

ブログに記述された不満表現からの潜在ニーズの発見

坂井 俊之^{†1} 藤村 考^{†1}

本研究では、ユーザが記述したテキストから、ユーザの不満を解決する商品を見出すことを目的としている。そのために、ユーザの実体験が記述されたブログのテキストを大量に収集し、その中から不満表現の抽出と、商品と不満との関係の抽出を行なう。この実体験に基づいた知識を利用することで、ユーザの様々な不満に対する商品の提示が可能となる。そして、これら不満表現抽出と関係性抽出について検討した結果、それぞれの抽出において、「解消」、「防止」等の、不満をなくす、あるいは、防ぐといった意味を持つ動作である「改善動作」との共起情報を利用することで、精度よく抽出できるという知見を得た。この知見に基づき、ユーザの不満に対して、その不満を解決する商品を見出すシステムのプロトタイプを作成した結果、いくつかの不満に対して妥当な商品を見出すことができた。

Discovering Latent Solutions from Expressions of Dissatisfaction in Blogs

TOSHIYUKI SAKAI^{†1} and KO FUJIMURA^{†1}

This paper aims to find the techniques or goods that solve problems in texts created by many users. We collected a large number of texts describing user experiences to extract expressions of dissatisfaction (indicative of problems). These texts can also contain information about the effectiveness of techniques or goods that address the dissatisfaction. We found that the co-occurrence frequency of words that indicate problem prevention or solution, such as “protect” and “cure”, is an effective technique for realizing this extraction. We implemented a prototype system on the proposed method and tested it. The results show that the system did indeed identify some useful goods for problem solving.

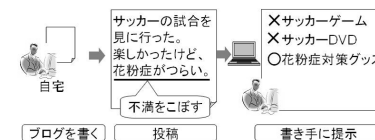


図1 書き手に対する広告配信
Fig. 1 Advertisement to the writer

1. はじめに

近年、インターネットの普及と検索技術の向上により、消費者の購買行動に対し、インターネット上の情報や広告が大きな影響を与えるようになってきた。インターネット広告の代表としては Adwords 広告や Adsense 広告がある。その中で、Adsense 広告は、ウェブコンテンツの閲覧者に対し、閲覧しているコンテンツのトピックに関連した広告提示を行っており、これは非常に有益である。ただし、Adsense 広告の対象は主にコンテンツの閲覧者であり、コンテンツの書き手に対する広告ではない。本研究では、書き手に対する広告の提示を考えた場合、ライフログの一種である、ブログやメール等の日々の記録が記述されたテキストコンテンツを詳細に解析することで、より効果的な広告を配信できるのではないかと考える。例えば、以下のような例を考える。

昨日、サッカーの試合を見に行った。自分の応援していたチームが 2-0 で勝ったこともあり、とても楽しかったが、花粉症がつかった。

上の文章はサッカーのトピックであり、トピックをマッチングする方式では、サッカーゲームや、サッカーの試合の DVD 等が提示されると思われる。しかし、文章をよく見ると、最後の「花粉症がつかった」という不満からは、「花粉症対策グッズ」という潜在ニーズが読み取れる。本研究ではこの潜在ニーズに即した商品広告を提示することで、効果的な広告配信が可能と考えている。まず、そこで、本研究では、コンテンツから書き手の潜在ニーズを見出すことを目的とする。ただし、ここで述べている「潜在ニーズ」とは、コンテンツの書き手に必要と考えられる商品やサービスであり、コンテンツには明示的に言及されていないものを指すものとする。上に示したコンテンツの書き手が言及した不満を解決するものはその一例である。ただし、当面の目標としては、このタイプの潜在ニーズにフォーカスし、コンテンツに記述された不満を解決する商品を見出すことを目指す。

不満を解決する商品を見出すにあたり、本稿では、Web 上に記述された人々の実体験

^{†1} 日本電信電話株式会社 NTT サイバーソリューション研究所
NTT Cyber Solutions Laboratories, NTT Corporation

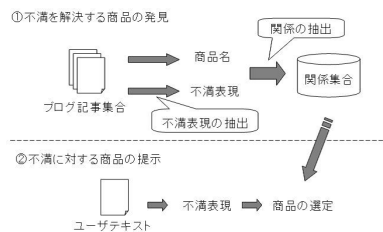


図 2 システム概要
Fig. 2 Outlines of the system

を利用する。つまり、ユーザが記述した Web テキストの中には「花粉症が辛い」といったような不満に対し、既に何らかの解決策にたどり着いた人が発信した情報も存在するため、この情報を抽出することで不満を解決する商品を見ることができるということである。まず、本稿では、不満を解決した人が発信した情報が記述された CGM (ブログ) から { 不満, 不満を解決する商品名 } という関係性を発見する。そして、発見した関係性を用いて、不満は持ってはいないが解決策を持たない人に対して適切な商品提示を行うシステムを構築する。

