Ontologies in Enterprise Applications: Dimensional Comparison

Valeria de Paiva, William Jarrold, David Martin,
Peter F. Patel-Schneider, Karen Wallace, Peter Z. Yeh

Nuance Communications
Sunnyvale, CA 94085 U. S. A

Abstract

Many important categories of applications such as information integration, data analytics, and personal assistance require access to a general store of knowledge. The usefulness of such a store depends not only on its factual content, but also on its conceptual framework, or ontology. Ontologies vary markedly in their characteristics, and the value of an ontology varies according to the purpose for which it will be used. Thus, selecting one or more ontologies suitable for an application can be challenging. We have created a set of dimensions to consider when making this selection. Although these dimensions are motivated by the needs of conversational assistance applications, they can benefit the development of a wide variety of enterprise applications. We applied these dimensions to four large, general-purpose ontologies—Cyc, Freebase, SUMO, and YAGO—and made qualitative comparisons between them.

General-purpose ontologies are playing an increasingly central role in many important categories of applications such as information integration, distributed knowledge management, data analytics, and personal assistance. These ontologies aim to provide a framework for better organization and access of data, effective information and knowledge sharing, reliable information exchange, and improved coordination between distinct organizations or among members of the same organization. Consequently, companies face an increasing need to be able to choose the most suitable ontologies for their applications. This paper addresses that need.

We aim to provide an evaluative framework that may be applied to assess the usefulness of an ontology for any particular commercial application. Specifically, we set out to define the *dimensions* under which ontologies can be compared, and focus on those dimensions that assess *terminological knowledge* (definitions of concepts and their properties and relationships), not *assertional knowledge* (instances of concepts and facts about them). This focus is warranted because treatments of terminological knowledge are often overshadowed by assertional knowledge, or overlooked altogether. We believe our dimensions can provide an evaluative framework for comparing the strengths and weaknesses of different ontologies, which may prove useful for others investigating a robust use of semantic knowledge.

Our dimensions have resulted from our investigation of different ontologies suitable for conversational systems, as described in Kaplan's "Beyond the GUI: It's Time for a Conversational User Interface" [6], and are exemplified by our end-to-end speech-driven *second screen* application for television program discovery, described in [14]. Despite our focus application, we believe that our dimensions and our preliminary results from applying them should be applicable to a wide variety of companies that need to make the best choice for their applications.

In: Vinay Chaudhri, Alexandre Rademaker (eds.): Proceedings of the Workshop on Formal Ontologies meet Industry (FOMI), Rio de Janeiro, Brazil, 22-Sep-2014, published at http://ceur-ws.org

This is because of the extremely broad range of requirements of conversational systems. A conversational system is more than just speech recognition and synthesized speech; it must combine these voice technologies with natural-language understanding of the intention behind those spoken words. The intelligence comes from contextual awareness (who said what, when and where), perceptive listening (automatically waking up when you speak), and artificial intelligence *reasoning*. Moreover, a conversational system needs to be capable of question answering, intent recognition, semantic database integration, proactive behavior and even social intelligence.

Much work has been done in comparing and evaluating ontologies, especially devising quantitative metrics for ontologies using OWL as their representation language [13, 3, 12]. [13] characterizes most prior work as focusing on the mere structural aspects of an ontology and not considering the *semantics* of the ontology. Our paper shares the goal of semantic focus, but because our scope is more general in the kinds of languages we want to compare (i.e.not only OWL-based) it is less formal, and we actually apply our evaluation strategy to four ontologies. The goals and motivations of [10] are similar to ours, they provide a broad overview and discussion of issues in ontology evaluation, and approaches that may be used. While [10] touches on some of the same dimensions that we identify, it does not propose a specific set of dimensions that make up an evaluative framework. Since its focus is on approaches and specific evaluations from the life sciences literature, it does not give a comparative evaluation of specific ontologies as we have done. Similarly, we share with the work on OntoQA [12] the goal of general-purpose ontologies, the idea that one should have independent metrics for schemas, knowledge bases (KBs) and classes, but we concentrate efforts on qualitative dimensions for the schemas, while they concentrate on a specific system to measure OWL-based ontologies.

