Extracting Inter-Firm Networks from World Wide Web

Yingzi Jin
University of Tokyo
Hongo 7–3–1, Tokyo 113-8656,Japan
eiko-kin@mi.ci.i.u-tokyo.ac.jp

Yutaka Matsuo
National Institute of Advanced
Industrial Science and Technology
y.matsuo@aist.go.jp

Mitsuru Ishizuka
University of Tokyo, Japan
Hongo 7–3–1, Tokyo 113-8656, Japan
ishizuka@i.u-tokyo.ac.jp

Abstract

*Social relations play an important role in a real communities. Interaction patterns reveal relations among actors (such as persons, groups, firms), which can be merged to produce valuable information as a network structure. This paper presents a new approach to extract inter-firm networks from the web for further analysis. Extraction of relations between a pair of firms is realized using a search engine and text processing. Because names of firms co-appear coincidentally on the web, we propose an advanced algorithm, which is characterized by addition of keywords (*relation keywords*) to a query. The relation keywords are obtained from the web using a Jaccard coefficient. We present some examples and comprehensive evaluations of our approach.*

1. Introduction

Various relationships exist among firms such as mergers, acquitions and partnerships. Together, these relationships define a network between firms. Such networks are useful in analyzing a firms' competitiveness and helps in determining its marketing strategy. Furthermore, overall network features can assist us in analyzing the stability and growth of the industry. Numerous studies of social network analyses have been conducted in the fields of economics and other social sciences [3, 19, 1, 23].

Many studies have investigated methods to extract social networks from the web while targeting people (particularly researchers or students). For example, using social networking services (SNS), aggregating Friend-of-a-Friend (FOAF) documents [6, 16]. Particularly, several studies have been undertaken to use a search engine to extract social networks [10, 15, 13, 14]. Co-occurrence of names on the web is commonly used as proof of relational strength. However, the co-occurrence methods can not apply directly for some entities such as famous people, organization or firms, which have multiple relations, and relational information on the web affected by media effect. Many economic analyses of inter-firm networks have been obtained relational data only from the stock market or shareholding information in business magazines that are much less diverse [1, 22].

Many relations among firms are published on the web in news articles or news releases (Fig.1). Our work emphasizes the investigation of such published relations on the web to address the relation extraction problem. Given a list of firms V=$\{v_1, v_2, ...\}$, our goal is to retrieve and extract relations among them to construct inter-firm networks $G(V, E)$, in which each edge e=$(v_1, v_2) \in E$ represents a relationship between v_1 and v_2. We specifically seek to develop methods that acquire relationships from the web, the largest available resource that deals with all firms. For each pair of firms (v_1, v_2), our system address two problems: (a) collecting information about target relations, such as "Company v_1 merged with Company v_2"; and (b) relation extraction, such as extract capital alliance (*merge*) from above sentence. For collecting information from entire web, we use a general-purpose search engine. Query expansion and modification techniques are applicable in this case [8, 18]. Research on relation extraction has been promoted by Message Understanding Conferences (MUCs) and Automatic Content Extraction (ACE) programs. Numerous techniques to address this task have been proposed in the literature, such as pattern matching [5], kernel methods [24], and logistic regression [9]. For the firm case, our extraction task is to detect relations among same types (i.e., *COM* type) of entities.

