Wikidata: A New Platform for Collaborative Data Collection

Denny Vrandečić

Wikimedia Deutschland e.V. / KIT,

denny.vrandecic@wikimedia.de

ABSTRACT

This year, Wikimedia starts to build a new platform for the collaborative acquisition and maintenance of structured data: Wikidata. Wikidata's prime purpose is to be used within the other Wikimedia projects, like Wikipedia, to provide well-maintained, high-quality data. The nature and requirements of the Wikimedia projects require to develop a few novel, or at least unusual features for Wikidata: Wikidata will be a secondary database, i.e. instead of containing facts it will contain references for facts. It will be fully internationalized. It will contain inconsistent and contradictory facts, in order to represent the diversity of knowledge about a given entity.

Categories and Subject Descriptors

H.3.5 [**Information Storage and Retrieval**]: Online Information Systems; H.5.3 [**Information Interfaces**]: Group and Organization Interfaces—Web-based interactions; I.2.4 [**Artificial Intelligence**]: Knowledge Representation; K.4.3 [**Computers and Society**]: Organizational Impacts—Computer-supported collaborative work

General Terms

Human Factors, Documentation, Languages.

Keywords

Semantic Web, Wikipedia, Linked Data, DBPedia.

1. Biography

Denny Vrandecic is director of the Wikidata project at Wikimedia Deutschland e.V. and a post-doctoral researcher at the Institute AIFB at KIT Karlsruhe Institute of Technology. His research interests include web-based collaboration of massive numbers of users. He was involved in the EU-funded projects SEKT, ACTIVE, and RENDER, the project Halo funded by Vulcan Inc., Seattle, and the BMBF-funded project SMART that deals with improving water resource management in the lower Jordan valley, partnering, among others, the state of Israel, the Kingdom of Jordan, and the Palestinian Authority. He is the founding administrator of the Croatian Wikipedia, which today has more than 80.000 articles. He consulted Metaweb's freebase.com service on creating their RDF export (now acquired by Google) and is co-inventor of the Semantic MediaWiki system (SMW). His contributions to Open Source software have received a number of prizes (Theseus Idea Competition, do.it award, two time Semantic Web Challenge finalist, and others).