Available online at www.sciencedirect.com

ScienceDirect

Procedia Technology 9 (2013) 573 – 583

Procedia Technology

CENTERIS 2013 - Conference on ENTERprise Information Systems / PRojMAN 2013 - International Conference on Project MANagement / HCIST 2013 - International Conference on Health and Social Care Information Systems and Technologies

Putting "human crowds" in the loop of bibliography evaluation: a collaborative working environment for CSCW publications

António Correia[a], Jorge Santos[a], Diogo Azevedo[a], Hugo Paredes[a,b], Benjamim Fonseca[a,b,*]

[a]*Department of Engineering, School of Sciences and Technology, UTAD – University of Trás-os-Montes e Alto Douro, Quinta de Prados, Apartado 1013, Vila Real, Portugal*

[b]*INESC TEC, Campus da FEUP, Rua Dr. Roberto Frias, 378, 4200 – 465, Porto, Portugal*

Abstract

The current impact of financial crisis on societal and scientific frameworks has raised the need to harvest and evaluate vast volumes of data in a socially-mediated interaction context to reduce knowledge gaps and accelerate innovation at a global scale. Existing mechanisms are inefficient for a single human to classify and transform data into knowledge patterns from a large number of publications. This time-consuming and computationally difficult activity requires a substantial cognitive effort grounded on scientific metrics and theoretical foundations to produce quality metadata through different knowledge representations. This paper reports on a work in progress community self-organizing bibliographic information system for semantic analytics focused on what scientific research data mean, and how they can be best interpreted through a division of intellectual labor among social, computer, and citizen scientists. Such a crowd labor ecosystem should be not restricted to traditional bibliometric approaches, attempting to uncover patterns and trends in publication data sets whilst intellectual connections can be examined to demonstrate how a scientific field is conceptually, intellectually, and socially structured.

Selection and/or peer-review under responsibility of SCIKA – Association for Promotion and Dissemination of Scientific Knowledge

Keywords: Crowdsourcing; CSCW; groupware; human computation; bibliographic information systems; scientific data repositories; scientometrics; socially-mediated bibliography evaluation.

* Corresponding author. Tel.: +351-259350369; fax: +351-259350356.
E-mail address: benjaf@utad.pt.

Selection and/or peer-review under responsibility of SCIKA – Association for Promotion and Dissemination of Scientific Knowledge
doi:10.1016/j.protcy.2013.12.064

1. Introduction

Collaborative working environments can be effective instruments for dealing with high-change conditions. Cooperative work endeavors among researchers have originated complex structures that require an exhaustive examination of their role and mental models in the processes of knowledge interpretation, evaluation, creation, consumption, and distribution [1]. Data gathering and analysis processes comprise labor-intensive activities to uncover knowledge patterns and trends. Scientists can interpret such evidences in distinct ways with emphasis on their background, and a recurrent challenge relies on keeping track of advancements, revealed assumptions, disciplinary boundaries, research fields lacking examination, and ghost theories [2] by researchers and general public spending a lot of time and cognitive effort in mining and analyzing scientific corpora with inefficient techniques [3]. How such data is gathered, classified, and visualized differs from field to field [4], and justifies further exploration for Computer-Supported Cooperative Work (CSCW) community due to its fragmented [5], multi-disciplinary nature, and lack of detailed understanding concerning its scope at the activity level.

Exploiting semantics of the published data from as many of sources as possible becomes an essential issue to allow a finely textured and up-to-date portrait of scientific output, a research task that has been performed manually examining variances, correlating evidences, and compiling descriptive statistics [6]. Existing metrics and measurement systems are inadequate to capture the intellectual structure of a scientific field, which can be understood as an abstraction of the collective knowledge of its researchers, as well as to trace the full range of socially-mediated activities that support and transmit scientific contributions across different disciplines [7]. In this perspective, scientometrics needs to be conducted in a social context [6] to evaluate the increasing amount of scientific digital artifacts taking into account human factors as a field of carefully examination.

Current bibliographic information systems lack semantic evidences that can be achieved from human labor at a massive scale using various kinds of classification mechanisms. This approach has been applied for image labeling (e.g., ESP Game), protein folding (e.g., Fold.it, and RCSB Protein Data Bank), galaxy classification (i.e., Galaxy Zoo), and question answering about surrounding obstacles (e.g., VizWiz) [8]. Crowdsourcing has been established as a development industry "employing over 2 million knowledge workers, contributing over half a billion dollars to the digital economy" [9]. Human crowds can act as cognitive operators with different but complementary skills for solving problems that are beyond the current abilities of automated approaches, combining Human Intelligence Tasks (HITs) with large-scale database systems that indirectly coordinate joint efforts to examine big data volumes. Our hypothesis relies on conducting semantic analytics by crowdsourcing HITs between multiple contributors, whilst research on whether manual data gathering and evaluation can be scalable to a large set of publications and knowledge workers remains unclear [10].

