

Wikipedia を用いたマイナースポーツ 検索手法の提案

服部 祐基^{†1} 灘本 明代^{†2}

インターネットの普及に伴い、インターネット上には様々な情報が存在している。本研究では認知度や知名度が低く見つけにくい情報をマイナー情報とし、このマイナー情報をインターネットから検索する手法の提案を行う。本論文ではマイナー情報検索の第一歩としてドメインをスポーツに絞り Wikipedia からマイナースポーツを検索する手法の提案を行う。具体的には、ユーザの興味や関心のあるスポーツを入力として、それと類似し、且つマイナーであるスポーツをたとえ表現に基づくマイナー検索手法と類似情報に基づくマイナー検索手法の2つの手法を用いて検索するシステムを提案する。ユーザの興味や関心のあることから検索することでマイナー情報を見つけやすくすることを目的としている。

The Minor Sports Search System based on Wikipedia

YUKI HATTORI^{†1} and AKIYO NADAMOTO^{†2}

In this research, we propose the method of searching for the minor information which has low acknowledged-degree and low popularity. In this paper, the first step in searching for the minor information from the Internet, we propose the system which searches for the minor sports from Wikipedia. In our proposed system, user first inputs the name of the sports which he/she wants to know and the system searches for it which is similar to the user's input and minor sports. We propose two types of methods to extract minor sports from Wikipedia. One is the method of calculating similar content, the other is the method of based on example sentence.

1. はじめに

近年インターネットの普及により、Web 上には大量かつ様々な情報が溢れている。その中で現在の検索システムでは、認知度、知名度の高い情報は比較的簡単に見つけることができるのに対し、認知度、知名度の低い情報は見つけにくいという問題がある。原因として認知度、知名度の低い情報は、情報が少なく、どのように探せば良いかわからないという点が上げられる。しかしそのような情報の中にも有益な情報が多く含まれている。例えば、スポーツはメジャーなスポーツとマイナーなスポーツと分けられる。我々が良く知っているスポーツというのはメディアなどによく取り上げられているものが多く、多数あるスポーツの種類のごく一部にすぎない。多くの種類があるスポーツにおいてマイナーなスポーツというのは、知るきっかけがほとんどない。しかしそのようなマイナーなスポーツの中にも知らないだけで興味のあるものが含まれていると考えられる。そこで本研究では、認知度、知名度が低い情報をマイナー情報とし、マイナー情報を検索し提示するシステムの提案を行う。具体的には、ユーザの興味や関心のあることを入力とし、その入力されたキーワードと類似し且つマイナーな情報の検索を行う。ユーザの興味や関心のあることを入力することで、見つけにくいマイナー情報を探しやすくできる。これにより、自分の知らなかった情報や気づいていない情報を取得することができ、新たな知識や興味の発見が可能になる。本論文ではマイナー情報の検索の第一歩として、ドメインをスポーツに絞り、ユーザが入力した興味や関心のあるスポーツを用いて、そのスポーツに類似しているが、マイナーなスポーツを Wikipedia から検索し提示するマイナースポーツ検索の提案を行う。具体的には、「たとえ表現に基づくマイナー検索」と「類似情報に基づくマイナー検索」の2つの手法を用いてマイナースポーツを検索する。マイナースポーツ検索の大まかな流れを以下に、システムの流れを図1に示す。

- (1) ユーザは興味や関心のあるスポーツ名を入力する。
- (2) 「たとえ表現に基づくマイナー検索」として、ユーザの入力したスポーツ名をたとえている Wikipedia の記事を取得し、これをマイナー情報の候補とする。
- (3) 「類似情報に基づくマイナー検索」として、ユーザの入力したスポーツ名の記事と類

†1 甲南大学大学院 自然科学研究科
Konan University

†2 甲南大学 知能情報学部
Konan University

似する Wikipedia の記事を取得する。この時、類似する Wikipedia の記事はリンク解析を行うことにより決定する。そして、この Wikipedia の記事の編集回数、編集人数が少ない物をマイナー情報と仮定し、抽出する。抽出された Wikipedia の記事をマイナー情報の候補とする。

