固有名の意味カテゴリの曖昧性解消における Wikipedia の利用に関する考察

村本　英明† 　鍛治　伸裕†† 　吉永　直樹†† 　喜連川　優††

† 東京大学大学院　情報理工学系研究科　〒 113–0033 東京都文京区本郷 7-3-1
†† 東京大学生産技術研究所　〒 153–8504 東京都目黒区駒場 4-6-1
E-mail: †{muramoto,kaji,ynaga,kitsure}@tkl.iis.u-tokyo.ac.jp

あらまし　ウェブテキストから製品や人物などに対する言及を抽出する際には，同一名が異なる実体に対して与えられることがあるため，固有名の曖昧性を解消する技術が必要となる．本稿では，テキスト中の固有名を予め定義された意味カテゴリに分類する分類器を，教師あり学習により構築することでこの問題の解決を図る．分類器の学習に用いる正解ラベル付きデータを，人手で作成するのにはコストがかかるため，Wikipedia の記事間リンクを用いて，正解ラベル付きデータを半自動的に生成する手法について検討する．

キーワード　固有表現分類，情報抽出

Coarse-grained Sense Disambiguation of Named Entities using Wikipedia

Hideaki MURAMOTO†, Nobuhiro KAJI††, Naoki YOSHINAGA††, and Masaru KITSUREGAWA††

† Graduate School of Information Science and Technology, University of Tokyo
7–3–1 Hongo, Bunkyo-ku, Tokyo, 113–0133 Japan
†† Institute of Industrial Science, University of Tokyo
4–6–1 Komaba, Meguro-ku, Tokyo, 153–8504 Japan
E-mail: †{muramoto,kaji,ynaga,kitsure}@tkl.iis.u-tokyo.ac.jp

Abstract　In mention extraction from Web text, named entity disambiguation technique is necessary, because the same name can be assigned to different entities. In this paper, we define semantic categories and construct a supervised classifier which classifies entity names into proper categories. Constructing labeled data for supervised learning by human annotation is costly, so, we generate labeled data semi-automatically by using hyperlinks in Wikipedia articles.

Key words　named entity categorization，information extraction

1. はじめに

ブログ等のウェブテキストには，多くの人々の意見や考えが書かれている．そこから製品や人物等についての言及を抽出することできれば，その評判や中心的話題を知るための手掛かりとなり，マーケティングをはじめとする社会分析に対する応用が期待される．

こうした固有物に関する言及抽出において，同じ名前が異なる固有物に対して与えられることが問題になりうる．例えば，「ライオン」という名前は化学メーカの「ライオン」と動物の「ライオン」の異なる 2 実体を指す．そのため，化学メーカの「ライオン」ついての言及をウェブから抽出したいとき，テキスト中の「ライオン」という単語が化学メーカか動物なのかを周辺のテキストを手掛かりとして識別する必要がある．

我々は**意味カテゴリ**を導入し，識別対象語を適切な意味カテゴリに分類することでこの問題の解決を図る．例えば，「法人」や「生き物」といった意味カテゴリを導入し，テキスト中の「ライオン」という語句を適切な意味カテゴリに分類することで，「ライオン」という単語がテキスト中で化学メーカの「ライオン」なのか動物の「ライオン」なのかの識別が可能になる．

こうした語句を意味カテゴリに分類する技術は，従来，人手でラベル付けしたデータを用いた教師あり学習による手法が用いられてきた [7] [3] [6]．高性能の分類器を構築するには多量のラベル付きデータが必要となるが，人手でそれを作成するのに

はコストがかかるため，十分な量のラベル付きデータを入手することが困難であることが多い．そのため，ラベル付きデータを廉価に生成する技術は必要不可欠である．

そこで，我々は Wikipedia から意味カテゴリが付与されたラベル付きデータを半自動生成する手法を提案する．具体的には，Wikipedia の記事間のハイパーリンクと，半自動生成した記事と意味カテゴリとの対応表を用いて，意味カテゴリがラベル付けされたデータを生成する．そして，得られたラベル付きデータを用いて，教師あり学習により分類器を構築する．

本稿の構成は下記の通りである．まず，関連研究について述べる．次に，Wikipedia から正解ラベル付きデータを半自動生成する手法，分類器を学習する手法について述べる．次に，分類器の評価実験について述べ，最後にまとめと今後の課題について述べる．

