Original scientific paper

New Access Structures to Scientific Information:

Metadata, citation and similar papers
ozitorij Filozofskog fakulteta u Zagrebu' at University of Zagreb

Sonja Špiranec
Department of Information Sciences
Faculty of Humanities and Social Sciences, University of Zagreb
Ivana Lučića 3, 10 000 Zagreb, Croatia
sspiran@ffzg.hr

Ana Babić, student
Department of Information Sciences
Faculty of Humanities and Social Sciences, University of Zagreb
Ivana Lučića 3, 10 000 Zagreb, Croatia
anbabic@ffzg.hr

Ana Lešković, student
Department of Information Sciences
Faculty of Humanities and Social Sciences, University of Zagreb
Ivana Lučića 3, 10 000 Zagreb, Croatia
aleskovi@ffzg.hr

Summary

Since the early 1980s, the scholarly community has been witnessing a considerable increase in the use of information and communication technologies (ICT). Specifically the use the Web has led to qualitative changes in the research community. With the advent of the Web 2.0 a new level of possible functionalities for science has been reached, leading to the concept of Science 2.0. Will the new research technology 2.0 change the way research is done and what aspects are already visible in current structures of scientific communication are questions this paper tries to answer.

Several clusters of expectation emerge from the prospect of applying the principles of Web 2.0 to scientific communication, like the opening of science communities towards public and the acceleration of dissemination of scientific research through new communication and collaboration tools. In the first part the authors will comment how the Web 2.0 challenges some traditional and known structures of scientific communication and explore possibilities of applying Web 2.0 principles (collaboration, collective validation, access and generation of information) to scientific work. In the second part the authors will present results gathered through analysis of Web 2.0 services that have been integrated

into academic databases and vice versa, the analysis of scientific information spaces that have been created within the Web 2.0.

Key words: Scientific communication, Web 2.0, Science 2.0, academic databases

Introduction

ICT and the Web have forced old-line institutions to adopt whole new ways of thinking, working and doing business. In the last decade a new version of the Web has emerged that has changed the premises of new developed models that comply with the Web 1.0 world, by transforming the very nature of knowledge, information and learning and by blurring the boundaries between using and producing, by connecting people and by expanding through user contributions like comments, texts, AV content, pictures, tags etc. This new cluster of practices, habits and business imperatives has influenced a wide range of domains and sectors and initiated a flood of buzzwords like *Business 2.0, Marketing 2.0, Libraries 2.0, Education 2.0* etc. Since scientific work has very much to do with information and knowledge, and since these are in the very heart of shifts initiated by the Web 2.0, it is time to analyze whether the restructuring of information spaces and creation of new communication patterns and information cultures do have an impact on scientific processes. What innovative configurations of scientific discourses are possible within this new and restructured Web 2.0 world and is it possible to talk about the concept of Science 2.0?

Web 2.0: new forms of online practices

The phenomenon known as Web 2.0 can be characterized by its technological aspects, but they are just a support for its conceptual nature which is what distinguishes it the most from the "old Web" (Banek Zorica et al, 2007, 93). Web 2.0 uses many new approaches for dealing with information including wikis, weblogs, aggregators, RSS and mashups. These often require active participation of users. Words that determine the Web 2.0 environment are participation, collaboration, sharing and communication. Applications like Instant Messaging and Peer-to Peer networks can be perceived as precursors of the Web 2.0 hype. But with the Web 2.0 the focus has shifted from communication between two users towards communication within groups of users (many-to-many relationships). New are also multiple communication paths, where the user contributions are not any more a collection of isolated monologues, but represent a matrix of dialogues (Maness 2006). Web 2.0 and social software tools are directed towards online cooperation and user communication; they facilitate communication in virtual networks and enable collaborative work on specific undertakings. With the application of widespread Web 2.0 services like blogs, wikis, social bookmarking services, media sharing or academic paper services – to name just a few – traditional educational and scientific institutions and information agencies transform themselves from a place of passive information consumption to a dynamic, participative, collaborative and creative knowledge production space. Many of the very words that describe Web 2.0 phenomenon's – communication, critiquing, suggesting, sharing ideas – are in the heart of science and bear the capacity to transform it.

