Communications of the ACM
Volume 57, Number 10 (2014), Pages 78-85

Wikidata: a free collaborative knowledgebase

Denny Vrandečić, Markus Krötzsch

DOI: 10.1145/2629489

Table of Contents

This collaboratively edited knowledgebase provides a common source of data for Wikipedia, and everyone else.

Unnoticed by most of its readers, Wikipedia continues to undergo dramatic changes, as its sister project Wikidata introduces a new multilingual "Wikipedia for data" (http://www.wikidata.org) to manage the factual information of the popular online encyclopedia. With Wikipedia's data becoming cleaned and integrated in a single location, opportunities arise for many new applications.

Key Insights

- Wikidata provides a free collaborative knowledgebase all can share.
- Wikidata has quickly become one of the most active Wikimedia projects.
- Wikipedia, as well as an increasing number of other sites, taps content from Wikidata in every pageview, magnifying the data's visibility and usefulness.

Originally conceived in 2001 as a mainly text-based resource, Wikipedia[1] has collected increasing amounts of structured data, including numbers, dates, coordinates, and many types of relationships, from family trees to the taxonomy of species. It has become a resource of enormous value, with potential applications across all areas of science, technology, and culture. This development is hardly surprising, given that Wikipedia is committed to "a world in which every single human being can freely share in the sum of all knowledge," according to its vision statement (https://wikimediafoundation.org/wiki/Vision). There is no question this must include data that can be searched, analyzed, and reused.

It may be surprising that Wikipedia does not provide direct access to most of it, through either query services or downloadable data exports. Actual use of the data is rare and often restricted to specific pieces of information (such as geo-tags of Wikipedia articles used in Google Maps). The reason for this striking gap between vision and reality is that Wikipedia's data is buried in 30 million Wikipedia articles in 287 languages from which extraction is inherently very difficult.

This situation is unfortunate for anyone wanting to use the data but is also an increasing threat to Wikipedia's main goal of providing up-to-date, accurate, encyclopedic knowledge. The same information often appears in articles in many languages and in many articles within a single language. Population numbers for Rome, for example, can be found in English and Italian articles about Rome but also in the English article "Cities in Italy." The numbers are all different.

Wikidata aims to overcome such inconsistencies by creating new ways for Wikipedia to manage its data on a global scale; see the result at http://www.wikidata.org. The following essential design decisions characterize the Wikidata approach.

Open editing. As in Wikipedia, Wikidata allows every user to extend and edit the stored information, even without creating an account. A form-based interface makes editing easy.

Community control. Not only is the actual data controlled by the contributor community, so, too, is the schema of the data. Contributors edit the population number of Rome but also decide whether there is such a number in the first place.

Plurality. It would be naive to expect global agreement on the "true" data, since many facts are disputed or simply uncertain. Wikidata allows conflicting data to coexist and provides mechanisms to organize this plurality.

Secondary data. Wikidata gathers facts published in primary sources, together with references to these sources; for example, there is no "true population of Rome" but rather a "population of Rome as published by the city of Rome in 2011."

Multilingual data. Most data is not tied to a single language; numbers, dates, and coordinates have universal meaning, so labels like "Rome" and "population" are translated into many different languages. Wikidata is multilingual by design. While Wikipedia has independent editions for each language, there is only one Wikidata site.

Easy access. Wikidata's goal is to allow data to be used both in Wikipedia and in external applications. Data is exported through Web services in several formats, including JavaScript Object Notation, or JSON, and Resource Description Framework, or RDF. Data is published under legal terms that allow the widest possible reuse.

Continuous evolution. In the best tradition of Wikipedia, Wikidata grows with its community of editors and developers and the tasks they give it. Rather than develop a perfect system to be presented to the world in a couple of years, new features are deployed incrementally and as early as possible.

