Exploring Scholarly Papers Through Citations

Angelo Di Iorio
Department of Computer
Science and Engineering
University of Bologna, Italy
angelo.diiorio@unibo.it

Raffaele Giannella
Department of Computer
Science and Engineering
University of Bologna, Italy
raffaele.giannella@studio.unibo.it

Francesco Poggi
Department of Computer
Science and Engineering
University of Bologna, Italy
fpoggi@cs.unibo.it

Silvio Peroni
Department of Computer
Science and Engineering
University of Bologna, Italy
silvio.peroni@unibo.it

Fabio Vitali
Department of Computer
Science and Engineering
University of Bologna, Italy
fabio.vitali@unibo.it

ABSTRACT

Bibliographies are fundamental components of academic papers and both the scientific research and its evaluation are fundamentally organized around the correct examination and classification of scientific bibliographies. Currently, most digital libraries publish bibliographic information about their content for free, and many include the citations (outgoing and in some cases even incoming) to the papers they manage. Unfortunately no sophistication is spent for these lists: monolithic pieces of text where it is even difficult to tell automatically the authors, the title and publication details, and where users are provided with no mechanisms to filter and access full context of each citation. For instance, there is no way to know in which sentence a work was cited (the *citation context*) and why (the *citation function*).

In this paper we introduce a novel environment for navigating, filtering and making sense of citations. The interface, called BEX, exploits data freely available in a Link Open Dataset about scholarly papers; end-user testing proved its efficacy and usability.

Categories and Subject Descriptors

H.5.2 [**Information Systems**]: Information Interfaces and Presentations—*User Interfaces (D.2.2, H.1.2, I.3.6)*

Keywords

Scholarly data visualization; Information interfaces and presentation; Semantic publishing; Citations

1. INTRODUCTION

Bibliographies are the most distinguishable peculiarity of the scientific communication. Researchers as a habit situate their work within a community of like-minded colleagues

whose contribution is revered, appreciated, criticized, or possibly even ridiculed, but in any case explicitly acknowledged and mentioned, and the accuracy and completeness of such mentions constitute a substantial means to evaluate the soundness and acceptability of the scientific communication itself.

In fact, researchers spend a lot of time exploring bibliographies: when approaching new research areas, when reviewing papers, when writing papers, they use bibliographies for exploring a discipline and for building the network of semantic connections of its topics. It is not a case that digital libraries of scholarly papers - such as ACM DL, IEEE Xplore DL, etc. - show bibliographic references and incoming citations in separate lists that can be freely examined often even when the full text of the paper is not publicly available.

Yet, digital libraries publish bibliographies as monolithic units that cannot be easily processed as separate and autonomous items. Consider, for instance, the frequent task of checking the freshness of references of an academic paper: there is no better way to do this than inspecting each reference and manually filtering outdated ones. Similarly, counting self-citations requires users to manually inspect the references and item by item compare the authors to those of the one providing the bibliography. While these tasks are not a big deal for a single paper, they become fairly time-consuming and difficult when dealing with a big number of papers, such as when dealing with the full set of submissions to a conference or a special issue of a journal. An aggravating factor is that these checks and evaluate tasks often have to be completed under some serious pressure: who cannot confess of having written a review right before the (extended) deadline? Or having made the initial selection of accepted papers for a conference just a few hours before delivering the acceptance letter? Or even arranged a related works section in short time?

In this paper we discuss how to support researchers and academics in these daily tasks, making it easy to access data in bibliographies that otherwise are hard to find, inspect and filter. We break the problem in two parts.

- ***Rich data***: on the one hand, we need to identify and classify bibliographics so that rich data can be published about papers (e.g. author list, publication year, venue, etc.) and citation networks, both incoming and outgoing references.

In particular, in the last few years we witnessed an ever increasing interest in making bibliographic data available as Linked Open Data (LOD): DBLP++[1], JISC OpenCitation corpus (OCC) [20] and Nature Publishing Group Linked Data Platform (NPG LDP)[2] are just some representative examples of such effort. Unfortunately, although a lot of information is available in these datasets, the landscape is still fragmented and a lot of useful data are still missing or unevenly provided by publishers.

For instance, none of the datasets we are aware of contains information about the *citation contexts* - defined in [16] as "the sentences of the original papers where a particular work was cited" - or the *citation functions*, defined in [22] as "the reasons why papers are cited". Having such data can be very helpful to understand the nature of citations and to give them more or less importance.

- **Intuitive interfaces** allowing users to read and make sense of the bibliographic data - for instance, to filter data, to navigate paper collections, to explore citations, along with their functions and contexts, in an easy and intuitive way.

This work builds upon a framework for extracting and describing rich data about bibliographies presented in previous works and focuses on the second aspect, introducing a novel environment for accessing rich data about citations. To do so, we briefly present the Semantic Lancet Triplestore, a LOD dataset used as backend for BEX.

The structure of the paper is as follows. Section 2 describes some related works; some issues in daily tasks on citations, along with possible improvements, are investigated in Section 3; Section 4 gives some background on the Semantic Lancet Triplestore; BEX is fully describe in Section 5; the evaluation of the current prototype is presented in Section 6 before concluding the paper.

2. RELATED WORKS

A large variety of systems support the exploration of scholarly data, some of them providing an interface to a specific repository of bibliographic information, others integrating multiple data sources to provide access to a richer set of data and navigation functionalities. The most famous academic search engine is probably Google Scholar[3], which provides search and citation services, supporting access to the scholarly literature. DBLP[4] is another well-known computer science bibliography portal. CiteSeerX[5] focuses instead on large-scale integration and indexing of research papers, and includes mechanisms for suggesting relevant papers. These systems mainly focus on providing a good interface for publication search and are not designed to support sensemaking tasks in the academic domain.

