# Social Knowledge Collection

**Yolanda Gil**
Information Sciences Institute and Department of Computer Science
University of Southern California
gil@isi.edu

June 9, 2013

*To appear in*
*"Handbook of Human Computation", P. Michelucci (Ed), Springer, 2013.*

## Abstract

Social content collection sites on the Web allow communities of interest to create and share information at unprecedented scale. As a point of reference, MediaWiki (the wiki that powers Wikipedia) has millions of installations that allow non-programmers to contribute content. However, because the content in these sites has very little structure the information cannot be easily aggregated to answer simple questions. In recent years several approaches have emerged for social knowledge collection, allowing a community of contributors to structure content so that information can be aggregated to answer reasonably interesting albeit simple factual queries. This chapter gives an overview of existing social knowledge collection research, ranging from intelligent interfaces for collection of semi-structured repositories of common knowledge, semantic wikis for organizing and structuring information, and collaborative ontology editors to create consensus taxonomies with classes and properties. The chapter ends with a reflection on open research problems in this area.

## 1. Introduction

In the early days of the Web, people contributed content in their individual Web pages and sites for the benefit of all. The turn of the millennium saw an emergence of social content collection sites as a new way to share information for the benefit of others. Social content collection sites range from wikis to blogs, and cover topics as broad as encyclopedias[1], health[2], and how to do things[3]. What characterizes social content collection? First, these are social sites where many individual contributors collaboratively synthesize a body of content. There may be different kinds of contributions, some simply suggesting extensions and others with actual content and updates to the shared collection. Another feature is that there is some degree of coordination among the contributors. It can be very light coordination, for example a simple set of rules to organize the content. Alternatively, it can be very process-heavy where a complex editorial process is in place and contributors play specific roles with different oversight and responsibilities. For example, in its first year the English Wikipedia had fewer than 300 project pages (i.e., pages devoted to describing editorial processes and conventions) to organize the contributions of 21,000 topics, and as of September 2010 it reported 582,000 project pages and 7.9M topics[4], quantifying the growth of bureaucracy in the editorial process from 1:70 to 1:13. Third, social content collection is organized around a coherent theme. For example, a wiki may be devoted to the theme of "how to do things". Finally, the content has a nascent structure. For example, wikis are organized so each page is devoted to a topic and may be related explicitly to other topics through hyperlinks. For example, a page about how to go camping could be linked to a page about how to set up a camping tent.

Social content collection sites are incredibly popular. A search for "Powered by MediaWiki", the wiki software underlying Wikipedia that was developed by the Wikimedia Foundation and distributed under a Creative Commons license [Barrett 2008], showed 87M hits in September 2010 and 150M in March 2013. Myriads of other sites use other wiki software or different frameworks for web content management. Masses of volunteers are collaborating daily to create millions of formidable resources. They contribute content, play well-defined editorial roles, and organize the content around useful topic pages and categories.

Despite their popularity, social content collection sites have important limitations for search and query answering. Because the content has very little structure, they cannot aggregate information to answer simple queries. For example, Wikipedia content is well organized, but it is not structured to answer simple queries such as "What US Congress representatives own a business?", "What major cities in Europe have soccer teams that play in a national league?", or "What are all the versions to date of the Android software for cell phones?".

---

1 http://www.wikipedia.org

2 http://www.healthnet.org

3 http://www.wikihow.com

4 http://stats.wikimedia.org/EN/TablesWikipediaEN.htm

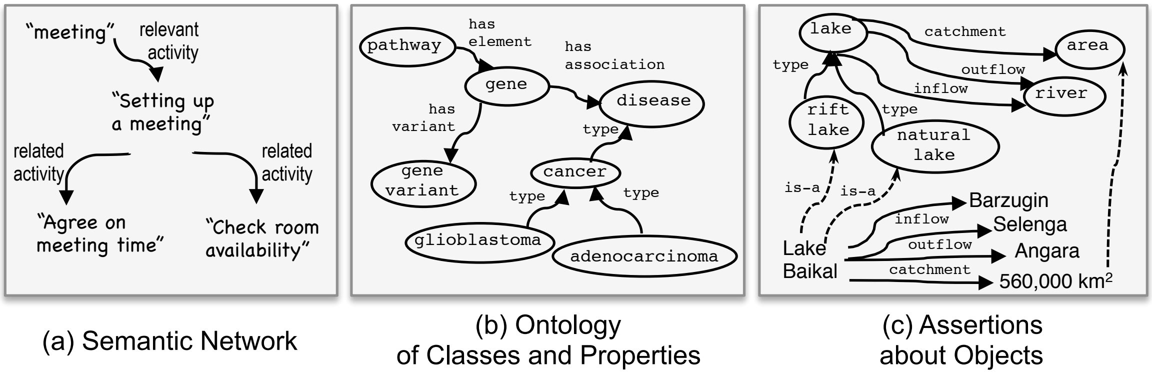


Figure 1. Social knowledge collection can target different kinds of internal representations, which have implications on the kinds of users that can contribute and the kinds of reasoning that the system can do about its knowledge: (a) semantic networks of semi-structured knowledge, (b) ontologies of classes and their properties, (c) assertions about objects.

