

# Adaptive Information Searching on Internet Based on User Personal Information(ユーザの個人 情報に基づいたインターネットにおける適応型情 報検索)

|     |   |
|-----|---|
| 著者  | OKADA SHIMOE ROBERTO SADA AKI   |
| 号   | 101   |
| 発行年 | 1998  |
| URL | <a href="http://hdl.handle.net/10097/12783">http://hdl.handle.net/10097/12783</a> |

|         |   |         |         |      |
|---------|---|---------|---------|------|
|         | オカダ シモエ ロベルト サダアキ   |         |         |      |
| 氏名 (本籍) | OKADA SHIMOE ROBERTO SADA AKI (パラグアイ)   |         |         |      |
| 学位の種類   | 博士 (情報科学)   |         |         |      |
| 学位記番号   | 情博 第 1 0 1 号  |         |         |      |
| 学位授与年月日 | 平成 11 年 3 月 25 日  |         |         |      |
| 学位授与の要件 | 学位規則第 4 条第 1 項該当  |         |         |      |
| 研究科、専攻  | 東北大学大学院情報科学研究科 (博士課程) 情報基礎科学専攻  |         |         |      |
| 学位論文題目  | Adaptive Information Searching on Internet Based on<br>User Personal Information<br>(ユーザの個人情報に基づいたインターネットにおける適応型情報検索) |         |         |      |
| 論文審査委員  | (主査)  |         |         |      |
|         | 東北大学教授  | 白鳥 則郎   | 東北大学教授  | 根元義章 |
|         | 東北大学教授  | 阿曾 弘具   | 東北大学助教授 | 木下哲男 |
|         |   | (工学研究科) |         |      |

## 論 文 内 容 要 旨

### 1 Introduction

The explosion in the amount of electronically available information and the powerful computers at lower cost are causing the computer-for-everybody trend, extending the range of users to non-experts. These phenomena create an information superhighway, which would be difficult to approach without help. Intelligent software agents will provide this necessary help to navigate in the world-wide network of information. In this thesis, we first review the agent-based information locating systems and present our view of the trends of agent technologies supporting users in finding information on distributed heterogeneous environments.

Next, we tackle the problem of customization of information gathering results to user personal tastes. For this, we propose a new method to customize the information searching on Internet. Existing search tools which evaluate the relevance of documents to the user query give many false positives. One reason is that at query formulation, the information provided by the user is too few to provide good results. In order to complement works on Web agents for autonomously browsing or filtering on behalf of the user, we focus on adaptive agents for providing support at query formulation and retrieval. These adaptive or learning agents makes possible to add a layer of personalization and collaboration between the users and the existing Web engines. The personalization is realized by learning the user model and using it at query formulation. We have implemented and experimented our web search method. The results of the experiments show that our approach provide an effective support for customized search and the quality of returned results increase.

Finally, we investigate the issues associated with the exchange of information learned by agents belonging to different users, in order to improve the slow initial learning curve problem for the learning systems.

## 2 Society of Agents for Information Gathering on Internet

The explosion in the amount of electronically available information put additional burden for the designers of information systems as well as for the people who use them. We decided to work on Internet, as it is one of the largest publicly available source of information and it constitutes a good testing ground.

Some of these drawbacks, which put a big burden on the human users are: (p1) they should know from where to obtain the desired information and how to interact properly with the heterogeneous information servers on a world-wide distributed information network, (p2) they should navigate on this huge information space in search for the desired data or put together a set of related pieces of information, acquired from several sources with data in a variety of forms and with different protocols - this is boring, difficult, time consuming and inefficient, (p3) they should figure how to express the query properly, which can be a difficult task in unfamiliar domains. Additional drawbacks related to the quality of information include the needs to maintain consistent and updated information and support for security and negotiation for obtaining information.

To overcome these drawbacks, our approach is based on one of the most promising software paradigms for the next generation information network: agents. They are attractive - both in academia and industry - and the interest will increase as computing systems become more distributed, interconnected, and open. In order to operate efficiently in such environments, agents have to autonomously plan and pursue their actions and goals, to cooperate, to coordinate, to negotiate with others, and to respond flexibly and intelligently to dynamic and unpredictable situations.

