

レファレンス事例に対するキーワードの自動付与

Automatic assignment of keywords to the reference records

◎樋澤光紀(慶應義塾大学), 原田隆史(慶應義塾大学), 江藤正己(亜細亜大学非常勤講師)
hizawa7@slis.keio.ac.jp, ushi@slis.keio.ac.jp, eto@fw.ipsj.or.jp

本研究の目的は、レファレンス事例の質問文と回答文を元にして、その事例に適切なキーワードを自動的に付与することである。これまでレファレンス事例中の語をキーワードとして抽出する実験は行われていたが、この方法では事例中に出現しない語はキーワードとして付与できないという欠点があった。そこで本研究では、ウェブ検索エンジンの検索結果リストに含まれるタイトルとスニペットを利用し、キーワードを自動的に付与する手法を提案する。具体的には(1)事例中の語からキーワードを機械学習手法により抽出し、(2)抽出したキーワードを検索語として検索エンジンからタイトルとスニペットを取得し、(3)その中に出現する語を対象として重要度計算によってキーワードを決定する、という手順でキーワードの自動付与を行った。日本史分野のレファレンス事例 120 件を使用して、付与実験を行った結果、キーワードの候補として 10 語を付与した時に精度 31.4%、再現率 61.4% で正しいキーワードを付与することができた。

The purpose of this research is to assign automatically appropriate keywords to the reference records on the basis of the sentences of questions and answers in the reference records. Former researches extracted appropriate words automatically as keywords from the reference records, but this method could assign only the words in the texts. This research proposes a method of assigning of keywords using titles and snippets of web pages. The method consists of three processes: (1) the process of extracting keywords as search terms from the reference records by the machine learning, (2) the process of acquiring titles and snippets from Google, and (3) the process of acquiring the keywords by estimation of weights. The result showed that the keywords were assigned successfully with precision ratio of 31.4% and recall ratio of 61.4%.

1. レファレンス事例におけるキーワード

レファレンスサービスとは“何らかの情報あるいは資料を求めている図書館利用者に対して、図書館員が仲介の立場から、求められている情報あるいは資料を提供ないし提示することによって援助すること、およびそれにかかわる諸業務”であり¹⁾、図書館の調査支援機能として重要な位置づけを持つ。このようなレファレンスサービスでは図書館員に高度な専門性が要求されるため、レファレンスを記録することが図書館員のレファレンスサービスに対する知識や経験の共有につながり、有効であるとされている²⁾。レファレンスサービスで受け付けた質問とそれに対する回答の記録を

レファレンス事例といい¹⁾、上述の用途以外にも質問傾向の把握、類似質問の調査ツールとしての活用方法もある。

この全国の図書館で記録されたレファレンス事例を共同で収集、保存しようとする試みが国立国会図書館のレファレンス協同データベース事業である³⁾。この事業は、“全国の図書館で行われているレファレンスサービスの記録や、そこで蓄積された調べ方に関する情報などをデータベース化し、図書館におけるレファレンス業務や、一般の人々の情報検索に役立てること”を目的としている⁴⁾。レファレンス協同データベースには、平成 21 年 1 月末の時点で 21965 件（一般公開分）⁵⁾のレファ

レンス事例が登録されている。

レファレンス協同データベースに登録する項目には、必須項目である「質問文」「回答文」「管理番号」「公開レベル」の4項目と、任意項目として「キーワード」「NDC」「回答プロセス」「参考資料」など15項目とがある⁴⁾。これらの項目のうち「キーワード」の項目は、大量のデータを対象とした検索の効率化や⁶⁾、レファレンス記録のグループ化などに有効であると考えられる。しかし、現在収録されているレファレンス事例の約4割にはキーワードが付与されていない。

