

# 意味クラス解析と意図推定に基づくインタラクティブな情報検索インタフェース

石谷 康人<sup>†</sup> 鈴木 優<sup>†</sup> 布目光生<sup>†</sup>

本論文では、Web ブラウザ上で閲覧している文書を起点として関連文書を次々とたどる連鎖検索 (Chaining search) のための新しいインタフェース技術を提案する。本インタフェースでは、ユーザがペンやマウスなどのポインティングデバイスにより閲覧文書中の語句を選択すると、意味クラス解析技術により語句の意味を決定し、意図推定技術により語句の意味から検索方法を列挙およびメニュー化してユーザに提示するようになっている。ユーザは提示された検索方法のメニューにおいて適切なものを選択するだけで、選択した語句に関連する情報を検索できるようになっている。このとき本インタフェースでは、検索方法に応じて、ユーザが選択した語句に補助キーワードを追加して絞り込み検索を行う。このようなアプローチにより、ユーザは直感的なペン/マウス操作だけで目的とする情報に効率良く的確にアクセスできるようになる。30人の被験者を対象とした実験で本インタフェースの検索性能とユーザ満足度を評価したところ、ユーザビリティの高さと絞り込み検索の有効性の両方を確認することができた。

## Interactive Human-interface for Information Retrieval Based on Semantic Pattern Analysis and Intention Understanding

YASUTO ISHITANI,<sup>†</sup> MASARU SUZUKI<sup>†</sup> and KOSEI FUME<sup>†</sup>

This paper proposes a new human-interface technology for chaining search in order to realize efficient and accurate Web search. When a user selects a word or a phrase on a Web page using a pointing device such as a pen or a mouse, the proposed interface determines its meaning by semantic pattern analysis and provides a menu with a set of appropriate search methods to the user by intention understanding. If the user selects an adequate search method from this menu, the user is able to access required information relevant to the selected keyword. In this web search, the proposed interface adds extra keywords to the selected keyword according to the search method selected from the menu for accurate information retrieval. Experimental results obtained from 30 trial subjects show that the proposed interface is effective for various types of Web access.

### 1. はじめに

コンピュータネットワークの発達、記録メディアの大容量化、WWWの普及、Webコンテンツの充実などにもない、我々は世界中の多種多様な情報にアクセスできるようになった。その結果、知識獲得、文書作成、計画立案などの知的生産において情報検索が必須となり、頻繁に実施するようになった。

現状の情報検索では、ユーザは、(1) 検索サイトへのアクセス/検索エンジンの起動、(2) 検索クエリの検討および入力、(3) 検索サイト/検索エンジンが出力した検索結果の吟味、(4) 目的に合致する情報の選択お

および閲覧という手順で必要な情報を入手している。

World Wide Webを情報検索の対象とする場合には、簡単な検索クエリを用いるとたくさんの検索結果が出力されてしまうので、上記(2)で検索クエリに補助的なキーワードを追加するとともにキーワード間の関係をブール式で定義して検索結果を絞り込む必要がある。それでも目的とする情報が検索結果の上位に出現しない場合には、上記(3)と(4)において検索結果をスクロールしたり、次の検索結果のページへ移動したりして情報を探ることになる。

このような情報アクセスでは、たとえばGoogle<sup>TM,5)</sup>のような優秀なWeb検索エンジンを使用する場合でも、目的とする情報への確にアクセスできる検索クエリを設定することは難しい。また、上記(1)~(4)の作業はキーボードとポインティングデバイスを切り替

<sup>†</sup> 株式会社東芝研究開発センター  
Corporate Research & Development Center, Toshiba Corporation

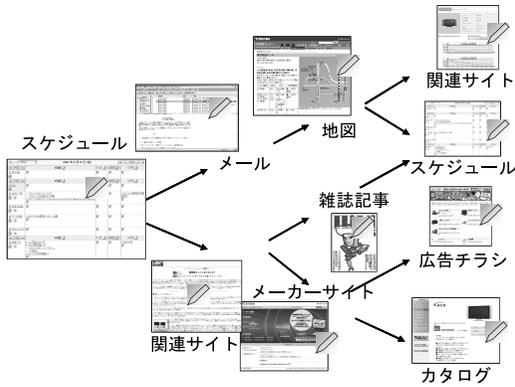


図 1 連鎖検索の概念図

Fig. 1 Concept of chaining search.

えながら行うため検索行為そのものが煩雑となっている。このように情報アクセスのプロセスが煩雑になると、本来の知的生産において思考がたびたび中断され、知的生産のトータルスループットが低下するという問題があった。

そこで、本論文では、検索サイト/エンジンの立ち上げを不要とし、検索クエリのキー入力を省き、検索結果を吟味する時間を短縮し、入力デバイスの切替えを必要としないことを特長とする連鎖検索 (Chaining search) のための新しいヒューマンインタフェース技術を提案する。本インタフェース技術では、まず、ユーザが文書中のキーワードや語句をマウスやペンなどのポインティングデバイスで選択する。次に、意味クラス解析技術 (semantic pattern analysis) により、ユーザが選択したキーワードの意味を推定するとともに、意図推定技術 (intention understanding) によりキーワードの意味に基づいて検索方法を列挙およびメニュー化してユーザに提示する。このように提示されたメニューからユーザは適切な検索方法を選ぶことになるが、このとき検索方法に応じたキーワード拡張を実施することにより、ユーザが目的とする文書に的確にアクセスできるようにする。本研究では、このようなコンセプトに基づく情報検索インタフェース技術を採用することにより、ユーザが閲覧文書の着目箇所のすぐ近くで情報を検索できるようにして、関連文書を次々とたどる連鎖検索 (図 1 参照) を簡便に実施できるようにする。そして、最終的には、ユーザが普段よく利用する Web ブラウザ、メーラ、ワープロソフト、表計算ソフト、プレゼンテーションソフト、ドキュメントビューワなどのアプリケーションで、上述したコンセプトによる情報検索インタフェースを実現し、情報検索を駆使する知的生産のトータルスループットを

向上させることを目指す。

以下、本論文では、Web 検索を目的とした Web ブラウザ向け情報検索インタフェースにフォーカスして議論する。まず、2 章で関連研究について概観する。次に、3 章で情報検索インタフェースの基本コンセプトについて述べ、4 章で本インタフェースの構成について説明する。そして、5 章で本インタフェースを用いた実験結果を示し、その有効性について評価する。最後に、6 章で本インタフェースの実世界での長期間利用について考察する。

## 2. 関連研究

従来の情報検索インタフェースは、「サーチボックスにキーワードを入力するキーワード検索タイプ」と、「分類に従って細分化されたカテゴリを上位から下位へとたどるディレクトリ探索タイプ」の 2 種類に分けることができる。これまでに、それぞれのタイプにおいて、情報検索時の負担の軽減を目的とするいくつかの研究結果が報告されている。

検索クエリの入力を軽減するために、シソーラスを用いてユーザが入力した検索クエリを自動的に拡張する方法が提案されている<sup>11)</sup>。この方法では、シソーラスにある類義語や上位/下位概念を用いて、ユーザが入力した検索クエリを拡張することにより情報検索の再現率を高めている。

内容の類似性に基づいて検索結果をグループ化して閲覧性を高めたり<sup>7),10)</sup>、検索時に関連情報を提示して情報検索の利用性を高めたり<sup>13)</sup> する方法がいくつか提案されている。いずれの方法も、検索結果を閲覧しながら徐々に目的とする情報にたどり着くようになっており、ユーザの検索目的が漠然としている場合に有効である。

XLibris<sup>15)</sup> では、ペンインタフェースによる簡便な情報検索インタフェースを提供している。このインタフェースでは、文書の読解を支援するために閲覧文書上にアノテーションやマーキングができるようになっており、語句にアンダーラインを引く、語句を丸で囲む、複数の語句や文を囲むといったペン操作により関連文書を検索できるようになっている。また、ペン操作によりキーワード検索と類似文書検索を簡単に切り替えられるようになっている。

以上のほかに、メタ情報を用いて情報検索を簡便に実施する方法がいくつか提案されている。たとえば Haystack<sup>6)</sup> では、ユーザの個人情報を用いて検索結果を的確に絞り込むことが可能となっている。また、Time-Machine Computing<sup>14)</sup> では、時間情報に基づ

