

# インタラクティブなペン操作を可能とする検索意図に基づく連鎖情報検索

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

閲覧中の文書に関連する様々な情報を簡単なペン操作のみで検索できる新しい検索方法（連鎖情報検索）を提案する。閲覧している文書中の任意のキーワードにペンでアンダーラインを引くと、固有表現抽出技術によりキーワードやその周辺の語句の意味を解析してユーザの検索意図を推定し、検索意図に沿った適切な検索クエリを生成して的確な絞り込み検索を実行するという特徴を備える。これにより検索クエリの検討と入力の手間を省き、検索結果の吟味や探索を必要としないペンインタフェースによるインタラクティブな検索方法の実現を目指す。本論文では提案手法に基づくウェブ検索アプリケーションを試作し、ユーザに自由に試用してもらいながら検索精度の評価を実施した。評価の結果、実行された検索要求の80%について検索結果の上位5位以内にユーザの所望の情報が出力され、連鎖情報検索が簡便な操作性と的確な情報検索を両立する検索方法であることが示された。

## Successive Information Retrieval Method with Pen-based Interaction based on Estimation of User Intention

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

We propose a novel information retrieval method with a pen-based simple interaction by which the user can acquire various kind of information that relates to the document which he/she is reading. It is designed to enable a precise retrieval by the query which is expanded in consideration of the user's intention estimated by the meaning of the keyword specified by the underlining operation and the surrounding words. Named entity extraction technology is applied to it to obtain the meaning of words. It aims at the achievement of the pen-based retrieval method that saves time to formulate and input the query and does not need the close examination of the retrieval results. We implemented the web search system on the existing web browser as an application of our proposed method and conducted evaluation to confirm the effectiveness of it. The result shows that the users can acquire relevant information in the top 5 retrieval results in the case of 80% of the tested queries. We can conclude that our proposed method enables both an easy operation and a precise retrieval.

### 1. はじめに

コンピュータネットワークの発展や記録媒体の大容量化に伴い、大量の情報にアクセスできるようになった。その結果、文書作成や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、頻繁に実施されるようになった。現状では、ユーザは以下の手順で情報検索を行い、必要とする情報にアクセスしていると考えられる。

- (1) 検索クエリの検討
- (2) 検索クエリを入力
- (3) システムが提示した検索結果の吟味
- (4) 目的に合致する情報の選択および閲覧

このような情報検索のプロセスが煩雑であれば、本来

の知的生産において思考がたびたび中断されるようになり、作業効率が低下してしまうという問題があった。

大量の情報から所望の情報を見つけ出してアクセスする場合、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ることがある。この場合、上記(1)では、検索クエリに補助的なキーワードを追加したり、キーワード間の関係をブール式で定義したりして検索結果の絞り込みを行う必要がある。その結果、上記(2)において、長く複雑な検索クエリをキーボードを介して検索システムに入力することになる。それでも所望の情報が検索結果の上位に出力されない場合には、上記(3)において検索結果をスクロールしたり、次の検索結果へ移動したりすることにより所望の情報を探すことになる。上記(3)及び(4)の作業は、一般的に、マウスなどのポインティングデバイスを用いて実施されることが多いため、検索のたびに入力デバイスの切り替えが頻繁に生じることになる。そこで筆者らは、

<sup>†</sup> (株) 東芝 研究開発センター 知識メディアラボラトリー  
Knowledge Media Laboratory,  
Toshiba Corporate Research & Development Center

上述した問題を解決するために、ペンインタフェースによる簡便な操作性を備え、閲覧している文書に関連する情報をユーザの思考や知的生産を妨げることなく的確に検索することができる新しい検索方式（以後、連鎖情報検索と呼ぶ）を提案する。

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、その箇所にアンダーラインを引くことがある。これは多くの人が長年にわたって慣れ親しんだ行為であり、その度に思考が中断されるようなことはない。そこで提案方式では、閲覧している文書の任意の語句にアンダーラインを引くとその語句に関連する情報を検索することができるペンベースのユーザインタフェースを採用することにする。

アンダーラインが引かれた語句のみを用いて検索を行うだけでは、上述したように検索システムにおいて大量の検索結果が出力される恐れがある。そこで本研究では、ユーザとの簡単なインタラクションによってユーザの検索意図を推定すると共に、検索意図に応じた検索クエリを適切に設定して検索結果を的確に絞り込むこととする。さらに、これらのユーザインタラクションをすべてペン操作で実施するものとする。ここで検索意図とは、企業名を手がかりとしてその企業の株価を検索したい、人物名を手がかりとしてその人の著作を検索したいなど、検索対象が明確となっている場合のユーザの検索要求を意味するものとする。

