

# Social Network Mining と人工知能

## Social network mining and artificial intelligence

松尾 豊\*<sup>1</sup> 濱崎 雅弘\*<sup>1</sup>  
Yutaka Matsuo Masahiro Hamasaki

\*<sup>1</sup>産業技術総合研究所  
AIST

Social networks play important roles in the Semantic Web: knowledge management, information retrieval, ubiquitous computing, and so on. We propose a social network extraction system called *POLYPHONET*, which employs several advanced techniques to extract relations of persons, detect groups of persons, and obtain keywords for a person. Several studies have used search engines to extract social networks from the Web, but our research advances the following points: Finally, a novel architecture called *Super Social Network Mining* is proposed; it utilizes simple modules using Google and is characterized by scalability and *Relate-Identify processes*: Identification of each entity and extraction of relations are repeated to obtain a more precise social network.

### 1. はじめに

情報検索や Semantic Web の分野で、社会ネットワークを抽出する多くの研究が行われている。近年では、mixi や Orkut といった Social Networking Service (SNS) が広まっており、人のネットワークを情報の流れの基盤に置く研究 [29, 26] は、これからの情報システムにおいて重要な方向性であろう。

著者らは、2002 年末から Web 上からの社会ネットワークの抽出の研究を行ってきた。2003 年度の人工知能学会全国大会では「人間関係ネットワーク支援システム」「スケジューリング支援システム」として、Web から得られた / ユーザ自ら登録した個人間のネットワークの表示を行い、その後、Polyphonet として 2006 年度の全国大会まで継続してサービスを行っている [33]。Web からの関係抽出は、もともとは 1997 年の H. Kautz らの Referral Web の研究 [13] にヒントを得て発展させたものであるが、近年では世界的に同様の研究が多く行われている。

本稿では、検索エンジンを用いて関係性を抽出する研究を広範囲にまとめて整理するとともに、そのプロセスを簡単なコードで書き下す。この研究は、簡単な実装で行うことができ（もちろんその後、さまざまなノウハウが存在するが）、Web を使った大規模な情報統合の研究が気軽にできる。知識表現や知識統合、情報の価値といった人工知能の根本的なテーマと関わっており [19]、今後さまざまな研究が展開する一助になれば幸いである。

### 2. 関連研究

Web からの Social Network Mining に関連する研究グループとしては、著者らの他に、いくつかの主要なグループがある。ここでは、人に着目して関連文献を紹介しよう。

MIT の A. McCallum らを中心としたグループでは、e-mail のメッセージの中から名前を見つけ、対応するホームページを見つけ、コンタクトアドレスを埋めるシステムを作っている [10]。この中で、他の人のホームページに現れる名前はその人と関係があるとしてネットワークを抽出する。最近のシステムでは、Web 全体の共起を調べており、同姓同名の解決の手法も提案している [5]。

オランダの Free 大学の P. Mika らは、Web 上の名前の共起関係や FOAF ファイルから社会ネットワークを抽出し図示する Flink ([www.semanticweb.org](http://www.semanticweb.org)) というシステムをつくっている。Google scholar を使って論文情報等も参照することができる。P. Mika らは、社会ネットワークに着目して、オントロジを抽出する手法を提案しており [22]、Flink が ISWC2004 で Semantic Web Challenge Award、オントロジ抽出が ISWC2005 ではベストペーパーを取るなど、この研究に対する Semantic Web コミュニティの評価も高い。Web 上から名前の共起関係を調べる部分では、著者らのアルゴリズムとほぼ同様の手法を用いているが、関係の種類や判別や同姓同名の問題など、複雑な処理は行っていない。