いて情報の蓄積と可視化を行うことにより、直感的な情報アクセス手段を提供している。さらに、近傍検索<sup>12)</sup>では、メタ情報に基づいて情報間の距離を定義するとともに、近さに基づいた直感的な情報アクセスを実現している。

本論文では、情報検索のユーザビリティと検索性能の向上を目的として、上述した従来研究の良いところである検索クエリ拡張、検索方法の提示、直感的な操作性を兼ね備えた新しい情報検索インタフェースを提案する。

### 3. 連鎖検索インタフェースの基本コンセプト

本章では、本論文で提案する連鎖検索のための情報検索インタフェースの基本コンセプトについて説明する。

紙に印刷された文書を読んでいる際、意味が分からない語句や重要な語句に出会うと、それらにペンでアンダーラインを引いたり、マーキングしたりすることがある。これは多くの人が慣れ親しんでいる行為であり、そのたびに思考が中断されるようなことはない。このことから、XLibris<sup>15)</sup>のように、マウスやペンなどのポインティングデバイスを用いて、Webブラウザ上で語句をマーキングもしくは選択することにより検索クエリを設定することができれば、直感的な操作による情報検索を実現できるようになる。しかし、選択された語句のみをキーワードとして用いて検索を行うだけでは、検索エンジンが大量の検索結果を出力する恐れがある。このため、選択された語句に適切な補助キーワードを加えて検索結果を絞り込む必要がある。

そこで本研究では、ポインティングデバイスを用いたキーワード選択を可能とするとともに、ユーザの検索意図に応じたキーワード拡張を可能とする情報検索インタフェースを提案する。以下に、本インタフェースに基づく情報検索の流れを示す。

**Step 1:** ユーザが、Webブラウザで閲覧している文書において、ポインティングデバイスを用いてキーワードを選択する。

**Step 2:** システムが、ユーザによって選択されたキーワードの意味を決定するとともに、キーワードの意味に応じて検索方法を列挙およびメニュー化してユーザに提示する(図2参照)。

**Step 3:** ユーザはメニューの中から適切な検索方法を選ぶ。

**Step 4:** システムは、ユーザが選択した検索方法に応じて適切な補助キーワードを追加したあと検索エンジンに検索要求を出す。

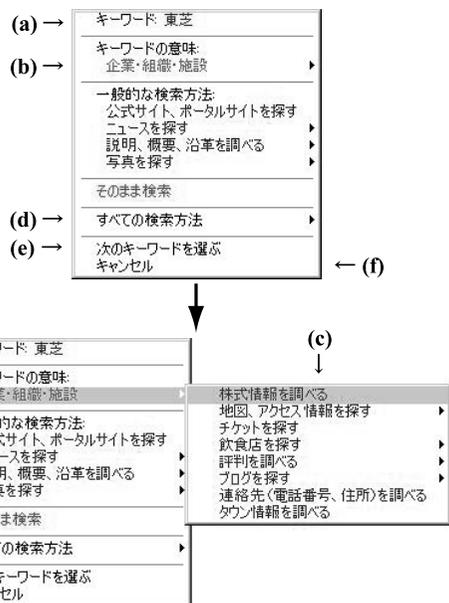


図2 検索方法のメニューの例

Fig.2 Examples of the menu of search methods generated by the proposed interface.

**Step 5:** ユーザは、検索エンジンが出力した検索結果の中から目的とする情報を選択し、閲覧する。

この場合の検索方法とは、たとえば、「企業名を手がかりとしてその企業の株式情報を検索する」とか、「人物名を手がかりとしてその人のプロフィールを検索する」のような検索を指すものとする。本研究では、上記 Step 1 から Step 4 までのユーザインタラクションによる検索過程を意図推定に基づく情報検索と見なしている。

また本インタフェースでは、たとえば「本郷」という言葉が「地名」「駅名」「人名」という3つの意味を持つように、ユーザが選択したキーワードが多義性を有する可能性を考慮して、検索方法を意味ごとにまとめるとともに、「キーワードの意味」→「検索方法」という順序で検索方法を階層的に選択できるようにしている(図2参照)。

以下に、図3を用いて、本インタフェースに基づく情報検索の具体例について説明する。この例ではまず、ユーザがWebブラウザを起動し、Webブラウザのブックマーク機能により東芝レビューのWebサイトに移動して記事を閲覧しているものとする(図3の「閲覧ページ」)。次にユーザは、閲覧しているページでふと東芝の株式情報について知りたくなり、ポインティングデバイスを用いて文中の「東芝」を選択したとする(図3の「キーワード」)。そうすると本インタフェースは、東芝が「企業・組織・施設」であると見なすとともに



検索エンジンの出力結果



図3 情報検索インタフェースによる連鎖検索の流れ  
Fig. 3 Chaining search process conducted by the proposed interface.

(図2の(a)),「株式情報を調べる」「連絡先(電話番号,住所)を調べる」などの検索方法を列挙およびメニュー化してユーザに提示する(図2の(c)).このときユーザが検索方法のメニューから「株式情報を調べる」を選択すると,本インタフェースはユーザが選択したキーワード「東芝」に「投資家情報」「個人投資家」「株主優待」「株価」「stock price」「IR」「investor」というキーワードを追加してWeb検索エンジンに検索要求を出す.ユーザは,検索エンジンが出力した検索結果から「東芝 投資家情報(IR)」のページを選択・閲覧して,目的としていた東芝の株式情報を入手することになる(図3の「検索結果」).

### 4. 情報検索インタフェースの構成

本研究では,普段使い慣れているWebブラウザから直接情報検索を実施できるようにするため,本インタフェースをWebブラウザのプラグインソフトとして実現した.本プラグインソフトは,図4に示すように,管理モジュール,UIモジュール,意味クラス解析モジュール,意図推定モジュールで構成されている.管理モジュールは,Webブラウザからのコンテンツの受け取り,モジュールの制御,モジュール間のデータの受け渡しを行うものである.UI(User Interface)モジュールは,キーワード選択,メニュー生成,ユー

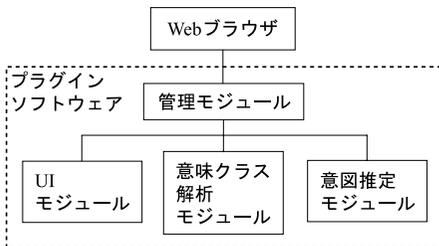


図4 情報検索インタフェースの構成  
Fig. 4 System architecture.

ザインタラクションなどに関するイベント処理を行うものである.意味クラス解析モジュールは,ユーザが選択したキーワードの意味を決定するものである.意図推定モジュールは,意味クラス解析モジュールが決定したキーワードの意味に基づいて検索方法を列挙し,ユーザが選択した検索方法に基づいて検索エンジンに検索要求を出すことを特徴としている.

以下では,インタフェース,意味クラス解析,意図推定についてそれぞれ説明する.

#### 4.1 インタフェース

本研究では,図2に示すメニュー形式のユーザインタフェースを採用している.ユーザがWebブラウザ上でキーワードを選択すると,キーワード付近に本メニューが提示されるようになっており,ユーザは本メニューと対話することによりキーワードに関連する情報を検索できるようになっている.

本インタフェースでは,図2(a)の箇所にユーザが選択したキーワードが表示される.これにより,ユーザはキーワードが適切に選択されているかどうかを確認することができる.図2(b)には,キーワードの意味が表示され,キーワードが多義性を持つ場合には適切な候補を選択できるようになっている.本メニューにおいて,キーワードの意味をさらにたどると,図2(c)のように,キーワードの意味に対応している検索方法の候補が提示されるようになっており,ユーザは検索方法の候補から目的に合致するものを選択できるようになっている.本メニューにおいて,提示されたキーワードの意味が誤っていたり,目的と合致する検索方法がなかったりする場合には,図2(d)の「すべての検索方法」から目的とするものを選択できる.また,2つ以上のキーワードを用いて情報検索を実施する場合には,図2(e)の「次のキーワードを選ぶ」を選択することにより,Webブラウザ上で次のキーワードをポインティングデバイスにより選択できるようになっている.