- (4) (2)と(3)で取得したマイナー情報の候補からスポーツ名のフィルタリングを行う事により、スポーツの記事のみを抽出し、それらをマイナースポーツとして提示する。

以下、第2章に関連研究を、第3章にたとえ表現に基づくマイナー検索、第4章に類似情報に基づくマイナー検索、第5章にスポーツ名のフィルタリングについて提案し、第6章に提案手法を用いて作成したプロトタイプシステムについて、第7章に評価実験、第8章でまとめと今後の課題を述べる。

2. 関連研究

大島ら¹⁾は文書群をクエリとし、それと似ているが異なる文書の検索を提案している。我々の提案はクエリから、類似し異なるものを検索する点においては似ているが、本研究では一つの文書をクエリとし、それと類似しているが異なり、かつ認知度、知名度が低いものを見つけてくる点においては異なる。

情報検索の際に Google や Yahoo!などを用いると検索結果が下位に出てくるページというのは見つけるのが難しくマイナーな記事となるが、そのようなページにおいても有益なページは含まれている。そこで近年そのようなページを上位に上げるためのリランキング手法^{2) 3)}が重要となっている。多くのリランキング手法を用いた情報検索の研究ではユーザーの嗜好に合った情報を検索できるように、Web 閲覧履歴や検索結果に対するユーザーの操作などを用いている場合が多い。ランキングが下位のページを上位に上げることにより、マイナーな記事を取得できる。それに対してマイナーな記事を取得するところでは類似しているが、本研究では対象ページを Wikipedia に絞る、またパーソナライズを目的としない点異なる。また竹原ら^{4) 5)}は Web 検索エンジンでページのランク付けのためによく用いられているリンク構造を解析した手法では、マイナーなページが表示されにくいという点を問題にしている。リンク構造を解析する手法では、より多くの他のページからリンクされているページがより上位にランク付けされるという点が挙げられる。より多くの他のサイトや人に知られている有名なページが検索結果の上位に表示されるということになる。ユーザーは有名なページを欲しているとは限らず、マイナーだがそのユーザーにとって有用であ