Another facet of the Web 2.0 paradigm with the potential to transform classical perspectives in science is the notion of collective intelligence. The application of collective intelligence in science results in the breakdown of traditional assumptions about scientific expertise and the transformation of rigid scientific processes by more open-ended processes of communication in cyberspace. P. Walsh suggests that the expert paradigm (that dominates traditional "Science 1.0"), uses rules about how you access and process information, rules that are established through traditional disciplines. By contrast, the strength and weakness of a collective intelligence (that determines Web 2.0) is that it is disorderly, undisciplined and unruly. This certainly affects the way science is done (Jenkins, 2006, 53). The above discussion shows that central conceptions of the Web 2.0 coincidence with ideas that determine science, yet the question remains how can science harness this new possibility. Several clusters of expectation emerge from the prospect of applying the principles of Web 2.0 to scientific communication; according to Weller et al., the relations between Web 2.0 and scientific work require a differentiation into several dimensions:

- new ways of public relations for scientific and research activities (blogs, podcasts etc)
- collective knowledge generation and management
- new structures of scientific communication (dissemination und discussion of scientific contents, finding and access of scientific information (2007).

The new anatomy of science: Science 2.0

Science has always been looking for solutions which will minimize organizational and technical routine tasks occurring during research processes and simplify scientific workflows. Recent endeavors undertaken in this direction are known by the generic term e-science.[1] The term e-science, as it was designed by John Taylor, refers to "global collaboration in key areas of science, and the next generation of infrastructure that will enable it". (Taylor, 2001) This original definition implies that e-science consists not only of tools and technologies, but depends on pooling resources and connecting ideas, people and data. It has to do with information management as much as with computing.

Therefore, the concept *Science 2.0* is complementary with the idea of e-science and may be defined as a mean for realizing the principles of e-science. Although efforts in this direction are still too scattered to be called a movement, an

[1] The terminology is not globally uniform, while in Europe the use of the term e-science is quite common, the equivalent term in the USA is "Cyberscience".

ever growing number of researchers are beginning to harness wikis, blogs and other Web 2.0 technologies as a potentially transformative way of doing science.

Potentials

Unifying the roles of scientific producers and consumers

The common perception of the user – in our case a scientist – is mostly defined by passive and one-directional information consumption. In this perception, the scientist can participate in his or her science communication system by following two options: (a) using informal channels (personal communication) or (b) preparing a critical article and publishing it via a formal channel, i.e. a scientific journal. (a) is effective in deed, but limited, and (b) is a time-consuming process. Through diverse Web 2.0 services it is possible to formalize informal channels and transform the scientist as a reader into the scientist as a prosumer, a person who simultaneously *pro*duces and con*sumes*. (Stock, 2007, 97)

Acceleration and multiplying of scientific communication processes

An often critiqued aspect of scientific work relates to its traditional communication tool, the journal. Journals are either about one-way communication, indirect communication (references, letters to the author or editor) or one to one communication (author and reader). This static structure of journals forces the inherently dynamic scientific discourse to decelerate. Assimilating Web 2.0 tools with scientific work helps to overcome the shortcomings of linear and static processes provided by journals and leads to the creation of dynamic, circular, modular, multidirectional and decentralized contexts.

New communication channels: blogs, wikis etc.

Communication is the driver of scientific processes. Only published (i.e. communicated) research results "exist" for the scientific community. Therefore, blogs as an online communication infrastructure that provides bi-directional communication and real dialogs are the perfect mean to speed up and revitalize existing static structures provided by journals. Wikis on the other hand are often used to create collaborative websites or to power community websites. Within a research group wikis can function as information- document- or project management systems. They are also suitable for discussion lists. Constructing a wiki within a community allows real and high-quality teamwork and constant updating. It creates conversations between researchers, lets them discuss the data and connect it with other data that might be relevant. Blogs and wikis permit users to make information available in ways that create a conversation.