この場合,本インタフェースでは,最初に選んだキー

ワードの意味クラスのみが有効となっている。連続して選択した複数個のキーワードを無効としたい場合には、図 2 (f) の「キャンセル」を選択すればキーワードを新たに選択できるようになっている。

本インタフェースでは、キーワードに付与する意味クラスとして「人名」「地名」「企業・組織・施設」「交通機関」「駅・空港」「イベント」「その他(商品, 専門用語など)」の 7 種類を採用している。また、合計で 26 種類の検索方法を採用しており、それぞれの検索方法は 1 つ以上の意味クラスに対応している。本研究では、以下のプロセスを経て、意味クラスと検索方法を策定した。

- (1) 市販のホームページガイドやよく利用されているディレクトリ探索型 Web サイトを調査して有益な Web サイトを洗い出すとともに、各 Web サイトへの検索に対して思いのままに検索方法のラベルを付与する。
- (2) 各 Web サイトを高精度に検索できるキーワード(補助キーワードを含む)を発見するとともに、各キーワードに対して思いのままに意味クラスを付与する。
- (3) KJ 法のような階層的グループ化およびラベル付けを行うことにより、上記(1)で抽出したもののうち同様の検索方法をまとめて検索方法の数を減らす。
- (4) 上記(3)と同様のアプローチにより、上記(2)で抽出したもののうち同様の意味クラスをまとめるとともに、意味クラス解析性能とのトレードオフから対象となる意味クラスの種別を絞る。
- (5) 意味クラスと検索方法のすべての組合せを作り、アンケート調査を実施してユーザにとって有意義な組合せを洗い出す。
- (6) アンケート調査に基づいて本インタフェースを試作するとともに、ユーザテストを実施して検索方法を洗練する。

上記(5)のアンケート調査では、社内外のユーザを対象として、キーワードの意味クラスごとに日頃実施する検索の頻度を回答してもらうことにより、意味クラスと検索方法の対応関係を明確化した。本アンケート調査の結果から、意味クラスに対応する検索方法は、人名で 6 種類、地名で 9 種類、企業・組織・施設で 9 種類、交通機関で 6 種類、駅・空港で 8 種類、イベントで 8 種類、その他(商品, 専門用語など)で 8 種類となった。各検索方法と対応する意味クラスの一覧を付録 A.1 に示す。

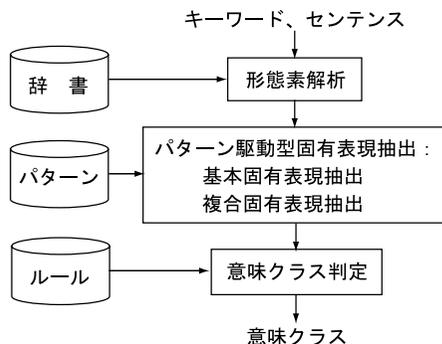


図 5 意味クラス解析処理の構成

Fig. 5 Semantic pattern analysis sub-system.

## 4.2 意味クラス解析

本インタフェースで実施する意味クラス解析処理では、ユーザが選択したキーワードの意味クラスをロバストに決定するため、当該キーワードを含むセンテンス全体を考慮することになっている。したがって、意味クラス解析モジュールは、ユーザがキーワードを選択した際、管理モジュールを介して Web ブラウザから当該キーワードを含むセンテンスを受け取るとともに、センテンスに対して形態素解析<sup>9)</sup>、パターン駆動型固有表現抽出<sup>9)</sup>、意味クラス判定を順次適用して、キーワードの意味クラスを決定している(図 5 参照)。

以下では、意味クラス解析処理を構成する形態素解析、パターン駆動型固有表現抽出、意味クラス判定のそれぞれについて説明する。

### 4.2.1 形態素解析

本形態素解析は、固有名詞 186,513 件を含む 334,797 件の語彙を有する辞書を用いて、センテンスから単語を抽出するとともに、単語に品詞を付与する。また、抽出した名詞に対して以下に示す属性を付与する。

- 名詞の意味カテゴリに関する情報  
合計：11 種類  
例：「犬：動物」、「ツツジ：植物」
- 他の名詞との共起情報  
合計：171 種類  
例：「大統領：人名と共起」、「区：地名と共起」
- 詳細情報  
合計：178 種類  
例：「ロンドン：首都名」、「守礼門：建物」

これらの属性はパターン駆動型固有表現抽出で利用される。

図 6 に、「東芝「RD-S600」は 600 GB の HDD を搭載した DVD レコーダーである。」というセンテンスに形態素解析を適用した例を示す。解析結果の例で

東芝「RD-S600」は600GBのHDDを搭載したDVDレコーダーです。

| ワード | 品詞(属性) | ワード      | 品詞(属性) |
|-----|--------|----------|--------|
| 東芝  | 名詞(企業) | GB       | 序数     |
| 「   | 記号     | の        | 付属語    |
| RD  | 英字     | HDD      | 英字     |
| -   | 記号     | を        | 付属語    |
| S   | 英字     | 搭載       | 両名     |
| 600 | 数字     | した       | 付属語    |
| 」   | 記号     | DVDレコーダー | 名詞     |
| は   | 付属語    | です       | 付属語    |
| 600 | 数字     | 。        | 句読点    |

図 6 形態素解析の例

Fig. 6 Examples of morphological analysis.

は、「東芝」という名詞に“企業”という属性が付与されている。

4.2.2 パターン駆動型固有表現抽出

パターン駆動型固有表現抽出では、あらかじめ作成した 500 個のパターン群を用いて、形態素解析結果から合計 114 種類の基本固有表現と複合固有表現を抽出する。基本固有表現抽出では、単語に付与された属性情報から個別の固有表現を抽出する。複合固有表現抽出では、既抽出の個別の固有表現の組合せから複合的な固有表現を抽出する(例：地名と番地の組合せから所在地を抽出する)。本論文では、紙面の都合上、すべてのパターンを列挙することと、それぞれのパターンについて説明することを省略する。また同様の理由により、すべての固有表現種別を列挙することを省略する。なお本方式では、各パターンが適用された際に、パターンにおいて定義されている固有表現種別と確信度を固有表現抽出結果に付与するようになっている。

参考までに、図 7 に基本固有表現抽出と複合固有表現抽出の例を示す。この例では、基本固有表現抽出により、「東芝：企業」「600 GB：量」「HDD：製品(部分)」「DVD レコーダー：製品」という抽出結果が得られている。また複合固有表現抽出で、“企業 + 「」→ 商品”というパターンを適用することにより、「RD-S600：商品」という抽出結果が得られている。

4.2.3 意味クラス判定

パターン駆動型固有表現抽出により、ユーザが選択したキーワードの固有表現種別が決定されれば、あらかじめ作成した対応表に基づいて当該固有表現種別を上述した 7 種類の意味クラスのいずれかに振り分ける。図 7 の例では、「東芝：企業」を“企業・組織・施設”に振り分け、「RD-S600：商品」「600 GB：量」「HDD：製品(部分)」「DVD レコーダー(製品)」をそれぞれ“その他(商品・専門用語など)”に振り分け

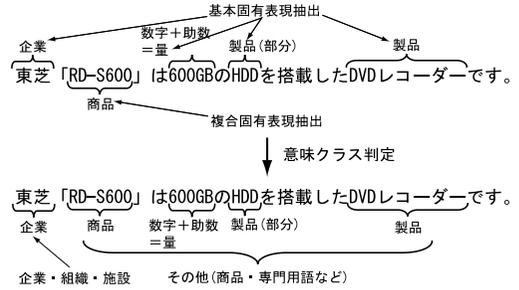


図 7 意味クラス解析の例

Fig. 7 Examples of semantic pattern analysis.

