Extracting Keyphrases to Represent Relations in Social Networks from Web

Junichiro Mori and **Mitsuru Ishizuka**
University of Tokyo
7-3-1 Hongo, Bunkyo-ku
Tokyo 113-8656, Japan
jmori@mi.ci.i.u-tokyo.ac.jp
ishizuka@i.u-tokyo.ac.jp

Yutaka Matsuo
National Institute of Advanced
Industrial Science and Technology
Sotokanda 1-18-13
Tokyo 101-0021, Japan
y.matsuo@aist.go.jp

Abstract

Social networks have recently garnered considerable interest. With the intention of utilizing social networks for the Semantic Web, several studies have examined automatic extraction of social networks. However, most methods have addressed extraction of the strength of relations. Our goal is extracting the underlying relations between entities that are embedded in social networks. To this end, we propose a method that automatically extracts labels that describe relations among entities. Fundamentally, the method clusters similar entity pairs according to their collective contexts in Web documents. The descriptive labels for relations are obtained from results of clustering. The proposed method is entirely unsupervised and is easily incorporated with existing social network extraction methods. Our experiments conducted on entities in researcher social networks and political social networks achieved clustering with high precision and recall. The results showed that our method is able to extract appropriate relation labels to represent relations among entities in the social networks.

1 Introduction

Social networks have recently attracted considerable interest. For the Semantic Web, there is great potential to utilize social networks for myriad applications such as trust estimation [Golbeck and Hendler, 2004], ontology construction [Mika, 2005b], and end-user ontology [Brickley and Miller, 2005].

Aiming at using social networks for AI and the Semantic Web, several studies have addressed extraction of social networks automatically from various sources of information. Mika developed a system for extraction, aggregation, and visualization of online social networks for a Semantic Web community, called Flink [Mika, 2005a]. In that system, social networks are obtained using Web pages, e-mail messages, and publications. Using a similar approach, Matsuo et al. developed a system called Polyphonet [Matsuo *et al.*, 2006b; 2006a]. In line with those studies, numerous studies have explored automatic extraction of social networks from the Web [Alema-Meza *et al.*, 2006].

Given social network extraction using the methods described above, the next step would be to explore underlying relations behind superficial connections in those networks. In the field of social network analysis, it has been shown that rich information about underlying social relationships engenders more sophisticated analysis [Scott, 2000]. However, most automatic methods to extract social networks merely provide a clue to the strength of relations. For example, a link in Flink [Mika, 2005a] is only assigned the strength of its relation. A user might wonder what kind of underlying relation exists behind the link.

One reason for the lack of information about underlying relations is that most automatic extraction methods [Mika, 2005a; Alema-Meza *et al.*, 2006; Matsuo *et al.*, 2006b] use a superficial approach (e.g. co-occurrence analysis) instead of profound assessment to determine the type of relation. Matsuo et al. defines four kinds of relations in a research community and classifies the extracted relations. They adopt a supervised machine learning method, which requires a large annotated corpus which costs great deal of time and effort to construct and administer. In addition, it is necessary to gather domain-specific knowledge a priori to define the extracted relations.

Our goal is to extract underlying relations among entities (e.g., person, location, company) from social networks (e.g., person-person, person-location network). Thereby, we are aiming at extracting descriptive labels of relations automatically such as affiliations, roles, locations, part-whole, social relationships. In this paper, we propose a method that automatically extracts the labels that describe relations among entities in social networks. We obtain a local context in which two entities co-occur on the Web, and accumulate the context of the entity pair in different Web pages. Given the collective contexts of each entity pair, the key idea is clustering all entity pairs according to the similarity of their collective contexts. This clustering using collective contexts is based on our hypothesis that entity pairs in similar relations tend to occur in similar contexts. The representative terms in context can be regarded as representing a relationship. Therefore, the labels to describe the relations among entities are extracted from the clustering process result. As an exemplary scenario for our approach, we address two kinds of social network. a researcher social network and a political social network.

The proposed method is entirely unsupervised. For that reason, our method requires neither a priori definition of relations nor preparation of large annotated corpora. It also requires no instances of relations as initial seeds for weakly supervised learning. Our method uses context information that is obtained during extraction of social networks. Consequently, the proposed method contributes to

- incorporating into existing methods of social network extraction and enriching the social network by adding relation labels.