本稿では、商品名が既知であるとした上で、「不満表現の抽出」と「関係の抽出」について検討を行なった。その結果、「解消」や「防止」等の不満をなくす、あるいは、防ぐといった意味を持つ動作である「改善動作」の共起を用いることが、不満表現の抽出と関係性の抽出に有効であるという知見を得ることができた。また、この改善動作を用い、2008 年 5 月 11 日から 2008 年 8 月 26 日までの健康商品に関するブログ記事 (35,267 件) を分析した結果、不満を解決する商品として、妥当な商品を見えることを確認した。

以下に、本稿の構成を示す。まず、2 章では、本研究における関連研究について述べる。次に、3 章では、「不満表現の抽出」と「関係の抽出」に関してアプローチの詳細を述べる。4 章では実験結果について述べる。最後に、5 章でまとめを述べる。

2. 関連研究

まず、ユーザに合った商品や情報を推薦するための技術として、ユーザの興味に基づいたユーザプロファイリング技術に関しては、既に多数の先行研究が存在する¹⁾。しかし、本研究における、ユーザの不満に対する商品の発見のような研究はほとんど無い。

また、本研究で扱う不満のように、Web 上のユーザの声を抽出する研究として、ユーザが記述した意見を抽出する研究がある²⁾。また、意見の中から、更に評判や要望等に特化して

抽出を行なう研究がある³⁾⁴⁾⁵⁾。これに対し、本研究では不満表現に特化して抽出を行なう。

関係の抽出については、単語の出現確率等をベースにしたマイニング技術や、関連語の抽出、あるいは、言語パターンを用いて“is-a”、“cause”、“effect”、“part-of”等の特定の関係を抽出する研究等がある⁶⁾⁷⁾⁸⁾⁹⁾。この中でも、文献 9) は、モノとその正常な利用や楽しみ方の妨げになるトラブルの関係として、“object-trouble pair”の抽出を行なっている。これに対し、本研究では「不満-不満を解決する商品」の抽出を行なう。

3. アプローチ

3.1 処理の流れ

不満を解決する商品を見える際の手順は以下になる。(図 2)。

手順 1 ブログ記事集合から商品名と不満表現を抽出する。

手順 2 抽出した商品名と不満表現の組み合わせの中から { 不満, 不満を解決する商品名 } という関係を抽出し、データベースに保存する。

手順 3 データベースに保存された不満表現が記述されており、かつ、商品名が記述されていない (解決策が無い) ブログ記事に対し、不満表現と関係のある商品名を提示する。次節からは、「不満表現の抽出」と「関係の抽出」について、アプローチの詳細を述べる。

3.2 不満表現の抽出

本研究では、不満表現とそれを解決する商品とを結びつけることを目的としている。そのため、本稿では不満表現を「商品が解決する対象であり、かつ、ネガティブな意味を持つ名詞 (複合名詞)」と定義する。このような不満表現は以下のようなタイプの文として記述されることが多い。

商品 A は花粉症 (不満表現) を解消 (動作) してくれます (モダリティ)。

本稿では、このタイプの文に注目して不満表現の抽出を行なう。上記のタイプの文は「不満表現」「動作」「モダリティ」の 3 要素で構成される。このうち「動作」は動作の受け手にとって利益性が高く、「不満をなくす、あるいは防ぐ」意味を持っている。以降、この動作を改善動作と呼ぶ。また、3 要素間の関係として以下の 2 つが成り立つ。

- 「不満表現」は「動作」の対象である。
- 「モダリティ (してくれる)」は「動作」が受け手にとって利益性が高いことを示す。後者については「あげる」「もらう」「くれる」等のモダリティによって、恩恵構文と呼ばれる、利害を表わす構文が確立されている (文献参照)。そこで、本研究では、これら 2 つの関係の内、特に恩恵構文に着目し、以下のように不満表現を抽出する。

