

# 地域情報における関連情報検索システムの開発とその実証

山本 太司<sup>†1</sup> 梅村 武久<sup>†2</sup> 鳥居 隆司<sup>†3</sup>

通信・放送機構<sup>††3</sup> NTT 西日本<sup>††2</sup> 椋山女学園大学<sup>†3</sup>

## 1. はじめに

現在、地域情報化が推進される中、自治体等において WWW を利用した地域情報の提供が実施されている。これらのサイトで提供される情報はさらに充実・増加することが予想され<sup>1)</sup>、地域住民が必要な情報を取り出すには高い情報リテラシーが必要となりつつある。地域情報においては、住民の目的とする情報を的確かつ速やかに取り出せるような情報提供が望まれる。

我々は、通信・放送機構の「愛知県マルチメディア・モデル市役所展開事業」において、地域情報におけるアクセス効率の向上を目的としたシステムを開発し、実証実験を実施した。本稿では、システムの開発・実証実験の概要について述べる。

## 2. 関連情報検索システムの開発概要

地域情報を検索する状況のひとつとして、ユーザが情報を閲覧中に、そのページ内の情報と関連性の高い情報を検索する、という場合を想定することができる。



図 2.1 関連情報検索システム概念図

今回我々は、この検索状況に対応した「関連情報検索システム」を開発した。システム構成は図 2.2 のとおりである。

本システムの開発ベースとなっているのは、ユーザが検索語を入力するタイプの検索システムであり、関連情

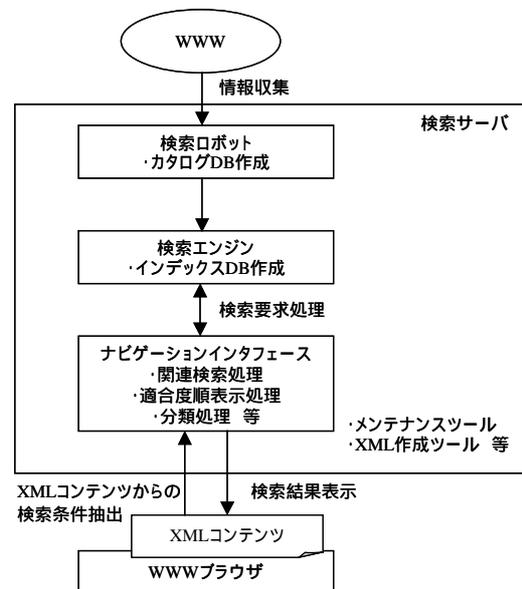


図 2.2 システム構成

報検索だけでなく、ユーザが任意に入力した検索語による検索も可能となっている。

また、本システムのユーザインタフェースは XML となっている。システムに含まれるツールによって、HTML コンテンツから以下のような XML コンテンツを作成することができる。作成した XML コンテンツには元の HTML コンテンツの情報内容を特徴付ける「関連キーワード」が埋め込まれている。XML コンテンツの例を図 2.3 に示す。

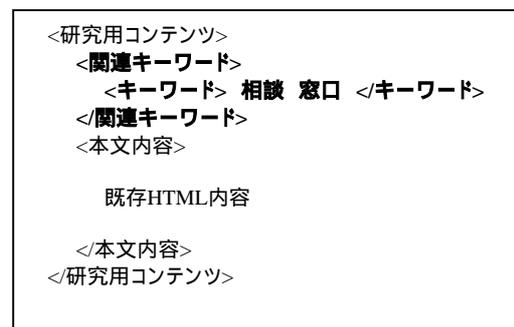


図 2.3 XML コンテンツの例

実際のサービスとしては、ユーザが情報にアクセスし、XML コンテンツ内の「関連情報の検索」ボタンを押すことで、この XML コンテンツに埋め込まれた関連キーワードを基にキーワード検索が実施され、元の HTML コンテンツの内容に関連した情報が表示される。情報は関連キーワードとの適合度順による一覧表示のほか、あらかじめ

Development and actual proof of a system for local residents to search for connected information

<sup>†1</sup> Futoshi Yamamoto (NTT-West Corporation)

<sup>†2</sup> Takehisa Umemura (NTT-West Corporation)

<sup>†3</sup> Takashi Torii (Sugiyama Jogakuen University)

定義しておいた分類基準による分類表示が可能である。

本システムは他の検索エンジンと比較して、探したい情報に関する知識が少ない場合においても、検索キーワードを入力する必要が無い、という優位性を持っている。また、各ページから直接関連情報へジャンプできるため、効率的に関連情報にアクセスすることができる。

