

## 説明の観点に基づくテキストの分類と要約

藤井 敦<sup>†</sup> 三條場 旭彦<sup>‡</sup>

<sup>†</sup> 筑波大学 大学院図書館情報メディア研究科

〒 305-8550 つくば市春日 1-2

<sup>‡</sup> 筑波大学 図書館情報専門学群

〒 305-8550 つくば市春日 1-2

科学技術や文化の急速な発展によって、言葉や事柄について World Wide Web 上のツールを用いて調べる機会が増えている。検索エンジンは情報の量が多いものの、情報が統制されておらず質が低い。人手で編集する事典は情報の質が高いものの、情報の量が制限される。両者の長所を統合するために、筆者らは、Web 情報や特許情報から説明テキストを抽出し、体系化する研究を行っている。本研究は、ある見出し語について説明した複数のテキストを観点に基づいて分類することで、多面的な要約を生成する手法を提案する。動物名や病名といった見出し語の種類によって説明に必要な観点が異なるため、人手による手法では大規模化が困難である。そこで、Wikipedia から見出し語の種類ごとに観点の構造に関するテンプレートを抽出する。さらに、Wikipedia の記事を機械学習のデータとして利用して、与えられた説明テキストを適切な観点に分類する。評価実験によって本手法の有効性を示す。

## Viewpoint-based Text Categorization and Summarization

Atsushi Fujii<sup>†</sup>, Akihiko Sanjouba<sup>‡</sup>

<sup>†</sup> Graduate School of Library, Information and Media Studies, University of Tsukuba

<sup>‡</sup> School of Library and Information Science, University of Tsukuba

Reflecting the rapid growth of science, technology, and culture, it has become common practice to consult tools on the World Wide Web for various words and things. Existing search engines provide an enormous volume of information, but retrieved information is not organized. Hand-compiled encyclopedias provide organized information, but the quantity of information is limited. To integrate the advantages of both tools, we have been proposing methods to extract and organize information on the Web and patent information. In this paper, we intend to produce a summary on multiple viewpoints and propose a method to categorize multiple expository texts for a single headword based on viewpoints. Viewpoints required for explanation are different depending on the type of a headword, such as animals and diseases, it is difficult to manually produce a large scale system. We use Wikipedia to extract a template of a viewpoint structure for each headword type. We also use articles in Wikipedia for a machine learning method, which categorizes a given text into an appropriate viewpoint. We evaluate the effectiveness of our method experimentally.

## 1 はじめに

科学技術や文化の急速な発展によって、言葉や事柄について調べる機会が公私を問わず増えている。そのため、World Wide Web 上の様々なツールを使うことが多い。代表的なツールには、Google<sup>1</sup> や Yahoo!<sup>2</sup>などの「検索エンジン」と Wikipedia<sup>3</sup>などの人手で編集された「事典」がある。両者には、情報の量と質という点において、それぞれ長所と短所がある。

検索エンジンは、億単位のページ集合が検索の対象であり、提供される情報の量が多いという利点がある。しかし、検索される情報が体系化されておらず、必要な情報も含まれるため、情報の質が低い。

他方において、事典は、説明を目的とした情報に限定され、かつ項目等によって情報が統制されているため、情報の質が高いという利点がある。しかし、人手による編集に依存しているため、調べたい言葉が必ず登録されているとは限らない。また、説明の内容が著者の視点に偏るという問題もある。すなわち、情報の量において問題がある。

検索エンジンと事典の長所を統合することができれば、有用性が高い調べ物のツールを実現することができる。この目的のために、筆者らは、Web 情報や特許情報から言葉や事柄に関する説明情報を抽出し、さらに複数の説明情報を組織化することで、百科事典的なコンテンツを自動構築する研究を行ってきた [3, 4, 5, 8, 10]。

さらに、構築したコンテンツに対して、見出し語、関連語、同義語、質問文、関連語マップといった多様な手段で検索する機能を開発した [9]。当成果は、事典検索システム CYCLONE で公開している<sup>4</sup>。最近の研究 [3, 8] では特許情報も対象とし、Web に存在しない高度な専門用語も調べることが可能になった。

