ICEB 2001 Proceedings
International Conference on Electronic Business (ICEB)

Winter 12-19-2001

A design of Personal Information Push-Delivery System on the Internet

Chentung Chen

Wei-Shen Tai

A DESIGN OF PERSONAL INFORMATION PUSH-DELIVERY SYSTEM ON THE INTERNET

Chen-Tung Chen and Wei-Shen Tai
Department of Information Management, Da-Yeh University,
112, Shan-Jiau Road, Da-Tsuen, Changhua, Taiwan, R.O.C
TEL: 886-4-8528469 Ext.3130 FAX: 886-4-8525982
EMAIL: chtung@mail.dyu.edu.tw

ABSTRACT

The Internet provides a powerful disseminative ability for users to acquire information more efficiently and quickly. However, an increasingly large scale of data induces certain problems as users face a more serious information overload situation. By using an information retrieval technique, information push-delivery provides a good solution for users to acquire rich information from the Internet. In fact, providing personal service for users is one of the most important issues in an electronic commerce (EC) environment. In order to increase interaction between themselves and customers, many enterprises provide personal services to improve management performance and competitiveness. However, since the customers have different preferences for information received from the Internet, it seems necessary to design a personal information system to guarantee that the customers can receive the desired information. In this study, the fuzzy retrieval and similarity measurement techniques are applied to design a personal information push-delivery system. The data resulting from testing a group of students at Da-Yeh University, Changhua, Taiwan, shows that the satisfaction degree for the received information for all participants was 70%. These results indicate that the proposed system can effectively provide correct and interesting information to users.

Key words: Information push-delivery, Information retrieval, Personal services, Similarity measurement.

1. INTRODUCTION

The Internet, with continuous progress in computer technology and the pervasion of network environments, has become a major channel for people to acquire information on demand by providing access to hundreds of software programs, documents, sounds, images and many other types of information [6]. It has been estimated that the amount of information stored on the Internet doubles every 18 months. However, the rate of increase in the number of home pages is even faster, doubling every six months or even sooner [32].

Users can acquire increasing amounts of information from Internet resources, but they also face the serious problem of information overload. Concurrently, web tools have become more important for developing better solutions to assist people in acquiring information easily and effectively [1]. Some web tools possess the ability to track information and changes within users' surfing on the Internet. Further more, by using these surfing data, web tools can assist users in acquiring and managing information or documents from the Internet. According to their intelligence and power, such tools can be divided into browsers, search engines, notifications, deductions and push-delivery systems [11].

The search engine is the most popular tool for locating and managing information on the Internet. However many unpolished data appear in the result obtained by using keywords to locate information. Users must refine and retrieve information that they actually need after utilizing a search engine. Such problems of information overload on the Internet have become increasingly serious [10][12]. An intelligent information agent not only provides the function of information retrieval and a recommendation for improving the drawbacks of a search engine, but also provides a profile to record the user's personal information. So it can build an individual push-delivery mode for each user to solve information overload effectively [10][11][14][23][30].

Recently, electronic commerce (EC) has become the most popular issue on the Internet. In addition to a swift growth in customers and the dissemination ability to provide services or information through time and space, a growing number of enterprises are acquiring customer information from the Internet. They expect to increase interaction with customers and improve their service level at the same time [20]. In other words, using the Internet to provide individual service and information to meet customers' preferences and requirements is the most important issue in an electronic commerce environment. Thus, we present personal information push-delivery architecture, and then design a system to demonstrate its performance. The results from testing at Da-Yeh University, Changhua, Taiwan, show that the degree of satisfaction from all participants reached 70%, thereby indicating that the proposed system can effectively provide information that meet users' requirements.

2. REVIEW OF THE LITERATURE

To obtain information meeting users' requirements is the main goal of web tools. Many methods and algorithms have been proposed to improve the information retrieval efficacy of the Internet, but we believe our proposed system is superior. In this section, some related work on push-delivery concepts and methods are reviewed.

2.1. Information push-delivery agents

Information push-delivery agents can be divided into three major types on the basis of their architectural and operational procedures. According to their characteristics, an introduction and comparison follow.