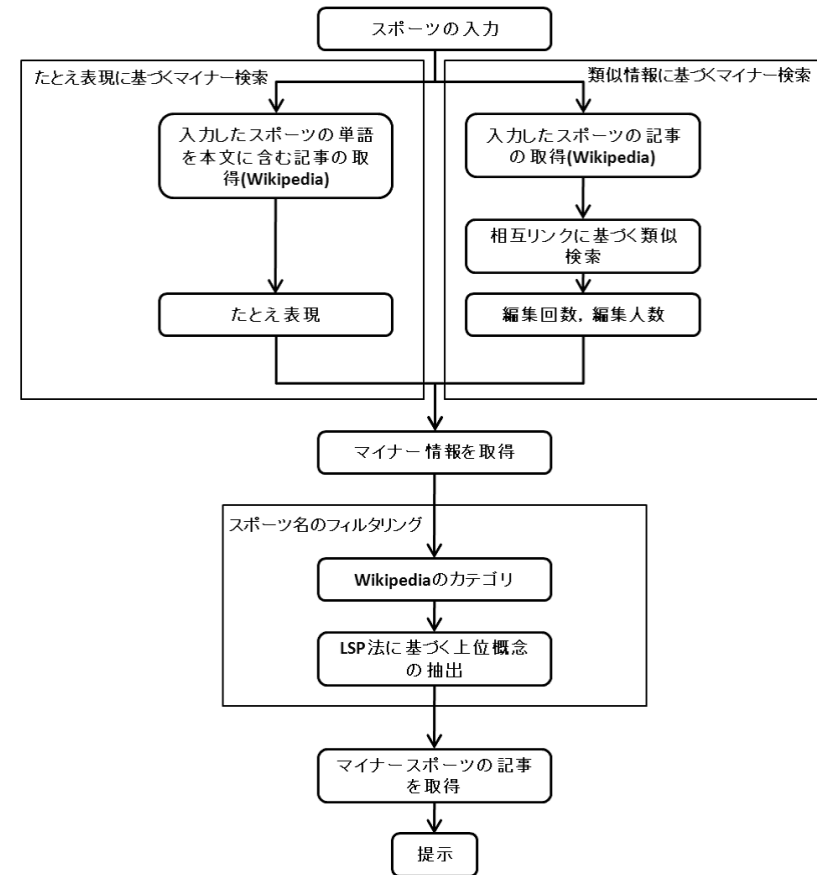


図1 システムのフロー
Fig.1 System Flow

るページが提示されないという問題を解決するために、Blog サイトの解析に基づいた検索システムの構築手法について提案している。マイナーで有益なページを検索する点においては非常に似ているが本研究では Blog ではなく Wikipedia を用い、ドメインをスポーツに絞っている点において異なる。

3. たとえ表現に基づくマイナー検索

本論文ではマイナー情報検索のはじめの一步として、Wikipedia からマイナースポーツを抽出する事を行う。マイナースポーツの Wikipedia の記事の特徴として、マイナースポーツを説明するために他の有名なスポーツに例えて表現している場合が多い。この特徴を我々は「たとえ表現」と呼び、この手法を用いて、ユーザの入力したスポーツ名からマイナースポーツを抽出する。「たとえ表現」の例として、「セパタクロー」というマイナースポーツの Wikipedia の記事では、「足のバレーボール」や「ルールはバレーボールと似て」というようにバレーボールに例えて表現がされている。また「ローンボウルズ」では「ボウリングの前身」と表記されており、このようにたとえ表現を用いた文を読むだけでどのようなスポーツであるかおおよそのイメージをすることができる。そこで「~と似て」や「~の前身」のような表記を本研究ではたとえ表現と呼び、この手法を用いてマイナースポーツの取得を行う。表 1 にたとえ表現の例と表記例を示す。たとえ表現を表す語はたくさんあり、「~に類似」や「~に似て」などの類似系や、「~の前身」や「~派生」などの起源を用いた表し方、「~と...を組み合わせた」や「~と...を融合した」などの複数のスポーツを組み合わせる方法、「~で行う...」などの特定の場所や特別な道具を用いて表現されている場合などがある。よってユーザが入力したスポーツに対して、たとえ表現である「に類似」などを拡張クエリとして付加し、検索を行う。例えば、入力が「バレーボール」であった場合には、「バレーボールに類似」や「バレーボールと似て」、「バレーボールの原型」のように拡張クエリを付加し、検索を行う。また中には「~と良く似て」などと強調語を用いて表記されている場合もあるため、「~と似て」と「~と良く似て」のように強調語がある場合の 2 パターンを用いて検索を行う。そして複数のスポーツを組み合わせている場合、例として「~と...を融合」の場合には、「バレーボールと*を融合」と「とバレーボールを融合」のように 2 通りのクエリを生成することができる。これは、入力されたスポーツが「~」に含まれている場合と「...」に含まれている場合の 2 パターンがあると考えられるためである。ここで「*」はワイルドカードを示す。これらのたとえ表現を検索する際には Yahoo!Api を用い Yahoo!検索により、たとえ表現を用いて書かれている Wikipedia の記事を取得する。これは Yahoo!Api を用いた方が検索の精度は良いためである。ここで取得された記事をマイナー情報の候補とする。