Maryland 大学の T. Finin らは、SWOOGLE (Semantic Web の検索エンジン) を使って集めた FOAF のネットワークの分析や、DBLP のデータと合わせて COI (Conflict of Interest: 利益相反) の検出を行っている [1]。社会ネットワークを用いた具体的な応用事例として興味深い。このように、複数のネットワークをどのように統合するか、それをどう用いるかは Semantic Web でも重要な課題のひとつである [20]。ドイツ Karlsruhe 大学の S. Staab らは、Web 上のテキストのパターンを用いて、エンティティ間のオントロジを抽出している [8, 9]。NTT 研究所の佐藤や原田らは、検索エンジンを用いてあるクエリーに関するページを取得し、その中で共起する人のネットワークを抽出している [34, 12]。

オントロジ抽出に関連して、自然言語処理の分野では、検索エンジンを使ったタキソノミーやレキシコンの構築が多く行われている。特に、A. Kilgarriff や M. Baroni (U. of Bologna) らは、Web as corpus という概念を提案し、継続的にワークショップを開催している [14]。英国 Sheffiled 大の Lapata らは、Web の検索ヒット数 (Web count) をさまざまな NLP タスクに適用し、そのうち 2 つのタスクで良い結果が得られたことを報告している [?]。Web の文書は、ある程度コントロールされた新聞記事などのコーパスと違って、その性質がはっきりしない、何の代表なのか分からないなどの問題があり、自然言語処理の分野では苦戦しているが、これからはますます重要な研究になるだろう。検索エンジンを使った言語処理の研究で最も興味深いもののひとつが、Turney (カナダ Natinal Research Council) らによる PMI の研究 [32, 30] である。TOEFL のシソーラスの同定問題を、Web の検索エンジンを用いることによって、平均的な学生を上回る精度で正解することを示してい

**Algorithm 3..1: GOOGLECOOC( $X, Y$ )**

```

氏名  $X, Y$  が与えられ, 共起の強さを返す
 $n_X \leftarrow GoogleHit("X")$ 
 $n_Y \leftarrow GoogleHit("Y")$ 
 $n_{X \wedge Y} \leftarrow GoogleHit("X Y")$ 
 $r_{X,Y} \leftarrow CoocFunction(n_X, n_Y, n_{X \wedge Y})$ 
return ( $r_{X,Y}$ )
    
```

図 1: *GoogleHit* を用いた共起の測定.

**Algorithm 3..2: GOOGLECOOCTOP( $X, Y, k$ )**

```

氏名  $X, Y$  が与えられ, 共起の強さを返す
 $D_X \leftarrow GoogleTop("X", k)$ 
 $D_Y \leftarrow GoogleTop("Y", k)$ 
 $n_X \leftarrow NumEntity(D_X \cup D_Y, X)$ 
 $n_Y \leftarrow NumEntity(D_Y \cup D_X, Y)$ 
 $n_{X \wedge Y} \leftarrow NumCooc(D_X \cup D_Y, X, Y)$ 
 $r_{X,Y} \leftarrow CoocFunction(n_X, n_Y, n_{X \wedge Y})$ 
return ( $r_{X,Y}$ )
    
```

図 2: *GoogleTop* を用いた共起の測定.

る. 日本では, 佐藤氏・宇津呂氏が検索エンジンを使った研究を行っている [28].

### 3. 基本的なアルゴリズム

#### 3.1 ネットワークの抽出

基本的なネットワーク抽出のアルゴリズムはシンプルである. まずノードを定め, 次にエッジを同定する. 我々の研究や Flink では, ノードは所与であるが, Referral Web や McCallum の研究では, ノードを順次増やしていく.

ノード間のエッジは, 検索エンジンを使って付与する. 例えば, 松尾豊と石塚満であれば「松尾豊 AND 石塚満」を検索エンジンにクエリーとして与え, そのヒット件数が多ければ関係が強い, そうでなければ関係が弱いとする. この関係の強さの測定には, さまざまな共起尺度を用いることができる. 我々は, Simpson 係数 (overlap 係数) を用いている [17] が, Flink や Referral Web では Jaccard 係数が用いられている. 共起の強さを測る処理を, アルゴリズムとして書き下すと図 1 となる.