手順1 少数の不満表現サンプルと利益を表わすモダリティとの共起情報から改善動作を抽出する。

手順2 抽出した動作名詞の直前の10形態素から、動作名詞の対象であり、かつ、ネガティブな名詞を人手で抽出する。

手順1について詳細を述べる。手順1では、まず以下に示す式に従い、動作名詞の利益性の高さ P_{kaizen} を算出する。

$$P_{kaizen} = P_{human} \cdot P_{modality} \quad (1)$$

ここで、 P_{human} は不満表現サンプルと動作との共起確率であり、 $P_{modality}$ は利益性を表わすモダリティと動作との共起確率である。次に、 P_{kaizen} が高い順に、「不満をなくす、あるいは防ぐ」意味を持った改善動作を人手で選定する。不満表現はこの改善動作の対象から人手で抽出する。

3.3 関係の抽出

{不満, 不満を解決する商品} という関係を抽出するにあたり、まず、テキスト中に記述された商品と不満の関係を分類する。分類としては以下のようなものが存在する。

- (1) 商品と不満に関係が無く、別の話題として記述される (例. この前インフルエンザ [不満] にかかった。やっと治ったので商品Aで遊んだ。)
- (2) 商品の購入 (使用) 理由として不満が記述される (例. 運動不足 [不満] 解消のために商品Aを買った。)
- (3) 商品購入後 (使用後) に解決した不満として記述される (例. 商品Aで運動不足 [不満] を解消した。)
- (4) 商品購入後 (使用後) に新たに発生した不満として記述される (例. 商品Aで筋肉痛 [不満] になった。)

(1)~(4) の内 {不満, 不満を解決する商品} という関係になるのは、(2) と (3) である。そこで、(1) と (4) の関係を排除し、(2) と (3) の関係のみを抽出する方式について検討を行なった。以下に詳細を述べる。

本稿では、上記の関係を表わす言語パターンを網羅することが困難であると考え、関連語の抽出において頻繁に用いられる共起情報によって、上記の関係の抽出を行なう。関連語の抽出においては、一般的に2つの語の共起頻度が高い程、2つの語の関係性は強いとされ、これは本稿における商品と不満の関係にも当てはまると考えられる。そこで、まず、商品と不満の共起頻度の高さによって、商品と不満が別の話題として記述されているような(1)の関係を排除する。しかし、それだけでは4のような、「関係はあるが、商品が不満を解決し

ていない」場合に対応することはできない。(4)を排除するためには、商品と不満が記述された文脈を考慮する必要がある。

本研究の目的は {不満, 不満を解決する商品} という関係を抽出することである。そのため、商品と不満が記述される場合の文脈としては、「筋肉痛 (不満) が解消した」、あるいは、「花粉症 (不満) を予防できた」といったような、「不満をなくす、あるいは防ぐ」意味を持った語を伴うことが望ましい。これは前節で用いた改善動作と同様である。そこで、提案方式として、関係の抽出においても改善動作を再利用する、つまり {不満表現, 商品名, 改善動作} の3つの共起頻度を算出する方式で {不満, 不満を解決する商品} という関係を抽出する。

4. 実験

4.1 実験方法

前章で述べた方式を用い、不満の抽出と関係の抽出を行なった。まず、改善動作を抽出するため、無作為に収集したブログ記事1,281,765件を形態素解析し、動作名詞を抽出した。形態素解析にはJTAG¹⁰⁾を用いた。次に、抽出した動作名詞と、不満サンプルの共起確率、そして、動作名詞と「してくれる」の共起確率をそれぞれ算出した。不満サンプルとしては、goo ランキング¹¹⁾の「治したい! 女性の身体の悩みランキング」、「30歳を過ぎて気になり始めた体の変化ランキング」、「2007年年間病名検索ワードランキング」から人手で抽出した90個のサンプルを用いて2形態素以内を共起範囲とし、「してくれる」に関しては、1形態素以内を共起範囲とした。その後、両共起確率の高い順に、改善動作の定義に当てはまる動作を10個選定した。改善動作の定義に当てはまらなかった動作の例としては、「応援」、「心配」などが挙げられる。

不満表現の抽出に関しては、まず、選定した改善動作とその直前の10形態素を同じブログ記事集合から抽出した後、改善動作の対象のうち、健康に関する不満表現を人手で980個抽出した。

抽出した改善動作、不満表現を用いて関係の抽出を行なった。商品名としては「WiiFit」を用い、関係の抽出に用いたブログ記事としては「WiiFit」または「Wii フィット」をクエリとして収集したブログ記事46,349件を用いた。共起範囲としては、不満表現と商品名の共起範囲が、100Byte~400Byte、1記事内であり、不満表現と改善動作が3形態素以内である。

