Voluntary Contributions of Unaware Internet Users ? On Automatic Knowledge Retrieval from the WWW

Marcin Skowron and Kenji Araki
Graduate School of Information Science and Technology, Hokkaido University
Kita-ku, Kita 14, Nishi 9
Sapporo, Japan 060–0814
{ms,araki}@media.eng.hokudai.ac.jp

Abstract

This paper presents the method of automatic knowledge retrieval from the Web. In this approach, millions of Internet users who naturally and constantly create and update the content of the Web are treated as the "voluntary contributors" who provide knowledge concepts without any need for additional motivation or coordination of their efforts. Obviously, only a small part of the statements accessible on the Web are valid knowledge axioms. Below, the "web knowledge concepts" filtering and verification methods are described, based on the similarity measurements with the concepts found in the manually created knowledge database. The results obtained with the system demonstrated that it is able to automatically retrieve semantic equivalents of several concepts submitted by the volunteer contributors, as well as the concepts that while being valid entries to a knowledge database, provide more details compared to the ones found in the manually developed database.

Introduction

Despite the years of research in the field of Artificial Intelligence, the creation of a machine with the ability to think is still far from realization. Although computer systems are capable of performing several complicated tasks that require human beings to extensively use their thinking capabilities, machines still cannot understand what people talk about and why they get angry or happy. One of the main unresolved problems is the lack of machine usable knowledge. Without it, machines cannot reason about the everyday world in a similar way to human beings. This slows down spread of AI systems.

Although there is no agreement on the precise amount of knowledge the average human adult has, researchers usually estimate it to include at least several million axioms. One of the most convincing studies in this respect measured the amount of information a human being was capable of remembering and retrieving after a long time. This information was acquired across three different channels: verbal, visual, and musical (Landauer 1986). The study demonstrated that a human being is able to absorb and retain information for a long time - around two bits per second. This implies a rate of 8 megabytes or a million short statements per year. Based on this, one can estimate the total amount of knowledge an adult human being possess to be several million axioms. Similarly, in work of Lenat (Lenat, Guha, Pittman, Pratt and Shepherd 1990) a study on the amount of knowledge necessary for human beings for commonsense reasoning was estimated to include approximately 10^8 axioms.

In the last decade we have witnessed a few attempts to create such knowledge databases using various approaches: manual, machine learning and mass collaboration of volunteer contributors. So far, none of these approaches can be considered as fully reliable and/or efficient. To a various extent, each of the presented methods has some problem areas and limitations that are inherent to them. We discuss this issue in more detail in the next section.

Compared to the efforts taken to input knowledge concepts manually, either as an effort of one entity or distributed among several parties, surprisingly little has been done to develop the methods for the automatic retrieval of knowledge concepts for such databases. This paper presents the idea of a language independent method, capable to automatically retrieve general and commonsensical knowledge for the open-domain applications from the Web. In this approach millions of Internet users who naturally and constantly create and update the content of the Web can be treated as "voluntary contributors", who provide knowledge concepts without any need for additional motivation or coordination for their efforts. Obviously, only a small part of the "submitted" entries are valid knowledge axioms. Below, we describe the methods for "web knowledge concepts" filtering and verification, which are based on the similarity comparison with concepts found in a manually created knowledge database. The paper also presents the implementation of these methods to the developed system and preliminary results obtained from it.

Knowledge for Machines

The most well known example of the manual approach to knowledge base construction is the CYC project, which contains 1.5 million assertions build over 15 years (Lenat 1995). The aim of the project was to create a database that could provide knowledge from a large scope of domains, along with the means to effectively use this knowledge for systems that engaged in the reasoning of human affairs. We learn

from the experiences of this and similar projects that building a database in this way was laborious, time-consuming and costly. We argue that in an open domain where new information appears and becomes obsolete on a daily basis, a complete knowledge base build using this approach is out of reach.

