Advanced Annotation for Research Area

A position paper for W3C Workshop on Annotations, 2 April 2014, San Francisco

Sergey Parinov, CEMI RAS, Moscow, Russia

1. Research area specifics

Typically doing research scientists create different types of relationship between research outputs, outcomes and other objects of scientific data and information space (DIS). They mentally manipulate the research artifacts, extracted from reading materials, and discover scientific relationships between the artifacts and their own outputs. Some of these relationships become visible as citations within their research outputs. Another part of the relationships is directly not observable since the existed citation technique does not allow researchers to express it explicitly and correctly.

The research community will benefit, if these relationships will be more visible and also provide computer readable data about its specific characteristics, like a scientific inference of one research results from another, motivations to use research outputs, its impact, hierarchical and association relationships, and so on. The initial classes of scientific relationships and a set of semantic vocabularies were proposed in [1], [2]. Initially we used semantic linkage technique ([1], [2]) to allow scientists expressing of research relationships over DIS content. The open annotation technique could be also very useful here.

2. Advanced Annotation Use Cases

An annotating activity as a "creating association between distinct pieces of information" has some important specifics when it is used by scientists for making research. Some additional use cases can be suggested to the list at http://www.openannotation.org/usecases.html:

2.1. A scientist wishes to annotate a scientific relationship that on his/her opinion (as a fact or hypothesis) exists between a pair of research artifacts. The artifact here is some part (fragment, segment, etc.) or a whole research publication.

2.2. A scientist wishes to annotate that some research artifacts were used by him to produce a new scientific knowledge published in some materials. And to specify by the annotation which motivations he had in using each artifact, how it was used, how it impacted on his/her research.

Note: This case is a demand on the advanced citation technique which could replace a traditional one for using it in digital research publications.

2.3. A scientist wishes to annotate details on his/her tries and fails to use some research artifacts for own research by linking the annotation with their outputs (as a feedback about loss citations).

Note: It is also can be a part of advanced citation technique.

2.4. A scientist wishes to receive immediate notifications: (1) when other scientists annotated his/her research outputs (e.g. because they used it in they research), including data about a character of using; (2) when other scientists changed materials which were annotated by the scientist (since there is a need to check and may be revise made annotations).

2.5. A scientist wishes to have an ability to react on cases of annotating (using) his/her research outputs, e.g. to help/assist with proper using of the outputs, or to protest against wrong using, etc.

2.6. A scientist wishes to have a statistical "portrait" (e.g. on some date) of how a research artifact, or a publication, or a set of artifacts/publications of one author/group/organization was annotated (used) by the scientific community. Or how a scientist/group/organization annotated (used in their work) research outputs of other scientists.

3. Requirements/suggestions

Additionally to the list at http://www.openannotation.org/usecases.html

3.1. Annotation Model

3.1.1. Identity of a pair of related publications/artifacts: Source Resource and Target Resource. See more in [2].

3.1.2. Identity of an annotation as a pair of outgoing and ingoing linkages between the Source Resource and Target Resource [2].

3.2. Services

3.2.1. Central repository to collect existed annotations and API to get annotations existed for the specified resource (both outgoing and ingoing types) [1].

3.2.2. Managing of taxonomy of the scientific relationship classes and related semantic vocabularies. Central repository of semantic vocabularies submitted to be used for annotating of research artifacts [1].

3.2.3. Notifications of authors of: (1) annotations, and (2) resources linked by the annotations [2], [4].

3.2.4. Collecting and processing of statistics about annotations, its semantics and linked resources. Building of scientometric indicators including statistical "portraits" for each research artifact, publication, author and organization [3], [4].

4. Examples of Implementation

Socionet research information system (http://socionet.ru/) which is driven by the community of scientists speaking and communicating in Russian language [4] provides a prototype of emerging annotating technology that could benefit from global Web standards. Currently it works as interdisciplinary research data and information space (DIS) aggregating data from about 2000 providers with about 2m information objects and about 7m semantic linkages (annotations) between them.

It also includes an online workbench to create, manage and submit to DIS single materials, whole collections and archives, and to create/manage networks of semantic linkages (annotations) between DIS objects. A monitoring and scientometric services provide for research community useful scientometric database (updated daily) and notifications. All counted scientometrics indicators are public and can be used for research assessments and scientometric studies [3], [4].

5. References

1. Parinov, S. (2012). Open Repository of Semantic Linkages. In: Proceedings of 11th International Conference on Current Research Information Systems e-Infrastructure for Research and Innovations (CRIS 2012), Prague (2012), http://socionet.ru/publication.xml?h=repec:rus:mqijxk:29&l=en

2. Parinov, S. (2013). Towards a semantic segment of a research e–infrastructure: necessary information objects, tools and services. International Journal of Metadata, Semantics and Ontologies, 8(4), 322-331, http://socionet.ru/pub.xml?h=RePEc:rus:mqijxk:30&l=en

3. Parinov, S., Kogalovsky M. (2014). Semantic Linkages in Research Information Systems as a New Data Source for Scientometric Studies. Scientometrics, Volume 98, Issue 2 (2014), Page 927-943, http://socionet.ru/pub.xml?h=RePEc:rus:mqijxk:31&l=en

4. Parinov S., Kogalovsky M. and Lyapunov V. (2014): A Challenge of Research Outputs in GL Circuit: From Open Access to Open Use. In Proceedings of the Fifteenth International Conference on Grey Literature: The Grey Audit: A Field Assessment in Grey Literature. Bratislava, Slovak Republic (February 2014), http://socionet.ru/pub.xml?h=RePEc:rus:mqijxk:31&l=en