ている。なお本論文では、紙面の都合上、固有表現種別と意味クラスの対応関係の説明を省略する。

ユーザが選択したキーワードに複数の確信度付き固有表現種別が付与されていたら、確信度の高い順に上位 N 位までの固有表現種別を抽出する。たとえば、ユーザが閲覧文書上で「青山スパイラル」という語句をポインティングデバイスで選択したとする。このとき、基本固有表現抽出により「青山：地名・確信度 0.7」、複合固有表現抽出により「青山スパイラル：施設・確信度 0.8」および「青山スパイラル：人名・確信度 0.3」という結果がそれぞれ得られたとすると、意味クラス判定では、1 位：企業・組織・施設、2 位：地名、3 位：人名という結果を出力することになる。

キーワードに確信度付き意味タグが付与されていない場合には、あらかじめ作成した 42 個のルールを用いて、キーワードを構成する文字列パターンに基づいてキーワードの意味クラスを判定する。たとえば「大洗ホールディングス」は、漢字 + カタカナで構成される文字列と見なすことができるので、本意味クラス判定では当該文字列を“企業・組織・施設”と判定することになる。また「Yasuto Ishitani」は、先頭が大文字、以降が小文字の英字で構成される単語が 2 語以上続いているので、本方式では当該文字列を“人名”と判定することになる。なお本論文では、紙面の都合上、すべてのルールを列挙することと、それぞれのルールについて説明することを省略する。

以上から、本意味クラス解析処理では、ユーザが選択したキーワードに対して 1 つ以上の意味クラスとその優先順位を出力する。

4.3 意図推定

意図推定処理では、意味クラス解析により決定されたキーワードの意味を手がかりとして、あらかじめ定義されている検索意図知識群を探索することにより検索方法を列挙し、メニュー化する。

本インタフェースで用いる検索意図知識は「キー

ワード意味クラス」「検索方法」「補助キーワード」で構成されており、以下に示すように XML で記述されている。

検索方法「株式情報を調べる」の検索意図知識：

- < 検索意図知識 >
  - < キーワード意味クラス >
    - 企業・組織・施設
  - < / キーワード意味クラス >
  - < 検索方法 >
    - 株式情報を調べる
  - < / 検索方法 >
  - < 補助キーワード >
    - “投資家情報”|“個人投資家”|“株主優待”|“株価”|“stock price”|“IR”|“investor”
  - < / 補助キーワード >
  - < / 検索意図知識 >

本研究では、上述したアンケート調査で得られた 26 種類の検索方法それぞれに対して検索意図知識を作成するとともに、300 個のサンプル文書を用いた検索実験を実施することにより、それぞれの検索意図知識において補助キーワードを設定している。

次に、上述した検索意図知識に基づいて検索方法の候補を列挙する手順を示す。

**Step A:** 管理モジュールを介して、ユーザが指定したキーワードの意味クラスを受け取る。

**Step B:** 検索意図知識を 1 つ読み出す。

**Step C:** 意味クラス解析が決定したキーワードの意味と検索意図知識のキーワード意味クラスが合致すれば、この検索意図知識で定義されている検索方法をユーザに提示する検索方法の候補と見なす。

**Step D:** すべての検索意図知識について Step B と Step C を実施する。

**Step E:** 検索方法の候補を検索方法の選択頻度に関する長期学習と短期学習に基づいてランキングする。

本インタフェースでは、メニューにおいて検索方法を選択する際のユーザビリティを向上させるために、上記 Step E において検索方法の候補をランキングしている。このため本研究では、ユーザが検索方法を選択する頻度に関して長期学習と短期学習を実施している。長期学習では、本インタフェースが初めて利用されてから直前の検索まで、ユーザが選択した検索方法の頻度（以後、「長期頻度」と呼ぶ）をキーワードの意味ごとにカウントしている。短期学習では、本インタフェースを立ち上げたときにキューを初期化するとともに、本インタフェースによる検索が始まってから

当該キューにユーザが選択した検索方法を蓄積している。なお本研究では、このキューを、先入れ先出し方式でデータを操作するためのデータ構造であるとする。以下に、上記 E で実施している長期学習と短期学習に基づく検索方法のランキングについて説明する。

**Step E-1:** 上記 Step D で抽出した検索方法を、長期頻度の値の降順にソートする。

**Step E-2:** 短期学習で作成したキューから 1 つずつ検索方法を取り出す。

**Step E-3:** Step E-1 でソートした検索方法の中に、Step E-2 で取り出した検索方法と同じものがあればそれをソート結果の最上位に配置する。

**Step E-4:** 短期学習で作成したキューに蓄積されているすべての検索方法について Step E-3 を適用する。

このようにして列挙およびランキングされた検索方法は、管理モジュールを介して UI モジュールに渡され、UI モジュールでメニュー化される。ユーザが検索方法のメニューから適切な検索方法を選択すると、本インタフェースでは検索意図知識を特定できるようになり、当該検索意図知識に基づいてユーザが選択したキーワードに補助キーワードを追加して、検索エンジンに検索要求を出すことになる。

## 5. 実 験

本研究では、Web 検索を対象とした連鎖検索のための情報検索インタフェースを試作するとともに、本インタフェースを用いた実験を行うことによりその有効性を評価した。本試作では、Web ブラウザとして Microsoft Internet Explorer™ を、検索エンジンとして Google™<sup>5)</sup> をそれぞれ利用した。また事前に、開発サンプルとして Web コンテンツ約 300 ページを用いて試行錯誤により 26 種類の検索意図知識を作成した。

本インタフェースの有効性の評価では以下に示す実験を実施した。

- (1) 意味クラス解析精度と Web 検索精度の計測
  - (2) 被験者による検索性能とユーザビリティの満足度評価
  - (3) 被験者によるインタフェースの満足度評価
- 以下に、それぞれの実験結果について述べる。

**5.1 意味クラス解析精度と Web 検索精度の計測**  
試作したインタフェースを用いて、実際の Web コンテンツを対象として検索試行を 100 回以上実施するとともに、意味クラス解析精度と Web 検索精度をそれぞれ計測した。この検索試行では、付録 A.1 に示

表 1 意味クラス解析と Web 検索の実験結果

Table 1 Experimental results of semantic pattern analysis and Web search.

|                  | 日本語 Web 検索 (%) | 英語 Web 検索 (%) |
|------------------|----------------|---------------|
| 1 位再現率 (意味クラス解析) | 88             | 82            |
| 再現率 (意味クラス解析)    | 99             | 97            |
| 適合率 (意味クラス解析)    | 57             | 51            |
| 1 位正答率 (検索)      | 81             | 72            |
| 3 位以内正答率 (検索)    | 96             | 96            |

している「意味クラスと検索方法の組合せ」のすべてが網羅されている。意味クラス解析結果と検索結果を表 1 に示す。

表 1 で用いられている評価基準について以下に説明する。

1 位再現率 (意味クラス解析): 検索メニューの 1 位に正しい意味クラスが表示される確率

再現率 (意味クラス解析): 検索メニューに正しい意味クラスが表示される確率

適合率 (意味クラス解析): 検索メニューに表示された意味クラスのうち正しい意味クラスの割合

1 位正答率 (検索): 検索結果の 1 位に所望の情報が検索される確率

3 位以内正答率 (検索): 検索結果の 3 位までに所望の情報が 1 つ以上検索される確率

本実験結果により、意味クラス解析で用いる辞書/パターン/ルールが適切に定義されており、意図推定で用いる検索意図知識が適切に作成されていることを確認することができた。

## 5.2 被験者による検索性能とユーザビリティの満足度評価

### 5.2.1 評価条件

本評価では、下記条件のもとで日本語 Web 検索と英語 Web 検索をそれぞれ実施して、本インタフェースの検索性能とユーザビリティに対するユーザ満足度を測定した。

実施期間:

日本語 Web 検索: 2006 年 8 月 5, 6 日

英語 Web 検索: 2007 年 1 月 26, 29 日

対象ユーザ:

日本語 Web 検索: 国内日本人モニタ 15 名

英語 Web 検索: 国内米国人モニタ 15 名

実験内容: 各モニタ (以後、被験者と呼ぶ) が、スタートページから本インタフェースを用いて検索課題を解決する (ただし、スタートページと検索課題は実験担当者によってあらかじめ設定されるものとする)

表 2 日本語 Web 検索の課題とスタートページの一覧

Table 2 Overview of tasks for Web search in Japanese Websites.

| 検索課題            | スタートページ      |
|-----------------|--------------|
| 保土ヶ谷公会堂の地図を探す   | プロフィール       |
| 新富町駅の時刻表を調べる    | ブログ          |
| 洗濯乾燥機のレビューを探す   | メーカーサイトの商品紹介 |
| DVD レコーダの価格を調べる | メーカーサイトの商品紹介 |
| ノート PC の仕様を調べる  | ブログ          |

表 3 英語 Web 検索の課題とスタートページの一覧

Table 3 Overview of tasks for Web search in English Websites.