In recent years, new approaches for social content collection have emerged that are more focused on structuring contributions. These approaches support social knowledge collection, representing content in such a way that it can be aggregated in meaningful ways to answer reasonably complex questions. They share the characteristics discussed above for social content collection sites: many individual contributors, there is some coordination among contributors, and contributions revolve around a theme. A unique feature of social knowledge collection is that the content is structured. Figure 1 illustrates some useful distinctions in the way that knowledge can be structured, using different approaches to knowledge representation [Brachman and Levesque 2004]. One possibility, shown in Figure 1(a), is to use semantic networks to link abstract concepts, but no reasoning is possible since the links and the concepts are not related to similar ones. In the figure, "Agree on meeting time" is a concept that has no relation to temporal representations of what time is, and therefore the system cannot answer questions about duration for example. Another possibility, shown in Figure 1(b), is to structure knowledge by defining ontologies, where classes of objects are created as well as properties of the objects in each class. In the figure, the class "gene" has a property of having an association with another class "disease", which in turn has several subclasses such as "cancer" and "glioblastoma". A third kind of knowledge concerns assertions about objects. Shown in Figure 1(c) are several assertions about the object Lake Baikal, for example that it has inflow from Barzugin and Selenga and has an area of catchment of 560,000 km2. Note that these assertions can be linked to ontologies, in this example there is an ontology of classes of lakes and their properties. The choice of knowledge structures determine the kinds of automated reasoning that can be performed on the knowledge collected, and therefore the kinds of questions that the system can answer about its knowledge. For example, since the inflow and outflow of lakes are to rivers, the system can infer that Barzugin, Selenga, and Angara are all rivers. It can then answer questions about rivers that flow into Lake Baikal.

Although the acquisition of structured knowledge has been an active area of research in artificial intelligence, the advent of the Web and the opportunity for collaborative knowledge capture presents new challenges [Gil 2011]. How should the interface be designed to guide contributors appropriately? What would be appropriate internal representation of the knowledge? What are successful approaches to attract and incentivize a healthy community of contributors? How can the quality of the knowledge collected be improved?

This chapter gives an overview of research to date and future challenges in social knowledge collection. Three major approaches are presented. The next section describes approaches to collect semi-structured repositories focused on common sense knowledge. The following section describes semantic wikis, extensions of traditional wikis that allow contributors to give more structure to topic pages and the links among them. After that, collaborative ontology editors are discussed as approaches to collect structured definitions of classes and properties. The chapter closes with a discussion of the research challenges ahead in this still nascent research area.

## 2. Collecting Semi-Structured Knowledge Repositories

An interesting area of research in social knowledge collection targets the creation of semi-structured repositories of knowledge. The knowledge is organized as semantic networks that, as we mentioned above, relate concepts that have no formal definitions and that do not fully support reasoning. Creating semi-formal repositories is easier for contributors with no expertise in logic or knowledge engineering, because they provide simple English statements that the system then tries to organize into more formal knowledge structures. The research in this area has focused on the collection of common knowledge, including common sense knowledge about world objects as well as daily and routine activities that require no particular expertise and are known by everyone but are not known to computers.

An analogy-based approach to collect knowledge about common objects was used in LEARNER [Chklovski 2003a; Chklovski 2003b]. LEARNER prompted volunteers for common objects, and upon an entry such as "newspaper" LEARNER would ask for useful things to know about newspapers. Contributors would respond with short sentences, for example "a newspaper is made of paper," "you can read a newspaper," and "you can carry a newspaper in your briefcase". LEARNER used simple natural language processing techniques to create a semantic network that made connections among the statements. LEARNER also used a novel analogical reasoning algorithm to detect commonalities among objects. So if a user entered "magazine" and said "a magazine is made of paper" and "you can read a magazine" then LEARNER would detect that magazines and newspapers seemed to have some things in common, and would ask whether "you can carry a magazine in your briefcase" was true along with other things it already knew about newspapers.

LEARNER2 [Chklovski 2005] was an extension of LEARNER focused on the collection of specific types of knowledge, originally designed to assist users with to-do lists [Gil et al 2012]. LEARNER2 toured for several years as an interactive kiosk at a science museum as part of a traveling exhibit called "Robots and Us" to raise