These properties characterize Wikidata as a specific kind of curated database.[8]

↑ Data in Wikipedia

The value of Wikipedia's data has long been obvious, with many efforts to use it. The Wikidata approach is to crowd-source data acquisition, allowing a global community to edit the data. This extends the traditional wiki approach of allowing users to edit a website. Wiki is a Hawaiian word for fast; Ward Cunningham, who created the first wiki in 1995, used it to emphasize that his website could be changed quickly.[17]

The most popular such system is Semantic MediaWiki, or SMW,[15] which extends MediaWiki, the software used to run Wikipedia,[2] with data-management capabilities. SMW was originally proposed for Wikipedia but was quickly used on hundreds of other websites as well. Unlike Wikidata, SMW manages data as part of its textual content, thus hindering creation of a multilingual, single knowledgebase supporting all Wikimedia projects. Moreover, the data model of Wikidata is more elaborate than that of SMW, allowing users to capture more complex information. In spite of these differences, SMW has had a great influence on Wikidata, and the two projects share code for common tasks.

Other examples of free knowledgebase projects are OpenCyc and Freebase. OpenCyc is the free part of Cyc,[16] which aims for a much more comprehensive and expressive representation of knowledge than Wikidata. OpenCyc is released under a free license and available to the public, but, unlike Wikidata, is not editable by the public. Freebase, acquired by Google in 2010, is an online platform that allows communities to manage structured data.[7] Objects in Freebase are classified by types that prescribe what kind of data an object can have; for example, Freebase classifies Einstein as a "musical artist" since it would otherwise not be possible to refer to recordings of his speeches. Wikidata supports the use of arbitrary properties on all objects. Other differences from Wikidata are related to multi-language support, source information, and the proprietary software used to run the site. The latter is critical for Wikipedia, which is committed to running on a fully open source software stack to allow all to fork, or copy and create one's own version of the project.

Wikipedia's data is buried in 30 million Wikipedia articles in 287 languages from which extraction is inherently very difficult.

Other approaches to creating knowledgebases from Wikipedia have aimed at extracting data from Wikipedia, most notably DBPedia[6] and Yago,[13] that extract information from Wikipedia categories and from infoboxes, the tabular summaries in the upper-right area of many Wikipedia articles. Additional mechanisms help improve extraction quality. Yago includes temporal and spatial context information, but neither DBpedia nor Yago extract source information.

Wikipedia data, obtained from these projects or through custom extraction methods, has been used to improve object search in Google's Knowledge Graph (based on Freebase) and Facebook's Open Graph and in answering engines, including Wolfram Alpha,[24] Evi,[21] and IBM's Watson.[10] Wikipedia's geo-tags are also used by Google Maps. All these applications would benefit from up-to-date, machine-readable data exports (such as the way Google Maps shows India's Chennai district in the polar Kara Sea, next to Ushakov Island). Among these applications, Freebase and Evi are the only ones that also allow users to edit or to at least extend the data.

A Short History of Wikidata

Wikimedia launched Wikidata in October 2012. Initially, features were limited, with editors only able to create items and connect them to Wikipedia articles. In January 2013, three Wikipedias—first Hungarian, then Hebrew and Italian—began to connect to Wikidata. Meanwhile, the Wikidata community had already created more than three million items. The English Wikipedia followed in February, and all Wikipedias were connected to Wikidata in March.

Wikidata received input from more than 40,000 contributors as of February 2014. Since May 2013, it has continuously had more than 3,500 active contributors, those making at least five edits per month. These numbers make it one of the most active Wikimedia projects.

In March 2013, Wikimedia introduced Lua as a scripting language for automatically creating and enriching parts of articles (such as the infoboxes mentioned earlier). Lua scripts can access Wikidata, allowing Wikipedia editors to retrieve, process, and display data.

Many other features were introduced in 2013, and development is planned to continue for the foreseeable future.

Out of Many, One

The first challenge for the Wikidata community was to reconcile the 287 language editions of Wikipedia; for example, for Wikidata to be truly multilingual, the object representing "Rome" must be one and the same across all languages. Fortunately, Wikipedia already has a closely related mechanism: language links, displayed on the left side of each article, connecting articles in different languages. These links were created from user-edited text entries at the bottom of each article, leading to a quadratic number of links; for example, each of the 207 articles on Rome included a list of 206 links to all other articles on Rome—a total of 42,642 lines of text. By the end of 2012, 66 of the 287 language editions of Wikipedia included more text for language links than for actual article content.