On the contrary, Microsoft Academic Search[6] provides a variety of visualizations, including co-authorship graphs,

publication trends, and co-authorship paths between authors. In a similar way Arnetminer[7] also offers different visualizations and provides support for expert search and trend analysis. [10] is another visual analytics tool that provides multiple coordinated views in order to reveal trends, support the investigation of connections, and describe activities throughout conference communities.

BiblioViz [18] is based on the principle of minimal visual representation, and provides a compact, comprehensive and extensible system for visualizing bibliography information based on two techiniques: a bidimensional tabular view, and a tridimensional network representation. Action Science Explorer [4] is another interesting example of visual analytic tool that leverages statistics, citation text extraction, natural language summarization, advanced filtering capabilities and network visualization to see citation patterns and identify clusters.

3. RESEARCH-RELATED TASKS ON BIBLIOGRAPHIES

Several experimental studies have confirmed their importance of bibliographies for the research community. Just to name a few: [13] surveyed more than two thousand American researchers working on natural science, engineering and medical science and showed that their primary search tools are citation databases, followed by general search engines; [12] showed similar results by interviewing social science faculty members in Taiwan.

The community has also investigated how citations and papers are read, and how they are actually exploited for research-related tasks. [2] surveyed thousands of faculty members for a long period (28 years, since 1977 to 2005). The results were very interesting: the number of papers being read is constantly increasing but people tend to cite a limited set of papers. One of the reasons is that they rely more on searching than on browsing and give priority to those papers highly ranked in digital libraries. The same paradox was noticed by [5]: ranking papers by their incoming citations made researchers cite the same papers that others have cited, thus narrowing the overall citation patterns. The research also showed how the amount of time spent for reading each paper is ever decreasing. This is also connected to the large number of digital publications available today: the overall time for reading increases but people have to deal with a huge amount of information and cannot help reading them rapidly [7].

In this section we focus on some of these tasks, in particular to those that researchers perform on bibliographies. This set of tasks is the results of an analysis based on the experience of our research group and some discussions with colleagues of our department. We organized our discussion around the roles played by a researcher and the tasks performed for each role. The researcher is first of all a *reader*, who reads articles and navigate them through citations. The process of finding relevant works is very common when writing new papers and project proposals: indeed, a researcher is also an *author*. The research community also relies on peer-review processes, which require researchers to also act as *reviewers* or *editors*. When reviewing papers, a researcher uses citations and checks, for instance, the number of self-citations or the publication year of cited papers.

[1]http://dblp.l3s.de/dblp++.php

[2]http://data.nature.com

[3]http://scholar.google.com

[4]http://www.informatik.uni-trier.de/~ley/db/

[5]http://citeseerx.ist.psu.edu/

[6]http://academic.research.microsoft.com/

[7]http://arnetminer.org/

Table 1: Common tasks on bibliographies, for different classes of users.

No.	Role	Goal/Task
1	reader/author	**Goal (G1) :** build/update a bibliography [of recent papers] on a given topic • **T1a:** find authoritative papers on a given topic • **T1b:** find recent papers on a given topic • **T1c:** navigate citation networks to find other interesting/relevant papers
2	reviewer	**Goal (G2) :** evaluate if a paper is up-to-date • **T2a:** check the publication year of the cited papers • **T2b:** check if relevant papers are missing in the bibliography
3	reviewer, editor	**Goal (G3):** evaluate how much a work is self-referential • **T3:** count self-citations
4	reviewer, editor	**Goal (G4):** evaluate if a paper fits the scope of a journal/conference • **T4a:** read the abstract; if possible, the whole paper • **T4b:** check the bibliography and search cited papers published in the same venue • **T4c:** check the relevance of the other cited papers (published in other venues)
5	evaluator	**Goal (G5):** evaluate the impact of a paper • **T5a:** check how many times a paper is cited • **T5b:** how citations have evolved • **T5c:** check why a paper is cited
6	evaluator	**Goal (G6):** evaluate the impact of a researcher • **T6a:** check how many times the work of a candidate is cited • **T6b:** check how citations have evolved • **T6c:** check why a candidate is cited
7	event organizer	**Goal (G7):** find potential participants • **T7a:** find papers about topics relevant to the event • **T7b:** check authors of relevant papers

Citations are also increasingly used for evaluation purposes. Researchers are then asked to be part of academic boards in charge of evaluating research products. Our analysis covers such a scenario with a specific role: *evaluator*. Researchers are involved in organizing scientific events, thus we added the role of *event organizer*. Note that there are many other tasks characterizing each of these roles but they are left out of discussion, since they are not specific on bibliographies.

Table 1 summarises our analysis. The discussion uses the well-known terminology of [8]: the term *goal* is used to indicate a final objective a user wishes to achieve, a *task* is a sequence of one or more activities the user thinks are required to achieve a goal. Each task is specific of a peculiar user *role*. Note also that we did not decompose each task in hierarchical subtasks, for the sake of simplicity.

The first goal (G1) consists in building or updating a bibliography on a given topic. A common approach is to use keyword-based search tools, scan the list of returned papers and filter potentially relevant ones (T1a, T1b). Researchers might read abstracts to have a clearer idea of each paper. Further candidates are often searched by analysing citation networks (T1c), as shown by several experimental studies [12] [23]: starting from a list of relevant papers, incoming and outgoing citations are scanned and, in turn, new candidates are analysed and, if relevant, added to the list. The criteria used to discern among papers may vary a lot: considering the total number of citations may be a suitable indicator to identify seminal papers, whereas it may fail when searching for recent developments or cutting edge works. In the second case, other criteria such as the publication date are more convenient. In all cases, the researchers have to

inspect the list of citations and to manually find relevant information.

The second group of tasks are performed by researchers as reviewers. To evaluate if a paper is up-to-date (G2), they inspect the bibliography and look at the publication year of each item (T2a). Furthermore they check if any relevant recent paper is not cited (T2b). These operations are still manual. One of the reasons is that, in almost all cases, papers under revision are PDFs and the bibliographic references cannot be processed as separate units.