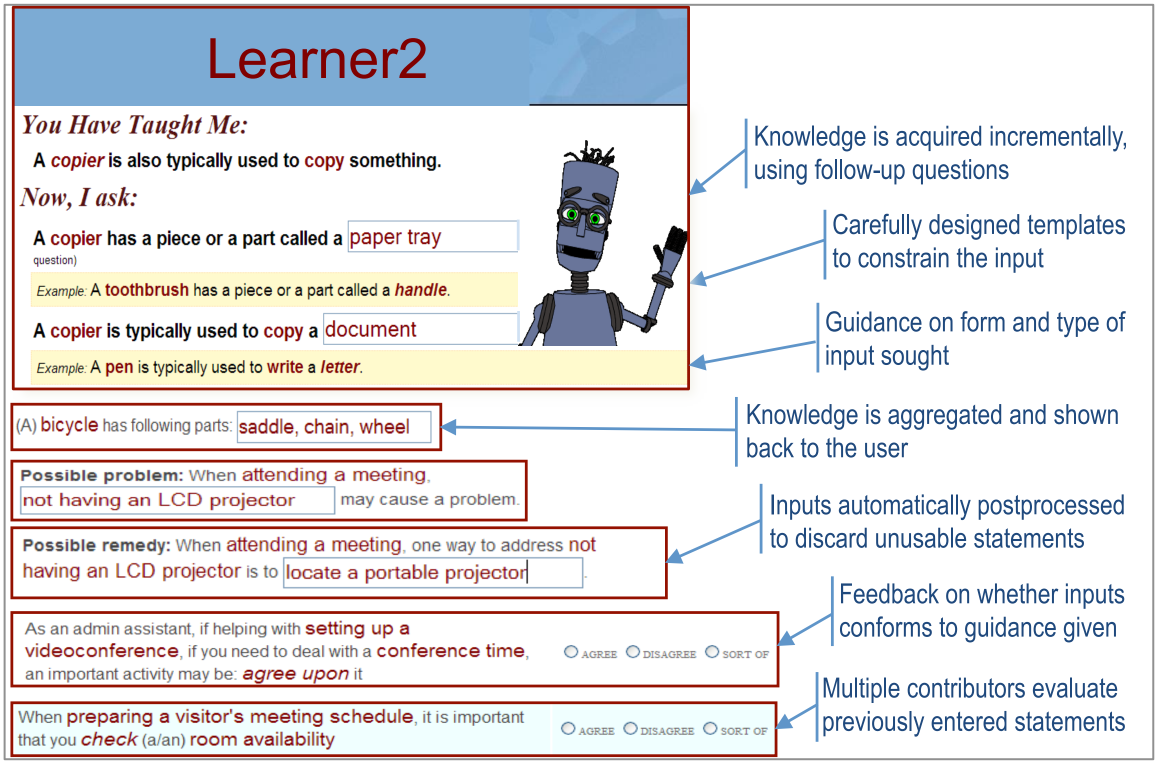


Figure 2. Learner2 collected semi-structured statements from volunteers about common objects and tasks. Its user interface was designed to guide contributors and improve the quality and breadth of the knowledge collected.

public awareness of the challenges of teaching common sense to computers. It collected more than 600,000 raw entries concerning task-oriented knowledge, such as objects relevant to a task, repairing task failures, descriptions of tasks in natural language, and decompositions of tasks into subtasks.

A detailed analysis of the statements collected with LEARNER2 revealed important findings [Chklovski and Gil 2005a]. First, redundancy of contributions helps identify high quality statements, so that if several contributors enter the same statement it is more likely to be correct. However, some of the statements also have overly high redundancy, drawing contributor effort away from areas where increasing coverage and increasing redundancy are more needed. That is, a large amount of contributors will think of entering the most common statements that are likely been already collected. This has consequences for the design of the user interfaces, so that contributors are enticed to make novel statements to the system [Chklovski and Gil 2005b; Chklovski and Gil 2005c]. Figure 2 illustrates key aspects of the design of the user interface. The user was asked follow up questions using templates designed to collect additional knowledge piecemeal. The user would get guidance on the type of input that the system was expecting, and would tend to enter simple statements. The knowledge entered was analyzed with simple natural language techniques to discard unusable statements that would not conform to the simple structure expected. The knowledge was also aggregated and shown back to the user for confirmation, and as a way to detect whether the user had understood what was expected. Finally, the statements acquired were shown to other contributors for validation. These user interface features can significantly improve the quality and coverage of the knowledge collected.

The Cyc FACTory [Matuszek et al 2005] allowed contributors to add facts to the Cyc knowledge base [Lenat and Guha 1990], which was designed to contain encyclopedic knowledge including common sense knowledge. Like LEARNER2, contributors were prompted with a template to fill, in this case a pre-defined schema based on the contents of the Cyc ontologies.

The Common Sense Computing Initiative (http://csc.media.mit.edu/) constellation of projects has been collecting common sense knowledge to create structured repositories [Havasi et al 2007]. Volunteers are prompted with objects that are mentioned in the contributions of others. A novel feature-based clustering technique was used to organize the contents collected [Speer et al 2008]. Specific collection efforts have been set up to collect knowledge about indoor objects to help with robot navigation [Gupta and Kochenderfer, 2004], about common tasks and events [Lieberman et al 2007], and about common objects and their properties [Havasi et al 2007]. These repositories have been used in a variety of contexts to assist users with tasks such as organizing pictures [Lieberman et al 2004] and personal task management [Smith and Lieberman 2010]. The site has collected over a million sentences from over 15,000 contributors.

