

【パート I：情報爆発時代における新しい基盤技術】

# 1. キーワードサーチを超える情報爆発サーチ —自然言語処理で価値ある未知をマイニング—

鳥澤健太郎\*<sup>1</sup>

黒橋 禎夫\*<sup>3</sup>

吉岡 真治\*<sup>5</sup>

喜連川 優\*<sup>2</sup>

中川 裕志\*<sup>2</sup>

乾 健太郎\*<sup>4</sup>

藤井 敦\*<sup>6</sup>

\* 1 (独) 情報通信研究機構

\* 2 東京大学

\* 3 京都大学

\* 4 奈良先端科学技術大学院大学

\* 5 北海道大学

\* 6 筑波大学

経緯 —情報爆発 NLP/IR 研究会—

本稿では、文部科学省科学研究費特定領域「情報爆発 IT 基盤」で開発されている、Web を対象とした一連の検索技術を解説する。これらは主として、情報爆発 IT 基盤の支援のもとで開催されてきた「情報爆発 NLP/IR 研究会」という会合における議論の中で生まれた技術であり、以下では総称して「情報爆発サーチ」と呼ぶ。研究会では、Web 上のテキスト情報の爆発に現在主流のキーワードサーチが対応できているのか否かの検討を行い、その過程で洗い出された問題点を分析し、特に自然言語処理技術を利用することでそれらの問題点を解決することを狙った。その議論における 1 つの突破口は、現在のキーワードサーチを使っている限り、ランキングの下位にしか存在しない情報はユーザから見てもたたく存在しないに等しいという事実である。この見かけ上不在の情報は、しばしば非常に価値のある重要な情報であり、これらを見つけ出せる技術には重大な意義がある。

実は、情報爆発サーチでのこの問題に対する一連の対処策は、表面上、単純な枠組みに収束しつつある。それは結局のところ、ランキング下位の情報へのインデックスとして、当初入力されたキーワード、つまりトピックの関連語の集合を提示するというものである。ユーザはその関連語をクリックするなどして、ランキング下位の情報にアクセスすることが可能となる。もちろん、関連語はそれ自体で意味を持っており、ユーザはその関連語を吟味することで価値ある情報を効率良く探し出すことができる。ただ、単に関連語の提示であるならば、現在利用可能なサービスが多数存在する。しかしながら、情報爆発サーチでの力点は、単に関連語を提示するだけでなく、それらを一定の価値観、意味において整理／分類したり、少数の関連語に要約／縮約する、さらには、関連語から文書へのリンケージを精密に求める、というところにある。

つまり、想定される多様なサーチの状況において、「価値ある情報」とはいかなるものであるべきかを考察し、つまるところ、「価値観」と呼べるものにまとめあげて、その「価値観」から見て価値ある情報へのリンクである関連語や、関連語から文書へのリンケージを自然言語処理技術を用いて探し出すということである。この関連語やリンケージは後の具体例で示されるように、多くの場合、自明なものではない。

ここで重要なのは、自然言語処理技術は、すでに「価値観」とでも呼べるものを扱えるまでに成熟しつつあるという点である。もちろん、価値観という言葉は通常、非常に複雑で多様な人の行動、認識の指針一般を指し、そのすべてが自然言語処理の射程に入っているわけではない。しかしながら、これから解説するいくつかのサーチが具体化しているものをまとめて形容するにあたり価値観という以外の語が見つからなかったのもまた事実である。今後はこのような自然言語処理で対処可能な価値観を次第に増やしていくことが重要な研究テーマとなろう。また、この自然言語処理の成熟の一因は利用できる文書量が飛躍的に増大したという事実であり、これには開放型検索エンジン基盤 TSUBAKI が多大な貢献をしている<sup>1)</sup>。TSUBAKI は、次世代サーチ研究のための基盤として情報爆発 IT 基盤で構築された検索エンジンであり、約 1 億件の日本語 Web ページについて、形態素・構文解析の結果を保持し研究者に提供している。以下で解説する情報爆発サーチのいくつかは TSUBAKI が提供しているデータをもとに開発されたものである。