The same applies to following two tasks. The first one covers the analysis of self-citations in order to understand how much a work is self referential (G3): in fact, the reviewers/editors check how many cited papers are written by the same authors, or some of them (T3). The editors/reviewers might also want to check the number of papers cited by a paper submitted to a journal and published in the same journal (T4b). That might be an indicator of the suitability of that paper for that journal (G4).

The evaluation of research works (G5) and researchers (G6) is often performed by counting the incoming citations to each research work and by aggregating them for further analysis. Thus, evaluators need to easily access information about the number of times each paper is cited (T5a). It is also interesting to know how citations are distributed over time (T5b). Tasks 5 and 6 are widely supported today. One aspect that is still under-estimated is the nature of the citations. In fact, not all citations are equal: some are given just for information, some as necessary background, some refers to a work the citing paper is an extension of. It would be useful to store and be able to analyse citations together with

the sentence to which they belong to (*citation context*), or together with information about the reason why a paper is cited (*citation function*, T5c). A time perspective is helpful here. For instance, we can expect that an important work is referenced for information or used as background for many years, but if it keep being extended, or some methods therein contained are used for a long time, it means it still plays an active and important role.

There is a further common task that researchers perform by exploring papers and citations: finding experts in a given topic or simply people who are interested in that topic (T7a, T7b). These people usually are searched among the authors of relevant papers — that, as discussed so far, are found by exploiting citations too. Such expert finding task might be useful, for instance, when searching for potential participants to a scientific events (G7).

3.1 Performing tasks with existing tools

These tasks are successfully performed today by accessing digital libraries and databases of bibliographic records. Nonetheless searching the relevant information for a given task is not always simple and direct. The problem can be studied from two orthogonal perspectives: *availability of data* and *interface.*

First of all, it depends on the way data are structured and made available. In fact, most of the existing repositories do not offer APIs to access data (e.g. Google Scholar) or offer partial APIs that do not cover all needed information (e.g. ScienceDirect and Scopus). Consider, for instance, the case of citation contexts and functions. Such data would be valuable tools to better characterize citations and to assess their impact. Though, none of the platforms we are aware of make these data available.

Even if some information is stored in the repositories, it is not directly processable by the users. Let us consider the list of bibliographic references. These lists are often made available as separate content from the full-text of the papers, for instance in ACM or IEEE Xplore DLs (where they can be freely accessed even if the PDF is subject to payment); on the other hand, they are treated as 'monolithic' units: the users can export citations and navigate to the cited/citing papers but they are not allowed to show only some references or to quickly access some information. If a researcher wants to read the abstract and the bibliography of a cited paper, for instance, he/she has to open it a separate tab and to interrupt the navigation. Similarly, there is no way to filter items in bibliographies automatically.

Some filtering capabilities on bibliographies are actually already available to the users. For instance, Elsevier's Scopus allows users to order the lists of references and to apply various filters. Such ordering options, on the other hand, are available only on lists of papers returned by a search (for instance, papers on a given topic, or written by a given author, etc.) or lists of papers citing a given one. It is not possible to work on the whole bibliography of a paper or to keep filters active while surfing bibliographies.

In the next sections we present a novel interface, called BEX, that makes all these actions straightforward and that helps researchers in completing most of the tasks discussed so far. To do so, we also need to introduce the Semantic Lancet Triplestore, a rich knowledge-base on scholarly papers and citations that is used as backend for BEX.

4. EXTRACTING AND MODELLING CITATIONS: SPAR AND SEMANTIC LANCET

In the last few years, different research groups in both academia and industry have started to work for pushing the research communication and the scholarly publishing to the next step of evolution of the whole publishing domain, characterised by the *Semantic Publishing* movement. Basically speaking, Semantic Publishing stands for the use of Web and Semantic Web technologies to enhance a published document such as a journal article so as to enable the definition of formal representations of its meaning, facilitate its automatic discovery, enable its linking to semantically related articles, provide access to data within the article in actionable form, and allow integration of data between papers [14].

Along the lines of the Semantic Publishing movement and guidelines, we have started the *Semantic Lancet Project* (`http://www.semanticlancet.eu`), which is focused on building a Linked Open Dataset of scholarly publications. The aim of the project is twofold. On the one hand, we have developed a series of scripts that allow us to produce proper RDF data compliant with the *Semantic Publishing and Referencing (SPAR) Ontologies* (http://www.sparontologies.net) [14], which are a suite of orthogonal and complementary ontology modules for creating comprehensive machine-readable RDF metadata for all aspects of publishing domain. On the other hand, we have made publicly-available the *Semantic Lancet Triplestore (SLT)*, i.e., a freely available LOD dataset that includes rich data about scholarly papers, that range from a large network of citations (that also includes citation contexts and functions) to semantically-enriched abstracts, from provenance data to time-aware descriptions of the scientific production. Currently the STL contains data about all papers in the Web Semantics journal published by Elsevier, but our plan is to extend it incrementally.

The SPAR Ontologies that are used in the SLT for describing the particular aspects of interest for this paper (i.e., citations and abstracts of papers) are the following:

1. the *Citation Typing Ontology (CiTO)*[8] is an ontology that enables the characterisation of the nature or type of citations, both factually and rhetorically;

2. the *Citation Counting and Context Characterisation Ontology (C4O)*[9] is an ontology that permits the number of in-text citations of a cited source to be recorded, along with the number of citations a cited entity has received globally on a particular date;

3. the *Document Components Ontology (DoCO)*[10] is an ontology that provides a structured vocabulary of document components, both structural (e.g., block, inline, paragraph, section, chapter) and rhetorical (e.g., abstract, introduction, discussion, acknowledgements, reference list, figure, appendix), enabling these components, and documents composed of them, to be described in RDF.

As shown in Fig. 1, several entities defined in the aforementioned ontologies have been used in the SLT for modelling citations and abstracts of papers in a semantic fashion.