The machine learning approach showed that it was feasible to automatically induce rules from data and to overcome some problems characteristic for the manual approach presented above. However, as of yet the machine learning approach has not resulted in creation of a large, open-domain knowledge base. As explained in (Richardson and Domingos 2003), a typical learning program has only weak assumptions about the world; consequently, the learned rules are relatively shallow as they refer only to correlations between observable variables. To address this problem the researchers attempted to incorporate pre-existing knowledge, combining the automatic and manual approaches. However, they soon faced the bottleneck related to collecting knowledge known from the manual approaches.

In recent years researches have demonstrated that it is possible to populate knowledge databases directly using normal means of communication, such as written natural language (Katz 1990), and that the items imputed in this way can be used for reasoning (Borchardt 1992) (Singh 2002). The advent of the Internet made the real-time communication and collaboration of the thousands of people working together on one project possible. An example of such undertakings include Wikipedia - the free Internet encyclopedia containing one million articles in one hundred languages[1], as well as thousands of open-source programs and systems written, bug-fixed, and improved by developers and users dispersed across the Web. When properly designed and coordinated, an approach based on massive collaboration using the Internet allows knowledge databases as large as the one created in the CYC project to be built. This can be achieved much faster and at a fraction of the cost. The challenges frequently faced in this approach include the need to ensure (Richardson and Domingos 2003):

- High quality of contributed knowledge,
- Consistency between knowledge entered by different contributors and at different times,
- Relevance of inputted knowledge to a given task,
- Scalability of the project,
- Motivation of contributors to start and consistently work on the project.

Knowledge acquisition from Internet users is the approach used in the construction of the Open Mind Common Sense database (OMCS), allowing mass collaborations of thousands of volunteer contributors (Singh 2002). The ability to input knowledge concepts directly in the natural-language form simplified the process of database building compared to the attempts that used semi-programming languages. Thanks to this, theoretically every (English speaking) Internet user, without any special training can input knowledge concepts using plain sentences. While significantly decreasing the amount of money spent, this approach still requires large investment of human labor distributed among thousands of Internet users. Compared to the manual approaches, the time requirement for creation of the knowledge database is reduced; but, this factor still cannot be overlooked. So far the OMCS project has gathered more than 700,000 items from more than 15,000 users. The OMCS database was evaluated (Singh 2002) by human judges using a sample of the knowledge concepts. After elimination of 12,3% of the items, which were marked by the judges as being garbage, the following conclusions were presented for the remaining set:

- 75% of the items are largely true,
- 84% is knowledge someone would have by high school,
- 82% are largely objective,
- 85% were judged as largely making sense.

Tables 1 and 2 show the examples of knowledge concepts related to the selected nouns “apple” and “water” from the OMCS database.

Table 1: Examples of Knowledge Concepts Related to “apple” from the OMCS Database.

No.	Knowledge Concept
1	Yellow apples are soft and sweet
2	The first thing you do when you eat an apple is rub it
3	An apple contains seeds
4	apples are edible
5	The Michigan state flower is the apple blossom
6	When you drop an apple, it gets bruised
7	You are likely to find an apple tree in an orchard
8	apple juice is delicious

Table 2: Examples of Knowledge Concepts Related to “water” from the OMCS Database.

No.	Knowledge Concept
1	The last thing you do when you take a shower is turn off the water
2	Clean air, clean water, organic food, and proper sanitation are vital to good health
3	Human beings need water to survive
4	When animals need water, they feel thirsty
5	Clouds are made up of water vapor
6	lakes are usually filled with fresh water
7	People need to drink water verey day
8	water would make you want to have a pee

Automatic Knowledge Retrieval from the Web

Analyzing the content of manually created knowledge databases, one can discover that several of the statements found there exist also on freely available Internet pages. Additionally, the number of “web knowledge concepts” while using slightly different words and/or syntax, semantically

[1] http://en.wikipedia.org/wiki/Wikipedia, as of 25.12.2004

provides equivalents for a large part of the entries from the manually constructed knowledge bases.