CYCLONE の基本機能では、調べたい言葉を入力して、その言葉に関して説明された 1 つ以上段落を検索することができる。しかし、検索された複数の説明は個別の Web ページや特許公報から抜粋した段落であるため、重複を含み冗長である。この問題を解消するために、自動要約手法 [4] を提案し、実装した。本研究は、自動要約をさらに発展させる手法を提案し、評価実験によって有効性について検討する。

<sup>1</sup><http://www.google.co.jp/>

<sup>2</sup><http://www.yahoo.co.jp/>

<sup>3</sup><http://ja.wikipedia.org/>

<sup>4</sup><http://cyclone.slis.tsukuba.ac.jp/>

## 2 本研究の位置付け

### 2.1 Cyclone における要約手法との比較

筆者らがすでに提案し、CYCLONE に実装されている要約手法 [4] は、説明の「観点」に着目し、ある見出し語について説明している複数の段落から、観点ごとに代表文を抽出する。

観点の設定と観点ごとに代表文を抽出する規則の作成は人手で行っている。現在は、情報処理用語を対象とし、岩波情報科学辞典 [7] などを参考にして、「定義」、「例示」、「目的」など 12 の観点を用いている。

図 1 に、見出し語「XML」について要約された説明情報を示す。図 1 に示された観点ごとに代表文を読むことによって、必要な情報を落とすことなく、複数の段落に含まれる冗長性を排除することができる。

しかし、上記の要約手法には 2 つの問題がある。1 つ目の問題は、観点の設定を人手で行う点にある。さらに、用意すべき観点は見出し語の種類によって異なる。例えば、病名について説明する場合は、「原因」や「症状」といった観点が必要であり、用意すべき観点の種類が情報処理用語と根本的に異なる。そこで、情報処理用語以外の見出し語に幅広く対応するためには、観点を設定する負荷が大きい。

2 つ目の問題は、観点ごとの代表文を抽出する規則を人手で作成する点にある。抽出規則は、特定の単語や表現を手掛かりとしている。例えば、見出し語の後に「とは」という表現があれば、「定義」に関する文として抽出する。しかし、1 つ目の問題で議論したように、見出し語の種類によって必要な観点が異なるため、新しい観点に対して常に規則を人手で作成しなければならず、負担が大きい。

以上より、現状の CYCLONE では、情報処理用語の要約に限定されている。本研究は、この問題を解決するために、拡張性が高い要約手法を提案する。本研究の特長は、Wikipedia から様々な見出し語の種類について観点の構造を抽出することで、見出し語ごとに観点を設定する手間と観点ごとに抽出規則を作成する手間を削減する点にある。

ここで、「Wikipedia の存在が前提であれば、自動要約を行わなくても Wikipedia の記事を読めばよいのではないか?」という疑問が生じるかもしれない。この問い合わせに対する答えは「No」であり、本研究には 2 つの意義がある。まず、Wikipedia の未登録語について、Wikipedia と同じような観点の構造で説明を

見出し語	要約
定義	本研究では、Web上のデータ記述言語であるXML(eXtensible Markup Language)を利用して、検索を行うシステムを分散させることで、元ページキヤッショウ
例示	XMLとは？Extensible Markup Languageのこと、情報を記述するための言語、例えばHTMLやCSSなどを、作成、記述する言語です。元ページキヤッショウ
同義	その内容はドキュメントの実データと体系情報が含まれている。元ページキヤッショウ
略語	XML(eXtensible Markup Language)はインターネット上で文書やデータを交換したり、配布したりするときの汎用のデータ記述言語です。元ページキヤッショウ
目的	さらに、XMLに基づくWebデータベースを実現するために必要な要素技術である、XMLデータの格納、問合せ言語、利用者インターフェースなどについて説明する。元ページキヤッショウ
歴史	XML-AQ: 2001年8月29日には、XMLによるマークアップ言語設計に際してのアクセシビリティ指針であるXML Accessibility Guidelinesの草案が公開されました。元ページキヤッショウ
機能	XMLを利用することで、従来のHTMLベースのWebアプリケーションとくらべ、より柔軟でスケーラビリティの高いシステムを構築することが可能になります。開発メンテナンス工数を削減できます。元ページキヤッショウ
要素	文書に関する基本的な要素を文書中に盛り込んでもらうことで、より検索しての精度を高められるのではないかと思われるためである。元ページキヤッショウ
製品	書誌情報、製品カタログ、スケジュール、アプリケーションの設定、ペクトルグラフィックス、数式、表なども、文書型と捉えることができる。元ページキヤッショウ
書籍	本書は、次世代のインターネット標準フォーマットとして、今、最も注目を浴びているXMLとその周辺技術について、わかりやすく解説した入門書です。元ページキヤッショウ
その他	XMLでは、タグは“<”と“>”で囲んで表します。元ページキヤッショウ