---

[8]http://purl.org/spar/cito
[9]http://purl.org/spar/c4o
[10]http://purl.org/spar/doco

In particular, they allow us to describe four different kinds of objects that are relevant for this work:

- *in-text reference pointer*, i.e., the entity present in the body text (e.g., "[5]") of a citing work that denotes a particular bibliographic reference in the reference list or a footnote. In scientific literature, this in-text reference pointer can be presented in different forms;

- *citation context*, i.e., the textual content of that component of the published paper (i.e., the sentence) within which an in-text reference pointer appears, which provides the rhetorical rationale for the existence of that citation;

- *citation function*, i.e., the author's reason for citing a given paper, that can be either factual (e.g., if the author of the citing entity describes work that uses a method detailed in the cited entity) or rhetorical (e.g., if the author of the citing entity agrees with statements or ideas presented in the cited entity);

- *abstract*, i.e., the text of papers abstract and its representation in RDF according to ontology design patterns and linguistic frames.

Figure 1: The Graffoo diagram [6] of the ontological entities used in the SLT for describing citation functions, in-text reference pointers and papers abstract.

As example, the RDF representation (in Turtle) of the abstract of a paper included in the SLT and of one of its citation (including its function) is available on-line[11].

The semantic characterisation of abstracts and the specification of citation functions are created automatically starting from the actual content of the paper in consideration. The semantic abstracts are generated by a module of the Semantic Lancet Project developed for this purpose. In particular it relies on FRED[12] [15], which is a tool that implements deep machine reading methods based on Discourse Representation Theory, Linguistic Frames, and Ontology Design Patterns, for deriving a logical representation (expressed as OWL) of natural language sentences. The output of this module is then used by the *Abstract Finder* (`http://www.semanticlancet.eu/abstractfinder`), which is a service for searching relevant papers according to their textual and semantic abstracts, by exploiting the semantic information about concepts, events, roles and named entities included in the semantic abstract.

The citation functions are extracted by another module that allows one to capture them by means of *CiTalO* [3]. CiTalO is a chain of tools for identifying automatically the nature of citations according to CiTO, in a way that is comparable with humans. A run of this module queries the SLT by finding all the in-text reference pointers (and the related citation sentences) and links them to their related citation functions as returned by CiTalO. The output of this module together with the Abstract Finder service are used by BEX for supporting the navigation of the SLT.

5. EXPLORING SEMANTIC LANCET DATA ON CITATIONS: BEX

The Bibliography EXplorer (BEX)[13] is an interactive web-based tool aimed at supporting the analysis, exploration and sense-making process of papers and citations available in the Semantic Lancet Triplestore. The BEX design and the consequent user interaction are driven by Shneiderman's Information Seeking Mantra [19]: "Overview first, zoom and filter, then details-on-demand".

The navigation starts with two search functionalities: besides searching a title, the user can also search relevant papers according to their content. This search is performed by calling the Abstract Finder service[14] described above. Other services, for instance based on full-text search or other mining techniques, can be integrated with a few modifications to BEX. Thus, through BEX a researcher can write in the search box a tentative abstract for her/his paper to retrieve meaningful papers that match with it from a pure textual but also semantic point of view. Fig. 2 shows the main interface of BEX and the output of a search.

Search results are organized as a list of papers, ordered by default from the most recent to the oldest one. Through the sorting box at the top of the interface, the user can easily define custom criterion to order the results (i.e. year, number of citations) and the order type (i.e. ascending or descending). For each returned paper, BEX shows a summary of basic information (e.g. title, publication year, author list,

---

[11]http://eelst.cs.unibo.it/DOCENG-2015/slt-example.ttl
[12]FRED: http://wit.istc.cnr.it/stlab-tools/fred
[13]http://eelst.cs.unibo.it:8089/
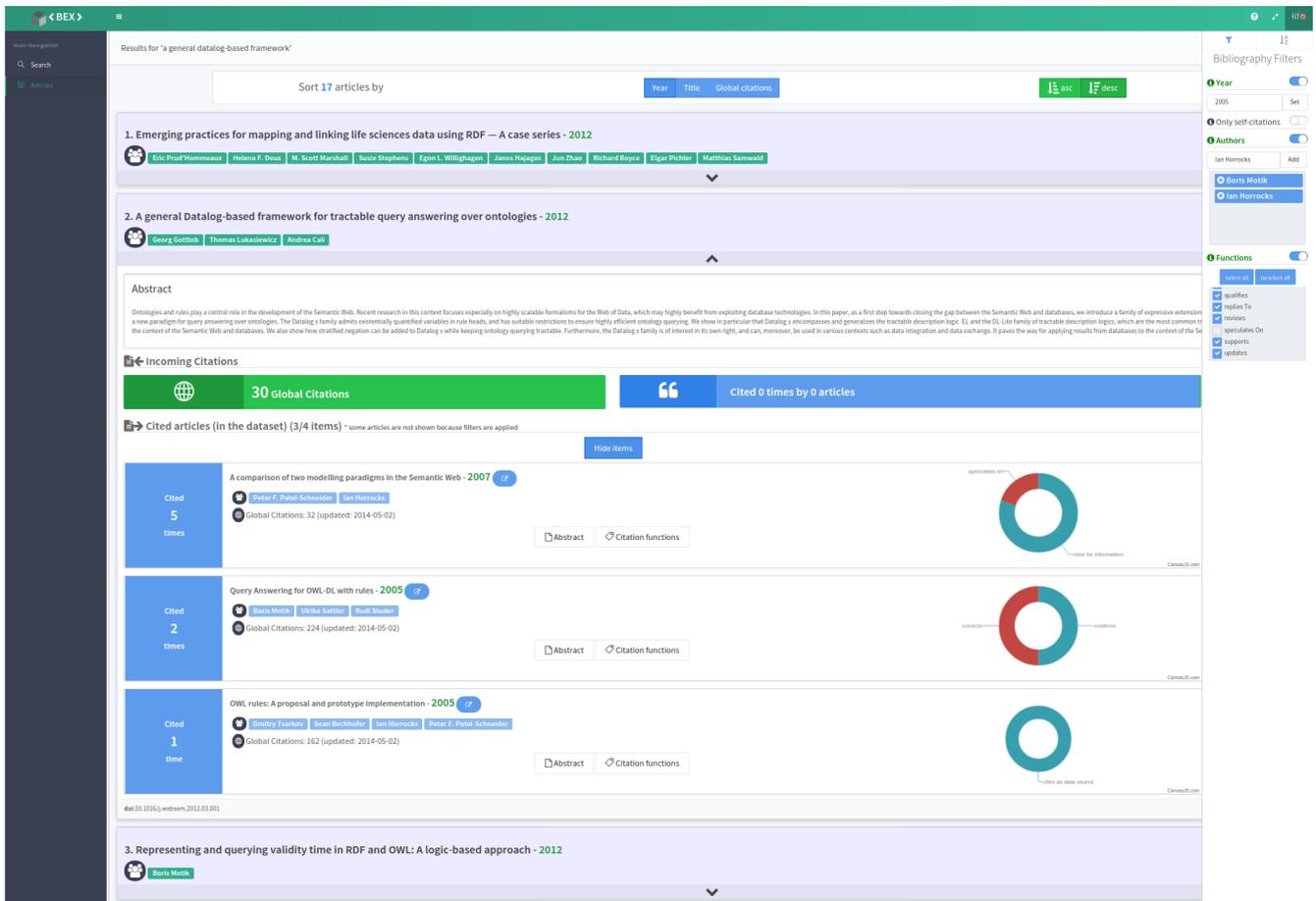[14]http://www.semanticlancet.eu/abstractfinder

