

ハイパーリンクによる新聞記事情報構造化の一手法

3G-10

紺谷 精一

平野 泰宏

武田 英昭

井上 潮

NTT 情報通信研究所

e-mail: konya@syrinx.dq.isl.ntt.jp

1 はじめに

パソコン通信で新聞記事情報を提供するサービス [1] などが行なわれ、オンラインで情報をアクセスできるネットワークサービスの有効性が広く認められてきている。ネットワークサービスは、コンピュータに不慣れた利用者のために、簡易な形で操作できることが望ましい。ハイパーテキストの形式で FTP, NNTP, Gopher 等の多種の情報にアクセスできる World-Wide Web (WWW)[2] により、ユーザフレンドリな情報システムが構築されてきている。このようなシステムにおいて、ユーザが関心ある情報を獲得するためには、ナビゲーションによって目的とする情報にたどり着けるように情報を構造化したり、柔軟な検索機能を備えておく必要がある。本稿では、ユーザの視点に沿った柔軟な情報検索システムを実現する手法として、概念に基づいて情報をハイパーテキスト形式に構造化する手法を提案し、プロトタイプとして試作した新聞記事情報システムについて述べる。

2 概念に基づく情報構造化手法

情報を系統的、網羅的に結び付けて構造化することで、人間が知識として把握しやすい形式となる。情報と情報との間を関連(参照関係)で結び付けるためには、どの情報とどの情報を結び付けるかの判断材料としてメタな情報が必要となる。

本研究では、このメタな情報として単語間あるいは文と文との間の概念的な関係を利用することにした。具体的には、表示情報においてアンカー(参照ポイント)として利用されるキーワードやフレーズを概念的に展開して検索を行なう。さらに、概念の構造を表示して概念的

な展開の範囲をユーザが指定できるようにすること、および、検索結果を概念によってグループ化することにより、ユーザの意図を反映した柔軟なナビゲーションを実現する。

3 新聞記事情報システム

本手法のプロトタイプとして、新聞記事情報システムを試作した。概念構造としては、シソーラスを用いた。また、ハイパーテキスト形式のユーザインタフェースが利用できること、オープンなシステムで他の情報との統合が図りやすいことから、WWW を利用した。

3.1 システム構成

新聞サーバは、商用データベース(日経テレコン BiZ[1])から新聞記事をダウンロードし、HTML フォーマットに自動変換し、蓄積する。新聞サーバは、約 2,400 行の perl script で記述されている。また、WWW サーバには、NCSA httpd 1.2 を使用した。

新聞記事の表示・検索を行なうクライアントには、NCSA Mosaic を用いた。

3.2 新聞記事情報の構造化

プロトタイプでは、日経シソーラス [3] から情報処理関連の約 700 語のキーワードとその階層構造を概念として利用した。

記事の構造化は、記事の追加に際して既に構造化された記事の変更を行なわないように、1日単位で行なっている。記事の構造化の形式を以下に示す(図1):

- 各日の記事へのリンク
(メインメニュー)
- キーワードからタイトルへのリンク
(インデックス)
- タイトルから記事へのリンク
(タイトル一覧)
- 記事中のキーワードからインデックスへのリンク
(逆インデックス)

A hyperlinking technique for structuring newspaper information

Seiichi KON'YA, Yasuhiro HIRANO,

Hideaki TAKEDA, Ushio INOUE

NTT Information and Communication Systems Labs.