図 1: 見出し語「XML」について要約された説明情報

得ることができる。さらに、Wikipedia の登録語に 対しても、別の見出し語を説明するために使われた 観点を補ったり、一般の Web ページや特許情報から 幅広く説明を収集することができる。

## 2.2 Wikipedia を用いた要約手法との比較

Biadsy ら [1] は、人物情報の要約に Wikipedia を 利用した。 Wikipedia から人物に関する記事を収集し、人物の説明に使われる文のモデルを学習した。この手法は、 Wikipedia を用いて要約に使用する情報を 選別する点では本研究と類似する。しかし、 Biadsy らは説明の観点を使用していない点が本研究と異なる。また、 Biadsy らが人物情報の要約に限定していることに対して、本研究は見出し語の種類を制限せず、見出し語の種類に応じて異なる観点の集合を使い分ける点も異なる。

## 2.3 複数文書要約手法との比較

本研究の目的は、自動要約の中でも「複数文書要 約」 [6] に属する。複数文書要約に関する既存の研究において、観点に基づく研究の事例は少ない。サー

ビスに関する評判情報の要約 [2] では、例えば、ホテルに関する評判情報から、「立地」や「価格」などの 観点に基づいて要約が生成される。しかし、観点や 観点の名称は評判情報の内容から抽出される。それ に対して、本研究は、 Wikipedia を外部情報として 利用することによって、要約対象の情報からは得る ことができない情報を補間する点が異なる。同様に、 本研究は、 Clusty<sup>5</sup> のように Web 検索結果の内容を 分析してクラスタリングする手法とも異なる。

本研究で作成した観点の構造は、 CYCLONE に限らず、汎用的な応用が可能である。4.4 節では、既存の 検索エンジンと組み合わせて、検索結果を観点に基 づいて分類する応用方法を提案する。

## 3 提案する要約手法

本研究で提案する要約手法は、 Wikipedia から説明 の観点に関する構造のテンプレートを抽出し、 Web や特許情報から抽出したテキストを観点構造のテン プレートに基づいて体系化する。さらに、「病名」や

<sup>5</sup> <http://clusty.jp/>

「動物名」といった見出し語の種類に応じて、異なるテンプレートを作成する。ある見出し語について書かれたテキストが文や段落の単位で与えられると、見出し語の種類を特定し、その見出し語に対応する観点の候補から適切な観点を特定する。見出し語の種類や観点の候補は事前に決められているため、この問題は、分類先のカテゴリが事前に決められた文書分類(categorization)に帰着する。

本手法の概要を図2に示し、この図に基づいて手法について説明する。図2は、事前に実行するオフライン処理と、分類対象のテキストが与えられた段階で実行するオンライン処理に大別される。

オフライン処理では、「人名」、「動物名」、「病名」といった見出し語の種類ごとに、Wikipediaの記事から観点構造のテンプレートを作成する。ここでは、Wikipediaの記事にある「目次」に着目する。図3は「破傷風」に関する Wikipedia の記事である。図3の「目次」には、「1. 原因」、「2. 症状」、「3. 治療」などの項目が並んでいる。本研究では、1つの項目を1つの観点として使用する。