Figure 2: The main interface of BEX and the list of papers returned by a search. Abstracts and data on citations can be collapsed/expanded on the user's demand.

etc.) and a link to the paper official page on Elsevier's ScienceDirect.

In order to gather more information about a paper in the list, the user can open a sliding box showing the full abstract and data about citations, organized in two separate sections: *outgoing* and *incoming*.

By clicking on the "Show Items" button, the user can get access to the information about the *outgoing citations*, as shown in the central part of Fig. 2. BEX organizes the cited papers in a vertical list, and for each paper it shows, from left to right, the following information: the number of times in which the paper is referenced by the paper under examination, some general information about the paper, and a piechart summarizing the number and type of citations received from the focus paper. Moreover, abstract and citation contexts are shown in popups windows.

In the *incoming citation* section, two counters show the number of *global* and *internal* citations received by the paper under examination. The term 'global' here indicates citations for a paper as counted by external services (Scopus); the term 'internal' indicates the citations given by papers described in our dataset (published in the Journal of Web Semantics).

Further details about the citation functions of incoming internal citations are available. This information is pre-sented in a popup window organized in three parts: a pie chart gives an overview of the number and type of incoming citations (top left), a column chart shows the distribution of the citation functions on a time axis (top right), and details about the citation contexts are presented in the bottom. In the two charts at the top of the page, different colors are used to encode the function of each single citation, and citations with the same function are grouped together. Finally, the last component shows, for each paper citing the paper under focus, the list of the citation contexts.

6. EVALUATION

We performed an experimental evaluation in order to assess the efficacy and users' satisfaction in using BEX for research tasks. The tests involved 19 users, with different background and skills.

The testers were asked to use the system[15] to complete four assignments and to fill some questionnaires about their experience. The real test was preceded by a warm-up task, in which the testers were asked to read a short-guide about BEX and to explore it. They were given exactly 5 minutes to become confident with the system.

_____

[15]Testers were asked to only use BEX, without any other external tool.

The test included four assignments, each mapped to one of the goals discussed in Section 3 and summarized below:

A1. **Build a bibliography of recent papers on a given topic (Goal G1) :** the testers were asked to find the most cited and the most recent paper about a given topic, together with a list of five papers citing/cited by those two and relevant to the same topic. We expected users to search papers by abstract similarity and to read their bibliographic information (title, publication year and abstract) to complete the assignment.

A2. **Check if a paper is up-to-date (G2):** the testers were given the title of a paper and were asked to indicate how many of the papers cited by that one have been published since a given date. The assignment could be then completed by searching the paper in BEX and inspecting its bibliography. As discussed in Section 3, such a task is very common when reviewing papers. On the other hand, papers are usually submitted as PDF and BEX cannot be used directly. This is not an issue for the purpose of our test, since we aimed at evaluating the interface only. We used one random paper, already loaded in the dataset. The issue is anyway very interesting and challenging. It could be addressed by implementing automatic processes that extract relevant information on-the-fly and produce a set of (semantic) data to be shown in BEX. We are investigating the integration of existing services like PDFX[16] towards this goal. Note also that these observations are valid for the next assignment as well.

A3. **Evaluate how much a work is self-referential (G3):** given the title of a paper, the testers were asked to indicate how many self-citations it contains and how many times it was cited by papers written by the same authors (or some of them). Even this assignment can be completed by searching the paper by title and inspecting both the bibliography and the list of papers citing it or using filters.

A4. **Evaluate the impact of a paper (G5):** the testers were asked to indicate how many times a paper with a given title was cited and for which reasons (in some specific years and overall). The citation counters and functions provided by BEX were expected to help users to complete this assignment.

Note that we have not included any assignment for goal G4 (evaluate if a paper fits the scope of a journal) since SLT only contains the papers published in one single journal at this stage. We also decided to exclude goals G5 and G6 for this test session. The reason was that the corresponding assignment would be too long, requiring users to iterate the same operations on all papers written by a given author, and we wanted to reduce the effort and time required to the testers. The current implementation of BEX, in fact, is mainly focused on the exploration of papers instead of authors. That exploration is anyway very interesting and one of future directions of our research.

Note also that the assignments here are numbered and presented in the same order of their corresponding goals. They were actually given in a different order (by increasing

---

[16]http://pdfx.cs.man.ac.uk/

difficulty and duration) but this does not affect the final result.