Access points to knowledge, collaborative organization of resources

An achievement of the Web 2.0 is that users do not only actively participate in the provision of content but also are employed to organize and index (or tag) it. The set of terms a group of users tagged content with are not a predetermined set of classification terms but rely on user experience. No longer do the experts and professional catalogers have the monopoly in this domain. As users continue to add tags, grassroots organizational scheme begins to emerge which is usually referred to as folksonomy. Such systems currently predominate in the private area for users to manage photos, videos, link collections etc. For scientific use the collaborative managing of web links and bibliographic data could and already does represent a special niche of interest. In most cases, users define the tags primarily for their own use and for organizing and accessing their own collection, but at same time there is the possibility to exploit approaches and alternative perspectives of other users.

With the aid of folksonomies the end user in formal science communication, the reader, is able to contribute to the indexing of scientific documents. This is a significant and up to this moment insufficiently explored area, at least in the science domain[2]. Prior to the Web 2.0 and folksonomies, scientific databases, the common access points to scientific information, lacked effective mechanisms to gain insight into user perspectives. This inhibited the process of retrieving information since the ones who usually created the databases were not the ones who actually used them. The information users were not able to easily question the information- gatherers or organizers. On the other hand, user-created tags are searchable for everyone beside the interpreter-created controlled terms and the author-created text words and references.

The described enrichment of scientific databases through folksonomies shows a lot of benefits: tagging represents authentically the use of language inside scientific communities, it allows for multiple interpretations from different disciplines or different schools, it can help to recognize neologisms and new scientific results fast. Folksonomy is by no means opposing controlled vocabularies; it should be clear that the development will profit from tagging, because tagging provides a rich source of authentic term material (Stock, 2007, 99).

Alternative forms of evaluating and pre- reviewing scientific works

Besides folksonomies the Web 2.0 offers further opportunities for accessing content, like social navigation or collaborative filtering. These are based on directly or indirectly derived user judgments, reviews or comments. The most famous expression of social navigation mechanisms are probably recommendation systems. With their help users can determine most popular or best rated articles, or within a bookmarking system identify those websites that users with similar profile have bookmarked and tagged.

Those systems could also be very valuable in a reconceptualized scientific peer review or a quality-control system. Classical peer review has many shortcomings (quality, delay, bias). The described recommendation systems provide an

[2] In other areas a critical mass of solutions and applications already do exist, e.g. Library 2.0, Catalog 2.0

useful alternative tool for evaluating scientific work, as well as some other popularity markers (the equivalent of citations), like download counts, number of tags etc. The more scientists tag a document, the more relevance does this article seem to have for this people. This would lead to a new scientific “currency” besides citations, voting systems etc.

Risks

Beside the analyzed potentials of Science 2.0 there are as many risks and controversial points. Diverse issues that arise within Web 2.0 discussion refer to the domain of science as well. The collaborative model of knowledge production, mash-up practice and anonymity creates information spaces where authenticity, trustworthiness, authority and reliability have to be continually questioned. Misinformation emerges, is worked over, refined or dismissed before a new consensus emerges. A particular risk for scientists refers to the global *copy-past culture*: scientists who put preliminary findings online risk having others copy or exploit the work to gain credit or even patents; particularly in hypercompetitive fields where patents, promotion and tenure can hinge on being the first to publish a new discovery (Waldrop, 2008).

Another problem of current social networks and other community portals are that many of the systems depend on active participation (starting with the registration) of community members. Trustworthiness and diligence in dealing with personal data are the minimum requirements to ensure this participation. To ensure the exchange of knowledge within scientific communities a basis of mutual trust must be created in order to prevent misuse or abuse of private data.

Science 2.0: how far are we?

A small but growing number of researchers (and not just the younger ones) have begun to carry out their work via the wide-open tools of Web 2.0. And although their efforts are still too scattered to be called a movement–yet–their experiences to date suggest that this kind of web-based “Science 2.0” is not only more collegial than traditional science but considerably more productive (Waldrop, 2008).