以下では、各々独自の価値観を持つ情報爆発サーチを 5 システム見ていく。

情報を俯瞰する

Web 情報に対する検索はナビゲーション的な問合せ

とインフォメーションな問合せに大別することができる。前者は「トヨタ」のような組織のトップページなどの検索であり、その場合にはランク付けされたページのリストを返す既存のリスト型検索エンジンが有効に機能する。一方、「ゆとり教育」のようなインフォメーションな問合せ、すなわち、ある話題についてその概要を知りたい場合や、それに関連する Web 情報の全体像を把握したい場合には、リスト型検索エンジンで満足な結果を得ることは難しい。

この問題の解決策として、クラスタリングによって検索結果を集約する方法が考えられ、Clusty などの商用システムも存在する (<http://clusty.jp/>)。ここでは、まずリスト型検索エンジンの検索結果としてスニペット(各ページの 2, 3 文の要約)を収集し、そこから適切な関連語(ラベル)を選択し、各ラベルを含むスニペットをまとめてクラスタとする。しかし、ラベル抽出の対象が数百文であるため、与えられたトピックに対して関連語を十分に抽出できているとはいえず、そこで提示されるラベル集合はかなりおそまつなものである。黒橋らは、このような問題を解決するために、検索エンジン基盤と連携することによって数千件の Web ページの数万文を対象としてラベル抽出を行うクラスタリングシステムを構築している<sup>2)</sup>。大量のテキストから単純に高頻度の表現を取り出すだけでは同じような意味の関連語が冗長に抽出されてしまう。この研究の主眼は、入力されたトピックに関する関連語を少数のものに要約/縮約する(蒸留する)ことにある。

このシステムは、次のような処理からなりたっている。

- (1) TSubaki の利用：まず TSubaki を用いて与えられたトピックに対する数千件の検索結果を取得する。
- (2) 各ページからの関連語抽出：次に、検索結果の各ページから関連語を抽出する。この処理は TSubaki が持つ言語解析結果や並列処理を利用することで高速に行うことができる。各ページから入力トピックに関する 15 文程度の重要文を選択し、そこから複合名詞、その部分単語列、括弧で囲まれた表現などを関連語とする。
- (3) 関連語の蒸留：従来のラベルベースクラスタリングでは、単純に高頻度の表現をラベルとして選択していた。しかし、本システムのように数万文を対象として網羅的に関連語抽出を行った場合、それでは質の高い情報提示とはならない。それは、「つめこみ教育」と「詰込み教育」、「カリキュラム」と「教育課程」、「IWC」と「IWC 総会」など、同じ内容を表す表現や包含関係にある表現が関連語中に多数存在し、それらが最終的なラベル集合の中に散在する



図-1 「ゆとり教育」に関する情報を関連語を通して俯瞰した例

と人間の全体把握のプロセスを大きく阻害するからである。そこで本システムでは、形態素解析による表記揺れの吸収、国語辞典や Web から自動獲得した同義表現のマージ、部分全体関係にある表現のマージなどを段階的に行うことによって、高純度に関連語を集約する。

このような処理によって、「ゆとり教育」について 1000 件の検索結果のクラスタリングを行った結果を図-1 に示す。画面左側に関連語蒸留の結果得られた 30 個程度のラベルが表示されている。各ラベルは、そのラベルを含む文書集合のクラスタに対応付けられており、ラベルをクリックすると、画面右側にそのラベルを含むページのリストが表示される。ここで提示されるラベル集合は「ゆとり教育」を総合的に理解する上で重要な観点をかなりの程度網羅しており、これによってクエリの関連概念および検索結果を鳥瞰図的に把握することができる。この処理時間は検索を含め約 40 秒である。既存のリスト型検索エンジンを用いてその結果を眺めていく方法では、このような全体把握を同程度の時間で行うことはきわめて難しい。

今後の課題としては、抽出されたラベルのより高度な組織化がある。現在は、出現するページの重複の割合が大きなラベルをまとめて提示しているが、固有表現解析やシソーラス・オントロジーの利用によってより高度な組織化が可能であると考えている。

## 意見の分布を見る

Web には、報道記事や学術論文のように客観性が高い情報だけでなく、意見、批評、感想のように主観性が高い情報も含まれる。ここでは、客観的な情報と区別

するために、主観性が高い情報を「主観情報」と総称する。入力されたトピック（話題）に関する主観情報の分布を「論点」と呼ばれる一連の関連語を介して可視化し、全容の把握を容易にすることが藤井らの狙いである。

