

Wikipedia を利用した日本語作文支援システムの開発

岡田 仁之[†] 島田 諭[†] 福原 知宏[‡] 佐藤 哲司[†]

[†]筑波大学大学院 図書館情報メディア研究科

[‡]東京大学人工物工学研究センター

Wikipedia のカテゴリ名を用いた日本語作文支援システムを提案する。Web の発展に伴い、メールやブログなどテキストを書く機会が格段に増えた。テキストを書く際、意図を的確に伝える言葉選びには誰もが苦慮している。これを解決するために、本研究ではテキスト入力中に関連する語を提示する事で作文を支援するシステムの実現を目指す。今回、使い慣れたかな漢字変換機構の上に、入力中の語と関連する語彙を変換候補として提示するシステムを開発した。システムは、Wikipedia の見出し語から pfibf を用いて単語間の関連度を計算する。本稿では、システムの基本的な課題とアプローチについて述べると共に、試作したシステムを紹介する。

A Support System for Writing Articles using Wikipedia

Hitoyuki Okada[†] Satoshi Shimada[†] Tomohiro Fukuhara[‡] Tetsuji Satoh[†]

[†]Graduate School of Library Information and Media Studies, University of Tsukuba

[‡]Research into Artifacts, Center for Engineering, The University of Tokyo

We propose a support system for writing articles using Wikipedia in this paper. The necessity for writing sentences by using E-mail and the blog has increased because the Internet spread. In such media, it is required that an appropriate term be carefully chosen to avoid misunderstanding. For a solution of this problem, we propose a system that presents related words at the same time as inputting the text. This system makes IME(Input Method Editor) a front end, interrupts the Japanese character conversion from Kana to Kanji, and presents related words alternatively. A related level for headwords of Wikipedia is estimated in this system by using the index of link analysis that is called pfibf. In this paper, we explain the basic problem and our approach, and show the output example by the prototype system.

1. はじめに

今日、インターネットの普及によりメールやブログなど、書くことで考えを伝えたり、記録に残したりする機会が増えている。国内のブログのデータ量は毎月 1.6TB 増え続けている[1]。また、近年は特に携帯電話からの Web 利用の増加が著しい。

日本語で文章を書くときの特徴として、かな漢字変換と呼ばれる入力支援プログラムを介してテキストを記述することがある。これにより、限られた入力機器でも簡単に数千文字の漢字を使った日本語の入力が可能になった。特に、携帯電話等で採用されている「予測変換」と呼ばれるかな漢字変換方式は、ユーザの文字入力と共に、頻繁に使われる語やフレーズの候補が次々と提示され、ユーザは候補を選択することで文章を書くことができる。携帯電話の数字キーで日本語を入力することは、パソコンのキーボードと比べても遙かに多くのキータッチ数を必要とする。このため、最初の 1, 2 文字を入力した時点で変換候補が提示されると少ないキー

タッチ数で漢字仮名交じり文を書くことができ、多くの携帯電話で採用されている。

一方、予測変換は便利に使える優れた変換方法であると同時に、過去に使ったことがある語彙だけで文章を書いてしまう可能性がある。出てきた候補が予想した語彙と多少違って、気にせず採用して書き進める事も十分に起こりうる。かな漢字変換における予測変換方式の登場により、ユーザの日本語入力は使いやすくなったとはいえ、学習していない語彙を出すためには、モードを切り替えたり多くのキーを打たねばならないなどの敷居が高く、ユーザはシステムの提示する候補を採用しながら文章の方を作り替えてしまうという「文章を書かされている」状態になる恐れがある。

本研究では、かな漢字変換機構の上に、入力中の語と関連がある多様な語彙を変換候補として提示することで、ユーザがより適切な語を用いて文章を作成できる作文支援方法を提案する。一般的かな漢字変換が読みとなる平仮名から対応する漢字への変換を意図していることに対して、本提案法では全く異なる読みであっても、

内容的に関連があると思われる語彙も提示する。例えば、

例文：インフルエンザが広まっている

の変換候補として「広まって」だけではなく「パンデミック」という語彙も変換の候補としてユーザに提示する。例文において、単にインフルエンザが広まることを述べているのか、危機的なレベルでの拡大を述べるのかは、書き手の選択に委ねられるが、少なくとも、パンデミックという選択肢もあることを提示することで、より柔軟な文章作成を可能にすると考える。