| 検索課題          | スタートページ       |
|---------------|---------------|
| イベントの予定を調べる   | ブログ           |
| 企業の沿革を調べる     | ニュースサイトの記事    |
| 用語の意味を調べる     | 口コミサイトの商品説明   |
| 人物のプロフィールを調べる | 公共機関のサイトのニュース |
| 自動車の評判を調べる    | メーカーサイトの商品紹介  |

使用機器: TabletPC (東芝 dynabook SS 200 166L/2X)

検索課題の概要を表 2 と表 3 に示す。

### 5.2.2 評価方法

実験では、実験担当者が立ち会いのもと、下記手順に従って被験者による評価を実施した。

Step 1: 評価担当者が被験者に操作方法を説明する。

Step 2: 被験者はシステムを 30 分間試用して本インタフェースの使い方に習熟する。

Step 3: 被験者は、各検索課題において、スタートページから検索キーワードを探して本インタフェースによる Web 検索と Google 検索をそれぞれ実施する。

Step 4: 被験者は、本インタフェースによる検索結果と Google 検索の結果をそれぞれ評価し、目的とする検索結果を特定するとともに、その検索順位を記録する。

Step 5: 各被験者は、すべての検索課題を実施したあと、本インタフェースによる検索と Google 検索に対して総合的な満足度を付与する。

Step 6: 各被験者が本インタフェースに対して自由に意見・感想を述べる。

なお本実験では、あらかじめ Google ツールバーをインストールすることにより、検索ボックスが Web ブラウザに設定されている状態で Google 検索を実施した。この場合、被験者による補助キーワードの追加を認めないこととした。これは、提案方式による絞り込み検索の有効性すなわち補助キーワードにより検索精度がどの程度向上するかについて厳密に測定することを目的としたためである。

表 4 検索性能の計測結果  
Table 4 Accuracy of information retrieval.

|          | 日本語 Web<br>検索 (%) |          | 英語 Web<br>検索 (%) |          |
|----------|-------------------|----------|------------------|----------|
|          | Google<br>検索      | 提案<br>方式 | Google<br>検索     | 提案<br>方式 |
| 1 位正答率   | 55.0              | 77.5     | 50.7             | 69.2     |
| 3 位以内正答率 | 75.0              | 91.3     | 59.4             | 83.1     |
| 5 位以内正答率 | 90.0              | 93.0     | 69.6             | 87.7     |

表 5 ユーザ満足度の計測結果  
Table 5 User evaluation results.

|       | 日本語 Web<br>検索 (人) |          | 英語 Web<br>検索 (人) |          |
|-------|-------------------|----------|------------------|----------|
|       | Google<br>検索      | 提案<br>方式 | Google<br>検索     | 提案<br>方式 |
| 非常に満足 | 0                 | 3        | 2                | 6        |
| 満足    | 8                 | 8        | 9                | 9        |
| やや満足  | 6                 | 4        | 4                | 0        |
| 不満    | 1                 | 0        | 0                | 0        |
| 計     | 15                | 15       | 15               | 15       |

### 5.2.3 評価結果

評価実験では、本インタフェースの検索性能と操作性を評価するため、1 位正答率のほかに、検索エンジンが出力する検索結果の上位 3 位 (5 位) 以内に目的とする検索結果が 1 つ以上含まれている確率: 3 位 (5 位) 以内正答率も計測した。目的とする情報が検索結果の上位 3 位もしくは 5 位までに含まれていれば、検索結果のスクロールやページ移動をしなくても目的とする情報にアクセスできることから、本研究では、情報検索の操作性を評価する基準として 3 位 (5 位) 以内正答率を採用することにした。また、本インタフェースによる絞り込み検索の有効性を確認するために、ユーザが指定したキーワードのみを用いた Google 検索 (5.2.2 項の Step 3) を実施するとともに、それぞれの検索結果を比較した。検索性能に関する実験結果を表 4 に、ユーザ満足度に関する実験結果を表 5 にそれぞれ示す。

検索性能に関する実験結果では、いずれのケースでも単独キーワードを用いた Google 検索より本インタフェースの方が高い検索精度を示しており、補助キーワードによる絞り込み検索が効果的に実現されていることを確認できた。また、ユーザ満足度に関する実験結果では、「非常に満足」もしくは「満足」と回答した人の総数が、Google 検索に対する場合よりも本インタフェースに対する場合の方が上回っていることから、本インタフェースのユーザビリティが良好であることを確認できた。

本実験では、被験者から本インタフェースに対する

意見・感想もいただいた。その結果、「キーワードをなぞるだけで検索できるので楽である」「検索時に検索ボックスに移動しなくてよいので便利である」「補助キーワードを思いつけないときに助かる」などの肯定的な感想を得ることができた。その一方で、本インタフェースにおいて検索性能が悪化したケースがあったという意見が得られ、補助キーワードの洗練化に課題があることが分かった。

本インタフェースにより検索性能が悪化したケースの概要について以下に示す。

#### ケース 1

キーワード: イベント名

検索方法: イベントスケジュールを調べる

検索結果: Google 検索では当該イベントのホームページが 1 位として検索されたが、本インタフェースによる検索ではこのホームページよりも昨年のイベントスケジュールが上位に検索された

#### ケース 2

キーワード: 人名

検索方法: プロフィールを調べる

検索結果: Google 検索では当該人物について解説したフリー百科事典の記事が 1 位として検索されたが、本インタフェースによる検索では当該人物のプロフィールページが 2 位以下で検索された。

今回のユーザ満足度評価では、被験者への負荷やコストの面から、一部の検索意図知識を用いた検索課題に限定せざるをえなかった。実験で用いた検索課題は典型的なケースであることがアンケート調査によって明らかになっていたものの、現実世界におけるユーザの幅広い検索要求をカバーするものではなかった。このため今回の実験では、本インタフェースに対するユーザ満足度評価が実用レベルで十分になされたとはいえないだろう。今後は、多くのユーザを対象とするとともに、長期間にわたる PC 利用を前提とする本インタフェースのユーザビリティ評価を実施する予定である。

### 5.3 被験者によるインタフェースの満足度評価

ConQuery<sup>2)</sup> や日経テレコン<sup>3)</sup> などが提供している情報検索サービスのインタフェースでは、閲覧している文書において語句やキーワードを選択すると、検索方法が固定型メニューとして提示される。日経テレコンのインタフェースでは検索方法は 3 つしか提示されないが、ConQuery ではユーザが自由に検索サービスを登録できるため、たくさんの検索方法が提示され

る可能性がある。検索方法の数が少ない場合には、どちらかという固定型メニューの方が検索方法を選びやすいが、検索方法の数が増えてしまうと、固定型メニューから目的とする検索方法を選び出すのに手間がかかってしまう。特に、固定メニューの項目数が短期記憶に関するマジックナンバ7(±2)を大きく超えてしまうと、固定メニュー型インタフェースの使い勝手が急速に悪くなってしまふことが予想される。

これに対して本インタフェースでは、検索方法を意味クラスごとにグループ化して、「意味クラスの選択」⇒「検索方法の選択」という階層的な選択を可能とすることにより、1つのメニューにおける検索方法の提示を最大で9個(すなわちマジックナンバの最上限)にとどめている(交通機関:6個,企業・組織・施設:9個,地名:9個,イベント:8個,駅・空港:8個,人名:6個,その他:8個)。また本インタフェースでは、4.2節で述べたように、性質の異なるヒューリスティック処理を相補的に組み合わせることにより、キーワード選択時に出力される意味クラスを平均2個程度としており、さらにユーザが意図するものを高い確率で1位に提示するようにした(5.1節参照)。また、ユーザの利用頻度に関する長期学習と短期学習に基づいてメニューにおける検索方法の配置を並べ替えることにより、ユーザは提示された検索方法を選択する際にメニューの上位に着目できるようにした。

そこで本実験では、被験者10名を対象として、本インタフェースによる適応的なメニュー提示に対するユーザ満足度評価を実施した。被験者は、まず本インタフェースを用いて「適応的なメニュー提示による検索」と「固定的なメニュー提示による検索」の両方を十分に実施し、次にどちらのメニュー提示の方が利用しやすいかを回答した。この場合、「固定的なメニュー提示による検索」では、図2(d)の「すべての検索方法」を選択して25個の検索方法をすべて表示し、その中から被験者にとって適切なものを選ぶようにした。その結果、8名が適応的なメニュー提示の方が利用しやすかったと回答し、2名が固定的なメニュー提示の方が利用しやすかったと回答した。