本研究では、以上のようなアプローチを採用することにより、検索クエリの検討と入力の手間を省き、検索結果の吟味や探索を必要とせず、入力デバイスの切り替えが生じない検索方法を実現することを目指す。

本論文では、まず、2章において関連研究について述べ提案方式の位置づけを明らかにする。次いで、3章で連鎖情報検索の基本コンセプトについて述べ、4章で連鎖情報検索方式の構成について述べる。5章では、連鎖情報検索の応用として試作したインタラクティブなウェブ検索アプリケーションについて説明する。そして、6章でウェブ検索アプリケーションのユーザ評価について述べ、提案方式の有効性を示す。

## 2. 関連研究

情報検索の分野では、検索時のユーザの負担を軽減することを目的とした様々な方法が提案されている。

検索精度を高める方法の一つとして、ユーザが入力した検索クエリをシソーラスを用いて拡張する研究<sup>1)</sup>がある。この方式では、シソーラスにある類義語や概念的に上位もしくは下位の語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。一般的に、再現率と適合率はトレードオフの関係にあるため、この方式では、適合率が下がり検索結果中に占める正解の割合が低くなってしまうことがある。連鎖情報検索もシソーラスを用いて検索クエリを拡張するが、ユーザの検索意図に沿って検索クエリを拡張するため、適合率を向上することが可能となっている。

情報検索のユーザインタフェースに関する研究としては、クラスタリング技術を用いて類似した検索結果をグループ化することにより検索結果の閲覧を容易にする研究<sup>2),3)</sup>や、関連する情報をシステムが自動的に選択候補としてユーザに提示する提示型ユーザインタフェースに関する研究<sup>4)</sup>などがある。いずれの方式も、検索結果を閲覧しながら徐々に所望の情報を絞り込んでいくことができ、ユーザの検索意図が漠然としている場合に有効である。しかし、クラスタリング技術を用いた方式では、検索結果のクラスタが必ずしもユーザの検索意図に対応しているとは限らず、ユーザは所望の情報を含むクラスタを探すために各クラスタを評価しなくてはならない。また、提示型ユーザインタフェースでも、ユーザに提示された選択候補がユーザの検索意図に対応しているとは限らないため、ユーザは所望の情報が提示されるまで待たなければならないという問題がある。連鎖情報検索では、検索を実行する前にユーザとの簡単なインタラクションによりユーザの検索意図を明確化し、検索意図に基づいて検索クエリを設定しているため、所望の情報に効率良くアクセスできるようになっている。

連鎖情報検索と同様にペンインタフェースを用いて情報検索を行うシステムとして XLibris<sup>5)</sup> が知られている。これは、文書の読解支援を目的としたシステムであり、ユーザはタブレット型の PC とデジタルペンで構成されたインタフェースを用いて、文書上にアンノテーションやマーキングを行うことができる。さらに、この方法はペンインタフェースによる関連文書の検索を可能としており、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方法で検索クエリを指定することができ、キーワード検索と類似文書検索の切り替えが容易になっている。この方式では、検索クエリの入力の省力化にとどまっているため、検索結果が大量に出力される場合には、そこから所望の情報を探索するコストが増大する。連鎖情報検索は、簡便なペン操作で、検索クエリの設定・入力と、所望の情報への効率的なアクセスの両方を可能としており、XLibris に比べてより多くのユーザベネ

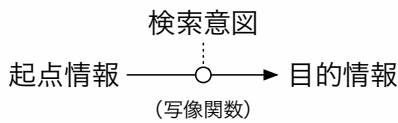


図 1 連鎖情報検索における検索処理の模式図

フィットを提供することができる。

### 3. 基本コンセプト

連鎖情報検索は、図 1 に示すように、「ユーザの検索意図に基づいて、閲覧文書中の起点となる語句から目的とする有益な情報へとナビゲートする機能」と定義することができる。本研究では、この定義に基づいて、連鎖情報検索を以下の手順に従うインタラクティブな検索として定式化する。