## 3. Semantic Wikis

Semantic wikis are wikis with extensions that support the creation of structured content, and have reasoning capabilities that exploit that structure to organize the wiki's knowledge. Traditional wikis support some ways to structure content, for example by assigning categories to topic pages. Wikipedia has infoboxes for athletes, politicians, and countries. Infoboxes are essentially just a form for users to organize content, and are often used to extract knowledge bases from wikis (notably from Wikipedia) [Auer et al 2007; Weld et al 2008]. However, the system cannot reason about their content to answer questions, such as what rift lakes are in Russia. In contrast, a semantic wiki allows users to organize topic page categories as classes (or concepts) in a taxonomy, and to define properties that apply to each class. For example, the Wikipedia page for Lake Baikal would be linked to the page for Russia through a regular hyperlink such as Lake Baikal is in `[[Russia]]`, while in a semantic wiki the hyperlink would be Lake Baikal is in `[[country Russia]]` where country is a property. This enables the system to answer questions about lakes in Russia. Semantic wikis allow users to constrain properties by the range of values that they can take, which are called structured properties. As content is added using these structured properties, the semantic wiki can use reasoning and inference. Users can then query the content to generate dynamic content for wiki pages. Visualizations can be created automatically by overlaying semantic information in maps or charts.

An important feature of semantic wikis is their integration with semantic web standards. Each assertion is turned into a triple of the form <object property value> that can be expressed in the Resource Description Framework (RDF) standard

[Brickley and Guha 2004]. This makes the knowledge collected through semantic wikis compatible with the data already captured in many billions of interlinked RDF triples that are accessible on the Web and are known as the Web of Data or Linked Open Data[5] [Heath and Bizer 2011; Auer et al 2007].

Semantic wikis are becoming very popular, as they offer the simplicity of a wiki with additional capabilities to help contributors organize content. There are several implementations of semantic wikis. Semantic MediaWiki [Krotzsch et al 2007] is a diverse set of extensions for the popular MediaWiki wiki platform, and allows users to easily create new concepts and structured properties without enforcing consistency up front. OntoWiki [Auer et al 2006] is another semantic wiki that requires that a schema be defined before users enter content to populate it through a form-based web-interface. AceWiki [Kuhn 2009] provides a more powerful knowledge representation formalism than most other semantic wikis, with the cost of requiring the contributors to learn and use a semi-formal logical language designed for them by the wiki developers/administrators. [Bry et al 2012] give a detailed overview of semantic wikis and a thorough comparison of semantic wiki frameworks. Perhaps because of its more permissive and organic approach to structuring knowledge collaboratively, Semantic MediaWiki has been adopted by hundreds of disparate communities for a variety of purposes such as science (e.g., organizing genomic knowledge), engineering (e.g., coding software), and hobbies (e.g., organizing gardening tips). A notable semantic wiki is Wikidata[6], a project by the Wikimedia Foundation to build a comprehensive multilingual collection of facts that would complement their Wikipedia effort. Wikidata is built with Semantic MediaWiki, which extends the MediaWiki platform used by Wikipedia.

Shortipedia is a semantic wiki designed to collect structured knowledge about objects [Vrandečić et al 2011]. It is based on Semantic MediaWiki, and extends it to allow users to add new properties and values together with their provenance. Figure 3 illustrates its user interface. On the top left, a page for Lake Baikal is shown, including properties such as its area of catchment, elevation, inflow and outflow, volume, and islands. Users can add new properties, together with the sources that support them. When the user adds a property, the system uses a command completion search to find existing properties that match what the user is typing. This encourages reuse and normalization of properties across contributors. Another feature of Shortipedia is that it allows contributors to state alternative values for a property. For example the area of catchment is different in the Russian and the English Wikipedia pages for Lake Baikal, so users can add both values with their respective sources. Shortipedia also enables users to add multilingual labels that allow the system to map assertions in different languages, shown on the top right in the figure. Shortipedia also allows users to easily include other known assertions on the Web of Data, by automatically retrieving them and allowing the user to select them as shown on the middle right of the figure. Users are also shown the original Wikipedia page for reference, and the properties that are contained in Wikidata so they can be included as well as shown in the bottom of the figure.