We subjected four general-purpose ontologies to our dimension-based comparison: ResearchCyc [8], Freebase [2],[1] SUMO [9], and YAGO [11, 5]. Resource constraints limited the number to four. Our analysis, and the numbers we report, are for these systems as they existed in early 2014. We chose these particular four systems because they are among the most prominent large, general-purpose, broad-coverage ontologies. They also take quite different approaches to modelling the world. Cyc is a long-standing project to assemble a comprehensive ontology and knowledge base of everyday common sense knowledge, with the goal of enabling AI applications to perform human-like reasoning. The Suggested Upper Merged Ontology (SUMO) and its domain ontologies form a large formal, open-source ontology with significant numbers of definitional axioms and rules. YAGO is a knowledge base developed at the Max Planck Institute for Computer Science in Saarbrücken, automatically extracted from Wikipedia, WordNet and other sources. Freebase is a large collaborative knowledge base consisting of metadata composed mainly by its community members. It is an online collection of structured data harvested from many sources, including individual 'wiki' contributions. Freebase aims to create a global resource which allows people (and machines) to access common information more effectively. Google's Knowledge Graph is powered, in part, by Freebase.

We also looked at the DBpedia [7] ontology and schema.org, http://schema.org, two other efforts that might be considered to be general-purpose ontologies. However, both of these were immediately seen to be unsuitable for our needs, as they are both comparatively small and inexpressive. As well, the meaning of constructs in schema.org is very hard to determine. This is not to say that the data available from DBpedia or the data described by schema.org markup would not be valuable, just that their ontologies are too limited to form the background ontology necessary for general conversational systems.

We first discuss the dimensions of comparison that we found useful when comparing the ontological schemas exemplified by the four chosen knowledge repositories. How the four compared in each one of these dimensions forms the basis for the matrix shown in table 1. We then discuss the main points or lessons learned from these one-by-one comparisons.

1 Ontology Evaluation Dimensions

Most knowledge repositories consist of three main components that are of interest to real-world applications: the ontology (i.e., schema, axioms, etc.), the representation language, and the data (i.e., instances of entity types in the ontology). These components are equally important, but our focus in this paper is on the ontology, which

- provides a scaffolding to organize data of interest to an application, and

[1] After we did our comparison, Google announced that Freebase will be discontinued in mid-2015, and that Google will provide a tool to help users add Freebase content to Wikidata (http://www.wikidata.org). We hope that users will transition most Freebase content to Wikidata and that the ontology supporting Wikidata will also grow into a general-purpose ontology on par with the ontologies in SUMO and ResearchCyc.

- supports the inferences over the data required by the application.

We propose ten dimensions for evaluating the utility of an ontology with respect to the above roles. These dimensions are inspired by those from data quality [1], but have been modified (and extended) for ontologies. We intend for these dimensions to provide a common yardstick for comparison across the ontologies of different knowledge repositories.

Domain Breadth: What is the breadth of domains covered by the ontology? For example, does the ontology cover a single domain such as geography or multiple domains ranging from entertainment to pathology? Broad domain breadth is important for applications that need to support a wide range of domains, such as virtual assistants.

Axiomatic Depth: Is the ontology complete in terms of the entity types, relations, and axioms for each domain covered? Axiomatic depth is orthogonal to *domain breath*, and is important for applications that focus on specialized domains such as healthcare, finance, etc.

Accuracy: How accurately do the entity types, relations, and axioms in the ontology reflect the real-world objects and domains they represent? Accuracy directly impacts the organizational quality of the data used by an application and the validity of any inferences over the data.

Consistency: Are there contradictions or incompatibilities in the ontology, e.g. contradicting axioms, incompatible generalizations, etc. If so, then how pervasive are these contradictions and incompatibilities? Like *accuracy*, consistency impacts data organization and inference validity. We note that it is impractical to expect a large ontology to have no inconsistency. Hence, this dimension is not intended to be binary. Instead, the intent is to provide a measure on the degree of consistency observed.

Integrity: Are the basic relations and axioms (e.g., generalization, instance, disjointness, selectional restrictions, etc.) present in the ontology? Have they been codified, or are they captured only in documentation and comments? Integrity is critical in enabling many common types of inference (e.g. subsumption, constraint violation, etc.) required by real-world applications, and can be viewed as *axiomatic depth* applied to the basic relations and axioms of an ontology.