A collaborative working environment is proposed for bibliography classification, engaging scientists and general public around metadata and semantically annotated textual elements for mapping science. This system aims to evaluate all kind of digital artifacts and other forms of intellectual assets that are produced, shared, and maintained by researchers (e.g., journal papers, conference proceedings, posters, tutorials, slides in electronic forms, images, videos, websites, blogs, web services, downloadable tools, research datasets, and scientific workflows [11]). This community data system can detect collective intelligence indicators at a massive scale in a relatively novel architecture of participation that gain value as more users cooperate.

The paper is organized as follows. Section 2 presents some background information on the social-technical challenges of data evaluation through bibliographic information systems. A comparative analysis is made as a complement for understanding features and capabilities of scientific information repositories. We then proceed to formulate our conceptualization of the requirements for a human-centered collaborative working ecosystem in Section 3. Following this, implementation issues are discussed with emphasis on the platform's architecture and features. The paper finishes with some limitations and future working directions that require consideration in the development of this community self-organizing bibliographic information system.

2. Background

Research on scientific collaboration effects has been growing over time [12], representing more than half of all research activities in many countries and organizations [27]. The increasing scale at which scientific collaboration has been performed can be associated with the growth of incentives to produce more, and the "greater facilitation of collaborative work" [26]. Scientists read approximately 50 percent more papers than in the 1970s, spending less time with each one [24]. Despite these valid indicators, team size in inter-disciplinary research endeavors is not growing so rapidly, and complexities are revealed in terms of the allocation of work, risk, specialization, role definition, willingness to engage, and resources (i.e., time, and cognitive effort) spent on knowledge tasks from different fields and sub-fields [25], considering the current long publication periods. Supporting scientific coordination, communication and cooperation in large-scale, multi-disciplinary research settings is critical to promote hybrid collaboration solutions for problem-solving in the context of scientific, technological and societal needs [43], and new geographically distributed knowledge workers are contributing overcoming political, economic and cultural barriers that make too expensive for scientists to work together.

2.1. Scientific research collaboration, scientometrics, classification models, and semantic analytics

At the light-end of scientific collaboration spectrum, researchers have made several attempts to understand social and individual information foraging behaviors, and crowd wisdom effects [23]. Knowledge analysis is a complex process that requires alternative metrics, and open workflows [31] can be functional instruments for supporting metadata contributions from cognitive operators evaluating scientific outputs. Such a framework allows combining hypotheses from distinct theoretical perspectives, methodologies, and units of analysis into an integrated model [6]. Concerning the inherent difficulty of maintaining large-scale scientific collaboration endeavors, this approach can expand classical scientometrics to solve complex problems, misunderstandings, trust building, and tacit and transactive knowledge communication [32].

Scientometrics is undergoing a renaissance since a continuous change occurs across all scientific domains [33], and it is a fact that "parameter choices for observing trends are often made ad hoc" [34]. Scientometrics represents a valuable instrument for identifying inter-relationships between research topics, prolific scientists or research groups, research performance by country and institution, collaboration patterns, and the prediction of future trends and developments [35]. Interweaving and co-evolving communities that grow around clusters of publications, disciplines, and subjects can be further studied to assess impact and build a science of science measurement [36] sustained on intellectual labor representations. Significant advances have been witnessed in the 21st century crossing the fields of webmetrics, text and data mining, semantic analytics, and open access, and the increasing use of sophisticated mapping and visualization approaches is probably the most significant development in relational bibliometrics [37].

In the scientific context, only a mass collaboration effort among researchers can help delimit and legitimate both categorization labeling and classification protocols, and a social evaluation approach denotes a "semiotic embodiment to ongoing collective learning processes". However, any evaluation process suffers from errors, which occurs not only due to possible conflicts of interest, depending on homogeneity and standardization of a field, Taxonomies are cognitive structures "aiming at providing some generic and robust classification rules via the construction of a fixed set of categories which hopefully encompass the existing content diversity and organize it" [38]. Open classification tools and free annotation (e.g., folksonomies) can support nomenclatures provided by cognitive operators for any given content by using keywords or "tags" that reflect their thesaurus and knowledge background without a hierarchical, taxonomy semantic dependency for detecting weak signals of past or incoming changes. However, manual analysis is labor-intensive and subjective [39], and automated systems (e.g., RSTTool) involve expensive training to work automatically being prone to errors.

2.2. Collective intelligence, crowdsourcing, and human computation at the service of science

The intellectual structure of a field can be understood as an abstraction of the collective knowledge of its researchers [13], and different forms of collective intelligence can emerge from cooperative work ensembles. The study of human intelligence at a massive scale is a relatively new field of analysis, and there is no known theory and/or model capable of explaining how it really works [28]. Collective intelligence can be conceived as "a form of universal, distributed intelligence, which arises from the collaboration and competition of many individuals" [29] dispersed by groups and other collective living systems performing a wide diversity of tasks such as "gathering, formulating, modifying, and applying effective knowledge" [30]. Another approach relies on Augmented Social Cognition [23], which aims at enhancing the ability of a group "to remember, think, and reason", augmenting knowledge achievement, production, communication, and use, and evolving collective and individual intelligence in socially mediated information environments.