Several clusters of expectation emerge from the prospect of applying the principles of Web 2.0 to scientific communication, like the opening of science communities towards public and the acceleration of dissemination of scientific research through new communication and collaboration tools. In order to determine the actual degree of integration of Web 2.0 services into formal and informal scientific communication channels, an ad-hoc analysis of potential forms of interlinking 2.0 services with science has been conducted in the first half of year 2009. The review was made in two directions:

1. Have Web 2.0 services have already been integrated into academic databases?

2. To what degree have scientific information spaces have been created within the world of Web 2.0?

Integration of Web 2.0 tools into scientific databases

Currently professional science databases of the "old" information industry (e.g., CAS, INSPEC, MEDLINE, BIOSIS, Web of Science) are at the beginning of exploring potential interactions between Web 2.0 tools and the services they provide. Existing features of this interaction will be illustrated through several databases and hosts: EBSCOhost, Project MUSE, Citeseer and EiVillage.

EBSCOhost

EBSCOhost provides the functionality of bookmarking. The system enables users to collect and save references on some social bookmarking site such as del.icio.us, Technorati, dig etc. A further and important step EBSCO host has undertaken in the direction of Web 2.0 is the new interface named EBSCOhost2.0. This interface incorporates Web 2.0 user controls such as a date slider, mouse-over previews, and modals making the interface more dynamic and complying with the 2.0 philosophy of rich-user experience.

Project MUSE

MUSE supports bookmarking/tagging for sites including CiteULike, Connotea, del.icio.us, Facebook. In that way it provides the community an insight into its resources. Syndication via RSS Feeds provides a mechanism for users to subscribe to important information regarding Project MUSE journals, journal issues, and announcements.

Citeseer

CiteSeer^X offers personalization and Web 2.0 features such as personal collections, tagging for articles, error correction and document submission (user-created content). Users can monitor specific papers for metadata updates via email and create bibliographies by marking and downloading specific records. The system itself offers bookmarking through several services such as Connotea and Bibsonomy, so in that way it includes its resources into social web.

EiVillage

EiVillage (includes the databases Compendex and Inspec) is probably the first host that has started to work with folksonomies. EiVillage enables the user to assign tags to documents. He can choose to make those tags accessible to colleagues, peer groups and even to all the users in the Engineering Village community. Users may also decide to keep tags private for personal use. Engineering Village databases have traditionally relied upon records being classified by experts using structured indexes. Now, by adding record tagging the power to classify records and create content has been extended to users. Users can tag records based on how they define a record's relevance and importance. By choosing to expose those tags, Engineering Village users' community is provided with a powerful way to identify engineering content other users find meaningful.

Science in Web 2.0 information spaces

From this part of analysis individual blogs written by scientists where excluded since they are already common and exist in a great number. Attention was paid to more complex and sophisticated services as well as bookmarking services.

2coolab

2coolab[3] represents a platform for cooperation. It allows using bookmarks or tags and sharing Internet resources, from articles to video clips. It allows building networks and find, evaluate and initiate contact with new people The major characteristic is that 2coolab is a completely free, open and independent service. Anybody can open a user account and add or save bookmarks, or even create groups. Members of groups can evaluate these resources (by adding ratings and comments. Also, it is possible to use RSS feed to receive notification's about members, groups and users.

SciTopics

SciTopics[4] is a free expert-generated knowledge-sharing service for the scientific community. It serves as an information and collaboration tool for researchers. It is designed as a wiki that invites the scientific, technical and medical community to participate by posting comments, feedback, questions and discussion items. In order to contribute, interested users need to register as a SciTopics member and identify themselves with name and affiliation. In addition, users may find contact information for authors who they have identified as being potential collaborators on future research projects. Quality of the contributions is safeguarded by non-anonymous postings.

Bookmarking services

Bookmarking services are one of the mostly used 2.0 applications in scientific work. As previously mentioned the most known bookmarking services in the scientific community are Connotea, Bibsonomy and CiteUlike[5]. It is interesting to note that all these services have been created by the scientific community for

---

[3] http://www.2collab.com/nonLoggedInHomePage

[4] http://www.scitopics.com/

[5] http://www.connotea.org/, http://www.citeulike.org/, http://www.bibsonomy.org/

the scientific community.[6] They are usually based on already existing popular services like del.icio.us. Since they where primarily created for scientific and research needs these services have special features for this particular audience.
Every of the analyzed service has its own distinctive features, like CiteUlike which combines Web 2.0 tools with traditional software for the bibliographic handling of information, or Bibsonomy which aims to integrate the features of bookmarking systems with team-oriented publication management. All of the services continuously expands existing functionalities and includes new ones, like new resources for the systems automatic recognition and processing (Connotea), the improvement of tagging tools, group and private bookmarks, or work on the convergence with the Semantic web by concentrating on machine understandable tagging and new methods for extracting semantics from folksonomies (Bibsonomy).