レファレンス事例に対してキーワードを自動的に設定する研究としては、原田らによる機械学習手法を用いたレファレンス事例中からの語の抽出がある⁷⁾。この実験では、キーワードが持つ表現上の特徴などをもとにして、レファレンス事例中に出現するキーワードを抽出し、最大で精度 54.0%、再現率 55.3%という結果を出している。ただし、原田らの手法はレファレンス事例の文章中に出現した語句の中でキーワードとして適切な語を抽出するものであって、レファレンス事例に新たな語を付与するものではない。そのため、新たなアクセスポイントの増加を行うことはできず、キーワード付与の効果も限定的であった。

本研究はレファレンス事例中に出現しない語も含めたキーワードの自動付与を行うために、レファレンス事例と内容が類似する Web ページを情報源にキーワードの自動付与手法を提案することを目的とする。具体的には、レファレンス事例を対象に、内容が近いウェブページ中の語をもとにして、レファレンス事例中にないキーワードを付与する手法を提案する。なお、関連するウェブページ中から語を選択する際には、ウェブページ中の全ての語を対象とするのではなく、検索エンジンで表示されるスニペット中の語を用いた。

2. キーワードの自動付与に関する関連研究

2.1 キーワードの自動付与

日本語テキストに対してキーワードを自動的に設定しようとする試みは昔から行われており、大きく分けて(1)テキスト中からキーワードとして

適切な語を抽出する抽出索引法と、(2)辞書を用いて付与するキーワードを選択する付与索引法の2種類の手法がある。テキスト中の語を対象とした抽出索引法の研究は特に活発に行われており、原田らによる研究もこの手法によるものである。

一方、付与索引法を実装したシステムはあまり多くない。付与のために使われる情報源としても、日本経済新聞社の新聞記事データに対する自動索引付けシステムのように⁸⁾、上位語、同義語、親子関係などを記録した辞書やシソーラスをもとにして追加する例がほとんどである。しかし、このように辞書やシソーラスを使用したキーワードの自動付与手法では、辞書やシソーラスを更新する作業を人手で行うため多額の費用が必要であり、また新しい語には対応できないなどの問題もある。

2.2 スニペットの利用

スニペットとはウェブページ中における検索語を含む前後の文章の抜粋のことを指し、ウェブページを全文読まずとも大まかな内容を把握することができる。つまりスニペットに含まれる語句は、検索語となんらかの関係性を持っていることが推測できる。

スニペットと検索語の関係性に着目した研究の1つに、Olga Vechtomova らによる検索語拡張についての研究が挙げられる⁹⁾。この研究では TREC2005 の HARD Track データを対象に、検索語とスニペットを用いた場合の方が、文書全体から検索拡張語を得る場合よりも効果的であることを明らかにしている。

また、大島らはタイトルやスニペットをもとに、ある語に対する同位語は並列助詞の「や」で接続されることを利用し、同位語とそのコンテキストを得る手法を提案している¹⁰⁾。従来の研究では同位語を求めるために、膨大な文書集合の解析を行う必要があったが、提案手法ではタイトルやスニペットのようなウェブ検索エンジンの持つ情報のみで効率的に同位語を得ることが出来ると考えられる。

3. 本研究で提案するキーワード自動付与の手法

本研究ではレファレンス事例に対するキーワー

ドの付与を以下の手順で行った。

(1) レファレンス事例からの検索語の抽出

レファレンス事例中に出現する語のうち、レファレンス事例の内容を示すと考えられる語を機械学習手法により抽出する

(2) 検索語をもとにしたスニペットの取得

(1)で抽出した語を検索語として、検索エンジン Google¹¹⁾を用いてウェブページのタイトルとスニペットを取得する

(3) スニペット中の語を基にしたキーワード付与

(2)で取得したウェブページのタイトルとスニペット中に含まれる語について重要度を計算することでキーワードの付与を行う。

3.1 レファレンス事例からの検索語の抽出

機械学習による自然言語処理では、特徴分析が重要であることが指摘されている¹²⁾。本研究では、

レファレンス記録中のキーワードが持つ特徴と人手に基づく正解判定の対を学習させ、自動判定を行う。具体的には原田らがレファレンス事例から重要語を自動抽出した手法⁷⁾を改良する方法をとった。機械学習システムは、ワイカト大学で開発されたWeka¹²⁾を使用し、判定器にはナイーブベイズ(以下、「ベイズ」とする)を用いた。