---

[5] http://www.w3.org/standards/semanticweb/data

[6] http://www.wikidata.org

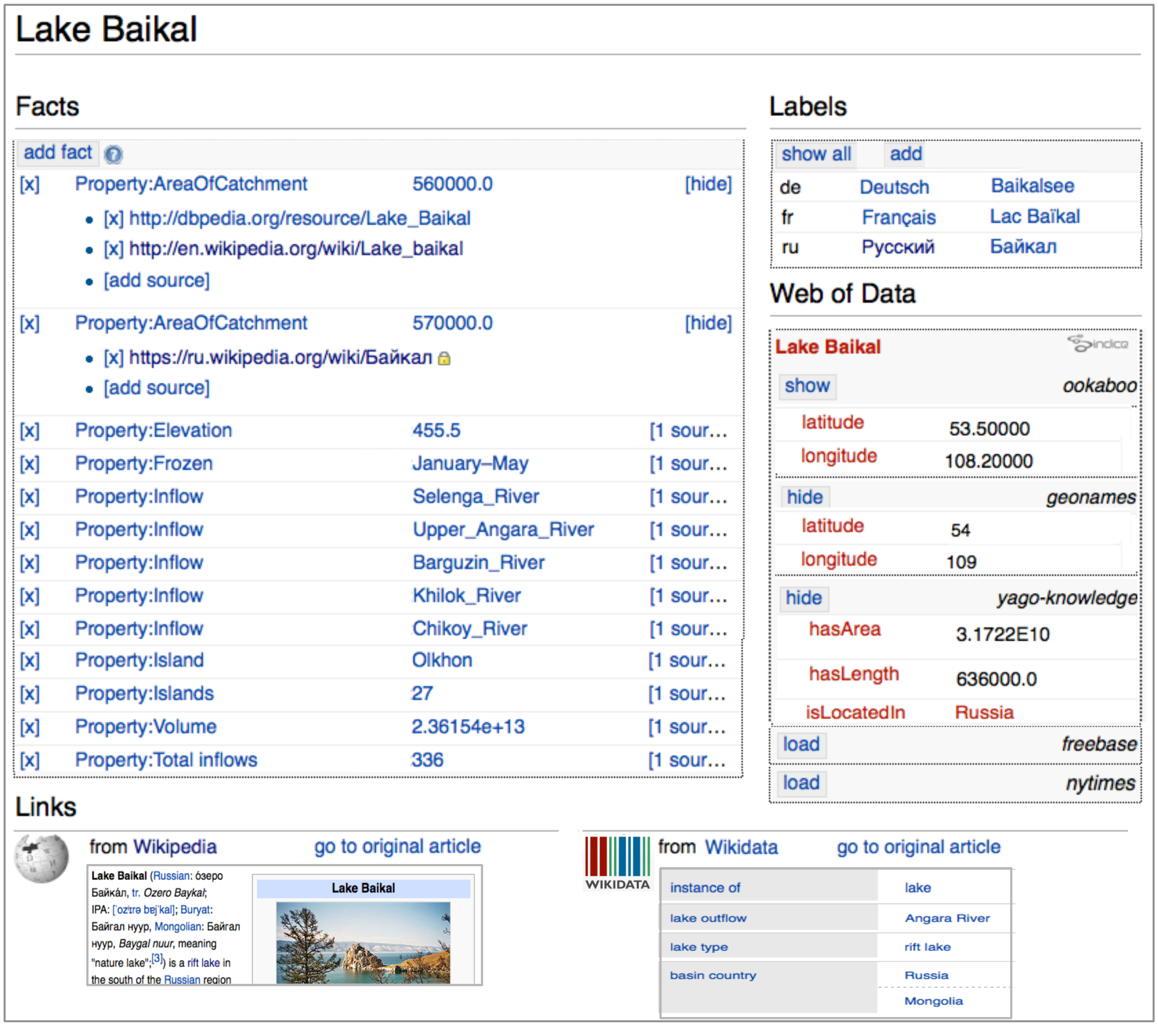


Figure 3. Shortipedia was designed to collect structured knowledge about objects. On the top left, a page for Lake Baikal is shown, including properties such as its area of catchment, elevation, inflow and outflow. Note that each assertion is annotated with sources that support it. The figure shows that the area of catchment is different in the Russian and the English Wikipedia pages for the lake. On the top right, multilingual labels are shown. On the middle right, other known assertions on the Web are retrieved and shown to the user. Here, the latitude and longitude are different depending on the source. At the bottom, the original Wikipedia page is shown for reference, as well as the properties that appear in Wikidata.

In order to understand how the semantic aspects of the wiki are used to structure the contributions, we carried out an analysis of more than two hundred semantic wikis [Gil and Ratanakar 2013]. The analysis showed the concepts and properties created in each wiki, and the amount of editors involved in creating them compared to the total amount of editors of the wiki. We found that concepts are not defined very often. In contrast, properties are very widely used. Large numbers of property assertions are used in almost every wiki. We also found that very small numbers of users edit properties. An important challenge is to understand the limited use of some semantic features of the wiki, such as concept definitions, as well as why there are relatively small amounts of users who create any definitions. One hypothesis is that this is due to the lack of support to the contributors in coordinating semantic edits, although further research is needed to understand this. In addition, semantic wiki communities might benefit from additional capabilities that make the system more proactive in making suggestions to contributors regarding the creation of new concepts, encouraging the reuse of properties created by other contributors, and resolving inconsistencies and missing knowledge.

## 4. Collaborative Ontology Development

For many years, ontology editors were used only by knowledge engineers, enabling them to create sophisticated ontologies of classes and properties either individually or in small well-orchestrated teams. Recently, ontology editors have been augmented to support the collaborative development of ontologies with contributors lacking prior training or prior knowledge about which specific areas each might be able to contribute to. Collaborative ontology development requires a framework that solicits and organizes contributions from people who might have different expertise and different views on the subject matter.

Collaborative Protégé is a framework for collaborative ontology development based on the widely used Protégé ontology editor [Tudorache et al 2011]. It has been used to develop biomedical ontologies of thousands of terms with dozens of contributors [Tudorache and Musen 2011], including the International Classification of Disease revision 11 (ICD-11) and the National Cancer Institute's Thesaurus (NCI Thesaurus). The system enables users to add new subclasses and properties, but it also allows them to override specific contributions made by others and post notes explaining disagreements that need to be discussed.