本論文の構成は次の通りである。2 節では筆者らの考える作文支援方法について述べ、関連する先行研究との比較を行う。3 節ではシステムの実現方法について述べる。4 節で試作したシステムとその動作例について述べる。5 節では本論文のまとめを行い、今後の課題について述べる。

2. 語彙候補を提示する作文支援手法

2.1 課題とアプローチ

作文を行う過程におけるかな漢字変換で提示する語彙のタイプによって以下の類型化できる。

(1) よみに相当する漢字を提示

通常のかな漢字変換、同音異義語など選択を迷うケースでは、語義や用例を提示するシステムもある。ただし、辞書ベースであるため、新語や未知語に弱い。

(2) よみに相当する語彙を提示

新語など、利用者はある程度は知っているが、正確な語義や用法を知らない語彙。上記と同様に語義や用例を合わせて提示することが有効。

(3) よみに関連する語彙を提示

「広まる」から「パンデミック」を提示するような、一般化しつつある専門用語や略語、新語を入力された「よみ」から提示。

本研究では、上記の内、特に(3)の実現を目指すシステムの実現を目指す。関連語彙の抽出には Web 上のテキストデータを用いる。具体的には、Wikipedia[2]を用いて構築した連想関係から、作文時に入力する語の連想語を取得して推薦する。

2.2 関連研究

作文を行う過程で、類語辞書や文例集を参照する代わりに、Web 検索や電子化辞書を用いる事で負担は軽減される。これを更に拡張して、関連語や類字文例を推薦することで作文支援を行うというアプローチは従来から行われている。

商用のかな漢字変換プログラムである ATOK は、「連想変換」機能を搭載している。連想辞書を内蔵し、変換する語の連想語を変換候補として提示する機能である。

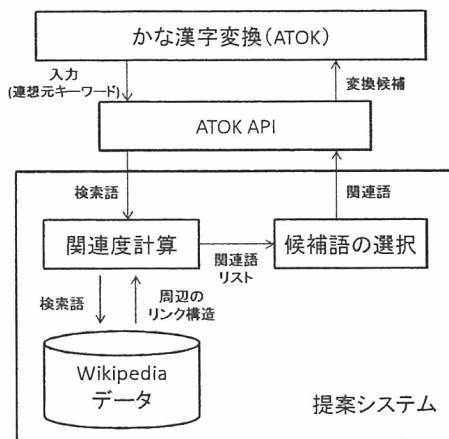


図1 提案システムの概要

中田らは、頭の中のイメージを断片的に表すキーワードから、そのイメージに類似する文例を求める試みを行っている[4]。この研究では、複数のキーワードから連想語を求め、その連想語を用いた文例を提示するシステムについて考察と提案を行っている。連想関係を求めるために、類義語辞書等に加えてシソーラスを用いて連想語辞書を構築する方針について説明している。

鄭らは連想作文支援システム ACSS (Associative Composition Support System) を提案している[5]。ACSS は、単語の連想情報を検索するための Web ベースの言語学習システムである。連想辞書を元に連想意味ネットワークを構築し、任意の 2 単語間の連想をグラフで出力する。具体的には、「グラフクラスタリング方法を利用し、単語レベルばかりではなく、全ての単語を概念別にグループ化し、概念レベルでのネットワークを作成」している。ユーザが 2 つの単語を入力すると、連想意味ネットワーク上の 2 単語間の最短パスを求め、パス上に存在する単語を連想語として提示する。

これらの研究では、人手によって編纂された辞書を用いて連想関係を構築しているが、連想関係を自動的に構築する手法が研究されている。中山らは、Web 上の事典である Wikipedia を用いて、連想シソーラスを構築している[6]。Wikipedia では各概念がページ毎にまとめられている事を利用して、ページ間のリンクを用いて概念間の関連性の強さを計測する。中山らはこの手法で、精度の高い連想シソーラスを構築した事を報告している。