1-2356 Take Yokosuka-Shi, Kanagawa 238-03, Japan

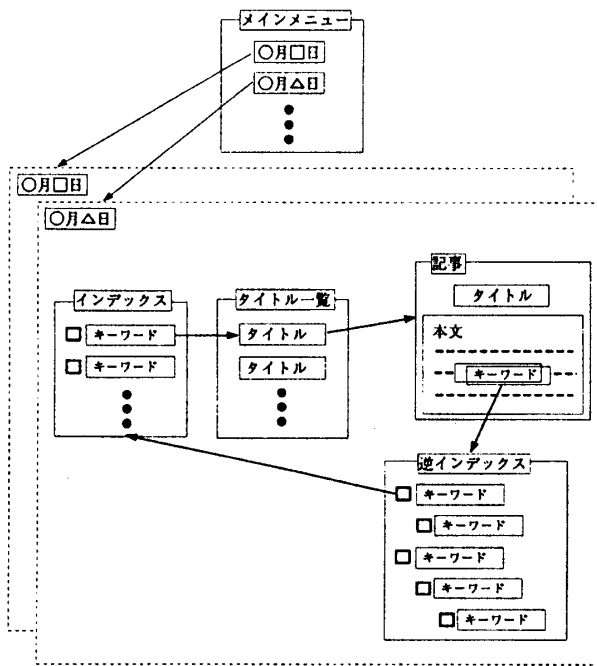


図 1: 記事情報の構造

逆インデックスは、キーワードを概念的な構造を表現したものであり、ユーザが概念的な展開の範囲を指定するのに使用する。

3.3 機能概要

プロトタイプでは、関連記事のナビゲーションをシソーラス階層による記事検索で実現した。記事中のキーワードは、ハイライト(強調)がなされており、キーワードをクリックすると、そのキーワードのシソーラス構造(上位概念キーワード、並列概念キーワード、下位概念キーワード)と各キーワードが含まれる記事の件数が表示される。上位概念が複数存在する場合は、各々のシソーラス構造が表示される。表示されたキーワード集合から check-box 形式でキーワードを選択し、選択したキーワード集合を持つ記事のタイトルを表示する。複数のキーワードを指定することで、類似キーワードの検索を一括して行なうことができる。

関連記事のナビゲーション例を図2に示す。まず、図2奥の画面で示す記事が表示されているとする。ユーザが“パソコン通信”をクリックすると、図2手前の画面に切り替わり、“パソコン”および“パソコン通信”についてキーワードの概念構造を表示する。そして、

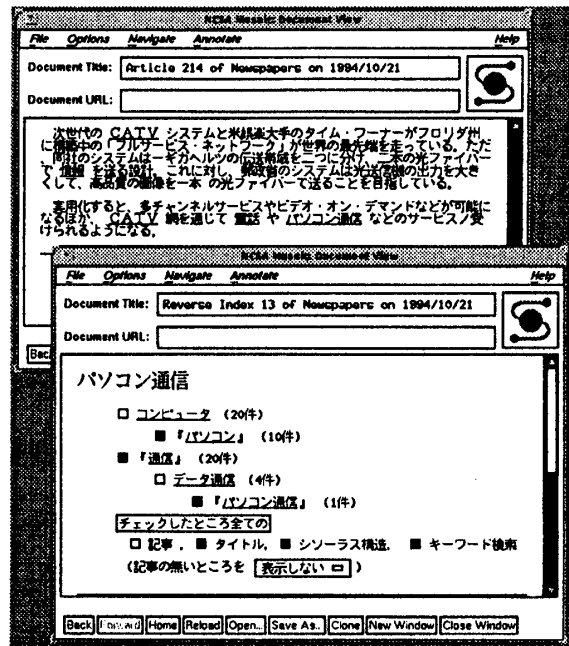


図 2: 画面イメージ

check-box 形式で検索範囲を指定することで、柔軟なナビゲーションが行なえる。

4 まとめ

柔軟な情報検索システムを実現するため、概念に基づいて情報をハイパーリンク形式に構造化する手法を提案した。さらに、本手法のプロトタイプとして、新聞記事情報システムを試作し、新聞記事のナビゲーション、概念を利用した記事検索、記事情報の自動構造化を実現した。

本プロトタイプでは、キーワードが約 700 語、概念的な関係も is-a 関係に限られている。今後は、キーワードの追加への対応や、part-of 等の関係の表現が行なえるように概念モデルの拡張等を検討する予定である。

参考文献

- [1] 「日経テレコン BIZ 操作ガイド」, 日本経済新聞社データバンク局
- [2] R. J. Vetter, “Mosaic and the World-Wide Web,” IEEE Computer, Vol.27, No.10, Oct. 1994
- [3] 「日経シソーラス」, 日本経済新聞社