Harnessing crowds to tackle complex problems for a single expert or group has been subject of meticulous research. A psychological study examined the cognitive aptitudes of a crowd (N=699 individuals, distributed by 192 groups of two to five members) performing tasks based on McGrath's taxonomy [14] (e.g., sharing typing assignments in a shared editor) to identify group genetic structures inferring on collective intelligence. Some studies have emphasized crowdsourcing attempts such as taxonomy construction from collective efforts provided by crowd workers through automated workflow systems (e.g., Cascade [15]), games with a scientific purpose [20], ubiquitous spatial information maintenance on urban elements [17], collaboration workflows for crowdsourcing procedures (e.g., Turkomatic [19]), massively distributed authorship of academic papers [18], users' behavior and profiles in Question-and-answer (Q&A) platforms (i.e., Stack Exchange [16]), processing queries that neither database systems nor search engines can reasonably answer (i.e., CrowdDB) [21], as well as online collaboration platforms for citizen scientists without requiring programming skills (e.g., Pathfinder [22]). Human Processing Units (HPUs), as powerful cognitive workforces, must be combined with computer science principles and practices to enhance application design, giving rise to an advanced class of software applications enabled by human crowds performing complex tasks.

2.3. Scientific information systems, digital libraries, repositories, observatories, and collaboratories

Suitable collaborative information systems and technologies are required to produce, share, filter, combine, and present scientific discoveries. An effective way to create information systems has been approached from a design science perspective considering goal-directed collective activities, routines, and resources to produce a linguistic sign, organizational scheme, or technical artifact that represent a social practice [40]. Web 2.0 tools and digital libraries could be integrated to design personalized bibliographic information systems and support scientific convergence through socially mediated interaction around semantically annotated elements.

As argued by Farooq et al. [4], digital libraries (understood as online repositories that enable scientific discovery through search and retrieval of intellectual resources) lack scientific collaboration capabilities. The premise relies on a direct collaboration between peers in a scientific community around meaningful digital artifacts, long-term endeavors, and scientific outcomes. Collaboratories support distributed scientists working together with features to access, view, manipulate, and have discussions on intellectual artifacts [45], creating a vast set of research possibilities (e.g., shared meaning) through cyberinfrastructure support. Community data systems (e.g., RCSB Protein Data Bank) support information resources that are created and manipulated by a geographically-distributed body of contributors [46]. Some attempts have been made suggesting methods for classifying full text content of scientific publications [44] through the enrichment of bibliographic metadata harvested by the Open Archives Initiative protocol.

Open science could foster the establishment of advanced collaboration models among researchers. Several communities have adopted development platforms such as HUBzero to construct collaboratories able to share ideas, publications, models, and data. Pegasus Workflow Management System has been applied to manage complex analyses running on campuses and large-scale cyberinfrastructures (e.g., Open Science Grid) [47]. Talkoot is a software toolkit and knowledge management environment designed for collaboration on Earth Sciences that "allows researchers to systematically gather, tag and share their data, analysis workflows and research notes" within a virtual community [48]. Concerning ubiquitous bibliographic data, DeaiExplorer [49] was introduced as a "data-centric community mapping tool that extracts and visualizes hundreds of research communities in Computer Science, based on the DBLP publication database". Figshare offers a repository for data, methods and materials for private archiving or public sharing. Open Science Framework is a Web-based project management framework that allows documenting and archiving research materials, analyzing scripts, and empowering users to keep materials private or public [31]. WikiDashboard [23] was presented as a social dynamic analysis tool for Wikipedia, and Alpha [50] is built around a vast repository of curated data.

Several limitations have been discussed comprehensively in the literature concerning the value of scientific databases (e.g., ACM Digital Library, IEEE Xplore, DBLP, PubMed, Web of Knowledge, Scopus, CiteSeer, Google Scholar, and arXiv [41]). Table 1 shows a complementary comparison between repositories and their overall characteristics. Web-based reference managers such as Mendeley allow users to save PDF files to their desktop application, automatically extract bibliographic information, and sharing data with other collaborators [42]. Awareness mechanisms have been studied in CiteSeer [4], a scholarly digital library that provide users with a set of notification mechanisms for publication events using feeds. Social bookmarking and publication-sharing systems such as BibSonomy provide users with the ability to store and organize their bookmarks and publication entries, supporting community and group creation trough a social platform for literature exchange.