(1) Client

A client can analyze a user's track while surfing the web without a server and determinate the user's access pattern automatically. After acquiring information such as a hyperlink trend or possible target, a search engine or other tools are used to locate related information on the Internet. Then some web pages or hyperlinks are recommended, those tools such as Letizia, WebWatcher, etc [4][10][11] [21][32].

(2) Server

A server is used to collect a user's track while surfing the web, in order to provide user-related information from a database [10][32]. Fuzzy clustering or other social-filter methods are usually used to analyze the user's Travel Path Graph (TPG) or the content of documents to determinate the user's cluster of similar preferences. Then related information or web pages are recommended to the members of a common cluster. Some severs use a database or search agent to locate related information and provide recommended information, those tools such as GroupLens, Ringo, etc. [4][11][18][22][26].

(3) Virtual proxy server

The per-fetch concept is used to build a virtual hierarchy proxy server, which collects similar contents from documents or data into different container levels. The Global Resource Index (GRI) is used to manage the location of information in a proxy server. When a user becomes a member of a special proxy server community, he/she can acquire related information from other members in the same community [2][28].

Table 1. The functional comparison of information agent

Item \ Type	Client	Server	Virtual proxy server
Using server	No	Yes	No
User community	No	Yes	No
Information category	No	No	Yes
Topical category definition	No	No	No
Push by category	No	No	Yes
Hierarchy structure	No	No	Yes

The recommended information restricted in the results of search engines locate on the Internet, couldn't provide related information from other users, is a prime drawback in the usage of a client agent. For all, a server agent can provide related information from other members in the same community. Fuzzy clustering method is used to build several user sets with similar attributes in a server agent. Even though a user set has almost same attributes, still no category topic could represent what it is. A virtual proxy hierarchy structure is used in a virtual proxy sever agent to replace social clustering method in a server agent. But each proxy server merely collect the most same documents still couldn't clarify the category topic in it. The functional comparison of information agent is shown in table 1. For the sake of improving those drawbacks of information push-delivery systems, we advocate a topical category concept to build an information push-delivery system. Both fuzzy information retrieval and fuzzy similarity measurement are used to classify topical category of documents and provide appropriate recommended information to user.

2.2. Fuzzy information retrieval

According to a user's prefer topics to locate appropriate documents in word representation is the definition of the information retrieval [19]. Term is usually used in automatic information retrieval to be the index of a documents set. In order to discriminate the important degree of terms in a documents set, the weight value is used to describe each term's important level in each document. There are three factors should be considered in the process of determining the term weight value. Those are term frequency in the document, prevalence of the term in the entire collection and the length of the document. And two key stages should be talked about; extracting term from the document and term weight adjustment, in the information retrieval processes [27].

(1) Extracting terms from the document and obtaining its initial weight value

Filtering and selecting terms related topic from each textual document, the process of extracting keywords and determining its weight value described as follow [4][11][12] [29][32].

i. Filtering out the duplicated documents, not-existed web pages or documents on the Internet.
ii. Using stop-word list to delete the words, which couldn't represent the content of the document.
iii. Counting the frequency of core words.
iv. Extracting keywords from core words by the frequency order.
v. Obtaining the initial weight value of each keyword by the normalized frequency.

(2) Term weight adjustment

Term weight adjustment is one of the most important processes in the information retrieval method. If a term has higher frequency in a document and lower frequency in the other documents, it should possess higher weight value in the document. Using the concept, Term Frequency Inversed Documents Frequency (TFIDF) has advocated in an algorithm to adjust a term's weight value in a documents set. There are many similar algorithms such as Okapi, INQUERY and Smart has proposed recently [3][5][10][24] [25][27].

2.3. Fuzzy similarity measurement

Both keyword and its weight value are used to compose a term vector. After a document translates to a term vector, every document can compare with each topical category and classified into appropriate category. There are many fuzzy similarity measurement methods include [8]:

(1) On the Basis of the operation of union and intersection.
(2) On the Basis of the maximum difference.
(3) On the Basis of the differences and the sum of degree of membership.

