Formal Ontology in Information Systems
P. Garbacz and O. Kutz (Eds.)
IOS Press, 2014

doi:10.3233/978-1-61499-438-1-251

Crowd-Sourcing Ontology Content and Curation: The Massive Ontology Interface

Samuel Sarjant [1], Catherine Legg, Matt Stannett, and Duncan Willcock
The University of Waikato, Hamilton, New Zealand

Abstract. Crowd-sourcing is an increasingly popular approach to building large, complex public-interest projects. The ontology infrastructure that is required to scaffold the goals of the Semantic Web is such a project. We have been thinking hard about what 'crowd-sourced ontology' might look like, and are currently advancing on two fronts: user-supplied *content* and user-supplied *curation*. We achieve the former by mining 90% of the concepts and relations in our ontology from Wikipedia. However other research groups are also pursuing this strategy (e.g. DBpedia, YAGO). Our claim to be on the cutting edge is in our latter goal. We are building a web portal: The Massive Ontology Interface, for users to interact with our ontology in a clean, syntax-light format. The interface is designed to enable users to identify errors and add new concepts and assertions, and to discuss the knowledge in the open-ended way that fosters real collaboration in Wikipedia. We here present our system, discuss the design decisions that have shaped it and the motivation we offer users to interact with it.

Keywords. ontology, crowdsource, interface

Introduction

The online public is drowning in information. The World Wide Web is full of text, and it is now also full of data, since the RDF triple standard has opened up data-sharing to an unprecedented degree, and the number of assertions now assembled in the Linked Data Cloud is staggering: estimated at 26 billion in 2011 [1]. However the text is not yet machine-readable, and the RDF data model can only support a shallow semantics, creating problems of underspecification and ambiguity in the gathered knowledge [2].

By contrast to raw text and RDF, an ontology is a highly structured knowledge base which supports complex assertions and reasoning. Such machine-readable representations of concepts have been described as the 'silver bullet' [3] for solving problems in information extraction, machine translation, database integration [4,5], and the Semantic Web [6]. Thus a great deal of work has been done to establish ontology as a research field, producing many languages and tools. But many of the tools are less widely used than initially hoped and there is currently little consensus on shared ontology frameworks. Although the Linked Data movement is teeming with instance-level data, as yet very little ontological structure has been put in place around them [1].

[1] Corresponding author email: `sarjant@waikato.ac.nz`

One might speculate as to why this is. It has been suggested that the success of the WWW was the "nearly embarrassing simplicity" [7] of its protocols, paired with its extremely modular design [8]. By contrast, ontologies are large and complex artifacts that take time to learn. They are also very holistically integrated: the introduction of a logical inconsistency or the loss of an axiom can affect the whole system in unpredictable ways. There is now a general consensus that manual creation of general-purpose ontologies is unworkable — the problem is too large and complex [9,10]. The next step would seem to be automated ontology construction, but this has its own challenges. In the following section we offer an anatomy of the problems we see in the field as it currently exists.

1. The Problem: Issues specific to Manually Built Ontologies

i) *Buy-in of Syntax / Philosophical knowledge (**Intelligibility**):* Ontologies have their roots in an older tradition of knowledge representation which derives in turn from the expert systems of classical AI. Such systems were designed to be reasoned over by theorem-provers, thus most ontology languages are built on *formal logic*, which is not easy for ordinary people to understand or work with. The original Cyc ontology project was famous for taking PhDs in philosophy 6 months to fully understand.[2]

ii) *Ontologies static — Don't Evolve (**Temporal Fidelity**):* Even if manual ontological representation of the world's knowledge could be completed and released, it would instantly go out of date as the world changes: countries change government, celebrities divorce, etc. The problem here arguably stems from viewing ontology as a packaged deliverable. Here the field arguably needs to learn from the evolution in the Western concept of an encyclopedia — somewhat painfully — as the once prized Encyclopedia Britannica lost its business model to Wikipedia where edits after significant events are near-instantaneous.

iii) *Perceived Epistemic Imperialism:* If a general-purpose ontology is manually created in a first-world country which contains many 'knowledge workers' and its concepts therefore stem from, for e.g., mainstream American life, this risks slighting other ways of seeing the world. To exactly the degree that we claim universal applicability and great usefulness for ontologies, it would seem that we should ensure that the knowledge is as general as possible. For example, OpenCyc asserts *"June solstice is a kind of summer solstice"*. This example is relatively benign, but one can imagine the disputes to which concepts with more political overtones might give rise. The history of 'edit wars' in Wikipedia[3] gives some sense of the feelings that can be tapped here.