Table 1. Comparison between CONTENTdm, DSpace, and DBLP (adapted from [41, 43])

	CONTENTdm	DSpace	DBLP
License	Proprietary	Free	ODC-BY 1.0
Product type	Software	Software	Host
Supported item types	JPEG, GIF, or TIFF images; WAV or MP3 audio files; AVI or MPEG video files; PDF files; EAD Finding Aids and URLs	PDF, JPEG, MPEG, TIFF). But DSpace will accept files of any format	-
Metadata formats	Unicode; Z39.50; Qualified Dublin Core; METS; VRA; XML; JPEG2000; OAI-PMH; and METS/ALTO	Qualified Dublin Core, MARC/MODS	-
Format conversion	PDF files; PDF compound objects; and XML	BibTex, RIS, TSV, CSV	XML, BibTex, Google Scholar, CiteSeerX, pubzone.org and Electronic Edition
Searching	Advanced search; All words; The exact phrase; Any word; None of the words; Search by title, subject, description, creator, coverage, format, and publisher	Keyword, Author, Title, Subject, Abstract, Series, Sponsor, Identifier	Browse by conferences, journals and series - Search by author, type, year, Coauthor Index - CompleteSearch, Faceted Search @ L3S, Free Search @ isearch
Web 2.0 features	Share (via e-mail, Facebook, Twitter, Flickr); tagging; Comment and rate	-	-
Statistical reporting	Downloads and item view summary; Top searches; Monthly summary; Daily summary; and hourly summary	Total visits of the current community home page; Visits of the community home page over a timespan of the last 7 months; Top 10 country from where the visits originate; Top 10 cities from where the visits originate; Total visits of the item; Total visits for the bitstreams attached to the item; Visits of the item over a timespan of the last 7 months	Distribution of publication types, Publications per year, Number of authors per publication, Number of publications per author, Number of coauthors per author, Records in DBLP (grouped by year), Records in DBLP (grouped by date of last change), Number of edits per publication, New records per year, New records per month, New records in year 2012, 2011, 2010, 2009, 2008, 2007, 2006, 2005, 2004, 2003, 2002

The analysis presented in Table 1 is complementary to previous comparative studies [41, 43], representing an attempt to identify a set of requirements for a collaborative working environment supported by an open and participatory model [6] that crosses several collaboration features from Web 2.0 with metadata visualizations. Synoptically, all systems support information retrieval in many formats (e.g., XML, and BibTex). DBLP is an automatically updated repository for Computer Science publications that is too static for metadata enrichment

since it is maintained by DBLP team, denying access for a common user. DSpace and DBLP systems present some similarities, but the first provides infrastructure customization allowing institutions to build or modify a personalized repository without additional costs. CONTENTdm is a proprietary system that contains multiple features from Web 2.0 tools, and an open participation model for registered users. However, existing systems do not have the required features for an open crowd-enabled environment focused on scientometric indicators (e.g., co-authorship data), and semantic analytics through distinct classification approaches.

3. Community self-organizing bibliographic information system for CSCW publications

An approach for the lightweight development of a bibliographic information system is based on the idea of involving crowds in the underlying engineering and design processes providing new features that correspond to specific user needs. Data should be collected on the full range of scientists' work, developing metrics and methods testing their validity through communication, coordination and cooperation features. Feedback loops, reflection, and concept formation features may allow actors to generate explicit knowledge. Semi-structured messages support asynchronous communication for non-routine information without rigid constraints, and a discussion forum can be also used to opinion sharing between users. Comments and textual annotations added on classified paper attributes can serve functions such as communicate a specific reasoning, criticism, requests for clarification, and supplemental evidences. A RSS feed may integrate the system aggregating content from many sources, allowing users receiving updates about new or changed items.

The success of many collaborative systems hinges on effectively supporting awareness of collaborators and their actions (e.g., assigning roles, making decisions, negotiating, or prioritizing tasks), and the process of co-creating shared work artifacts. Currently, such functionality is primarily limited to alert services in digital libraries through email notifications. Defining work roles for registered group members (e.g., research lab), and specifying profile information and contributor skills can represent possible implementations. A workflow management system will organize and determine task series to be made by community members.

Content search must be supported by a customized search engine semantically empowered. Classification features must be supported by semantic data and visualization mechanisms allowing users to infer statistically on the information correlations. Metadata can be stored, edited and requested for research, and the software must allow distinguishing between user groups assigning for each specific rights (e.g., administrator, metadata producer, or common user). A paper recommendation mechanism will be implemented, and the openness with data, methods, and tools will make them citable through a referring approach where each contributor selects his/her specialized fields evaluating papers in accordance. A feature would support the rating of papers and all relevant agents in the classification process (i.e., authors, and referees) visible to everyone and carried out by registered users through a knowledge reputation system.