Uniformity: Are the entity types and relations in the ontology used in a uniform manner? Do these uses conform with the semantics of the types and relations? Uniformity reduces the likelihood of modeling errors and hence the organizational quality of the data. Uniformity also improves the maintainability of an ontology and hence the maintainability of applications that use the ontology.

Redundancy: Are there redundancies in the ontology? For example, an ontology may have multiple entity types and relations for representing time, or generalizations that share similarities as well as differences. Redundancy is orthogonal to *uniformity*, and hence increases the difficulty in maintaining and using the ontology. Moreover, redundancy can result in unanticipated inference behaviors or invalid inference results.

Granularity: How granular are the entity types and relations in the ontology? An ontology that is too coarse (e.g. every type is a specialization of *Thing*) is not ideal. Neither is an ontology that is too fine-grained, e.g. an ontology with types like *Person-from-California*, a subclass of *Person*. Ontologies that are too coarse or fine-grained are difficult to extend, and hence less desirable for applications requiring additional customization.

Timeliness: How frequently is the ontology updated to reflect changes in the real-world objects and domains it represents? For example, how much time typically elapses for a pathology ontology to be updated when a new virus is discovered? Timeliness can be viewed as *accuracy* over time, and is important for applications that operate in dynamic domains.

Stability: How frequently is the ontology changed to fix mistakes, re-organized for efficiency purposes, etc? Stability directly impacts the maintainability of the applications that use the ontology.

2 Comparison Between Ontologies

We applied the dimensions outlined above to the ontologies (i.e. schemata) of four well-known knowledge repositories (i.e., YAGO, SUMO, Freebase, and ResearchCyc) as part of a preliminary comparison. We emphasize that our comparison focuses on the *ontologies* of the selected knowledge repositories, not the data or representation languages. Table 1 shows the qualitative findings that we observed from this comparison. Additional qualitative observations and remarks along the most salient dimensions for each ontology can be found in their respective subsections below.

We note that the findings presented are not based on a formal, rigorous evaluation and should be treated as qualitative observations suitable for high-level guidance. Thus, given the preliminary nature of our work, additional effort is needed such as providing more quantitative characterizations of the ratings for each dimension.

In addition, we deliberately have chosen not to prioritize these dimensions. The relative weights given to the dimensions should be established in the context of a particular application.

	YAGO	SUMO	Freebase	ResearchCyc
Domain Breadth	excellent	good	good	excellent
Axiomatic Depth	weak	average+	average	excellent
Accuracy	good	good+	good	excellent
Consistency	good	excellent	average+	good+
Integrity	weak	good	week	good+
Uniformity	good	excellent	average	excellent
Redundancy	average	excellent	weak	excellent
Granularity	poor	average+	average	good
Timeliness	weak	weak	average+	average
Stability	weak	good	average	good

Table 1: Qualitative findings from applying the dimensions outlined in the previous section.

2.1 The ResearchCyc ontology

Work on Cyc began in 1984 with the original goal of ontologizing human common sense reasoning and has continued until the present day. Currently, the Cycorp website claims adoption of Cyc-based products by a major database company, a Fortune 100 investment bank, and an oil company, among others.

Our focus, ResearchCyc (RCyc; in particular, CycL Version 10.148440, KB 7164), is one of three main offerings. The other two are OpenCyc (an Apache-licensed free version with almost all of the defining axioms removed) and EnterpriseCyc (a business-focused version). Cyc's *microtheories* are one of the means by which assertions are contextualized. A given microtheory may contain assertions made from a particular point of view or belief system or topic area. In addition a microtheory may define the time range of the assertions contained.