次に、本インタフェースの適応的なメニュー提示において、検索方法の並べ替えを行う場合と行わない場合についてのユーザ満足度評価を実施した。この場合も両方のケースで、被験者が検索試行を十分に実施し、どちらの方が利用しやすいかを回答した。その結果、6名が並べ替えを実施する場合の方が利用しやすかったと回答し、1名が並べ替えを実施しない場合の方が利用しやすかったと回答した。また、3名がどちらで

もないと回答した。

なお、本実験において固定的なメニュー提示を支持したユーザからは、「たとえ項目数が増えても、項目の配置を覚えてしまえば使いやすくなると思う」というコメントを得ることができた。このようなユーザからも支持される適応的なメニュー提示の実現については今後の課題としたい。

## 6. 考 察

本論文で提案した情報検索インタフェースは、ユーザが選択した語句やキーワードの意味に応じて、ユーザにとって適切な検索方法を適応的にメニュー提示するものであり、情報検索における手間を削減することを主な目的としている。一般的に、このような適応型メニューを備えたインタフェースは限定されたユーザ評価環境においてその有効性が認められても、長期間にわたって継続的に利用されるケースは少ないといわれている<sup>8)</sup>。すなわち、本論文で提案したインタフェースでは、検索方法の適応的な提示の精度がユーザの満足度を得られなければ、だんだん利用されなくなってしまふということが考えられる。この場合、結局、キーワードをカット&ペーストして検索ボックスに入力したり、固定メニュー型のインタフェース<sup>2),3)</sup>を利用したりする方が使いやすいということになるだろう。そこで本章では、本インタフェースを長期間にわたり継続利用する場合に対する課題を列挙するとともにその解決方法について考察する。

### 6.1 利用シーンの拡大

ユーザの知的生産として、Webブラウザを用いたWebページの検索・閲覧だけでなく、ワードプロセッサを用いた文書作成、スプレッドシートを用いた表計算、メーラを用いたメールの読み書き、ドキュメントビューワを用いた文書の閲覧などさまざまなケースが考えられる。本論文では主に、Webブラウザに本インタフェースを適用したケースについて紹介しているが、本研究では、ワードプロセッサ、スプレッドシート、メーラ、ドキュメントビューワでも同様にプラグインソフトによる情報検索インタフェースを実現している。これに加え、本インタフェースからデスクトップサーチエンジンへの接続を可能とすることにより、上述したさまざまなアプリケーションにおいて同一のインタフェースからWeb検索とデスクトップ検索の両方が可能となっている。このように、本研究では、知的生産の多くのケースで本インタフェースを利用できるようにしてあり、情報検索を駆使する知的生産のトータルスループットの向上を可能としている。

## 6.2 意味クラス解析の未知語対応

本インタフェースでは、意味クラス解析によりユーザが選択したキーワードの意味を推定するとともに、意味に応じた検索方法を適応的にメニュー提示するようになっている。このため、意味クラス解析の精度が低い場合には、ユーザに対して適切な検索方法を提示できないという問題が生じる。また、意味クラス解析では、未知語や複合語などに対しても高い解析精度を維持する必要があるものの、それはすぐに解決できない難しい課題であると見なされている。

本研究では、(1) 辞書照合による意味クラスの推定、(2) 前後関係による意味クラスの推定、(3) 部分推定結果に基づく意味クラスの推定、(4) 文字並びのパターンに基づく意味クラスの推定など、性質の異なるヒューリスティック処理を相補的に実施することにより、未知語や複合語に対してもロバストな意味クラス解析精度を実現している(4.2節参照)。この場合、本インタフェースでは、(1)→(4)の順に推定結果の確信度が低くなると考えており、確信度が低いヒューリスティック処理で意味クラスが推定された場合には、意味クラスの推定結果としてより多くの候補を出力して未知語や複合語に対するロバスト性を高めている。

本研究ではさらに、適応型メニューが提示した意味クラス解析結果が誤っており、ユーザが同メニューで「すべての検索方法」(図2(d)参照)を選んだ場合には、ユーザが選択したキーワードと正しい意味クラスの組合せを形態素解析辞書に登録できるようになっている。本インタフェースでは、このようなしくみによりユーザが性能向上に寄与することを可能としており、ユーザ満足度の低下をできるだけ回避するような工夫をしている。ただし現時点では、上述した7クラス以外の新しい意味クラスを本インタフェースで定義できないようになっている。これは、本インタフェースでは、新しい意味クラスが定義された場合に、その意味クラスを推定するためのヒューリスティック処理も同時に定義することは難しいためである。したがって本研究では、新しい意味クラスの定義と、それを可能とするためのヒューリスティック処理の新規生成については今後の課題として位置づけている。

## 6.3 検索意図知識のカバレッジ性

本インタフェースがユーザに提示する検索方法のバリエーションは、4.1節で述べたように、事前に不特定の被験者に対して実施したアンケート調査の結果を分析することにより抽出されたものである。本アンケート調査では、被験者が普段よく実施する検索に該当する「キーワードの意味クラスと検索方法の組合せ」

を選択してもらった。本アンケート調査では被験者に対して十分な時間が与えられていたが、被験者が普段漫然と実施している情報検索までの確に汲み取ることは難しかった。また、本インタフェースのユーザビリティを高めるために1つの意味クラスに対応する検索方法の数を限定する必要があったので、あらゆるユーザにとって適切な検索メニューを構成することも難しかった。

そこで本研究では、本インタフェースにおいて、意味クラスと検索方法の組合せや、検索方法における補助キーワードの設定をユーザが自由にカスタマイズできるようにすることにより、上述したアンケート調査の不十分性に対応するようにした。すなわち、本インタフェースでは、それぞれの意味クラスで、既存の検索方法を追加・削除できるようになっている。さらに、意味クラスごとに新しい検索方法を定義できるようになっている。このときユーザは、新しい検索方法において補助キーワードの組合せを自由に定義できるようになっており、メニューにおいて新しい検索方法の順位を任意に設定できるようになっている。以上から本インタフェースでは、それぞれのユーザが自分の検索行動に適した形で検索メニューをアレンジできるようになっている。本論文では、紙面の都合により、これらのカスタマイズを可能とするツールの詳細な説明を省略する。

なお今回の研究では、ユーザの実際の検索行為から、新しい検索方法を自動的に獲得する機能については今後の課題とした。今後は、本インタフェースの枠組みでそのようなことが可能であるかどうか検討し、可能であればどのようなアプローチが有効であるか見極める予定である。

## 6.4 サーベイ検索への対応

NTCIR<sup>1)</sup>やTREC<sup>4)</sup>などの情報検索システムの比較評価プロジェクトでは、情報検索のタスクとしてターゲット検索とサーベイ検索が定義されている。これを本研究に照らして考えてみると、ユーザの知りたいことがよく現れている文書を検索するタスクがターゲット検索であり、ユーザの知りたいことを含む文書を網羅的に検索するタスクがサーベイ検索であると見なすことができる。このような観点に立つと、本研究では、ターゲット検索にフォーカスしてそれぞれの検索方法で補助キーワードを設定するとともに、ターゲット検索を目的とした評価実験を実施したことになる。

これとは別に、本インタフェースでサーベイ検索に対応するためには、サーベイ検索に適した補助キーワードを策定するとともに、ターゲット検索とサーベ

イ検索で補助キーワードが異なる場合にはユーザの目的に応じてどちらかに切り替えられるインタフェースを構成する必要があるだろう。本インタフェースのコンセプトはこのような形への拡張に適していると思われるが、サーベイ検索への対応とサーベイ検索に適した評価方法の検討については今後の課題としたい。

#### 6.5 長期継続利用のための工夫

本インタフェースは、ユーザが普段実施する検索行為のうち典型的なケースを検索方法としてラインアップするとともにメニュー提示するものである。したがって、本インタフェースでユーザのあらゆる検索行動をカバーすることは難しいので、実際にはユーザは、本インタフェースを用いた検索と従来の検索を併用することになるだろう。この場合、併用の手間が煩わしいようであれば、本インタフェースは使われなくなり、結局従来の検索が利用され続けることになってしまうことが予想される。