Legal barriers (e.g., copyright), trust, anonymity, and ethical issues must be carefully considered to assure confidentiality of participant identities. Indexing scientometric indicators from citation indexes (e.g., Google Scholar) and metadata from XML-based libraries (e.g., DBLP) can be effective for reducing human efforts in cataloguing processes. Paper classification mechanisms with emphasis on community thesauri, taxonomies, folksonomies, and distinct kinds of classification models are delegated to contributors under the responsibility of defining and managing procedures. Qualitative papers are expected to include not only the interpretation of the data (findings) but also the ways (processes) in which the authors derived these interpretations from the qualitative data through semantic annotations. Considering the storage mechanism of a decentralized archive, papers that can have not been accessed for a long time would lose visibility, while frequently accessed papers would gain position. Data can only be modified based on predefined constraints, and conflicting users cannot modify data. Regarding accessibility concerns, blind people need meaningful text equivalents for images so a

screen reader can "read" the information they need to navigate. Humans with low vision require larger fonts and higher contrast, whilst deaf people need visual representations for audio.

3.1. System architecture

Information storage, collection, and visualization are central concepts related with this CSCW environment since scientific data are putted in a query state providing user community with reading, discussing, annotating, tagging, reviewing, and other actions supported by collaboration features. The system architecture (Figure 1) provides a separation into three layers for better code understanding and optimizing. This typology allows modifying, for example, the presentation layer without critical changes on support code. An advantage of this architecture relies on layers' independence concerning the same physical machine (e.g., user interface layer is on the user side, and server layer can be inserted on a dedicated server whilst database is putted on another server without database systems dependencies). System architecture is guided by principles such as clearly defined functional layers, abstraction, encapsulation, high cohesion, reusability, and loosely coupling [51].

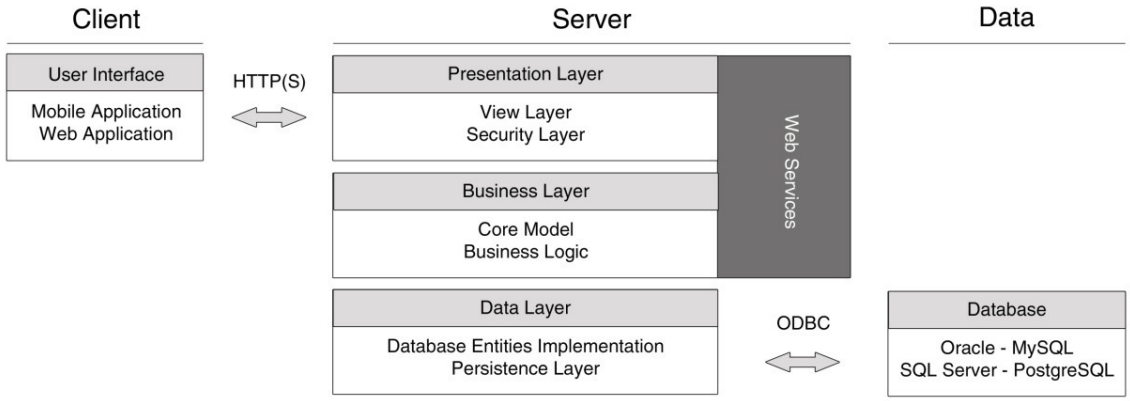

Fig. 1. Collaborative working system architecture

The architecture layers presented in Figure 1 are responsible for handling data between existing interfaces (database, and application prototype) [43]. A presentation layer links client interface and prototype, data layer supports the database information handling, and business layer is a controller between presentation layer, data layer, and web services. Web services are an experimental attempt to introduce new publications in external applications. This framework could be realized through the use of standardized building blocks explicitly created to communicate with a large information platform for metadata creation across several contributors. Architecture modules can include an incentive-based crowdsourcing system that collects data to address the most important scientific challenges; a many-to-many communication feature; a public dashboard; a social networking career manager system; a decentralized co-creation and evaluation system; an annotation tool that would store and retrieve metadata that could be extracted in the form of graphics, statistics, links, videos, scientific references, and other important data; a semi-automated reporting system; a networked knowledge manager; a context-aware reputation system; and a privacy settings panel [3].

3.2. System prototype

The collaborative system prototype (Figure 2) represents the user interface, publications list (which shows a set of publications by title, author metadata, and supported actions), and its corresponding feed, main page, administration panel, contact, and project information. Users can visualize details of each publication pressing "show" button. Title, year, series volume, pages, publication name, keywords, DOI, publication type, ISBN, author(s), publisher, and editor(s) represent the current metadata fields. Furthermore, system prototype allows users to list not only by publication but also by author, editor, institution, publisher, TypePub, and ISNType. The collaborative working system is now able to store information from scientific publications incorporated in several digital libraries, automating the process by indexing elementary metadata.

