

企業ポータルにおける関連ディレクトリ検索

4Y-01

高山 泰博 永井 明人 増塩 智宏 鈴木 克志

三菱電機株式会社 情報技術総合研究所

1. はじめに

インターネット上の企業ポータルは、その企業の宣伝や顧客誘導のために重要な役割を担っている。顧客が企業ポータル内の情報を探し出すためには検索機能が必須である。従来の企業ポータルにおける検索は単純な全文検索機能がほとんどであり、顧客が所望する情報を検索結果から探し出すのを十分に支援できていない。本稿では、入力文に関連する、企業ポータル内のディレクトリを、関連文書検索機能によって順位付け、全文検索結果を各ディレクトリ毎に分類することにより、所望の検索結果を得やすくする関連ディレクトリ検索方式を提案する。

2. 企業ポータルにおける検索機能の課題

ポータルには、インターネット全体の雑多な情報を検索する検索ポータルと、各企業が宣伝や顧客誘導のために自社ビジネスの製品やサービスに関連する情報を集約したホームページを作成して各種サービスを提供する企業ポータルがある。

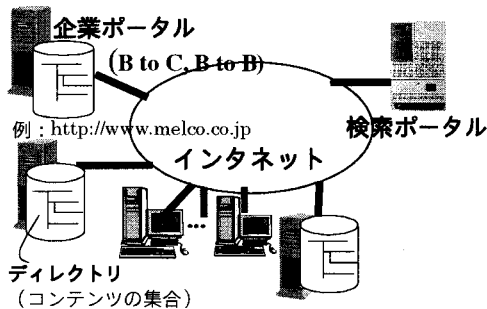


図1 インターネット上の企業ポータル

従来の企業ポータルにおける検索機能は、キーワード入力による全文検索が中心であり、ポータル内の情報が増加してくると、一覧表示された検索結果から所望の情報を取り出すのが困難になってくる。また、検索対象の文書が持つ時間情報を検索

結果の選択に充分活用できていない。

ベクトル空間モデルによる文書毎のベクトル索引を用いる検索システム⁽²⁾は、全文検索に比して索引のサイズ・更新時の速度が大きくなることから、企業の特定部門内の情報を検索するのに用いられることが多い。企業ポータルにおいて、自然言語検索機能を提供するもの⁽³⁾があるが、そのポータル全体に渡る情報を検索するのではなく、サポート情報など一部の情報に限られている。

3. 企業ポータルに必要な検索機能

この章では、前節で述べた企業ポータルの従来の検索機能での課題に対処するため、企業ポータルの検索に必要とされる機能を検討する。

(1) 検索結果の分類表示

企業ポータル内の文書は、製品やサービス毎にディレクトリと呼ばれるグループに分けられていることが多い。顧客が検索したときに、入力に関連する複数のディレクトリが表示されれば、関連する製品やサービスが容易に把握できるようになる。

例えば、個人向け、法人向け両方の製品に関する情報を提供している企業ポータルにおいて、「テレビ」という入力に対して、「カラーテレビ」および「映像・画像システム」等が分類して表示されれば、個人ユーザであれば「カラーテレビ」の欄を、法人ユーザであれば「映像・画像システム」の欄を選択すれば良いことになる。

従来の全文検索では、これらの文書が混在して一覧表示されるが、検索した文書をディレクトリ毎に分けて表示すれば、従来の全文検索で埋もれたり混在したりしていた文書を探しやすくなる。

(2) 自然言語入力と同義語展開

従来の企業ポータル検索では、キーワード入力が多であるが、検索条件の入力には、キーワードに限定されない自然言語による入力が望まれる。

企業ポータルと検索ポータルの大きな違いは、コンテンツが管理されているか否かである。企業ポータルでは、この管理されたコンテンツからあらかじめ

Relevant Directory Retrieval in Corporate Portals
Yasuhiro TAKAYAMA, Akito NAGAI, Tomohiro MASUSHIO,
Katsushi SUZUKI
Information Technology R&D Center, Mitsubishi Electric Corp.
5-1-1 Ofuna, Kamakura, Kanagawa 247-8501, JAPAN

