

連鎖検索インタフェース“ササッとサーチ”TM

"SASATTO Search" Human Interface Technology for Information Retrieval

鈴木 優 石谷 康人 坂本 圭

■ SUZUKI Masaru ■ ISHITANI Yasuto ■ SAKAMOTO Kei

東芝は、Webブラウザで閲覧している文書において、マウスやペンなどでキーワードをなぞり、メニューを選ぶだけで関連情報を効率良く的確に検索できる、連鎖検索インタフェース“ササッとサーチ”TMを開発した。ササッとサーチTMでは、意味クラス解析によりキーワードの意味を決定した後、検索意図判定によりキーワードの意味に応じた検索方法を列挙することで、使い勝手の良いインタフェースを実現している。一般ユーザーを対象とした評価実験を行ったところ、高い検索精度とユーザー満足度を得ることができた。

To realize easy and accurate information retrieval, Toshiba has developed a pen/mouse-based human interface called "SASATTO Search" for chaining searches of the Web search system. If a user selects a keyword from a document that he/she is reading, documents related to the keyword can be obtained simply by selecting the desired search method from a display context menu. It is easy to accomplish such a search because the meaning of the keyword is determined by semantic pattern analysis and the menu contains search methods suitable for the meaning. In an experiment involving 15 users, it was confirmed that the proposed interface is more effective in terms of easier and more accurate information retrieval compared with the conventional method.

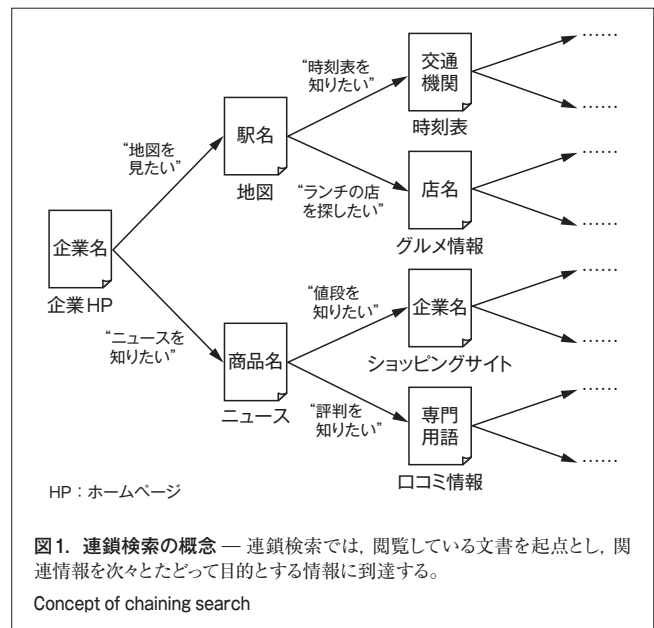
1 まえがき

コンピュータネットワークの発達や記録メディアの大容量化、WWW (World Wide Web) の普及、及びWebコンテンツの充実などに伴い、人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、オフィスや家庭、あるいは街中でも、パソコン (PC) や携帯端末を用いて頻繁に情報検索を行うようになった。

このような情報検索では、関連する情報を芋づる式にたどっていく“連鎖検索”が頻繁に実施されている(図1)。このときユーザーは、それぞれの検索において、以下の手順で必要な情報を入手している。

- (1) 検索サイトへのアクセス
- (2) 検索キーワードの検討及び入力
- (3) 検索サイト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及び閲覧

しかし、WWWでの情報検索では、簡単な検索キーワードを用いるとたくさんの検索結果が出力されてしまう。それを避けるため、手順(2)で検索キーワード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(条件式)で定義して検索結果を絞り込む必要があった。それでも目的とする情報が検索結果の上位に現れない場合には、手順(3)と(4)で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していた。このため、前述した連鎖検索の手順が非常に煩わしいものとなっていた。



東芝は、情報検索でのこのような煩わしさを軽減することを目的として、連鎖検索のための新しいインタフェース技術“ササッとサーチ”TMを開発した⁽¹⁾。ササッとサーチTMは、Webブラウザに表示された文書で気になるキーワードがあれば、それをマウスやペンなどのポインティングデバイスでなぞると、関連情報を的確に検索できるインタフェースである。このインタフェースでは、ユーザーがキーワードを選択したとき、キーワードの意味に応じて検索方法の候補をユーザーに提示すること