キーワードが持つ特徴としては、以下の5つの観点に着目した。①出現頻度や質問文と回答文に共通して出現したかどうか、②助詞や記号との前後関係、出現位置や係り受け関係などの文章構造的な特徴、③特定性が高い固有名詞や動詞化するサ変名詞などの品詞情報、④語が人名を表すのか地名を表すのかなど意味情報、⑤特定性が低い不要語であるかどうかやNDC中で使用されている語であるかなど辞書情報との照合。具体的には第1

第1表 キーワードの特徴分析

観点	特徴	観点	特徴
出現頻度	2回以上出現	係り受け関係	特定名詞から一定距離
	3回以上出現		特定述語から一定距離
共起関係	質問文と回答文の共通語	意味	人名(質問文と回答文別)
辞書	不要語		地名(質問文と回答文別)
	NDC細目	出現位置	質問文から抽出
に～	回答文から抽出		
～に	質問文の文頭		
～について	回答文の文頭		
～における	～(ふりがな)		
～に関する	・の前		
助詞及び助詞相当表現	～とは	・の後	
	～を	(名詞), (名詞)	
	～の	「」前・中・後(質問文と回答文別) 以下同様	
	の～	『』の前・中・後	
	～で	()の前・中・後	
	～は	品詞	固有名詞
	～での		サ変名詞
	～が		形容詞

表に示す 53 個の特徴を設定した。

原田らの手法からの改良点としては、日本語係り受け解析器 CaboCha¹³⁾を用いて係り受け関係に関する特徴、地名や人名に関する特徴、さらに共起関係やサ変名詞、形容詞に関する特徴を用いたことが挙げられる。その結果、精度・再現率ともに向上がみられた。比較のため原田らが行った実験と同じ実験データを用いた場合のレファレンス事例からの重要語抽出結果を第 2 表に示す。

第 2 表 レファレンス事例中からの重要語の抽出

	原田ら	本研究
重要語と判定した語数	1238	1508
正しく判定できた語数	668	818
誤って判定した語数	570	690
精度	54.0%	54.2%
再現率	55.3%	67.7%
F 値	54.6%	60.2%

学習対象語数 9490 語, キーワード数 1208 語

3.2 検索語をもとにしたスニペットの取得

レファレンス事例中の語を利用し、検索エンジン Google を用いて検索をおこなう。このとき、レファレンス事例の内容を的確に示す語を用いる、レファレンス内容を示す多くの検索語を組み合わせるなどで検索式の特定性を高める方が、よりレファレンス事例と関連が高いウェブページの検索に結びつくと考えられる。一方で特定性を高めすぎると、結果として表示されるウェブページ数が過小になり、キーワードの候補となる語が少なくなってしまうことが懸念される。

そこで本研究では、より特定性の高い語の組み合わせで必要なキーワード候補語を得るために、1) その語だけを検索語とした検索で表示されるウェブページ数が少ない語から順に、2) 語を論理和で組み合わせた検索を行い、3) 検索結果が 500 件を最初に下まわった語の組み合わせで、検索を行った結果を用いることとした。

たとえば、自動抽出により、[オランダ船][積荷][江戸時代][品目][日本]というキーワードが判定された場合、それぞれの語の検索結果数は第 3 表

であり、これを検索結果の数が少ない順に組み合わせた結果が第 4 表である。この場合には、第 4 表で 3 番目に示す[オランダ船]AND[積荷]AND[江戸時代]で検索した結果を用いることとした。

第 3 表 各語単独で検索した場合のページ数

オランダ船	23,500
積荷	293,000
江戸時代	4,440,000
品目	4,750,000
日本	824,000,000

第 4 表 語の組み合わせによる検索結果の変化

[オランダ船]	23,500
[オランダ船]AND[積荷]	579
[オランダ船]AND[積荷]AND[江戸時代]	204
[オランダ船]AND[積荷]AND[江戸時代]AND[品目]	39

