

インターネット・タウンページの構築 (1) ～概要～

7N-7

島 健一 高橋 克巳 三浦 信幸

{kshima, takahasi, miura}@slab.ntt.co.jp

NTT ソフトウェア研究所

1 はじめに

NTTの発行する職業別電話帳タウンページをもとにインターネットWWWとオブジェクト指向データベースを結合した構成を持つインターネット・タウンページシステム¹を構築した。本論文では、その概要について述べる。本システムの構築にあたっては、電話番号、住所、名前の他、地図、広告などマルチメディア情報をオブジェクト指向データベースを用いて有機的に結合している。これにより多彩な検索、ナビゲーションを可能とするWWWサーバの構築が可能となった。また、本システムは日本語の曖昧検索の研究成果にもとづく人名/会社名や地名などの漢字表記のゆれ等を吸収する機能を持つ。さらに、過去の検索結果、個人の興味等を反映した検索等に関わるユーザモデルの研究成果にもとづくカスタマサービス機能を持つ。

2 基本的考え方

昨年度、英語版電話帳TOWNPAGE²のインターネットを介した検索を行うための以下の各種技術を確立した³。

1. オブジェクト指向データベースと連携したWWWサーバの構築法
2. エリアマップなどマルチメディア情報とのハイパーリンクの統合管理法
3. ローマ字表記のゆれに対処する曖昧検索技術
4. 事例ベース検索技術にもとづく個人別電話メモ

東京都23区内100万件掲載のタウンページ電話帳をインターネット時代に適したWWWサーバとして構築するための課題について検討した。インターネットタウンページサーバとして構築するためには、英語版TOWNPAGEの成果を生かしつつ、以下の技術を実現する必要がある。

- 英語版TOWNPAGEのデータ件数と比較すると、日本語版タウンページでは、大量データを効率良く蓄積する技術、多様な検索技術がデータベース構成上必要となる。

- 本格的な地図と電話帳データとの連動を検討する必要がある。そのため掲載の所在地を示す座標値を自動的に求める機能が必要である。
- 日本語の検索は、人名、地名などにさまざまな曖昧性がある。これに対処できる曖昧検索技術を確立する必要がある。
- ユーザフレンドリーな検索を実現するためには、個人毎のユーザモデルに基づく事例ベースナビゲート技術を実現する必要がある。

3 インターネットタウンページの特徴

1. 東京都区内100万件的タウンページ電話帳をオブジェクト指向データベース⁴)に格納し、多彩な検索をできる技術を確立した。

東京都区内100万件的タウンページ電話帳、および本格的な地図情報、および広告などさまざまな形式のマルチメディア情報とともにインターネットを介して電話帳データを検索できる技術を確立した。従来、国内での大規模システムの構築にあたってオブジェクト指向データベースを適用した例はなく、電話番号、住所、掲載名、地図情報などをハイパーリンクで結合するためオブジェクト指向技術（概念階層、部分全体関係など）を活用した。

図1に、インターネットタウンページシステムの構成を示す。

2. 住所の検索とあわせて本格的な地図と電話帳データとの連動技術を開発した。

電話帳掲載結果を地図上に表示するため、検索した電話帳掲載情報に基づき、実際の所在地、最寄り駅などを地図上に表示する技術を開発した。従来のカーナビのような地図検索では、検索が一方向であるが、今回は地図上でユーザが指定したポイント周辺の電話帳掲載情報を電話帳サーバから検索する技術も開発した。所在地確認、道案内、駅の回りの検索などに利用している。

3. 日本語の曖昧検索におけるマッチング処理技術を確立した¹)。

日本語の固有名詞のゆれや誤りのパターンに基づいた正規化を行ない、高速に類似した文字列を検索できる技術を開発した。英語の場合は、語単位で分離されているため日本語に比べれば簡単であるが、日本語の場合は、分かち書きが必要であった。また、従来濁音のみであった日本語での曖昧検索技術を漢

Internet TOWNPAGE (1)

Ken-ichi SHIMA, Katsumi TAKAHASHI, Nobuyuki MIURA
NTT Software Laboratories<http://townpage.isp.ntt.co.jp/><http://www.pearl.org/jtd/>

字、読みの音韻変化、外来語、異体字にまで拡張する必要があり、新たに研究開発した。

図2に、曖昧検索の例、および地図との連動の例を示す。

4. 個別ユーザモデル適用によるユーザインタフェース技術を確立した²⁾。

高品質なサーバ構築のための一手法として、ユーザの性質、行動分析を把握し、各ユーザのユーザモデルを構築し、WWWサーバにFeed Backする技術を確立した。今後、ユーザモデルの構築の一手法として、WWWサーバのアクセス履歴から個人のアクセス履歴を抽出し、それぞれの特性に応じた検索手法を提示する技術を開発する。