We also asked testers to indicate which features of BEX they used to complete each assignment, by filling a list of options shown at the end of each assignment – equal for all assignments, thus containing even non-required features, with no suggestion to the testers. The answers to these questions gave us very interesting indications about the use of the interface. These results will be discussed at the end of the section, after investigating BEX efficacy and usability.

6.1 Efficacy

We first studied if testers could complete the assignments with BEX, by manually comparing the answers given by each tester to the expected answers. Though some participants failed to complete some assignments, the overall efficacy was high. Results are summarized in Table 2.

Table 2: Testing efficacy on all four assignments. The values for assignment A1 are partial since there was no fixed expected output (open assignment).

Assign.	Success	Partial	No answer
A1	75,33% *	9,00% *	15,67%
A2	79,00%	21,00%	0,00%
A3	84,50%	10,50%	5,00%
A4	80,67%	12,33%	7,00%

The efficacy on assignment A2 (check if a paper is up-to-date) was lower than we expected. Some users, in fact, listed all papers in bibliography without taking into account their publication year. Only one tester listed one single paper, probably due to a wrong interpretation of the question. Nonetheless, no one complained that the available information was not enough to complete the assignment.

The efficacy on assignment A3 (check if a work is self-referential) was very high. Though not shown in the table, in fact, all users apart from one identified correctly the number of self-citations given by the input paper. The testers confirmed that the presence of visual hints and clear information about self-citations was very helpful to complete the task. The overall score decreases to 84.50% since some users failed in identifying self-citing papers among those citing the input one. Since 4 testers over 19 gave the same value, lower than the expected one, we believe that the testers did not wait for having the full list of citing papers or received only a partial list from the server. A clearer hint about the loading process and some more control is then needed, as discussed in the next section.

The efficacy on assignment A4 (evaluate the impact of a paper) is worth discussing in detail. The overall score is quite high (80.67%). Nonetheless some testers misinterpreted the question about the overall number of citations and used different counters to answers: some reported the number of global citations, others the number of internal citations. Only 13 testers over 19, in fact, answered correctly. This affected the overall efficacy for this assignment. Considering only the questions about citation functions, in fact, the success is higher than 90%.

We discuss A1 at the end since it is the most critical one. The efficacy was acceptable, though not very high. One user

experienced some problems in completing the assignment, due to technical problems in loading the content. Apart from that, all others were able to find the most cited and the most recent papers. Limiting to these two questions, in fact, the efficacy score is 90%. Nonetheless, 8 users had some difficulties in searching relevant papers on a given topic and navigating the large amount of information available in the platform. In the next section we will discuss these usability issues in more detail.

6.2 Usability

After the completion of all assignments, the testers were asked to fill two questionnaires. The first one was a *System Usability Scale* (*SUS*) [1], a well-known questionnaire used for the perception of the usability of a system. It has the advantage of being technology independent and it is reliable even with a very small sample size [17]. The mean SUS score was 64.5 (in a 0 to 100 range). The value shows how the users' satisfaction was quite high though there are still some open issues as discussed in the rest of this section.

In addition to the main SUS scale, we also examined the sub-scales of pure *Usability* and pure *Learnability* [11]. They gave us a more precise characterization of the users' feedback. The mean values for Usability was 62.5 and Learnability was 72.4. This means that the system was perceived as easy to learn but some of its functionalities were not completely clear and users experiences some difficulties in completing the tasks.

In order to go deeper, we also included four open questions in the test:

- What were the most useful features of BEX to help you realise your assignments?

- What were the main weaknesses that BEX exhibited in supporting your assignments?

- Would you suggest to modify any part of BEX?

- Can you think of any additional features that would have helped you to accomplish your assignments?

We subjected the text answers to a qualitative analysis based on *grounded theory* methods. Grounded theory [21] is a method often used in social science to extract relevant concepts from unstructured corpora of natural language resources. In opposition to traditional methods aiming at fitting (and sometimes forcing) the content of the resources into a predefined model, grounded theory aims at having the underlying model emerge "naturally" from the systematic collection by rephrasing, reorganising and interpreting the actual sentences and terms of the resources. We thus believe it is a reasonable tool to examine our questionnaires in order to let important concepts emerge from the text.

We proceeded first with *open coding*, with the purpose of extracting actual relevant sentences – called *codes* – from the texts, and subsequently performed the so-called *axial coding*, which is the rephrasing of the original codes so as to have semantic connections emerge from them and generate concepts. We finally analysed the respective frequency of each emerged concept (defined as the number of codes which contributed to the concept's existence) so as to consider the most important issues arising from the answers. Figure 3 shows the codes that were mentioned by the users.
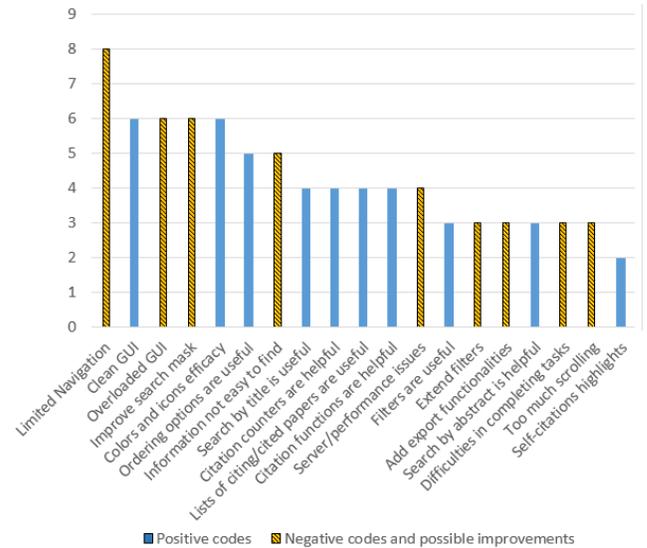

Figure 3: The codes mentioned by at least two users, ordered by their frequency and grouped in positive and negative codes. Negative codes also include suggestions for possible improvements.

