

Title	教官公募情報のダイジェスト自動生成
Author(s)	見館, 潔
Citation	
Issue Date	1999-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1251
Rights	
Description	Supervisor:佐藤 理史, 情報科学研究科, 修士

教官公募情報のダイジェスト自動生成

見舘 潔

北陸先端科学技術大学院大学 情報科学研究科

1999 年 2 月 15 日

キーワード： 自動編集、自動ダイジェスト生成、ワールドワイドウェブ、教官公募情報.

近年、インターネットの普及により電子的にアクセスできる情報量が爆発的に増加した。これに伴い、ユーザが欲しい情報を素早く見つけることが困難になり、これを支援するシステムが不可欠となっている。

これを支援する情報検索には、(1) サーチエンジンに代表されるような、情報が必要になった時に探索する方法と、(2) あらかじめ、後の利用を想定して情報を編集しておき、これを利用する方法、の二つの方法がある。

サーチエンジンは、ユーザに「得たい情報に関するキーワード」を指定してもらい、そのキーワードを含むテキストを提示するシステムであり、多くのサーチエンジンが実用に使われている。しかし、サーチエンジンは以下の理由により万能とは言えない。

- ユーザ側から見れば、適切なキーワードの入力が必要である。
キーワード検索システムは、一つの語、または、複数の語を論理演算子で統合したものを、検索質問として受け付ける。そのため、検索者は検索要求をこのような形式の質問に適切に変換する必要がある。
- システム側から見ればキーワードから分かることが少なすぎる。
システムは、検索質問が適切なものであることを前提としているため、ユーザが知りたい情報とかけはなれたことばが検索要求された場合どうしようもない。

そこで、本研究では、(2) の方法を採用し、ワールドワイドウェブ上の大学などの教官公募情報を対象に、情報を自動的に収集、ダイジェスト形式に編集しユーザに提供するシステムの作成を行った。この方法は、上記の問題に対して、あらかじめ情報をフィルタリングして保存しておくため、ユーザの適切なキーワードの入力を省くメリットがある。

教官公募情報をワールドワイドウェブを用いて見たいと考える人々は、全ての教官公募情報を網羅的に読むのではなく、自分の興味のある公募情報だけを取捨選択して読みたいと考える。このため教官公募情報がダイジェスト形式で提示され必要に応じて、詳細な情報が参照できるようにすることが望ましい。

しかし、学術情報センターの「研究者公募情報」のように、公募情報を人手により編集してダイジェストとして提供するには、作成に時間がかかり、またコストがかかる。そこで、ワールドワイドウェブ上の教官公募情報を機械処理によりダイジェストを自動生成し、ユーザに提供する方法を検討する。これを実現することは、以下のようなメリットがある。

(1) 機械処理により、ダイジェスト作成、更新が短時間で大量にできる。

(2) ダイジェスト作成コストがほとんどかからない。

以上の考えにより、システムを作成した。本システムは、以下の4つのモジュールと2つのデータベースから構成される。

(1) 収集モジュール

ワールドワイドウェブ上に存在する、公募ページらしきページを定期的に収集する。

(2) 選別モジュール

収集モジュールが取得してきたページを、公募ページ、公募リストページ、その他のページ、に選別する。

(3) 情報抽出モジュール

公募ページから、「公募ページのURL」「公募機関名」「所属先」「公募職名」「分野・職務内容」「公募締切日時」「公募機関の地域」を抽出する。

(4) 公募情報データベース

情報抽出モジュールで抽出された情報を「公募ページのURL」「公募機関名」「所属1」「所属2」「所属3」「所属4」「公募職名」「分野・職務内容」「公募締切日時」「公募機関の地域」の10フィールドからなるレコードとして格納したデータベース。

(5) 検索モジュール

ユーザの要求に応じて公募情報データベースを検索し、その結果をダイジェスト形式で表示する。「公募機関名」「公募機関の地域」「分野・職務内容」「公募職名」の4つの条件を指定することができる。ダイジェスト表示の所属先には、その大学のホームページへのハイパーリンクが付加される。

(6) 公募リンクデータベース

公募リンクページと判定されたページのURLのデータベース。収集モジュールで利用される。

本システムは、プログラム言語 Perl5 を用いて実装されており、その規模は、約 3000 ステップである。なお、本システムは、現在、JAIST において試験運用しており、ワールドワイドウェブを通して、アクセスすることができる。

本システムの中心的部分は、収集モジュールと情報抽出モジュールである。収集モジュールは、教官公募ページらしきページを収集するために、サーチエンジンを用いる方法、自動巡回ロボットを用いる方法、公募リンクデータベースを用いる方法、の3つの方法を実装した。情報抽出モジュールでは、教官公募ページによくみられる見出し語を用いてページをいくつかの部分に分割し、分割された各部分から、抽出する情報に特有な言語表現パターンを用いて情報を抽出する。抽出する情報は、所属先、公募職名、分野・内容、公募締切日時、公募機関の地域である。分野・内容の情報は、書き手により表現がさまざまであり、かつ書かれている文章から専門分野を特定することが現段階では困難なため、80-85 %と精度はそれほど高くない。これ以外の情報は、90 %以上の抽出精度を得られた。

ワールドワイドウェブ上で教官公募情報を検索できるサービスに学術情報センターの「研究者公募情報」がある。このサービスは現在、学術情報センターに公募掲載依頼があったもののみを人手により編集しているので、掲載依頼がない限り掲載されることはない。これに対して、本システムでは、ワールドワイドウェブ上に存在する教官公募ページを自動的に見つけ出すことができる。また現在、紙の媒体による教官公募案内が主流であるが、今後さらにインターネットが普及することにより、紙の媒体と同時に各公募機関がワールドワイドウェブ上で公募情報を発信していくと思われる。そうなると、本システムの重要性がさらに高まると考えられる。今後、学術情報センターと関係を図り、本システムで得られた教官公募情報を学術情報センターに提供していく予定である。