A Framework for Performance Evaluation of Intelligent Search Agents

Nurul I. Sarkar, Auckland University of Technology, Auckland 1142, New Zealand; E-mail: nurul.sarkar@aut.ac.nz
Naveed Iqbal, Auckland University of Technology, Auckland 1142, New Zealand

ABSTRACT
Because of the intelligent search agents (ISAs) can automate the process of searching through and evaluating reams of information on the Web, these agents are becoming increasingly popular in applications such as e-commerce and online businesses worldwide. While efforts have been made recently to develop various ISAs and multi-agents systems, very little is understood and known about the performance evaluation of such agents. A robust framework is required to assist efficient performance evaluation of ISAs. In this paper, we propose a framework for performance evaluation and comparison of ISAs. The proposed framework allows experimentation in which search agents can be evaluated based on a performance metric. This paper provides an in-depth performance evaluation of five selected ISAs to verify the robustness of the framework. The proposed framework is simple and can easily be adapted to applications requiring performance evaluation of similar intelligent agents.

Keywords: Intelligent search agents, performance evaluation, World Wide Web

INTRODUCTION
The competitive business environment and the growing complexity of work and personal lives create demands for performing many (often simultaneous) tasks more efficiently and promptly. To support these imperatives, new sophisticated and powerful intelligent software tools are developed. One such emerging software tool is the notion of an ISA (Boudriga & Obaidat, 2004).

ISAs are becoming increasingly popular in applications such as e-commerce and online businesses worldwide. This popularity results from the availability of various sophisticated and powerful intelligent agents that can automate the process of searching through and evaluating reams of information on the Web. For example, an ISA can be used to search on the Web to find a car matching a list of criteria, as tracking down the best price for purchasing grey Toyota Camry 2006.

A detailed discussion of intelligent agents, in general, can be found in (Boudriga & Obaidat, 2004; Dragan, 1998). Although many sophisticated and powerful intelligent agents and multi-agents systems have been developed in recently years, the performance evaluation of such agents is still an unexplored area in the field of agent technology (Saracevic, 1995; Srinivasan, Menczer, & Pantt, 2003).

This paper emphasizes that a good framework is required to assist efficient performance evaluation and comparison of ISAs. In summary, the main contributions of this paper are:

- We present a framework for performance evaluation and comparison of ISAs.
- We introduce a new performance metric called "search speed" for a better performance evaluation of ISAs.
- We provide an in-depth performance evaluation of five selected ISAs to verify the robustness of the proposed framework.

This research is exploratory in the sense that there was very limited prior research in the area of performance evaluation of ISAs to guide this research endeavour. Therefore, we adopted an empirical investigation methodology for the performance evaluation of ISAs (Dora, 2004; Serenko, 2006). The following research question was proposed:

How to select an ISA from a pool of search agents for a particular application?

The remainder of this paper is organized as follows. We first review past research on intelligent agents. We then introduce a framework for performance evaluation and comparison of ISAs. The experiment details and the performance evaluation of five selected ISAs are described. Experimental results are presented, and a brief conclusion ends the paper.

LITERATURE REVIEW
A number of recent studies demonstrate the usefulness and viability of using agent-based technologies in various applications; for example, in manufacturing automation (Heragu, Graves, Byung-In, & St Onge, 2002; Weiming, 2002), managing relational database (Rudowsky, Kulyba, Kunin, Ogarodnikov, & Raphan, 2005), network payment security (Hui-Zhang & Ji-di, 2004), e-business (Zi-Ming, Bo, & Yuan-Yuan, 2005), knowledge management (Houari & Far, 2004), fault diagnose system (Bo, Changhong, & Xiue, 2006), education (Pan & Hawrysiewycz, 2006), and web content filtering (Salter & Antonopoulos, 2006). In this section, we summarize previous work on performance evaluation of intelligent agents.

Mawlood-Yunis et al. (2004) experimented using two mobile agents and analyzed their platform's performance behaviour in distributed search. An interesting finding from their research is that in small networks multi agents systems do not provide any advantages and the associated overhead is high.

Camacho et al. (2002) evaluated various agent platforms (e.g., JADE, ZEUS, and SKELETON AGENT) based on the following metrics: number of agents used; number of requested documents; request time; and number of articles retrieved. Qi and Sun (2004) identified the operators that may increase the performance of a multi-agent system.