As seen in the table above, RCyc's greatest strengths evaluation were in **accuracy**, **domain breadth**, and **axiomatic depth**. Since we started use of RCyc in the conversational prototype [14] in the Spring of 2013, we have collectively looked at thousands of assertions and have only found between 5 and 10 factually incorrect sentences. That said, these assertions were safely encapsulated in crowdsourced microtheory and thus not subject to the typical level of review by Cycorp's professional ontologists. **Domain breadth** was deemed high because of the broad range of content and the raw numbers. In terms of "raw numbers", there is a large number of classes (52K, contrasted with Freebase's 25K types) and predicates (26K). Breadth is high, for example because there are 366 instances of #$GeneralMicrotheory (against a background of all 21K Microtheories) contrasted with 40 domain ontologies listed on the SUMO website. Additionally, there are not only content areas from business and economics, sciences, arts, etc., there is also a meta-knowledge ontology that enables expressing facts about provenance, reflection, and database integration. **Axiomatic depth** is high because RCyc has more definitional axioms than any of the sources we have looked at.

Timeliness was one of RCyc's weaker areas. Unlike the vast army of volunteers that a resource like Freebase has, the RCyc Ontology gets updated depending on what contracts Cycorp has. **Granularity** was also a relative weakness largely due to the relatively large number of predicates with highly specifc meaning or non-obvious distinctions. In **stability** the RCyc is tied with SUMO for first place. Nonetheless it is another relative weakness. RCyc is under active development. New terms are added regularly, and sometimes existing term names are changed.

2.2 The Freebase Ontology

Freebase [2] is a large-scale knowledge graph of topics and relations organized around a schema that provides typing (approx. 25K types), domain and range constraints, etc.

We observed that the Freebase ontology (which we will refer to as just Freebase for brevity) is strongest along the dimensions of **domain breadth** and **accuracy**. Freebase covers a wide variety of domains ranging from popular culture (e.g. movies, music, etc.) to the sciences (e.g. physics, geology, etc.). Its breadth is nearly comparable to the other ontologies examined, but many concepts that are types in the other ontologies examined are instances in Freebase. For example, specializations of profession (e.g., lawyer and doctor) and organism (e.g., dog and cat) are all instances in Freebase, not types.

We note that the instance-level data in Freebase (although outside the scope of this comparison) is more comprehensive than the other systems examined. Freebase has over 40 million topics (i.e. entities), and over 2 billion triples, corresponding to semantic relationships between these topics. This is an important consideration if the target application has significant data requirements.

Moreover, most types (and their properties) in Freebase accurately reflect the real-world objects they represent. We reached this observation by examining different types, their properties, and the expected values of these properties across multiple domains such as entertainment, biology, etc. We note that our observations on accuracy are on the schema only, not the data.

On the other hand, we observed that Freebase is weakest along the dimensions of **integrity** and **redundancy**. Although Freebase has generalization and type relations, they are largely absent. For example, Freebase does not have a unified type hierarchy like Cyc, SUMO, and YAGO. Instead, generalization relations only exist between select types, via the *included_types* relation, resulting in noticeable gaps in Freebase's type hierarchy. Similarly, an instance in Freebase can have multiple, incompatible types (e.g. an instance can be of type person and book subject). However, Freebase lacks sufficient disjointness constraints to prevent these integrity issues.

Freebase also has significant redundancies. For example, the *instance* relation in Freebase is used to assert that an instance X is of type Y. However, in the biology domain, the *organisms_of_this_type* relation is also used to capture the same relationship. Similarly, the profession domain uses the *specialization_of* relation to encode that one profession is a subclass of another. However, the biology domain uses the *higher_classification* relation to encode the same relationship between organism classifications. These redundancies cause significant overhead in using, extending, and reasoning with the ontology.

2.3 SUMO

The Suggested Upper Merged Ontology (SUMO) [9], http://www.ontologyportal.org/, together with its domain ontologies, form a large formal ontology licensed under several open-source licenses. As its name suggests, SUMO itself is an upper ontology with broad coverage of various categories including time and space, measures, sets, and classes. SUMO also includes a mid-level ontology (MILO) that expands the upper ontology into general groupings of domains. SUMO comes with a set of domain ontologies covering a variety of domains including political regions, people, and transportation. Domain ontologies are mostly contributed, but there is a vetting process for domain ontologies.

SUMO is written in a variant of KIF [4], which makes SUMO quasi-higher order. SUMO ontologies provide not just domain hierarchies, but also have a rich axiomatization of the domains. SUMO has tools to extract standard first-order logic and OWL from SUMO ontologies. SUMO has over 25,000 terms (including classes and properties) if the domain ontologies are included. However, this is a slightly inflated number, since some of these terms are nationalities and regions.

