Extended Summary<

A development of an efficient information collecting and retrieval system using an Agent Technology –for infectious disease-

Takemura T, Ashida N, Makimoto K, Kirikae T, Okamoto K, Kuroda T, Nagase K, and Yoshihara H
Department of Medical Informatics, Kyoto University Hospital,
Department of Medical and Welfare, Koshien University,
School of Allied Health Sciences, Osaka University,
Research Institute, International Medical Center of Japan

I. INTRODUCTION

At present, a lot of information exists on World Wide Web (WWW) because of spreading of the Internet. These information is, naturally, useful or not a users has an information need. It is very difficult that we judge information that we search is most effective on WWW and we always catch up the newest information if we find a web site including effective information. Moreover, we can not prospect of information retrieval system's updating and providing newest information, even if we can check the useful web site every day.

However, a method of using software agent system is proposed to these problems. The agent is a program that performs some information gathering or processing task autonomously. The internet agent is just as "agent", collect, retrieve, check objective information and so on. The Internet agent has to have agentive, autonomous and adaptive, but there is no internet agent systems that has all specification sufficiently, and multi-purpose agent system hasn't shown up yet.

However, we can consider that multi-purpose agent is not only easy-to-use. If an agent system have gather and check up some web sites decided, the agent system will be very useful.

By the way, infection disease information is very important to prevent the spread of disease. This information includes epidemiological, pathogenic, serum and so on, and it must be provided public as knowledge base that many people can use. Especially, it is desirable that physician, nurse or person that devises a countermeasure of infection disease can read it on time. The internet shows a big effect in this case as well and Infectious disease surveillance center, National Institute of Infectious Diseases in Japan provides a lot of infection disease information in the country level. Moreover prefectures health care center also provide the information in prefectures and smaller level, and it is using the infection disease prevention and the strategy of the control.

However, these sites hold and update information individually, so persons who want to get the information of infection disease have to check and search the every site. Therefore, if the internet agent system works as their "agent", it is very effective for persons who want the information.

In Japan, the research of "a development of hospital infection prevention network" has been conducted to reduction to infection disease information scattering.

In this paper, we developed the collect and retrieval system using an agent technology for the infection disease information efficiently.

Concretely, user can get the infection disease information from keyword retrieval and update page list based on several agents. The agents have three functions mainly as follows.

1. An agent judges information about the infection disease or not automatically.
2. It is as early as possible to catch up the updating information, so collected data make always update.
3. Automatic downloading.

II. METHODS

1. The characteristics of infection disease information on WWW

The infectious disease information on WWW is provided by National Institute of Infectious Diseases, Ministry of Health, Labor and Welfare, prefectures health center, division of Hospital Infection control team and so on, as administrative service, or large hospital provide information of method of measures infection. Several web sites have other issue different from infectious disease. And it is difficult that user grasp the update information because almost every site is very large-scale.

Therefore we consider that we list several web sites is judged useful for infectious disease and we have to deal with the information relates to it in these sites.

Figure 1 shows general structure of web sites.

Extended Summary<

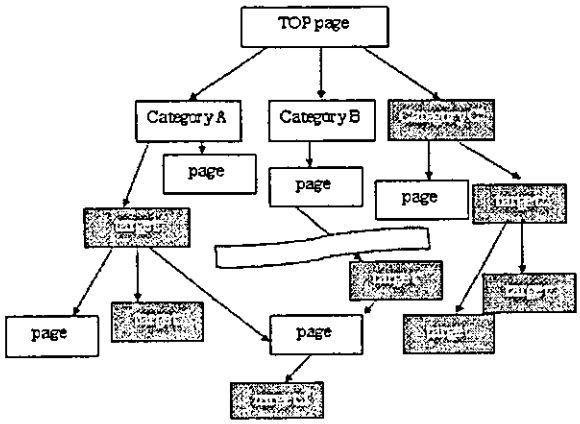

Figure 1. web site structure

2. Agent framework

The basic concept of the agent in this research is three agents work in interaction. One agent is "download agent". This agent's main task is downloading from objective pages according to new URL that given other agents. Second agent is "search agent". This agent rounds web sites according to URL list and confirms the page is updated or checked new pages. Third agent is "judgment agent". This agent judges the web page that "download agent" get, is infection disease page or not. This concept is shown in figure 2.