Lau and Zhang (2004) investigated how agents in a multi-agent system cooperate with each other. Results show that partial cooperation between agents is better than the fully cooperation of agents. Mao et al. (2004) point out that the coordination amongst the agents in a multi-agent system effects the performance of the system.

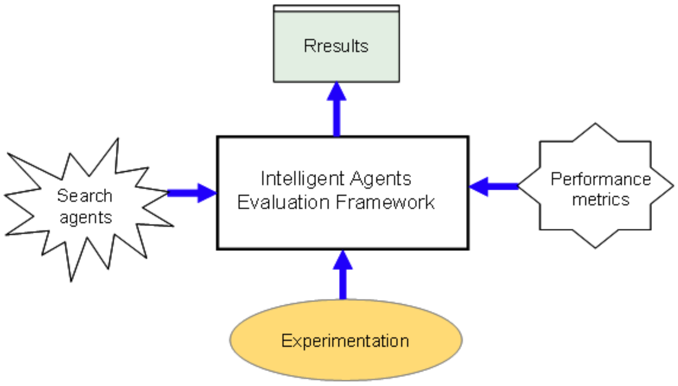
Samaras et al. (1999) stress the importance of performance benchmarking in the case of multi-agent distributed systems. O'Malley et al. (2000) examined the performance of static agents versus mobile multi-agent systems. The experiment involved searching for documents using static and multi-agent systems on a number of machines and then calculated the transmission control protocol/internet protocol (TCP/IP) traffic generated by the agents.

In summary, the proposed framework is almost alone in its goals and capabilities. The framework is described next.

THE PROPOSED FRAMEWORK
Figure 1 shows a simple framework for the performance evaluation of ISAs.

As seen in Fig. 1, the proposed framework allows experimentation in which search agents can be evaluated based on a performance metric. It is important to use an appropriate metric for the performance evaluation of ISAs. Once a performance metric is identified, the experiments can be conducted and results can be obtained using the proposed framework (Fig. 1).

Figure 1. The proposed framework



While various metrics for the performance evaluation of intelligent agents are reported in the software engineering literature, we briefly describe the following commonly used performance metrics:

Accuracy: In information retrieval, accuracy is measured in terms of recall and precision, where recall is relevance of the results obtained and precision is the cleanliness of the results (Kaki, 2004). This metric is not considered in our study since we did not have any human participants in the study who could rank the results obtained for accuracy measurement.

Throughput: This metric is commonly used for the performance measurement of communication networks. Throughput can be measured in bytes per second or packets per second or bits per second (bps). We ran all the experiments for the evaluation of ISAs under a control environment and measured the network throughput for every iteration performed, ensuring that the network throughput was not affecting the performance measures.

Round trip time: The round trip time (RTT) is the time required by a process to complete a task. For example, RTT is measured from the moment a query is placed on the source machine until the last packet arrives. The RTT can be affected by the size of a query.

Memory consumed: To measure the amount of computer memory consumed by a search agent, the MS Windows® default task manager utility can be used. This utility measures the memory usage dynamically when an application or process runs on the system.

Results returned: This parameter gives us the total number of results returned by an ISA per search.

Both “RTT” and “Results returned” can be used independently for performance comparison of ISAs. However, for simplicity and better performance comparison, it is useful to be able to compare the ISAs based on a single metric. Therefore, we define a new metric called “search speed” by combining both Results returned and RTT. The search speed (S_{speed}) is defined as

$$S_{speed} = \frac{N_{srr}}{T_{rtt}} \quad (1)$$

Where, N_{srr} is the total number of results returned in one search; and T_{rtt} is the round trip time (in seconds).

The search speed tells us about the capacity of an ISA (i.e., how many results returned by an ISA per second per query). An agent is said to be more powerful compared to another agent if it has a higher search speed.

EXPERIMENT DETAILS

The experimental setup consists of a desktop PC (Intel Pentium 4, 2.8 GHz; 512 MB RAM; 60 GB Hard disk; MS Windows XP Professional) and an ADSL modem (upload and download at 128 Kbps and 2 Mbps, respectively) linked to the Internet as shown in Fig. 2.

To measure the RTT and throughput, an open source software tool called Ethereal (www.ethereal.com) was used to captured packets from the live network and save them in a log for analysis.

