

固有表現抽出手法を用いたレストラン属性情報の自動認識

新里圭司[†] 関根聡[‡] 吉永直樹^{##†} 鳥澤健太郎[†]

[†] 北陸先端科学技術大学院大学 情報科学研究科

[‡] ニューヨーク大学 ^{##} 日本学術振興会

{skeiji, n-yoshi, torisawa}@jaist.ac.jp sekine@cs.nyu.edu

1 はじめに

本稿では、レストラン¹について書かれた文書中に現れる、レストランの“住所”や“ジャンル”，提供している“料理名”，店の“雰囲気”といったレストランの属性情報を自動的に認識/抽出する手法について述べる。近年ウェブ上では、個人レベルでの情報発信が盛んに行われており、レストランに限らず様々な事物に関する情報、意見/評価を得ることができるようになってきた。いわゆるカタログページ（レストランであれば“ぐるなび”など）と比べ、個人レベルで発信される情報には、(1) 事物の網羅性が高い（地元の隠れた名店や新店情報など）、(2) 事物の詳細な情報・ユーザの意見が得られる（立地場所、料理の質）といった利点がある。従って、個人レベルで発信されている情報を自動的に抽出できれば、非常に有用である。本研究では、個人レベルでの情報発信が盛んに行われているドメインの一つであるレストラン・ドメインに焦点を絞り、レストランの属性情報を自動抽出するシステムの開発を行った。

本研究ではレストランに関する属性情報の抽出を、ドメインを限定した一種の固有表現抽出タスクとみなし、既存の固有表現抽出手法 [4] を用いてその属性情報の抽出を行った。これにともない、まず約 60 種類のレストランの属性情報タグを定義し、ウェブから収集した文書に対してタグ付けを行い、属性情報タグ付きコーパスを作成した。属性情報の抽出には、タグ付きコーパスの不足を補助する目的で、大量の HTML 文書から自動的に構築した辞書を用いている。本稿では、レストラン属性情報タグ付きコーパスに加え、このドメインに特化した辞書の構築手法についても報告するとともに、構築された辞書がレストランドメインにおける属性情報抽出の精度向上に有効に働くことを実験により示す。

2 SVM を用いた固有表現抽出

ここでは、山田ら [4] による Support Vector Machine (SVM)[3] を用いて固有表現抽出を行う手法について簡単に述べる。この手法は実装が簡単なため、属性情報を抽出するためのモデルとして本研究でも採用している。

一般に固有表現抽出は、入力として与えられた文を適当な解析単位（トークン）に分割し、各トークンをあらかじめ定義された数種類の固有表現クラスもしくは固有表現以外のいずれかに分類する多値分類問題として解かれる。そのため、2 値分類器である SVM を使って固有

¹本研究では普通の意味での“レストラン”に限らず居酒屋や蕎麦屋など飲食店全般を指してレストランと呼ぶ。

<p>文 1 私の <料理><食材> 紅茶 </食材> のムース タルト </料理> も <料理の質> おいしかった♪ </料理の質></p> <p>文 2 <予約> できましたら、事前に予約を入れて頂くことにより、より満足頂けるサービスをご提供出来るかと存じます。フォームをご用意させて頂きましたが、お電話やFAXでもかまいませんので、よろしくお願ひ申し上げます。</予約></p>
--

図 1 レストランコーパスの例

表現抽出を行う場合は、SVM を多値分類に拡張する必要があり、山田らは Pairwise 法を用いてその拡張を行っている。あるトークンがどの固有表現クラスに属するかを判断する場合、前後に現れる数トークン、表層文字、文字種、品詞情報、固有表現タグといった情報が有効であることが知られている。最近では、これらに加え形態素解析結果の n 次解を用いる手法 [7] や、前後の文節を考慮する手法 [6] が提案されており、それらの素性が抽出精度の向上に役立つことが報告されている。

また、辞書を用いることで抽出精度が向上すると報告されているが [7, 6]、そのコストの高さから満足のいく辞書を準備することが難しいため、現状では既存のものを用いるのに留まっている。