しかし、あらゆる病名の記事で目次の項目が完全に一致するわけではない。そこで、ある見出し語の種類(例えば「病名」)ごとに複数の記事を収集し、使用頻度が高い目次項目を観点として選択する。観点の選択基準として、使用頻度に関する閾値を設定する方法や、使用する観点数の上限を決めておく方法があり、目的等に応じて適宜使い分ける必要がある。5章の評価実験では、使用頻度が高い上位5件の目次項目を観点として使用した。

さらに、見出し語や観点を分類するために、該当する Wikipedia の記事集合から複数の分類器を学習する。ただし、分類の目的によって使用する学習データが異なる。見出し語の種類を分類するためには、見出し語の種類ごとに記事集合をまとめて、1つのカテゴリに対応する学習データを作る。他方で、観点を分類するためには、観点ごとに記事集合をまとめて、1つのカテゴリに対応する学習データを作る。

なお、前処理として、「概要」や「事件の概要」のような目次项目的異表記を統一する必要がある。現在は、表層に基づく簡単な照合しか行っていない。今後は意味的な異表記に対する検討が必要である。

本研究では、機械学習にサポートベクターマシン(SVM)を使用する。SVM light<sup>6</sup>を使用し、さらに

<sup>6</sup><http://svmlight.joachims.org/>

多値分類に拡張した。多値分類への拡張方法、素性、値にはそれぞれ複数の選択肢がある。予備実験の結果、One-Vs-Rest法を使って多値分類に拡張し、単語ユニグラムを素性として、値は全て定数とした場合に分類精度が最も良かったため、これらを採用する。この他に、多値分類としてPairwise法、素性として単語バイグラム、値としてTF.IDFなども試した。

オンライン処理では、「りんご病」などの見出し語について書かれたテキストを文や段落の単位で入力し、「見出し語分類」と「観点分類」を順番に実行する。見出し語分類では、見出し語に関する分類器を用いて、「りんご病」の説明テキストを「人名」、「動物名」、「病名」のいずれかに分類する。観点分類では、分類された見出し語の種類に応じて観点を分類する。図2の例では、「病名」と分類された場合、説明テキストを病名に対応する「原因」、「症状」、「治療」のいずれかに分類する。

## 4 応用方法

### 4.1 概要

本研究の目的は、Webや特許情報から収集した種々の情報を説明の観点に基づいて体系化する点にある。しかし、3章で提案した手法の主旨は、与えられた文字列を何らかの観点に分類するという抽象度の高い処理である。本研究の目的を達成するために3章の手法を応用する方法は1通りではない。運用状況等に応じて使い分ける必要がある。4.2～4.4節では、3通りの応用方法を提案する。後になるほど、CYCLONEでの運用方法から逸脱する度合いが大きくなる。

### 4.2 CYCLONEの要約手法を自動化する

CYCLONEに実装されている要約手法[4]において、現在は人手に依存している部分を3章で提案した自動手法に置き換える方法がある。具体的には、ある見出し語に関する説明の段落を文に分割し、文単位で観点に分類して、最後に観点ごとに代表文を選択する。CYCLONEの枠組みでそのまま利用できるという利点があるものの、文単位では素性が少ないため分類精度が高くならない可能性がある。また、1つの観点が複数の文にわたって説明されている場合には、文単位での分類は本質的に困難である。