In this chapter two, we first survey the existing partially-agent based approaches for locating information on a highly distributed, heterogeneous environment with a huge number of information sources like Internet. It is not possible to build a single all-mighty agent having knowledge about different domains. Also, if all the tasks are assigned to a single agent, it constitutes a processing bottleneck and a single point of failure. Additionally, because of the complexity of the information finding and filtering task, and the large amount of information, the required processing would overwhelm a single agent. Thus it is more natural to have several agents, each with expertise in a given domain, and make them cooperate.

We believe that the next step is to integrate these researches in order to build a *Society of Cooperative Agents*.

In this chapter, we propose a completely-agent based framework to realize a Society of Cooperative Agents for Information Gathering on Internet. in order to support users in access and composition of associated information spread across a world-wide network of information servers. The basic idea is to build a layer of cooperative agents between the low-level network layer and the user layer, where the agent layer is composed of user-specific, task-specific and information-specific agents, named User Agent, Manager Agent and Machine Agent, respectively. Our view of the future is a society of such agents to bridge between the upper layer of human users and the lower layer of distributed heterogeneous network. These artificial assistants will work together and at the same time improve their usefulness, to reduce the existing gap between these two layers.

### 3 User-adaptive Information Searching on Internet

In this chapter three, we will focus on the problem (p3) regarding the difficulty for proper query formulation.

As it has been arguing that the standard Web search services are far from ideal, many researches are seeking for a better way to tackle the ever growing WWW. Various ranking algorithms used to evaluate the relevance of documents to the query are impractical. This is because the information given by the user is too few to give good estimation. Most search tools still use keywords to specify queries. One factor limiting the precision of queries is that users do not submit well-focused queries. In general, queries get more precise as more words are added to them. Unfortunately, the average number of words in a query submitted to is 1.5, barely enough to narrow in on a precise set of documents. One way to improve effectiveness is to better represent the information need by adding useful terms to the query, e.g. by relevance feedback, a well known technique in information retrieval, where terms occurring in known relevant documents are added to the query. Existing systems with query reformulation support work well for closed information spaces. However, the Internet information space cannot be considered in this category. Another drawback of existing query reformulation is that the support they provide are not customized to user's tastes.

In this chapter, we present a new query reformulation support, in order to provide customized search results. Our proposed support consists of : (1) suggestion of related words for query reformulation, (2) feedback of user interest to refine the query and (3) a clustering of the search results.

Our approach is based on the assumption that *the information about the query that are not specified in the user's input may be obtained from the results of the previous or past queries.*

In this chapter, the objective is to assist users in personalized query reformulation for information gathering adapted to user preferences. The basic idea is: (1) build a user profile expressed as a set of terms expressing his interest, from acknowledged documents, and (2) provide support for query reformulation based on the terms contained in his profile.

The User Agent is in charge of learning user preferences, by looking at the user search behaviour and extracting terms representing user interests from documents acknowledged by the user.

The experimental results shows that: (1) features representing user interests are generated, and (2) by adding these related terms to the user input query, we can have a customized service with the general-purpose web search engines

### 4 Collaborative Agents for Information Searching on Internet

In chapter 3, we have focus on the interaction "one user  $\leftrightarrow$  one User Agent  $\leftrightarrow$  search engines", with focus on the user profile learning mechanism. Machine learning and information retrieval techniques were utilized to learn the user preferences and to provide support for well-formed personalized query reformulation. However, for learning agents working individually, they face two problems: (i) *serendipity*, i.e. they cannot deal properly with situations previously unseen in the past; and (ii) *cold-start*: they spend some time to relearn about new situations.

In order to deal with these problems, we add a layer of collaboration between the agents, where the

selection of peers is based on the trust relationship among them. In this chapter 4, our aim is to assist the information search process by collaborating with trusted peers, in order to avoid re-learning what others already have learned, in order to improve the learning curve.

When faced with an unfamiliar situation, an agent consult its “trusted” peers, instead of broadcasting the request to every known agents. This trust relationship between agents is also learned through the successive interactions.