We argue that the Web is a rich resource of commonsensical and general knowledge and that this resource is usable in the process of automatic creation of knowledge databases. For automatic knowledge retrieval the Web has important advantages, including real-time updates of content, wide coverage of various domains, and diversity of presented opinions. At present, a popular search engine indexes 8,058,044,651 web pages[2]; this is just a fraction of the total number of pages available in the WWW. Assuming only a small portion of them include statements that can be treated as valid entries to a knowledge database, the Web still hosts an immense number of knowledge concepts that can be automatically retrieved.

The estimated number of Internet users in 2004 was 937 million[3], with projected number to exceed 1.2 billion in 2006. A fraction of the users are active creators of the Web content including writing and publishing personal home pages, writing blog entries, submitting comments under the articles, and participating in mailing-lists discussions. Additionally, WWW hosts millions of pages displaying online versions of newspapers, magazines, technical reports, manuals, books, FAQ lists, encyclopedias etc. We think that in every moment, in various part of the world Internet users/WWW creators contribute the knowledge that can and ought to be used for the building of the knowledge databases and supporting several AI applications. Obviously, there is a need to filter out statements that cannot be considered as reliable and valid entries to a knowledge database. The main challenge in this approach is to ensure a high recall rate of knowledge concepts from various domains and precision of concepts filtering and selection.

Relevant Research

Some relevant research adapting different approaches, and to some extent having different aims, is included in the following works. Studies on knowledge representation and classification were done by Woods, followed by work on automatic construction of taxonomies by extracting concepts from texts (Woods 1997). Garcia (Garcia 1997) use verbs as casual indicators for acquiring commonsensical knowledge in French. Satoh (Satoh 1998) use connective markers to acquire casual knowledge similarly to the later work of Inui (Inui, Inui and Matsumoto 2003), where the method for classification of casual relations was presented. The research of Rzepka (Rzepka, Itoh and Araki 2004) described the methods for automatic word categorization and script generation using the statistic of words obtained from a Web-sized corpus and by exploring Japanese grammar features.

Substantial work in information extraction aiming at automatic generation of the large collection of facts from the Web using extraction patterns and relations has been described by Brin (Brin 1998), Agichtein (Agichtein and Gravano 2000) and Etzioni (Etzioni, Cafarella, Downey, Kok, Popescu, Shaked, Soderland, Weld and Yates 2004).

[2]www.google.com, as of 25.12.2004

[3]Computer Industry Almanac, http://www.c-i-a.com/

The work of Moldovan (Moldovan 2000), described a system that gathers knowledge from a financial domain. However, this system was tailored to a specific domain and its processing could not be done automatically, as the system required an interaction with a user to verify the acquired concepts, relationships and extraction patterns.

System Processing

Below we present the "KnowY" system that implements the idea of automatic retrieval of open-domain knowledge concepts from the Web. There are two aims of this system: to automatically create a knowledge database similar to ones that are being built manually through the mass collaboration of Internet users, and to support various AI applications with the knowledge required in their processing. At present, the system utilizes the OMCS database to learn what constitutes a valid entry to a knowledge database. This information is necessary for filtering out statements that are unlikely to form such an entry and to rank discovered concepts depending on their similarity to ones found in the OMCS database. The adaptation to any other knowledge database written in natural language form is also feasible.

"KnowY" is capable of working both in offline and online mode. In offline mode, the system automatically creates entries to a knowledge database from a submitted list of nouns. The online mode is designed to serve as a helper application for the dialog system or any other application that could benefit from access to knowledge concepts. In this mode "KnowY" uses additional features not present in the offline mode, which ensure faster response time by retrieving only that knowledge which is required by the application in a given situation and related to a discussed domain. In order to present these additional features we decided to describe system processing when used in the online mode. Figure 1 presents the system flowchart.