### 4.3 CYCLONEの段落を観点に分類する

CYCLONEで検索される段落の単位で観点に分類する方法がある。CYCLONEでは、段落を「機械」や

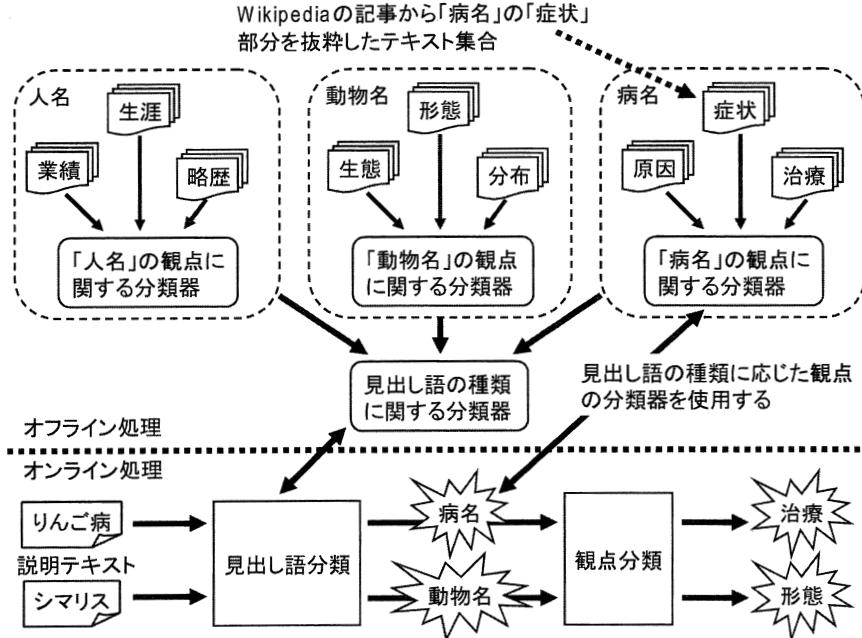


図 2: 観点に基づいてテキストを分類する手法の概要

「化学」といった技術分野や「スポーツ」や「芸能」といったジャンルに分類し、ユーザは情報を絞り込むことができる。これと同じ発想で、段落を観点に分類することで、ユーザは自分が知りたい観点に基づく説明だけを選択的に取得することが可能になる。ただし、1つの段落に複数の観点が混在する場合があるので、観点に順位を付けて、上位から一定数の観点を段落に付与する必要性が生じる可能性がある。

具体例を用いて説明する。ここで示す例は、5章の評価実験で構築した分類器に基づいて得られた結果である。以下は、CYCLONEで「ハクビシン」について検索された段落である<sup>7</sup>。説明の都合上、句点で文に分割し、文単位で箇条書きする。

1. 先日から家の屋根裏で物音がしています。
2. 一昨日、ふと見上げたら天井のすきまから「ハクビシン」が顔を出していました。
3. ハクビシンとはジャコウネコ科の動物で全長76 cm、体重約6 kg。

4. 鼻から頭にかけて白い筋があるのが特徴。
5. 東アジア中南部と日本に分布。
6. 地元では「むじな」ともいう。
7. ..... だそうです。

ここで、3文目はハクビシンの「形態」に関する説明であり、5文目は「分布」に関する説明である。この段落は、見出し語分類によって「動物名」に分類され、さらに観点分類によって「形態」、「分布」、「分類」、「人間との関係」、「生態」という順番に観点が付与された。上位2つの観点は、当該段落の観点を適切に捉えており、どちらも付与することが好ましい。なお、分類器を学習したデータには「ハクビシン」に関する Wikipedia の記事は含まれていなかった。

#### 4.4 一般の検索エンジンと組み合わせる

一般的の Web 検索エンジンで検索されたページ集合を観点に基づいて分類する方法がある。CYCLONE とは独立しているものの、収集した種々の情報を説明の観点に基づいて体系化するという本研究の目的は達成される。

<sup>7</sup>[http://www.irem.co.jp/omake/koma4/mess\\_067.html](http://www.irem.co.jp/omake/koma4/mess_067.html)