具体的には、対象の話題に関する「肯定／否定」や「賛成／反対」といった対極の観点に基づいて論点を可視化する。このような可視化は、個人や組織の意思決定に役立つ場合がある。たとえば、商品に関する批評を読んで購入する商品を決めるような個人的な意思決定から、ある社会問題に対する賛成と反対の意見を読んで個人や組織としての態度を決定する場合がある。これらの例における意思決定は、以下に示す手順に分解することができる。

- (1) 対象の話題（商品や社会問題）に関する文書を Web から収集する。
- (2) 収集した文書から主観的な記述を抽出する。
- (3) 抽出した主観的記述を「肯定／否定」や「賛成／反対」などの観点に応じて分類する。
- (4) 主観的記述を集約し、全容を把握しやすい形式で可視化する。
- (5) 可視化された内容を吟味して、「肯定／否定」（あるいは「賛成／反対」）から一方を選択する。

既存の Web 検索エンジンが支援するのは上記 (1) だけであり、それ以外の作業はユーザの労力や能力に依存している。藤井らは、上記 (2)～(4) を自動化することで、キーワードサーチを超える検索機能の研究を行っている<sup>3)</sup>。具体的には、論争の余地がある社会問題を対象として、主観情報をサーチするシステム「OpinionReader（オピニオンリーダー）」を開発している。これは、意見を読むための道具（reader）と統率者（leader）の 2 つをかけた名称である。

ある話題について賛否両論が存在する場合は、論点が存在する。たとえば、「赤ちゃんポスト」の是非について、「人命を救う」や「育児放棄の助長」などの論点がある。意思決定とは、論点に基づいて賛否両論を洗い出し、より合理的な立場を採用する過程と捉える。対象とする社会問題によっては、答えが簡単に出ない場合がある。そのような場合でも賛否両意見を把握して問題の解決に向けた努力を続けることには意義がある。

OpinionReader は賛否両意見の分布を論点に基づいて可視化する。その結果、ユーザは大量の主観情報を読まなくても、その話題に対する議論の全容を把握することが可能になる。議論に途中から参加したユーザにとっても有用性が高い機能である。図-2 は「赤ちゃんポスト」という話題に対する論点の分布を可視化した例である。図-2 において、「育児放棄」や「最終手段」などの語句が論点である。名詞句と動詞句を論点として抽出す

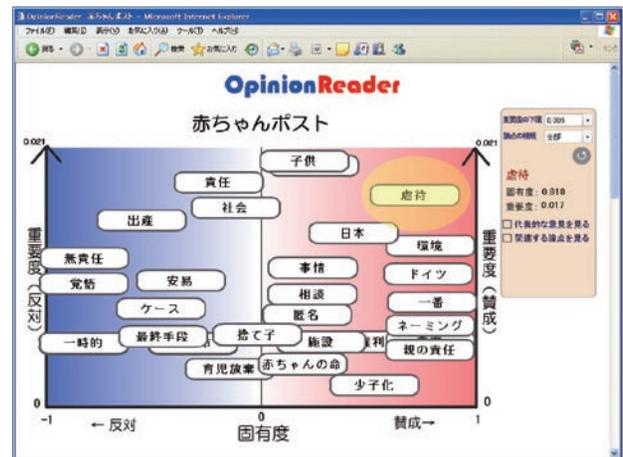


図-2 「赤ちゃんポスト」に対して論点の分布を可視化した例

る。横軸は、右にいくほど賛成派に固有の論点であり、左にいくほど反対派に固有の論点であることを表す。縦軸は、上にいくほど各立場で頻繁に使われる論点であることを表す。

一方の立場に偏って論じられている論点は、逆の立場にとっては不利な論点であるか、もしくは議論が不十分なために逆の立場では論じられていない可能性がある。論点の分布を可視化することで、各立場で何が論拠となっているのかが分かると同時に、各立場にとって強みになり得る論点や議論が不十分な可能性がある論点を発見することができる。