以上の条件を元に関係の抽出を行ない、精度、再現率によって評価を行なった。正解デー

改善動作	P_{kaizen}
解消	2.543870e-04
改善	7.044562e-05
治療	6.087893e-05
克服	2.815650e-05
復活	1.478488e-05
予防	1.092559e-05
防止	2.989590e-06
緩和	2.446028e-06
回復	1.576329e-06
除去	1.386083e-06

Table 1 Extracted “kaizen-dousa”

不満表現例			
ストレス	腰痛	ガン	トラブル
運動不足	虫歯	紫外線	ワキガ
違和感	夏バテ	疲れ	メタボ
アレルギー	老化	症状	寝不足
生活習慣病	便秘	風邪	角質

Table 2 Examples of extracted dissatisfactions

不満表現	共起頻度
筋肉痛	2426
メタボ	546
風邪	392
ストレス	373
脂肪	367
運動不足	304
疲れ	298
内臓脂肪	210
メタボリック	208
リバウンド	185

Table 3 Relation ranked by frequency of co-occurrence (“kaizen-dousa” was’t applied)

不満表現	共起頻度
運動不足	208
ストレス	73
メタボ	40
便秘	24
太り	21
にきび跡	14
メタボリック	14
肥満	8
抜け毛	8
疲労	8

Table 4 Relation ranked by frequency of co-occurrence (“kaizen-dousa” was applied)

タは商品名と不満表現が記述されているブログ記事を人手で確認して作成した。ただし、不満表現は「WiiFit」との共起頻度が5回以上の不満表現のみを用い、評価者は1名である。正解としては、以下のいずれかの項目を含む記事が1つでも存在した場合、その不満表現を正解とした。

- 商品によって不満がなくなった、あるいは防ぐことができたことが明記されている。
- 商品によって、不満をなくす、あるいは防ぐことを目的としていることが明記されている。

4.2 結果と考察

まず、今回抽出した改善動作と不満表現をそれぞれ表1、表2に示す。表2には、「トラブル」、「症状」、「違和感」等の、単独では意味があいまいな表現も含まれているが、これは「～のトラブル」、「～の症状」、「～の違和感」といった表現を抽出することを考慮してのことである。ただし、本稿における評価からはこれらのあいまいな表現は排除した。

実際に不満と商品の関係を抽出した結果を表3、表4、図3に示す。表3、表4はそれぞれ{不満表現, 商品名}の共起頻度上位10件と、{不満表現, 商品名, 改善動作}の共起頻度上位10件であり、図3は横軸に共起頻度のランキング、縦軸に精度をとったグラフである。ただし、不満表現と商品名の共起範囲は1記事内である。図3を見ると、上位6件までは改善動作を共起条件に加えた方が精度は高くなっている。これは、改善動作を用いることで表3における「筋肉痛」(商品購入後(使用後)に新たに発生した不満)を取り除くことができた結果である。また、「風邪」は表3では3位であったのに対し、改善動作を共起条件に含めた場合では31位となり、大きく順位を落としている。「風邪」に関して

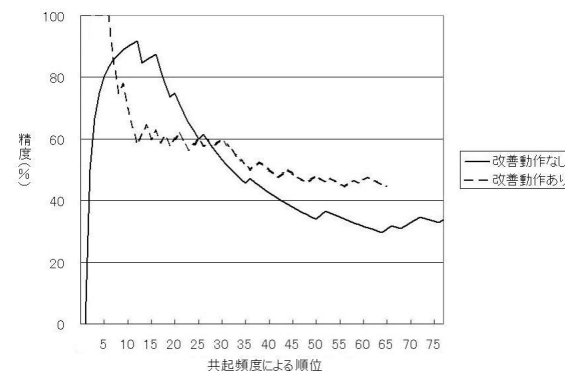


図3 関係抽出結果の精度(共起範囲1記事内)

Fig. 3 Accuracy of extracted relation (the range of co-occurrence is within a blog)

は、確かに「体力をつけて風邪を引きにくくする」といったような効果の記述も存在するが、「風邪を引いたのでWiiFitをサボった」といったような、WiiFitの効果とは何の関係もない記述も多い。このことから、効果である場合と、効果でない場合の両側面を持つ不満表現に関しては、改善動作のように、効果であるということを示すような語を利用することで、正しくランキングしていくことが重要であると思われる。