表 1 たとえ表現と表記例
 Table 1 The Expression Of Example and The Notation Example

たとえ表現の例	表記例 (記事)
~に類似	サッカーに類似して (バンディ)
~と似て	バレーボールと似て (セパタクロー)
~の原型	ゲートボールの原型 (クロッキー)
~の前身	ボウリングの前身 (ローンボウルズ)
~の派生	テニス等の派生 (スカッシュ)
~と...を組み合わせた	クロスカントリーとライフル射撃を組み合わせた (バイアスロン)
~と...を融合	チェスとボクシングを融合したスポーツ (チェスボクシング)
~で行う...	自転車で言うサッカー (サイクルサッカー)

4. 類似情報に基づくマイナー検索

すべてのマイナースポーツが Wikipedia 上でたとえ表現を用いているとは限らないため、たとえ表現に基づくマイナー検索だけではマイナー情報を取得しきれない。そこで我々は新たにユーザの入力したスポーツと類似しているがマイナーであるスポーツを取得することを行う。入力したスポーツと類似した記事を抽出することを提案する。一般に類似検索には文書間の類似度を用いる場合が多いが、ここではマイナー情報検索の計算スピードを上げるため、記事間のリンク解析と編集回数・人数に注目する。具体的には、記事間の相互リンクに基づく類似検索と類似検索から得られた記事からマイナー情報を取得する編集回数、編集人数の 2 つの手法を用いて行う。

4.1 相互リンクに基づく類似検索

スポーツの記事と相互リンクの関係にある記事は、関係が強く類似している可能性が高いと考え、リンク構造の中で相互リンクに着目し、入力したスポーツの記事と相互にリンクしている記事を抽出する。しかし相互リンクのみを用いて類似検索を行うと、そのスポーツの選手や用語、道具等の必要のない情報まで多く取得してしまう問題がある。実際に相互リンクをしていて類似しているページを見てみると、お互い同じページをリンクしている場合が多い。そこで相互リンクをしており、且つ共通のリンクのページを持つ割合が高い記事は類似していると仮定し、相互リンクだけでなく共通のリンクにも着目する。ここで共通のリンクの数が多記事を類似した記事とするのではなく、共通のリンクを持つ割合が高い記事を類似した記事とする理由として、記事の文章量が多くなればなるほどリンクの数が増え共通のリンクが多くなる問題がある。そこで共通のリンクを持つ割合が高い記事を類

似た記事とすることで、マイナー情報の特徴である文章量が少ない記事でも抽出することが可能となる。本研究では相互リンクの関係にあり共通のリンクの割合がある閾値 α 以上のものを類似した記事としマイナー情報の候補とする。ここで取得したマイナー情報の候補の中にはスポーツ名以外の記事も含まれている。例えばスポーツ選手やスポーツ用語、道具などその入力したスポーツと関わりのある記事が取得されている。

4.2 編集回数，編集人数

類似した記事の中からマイナー情報を抽出するために、本研究では Wikipedia の編集回数と編集人数に着目する。メジャーなスポーツは知っている人も多く、Wikipedia の記事では編集する回数や人数が多くなるが、マイナーなスポーツは知っている人が少ないため Wikipedia の記事では編集する回数や人数が共に少なくなる場合が多いと考えられる。そこで本研究ではマイナーなスポーツの特徴として、編集回数，編集人数が共に少ない傾向があるという仮説を立て、この仮説を実証するために実験を行った。150 個のスポーツの Wikipedia の記事を用意し、各々の記事の編集回数と編集人数を調べる。また被験者 13 名に対し、150 個のスポーツを見せ、各スポーツに対して以下 1~3 の 3 段評価をしてもらった。

- (1) 知っている，だいたい知っている。
- (2) 名前は知っている（名前は知っているが何をするのかわからない）
- (3) 知らない（名前すら聞いたことがない）

この被験者実験の例を表 2 に示す。各々のスポーツに対して編集回数と編集人数，3 段評価の平均，各評価の人数を載せている。スポーツの Wikipedia の記事の編集回数，編集人数と被験者実験から得られた評価の平均との関係の結果を図 2，図 3 に示す。X 軸にそれぞれ編集回数と編集人数を，Y 軸には評価の平均としている。