図-2 では、特定の論点を選択することで、その論点に関する代表的な意見を賛成と反対で読み比べることができる。ある論点 A に関する代表的な意見とは、その意見を読めば論点 A に関する他の意見を読む必要がないような意見である。具体的には、論点 A を含む意見の集合によく表れる単語を網羅する度合いが高い意見ほど、論点 A の代表的な意見になりやすい。

さらに、「赤ちゃんポスト」と同じような論点に基づいて議論される「代理出産」や「不妊治療」などの関連する話題を検索し、議論の参考にすることができる。今後は、時間によって議論が変化の様子を可視化することを予定している。

## 海外の事情を探る

各国の新聞、放送などのメディアが Web での情報発信を始めたことにより、広く世界中の情報が得られるようになった。ただ、すべてのニュースを読むことは困難であるため、記事を読むための指針が必要である。そこで、中川、吉岡らは、日本と海外でのニュースの取り扱い方の違いに注目することにより、日本のニュースだけでは分かりにくい、各々の国の興味の違いを明示化する

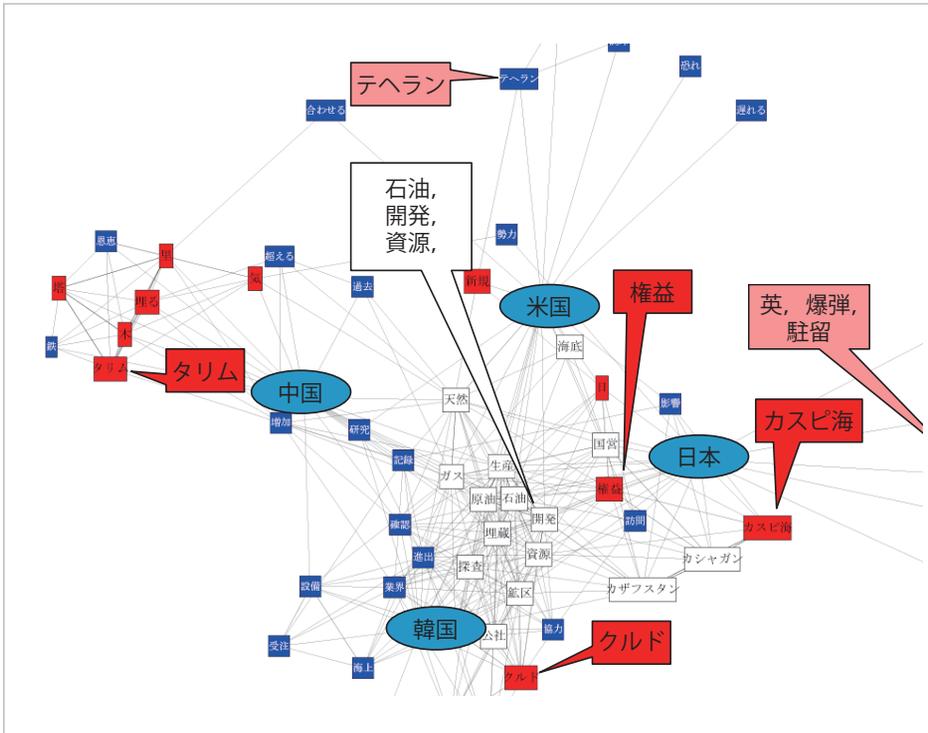


図-3 日米中韓における「油田」と共起する単語を提示した例

方法の研究を行っている。この違いを基準にニュースを読むことにより、環境問題・国際紛争などの全世界的なトピックに関する分析、世界的なビジネスチャンスの発見などが期待できる。

上述の目的に対し、トピックと共起する関連語の違いに注目して、各国のニュースの取り扱い方の違いを明示化するシステムを構築している<sup>4)</sup>。このシステムは、次の2つの要素で構成される。

- (a) 情報検索部：各国の新聞社サイトをクロールして得た記事から世界ニュースデータベースを作成し、それらに対する検索APIを提供している。ただし、日本語の記事を対象として分析を行うため、各国メディアの日本語版を利用した。また、韓国語については機械翻訳の質が、単語レベルの分析を行うという観点からは、十分な性能を持っていると考えたため、機械翻訳した記事も日本語版と合わせて利用することにした。
- (b) 対照分析部：与えたキーワードと関連語の共起度を、さまざまな新聞記事群（たとえば、発信国ごとにまとめた記事群）に対して計算するとともに、その違いを分析する。