It would clearly be better to store and manage language links in a single location, and so became Wikidata's first task. For every Wikipedia article, a page has now been created on Wikidata for managing links to related Wikipedia articles in all languages; these pages are called "items." Initially, only a limited amount of data could be stored for each item: a list of language links, a label, a list of aliases, and a one-line description. Labels, aliases, and descriptions can now be specified individually for up to 358 languages.

The Wikidata community has created bots to move language links from Wikipedia to Wikidata, and more than 240 million links were removed from Wikipedia. Most language links displayed on Wikipedia are served from Wikidata. It is still possible to add custom links in an article, as needed in the rare cases where links are not bi-directional; some articles refer to more general articles in other languages, while Wikidata deliberately connects pages that cover the same subject. By importing language links, Wikidata gained a huge set of initial items "grounded" in actual Wikipedia pages.

Simple Data (Properties and Values)

To store structured data beyond text labels and language links, Wikidata uses a simple data model. Data is basically described through property-value pairs; for example, the item for "Rome" might have a property "population" with value "2,777,979." Properties are objects and have their own Wikidata pages with labels, aliases, and descriptions. Unlike items, however, these pages are not linked to Wikipedia articles.

On the other hand, property pages always specify a datatype that defines which type of values the property can have. "Population" is a number; "has father" relates to another Wikidata item; and "postal code" is a string. This information is important for providing adequate user interfaces and ensuring the validity of inputs. There are only a small number of datatypes, mainly quantity, item, string, date and time, geographic coordinates, and URL. Data is international, though its display may be language-dependent; for example, the number 1,003.5 is written "1.003,5" in German and "1 003.5" in French.

Not-So-Simple Data

Property-value pairs are too simple in many cases; for example, Wikipedia says the population of Rome was 2,651,040 "as of 2010" based on "estimations" published by the National Institute for Statistics, or Istat, in Italy (http://www.istat.it/); see Figure 1 for how Rome statistics can be represented in Wikidata. Even leaving source information aside, the information cannot be expressed easily in property-value pairs. One could use a property "estimated population in 2010" or create an item "Rome" in 2010 to specify a value for its "estimated population." However, either solution is clumsy and impractical. As suggested by Figure 1, we would like the data to contain a property "as of" with value "2010" and a property "method" with value "estimation." These property-value pairs do not refer to Rome but to the assertion that Rome has a population of 2,651,040. We thus

arrive at a model where the property-value pairs assigned to items can have additional subordinate property-value pairs we call "qualifiers."

Qualifiers can be used to state contextual information (such as the validity time for an assertion). They can also be used to encode ternary relations that elude the property-value model; for example, to say Meryl Streep played Margaret Thatcher in the movie *The Iron Lady*, one could add to the item of the movie a property "cast member" with value "Meryl Streep" and an additional qualifier "role = Margaret Thatcher."

Such qualifiers illustrate why we adopted an extensible set of qualifiers instead of restricting ourselves to the most common qualifiers (such as for temporal information). Qualifiers in their current form are indeed an almost direct representation of data found in Wikipedia infoboxes. This solution resembles known approaches to representing context information.[11,18] It should not, however, be misunderstood as a workaround to represent relations of higher arity in graph-based data models, since Wikidata statements do not have a fixed (or even bounded) arity in this sense.[20]

Wikidata also allows for two special types of statements: First, it is possible to specify that the value of a property is unknown; for example, one can say Ambrose Bierce's day of death is unknown rather than not say anything about it, clarifying he is certainly not among the living. Second, one can say a property has no value at all (such as in asserting Australia has no countries sharing its borders). It is important to distinguish this situation from the common case that information is simply incomplete. It would be wrong to consider these two cases as special values, becoming clear when considering queries that ask for items sharing the same value for a property; otherwise, one would have to conclude Australia and Iceland have a common neighbor.