次に、不満表現と商品名の共起範囲を変化させた結果(改善動作あり)を図4、図5に示

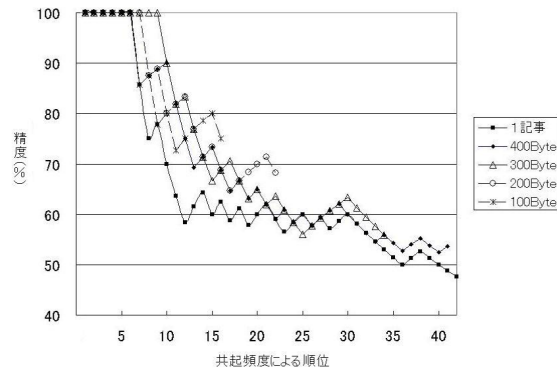


図 4 関係抽出結果の精度 (改善動作あり, 共起範囲変化)

Fig. 4 Accuracy of extracted relation (The “kaizen-dousa” is co-occurring and the range of co-occurrence is changed)

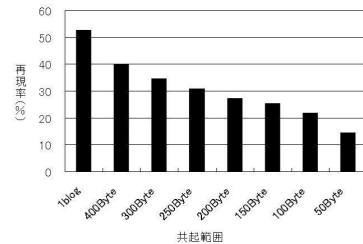


図 5 再現率の変化

Fig. 5 recall of extracted relation

す。ただし、図 4 は横軸に共起頻度のランキング、縦軸に精度をとったグラフ、図 5 は横軸に共起範囲、縦軸に再現率をとったグラフである。図 4、図 5 を見ると、共起範囲を絞ることにより精度の向上が見られたが、再現率が大きく低下した。また、共起範囲を変化させない場合でも、改善動作を共起条件に加えるだけで再現率が大きく低下している。このことから、改善動作や効果を表わす語を増やすことが今後の課題で重要であると思われる。

4.3 プロトタイプシステム

3.2 節で抽出した不満表現を用い、不満表現、商品名、改善表現の共起頻度によって商品と

不満との関係性を抽出する方式を用いて、プロトタイプシステムの実装を行なった。このシステムでは、不満表現、商品名、改善表現の共起頻度をあらかじめ計算しておき、ユーザから不満を表わすキーワード、あるいは、不満が記述されている文章が入力された場合に、その不満を解決する商品をリアルタイムで選定し、出力する。商品の選定には、関係する不満に共起頻度によるランキング順位の逆数を得点として与え、その得点が高い順に商品を選定した。

実装のために用いたブログ記事としては、「健康グッズ」や「健康商品」等のクエリで収集した 35,267 件のブログ記事を用いた。ただし、商品名に関しては、これらのブログ記事と、無作為に収集した 1,281,765 件のブログ記事から「～を買いました」等の購買表現に繋がる名詞等を抜き出し、その中から健康に関して効果のある 289 個の商品を人手で選択している。また、不満表現と商品名の共起範囲は 1 記事内とした。

実際に、不満を文章として入力した場合の結果と、キーワードとして入力した結果をそれぞれ図 6、図 7 に示す。図 6 では、ユーザが入力した文章が上部に表示されており、その下に、文章から抽出したユーザの不満と、不満を解決する商品が示されている。これにより、ユーザはどんな不満に対して、それを解決するどんな商品があるのかが一目でわかるようになっている。また、図 7 を見ると、「運動不足」に対して「ロデオボーイ」や「ジョーバ」等「疲れ目」に対して「ブラックベリー」や「ブルーベリー」等「頭痛」に対して「パファリン」など、不満を解決する商品として妥当な商品が提示されることが見て取れた。

5. ま と め

本研究では「不満、不満を解決する商品」という関係を発見するため、「不満表現の抽出」と「関係の抽出」について検討を行なった。不満表現の抽出では、商品名、動作、利益を表わすモダリティの 3 要素で構成されるタイプの文に着目し、「不満をなくす、あるいは防ぐ」意味を持つ改善動作を定義し、利益を表わすモダリティを利用してその抽出を行なった。そして、改善動作の対象を人手で抽出することによって、不満表現の抽出を行なった。

更に、関係の抽出においては、改善動作を再利用することで、目的とする文脈に沿った共起頻度を算出し、商品と不満の関係性を抽出することができた。また、改善動作を利用した方式を用いてプロトタイプシステムを実装したところ、いくつかの不満に対して、妥当な商品を見出すことができた。