2007年8月から2008年2月に報道された日米中韓の新聞記事（日本：朝日、読売、日経、韓国：朝鮮日報、朝鮮日報（韓国語翻訳版）、中央日報、中国：人民日報、アメリカ：CNN）を利用して、「油田」をキーワードとした場合の出力結果を図-3に示す。このグラフ中で関連語は長方形ノードで表され、国は、楕円ノードで表さ

れる。また、ノード間のリンクは、共起度の強さに応じて設定され、このリンクを考慮したノードの配置が行われる。

関連語は、次の2種類の基準で選択され、そのタイプに応じた形でノードの色分けがされている。

- (1) 1国で共起度の高い関連語
  - 1国におけるメジャーなトピックを表す関連語
- (2) 1国での共起度と残りの国々の新聞群全体における共起度の比を取り、その値が高い関連語
  - 1国で、他の国々よりも注目されている関連語

(1)と(2)の両方の基準を満たす語（赤地）は、1国のみでメジャーであり、他の国ではあまり注目されていない、その国の特徴をよく表した語と考えられる。(1)のみを満たす語（白地）は、多くの場合、複数の国で共通に興味を持たれている語であり、複数の国での共通の興味を示す語と考えられる。(2)のみを満たす語（青地）は、その国において、メジャーなトピックとはいえないまでも、比較的注目されている語を示す。これらの語は注目されている国の近くに配置され、すべての国で注目されている場合は中心部に配置される。このグラフの中央に存在する白地の語に注目すると、各国共通の話題として、「石油、開発、資源」などが読み取れる。次に青字の単語を見ると、日本では、「英、爆弾、駐留」があり、自衛隊のイラクの油田に近い地域への派遣を示している。また、米国では、「テヘラン」があり、イランとの関係に絡む事情が注目されている。これらは、各々の国の興味の違いを示す語と考えられる。最後に赤字の単

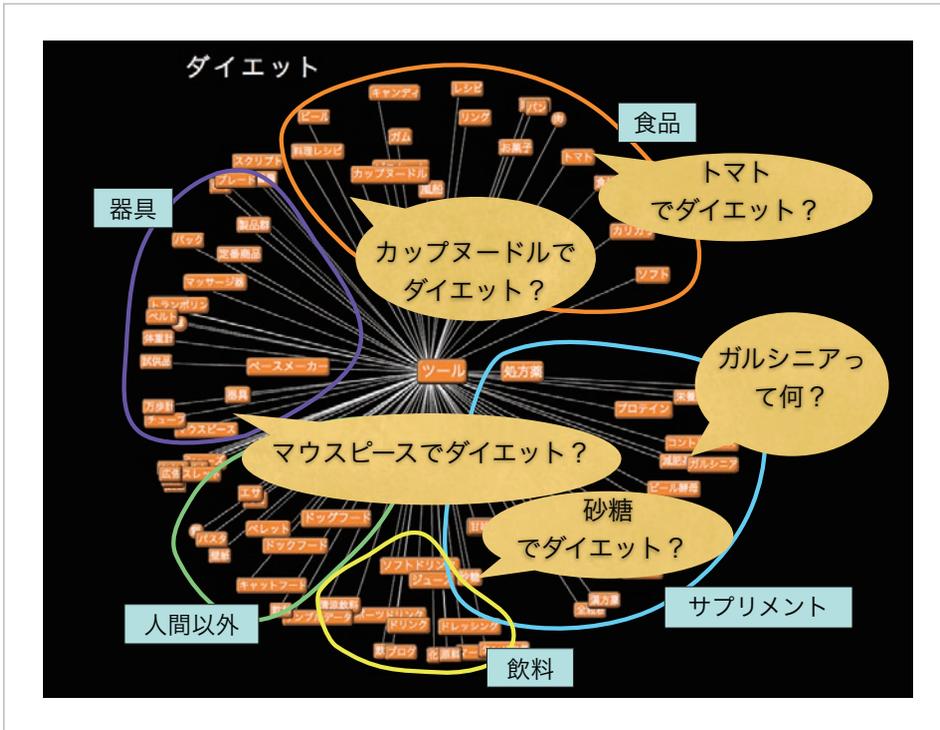


図-4 ダイエットのツールを鳥式で提示した例

語を見ると、中国では「タリム」、韓国では「クルド」、日本では「カスピ海」という語がある。これらは、各々の国の中心的な油田開発地であり、他国との興味の差が大きい語である、また日本にはさらに「権益」をめぐる問題があることがうかがえる。