この研究では, 検索エンジンを使った 2 つのモジュールを用いる.

*GoogleHit* クエリーの検索ヒット件数を得るものである.

*GoogleTop* クエリーにヒットした上位文書を得る.

共起の強さは, *GoogleTop* を用いた処理 (図 2) でも測ることができる. 実際, McCallum や原田らはこの方法で測っている. この方法は, 詳細なテキスト解析を用いることができるという利点があるが, スケーラビリティの観点からどこまで適用可能なのかが, 明確ではない.

社会ネットワークを抽出する処理を図 3 に示す. エッジの関係の強さを測定しながら, ネットワークを構成していく. なお, 図 4 のモジュールを用いれば, 氏名を拡大していくことができる. 既存の研究はほぼ, これらのモジュールの組み合わせで整理することができる.

#### 3.2 同姓同名の問題

Web 上には, 多くの人の情報があるので, 求めたい人物を正しく特定することが重要である. 例えば, Flink では, Semantic Web コミュニティを対象としているので「Semantic Web OR ontology」というクエリーを名前に付加している. 著者らの方法では, 所属名をクエリーに付与している.

**Algorithm 3..3: GETSOCIALNET( $L$ )**

```

氏名リスト  $L$  が与えられ, ネットワーク  $G$  を返す
for each  $X \in L$ 
do set a node in  $G$ 
for each  $X \in L$  and  $Y \in L$ 
do  $r_{X,Y} \leftarrow GoogleCooc(X, Y)$ 
for each  $X \in L$  and  $Y \in L$  where  $r_{X,Y} > threshold$ 
do set an edge in  $G$ 
return ( $G$ )
    
```

図 3: *GoogleCooc* を用いた社会ネットワークの抽出.

**Algorithm 3..4: EXPANDPERSON( $X, k$ )**

```

検索されたページから氏名の抽出.
 $D \leftarrow GoogleTop("X", k)$ 
 $E \leftarrow ExtractEntities(D)$ 
return ( $E$ )
    
```

図 4: 氏名の拡張.

Bekkerman らは, Web 上の同姓同名問題に対する確率モデルを提案している. Web ページはクラスターにまとめられ, 関連するものだけ用いられる [5]. LI らは, AI magazine で, 複数の文書における名前の同定の問題を取り上げ [16], その解決のため, 氏名が文書に散りばめられる generative model を提案している. 同姓同名問題は, 自然言語処理では word-sense disambiguation のひとつのクラスとして, さまざまな研究が行われている. 同じ名前の別人は, namesake と呼ばれ, namesake をいかに排除するかが鍵である.

著者らは, Bekkerman の手法を進展させ, 名前で検索してヒットした文書をクラスタリングし, 関連するクラスター (対象となる人物) を最も同定する検索語は何かを調べる方法を提案している [6]. この検索語を付与することで, よりの確にその人物が同定できるようになる. この人物を同定する検索語を加えたときの処理は, 図 5 である.

### 4. 拡張

Polyphonet では, 前節の基本的なネットワーク抽出以外に, 関係の同定, スケーラビリティ, 語と氏名の共起といった要素を扱っている. ここでは, これらを簡単に紹介する.

#### 4.1 関係の同定

三木らは, 研究者の関係をオントロジカルに考察している [23]. FOAF では, 知人であるという関係を記述することができるが, RELATIONSHIP[11] では, その他に 30 種類近くの関係を定義している. 一般的に人間関係とは何かを考えると, 人間関係の原始的な要因 (イベントへの共参画) に立ち返らなければならない [18].