実験結果より、編集回数，編集人数共に回数，人数が増えると、知っている人が多くなり、回数，人数が減ると共に、知らない人が多くなっていることがわかった。これにより、メジャーなスポーツは編集回数と編集人数が多く、マイナーなスポーツは編集回数，編集人数が共に少なくなることが判明した。この結果より、相互リンクに基づく類似検索により得られたマイナー情報の候補の記事の編集回数，編集人数を調べ、共に閾値 β 以上の記事はメジャーであるとして削除する。ここで実験より閾値 β は編集回数を 180，編集人数を 80 とする。そして残った記事をマイナー情報として取得する。ここで得られたマイナー情報の中にもスポーツではない情報が含まれる。

表 2 実験の例

Table 2 The Example Of Experiment

No	スポーツ名	編集回数	編集人数	平均	評価値 1(人)	評価値 2(人)	評価値 3(人)
1	フットサル	299	152	1	13	0	0
2	ハンドボール	601	236	1	13	0	0
3	水球	248	145	1.1538	11	2	0
4	ボロ	70	52	2.6923	2	0	11
5	ロデオ	94	54	1.3846	8	5	0
6	カボエイラ	288	178	1.7692	5	6	2
7	フェンシング	186	127	1	13	0	0
8	パイアスロン	94	67	2.3846	3	2	8
9	アーチェリー	162	106	1	13	0	0
10	ダーツ	238	159	1	13	0	0
11	ラクロス	266	156	1.1536	11	2	0
12	スカッシュ	22	16	1.2308	10	3	0
13	リュージュ	53	43	2.5385	2	2	9
14	スノーカイト	10	8	2.9231	0	1	12
15	カヌーボロ	78	46	2.9231	0	1	12
:	:	:	:	:	:	:	:

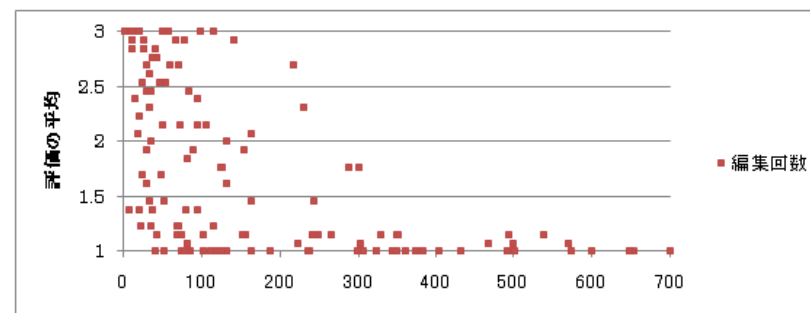


図 2 編集回数と評価平均の関係

Fig. 2 The Relations With The Editing Number of Times and The Average Of The Evaluation

5. スポーツ名のフィルタリング

3, 4 章で得られたマイナー情報の中には、スポーツ選手やスポーツ用具などのスポーツ名でない記事も取得されている。よってマイナー情報からスポーツの記事のみを抽出するこ

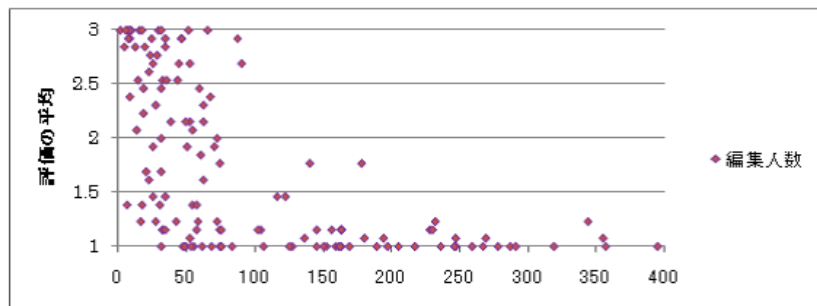


図3 編集人数と評価平均の関係

Fig.3 The Relations With The Editing Number of People and The Average Of The Evaluation

