

インターネット多角的検索システムOTROS — 全体の概要と構成 —

3 L-1

山田洋志 松田勝志 竹元義美 赤 峯 享 福島俊一

NEC ヒューマンメディア研究所

1 はじめに

インターネット、イントラネットの普及に伴い、大量の文書の中から必要な情報を見つけ出すための検索技術の必要性が高まっている。インターネット上では多くの検索サイトが検索サービスを行っており、また、文書管理ソフトウェアでも検索機能が必須になっている。

現在の検索機能は、指定した単語をテキスト中から探し出す単語検索を基本としている。単語による検索では、検索結果が大量になった場合の絞り込みに限界がある、適切な単語の指定が難しいなどの問題がある。さらには、ユーザの求める情報が単語だけでは表現できない場合も多い。これらの問題を解決するために、類似文書による検索や単語以外の情報（URLの種類やイメージの有無など）による絞り込み機能を提供しているシステムがある[6, 7]。

筆者らは、上記のようなアプローチをさらに発展させ、多彩な条件指定の方法を提供することでユーザの多様な検索要求に応える、多角的検索システムOTROSを開発している。OTROSは、単語あるいはテキストによるテキスト内容の検索に加え、内容とは別の視点からの検索支援機能を備えている。

本稿では、OTROSがめざす多角的な検索機能について述べ、システムの機能を紹介する。

2 多角的な検索機能

OTROSでは、ユーザの多様な検索要求に応えるために、以下のような検索機能の実現をめざしている。

2.1 テキスト内容による検索機能

- 単語検索: 検索したいテーマについて広範に情報を集められる。一方で、テキストの主題も、補足的な記述も検索されるため、余計なテキストが検索される。逆に、同じ内容であっても同義語の存在など表現が異なるために検索に漏れが生じる場合もある。むだや漏れの軽減のため、後述する検索語展開や類似テキスト検索を利用する。

- 係り受け検索: より高度な検索としては、単語同士の掛かり受け関係や5W1Hを指定した検索方法がある[4]。これによって、特定の事実や出来事について知りたい場合に、目的のテキストだけを検索できる。精度の高い検索を実現するためには、構文解析・文章理解などの言語処理技術が必要である。
- 数値検索: 数値を検索する場合は、テキストに現れた文字列を検索するのではなく、値として抽出することで、範囲指定や以上・以下といった検索条件に対応でき、より柔軟に利用できる。単語の検索と組み合わせることで検索結果を絞り込める。さらに、前記の係り受け検索と組み合わせることで、何を表す数値なのか（「自動車の値段」、「〇〇県の人口」など）まで判断することで、さらに強力な機能となる。

2.2 テキストの形式による検索機能

内容とは別の視点からの検索、分類機能を利用することで、直接必要な結果を絞り込む。

- 文書タイプ検索: テキストは目的や形式でいくつかのタイプ、たとえば「カタログ」、「研究室紹介」、「求人情報」などに分けられる。タイプを指定することで、目的に合わせて検索結果を絞り込める。たとえば、特定の製品の購入のためにはメーカーの製品紹介やカタログが役に立つ。研究論文の参考には、大学や研究所を重点的に探してみる。ことばの意味が知りたいときは用語集や辞書を探せばよい。
- メディア検索: 画像や音といった文字以外の情報を利用した検索は、マルチメディア化の進む中でますます重要になってきている。各情報の表題を利用したり、実際にそれぞれの内容を検索するなど幅広いレベルの利用が考えられる。

2.3 ユーザ支援機能

検索機能が多様化、高度化しても、操作が複雑になってはユーザに使ってもらえない。機能が増える

Internet Diversity Retrieval System OTROS - Search Functions and Structure -

Hiroshi Yamada, Katsushi Matsuda, Yoshikazu Takemoto,
Susumu Akamine and Toshikazu Fukushima
Human Media Research Labs., NEC Corp.

多角的検索システムOTROS

View Go Bookmarks Options Directory Window Help

OTROS

Internet Diversity Retrieval
OTROS means Variable
can lookup desired Web
variable method. This sys
Human Media Research L

テキストを指定して検索→類似テキスト検索

基本検索

ワゴン

単語の組み合わせ: ☐ すべてを含む(AND検索) ☐ いずれかを含む(OR検索)

関連語の追加: ☐ しない ☐ する

拡張検索

ページの種類の指定してしぼりこむ 適合度 30 以上を選択

☐ 指定しない ☐ 商品カタログ ☐ リンク集 ☐ 研究室 ☐ 求人情報 ☐ F.A.Q.