3 レストラン属性情報タグ付きコーパス

本節では作成したレストランの属性情報タグ付きコーパスについて述べる。本稿ではこの属性情報が付与されたコーパスのことを単にレストランコーパスと呼ぶ。

コーパスの作成に先立ち、まずレストランの属性情報を定義した。具体的には、ウェブ上の掲示板サイトなどに寄せられているレストランに関する質問文を集め、その中で頻繁に尋ねられている事柄を中心に 64 種類のレストランの属性情報を定義し、各々の属性情報に対応するタグを設計した。表 1 に定義した属性情報とその例を示す。属性情報タグは大きく分けて、レストランに関するタグ、訪れた客に関するタグ、論理条件タグの 3 種類からなる。レストランに関するタグは、さらに住所や電話番号などのレストランの基本的な情報を表すタグ、レストランのジャンル、提供している料理やその質といった主に料理に関する情報を表すタグ、レストランの内装や設備、雰囲気などレストランの印象に関するタグなどに細かく分別される。

次いで、タグ付けの対象となる文書をウェブより収集した。具体的には、まず既存のレストラン検索エンジンを使って自由が丘周辺にあるレストラン名を収集した。そして収集したレストラン名を検索語として既存の検索エンジンに与え、検索結果として得られたウェブ文書の

表 1 本研究で定義したレストラン属性情報一覧

属性名	要素数	要素の例	属性名	要素数	要素の例
店名	736	吉野家, ドトール	営業日	397	不定休, 無休, 祝祭, 水～土
住所	248	目黒区自由が丘 X-X Y ビル B1F	営業時間	373	夜も遅くまで, 5時半より11時
電話	247		メディア	15	グルメジャーナル2004年5月号
FAX	51		経営者	99	下北沢“X”の初代マスターだった店長
URL	47		シェフ	76	本場上海から招いた点心職人
EMAIL	13		従業員	103	店員さんもとても親切
地域	251	東京, 自由が丘, 首都圏	予約・貸切	55	貸切可能, 当日予約で大丈夫
最寄路線	162	大井町線, 東急東横線	エンターテイメント	17	生演奏, 野球観戦ができる
最寄駅	230	田園調布駅, 自由が丘駅正面口	ドレスコード	2	普段着, 普段着のまま
最寄施設	79	六本木ヒルズ, みずほ銀行	クレジットカード	223	AMME X, カード使用不可
距離	212	徒歩8分, 歩いて10分くらい	身体障害者	0	
立地場所	68	素朴な商店街の中, 路面店	子供	39	子どもを遊ばせることができ
ジャンル	307	トスカナ料理, イタリアン	喫煙	5	禁煙, 喫煙可
料理	2,064	帆立と三つ葉の挟み揚げ, 日本酒	ペット	13	犬を連れて行ってもOK
料理の質	1,188	かなりヘルシーな感じでした, 辛い	その他のサービス	159	送迎バス有り
食材	974	旬の素材, 日本一のマグロ, 豚	価格設定	474	一人5000円
食事形式	261	フリードリンク, ディナー	評判・知名度	90	お客さんが入れ替わるが常に満席
品揃え	156	充実しています, 多数!!	ビジネススタイル	27	全国チェーン
形態	260	ファミレス, オープンカフェ風	客層	182	スポーツ好き, 主婦
歴史	81	老舗, 今年4月にオープンしました	その他の特徴	24	ベアカーが併設されている
雰囲気	364	穴場的, ちょっと小洒落た	プロフィール	47	友だち同士, 家族, 50代, 男性
内装	69	凝っている, 黒を基調とした空間だ	人数	107	5人, 数人
照明	23	暖かい灯り, 夜はキャンドルが光る	希望日・利用日	31	10月5日, 正月
清潔さ	16	厨房も清潔感がある, 清潔感に溢れ	利用目的	193	飲み会, 宴会, 送別会
静けさ	7	中はPUBならではの賑やかさ	評価・印象	140	足繁く通いたくなる店です
眺め	2	眺めがいい	もしくは	0	
BGM	26	ジャズ, 60～80年代の音楽	できることなら	1	
広さ	43	中も結構広いです, こじんまり	でも可	0	
席数	63	テーブル35席, 15席くらい	例えば	2	
設備	211	テラス, 席から見えるピザ窯	のような	2	
食器	22	真っ白な食器, 手作りのデカンタ	除外	124	
外観	43	かわいらしい一軒家, 看板が無い			
駐車	35	駐車場サービス券有, Pなし			