検索結果がどの程度まで少なくなる場合に適切なキーワード候補語が得られるかについては、何度も試行錯誤を繰り返し、レファレンス事例と関係のない語がそれほど多く含まれない範囲ということで 500 件という基準を定めた。ただし、この点については考慮の余地があると考えられる。

なお、500 件以上の検索結果であった組み合わせに新たな語を追加した場合の検索結果が 5 件を下回った場合は、新たな検索語の追加を行わなかった場合の検索結果を用いることとした。

3.3 スニペット中の語を基にしたキーワード付与

前節で収集されたウェブページのタイトルおよびスニペット中の文章をもとに、パープレキシティによる重要度計算によって、レファレンス事例にとって適切なキーワードの候補を決定する。

パープレキシティは情報理論で使われる指標で、各単名詞に「情報理論的に見ていくつの単名詞が接続可能か」を示す値である。本研究では、重要語抽出プログラムである TermExtract¹⁴⁾が算出する値を用いた。TermExtract では計算コストの低減のために、単名詞のエントロピーを元に 2 のべき乗を行うのではなく、2 を底にしたパープレキ

シティの対数を出すことで求めている。

キーワードとなりえないようなキャッシュ、日時などの数字、矢印などの記号は予め取り除き、重要度の上位 100 語をキーワードの候補とした。なお、検索結果によっては取得できた語数が 100 語に満たない場合もあったため、その場合は取得できた限りの語を判定対象とした。

4. レファレンス事例へのキーワード付与結果

4.1 付与対象となるレファレンス事例

キーワード付与実験の対象として、2007 年 7 月 19 日時点でレファレンス協同データベース上に一般公開されていたレファレンス記録 14,524 件のうち、日本史分野(NDC2 類 1 項)のレファレンス記録で、かつキーワードが 1 個以上付与されているレファレンス記録 504 件中から無作為に抽出した 120 件を用いた。

4.2 レファレンス事例に対する正しいキーワード

キーワード付与の結果を評価する際には、各レファレンス事例に付与すべきキーワードが何であるのかを定めた正解集合が必要となる。しかし、正解の候補となる語が提示されない状況でレファレンス事例中に付与すべき全てのキーワードを決定することは困難である。

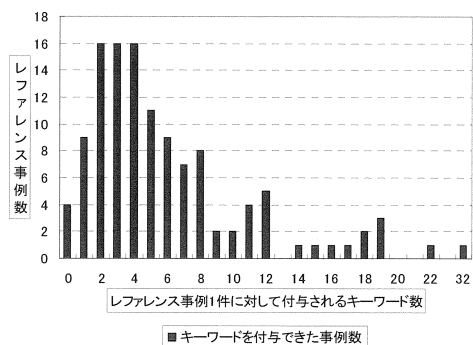
そこで、本研究では各レファレンス事例に対して付与候補とされた重要度上位 100 語を対象に人手によって判定し、キーワードと判定された語を正しいものとして評価を行った。人手による判定は 3 名が行い、判定者のうち 2 名以上がキーワードと判定した語を正しいキーワードとして扱うこととした。

その結果、120 のレファレンス事例に対し 726 のキーワードが付与された。付与されたキーワード数ごとのレファレンス事例の数を第 1 図に示す。第 1 図に示すように最も多くのキーワードが付与されたレファレンス事例では 32 個のキーワードが付与され、平均は 6.1 個であった。実際のレファレンス協同データベースにおいて、日本史分野のレファレンスで最も多くのキーワードが付与されたものでも 6 個のキーワードであり、上位 100 語を対象としてキーワード判定を行うことで十分