In addition, as a Web mining approach our method contributes to

- extracting information from the Web and bootstrapping the Semantic Web by annotating relation information to social networks and Web contents.

The remainder of this paper is structured as follows. Section 2 compares our approach to other ongoing relevant research in social network extraction, relation extraction, and ontology population. Section 3 describes basic ideas of our approach and detailed steps of the proposed method. Section 4 describes our experiment. Section 5 describes results and evaluation. We end our presentation with a discussion of future work, after which we provide concluding remarks in section 6.

2 Related Work

Aiming at extracting underlying relations in social networks from the Web, our method is related closely to existing extraction methods of social networks. Several studies have addressed extraction of social networks automatically from various sources of information such as the Web, e-mail, and contacts [Alema-Meza *et al.*, 2006; McCallum *et al.*, 2005; Mika, 2005a; Matsuo *et al.*, 2006a]. While most approaches for social network extraction have focused on the strength of the relation, few studies have addressed automatic identification of underlying relations. Matsuo et al. employed a supervised machine learning method to classify four types of relations in a research community [Matsuo *et al.*, 2006b]. There have also been several important works that have examined supervised learning of relation extraction in the field of natural language processing and information extraction [Zelenko *et al.*, 2003; Culotta and Sorensen, 2004; Kambhatla, 2004]. However, a supervised method requires large annotated corpora, which cost a great deal of time and effort. In addition, it is necessary to gather the domain specific knowledge a priori to define extracted relations.

Identifying underlying relations has been also addressed in ontology population [Buitelaar *et al.*, 2005]. Particularly, the current approaches for relation extraction in ontology population are classifiable into two types: those that exploit certain patterns or structures, and those that rely on contextual features. Pattern-based approaches [Cimiano *et al.*, 2005] seek phrases or sentence structures that explicitly show relations between instances. However, most Web documents have a very heterogeneous in structure, even within individual web pages. Therefore, the effectiveness of the pattern-based approach depends on the domain to which it is applied.

Rather than exploiting patterns or structures, context-based approaches [Kavalec *et al.*, 2004; Schutz and Buitelaar, 2005]

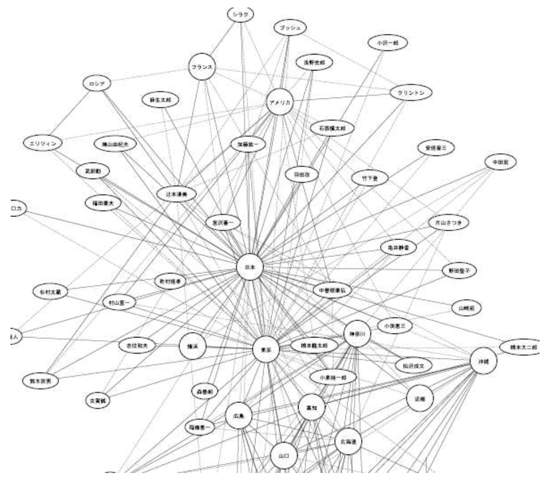

Figure 1: Political social network extracted from the Web: a circular node represents a location entity and a ellipse node represents a person entity. Each edge in the network implies that there is relation between entities.

assess contextual features such as syntactic, semantic, and co-occurrence. Several studies have employed contextual verb arguments to identify relations in text [Kavalec *et al.*, 2004; Schutz and Buitelaar, 2005], assuming that verbs express a relation between two ontology classes that specify a domain and range. Although verbs are relevant features to identify relations, it assumes that syntactic and dependency analyses are applicable to text collections. Because the Web is highly heterogeneous and often unstructured, syntactic and dependency structures are not always available. For that reason, we employed context model that uses a bag-of-words to assess context. Therefore, the method is applicable to any unstructured documents in the Web. As shown in our experiment, the simple context model performed well to extract descriptive relation labels without depending on any syntactic features in text. In the field of NLP, several studies have proposed relation extraction from a large language corpus using a bag-of-words of context [Turney, 2005; Hasegawa *et al.*, 2004]. Our method can be considered as an application of relation extraction methods in NLP to social networks and a Web mining. We are aiming at easily incorporating into extraction methods of social network from the Web. Therefore, our method uses context information that is obtained during extraction of social networks. Consequently, it serves to enrich such networks by adding relation labels.