Some drawbacks maybe appear after both documents similarity measuring like [7][8][9][13][15][16][17][31].

(1) When the same elements in both documents have the same degree of membership overall, the degree of similarity should be one. In some formulas, the degree of similarity will be the highest degree of membership in the same elements seems unreasonable obviously.
(2) In some formulas, if one pair of the same elements whose degree of membership is one in both documents, even the other same element pairs have obvious difference, the degree of similarity become one. Obviously, the result is un-reasonable and un-acceptable to us.
(3) Using the maximum of difference method, when one of the same element pairs appears obvious difference in degree of membership, even the degree of membership of others element pairs are nearly equal, the degree of similarity will still be excessively low.
(4) Using the operation of complement method, when one of element in both documents that's the degree of membership is one, then the degree of similarity will be zero. Obviously, the result is un-reasonable and un-acceptable also.

We use Microsoft Excel generator to build 60 random testing data sets, in order to make a comparison in such similarity measurement methods. After testing of similarity measurement, the absolute value of the difference between each formula and the average degree of similarity of all formulas is used to evaluate the similarity measurement effect of each formula. According to the difference between average value and degree of similarity, the similarity measurement in Chen [7] should be the better one and its formula is described as

$$S(A,B) = 1 - \sqrt{\frac{\sum_{i=1}^{n}(m_A(x_i) - m_B(x_i))^2}{n}}$$

where $\mu_A(x_i)$ and $\mu_B(x_i)$ are the degree of membership of the element x_i in A and B sets respectively. The symbol $S(A,B)$ is the degree of the similarity between A and B sets. n is the number of elements in a set.

3. SYSTEM STRUCTURE

3.1. Component introduction

According to the related literature of information push-delivery agents, we propose architecture to design an active personal information push-delivery system (shown in figure 1). The major components of personal information push-delivery system are described as follow:

(1) Raw data and information

The resources of push-delivery information could be web pages on the Internet or other types of data like goods information from advertising agency. All information should be translated to plain text, and extracted or managed in the operational processes.

(2) Document database

After information translated to plain text, filtering out unsuitable information by date, length or other conditions. The data meet restricted condition would be stored to document database.

(3) Extracting term

Extracting terms in each document, then computing the weight value of each term.

(4) Category knowledge database

Constructing the category domain knowledge according to the key term vector set, and then building the centric set for each category of domain knowledge. Using the centric set to provide similarity measuring with each document.

(5) Similarity measurement

Using fuzzy similarity measurement to measure the similarity between the centric set and each document. And then categorize the document into push-delivery database by its degree of similarity of each category.

(6) Push-delivery database

Storing the document after similarity measured to push-delivery database by category. According to the similarity of both each document and each category, the proposed system will provide users the information, which meets their preferences.

(7) Information push-delivery

According to conditions such as the threshold values or the maximum number of push-delivery documents, this push-delivery information system will send each user the documents by his/her preferences respectively.

(8) User database

User could register his/her profile into user database such as account, password, name and education background. The proposed system will provide some categories for user to fill his/her preference.

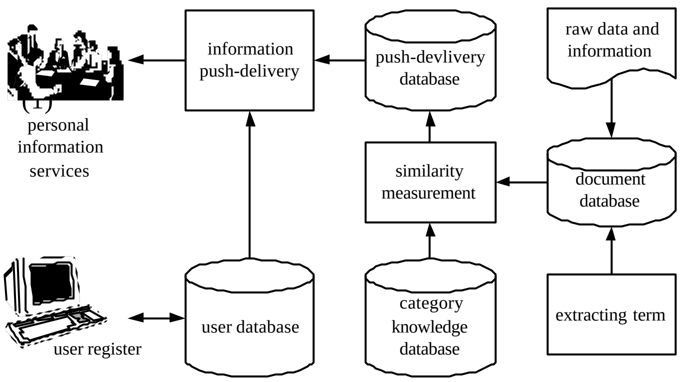

Figure 1. Architecture of personal information push-delivery system

3.2. Satisfaction analysis

In order to indicate the proposed system can provide users the personal information to meet users requirement effectively, we present two indices, response ratio and degree of satisfaction of push-delivery information in a period, to describe the effectiveness of the proposed system. In this study, user can represent his/her satisfaction with semantic variable or score. The semantic variables can be converted into score of satisfaction by mapping table (shown in table 2). The two indices are described as follow.