関係の同定は, 基本的にはテキスト分類として扱うことができる [17]. 図 6 に処理を示す. テキスト分類は, 情報検索の分野で盛んに研究が行われており, 例えば, ラベルなしデータによる精度の向上 [25] など重要な技術である. なお,  $k$  を大きくすると再現率を上げることができるが, Web サイズの拡大に対して, 一部の文書を調べることで不十分になる. そこで, 検索クエリーに関係を同定する語を加えることが重要であ

**Algorithm 3..5: GOOGLECOOCCONTEXT( $X, Y, W_X, W_Y$ )**

```

氏名  $X, Y$ , 語  $W_X, W_Y$  が与えられ, 共起の強さを返す
 $n_X \leftarrow GoogleHit("X W_X")$ 
 $n_Y \leftarrow GoogleHit("Y W_Y")$ 
 $n_{X \wedge Y} \leftarrow GoogleHit("X Y W_X W_Y")$ 
 $r_{X,Y} \leftarrow CoocFunction(n_X, n_Y, n_{X \wedge Y})$ 
return ( $r_{X,Y}$ )
    
```

図 5: 同姓同名を解決するクエリーを加えた共起の測定.

**Algorithm 4.1:** CLASSIFYRELATION( $X, Y, k$ )

```

名前 X, Y が与えられ, 関係のクラスを返す
 $D_{X \wedge Y} \leftarrow GoogleTop("X Y", k)$ 
for each  $d \in D_{X \wedge Y}$ 
do  $c_d \leftarrow Classifier(d, X, Y)$ 
 $class \leftarrow determine\ on\ c_d \in D_{X \wedge Y}$ 
return ( $class$ )
    
```

図 6: 関係の分類 .

**Algorithm 4.2:** GETSOCIALNETSCALABLE( $L, k$ )

```

氏名のリスト L が与えられ, ネットワーク G を返す
for each  $X \in L$ 
do set a node in G
for each  $X \in L$ 
do {
 $D \leftarrow GoogleTop("X", k)$ 
 $E \leftarrow ExtractEntities(D)$ 
for each  $Y \in L \cap E$ 
do  $r_{X,Y} \leftarrow GoogleCooc(X, Y)$ 
for each  $X \in L$  and  $Y \in L$  where  $r_{X,Y} > threshold$ 
do set an edge in G
return (G)
    
```

図 7: スケーラブルな社会ネットワーク抽出 .

る . 企業間関係の抽出 [15] では, この技術が用いられている .

また, 関係をあらかじめ与えるのではなく, 教師なしで複数の関係を捉えようという試みも行っている [31] . コーパスを対象とした研究では, 教師なしの関係の同定の研究があるものの, Web を対象とした研究でこういった関係が現れるかは非常に興味深い .

#### 4.2 スケーラビリティ

スケーラビリティを考慮することは重要である .  $n$  人のリストであれば,  $O(n^2)$  の検索クエリーが必要である . 例えば 500 人のネットワークを抽出するには,  ${}_{500}C_2 = 74750$  回のクエリーが必要である ( Google API の検索数の制限は 1 日 1000 件である . ) そこで, [3] では, 検索クエリーを減らす方法を提案している . 基本的には,  $GoogleTop$  を用いてフィルタリングを行うことで,  $GoogleHit$  の回数を減らす . そのアルゴリズムを図 7 に示す .

#### 4.3 語と氏名の共起

氏名同士の共起ではなく, 氏名と語の共起を考えることもできる . 著者らは, 氏名と語の共起に着目した人物のキーワード抽出 [24], また語の共起の分布に着目した研究者の分類 [4] を提案している .

これらのアルゴリズムを図 8, 9 に示す . 複数の語と複数の氏名の共起は, 氏名-語の行列を構成する . これは, 社会ネットワーク分析では, アフィリエーション行列と呼ばれるものに相当し, 2 分グラフで表すことができる .

### 5. Super Web マイニング

社会ネットワーク抽出の精度を上げていくために, 我々は Super Web マイニング ( 仮 ) と呼ぶアーキテクチャを提案している .