3 Method

3.1 Concept

In this paper, as an exemplary scenario for our approach, we use two types of social network: a researcher network that is composed of researcher entities and a political social network that is composed of two types of entities: politicians and geopolitical-entities. Figure 1 shows a political social network that was automatically extracted from the Web using Mika and Matsuo's method [Mika, 2005a; Matsuo *et al.*, 2006b].

Table 1: Keywords obtained from each local context of four kinds of entities pairs: Junichiro Koizumi-Japan, Yoshiro Mori-Japan, Junichiro Mori-Kanagawa, and Yoshiro Mori-Ishikawa

Entity pair	Keywords
(1) Junichiro Koizumi-Japan	pathology, Fujiwara, **prime minister**, Koizumi, Kobun-sha, politics, visit, page, products, cabinet,...
(2) Yoshiro Mori-Japan	rugby, **prime minister**, chairman, bid, minister, association, science, administration, director, soccer, Africa,...
(3) Junichiro Koizumi-Kanagawa	**election**, **prime minister**, Yokosuka, **candidate**, **congressional representative**, Saito, **Liberal Democratic Party**, Miura,..
(4) Yoshiro Mori-Ishikawa	Ichikawa, Yasuo, **prime minister**,**election**, **Liberal Democratic Party**, Okuda, **candidate**, komatsu, **congressinal representative**,...

The social network was extracted according to co-occurrence of entities on the Web.

Given entity pairs in the social network, our present goal is to extract labels to describe the relations of respective entity pairs (to discover relevant keyphrases that relate entities). The simple approach to extract the labels that are useful for describing relations in social networks is to analyze the surrounding local context in which entities of interest co-occur on the Web, and to seek clues to describe that relation. Local context is often used to identify entities or relations among entities in tasks of natural language processing or information extraction [Grefenstette, 1994; Lin, 1998; Schutze, 1998].

Table 1 shows keywords [1] that were extracted from local contexts of four entity pairs (Junichiro Koizumi-Japan, Yoshiro Mori-Japan, Junichiro Koizumi-Kanawaga, Yoshiro Mori-Ishizuka) using keyword extraction method [Matsuo *et al.*, 2006b]. The keywords were extracted from the collective local contexts where co-occurrence of each entity pair is found. For each entity pair, the local contexts from 100 Web pages were collected. The keywords are ordered according to TF-IDF-based scoring, which is a widely used method in many keyword extraction methods to score individual words within text documents to select concepts that accurately represent the documents' contents. The keywords scored by TF-IDF can be considered as a bag-of-words model to represent the local context surrounding an entity pair.

If we examine the common keywords (shown in bold typeface in the table) shared by (1) and (2) or (3) and (4), we note that the keywords that describe the relations of each entity pair, such as "prime minister" and "candidate", are commonly shared [2]. In contrast, if we compare Koizumi's keywords (1) with another of his keywords (3), we find that different keywords appear because of their respective links to different locations: Japan and Kanagawa. (although both keywords are Koizumi's.)

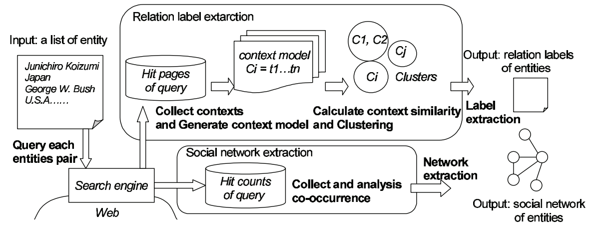

Figure 2: Outline of the proposed method

Based on the observations described above, we hypothesize that if the local contexts of entity pairs in the Web are similar, then the entity pairs share a similar relation. Our hypothesis resembles previously tested hypotheses related to context [Harris, 1968; Schutze, 1998]: words are similar to the extent that their contextual representations are similar. According to that hypothesis, our method clusters entity pairs according to the similarity of their collective contexts. Then, the representative terms in a cluster are extracted as labels to describe the relations of each entity pair in the cluster, assuming that each cluster represents different relations and that the entity pair in a cluster is an instance of a certain relation. The key point of our method is that we determine the relation labels not by examining the local context of one single entity pair, but by the collective local contexts of all entity pairs of interest. In the following section, we explain the precise steps of our proposed method.