(1) Response ratio

When response ratio of the information is higher, it represents users are more interested in the information system push-delivery obviously. The response ratio can be computed as

$$R_i = \frac{r_i}{s_i}$$

where R_i is the response ratio of the i-th push-delivery information, r_i is the response number of users when they received i-th information, and s_i is the number of users who received i-th push-delivery information.

(2) Degree of satisfaction of push-delivery information in a period

In a period, the degree of satisfaction of push-delivery information in each category can be computed as

$$S_c = \frac{\sum_{i=1}^{n}\sum_{t=1}^{T} S_{cti}}{n * T_c * E_{\max}}$$

where S_c ($0 \leqq S_c \leqq 1$) is the degree of stratification of c-th category, S_{cti} is the i-th user's score of satisfaction of the t-th information in the c-th category, n is number of users in the c-th category, T_c is number of push-delivery information for the c-th category in a period, E_{max} is the maximum value in the semantic translation mapping table (in this case, E_{max}=5).

Table 2. Score of semantic variable mapping

Semantic variable	VS	SA	NR	US	VU
Score	5	4	3	2	1

VS: Very satisfying, SA: Satisfying, NR: Normal, US: Unsatisfying, VU: Very unsatisfying.

4. TEST AND RESULT

In this paper, we design a personal information push-delivery system to indicate the proposed architecture is feasible and effective.

4.1. System design

Our system divided into web site and main system. The functions of the proposed system can be described as follow.

(1) Web site

In this web site, the "Da-Yeh University information center" we named, includes main page, member region, questionnaires and message board (shown in figure 2 and figure 3).

(2) Main system

Main system is in charge with information maintenance and management about entire system. The functions of main system include member profile, topical category, and questionnaires, information push-delivery and satisfaction measurement (shown in figure 4 ~ figure 9).

4.2. Test process and result

(2)

In this test, our participants are the students of Da-Yen University, Changhua, Taiwan. The topical categories are divided into "internet", "computer science", "wireless communication" and "electronic commerce". The test procedure is described as follow (shown in figure 10).