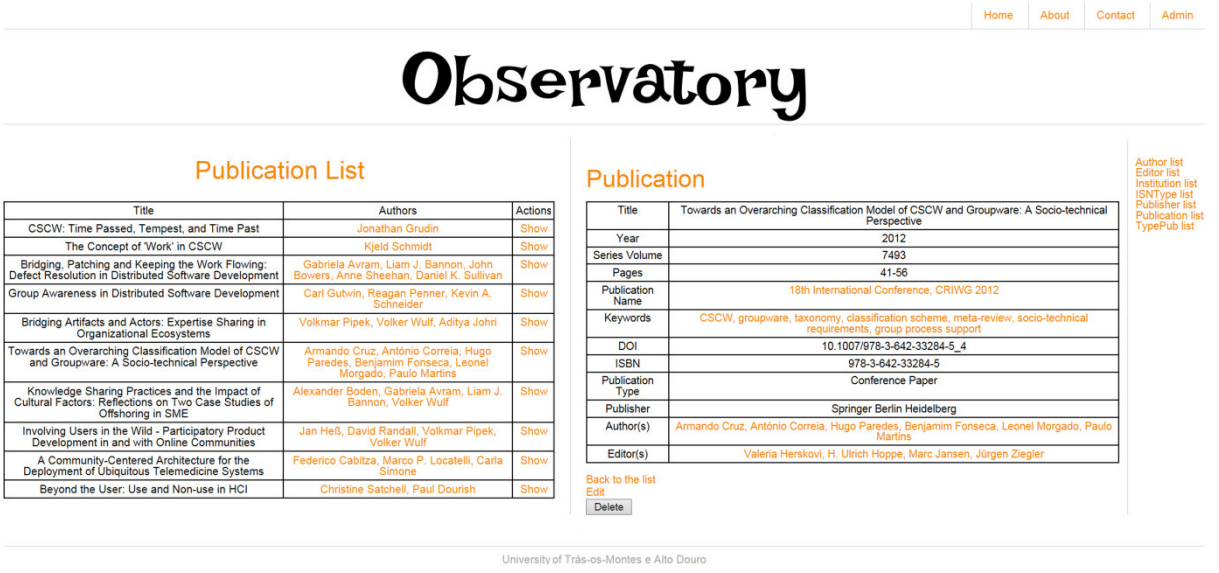

Fig. 2. System prototype

A total of 28,991,154 indexed publications were uploaded from DBLP community, which structure comes from two bibliographic reference management systems (i.e., EndNote, and BibTeX). In this context, database represents a set of possible relationships between data fields that will be handled in the system and subsequent data cataloguing processes. In order to perform scientometrics and semantic analytics, a crowd-enabled model [6] is being validated based on data crossing between publications. These data will be provided by crowds and include citation indicators, keywords, tags, and annotations. Each user can evaluate a publication according to his/her background, experience, and personal vision associated with a research topic (e.g., a sociologist can analyze a paper related with ethnography, whist a computer scientist can prefer evaluate system attributes). If authenticated users want to download or purchase a specific paper or book, they will be redirected to the database system in which the publication is hosted (e.g., ACM Digital Library, or IEEE Xplore).

4. Conclusions and future directions

This paper emphasizes a community self-organizing bibliographic information system prototype conceived to support a higher level of engagement by CSCW researchers and general public. Further developments will

be focused on providing groupware functionalities aggregating contributions from several users. This system prototype aims to boost the scientific publications' storage and handling mechanisms, harnessing the wisdom of crowds. The main target relies on improving the effective use of scientific data to solve complex problems at a large scale (e.g., a literature review on a particular topic), and this collaborative working environment can be further adapted for other scientific fields using distinct datasets and classification models.

Limitations are unfilled in the lack of semantic data fields, user involvement in evaluation tasks, absence of collaboration features, scarcity of experimental research to validate the system usefulness, doubtful reliance on human contributions classifying large volumes of bibliographic data, static taxonomic granularity, possible evaluation errors, subjectivity, and occasional conflicts of interest. HCI tasks and design, implementation, and evaluation principles will be explored. Connections that can emerge from user interaction logs and semantic analytics can be represented by nodes. Visualization mechanisms and graphical navigation may be integrated, and cognitive task analysis suggests a promising research approach to conduct in this space entirely dependent on the wisdom of crowds. Usability tests, case studies, and field work (i.e., ethnography) with research teams are required to gain insight on work practices that can better inform development issues. Despite the prospects for crowd labor research in scientometrics and semantic analytics, ethical constraints, motivational factors for contributing actively and systematically, and social-technical barriers should be considered for "closing the loop" in developing complementary, hybrid human-machine information systems.

Acknowledgements

This work is funded (or part-funded) by the ERDF – European Regional Development Fund through the COMPETE Programme (operational programme for competitiveness) and by National Funds through the FCT – Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) within project «FCOMP - 01-0124-FEDER-022701».