以上のことにより、改善動作のような、効果を表わす語が有効であるという知見を得た。今後の予定としては、不満表現として、今回用いたような名詞だけでなく、評判抽出で得

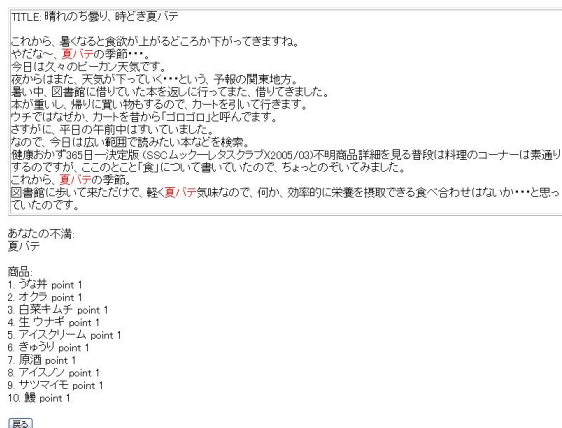


図 6 文章入力結果
Fig. 6 Output of the system (input is blog)

<p>運動不足:</p> <ol style="list-style-type: none"> 1. ロデオボーイ 2. ヨガマット 3. ジョーバ 4. ピラティス 5. サングラス 	<p>アトピー:</p> <ol style="list-style-type: none"> 1. ネイチャーメイド 2. オリーブオイル 3. 石けん 4. レメディ 5. オーガニックコットン 	<p>疲れ目:</p> <ol style="list-style-type: none"> 1. ブラックベリー 2. 黒豆 3. スイカ 4. ブルーベリー 5. いも
<p>メタボ:</p> <ol style="list-style-type: none"> 1. ライ麦 2. スポーツ自転車 3. 草履 4. やせ薬 5. 黒烏龍茶 	<p>抜け毛:</p> <ol style="list-style-type: none"> 1. シャンプ 2. マッサージオイル 3. ケラスターゼ 4. 育毛剤 5. アロエ 	<p>頭痛:</p> <ol style="list-style-type: none"> 1. バファリン 2. 頭痛薬 3. 風邪薬 4. ぶり 5. いも

図 7 キーワード入力結果
Fig. 7 Output of the system (input is a keyword)

られる「名詞+形容詞」のような不満表現についても検討していくことで、より多くの商品、不満に対応していく予定である。また、効果を表わす語についても広く抽出を行なっていく予定である。

参考文献

- 1) 土方嘉徳：情報推薦・情報フィルタリングのためのユーザプロファイリング技術，人工知能学会誌，Vol.19, No.3, pp. 365-372 (2004)
- 2) 古瀬蔵，廣嶋伸章，山田節夫，片岡良治：ブログ記事からの意見文検索，情報処理学会研究報告，NL-176 (2006)
- 3) 小林のぞみ，乾健太郎，松本祐治，立石健二，福島俊一：テキストマイニングによる評価表現の収集，情報処理学会研究報告，NL-154, pp. 77-84 (2003)
- 4) 廣嶋伸章，山田節夫，古瀬蔵，片岡良治：評判検索におけるクエリ依存型の評価極性付与，情報処理学会研究報告，NL-176 (2006)
- 5) 金山博，那須川哲哉：要望表現の抽出と整理，言語処理学会第 11 回年次大会，pp. 660-663, (2005)
- 6) LIQIANG GENG, HOWARD J. HAMILTON: Interestingness Measures for DataMining: A Survey, ACM Computing Surveys, Vol. 38, No. 3 (2006)
- 7) 渡部啓吾, Danushka Bollegala, 松尾豊, 石塚満: 検索エンジンを用いた関連語の自動抽出, 人工知能学会全国大会論文集, Vol.22nd Page.3B1-04 (2008)
- 8) 乾孝司, 奥村学: 文書内に現れる因果関係の出現特性調査, 情報処理学会研究報告, 2005-NL-167 (2005)
- 9) De Saeger Stijn, 鳥澤健太郎: トラブルを見つける, 言語処理学会, 第 14 年次大会論文集, pp.1073-1076 (2008)
- 10) Takeshi Fuchi, Shinichiro Takagi: Japanese morphological analyzer using word co-occurrence -JTAG-, In Proceedings of COLING-ACL '98: 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, volume 1, pages 409-413, Montreal (1998)
- 11) goo ランキング: <http://ranking.goo.ne.jp/>