In this study, we use a search engine to collect target

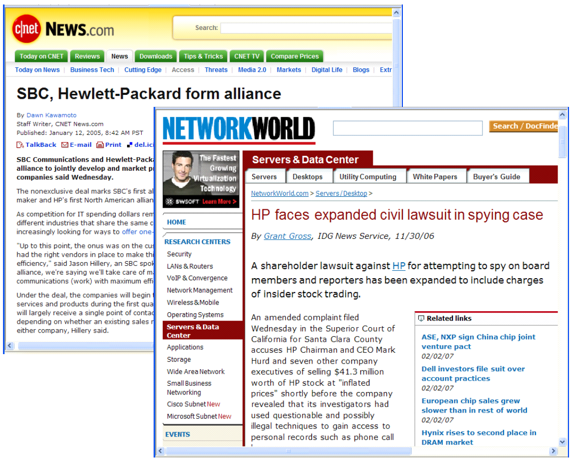


Figure 1. News about firms' relationships on the web

relational pages from the web. Since names of firms co-oppear coincidentally on the web, we propose to add additional words (call *relation keyword*) to name pairs of firms as a query. We then apply a simple pattern-based approach to extract the relations. We extract alliance relations as a positive relation and lawsuit relations as a negative relation. Much of this daily information is obtainable from the web. Examination of daily changing and complex social relationships is important for analyzing social trends, understanding social structures, and for formulating new industrial activities. Our method is a first attempt to extract inter-firm networks from the web using a search engine. Our approach is applicable to other entities, such as famous persons, or other multiple relational entities.

The remainder of this paper is organized as follows. We show related work in section 2. Our proposed idea and methods are described in section 3. We also show experiments and evaluations in section 4 before we conclude this paper.

2 Related works

Many studies have used search engines to extract social networks automatically from the Web [10, 15, 14]. Co-occurrence of names on the Web is commonly used as evidence of relational strength [10]. Related to the Semantic Web community, P. Mika developed a system called *Flink*, which extracts relational information from web pages, e-mail messages, publications, and self-created FOAF profiles [15]. The web mining component of the system uses a search engine to measure the strength of relations among researchers. Comparably, Y. Matsuo and his colleagues developed a system called *POLYPHONET*, mainly for use by Japan's AI community [14]. However, the co-occurrence-based methods become ineffective when two target entities co-occur universally on many Web pages. We take two persons to explore this problem: Bill Gates and George Bush. Those two names "coincidentally" co-occur on the Web very often: They may be on the same news pages just because they made some statements on the same day. They may be on the pages that list "the most famous persons in the world." For that reason, it is not a good idea to measure the strength of relations simply through the use of co-occurrence measures. This problem is commonly confronted for firms: a firm name is similar to a famous person's name, and they often co-occur for various reasons, even though no formal relations exist among them. When the relation between firms attracts attention by media services (such as a lawsuit relation), many pages describe and comment on it; in contrast, only a few pages exist on the web if the relation gets no attention. Considering that media effects influence the number of web pages that appear, co-occurrence of names on the web is not always useful to represent the actual relations linking two firms.

Web search by query modification and expansion is described in [8]; they extracted query modification rules for finding personal homepages and calls for papers. For information retrieval and query expansion, S. Oyama's work is more closely related to ours [18]. They added keywords called "keyword spices" to the user's input query with a Boolean expression for a domain-specific web search. They sampled web pages using initial keywords and classified them manually as either relevant or irrelevant, thereby producing a training corpus. Subsequently, they applied a decision-tree learning algorithm to discover keyword spices. Our system sets relation keywords that are added by query as combinations of one or two terms. Therefore, a Jaccard coefficient is used simply to measure the scores. Other studies such as Flink use a phrase "*Semantic Web* OR *Ontology*", POLYPHONET adds affiliation information together with a name for disambiguation. To extract characteristic key phrases for a person automatically, D. Bollegara clusters web pages that are related by each name into several groups using text similarity [4].

Battiston et al. extract shareholding relationships from stock market information (MIB, NYSE and NASDAQ) to analyze characteristics of market structure [1]. Souma et al. extract data published by Tokyo Keizai Inc. to construct Japanese shareholding networks to analyze features of Japanese companies' growth [22]. Our work specifically addresses alliance and lawsuit relations among firms from published resources on the web. Consequently, we can obtain relations easily and can track down daily changing and social trends. Dealing with time series changes of relations is one of our interests for future work.

Name disambiguation is an important problem for social network mining. To date, several studies have produced attempts at personal name disambiguation on the Web [2, 4, 11]. However, ambiguity in firm (or organization) names is lesser compared to personal names. We intend to explore ambiguities in company names in our future work.

3 Social Network Extraction for Firms

3.1 Basic Concept

In social sciences, the definition of a weak or strong tie might vary among contexts [12]. For example, the frequency or degree of relations affects that strength; multiple relations between two actors also can imply a stronger tie. In the firm case, the types of relations define the strength: For example, a capital alliance relation is stronger than a business alliance relation. Consequently, to present a tie among firms, it is appropriate that we identify the concrete relations of firms.