な正解キーワードを得ることができると判断した。なお、上位 100 語中にキーワードとなる語が全く含まれなかった事例が 4 件あった。

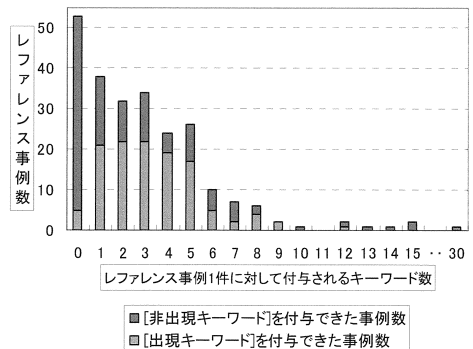
評価は、人手により 1 つ以上のキーワードが付与できた残りの 116 件を対象に行う。

第 1 図 レファレンス事例に対するキーワード数



この正解集合中のキーワード数を、レファレンス事例中に出現する語と出現しない語に分けて第 2 図に示す。

第 2 図 付与キーワード数ごとのレファレンス事例数



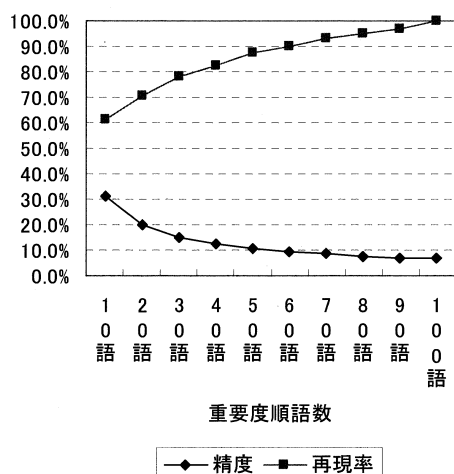
第 2 図に示すように、116 件中 78 件(67.2%)のレファレンス事例に対して、事例中に出現しないキーワードを 1 つ以上付与することができた。最も多くの新しいキーワードをできた事例では 30 個の出現しないキーワードが付与可能であった。

4.3 レファレンスに対するキーワードの付与

各レファレンス事例に対するキーワード候補の

語を重要度の高い順にならべた時、精度(事例ごとのキーワード候補語中の正しいキーワードの割合の平均)と再現率(事例ごとの正しいキーワード数中の語の割合の平均)を第3図に示す。本研究では候補語100語中の語を対象に正しいキーワードを判定しているため、上位100語までで再現率は100%となる。

第3図 重要度上位候補語数ごとの精度、再現率



第3図に示すように、上位10語の候補語中に約60%のキーワードが含まれており、上位40語では約80%のキーワードが含まれていた。

また、各レファレンス事例に対してキーワード候補語を上位から提示した場合に、レファレンス事例中出现するキーワードと、事例には出現しないキーワードとの語数を第5表に示す。

第5表に示すように、レファレンス事例116件に対して、重要度上位の10語の候補語全1160語中にはレファレンス事例中に含まれる語が404個含まれており、そのうち283語(レファレンス1件あたり2.4個/候補語のうち70.0%)がキーワードであった。また、レファレンス事例中出现しない語は756語含まれており、そのうち89語(レファレンス1件あたり0.77個/候補語のうち11.8%)がキーワードであった。

さらに、上位40語までの候補語中全4602語中にはレファレンス事例中に含まれる語が615個含

まれており、そのうち354語(レファレンス1件あたり3.10個/候補語のうち57.6%)がキーワードであった。また、レファレンス事例中出现しない語は3987語含まれており、そのうち218語(レファレンス1件あたり1.88個/候補語のうち5.5%)がキーワードであった。

第5表 レファレンス事例中に含まれていたキーワードと含まれていないキーワードの数

重要度順語数	含まれていた KW	含まれなかった KW	合計
10 語	283/404	89/756	372/1160
20 語	322/516	140/1800	462/2316
30 語	343/582	182/2884	525/3466
40 語	354/615	218/3987	572/4602
50 語	368/647	242/5066	610/5713
60 語	375/650	262/6158	637/6808
70 語	383/642	283/7236	666/7878
80 語	389/635	300/8313	689/8948
90 語	394/629	312/9366	706/9995
100 語	398/618	328/10381	726/10999