Understanding the processes or workflows that arise from different ontology editing patterns is helpful for developing new techniques that can support common patterns. For example, a recent analysis found a strong correlation between the amount of changes that a given contributor makes and the amount of notes that the contributor posts [Strohmaier et al 2013]. To provide a global view of the status of the ontology, visualization tools enable monitoring progress over time, expose areas of major disagreements, and measure the quality of the contributions [Walk et al 2013]. This exposes the breadth of expertise of specific contributors, and the most heavily edited areas of the ontology over time.

Further research is needed for supporting different editing patterns, different contributor skills, and managing the dynamic evolution of the ontology and its user community over time.

## 5. Research Challenges in Social Knowledge Collection

Social knowledge collection approaches have been demonstrated to create useful repositories of knowledge for a variety of purposes. However, further research is needed in designing systems that take a more active role in guiding the acquisition process, manage the knowledge collected, and coordinate contributions from different users. Research challenges in social knowledge collection include:

- **User interface design**: How can people detect errors and misconceptions in the system and fix them? How can contributors enter knowledge with minimal burden or prior training?
- **User feedback and prompting**: How can the system generate follow up questions that complement knowledge that users contribute on their own accord? How can users be assigned follow up questions based on their demonstrated expertise?
- **Coordination among contributors**: What are the most effective editorial processes to organize contributors? How can systems learn from several people who are providing overlapping and perhaps incompatible or even contradictory information?
- **Incentives**: What are successful ways to reach and recruit potential contributors to maintain a reasonable community over time? What are the right incentives and rewards to retain contributors?
- **Provenance**: How can users document the knowledge they enter so that the system can justify the sources of its knowledge to other users and be trusted?
- **Quality of the knowledge**: What mechanisms can be used to validate contributions?
- **Purpose**: What kinds of knowledge can we collect effectively through crowdsourcing approaches? What are appropriate knowledge acquisition tasks that contributors can handle?
- **Nature of knowledge collected**: What kinds of knowledge can be collected through volunteer contributors? What are appropriate uses of the knowledge collected? What knowledge formalism is adequate for a given use and kind of knowledge targeted?
- **Managing updates over time**: What are appropriate mechanisms to manage updates and changes, particularly when other systems may have been designed to use the knowledge being collected?
- **Combining interactive and automatic extraction**: How can we combine volunteer contributions with automatic extraction of knowledge from text? Can volunteers validate and extend knowledge automatically extracted that with varying accuracy?

Some of these issues have been studied in social content collection frameworks, notably Wikipedia [Adler and de Alfaro 2007; Almeida et al 2007; Benson et al 2010; Erickson 2008; Hoffmann et al 2009; Hsieh et al 2010; Kittur et al 2008; Kittur and Kraut 2008; Kittur et al 2009; Kittur and Kraut 2010; Lam et al 2010; Leskovec et al 2010; Panciera et al 2010; Raban et al 2010; Spinellis and Louridas 2008]. However, the applicability of these results for social knowledge collection should be carefully considered. In addition, social knowledge collection presents its own set of challenges that need to be addressed.

We foresee in the not too distant future that knowledge repositories created through social knowledge collection could be interlinked through semantic web infrastructure, enabling knowledge sharing across communities of contributors. For example, a repository of genomics knowledge and a repository of biodiversity knowledge could be interconnected to relate genomic information to specific species. The provenance of knowledge sources will be crucial to propagate updates throughout the knowledge bases and to assess trust and resolve conflicting views.


### Acknowledgments

We gratefully acknowledge support from the National Science Foundation with grant IIS-1117281. We also thank Kshanti Greene and Pietro Michelucci for their comments on earlier drafts of this chapter.


## References


[Adler et al 2007] B. Thomas Adler, Luca de Alfaro. 2007. "A content-driven reputation system for the Wikipedia." Proceedings of the 16th International Conference on the World Wide Web (WWW).

[Almeida et al 2007] Rodrigo Almeida, Barzan Mozafari and Junghoo Cho. 2007. "On the Evolution of Wikipedia." Proceedings of the International Conference on Weblogs and Social Media (ICWSM).

[Auer et al 2006] Soren Auer, Sebestian Dietzold, and Thomas Riechert. 2006. "OntoWiki - A Tool for Social, Semantic Collaboration." Proceedings of the 5th International Semantic Web Conference (ISWC).

[Auer et al 2007] Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, Zachary Ives. 2007. "DBpedia: a nucleus for a web of open data." Proceedings of the 6th International Semantic Web Conference (ISWC).

[Barrett 2008] Daniel J. Barrett. 2008. "MediaWiki," O'Reilly Media Inc.

[Benson et al 2010] Benson, Edward, Marcus, Adam, Howahl, Fabian, and Karger, David. 2010. "Talking about data: sharing richly structured information through blogs and wikis." Proceedings of the 19th International Conference on the World Wide Web (WWW).

[Brachman and Levesque 2004] Ronald J. Brachman, Hector J. Levesque. 2004. Knowledge Representation and Reasoning. Elsevier.