2. 関 連 研 究

テキスト中の語句に対して，語句横断的な意味カテゴリを付与する技術として，固有表現認識（Named Entity Recognition, NER）がある．NER では主に，人手でラベル付けしたデータを用いた教師あり学習による手法が高い性能を発揮している [7]. しかし，教師あり学習による手法は，ラベル付きデータの作成のコストが問題となることがある．例えば，意味カテゴリを変更する際には，ラベル付きデータを作り直さないといけないためコストがかかる．また，ウェブテキストのような多様なテキストにこうした技術を用いるには，多量のラベル付きデータが必要となることが想定されるため，それを人手で作成するコストは大きい．

[9] [10] [12] では，ラベル付きデータを自動生成することで，上記の問題の解決を図っている．これらの研究は，ラベル付きデータの自動生成に，gazetteer と呼ばれる固有表現とその意味カテゴリの対応関係を示した辞書を用いている．ラベルなしテキスト中の gazetteer とマッチした語句に対して正解ラベルを付与し，これをラベル付きデータとして用いている．これらの研究は，我々の研究と問題意識を少なからず共有しているが，ラベル付きデータの自動生成の際に言語依存のヒューリスティクスを用いているため，あらゆる言語にそのまま応用することはできない．

Wikipedia の記事間のハイパーリンクを手掛かりにラベル付きデータを自動生成した研究として [1] [2] がある．これらの研究は，分類対象語が Wikipedia 中にアンカーテキストとして出現することを前提としているため，Wikipedia 中に一度も出現しない語句に対する分類器を構築することができない．一方，我々の手法は，そうした語句も分類の対象として扱っている点が異なっている．

3. 提 案 手 法

本節では，意味カテゴリ分類器を廉価に構築する手法の概要について述べる．我々は，文中の語句を意味カテゴリに分類するタスクを扱う．ただし，分類対象語は予め与えられるものとする．

(1) a. シマウマの群れでは見張り役がいて，ライオンの接近を鳴き声や身振りで群れに知らせるという．

b. ブライトは，ライオンが 6 月に発売した白物衣類用の塩素系液体漂白剤の商品名である．

例えば，分類対象語を「ライオン」として，(1)a.，(1)b. の文が与えられたする．このとき，(1)a. の「ライオン」については意味カテゴリ「生き物」に分類し，(1)b. は意味カテゴリ「法人」に分類する分類器の構築を目指す．なお，意味カテゴリの定義については，3.1 で述べる．

意味カテゴリ分類器は教師あり学習により構築する．教師あり学習に用いるラベル付きデータは，Wikipedia から半自動生成を行う．具体的には，下記のようなラベル付きデータを半自動生成する．

(2) a. 多くの**魚<生き物>**は群れて行動する習性をもっている．

b. プリウスは，**トヨタ<法人>**が発売したハイブリッドカーである．

なお，(2)a. の「魚」については「生き物」の意味カテゴリが付与されており，(2)b. の「トヨタ」については「法人」の意味カテゴリが付与されている．

以下に提案手法の詳細について説明する．まず，導入した意味カテゴリについて述べる．次に Wikipedia からラベル付きデータを自動生成する手法について述べ，最後に分類器の学習手法について述べる．

3.1 意味カテゴリ

提案手法は，特定の意味カテゴリ集合を前提としないが，本節では実際に実験で用いた意味カテゴリ集合を例に議論を進める．我々は，Sekine らが設計した拡張固有表現 [13] を元に，具体物を対象とした意味カテゴリ集合を構成した．具体的には，時間表現・数値表現を除いた固有名に関する拡張固有表現階層において，第二層の各クラス (例: 組織名，地域名，材料名，自然現象名) を一つの意味カテゴリとみなした．また，拡張固有表現のどのクラスにも対応しない意味カテゴリ（例えば一般名詞）の分類先として,「その他」という意味カテゴリを準備した．表 1 が本研究で用いた 45 の意味カテゴリである．なお，意味カテゴリの導出に用いた拡張固有表現階層の詳細については，関根の拡張固有表現階層-7.1.0-(注1) を参照されたい．

