A Web Searching Guide: Internet Search Engines & Autonomous Interface Agents Collaboration

Ahmed Al-Nazer
Saudi Aramco, Dhahran, Saudi Arabia
ahmed.nazer@aramco.com

Tarek Helmy
King Fahd University Of Petroleum & Minerals, Dhahran, Saudi Arabia
helmy@kfupm.edu.sa

Abstract

The Internet represents the biggest communication media and its dimension increases every day. This continuous growth of information makes the Internet more and more interesting, but also the task of finding selected information becomes more complex and hard. Finding exactly what a user needs is not always an easy task: for example common search engines provide thousands of links for every search. Obviously not all these links are related to what the user really needs. In this paper, we present a Collaborative Autonomous Interface Agent (CAIA) that collaborates with the Internet search engines and supports the user in finding exactly the information consistent with his/her interest. A system has been designed, fully implemented and tested. The testing results shows a big improvement in the relevancy of the retrieved links and of the user's satisfaction by using CAIA+Google compared to using only Google.

1. Introduction

While the Internet is a great medium for information and services, finding adequate results to a given query is a challenging task for most users. The Internet unlike any other computerized application covers diverse data structures, formats, and generalized contents [1]. The lack of uniform organization is significant because Internet users need to apply their own value system to scrutinize relevant results. Similarly, having multiple sources of information can be misleading and result in inaccurate decisions if those sources do not quite fulfill the needs of the user. Quite often a single search could display numerous results, which may impede the user in remembering the original details of a task [2] [3]. To solve this problem, we need to customize the Internet search engines to our need and interests. Customization of the Internet search engines is to carry out retrieval for each user incorporating his/her interests and retrieve more relevant information consistent with the user's intention. Autonomous, intelligent agents may prove to be the needed item in transforming passive search and retrieval engines into active, and personal assistants. Moreover, if those agents are collaborative, they will help the community's members to share what they found in previous searches.

Through the adoption of software agents, many attempts have been made to enable Internet users to quickly scan a variety of information. Dignum [4] further explained that Internet agents use built-in knowledge learned from the user or a process to accomplish a task. The use of Internet agents is often dispersed among search engine technology and approaches found on online services. Several new approaches in search engines are beginning to adopt intelligent techniques for collaborative search. For example, Choi and Yoo [5] have endorsed the use of neural networks to capture user's interest and to communicate this representation to other users by compiling knowledge bases and generalizations. Agent learning and social interaction are key issues independent of how the agent functionality assists the user in executing a task. Lewis [6] determined that unlike desktop applications, the user interaction with agents requires a higher cognitive effort to express activities or goals. Therefore, routine tasks to search queries on the Internet are not necessarily executed with routine knowledge since each task may not be identical to the previous one. Erickson [7] stressed that research is needed to understand how users react to agents, particularly in the area of user collaboration to increase user satisfaction. Nwana [8] posted a step in this direction by the need to conduct experiments using various agent learning techniques (fixed and evolutionary) over several domains. Under fixed agent

0-7695-3028-1/07 $25.00 © 2007 IEEE
DOI 10.1109/WI-IATW.2007.118

learning, the agent is static and cannot adapt to changes. Evolutionary agent learning, on the other hand, is ongoing and dynamic. Evolutionary learning is the adaptation of existing knowledge with new collective learning through the interaction of individuals over time [9] [10]. In this paper, we present a Collaborative Autonomous Interface Agent (CAIA) that collaborates with the Internet search engines and supports the user in finding exactly the information consistent with his/her interest.

The rest of this paper is organized as follows. Section 2 provides brief description and main objectives of CAIA system. Section 3 gives a brief study of current efforts and related work. An overview of CAIA is described Section 5 followed by the CAIA components. The implementation details are discussed in Section 6. Section 7 analyzes the performance of the CAIA system by showing the experimental results. Finally the paper is concluded in Section 8 with the benefits of CAIA and the future work.