For using a search engine to retrieve and extract relations, a proper query is necessary. The intuitive query is the names of the two firms. For example, we issue a query such as "*Matsushita* AND *JustSystem*"[1] to discover data containing their relationships. Thereby, we obtain as many as 425,000 pages. Many top-ranked pages are lawsuit-relation pages[2] which drew much attention during the last year. Therefore, analyzing these pages, we were able to identify lawsuit relations among them. However, two companies showed a collaboration relation in knowledge management in 2001, which pages are in lower ranks of 124^{th}, on account of the collaboration relation occurred years ago, it might be lost. Of course, we can download and analyze all the returned pages from a search engine to find all possible relations, but that is both time consuming and costly.

As a solution, we can add some word or combination of words (called *relation keyword*) to a search query and apply text processing to confirm the existence of fact. Using this strategy, we can efficiently identify relations among firms. For example, when we wish to extract lawsuit relations, we add a term "*lawsuit*". We issue a query "*Matsushita* AND *JustSystem* AND *lawsuit*" so that the search engine will return the lawsuit pages that are associated with the two firms. Then we can conduct text processing to these pages to validate the relation's existence. This idea is similar to keyword spices [18], which extend queries for domain-specific web searches. Question-answering systems also construct elaborate queries for using a search engine [20]. Requirements of relation keywords are identifying the relations more precisely and reducing the leakage of relation pages if they exist. Therefore, both precision and recall are important for relation keywords.

[1] Both are names of famous Japanese corporations.
[2] http://pc.watch.impress.co.jp/docs/2005/0201/just2.htm

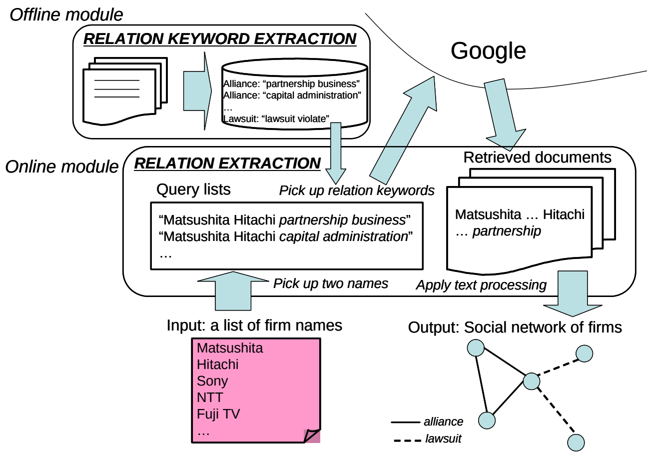


Figure 2. System flow to extract a firm network.

Our system has two major procedures: an online procedure and an offline procedure. In the offline, relation keywords for each relation types are obtained beforehand using our proposed method. In the online, a list of firms and specific relation types are given as an input and the output is a social network of firms. In the following, we will first consider relation types described in our study; then we propose relation keyword extraction. Finally, we will describe online processes of our system. The entire system is depicted in Fig. 2.

3.2 Relation type

Relationships among firms are various. For example, capital combinations such as mergers, acquisitions, joint ventures, and business partnerships, such as business alliances, co-development, service provision, and dispatching personnel, competition, lawsuit, etc. It is considerable that pairs of firms have multiple relations. For example, two firms have alliance and lawsuit relations. Each relation is typed in a more detailed manner. Alliance relations between firms include capital alliances and business alliances, where the former usually represents a stronger relation than the latter. A lawsuit relation has multiple stages; at some time, the dispute will be settled by mutual accommodation or by final judgment. Therefore, the relation can be typed as either on in a claim phase or in an accommodation phase. For dynamic and complex relational networks, it is important to distinguish such typical and temporal relations for detailed analyses of social networks [12, 21].

