

## タグによる Web 検索の絞込み

鈴木 健太 松田 朋子 野崎 政人 濱川 礼

中京大学 情報科学部 情報科学科

### 1. はじめに

本研究は Web サイトの検索結果をフィルタリングし、ユーザの検索結果の選択を補助することを目的とする。

一般的に Web サイト検索の結果の件数はユーザの入力したキーワードによっては数千件以上になることもあり、ユーザはその中から目的の Web サイトとそうでない Web サイトを判別しなければならない ([1])。しかし数千件以上の検索結果のすべてを判別する事はなく、ほとんどのユーザは上位数件から数十件を判別して目的の Web サイトが見つからない場合は、その検索結果をそれ以上使用せずに別のキーワードで再度検索を行う。このように現状ではユーザは検索結果のすべての情報を判別するには時間と手間がかかるために一部の情報しか活用していないという問題点がある。

これらの問題に対し、本研究では次に示すような手法での解決を図った。

### 2. 提案手法

本研究は Web サイトの検索結果を対象とする。Web サイトの検索結果に、各 Web サイトの内容と関連したキーワードを合わせて表示する。それらのキーワードを一覧で選択できるように表示し、ユーザの選択したキーワードと検索された各 Web サイトに合わせて表示されるキーワードで論理演算し検索結果を任意にフィルタリング出来るようにすることで、ユーザが Web サイトの検索結果群を検索しやすいようにした。

この時の検索された各 Web サイトに合わせて表示されるキーワードを「タグ」と呼ぶこととし、この検索方法を実現するために以下の点に着目した。

**検索結果** 本研究は検索結果にタグを付加しそれらによって検索結果を絞り込む事を目的としているため Web 検索結果の取得は Yahoo!検索から得た情報を利用する ([2])。

**タグ登録** 本研究ではタグを「検索ワード登録」と「タグクローラ」を元に登録する。ユーザの入力した検索ワードと検索結果の各 Web サイトは最も関連性のある内容のキーワードであるが、それだけを Web サイトを表すタグとして用いるのはタグの数が少なくタグによる検索結果の絞込みが有効に出来ない場合があるので、個々の Web サイトごとに内容と関連するタグを複数取得してくる「タグクローラ」も合わせて使用する。

**タグの一覧** タグを検索結果のすべての各 Web サイトから集計し一覧で表示する。同じタグは数をカウントし、カウントの多い順にソートして表示される。これにより各 Web サイトのタグが一定でも、検索されたキーワードの内容によってより関連する内容のタグが上位に変動して一覧で表示されるようになると考えた。

### 3. システムの全体構成

本システムは、以下の 2 つのシステムから構成されている。

- ・ タグの登録処理
- ・ 検索処理

#### 3.1. タグの登録処理

タグの登録処理には

- ・ 検索ワード登録
- ・ タグクローラ

の二つの登録方法がある。

##### 3.1.1 検索ワード登録

ユーザの入力した検索ワードで検索された各 Web サイトの URL と関連付け、タグとして登録する。入力したキーワードで検索された Web サイトであるので検索された Web サイトを表すタグとして関連があると考えこの方法を採用した。また登録された回数をカウントし、検索結果に表示されるタグの一覧ではこのカウントの順にソートをして表示する。

A website search system using the tag that keyword of contents related to the website

Kenta Suzuki, Masato Nozaki, Tomoko Matsuda and Rei Hamakawa

### 3.1.2 タグクローラ

検索ワード登録で蓄積されている URL を取得し、その Web サイトの HTML からタグとして適切な単語を抽出し、それらをその URL に関連するタグとしてデータベースに登録する。タグとして適切かどうかの選択方法は HTML のデータから HTML タグを削除し、残ったデータの形態素解析をして品詞が名詞となるものをタグとして用いた。またそこからタグとして不要だと思われる単語のリストも別途用意し、それと重複する語も削除した。

## 3.2. 検索処理

### 3.2.1 キーワード検索

ユーザから得た検索ワードを Yahoo! 検索へリクエストを出し、取得した Web 検索結果の各 URL が本システムのデータベースに登録されているものであればユーザに表示する Web 検索結果にタグ情報を合わせて表示する。

### 3.2.2 タグによる検索結果群の絞込み

検索結果に付加されているタグを選択する。そこで選択されたタグで論理演算を行い検索結果の絞り込みをする (図 1)。

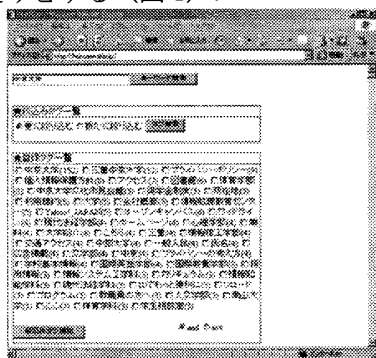


図 1. タグ選択画面

ここでは各 Web サイトに付加されているタグを集計し一覧で表示する。and, not 論理演算のどちらかで処理を行うかを選択し、その条件で検索された各 Web サイトに付加されているタグと比較し検索結果を絞り込み表示する。

## 4. 実例

「中京大学」の検索結果より「一般入試」の内容を含む検索をタグにより絞り込んだ結果このようになる (図 2)。

Yahoo! 検索で「中京大学」を検索した場合約 333 万件の検索結果がありユーザは 1000 件利用できる。例えば本システムでその検索結果の中から「一般入試」のタグで絞り込むと 4 件となりユーザの検索結果の判別の時間と手間を削減することが出来た。

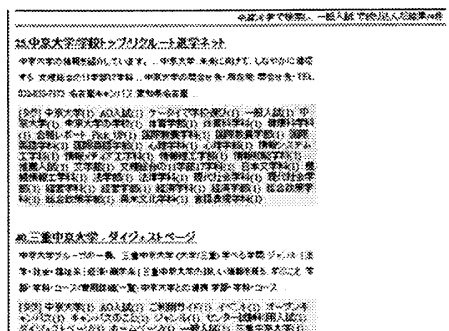


図 2. タグ選択画面

## 5. 評価

本研究の評価を研究室の生徒にしてもらったところ検索時のステップ数が減少したという評価が多かったが、最初にタグの付加された状態の検索結果を得る時間がかかるという意見が多かった。そこで 1 回の Web 検索にかかる処理時間を計測したところ従来の Web 検索より処理に時間がかかってしまう結果が出た (表 1)。より快適な検索を実現するためサーバマシンの性能を上げる事やデータベース設計の更なる効率化が必要である。

表 1. 検索結果の出力にかかる平均時間

	表示件数	処理時間
Yahoo!検索	10 件	0.02s
	100 件	0.06s
本システム	10 件	4.8s
	100 件	7.5s

## 6. まとめと今後の課題

本研究では Web サイトの検索結果にタグを付加し、それらを用いて検索結果群を絞り込んで表示するシステムを開発した。

現在、本システムのデータベースに登録されている Web サイト数は約 16 万で、どのような検索結果にも十分なタグ情報を付加した状態で検索結果を返すには数が少ない問題がある。そのためどのキーワードで検索されてもタグの絞り込み検索ができるように Web サイトを別途登録するクローラを作成する必要がある。またタグの精度についてはタグクローラの部分を改善し、今後更に細分化したタグの選択基準を設ける必要があると考える。また処理速度向上のためにデータベースの設計も見直す予定である。

## 参考文献

- [1] Google ウェブ検索  
<http://www.google.co.jp/>
- [2] Yahoo! 検索  
<http://search.yahoo.co.jp/>