Trends

Big databases are starting to incorporate in Web 2.0 trends and work primarily with services that focus on information discovering and handling: RSS, bookmarking, tagging, annotation and reviews. On the other hand, the web itself consist of the various number of different Web 2.0 services which are created specifically for (and by) the scientific community. Unlike the nonscientific version of these services which are based on anonymity, more and more attention is dedicated to the prevention of the aforementioned risks that are usually caused by anonymity. Therefore, such services are oriented towards developing mechanisms for the identification of scientist, user groups/scientific communities etc.

Conclusion

The Web 2.0 with its tools and technologies has begun to alter science. The comparison of older or classical scientific practices with those carried out in the Web 2.0 environment shows a transformation and shift that improves many facets of scientific work and even allows science the return to values that where the supposed hallmarks of science, but bring about some ambiguous and controversial aspects.
The ad hoc analysis of scientific information spaces conducted here shows that today the technical and infrastructural preconditions that have the potential to raise the functionalities and improve the features of scientific processes are available. Anyway, as with many other domains, technology does not drive change as much as our cultural response to technology does. Therefore the ac-

[6] Connotea was created by the Nature Publishing Group, Bibsonomy by a team of experts and students at the Institute of Knowledge and Data Engineering in Kassel, Germany and CiteUlike at the University of Manchester.

ceptance and emergence of a Science 2.0 would require a big change in academic culture.

Despite the discussed pros and cons, Science 2.0 is beginning to proliferate; current applications and articulations of Science 2.0 allow scientists to built networks and communities, market themselves, improve multidirectional communication, publish or disseminate materials at the point of need, criticize and discuss, find and access better, organize better. But for Science 2.0 advocates, the real significance is the technologies' potential to move researchers away from a focus on priority and publication and overcome the shortcomings of the traditional review process. The conducted analysis of Web 2.0 services within scientific communication processes has shown that 2.0 tools have been integrated into academic databases and vice versa, that scientific information spaces have been created within Web 2.0 environments. Having in mind that prominent and respectable research organizations, commercial and non-commercial, have begun to effectively leverage Web 2.0 practices, this domain is certainly going to advance and prosper.

References

Banek Zorica, M; Špiranec, S.; Zauder, K. Collaborative tagging: providing user created organizational structure for Web 2.0. // *INFuture 2007: Digital information and heritage* / Seljan, S; Stančić, H. (ed). Zagreb: Odsjek za informacijske znanosti, 2007, 193-203.

Jenkins, H. Convergence culture: where old and new media collide. New York, London: New York University Press: 2006.

Maness, J.M. Library 2.0 Theory: Web 2.0 and Its Implications for Libraries //,*Webology* 3(2006)2. http://www.webology.ir/2006/v3n2/a25.html (2009-05-15)

Stock, W.G. Folksonomies and science communication: a mash-up of professional science databases and Web 2.0 services. // *Information Services and use*. 27 (2007), 97-103.

Taylor, J. Defining e-science. http://www.nesc.ac.uk/nesc/define.html (2009-08-04).

Waldrop, M. Science 2.0: great new tool or great risk? // Scientific American. Jan, 2008. http://www.scientificamerican.com/article.cfm?id=science-2-point-0-great-new-tool-or-great-risk (2009-08-06)

Weller, K et al. Wissenschaft 2.0? Social Software im Einsatz fuer die Wissenschaft // *Information in Wissenschaft, Bildung und Wirtschaft. Proceedings der 29. Online-Tagung der DGI* / M. Ockenfeld (ed). Frankfurt am Main: DGI, 2007, 121-136.