In this study, we address an alliance and lawsuit relation, which respectively represent positive and negative relations. The alliance relation is distinguished by business alliances and capital alliances; in addition, the lawsuit relation is separated into a claim phase and an accommodation phase. In the following, we designate these separate relations as *detail relations*.

3.3 Relation Keyword Extraction

In this section, we describe relation keyword extraction methods which are useful to collect relation pages from the web, and which are useful for the relation extraction procedure. Good relation keywords should satisfy a proper balance between specificity and generality.

The intuitive method for finding relation keywords is to select terms that appear often in the target pages (where the target relation is described) and which do not appear in other pages. Therefore, we need to collect annotated web pages of specific relations of the firms as a training corpus. Then we estimate the classification features of each word and word combination. We simply measure the F-value for each word (or word combination) w to see how the training documents can be classified correctly. However, collecting and annotating the training corpus requires many hours of tedious work.

In our study, we propose to use a search engine to extract relation keywords. This method is identical to that of Mori's work [17], in which a specific word w_c is assigned, which can represent the relation most precisely. In our work, we regarded w_c as seeds of relation keywords. If we want to retrieve an alliance relation, we add w_c such as "*alliance*" to a search query; words that co-occur frequently with it also become good clues to discern the relation. We use the Jaccard coefficient, to measure the relevance of word w to word w_c.

$$J_{w_c}(w) = |w_c \cap w| / |w_c \cup w|, \tag{1}$$

Where, $|w_c \cap w|$ represents the number of hits yielded by the query w_c AND w and $|w_c \cup w|$ represents the number of hits by the query w_c OR w. Words w with large Jaccard coefficients are also used as relation keywords aside from w_c. It would save costs of annotating training data with relevance or non-relevance manually. For choosing candidate words, it is necessary to prepare some target pages. However, they are easily obtainable from several news articles such as Yahoo! News, about target relations.

As preprocessing, we first eliminate all html tags and scripts from these web pages. Then we extract the body text of pages and apply a part-of-speech tagger Chasen[3] to choose nouns and verbs (except stop words). Then we select

[3] http://chasen.naist.jp/hiki/ChaSen/

```
function RELATIONEXTRACTION(D, x, y, W)
score_xy ← 0
S ← GetSentences(D,x,y)
for each s ∈ S do
  if s contains "x" and s contains "y" then
  score_s ← Σ_{w_i(∈W) contained in s} t_{w_i}
  if score_s > score_xy then
    score_xy ← score_s
done
if score_xy > score_thre then
  do set an edge between x and y in G
done
```

Figure 3. A procedure to extract relations using text processing.

the top N words with highest $tf * idf$ score[4]. These words are candidates of relation keywords. We also use two-word combinations as candidates. We measure the score of each candidate word / phrase by calculating the Jaccard coefficient with a seed of relation keywords w_c (We used *alliance* AND *corporate* as w_c for alliance relations. In addition, we use the word that appeared in the first lines in Table 1 as w_c for each relation: We determine these words through preliminary experiments.). Candidates with the highest scores are recognized as relation keywords.

Choosing the relation keywords can be treated as feature selection for classifying relation pages, but a combination of complex queries does not work well for a search engine. Therefore, we simply consider words or combinations of words as relation keyword candidates. It is explicit that the weight of w varies according to the relation types r. Once we find the relation keyword, we can extract the relations among many firms. For detailed relations, it is necessary to prepare relation keywords for each detailed relation, but extraction methods for relation keywords are similar.

3.4 Relation Extraction

Online, a list of firms and specific relation types are given as an input and the output is a social network of firms. Three steps exist: making queries, Google search, and network construction. First, we make queries by adding relation keywords to each pair of firms. We use top n_q relation keywords from Table 1. Then, we put these queries into the Google search engine to collect top-n_p web pages. For

[4] Here, $tf * idf = tf(w) * log(N/|w|)$, where $tf(w)$ is the number of occurrences in news articles containing w. In addition, N is the total number of Web documents, and $|w|$ is the number of web pages containing w

this experiment, we set $n_q = 2$ and $n_p = 5$. Finally, for each downloaded document D, we conduct text processing to judge whether or not the relation actually exists. A simple pattern-based heuristic (as described in Fig. 3) has been useful in our experience. We first select all sentences S that include the two firms' names (x and y) and assign each sentence the sum of relation keyword scores t_w in the sentence. The score of firms x and y is the maximum of the sentence scores. An edge is invented between the two firms if $score_{xy}$ is greater than a certain threshold, i.e., if the two firms seem to have the target relation with high reliability.