- スケーラビリティ: 検索エンジンを使ったシンプルなモジュールでスケーラビリティを備える .

**Algorithm 4.3:** EXTRACTKEYWORDS( $X, k_1, k_2$ )

```

 $D \leftarrow GoogleTop(X, k_1)$ 
 $words \leftarrow ExtractWords(D)$ 
for each  $W \in words$ 
do  $score_W \leftarrow GoogleCooc(X, W)$ 
 $K \leftarrow \{W | score_W\ is\ top\ k_2\}$ 
return (K)
    
```

図 8: 人物のキーワード抽出 .

**Algorithm 4.4:** CONTEXTSIM( $X, Y, W_L$ )

```

氏名 X, Y と語のリスト  $W_L$  が与えられ, 類似度を返す
for each  $W \in W_L$ 
do {
 $a_W \leftarrow GoogleCooc(X, W)$ 
 $b_W \leftarrow GoogleCooc(Y, W)$ 
 $s_{X,Y} \leftarrow similarity\ of\ vectors\ a = \{a_W\}$  and  $b = \{b_W\}$ 
return ( $s_{X,Y}$ )
    
```

図 9: 2 人の人物のコンテキストの類似度 .

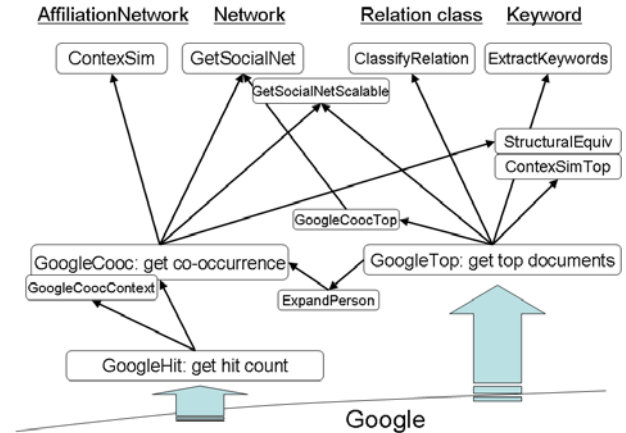


図 10: モジュールの依存関係 .

- Relate-Identify プロセス: エンティティの同定と関係の抽出が交互になり, 精度が向上する .

図 10 に,  $GoogleTop$  と  $GoogleHit$  を基盤としたモジュールの構成を示す . 簡単なモジュールの積み上げでいろいろな処理を行うことが可能である .  $ContextSimTop$  は  $GoogleTop$  を用いたコンテキスト類似度を表し [27],  $StructuralEquiv$  は構造同値による類似性である .

我々が特に注意しているのは, Web の規模が今の 10 倍, 100 倍になったときに有効なアルゴリズムであるかどうかである . 例えば, Web ページを一定数調べるという  $GoogleTop$  を使った処理は, Web の規模が 10 倍, 100 倍になったときに, 調べられる文書の量は相対的に 10 分の 1, 100 分の 1 になる . 一方で,  $GoogleHit$  を使った処理は, Web の規模が多くなればなるほど, 低頻度語の問題が解消され精度が良くなる . したがって,  $GoogleTop$  と  $GoogleHit$  をいかにうまく組み合わせるかが重要である . Web ページのサンプリングをうまく行う方法 [2] や NLP 独自の検索エンジンを作る方法 [7] も提案されている .

図 11 に, Relate-Identify プロセスを示す .  $ExtractKeyword$  モジュールでキーワードを得て, それを個人のメタデータとして用いることができる . RELATE ステップでは, 個人間の関係が抽出され, 最終的に 2 種類の行列 ( 隣接行列, アフィリエーション行列 ) が得られる . IDENTIFY のステップでは, 全体の関係性を使って個人を同定するクエリーが改良される . ここでは 2 つの可能性があり, ひとつのエンティティを複数のに分割する ( 同姓同名問題 ) , もしくは複数のエンティティをひとつにまとめる ( 愛称問題, 転職問題 ) ことである . この際, ネットワークを用いた構造同値の関係が重要な役割を果たす . 例えば, 2 つの名前が同じエンティティを指すのであれば, 他のエンティティとの分布は同様になるであろう .