このように、レファレンス事例中出现する語の多くだけではなく、レファレンス事例中出现しないキーワードも多数付与することができることは、本手法の有効性を示しているといえよう。

4.4 レファレンス事例中出现しないキーワード
レファレンス事例中出现しない候補語の中で、キーワードであるものとしては、いくつかのタイプが見られた。

このうち、同義語・類義語や上位語・下位語については、その分野に関わる辞書を整備することで比較的キーワードを追加することが容易であるが、関連語については語幹の関わり方の種類も多く、また新しい分野の語を中心として追加が困難な例であると考えられる。

1) レファレンス事例中の語の同義語・類義語

自然言語においては、同じ語の反復を避けるために同じ意味の語を別の表現で置き換えることが行われる。そのため、スニペット中にレファレン

ス事例中に出現する語の同義語や類義語が出現することなどが考えられる。

たとえば、「太平洋戦争」という語の類義語として「大東亜戦争」や「第二次世界大戦」が用いられる例や、人名で「大岡越前守」「大岡忠相」「大岡越前守忠相」などの省略形が使われるのはこの例である。

2) レファレンス事例中の語の上位語, 下位語

たとえば【奈良時代、和氣清麻呂が宇佐八幡宮で受けた神託の言葉は何か。】というレファレンス質問で、宇佐八幡宮の上位語である[神社]を付与した例や、【源氏長者に関する文献を探している。】というレファレンス質問で、源氏の下位語として[足利義満][清和源氏]などをあげた例などがある。

3) レファレンス事例中の語の関連語

レファレンス事例中の語の関連語としては、多様な語を候補とすることができた。

関連語が候補としてあげられた例としては、【太平洋戦争に関する文献にどのようなものがあるか調べたい。】というレファレンス質問に対して、[真珠湾攻撃]という太平洋戦争の開戦となった作戦名をあげた例、【親鸞聖人の越後の七不思議を知りたい。】という検索質問に対して親鸞が開祖である「浄土真宗」という宗教名が候補とされた例、【江戸時代のオランダ船の積荷リストがみたい。】という質問でオランダ船が入港する「出島」があげられた例、などがあげられる

4.5 キーワードが提示できなかった例

本実験において、新しいキーワードを付与することができなかったレファレンス事例としては、以下の2つがあげられる。

1) レファレンス事例中から特定性の高い語が抽出できなかった場合

本研究の手法は、レファレンス事例中の語をもとにして検索エンジンで検索を行い、その結果を用いてキーワードを拡張しようとするものである。したがって、レファレンス事例中から適切な検索語が抽出できない場合には、適切な候補語にたどりつくことができない。

たとえば、【香川との県境を決めた話について知

りたい。】【「玉野市史続編」「玉野市史」に記載がある。】というレファレンス事例の場合、[香川]および[玉野市]は検索語と判定されたが[県境]は検索語とは判定できなかった。ここで[香川]と[玉野市]だけから検索を行った場合には、県境の確定という観点とは別のウェブページが多数表示されることになってしまい、適切な語を選択することができなかった。

2) 専門用語などの特殊な語が使われている場合

非常に珍しい固有名詞など、レファレンス質問中に特殊な語が使われていた場合には、その語を検索語とする検索で十分な数の検索結果を得ることができず、そのために適切な候補語が表示できない場合がある。

たとえば、【『播陽多我身飢』を探している。】というレファレンス事例では、検索語として[播陽多我身飢]が抽出されるが、この語を用いた検索結果としてウェブページは1件しか表示されず、そのために適切な候補語をあげることができなかった。特に古典書籍などの固有名詞では、レファレンス事例の内容以上に多様な語を獲得することが難しい場合が多かった。

5. まとめと今後の課題

本研究ではレファレンス事例中に含まれる語を抽出した結果を用いてウェブ検索エンジンの持つ情報からキーワードを獲得する手法を提案した。

その結果、1つのレファレンス質問に対して10語の候補語を提示した場合に平均0.77個、また30個の候補語を提示した場合には1.92個の新しいキーワードが含まれるという結果を得ることができた。