References

[1] Inzelt, A., Schubert, A., Schubert, M., 2009. Incremental Citation Impact due to International Co-authorship in Hungarian Higher Education Institutions, Scientometrics 78, 1, pp. 37-43.
[2] Evans, J., Foster, J., 2011. Metaknowledge, Science 331, 6018, pp. 721-725.
[3] Helbing, D., Balietti, S., 2011. How to Create an Innovation Accelerator, European Physical Journal Special Topics 195, 1, pp. 101-136.
[4] Farooq, U., Ganoe, C. H., Carroll, J. M., Giles, C. L., 2009. Designing for e-Science: Requirements Gathering for Collaboration in CiteSeer, International Journal of Human Computer Studies 67, pp. 297-312.
[5] Schmidt, K., 2009. "Divided by a Common Acronym: On the Fragmentation of CSCW," 11th European Conference on Computer Supported Cooperative Work. Vienna, Austria, pp. 223-242.
[6] Correia, A., Fonseca, B., Paredes, H., 2013. "Exploiting Classical Bibliometrics of CSCW: Classification, Evaluation, Limitations, and the Odds of Semantic Analytics," 1st International Conference on Human Factors in Computing & Informatics. Maribor, Slovenia, pp. 137-156.
[7] Lane, J., 2010. Let's Make Science Metrics More Scientific, Nature 464, pp. 488-489.
[8] Quinn, A. J., Bederson, B. B., 2011. "Human Computation: A Survey and Taxonomy of a Growing Field," 29th ACM SIGCHI Conference on Human Factors in Computing Systems. Vancouver, Canada, pp. 1403-1412.
[9] Vukovic, M., Bartolini, C., 2010. "Crowd-Driven Processes: State of the Art and Research Challenges," 8th International Conference on Service-Oriented Computing. San Francisco, USA, p. 733.
[10] Eysenbach, G., 2011. Can Tweets Predict Citations? Metrics of Social Impact Based on Twitter and Correlation with Traditional Metrics of Scientific Impact, Journal of Medical Internet Research 13, 4, e123.
[11] Tan, W., Zhang, J., Foster, I., 2010. Network Analysis of Scientific Workflows: A Gateway to Reuse, IEEE Computer 43, 9, pp. 54-61.
[12] Boyack, K. W., 2009. Using Detailed Maps of Science to Identify Potential Collaborations, Scientometrics 79, 1, pp. 27-44.