We characterized codes as 'positive' (highlighting strengths and functionalities appreciated by the users) and 'negative' (highlighting current limitations and possible improvements).

The overall impression of the testers is encouraging, as shown by the distribution of positive/negative codes.

Some other interesting observations can be distilled from our analysis:

- **Clean GUI with several functions**. Several users appreciated the interface and the set of functionalities made available. In fact they classified as 'useful' most of the features: self-explaining icons (in particular, those on self-citations), citation counters and lists of citations, filters and search options, etc.. The presence of citation contexts and functions was also identified as a key and innovative contribution.

- **Information not easy to find.** The great amount of data and functionalities, on the other hand, confused some testers. Some of them complained that they had to do a lot of actions in order to get the information they were looking for. The fact that some boxes are collapsed by default (so that users have to manually open them) and some content is shown with a large font (and users have to scroll down and risk to get lost) was perceived as a strong limitation.

- **Limited navigation.** One of the most critical issues was the fact that tabs are disabled, so that users are not allowed to open multiple views and to freely surf the knowledge-base. The fact that details of the papers citing a given one are shown in a fixed modal window (that cannot be moved and resized, and requires users to scroll) was also criticized. Overall, the navigation was considered too basic by some users. The possibility of bookmarking papers and surfing the history were also suggested as possible improvements. This also

FEATURE	A1			A2			A3			A4			TOTAL	
search by title	10	53%		19	100%	X	18	95%	X	17	89%	X	64	112%
search by abstract	12	63%	X	0	0%		0	0%		0	0%		12	63%
ordering options on search results	16	84%	X	2	11%		2	11%		3	16%		23	121%
icons and tooltips	11	58%	X	4	21%	X	13	68%	X	12	63%	X	40	53%
paper publication info box	7	37%	X	3	16%	X	2	11%	X	3	16%		15	26%
abstract	3	16%		0	0%		0	0%		0	0%		3	
clickable list of authors	0	0%		0	0%		1	5%		0	0%		1	
bibliography filters	2	11%		2	11%	X	2	11%	X	0	0%		6	16%
bibliography ordering options	3	16%		0	0%	X	0	0%	X	0	0%		3	8%
breadcrumbs	0	0%	X	0	0%		0	0%		0	0%		0	0%
charts	3	16%		0	0%		0	0%		9	47%	X	12	63%
motivations for citations in textual form	2	11%		0	0%		0	0%		8	42%	X	10	53%
citation counters	12	63%	X	4	21%		6	32%		10	53%	X	32	84%
in-line help guide	0	0%	X	1	5%	X	0	0%	X	0	0%		1	1%
other	0	0%		2	11%		1	5%		0	0%		3	

Figure 4: The features of BEX used by the testers to accomplish the assignments.

tell us that the breadcrumb is not evident and should be made available during all navigation phases.

- **Improve search and filters.** Search capabilities were largely used during the test sessions. Nonetheless several users confirmed that the search mask needs to be extended (as a matter of fact we have already implemented further search options, not available at the time of the test). The suggestion is to include a *faceted search,* as already available in most of the existing digital libraries. Filters were also well appreciated and users suggested to extend them to also work on the list of papers citing a given one, the list of papers returned by a search or written by a given author.

- **Add export functionalities**. An interesting extension – that we admittedly had not taken into consideration – is to give users the possibility of exporting information about citations. Not only the lists of papers (both citing and cited by a given one) but also the citation functions and contexts. The idea is to process such information separately and to integrate it with other sources of data. The availability of a public SPARQL end-point for Semantic Lancet LOD, though not visible and relevant for the test, goes exactly in that direction.

- **Server/performance issues.** During some tests the system was particularly slow in loading content, making some tasks very difficult – and in a few cases impossible – to complete. It was due to some technical problems on the SPARQL end-point, since the current installation is not optimized to handle a large volume of requests (indeed we are migrating to a different platform).

6.3 Use of BEX features

The participants were also asked to indicate which features of BEX they used to complete the test. Fig. 4 summarizes their answers, indicating how many testers used each feature for each assignment. For each assignment we also indicate, in the third column, which features we expected testers to use. Note that some of these features were strictly required to complete the assignment (for instance, the "search by title" for assignment A2) while others were optional (for instance, the in-line guide might be helpful but the assignment could be completed without accessing it).

It is interesting to notice that the behaviour of the testers was not uniform. Let us consider for instance the assignment A2: although all users started searching by title, some of them exploited only the information provided by icons and tooltips, while a few others activated filters. Furthermore, it is interesting to investigate the cases in which testers did not behave as we expected. In the assignment A3, for example, all users searched the paper and most of them looked at the self-citation icon to recognize those written by the same authors, but only two testers used filters. An explanation of such a behavior is that, since the list of paper under examination was very short (only 6 items), the testers have preferred to scan it manually. It is also interesting to notice that the citation functions have been read in two alternative ways (for assignment A4): some users preferred the aggregated information available in charts, while others have accessed the citation contexts and motivations in textual form.

Finally, the last column of the table gives us very useful clues about the overall use of BEX features in performing different tasks. The first comforting consideration concerns BEX functionalities: citation counters, icons, charts and motivations for citations have been largely used by most of the testers. Another consideration is about some possible improvements of BEX interface. For example, none of the testers used bredcrumbs, that are the only mechanism provided by BEX to surf through the navigation history. This is a clear weakness that we have to take in consideration in the next version of the tool. Moreover, by comparing the wide use of the ordering options on search results (available by default in the center of the interface) to bibliography filters and ordering options (that have to be activated by clicking on a small icon), we understand that we should give more emphasis to these last useful functionalities.