The upper and middle levels of SUMO provide excellent **breadth** for high-level organization of knowledge. At first glance the domain ontologies provide reasonable **depth** in their domains, but there are quite a few holes on closer investigation. For example, movies and tv shows are only lightly covered. There is the class MotionPicture for representing the class of all physical objects that have motion picture content (e.g. my DVD copy of "Gone with the Wind") but no class for representing instances of the conceptual creative works, e.g., "Gone With the Wind" itself. As well, there are only a limited number of domain ontologies, so many areas are only covered by the upper and middle levels of SUMO.

The upper and mid-level ontologies have been carefully vetted, and we have not found any problems with **accuracy**. The domain ontologies are less well vetted, but neither have we found any problems in them. SUMO has been machine-validated, meaning that there are no discoverable **inconsistencies** in its ontologies, nor classes that are necessarily empty.

The upper and mid-level ontologies are well-**integrated** and well-axiomatized, forming an excellent, **uniform** basis for the domain ontologies. The domain ontologies themselves are somewhat varied, but are generally well-axiomatized.

The upper and mid-level ontologies are parsimonious, with no observed problematic **redundancies**. The domain ontologies are well-separated, with no observable redundancies betwen them. The upper and mid-level ontologies take a middle, reasonable road on **granularity**. The domain ontologies are a bit mixed, with some providing only a coarse specification for their domain.

There is an update mechanism but it largely depends on submissions, and there is not that much use of SUMO, leading to questions of how **timely** updates are generated and applied to SUMO. There is a vetting mechanism for all updates to the ontology, leading to good **stability**. However, there is not a full formal process for checking that updates are reasonable.

In summary, SUMO and its domain ontologies form a good, well-axiomatized broad-coverage ontology. We were most impressed with the care that has gone into the upper and mid levels of SUMO, which provides an excellent scaffolding for general organization of knowledge. However, SUMO is much less useful at lower levels. There are not that many domains covered and even those that are covered are not always covered in sufficient detail for use as a core domain in a system.

2.4 The YAGO Ontology

YAGO, of which YAGO2s is the current release, is derived from Wikipedia, WordNet, and GeoNames. It is a huge semantic knowledge base in terms of the numbers of classes, entities, and facts, but defines a relatively small number of properties — about 100 properties for facts extracted from Wikipedia, a small number for capturing temporal and geospatial data, and several for capturing contextual (provenance) information. Roughly speaking, the upper level classes of YAGO correspond to synsets from WordNet, and the lower level classes correspond to Wikipedia categories. As of March 2011, YAGO had 292,070 classes based on Wikipedia categories, 68,446 classes based on WordNet synsets, 642 GeoNames-based classes, and 53 of its own classes. (These statistics refer to the YAGO release available at that time.)

For our purposes, the characteristics of YAGO are dependent on the nature of its three sources, and the limited ontological structuring provided by those three sources. YAGO naturally ranks high on **domain breadth**, due to the encyclopedic breadth of both Wikipedia and WordNet. It ranks high on **accuracy**, **consistency**, and **uniformity**, partly because of the high quality of those sources, and partly because of the high quality of its approach to data extraction. However, these high rankings are also partly due to the simplicity of what YAGO tries to express. In other words, because the bulk of the YAGO ontology is the class taxonomy, along with the small number of properties mentioned above, there aren't too many opportunities for mistakes along these three dimensions.

At the same time, this limited expressiveness of the ontology is reflected in a low score on the dimensions of **axiomatic depth** and **integrity**. The Wikipedia categories are often idiosyncratic, e.g., "Catalan handball clubs" and "Hotels established in 1806". Except for the taxonomic relations that YAGO determines for them, they are not axiomatized (that is, they have no properties directly associated with them). Although it is true that such categories carry information understandable by a human reader, that information is not accessible to a computational system without the creation of additional axioms. This could involve a large effort in manual annotation and/or further research on algorithms for extracting meaning from Wikipedia text. WordNet is not well structured for reasoning purposes, and has only minimal axiomatization. (Furthermore, some WordNet properties, such as meronym and holonym, are not present in YAGO.)