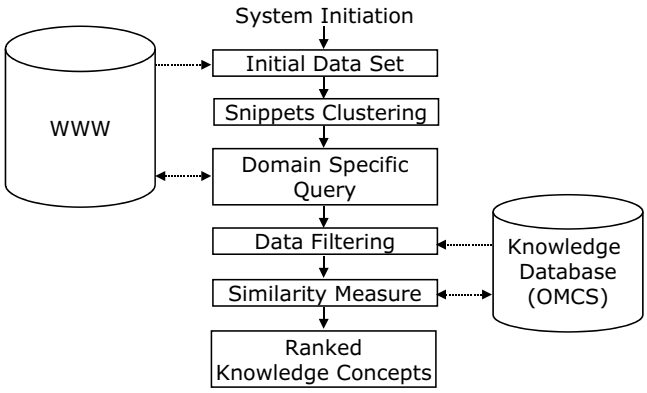

Figure 1: System Flowchart

Domain Recognition

"KnowY" is initiated by receiving a request for a noun or list of nouns for which knowledge concepts are to be discovered. Since, nouns frequently have several distinct senses and are

related to different domains, it is often difficult to decide beforehand which sense/domain is the most suitable in a given situation. If this information cannot be acquired from a user utterance to the dialog system, "KnowY" obtains a diverse set of information related to a requested noun, where a noun sense/domain intended by a user could be found. To provide such a set and to ensure the obtained set of snippets do not include only commercial information, the system retrieves the 300 highest scored snippets for each of the 5 different domains (.gov, .edu, .net, .org, and .com) using a submitted noun as a query. Next, a set of the unique snippets is clustered (Karypis 2002), and for each cluster the 3 most descriptive words are discovered. Initially the number of clusters is set to 4. For example for the noun "apple" the following most descriptive words characterizing the content of the clusters have been found:

1. tree, fruit, apples
2. juice, production, cider
3. computer, mac, macintosh
4. iTunes, music, server

Using a query composed of the most descriptive words for each cluster, and the requested noun, "KnowY" searches the Web and counts the frequency of the statements that consist of the following patterns: "noun[s] $\langle BE \rangle$ $\langle a/the/\ / \rangle$ $\langle descriptive\ word \rangle$". The most frequent statements are used to form questions, which are then returned to the dialog system to find the noun sense/domain intended by a user. If a user does not find the one he/she is interested in, the clustering of snippets is performed again, with the number of clusters increased by one. This procedure is repeated until the requested domain/noun sense is found.

Domain Specific Data Retrieval

Once the intended domain is obtained, "KnowY" executes the new search using a query consisting of the requested noun along with the 3 most descriptive words discovered for a selected cluster. This time the system retrieves a set of 300 web pages, which provides a wide selection of information concerning the noun. Extension of a used query with the 3 most descriptive words for a selected cluster ensures that domain limitations are respected. After dividing the text into sentences, the ones that do not include the requested noun (either in singular or plural form) are excluded. For example, for the noun "apple" within the domain (tree, fruit, apples), 4,140 sentences were obtained, with 2,751 unique ones. The concurrence of the sentences is used later in the system processing to assign the commonness rate to the knowledge concepts.

Knowledge Concepts Filtering and Selection

In the data filtering and similarity measure stages, the OMCS database was used as training data that revealed the most frequent ways of describing knowledge concepts, as well as commonly used words and grammatical constructions[4]. To ensure reasonable processing time the system uses only a set of concepts from the OMCS database. This set is composed of the concepts that include the requested noun (either in singular or plural form). If the number of such concepts is less than 100, the system randomly adds additional concepts to obtain such a set. As our preliminary experiments show, for a well defined concept (high quality and wide coverage of a given noun in the OMCS in the database), the former method is likely to find a slightly better set of automatically retrieved "web knowledge concepts". On the other hand, the later approach provides the means to discover new concepts, and allows "KnowY" to generate a database for terms that are not covered by the OMCS at all. We think that this feature is of prime importance for many applications in the open domain. For the similarity score ranking we decided to include only "web sentences" and OMCS concepts with the number of words between 3 and 20. In the filtering process "KnowY" also uses information on the proportion of alphanumeric and special characters. To exclude the sentences that are unlikely to be valid entries to the knowledge database, "KnowY" compares them to the OMCS knowledge database concepts and ranks using the highest similarity score obtained, calculated with the following formula:

$$Similarity = \frac{\sum_{n=1}^{N} Wn * Wn_{IDF}}{L1 + L2} + \log Np * \alpha \quad (1)$$

,where: Wn - matching word found both in a OMCS concept and a "web sentence", Wn_{IDF} - inverted document frequency for a matching word (OMCS), L1 - number of words in a "web sentence", L2 - number of words in a OMCS concept, Np - number of concepts from OMCS where a noun was found in a position from a "web sentence", α - parameter (the value of α was set to 0.2). The formula takes account of the number of matching words between a "web sentence" and an OMCS concept as well as the importance of the given word in a used set of OMCS terms by the means of the IDF. The Np value promotes "web sentences", where a noun appears on a position, which is frequent for many OMCS concepts.

Experiment Results

The preliminary experiments with the system were performed using a set of nouns, including ones that are frequently described in OMCS, as well as ones that are not covered at all in this database. Tables 3 and 4 show the examples of the knowledge concepts automatically discovered from the Web; for each noun, only the first eight highest ranked statements are presented.

As the results demonstrate, "KnowY" is able to automatically retrieve the knowledge concepts for a domain selected by a user with relatively high accuracy. Furthermore, the experiments proofed that the system is capable of finding and automatically retrieving from the Web the semantic equivalents of several of the entries that were inputted manually to

[4]In our preliminary experiments we used the snapshot of the OMCS database consisting of 700,000 sentences, available at http://commonsense.media.mit.edu/cgi-bin/download.cgi. From this set the sentences describing pictures and stories were removed.

Table 3: Highest Ranked Knowledge Concepts Related to "apple", Domain (tree, fruit, apples).

No.	Discovered Knowledge Concept	Sim. Score
1	Apples bruise easily and must be hand picked.	1.61
2	Apple blossom is the state flower of Michigan.	1.58
3	Provide small samples different types of apples.	1.28
4	Apples have 5 seeds.	1.22
5	Apples are a member of the rose family.	1.21
6	Some apple varieties such Red Delicious and Golden Delicious are also available in various strains.	1.15
7	Apples that don't have enough pollinated seeds inside usually drop from the tree before ripening.	1.13
8	Apples are soft, medium size and easily bruised.	1.11

Table 4: Highest Ranked Knowledge Concepts Related to "water", Domain (drinking, epa, treatment).

No.	Discovered Knowledge Concept	Sim. Score
1	Sanitation means not only clean water but also clean air and clean soil.	1.33
2	Clouds are made when water vapor condenses into tiny droplets.	1.29
3	Most lakes are filled with fresh water, but there are a few lakes that are filled with salt.	1.25
4	When animals need water, they should not have to stand and wait.	1.17
5	Overwatering your yard can also cause water to run into the streets and into storm drains.	1.12
6	Safe, clean drinking water is what we expect when we turn on our faucets.	1.09
7	1 pound of water requires 1 BTU to raise its temperature 1 degree farenheit.	1.08
8	A person can live several days without food, but just a few days without water.	1.04

the OMCS database. An example of such a statement discovered for the noun "apple", which is present among first eight concepts with the highest similarity score is "Apple blossom is the state flower of Michigan", and its counterpart from the OMCS, "The Michigan state flower is the apple blossom". Some of the knowledge concepts obtained from the Web provide more detail compared to the OMCS entries. The instances of such "web knowledge concepts" include: "Apples have 5 seeds", "Clouds are made when water vapor condenses into tiny droplets", "A person can live several days without food, but just a few days without water" and the corresponding OMCS database entries, "An apple contains seeds", "Clouds are made up of water vapor", and "Human beings need water to survive".