中から、検索語として与えたレストランについて書かれている部分を人手により抜き出した。これにより、自由が丘近辺に位置する118のレストランについて書かれた745文書(6,080文)が収集され、これらに対してタグ付けを行った。図1にレストランコーパスの例を示す。IREX日本語固有表現抽出タスク[1]で用いられているコーパスと比べると、(1)文1の<料理>、<食材>タグのようにタグがネストすることがある、(2)文2の<予約>タグのように複数の文に跨ってタグ付けされることがある、(3)単語に限らず、述語を含む文表現もタグ付けの対象となる、という点が異なる。

4 ウェブを用いた辞書の構築

2節で述べたように、固有表現抽出では辞書を用いることでその抽出精度の向上が期待できる。しかしながら、ある特定のドメインに特化した辞書を整備することは、コスト面からみて容易にできることではない。そこで本研究ではHTML文書において、箇条書きや表などのレイアウトに含まれている自然言語表現同士は意味的に類似しやすいという観察[8, 5]に従い、タグ付きコーパスと大量にダウンロードしたHTML文書を使ってドメインに特化した辞書を自動的に構築することを試みた。本節ではその構築手法について述べる。

まず、辞書を構築する前準備として、ウェブ上より大量のHTML文書をダウンロードし、同じレイアウトの要素となっている表現の集合を抽出した。以下、この同じレイアウトに含まれている表現の集合を**関連表現集合**と呼ぶ。本研究では、論文[5]に従いHTML文書中から**関連表現集合**を抽出した。本稿では紙面の都合によ

りこの抽出方法の詳細にはふれないが、簡単に説明すると、HTML文書中に現れる各表現の持つ**パス**に注目し、HTML文書中で同一のパスを持つ表現同士を**関連表現集合**として抽出している。

次いで抽出された**関連表現集合**から構築したい辞書のエンタリを獲得する。ここでは例として“食材”に関する辞書を構築することを考えよう。まず、コーパス中から“食材”としてタグづけされている表現を抽出する。この抽出された表現の集合を U とする。そして、この U 中の表現(つまり“食材”としてタグづけされている表現)を含む**関連表現集合**中に現れている表現を全て収集する。この収集された表現の集合を V とする。基本的には、この V に含まれている表現 v を辞書のエンタリと見なすが、 V の中には“食材”として考えにくい表現が多く含まれ得る。そこで、**箇条書きや表などのレイアウト**中で、**タグTでタグ付けされた多くの表現とだけ頻繁に共起する表現は、タグTに関する辞書のエンタリになりやすい**という仮説を設け、この仮説を反映した以下のスコアを導入した。

$$score(v) = \sum_{u \in U} P(v, u) \log_2 \frac{P(v, u)}{P(v) \cdot P(u)}$$

ここで $P(x)$ は、表現 x が**関連表現集合**に現れる確率であり、あらかじめ抽出された**関連表現集合**の総数を N とすると(表現 x を含む**関連表現集合**数/ N)で求めることができる。同様に、 $P(x, y)$ は表現 x, y が**関連表現集合**中で共起する確率である。上のスコアは v と U に含まれる全ての表現(“食材”としてタグ付けされてい

料理 八宝菜, エビフライ, 抹茶, ウスターソース*, インゲン*, モンブラン, キャラメル, ロールパン, 米*, うなぎ料理, ミネストローネ, 水稲*, 卵*, 鉄板焼, 豚汁, 鶏卵*, はちみつ*, 和え物
食材 砂糖*, にんじん, ビーマン, 牛乳, きゅうり, レタス, こしょう*, サラダ油*, なす, しょうゆ, ねぎ, 小麦粉*, 材料*, かぼちゃ, パセリ, じゃがいも, 大根, 人参, 白ワイン, ナス

*がついてる表現は望ましくないエントリ

図 2 構築された辞書の例

表 2 実際に構築された辞書のサイズ

属性名	構築した辞書のエントリ数	学習データ中での固有表現の異なり数
店名	450	173
地域	2,573	23
BGM	183	16
クレジットカード	212	27
利用目的	1,008	39
ジャンル	3,100	33
営業日	4,047	24
料理	18,667	416
食材	12,302	242
形態	1,428	67
食事形式	541	15
最寄路線	440	10
最寄施設	619	37
最寄駅	307	38

る表現) との間の相互情報量の和を求めていることになり, $P(u, v)$ が $P(u) \cdot P(v)$ より大きい表現 v , つまり関連表現集合中で U に含まれる表現とより共起しやすい v 程大きなスコアを得る. 全ての v について上記のスコアを計算し, 閾値 (本研究では 0.00001 としている) 以上の v を辞書のエントリとした.