Figure 2. Experimental setup

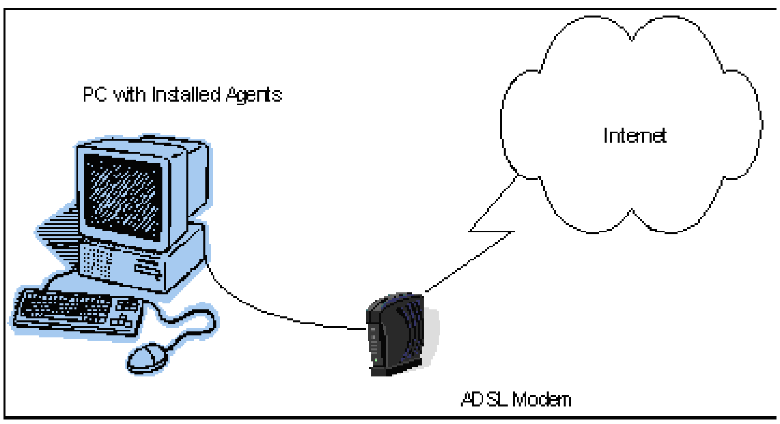


Table 1. Queries used in the experiments

Number of terms	Query
4	nicotine level smokeless tobacco
3	attention deficit disorder
2	dog crate

The selection of queries was very crucial in our experimentation. According to Jansen (2000) most of the web queries are only of two terms. A term can be a word, number, symbols, or even a URL. A query also contains one or more search terms and logical operators. In the experiments, we have used queries with 2, 3, and 4 terms as shown in Table 1 (Jansen, 2000).

Selection of ISAs

The selection of ISAs was a multifaceted and complex task since there were very limited number of ISAs available at free of costs for download and experiment in a control environment. The following five ISAs were selected for experimentation to test the robustness of the proposed framework:

FirstStop WebSearch™: This search agent provides a comprehensive search results and run under MS Windows. It can search on multiple search engines and websites concurrently.

WebSeeker™ 5.0: This agent has an indexer built into the program that keeps track of the browsed pages and monitors for new information. It has the ability to check whether the results can be reachable or not.

WebFoil 2.1: This agent provides multiple search sources at a time and search results provided through a detailed report.

Copernic Agent: This search agent can search multiple resources, combine the results, and stores the search history for later use.

WebFerret 5.0: This agent searches multiple sources, and sorts the results.

To facilitate an accurate data collection, it was vital to eliminate all the extraneous differences among the ISAs. Therefore, the advanced options provided by the ISAs were not selected in our experiments and only the default settings were used. To obtain an accurate and unbiased result, each query was submitted 10 times to each of the five ISAs in turn and all the measurements were recorded. This process was repeated at three different times of the day (at 0900, 1600, and 2100 hrs) for each of the five ISAs. In summary, we took the mean value for 30 observations per query. The experimental results are presented next.

RESULTS AND DISCUSSION

The experimental results for the performance comparison of five selected ISAs are summarized in Table 2. The ISAs are shown in the first column. The mean results returned by each ISA, and the corresponding mean RTT (in seconds) are shown in column 2 and 3, respectively. In column 4, we present mean throughput (bytes/second) which was measured for all iteration performed during experimentation. The mean search speed and the standard deviation are shown in the column 5 and 6, respectively. The 95% confidence interval (C.I) of the mean search speed is shown in the brackets.

As seen in Table 2, Copernic performs better (in terms of search speed) than the other four agents. The standard deviation of search speed varies significantly from its mean. We observe that the mean throughput performance of the ISAs differs slightly (from one another), but this variation is not very significant and does not affect the search speed.

We also found that as the complexity of the queries increases, ISAs tend to lose their search speed. For example, in the case when we used a quotation in the first query "nicotine level smokeless tobacco" to match the exact phrase, almost all five ISAs performed very low (results are not shown here due to the space limitation).

The mean search speeds of the ISAs are compared to see whether they differ significantly. The one-way ANOVA (also called F test) was carried out for the analysis of variance for the five ISAs under study. The F value after the ANOVA test was found to be 20.99, which is the ratio of the two variances: (1) the mean square between agents; and (2) the mean square within the agents. The degree of freedom was 4, which is the variation between the agents.