iv) *Lack of user input into ontology application design (**Deployability**):* It was noted earlier that ontologies have been advertised by leaders in the field as the 'silver bullet' for solving problems in information extraction, machine translation and database integration, amongst others. If ontologies really have such a remarkable range of fundamental applications, why isn't everybody using them? Part of the problem is arguably precisely these very general ambitions of the field. Has there ever been an IT resource so potentially useful for everything, yet which in the short term people are so unclear what to do with? We believe that more engagement with real-world users, offering ontology applications specifically designed for them, is needed here.

[2] From author's experience. [3] `http://en.wikipedia.org/wiki/Wikipedia:Edit_war`

2. The Problem: Issues specific to Automatically Built Ontologies

v) *Accuracy of fully automated methods (**Semantic Fidelity**):* A significant amount of mapping between concepts in different ontologies can be done by straight label-matching. Strings such as 'Cate Blanchett' typically pick out the same entity in any knowledge resource. Matching relations is more difficult since they can partially overlap (e.g. `mother` and `parent`), but the vast majority of concepts in large-scale general ontologies are individuals and collections. However semantic disambiguation of the last ∼15% is difficult to achieve. For instance, if the car `Nissan Forum` is labelled 'forum' (as in Wikipedia), it will map to a concept representing the Roman monument, and despite appearances, `Silver Bank` is not a financial institution. Efforts to refine ontology mapping algorithms such as the Ontology Alignment Initiative[4] seem only to be chipping away at this issue (no winner has yet got above an F-measure of 93%).

vi) *Lack of agreed resources for evaluation (**Fidelity Standards**):* Formal evaluation is generally considered vital to rigorous research in information science. However if we examine the literature in automated ontology building, we find a lack of shared or systematic evaluation [4,11,12]. This seems to be caused in large part by a lack of suitable resources for comparison. Surveys of human subjects are of course possible, but they are not only generally small and extremely labour intensive, but also subject to variations in inter-rater agreement which are often alarmingly close to the improvements in ontological accuracy being evaluated. Human subjects are also not necessarily the best judges of taxonomic questions, e.g., *"Is the New Zealand Army an individual or a collection?"*, nor possessed of sufficient general knowledge to assess questions such as, *"Were the Kipchaks an ancient Turkish race?"*. It is sometimes stated that what is required is some kind of ontology gold standard dataset, however its possibility has been questioned [13] (p. 228)).

3. Our Solution

We address issue i) **Intelligibility** in a number of ways. Our hyperlinked browser interface is easily navigable and searchable, and allows public discussion of every concept and assertion in the ontology. It displays as much information as possible in natural language so that users can understand and contribute to the ontology without prior experience with formal languages. At the same time the fact that the interface is usable at a variety of levels of sophistication scaffolds users' development towards 'extreme knowledge engineer mode' as their contributions are checked and corrected by moderators and other users. We aim to address ii) **Temporal Fidelity** and iii) perceived epistemic imperialism by throwing the interface open to users to interact with at all times.

Regarding the issue of alignment accuracy in fully automated methods v) **Semantic Fidelity**, we along with others (e.g. [10]) believe that the only solution to getting the 'high-hanging fruit' is to relax to semi-automated methods. One must combine machine-generated knowledge discovery for scale with human correction for exactness. This raises the issue of how to ensure the accuracy of human contributions. We address this insofar as our interface is designed for users to work collaboratively, checking each other's contributions, and log all interactions so that, as with Wikipedia, they can be reverted.

[4] `http://oaei.ontologymatching.org`

Our final issue vi) was evaluation (**Fidelity Standards**). It was noted above that a problem of genuinely independent evaluation plagues automated ontology-building research projects. No ontology gold standard, such as an equivalent of the popular TREC[5] series of datasets and contests, has yet been developed. We suggest that the main reason for this is that if an automated ontology developer were to encounter a sufficiently large and accurate new knowledge resource, it would be only natural to add it to — rather than using it to evaluate — their ontology, and this is in fact what is happening. Therefore some other model of 'success' is needed in this field.