とを行う．ここで (1)Wikipedia のカテゴリ, (2)LSP 法に基づく上位概念の抽出の 2 つの方法を用いてスポーツ記事の抽出を行う．以下それぞれの方法について説明していく．

5.1 Wikipedia のカテゴリ

Wikipedia の記事にはすべてカテゴリ分けがされている．それを用いてマイナー情報からスポーツ記事の抽出を行う．本研究ではスポーツと判断するカテゴリを「スポーツ」、「スポーツ競技」、「オリンピック競技」、「団体競技」などとする．例えば「サッカー」のカテゴリは「サッカー」、「オリンピック競技」と分けられている．「オリンピック競技」からサッカーはスポーツであると判断することができる．しかし「セパタクロ」のカテゴリは「セパタクロ」となっている．これでは本来スポーツであるものがスポーツと判断することができない．そのため記事のカテゴリのもう一つ上の上位 2 階層のカテゴリを見ることとする．「セパタクロ」の上位 1 階層のカテゴリは「セパタクロ」だが，その上の上位 2 階層のカテゴリを見ると「球技」、「タイのスポーツ」、「団体競技」と分けられ「団体競技」からセパタクロはスポーツと判断することができる．多くのスポーツの記事を調べた結果上位 2 階層のカテゴリまででスポーツかどうかを判断できる．スポーツに対しては，この方法によりスポーツと判断することができたが，スポーツの道具や用語などスポーツに深く関係のある記事にも上位 2 階層のカテゴリにスポーツと判断するカテゴリが含まれてしまう問題があり，これらを削除する必要がある．例えば「ボール」のカテゴリには「球技」が含まれており，球技の上位のカテゴリには「スポーツ競技」とスポーツと判断するカテゴリが含まれていたため，ボールはスポーツと判断されて取得してしまう問題がある．そこでスポー

ツの道具や用語などを削除するために，LSP 法を用いる．

5.2 LSP 法に基づく上位概念の抽出

Wikipedia のカテゴリを用いて取得されたスポーツ以外の記事を削除するために，本研究では中山ら^{6) 7)}の提案している LSP 法を用いる．中山らは Wikipedia 内の記事における文法に着目し，記事のリード部分を重要文と見なして解析する手法である LSP 法を提案している．Wikipedia 内の文章は多くの場合冒頭文が is-a 関係にある．この関係を利用し単語の上位概念を抽出することを行う．具体的には，記事の 1 文目の最後に出てきた名詞をその記事の上位概念とする．この手法を用いると例えば「ボール」の Wikipedia の記事の 1 文目は「ボール (ball, 玉) はゲーム (球技や遊戯) などに使う球形の用具」となっており，最後に出てきた名詞である「用具」がボールの上位概念となる．この「用具」はスポーツではない．よって最後に出てきた名詞が「用具」や「大会」、「道具」、「協会」、「連盟」などのスポーツではない単語が上位概念として取得された場合には，スポーツではないと判断し削除し，残った記事をマイナースポーツの候補とし取得する．ここで「スポーツのこと」などのように表記されていた場合，最後の名詞は「こと」になる．これではスポーツである記事がスポーツと判断できず，削除されてしまう問題がある．そのため最後に出てきた名詞が「こと」、「事」、「一つ」、「ひとつ」、「名称」などであった場合は，一つ前の名詞を上位概念とする．先ほどの「スポーツのこと」という場合には，最後の名詞が「こと」になっているため，一つ前の「スポーツ」が上位概念となり，スポーツと判断できるため削除を行わない．これらの手法により取得した記事をマイナースポーツの記事とし，ユーザに提示する．