With the exception of the statement "Provide small samples different types of apples" (Table 3, pos. 3) all remaining, automatically retrieved statements can be treated as valid entries to a knowledge database[5].

Tables 5 and 6 present the results obtained for the nouns "php" and "Kendo". These nouns are covered only to a very limited extend in the OMCS ("php" - 4 entries) or do not appear in this database ("Kendo"). For the similarity score calculation the comparison set included 96 and 100 randomly selected statements from the OMCS database, respectively, for the nouns "php" and "Kendo". The average similarity score is considerably lower, compared with the score obtained for the automatically discovered knowledge concept related to "water" and "apple". However, although the comparison set was composed of mostly randomly selected concepts from OMCS, it did not significantly compromise the quality of the discovered concepts. The majority of retrieved statements related to "php" and "Kendo" could be included in a knowledge database. The exceptions are sentences 3 and 7 from Table 5, and 3 and 8 from Table 6. These demonstrate that there is a need to extend the similarity measurements formula with the means to penalize the statements that include words frequently used to describe personal opinions and information of a strictly commercial nature.

[5]In the assessment of the "web knowledge concepts", we adapted the approach used to evaluate OMCS-1 database as described by Singh (Singh, Lim, Lin, Mueller, Perkins, Tompkins and Zhu 2002).

Conclusions and Future Work

This paper presented the method of automatic knowledge retrieval from the Web. The preliminary experiments performed with the system demonstrated that it is capable of automatically discovering several knowledge concepts in a user-selected domain with relatively high accuracy. Some of the automatically retrieved knowledge concepts provided semantic equivalents of the statements that were manually inputted to the OMCS. Others, while including more details compared to the OMCS entries, could also become a part of a knowledge database. The results confirmed that the system was able to retrieve high quality knowledge concepts, even for the terms that were not described in the knowledge database built by mass collaboration of Internet users.

Our future work includes further refining of the presented method and the application of other knowledge databases for creating a comparison set for data filtering and similarity measurements. We are focusing on extending the similarity measurement formula with the means that allow to promote the statements written using a syntax, which is frequently used in manually inputted knowledge concepts and to penalize the statements that include strictly personal opinions or information of only a commercial nature. Our future work includes also a complete evaluation of the proposed method, using the criteria similar to the ones used to evaluate the OMCS-1 database. Given that the described method is lan-

Table 5: Highest Ranked Knowledge Concepts Related to "php", Domain (language,scripting,server).

No.	Discovered Knowledge Concept	Sim. Score
1	PHP is a free server side scripting language.	0.81
2	PHP is short for PHP Hypertext Preprocessor.	0.76
3	I personally however find PHP more than adequate for my Web programming needs.	0.72
4	Also developing Web applications using PHP language required some knowledge of programming concepts.	0.69
5	PHP and ASP are essentially very similar with PHP being free and ASP not.	0.68
6	PHP is best used to build user interfaces and dynamic Web presentations, according to Murphy.	0.68
7	Free Website templates, free php scripts, and free Website content analysis for search engines.	0.65
8	However, function names in PHP are not case-sensitive.	0.65

Table 6: Highest Ranked Knowledge Concepts Related to "Kendo", Domain (sword,art,japanese).

No.	Discovered Knowledge Concept	Sim. Score
1	Anger and true aggression has nothing to do with Kendo.	0.73
2	The bokken is used in modern kendo for kata practice.	0.67
3	This being the modern day Kendo's source of philosophy.	0.65
4	There are eight striking points in Kendo used for scoring.	0.62
5	Shiai geiko is the most competitive part of Kendo.	0.61
6	The competitions rules are defined by the Kendo Federation of Japan.	0.61
7	The primary goal of kendo is to improve oneself through the study of the sword.	0.58
8	Book's, DVD's, and Video's about Kendo and Samurai.	0.59

guage independent, we intend also to apply it to languages other than English.