YAGO's ratings on **granularity** and **redundancy** are also directly related to the nature of its sources. For our purposes, **granularity** is particularly problematic. With over 360,000 concepts, there is a great deal of clutter; that is, concepts that are unlikely to be directly employed, such as the Wikipedia categories mentioned above.. In addition, most instances in YAGO have many types — some corresponding to Wikipedia categories and others corresponding to WordNet senses — and there appears to be no robust automated way to identify the most useful type(s) for the kinds of use cases we have in mind.

Finally, YAGO's low ratings on **timeliness** and **stability** are simply due to the small number of releases since its inception, the possibility of non-incremental exploration of future directions, and its status as an academic research effort with only a small community of contributors.

3 Analysis

The four ontologies we examined fall into two camps. The Freebase and YAGO ontologies are relatively inexpressive, containing little more than generalization relationships and role typing. The Cyc ontology and SUMO are much richer, with disjointness and many other axioms that define the concepts and relationships in the ontologies.

Of course, if you don't care too much about this extra information, or cannot utilize it, the added computational costs in working with expressive ontologies may be daunting. For example, if you are concerned with simple access to information (i.e., your application performs retrievals over data), then you probably only need simple typing, and the Freebase ontology would be a reasonable choice, particularly as Freebase contains a large amount of data. In this scenario, the Freebase ontology's issues with redundancy, integrity, and uniformity become less critical, because you are only returning retrievals for further analysis by humans, or for import into some other ontological framework. However, even in this scenario you can run into difficulties; for example, the two different relationships for birth date in Freebase—one for persons and one for non-human animals—makes even simple information access more difficult, showcasing a problem caused by Freebase's lack of uniformity. YAGO can be similarly used as a simple information source, but is less versatile in that, although it covers many domains, its ontology says very little about each of them (that is, it defines a much smaller number of properties).

However, if you need to perform reasoning over information, such as we envision in a conversational system to implement general user requests in terms of a knowledge repository, then the extra organizational power of Cyc and SUMO is extremely useful. Along with this extra power, Cyc and SUMO are better in terms of accuracy, consistency, and integrity, all important in a reasoning setting because incorrect information is magnified during reasoning. Between SUMO and Cyc, we prefer the Cyc ontology because it covers more domains and provides a more complete organization of many of these domains.

In some sense, any large system is going to have to interact with several ontological setups, as it will be pulling information from several sources, and the different sources are very likely to have different organizations of their information. This again argues for a powerful ontological system, one that can axiomatize the relationships between its organization and the organization of the other ontology.

On a final note, different ontologies, and data sources, come with different licenses. The licenses of some of the freely-available ontologies may cause problems in an industrial setting and these need to be considered separately.

4 Conclusion

We have defined 10 dimensions for use in evaluating general-purpose ontologies, and applied them in a qualitative, informal comparison of four such ontologies – SUMO and the ontologies associated with ResearchCyc, Freebase, and YAGO. We have discussed these particular ontologies so as to illustrate how the dimensions may be used in studying a broad range of ontologies. Because ontologies are so varied in their underlying philosophy, formalisms, and approach, we believe these dimensions provide a structure that can facilitate the selection of ontologies for many different kinds of enterprise applications.

As stated in Section 2, these results do not constitute a formal, rigorous evaluation. This should be addressed in the future by a statistically-meaningful evaluation of ontology content.

Our dimensions currently do not cover connections between an ontology and the the natural language expressions that may be used to organize, search or populate it. This is a drawback of our comparison, as ontologies such as SUMO have a full mapping from WordNet synsets to SUMO concepts, which is beneficial in practice. Moreover, Cyc and Yago also have similar mappings. Additional investigation is required to assess their strengths and weaknesses.

This refinement will also enable us to extend our work to comparing lexical ontologies. Clearly some of the concerns are similar, in that lexical ontologies can be domain-specific or aim for the whole language; lexical ontologies can pay attention to their consistency or not, etc. We hope that this broadening of the investigation may prove that our findings are robust. We also hope to help any application developer in need of guidance when it comes to the many choices offered by Linguistic Linked Open Data (http://linguistics.okfn.org/resources/llod/).