このように、各国の差に注目することにより、一般的な話題だけではなく、国々の特徴などを分析することが可能になる。結果として、日本の新聞を読んでいただけでは、気づきにくい情報を探すための手がかりを提供できると考えている。

### キーワードの想起を支援する

適切な行動をとるための情報収集に、検索エンジンを利用するのはもはや常識である。つまりは、さまざまなトピックに関する問題回避、あるいは行動に関する未知のアイデア、Tips について情報を求めるため、検索エンジンを利用するというところである。ところが、そうした情報を得るにはユーザから見て「意外」なキーワードを入力する必要がある。たとえば、執筆者の1人が常宿としていたホテルがいわゆる建築偽装疑惑に関係した会社によって建設されていたという事実は、その執筆者にとってはまったく未知であった。予約のためサーチを行っても検索結果の上位にはそうした情報はなく、実際そのホテルに何度も宿泊した。ところが、実は通常の検索エンジンでも、ホテル名に加えて「落とし穴」という意外なキーワードを与えると、検索結果のトップ

に問題の事実が見つかる。重要な点は、こうした意外なキーワードはユーザの「意識にのぼっていない」以上、システム側から提示する必要があることである。

鳥澤らはこうしたキーワードの想起を支援するため、「鳥式」という検索ディレクトリを開発している<sup>5)</sup>。これは、ユーザが最初に入力したキーワード、つまり、トピックに対して、関連語を意外なものまで含めて提示し、検索に利用できるようにする。なお、鳥式の第1の特長は鳥式がWeb 文書に自然言語処理技術を適用することで自動生成されており、現在128万語という大量のトピックをカバーしていることである。第2の特長は価値ある情報を効率良く検索できるようにするため、いくつかの意味的カテゴリに属する関連語のみを提示することである。ホテルの「落とし穴」は「トラブル」というカテゴリ中の関連語として提示される。現時点では、トピックを利用する行為（例：ホテルならば「宿泊」）あるいはトピックに対処する行為（例：トピック「花粉症」に対して「治療」）に関する情報収集が検索ニーズの一定部分を占めていると仮定し、それらの行為を行う上で有用なカテゴリが設定されている。具体的には、利用/対処の行為自体、利用/対処を行うための「準備の行為」（例：ホテルの「予約」）、利用/対処/準備といった行為を阻害する要因としての「トラブル」、それら行為を行う際のTipsとしての具体的「方法」、有用な「ツール/材料」がある。図-4は、トピック「ダイエット」の対処に利用できるツール/材料を、「トマト」のような意外なものも含め提示した例であるが、意味的に類似



図-5 「iPod touch<sup>TM</sup>が欲しい」と書いたブロガーの検索例

した関連語がまとめて表示され、欲しい関連語を探すのを容易にしている。

また、鳥式の第3の特長は、広範な関連語を提示するため、トピックの上位概念の名称（例：トピック「東京大学」に対する「大学」）を自動的に獲得し、大量に保持していることである。今年はじめに話題になった農薬ぎょうぎ事件を例にとると、これまで開発した手法では、昨年、つまり、事件以前のWeb文書から、トピック「ぎょうぎ」の関連語として「農薬」を直接認識することはできなかった。しかしながら、「ぎょうぎ」の上位概念になる可能性のあるものに「冷凍食品」があり、「冷凍食品」のトラブルとして「残留農薬」が認識できていることから、「残留農薬」を「ぎょうぎ」のトラブルとして提示できる。つまり、騒ぎになる以前にぎょうぎ事件をあたかも「予測」していたことになる。実際にぎょうぎに付着していたものが「残留」農薬なのか意図的なものであるのかは今もって不明であるが、問題のぎょうぎにかかわった人々に「残留農薬」の可能性が事件の早い段階で示唆されていたとすれば、状況は改善されたかもしれない。鳥式はトピックに対して関連語を提示するという一見単純な処理しか行わないが、このぎょうぎの例などは、そのような単純な処理ではあっても実社会でインパクトを持ち得ることを示唆しているものと考えている。