3.2 Procedure

Our method for extraction of relation labels in social networks includes the following steps.

1. Collect co-occurrence information and local context of an entity pair
2. Extract a social network that is composed of entity pairs.
3. Generate a context model of each entity pair.
4. Calculate context similarity between entity pairs.
5. Cluster entity pairs.
6. Select representative labels to describe relations from each cluster.

Figure 2 depicts the outline of our method. Our method requires a list of entities (e.g., personal name, location name) to form a social network as the input; it then outputs the social network and a list of relation labels for each entity pair. Although collection of a list of entities is beyond the scope of this paper, one might use named entity recognition to identify entities and thereby generate a list of entities of interest.

As shown in Fig. 2, our method (upper box) can be processed in parallel with existing methods of social network

[1] In our experiment, we mainly used Web pages in Japanese. Therefore, keywords in the table are translated from their original Japanese.

[2] Junichiro Koizumi is the current Prime Minister of Japan and Yoshiro Mori is a former Prime Minister. Kanagawa is the prefecture where Koizumi was elected and Ishikawa is the prefecture where Mori was elected.

extraction (bottom box). The first step is to collect co-occurrence and local contexts of each entity pair from the Web. Many existing methods of social network extraction use a search engine and its query hit counts to obtain co-occurrence information of entities from the Web [Matsuo, Mika]. In line with such methods, we use Google [3] to collect co-occurrence information and generate a social network, as shown in Fig. 1.

Using co-occurrence information, we also collect local contexts in which elements of an entity pair of interest co-occur within a certain contextual distance of one another within the text of a Web page. For this, we downloaded the top 100 web pages included in the search result of corresponding search query to each entity pair (in our example of a politician and location name, the query is "Junichiro Koizumi AND Japan"). This can be accomplished in the process of collecting co-occurrence information, which uses search query hit counts.

3.3 Context Model and Similarity Calculation

For each entity pair, we accumulate the context terms surrounding it; thereby, we obtain the contexts of all entity pairs. As the next step, to calculate the similarity between collective contexts of each entity pair, we require a certain model that represents the collected context. In our method, we propose a context model that represents the context using a bag-of-words and a word vector [Raghavan and Wong, 1998]. We define the context model as a vector of terms that are likely to be used to describe the context of an entity pair (e.g., the keywords list shown in Table 1 can be considered as an example of the context model.). A context model $C_{i,j}$ of an entity pair (e_i, e_j) is defined as the set of N terms $t_1, ..., t_N$ that are extracted from the context of an entity pair as $C_{i,j}(n, m) = t_1, ..., t_N$, where both n and m are parameters of the context window size, which defines the number of terms to be included in the context. In addition, m is the number of intervening terms between e_i and e_j; n is the number of words to the left and right of either entity.

Each term t_i in the context model $C_{i,j}(n, m)$ of an entity pair (e_i, e_j) is assigned a feature weight according to TF-IDF-based scoring defined as $tf(t_i) \cdot idf(t_i)$. Therein, $tf(t_i)$ is defined by the term frequency of term t_i in all the contexts of the entity pair (e_i, e_j). Furthermore, $idf(t_i)$ is defined as $log(|C|/df(t_i))$+1, where $|C|$ is the number of all context models and $df(t_i)$ is the number of context models including term t_i. With the weighted context model, we calculate the similarity $sim(C_{i,j}, C'_{i,j})$ between context models according to the cosine similarity as follows: $sim(C_{i,j}, C'_{i,j}) = C_{i,j}C'_{i,j}/(|C_{i,j}||C'_{i,j}|)$.

In our exploratory experiment we tried probability distribution-based scoring and several similarities such as L1 norm, Jensen-Shannon and Skew divergence [Lin, 1998]. According to that results, TFIDF-based cosine similarity performs well.

[3]http://www.google.com

3.4 Clustering and Label Selection

Calculating the similarity between the context models of entity pairs, we cluster all entity pairs according to their similarity. This is based on our hypothesis: the local contexts of entity pairs in the Web are similar, the entity pairs share a similar relation.