4 Experiments

A network of 60 firms in Japan including IT, communication, broadcasting, and electronics firms, is extracted. For the dataset, we manually created a dataset for these 60 firms. The annotators decided the relations among the firms based only using the information available on the web. In our experiments, we will first show the extracted relations and networks about alliance and lawsuit (and detail relations) among these firms, and indicate the overall performance of our system. Then we will represent extracted relation keywords and show their effectiveness. Finally, we will show the application of our system to Semantic Web.

4.1 Extracting Relation Keywords

To extract relation keywords for each concrete relation, we gathered 456 pages and 165 pages, respectively, for alliance and lawsuit relations from Nikkei Net and IP News sites [5]. After preprocessing and scoring, we obtained the highest scores as relation keywords. Table 1 shows the top five relation keywords and their Jaccard scores denoted as t_w [6].

To evaluate the effectiveness of relation keywords, we compared information contained in retrieved pages merely by putting a pair of names as a search query to adding relation keywords to the query. We compared five methods as follows:

- **noW:** A firm pair (without relation keywords) is used as a query.
- **W1:** A firm pair and the top-weighted relation keyword (w_1) are used as a query.
- **W2:** A firm pair and the second-weighted relation keyword (w_2) are used as a query.
- **W1+ W2:** It generates two queries: W1 and W2.
- **W1+W2+noW:** It generates three queries: W1, W2, and noW.

The **noW** query is considered as the existing method (i.e. Mika and Matsuo's method) as baseline of this evaluation; the others are proposed method variations. In all cases, we downloaded the same number of web pages. The other conditions are all same. For instance, one of the variations of our method **W1+W2+noW** generates three queries W1, W2, noW, and download 10 pages in total for the three queries. For example, using W1 as the query we download 3 pages, 4 for W2, and 3 for noW.

Figure 4 shows the results. Overall, the proposed methods perform better than the existing method (**noW**) with respect to precision. The precision and recall are 65.7% / 95.0%, respectively, if we do not use relation keywords at all. Relation keywords improve the precision using the same number of downloaded documents. By integrating multiple queries (as **W1+W2+noW** case), we can achieve the highest precision as 71.9% while maintaining a high recall (92.5%).

[5]Nikkei Net (http://release.nikkei.co.jp/) is a famous online business newspaper. IP News (http://news.braina.com/judge.html) is an online news archive of intellectual property issues.

[6]For our experiment, we mainly used web pages in Japanese. Therefore, relation keywords are translated from Japanese.

Table 2. Precision and recall of the system.

Target relation	Precision	Recall
Alliance	60.9% (70/115)	62.0% (70/113)
Capital alliance	75.0% (9/12)	42.9% (9/21)
Business alliance	67.4% (60/89)	60.0% (60/100)
Lawsuit	61.5% (16/26)	100% (16/16)
Claim phase	63.6% (14/22)	87.5% (14/16)
Accommodation	72.7% (8/11)	88.9% (8/9)

Table 3. Precision and recall in a particular web site.

Target relation	Precision	Recall
Alliance	100.0% (27/27)	23.8% (27/113)
Capital alliance	100.0% (6/6)	28.6% (6/21)
Business alliance	100.0% (21/21)	21.0% (21/100)
Lawsuit	100.0% (11/11)	68.8% (11/16)
Claim phase	100.0% (11/11)	68.8% (11/16)
Accommodation	100.0% (6/6)	66.7% (6/9)