**Step 1:** 閲覧中の文書においてユーザがキーワードにアンダーラインを引く。

**Step 2:** キーワードとその周辺の語句の意味からシステムがユーザの検索意図を推定し、検索方法の候補を提示する。

**Step 3:** 提示された検索方法の候補からユーザが適切な検索方法を選択する。

**Step 4:** ユーザが選択した検索方法に基づいてシステムが検索処理を実行する。

**Step 5:** システムが出力した検索結果からユーザが所望の情報を選択・閲覧する。

連鎖情報検索では、文書を表示する際にバックグラウンドで文書の意味を解析しており、Step 1 で指定されたキーワードとその周辺の語句の意味を獲得できるようになっている。文書の意味解析結果の例を図 2 に示す。そして、これらの語句の意味の分布に対して検索意図知識をあらかじめ定義できるようになっている。検索意図知識は、図 3 に示すように、意図推定知識と検索知識で構成されている。意図推定知識では、ユーザが指定したキーワードの意味種別 (図 3 中の <キーワード意味種別>)、その周辺語句の意味種別 (図 3 中の <周辺語句意味種別>)、検索方法 (図 3 中の <検索方法>) を定義できるようになっている。検索知識では、WWW 上でウェブコンテンツを検索する「ウェブ検索」とクライアント PC 内のローカルコンテンツを検索する「ローカル検索」のいずれかの検索カテゴリを定義でき、検索クエリの拡張を目的とした追加キーワードを定義できるようになっている。

Step 2 では、Step 1 で指定されたキーワードとその周辺語句の意味情報に基づいて検索意図知識を参照

### 入力例

```

技術委員会の開催について
標記会議を下記により開催いたしますので、万障お繰り合わせのう
えご出席下さいますよう、ご案内申し上げます。
記
日時：平成 18 年 3 月 7 日(火)午前 10 時～11 時 30 分まで
場所：(株)東西(東京都港区芝浦 1-1-1)本社ビル 3 9 階大会議室
議題：インタラクシオン 2006 参加報告
/鈴木一郎(東西研究開発センター)
以上

```

### 出力例

```

<ORG>技術委員会</ORG>の開催について\\<EVENT>標記会
議</EVENT>を下記により開催いたしますので、万障お繰り
合わせのうえご出席下さいますよう、ご案内申し上げます。\\
記\\日時：<DATE>平成 18 年 3 月 7 日</DATE><(火)<TIME>
午前 10 時</TIME>～<TIME>11 時 30 分</TIME>まで\\
場所：(株)<COMPANY>東西</COMPANY><(JAPAN_CITY>東京
都港区芝浦</JAPAN_CITY>1-1-1) 本社ビル<NUM>3
9 階</NUM>大会議室\\議題：インタラクシオン 2006 参
加報告/<PERSON>鈴木一郎</PERSON><(COMPANY>東
西</COMPANY>研究開発センター</ORG>)\\以上

```

図 2 固有表現抽出処理の例

```

<検索意図知識>
<意図推定知識>
<キーワード意味種別>
COMPANY
</キーワード意味種別>
<周辺語句意味種別/>
<検索方法>日経:株価</検索方法>
</意図推定知識>
<検索知識>
検索カテゴリ="ウェブ検索"
追加キーワード="株価 site:nikkei.co.jp"
</検索知識>
</検索意図知識>

```

図 3 検索意図知識の例

すると共に、意図推定知識を用いてユーザに対して適切な検索方法を提示する。Step 3 でユーザが検索方法を選択した場合、Step 4 でその検索方法に対応する検索意図知識の検索知識を参照して検索クエリを適切に拡張し検索を行う。

以下に、図 4 に示す「ウェブ文書を閲覧している際にウェブ検索をインタラクティブに行う例」を用いて連鎖情報検索の手順を具体的に説明する。図 4 の例では、ユーザが閲覧している文書中の「東芝」という語句にアンダーライン (図 4(a)) が引かれている (Step 1)。この場合システムは、アンダーラインが引かれている語句の意味を「企業」あるいは「建物」と解析しており、これに基づいて検索意図知識を参照して、ユーザの検索意図が「地図を見たい」「関連するニュースを探したい」「プレスリリースを読みたい」「ホームページ