また、関連語を推薦するという点で、検索クエリ拡張とも関連がある。文書検索においてキーワードマッチを用いると、クエリ中の単語を含まない文書は探し出せない。そこで、別の情

報源を用いてクエリ中の単語に関連する単語を新たにクエリに付け加えるのがクエリ拡張である。

堀ら[7]は、Wikipedia を用いたクエリ拡張手法について報告している。クエリ中の単語でまず Wikipedia 内を検索し、そこで得た関連単語をクエリ拡張のために出力する。具体的には、元のクエリで Wikipedia 内を検索し、クエリが使用されている段落を抜き出す。その段落を用いて、単語間の距離に基づいて単語間の関連度を求め、関連度に応じてクエリに関連する単語を抽出する。システムはアンケートを用いて被験者評価を行い、従来の手法に比べてより満足度の高い関連語リストを出力できた事を報告している。

かな漢字変換プログラムに搭載される等、作文中に語彙を補助する需要は存在する。これまでの単語推薦では、関連語を求める際に人手による辞書をベースにしていた。一方、Wikipedia に関する研究の進展により、自動的に関連度を計算し関連語を求める事が出来る様になっている。これらを組み合わせる事で、収録単語数や新語への対応の面で優れた単語推薦を行えると考えられる。これをかな漢字変換機構の上に実装する事で、単語推薦を受けながら負担なく文章を作成できる作文支援システムを目指す。

3. Wikipedia を用いた変換候補語の提示方法

3.1 Wikipedia

Web 上で無償で公開され、誰でも自由に編集することができる百科事典として、Wikipedia[2]がある。日本語版 Wikipedia の見出し語数は、2009 年 8 月時点で約 62 万である[3]。事典に掲載する項目の選定を含め、すべての作業が利用者に開放されているため、一般の利用者が興味を持ちやすいポップカルチャーなどの分野の記事が多く、科学など専門分野の記事が少ないという傾向がある。

Wikipedia では、MediaWiki と呼ばれる CMS(Contents Management System)が利用されており、記事本文だけでなく、編集履歴や項目間のリンクなど多種多様なメタデータがシステムティックに蓄積されている。そのため、Wikipedia は紙媒体の百科事典と異なり、また一般の Web ページとも異なる、以下に示すような特徴を持つ。

(1) 概念の独立

概念がページによって一意に区別できる。1つの記事(1つの URL)には一つ概念について記述される。多義語は概念ごとに分解され、複数の記事に記される。これは次の(2)の特徴と関連して、関連語の取得を容易にする。

(2) 多数の内部リンク

2009 年 8 月時点で、記事数は約 62 万、内部リンク数は約 2200 万リンク存在する。平均すると 1 記事あたり約 35.5 リンク含む計算になる。リンクは、リンク元の概念からリンク先の概念への関連を表すと考えられるが、一般的な Web ページでは、スパムや広告等、ページ内容との直接的な関係を認めがたいリンクも数多く存在する。Wikipedia の内部に向かって密なリンク構造からは、良質な関連を容易に多数取得することが出来ると考えられる。また、(1)の特徴によって概念の独立が保証されているため、多義語の関連も、概念ごとに区別して取得することが出来る。

(3) 即時性に富んだ記事

誰でも更新できるというその特性から、多数のユーザによって頻繁に更新される。これによって、一般的な事典よりも新語・専門語への対応が柔軟であると考えられる。例えば、「パンデミック」という語は、2005 年 3 月にエントリが出来て以来、2009 年 9 月現在まで 148 回更新されている。しかし、そのうち 84 回は新型インフルエンザ騒動が始まった 2009 年 4 月以降の更新である。使用頻度の低い専門用語であったため、以前は簡素な記事であったが、新型インフルエンザ騒動で繰り返し使用され、関心が高まると共に頻繁に更新されて記事も充実して来ている。この様に、Wikipedia の記事は最新の時勢に対応しやすい事から、新語や専門用語を豊富に取得できると考えられる。