Acknowledgments

We would like to thank the anonymous reviewers for their valuable suggestions and comments. Marcin Skowron is supported in part by Japanese Ministry of Education, Culture, Sports, Science and Technology.

References

Agichtein E. and Gravano S. Snowball: Extracting Relations from Large Plain-Text Collections. In Proceedings of the 5th ACM International Conference on Digital Libraries, 2000.

Borchard G. Understanding Casual Descriptions of Physical Systems. *In Proceedings of the Tenth National Conference on Artificial Intelligence*, pages 2-8, 1992.

Brin S. Extracting Patterns and Relations from the World Wide Web. *In Proceedings of the 1998 International Workshop on the Web and Databases*, 1998.

Chklovski T. and Pantel P. Large-Scale Extraction of Fine-Grained Semantic Relations between Verbs. *Int. Workshop on Mining for and from the Semantic Web*, pages 12-23, 2004.

Etzioni O., Cafarella M., Downey D., Kok S., Popescu A.-M., Shaked T., Soderland S., Weld D. S. and Yates A. *Web-Scale Information Extraction in KnowItAll (Preliminary Results)*. In Proceedings of the 13th International World Wide Web Conference, pages 100-110, 2004.

Garcia D. COATIS, an NLP system to locate expressions of actions connected by casuality links. *In Proceedings of the 10th European Knowledge Acquisition Workshop*, pages 347-352, 1997.

Inui T., Inui K. and Matsumoto Y. What Kind and Amount of Casual Knowledge Can Be Acquired from Text by Using Connective Markers as Clues? *In the 6th International Conference on Discovery Science*, pages 179-192, 2003.

Katz B. Using English for Indexing and Retrieving. *Artificial Intelligence at MIT: Expanding Frontiers, volume 1*, 1990.

Landauer T. How much do people remember? Some estimates of the quantity of learned information in log-term memory. *Cognitive Science, 10(4)*, pages 477-493, 1986.

Lenat D. CYC: A large-scale investment in knowledge infrastructure. *Communications of the ACM, 38(11)*, pages 33-38, 1995.

Lenat D., Guha K., Pittman K., Pratt D. and Shepherd M. CYC: towards programs with common sense. *Communications of the ACM, 33(8)*, pages 30-49, 1990.

G. Karypis. *A Clustering Toolkit*. http://www.cs.umn.edu/~cluto. 2003.

Moldovan D., Girju R. and Rus V. Domain-Specific Knowledge Acquisition from Text. *Proceedings of the Applied Natural Language Processing Conference*, pages 268-275, 2000.

Richardson M. Domingos P. Building Large Knowledge Bases by Mass Collaboration. *Proceedings of the Second International Conference on Knowledge Capture*, pages 129-137, 2003.

Rzepka R., Itoh T. and Araki K. Rethinking Plans and Scripts Realization in the Age of Web-mining. *IPSJ SIG Technical Report 2004-NL-162*, pages 11-18, 2004.

Satoh H. Retrieval of simplified casual knowledge in text and its applications. *In Proceedings of the IEICE, Thought and Language*, 1998. (In Japanese).

Singh P. The public acquisition of commonsense knowledge. *In Proceedings of AAAI Spring Symposium: Acquiring (and Using) Linguistic (and World) Knowledge for Information Access*, 2002.

Singh P., Lim G., Lin T., Mueller E., Perkins T., Tompkins M. and Zhu W. Open Mind Common Sense: Knowledge Acquisition from the General Public. *Proceedings of the Fifth International Conference on Ontologies, Databases, and Applications of Semantics for Large Scale Information Systems*, 2002.

Woods W. A Better way to Organize Knowledge. *Technical Report of Sun Microsystems Inc.*, 1997.