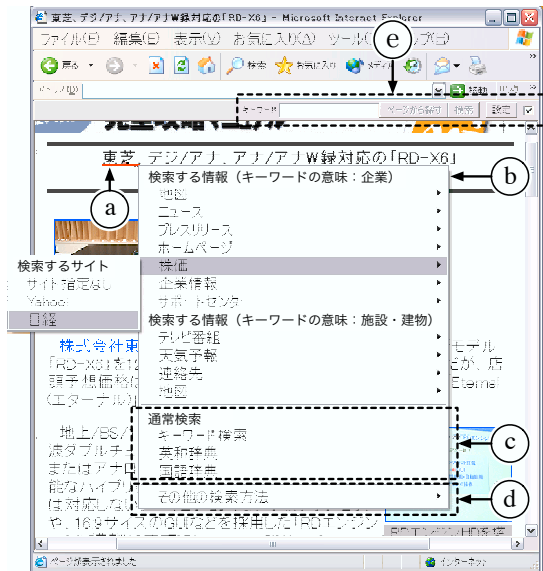


図4 ユーザインタフェースの例

を閲覧したい」「株価を知りたい」「企業情報を調べたい」などのいずれかであると推定している。そしてシステムは、該当する検索意図知識に基づいて、図4(b)に示すコンテキストメニューにより「地図」「ニュース」「プレスリリース」「ホームページ」「株価」「企業情報」などの検索方法をユーザに提示する(Step 2)。もしユーザが、特定のウェブサイトから情報を得たいと考えている場合には、提示されたコンテキストメニューにある特定のウェブサイトから所望のものを選択できるようになっている(Step 3)。ここで特定のウェブサイトとして「日経」が指定された場合、システムは検索意図知識の検索知識を参照して、検索クエリを「東芝 株価 site:nikkei.co.jp」として設定する(Step 4)と共に、この検索クエリを用いてウェブ検索を実施する。最後に、ユーザは検索結果の中から所望の情報を探し出して閲覧することになる(Step 5)。

#### 4. 連鎖情報検索の構成

連鎖情報検索は、図5に示すように、紙文書、ウェブコンテンツ、メールなどをクライアントPCに蓄積する(ローカル化する)情報蓄積部と、ローカル化された情報とWWW上にあるウェブコンテンツを検索する情報検索部で構成されている。以下に、情報蓄積部と情報検索部のそれぞれについて説明する。

##### 4.1 情報蓄積部

情報蓄積部は、ローカル文書を検索するためのインデクス情報を作成および蓄積するものであり、OCR・文書構造解析、固有表現抽出、XML文書変換、イン

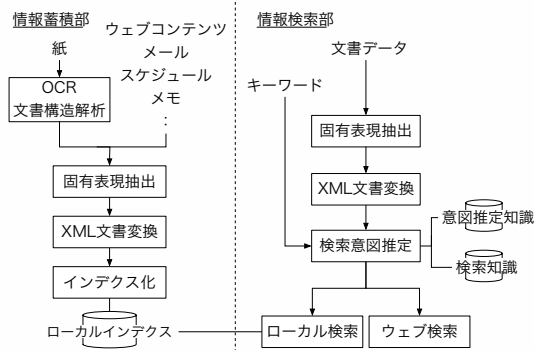


図5 連鎖情報検索システムの構成

デクス化で構成されている。情報蓄積部は、入力文書に対して固有表現抽出処理とXML文書変換処理を順次実施して、入力文書から意味情報や文書構造情報を抽出しインデクス化する。これにより、情報検索部のローカル検索では、意味や構造を手がかりとした絞り込み検索が可能となる。以下に、各構成要素について説明する。

**OCR・文書構造解析** 紙文書が入力された場合、文献<sup>6)</sup>に基づくOCR処理と文書構造解析処理を適用して、紙文書をプレーンテキストもしくは章節構造や箇条書き構造を保持したXHTML文書に変換する。

**固有表現抽出** 入力文書に対して文献<sup>7)</sup>の固有表現抽出処理を適用することにより、文書中の語句に対して固有表現クラスを付与する。ここで固有表現クラスとは、人物名(PERSON)、企業名(COMPANY)、地名(JAPAN\_CITY)などの固有名詞や、E-mailアドレス(EMAIL)、URL(URL)、日時(DATE, TIME)、数量(NUM)、金額(MONEY)など文中の表層表現の意味を表すラベルである。付与された固有表現クラスは文書中にXMLタグとして埋め込まれる。固有表現抽出結果の例を図2に示す。

**XML文書変換** 固有表現抽出結果に対して文献<sup>8)</sup>のXML文書変換処理を適用することにより、固有表現のグループやグループの階層構造などを抽出してXML化する。

**インデクス化** 入力文書に対して、形態素、固有表現と固有表現クラス、固有表現のグループと階層構造、文書の種別、書誌情報などのインデクスを作成する<sup>9)</sup>。

##### 4.2 情報検索部

情報検索部は、ユーザとのインタラクションによっ

てユーザの検索意図を推定し、検索意図に基づいた情報検索を実行するものであり、固有表現抽出、XML 文書変換、検索意図推定、ウェブ検索、ローカル検索で構成されている。情報検索部では、固有表現抽出結果と XML 文書変換結果に基づいて、ユーザが指定したキーワードとその周辺語句の意味を抽出すると共に、抽出結果に基づいてユーザの検索意図を推定するようになっている。そして、検索意図推定結果に基づいて検索方法を決定したあと、検索クエリを拡張して検索を行う。以下に、各構成要素について説明する。