5 実験

5.1 準備

実験に先立ち, 文を跨いでタグ付けされているタグ, 論理項目に関するタグ, レストランコーパス中での頻度が 10 回未満のタグについては学習が難しいと考え, コーパスから削除し, 今回は評価の対象としなかった. 上記のタグを削除した上で, コーパスから 24 軒のレストランに関するデータ (138 文書, 1,124 文) を無作為に抜き出し評価用のデータとした. そして, 残りの 94 軒分のデータ (607 文書, 4,932 文) を抽出規則を学習するための訓練データとした.

属性情報抽出の際の解析単位は形態素とし, 形態素解析器には MeCab² を使った. その際, “土日” のようなタグづけされた表現と形態素区切りとの不一致が 100 個見つけた. しかし, 全属性情報数 (11,386 個) と比べると十分に少ないため, 今回はそれらを全て無視し評価の対象から除外した. SVM へ与える素性としては, 前後の文脈 (2 単語), 単語 (基本型), 文字種³, 品詞情報, 4 節で述べた手法により構築された辞書の情報を用いている. 辞書素性は, 入力文の先頭から最長文字列マッチングを行い, 浅原ら [7] にならい, 1 つの形態素の場合は S タグを, 複数の形態素の場合は, 先頭の形態素に B タグ, 末尾に E タグ, 中間に I タグをそれぞれ辞書名と共に付与した. SVM の多値分類への応用手法は, 本研

²<http://chasen.org/~taku/software/mecab/>

³アルファベット, 数字, 平仮名, 片仮名, 漢字, これら以外の 6 種類を定義し, これらの組み合わせを考慮した.

究では抽出対象となる属性の種類が多いため, 生成する分類器の数が少なく済む One vs. Rest 法を採用した. One vs. Rest 法を用いたこと以外は, 山田ら [4] の実験で最も高い性能が得られたモデルに従っている. つまり, カーネル関数には 2 次の多項式カーネルを, 固有表現の状態の表現には IOB2 を用いており, 解析の方向は「文末から文頭」に固定して実験を行った.

次にレストランドメインに特化した辞書を構築するため, ウェブより収集した 1.00×10^7 件の HTML 文書 (103GB, タグ付き) から関連表現集合を抽出した. その結果, 3.01×10^6 個の関連表現集合 (重複なし, 含まれている表現の異なり数は 8.98×10^6 個) が得られた. そして表 2 に示す属性情報について辞書の構築を行った. 実際に構築された辞書の例を図 2 に示す. 表 2 にあげた属性タグは, タグで囲まれた表現が名詞句になりやすく, さらにコーパス中での出現頻度が割と高く, 精度の向上に繋がりやすいものという基準で選んだ.

5.2 HTML 文書から自動獲得した辞書の効果

辞書のサイズによりレストラン属性情報の抽出精度がどのように変化するかをみるため, 辞書を素性として用いない場合, 辞書に含まれる表現のうち, スコアの上位 10%, 20%, 50%, 100% にあたるエントリを用いた場合のそれぞれについて属性情報抽出実験を行った. その結果を表 3 に示す. 表 3 の「平均」の行は表に挙げた 14 種類の属性タグの精度, 再現率, F 値の平均値であり, 「全体」の行は全ての属性タグを考慮した場合の精度, 再現率, F 値の値である. 表より, スコアが上位 30% のエントリまでを用いた場合が最も F 値が高いことがわかる. その F 値は, 辞書を導入した属性情報に限った場合についてみると 56.9 であり, 辞書を素性として用いることで 4.4 向上していることがわかる. また全ての属性情報についてみた場合の F 値は 48.7 であり, 辞書を用いない場合と比べると 2.4, やはり向上している.

その一方で“クレジットカード”と“営業日”に関しては, 辞書を導入しても性能の向上が見られなかった. これらの属性情報の値はあまりばらつきがないため, 辞書を用いるまでもなく表層文字列から簡単に推定できると考えられる. そのため, ある表現が“クレジットカード”もしくは“営業日”かどうか認識する際に, 辞書を用いたことで辞書素性が必要以上に重要視されてしまい, 抽出精度が下がったのではないかと考えられる.