Figure 3 shows the difference (in magnitude) among the search speeds of the five ISAs (A = Copernic Agent, B = FirstStop, C = WebFerret, D = WebFoil, and E = WebSeeker). As seen in Fig. 3, Copernic has the highest search speed while the WebFoil has lowest.

LIMITATIONS OF THE STUDY

Firstly, a very limited number of ISAs was available (free of costs) for download and install on a local machine for experimentation. Due to the budget constraint we could not get hold of commercial agents to be included in our study. Another limitation was the time constraint as we had to complete the project within the given time frame.

CONCLUSION AND FUTURE WORK

In this paper, we proposed a framework for performance evaluation and comparison of intelligent search agents (ISAs). The framework allows experimentation in which search agents can be evaluated based on a performance metric. For a

Figure 3. Variation of search speed (magnitude) among the five ISAs

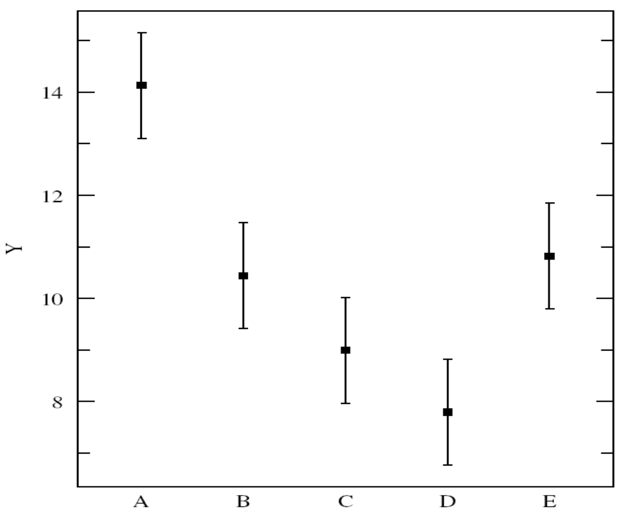


Table 2. Performance comparison of ISAs

Search agents	Mean results returned	Mean RTT (second)	Mean Throughput (bytes/second)	Search speed (results returned/second/query)	
				Mean (95% C.I)	Standard deviation
FirstStop	229.05	19.11	136190.83	10.40(9.41,11.46)	4.95
WebSeeker	218.62	19.40	125482.32	10.80(9.80,11.85)	6.11
WebFoil	166.22	21.07	128475.43	7.79(6.77,8.82)	4.35
Copernic	237.05	16.50	129326.87	14.10(13.10,15.15)	7.74
WebFerret	172.97	18.55	121421.73	8.99(7.97,10.02)	4.75

better performance comparison of ISAs, we introduced a new metric "search speed" by combing the number of results returned and the RTT. Through various experiments and measurements under a control environment, we gained an insight into the performance comparison of the ISAs by using search speed as one of the performance metrics.

We found that not all the five ISAs can perform equally well in terms of their search speed. For example, among the five ISAs evaluated, Copernic is up to 36.2% $(\frac{14.1-8.9}{14.1}\times 100\%)$ faster than the WebFoil.

The proposed framework is simple, easy to implement and can be used for performance evaluation of similar intelligent agents from other domains, such as online shopping. We have tested the framework and found to be robust.

This study can be further extended by using a case study in evaluating the performance of similar ISAs. Further research on the impact of query length and complexity (e.g., more terms using Boolean operators) on the performance of ISAs is suggested. The performance of search agents under different data types, such as text, images, and multimedia data is also planned as an extension of the present study.

REFERENCES

Bo, C., Changhong, W., & Xiue, G. (2006, January 19-21). *Research on the intelligent agent of distributed fault diagnose system.* Paper presented at the First International Symposium on Systems and Control in Aerospace and Astronautics. pp. 1-4.

Boudriga, N., & Obaidat, M. S. (2004). Intelligent agents on the Web: a review. *Computing in Science & Engineering [see also IEEE Computational Science and Engineering], 6*(4), 35-42.

Camacho, D., Aler, R., Castro, C., & Molina, J. M. (2002). *Performance evaluation of ZEUS, Jade, and SkeletonAgent frameworks*, IEEE International Conference on Systems, Man and Cybernetics. pp. 1-6.

Dora, J. (2004). *Evaluating reactive scheduling systems.* Paper presented at the IEEE/WIC/ACM International Conference on Intelligent Agent Technology (IAT 2004). pp. 458-461.