表 1　拡張固有表現から構成した 45 の意味カテゴリ

人物，神，国際組織，公演組織，家系，民族，競技組織，法人，政治的組織，温泉，GPE (注2)，地域，地形，天体，遺跡，GOE(注3)，路線，製品その他，材料，衣類，貨幣，医薬品，武器，賞，勲章，罪，キャラクター，乗り物，食べ物，芸術作品，出版物，主義方式，規則，称号，言語，単位，催し物，事故事件，自然災害，元素，化合物，鉱物，生物，生物部位，動物病気，自然色，その他

(注1)：http://sites.google.com/site/extendednamedentityhierarchy/

3.2 Wikipedia を用いたラベル付きデータの半自動生成手法

本節では，Wikipedia からラベル付きデータを廉価に構築する手法について述べる．

3.2.1 Wikipedia の記事間リンク

[1] [2] と同様，Wikipedia の記事間リンクを手掛かりに，ラベル付きデータを半自動生成する．

Wikipedia のテキストは (3) のようなハイパーリンクにより結ばれている．

(3) a. 多くの [魚（動物）| 魚] は群れて行動する習性をもっている．

b. プリウスは，[トヨタ自動車 | トヨタ] が発売したハイブリッドカーである．

ただし，(3)a. の「魚」については記事「魚（動物)」にリンクが張られ，(3)b. の文中の「トヨタ」については記事「トヨタ自動車」にリンクが張られていることを示している．こうしたハイパーリンクにより記事と結びついた語句は，曖昧性が解消されているため，正解ラベルが付与されたラベル付きデータの自動生成の手掛かりとして用いることができる [2] [1].

記事と意味カテゴリとの対応関係があれば，意味カテゴリがラベル付けされたデータを自動生成することができる．例えば，記事「魚（動物)」と意味カテゴリ「生き物」，記事「トヨタ自動車」と意味カテゴリ「法人」が対応付けられているとする．この対応関係を用いることで，(3) のような Wikipedia のテキストから (2) のような意味カテゴリが付与されたラベル付きデータを自動生成することができる．本稿では，記事と意味カテゴリの対応関係を示した表を**記事-意味カテゴリ対応表**と呼ぶこととする．多量の記事に対して，意味カテゴリとの対応関係を一つ一つ人手で記述するのはコストがかかるため，記事-意味カテゴリ対応表を廉価に構築する手法が必要である．

3.2.2 記事-意味カテゴリ対応表の廉価な構築手法

記事タイトルの上位語を手掛かりに，記事-意味カテゴリ対応表を廉価に構築することができる．例えば，表 2 のような，記事タイトルの上位語が分かったとする．

表 2　記事タイトルの上位語と記事の対応表

記事タイトルの上位語	記事タイトル
化学メーカー	ライオン（企業），花王，資生堂，...
総合家電メーカー	シャープ（企業），ソニー，東芝，...
哺乳類	ライオン（動物），うさぎ（動物），...

記事タイトルの上位語と意味カテゴリとの対応ルールを記述することで，記事と意味カテゴリとの対応関係を記述するよりも少ないルールで，記事-意味カテゴリ対応表を構築することができる．例えば，記事タイトルの上位語「化学メーカー」が意味カテゴリ「法人」に属するというルールを人手で記述する．すると，表 2 を手掛かりに「ライオン（企業）」や「花王」,「資生堂」などが意味カテゴリ「法人」に属することが分かる．

記事タイトルの上位語は記事の最初の文から自動抽出できる [4]．Wikipedia の記事の最初の文には，その記事に対応する固有物の簡潔な定義が述べられれている．なお，最初の文のことを定義文と呼ぶ．例えば，記事「ライオン（動物)」,「ライオン（企業)」,「シャープ（企業)」の定義文はそれぞれ，(4)a.-c. のようになっている．

(4) a. ライオンはネコ目（食肉目）ネコ科ヒョウ属に分類される**哺乳類**である．

b. ライオン株式会社は，洗剤，石鹸，歯磨きなどトイレタリー用品，医薬品，化学品を手がける日本の**化学メーカー**．

c. シャープ株式会社は大阪府大阪市阿倍野区長池町に本社を構える**総合家電メーカー**である．

例えば，上記の定義文から,「ライオン（動物）」の上位語は「哺乳類」,「ライオン（企業）」の上位語は「化学メーカー」,「シャープ（企業)」の上位語が「総合家電メーカー」であることが分かる．