さらに表より, 各属性情報の抽出精度が最大となる辞書のサイズにばらつきがあることから, 各属性情報ごとに用いる辞書のサイズを変更することでより高い精度で抽出が行えると考えられる.

5.3 入れ子を除いた場合の抽出精度

レストランコーパスではタグの入れ子を許しているが, 今回作成した情報抽出器では 1 つの形態素に対して唯一のラベルしかふることができない. そのため, 入れ子がどの程度悪影響を及ぼしているのかをみるため, コーパス中からタグの入れ子を含んでいる文を全て除いた上で, 再度抽出器の学習, 評価を行った. 抽出器としては, 先程の実験で最も高い F 値を示した辞書エントリを上位 30% まで用いるモデルを用いた. 入れ子を除いた場合の属性情報抽出の結果を表 4 に示す. 入れ子を除くことで

表 3 辞書のサイズを変化させた時の属性情報の抽出精度

属性名	固有表現の出現数	なし			TOP 10%			TOP 30%			TOP 50%			TOP 100%		
		Prec.	Recall	F 値	Prec.	Recall	F 値	Prec.	Recall	F 値	Prec.	Recall	F 値	Prec.	Recall	F 値
店名	124	56.1	29.8	39.0	54.7	23.4	32.8	62.1	29.0	39.6	68.4	31.5	43.1	67.3	28.2	39.8
地域	44	54.6	54.6	54.6	64.3	61.4	62.8	62.2	63.6	62.9	63.6	63.6	63.6	65.2	68.2	66.7
BGM	3	0.00	0.00	0.00	50.0	33.3	40.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
カード	47	95.7	93.6	94.6	93.6	93.6	93.6	93.6	93.6	93.6	91.7	93.6	92.6	93.6	93.6	93.6
ジャンル	68	66.7	35.3	46.2	68.6	35.3	46.6	70.6	35.3	47.1	61.8	30.9	41.2	67.7	33.8	45.1
利用目的	47	46.7	29.8	36.4	41.2	29.8	34.6	45.2	29.8	35.9	50.0	29.8	37.3	52.2	25.5	34.3
営業日	76	85.1	82.9	84.0	84.0	82.9	83.4	82.7	81.6	82.1	81.1	79.0	80.0	81.6	81.6	81.6
料理	282	45.3	46.1	45.7	51.6	53.2	52.4	53.2	59.2	56.0	49.2	56.4	52.6	49.1	57.1	52.8
食材	210	47.1	19.1	27.1	54.6	28.1	37.1	54.5	26.2	35.4	51.6	23.8	32.6	51.9	26.2	34.8
形態	40	69.7	57.5	63.0	70.3	65.0	67.5	57.5	67.5	62.1	62.8	67.5	65.1	60.9	70.0	65.1
食事形式	58	80.0	62.1	69.9	80.9	65.5	72.4	79.2	65.5	71.7	80.4	63.8	71.2	76.9	69.0	72.7
最寄路線	36	91.2	86.1	88.6	89.2	91.7	90.4	89.2	91.7	90.4	91.9	94.4	93.2	91.2	86.1	88.6
最寄施設	11	40.0	18.2	25.0	33.3	18.2	23.5	25.0	9.10	13.3	33.3	18.2	23.5	50.0	18.2	26.7
最寄駅	46	81.8	78.3	80.0	74.0	80.4	77.1	69.2	78.3	73.5	75.5	80.4	77.9	80.4	80.4	80.4
平均	1,092	60.7	46.2	52.5	63.3	50.1	55.9	63.2	51.7	56.9	62.2	50.5	55.8	62.6	51.3	56.4
全体	1,973	53.2	41.0	46.3	54.8	43.5	48.5	54.4	44.0	48.7	53.8	43.3	48.0	54.6	43.7	48.6