In our clustering process, we do not know in advance what kinds of relation pertain and therefore how many clusters we should make. Therefore, we employ hierarchical agglomerative clustering. Several clustering methods exist for hierarchical clustering. According to our exploratory experiment, complete linkage performs well because it is conservative in producing clusters and does not tend to generate a biased large cluster. In complete linkage, the similarity between the clusters CL_1, CL_2 is evaluated by considering the two most dissimilar elements as follows. $min_{C_{i,j} \in CL_1, C'_{i,j} \in CL_2} sim(C_{i,j}, C'_{i,j})$.

The clustering process terminates when cluster quality drops below a predefined threshold. The cluster quality is evaluated according to two measures [Kannan *et al.*, 2000]: the respective degree of similarity of entity pairs within clusters and among clusters. After the clustering process terminates and creates a certain number of clusters, we extract the terms from a cluster as labels to describe the relations of each entity pair in the cluster. This is based on our assumption that each cluster represents a different relation and each entity pair in a cluster is an instance of similar relation. The term relevancy, as a cluster label, is evaluated according to a TFIDF-based measure in the same manner as weighting the terms in a context model. However, in this process, the term frequency is determined for all contexts of a cluster. The underlying idea is to extract terms that appear in the cluster, but which do not appear in other clusters. With a cluster CL's labels $l_1, ..., ln$ scored according to the term relevancy, an entity pair, e_i and e_j, that belongs to the CL can be regarded as holding the relations described by $l_1, ..., ln$.

4 Experiment

Using our proposed method, we extracted labels to describe relations of each entity pair in a social network. We chose 143 distinct entity pairs (pair of a politician and a geo-political entity) from a political social network and 421 entity pairs (pair of Japanese AI researchers) from a researcher network [Matsuo *et al.*, 2006a].

We created a context model of each entity pair using nouns and noun phrases from part-of-speeches (POS) of surrounding entity pairs in a Web page. We exclude stop words, symbols and highly frequent words. For each entity pair, we download the top 100 web pages in the process of collecting co-occurrence information for extraction of social network. For the context size, we used two parameters, m and n, as explained in Sect. 3.3. As a baseline of the context size, we assigned 10 and 5, respectively, to m, n.

We used complete-linkage agglomerative clustering to cluster all entity pairs. Thereby, we created five distinct clusters for the political social network and twelve distinct clusters for the researcher network according to the predefined thresholds of two quality measures within the clusters and

Table 2: Cluster label (left) and automatically extracted relation labels from a cluster (right)

political social network (5 clusters)

1	mayor	mayor, citizen, hosting, president, affairs, officer, mutter, answer, city, conference
2	president	president, administration, world, Japan, economics, policy, war, principle, politics, Iraq
3	prime minister	prime minister, administration, politics, article, election, prime minister, government, peace
4	governor	prefectural governor, president, prefectural government, committee, Heisei, prefectural administration, mayor
5	congressional representative	congressional representative, election, Liberal Democratic Party, candidate, lower house, Democratic Party, proportional representation

researcher social network (6 representative clusters among 12 clusters)

1	co-authorship of conference paper	paper, author, conference venue, presentation, title, program
2	co-authorship of book	edit, book, publishing, programming, recommendation, co-author
3	co-edit of book	edit, revision, article, publishing, educational material, editor
4	collaborative project	representative person, contributor, minister, acceptance
5	co-authorship of journal paper	journal, Shogi, distribution, computer, information processing society of Japan
6	same affiliation	University of Tokyo, metropolitan, techonolgy, University, science

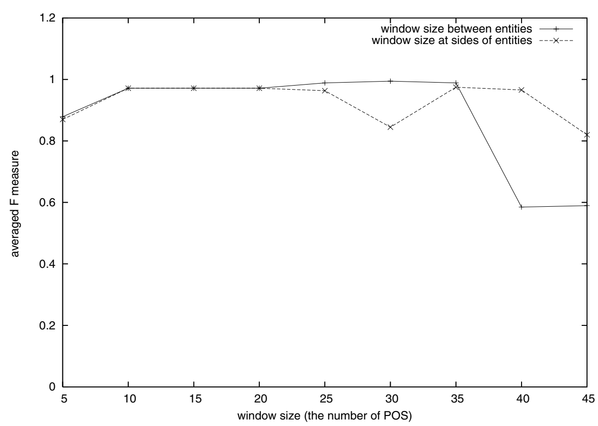

Figure 3: F measure of clustering results vs. Context window size

among the clusters explained in Sect. 3.4. To evaluate the clustering results and the extracted labels, two human subjects analyzed the context terms of each entity pair and manually assigned the relation labels (three or fewer possible labels for each). Then, a cluster label was chosen as the most frequent term among the manually assigned relation labels of entity pairs in the cluster. The manually assigned relation labels are used as ground truth in the subsequent evaluation stage.