[13] Chen, C., Chen, Y., Horowitz, M., Hou, H., Liu, Z., Pellegrino, D., 2009. Towards an Explanatory and Computational Theory of Scientific Discovery, Journal of Informetrics 3, 3, pp. 191-209.
[14] Woolley, A., Chabris, C., Pentland, A., Hashmi, N., Malone, T., 2010. Evidence for a Collective Intelligence Factor in the Performance of Human Groups, Science 330, pp. 686-688.
[15] Chilton, L. B., Little, G., Edge, D., Weld, D. S., Landay, J. A., 2013. "Cascade: Crowdsourcing Taxonomy Creation," 31st ACM SIGCHI Conference on Human Factors in Computing Systems. Paris, France.
[16] Furtado et al., 2013. "Contributor Profiles, their Dynamics, and their Importance in Five Q&A Sites," 16th ACM Conference on Computer Supported Cooperative Work. San Antonio, USA, pp. 1237-1252.
[17] Mashhadi, A., Quattrone, G., Capra, L., 2013. "Putting Ubiquitous Crowd-sourcing into Context," 16th ACM Conference on Computer Supported Cooperative Work. San Antonio, USA, pp. 611-622.
[18] Tomlinson et al., 2012. "Massively Distributed Authorship of Academic Papers," 30th ACM Annual Conference Extended Abstracts on Human Factors in Computing Systems. Austin, Texas, USA, pp. 11-20.
[19] Kulkarni, A., Can, M., Hartmann, B., 2012. "Collaboratively Crowdsourcing Workflows with Turkomatic," 15th ACM Conference on Computer Supported Cooperative Work. Seattle, Washington, USA, pp. 1003-1012.
[20] Good, B. M., Su, A. I., 2011. Games with a Scientific Purpose, GenomeBiology 12, 135.
[21] Franklin, M. J., Kossmann, D., Kraska, T., Ramesh, S., Xin, R., 2011. "CrowdDB: Answering Queries with Crowdsourcing," ACM SIGMOD International Conference on Management of Data. Athens, Greece, pp. 61-72.
[22] Luther, K., Counts, S., Stecher, K. B., Hoff, A., Johns, P. 2009. "Pathfinder: An Online Collaboration Environment for Citizen Scientists," 27th ACM SIGCHI Conference on Human Factors in Computing Systems, pp. 239-248.
[23] Chi, E. H., Pirolli, P., Suh, B., Kittur, A., Pendleton, B., Mytkowicz, T., 2009. "Augmented Social Cognition: Using Social Web Technology to Enhance the Ability of Groups to Remember, Think, and Reason," 35th SIGMOD International Conference on Management of Data Providence. Providence, USA.
[24] Renear, A. H., Palmer, C. L., 2009. "Strategic Reading, Ontologies, and the Future of Scientific Publishing," Science 325, 5942, pp. 828-832.
[25] Rigby, J., 2009. Comparing the Scientific Quality Achieved by Funding Instruments for Single Grant Holders and for Collaborative Networks within a Research System: Some Observations, Scientometrics 78, 1, pp. 145-164.
[26] Wagner, C. S., 2005. Six Case Studies of International Collaboration in Science, Scientometrics 62, pp. 3-26.
[27] Hoekmana, J., Frenkena, K., Tijssen, R. J. W., 2010. Research Collaboration at a Distance: Changing Spatial Patterns of Scientific Collaboration within Europe, Research Policy 39, 5, pp. 662-673.
[28] Schut, M. C., 2010. On Model Design for Simulation of Collective Intelligence, Information Sciences 180, 1, pp. 132-155.
[29] Levy, P., 1997. Collective Intelligence: Mankind's Emerging World in Cyberspace, Basic Books.
[30] Atlee, T., Pór, G., 2000. Collective Intelligence as a Field of Multi-disciplinary Study and Practice.
[31] Nosek, B. A., Spies, J. R., Motyl, M., 2012. Scientific Utopia : II. Restructuring Incentives and Practices to Promote Truth Over Publishability, Perspectives on Psychological Science 7, 6, pp. 615-631.
[32] Hennemann, S., Rybski, D., Liefner, I., The Myth of Global Science Collaboration - Collaboration Patterns in Epistemic Communities, Journal of Informetrics 6, 2, pp. 217-225.
[33] Kurtz, M. J., Bollen, J., 2010. Usage Bibliometrics, Annual Review of Information Science and Technology 44, 1, pp. 1-64.
[34] Tseng, Y.-H., Lin, Y.-I, Lee, Yi-Y., Hung, W.-C., Lee, C.-H., 2009. A Comparison of Methods for Detecting Hot Topics, Scientometrics 81, 1, pp. 73-90.
[35] Vinkler, P., 2010. Indicators are the Essence of Scientometrics and Bibliometrics, Scientometrics 85, 3, 861-866.
[36] Priem, J., Hemminger, B., 2010. Scientometrics 2.0: Toward New Metrics of Scholarly Impact on the Social Web. First Monday 15, 7.
[37] Smith, D. R., 2012. Impact Factors, Scientometrics and the History of Citation-based Research, Scientometrics 92, 2, pp. 419-427.
[38] Glassey, O., 2012. Folksonomies: Spontaneous Crowd Sourcing with Online Early Detection Potential?, Futures 44, 3, pp. 257-264.
[39] van Eck, N. J., Waltman, L., Noyons, E. C. M., Buter, R. K. 2010. Automatic Term Identification for Bibliometric Mapping, Scientometrics 82, 3, pp. 581-596.
[40] Rohde, M., Stevens, G., Brödner, P., Wulf, V., 2009. "Towards a Paradigmatic Shift in IS: Designing for Social Practice," 4th International Conference on Design Science Research in Information Systems and Technology. Philadelphia, USA.
[41] Hull, D., Pettifer, S. R., Kell, D. B., 2008. Defrosting the Digital Library: Bibliographic Tools for the Next Generation Web, PLoS Computational Biology 4, 10, e1000204.
[42] Li, X., Thelwall, M., Giustini, D., 2011. Validating Online Reference Managers for Scholarly Impact Measurement, Scientometrics 91, 2, pp. 461-471.
[43] Santos, J., Paredes, H., Fonseca, B., Correia, A., 2012. "Preliminary Study of an Observatory for CSCW Scientific Publications," 18th International Conference of European University Information Systems organisation, E-science and E-repositories, virtual libraries, virtual laboratories. Vila Real, Portugal.
[44] Bertin, M., Atanassova, I., 2012. Semantic Enrichment of Scientific Publications and Metadata, D-Lib Magazine, 18, 7/8.
[45] Finholt, T. A., Olson, G., 1997. From Laboratories to Collaboratories: A New Organizational Form for Scientific Collaboration, Psychological Science 8, 1, pp. 28-35.
[46] Bos, N., Zimmerman, A., Olson, J., Yew, J., Yerkie, J., Dahl, E., Olson, G., 2007. From Shared Databases to Communities of Practice: A Taxonomy of Collaboratories, Journal of Computer-Mediated Communication 12, 2, a16.

[47] Deelman, E., McLennan, M., 2012. Sharing Science in the 'Collaboratory', International Science Grid This Week. Retrieved from http://www.isgtw.org/feature/sharing-science-%E2%80%98collaboratory%E2%80%99.
[48] Ramachandran, R., Maskey, M., Kulkarni, A., Conover, H., Nair, U. S., Movva, S., 2012. Talkoot: Software Tool to Create Collaboratories for Earth Science, Earth Science Informatics 5, 1, pp. 33-41.
[49] Konomi, S., 2011. "Community Mapping for Cross-Boundary Research Collaboration," Creating, Connecting and Collaborating through Computing, pp. 11-16.
[50] Abercrombie, R. K., Udoeyop, A. W., Schlicher, B. G., 2012. A Study of Scientometric Methods to Identify Emerging Technologies via Modeling of Milestones, Scientometrics 91, 2, pp. 327-342.
[51] Microsoft Patterns & Practices Team, 2009. Microsoft Application Architecture Guide, 2nd Edition (Patterns & Practices), Microsoft Press.