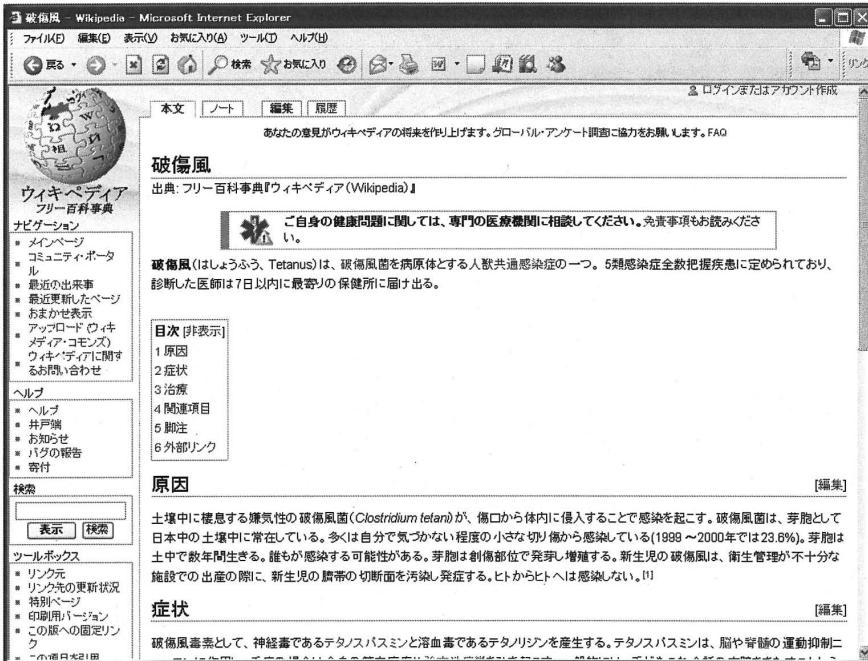


図 3: 「破傷風」に関する Wikipedia の記事

具体例を用いて説明する。Yahoo!で「ハクビシン」を検索した結果の上位 30 件からスニペット（検索結果一覧に示されるテキストの断片）を分類した。分類には、4.3 節と同じ分類器を用いた。SVM に基づく One-Vs-Rest 法では、分類結果ごとにスコアが計算されるため、観点ごとにスコアが最大のスニペットを抽出した。実行結果を付録 A に示す。付録 A の結果では、スニペットは適切な観点に分類され、5 つのスニペットを読めばハクビシンについて多面的な説明が得られることが分かった。

Clusty で「ハクビシン」を検索し、クラスタリングした結果、「ジャコウネコ科」や「被害、防止」などの事典項目らしいクラスタが少数生成された一方で、「ハクビシン・アライグマ」や「日記」のように事典と無関係なクラスタも生成された。本手法の方が事典的な利用に有利であることを示す例である。

## 5 評価実験

2008 年 7 月 24 日にダウンロードした Wikipedia の日本語記事を用いて実験を行った。Wikipedia で

は、見出し語が階層的なカテゴリで一部整理されているものの、完全ではない。また、階層の中から適切な粒度のカテゴリを決定しなければならない。

今回の実験では、見出し語の種類として表 1 に示す 4 種類を使用した。次に、見出し語の種類ごとに記事を収集し、さらに目次項目に従って観点に分割した。最後に、見出し語の種類ごとに使用頻度が高い目次項目の上位 5 件を観点として使用した。

Wikipedia は、「主要カテゴリ」や「スタブ」などのページで、カテゴリごとに見出し語を一覧している。これらのページから各見出し語の記事を収集した。表 1 では、見出し語の種類が互いに排他的になるように記事を収集した。表 1 には、見出し語の種類ごとに記事数や観点数も示している。本実験において、1 つの「記事」とは、ある見出し語の記事から 1 つの観点について抜粋したテキストである。そのため、表 1 では見出し語数よりも記事数が多い。

本実験では、Wikipedia の記事が本来のカテゴリ（見出し語の種類や観点）に正しく分類される正解率を評価した。実際の運用では、説明でないテキスト

表 1: 見出し語分類と観点分類の実験結果

見出し語の種類	見出し語数	記事数	見出し語分類の正解率	観点分類の正解率
動物名	468	1645	99.0%	95.4%
日本の事件	436	513	96.1%	66.1%
病名	403	947	93.5%	85.0%
数学用語	391	532	93.0%	62.4%

も含む雑多な情報が対象となるのに対して、本実験では説明として整ったテキストを対象としており、限定的な評価である。しかし、一度に全ての項目について評価することは困難であり、研究の第1段として妥当な評価実験であると考える。