6. プロトタイプシステム

我々の提案手法を用いてプロトタイプシステムを作成した．プログラミング言語には Ruby を，ユーザインターフェースには CGI を用いた．システムの起動画面と出力例を図 4 と図 5 に示す．まずユーザに好きなスポーツを入力してもらい，決定をクリックするとシステムは手法を用いて検索を行う．そして入力に対して得られたマイナースポーツのリストを表示している．また Yahoo!検索により得られたサマリを同時に載せることでそのスポーツの大体の意味がわかるようになっている．ここでマイナースポーツを選択すると Wikipedia の記事を見ることができる．これによりそのスポーツについてより詳しく知ることができる．



図4 プロトタイプシステムの画面
Fig. 4 Display of Prototype System

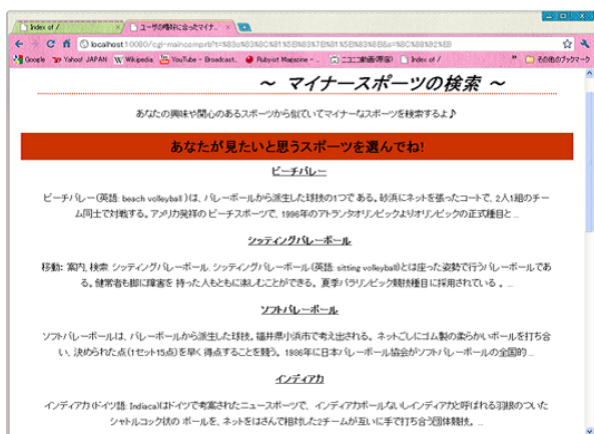


図5 出力画面
Fig. 5 Output

7. 実験

7.1 編集回数, 編集人数によるたとえ表現の評価

たとえ表現で得られた記事がマイナー情報であるかを検証するために編集回数と編集人数

を用いて実験を行った。ここでは編集回数と編集人数を用いて、たとえ表現の評価を行う。図2と図3よりマイナー情報と判定する編集回数と編集人数の閾値をそれぞれ180, 80とし、提案手法であるたとえ表現から取得された記事のマイナー度の評価を行う。任意の3種類のスポーツを入力とし、たとえ表現より取得されたスポーツ記事の編集回数と編集人数との関係を表3に示す。表3よりたとえ表現から得られたスポーツの記事の多くがそれぞれの閾値を下回る結果となったため、提案手法であるたとえ表現によるマイナー情報の取得方法は有用だと示すことができた。

表3 たとえ表現の検証

Table 3 The Inspection Of The Expression Of Example

入力	マイナー情報	編集回数	編集人数
サッカー	バンディ	38	28
	フットサル	299	152
	サイクルサッカー	26	12
	電動車いすサッカー	33	21
	フリースタイルフットボール	5	3
バレーボール	セバタクロ	77	47
	インディアカ	43	25
	シッティングバレーボール	41	24
	ソフトバレーボール	20	14
バスケットボール	ビーチバレー	135	80
	ストリートバスケットボール	44	24
	ビーチバスケットボール	6	5
	ウォーターバスケットボール	6	6
	カヌーボロ	79	47

7.2 評価実験と考察

本提案手法の有用性を示すため、プロトタイプシステムを用いて評価実験を行った。任意の15種類のスポーツに対してマイナースポーツ検索を行った結果、適合率は71パーセントと良い結果が得られた。しかしスポーツによりマイナースポーツがない場合もあり、スポーツにより適合率は大幅に差が出る結果となった。また出力されたマイナースポーツの中にもまだスポーツではない記事が含まれていた。これは提案手法であるスポーツ名のフィルタリングの精度に問題があったためである。中でもLSP法に基づく上位概念の抽出において、Wikipediaの1文目の最後の名詞を上位概念として抽出し、スポーツではないものを削除することを行ったが、Wikipediaの文の構成上削除しきれなかったことが原因として