Further details on the Wikidata data model and its expression in Web Ontology Language in Resource Description Framework, or OWL/RDF, can be found in Erxleben et al.[9]

Citation Needed

Property assertions, possibly with qualifiers, provide a rich structure for expressing arbitrary claims. In Wikidata, every such claim can include a list of references to sources that support the claim. Including references agrees with Wikipedia's goal of being a secondary (or tertiary) source that does not publish its own research but rather gathers information published in other primary (or secondary) sources.

There are many ways to specify a reference, depending on whether it is a book, a curated database, a website, or something else entirely. Moreover, some sources may be represented by Wikidata items, while others are not. In this light, a reference is simply a list of property-value pairs, leaving the details of reference modeling to the community. Note Wikidata does not automatically record provenance[19] but does provide for the structural representation of references.

Sources are also important as context information. Different sources often make contradictory claims, yet Wikidata intends to represent all views rather than choose one "true" claim. Combined with the context information provided by qualifiers (such as for temporal context), many statements could be stored about a single property (such as population). To help manage this plurality, Wikidata allows contributors to optionally mark statements as "preferred" (for the most relevant, current statements) or "deprecated" (for irrelevant or unverified statements). Deprecated statements may be useful to Wikidata editors, to record erroneous claims of certain sources or to keep statements that still need to be improved or verified. As with all Wikidata content, these classifications are subject to community-governed editorial processes, similar to those of Wikipedia.[1]

Wikidata by the Numbers

Wikidata has grown significantly since its launch in October 2012; see the table here for key facts about its current content. It has also become the most edited Wikimedia project, with 150–500 edits per minute, or a half million per day, about three times as many as the English Wikipedia. Approximately 90% of these edits are made by bots contributors create for automating tasks, yet almost one million edits per month are still made by

humans. Figure 2a shows the number of human edits during 14-day intervals. We highlight contributions of power users with more than 10 or even 100,000 edits, respectively, as of February 2014, as they account for most of the variation. The increase in March 2013 marked the official announcement of the site.

Figure 2b shows the growth of Wikidata from its launch until February 2014. There were approximately 14.5 million items and 36 million language links. Essentially, every Wikipedia article is connected to a Wikidata item today, so these numbers grow slowly. In contrast, the number of labels, 45.6 million, as of February 2014, continues to grow; there are more labels than Wikipedia articles. Almost 10 million items have statements, and more than 30 million statements were created using more than 900 different properties. As expected, property use is skewed; the most frequent property is "instance of" P31 (5.6 million uses) for classifying items; one of the least-frequent properties is P485 (133 uses), which connects a topic (such as Johann Sebastian Bach) with the institution that archives the topic (such as the Bach-Archiv in Leipzig).

The Web of Data

One promising development in Wikidata is the volunteer community's reuse and integration of external identifiers from existing databases and authority controls, including the International Standard Name Identifier, or ISNI, China Academic Library and Information System, or CALIS, International Air Transport Association, or IATA, MusicBrainz for albums and performers, and North Atlantic Basin's Hurricane Database, or HURDAT. These external IDs allow applications to integrate Wikidata with data from other sources that remain under the control of the original publisher.

Wikidata is not the first project to reconcile identifiers and authority files from different sources. Others include the Virtual International Authority File, or VIAF, in the bibliographic domain,[3] GeoNames in the geographical domain,[22] and Freebase.[7] Wikidata is linked to many of these projects yet also differs in terms of scope, scale, editorial processes, and author community.

The collected data is exposed in various ways; for example, current per-item exports are available in JSON, XML, RDF, and several other formats. Full database dumps are created at intervals and supplemented by daily diffs. All data is licensed under a Creative Commons CC0 license, thus putting the data in the public domain.