Another direction we would like to take this work is to multilingual ontologies. Many complain about the fact that hand-curated knowledge repositories can be fragile, as they represent some individual (or group of individuals) conceptualization of the world, with their social and cultural biases. If our ontologies can be automatically created in many languages in parallel, some of this criticism is curtailed. There are already some attempts at such multilingual ontologies (for example UWN, Menta, BabelNet and UBY) and presumably our dimensional criteria and/or a suitable modification will allow us to compare those.

References

[1] Carlo Batini, Cinzia Cappiello, Chiara Francalanci, and Andrea Maurino. Methodologies for data quality assessment and improvement. *ACM Computing Survey*, 41, 2009.

[2] Kurt D. Bollacker, Robert P. Cook, and Patrick Tufts. Freebase: A shared database of structured general human knowledge. In *Proceedings of the Twenty-Second AAAI Conference on Artificial Intelligence*, 2007.

[3] Juan Garca, FranciscoJose Garca-Pealvo, and Roberto Thern. A survey on ontology metrics. In MiltiadisD. Lytras, Patricia Ordonez De Pablos, Adrian Ziderman, Alan Roulstone, Hermann Maurer, and JonathanB. Imber, editors, *Knowledge Management, Information Systems, E-Learning, and Sustainability Research*, volume 111 of *Communications in Computer and Information Science*, pages 22–27. Springer Berlin Heidelberg, 2010.

[4] Michael R. Genesereth and Richard E. Fikes. Knowledge interchange format version 3.0 reference manual. Report Logic-92-1, Logic Group, Computer Science Department, Stanford University, 1992.

[5] Johannes Hoffart, Fabian M. Suchanek, Klaus Berberich, and Gerhard Weikum. YAGO2: A spatially and temporally enhanced knowledge base from Wikipedia. *Artificial Intelligence*, 2012.

[6] Ron Kaplan. Beyond the GUI: Its time for a conversational user interface. *Wired*, 21 March 2013.

[7] Jens Lehmann, Robert Isele, Max Jakob, Anja Jentzsch, Dimitris Kontokostas, Pablo N. Mendes, Sebastian Hellmann, Mohamed Morsey, Patrick van Kleef, Sören Auer, and Christian Bizer. DBpedia - A large-scale, multilingual knowledge base extracted from Wikipedia. *Semantic Web Journal*, 2014.

[8] D. B. Lenat. Cyc: A large-scale investment in knowledge infrastructure. *Communications of the ACM*, 38, 1995.

[9] I. Niles and A. Pease. Towards a standard upper ontology. In Chris Welty and Barry Smith, editors, *Proceedings of the 2nd International Conference on Formal Ontology in Information Systems (FOIS-2001)*, 2001.

[10] Leo Obrst, Werner Ceusters, Inderjeet Mani, Steve Ray, and Barry Smith. The evaluation of ontologies. In ChristopherJ.O. Baker and Kei-Hoi Cheung, editors, *Semantic Web*, pages 139–158. Springer US, 2007.

[11] Fabian M. Suchanek, Gjergji Kasneci, and Gerhard Weikum. YAGO: A Large Ontology from Wikipedia and WordNet. *Journal of Web Semantics*, 2008.

[12] Samir Tartir, I Budak Arpinar, Michael Moore, Amit P Sheth, and Boanerges Aleman-Meza. Ontoqa: Metric-based ontology quality analysis. *IEEE Workshop on Knowledge Acquisition from Distributed, Autonomous, Semantically Heterogeneous Data and Knowledge Sources*, 9, 2005.

[13] Denny Vrandečić and York Sure. How to design better ontology metrics. In *The Semantic Web: Research and Applications*, pages 311–325. Springer Berlin Heidelberg, 2007.

[14] Peter Z. Yeh, Ben Douglas, William Jarrold, Adwait Ratnaparkhi, Deepak Ramachandran, Peter F. Patel-Schneider, Stephen Laverty, Nirvana Tikku, Sean Brown, and Jeremy Mendel. A speech-driven second screen application for tv program discovery. In *Proceedings of the Twenty-Sixth Annual Conference on Innovative Applications of Artificial Intelligence (IAAI-14)*, 2014.