以上の特性から、Wikipedia は関連語の取得に最適であると考えられる。

3.2 用語間の連想関連の抽出法

Wikipedia からの連想関係の抽出法には、中山らの pfibf[6]を用いる。

pfibf は中山らの提案する、Wikipedia 内のリンク構造を用いて、2 つの記事間の関連性を数値化する手法である。関係の強さは、

- 2 つの記事間のパス数
- 2 つの記事間のパスの長さ

から計算する。2 記事間のパスが多ければ多いほど(共通のリンク先や共通の参照元が多いほど)、関連性は強くなり、そのパスが短ければ短いほど、関連性は強くなる。この考えの元に、次の式(1)で計算されるのが pf(Path Frequency)である。[5]

$$pf(v_i, v_j) = \sum_{k=1}^n \frac{1}{d(t_k)} \quad (1)$$

n ホップ以内リンクで繋がる記事 v_i と記事 v_j の関連度は、その間の全経路 $T = \{t_1, t_2, \dots, t_n\}$ の長さ $|t_k|$ の合計によって計算される。関数 $d()$ は単調増加関数である。この手法は再帰的にリンクを計算するため、ホップ数の上限を設けないと計

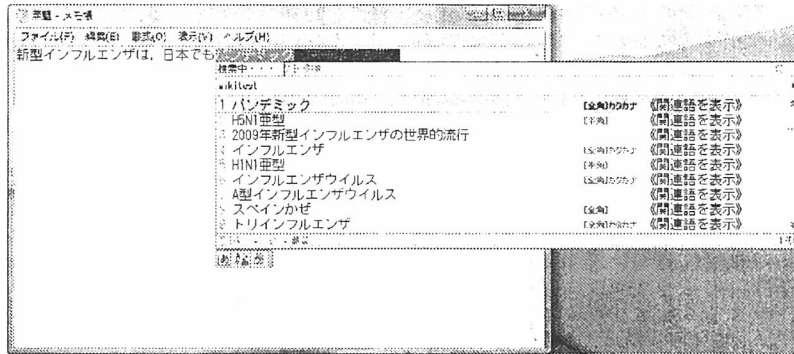


図2 提案システムのスクリーンショット

算量が膨大になる．今回はホップ数の上限を 2 として計算した．

様々な語からリンクされ、無数のリンクを持っている語がある．pf では、これらの一般語は、どの語に対しても強い関連を持つ事になり、関連抽出のノイズになると考えられるそこで、記事の被リンク数を考慮した ibf (Inversed Backward link Frequency)を導入している．

$$ibf(v_j) = \log \frac{N}{df(v_j)} \quad (2)$$

N は全記事数、 $df(v_j)$ は記事の被リンク数である． ibf は被リンク数 (バックワードリンク) が増えるに従って減少する．これをもって、関連度の急激な増加を和らげる．

関連度 $pfibf$ は、(1)式ならびに(2)式を合算して得る．

$$pfibf(v_i, v_j) = pf(v_i, v_j) \cdot ibf(v_j) \quad (3)$$

つまり、 $pfibf$ は「多くのリンク先を共有するが、他の記事とはリンク先を共有しない記事により高い値を示す」ように作られた指標である．

関連度の計算方法として他に、TFIDF を用いる手法、リンク共起性解析がある．これらの手法と比べ、 $pfibf$ は解析に記事本文を全く使わず、リンク情報だけを用いる．また、他の方法は Wikipedia 全体で予め関連度を計算しておく必要があるが、 $pfibf$ はリンク情報さえあれば求める単語間の関連度のみを求めることができる．随時更新される Wikipedia のデータをすぐに利用する事ができ、新語・未知語も素早く取り込むことができる．他と比べて、 n ホップ内で再帰的に計算を繰り返すため計算量多いが、今回設定したホップ数 2 ではレスポンスに大差はないと考えた．

3.3 変換候補語の提示法

かな漢字変換プログラムの UI を使用する．日常的に使用するかな漢字変換プログラムからシームレスに呼び出せるインタフェースを目指す．

表 1 使用したデータ

ファイル名	page.sql.gz	pagelinks.sql.gz
サイズ(MB)	53	342
内容	記事の見出しやID等、記事の属性を含む	見出し語とIDの対でリンクを表す

