

claim identification tree into the argumentative discourse network and finally introduce Latent Semantic Indexing (LSI).

4.1 SALT: Semantically Annotated LaTeX

SALT is a semantic authoring framework targeting the enrichment of scientific publications with semantic metadata. The result of a SALT process is a semantic document, built by using the PDF file format as a container for both the document content and the semantic metadata. SALT comprises two layers: (i) a syntactic layer, and (ii) a semantic layer. The syntactic layer represents the bridge between the semantic layer and the hosting environment, i.e., LaTeX. It defines a series of new LaTeX commands, while making use of some of the already existing ones to capture the logical structure of the document and the semantics in the publication. We chose LaTeX because it is one of the most widely used writing environments in the scientific community. In a similar way, the SALT framework can be used also together with, for example, Microsoft Word.

The semantic layer is formed by a set of three ontologies: (i) The Document Ontology, which models the logical structure of the document, (ii) The Rhetorical Ontology, which models the rhetorical structure of the publication and (iii) The Annotation Ontology, which links the rhetorics captured in the document's content and the physical document itself. The Annotation Ontology also models the publication's shallow metadata. The Rhetorical Ontology has three sides: (i) one side is responsible for modeling the claims and their supporting argumens in scientific publications, as well as the rhetorical relations connecting them, (ii) a second side models the rhetorical block structure of the

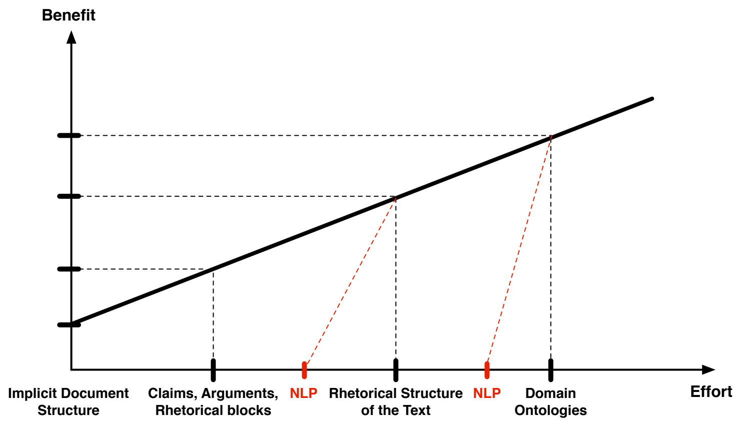

Fig. 4. SALT Process incremental approach

publication, while (iii) the last one, modeling the argumentative discourse over multiple publications, can be seen as a communication between different authors.

One of the main differences that distinguishes SALT from other semantic authoring solutions is its incremental approach. The SALT process was designed from the beginning to be flexible and entirely driven by the author. As shown in Fig. 4 the benefit that an author receives by using different parts of SALT is increasing proportionally with the amount of effort it is involved into the authoring process. While we do believe in a quasi-linear increase in benefit, note that this figure is only an approximation of the effort/benefit ratio we conjecture; the figure does not show an actual measurement.

For example, without investing any effort at all, the author still gains value from SALT, because it automatically extracts the shallow metadata of the publication together with its logical structure and references. Used within KonneX$^{\mathrm{SALT}}$, using only this information would make it equivalent to CiteSeer or DBLP in terms of functionality. After this, adding simple annotations like the claims, arguments or rhetorical blocks the value brought to the user increases considerably. As a result, KonneX$^{\mathrm{SALT}}$ provides specialized search and navigation of the argumentation discourse. Continuing in the same direction, the effort would increase, but in return we would improve the existing functionalities and maybe even discover unexpected features. At the same time, we would argue that by automating the process with the help on NLP techniques, the authors' effort would decrease substantially[5].

4.2 Argumentative Discourse Networks

In [2] we defined a semiotic approach for modeling claims and their textual representations, called *claim identification tree*. The tree has three levels: the root

[5] By introducing NLP techniques, the effort/benefit ratio would show a non-linear development.