Consider how Wikipedia works, without consulting any external experts. Insofar as Wikipedia is a success [14], this is not because its knowledge has been evaluated against an external knowledge base, found to have X% accuracy, and thereby warranted to the world. For better or worse, Wikipedia has bypassed any such stamp of approval and is arguably now "too big to evaluate". So why is Wikipedia such a success? Arguably because so very many people find it useful enough to visit, read, and improve it further. It therefore seems that the objectivity and reliability of a knowledge resource might be determined by orders of magnitude of internal contributions as well as by assessment by an external standard. In fact, when the knowledge concerned is broad and general enough, this might be the only evaluation possible, as well as the one that really matters (for an extended argument for this claim that draws on formal sign theory, see [15]).

4. Crowdsourcing

A system may be defined as crowdsourcing if it openly enlists an indefinitely large number of humans to help solve a problem. It is increasingly being engaged in for purposes of information curation as data volumes increase, even by private enterprises [16]. Doan et al. [17] present a useful overview of the key challenges for crowdsourcing projects, namely: i) recruiting and retaining users, ii) determining what contributions users can make, iii) working out how to combine user contributions to solve the target problem, and iv) evaluating users and their contributions.

With respect to i), Doan et al. suggest that the five main strategies are: to *require* (only effective when one is in a position of authority over the user, which we are not), to *pay* (a strategy we are not pursuing), to *ask for volunteers*, to *offer a service* in return, and to *piggyback* on user traces from other systems. Our strategies are essentially to piggyback in our content and ask for volunteers in our curation.

With respect to ii), Doan et al. divide user contributions into: *evaluating* (users giving expertise and judgements to each other, e.g. book reviews on Amazon), *sharing* (users offering items to each other, e.g. photos on flickr), *networking* (users forming some kind of online community which is its own goal, e.g. Star Trek fans) and *building artifacts* (e.g. Wikipedia). Our system is explicitly focused on the first and the fourth of these goals, but has the potential to offer the other two as well.

With respect to iii), how do we combine user contributions to solve our target problem of ontology curation? This is a delicate issue given that ontologies are so holistically organized. We envision an evolving mix of self-contained tasks which may be performed by individuals and more complex work requiring collaboration enabled by discussion pages. This will be discussed in detail in Section 6.

[5] `http://trec.nist.gov`

Finally for iv), how do we evaluate users and their contributions? We have thought carefully about this, and plan to step users through three levels, based on the quality and quantity of their input: *normal user*, *moderator*, and *administrator*. Moderators can remove assertions, message users regarding their additions, and adjudicate whether a reported comment should be removed. Administrators have the same power as moderators, but may also change user levels, and view the log of moderation events.

4.1. Crowd-Sourced Ontological Content

An early example of formalizing general-purpose knowledge gathered from Web volunteers is OpenMind. In this project, hosted by MIT in the early 2000s, people entered common-sense statements in ordinary English, such as *"People pay taxi drivers to drive them places"*, producing 450 000 facts provided by over 9000 people [18]. This then became the basis for ConceptNet [19]. ConceptNet (now in its 5th iteration) is a rich resource, multilingual and organized in hypergraph structure. However it does have some limitations. Its nodes are indexed solely by name, creating semantic ambiguity (e.g. `Kiwi` refers to both bird and fruit). Its aim of integrating knowledge "from sources with varying levels of granularity and varying registers of formality" [19], (p. 3679) render its structure sufficiently loose and associational that it does not support logical consistency checking. Relatedly, its 'justifications' for its assertions are more provenance than proof.

As Wikipedia blossomed it became a natural resource for harvesting ontological content. In fact Wikipedia can be viewed as already constituting some kind of ontology whose nodes are its articles, for which the URLs serve as unique IDs. It is full of semi-structured knowledge which can be mined to provide full-blown ontological structure in a host of ways [5]. The DBpedia project [12] has transformed Wikipedia's semi-structured information (primarily infoboxes) into RDF triples to provide a giant open dataset, which has since become the hub of the billion-triple Linked Data Movement. A related effort is Wikidata [20], a giant open data repository built by harvesting Wikipedia's interwiki links, then its infoboxes and (to come shortly) its lists.