Each of our user agents maintains: (1) history information (i.e. where to find information related to a given topic), (2) personal dictionary information (i.e. the query and related words, as explained in Chapter 3 for customizing information search). In this chapter, in order to allow collaboration, we have added the trust relation information to the profile. Another important feature of our approach is the use of *referrals* i.e. advices from known agents about other agents which might have similar interests in order to build communities of interests.

We have implemented and experimented with our collaborative search. We have shown that in our approach, the user agents of different users - by sharing both types of information - improve the learning curve and are able to face new situations effectively.

## 5 Conclusions

In this thesis we first review one of the most promising software paradigms for the next generation information network: agents. They are attractive - both in academia and industry - and the interest will increase as computing systems become more distributed, interconnected, and open. Our view of the future is a society of such agents to bridge between the upper layer of human users and the lower layer of distributed heterogeneous network. These artificial assistants will work together and at the same time improve their usefulness, to reduce the existing gap between these two layers. Next, we have proposed a method for query reformulation adapted to user interests. The experiments show that with our proposed method, (1) features representing user interest are generated, and (2) by adding these related terms to the user input query, we can have a customized service with the general-purpose web search engines.

Despite the success reported for learning user’s interests and providing support at query formulation, some drawbacks remains. The problems with the learning approach are that they require a sufficient amount of time (the training phase) before they can start being useful, i.e. it present a slow increasing learning curve. Even worse, their competence is limited to situations similar to those they have encountered in the past. The agents of different users thus have to go through similar experiences before they can achieve a minimal level of competence, although there may exist other agents that already possess the necessary experience and confidence. In chapter 4, we have proposed a method for collaborative information searching, in order to solve the problem of the slow learning curve when an user agent faces a situation unseen before.

Our approach is based on the fact that agents build a trust relationship among them and use this trust weight in order to decide collaborative peers.

## 論文審査の結果の要旨

インターネットを利用して情報検索を行うユーザにとって、様々な情報源の中から適切な検索対象を同定し、必要とする情報を得ることは困難な作業であり、情報源の増大に伴ってその負担は一層大きくなっている。これまで情報検索については、検索アルゴリズムなどを中心として研究開発が行われてきた。しかしながら、多様なユーザ要求に柔軟に対処する手法は確立されておらず、ユーザの視点から情報検索を支援する効果的な手法が必要とされている。そこで著者は、エージェント技術を適用し、エージェントがユーザの要求や特性を反映して作業を代行することにより、ユーザの情報検索作業を柔軟に支援する手法に関する詳細な研究を行った。本論文はその成果をまとめたものであり、全編5章からなる。

第1章は序論である。

第2章では、ユーザが情報検索を要求するユーザ層と検索対象が分散配置されるネットワーク層の間に、新たに検索支援のためのエージェント層を導入したモデルを提案している。本モデルはエージェント層において、ユーザエージェントなどの3種類のエージェントが協調して検索支援を行う方式を与えるもので、従来の検索ロボットなどに比べて、より柔軟な支援機能が実現できる。

第3章では、ユーザエージェントに着目して、ユーザの特性に合致した検索結果が得られる個人適応型検索支援の手法を提案し、プロトタイプシステムの試作と実験により有効性を検証している。提案手法は、過去のユーザの検索結果から検索対象に関するユーザの特性を表わすユーザプロフィールを学習し、これを検索時に利用することにより検索範囲の絞り込みや検索ノイズの軽減を図っている。これは、情報検索支援の実用上、興味深い成果である。

第4章では、第3章で提案した手法を拡張して、複数のユーザエージェントの協調に基づく検索支援の手法を提案・実装し、処理可能な検索要求の範囲が拡大できることを示している。提案手法は、複数のユーザエージェント間でユーザプロフィールに関する情報を交換し、単独のユーザエージェントでは困難であった未経験領域での検索支援を可能とする。これは、より高度な情報検索支援を実現する上で重要な成果である。

第5章は結論である。

以上要するに本論文は、エージェント技術に基づいて、インターネット上での情報検索支援に関する新しい手法を提案し、ユーザの特性を反映した効果的な情報検索支援を実現する基礎を与えたものであり、情報基礎科学の発展に寄与するところが少なくない。

よって、本論文は博士（情報科学）の学位論文として合格と認める。