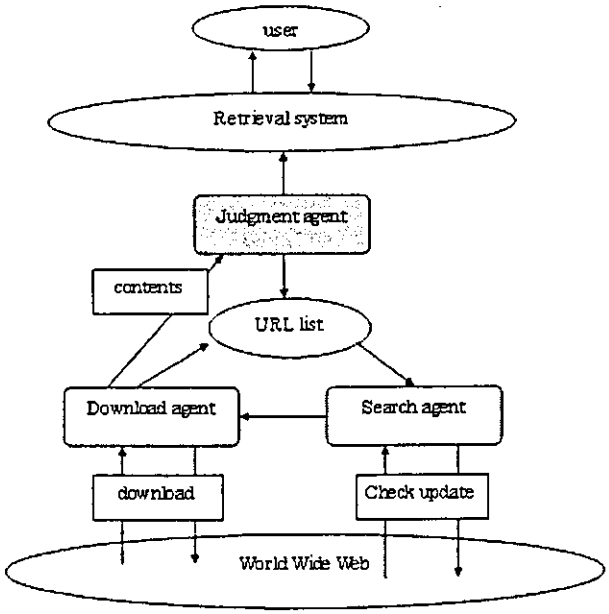

Figure 2. a concept of agent framework

3. Automatic judgment of "judgment agent"

In this concept, it is necessary that the judgment agent judge whether a download page related to infectious disease or not. This judgment ability is very important that if the agent include too mach download pages, the gathering information contains a lot of noises, on the contrary if it underestimate, useful information falls off.

We adopted the technique of Support Vector Machine (SVM) on this time, though there were various methods to judge the contents of a certain document mechanically. SVM is one of the pattern recognition techniques and is expanded by the method which is called a kernel trick is used as the mathematical technique to construct the distinction function of the non-line form. So SVM is one of the best learning models in other many recognition techniques well known. SVM is, to put briefly, a learning method to distinguish two classes. If many objects which have each attributes (attribute A and attribute B) can be expressed by a vector as some amounts of characteristics and mapped on the vector field, SVM choices the nearest elements in two attribute space, and draws a boundary line so that a distance from that both may be the longest. This boundary line makes divide any elements into either. Conceptual figure shows as follows.

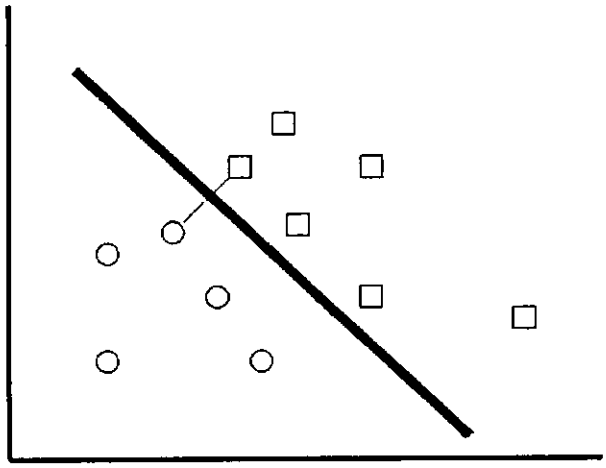

Figure 3. Support Vector Machine

On this time, we made SVM learn medical document or not from the point of medical terminology of which the documents consisted, and SVM judges an objective document whether infections or not with the knowledge. And the knowledge has constructed continuously by learning correct answers judged by human, the data which the download agent corrects, to put it concretely, are judged by us, the judgment agent improve the precision of judgment because the it learn the data with SVM.

III. Result and Discussion

We developed the system that can collect and judgment the only infection information with an agent technology. This system can collect all web sites which contain infection data without omission, and we can efficiently use the information which is scatted and loss on WWW.

Extended Summary<

Naturally, we can get a lot of infection information crossly and functionally if this system works within a certain period of time. Therefore, we can try to analyze and mining data from the point of various view. In this case, it is very important that we use natural language processing technology because almost every data is expressed by free text.

An agent technology makes us help very efficiently when we want to collect and retrieval information as "agent" literally.

厚生労働科学研究費補助金　医療技術評価総合研究事業

院内感染地域支援ネットワーク及び相談体制の改善・普及や、
データベースおよびバックアップ体制の構築に関する研究

平成16年度　総括研究報告書

発行日　平成17年4月10日
発行者　倉辻忠俊
〒162-8655　東京都新宿区戸山1－21－1
国立国際医療センター研究所　☎(03) 3202－7181
製　作　株式会社メヂカルフレンド社