数値の条件でしぼりこむ

円 の範囲

データベースを利用してキーワードを追加

自動車 パソコン

図 1: 検索画面

ほど簡単に利用するための工夫が必要になる。

- 検索語展開: 同義語辞書で検索語を増やすことで検索漏れを減らせる。従来、展開の対象は普通名詞で、製品名などは対象外であったが、特定の条件にかなう製品や、類似した製品の名前に展開して検索できるように、固有名詞を展開する機能を備える。購入や障害情報の検索に有効である。
- 単語入力支援: 検索単語を組み合わせることで検索結果を減らしたり、適当な順位付けを行えるが、適切に単語の組み合わせを選ぶのは容易でない。そこで、サンプルとなるテキストを入力して検索する類似テキスト検索や、検索結果のテキストを使って再検索するフィードバック検索、検索結果から重要語を取り出して利用する関連語検索(リファイン)によって、ユーザの負担を軽減する。

3 試作システム

前節にあげた機能の一部を実現したOTROSの試作システムを開発した。本システムはNEC製のEWS4800上に構築し、単語の検索にはNECで開発した文字ベースの検索エンジンRetrievalExpress[5]を使用した。

OTROSの検索画面と検索結果の例を図1,2に示す。

開発にあたっては、利用場面が多く、かつ、実用的な精度の得られる機能を優先して選択した。

新たな検索機能として、数値検索[1]、タイプ指定検索[2]、データベースを利用した製品名展開[3]を開

ワゴンの検索結果

View Go Bookmarks Options Directory Window Help

OTROS

Internet Diversity Retrieval
OTROS means Variable
can lookup desired Web
variable method. This sys
Human Media Research L

検索結果を利用して再検索

☐ 絞り込み検索 ☐ 新規検索

フィードバック検索

関連語検索

検索結果の関連語

- ☐ 仙台
- ☐ 用
- ☐ 情報
- ☐ 宿泊
- ☐ 設置
- ☐ セダン
- ☐ 車

検索結果

検索語 [ワゴン]

ページタイプ [商品カタログ] (適合度
検索数 5/5)

1. ☐ caarygracia.html
<http://autoshow.path.ne.jp/>
グラシア(カムリ)トヨタ自動車
2. ☐ passatvariant.html
<http://autoshow.path.ne.jp/>
パサートワゴン(フォルクスワー
3. ☐ mu.html
<http://autoshow.path.ne.jp/>

図 2: 検索結果

発・実装した。また、従来の単語検索に基づく機能として、同義語展開、類似テキスト検索、フィードバック検索、関連語提示機能、抄録表示機能も備えている。各テキストの数値情報と文書タイプ、重要語は、登録時にインデックスに格納している。

4 おわりに

ユーザの多様な検索要求に応えるための高度で使いやすい検索システムをめざし、多角的検索システムOTROSを開発した。

本システムは、従来とは異なった検索要求に応えるため、数値検索、文書タイプ検索、データベースによる製品名展開の各機能を実現し、単語検索や類似検索とともに提供している。

今後各検索機能の精度向上と未実装機能の実現を進めるとともに、複数の機能を組み合わせて使うためのインタフェースを検討する。

参考文献

- [1] 山田, 福島, “インターネット多角的検索システムOTROS-数値情報の抽出と検索-”, 情処57回大会, 3L-02
- [2] 竹元ほか, “インターネット多角的検索システムOTROS-データベースを利用したクエリ展開方式の提案-”, 情処57回大会, 3L-03
- [3] 松田, 福島, “インターネット多角的検索システムOTROS-構造的特徴量によるタイプ分類と検索-”, 情処57回大会, 3L-04
- [4] 岸本ほか, “テキストの構造化に基づく検索システム”, 情処論文誌, 1994
- [5] 赤峯, 福島, “高速全文検索のためのフレキシブル文字列インバージョン法”, 情処ADBS, 1996
- [6] Infoseek, <http://japan.infoseek.com/>
- [7] goo, <http://www.goo.ne.jp/>