Dragan, R. (1998, June 9). Software Agents. *PC Magazine*, 166.

Heragu, S. S., Graves, R. J., Byung-In, K., & St Onge, A. (2002). Intelligent agent based framework for manufacturing systems control. *IEEE Transactions on Systems, Man and Cybernetics, Part A, 32*(5), 560-573.

Houari, N., & Far, B. H. (2004, August 16-17). *Application of intelligent agent technology for knowledge management integration.* Paper presented at the Third IEEE International Conference on Cognitive Informatics. pp. 240-249.

Hui-Zhang, S., & Ji-di, Z. (2004, November 16-18). *Application of intelligent agent in network payment security.* Paper presented at the 29th Annual IEEE International conference on Local Computer Networks. pp. 437-440.

Jansen, B. J. (2000). The effects of query complexity on Web searching results. *Information Research, 6*(1), 30-54.

Kaki, M. (2004). *Proportional search interface usability measures.* Paper presented at the Proportional search interface usability measures, Tampere, Finland. pp. 365-372.

Lau, H. C., & Zhang, L. (2004). *A two-level framework for coalition formation via optimization and agent negotiation.* Paper presented at the IEEE/WIC/ACM International Conference on Intelligent Agent Technology (IAT 2004). pp. 441-445.

Mao, X., Wang, J., & Yu, E. (2004). *A framework to specify and evaluate coordination in MAS with logic.* Paper presented at the IEEE/WIC/ACM International Conference on Intelligent Agent Technology (IAT 2004). pp. 325-328.

Mawlood-Yunis, A., Nayak, A., Nussbaum, D., & Santoro, N. (2004). *Comparing performance of two mobile agent platforms in distributed search.* Paper presented at the IEEE/WIC/ACM International Conference on Intelligent Agent Technology (IAT 2004). pp. 425-428.

O'Malley, S. A., Self, A. L., & Deloach, S. A. (2000). *Comparing performance of static versus mobile multiagent systems.* Paper presented at the IEEE National Aerospace and Electronics Conference (NAECON 2000). pp. 282-289.

Pan, W., & Hawrysiewycz, I. (2006). Assisting learners to dynamically adjust learning process through software agents. *International Journal of Intelligent Information Technologies, 2*(2), 1-15.

Qi, D., & Sun, R. (2004). *A comparison of team evolution operators.* Paper presented at the IEEE/WIC/ACM International Conference on Intelligent Agent Technology (IAT 2004). pp. 369-372.

Rudowsky, I., Kulyba, O., Kunin, M., Ogarodnikov, D., & Raphan, T. (2005, April 4-6). *Managing a relational database with intelligent agents.* Paper presented at the International Conference on Information Technology: Coding and Computing. pp. 238-242.

Salter, J., & Antonopoulos, N. (2006). CinemaScreen recommender agent: combining collaborative and content-based filtering. *IEEE Intelligent Systems [see also IEEE Intelligent Systems and Their Applications], 21*(1), 35-41.

Samaras, G., Dikaiakos, M. D., Spyrou, C., & Liverdos, A. (1999). *Mobile agent platforms for Web databases: a qualitative and quantitative assessment.* . Paper presented at the First International Symposium on Agent Systems and Applications, and Third International Symposium on Mobile Agents. pp. 50-64.

Saracevic, T. (1995). *Evaluation of Evaluation in Information Retrieval.* Paper presented at the ACM SIGIR, Seattle, USA. pp. 138-146.

Serenko, A. (2006). Importance of interface agent characteristics from end-user perspective. *International Journal of Intelligent Information Technologies, 2*(2), 49-60.

Srinivasan, P., Menczer, F., & Pantt, G. (2003). *Defining evaluation methodologies for topical crawlers.* Paper presented at the ACM SIGIR, Toronto, Canada. pp.

Weiming, S. (2002). Distributed manufacturing scheduling using intelligent agents. *IEEE Intelligent Systems [see also IEEE Intelligent Systems and Their Applications], 17*(1), 88-94.

Zi-Ming, Z., Bo, M., & Yuan-Yuan, Z. (2005, August 18-21). *An Intelligent Agent-Based System in Internet Commerce.* Paper presented at the International Conference on Machine Learning and Cybernetics. pp. 299-303.