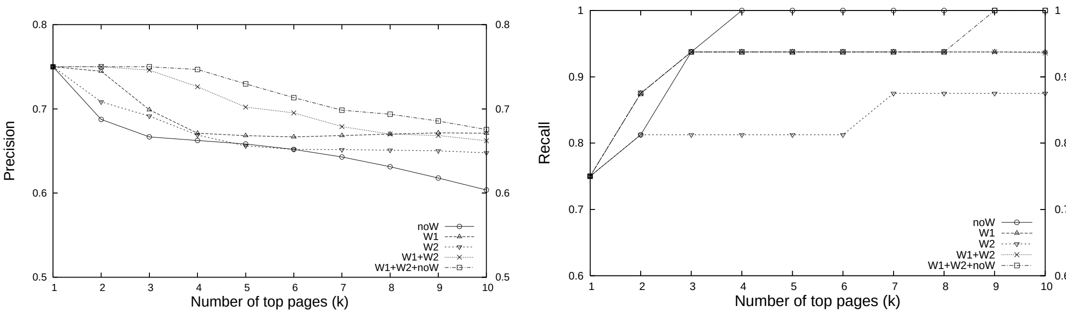
4.2 Extracting Relations and Networks

The obtained network for 60 firms in Japan is shown in Fig. 5. Bold lines represent capital alliances, thin lines are business alliances, dashed lines represent the claim phases

Table 1. Relation keywords extracted from the web using a Jaccard coefficient.

Alliance relation	t_w	Capital alliance	t_w	Business alliance	t_w
alliance AND *corporate*	1.000	*operation* AND *capital*	1.000	*alliance* AND *business*	1.000
alliance AND *stock*	0.878	*capital* AND *manage*	0.553	*alliance* AND *company*	0.475
alliance AND *company*	0.704	*capital* AND *company*	0.548	*alliance* AND *operation*	0.459
alliance AND *system*	0.565	*capital*	0.543	*alliance* AND *develop*	0.437
alliance AND *business*	0.534	*capital* AND *manage*	0.533	*alliance* AND *company*	0.432

Lawsuit relation	t_w	Claim phase	t_w	Accommodation phase	t_w
violate AND *lawsuit*	1.000	*violate* AND *sue*	1.000	*lawsuit* AND *accommodate*	1.000
violate AND *claim*	0.514	*patent* AND *sue*	0.533	*accommodate* AND *company*	0.648
violate AND *judge*	0.490	*sue* AND *technology*	0.486	*accommodate* AND *announce*	0.646
violate AND *court*	0.458	*sue* AND *develop*	0.483	*accommodate* AND *develop*	0.641
violate AND *indemnify*	0.444	*sue* AND *relevance*	0.469	*accommodate* AND *product*	0.640



(a) Precision of retrived pages

(b) Recall of relations

Figure 4. Evaluation of relation keywords for lawsuit relations.

in lawsuit relations and dotted lines are accommodation phases in the lawsuits.

Using our system described in Section *4*, we extract relationships among 60 firms. The precision and recall of our system are shown in Table 2. For ${}_{60}C_2 = 1770$ pairs of firms, 113 pairs actually show alliance relations. Our system correctly extracted 70 pairs. There were actually 21 and 100 pairs of capital and business alliances; our system extracted 9 and 60, respectively. Compared to alliances, the lawsuit relations show higher recall, probably because lawsuit relations are described in rather common formats using words such as *judgment*, *lawsuit*, or *accommodate*.

The simple pattern-based rule can extract relations between firms efficiently. Sometimes, it is unable to deal with complex meanings of sentences. Applying advanced relation extraction approaches, such as conversion of sentences into syntactic tree, might improve future results.

Although they are not comparable technically, we compared the data set against Nikkei Net and IP News, using the search functionality provided in these sites. We collected all alliance and lawsuit relations from each firm's news articles appeared in these sites, and compared those relations to our results. The precision values at these sites are 100%, but the respective recall rates of alliance and lawsuit relations among 60 firms are low, at 22.8% and 68.8%, respectively, because these sites deal little with information related to small companies and foreign corporations. The alliance and lawsuit relations are easily obtainable from the web using our algorithm.

4.3 Application

The obtained network is useful for Semantic Web studies in several ways. We might find a cluster of firms and characterize a firm by its cluster. Business experts often make such inferences based on firm relations and firm groups. For that reason, the firm network might enhance inferential abilities on the business domain. As a related work, F. Gandon et al. build a Semantic Web server that maintains annotations about the industrial organization of Telecom Valley to partnerships and collaboration [7].