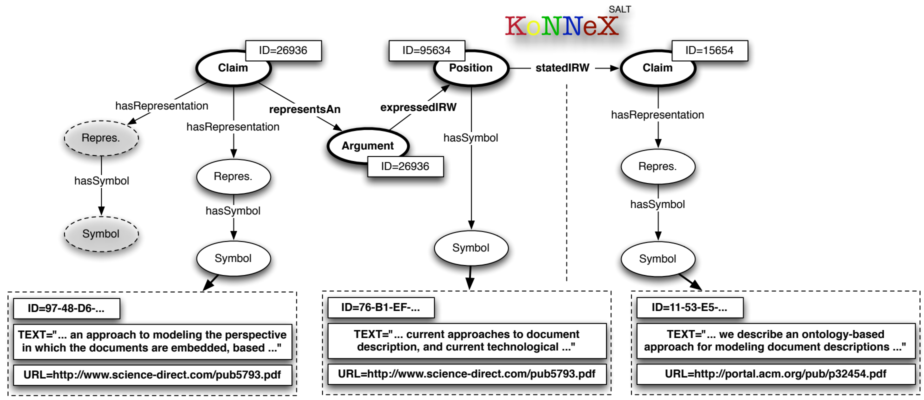

Fig. 5. Argumentative discourse network in KonneX$^{\text{SALT}}$

of the tree, representing a claim at an abstract level, a representation level linking multiple representations to the abstract claim and a symbolic level pointing to the actual textual claim in different publications on the Web.

Continuing in the same direction, we take the modeling of claims one step further and consider the associative trails (see Sect. 2) captured in the argumentative discourse between authors in different publications. Fig. 5 depicts the way in which the example introduced in the previous section is modeled as an argumentative discourse network.

4.3 Latent Semantic Indexing

Latent Semantic Analysis (LSA) represents a technique used in natural language processing (NLP) which takes advantage of implicit higher-order structure in the association of terms with documents, in order to improve detection of relevant documents on the basis of terms found in queries [14]. Often, in its application in information retrieval, this technique is also called Latent Semantic Indexing (LSI). The basic LSI process represents documents as vectors in a multidimensional space and calculates relevancy rankings by comparing the cosine of the angles between each document vector and the original query vector.

Some of the proved limitations of LSA motivated us in using this technique for federating claims in KonneX$^{\text{SALT}}$: (i) long documents are poorly represented, due to the small scalar product and the large dimensionality – in our case each document is represented by a claim, i.e., a phrase or sentence with a maximum length of around 20 words; or, (ii) documents with similar context but different term vocabulary will not be associated – the different representations of the claim usually share a common (or similar) term vocabulary, due to the fact that they all describe the same abstract claim.

5 Semantic Claim Federation Infrastructure

This section gives an overview of the design of KonneX$^{\mathrm{SALT}}$, including a brief description of the publication registration process, which is respondible for the linking data use-case described in Sect. 3.2 and of the data visualization issue in the context of a pure Semantic Web application.

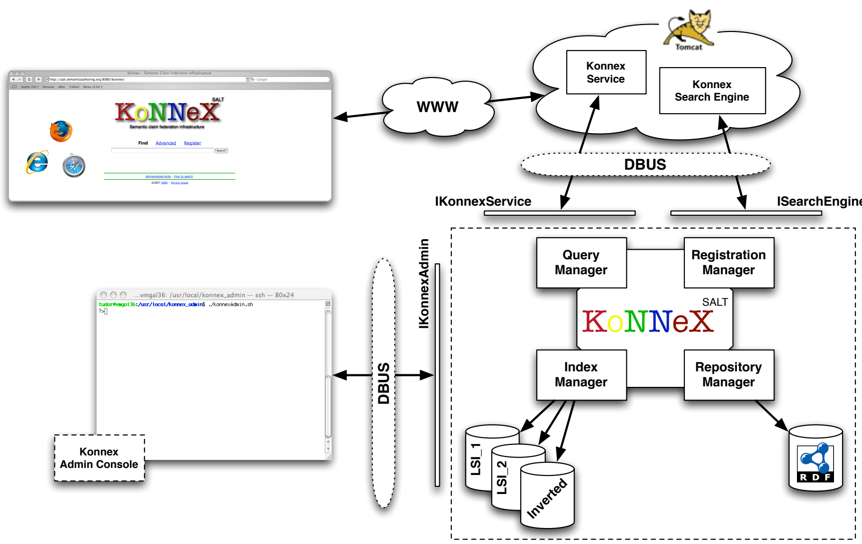

Fig. 6. Overall architecture of KonneX$^{\mathrm{SALT}}$

Fig. 6 depicts the overall architecture of the semantic claim federation infrastructure. The core part is composed by the four managers (*Query*, *Registration*, *Index* and *Repository*), together with the associated indexes and the RDF repository. Our goal was to have a clean design and clearly distinguish between the functional modules placed outside the core. Thus, due to the existing overlaps between some of the exposed functionalities, we decided to use DBus [15] as a communication infrastructure. As a consequence, all the modules reside on the server, but each module deals with its own particular functionality. The core of KonneX$^{\mathrm{SALT}}$ acts a as daemon and exposes three different interfaces over DBus, thus limiting the access to a particular set of functionalities for a particular module.