7. CONCLUSIONS

In this paper we have introduced the *Bibliography Explorer* (*BEX*), i.e., an interactive web-based tool aimed at supporting the analysis, exploration and sense-making of papers and citations available in the Semantic Lancet Triplestore. In particular, after discussing the most common tasks that researchers perform on bibliographies of a research paper (according to their actual role, e.g., author, reader, reviewer),

we have introduced the main features of BEX and we have described the results of an user testing session. The evaluation of such results had shown that BEX is quite good in supporting researchers to address the aforementioned tasks, even if a few of issues still have to be solved in order to use the full potential of the tool appropriately. We are working to refine the tool, and we plan to perform a full comparative analysis with existing solutions.

We also plan to focus on two new development directions. On the one hand, we want to allow the use of BEX also with (RDF-based) datasets that do not comply with the models used in the Semantic Lancet triplestore, such as DBLP++, OpenCitation Corpus and NPG Linked Data Platform. On the other hand, we want to focus on enabling users to address the tasks typically associated with the evaluation of authors (related to goals G6 and G7).

8. REFERENCES

[1] Brooke, J. (1996). SUS: a "quick and dirty" usability scale. In Usability Evaluation in Industry: 189-194.

[2] Carol Tenopir, Donald W. King, Sheri Edwards, Lei Wu, 2009. Electronic journals and changes in scholarly article seeking and reading patterns, Aslib Journal of Information Management, 61:1, pp. 5-32.

[3] Di Iorio, A., Nuzzolese, A. G., & Peroni, S. (2013). Towards the automatic identification of the nature of citations. In A. Garcia Castro, C. Lange, P. Lord, & R. Stevens (Eds.), Proceedings of 3rd Workshop on Semantic Publishing (SePublica 2013), CEUR Workshop Proceedings 994: 63–74. Aachen, Germany.

[4] Dunne, C., Shneiderman, B., Gove, R., Klavans, J., Dorr, B. (2012). Rapid understanding of scientific paper collections: Integrating statistics, text analytics, and visualization. JASIST, 63(12), 2351-2369.

[5] Evans, James A. "Electronic Publication and the Narrowing of Science and Scholarship." Science 321, no. 5887 (2008): 395-99.

[6] Falco, R., Gangemi, A., Peroni, S., & Vitali, F. (2014). Modelling OWL ontologies with Graffoo. In ESWC 2014 Satellite Events - Revised Selected Papers, Lecture Notes in Computer Science 8798: 320–325. Berlin, Germany: Springer.

[7] Hélène de Ribaupierre and Gilles Falquet. 2011. New trends for reading scientific documents. In Proceedings of the 4th ACM workshop on Online books, complementary social media and crowdsourcing (BooksOnline '11). ACM, New York, NY, USA, 19-24.

[8] J. Preece, J., Carey, T. , Rogers, Y., Holland,S. , Sharp, H. Benyon,D. Human-Computer Interaction. Ics Series. Addison-Wesley Publishing Company, 1994.

[9] Juliane Franze, Kim Marriott, and Michael Wybrow. 2014. What academics want when reading digitally. In Proceedings of the 2014 ACM symposium on Document engineering (DocEng '14). ACM, New York, NY, USA, 199-202.

[10] Lee, B., Czerwinski, M., Robertson, G., Bederson, B. (2005). Understanding research trends in conferences using PaperLens. In CHI extended abstracts on Human factors in computing systems (pp.1969-1972). ACM.

[11] Lewis, J. R., & Sauro, J. (2009). The Factor Structure of the System Usability Scale. In Proceedings of the 1st International Conference on Human Centered Design (HCD 2009): 94-103.

[12] Mei-Ling Wang, Scholarly journal use and reading behavior of social scientists in Taiwan, The International Information & Library Review, Volume 42, Issue 4, 2010, Pages 269-281, ISSN 1057-231.

[13] Niu, X., Hemminger, B. M., Lown, C., Adams, S., Brown, C., Level, A., McLure, M., Powers, A., Tennant, M. R. and Cataldo, T. (2010), National study of information seeking behavior of academic researchers in the United States. J. Am. Soc. Inf. Sci., 61: 869–890.

[14] Peroni, S. (2014). Semantic Web Technologies and Legal Scholarly Publishing. Law, Governance and Technology Series 15. Springer. http://dx.doi.org/10.1007/978-3-319-04777-5

[15] Presutti, V., Francesco D., & Aldo G. (2012). Knowledge extraction based on discourse representation theory and linguistic frames. In Proceedings of the 18th International Conference on Knowledge Engineering and Knowledge Management (EKAW 2012): 114–129. Berlin, Germany: Springer.

[16] Qazvinian, V., Radev, D. (2010). Identifying Non-explicit Citing Sentences for Citation-based Summarization. In Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL 10): 555–564.

[17] Sauro, J. (2011). A Practical Guide to the System Usability Scale: Background, Benchmarks & Best Practices. CreateSpace. ISBN: 1461062707

[18] Shen, Z., Ogawa, M., Teoh, S. T., Ma, K. L. (2006). BiblioViz: a system for visualizing bibliography information. In Proceedings of the 2006 Asia-Pacific Symposium on Information Visualisation-Volume 60 (pp. 93-102).

[19] Shneiderman, B. (1996). The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In Proceedings of the IEEE Symposium on Visual Languages, pages 336-343,Washington. IEEE Computer Society Press, 1996.

[20] Shotton, D. (2013). Publishing: Open citations. Nature, 502(7471): 295–297.

[21] Strauss, A. Corbin, J. (1998). Basics of Qualitative Research Techniques and Procedures for Developing Grounded Theory (2nd edition). Sage Publications. ISBN: 0803959408

[22] Teufel, S., Siddharthan, A., Tidhar, D. (2006). Automatic classification of citation function. In Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing (EMNLP 06): 103–110.

[23] Yichen Jiang, Aixia Jia, Yansong Feng, and Dongyan Zhao. 2012. Recommending academic papers via users' reading purposes. In Proceedings of the sixth ACM conference on Recommender systems (RecSys '12). ACM, New York, NY, USA, 241-244.