Every Wikidata entity is identified by a unique URI (such as http://www.wikidata.org/entity/Q42 for item Q42, Douglas Adams). By resolving this URI, tools are able to obtain item data in the requested format (through content negotiation). This follows Linked Data standards for data publication,[5] making Wikidata part of the Semantic Web[4,9] while supporting integration of other semantic Web data sources with Wikidata.

Wikidata Applications

The data in Wikidata lends itself to manifold applications on very different levels of data integration.

Language labels and descriptions. Wikidata provides labels and descriptions for many terms in different languages, possibly using them to present information to international audiences. Unlike common dictionaries, Wikidata covers many named entities (such as for places, chemicals, plants, and specialist terms) that may be very difficult to translate. Many data-centric views can be translated trivially term by term—think maps, shopping lists, and ingredients of dishes on a menu—assuming all items are associated with suitable Wikidata IDs. The open source JavaScript library qLabel (http://googleknowledge.github.io/qLabel/) provides this functionality for any website.

Identifier reuse. Item IDs can be used as language-independent identifiers to facilitate data exchange and integration across application boundaries. Referring to Wikidata items, applications can provide unambiguous definitions for the terms they use that are also the entry points to a wealth of related information. Wikidata IDs thus resemble digital object identifiers, or DOIs, but emphasize (meta)data beyond online document locations and use another social infrastructure for ID assignment. Wikidata IDs are stable: IDs do not depend on language labels; items can be deleted, though IDs are never reused; and the links to other datasets and sites further

increase stability. Besides providing a large collection of IDs, Wikidata also provides the means to support contributors in selecting the right ID by displaying labels and descriptions; external applications can use the same functionality through the same API.

Wikidata allows conflicting data to coexist and provides mechanisms to organize this plurality.

Accessing Wikidata. The information collected by Wikidata is interesting in its own right, and many applications can be built to access it more conveniently and effectively. Applications created as of early 2014 included generic data browsers like the one in Figure 3 and special-purpose tools, including two genealogy viewers, a tree of life, a table of the elements, and various mapping tools. Applications can use the Wikidata API to browse, query, and even edit data. If simple queries are not enough, a dedicated copy of the data is needed; that copy can be obtained from regular dumps and possibly be updated in real time by mirroring edits on Wikidata. The Wikidata Toolkit, an open source Java library (https://www.mediawiki.org/wiki/Wikidata_Toolkit), provides convenient access to the dumps.

Enriching applications. Many applications can be enriched by embedding information from Wikidata directly into their interfaces; for example, a music player might want to fetch the portrait of the artist just being played in the audio file. Unlike earlier uses of Wikipedia data (such as in Google Maps), application developers need not extract and maintain the data themselves. Such lightweight data access is particularly attractive for mobile apps. In other cases, application developers preprocess data to integrate it into their applications; for example, it would be easy to extract a file of all German cities, together with their regions and post-code ranges, that could then be used in an application. Such derived data can be used and redistributed online or in software under any license, even in commercial contexts.

Advanced analytics. Information in Wikidata can be further analyzed to derive new insights beyond what is already revealed on the surface. An important approach in this regard is logical reasoning, where information about general relationships is used to derive additional facts; for example, Wikidata's property "grandparent" is obsolete, since its value can be inferred from values of properties "father" and "mother." If an application developer is generally interested in ancestors, then a transitive closure must be computed. Such a closure is relevant for many hierarchical, spatial, and partonomical relations. Other types of advanced analytics include statistical evaluations of both the data and the incidental metadata collected in the system; for example, a researcher can readily analyze article coverage by language,[12] as well as the gender balance of persons described in Wikipedia articles.[14] As in Wikipedia, Wikidata provides plenty of material for researchers to study.

These are only the most obvious approaches to exploiting the data, and many as-yet unforeseen uses should be expected. Wikidata is young, and its data is far from complete. We look forward to new and innovative applications due to Wikidata and its development as a knowledgebase.[23]

↑ Prospects

Features still missing include support for complex queries, which is now under development. However, in trying to predict the future of Wikidata, the development team's plans are probably less important than one would expect; for example, the biggest open questions concern the evolution and interplay of the many Wikimedia communities. Will Wikidata earn their trust? How will each of them, with its own language and culture, access, share, and co-evolve the way Wikidata is structured? And how will Wikidata respond to the demands of the communities beyond Wikipedia?