The four managers comprising the core are: (i) the *Query Manager*, which acts as a query distribution and results merging hub for the Index Manager and the Repository Manager; (ii) the *Registration Manager*, which analyzes the publications to be registered, creates links to other publication repositories and populates the local indexes and RDF repository. For each publication, it extracts the SALT metadata and, based on the title-authors information, for linking purposes, it requests more data from DBLP and Bibsonomy. Due to possible errors that might appear in the shallow metadata of a publication, the results received from the two sources are filtered to a certain degree of similarity; (iii) the *Repository Manager*, which manages the RDF repository, which in our case is a wrapped Sesame [6] instance. Finally (iv) , the *Index Manager* orchestrates a set of three indexes: two latent semantic indexes (InfoMap [16] and Semantic Engine [17]) and one inverted index. The inverted index is used to provide the link from the textual representation to the root of the RDF graph modelling the claim. The final ranking of a query is computed as a weighted average over the rankings coming from the LSIs.

Outside this core box there are three independent modules which provide separate functionalities to end-users: (i) the *Konnex Service*, which represents a RESTful Web service handling query requests, both full-text and URI-based. It can be used as an API for interacting with the semantic claim federation infrastructure, but also as a mechanism to retrieve URIs in usual web browsers. In the context of the URIs, we adhere to the Linking Open Data principle, and make them retrievable. (ii) The *Konnex Search Engine* acts as the main access point to the system for end-users. It is accompanied by a Web interface which allows searching for scientific claims and to register publications. (iii) The *Konnex Admin Console* is a utility for administrating the system. It provides capabilities for tuning the systems' parameters and querying the system's status.

As mentioned before, we argue that KonneX$^{\text{SALT}}$ represents a pure Semantic Web application. The data flow between the core of the infrastructure and the other modules is strictly based on RDF graphs. Here, we also include the flow between the core and the end-user Web interface of the search engine. As a consequence, when retrieving a URI or performing text-based queries, the result will always be represented as an RDF graph. Rendering such an RDF graph for a human user in a Web browser is handled by using an XSL stylesheet. In this way, the burden of transformation is moved from the server to the client: the actual process of transformation is executed by the local web browser via its built-in XSLT processor, the RDF result having attached an XSL style sheet. This method helps us to avoid duplication of data and provides flexibility for both machines and humans, because it can be easily read by a human, and at the same time be analyzed by an application or fetched by a crawler.

6 Discussion

In this section we present two interesting issues which appeared during our research. First, we deal with the sensitive issue of copyright and transcopyright. Then we discuss how domain knowledge can improve search results and navigation in KonneX$^{\text{SALT}}$.

[6] http://www.openrdf.org/

6.1 Copyright and Transcopyright

One of the main issues raised by implementing transclusion is the copyright of the included material. This was discussed for the first time by Ted Nelson and included in his list of 17 original rules of Xanadu [5]. The specific rules considering copyright were: (i) Permission to link to a document is explicitly granted by the act of publication; (ii) Every document can contain a royalty mechanism at any desired degree of granularity to ensure payment on any portion accessed, including virtual copies ("transclusions") of all or part of the document.

Nelson was the one to provide also a solution to the copyright issue, by defining *Transcopyright* [18], or the pre-permission for virtual republishing. He mainly proposed the introduction of a particular format to mark transcopyright, as an adaptation of the traditional copyright notice. For example, from (c) 1995 DERI to Trans(c) 1995 DERI.

In the case of KonneX$^{\mathrm{SALT}}$, due to the author-driven approach, the copyright of the text used in transclusion has an implicit assumption: the author of the publication, by registering it with our system, grants us the right to cite and re-use the text for achieving our purposes. This assumption follows the common author-driven publishing trend, which traditional search engines also take advantage of, i.e., the authors themselves are the ones who expose their publications on the open Web and thus give everyone the right to read[7] the publications. Still, in order to maintain and exhibit the copyright of the material, we will adopt and implement a solution based on the approach proposed by Ted Nelson.

6.2 Improving Results Based on Domain Knowledge

The functionalities KonneX$^{\mathrm{SALT}}$ provides are driven by the way in which the authors use the SALT framework. As previously discussed in Sect. 4, the benefit an author gains by using SALT is proportional to the effort spent on creating the annotations. Consequently, the more features of SALT are used, the more functionalities KonneX$^{\mathrm{SALT}}$ can offer. An important aspect of the semantic authoring process is the use of domain knowledge. In a previous publication [19] we have described the support for annotating rhetorical elements in scientific publications with domain knowledge. In this section, we present an example where domain knowledge helps to improve the search results and the navigation of the argumentative discourse network.