2. Objective

The main objective is to develop a Collaborative Autonomous Interface Agent (CAIA) that personalizes the Internet search engines and supports communities of people in finding the relative information to their interest. CAIA uses data mining and machine learning techniques in order to learn and discover user's preferences. It resides in a user's machine for his/her trust of privacy since it monitors the user's explicit and implicit browsing behavior, communicates with the many Web portals search engines (the experiments done in this paper show the cooperation between CAIA and Google only) to retrieve relevant information that user needs. This research aims to:

(1) Improve the information retrieval performance of Internet search engines based on specified, measurable attributes.
(2) Develop an autonomous, intelligent agent that depends on Meta search engines and monitors the user's actions to prioritize the search results. The agent will learn based on the user's preferences and information content of the search results.
(3) Implement a method for the agent to learn the user's preferences during searching and browsing.
(4) Build a user's profile based on his/her preferences. It turns over time to reflect the current interest.
(5) Filter and refine the query entered by the user and then filter the retrieved information based on the current user's preferences.
(6) Collaborate with other CAIA agents for exchanging the search results.

3. CAIA Overview

When the user enters new keywords to search for, CAIA will refine the user's query then it will start the search process. After that, it will display the search results and it will issue a process to monitor the user's behaviors on the search results. So, there are three main processes as in the following:

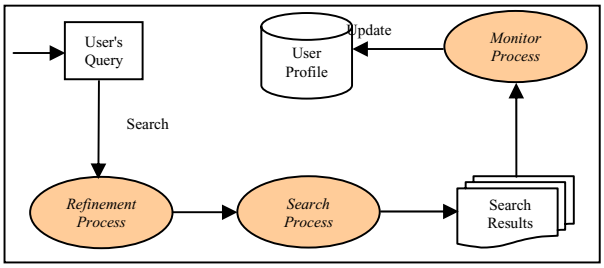

Figure1: Overview of CAIA

Refinement Process: CAIA starts refinement process immediately after the user enters new keywords. There are three main steps and two addition steps as following:

1. Filter the query by removing any noise words. Noise words are defined in a table in the database and can be grown based on the user's preferences.
2. Spell-check the query word by word and suggest the nearest words. Google API is used as spell checker.
3. Refine the query by checking the user profile (UP) if there are any relevant keywords to any word in the query. Then, CAIA will suggest new keywords to the user and give the flexibility to take or leave them.
4. Extra step done in first step in the search process when CAIA looks up in the UP. CAIA looks for thesaurus for any word in the query and then searches CAIA against those thesaurus words in addition to the original search to the user's keywords.
5. Extra step done in the second step in the search process when CAIA forward the user's query to the search engine. It looks for the most frequent word in the URLs of the relevant Websites found in first step of the search process. And then, suggest adding it as part of the URL when forwarding the query to the search engine. This will improve the performance of the search process.

Search Process: After refinement process, CAIA starts the search process. There are three steps, or sources, in the search process as following:

1. CAIA looks up in the UP, check for relevant Websites and they will be displayed if any was found.
2. CAIA will then forward the keywords to the Google search engine through a special APIs and retrieve relevant Websites and display them to the user.

3. CAIA also collaborates with other agents and forwards the user's keywords to the community. It will receive from them the relevant Websites if any and will display them to the user. Other agents will deal with any coming request with reserving the user's privacy so it will provide only the shared Websites.

Monitor User's Behaviors Process: When CAIA receives the results from different sources, it will re-order them based on the preferences in the UP. Then, it will monitor the user's behaviors in order to determine the Websites that s/he is interested in. The user's behaviors could be noticed explicitly by allowing him to choose specific value (between 0 to 1) for the Website s/he is reading, or implicitly by monitoring the user's actions that indicate his/her interest about it; like saving the Website, bookmaking, printing, and so on. All users' behaviors will be saved in the UP to be referred to by CAIA every time the user searches again.