The influence of the volunteer community extends to technical development of the website and its underlying software. Wikidata is based on an open development process that invites contributions, while the site itself provides many extension points for user-created add-ons. The community has designed and developed features (such as article badges for featured articles, image embedding and multi-language editing). The community has also developed ways to enrich the semantics of properties by encoding (soft) constraints, as reflected in the

guideline "Items should have no more than one birthplace." External tools gather this information, analyze the dataset for constraint violations, and publish the list of violations on Wikidata to allow editors to check if they are valid exceptions or errors.

These aspects of the Wikidata development process illustrate the close relationships among technical infrastructure, editorial processes, and content and the pivotal role the community plays in shaping Wikidata. However, the community is as dynamic as Wikidata itself, based not on status or membership but on the common goal of turning Wikidata into the most accurate, useful, and informative resource possible. This goal promises stability and continuity, even as it allows anyone to take part in defining the future of Wikidata.

Wikipedia is by all accounts one of the most important websites today, a legacy Wikidata must live up to. In only two years, Wikidata is already an important platform for integrating information from many sources. In addition, it also aggregates large amounts of incidental metadata about its own evolution and contribution to Wikipedia. Wikidata thus has the potential to be a major resource for both research and development of new and improved applications. Wikidata, the free knowledgebase anyone can edit, may thus bring us all one step closer to a world that freely shares in the sum of all knowledge.

Acknowledgments

The development team's work on Wikidata is funded through donations by the Allen Institute of Artificial Intelligence, Google, the Gordon and Betty Moore Foundation, and Yandex. Markus Krotzsch's research is supported by the German Research Foundation through the Data Integration and Access by Merging Ontologies and Databases, or DIAMOND, project (Emmy Noether grant KR 4381/1-1).

References

1. Ayers, P., Matthews, C., and Yates, B. *How Wikipedia Works: And How You Can Be a Part of It*. No Starch Press, San Francisco, CA, 2008.

2. Barrett, D.J. *MediaWiki*. O'Reilly Media, Inc., Sebastopol, CA, 2008.

3. Bennett, R., Hengel-Dittrich, C., O'Neill, E.T., and Tillett, B.B. VIAF (Virtual International Authority File): Linking Die Deutsche Bibliothek and Library of Congress name authority files. In *Proceedings of the World Library and Information Congress 72nd General Conference and Council* (Seoul, South Korea, Aug. 20–24). IFLA, Den Haag, The Netherlands, 2006.

4. Berners-Lee, T., Hendler, J., and Lassila, O. The Semantic Web. *Scientific American* (May 2001), 96–101.

5. Bizer, C., Heath, T., and Berners-Lee, T. Linked data: The story so far. *International Journal on Semantic Web and Information Systems 5*, 3 (2009), 1–22.

6. Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R., and Hellmann, S. DBpedia: A crystallization point for the Web of Data. *Journal of Web Semantics 7*, 3 (Sept. 2009), 154–165.

7. Bollacker, K., Evans, C., Paritosh, P., Sturge, T., and Taylor, J. Freebase: A collaboratively created graph database for structuring human knowledge. In *Proceedings of the ACM SIGMOD International Conference on Management of Data* (Vancouver, BC, Canada, June 9–12). ACM Press, New York, 2008, 1247–1250.

8. Buneman, P., Cheney, J., Tan, W.-C., and Vansummeren, S. Curated databases. In *Proceedings of the 27th Symposium on Principles of Database Systems*, M. Lenzerini and D. Lembo, Eds. (Vancouver, BC, Canada, June 9–12). ACM Press, New York, 2008, 1–12.

9. Erxleben, F., Günther, M., Krötzsch, M., Mendez, J., and Vrandečić, D. Introducing Wikidata to the Linked Data Web. In *Proceedings of the 13th International Semantic Web Conference* (Trentino, Italy, Oct. 19–23). Springer, Berlin, 2014.