We present a prototypical example of applications using a network of firms. We calculate the *centrality*, which is a measure of the structural importance of a node in the network, for each firm on the extracted network (on alliance relations). Table 4 shows the top ten firms by eigenvector and betweenness centrality. These firms have remained large and reliable corporations in Japan for decades. Interestingly, IBM, Livedoor and Cisco are on the list. These firms might bridge two or more clusters of firms: IBM and

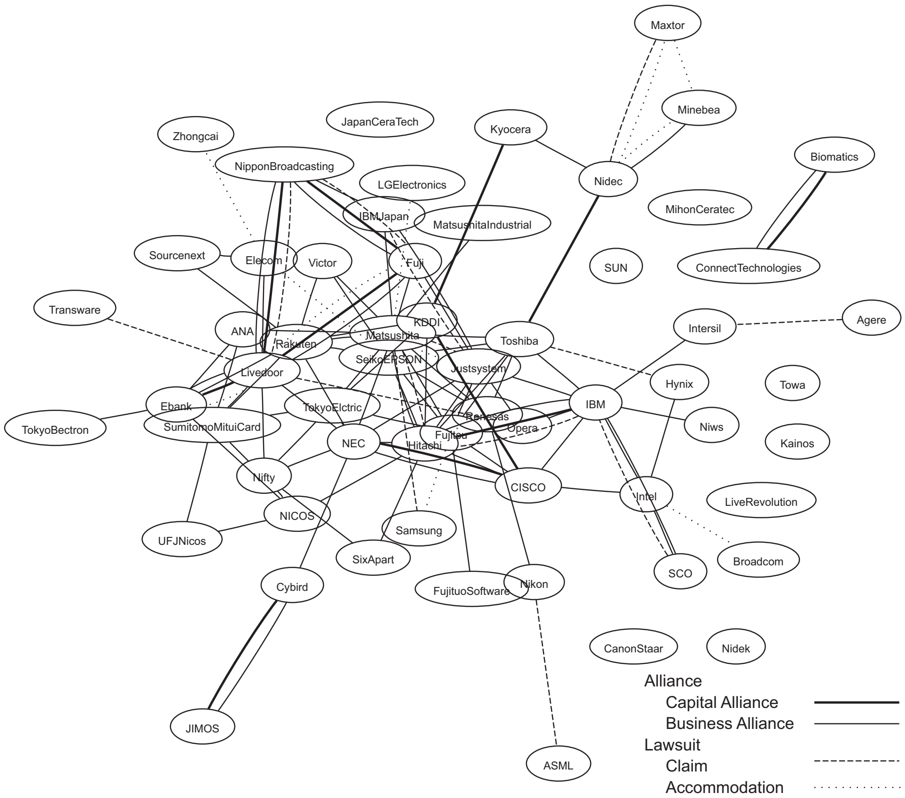


Figure 5. Network of 60 firms in Japan.

Table 4. Centrality

Rank	Eigenvector		Betweenness	
	Name	Value	Name	Value
1	Matsushita	0.366	Matsushita	168.981
2	Hitachi	0.351	IBM	149.192
3	NEC	0.289	NEC	144.675
4	Fujitsu	0.275	Hitachi	136.978
5	Toshiba	0.263	Toshiba	113.239
6	Rakuten	0.257	Rakuten	109.887
7	JustSystem	0.241	JustSystem	77.175
8	KDDI	0.208	Livedoor	74.141
9	Tokyo Electric	0.207	CISCO	64.558
10	Seiko Epson	0.204	Fujitsu	56.081

Cisco are United States firms and form alliances with firms in multiple clusters; Livedoor is famous for its aggressive M & A strategy in Japan. Such information can only be inferred after extracting a network. There seem to be many potential applications that can make use of social networks in various analysis.