## 人々の経験から学ぶ

以上に紹介したサーチではいずれも、ユーザに提示すべき関連語をどうやって選ぶかに焦点が当たっており、ユーザが関連語を指定した後は単純なキーワード検索で

関連文書を返すことが仮定されていた。これに対し、乾らが研究を進める「経験マイニング」<sup>6)</sup>は、関連語から文書へのリンケージを精密にする試みと位置づけることができる。経験マイニングは、ブログ等へ書き込まれた個人の行動、成功体験、トラブル、興味、感想など、膨大な経験の記述を意味的に解析し、データベース化するものである。前章までのサーチとのアナロジーで言えば、データベースに用いる意味的な索引がこれまで議論してきた「関連語」に相当する。

現在想定している索引は次の5項目である。説明のため、「戸塚駅では、11時を過ぎるとエレベータは止まってしまうので、西口には出られなくなります」という記述に対する索引付けの例を括弧で付記する。

- 経験者：経験の主体（著者）
- トピック：何に関する経験か（「戸塚駅」）
- 事態表現：経験の核となる事態表現（「11時にエレベータが止まる」あるいは「西口に出られない」）
- 事態タイプ：ポジティブ／ネガティブな出来事・状態、入手・利用等の行為など、経験情報の核となる事態の種類（ネガティブな出来事）
- 事実性：当該の事態が実際に起こったことなのか、可能性を述べただけなのかといった、事態の時間情報とそれに対する話者態度（過去から現在まで繰り返し起こっている事実の断定）

たとえば、事態タイプと事実性の情報を組み合わせることによって、上の「戸塚駅のエレベータ」の例のような〈実際に起こったトラブル〉と「(車が)横転しそうな気がする」のような〈可能性に対する心配〉、あるいは「薄めて飲むと痩せない」のような単なる仮定の話とを区別した検索が可能になる。経験マイニングのねらいは、商品（車、携帯電話など）や行政サービス（子育て支援制度、花火大会）、場所（病院、飲食店）など、さまざまなトピックに関する膨大な数の経験情報をWebのような広大な空間から収集し、事態タイプや事実性のような意味的索引で構造化することにより、個人や組織の意思決定やトラブルの回避解消に有用な「知」の宝庫に変えることである。

上のデザインをもとに乾らは、経験に基づくブロガー検索システムを1つの応用例として開発した。このシステムは、図-5のように特定の商品に関する人々の経験を〈興味あり〉〈買った〉〈満足〉などの経験クラスで検索し、結果をブロガー単位で表示する。これによって、たとえば「iPod touchに興味を持っている人」、「実際に買った人」、「満足している人」を集めることができる。図は経験クラス〈欲しい〉で検索した例で、「iPod touch」について何らかの経験を書いたブロガー7797人のうち、「欲しい／使ってみよう」と言った1250人

がヒットし、述べた経験の多い順に並ぶ。このほか「他の経験を参照する」をクリックすると、「iPod touch」に関するそのブロガーのすべての経験が時系列に表示される。技術的なポイントは、〈欲しい〉や〈よく使う〉といった経験の検索を、事態タイプや事実性といった一般性の高い意味情報の組合せで実現する点にある。図の例では、〈欲しい〉の検索を「入手または利用行為を未来において行う意志、欲求がある人」という条件で行っている。

こうして得られる情報は著者のバックグラウンドを知る良い手がかりになる。著者の経験が〈興味あり〉から始まって〈買った〉〈使った〉〈満足〉のように推移していればいわゆる“サクラ”でないことが分かるなど、記事の信頼性を判断する際にも有用である。また、「欲しいと思いながら未購入の人」、「愛用していたのに止めた人」のような複雑な検索も原理的には可能であり、個人の利用はもとより、企業のマーケティング、行政サービスの評価などの情報源として有効活用できると考えている。

