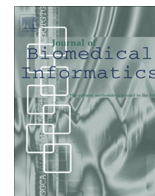




Contents lists available at ScienceDirect

# Journal of Biomedical Informatics

journal homepage: [www.elsevier.com/locate/yjbin](http://www.elsevier.com/locate/yjbin)



## Is the crowd better as an assistant or a replacement in ontology engineering? An exploration through the lens of the Gene Ontology

Jonathan M. Mortensen<sup>a,b</sup>, Natalie Telis<sup>b</sup>, Jake J. Hughey<sup>d</sup>, Hua Fan-Minogue<sup>c</sup>, Kimberly Van Auken<sup>e</sup>, Michel Dumontier<sup>a</sup>, Mark A. Musen<sup>a,\*</sup>

<sup>a</sup>Stanford Center for Biomedical Informatics Research, Stanford University, Stanford, CA 94305-5479, United States

<sup>b</sup>Biomedical Informatics Training Program, Stanford University, Stanford, CA 94305-5479, United States

<sup>c</sup>Department of Pediatrics, Stanford University, Stanford, CA 94305-5479, United States

<sup>d</sup>Institute of Computational Health Sciences, University of California, San Francisco, San Francisco, CA 94143, United States

<sup>e</sup>Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125, United States

### ARTICLE INFO

#### Article history:

Received 25 August 2015

Revised 6 December 2015

Accepted 3 February 2016

Available online xxx

#### Keywords:

Crowdsourcing

Ontology engineering

Gene Ontology

### ABSTRACT

Biomedical ontologies contain errors. Crowdsourcing, defined as taking a job traditionally performed by a designated agent and outsourcing it to an undefined large group of people, provides scalable access to humans. Therefore, the crowd has the potential overcome the limited accuracy and scalability found in current ontology quality assurance approaches. Crowd-based methods have identified errors in SNOMED CT, a large, clinical ontology, with an accuracy similar to that of experts, suggesting that crowdsourcing is indeed a feasible approach for identifying ontology errors. This work uses that same crowd-based methodology, as well as a panel of experts, to verify a subset of the Gene Ontology (200 relationships). Experts identified 16 errors, generally in relationships referencing acids and metals. The crowd performed poorly in identifying those errors, with an area under the receiver operating characteristic curve ranging from 0.44 to 0.73, depending on the methods configuration. However, when the crowd verified what experts considered to be easy relationships with useful definitions, they performed reasonably well. Notably, there are significantly fewer Google search results for Gene Ontology concepts than SNOMED CT concepts. This disparity may account for the difference in performance – fewer search results indicate a more difficult task for the worker. The number of Internet search results could serve as a method to assess which tasks are appropriate for the crowd. These results suggest that the crowd fits better as an expert assistant, helping experts with their verification by completing the easy tasks and allowing experts to focus on the difficult tasks, rather than an expert replacement.

© 2016 Published by Elsevier Inc.

### 1. Introduction

Ontologies enable researchers to specify, in a computational fashion, the entities that exist in the world, their properties, and their relationships to other entities. For instance, a researcher might encode in an ontology the kinds of cellular components that exist, such as a nucleus or ribosome. By leveraging such an ontology, a computer can recognize that a nucleus and a ribosome are, in fact, both a kind of cellular component and use that relationship when aggregating data. Further, ontologies allow everyone to “speak the same language” by creating a shared set of terms with clearly defined meanings. This property enables disparate parties to share data and to integrate them readily. For example, when

two data sources contain different information about cellular components (one focused on nuclei and the other on ribosomes) and use the same ontology to describe that information, a researcher is able to combine them with relative ease. These powerful properties enable ontologies to facilitate data integration, search, decision support, and data annotation [1]. Today, ontologies are ubiquitous. Indeed, the *Google Knowledge Graph* contains an ontology that supports an advanced understanding of entities on the Internet. With the Knowledge Graph’s ontology, Google provides additional information about an entity – a search for a movie also provides its star actors, director, budget, and so on [2]. Ontologies are latent in many of the technologies we encounter today. Given the important of ontologies, it is essential to ensure users are able to build and maintain them with minimal errors. In this work, we consider applying crowdsourcing to the task of ontology quality assurance – a task that is particularly challenging for biomedical ontologies.

\* Corresponding author.

E-mail address: [musen@stanford.edu](mailto:musen@stanford.edu) (M.A. Musen).

Biomedicine relies heavily on ontologies. In the clinic, they support electronic health records with tasks such as computerized physician order entry, alerting, and decision support [3]. In the life sciences, ontologies help combat the data deluge, giving researchers a tool to describe the intricate complexities of biomedicine and use that encoded knowledge to organize, annotate, and sift through data [4–7]. One of the most well known biomedical ontologies is the *Gene Ontology* [8]. By describing, in a computational fashion, experimental data and published literature with Gene Ontology (GO) terms, researchers are able to integrate results that are described with the same terms and gain insight about cellular components, biological processes, and molecular functions involved with a gene set of interest. One common use of these annotations and terms is *GO enrichment analysis*, wherein sets of differentially expressed genes are related, via a statistical overrepresentation analysis, to terms in GO [9]. These returned terms assist a researcher in developing hypotheses about the underlying biological phenomena that differentiates cases and controls. Of note, when one works with microarray data, GO enrichment analyses are standard practice. Such studies are pervasive in the literature.