5 Conclusion

This paper described a method to extract inter-firm networks from the web. Given a list of names of firms, our system uses a search engine to collect target pages from the web, and applies text processing to construct a network of firms. To retrieve target pages we append the query with keywords indicating the relation. Moreover, we proposed an automatic method to extract such keywords from the web. Although we focused on alliance and lawsuit relations, in future we plan to extend the propsed method to other types of relations between firms.

References

[1] S. Battiston. Inner structure of capital control networks. *Physica A*, 338:107–112, 2004.

[2] R. Bekkerman and A. McCallum. Disambiguating web appearances of people in a social network. In *Proc. WWW2005*, 2005.

[3] M. Bengtsson and S. Kock. Cooperation and competition in relationships between competitors in business networks. *Journal of Business & Industrial Marketing*, 14:178–194, 1999.

[4] D. Bollegara, Y. Matsuo, and M. Ishizuka. Disambiguating personal names on the web using automatically extracted key phrases. In *Proc. ECAI 2006*, 2006.

[5] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. In *Proc. 7th WWW Conf.*, 1998.

[6] T. Finin, L. Ding, L. Zhou, and A. Joshi. Social networking on the semantic web. *The Learning Organization*, 12(5):418–435, 2005.

[7] F. Gandon, O. Corby, A. Giboin, N. Gronnier, and C. Guigard. Graph-based inferences in a semantic web server for the cartography of competencies in a telecom valley. In *Proc. ISWC2005*, 2005.

[8] E. Glover, G. Flake, S. Lawrence, W. P. Birmingham, A. Kruger, C. L. Giles, and D. Pennock. Improving category specific web search by learning query modifications. In *Symposium on Applications and the Internet, SAINT*, pages 23–31, 2001.

[9] N. Kambhatla. Combining lexical, syntactic, and semantic features with maximum entropy models for extracting relations. In *Proc. ACL04*, pages 178–181, 2004.

[10] H. Kautz, B. Selman, and M. Shah. The hidden Web. *AI magazine*, 18(2):27–35, 1997.

[11] X. Li, P. Morie, and D. Roth. Semantic integration in text: From ambiguous names to identifiable entities. *AI Magazine Spring*, pages 45–68, 2005.

[12] P. Marsden. Measuring tie strength. *Social Forces*, 63:482–501, 1984.

[13] Y. Matsuo, M. Hamasaki, H. Takeda, J. Mori, D. Bollegala, Y. Nakamura, T. Nishimura, K. Hasida, and M. Ishizuka. Spinning multiple social networks for semantic web. In *Proc. AAAI-06*, 2006.

[14] Y. Matsuo, J. Mori, M. Hamasaki, H. Takeda, T. Nishimura, K. Hasida, and M. Ishizuka. POLYPHONET: An advanced social network extraction system. In *Proc. WWW2006*, 2006.

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

[16] T. Miki, S. Nomura, and T. Ishida. Semantic web link analysis to discover social relationship in academic communities. In *Proc. SAINT 2005*, 2005.

[17] J. Mori, M. Ishizuka, T. Sugiyama, and Y. Matsuo. Real-world oriented information sharing using social networks. In *Proc. ACM GROUP2005*, 2005.

[18] S. Oyama, T. Kokubo, and T. Ishida. Domain-specific web search with keyword spices. *IEEE Transactions on Knowledge and Data Engineering*, 16(1):17–27, 2004.

[19] N. V. Papakyriazis and M. A. Boudourides. Electronic weak ties in network organisations. In *4th GOR Conference*, 2001.

[20] G. Ramakrishnan, S. Chakrabarti, D. Paranjpe, and P. Bhattacharyya. Is question answering an acquired skill? In *Proc. WWW2004*, 2004.

[21] J. Scott. *Social Network Analysis: A Handbook (2nd ed.)*. SAGE publications, 2000.

[22] W. Souma, Y. Fujiwara, and H. Aoyama. Shareholding networks in japan. 776:298–307, 2005.

[23] H. Yeung. The firm as social networks: An organisational perspective. *Growth and Change*, 36:307–328, 2005.

[24] S. Zhao and R. Grishman. Extracting relations with integrated information using kernel methods. In *Proc. ACL 2005*, 2005.