## 今後の展開

以上、5つの情報爆発サーチについて紹介してきた。今後、これらのサーチ自体はより大規模なデータとより先進的な自然言語処理技術の投入によって深化、洗練されるであろう。また、今後考察しなければいけないのは、サーチで取り扱うべき「価値観」の範囲をより明確にすることと、価値観からサーチ技術へのリンケージを体系的に行う方法論である。これまで見てきたシステムでは、たまたま開発者が興味を持った価値観をやはり開発者の勘と経験によってシステムに取り込んでいる。このプロセスをより体系的にすることが、より実用性／利便性が高いシステムを実現する上での鍵であると思われる。

最後に、本稿著者以外の情報爆発 NLP/IR 研究会の参加者にこれまでの濃厚な議論、協力に対して心からの謝意を表し結びとしたい。

### 参考文献

1) 黒橋禎夫, 新里圭司: TSUBAKI: 深い言語処理を特長とするオープンサーチエンジン基盤, 情報処理, Vol.49, No.8 (Aug. 2008).

- 2) 馬場康夫, 新里圭司, 黒橋禎夫: 検索エンジン基盤 TSUBAKI を用いた大規模ウェブ情報クラスタリングシステムの構築, 情報処理学会 自然言語処理研究会 183-10, pp.67-74 (2008).
- 3) 藤井 敦: OpinionReader: 意思決定支援を目的とした主観情報の集約・可視化システム, 電子情報通信学会論文誌, Vol.J91-D, No.2, pp.459-470 (2008).
- 4) 吉岡真治: トピックの差異に注目した複数新聞の比較対照分析方法の提案, 言語処理学会第 14 回年次大会発表論文集, pp.592-595 (2008).
- 5) 鳥澤健太郎, 隅田飛鳥, 野口大輔, 風間淳一: 自動生成された検索ディレクトリ「鳥式」の現状, 言語処理学会第 14 回年次大会論文集, pp.729-730 (2008).
- 6) 乾健太郎, 原 一夫: 経験マイニング: Web テキストからの個人の経験の抽出と分類, 言語処理学会第 14 回年次大会論文集, pp.1077-1080 (2008).

(平成 20 年 4 月 30 日受付)

### 鳥澤健太郎 (正会員)

パートI「0. 情報爆発時代の研究動向」を参照

### 中川 裕志 (正会員): nakagawa@dl.itc.u-tokyo.ac.jp

東京大学情報基盤センター教授。工学博士。1980年横浜国立大学講師, 1999年より現職。自然言語処理, 統計的機械学習の研究に従事。ACL Exec Member, 言語処理学会会長などを歴任。現在, 本会 自然言語処理研究会主査。http://www.r.dl.itc.u-tokyo.ac.jp/~nakagawa/

### 黒橋 禎夫 (正会員): kuro@i.kyoto-u.ac.jp

京都大学大学院情報学研究所教授。博士(工学)。自然言語処理, 知識情報処理の研究に従事。言語処理学会 10周年記念論文賞, 同平成 17年, 19年論文賞等を受賞。http://www-lab25.kuee.kyoto-u.ac.jp/

### 乾 健太郎 (正会員): inui@is.naist.jp

奈良先端科学技術大学院大学情報科学研究科准教授。自然言語処理, 知識処理の研究に従事。言語処理学会, 人工知能学会, ACL 各会員。http://cl.naist.jp/~inui/

### 吉岡 真治 (正会員): yoshioka@ist.hokudai.ac.jp

北海道大学大学院情報科学研究科准教授。1996年学術情報センター助手, 2001年より現職(当時, 工学研究科助教授)。情報検索への知識処理技術の応用。人工知能学会, 言語処理学会, ACM 各会員。http://www-kb.ist.hokudai.ac.jp/~yoshioka/

### 藤井 敦 (正会員): fujii@slis.tsukuba.ac.jp

筑波大学大学院図書館情報メディア研究科准教授。自然言語処理, 情報検索, 音声言語処理の研究に従事。言語処理学会, 人工知能学会, ACL 等各会員。http://www.slis.tsukuba.ac.jp/~fujii/

### 喜連川 優 (正会員)

総括記事「特定領域研究「情報爆発 (Info-plosion)」: 本格稼働から 2 年を経過して」を参照