更に少ない対応ルールで，多くの記事-意味カテゴリ対応表やラベル付きデータを半自動生成するために，以下の 2 つの手続きを行う．まず,「化学メーカー」や「総合家電メーカー」,「電気メーカー」など主辞が同じ記事上位語に対して，一つ一つ意味カテゴリと対応付けていく作業は効率的でない．そこで，記事上位語の主辞に対して意味カテゴリとの対応関係のルールを記述するものとする．また，一つの対応ルールによって，できるだけ多くの訓練事例を得るために，Wikipedia 中でリンクされている頻度が高い記事に出現する上位語から順にルールを記述する．

3.3 分類器の教師あり学習

得られた訓練データから，教師あり学習に用いる特徴量を抽出する．特徴量は，正解ラベルが付与された語句の出現する文の bag of words と，前後 n-gram(n = 1〜3)，及び，係り受け先の動詞の原型を用いる．ただし，ストップワード（「こと」,「とき」,「する」等）は特徴量から削除する．また，全ラベル付きデータ中で 5 回未満しか出現しない特徴量は削除する．

機械学習のアルゴリズムには平均化パーセプトロン [8] を用い，一対他法 (one-versus-rest) で学習を行う．具体的には,「法人」の意味カテゴリについての学習を行う際には,「法人」カテゴリの訓練事例を正例とし,「法人」カテゴリ以外の全カテゴリの訓練事例は全て負例として 2 値分類器の学習を行う．我々が扱っている問題は分類先の意味カテゴリ数が 45 と多いため，1 対多法の学習において，正例のデータ数が負例のデータ数に比べて少ないことが分類器の性能上の問題となる．そこで，over sampling [5] を行うことで，この問題の解決を図る．

意味カテゴリ「その他」については，正例を準備するのが困難なため，45 カテゴリ全ての分類器が負例と判断した場合にのみ「その他」に分類するとする．また，分類対象語に対して，複数の意味カテゴリに対する分類器が正例と判断した場合は，分類器の出力したスコアが最も高い意味カテゴリに分類する．

4. 評価実験

4.1 実験の設定

Wikipedia の定義文から，記事上位語の抽出については，上位下位抽出ツール [4] を利用した．このツールはホームページ(注4)からダウンロードできる．

評価に用いるデータは，ラベル付きデータと同様に Wikipedia から作成する．まず，Wikipedia から 10 回以上リンクしている記事が 2 つ以上ある語句を抽出する．次に，記事を意味カテゴリに対応付ける．この作業は半自動では行わず人手で行った．30 語ランダムにサンプリングして評価に用いた．なお，評価に用いた語句は，学習器の訓練には用いないものとする．これは，新語，新用法に対しての頑健性を評価するためである．

4.2 実験結果

4.2.1 訓練データの半自動生成

評価実験では，600 個の記事上位語の主辞に対して，人手で意味カテゴリとの対応関係を記述した．600 個のルールを記述することで，281,188 個の記事と意味カテゴリの対応を得ることができた．意味カテゴリごとに得られるラベル付きデータ数は表 3 のようになった．

表 3　半自動生成されたラベル付きデータ数

意味カテゴリ	ラベル付きデータ数	意味カテゴリ	ラベル付きデータ数
人物	850859	賞	20842
神	14386	罪	6245
国際組織	3098	キャラクター	16647
公演組織	34158	乗り物	72835
家系	29623	食べ物	42699
民族	13053	芸術作品	380603
競技組織	76472	出版物	58790
法人	395581	主義方式	195007
政治的組織	181570	規則	55008
温泉	4294	称号	103713
GPE	1292878	言語	85028
地域	10616	単位	35909
地形	208570	催し物	84545
天体	15234	事故事件	99192
GOE	460652	元素	16887
路線	160564	化合物	48788
製品その他	171346	鉱物	5131
衣類	4730	生物	91740
貨幣	3869	生物部位	10962
医薬品	3040	動物病気	16503
武器	22568	自然色	5709
		合計	5409944

4.2.2 分類精度

評価実験の結果は表 4 のようになった．なお，表 4 の 2 行目のように正解意味カテゴリが,「生物 or 食べ物」と複数ある場合は，分類器の出力が「生物」,「食べ物」の両方の場合を正解としている．正答率のマクロ平均，マイクロ平均はそれぞれ，65.6%，71.3%となった．なお，マイクロ平均，マクロ平均は分類対象語と正解意味カテゴリとのペア（表 4 の 1 行に相当）を 1 つの試行として計算している．