そこで、ユーザの能動的な検索行為について検討してみたところ、本検索インタフェースが提供する連鎖検索のほかに、(1) 単独キーワードによる Web 検索、(2) 情報源を選別した検索（地図サイト、ショッピングサイト、乗換案内サイト、グルメ情報サイト、フリー百科事典サイトなど特定の Web サービスでの検索）、(3) 特定のテーマを掘り下げるための試行錯誤的な検索、などのバリエーションがあることが分かった。

このようなケースにも対応して本インタフェースの実用性を高めるために、本インタフェースでは次に示す 2 つの機能を採用した。

そのまま検索機能：閲覧文書においてユーザが選択したキーワードのみを用いた検索を可能とするため、メニューに「そのまま検索」の項目を追加した（図 2 参照）。ユーザがこの項目を選べば選択したキーワードに補助キーワードを設定せずに Web 検索を実施できるようになっている。これにより、上記 (1) に対応することが可能となる。

サイト指定機能：それぞれの検索方法において、補助キーワードのほかに、特定のサイトを指定できるようにした。この結果、たとえば、閲覧文書中である商品の名前を選択したとき、いつもよく利用しているショッピングサイトで価格情報を調べることが可能となる。

また本インタフェースでは、ユーザがメニューで検索方法を選択した後、検索エンジンに検索要求を出し、検索エンジンが検索結果を出力するという流れになっているので、検索エンジンの出力結果が表示された時点で自然な形で従来の検索に移行できるようになって

いる。したがって、検索エンジンの出力結果が表示された時点で、検索キーワードを吟味しながら検索ボックスに入力するという通常の検索行為が可能となるので、上記 (3) のような「特定のテーマを掘り下げた検索」を実施できるようになる。このため、本インタフェースを用いた検索からスタートしても、従来の検索に適宜合流できるようになっているので、ユーザは安心して本インタフェースを利用できると考えられる。

著者の 1 人が、情報検索を実施する際に「必ず本インタフェースからスタートする」というトライアルを数日間実施したところ、上述したような流れで情報検索を継続することが可能であることを確認した。ただし、連鎖検索結果から「深く掘り下げていく検索」に移行する場合には、すでに設定された補助キーワードを吟味しながらアレンジするのは煩雑であることが分かった。今後は、このような問題点を解決するとともに、上述した利用方法を許容できるかどうかについて不特定多数のユーザを対象とした評価を実施したい。

## 7. ま と め

本論文では、Web ブラウザで閲覧している文書から関連情報を次々とたどる連鎖検索のための新しい情報検索インタフェース技術を提案した。本インタフェース技術では、ユーザがポインティングデバイスを用いて Web ブラウザ上で語句を選択すると、意味クラス解析技術により語句の意味を決定し、意図推定技術により語句の意味に応じた検索方法を列挙およびメニュー化してユーザに提示するようになっている。そしてユーザがこのメニューから適切な検索方法を選べば、本インタフェースは選択された語句に補助キーワードを追加して効果的な絞り込み検索を行う。その結果、ユーザは、キーワード選択とメニュー選択という直感的で簡便な操作のみで、キーワードに関連する情報を効率良く的確に検索できるようになる。

このような情報検索では、検索クエリの検討、検索クエリの入力、検索結果の吟味に要する手間を大幅に減らすことができるので、移動しているとき、人と話しているとき、作業をしているとき、急いでいるときでも目的とする情報への確にアクセスすることが可能となる。今後は、意味クラス解析のロバスト性を向上させ、検索意図知識を拡充することにより、“どんな状態でも” 欲しい情報にアクセスできるインタフェース技術の実現を目指す。

## 参 考 文 献

- 1) <http://research.nii.ac.jp/ntcir/>

- 2) <https://addons.mozilla.org/ja/firefox/addon/231>
- 3) <http://telecom21.nikkei.co.jp/nt21/service/>
- 4) <http://trec.nist.gov/>
- 5) <http://www.google.co.jp/>
- 6) Adar, E., Kargar, D. and Stein, L.A.: Haystack: Per-user information environments, *Proc. 8th International Conference on Information and Knowledge Management*, pp.413-422 (1999).
- 7) Cutting, D.R., Karger, D.R., Pedersen, J.O. and Tukey, J.W.: Scatter/gather: A cluster-based approach to browsing large document collections, *Proc. 15th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp.318-329 (1992).
- 8) Findlater, L. and McGrenere, J.: A comparison of static, adaptive, and adaptable menus, *Proc. SIGCHI Conference on Human Factors in Computing Systems*, pp.89-96 (2004).
- 9) 市村由美, 齋藤佳美, 酒井哲也, 國分智晴, 小山誠: 質問応答と, 日本語固有表現抽出および固有表現体系の関係についての考察 (抽出), 情報処理学会研究報告 (自然言語処理研究会報告), Vol.2004, No.47, pp.17-24 (2004).
- 10) 小林拓海, 佐藤大介, 三末和男, 田中二郎: Web 検索結果の概観提示による情報収集支援インタフェース, 第 19 回人工知能学会全国大会予稿集, No.3C3-03 (2005).
- 11) Mandala, R., Tokunaga, T. and Tanaka, H.: The use of WordNet in information retrieval, *Proc. COLING/ACL Workshop on Usage of WordNet in Natural Language Processing Systems*, pp.31-37 (1998).
- 12) 増井俊之, 塚田浩二, 高林 哲: 近傍関係にもとづく情報検索システム, インタラクティブシステムとソフトウェア XI, 増井俊之 (編), 日本ソフトウェア科学会, pp.79-86 (2003).
- 13) 水口 充, 梅本あずさ, 柴尾忠秀, 浦野直樹: 提示型ユーザインタフェースの実装と評価, コンピュータソフトウェア, Vol.18, No.1, pp.13-27 (2001).
- 14) Rekimoto, J.: Time-machine computing: A time-centric approach for the information environment, *Proc. ACM Symposium on User Interface Software and Technology (UIST'99)*, pp.45-54 (1999).
- 15) Schilit, B.N., Golovchinsky, G. and Price, M.N.: Beyond paper: Supporting active reading with free form digital ink annotations, *Proc. SIGCHI Conference on Human Factors in Computing Systems*, pp.249-256 (1998).

## 付 録

### A.1 検索方法の一覧

本研究で用いる検索意図知識の一覧を以下に示す。各検索意図知識は検索方法, 対応意味クラス, 補助キーワードで構成されており, 日本語 Web 検索と英語 Web 検索で共通となっている。

- (1) 路線図 (Train route maps) を探す  
 対応意味クラス: 交通機関 (Transport)  
 補助キーワード: “路線図” | “運行路線” | “routes”
- (2) チケット (Tickets) を探す  
 対応意味クラス: 企業・組織・施設 (Company / Organization / Facility), 交通機関 (Transport), イベント (Event / Show)  
 補助キーワード: “観覧料” | “航空運賃” | “入場料” | “指定券” | “航空券” | “チケット” | “ticket” | “ticketing” | “round-trip” | “make reservation” | “Airfare” | “Travel Reservation” | “departure date” | “return date” - “hot ticket” - “blog” - “job opportunities”
- (3) 飲食店 (Restaurants) を探す  
 対応意味クラス: 地名 (Place), 企業・組織・施設 (Company / Organization / Facility), 交通機関 (Transport), イベント (Event / Show), 駅・空港 (Station / Airport)  
 補助キーワード: “レストラン” | “飲食店” | “グルメ” | “restaurant” | “restaurants” | “beverage” | “dining” | “Recommended Places”
- (4) 時刻表 (Timetables) を調べる  
 対応意味クラス: 交通機関 (Transport), 駅・空港 (Station / Airport)  
 補助キーワード: “時刻表” | “フライト情報” | “timetable” | “timetables” | “schedule” | “schedules” | “train number” | “flight schedule” | “flight schedules” | “Commuter Schedule” | “arriving from” | “departing to” | “arrivals” | “departures” | “flight information” | “flight #” | “flight tracker”
- (5) タウン情報 (Shopping mall, Stores) を調べる  
 対応意味クラス: 地名 (Place), 交通機関 (Transport), 駅・空港 (Station / Airport), その他: 商品, 専門用語など (etc)