[Brickley and Guha 2004] Dan Brickley and R.V. Guha. 2004. "RDF Vocabulary Description Language 1.0: RDF Schema." World Wide Web Consortium. Available from http://www.w3.org/TR/rdf-schema.

[Bry et al 2012] François Bry, Sebastian Schaffert, Denny Vrandečić and Klara Weiand. 2012. "Semantic Wikis: Approaches, Applications, and Perspectives." Lecture Notes in Computer Science, Reasoning Web. Semantic Technologies for Advanced Query Answering, Volume 7487.

[Chklovski 2003a] Chklovski, T. 2003 "Using Analogy to Acquire Commonsense Knowledge from Human Contributors." PhD thesis, Department of Computer Science. MIT Artificial Intelligence Lab technical report AITR-2003-002. Massachusetts Institute of Technology.

[Chklovski 2003b] Chklovski, T. 2003. "LEARNER: A System for Acquiring Commonsense Knowledge by Analogy." Proceedings of Second International Conference on Knowledge Capture (K-CAP).

[Chklovski 2005] Chklovski, T. 2005. Designing Interfaces for Guided Collection of Knowledge about Everyday Objects from Volunteers. Proceedings of the ACM International Conference on Intelligent User Interfaces (IUI).

[Chklovski and Gil 2005a] Chklovski, T., and Gil, Y. 2005. "An Analysis of Knowledge Collected from Volunteer Contributors." Proceedings of the Twentieth National Conference on Artificial Intelligence (AAAI).

[Chklovski and Gil 2005b] Chklovski, T., and Gil, Y. 2005. "Towards Managing Knowledge Collection from Volunteer Contributors." Proceedings of the AAAI Spring Symposium on Knowledge Collection from Volunteer Contributors (KCVC).

[Chklovski and Gil 2005c] Chklovski, T., and Gil, Y. 2005. "Improving the design of intelligent acquisition interfaces for collecting world knowledge from web contributors." Proceedings of the Third International Conference on Knowledge Capture (K-CAP).

[Erickson 2008] Thomas Erickson. 2008. "'Social systems': designing digital systems that support social intelligence." Journal AI & Society - Special Issue: Social intelligence design: a junction between engineering and social sciences, 23(2).

[Gil 2011] Yolanda Gil. 2011. "Interactive Knowledge Capture in the New Millenium: How the Semantic Web Changed Everything." Knowledge Engineering Review, 26(1).

[Gil et al 2012] Gil, Y.; Ratnakar, V.; Chklovski, T.; Groth, P.; and Vrandecic, D. 2012. "Capturing Common Knowledge about Tasks: Intelligent Assistance for To Do Lists." ACM Transactions on Interactive Intelligent Systems, 2(3).

[Gil and Ratnakar 2013] Yolanda Gil and Varun Ratnakar. 2013. "Knowledge Capture in the Wild: A Perspective from Semantic Wiki Communities." Proceedings of the International Conference on Knowledge Capture (K-CAP).

[Gupta and Kochenderfer 2004] Gupta, R., and Kochenderfer, M. 2004. "Common sense data acquisition for indoor mobile robots." Proceedings of the Nineteenth National Conference on Artificial Intelligence (AAAI).

[Havasi et al 2007] Catherine Havasi, Rob Speer and Jason Alonso. 2007. "ConceptNet 3: a Flexible, Multilingual Semantic Network for Common Sense Knowledge." Proceedings of the Conference on Recent Advances in Natural Language Processing (RANLP).

[Heath and Bizer 2011] Tom Heath and Christian Bizer. 2011. "Linked Data: Evolving the Web into a Global Data Space." Morgan and Claypool Publishers, Synthesis Lectures on the Semantic Web.

[Hoffmann et al 2009] Raphael Hoffmann, Saleema Amershi, Kayur Patel, Fei Wu, James Fogarty, Daniel S. Weld. 2009. "Amplifying community content creation with mixed initiative information extraction." Proceedings of the 27th International Conference on Human Factors in Computing Systems (CHI).

[Hsieh et al 2010] Gary Hsieh, Robert E. Kraut, Scott E. Hudson. 2010. "Why pay?: exploring how financial incentives are used for question & answer." Proceedings of the 28th International Conference on Human Factors in Computing Systems (CHI).

[Kittur et al 2008] Aniket Kittur, Bongwon Suh, Ed H. Chi. 2008. "Can you ever trust a wiki? Impacting perceived trustworthiness in Wikipedia." Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW).

[Kittur and Kraut 2008] Aniket Kittur, Robert E. Kraut. 2008. "Harnessing the wisdom of crowds in Wikipedia: Quality through coordination." Proceedings of the ACM conference on Computer Supported Cooperative Work (CSCW).

[Kittur et al 2009] Aniket Kittur, Bryant Lee, Robert E. Kraut. 2009. "Coordination in collective intelligence: the role of team structure and task interdependence." Proceedings of the 27th International Conference on Human Factors in Computing Systems (CHI).

[Kittur and Kraut 2010] Aniket Kittur, Robert E. Kraut. 2010. "Beyond Wikipedia: coordination and conflict in online production groups." Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW).