この結果から，多くの事例では，ラアベル付きデータに含まれていない語句に対しても，頑健に動作していることが見て取れる．この結果から新語や新用法に対しても，頑健に動作することが期待できる．しかし，以下の 2 つの場合は，分類精度が悪い．

1 つ目は文脈情報では，正解意味カテゴリの判別が難しい場合である．

(5) a. **オアシス**の音楽から影響を受けた．
　　b. **モーツアルト**の音楽から影響を受けた．

例えば，上記の文中の「オアシス」が「公演組織」で,「モーツアルト」が「人物」であることを文脈情報のみを手掛かりに分類することは困難である．こうした語句を正しく分類するためには,「オアシス」が「公演組織」であり「人物」ではないという事前知識が必要だと考えられる．

2 つ目は正解意味カテゴリがその他の場合である．意味カテゴリ「その他」に対する分類結果（「石」,「少年」,「羞恥心」,「忍者」など）の正答率が他と比べて低いことが見て取れる．今回の分類手法では，各意味カテゴリに対する分類器全てが，負例と判定した場合にのみ，意味カテゴリ「その他」に分類しているが，この手法は改善の余地があることが分かる．なお，正解意味カテゴリが「その他」のみの事例を，評価から除くと正答率のマクロ平均，マイクロ平均はそれぞれ，69.5%，74.4%にまで上昇した．

5. おわりに

本稿では，語句横断的な意味カテゴリを導入し，Wikipedia から訓練事例を半自動生成することで，新語や新用法に対して動作する分類器を，廉価に構築することが可能なことを示した．

今後の課題としては，分類精度の向上のために，ウェブテキストから得られた情報を活用することが上げられる．NER のタスクにおいて，ウェブテキストから得られた情報を機械学習の特徴量に用いることで，分類精度が向上することが示されている [11]．我々のタスクにおいても，こうした手法が有効であるかどうかについて検証を進めていきたい．

一方，4.2.2 で，我々の手法では，分類対象語の事前知識を用いないことが分類精度の悪化の原因であることを述べた．我々は，新語，新用法に対して廉価に適応可能な分類器の構築を目指しているため，分類対象語の事前知識は，自動で行いたい．そこで，大量に手に入るウェブコーパスを知識源として，事前知識を獲得することでこの問題が解決を図る．ウェブコーパスからの事前知識の獲得手法，及び，その利用方法については，今後，検討していきたい．

また，今回の評価実験では，600 個の主辞に対して人手で意味カテゴリとの対応付けを行った．より少ない人手での作業で，分類器を構築することは，工学的に価値のあることである．そのため，人手での作業を減らした際に，分類器の精度がどう影響を受けるかについても検証を進めていきたい．