In Table 2, the left column shows the label of each cluster. The right column shows the highly-scored terms[4] that are extracted automatically from each cluster, which can be considered as the labels to describe relations of each entity pair in the cluster. The terms are sorted by relevancy score.

5 Evaluation

We first evaluated the clustering results using a political social network. For each cluster cl, we counted the number of entity pairs $EP_{cl,correct}$ whose manually assigned relation labels include the label of cluster cl. We also counted the entity pairs $EP_{cl,total}$ in the cluster cl. Next, for each relation label l, we counted the number of entity pairs $EP_{l,correct}$ that have the relation label l whose cluster label is l. We also counted the entity pairs $EP_{l,total}$ that have the relation label l. Then, precision and recall of the cluster were calculated as:

$$precision = \Sigma_{cl \in CL} \frac{EP_{cl,correct}}{EP_{cl,total}}, \; recall = \Sigma_{l \in L} \frac{EP_{l,correct}}{EP_{l,total}}.$$

According to $precision$ and $recall$, we evaluated clusters based on the F measure.

[4]Terms in the table are translated from their original Japanese.

Table 3: Clustering performance in parameters of context window size

Context window size n, m	Precision	Recall	F-measure
$n = 10, m = 30$	0.992	0.995	0.994
$n = 5, m = 10$	0.88	0.85	0.86
All terms in a Web page	0.76	0.677	0.716

The graph depicted in Fig. 3 shows that the clustering results vary depending on the context size. Consequently, to find the optimal context size, we calculate the F-measure by changing two size parameters: m and n. Expanding the context size from the minimum, the F-measure takes an optimal value when m is around 30 and n is around 10 (Fig. 3 and Table 3) . We employed this optimal context size to extract the relation labels in our experiment. After reaching the peak, the value of the F-measure decreases as the context size increases. The wider context window tends to include noise terms that are not appropriate to represent the context, thus rendering the similarity calculation between the contexts irrelevant. The optimal context size depends on the structural nature of language. Consequently, we must choose the context size carefully when applying our methods to a different language.

To evaluate the automatically extracted relation labels, we compared the cluster label (left column of Table 2) with the automatically extracted relation labels (right column of Table 2). For a political social network, we found that the relation label that has the highest score is equal to the corresponding cluster's relation label. Precision of the clustering results in our experiment is quite high, as shown above. Therefore, we can say that each entity pair in a cluster is represented properly by the highest-scored relation labels. For a researcher social network, extracted relation labels are highly correlated with a manually assigned clustering label. Matsuo et al. defined four kinds of relations for a research social network: co-authorship, same affiliation, same project, and same conference [Matsuo *et al.*, 2006b]. We found that extracted clusters and relation labels are corresponding those relations.

6 Conclusions and Future Work

We propose a method that automatically extracts labels that describe relations between entities in social networks. The proposed method is entirely unsupervised and domain-independent; it is easily incorporated into existing extraction methods of social networks.

Recent important approaches of a Web mining toward the Semantic Web use the Web as a huge language corpus and combine with a search engine. The underlying concept of these methods is that it uses globally available Web data and structures to annotate local resources semantically to bootstrap the Semantic Web. In line with this, our approach utilizes the Web to obtain the collective contexts which engender extracting representative relations in social network. As pointed in [Matsuo *et al.*, 2006a], we claim that relations should be defined not by local information but rather by a global viewpoint of a network composed of individual relations.

Future studies will explore the possibilities of extending the proposed method to relations in other types of social networks. Enriching social networks by adding relation labels, our method might contribute to several social network applications such as finding experts and authorities, trust calculation, community-based ontology extraction, and end user ontology.

References

[Alema-Meza *et al.*, 2006] B. Alema-Meza, M. Nagaraja, C. Ramakrishnan, A. Sheth, I. Arpinar, L. Ding, P. Kolari, A. Joshi, and T. Finin. Semantic analytics on social networks: Experiences in addressing the problem of conflict of interest detection. In *Proc. WWW*, 2006.