(3

Figure 2. Main web page

Figure 3. Member region

Figure 4. Main system screen

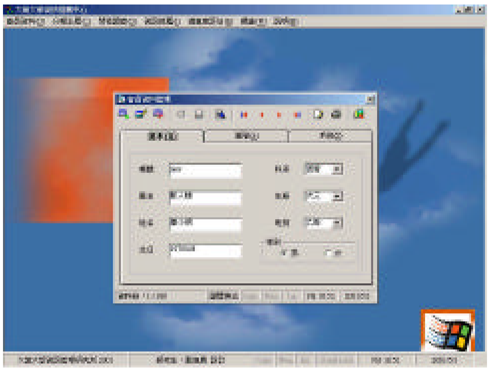

Figure 5. Member profile

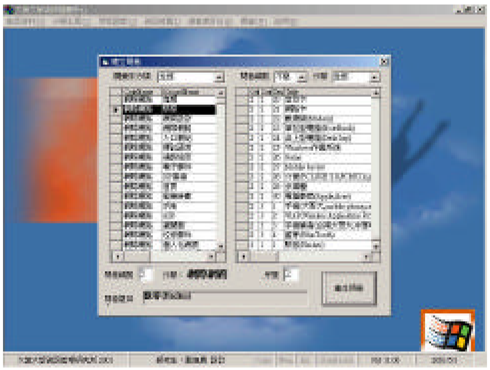

Figure 6. Questionnaire form

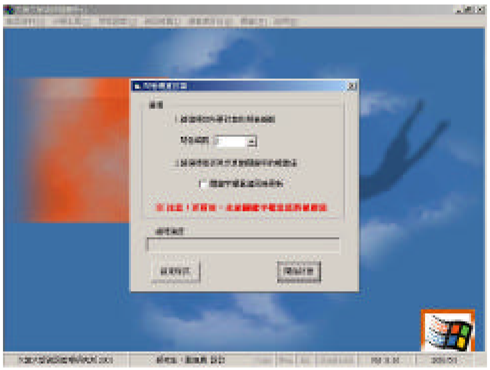

Figure 7. Calculation of term weight

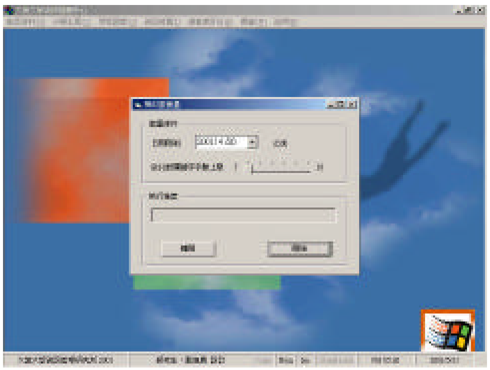

Figure 8. Similarity measurement

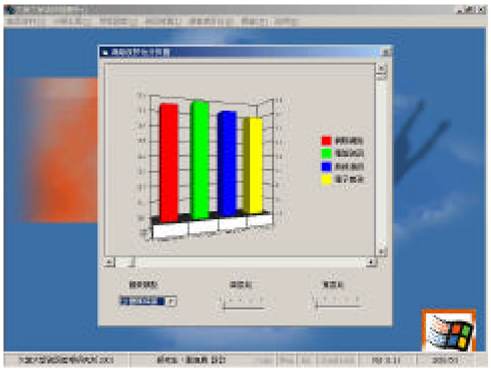

Figure 9. Graph of degree of satisfaction

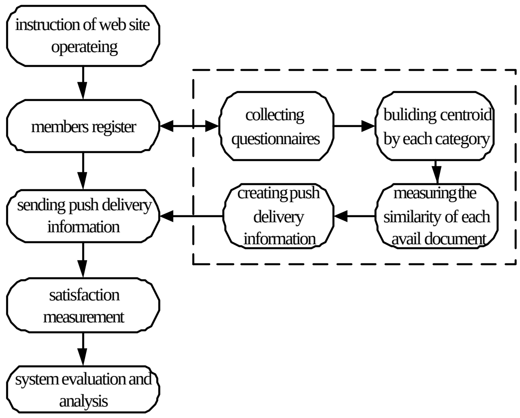

Figure 10. The process of information push-delivery

Step 1: Members register and the first investigation

All participants should register their profile into user database on the Internet. And then, they must fill out the keyword importance questionnaire and also propose some keywords, which they feel important but the questionnaire doesn't appear. According to the collection from the participants' response, the main system translates semantic variable into score by mapping table (shown in table 3) to calculate the weight value of each keyword (term). The weight value of each keyword (term) can be described as

$$w_i = \frac{\sum_{k=1}^{5} U_k * V_k}{U_T * V_{\max}}$$

where w_i is the importance weight value of i-th term, U_k is the number of users in the k-th semantic variable, V_k is the score of k-th semantic variable mapping, U_T is the number of users, V_{max} is the maximum of score of semantic variable mapping (in this case, V_{max} = 5)。 The weight value of each category is shown in table 4.

Table 3. Score of semantic variable mapping

Semantic variable	VI	IM	NR	UI	VU
Score	5	4	3	2	1

VI: Very important, IM: Important, NR: Normal, UI: Unimportant, VU: Very unimportant.

Table 4. Result of the first investigation

Weight / Category	Maximum	Minimum	Average	Standard deviation
Internet	0.938	0.382	0.739	0.126
Computer science	0.946	0.596	0.781	9.716
Wireless communication	0.882	0.548	0.721	0.101
Electronic commerce	0.972	0.574	0.808	9.210

Step 2:Aggregating proposed keywords and the second investigation

After filtering out the keywords of all participants proposed, the main system will aggregate the existed keywords and the new keywords of participants proposed to organize a new keywords list in each category. Then select some more important keywords in each category to form a new questionnaire. Repeat the process of step 1, the weight value in each category can be computed and shown in table 5.