Wikipedia-harvesting efforts that are less data-driven and more in the traditional (subsumption hierarchy) ontology format are the efforts of European Media Lab Research Institute (EMLR), who built an ontology from Wikipedia's category network in three stages: [21], [22], and [23], ultimately deriving 49M facts indexed on 3.7M entities. YAGO [4] mapped the contents of Wikipedia's leaf categories to the WordNet taxonomy to index 5M facts on 1M entities, then [24] produced a much larger resource indexing 447M facts on 9.8M entities.

It is also worth mentioining Freebase [25], a collaborative knowledge base which contains many concepts and relations mined from Wikipedia and other sources such as MusicBrainz, along with direct user input structuring and maintaining the content. It currently holds 43M topics organized in a graph structure.

Finally, we mention our own Knowledge Miner [26,27], which builds concepts and relations from Wikipedia on a taxonomic backbone of OpenCyc, adding a further 2.2M concepts and 20M assertions extracted from Wikipedia. The use of OpenCyc enables logical reasoning with the expressivity of full first-order logic, and also 'common-sense knowledge' ontological quality control techniques utilising disjointness knowledge.

These ontology building efforts are increasingly converging. Thus, Cyc has opened an API to Wordnet, EMLR researchers have expressed plans to "link to taxonomies in other languages" [22], DBpedia has made YAGO available through its interface [12], YAGO2 has incorporated Geonames, and Linked Data seeks to ultimately subsume all of the above. For this reason, we see the next frontier in large-scale general purpose ontology building to be moving beyond the stockpiling of assertions to involving users in more intimate relationships with the knowledge. For this, interface usability will be crucial. So far ontology interface design is a remarkably under-explored area in HCI research, with just a few exceptions, e.g. [28].

4.2. Crowd-Sourced Ontological Curation

Little work has been done so far in this research area. A team at Stanford has been investigating the crowdsourcing of ontological curation by breaking it into *microtasks*: individual judgements such as: *Does instance X belong in class Y*? So far they have applied their methods to ontology *alignment* [29] and ontology *verification* [30]. However, they used paid workers on Mechanical Turk, so this research is of limited relevance to us. We also question whether microtasks are sufficient to perform all ontology curation, or whether some tasks require higher-level oversight.

The EMLR ontology offers no user interface as yet. DBpedia offers one but it does not allow user input to the knowledge. YAGO2 has made a good start in allowing user input, developing both tree-based and graphical browser interfaces which allow users to explore their ontology and give individual assertions thumbs-up or thumbs-down. However they do not allow discussion of the knowledge, or attempt to motivate users to contribute in any further way. Freebase offers a richer suite of functionalities through its schema editor, which allows users to create, edit and share mini-ontologies across its instance-level data. It also allows public viewing of user accounts and contributions. Wikidata is actively calling for user input, which at the present time mainly consists in adding missing labels, resolving inconsistencies, and writing documentation. Our 'Massive Ontology Interface' is designed to push accessibility and user interactivity even further along a number of dimensions which we will now describe.

5. Interface Architecture

Our interface is designed to expose as much of the ontology as possible while representing it in a simple, organised format. Most information is shown via *concept pages*, each of which displays all the assertions on that concept. The *query builder* page allows users to inference over the ontology in two main ways: i) finding bindings for variables, ii) proving statements true or 'not proven.' Other pages include: a *user page* for each user, which summarises their interactions with the ontology; *popular* and *recent* concept pages, and utility pages for moderation tasks.

5.1. Addressing Intelligibility

The **Intelligibility** of the interface is addressed primarily by utilising the natural language information associated with the ontology's concepts. As every concept typically lists one or more strings as synonyms (inherited from OpenCyc or extracted from Wikipedia), as well as a canonical name, the majority of the ontology's information can be represented in natural language format. Our natural language generation assembles this in-

Table 1. Example natural language sentence schemas. Conditional elements change with variable arguments.