[Brickley and Miller, 2005] D. Brickley and L. Miller. Foaf vocabulary specification. namespace document. 2005.

[Buitelaar *et al.*, 2005] P. Buitelaar, P. Cimiano, and B. Magnini. *Ontology Learning from Text: Methods, Evaluation and Applications*. IOS Press, Amsterdam, 2005.

[Cimiano *et al.*, 2005] P. Cimiano, G. Ladwig, and S. Staab. Gimme' the context: Context-driven automatic semantic annotation with cpankow. In *Proc. of WWW*, 2005.

[Culotta and Sorensen, 2004] A. Culotta and J. Sorensen. Dependency tree kernel for relation extraction. In *Proc. ACL*, 2004.

[Golbeck and Hendler, 2004] J. Golbeck and J. Hendler. Accuracy of metrics for inferring trust and reputation in semantic web-based social networks. In *Proc. EKAW*, 2004.

[Grefenstette, 1994] G. Grefenstette. *Explorations in Automatic Thesaurus Construction*. Kluwer, 1994.

[Harris, 1968] Z. Harris. *Mathematical Structures of Language*. Wiley, 1968.

[Hasegawa *et al.*, 2004] Takaaki Hasegawa, Satoshi Sekine, and Ralph Grishman. Discovering relations among named entities from large corpora. In *Proc. of ACL*, 2004.

[Kambhatla, 2004] N. Kambhatla. Combining lexical, syntactic, and semantic features with maximum entropy models for extracting relations. In *Proc. ACL*, 2004.

[Kannan *et al.*, 2000] R. Kannan, S. Vempala, and A. Vetta. On clustering: Good, bad and spectral. *Computer Science*, 2000.

[Kavalec *et al.*, 2004] M. Kavalec, A. Maedche, and V. Svatek. Discovery of lexical entries for non-taxonomic relations in ontology learning. In *Van Emde Boas, P., Pokorny, J.,Bielikova, M.,Stuller, J. (eds.). SOFSEM 2004*, 2004.

[Lin, 1998] D. Lin. Automatic retrieval and clustering of similar words. In *Proc. COLING-ACL98*, 1998.

[Matsuo *et al.*, 2006a] Y. Matsuo, M. Hamasaki, H. Takeda, J. Mori, B. Danushka, H. Nakamura, T. Nishimura, K. Hashida, and M. Ishizuka. Spinning multiple social network for semantic web. In *Proc. AAAI*, 2006.

[Matsuo *et al.*, 2006b] Y. Matsuo, J. Mori, M. Hamasaki, K. Ishida, T. Nishimura, H. Takeda, K. Hashida, and M. Ishizuka. Polyphonet: An advanced social network extraction system. In *Proc. WWW*, 2006.

[McCallum *et al.*, 2005] A. McCallum, A. Corrada-Emmanuel, and X. Wang. Topic and role discovery in social networks. In *Proc. IJCAI*, 2005.

[Mika, 2005a] P. Mika. Flink:semantic web technology for the extraction and analysis of social networks. *Journal of Web Semantics*, 3(2), 2005.

[Mika, 2005b] P. Mika. Ontologies are us: A unified model of social networks and semantics. In *Proc. ISWC*, 2005.

[Raghavan and Wong, 1998] V.V. Raghavan and S.K.M. Wong. A critical analysis of vector space model for information retrieval. *Journal of the American Society for Information Retrieval*, 35(5), 1998.

[Schutz and Buitelaar, 2005] A. Schutz and P. Buitelaar. Relext: A tool for relation extraction from text in ontology extension. In *Proc. ISWC*, 2005.

[Schutze, 1998] H. Schutze. Automatic word sense dicrimination. *Computational Linguistics*, 24(1), 1998.

[Scott, 2000] J. Scott. *Social Network Analysis: A Handbook*. Sage Publications, London, 2000.

[Turney, 2005] P. Turney. Measuring semantic similarity by latent relational analysis. In *Proc. IJCAI*, 2005.

[Zelenko *et al.*, 2003] D. Zelenko, C. Aone, and A. Richardella. Kernel methods for relation extraction. *Machine Learning Research*, 2003(2), 2003.