Table 5. Result of the second investigation

Weight / Category	Maximum	Minimum	Average	Standard deviation
Internet	0.949	0.646	0.802	0.085
Computer science	0.949	0.669	0.804	7.739
Wireless communication	0.903	0.586	0.775	0.079
Electronic commerce	0.968	0.701	0.844	5.366

Step3:Information push-delivery and users satisfaction measurement

After filtering out several electronic newspapers on the Internet, we obtain 2275 avail-data samples in this test. Using similarity measurement methods, the proposed system can push users the information by conditions such as the maximum number of document, the threshold of degree of similarity and the expired published date. According to the collection of the feedback from each participant's score of satisfaction after the information, which he/she has read, the final degree of satisfaction of the proposed system is 0.709 (shown in table 6). In other words, the degree of satisfaction of all participants in this test is quite high. This result indicate that the proposed system is feasible and acceptable.

Table 6. Result of the user satisfaction measurement

Weight / Category	Maximum	Minimum	Average	Standard deviation
Internet	0.824	0.610	0.717	0.064
Computer science	0.818	0.637	0.747	0.069
Wireless communication	0.803	0.563	0.701	0.069
Electronic commerce	0.786	0.552	0.671	0.077

5. CONCLUSIONS

In this paper, we propose information push-delivery system architecture and design a personal information push-delivery system based on knowledge base. According to the result of testing investigation, it has indicated that the degree of satisfaction of participants acquire information from information push-delivery system is quite high. It means that the personal information push-delivery system can provide users personal information on the Internet more feasible and effective.

In an electronic commerce environment with increasing competitors, a growing number of enterprises emphasize to build the closeness relationship between themselves and customers. The personal service is just a best tool to provide information or services and meet their expectancy in increasing interaction with customers. Using the architecture of personal information push-delivery system would provide a better efficient solution to bring up interaction in an electronic commerce environment.

6. ACKNOWLEDGEMENTS

The authors gratefully acknowledge the financial support of National Science Council, Taiwan, R. O. C. under project number NSC 90-2218-E-212-009.

7. REFERENCE

[1] Adorf, H. M., "Resource Discovery on the Internet," *Vistas in Astronomy*, 1995, 39 (2), 243-253.

[2] Bogonikolos N., D. Fragoudis and S. Likothanassis, "ARCHIMIDES: an intelligent agent for adaptive-personalized navigation with WEB server," *System Science, Processing of the 32nd Annual Hawaii International Conference*, 1999, 1-9.

[3] Broglio, J., J. P. Callan, W. B. Croft and D. W. Nachbar, "Documents retrieval and routing using the INQUERY system," *Proceedings of the Third Text Retrieval Conference, NIST Special Publication*, 1995, 29-38.

[4] Bum, R. Y., J. Y. Jae and D. K. Soo, "COPEN: a COBRA-based intelligent push-engine," *Software Engineering Conference in Asia Pacific*, 1998,