上げられる．そのためスポーツの記事ではない記事がスポーツとして出力されてしまった．

8. まとめと今後の課題

本論文では見つけにくいマイナー情報をユーザの興味や関心のあるスポーツからマイナースポーツを検索する手法の提案を行った．マイナーなスポーツを検索し提示することを行うことで、ユーザの新たな知識の取得が可能である．また本システムを用いることでマイナースポーツの発展に繋がることを期待している．本研究を進めていきいくつかの問題点が見つかったので、今後の課題として考えていく．今後の課題として以下が挙げられる．

● スポーツ名のフィルタリングの精度の向上

出力されたマイナースポーツの中には、スポーツ以外の記事がまだ多く残っていた．この原因として、LSP 法を用いた手法で、うまく上位概念が取れていなかったことが上げられる．またスポーツ用語のフィルタリングにおいても、スポーツ用語を抽出する際に取り切れていない用語も多数あった．これらの手法を改善することでスポーツ以外の記事を削除することができ、適合率の改善ができると考えられるが、その方法は現在のところ検討中である．

● 出力結果の改善

本システムでの出力では、マイナースポーツのリストと Yahoo!検索により得られたサマリのみを提示しているが、マイナースポーツの画像を一緒に載せることでそのスポーツをよりイメージしやすかったのではないかと考えた．しかし画像検索によりマイナースポーツを検索した結果、マイナーであるために画像が取得できないということがわかった．そのため Wikipedia の記事に画像が使われていた場合にはその画像も一緒に提示することも考えている．またマイナーなスポーツにもそのスポーツをメジャーとして考えている人もいる．そのメジャーとしている人が集まっている SNS などのコミュニティでの話題などを載せることで、そのスポーツのことがより理解し易くなるのではないかと考えている．

● その他の分野への応用

本研究ではスポーツのみを対象として行っているため、今後は他の分野にも応用していくことを考えている．Wikipedia のみを用いてマイナー検索をしていたので今後は Wikipedia だけでなく、SNS なども用いていきたい．

● 評判情報の付加

出力されたスポーツに対してより興味を持ってもらうために、評判情報を付け加えるこ

とを考えている．SNS のコミュニティを用いることで、そのコミュニティ内でのそのスポーツに対しての評判を抽出し、ユーザに提示する．これによりユーザがそのスポーツに対してより興味や関心が増すと考えられる．

参 考 文 献

- 1) 大島裕明, 小山 聡, 田中克己: 文書群をクエリとした "似て非なる" 文書の検索, 日本データベース学会論文誌 (DBSJ Letters), Vol.5, No.1, pp.121-124 (2006).
- 2) 山本岳洋, 中村聡史, 田中克己: 編集操作を用いたウェブ検索結果の最適化, データ工学ワークショップ (DEWS2007), Vol.17 (2007).
- 3) 松岡研二, 范 薇, 小柳佑介, 渡邊豊英: Web ページに対するユーザの適・不適の評価を用いた検索結果のリランキング, *DEIM Forum 2010* (2010).
- 4) 竹原幹人, 中島伸介, 角谷和俊, 田中克己: Web 情報検索のための Blog 情報に基づくトラスト値の算出方式, *DBSJ Letters*, Vol.3, No.1.
- 5) 中島伸介, 竹原幹人, 日野洋一郎, 館村純一, 原良憲, 田中克己: blog 解析に基づく Web 情報検索の信頼性向上技術, 人工知能学会セマンティック Web とオントロジー研究会, Vol.6 (2004).
- 6) 中山浩太郎, 原 隆治, 西尾章治郎: 自然言語処理とリンク構造解析を利用した Wikipedia からの Web オントロジー自動構築, 日本データベース学会論文誌, Vol.7, No.1, pp.67-72 (2008).
- 7) Nakayama, K., Hara, T. and Nishio, S.: Wikipedia Mining - Wikipedia as a Corpus for Knowledge Extraction, *Proceedings of annual Wikipedia Conference (Wikimania)* (2008).