4 おわりに

東京都23区内100万件掲載のタウンページ電話帳を、インターネットを介して検索できる技術を確立し、平成8年12月6日より試行サービスを開始した。

オブジェクト指向データベースを用いた大規模事例として国内では初めてであり、本格的な地図情報、および広告などさまざまな形式のマルチメディア情報とともにタウンページ検索を提供した。

本システムの構築にあたっては、従来濁音のみであった日本語での曖昧検索技術を漢字、読みの音韻変化、外来語、異体字にまで拡張した。

今後の予定としては、タウンページ電話帳システムの設計管理、およびユーザ行動分析の研究を進めるとともに、さらに大規模な事例を用いたサーバの構築を進める予定である。

最後に、研究の機会を与えて頂いた伊土所長、日頃御指導いただく市川部長、後藤リーダ、ならびにグループ員に感謝する。また、電話帳事業推進部、マルチメディアビジネス開発部の各位に感謝する。

参考文献

- 1) 高橋克巳, 三浦 信幸, 島健一. “インターネット・タウンページの構築 (2) ~あいまい検索技術~” 第54回情報処理学会全国大会 (1997).
- 2) 三浦 信幸, 高橋克巳, 島健一. “インターネット・タウンページの構築 (3) ~個人適応型WWWの試み~” 第54回情報処理学会全国大会 (1997).
- 3) 島健一, 高橋克巳, 三浦 信幸. “インターネット版マルチメディア電話帳の構築” Japan WWW conference (1995).
- 4) “オブジェクト指向データベース”, 日経インテリジェントシステム別冊, 日経 BP 社, (1994).

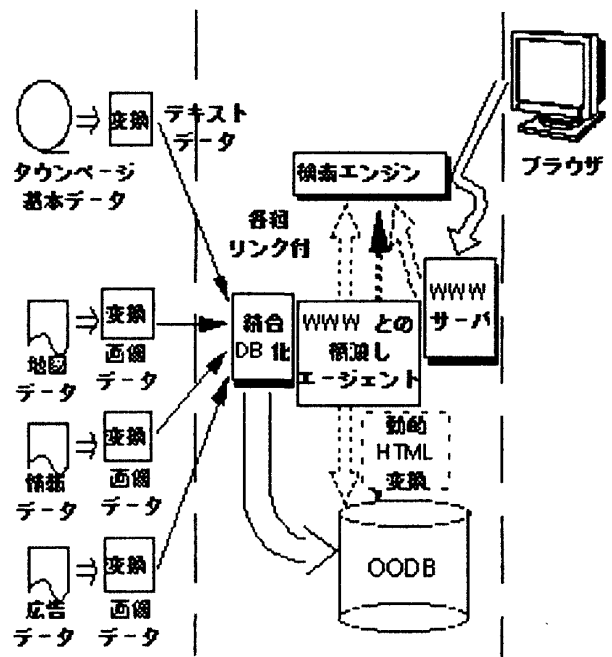


図1: インターネットタウンページシステムの構成

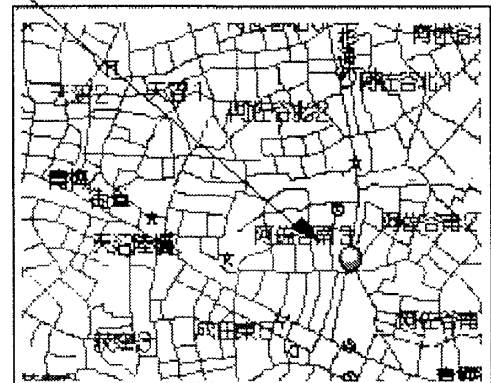
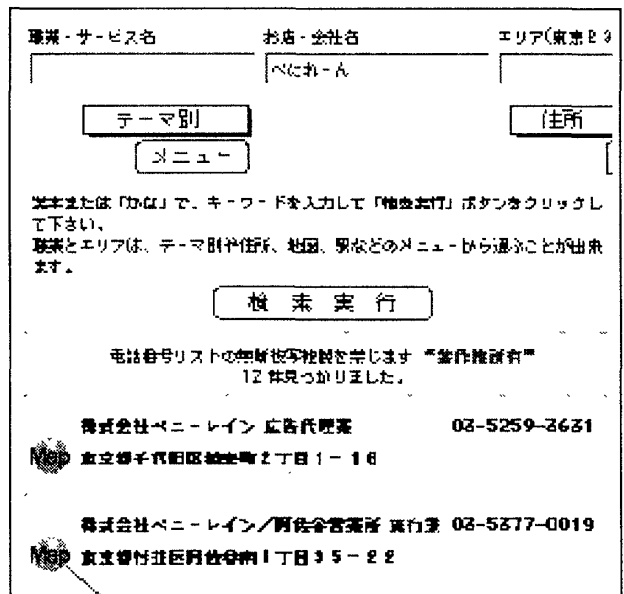


図2: 曖昧検索と地図との連動