4. Experimental Results

Precision Analysis: Precision is how much results related to the user, i.e. results that meet the user needs. It has a well-known equation. Precision is calculated by dividing the number of relevant URLs retrieved on the total number of URLs retrieved. Precision tells us how much improvements are the query expansion and filtration of the retrieved result. In this experiment, we submit twenty queries to Google and then submit the same queries to the CAIA with its advantages of enhancing the queries and filtering the results. The number of relevant URLs is assumed to be constant and the number of Google CAIA results was observed. So, the precision was calculated based on the total number of retrieved URLs and assume there are only twenty relevant URLs. Here is the result of first five queries.

Table1: Number of Results and Precision for 5 queries

Query	Number of Results		Precision	
	CAIA	Google	CAIA	Google
1	26	11600	76.92	0.17
2	27	12200	74.07	0.16
3	77	34100	25.97	0.06
4	68	30040	29.41	0.07
5	25	11250	80.00	0.18

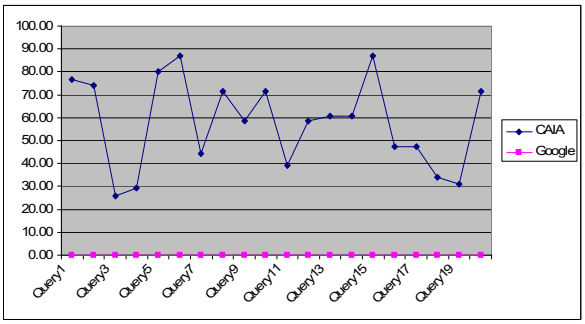

Figure 2: Chart of precision

Also, CAIA is improving over trials by decreasing number of results and hence increasing the precision. Next table and figure show the experiment of five queries with three trials and the improvement is clear.

Table2: Number of results and precision for three trials

Query	Number of Results			Precision		
	Trial 1	Trial 2	Trial 3	Trial 1	Trial 2	Trial 3
1	11600	302	26	0.17	6.62	76.92
2	12200	329	27	0.16	6.08	74.07
3	34100	2626	77	0.06	0.76	25.97
4	30040	2043	68	0.07	0.98	29.41
5	11250	281	25	0.18	7.12	80.00

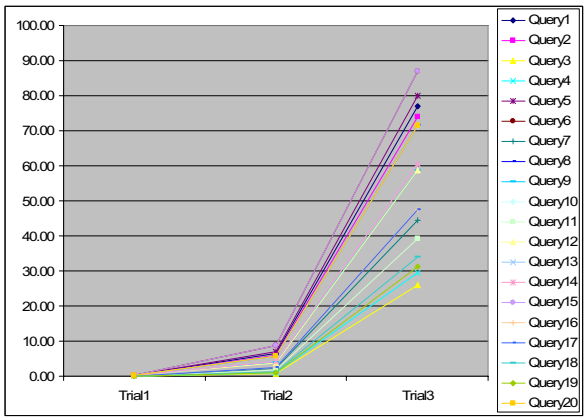

Figure 3: Chart shows the precision over trials

Adaptability Analysis: User Profile (UP) stores all URLs the user visits, and assigns keywords for each URL. We want to measure how CAIA is being able to adapt contents of the UP over time. In addition, we want to measure how good is the correlation between the keywords assigned to the URLs in the UP and the context of the URLs. We use the same formulas as in [11] that define a Fitness value to show the correlation between the weights of the URL's keywords calculated automatically by the CAIA

(*T)* and the weights of the URL's keywords calculated by the user (*S*), as follows:

Weight calculated by the user: $S_{URL} = \sum_{k=1}^{m} b_k . W_k$

Where W_k is the weight of attribute k, and b_k equals to 1 if the user judges the keyword k in the URL, as relevant for the context of the URL, otherwise b_k equals to zero.

Weight calculated by CAIA: $T_{URL} = \sum_{k=1}^{m} W_k$

Then, we define the Fitness value, which reflects the correlation between the two adaptations for URLj:
Fitness value: $F_j = S_j / T_j$

Fifty different queries have been entered. After frequent interactions of retrieval, we checked the correlation of each keyword with the context of the URLs in the UP and calculated S and T values, and then a Fitness value was calculated for the keywords of each URL in the UP. The following table and figures show the Fitness values as they are converging over time to each other. We conclude that CAIA is able to predict and adapt the URL's keywords to reflect well the context of the URLs in the UP over time. It is clear that CAIA needs to be used by the user over time and it then gives better correlation between the URL and the relevant keywords.