表 4 入れ子を除いた時の属性情報の抽出精度

属性名	入れ子あり		入れ子なし	
	固有表現の出現数	F 値	固有表現の出現数	F 値
店名	124 (600)	39.6	115 (522)	42.1
地域	44 (203)	62.9	34 (160)	59.4
BGM	3 (23)	0.00	3 (20)	0.00
カード	47 (178)	93.6	47 (178)	94.6
ジャンル	68 (238)	47.1	19 (160)	47.1
利用目的	47 (145)	35.9	35 (121)	42.6
営業日	76 (291)	82.1	44 (199)	89.4
料理	282 (1,733)	56.0	192 (1,204)	56.3
食材	210 (753)	35.4	114 (366)	37.3
形態	40 (218)	62.1	22 (104)	45.5
食事形式	58 (198)	71.7	55 (180)	69.9
最寄路線	36 (124)	90.4	31 (122)	92.1
最寄施設	11 (62)	13.3	11 (55)	22.2
最寄駅	46 (170)	73.5	45 (156)	70.3
平均	1,092 (4,936)	56.9	799 (3,547)	58.1
全体	1,973 (9,280)	48.7	1,534 (6,956)	50.4

() 内は学習データ中の固有表現の出現数

学習データの量が約 25%減っているにもかかわらず、全体的に精度が向上しており、「全体」でみると F 値で 1.7 ポイント高い。しかしながら、入れ子を除いたことにより F 値が下がっているタグも幾つか見受けられる。形態タグに関しては精度の減少が顕著に見られる。これは、入れ子を除くことにより学習データの量が減ってしまったため（正例の数が 218 個から 104 個に減少）、十分な学習が行えなかったことが原因であると考えられる。

6 おわりに

本稿では、良質なレストラン検索サイト構築の準備段階として、レストランについて書かれた文書から、レストランの“雰囲気”や“ジャンル”，提供される“料理名”といったレストランの属性情報を自動的に抽出するシステムについて述べた。また固有表現抽出には有効であるとされているながらもそのコストの高さから敬遠されがちであった辞書の構築を、簡単な方法で自動的に行う手法を提案した。そして、この手法により構築された辞書を用いることでレストラン属性の抽出精度を改善できることを実験により示した。

残念ながら現時点での精度は決して高いとは言えないため、精度の向上が当面の課題になる。精度向上のためには、辞書およびコーパスの拡充が必要である。辞書の

拡充に話を限れば、今回提案した辞書構築手法はとても単純なものであるため、他の技術と組み合わせることで辞書のブラッシュアップは可能だと考えている。例えば論文 [2] にあるように、ブートストラップ手法を組み込むことで、辞書のカバレッジを広げることが可能であろう。その他の課題としては、多くの単語からなる長い固有表現を精度良く認識／抽出する手法についても検討しなくてはならないと考えている。IREX が提供する固有表現タグ付きコーパスでは、1 つないし 2 つの単語からなる固有表現が大部分を占めていたが、レストランコーパスにおいては、複数の文節からなる長い固有表現も少なくない。このような固有表現を認識するためには、今までに提案されているシーケンシャルなデータにラベル付けをする手法ではなく、構文木のノードにラベルを割り当てるような手法の方が適していると考えられる。最後に別の課題として、タグの入れ子が許されている場合の固有表現抽出モデルについても検討する必要があると思われる。

参考文献

- [1] IREX 実行委員会（編）. IREX ワークショップ予稿集, 1999.
- [2] Michael Thelen and Ellen Riloff. A Bootstrapping Method for Learning Semantic Lexicons using Extraction Pattern Context. In *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP 2002)*, pp. 117–124, 2002.
- [3] Vladimir Vapnik. *The Nature of Statistical Learning Theory*. Springer, 1995.
- [4] 山田寛康, 工藤拓, 松本裕治. Support Vector Machine を用いた日本語固有表現抽出. 情報処理学会論文誌, Vol. 43, No. 1, pp. 44–53, 2002.
- [5] 新里圭司, 鳥澤健太郎. HTML 文書からの単語間の上位下位関係の自動獲得. 自然言語処理, Vol. 12, No. 1, pp. 125–150, 2005.
- [6] 中野桂吾, 平井有三. 日本語固有表現抽出における文節情報の利用. 情報処理学会論文誌, Vol. 45, No. 3, pp. 934–941, 2004.
- [7] 浅原正幸, 松本裕治. 日本語固有表現抽出におけるわかち書き問題の解決. 情報処理学会論文誌, Vol. 45, No. 5, pp. 1442–1450, 2004.
- [8] 野口正人, 廣川佐千男. Web からの同系統単語知識獲得についての実験. 情報処理学会第 65 回全国大会講演予稿集 第 5 分冊, pp. 223–226, 2003.