見出し語分類と観点分類とともに、収集した記事を5分割して交差検定を行った。見出し語分類と観点分類の正解率を表1に示す。正解率とは、自動手法によって1位に順位付けされたカテゴリが Wikipedia に記載されたカテゴリと一致した割合である。

表1より、見出し語分類の正解率は90%以上となり、良好な結果が得られた。しかし、見出し語の種類は本実験で使用した4種類以外にも存在するため、規模を拡張しながら今後も評価実験を行う必要がある。

次に観点分類の結果について考察する。実際の運用では、見出し語分類を誤れば適切な観点の候補が得られないため、観点分類も必然的に誤ってしまう。しかし、本実験では観点分類そのものの精度を評価するために、全ての入力記事に対して見出し語の種類が正しく特定されたと想定した。

表1より、見出し語の種類によって観点分類の結果にばらつきがあることが分かった。最も良かった「動物名」に対する観点分類の内訳を表2に示す。「分類」以外の観点は、96%以上の正解率で分類することができた。

観点分類を誤る原因は大きく3種類ある。まず、本質的に分類が難しい場合である。例として、「病名」に対する観点分類の内訳を表3に示す。「診断」と「検査」という観点は記述の内容が互いに類似する場合があるため、他の観点に比べて分類の正解率が低い。

次に、学習に用いる記事数が少ない場合である。表2と表3より、記事数が100件程度の観点は分類の正解率が低い。今回、ある見出し語の種類に該当する記事を Wikipedia から全て収集したわけではなく、網羅的な収集方法について検討する必要がある。

最後に、分類するための素性が十分に抽出できな

い場合である。例えば、「数学用語」では、数式や記号などを除くと、素性である単語の数が少なくなった。

表 2: 「動物名」に対する観点分類の正解率

観点の種類	記事数	正解率
生態	408	96.5%
形態	388	97.1%
分布	387	99.2%
人間との関係	336	96.1%
分類	126	72.6%

表 3: 「病名」に対する観点分類の正解率

観点の種類	記事数	正解率
症状	285	91.1%
治療	262	97.9%
原因	193	84.1%
診断	105	51.0%
検査	102	69.7%

## 6 おわりに

本研究の目的は、検索エンジンと事典の長所を統合して、大規模な情報を統制し要約する点にある。そのため、 Wikipedia から説明の観点に関する構造を抽出し、種々のテキストを観点に基づいて分類する問題に帰着させた。本手法の特長は、与えられたテキストがどのような見出し語について書かれているかを特定し、その結果に基づいて分類すべき観点の候補を変更する点にある。また、本手法を分類や要約に応用するための方法も提案した。さらに、 Wikipedia の記事を入力として評価実験を行った。今後は、Web や特許情報から抽出した雑多なテキストをどこまで正確に分類できるか評価する必要がある。

Wikipedia のように協調的な人手編集による事典は今後も発展するだろう。しかし、爆発的に増える情報を自動的に統制する技術も必要である。情報の統制において自動化が困難な事象を特定し、人手編集との棲み分けや共存について検討する必要がある。