Considering the bio-medical domain, we assume that the publications, were enriched with concepts present in a disease domain ontology. This could be performed either manually by the author, or semi-automatically by using a specialized tool. As a direct consequence, internally to KonneX$^{\mathrm{SALT}}$, the rhetorical elements can be clustered not only based on linguistic information but also based on the attached domain concepts. Obviously, we would need an ontology import mechanism to make the system aware of the particular ontologies. In terms of enhancements, we foresee two new features:

[7] In this particular case, when mentioning the read access, we also include citation, crawling and all other operations that re-use the text without modifying it.

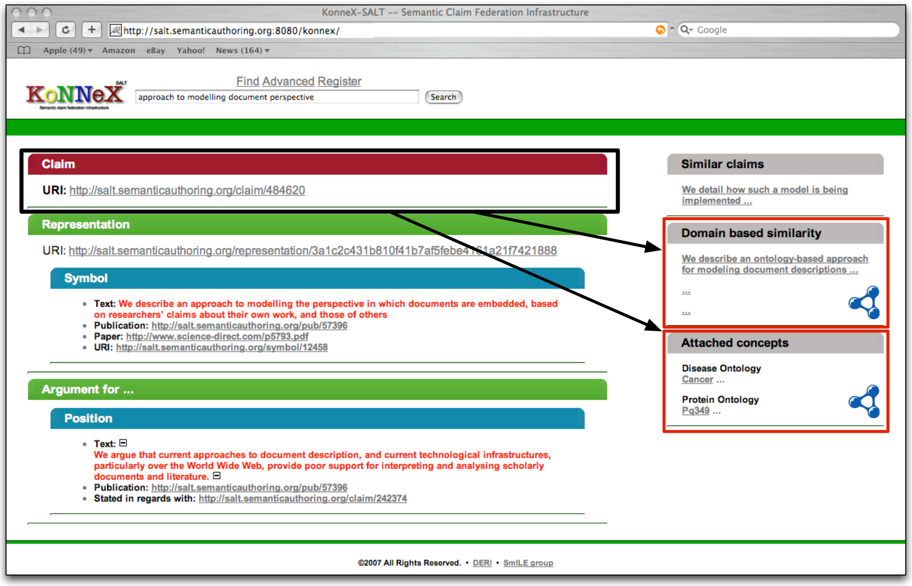

Fig. 7. Domain knowledge incorporated into the browsing process

- a special search syntax to take into account the concepts in the ontology. Such a feature could be used in two different ways: either to prune the search results based on the specified concepts (see the first example below), or to expand the search to a certain lens created around the specified concepts (the second example below).
  E.g. 1 $:< about : prot\#P3456\ \&\ about : prot\#P6734 > (\ldots\ text\ \ldots)$ — Search for *text* which has attached information only about the two proteins: `P3456` and `P6734`.
  E.g. 2 $:< about : prot\#P3456 \mid 1 > (\ldots\ text\ \ldots)$ — Search for *text* which has attached information about protein `P3456`, but consider also the direct parent concept and direct subconcepts from the ontology hierarchy.
- attached concepts to the argumentative discourse network. The dual navigation (depicted in Fig. 7) could improve the understanding of the reader not only in the line of argumentation but also following the covered domain concepts. This feature would open up to the possibility of integrating a faceted browser as part of KonneX$^{\mathrm{SALT}}$.

7 Conclusion

In this paper we presented KonneX$^{\mathrm{SALT}}$, a semantic claim federation infrastructure geared towards finding claims in scientific publications and facilitating the navigation of argumentative discourse networks, starting from a root claim. Our main goal is to provide a look-up service for scientific claims and not yet another traditional search engine. Hence, we distinguish our approach from others by adopting an author-driven method of populating the publication repository. In addition, we give the reader the possibility of following the argumentation line by realizing associative trails in conjunction with on-demand transclusion. Finally, we implemented KonneX$^{\mathrm{SALT}}$ as a pure Semantic Web application (i.e. all the data is being strictly modeled as RDF) and follow the Linking Open Data principle by linking instances of our data with data elsewhere on the Semantic Web.