め生成した同義語辞書を利用し、入力文を同義語展開することで、検索漏れを少なくすることができる。

(3) 新着情報の検索

企業ポータルを何度も訪れる顧客に対しては、常に新しい情報を提供することが重要である。そのためには、ポータル内で最近更新された文書(新着情報)をまとめて検索する機能が有効である。

また、検索要求に「最新」や「新しい」等の新着情報を想起させる言葉が含まれる際には、この新着情報のグループに対する検索結果を自動的に表示できることが望まれる。

4. 関連ディレクトリ検索方式

前章の検討を踏まえ、この章では、関連ディレクトリ検索の機能を提案する。企業ポータルでは、ユーザに切れ目なくサービスを提供するために、索引の更新時間を最低限にすることが重要である。そこで、個別の文書と比較して追加・更新までの期間が長いディレクトリに対して、そのディレクトリに属する代表文書を基にしたベクトル索引⁽⁴⁾を用意する。高々数百個のディレクトリに対する索引の作成は高速であり、また更新の頻度も少ない。

この時、個別の文書に対しては全文検索の索引を別に作成しておくことで、特定の製品型番など、ディレクトリを作成した際の代表文書に含まれていない語を含む文書は、全文検索機能により検索できる。我々の提案する企業ポータル検索システムの構成を図2に示す。

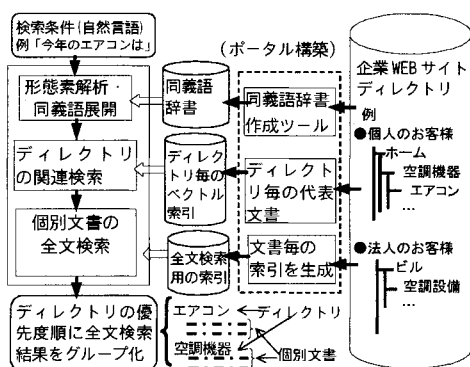


図2 企業ポータル検索システムの構成

図2の右側は関連ディレクトリ検索に関わるポータル構築処理を示す。企業ポータルの管理されたコンテンツから、ベクトル索引を基に同義語辞書

を作成する。また、ディレクトリ毎の代表文書によりディレクトリ毎のベクトル索引を生成しておく。文書の登録・更新毎に全文検索索引を生成する。

図2の左側は検索時の処理を示す。自然言語による検索条件の形態素解析・同義語展開を行ないディレクトリを関連文書検索し、関連ディレクトリを優先度付きで求める。全ての文書ではなく、ディレクトリ毎のベクトル索引を検索するので、高速な検索が可能である。次に、解析した検索条件を用い、個別文書に対する全文検索を行なう。この全文検索の結果をディレクトリ毎に分類して表示する。

5. 関連ディレクトリ検索の実験と評価

今回、3章および4章で述べた機能を備えた検索システムを試作した。この試作システムに対して、約2万HTML文書、約200個のディレクトリからなる文書データに対して、220文の検索条件(100個の単一キーワード、120個の自然言語文)を用いて検索実験・評価を行なった。

評価は、被験者の主観により検索されたディレクトリが、(1)期待した結果である、(2)期待した内容ではないが検索条件から許容できる、(3)妥当ではない、の3段階で実施した。その結果、(1)および(2)となる割合が、8割以上となる評価結果を得た。

試作システムでは、関連ディレクトリ検索の評価が、(3)の妥当ではない、となる場合にも、通常の全文検索結果を表示しており、従来と同等な個別文書の検索結果が得られることから、この結果は本アプローチの有効性を示していると考えられる。

6. まとめ

インターネットの企業ポータルにおいて、顧客が所望する企業ポータル内の情報を探し出すための検索機能として従来の全文検索に加え、入力文に関連するディレクトリを関連文書検索機能によって順位付け、全文検索結果を各ディレクトリ毎に分類することにより、所望の検索結果を得やすくする関連ディレクトリ検索方式を提案した。

参考文献

- (1)藤井洋一,他;"段落内共起情報を利用した文書自動分類方式",情報処理学会論文誌, Vol.42, No.3(2001).
- (2)野村直之;"ConceptBaseの言語処理と新しいソリューション",情報処理学会研究会資料 NL129-1(1999).
- (3)日本マイクロソフト:話し言葉によるサポート技術情報検索 <http://www.microsoft.com/japan/enable/nlsearch/>