Towards Web Intelligence through the Crowdsourcing of Semantics

Sören Auer
EIS, University of Bonn and Fraunhofer IAIS
Römerstrasse 164, 53117 Bonn, Germany
auer@cs.uni-bonn.de

Dimitris Kontokostas
AKSW, Computer Science, University of Leipzig
Augustusplatz 10-11, 04009 Leipzig, Germany
kontokostas@informatik.uni-leipzig.de

ABSTRACT

A key success factor for the Web as a whole was and is its participatory nature. We discuss strategies for engaging human-intelligence to make the Web more semantic.

Categories and Subject Descriptors

H.1.2 [**Information Systems**]: User/Machine Systems—*Human factors,Human information processing*

1. TALK SUMMARY

A key success factor for the Web as a whole was and is its participatory nature – in a truly distributed and democratic nature everybody can publish information on the Web and interlink it with related content. In order to make the Web more intelligent, we have to empower people to easily create and interlink structured information and semantics. Also, rich semantics can probably not emerge only by automated means. In physics, the law of conservation of energy states that the total energy of an isolated system cannot change – it is said to be conserved over time. A similar observation seems to hold for semantics: rich semantics can not appear miraculously out of nowhere, even if the most advanced and powerful algorithms are applied. In fact, two main strategies for 'increasing the intelligence' of the Web seem to be currently prevalent:

- Translating non-machine readable semantics into machine-readable semantics – this is, for example, what most natural language processing techniques aim at.

- Enriching existing semantic representations with additional or related information, for example, through link discovery.

However, the most progress can be made, if we can put the human intelligence in the loop. There were a number of strategies aiming to achieve that: games with a purpose, semantic (data) wikis and crowd-sourcing. Of course these strategies are not strictly separable, but overlap in many aspects. Semantic Wikis [2, 3] apply the wiki paradigm of 'making it easy to add and change information' to structured knowledge-bases. Small bits and pieces of information can be added to the semantic wiki in an agile manner. In this regard, semantic wikis can be considered as one technology to facilitate crowd-sourcing. However, other than voluntary contributions to a wiki, crowd-sourcing often involves monetary incentives and large groups of crowd-workers aiming to achieve a specified goal in short time [6, 1]. Games with a purpose (or human-based computation games) [5] on the other hand also implement some kind of crowd-sourcing, where the monetary incentive is replaced by some entertainment and recognition.

In this talk, we will present some approaches regarding the involvement of humans for the crowd-sourcing of semantics. One example is DBpedia [4] along with the DBpedia mappings wiki, a two tier crowd-sourcing endeavor, where the low level semantic contributions on the first tier (i.e. Wikipedia edits) are refined and organized by mappings created in the second tier on the DBpedia mappings wiki. We will discuss challenges and lessons learned as well as identify further promising directions for a combination of human and machine intelligence on the Web.

2. REFERENCES

[1] M. Acosta, A. Zaveri, E. Simperl, D. Kontokostas, S. Auer, and J. Lehmann. Crowdsourcing linked data quality assessment. In *12th International Semantic Web Conference, 21-25 October 2013*, 2013.

[2] P. Frischmuth, M. Martin, S. Tramp, T. Riechert, and S. Auer. OntoWiki - An Authoring, Publication and Visualization Interface for the Data Web. *Semantic Web Journal*, 2014.

[3] M. Krötzsch, D. Vrandecic, M. Völkel, H. Haller, and R. Studer. Semantic wikipedia. *Journal of Web Semantics*, 5:251–261, 2007.

[4] J. Lehmann, R. Isele, M. Jakob, A. Jentzsch, D. Kontokostas, P. N. Mendes, S. Hellmann, M. Morsey, P. van Kleef, S. Auer, and C. Bizer. DBpedia - a large-scale, multilingual knowledge base extracted from wikipedia. *Semantic Web Journal*, 2014.

[5] K. Siorpaes and M. Hepp. Games with a purpose for the semantic web. *IEEE Intelligent Systems*, 23(3):50–60, 2008.

[6] A. Zaveri, D. Kontokostas, M. A. Sherif, L. Bühmann, M. Morsey, S. Auer, and J. Lehmann. User-driven quality evaluation of dbpedia. In *9th Int. Conference on Semantic Systems, I-SEMANTICS '13*. ACM, 2013.