「ATOK ダイレクト API」はかな漢字変換プログラム「ATOK」の開発元が配布する API であり、ATOK が提示する変換候補を作成することができる¹．図 2 の様に、通常のかな漢字変換とほぼ同様の使い方で、様々な変換候補を提示させることができる．今回はこの API を利用し、Wikipedia から取得した関連語を提示する．

4. 実装

4.1 試作システムの実装

提案するシステムは、入力に応じて、入力した語の関連語を提示する．例えば、「パンデミック」という語の入力に応じて、「A 型インフルエンザ」「汎発流行」「バイオハザード」等の関連語を出力する．

システムにおける処理の流れは図 1 に示した．入力は通常のかな漢字変換プログラムから行う (図 2)．かな漢字変換プログラムから検索を行うと、提案システムに入力中の単語が検索語として渡される．システムは検索語を受け取り、それと最も近い Wikipedia 内の記事を取得する．Wikipedia データは予めデータベースに格納されている．取得した記事からリンクを辿り、関連語を取得する．このリンク構を基に各関連語の関連度を計算する．その後、関連度等の情報を用いて関連語を絞り込み、ユーザに提示する．用いた Wikipedia のデータの概要を示す．

¹ <http://www.atok.com/useful/developer/api/>

表 2 システムが提示する関連語の例

順位	インフルエンザ	pfibf	連想元キーワード							
			流行	流行	妖怪	妖怪	河童	河童	雪女	雪女
1	インフルエンザ	2363.5	流行	865.3	妖怪	5867.2	河童	2935.5	雪女	1293.6
2	インフルエンザウイルス	695.1	ファッション	130.2	日本の妖怪一覧	3300.2	妖怪	575.7	妖怪	376.7
3	2009年新型インフルエンザの世界的流行	675.9	1990年代	76.8	村上健司	1578.6	日本の妖怪一覧	435.9	ろくろ首	274.5
4	H5N1亜型	659.2	ブーム	76.4	水木しげる	1178.9	天狗	351.8	河童	266.9
5	A型インフルエンザウイルス	629.0	1980年代	76.1	多田克己	1134.7	ゲゲゲの鬼太郎の登場人物	326.8	産女	229.5
6	トリインフルエンザ	627.8	アメリカ合衆国	74.2	ゲゲゲの鬼太郎の登場人物	1134.1	雪女	308.0	牛鬼	227.3
7	オセルタミビル	622.3	髪型	72.2	鳥山石燕	1043.9	日本	287.7	鎌鼬	221.6
8	スペインかぜ	621.5	日本	71.3	毎日新聞社	1014.4	座敷童子	279.5	ゲゲゲの鬼太郎の登場人物	208.0
9	H3N2亜型	579.7	若者文化	67.5	伝説の生物一覧	984.7	鬼	259.2	座敷童子	205.9
10	H1N1亜型	578.1	1960年代	64.0	京極夏彦	841.9	タヌキ	255.6	猫又	194.1
11	豚インフルエンザ	559.9	ギャル	63.7	国書刊行会	757.1	牛鬼	252.6	化け猫	178.8
12	H2N3亜型	537.5	文化	63.1	江戸時代	712.2	村上健司	250.3	出身都道府県別漫画家一覧	178.1
13	猫インフルエンザ	535.5	日本の文化	62.3	新紀元社	711.6	人魚	233.2	ぬらりひよん	176.1
14	ライミニダーゼ阻害薬	533.8	1970年代	61.9	画図百鬼夜行	706.9	福岡県	225.6	火車(妖怪)	173.8
15	ザナミビル	531.2	服装の乱れ	61.4	河童	675.5	山童	224.8	犬神	173.4
16	C型インフルエンザウイルス	524.6	英語	60.4	幽霊	644.6	水木しげる	218.6	小豆洗い	173.2
17	B型インフルエンザウイルス	523.6	ファッション用語	59.8	日本	633.4	東京都	216.9	青坊主	168.5
18	アジアかぜ	521.3	ミーム	59.2	講談社	613.3	岩手県	211.7	天狗	167.3
19	H2N2亜型	517.1	2004年	58.9	百器徒然袋	591.0	九州	208.2	鳥天狗	164.3
20	犬インフルエンザ	517.1	ファッションショー	58.3	付喪神	585.4	カワウソ	208.1	都道府県の面積一覧	160.2
21	アマンタジン	516.9	服装	57.5	鬼	563.6	ろくろ首	198.3	九尾の狐	157.7
22	馬インフルエンザ	504.8	ディスコ	56.8	今昔画図続百鬼	560.7	毎日新聞社	196.8	のつべらぼう	156.3
23	新型インフルエンザ	450.1	衣類	56.0	ゲゲゲの鬼太郎	519.8	出身都道府県別漫画家一覧	195.3	山姥	156.1
24	パンデミック	438.7	2000年代	55.9	雪女	510.2	多田克己	192.7	わいら	156.0
25	感染症	413.7	2002年	55.3	座敷童子	500.1	2005年	192.8	日本	154.4
26	ウイルス	378.2	神戸コレクション	55.1	古神道	499.6	日本の漫画家一覧	190.2	都道府県の人口一覧	151.7
27	感染症の歴史	347.8	服飾	54.7	天狗	476.2	火車(妖怪)	190.0	釣瓶落とし	151.1
28	抗ウイルス薬	337.2	価値観	54.4	タヌキ	470.0	伝説の生物一覧	186.7	ぬっぺふぼふ	150.8
29	ワクチン	308.6	ストリートスナップ	53.8	お化け	468.8	北海道	185.4	おどろし	150.2
30	伝染病	281.6	経済	52.8	今昔百鬼拾遺	450.2	漫画	185.4	見越し入道	150.2