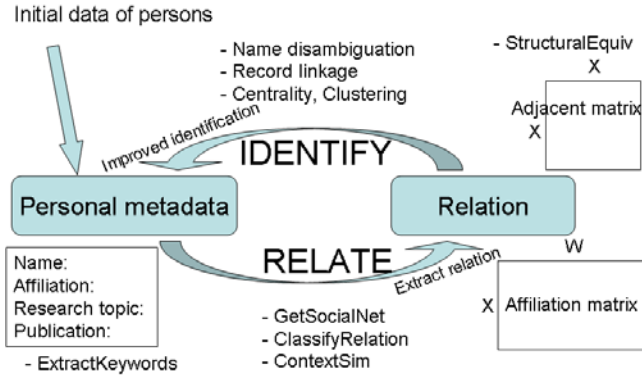


図 11: Relate-Identify プロセス .

## 6. おわりに

人工知能の分野では、エンティティ間の関係、概念間の関係をいかに記述し、それを推論に用いるかという研究が長い間行われてきた。一方、社会学の社会ネットワーク分析の分野では、構造主義からはじまる長い歴史の中で、関係性を総体として捉えたときにどのような分析を行うことができるのかという多くの知見が得られている。近年の複雑ネットワークの研究は、系全体に何らかの力が働いているときにスケールフリーやスモールワールドといった特徴的なネットワーク構造が得られることを示唆している。関係性が、全体としてのどのような意味を持つのか、関係の総体としてのネットワークが個々の関係性の認識にどのような影響を与えるのか、これは、人工知能、社会ネットワーク分析、複雑ネットワークをまたがった興味深い領域であろう。我々は、これを Web マイニングという題材を使って研究していきたいと考えており、その詳細については [19] で議論する。

## A 謝辞

本研究は、以下の各氏との共同研究、協力により行っている：森純一郎氏、ダヌシカ・ボッレーガラ氏、金英子氏、石塚満氏（東京大学）、武田英明氏（国立情報学研究所）、石田啓介氏、藤岡由季氏、西村拓一氏、橋田浩一氏（産業技術総合研究所）、松原仁氏、中島秀之氏（はこだて未来大学）。本研究の一部は、NEDO 産業技術研究助成事業により助成を受けて実施しているものである。ここに謝意を記す。

## 参考文献

[1] B. Aleman-Meza, M. Nagarajan, C. Ramakrishnan, A. Sheth, I. Arpinar, L. Ding, P. Kolari, A. Joshi, and Tim Finin. Semantic analytics on social networks: Experiences in addressing the problem of conflict of interest detection. In *Proc. WWW2006*, 2006.

[2] A. Anagnostopoulos, A. Z. Broder, and D. Carmel. Sampling search-engine results. In *Proc. WWW 2005*, pp. 245–256, 2005.

[3] 浅田洋平, 松尾豊, 石塚満. Web からの研究者ネットワーク抽出の大規模化. *人工知能学会論文誌*, Vol. 20, No. 6, 2005.

[4] Y. Asada, Y. Matsuo, and M. Ishizuka. A method to automatically find foaf:group based on the cooccurrence of people with keywords in the web. In *Proc. 1st Workshop on Friend of a Friend, Social Networking and the Semantic Web*, pp. 34–37, 2004.

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

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

[7] M. Cafarella and O. Etzioni. A search engine for natural language applications. In *Proc. WWW2005*, 2005.

[8] P. Cimiano, S. Handschuh, and S. Staab. Towards the self-annotating web. In *Proc. WWW2004*, pp. 462–471, 2004.