### 3. 関連キーワードの設定

本システムの運用にあたって、利用者に有益な関連情報を提供するためには、関連キーワードの設定が最も重要となる。関連キーワードの初期値は、HTML コンテンツの形態素解析<sup>2)</sup>によって得られた単語データを利用し、ページ内出現回数等を基にして設定する。

さらに本システムでは、検索語による検索において、ユーザが複数の検索語を入力して検索を行った場合に、それらの単語は関連があるととして関連語 DB に記録しておく。関連語 DB で関連キーワードと共起する単語のうち、一定の指定値以上の関連語を関連キーワードとして自動的に追加している。これにより、ユーザの志向をシステムに取り込み、検索結果の精度向上を図っている。関連キーワードの設定フローを図 3.1 に示す。

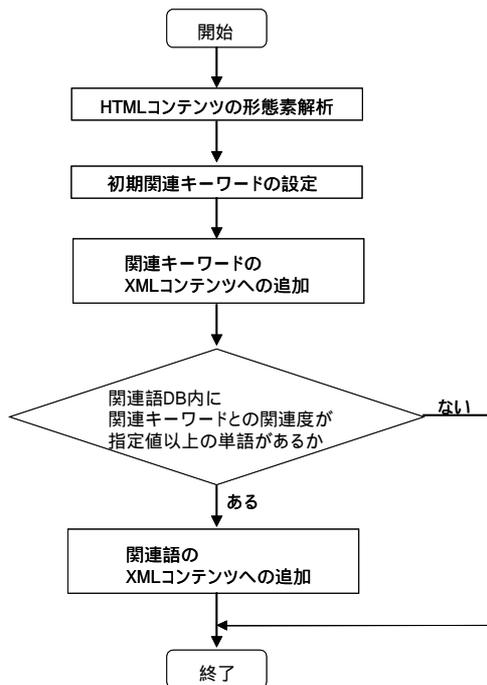


図 3.1 関連キーワード設定フロー

### 4. 関連情報検索システムの実証実験概要

実験用ページとして、関連情報検索サイト (<http://search.pref.aichi.jp>) を開設し、県内の生活情報を中心とした地域情報について、本システムを適用し、実証実験を行った。

関連キーワードの初期設定にあたっては、各 HTML コンテンツに対して形態素解析を実施し、ページ内出現回数上位のキーワードから数語を、ページ内容を考慮した上で、キーワードとして付加した。

運用状況としては、2002 年 12 月現在の登録コンテンツ数 (XML ファイル数) は 224、それぞれに付加した関連

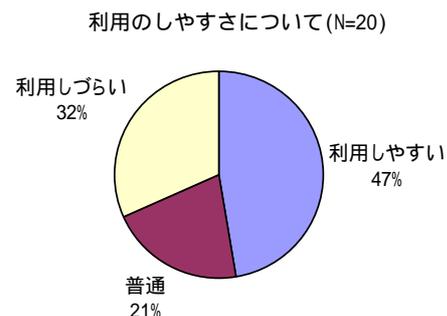
キーワード数は 2~5 となっている。また、2002 年 5 月 1 日~2002 年 12 月 31 日までの関連情報検索サイトトップページのアクセス数は 7,023 となっている。実証実験で実際に公開されている XML コンテンツを図 4.1 に示す。



図 4.1 実際の XML コンテンツ

### 5. ユーザヒアリング

利用のしやすさ等を調査するため、20 人の県民に対してのヒアリングを実施した。結果を図 5.1 に示す。



「利用しづらい」理由 (複数回答あり)  
 ・ 関連情報として欲しい情報が得られない(5 人)  
 ・ 処理速度が遅い(2 人) 他

図 5.1 ヒアリング結果

おおむね良い結果が得られているが、利用しづらいと回答したユーザの多くが、求める関連情報が得られない、という意見を持っており、関連キーワードの変更によりユーザに提示する関連情報の精度を向上させることで、利用しやすさも向上できることを示している。

### 6. おわりに

本稿では、地域情報において関連情報に容易にアクセスするためのシステムの開発と実証実験の概要について述べた。今後は、関連キーワード設定を中心に、ユーザに有益な関連情報を提示できるよう、システムの改良を行っていく予定である。

### 参考文献

- 1) e-Japan 戦略(首相官邸ホームページ), <http://www.kantei.go.jp/>
- 2) 田中穂積監修: 自然言語処理-基礎と応用-, 電気情報通信学会(1999)。