が大きな特長となっている。これにより、ユーザーはその中から適切な検索方法を選ぶだけで、目的とする情報に的確にたどりつけるようになっている。

ここでは、前述した特長を持つ連鎖検索インタフェース「ササッとサーチ_{TM}」について、インタフェースの概要、システム構成、及び評価とその結果について順に述べる。

2 ササッとサーチ_{TM}の概要

ササッとサーチ_{TM}では、メニュー形式のユーザーインタフェースを採用している(図2)。このインタフェースでは、ユーザーがWebブラウザ上でキーワードを選択すると、キーワード付近にメニューが提示されるようになっており、メニューと対話することでキーワードに関連する情報を検索できる。インタフェースの概要を図2を用いて説明する。

ササッとサーチ_{TM}では、ユーザーがWebブラウザ上でマウスによりまずキーワードを選択する(①)。その際、マウスを右クリックすればコンテキストメニューが表示されるようになっており、連鎖検索を実施する場合には、そこから「ササッとサーチで検索」を選ぶことになる(②)。すると、選択したキーワードの意味を示す「検索メニュー」が新たに表示される(③)。このとき、キーワードの意味として複数の候補が提示されている場合には、ユーザーはその中から適切なものを選ぶ必要がある。ユーザーによってキーワードの意味が選択されれば、更に、検索方法で構成される新しいメニューが表示される(④)。ユーザーは、このメニューの中から適切な検索方法を選ぶことで、目的とする情報にアクセスできるようになる。

このようにササッとサーチ_{TM}では、ユーザーが「キーワードをなぞる」、次に「メニューを選択する」という2段階の操作を行うだけで、目的とする情報に的確にアクセスできるようになっている。この場合、検索方法ごとに補助キーワードがひも付けられており、ユーザーが検索方法を選択すると、ササッと

サーチ_{TM}は、検索方法に応じた補助キーワードを先に選択されているキーワードに追加して検索サイトを呼び出すことにより、効果的な絞込み検索を実現している(3.2節参照)。

ササッとサーチ_{TM}では、ユーザーが選択できる検索方法として表1に示す25種類を用意している。ユーザーがキーワードを選択した際、25種類すべての検索方法が提示されると、一つの検索方法を選ぶ作業が煩雑になってしまうと考えられる。そこでササッとサーチ_{TM}では、検索方法をキーワードの意味でクラス分けし、意味クラスの選択、次に検索方法の選択という階層的なメニュー選択を可能にすることで、常に少数の候補から必要なものを選択できるようにした。意味クラスとしては、「企業・組織・施設」、「イベント」、「人名」、「地名」、「駅・空港・交通機関」、及び「その他(商品、専門用語など)」の6種類を定義している。そして、それぞれの意味クラスに8~13個の検索方法を割り当て、そのうち利用頻度の高い5個の検索方法をメニュー上で優先的に表示している。

表1. 検索方法の一覧

List of search methods

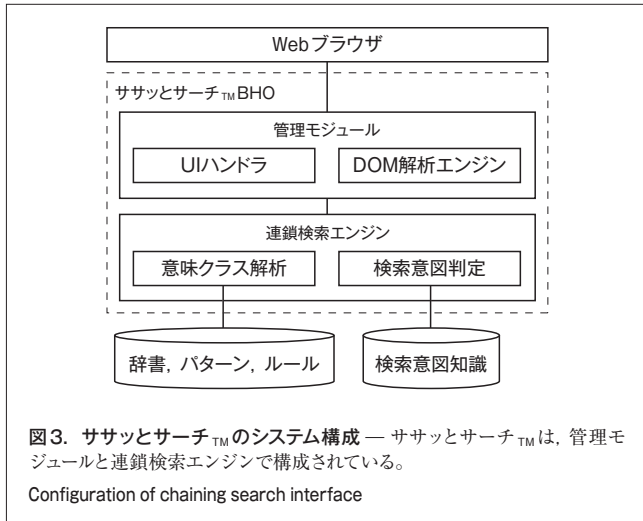
選択可能な検索方法	
(1) イベントスケジュールを調べる	(14) 地図、アクセス情報を探す
(2) 飲食店を探す	(15) ツアーを探す
(3) オークションで探す	(16) 天気予報を調べる
(4) 価格、購入情報を調べる	(17) ニュースを探す
(5) 求人情報を探す	(18) 評判を調べる
(6) 公式サイト、ポータルサイトを探す	(19) ブログを探す
(7) 時刻表を調べる	(20) プロフィールを調べる
(8) 施設内の案内図を探す	(21) ホテルを探す
(9) 写真を探す	(22) 役所のサイトを探す
(10) 書籍、文献を調べる	(23) レシピを調べる
(11) 説明、概要、沿革を調べる	(24) 連絡先(電話番号、住所)を調べる
(12) タウン情報を調べる	(25) 路線図を探す
(13) チケットを探す	