使用したデータは表 1 のものである。数字は 2009 年 9 月 27 日版のものである。page.sql が含む約 159 万ページから、管理用のページ等を取り除き、99 万記事¹を得た。pagelinks.sql からは約 4600 万リンクを取得し、うち取り除いたページに対するリンクを除き、約 3490 万リンクを得た。

関連度計算はこの 2 つのデータから行う。まず検索語に一致する記事を探し、ID を取得する。ID を元にリンクが繋がる見出し語を収集し、見出し語の pf を更新する。収集した見出し語から更にリンクするページを取得し、pf を更新する。その後、各 pf と ibf を合算し pfibf のランキングを作成、上位 50 件を関連語リストとして API に渡す。

4.2 システムの出力例

¹ リダイレクトページを含む為、前述の統計と数が異なる

本節では、本研究で提案する作文支援システムの一部分を構成する関連度計算のモジュールを用い、作文支援における主な提示例に沿ってクエリを入力し、関連語と pfibf 値を出力した結果を考察する。

4.2.1. 文脈に適した言い換え表現の推薦

これは、例えば「いんふるえんざのりゅうこう」と入力した時点で、そのまま「流行」と変換する代わりに「パンデミック」という語を提示するような関連提示である。インフルエンザに関して、やや専門的な文章を書く場合、「流行」という語を用いると一般的すぎる。インフルエンザなどの感染症が爆発的に拡大することを意味する専門用語である「パンデミック」を推薦すれば、よりインフルエンザに話題を絞った、専門的な記事であるという印象を読者に与えることができる。本システムでは、「インフルエンザ」から見た「パンデミック」の pfibf 値は 438.7 となり、「インフルエンザ」に関連す

る 198,429 語中、24 位になる。「インフルエンザ」の関連語を単純に pfbf 値で並べ替えると、「インフルエンザ」自体の同義語や下位概念である「H5N1 型」「鳥インフルエンザ」などの語が上位になる。「インフルエンザ」+「流行」という特定の組み合わせに対し、「流行」の言い換え表現として「パンデミック」を提示するには、以下に示すようなルールの追加が必要と考えられる。

- ・見出し語にインフルエンザを含まない
- ・上位概念が異なる
- ・説明文に「流行」も含む

4.2.2. 下位概念の列挙

これは、例えば「にほんのようかいといえは」と入力した時点で、続けて「河童や雪女など」と具体例を提示する場合である。このように具体例を挙げることで、読者にとって具体的なものをイメージしやすい文章とすることができる。また、下位概念が曖昧または多様である場合には、あえて特定の数例を例示することにより、著者の意図を明確化することができる。例えば「日本の妖怪といえは」に対し、「河童や雪女が有名である」という文脈ではなく、「最近はずみ男や猫娘など、水木しげるが創作したのも市民権を得てきている」といった文脈もあり得る。このような場合、具体例が例示されることで読者は安心して読み進めることができる。