**固有表現抽出** ユーザが文書データを閲覧する際に、バックグラウンドで文書から固有表現を抽出する。

**XML 文書変換** 固有表現抽出結果から固有表現のグループおよびグループの階層構造を抽出して XML 化する。

**検索意図推定** あらかじめ用意した複数の意図推定知識（図 3 参照）を順次参照しながら以下の手順に基づいて検索意図の推定を行う。

- (1) 入力文書の固有表現抽出結果と XML 文書変換結果から、ユーザが指定したキーワードの意味種別と、その周辺の語句の意味種別を抽出する。
- (2) 検索意図知識をひとつ読み出す。
- (3) (1) で抽出した意味種別が検索意図知識の意図推定知識で定義されているならば、この意図推定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。
- (4) 全ての検索意図知識について (2) 以降の処理を繰り返す。

こうして得られた検索方法の候補をコンテキストメニューに記載してユーザに提示する。一度に複数の検索方法が選出された場合には、これらの中からユーザが適切な検索方法を選択するものとする。

**ウェブ検索** ユーザが選択した検索方法に対応する検索意図知識において検索知識の検索カテゴリがウェブ検索と定義されている場合、ウェブ検索サービスに検索クエリを供給してウェブ検索を実施する。

**ローカル検索** ユーザが選択した検索方法に対応する検索意図知識において検索知識の検索カテゴリがローカル検索と定義されている場合、情報蓄積部がローカル化したコンテンツを対象とした情報検索を実施する。情報蓄積部では、形態素、固有表現と固有表現クラス、固有表現のグループと階層構造、文書の種別、書誌情報など様々な観点でインデクス化処理を実施しているため、メール、ス

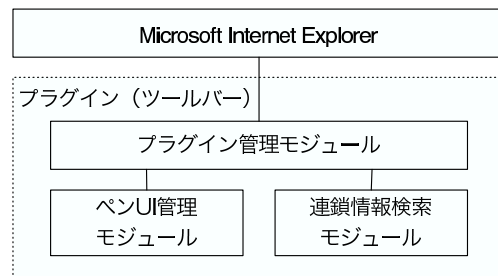


図 6 システム構成

ケジュール、メモ、紙文書などの文書種別による絞り込み検索、人物名、企業名、地名、日時などの固有表現クラスによる絞り込み検索、グループの階層構造による絞り込み検索のような高度な検索が可能となっている。

## 5. 連鎖情報検索によるインタラクティブなウェブ検索

ユーザが日常的に利用しているアプリケーションで文書を閲覧している際に、アプリケーションの画面上の語句にアンダーラインを引いて検索を行うことができればユーザにとって利便性が高くなる。そこで、多くのユーザが頻繁に利用するアプリケーションのひとつと見なすことができるウェブブラウザに連鎖情報検索を組み込んでインタラクティブなウェブ検索アプリケーションを試作した。

### 5.1 システム構成

ウェブ検索アプリケーションの構成を図 6 に示す。本研究では連鎖情報検索を組み込むアプリケーションとして Microsoft Internet Explorer™ を選択した。本アプリケーションは、ウェブブラウザのプラグインとして HTML データの読み出しやウェブ検索サービスへの接続などの処理を行うプラグイン管理モジュールと、4 章で説明した情報検索部の機能を提供する連鎖情報検索モジュールと、ペンによるユーザインタラクションを司るペン UI 管理モジュールから構成されている。

### 5.2 ユーザインタフェース

本アプリケーションは既存のウェブブラウザに組み込まれており、ユーザは普段通りにウェブブラウザを起動するだけでペンインタフェースによるウェブ検索アプリケーションを利用することができる。図 4 が IE に連鎖情報検索の機能を組み込んだ場合のユーザインタフェースの例になっており、ウェブブラウザ上に表示された文書に直接アンダーラインを引くことを可能としている。ハイパーリンクやブックマークメニュー、