図2. ササッとサーチ_{TM}による情報アクセス — ササッとサーチ_{TM}では、Webブラウザにおいてマウスやペンでキーワードを選択しメニューを選ぶだけで、関連情報を効率良く的確に検索できる。

Example of chaining search interface

3 システム構成

ササットサーチ™は、Webブラウザの機能を拡張するBHO (Browser Helper Object) として実現されており、管理モジュールと連鎖検索エンジンで構成される (図3)。



管理モジュールは、Webブラウザとのデータの入出力を行うことを主な目的とし、ポインティングデバイスの操作のイベントを処理するUI (User Interface) ハンドラと、Webブラウザに表示された文書を読み込み、ユーザーが選択したキーワードを含む部分領域を特定するDOM (Document Object Model) 解析エンジンで構成される。一方、連鎖検索エンジンは、連鎖検索を可能とするための主要な構成要素であり、意味クラス解析と検索意図判定で構成される。意味クラス解析は、ユーザーがポインティングデバイスにより選択したキーワードの意味を決定するものである。また、検索意図判定は、意味クラス解析で決定されたキーワードの意味に基づき、検索方法を列挙及びランキングしてメニュー表示するものである。

以下に、連鎖検索エンジンを構成する意味クラス解析と検索意図判定について述べる。

3.1 意味クラス解析

意味クラス解析では、形態素解析やパターン処理、及びヒューリスティック判定を相補的に組み合わせて、ユーザーが選択したキーワードの意味をロバストに推定している。このとき、未知語にも対応するため、前後の文脈を用いて選択されたキーワードの意味クラスを決定している。“東芝G40/97Dは最新のノートPCです。”という文章を例に、ユーザーが“G40/97D”をキーワードとして選択したときの意味クラス解析の動作を以下に説明する。

まず、“東芝G40/97Dは最新のノートPCです。”に対して、辞書を参照して形態素解析を実施し、単語を切り出すとともに品詞を付与する。この辞書には典型的な企業名や人名、地名

などを含む30万語の語彙 (ごい) が登録されているので、形態素解析の時点で“東芝:名詞 (企業)”という処理結果を得ることができる。この後、次のように、典型的な単語を6種類の意味クラスのいずれかに分類する。

東芝:名詞 (企業) → 東芝:企業・組織・施設

次に、形態素解析で得られた結果に対し、パターン処理を適用して固有表現を抽出する。例えば、“G40/97D”には“企業+「」→商品”というパターンが適用され、“G40/97D”は“商品”の固有表現であると判定される。ササットサーチ™では500種類のパターンが定義されており、114種類の固有表現の抽出が可能である。そして、次のように、固有表現を6種類の意味クラスのいずれかに分類する。

G40/97D:商品 → G40/97D:その他 (商品, 専門用語など)

形態素解析やパターン処理で意味クラスを判定できない単語については、ヒューリスティック判定を適用して最終的な意味クラスを決める。ヒューリスティック判定とは、キーワードを構成する文字列の並び方に基づいてキーワードの意味クラスを判定する方法で、例えば“Masaru Suzuki”は、先頭だけが大文字の英字で構成される単語が2語以上連続していることから“人名”と判定される。ササットサーチ™では、42個のルールを用いてヒューリスティック判定を実施している。

3.2 検索意図判定

検索意図判定は、意味クラス解析で判定されたキーワードの意味を手がかりに、あらかじめ定義されている複数の検索意図知識を探索して、検索方法を列挙するとともに検索メニューを構成する。検索意図知識は、図4に示すように“キーワード意味クラス”、“検索方法”、及び“補助キーワード”で構成され、XML (eXtensible Markup Language) 形式で記述されている。

検索意図判定では、まず、意味クラス解析で判定され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る場合、その検索意図知識を検索方法の候補として選

```
< 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 >
  求人情報を探す
< / 検索方法 >
< 補助キーワード >
  "採用情報" | "求人情報" | "新卒採用" | "中途採用" |
  "新卒サイト" | "Jobs at" | "Jobs offered" | "Job title"
< / 補助キーワード >
< / 検索意図知識 >
```

図4. 検索意図知識の概要 — 検索意図知識は、キーワード意味クラス、検索方法、及び補助キーワードで構成され、XMLで記述されている。
Knowledge of user's search intention