The *Gene Ontology* its application is just one of the many examples the rapid increase in ontology use. Demonstrating this trend, The National Center for Biomedical Ontology provides a repository, called the Biportal, of over 450 ontologies ranging from brain anatomy to medical procedures [10]. These ontologies vary in size from hundreds of concepts to tens of thousands concepts and contain even more relationships between those concepts. However, as the size and complexity of ontologies continue to grow, so too does the difficulty of their development and maintenance. It becomes difficult for any single engineer to grasp the entirety of the ontology.

As a consequence of the difficulty of ontology development and maintenance, ontologies, not surprisingly, contain errors. Rector [11], Ceusters [12,13], Mortensen [14], and others have all identified systematic issues in *SNOMED CT*, an ontology intended to describe clinical encounters, and the *National Cancer Institute Thesaurus*, a clinical ontology focused on cancer. *SNOMED CT* contained domain-specific errors such as *Short Sleeper* *SubClassOf Brain Disorder* (brain disorders are not the sole cause of short sleep) and *Diabetes SubClassOf Disorder of the Abdomen* (diabetes is not a disorder of the abdominal cavity but rather of the endocrine system). In this work, we refer to techniques that identify such errors as “ontology verification”. Speaking to the frequency of these errors, there have been entire journal special issues dedicated to ontology quality verification methods [15]. Unfortunately, these methods are limited in their ability to catch domain-specific errors. For instance, a common class of computational ontology evaluation methods is metrics-based. In these methods, metrics are calculated about various characteristics of an ontology, such as its structure (e.g., average number of children), its syntax (e.g., number of syntax errors), its content (e.g., number of definitions) or adherence to best practices (e.g., using fully defined concepts) [16–22]. These metrics serve as a proxy for ontology *quality*. However, quality alone does not point to specific errors, limiting these methods in their ability to find errors such as those highlighted above (i.e., domain-specific errors). As a result, the currently accepted approach for identifying ontology errors is expert review. Only domain experts can interpret the symbols in an ontology and determine whether they reflect their understanding of the domain. However, the use of experts is very expensive. Experts cannot verify the large ontologies now found in biomedicine simply by inspection. In short, there is a fundamental trade-off between scalability (computational) and accuracy (expert) in current ontology verification methods.

Crowdsourcing, the practice of taking work traditionally done by one person and outsourcing it to online, anonymous crowds

[23], is one approach to overcoming the limitations existing ontology quality assurance methods. Researchers have shown that crowdsourcing can solve certain intuitive, human-level intelligence tasks more accurately than computers. For example, crowds of online workers might annotate an image with properties such as whether it contains a ball or a cat. Performing this task computationally remains a challenge, but humans can complete it easily. As crowdsourcing has grown, online platforms have emerged that provide users (i.e., requesters) with access to crowds (i.e., workers). The most common form of crowdsourcing on these platforms is micro-tasking. Here, many workers complete small, short tasks (requiring only minutes) for small rewards, including monetary compensation [23]. With this model, large tasks are completed quickly by large crowds that scale dynamically. Crowdsourcing is a complement to many computational techniques.

Researchers have begun using crowdsourcing extensively [24–26]. One challenge that remains in crowdsourcing research is understanding how the crowd can contribute to solving expert-level, knowledge-intensive tasks. In the biomedical realm, for one such expert-level domain, MacLean and Heer developed a crowd-based methodology to extract medical entities from patient-authored text [27]. They used crowd workers to find and to label terms. They then used these labels as a training set for a statistical classifier. This classifier then identified relevant medical terms written by patients in online forums. This system was able to identify medical terms with significantly higher accuracy in comparison to common automated medical extraction methods and thus showed that the crowd can work reliably on certain medical topics.

The use of crowdsourcing in ontology engineering, a knowledge-intensive task, is still nascent. There has been beginning investigation into micro-task based ontology mapping and gaming-based ontology tagging [28–30]. The success of this work suggests that crowdsourcing is a candidate to solve various ontology engineering tasks. Building on these efforts, in our previous work, we have developed, refined, and applied methods to perform ontology verification with the crowd [31,32]. At a high level, the method asks crowd workers to read sentences reflecting natural language representations of relationships in an ontology and to decide whether a sentence is TRUE or FALSE based on their knowledge and provided definitions. We have already applied successfully this method to verify a sample of *SNOMED CT*, finding a number of errors (More detail in Section 2) [14].

In this work, we applied the same crowd-based verification methodology to another ontology, the *Gene Ontology*. We investigated how the crowd performed in various configurations and how their performance varied with task difficulty and the quality of concept definitions. Further, we developed a strategy to predict a task's difficulty based on Google search results. In doing so, we make the following contributions:

1. We replicated previous work on crowdsourced ontology verification.
2. We compared and contrast our results on verifying GO with those of *SNOMED CT*.
3. We identified the important factors required for successfully using crowdsourcing for ontology verification.
4. We described a system for a hybrid between crowd-based and expert-based ontology verification (i.e. “group-sourcing”).

## 2. *SNOMED CT* verification study summary

The current work is based on our previous work [14]. Here, we summarize the results of that work. Note that the methodology is the same for both studies, and therefore Section 3 details the methodology itself.

Download English Version:

<https://daneshyari.com/en/article/6927843>

Download Persian Version:

<https://daneshyari.com/article/6927843>

[Daneshyari.com](https://daneshyari.com)