「戻る」ボタンなど、ウェブブラウザが備える機能もそのまま利用可能となっている。

また、検索意図に基づく検索クエリの生成を必要としない単純な検索を行う場合や、検索意図の推定結果が誤っていた場合に備えて、コンテキストメニュー(図 4(b)) 上に「通常検索(図 4(c))」と「その他の検索方法(図 4(d))」という二種類の補助的な項目を毎回提示するようにした。「通常検索」は、キーワード検索や辞書引きなどのキーワード拡張を必要としない検索方法を提示するものである。また「その他の検索方法」は検索意図知識に含まれる全ての検索方法を提示するものである。

同様に提案方式の本質的な機能ではないが、アプリケーションとしての総合的な使い易さを考慮して以下のような補助的な機能を実装した。これらは図 4(e) に示したようなウェブブラウザのツールバーから利用できる。説明のため拡大したツールバーの表示例を図 7 に示した。

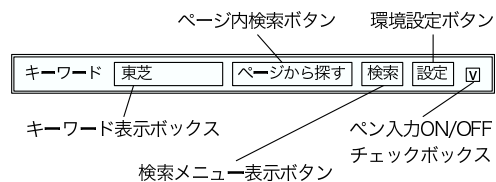


図 7 ツールバー

**キーワード表示ボックス** ユーザがアンダーラインを引いた語句を表示する。アンダーラインの範囲と語句の対応付けに曖昧さが生じることがある。指定した語句が正しく入力されていることを確認できるようにしている。

**ページ内検索ボタン** 表示している文書データの中から検索時にユーザがアンダーラインを引いた語句を探索する。長い文書中で参照したい部分が見つからず、画面スクロールなどの手間がかかってしまうことを回避できる。

**検索メニュー表示ボタン** 前回提示されたコンテキストメニューを再度表示する。同じ語句に関して複数の検索意図に基づく検索を順次実行する場合に、何度も同じ語句にアンダーラインを引く手間を省くことができる。

**環境設定ウィンドウ表示ボタン** 利用するウェブ検索サービスの URL などを設定する環境設定ウィンドウを表示する。

**ペン入力 ON/OFF チェックボックス** 提案手法が提供するペンインタフェース有効化/無効化する。

### 5.3 ユースケース

本アプリケーションの動作を具体的に説明するため、ユーザが業界動向調査のためにウェブ上のニュースサイトを閲覧している場合のユースケースについて述べる。

ユーザが、図 4 のニュースによる企業の株価への影響について知りたいと考え、文書中の「東芝」という語句にアンダーラインを引いた場合には、3 章で説明

したユーザインタラクションによりこの企業の株価を調べることができる。この時、アンダーラインを引いた語句がキーワード表示ボックスに表示され、ユーザは入力された語句を確認することができる。

株価を調べた後、ユーザがさらにこの企業のプレスリリースを検索したいと考えた場合には、株価の検索結果画面において文書中の「東芝」という語句にアンダーラインを引くか、ツールバー上の検索ボタンを押すことで、図 4(b) と同じコンテキストメニューが表示される。ユーザがここから「プレスリリース」を選択すると、システムは「東芝 プレスリリース | ニュースリリース」という検索クエリをウェブ検索サービスに送信し、ウェブ検索を実行する。

次に、図 4 の文書に戻り、ユーザが「RD-X6」という商品の評判を調べようとした場合を考える。ユーザは図 4 の場合と同様に「RD-X6」という語句にアンダーラインを引く。しかしシステムはこの語句を商品名であるとは解析できず、適切な検索方法を提示することができない。そこでユーザが「その他の検索方法(図 4(d))」を選択すると、サブメニューとして「企業名」「商品名」「人物名」などのキーワードの意味を指定するための選択肢が提示される。ここでユーザが「商品名」を選択すると、さらにサブメニューに「商品名」に関する検索方法を表す選択肢として「価格」「評判」「プレスリリース」「仕様」などが提示される。これらの選択肢からユーザが「評判」を選択すると、システムは「RD-X6 感想 | レビュー | クチコミ | 口コミ」という検索クエリをウェブ検索サービスに送信し、ウェブ検索を実行する。

このように本アプリケーションでは、ユーザが普段使用しているウェブブラウザ上で、キーワードにアンダーラインを引いてコンテキストメニューを選択するという簡単な操作だけで、企業の株価やプレスリリース、商品の評判や価格などの様々な情報を検索することが可能となる。

## 6. 評価

連鎖情報検索の有効性を応用に基づいて検証するた

め、試作したウェブ検索アプリケーションをユーザに自由に試用してもらい、試用中に実行された検索処理について検索精度の評価を実施した。具体的な応用としては、より多くのユーザに試用してもらうことを考え、筆者らにとってユーザの確保と評価結果の分析がしやすい IT 系業務を対象とした。