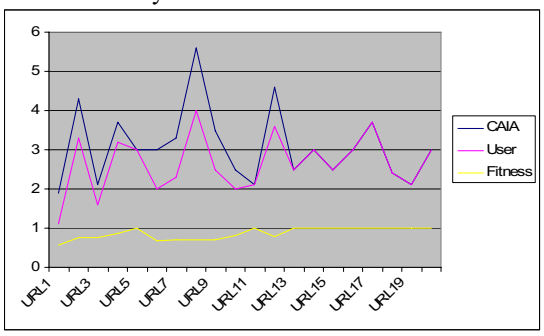

Figure 4: Calculated weights by the user and CAIA

5. Conclusion and Future Work

The paper is concluded with the benefits of CAIA. First, it is adaptable and customizable to the user preferences and this make it more personalized. The search history of the user is utilized for the future search so that only relevant information is retrieved. Second, CAIA has the ability to sense and react to the environment and socially communicate with other agents. This gives each agent access to a potentially vast body of experience that already exists. Over time each agent builds up a trust relationship with each of its peers analogous to the way we consult different experts for help in particular domains, with guaranteeing the privacy of the user. The collaboration increases the performance of the agent by improving the communication overhead. Finally, the precision, which is how much results related to the user, is improved using CAIA. The agent monitors the user's interactions and then improves and refines the next searches.

Although CAIA provides a great help in searching over the Internet, there are some future improvement areas. First, user behaviors' weight could be detected automatically and adopted based on the user reaction and the frequent user behaviors. Second, CAIA agents could be clustered based on the interests so that the related agents communicate effectively. Third, other search engines could be utilized in fining the search process and the results can be filtered and re-ordered based on the user profile. Finally, the user profile could be improved and enhanced by communicating with other applications like Outlook, an email application.

References

[1] K. Candan, H. Liu, and R. Suvarna, "Resource description framework: Metadata and its applications", SIGKDD Explorations, Volume 3, 2001, pp. 6-19.

[2] J. Lee, "Emerging challenges in cognitive ergonomics: Managing swarms of self-organizing agent-based automation", Theoretical Issues in Ergonomics Science, Volume 2, 2001, pp. 238-250.

[3] M. Eirinaki and M. Vazirgiannis, "Web mining for Web personalization", ACM Transactions on Internet Technology, Volume 3, 2003, pp. 1-27.

[4] G. Dignum, "Information management at a bank using agents: Theory and practice", Applied Artificial Intelligence, Volume 14, 2000, pp. 677-696.

[5] Y. Choi and S. Yoo, "Multi-agent learning approach to WWW information retrieval using neural networks", Proceedings of the Fourth International Conference on Intelligent User Interfaces, Redondo Beach, CA, 1998, pp. 23-30.

[6] M. Lewis, "Designing for human-agent interaction", AI Magazine, Volume 19, 1998, pp. 67-78.

[7] T. Erickson, "Designing agents as if people mattered," in Software Agents, J. Bradshaw, Ed, AAAI Press/The MIT Press, Menlo Park, CA, 1997, pp. 79-95.

[8] H. Nwana, "Software agents: An overview", Knowledge Engineering Review, Volume 11, 1996, pp. 1-40.

[9] W. Meng, C. Yu, and K. Liu, "Building efficient and effective Meta-search engines", ACM Computing Surveys, Volume 24, 2002, pp. 48-89.

[10] W. Lee and T. Tsai, "An interactive agent-based system for concept-based Web search", Expert Systems with Applications, Volume 24, 2003, pp. 365-363.

[11] T. Helmy, S. Amamiya, T. Mine and M. Amamiya, "A New Approach of the Collaborative User Interface Agents", Proceedings of IEEE/WIC/ACM International

Conference on Intelligent Agent Technology (IAT'03), pp. 147-153, October 13-17, 2003, (Halifax, Canada).