出する。そして、選出された検索意図知識をユーザーの検索方法の利用頻度に基づいてランキングすることにより、検索メニューの構成要素を形成する。利用頻度に応じたランキングでは、検索方法を利用頻度の長期学習によりいったんランキングした後、更に、利用頻度の短期学習に基づき再ランキングするようになっている。

このようにしてランキングされた検索方法の候補は、管理モジュールに送られ検索メニューとして表示されることになる。管理モジュールでは、ユーザーが検索メニューから検索方法を選択すると、その検索方法に対応する検索意図知識から補助キーワードを読み出すとともに、ユーザーが指定したキーワードに追加して検索サイトに検索要求を送信する。なお、ササッとサーチTMでは、2,700回以上実施した検索実験の結果に基づき補助キーワードを設定している。

4 評価

ササッとサーチTMの有効性を確認するため、意味クラス解析の精度とWeb検索の精度、及びユーザー満足度をそれぞれ評価した。

4.1 意味クラス解析の精度

ササッとサーチTMを用いて実際のWebコンテンツからキーワードを無作為に選択し、表示される検索メニューの内容を確認することで意味クラス解析の精度を評価した。この評価でキーワード選択を100回以上試行した結果、1位再現率（検索メニューの最上位に正しい意味クラスが表示される確率）88%、再現率（検索メニューに正しい意味クラスが表示される確率）99%、適合率（検索メニューに表示されたうち正しい意味クラスの割合）57%、という意味クラス解析精度が得られた。この評価結果から、意味クラス解析で用いる辞書やパターンが適切に定義されていることを確認した。

4.2 Web検索の精度及びユーザー満足度

一般モニター15名が、ササッとサーチTMを用いて検索実験を行うことにより、検索精度と検索結果に対する満足度をそれぞれ評価した。この評価では、検索の起点となる文書と検索課題の両方を実験担当者があらかじめ設定した。検索課題として、“保土ヶ谷公会堂の地図を探す”、“新富町駅の時刻表を

調べる”、“洗濯乾燥機のレビューを見つける”、“DVDレコーダの価格を調べる”、“ノートPCの仕様を調べる”という五つのケースを用いた。その結果、1位正答率（検索結果の1位に目的とする情報が検索される確率）77.5%、3位以内正答率（検索結果の上位3位以内に目的とする情報が検索される確率）91.3%、という検索精度が得られた。また、被験者による、ササッとサーチTM及び従来方式に対する満足度の評価と比較を行った（表2）。その結果、“非常に満足”又は“満足”と回答した被験者が従来方式よりもササッとサーチTMのほうが多く、ササッとサーチTMの有効性を確認することができた。

5 あとがき

Webブラウザでマウスやペンによりキーワードを選択し、メニューを選ぶだけで関連情報を効率良く的確に検索できる、連鎖検索インタフェース ササッとサーチTMについて述べた。ササッとサーチTMでは、当社の伝統的な自然言語処理技術を情報検索のインタフェースに取り込むことにより、簡単な操作にもかかわらず高精度な検索を実現するという、相反する課題を両立させることができた。

現在、ササッとサーチTMは、“インターネットカンタン検索”のお役立ちソフトとして、当社製PCの2007年度秋冬モデルに搭載されている。今後は、デスクトップ検索に対応するなど機能を拡張していくとともに、Webブラウザを利用する他のデジタルプロダクトへの搭載を目指していく。

文献

- (1) 鈴木 優, ほか. “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”. インタラクシオン2006シンポジウム予稿集. 情報処理学会編. 東京, 2006-03. 情報処理学会. 2006. p.101-108.

表2. ユーザー満足度の計測結果

Results of users' evaluations

満足度	従来方式	ササッとサーチ TM
非常に満足	0	3
満足	8	8
やや満足	6	4
不満	1	0
計	15	15



鈴木 優 SUZUKI Masaru

研究開発センター 知識メディアラボラトリー研究主務。
情報検索インタフェースの研究・開発に従事。
情報処理学会, 人工知能学会会員。
Knowledge Media Lab.



石谷 康人 ISHITANI Yasuto, D. Eng.

東芝ソリューション(株) IT技術研究所 研究開発部主任
研究員, 工博。ビジネスインテリジェンス技術の研究・開発
に従事。IEEE, 電子情報通信学会, 情報処理学会会員。
Toshiba Solutions Corp.



坂本 圭 SAKAMOTO Kei

PC&ネットワーク社 PC開発センター PCソフトウェア設計
第一部主務。PC向けソフトウェアの設計業務に従事。
PC Development Center