	Assertion	`(isa Aristotle HomoSapiens)`	`(genls Pony Horse)`
	Predicate Pattern	$1 |*1(is)*|*(are)*| *an instance of* $2	$1 |*1(is)*|*(are)*| *a kind of* $2
i	NL Assertion	*Aristotle is an instance of human*	*Pony is a kind of horse*
ii	NL Proof	*Is Aristotle is an instance of human?*	*Is pony is a kind of horse?*
iii	NL Variable Arg 1	*What things are an instance of human?*	*What things are a kind of horse?*
iv	NL Variable Arg 2	*Aristotle is an instance of what things?*	*Pony is a kind of what things?*

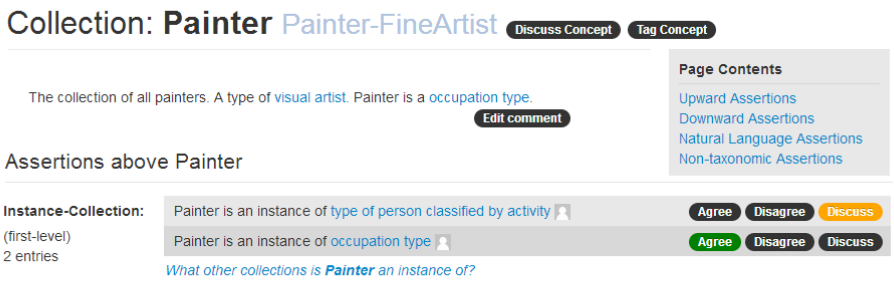

Figure 1. The concept page for `Painter-FineArtist` (top section).

formation using simple sentence schemas (Table 1). Cyc's functions use underscores to denote where the function's argument is shown (e.g. `DriedFn` defines *'dried* ___' for a synonym). The syntax is also displayed alongside the natural language in order to disambiguate identically named elements, and to scaffold user learning of the ontology structure. Although this solution is relatively simple, users in an earlier study we performed found it much easier to interpret than raw ontology syntax [31]. The Find-As-You-Type (FAYT) auto-complete search functionality throughout the interface also makes use of the natural language information.

5.2. *Concept Page*

Each concept page is structured in a top-down fashion, such that the most important information is near the top (concept name, comment, taxonomic information), followed by the rest of the information (lexical synonym assertions, non-taxonomic assertions). In Figure 1, a user can see that `Painter-FineArtist`, canonically named 'Painter,' is a collection with an editable comment and several assertions organised into groups. Each assertion is hyperlinked and represented in natural language format. Users can interact with every assertion via agreement, disagreement, or discussion (see Section 6.1), or adding their own assertions (see Section 6.2).

5.3. *Query Builder Page*

For broader views of the ontology, users can use the query builder page to submit queries in the form of either variable bindings (*'Who are all the Polish Actors?'*), or proofs (*'Is Bill Clinton a politician?'*) that return 'true' or 'not proven' due to the open-world assumption. Sample queries and proofs are provided to guide the user. We attempt to address **Intelligibility** here by using FAYT search boxes for locating the query's concepts, and apply our natural language generation to the query to further clarify it for the user.

Every query result is also displayed with a justification derived from the inferencing that produced it (a single justification for proofs and multiple for every valid variable binding substitution). Although many ontologies offer queries, the key difference with our application is the assistance that users receive in creating the queries, and their ability to view justifications for answers.

6. User Interaction

An important design principle of this interface is the ability for users to interact with and contribute to the ontology. Wikipedia is today's single biggest source of free knowledge, and a great example of what can be achieved when the web community collaborates online. We aim to emulate its processes for improving both the quality and quantity of its information, by enabling our users to vote on, discuss, and create concepts and assertions.

6.1. Voting and Discussion

An ontology is primarily defined by its assertions that link concepts. We encourage users to improve the quality (**Semantic Fidelity**) of these links through two operations: *voting* and *discussion*. Each user's profile page summarises their interactions.

Voting allows users to quickly agree or disagree with an assertion. Assertions with strong disagreement are treated as candidates for removal by moderators, while assertions with strong agreement increase the popularity of their respective concepts, increasing their visibility in the *Popular Concept* page. The aggregated operations improve the ontology by identifying weak assertions to remove, or strong assertions to base further automated growth upon. The second form of user interaction is *discussion*, which may be posted both on individual assertions or entire concepts. These discussions may involve users justifying their agreements or disagreements, proposing alternatives, or simply talking to other users within the context of the concept or assertion.