### 6.1 目的

連鎖情報検索ではペンインタフェースによる簡単なインタラクションによって情報検索を実行できる。しかし検索結果の上位に所望の情報が出力されなかった場合には、検索結果のリストをスクロールしたり、次の検索結果画面へ移動したりという手間が必要となってしまう。そこで提案方式が検索クエリの検討や入力を簡便化すると共に、検索結果の吟味や探索の手間を省力化できることを検証するため、連鎖情報検索が出力する検索結果の定量的な評価を実施した。評価尺度としては検索結果の上位 5 位以内にユーザの所望の情報が含まれる確率 (以後、5 位以内正答率と表す) を用いる。これは上位 5 位以内であればユーザが検索結果のリストをスクロールすることなく所望の情報に辿り着けるであろうとの仮説に基づいて定めた評価尺度である。

### 6.2 評価用検索意図知識の策定

IT 系業務に従事するユーザの具体的な検索意図に対応できるよう、以下のようなシミュレータによる予備実験を実施して、IT 系業務での利用を想定した検索意図知識を策定した。

まず図 1 の起点情報としてユーザが頻繁に閲覧する情報を知るため、IT 系業務に従事するユーザ 41 名からウェブブラウザのブックマークに登録しているウェブサイトの情報を収集した。次に有用な目的情報について知るため、インターネットに関する市販のガイドブックから一般に有用とされる 156 件のウェブサイトの情報を収集した。さらに、ユーザのブックマークからサンプリングした 50 個のウェブ文書について、文書中に含まれる様々な語句に関連する有用ウェブサイトの情報を手動で検索し、これらの語句から関連する情報への多くの写像を列挙した。こうして得られた写像を元のウェブ文書にリンクとして埋め込んだサンプルデータを作成し、サンプルデータ中の語句にアンダーラインが引かれると埋め込まれたリンクに従って関連する情報のリストを表示するシミュレータを作成した。

予備実験では評価実験とは異なる 10 名のユーザにシミュレータを試用してもらい、その操作ログから、ユーザがアクセスした起点情報と、目的情報への写像のサンプルを収集した。これはサンプルデータとして

意図推定知識		検索知識
キーワード意味種別	検索方法	追加キーワード
COMPANY	企業情報	企業情報   会社情報   企業概要   会社概要
COMPANY	株価	ファイナンス 株価
JAPAN_CITY	宿泊	ホテル トラベル
JAPAN_CITY	飲食店	グルメ
JAPAN_CITY	天気予報	天気予報
PERSON	プロフィール	プロフィール   自己紹介   profile
PERSON	ブログ	ブログ   blog

表 1 ウェブ検索アプリケーション用検索意図知識の例

用意した多くの写像の中で、対象分野のユーザにとって有用な写像のサンプルであると考えられる。これらの写像のサンプルを参照しながら 180 個の評価用検索意図知識を策定した。評価用検索意図知識の例を表 1 に示す。

### 6.3 環境

本研究では以下の環境で評価を実施した。

**実施期間** 2005 年 8 月～9 月。

**対象ユーザ** IT 系業務従事者 20 名 (研究職 7 名, 管理職 3 名, 開発職 5 名, 企画職 5 名)

**使用機器** タブレット PC ((株) 東芝 dynabook SS M200)

### 6.4 実験方法

評価は担当者が同席し以下の手順で実施した。

- (1) 実験担当者が操作方法のインストラクションを実施する
- (2) 被験者は日常業務で利用するサイトを起点として自由にウェブブラウジングを行う
- (3) 被験者は検索したいキーワードがあれば実験担当者に口頭で伝達する
- (4) 実験担当者がキーワードを用紙に記録する
- (5) 被験者がキーワードにアンダーラインを引きコンテキストメニューから適切な検索意図を選択して検索を実行する
- (6) 被験者が検索結果を評価し所望の情報の検索順位を実験担当者に口頭で伝達する
- (7) 実験担当者が正解の順位を用紙に記録する

1 名あたりの試用時間は 30 分程度を目安として 2～7 を繰り返す。評価中の音声及び操作ログは全て記録し、評価終了後に担当者が用紙と操作ログを照合して、検索の起点となったウェブサイトの URL、選択された検索意図、正解と判定されたウェブサイトの URL などをまとめて記録した。