330-337.
[5] Buckley, C., J. Allan, G. Salton, and A. Singhal, "Automatic query expansion using SMART," *Proceedings of the Third Text Retrieval Conference, NIST Special Publication*, 1995, 69-80.
[6] Bowman, C. M., P. B. Danzig, U. Manber and M. F. Schwartz, "Scalable Internet resource discovery research problems and approaches," *Communications of the ACM*, 1994, 37 (8), 98-107.
[7] Chen, C. T., "Extensions of the TOPSIS for group decision-making under fuzzy environment," *Fuzzy Sets and Systems*, 2000, 114 (1), 1-9.
[8] Chen, S. M., M. S. Yeh and P. Y. Hsiao, "A comparison of similarity measures of fuzzy values," *Fuzzy Sets and Systems*, 1995, 72(1), 79-89.
[9] Chen, S.M., "A new approach to handing fuzzy decision making problems," *IEEE Trans. Systems, Man, Cybernetics*, 1988, 18, 1012-1016.
[10] Chen, H. C., Y. M. Chung and C. C. Yang, "An intelligent personal spider (agent) for dynamic Internet/Intranet searching," *Decision Support Systems*, 1998, 23(1), 41-58.
[11] Cheung, D. W., B. Kao and J. Lee, "Discovering user access pattern on the World Wide Web," *Knowledge-Based Systems*, 1998, 10, 463-470.
[12] Fragouidis D. and S. D. Likothanassis, "Retriever: a self-training agent for intelligent information discovery," *Information intelligent and Systems International Conference*, 1999, 594-599.
[13] Gerstenkorn, T. and J. Man' ko, "Correlation if intuitionistic fuzzy sets," *Fuzzy Sets and Systems*, 1991, 44, 39-43.
[14] Hal B., "Cyberspace 2000 dealing with information overload," *Communications of the ACM*, 1997, 40(2), 19-24.
[15] Hyung, L. K., Y. S. Song and K.M. Lee, "Similarity measure between fuzzy sets and between fuzzy elements," *Fuzzy Sets and Systems*, 1994, 62, 291-293.
[16] Jiulun, F. and W., Xie, "Some notes on similarity measure and proximity measure," *Fuzzy Sets and Systems*, 1999, 101(3), 403-412.
[17] Kazem, S. Z., "Fuzzy genomes," *Artificial Intelligence in Medicine*, 2000, 18(1), 1–28.
[18] Kim, J. G. and E. S. Lee, "Intelligent information recommend system on the Internet," *Parallel Processing International workshop*, 1999, 376-380.
[19] Klir, G. J. and B. Yuan, *Fuzzy sets and fuzzy logic: theory and applications*, Prentice Hall, 1995.
[20] Levy, A. Y. and M. C. Rousset, "Verification of knowledge bases based on containment checking," *Artificial Intelligence*, 1998, 101(1-2), 227-250.
[21] Matsuura S., J. Ozawa, S. Araki and T. Imanaka, "An extension of ECA architecture and its application to HTML document browsing," *Systems, Man, and Cybernetics IEEE International Conference*, 1999, 1, 738-743.
[22] Mattthew C., K. Rodden and D. Brodbeck, "The order of things: activity-centered information access," *Computer Network and ISDN Systems*, 1998, 30, 359-367.
[23] Ouderkirk, J. P., "Technical services task assignment: from macros to collection management intelligent agents," *The Journal of Academic Librarianship*, 1999 25(5), 397-401.
[24] Robertson, S. E., S. Walker, S. Jones, M. M. Hancock-Beaulieu, and M. Gatford, "Okapi," *Proceedings of the Third Text Retrieval Conference, NIST Special Publication*, 1995, 109-126.
[25] Salton, G., "Another Look At Automatic Text-Retrieval Systems," *Communications of the ACM*, 1986, 29(7), 648-656.
[26] Seoyoung P. and C. Wu, "Intelligent search agent for software components," *Software Engineering Conference Sixth Asia Pacific*, 1999, 154-161.
[27] Singhal, A, G. Salton and C. Buckley, "Document Length Normalization," *Information Processing & Management*, 1996, 32(5), 619-633.
[28] Sun W. and C. C. Liao, "Virtual proxy server for WWW and the intelligent agent on the Internet," *System Science, Processing of the 30th Annual Hawaii International Conference*, 1997, 4, 200-209.
[29] Tak, W. Y, J., Mattew, G. M., Hector and D., Umeshwar, "From user access pattern to dynamic hypertext linking," *Computer Network and ISDN Systems*, 1996, 28, 1007-1014.
[30] Tu, H. C. and H. Jieh, "An architecture and category knowledge for intelligent information retrieval agents," *Decision Support Systems*, 2000, 28(3), 255-268.
[31] Wang, W. J., "New similarity measures on fuzzy sets and on elements," *Fuzzy Sets and Systems*, 1997, 85(3), 305-309.
[32] Yang, C. C., J. Yen and H. C. Chen, "Intelligent internet searching agent based on hybrid simulated annealing," *Decision Support Systems*, 2000, 28(3), 269-277.