10. Ferrucci, D.A., Brown, E.W., Chu-Carroll, J., Fan, J., Gondek, D., Kalyanpur, A., Lally, A., Murdock, J.W., Nyberg, E., Prager, J.M., Schlaefer, N., and Welty, C.A. Building Watson: An overview of the DeepQA project. *AI Magazine 31*, 3 (Fall 2010), 59–79.

11. Guha, R.V., McCool, R., and Fikes, R. Contexts for the Semantic Web. In *Proceedings of the Third International Semantic Web Conference, Vol. 3298 of LNCS*, S.A. McIlraith, D. Plexousakis, and F. van Harmelen, Eds. (Hiroshima, Japan, Nov. 7–11). Springer, Berlin, 2004, 32–46.

12. Hale, S.A. *Multilinguals and Wikipedia Editing*. arXiv:1312.0976 [cs.CY], 2013; http://arxiv.org/abs/1312.0976

13. Hoffart, J., Suchanek, F.M., Berberich, K., and Weikum, G. YAGO2: A spatially and temporally enhanced knowledge base from Wikipedia. *Artificial Intelligence (Special Issue on Artificial Intelligence, Wikipedia, and Semi-Structured Resources) 194* (Jan. 2013), 28–61.

14. Klein, M. and Kyrios, A. VIAFbot and the integration of library data on Wikipedia. *code{4}lib Journal 22* (Oct. 2013); http://journal.code4lib.org/articles/8964

15. Krötzsch, M., Vrandečić, D., Völkel, M., Haller, H., and Studer, R. Semantic Wikipedia. *Journal of Web Semantics 5*, 4 (Dec. 2007), 251–261.

16. Lenat, D.B. and Guha, R.V. *Building Large Knowledge-Based Systems: Representation and Inference in the Cyc Project*. Addison-Wesley, Boston, MA, 1989.

17. Leuf, B. and Cunningham, W. *The Wiki Way: Quick Collaboration on the Web*. Addison-Wesley Professional, Boston, MA, 2001.

18. MacGregor, R.M. Representing reified relations in Loom. *Journal of Experimental and Theoretical Artificial Intelligence 5*, 2–3 (1993), 179–183.

19. Moreau, L. The foundations for provenance on the Web. *Foundations and Trends in Web Science 2*, 2–3 (Oct. 2010), 99–241.

20. Noy, N. and Rector, A., Eds. *Defining N-ary Relations on the Semantic Web*. W3C Working Group Note, Apr. 12, 2006; http://www.w3.org/TR/swbp-n-aryRelations/

21. Tunstall-Pedoe, W. True Knowledge: Open-domain question answering using structured knowledge and inference. *AI Magazine 31*, 3 (Fall 2010), 80–92.

22. Unxos GmbH. GeoNames (launched 2005); http://www.geonames.org

23. Vrandečić, D. The rise of Wikidata. *IEEE Intelligent Systems 28*, 4 (July/Aug. 2013), 90–95.

24. Wolfram Research. Wolfram Alpha (launched 2009); https://www.wolframalpha.com

Authors

Denny Vrandečić (vrandecic@google.com) is an ontologist at Google, San Francisco, and was project director of Wikidata at Wikimedia Deutschland, Berlin, until September 2013.

Markus Krötzsch (markus.kroetzsch@tu-dresden.de) is lead of the Wikidata data model specification and research group leader at Technische Universität Dresden, Dresden, Germany.

Figures

Figure 1. Screenshot of a complex statement in Wikidata.

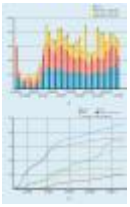

Figure 2. Growth of Wikidata: (a) bi-weekly number of edits for different editor groups and (b) size of knowledgebase.

Figure 3. Wikidata in external applications: the "Reasonator" data browser (http://tools.wmflabs.org/reasonator/)

Tables

Table. Basic statistics about Wikidata (August 2014).