Title	
Author(s)	,
Citation	
Issue Date	1999-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1251
Rights	
Description	Supervisor: , ,

Japan Advanced Institute of Science and Technology

Automated Editing of University Employment Bulletins.

Kiyoshi Mitate

School of Information Science,
Japan Advanced Institute of Science and Technology

February 15, 1999

Keywords: Automated Editing, Automatic Digesting, World Wide Web, University Employment Bulletins..

In recent years, the amount of information which we can electronically access increased explosively by the spread of the Internet. Therefore, Internet users feel difficulties in finding information that they want from electronic text mass media such as the USENET News (netnews) and World Wide Web (WWW). The system which supports them is indispensable.

There are two methods in the information retrieval which supports users. (1) The search method for when users look for particular information. (e.g., search engine), (2) The editing method for further retrieval of users use.

The search engine is the system which presents the text containing the keywords of a reference demand. Many search engines are now for practical use, however, search engines are not omnipotent because the following reasons.

- Users must input the suitable keywords for the system. A retrieval-by-keyword system receives one word, and two or more words that are unified by the logic operator as a reference question. Therefore, a retrever needs to transform his reference demand into such forms.

- Systems obtain little information from keywords. The system premise that a reference question is suitable. Therefore, if the words which was different from the information that a retriever wants to know is chosen as a reference demand, the system cannot cope with it.

In this research, I implement the automated editing system of university employment bulletins using the secound method. This system filter and store information previously,

1

so there is a merit that users do not need to input suitable keywords. Therefore, this method is effective to the previous problems.

People who see the university-employment-bulletins information on World Wide Web wants to select and read only interested information. It is desirable that university-employment-bulletins information is shown in digest form, and enable users to refer detailed information if necessary.

However, editing and carrying out digest of university-employment-bulletins information by hand requires a lot of time and cost. (e.g., ”*University-employment-bulletins information*” of NACSIS). Therefore, I concentrate on automatic generation of the digest of the university-employment-bulletins information on World Wide Web. Realization of this system has the following merits.

(1) Digest generation and its updating can be performed in large quantities for a short time.

(2) Digest generation cost is hardly needed.

A system configuration is shown below. The system consists of four modules and two databases: (1) collection module, (2) selection module, (3) information extraction module, (4) employment information database,(5) retrieval module, and (6) URL database.

(1) collection module
The collection module collects Web pages that may be related to university employment bulletins.

(2) selection module
The selection module categorizes the collected pages into three groups : (a) the pages that describe details of employment, (b) the pages that have links to other employment pages, and (c) other pages.

(3) information extraction module
The information extraction module extracts seven items from each employment page : URL, university name, division name, vacant post, research field, application deadline, and university location.

(4) employment information database
Employment information database is database with which one record consists of ten fields : URL, university name, division name 1, division name 2, division name 3, division name 4, vacant post, research field, application deadline, and university location.

(5) retrieval module
The retrieval module retrieves the database, and presents the result as digests. The following reference demand is received : university name, university location, research field, and vacant post.

(6) URL database

This database stores URL of the pages that have links to other employment pages; this is used by the collection module.

This system is implemented using programming-language Perl5, and the scale of the system is about 3000 steps. Examination execution of this system is carried out in JAIST now. It can be accessed via World Wide Web.

The central parts of this system are the collection module and the information extraction module. The collection module collects Web pages that may be related to university with three method : the method using a search engine, the method using a webrobot, and the method using URL of the pages that have links to other employment pages. The information extraction module divides a university-employment-bulletins page into several parts using an item label, and information that is extracted from using a language pattern. This module extracts seven items from each employment page : URL, university name, division name, vacant post, research field, application deadline, and university location. The accuracy of information extraction on *research field* is not so high as 80 to 85 %. Because it is difficult to extract *research field* from the sentences that contain various expressions by the different writers. The extraction accuracy of the other information is more than 90 %.

The "*University-employment-bulletins information*" of NACSIS is the service which peopele can search university-employment-bulletins information on World Wid Web. This service is edited by human power, only when a university requests on addition to the university-employment-bulletins information; that is, no information will be added to the university-employment-bulletins information if there is no request. This system can find out the university-employment-bulletins pages which exist on World Wide Web automatically. Though, the university-employment-bulletins guidance by the medium of paper is mainly used now, if the Internet spreads more each university opens its own university-employment-bulletins information using World Wide Web. The importance of this system will increase in the near future. It is planned that ,NACSIS will provid with the university-employment-bulletins information that is obtained by this system.

NACSIS : National Center for Science Information Systems