レファレンス事例に対するキーワードの付与を行う際には、候補となる語が提示されない状況よりも、提示された候補の中から選択する方法が有効であると考えられ、これを機械的に行うことができる本手法は実用システムを考える上で重要な成果をあげることができたと考えられる。

特に本手法では同義語・類義語などのような辞書中に収録されている語ではない多様な語句を候補としてあげることができており、レファレンス

事例に対するキーワードを自動的に付与する実用システム開発の基礎として重要な知見が得られたと考えられる。

ただし4.5節で述べたように、適切なキーワード候補語を選択できない例も見られた。特に、レファレンス事例中から検索語となる語が抽出できない場合、または、あまりにも特定性が高い語しか抽出できない場合には適切なキーワードの提示は困難となる。さらに、提示する候補語中のキーワードの割合が、本実験ではそれほど高くないことも課題のひとつであろう。

これらの問題の解決のためには、レファレンス事例からの検索語の抽出過程を再検討することが有効であると考えられ、検索語抽出のための手法の精緻化などが望まれる。

レファレンス事例の蓄積は、まだその端緒について段階にあり、今後蓄積されるデータの増加とともにキーワードの自動付与の必要性はますます高まっていくと考えられる。本研究が、今後のレファレンス事例の有効活用の基礎となることを願っている。

6. 注・引用文献

- 1) 日本図書館情報学会用語辞典編集委員会編. 図書館情報学用語辞典. 第3版. 東京, 丸善, 2007, 286p.
- 2) 斎藤文男, 藤村せつ子. 実践型レファレンスサービス入門. 東京, 日本図書館協会, 2004, 169p.
- 3) レファレンス協同データベース. [2009-2-21], <<http://crd.ndl.go.jp/jp/public/>>
- 4) 依田紀久. レファレンス協同データベース事業に見るデジタルレファレンスサービス. 情報の科学と技術. Vol. 56, No. 3, p. 90-95 (2006)
- 5) レファレンス協同データベース事業累積統計(平成20年11月末現在). [2009-2-21], <http://crd.ndl.go.jp/jp/library/documents/stats_200901.pdf>
- 6) 野末道子. "段落を対象とした日本語全文データベースの検索". Library and information science. No. 31, 1993, p. 79-93.
- 7) 原田隆史, 江藤正己, 瀬口真徳, レファレンス記録に対するキーワードの自動付与. 情報処理学会研究報告 自然言語処理研究会報告. Vol. 2008, No. 33, 2008, p. 119-124.
- 8) 石井昌之. "日経新聞社における自動索引システムの使用". 情報の科学と技術. Vol. 42, No. 11, 1992, p. 1058-1064.
- 9) Olga Vechtomova, Murat Karamuftuoglu. "Query expansion with terms selected using lexical cohesion analysis of documents". Information Processing and Management. No. 43, 2007, p. 849-865.
- 10) 大島裕明, 小山聡, 田中克己, "Web 検索エンジンのインデックスを用いた同位語とそのコンテキストの発見", 情報処理学会論文誌 データベース. Vol. 47, No. SIG_19(TOD_32), 2006, pp. 98-112.
- 11) Google. [2009-2-21], <<http://www.google.co.jp>>
- 12) 磯崎秀樹, 平尾努, 鈴木潤. "機械学習のための組み合わせ素性の選択基準について". 情報処理学会研究報告 自然言語処理研究会報告. Vol. 2003, No. 108, 2003, pp. 63-68.
- 13) Weka 3 - Data Mining with Open Source Machine Learning Software in Java. [2009-2-21], <<http://www.cs.waikato.ac.nz/ml/weka/>>
- 14) CaboCha: Yet Another Japanese Dependency Structure Analyzer. [2009-2-21], <<http://chasen.org/~taku/software/cabocho/>>
- 15) 専門用語 (キーワード) 自動抽出用 Perl モジュール "TermExtract" の解説. [2009-2-21], <<http://gensen.dl.itc.u-tokyo.ac.jp/termextract.html>>