6.2. Ontology Content Creation

An important goal of this research is to allow users to freely add information to the ontology, in the same way that Wikipedia is freely editable. We have adopted a philosophy of 'instant creation, moderation later'. With instant creation, users immediately see their contribution made public and available to other users, (thereby increasing **Temporal Fidelity**). We address the **Semantic Fidelity** of user contributions via our logical and semantic quality control constraints, reducing the possibility of vandalism and guiding users towards making useful assertions.

Users may add new assertions to a concept directly from its respective concept page. Taxonomic assertions can be quickly added by specifying a relevant concept in the respective 'upward' or 'downward' assertion sections and submitting the assertion. If the assertion is logically and semantically consistent with the existing information, it is accepted into the ontology immediately. For non-taxonomic assertions, users can specify the relation and are then prompted to enter the arguments. The process is guided by the semantic argument constraints defined on the relation, which are displayed to the user and limit the results returned by the auto-complete suggestions for each argument. When an assertion is created, it records the user that added it. This information is normally only viewed when discussing the assertion but if the user has enabled Gravatar,[6] their chosen image will be displayed beside the assertion as a form of recognition.

Users are not limited to adding assertions to existing concepts. They can also create entirely new concepts as instances or subtypes of existing concepts. When creating a concept, we require that users enter a minimum amount of data about it: a unique internal name, the type of concept (at this stage: `Individual` or `Collection`), a comment briefly explaining it (as with Wikipedia, comments can be marked up to provide links to other concepts), a canonical natural language name, and at least one parent concept from the current ontology. If all these assertions are consistent with the current ontology, and the concept is created, it is immediately integrated into the ontology.

7. Conclusion

Our interface has been under in-house development and testing for the past 18 months. We are now reaching the point of advertising it to potential users. Building a crowd-sourcing project such as this is risky insofar as its success will depend on the quantity and quality of users it manages to attract. Our current ideas for attracting users include offering functionality for them to freely download modular sections of the ontology, and highlighting the potential usefulness of the knowledge by means of specific applications, such as a 'conflict of interest detector' which will examine taxonomically-represented corporate ownership structures to determine whether, for instance, a film review is published on a website owned by the company that produced the film. Also, as has been well-noted by [16], when looking to create an online community, good governance and "social best practice" are also key.

The interface is freely available at `http://bit.ly/MOIwaikato`.

References

[1] Tom Heath and Christian Bizer. Linked data: Evolving the web into a global data space. *Synthesis lectures on the semantic web: theory and technology*, 1(1):1–136, 2011.

[2] Prateek Jain, Pascal Hitzler, Peter Z Yeh, Kunal Verma, and Amit P Sheth. Linked data is merely more data. In *AAAI Spring Symposium: linked data meets artificial intelligence*, 2010.

[3] Dieter Fensel. Ontologies: A silver bullet for knowledge management and electronic-commerce (2000). *Berlin: Spring-Verlag*.

[4] Fabian M Suchanek, Gjergji Kasneci, and Gerhard Weikum. Yago: A large ontology from Wikipedia and Wordnet. *Web Semantics: Science, Services and Agents on the World Wide Web*, 6(3):203–217, 2008.

[5] Olena Medelyan, David Milne, Catherine Legg, and Ian H Witten. Mining meaning from Wikipedia. *International Journal of Human-Computer Studies*, 67(9):716–754, 2009.

[6] Tim Berners-Lee, James Hendler, and Ora Lassila. The semantic web. *Scientific american*, 284(5):28–37, 2001.

[7] R Guha. Semantic issues in web-scale knowledge aggregation. *Knowledge Systems Laboratory*, 2003.

[8] Catherine Legg. Ontologies on the semantic web. *Annual review of information science and technology*, 41(1):407–451, 2007.

[9] John F Sowa. The challenge of knowledge soup. *Research trends in science, technology and mathematics education*, pages 55–90, 2006.

[6] Globally Recognised Avatar: `http://en.gravatar.com`

[10] Amal Zouaq. An overview of shallow and deep natural language processing for ontology learning. *Ontology Learning and Knowledge Discovery Using the Web: Challenges and Recent Advances, Hershey, PA*, pages 16–37, 2011.

[11] Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary Ives. Dbpedia: A nucleus for a web of open data. In *The semantic web*, pages 722–735. Springer, 2007.