### 6.5 結果

評価の結果を表 2 に示した。この表には、従来方式

	検索数	5 位以内正答数・率			
		連鎖情報検索		従来方式	
IT 業務向け検索	145	116	80.0%	87	60.0%
それ以外	73	42	57.5%	37	50.7%
計	218	158	72.5%	124	56.9%

表 2 5 位以内正答率

として既存のウェブ検索サービスを利用してキーワード検索を行った場合の 5 位以内正答率も示している。

ユーザによって実行された検索のうち、正解が存在しないなど評価対象外としたものを除いたのべ 218 回の検索において、提案方式の 5 位以内正答率は 72.5%であった。また検索意図知識の策定時に想定した IT 系業務向けの検索は 145 回実施され、この時の提案方式による 5 位以内正答率は 80.0%であり、従来方式の結果は 60.0%であった。IT 業務向けの検索のうち提案方式で 5 位以内に正解が検索され、従来方式で検索できなかった例は 41 例あり、逆のケースは 12 例であったことから、符号検定によればこれらの結果には有意な差があると言える。なお、提案方式と従来方式で同一の正解が検索されるとは限らないため、従来方式については担当者が目視で検索結果の再評価を行い、ユーザが指定した正解と同等と見なせる検索結果が上位に検索されている場合には上位の結果を正解とした。

## 6.6 考 察

従来方式の 5 位以内正答率 60.0%という結果は、5 回の検索のうち 2 回は検索結果が得られないか、画面をスクロールしなければ結果が探せないということを表しており、出力された検索結果から所望の情報を探す際の手間が情報検索において実際に問題となっていることを示していると考えられる。提案方式ではこのようなケースが 5 回中 1 回程度に削減されており、従来方式と比較して所望の情報への効率的なアクセスが可能となっていることを確認することができた。

一方、この結果は提案方式によっても上位 5 位までに所望の情報が検索できないケースが 20%程度残るということでもある。本方式は情報検索を実行する際の操作を簡便にする方法を提供するが、さらに所望の情報が上位に検索されなかった場合でも簡便な操作で検索結果から目的の情報を探し出すことを可能とする方法についても検討する必要があると考えられる。

## 7. ま と め

本論文では、ペンインタフェースを用いた簡単な操作で閲覧中の文書に関連する情報を的確に検索することができる連鎖情報検索のコンセプトを提案し、このコンセプトの応用であるウェブ検索アプリケーションの

試作と、IT 系業務を対象に実施したユーザ評価について述べた。評価の結果、対象業務に関する 80.0%の検索要求に対して画面のスクロールなどの手間をかけずに簡単なインタラクションのみで的確な検索結果を提示できることが示され、従来方式に対する提案手法の優位性が検証できた。これにより連鎖情報検索が、検索クエリの検討や入力の手間を省力化し、検索結果の吟味や探索を必要としない、ペンインタフェースによるインタラクティブな検索方法の実現に有効な方式であることがわかった。

本研究では IT 系業務向け検索意図知識の策定に 1 人月ほどのコストを要した。連鎖情報検索のコンセプトを他の幅広い用途に適用するためには検索意図知識構築の省力化が課題となる。

## 参 考 文 献

- 1) Rila Mandala, Tokunaga Takenobu, Tanaka Hozumi, The Use of WordNet in Information Retrieval, Proceedings of Coling-ACL '98, pp. 31-37, 1998.
- 2) Douglass R. Cutting, David R. Karger, Jan O. Pedersen, John W. Tukey, Scatter/Gather: A Cluster-based Approach to Browsing Large Document Collections, Proceedings of the Fifteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp.318-329, 1992.
- 3) 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, 3C3-03, 2005.
- 4) 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, 2000.
- 5) Bill N. Schilit, Gene Golovchinsky, Morgan N. Price, Beyond Paper: Supporting Active Reading with Free From Digital Ink Annotations, CHI98, pp. 249-256, 1998.
- 6) 石谷康人, 住田一男, 紙文書を対象としたビボット XML 文書に基づく XML 文書変換システム, 電子情報通信学会 技術研究報告, PRMU2003-216, pp.7-12, 2004.
- 7) 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, 情報処理学会 研究報告, NL-161-3, 2004.
- 8) 布目光生, 石谷康人, 表層表現抽出と文書構造解析に基づく XML 文書変換システム, 情報処理学会 研究報告, 2004-DD-46, pp.1-8, 2004.
- 9) (株)東芝, [http://pf.toshiba-sol.co.jp/prod/km2/index\\_j.htm](http://pf.toshiba-sol.co.jp/prod/km2/index_j.htm)