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

[10] A. Culotta, R. Bekkerman, and A. McCallum. Extracting social networks and contact information from email and the web. In *CEAS-1*, 2004.

[11] I. Davis and E. Vitiello Jr. RELATIONSHIP: A vocabulary for describing relationships between people. <http://vocab.org/relationship/>.

[12] M. Harada, S. Sato, and K. Kazama. Finding authoritative people from the web. In *Proc. Joint Conference on Digital Libraries (JCDL2004)*, 2004.

[13] H. Kautz, B. Selman, and M. Shah. The hidden Web. *AI magazine*, Vol. 18, No. 2, pp. 27–35, 1997.

[14] A. Kilgariff. Introduction to the special issue on the web as corpus. *Computer Linguistics*, Vol. 29, No. 3, 2003.

[15] 金英子, 松尾豊, 石塚満. Web 上の情報を用いた企業間関係の抽出. 第 70 回 知識ベースシステム研究会, 2005.

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

[17] 松尾豊, 友部博教, 橋田浩一, 石塚満. Web 上の情報からの人間関係ネットワークの抽出. *人工知能学会論文誌*, Vol. 20, No. 1E, pp. 46–56, 2005.

[18] 松尾豊, 武田英明, 森純一郎. 人間関係オントロジー. 第 10 回 セマンティックウェブとオントロジー研究会, 2005.

[19] 松尾豊, 山川宏. ネットワーク-予測性-属性生成. 第 20 回人工知能学会全国大会, 2006.

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

[21] Y. Matsuo, T. Sakaki, K. Uchiyama, and M. Ishizuka. Graph-based word clustering using web search engine. In *submitted*, 2006.

[22] P. Mika. Ontologies are us: A unified model of social networks and semantics. In *Proc. ISWC2005*, 2005.

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

[24] 森純一郎, 松尾豊, 石塚満. Web からの人物に関するキーワード抽出. *人工知能学会論文誌*, Vol. 20, No. 5, pp. 337–345, 2005.

[25] K. Nigam, A. McCallum, Sebastian Thrun, and Tom Mitchell. Text classification from labeled and unlabeled documents using em. *Machine Learning*, Vol. 39, pp. 103–134, 2000.

[26] 大向一輝, 松尾豊, 松村真宏, 武田英明. Community web プラットフォーム. *人工知能学会論文誌*, Vol. 3, No. 21, pp. 251–256, 2006.

[27] M. Sahami and T. Heilman. A web-based kernel function for measuring the similarity of short text snippets. In *Proc. WWW2006*, 2006.

[28] 佐々木靖弘, 佐藤理史, 宇津呂武仁. ウェブを利用した専門用語集の自動編集. *言語処理学会第 11 回年次大会発表論文集*, pp. 895–898, 2005.

[29] Jay Martin Tenenbaum. AI meets Web 2.0: Building the web of tomorrow today. In *Proc. AAAI05*, 2005.

[30] E. Terra and C. Clarke. Frequency estimates for statistical word similarity measures. In *Proc. HLT/NAACL 2003*, 2003.

[31] 辻下卓見, 森純一郎, 石塚満. Web からの関係情報の抽出. 第 5 回 Web インテリジェンスとインタラクション研究会, 2006.

[32] P. Turney. Mining the web for synonyms: PMI-IR versus LSA on TOEFL. In *Proc. ECML-2001*, pp. 491–502, 2001.

[33] 濱崎雅弘, 松尾豊, 中村嘉志, 西村拓一, 武田英明. 学会支援システムにおける実世界指向インタラクション. *日本知能情報ファジィ学会誌*, Vol. 18, No. 2, 2006.

[34] 佐藤進也, 風間一洋, 福田健介, 村上健一郎. 実世界指向 web マイニングの提案とその同姓同名人物分離問題への適用. *日本データベース学会 Letters*, Vol. 3, No. 4, pp. 21–24, 2005.