[12] Christian Bizer, Jens Lehmann, Georgi Kobilarov, Sören Auer, Christian Becker, Richard Cyganiak, and Sebastian Hellmann. Dbpedia-a crystallization point for the web of data. *Web Semantics: Science, Services and Agents on the World Wide Web*, 7(3):154–165, 2009.

[13] Klaas Dellschaft and Steffen Staab. On how to perform a gold standard based evaluation of ontology learning. In *The Semantic Web-ISWC 2006*, pages 228–241. Springer, 2006.

[14] GW Lovink and Nathaniel Tkacz. *Critical point of view: a Wikipedia reader*. Number 7. Institute of Network Cultures, 2011.

[15] Catherine Legg. Peirce, meaning, and the semantic web. *Semiotica*, 2013(193):119–143, 2013.

[16] Edward Curry, Andre Freitas, and Sean ORiáin. The role of community-driven data curation for enterprises. In *Linking enterprise data*, pages 25–47. Springer, 2010.

[17] Anhai Doan, Raghu Ramakrishnan, and Alon Y. Halevy. Crowdsourcing systems on the world-wide web. *Commun. ACM*, 54(4):86–96, April 2011.

[18] Push Singh, Thomas Lin, Erik T Mueller, Grace Lim, Travell Perkins, and Wan Li Zhu. Open mind common sense: Knowledge acquisition from the general public. In *On the Move to Meaningful Internet Systems 2002: CoopIS, DOA, and ODBASE*, pages 1223–1237. Springer, 2002.

[19] Robert Speer and Catherine Havasi. Representing general relational knowledge in conceptnet 5. In *LREC*, pages 3679–3686, 2012.

[20] Denny Vrandečić. Wikidata: A new platform for collaborative data collection. In *Proceedings of the 21st International Conference Companion on World Wide Web*, WWW '12 Companion, pages 1063–1064, New York, NY, USA, 2012. ACM.

[21] Simone Paolo Ponzetto and Michael Strube. Deriving a large scale taxonomy from Wikipedia. In *AAAI*, volume 7, pages 1440–1445, 2007.

[22] Căcilia Zirn, Vivi Nastase, and Michael Strube. Distinguishing between instances and classes in the Wikipedia taxonomy. In *The Semantic Web: Research and Applications*, pages 376–387. Springer, 2008.

[23] Vivi Nastase and Michael Strube. Transforming Wikipedia into a large scale multilingual concept network. *Artificial Intelligence*, 194:62–85, 2013.

[24] Johannes Hoffart, Fabian M Suchanek, Klaus Berberich, and Gerhard Weikum. YAGO2: A spatially and temporally enhanced knowledge base from Wikipedia. *Artificial Intelligence*, 194:28–61, 2013.

[25] Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. Freebase: A collaboratively created graph database for structuring human knowledge. In *Proceedings of the 2008 ACM SIGMOD International Conference on Management of Data*, SIGMOD '08, pages 1247–1250, New York, NY, USA, 2008. ACM.

[26] Samuel Sarjant, Catherine Legg, Michael Robinson, and Olena Medelyan. All you can eat ontology-building: Feeding Wikipedia to Cyc. In *Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology-Volume 01*, pages 341–348. IEEE Computer Society, 2009.

[27] Catherine Legg and Samuel Sarjant. Ontological quality control in large-scale, applied ontology matching. In *Proceedings of the The Eighth International Workshop on Ontology Matching*, 2013.

[28] J. Domingue. Tadzebao and webonto: Discussing, browsing, editing ontologies on the web. In *11th Knowledge Acquisition for Knowledge-Based Systems Workshop*, 1998.

[29] Cristina Sarasua, Elena Simperl, and Natalya F Noy. Crowdmap: Crowdsourcing ontology alignment with microtasks. In *The Semantic Web–ISWC 2012*, pages 525–541. Springer, 2012.

[30] Natalya F Noy, Jonathan Mortensen, Mark A Musen, and Paul R Alexander. Mechanical Turk as an ontology engineer?:Uusing microtasks as a component of an ontology-engineering workflow. In *Proceedings of the 5th Annual ACM Web Science Conference*, pages 262–271. ACM, 2013.

[31] Matt Stannett, Catherine Legg, and Samuel Sarjant. Massive ontology interface. In *CHINZ 2013: 14th Annual Conference of the New Zealand Chapter of the ACM Special Interest Group on Computer-Human Interaction*. ACM, 2013.