- 補助キーワード：“商店街” | “タウン情報” | “エリアガイド” | “town guide” | “shopping” | “shopping mall” | “shopping centre” | “shopping center” | “tenants”
- (6) ニュース (News articles) を探す  
 対応意味クラス：地名 (Place), 駅・空港 (Station / Airport)  
 補助キーワード：“ニュース” | “news” “copyright” | “All rights reserved”
- (7) 説明, 概要, 沿革 (Explanations, Overviews, Histories) を調べる  
 対応意味クラス：地名 (Place), 企業・組織・施設 (Company / Organization / Facility), イベント (Event / Show), 駅・空港 (Station / Airport), その他：商品, 専門用語など (etc)  
 補助キーワード：“とは” | “意味” | “用語” | “解説” | “仕様” | “スペック” | “概要” | “一般事情” | “基本情報” | “地域情勢” | “企業概要” | “会社概要” | “企業情報” | “企業概要” | “ガイド” | “タウン情報” | “プロフィール” | “百科事典” | “history” | “overview” | “explanation” | “definition” | “encyclopedia”
- (8) ホテル (Hotels, Accommodations) を探す  
 対応意味クラス：地名 (Place), 駅・空港 (Station / Airport)  
 補助キーワード：“宿泊予約” | “ホテル予約” | “宿泊施設” | “hotel” | “hotels” “予約” | “reservation” inurl:&
- (9) 連絡先 (Contact information): 電話番号, 住所を調べる  
 対応意味クラス：企業・組織・施設 (Company / Organization / Facility), イベント (Event / Show), 駅・空港 (Station / Airport), 人名 (Person)  
 補助キーワード：“電話” | “電話番号” | “連絡先” | “問い合わせ” | “問合せ” | “所在地” | “住所” | “本社所在地” | “カスタマースク” | “0120” | “TEL” | “FAX” | “telephone” | “phone” | “contact” | “contacts” | “fan mail” | “contact information” | “Neighbourhood” | “contact name” | “mailing address” - “contact lens” - “battery” - “phone hire” - “rental” - “携帯電話”
- (10) 施設内の案内図 (Floor map) を探す  
 対応意味クラス：駅・空港 (Station / Airport)  
 補助キーワード：“構内図” | “構内案内図” | “平面図” | “フロアマップ” | “ターミナルマップ” | “ターミナル案内図” | “フロア案内図” | “空港案内” | “terminal map” | “terminal maps” | “wayfinding information” | “Airport Layout” | “terminal layout” | “station map”
- (11) ツアー (Tours, Trips) を探す  
 対応意味クラス：地名 (Place)  
 補助キーワード：“ツアー” | “tour” | “tour package” | “package tour” | “tours” | “packages” | “travel”
- (12) 公式サイト, ポータルサイト (Official sites, Home pages, Portal sites) を探す  
 対応意味クラス：企業・組織・施設 (Company / Organization / Facility), イベント (Event/Show)  
 補助キーワード：なし
- (13) 写真 (Photos, Figures, Illustrations) を探す  
 対応意味クラス：企業・組織・施設 (Company / Organization / Facility), イベント (Event/Show), 人名 (Person), その他：商品, 専門用語など (etc)  
 補助キーワード：“写真” | “画像” | “近影” | “jpeg” | “portrait” | “photo” | “photograph” | “photo gallery”
- (14) イベントスケジュール (Event schedule) を調べる  
 対応意味クラス：イベント (Event / Show)  
 補助キーワード：“開催日程” | “開催日” | “日程” | “Jan” | “Feb” | “Mar” | “Apr” | “Jun” | “Jul” | “Aug” | “Sep” | “Oct” | “Nov” | “Dec”
- (15) 書籍, 文献 (Books, Literature) を調べる  
 対応意味クラス：人名 (Person)  
 補助キーワード：“ISBN” | “ASIN” “出版社” | “著” | “Publisher” | “Product Details”
- (16) 価格, 購入情報 (Prices, Purchase information) を調べる  
 対応意味クラス：その他：商品, 専門用語など (etc)  
 補助キーワード：“料金” | “利用料金” | “宿泊料” | “宿泊プラン” | “価格” | “会費” | “参加費” | “費用” | “購入” | “予約” | “通販”

- | “ショッピング” | “先行” | “前売” | “compare prices” | “shipping” | “reservation” | “rate” | “fee” | “room charge” | “deals” | “in stock” | “1-Click ordering”
- (17) 地図, アクセス情報 (Maps, Access information) を探す  
 対応意味クラス: 企業・組織・施設 (Company / Organization / Facility), イベント (Event / Show)  
 補助キーワード: “地図” | “行き方” | “へのアクセス” | “アクセスマップ” | “所在地” | “周辺地図” | “交通のご案内” | “来場案内” | “案内図” | “付近図” | “タクシー” | “電車” | “map” | “maps” | “direction” | “directions” | “location” | “locations” | “locate” | “venue” | “Map and Directions” | “directions to” | “directions from” | “driving direction” | “driving directions” | “street map” | “venue information” - “サイトアクセス” - “リモートアクセス”
- (18) 評判 (Reports, Evaluations) を調べる  
 対応意味クラス: 地名 (Place), 企業・組織・施設 (Company / Organization / Facility), 人名 (Person), その他: 商品, 専門用語など (etc)  
 補助キーワード: “感想” | “レビュー” | “クチコミ” | “口コミ” | “報告” | “レポート” | “report” | “review” | “statement” | “survey” | “rating” | “evaluation” | “Editorial Reviews” | “Customer Reviews” | “Book Review” - “privacy statement”
- (19) ブログ (Blog) を探す  
 対応意味クラス: 企業・組織・施設 (Company / Organization / Facility), 人名 (Person), その他: 商品, 専門用語など (etc)  
 補助キーワード: “ブログ” | “blog” | “track-back”
- (20) レシピ (Recipe) を調べる  
 対応意味クラス: その他: 商品, 専門用語など (etc)  
 補助キーワード: “レシピ” | “作り方” “材料”
- (21) オークション (Auctioned items) で探す  
 対応意味クラス: その他: 商品, 専門用語など (etc)  
 補助キーワード: “オークション” | “出品者” | “入札” | “end time” | “time left” | “place a bid” | “auction info” | “seller info”
- (22) 天気予報 (Weather forecast) を調べる  
 対応意味クラス: 地名 (Place)  
 補助キーワード: “天気予報” | “weather forecast” | “weather report”
- (23) 役所 (Government office) のサイトを調べる  
 対応意味クラス: 地名 (Place)  
 補助キーワード: “市役所” | “区役所” | “県庁” | “都庁” | “府庁” | “道庁” | “町役場” | “村役場” | “city of” | “public office” | “government” | “council”
- (24) 株式情報を調べる (日本語検索のみ)  
 対応意味クラス: 交通機関 (Transport)  
 補助キーワード: “投資家情報” | “個人投資家” | “株主優待” | “株価” | “stock price” | “IR” | “investor”
- (25) Jobs (英語検索のみ)  
 対応意味クラス: Place, Company / Organization / Facility  
 補助キーワード: “Jobs at” | “Jobs offered” | “Job title”
- (26) プロフィール (Profile) を調べる  
 対応意味クラス: 人名 (Person)  
 補助キーワード: “プロフィール” | “略歴” | “経歴” | “profile” | “biography”  
 (平成 19 年 3 月 30 日受付)  
 (平成 19 年 9 月 3 日採録)



石谷 康人 (正会員)

1988 年明治大学工学部電気工学科卒業。1990 年明治大学大学院工学研究科電気工学専攻博士前期課程了。同年 (株) 東芝入社。現在, 同研究開発センターにて, 文書解析・認識, 自然言語処理, 情報検索, ヒューマンインタフェースの研究開発に従事。博士 (工学)。平成 5 年度電子情報通信学会学術奨励賞受賞。電子情報通信学会, 言語処理学会, IEEE 各会員。



鈴木 優 (正会員)

1993年早稲田大学理工学部応用物理学科卒業。1995年早稲田大学大学院理工学研究科物理学及応用物理学専攻修士課程修了。同年(株)東芝入社。現在、研究開発センターにて知識共有システム、情報検索インタフェースの研究開発に従事。人工知能学会会員。



布目 光生 (正会員)

1999年北海道大学工学部電気工学科卒業。2001年北海道大学大学院工学研究科電子情報工学専攻修士課程修了。同年(株)東芝入社。現在、研究開発センターにて、テキストマイニング、情報抽出、文書構造化技術の研究開発に従事。人工知能学会会員。

---