## 謝辞

本研究の一部は、文部科学省科研費特定領域研究「情報爆発時代に向けた新しいIT基盤技術の研究」(課題番号: 19024007)によって実施された。

## 参考文献

- [1] Fadi Biadsy, Julia Hirschberg, and Elena Filatova. An unsupervised approach to biography production using Wikipedia. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics*, pp. 807–815, 2008.
- [2] Sasha Blair-Goldensohn, Ryan McDonald, George Reis, Tyler Neylon, Kerry Hannan, and Jeff Reynar. Building a sentiment summarizer for local service reviews. In *WWW2008 Workshop on NLP Challenges in the Information Explosion Era*, 2008.
- [3] Atsushi Fujii. Producing an encyclopedic dictionary using patent documents. In *Proceedings of the 6th International Conference on Language Resources and Evaluation*, 2008.
- [4] Atsushi Fujii and Tetsuya Ishikawa. Summarizing encyclopedic term descriptions on the Web. In *Proceedings of the 20th International Conference on Computational Linguistics*, pp. 645–651, 2004.
- [5] Atsushi Fujii, Katunobu Itou, and Tetsuya Ishikawa. Cyclone: An encyclopedic Web search site. In *Special Interest Tracks & Posters of the 14th International World Wide Web Conference*, pp. 1184–1185, 2005.
- [6] Inderjeet Mani. *Automatic Summarization*, chapter 7, pp. 169–208. John Benjamins, 2001.
- [7] 長尾真. 辞典形式での専門分野の知識の体系的構築法. 人工知能学会誌, Vol. 7, No. 2, pp. 320–328, 1992.
- [8] 藤井敦. 特許情報を用いた辞典検索システム. 情報処理学会研究報告, 2008-FI-91, pp. 9–15, 2008.
- [9] 藤井敦, 伊藤克亘, 石川徹也. Webマイニングによる事典的コンテンツの構築と多様なアクセス手法. 電子情報通信学会技術研究報告, DE2004-6, pp. 31–36, 2004.
- [10] 藤井敦, 石川徹也. World Wide Webを用いた事典知識情報の抽出と組織化. 電子情報通信学会論文誌, Vol. J85-D-II, No. 2, pp. 300–307, 2002.

## A 付録

Yahoo!で「ハクビシン」を検索した結果のスニペットを観点に分類した結果を示す。いずれのスニペットも見出し語分類によって「動物名」に分類され、さらに観点分類によって「分類」や「形態」など動物名固有の観点に分類された。各スニペットの先頭にある数字は、Yahoo!検索の順位である。

### ● 分類

(5位) ハクビシン. 分類. 哺乳類 食肉目裂脚  
亜目 ジャコウネコ科. 学名. Paguma larvata.  
国内分布. 北海道から九州 特に関東地方に多い.  
生息域. 通常山林の樹木の上に生息するが、近年では住宅やお寺の天井裏に巣を作り生活している。体重 ...

[hwpbc.gate01.com/i-tomo/newpage4.html](http://hwpbc.gate01.com/i-tomo/newpage4.html)

### ● 形態

(10位) ハクビシンと見分けられる部分、特徴と思われる部分を捕獲等の時の経験から掲載します。... ハクビシンの顔. 鼻から頭に向けて白い毛があります。漢字では「白鼻芯」と表現されています。... ハクビシンの手. 前足は、もみじのように見えます。...

[www.city.fujisawa.kanagawa.jp/khozen/page100031.shtml](http://www.city.fujisawa.kanagawa.jp/khozen/page100031.shtml)

### ● 人間との関係

(8位) ハクビシンは県内では1980年に初めて生息が確認されて以降、農作物被害等が発生しており、生息数の増加と県内全域への生息地の拡大が見られております。... また、野生動物であるハクビシンは「鳥獣の保護および狩猟の適正化に関する法律」...

[www.marusans.co.jp/home/hakubishin.html](http://www.marusans.co.jp/home/hakubishin.html)

### ● 生態

(28位) ハクビシンは雑食性であり、特に果実類を好むことから秩父地域の観光果樹園のぶどうを... ハクビシンは、食肉目ジャコウネコ科で東南アジア.を中心とした亜熱帯から熱帯地域に生息している。... これは、ハクビシン.が持つ ...

[www.pref.saitama.lg.jp/A06/BQ23/02/siken/H13/h13\\_06-13.pdf](http://www.pref.saitama.lg.jp/A06/BQ23/02/siken/H13/h13_06-13.pdf)

### ● 分布

(6位) 中国南東部、台湾、東南アジアなどに分布するジャコウネコ科の中型獣。... ハクビシンは「白鼻芯」と書き、顔の真ん中に白い線がある。参考文献. 落合啓二 (1998) 千葉県におけるハクビシンの分布と移入経路. 千葉中央博自然誌研究報告 5:51-54. ...

[www.chiba-muse.or.jp/NATURAL/special/mammal/chibamammal/civit.html](http://www.chiba-muse.or.jp/NATURAL/special/mammal/chibamammal/civit.html)