本システムでは、「妖怪」の関連語は 306,607 語あり、そのうち「河童」が 15 位、「雪女」が 24 位、「座敷童子」が 25 位、「天狗」が 27 位になる。一方、「水木しげる」が 4 位になるなど、必ずしも具体例として挙げるには適さない語が上位になる傾向がある。

これには、Wikipedia 上での関連項目数の多さが影響している。Wikipedia における項目間のリンクでは、上位下位関係は明示されていない。このため、リンク解析によって上位下位関係を推定する手法が提案されている。しかし、単に上位下位関係が推定できるだけでは、適切な下位概念を提示することは困難と考えられる。下位概念を適切に提示するためには、見出し語だけでなく記事本文も利用し、また入力中の文字列を遡って取り込み、クエリを「日本 AND 妖怪」とするなどの方法が必要である。ただし、例えば「村上健司」の記事本文には「日本」と「妖怪」が出現するが、この語は人物名であり「妖怪」の下位概念ではない。「○○は日本の妖怪である。」といった構文パターンの照合を行う必要もあると考えられる。

4.2.3. 上位概念の推薦

これは、例えば「かっぱやゆきおんななどの」と入力した時点で、続けて「妖怪」と入力することを推薦する場合である。「河童」や「雪女」

などの個別の概念に深く言及することを意図しない文章では、これらの総称を挙げて話を展開することが多い。文脈によって、どの程度まとめた呼称を用いるのが適切かは変わってくる。

本システムでは、複数語からなるクエリの入力に対応しないため、「河童」と「雪女」に分けて、それぞれをクエリとした。その結果、「河童」の関連語は 229,265 語あり、クエリと同一の語を除くと「妖怪」が 1 位になる。また「雪女」の関連語は 140,798 語あり、同じく「妖怪」が 1 位になる。

このように、上位概念の提示においては、pfbf を用いることで十分に適切な推薦が行えることが示唆された。

5. おわりに

本稿では、Web 事典である Wikipedia を利用した作文支援システムを提案した。関連度計算に pfbf を用いて関連語の取得し、出力例を作文支援の視点から考察した。今後の課題として、候補語を絞り込むルールの追加が必要である。それらを加え、利用者実験を行い、提示内容及び提示方法の妥当性を評価する予定である。

謝辞

本研究は科研費(21500091)の助成を受けたものである。ここに記して謝意を示す。

参考文献

- [1] 総務省情報通信政策研究所, ブログの実態に関する調査研究: <http://www.soumu.go.jp/iicp/chousakenkyu/data/research/survey/telemcom/2009/2009-02.pdf>
- [2] Wikipedia: <http://ja.wikipedia.org/>
- [3] Wikipedia 日本語版の統計: <http://ja.wikipedia.org/wiki/Wikipedia:日本語版の統計>
- [4] 中田充, 葛崎偉, 吉村誠: キーワードおよびその連想語による文例の検索に関する考察, 情報処理学会研究報告. 人文科学とコンピュータ研究会報告, Vol.2004, No.7, pp.5-12, 2007.
- [5] 鄭在玲, 三宅真紀, 馬越庸恭, 赤間啓之: グラフ理論を用いた日本語連想作文支援システムの開発と評価, 2006PC カンファレンス論文集, pp.27-30.
- [6] 伊藤雅弘, 中山浩太郎, 原隆弘, 西尾章治郎: Wikipedia からの連想ソーラス構築プロジェクト, 人工知能学会第 20 回セマンティックウェブとオントロジー研究会, SIG-SWO-A803-05, 2009.
- [7] 堀憲太郎, 大石哲也, 峯恒憲, 長谷川隆三, 藤田博, 越村三幸: Wikipedia への関連単語抽出アルゴリズムの適用とその評価, 情報処理学会研究報告, データベース・システム研究会報告, No.56, pp.81-88, 2008.