Future developments of KonneX$^{\mathrm{SALT}}$ include: (i) using domain knowledge as part of the argumentative discourse network, (ii) providing transclusion for rhetorical blocks, (iii) linking deeper into the Web of Open Data (e.g. authors' links), and (iv) introducing an advanced search method inside rhetorical blocks based on a special syntax. All the above mentioned features represent basic enhancements of the infrastructure. In addition, we are considering o provide a way for browsing the publications present in the repository, based on a faceted navigation methodology like the one offered by Exhibit [20].

Acknowledgments

The work presented in this paper was supported (in part) by the European project NEPOMUK No FP6-027705 and (in part) by the Lion project supported by Science Foundation Ireland under Grant No. SFI/02/CE1/I131. The authors would like to thank Diana Trif, Laura Dragan, Renaud Delbru, Alexander Schutz and Andreas Harth for their support and involvement in the the development of SALT.

References

1. Nonaka, I., Takeuchi, H.: The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation. Oxford University Press (1995)
2. Groza, T., Möller, K., Handschuh, S., Trif, D., Decker, S.: SALT: Weaving the claim web. In: ISWC 2007, Busan, Korea (2007)
3. Bush, V.: As We May Think. The Atlantic Monthly **176(1)** (1945) 101–108
4. Nelson, T.H.: The Heart of Connection: Hypermedia Unified by Transclusion. Communications of the ACM **8** (1996) 31–33
5. Nelson, T.H.: Literary Machines: The report on, and of, Project Xanadu concerning word processing, electronic publishing, hypertext, thinkertoys, tomorrow's intellectual... including knowledge, education and freedom. Mindful Press, Sausalito, California, 1981 edition: ISBN 089347052X (1981)
6. Mancini, C., Shum, S.B.: Modelling discourse in contested domains: a semiotic and cognitive framework. Int. J. Hum.-Comput. Stud. **64**(11) (2006) 1154–1171
7. Uren, V., Shum, S.B., Li, G., Bachler, M.: Sensemaking tools for understanding research literatures: Design, implementation and user evaluation. Int. Jnl. Human Computer Studies **64, No.5** (2006) 420–445

8. Mancini, C., Scott, D., Shum, S.B.: Visualising Discourse Coherence in Non-Linear Documents. Traitement Automatique des Langues **47(1)** (2006) 101–108
9. Kunz, W., Rittel, H.: Issues as elements of information system. Working paper 131, Institute of Urban and Regional Development, University of California (1970)
10. Conklin, J., Begeman, M.L.: gIBIS: A Hypertext Tool for Team Design Deliberation. In: HYPERTEXT '87: Proceeding of the ACM conference on Hypertext, ACM Press (1987) 247–251
11. Conklin, J., Begeman, M.L.: gIBIS: A Hypertext Tool for Exploratory Policy Discussion. In: CSCW '88: Proceedings of the 1988 ACM conference on Computer-supported cooperative work, ACM Press (1988) 140–152
12. Tempich, C., Pinto, H.S., Sure, Y., Staab, S.: An Argumentation Ontology for Distributed, Loosely-controlled and evolvInG Engineering processes of oNTologies (DILIGENT). In: ESWC 2005, Heraklion, Crete, Greece, May (2005) 241–256
13. Marshall, C.C., Halasz, F.G., Rogers, R.A., William C. Janssen, J.: Aquanet: A Hypertext Tool to Hold Your Knowledge in Place. In: HYPERTEXT '91: Proceedings of the third annual ACM conference on Hypertext, ACM Press (1991) 261–275
14. Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K., Harshman, R.: Indexing by Latent Semantic Analysis. Journal of the Society for Information Science **41(6)** (1990) 391–407
15. Palmieri, J.: Get on D-BUS. Red Hat Magazine **3(1)** (2005)
16. Infomap NLP Software: An Open-Source Package for Natural Language Processing. http://infomap-nlp.sourceforge.net/ (2007)
17. The Semantic Indexing Project: Creating tools to identify the latent knowledge found in text. http://www.hirank.com/semantic-indexing-project/index.html (2007)
18. Nelson, T.H.: Transcopyright: Dealing with the Dilemma of Digital Copyright. Educom Review **32:1** (1997) 32–35
19. Groza, T., Handschuh, S., Möller, K., Decker, S.: SALT - Semantically Annotated LaTeX for Scientific Publications. In: ESWC 2007, Innsbruck, Austria (2007)
20. Huynh, D., Karger, D., Miller, R.: Exhibit: Lightweight Structured Data Publishing. In: Proceedings of the 16th International World Wide Web Conference, WWW 2007, Banff, Canada, May 8-12. (2007)