[Krötzsch et al 2007] Markus Krötzsch, Denny Vrandecic, Max Völkel, Heiko Haller, Rudi Studer. 2007. "Semantic Wikipedia." Journal of Web Semantics, 5(4).

[Kuhn 2009] Tobias Kuhn. 2009. "AceWiki: A Natural and Expressive Semantic Wiki." Proceedings of the Fifth International Workshop on Semantic Web User Interaction (SWUI), CEUR Workshop Proceedings, Volume 543.

[Lam et al 2010] Shyong (Tony) K. Lam, Jawed Karim, John Riedl. 2010. "The effects of group composition on decision quality in a social production community." Proceedings of the 16th ACM International Conference on Supporting Group Work (GROUP).

[Lenat and Guha 1990] Lenat, D. B., and R. V. Guha. 1990. Building Large Knowledge-Based Systems: Representation and Inference in the Cyc Project. Addison Wesley.

[Leskovec et al 2010] J. Leskovec, D. Huttenlocher, J. Kleinberg. 2010. "Governance in Social Media: A case study of the Wikipedia promotion process." Proceedings of the AAAI International Conference on Weblogs and Social Media (ICWSM).

[Lieberman et al 2007] Henry Lieberman, Dustin Smith and Alea Teeters. 2007. "Common Consensus: A Web-based Game for Collecting Commonsense Goals." Intelligent User Interfaces (IUI-07), Honolulu, HI.

[Lieberman et al 2004] Henry Lieberman, Hugo Liu, Push Singh, and Barbara Barry. 2004. "Beating some common sense into interactive applications." AI Magazine, 25(4).

[Matuszek et al 2005] Cynthia Matuszek, Michael J. Witbrock, Robert C. Kahlert, John Cabral, David Schneider, Purvesh Shah, Douglas B. Lenat. 2005. "Searching for Common Sense: Populating Cyc from the Web." Proceedings of the 20th National Conference on Artificial Intelligence (AAAI).

[Panciera et al 2010] Panciera, K.; Priedhorsky, R.; Erickson, T.; Terveen, L. 2010. "Lurking? Cyclopaths? A Quantitative Lifecyle Analysis of User Behavior in a Geowiki." Proceedings of the 28th International Conference on Human Factors in Computing Systems (CHI).

[Raban et al 2010] Daphne R. Raban, Mihai Moldovan, Quentin Jones. 2010. "An empirical study of critical mass and online community survival." Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW).

[Smith and Lieberman 2010] Dustin A. Smith and Henry Lieberman. 2010. "The Why UI: Using Goal Networks to Improve User Interfaces." Proceedings of the International Conference on Intelligent User Interfaces (IUI), Hong Kong, China.

[Speer et al 2008] Robert Speer, Catherine Havasi, and Henry Lieberman. 2008. "AnalogySpace: Reducing the Dimensionality of Common Sense Knowledge." Proceedings of the Conference of the Association for the Advancement of Artificial Intelligence (AAAI).

[Spinellis and Louridas 2008] Diomidis Spinellis and Panagiotis Louridas. 2008. "The Collaborative Organization of Knowledge." Communications of the ACM, 51(8).

[Strohmaier et al 2013] M. Strohmaier, S. Walk, J. Poeschko, D. Lamprecht, T. Tudorache, C. Nyulas, M. Musen, and N.F. Noy. 2013. "How ontologies are made: Evaluation of the hidden social dynamics behind collaborative ontology engineering projects." Journal of Web Semantics, 20.

[Tudorache and Musen 2011] T. Tudorache, M. A. Musen. 2011. "Collaborative Development of Large-Scale Biomedical Ontologies." In Collaborative Computational Technologies for Biomedical Research, S. Ekins, M. A. Z. Hupcey and A. J. Williams, John Wiley & Sons, Inc., Hoboken, NJ.

[Tudorache et al 2011] T. Tudorache, C. I. Nyulas, M. A. Musen, N. F. Noy. 2011. "WebProtégé: A Collaborative Ontology Editor and Knowledge Acquisition Tool for the Web." Semantic Web Journal, 11-165.

[Vrandečić et al 2011] Denny Vrandečić, Varun Ratnakar, Markus Krötzsch, Yolanda Gil. 2011. "Shortipedia: Aggregating and Curating Semantic Web Data." Journal of Web Semantics, 9(3).

[Walk et al 2013] S. Walk, J. Pöschko, M. Strohmaier, K. Andrews, T. Tudorache, N. F. Noy, C. I. Nyulas, M. A. Musen. 2013. "PragmatiX: An Interactive Tool for Visualizing the Creation Process Behind Collaboratively Engineered Ontologies." International Journal on Semantic Web and Information Systems, IJSWIS, Special Issue on Visualization of and Interaction with Semantic Web Data, 2013.

[Weld et al 2008] Daniel S. Weld, Fei Wu, Eytan Adar, Saleema Amershi, James Fogarty, Raphael Hoffmann, Kayur Patel, Michael Skinner. 2008. "Intelligence in Wikipedia." Proceedings of the 23rd Conference of the Association for the Advancement of Artificial Intelligence (AAAI).