(注4)：http://alaginrc.nict.go.jp/hyponymy/index.html

表 4　評価実験の結果

分類対象語	正解意味カテゴリ	正答率	
オレンジ	GPE	9/10	90.0
オレンジ	生物 or 食べ物	102/104	98.1
オレンジ	自然色	0/22	0.0
レーダー	武器 or 製品その他	587/633	92.7
レーダー	神 or 人 or キャラクター	9/10	90.0
ボール	製品その他	105/146	71.9
ボール	その他	0/12	0.0
ボール	武器	25/26	96.2
レベル	その他 or 単位	4/92	4.3
レベル	法人	14/14	100.0
石	鉱物 or その他	25/127	19.7
石	単位	91/92	98.9
少年	出版物	11/12	91.7
少年	その他	1/148	0.7
スズキ	法人	294/303	97.0
スズキ	生物	36/39	92.3
サルサ	その他 or 芸術作品 or 主義方式	39/54	72.2
サルサ	食べ物	17/17	100.0
シカゴ	GPE	1050/1133	92.7
シカゴ	公演組織	20/22	90.9
長征	乗り物	0/12	0.0
長征	事故事件	17/59	28.8
ヴァルナ	GPE	45/45	100.0
ヴァルナ	神	12/13	92.3
エセックス	GPE	20/44	45.5
エセックス	乗り物 or 武器	19/20	95.0
羞恥心	その他	3/30	10.0
羞恥心	公演組織	1/93	1.1
羞恥心	芸術作品	19/21	90.5
ハイヒール	公演組織	16/21	76.2
ハイヒール	衣類	6/34	17.6
カサブランカ	GPE	60/70	85.7
カサブランカ	芸術作品	25/26	96.2
レンジャー	乗り物 or 武器	22/22	100.0
レンジャー	称号	6/32	18.8
ホーネット	乗り物 or 武器	52/52	100.0
ホーネット	製品その他	1/10	10.0
サラトガ	GPE	10/13	76.9
サラトガ	乗り物 or 武器	53/53	100.0
オセロ	製品その他	27/27	100.0
オセロ	公演組織	38/45	84.4
オセロ	芸術作品	22/27	81.5
ジャングル	芸術作品	11/11	100.0
ジャングル	地域 or 地形 or その他	24/51	47.1
ジャングル	その他 or 芸術作品 or 主義方式	13/14	92.9
オアシス	地域 or 地形 or その他	11/109	10.1
オアシス	公演組織	140/144	97.2
銀河	乗り物 or 路線	26/26	100.0
銀河	乗り物 or 武器	11/13	84.6
銀河	天体	162/169	95.9
クリーム	公演組織	64/67	95.5
クリーム	食べ物	33/54	61.1
ジェネシス	公演組織	4/46	8.7
ジェネシス	武器 or 乗り物	14/14	100.0
忍者	称号	6/193	3.1
忍者	公演組織	1/14	7.1
セオドア・ルーズベルト	人物	80/170	47.1
セオドア・ルーズベルト	乗り物 or 武器	9/14	64.3
タンポポ	公演組織	1/10	10.0
タンポポ	生物	24/31	77.4
安全地帯	公演組織	27/36	75.0
安全地帯	その他	0/41	0.0
イルカ	生物	112/205	54.6
イルカ	人物	39/41	95.1
ハンニバル	人物	82/90	91.1
ハンニバル	芸術作品	15/16	93.8
		マクロ平均：65.6，マイクロ平均：71.3	

文　　献

[1] Razvan Bunnescu, Marius Pasca “Using Encyclopedic Knowledge for Named Entity Disambiguation”, In Proceedings of EACL, 2006.

[2] Rada Mihalcea “Using Wikipedia for Automatic Word Sense Disambiguation”, In Proceedings of NAACL HLT, pp.196—203, 2007.

[3] Diana McCarthy “Word Sense Disambiguation:An Overview”, Language and Linguistics Compass3(2), pp.537—558, 2009.

[4] 隅田 飛鳥，吉永 直樹，鳥澤 健太郎　“Wikipedia の記事構造からの上位下位関係抽出”，自然言語処理 16(3), pp.3-24, 2009.

[5] Nitesh V.Chawla, Nathalie Japkowicz, Aleksander Kolcz, “Editorial: Special Issue on Learning from Imbalanced Data Sets””, SIGKDD Explorations6(1), pp.1-6, 2004.

[6] Massimiliano Ciaramita, Mark Johnson, “Supersense Tagging of Unknown Nouns in WordNet”. In Proceedings of EMNLP, pp.168-175, 2003.

[7] David Nadeau, Satoshi Sekine, “A Survey of Named Entity Recognition and Classification,” Journal of Linguisticae Investigationes30(1), pp.3-26, 2007.

[8] M. Collins, “Discriminative Training Methods for Hidden Markov Models: Theory and Experiments with Perceptron Algorithms”, In Proceedings of EMNLP, pp.1-8, 2002.

[9] Andrew Carlson, Scott Gaffney, Vasile Flavian, “Learning a named entity tagger from gazetteers with the partial perceptron” In AAAI Spring Symposium on Learning, 2009.

[10] Casey Whitelaw, Alex Kehlenbeck, Nemanja Petrovic “Web-scale named entity recognition””, In Proceedings of CIKM, pp.123-132, 2008.

[11] Jun’ichi Kazama, Kentaro Torisawa “Inducing Gazetteers for Named Entity Recognition by Large-Scale Clustering of Dependency Relations,” In Proceedings of ACL, pp.407-415, 2008.

[12] Ruihong Huang, Ellen Riloff “Inducing domain-specific semantic class taggers from (almost) nothing”, In Proceedings of ACL, pp.275-285, 2010.

[13] Satoshi Sekine, Kiyoshi Sato, Chikashi Nobata, “Extended Named Entity Hierarchy”, LREC-2002, 2002.