Journal of Universal Computer Science, vol. 14, no. 14 (2008), 2400-2415
submitted: 30/9/07, accepted: 30/4/08, appeared: 28/7/08 © *J.UCS*

An Application of Meta Search Agent System Based on Semantized Tags for Enhanced Web Searching[1]

Chonggun Kim
(Yeungnam University, Korea
cgkim@yu.ac.kr)

Jae-Youn Jung
(Yeungnam University, Korea
e-mail@ynu.ac.kr)

Hyeon-Cheol Zin
(Yeungnam University, Korea
cyberzin@ynu.ac.kr)

Jason J. Jung[2]
(Yeungnam University, Korea
j2jung@intelligent.pe.kr)

Abstract: Web searching techniques have been investigated and implemented in many aspects. Particularly, in case of personalization, more important issue is how to manipulate the results retrieved from search engines for better user understandability and satisfaction. Such manipulation processes are i) ranking the results in accordance with user relevance, and ii) exchanging the results between users who have similar tastes. Thus, our work has been mainly focusing on relevance-based ranking mechanism as well as sharing schemes for the results retrieved from heterogeneous web information sources. In this paper, we propose a hybrid model for meta search agent systems with three main functionalities, i.e., i) URL filtering method for preprocessing, ii) tag-based information conceptualization scheme for ranking, and iii) ontology-based standardization scheme for sharing. It means that the proposed meta search agent model exploits semantized tags to formalize and share heterogeneous information obtained from multiple search engines and to finally maintain the shared information. Within the tag-based information space, a conceptual distance between retrieval interest and search results can be efficiently computed. By conducting some experimentations, we have shown the semantized tag model can conceptualize the retrieved results, and make them sharable. We also compare performance of the proposed system with hyperlink-based methodologies.

Key Words: Ontology; Semantiization; Information searching; Meta searching;
Category: H.1.1, H.3.5, I.2.11

1 Introduction

There have been various ways of managing and publishing information through newly emerging web applications and environments. These information spaces and applica-

[1] This paper is significantly revised from an earlier version presented at The 2nd KES International Symposium on Agent and Multi-Agent Systems: Technologies and Applications (KES-AMSTA 2008) in March 2008.

[2] Corresponding author

tions on the web are strongly motivating several important issues on information searching and delivering services for web users. Especially, it is crucial to understand semantics and contexts of the information located in the web applications [Jung et al. 2007;Sliwko and Nguyen 2007;Nguyen 2008].

However, one serious problem is that web users do not have enough time to evaluate various search engines and to be more knowledgeable for selecting the best results in a certain situation (e.g., topic and context). In other words, it is difficult for them to submit each query to multiple search engines and to wade through the received flood of good, duplicated, or irrelevant information, and missing web contents. In order to meet the information requirements of users, many information services have been customized and developed, according to the various user interests and contexts.

Especially, it is important to manage multiple sets of results from the corresponding search engines, e.g., integrating and sorting out the results [Sliwko and Nguyen 2007]. Thereby, in this paper, we want to propose an integrated system which is capable of searching and sharing the resulted web documents by using three main features, i.e., i) link structures among the documents, ii) semantized tags, and iii) ontologies. The main hypotheses of this study can be noted as follows;

1. Performance of ranking based on the link structure among the web resources are better than that of simple keyword matching-based ranking,

2. Performance of results retrieved by the semantized tags and ontological sharing approaches are more personalized than searching results by single individual activities, and

3. Results can be integrated by combining both of results above.

Clearly, interoperability between semantized tags is a key factor to make the search results from different search engines to be effectively integrated with each other. The important assumption behind this ranking method is that contextual differences between the semantized tags and tag components for appropriate ranking should be solved. In practice, while most of search engines have shown a good *recall* performance (i.e., many web sites are successfully searched and filtered), they have not provided high *precision* results (especially, low precision of result ranking), given a certain query [Jung et al. 2007;Jung 2007]. By using appropriate rank aggregation strategies, we can prevent such results from appearing in the top results.

The outline of this paper is organized as follows. In the following Sect. 2, we discuss prior research in the area of search agents and practical paradigm in the context. In Sect. 3, we will present main methodologies of this work. In Sect. 4, we will demonstrate effectiveness and usefulness of the proposed strategies for web searches as well as ranking mechanisms by comparing to the existing methodologies. Finally, Sect. 5 will draw some conclusions of the paper with our future plan to extend this work.

2 Related Works

To search for relevant information from web is such a time-consuming and laborious task. Moreover, it might be a difficult task to automatically discriminate whether the retrieved result is precisely matched with user context or not [Convey 1992]. General search engines (e.g., Yahoo and Google) are the most common way to search for the information on the web. Users provide a set of terms to the search engines so that they can acquire a set of corresponding URLs of web pages that include the terms [Jansen et al. 1998]. However, these general purpose search engines are not always good at fulfilling user requirements, due to the lack of information about user interests. It is well known that users in the web search engines tend to use short queries that consist of only one or two words [Jansen et al. 1998; Silverstein et al. 1999].

However, short queries are often non-specific and ambiguous in nature. The results retrieved by the similarity-based method can thus bring very poor quality. Consequently, users are asked to spend a considerable amount of time on searching and retrieving quality information. In order to solve these problems, personalized web search personalization has been discussed from various viewpoints [Godoy and Amandi 2000;Ligon et al. 2005].

Search results must be the most reliable and relevant web site that people might be currently expecting. It means that the large-scale web search engines have to solve the slow speed problem caused by computing the relevance of web documents before user searching activities take place [Brin and Page 1998]. This approach seems to distill a smaller and more relevant topics from a large search topical candidates on the web that in order to make sense to a human user. It identifies authoritative and hub sources about a given user query. While authoritative and hub sources are calculated using link structure information, authoritative sources are the most reliable web site about specific topics and hub sources are documents that link to many authoritative sources [Kleinberg 1999].

Many agent systems have been developed in order to give a more appropriate solution to the problem of searching for the web [Mladenic 2002;Goldman et al. 1997]. Several approaches have been proposed for describing user interests by receiving different kinds of feedback. Mostly, this feedback is explicit, and it consists of direct rating scores that the user makes of specific contents [Chen et al. 2002]. Explicit relevance feedback could be annoying for users and that is why most recently developed approaches have tried to avoid it [Balabanovic 1998; Chen and Sycara 1998].

Furthermore, some studies [Klyne et al. 2004;SafetyOnline 2003] have applied content labeling and filtering schemes to web searching systems [Nguyen 2008]. This method can mainly provide a guide on how to block irrelevant sites. The relevant search results on web can be built on reconsidering the semantic approach from three directions: total architectural design, detailed specification, and deployment plan.

Search engines use several ranking algorithms to order the list, according to a certain criteria. Alternatively, Xue et al. has proposed to learn the ranking function from clickthrough data (the logfile of results the users clicked on) of a group of users [Xue et al. 2004]. An aggregation of the results obtained would be more useful than just dumping the raw results. For such an aggregation, Dwork et al. has suggested a local Kemenization technique [Dwork et al. 2001], which can bring the results highly ranked by the major search engines to be located in higher position in the aggregated ranking list.

In establishing an information-sharing platform, many technologies have been used. Not only COM (Component Object Model) and CORBA (Common Object Request Broker Architecture) but also ontology-based agents have been developed to solve the technical problem of information sharing [Filman et al. 2000].

3 The Framework of Proposed System

In this section, we present an architectural design and mechanism of our meta search agent model [Sliwko and Nguyen 2007] in order to increase the performance on web searching process., as shown in the Fig. 1

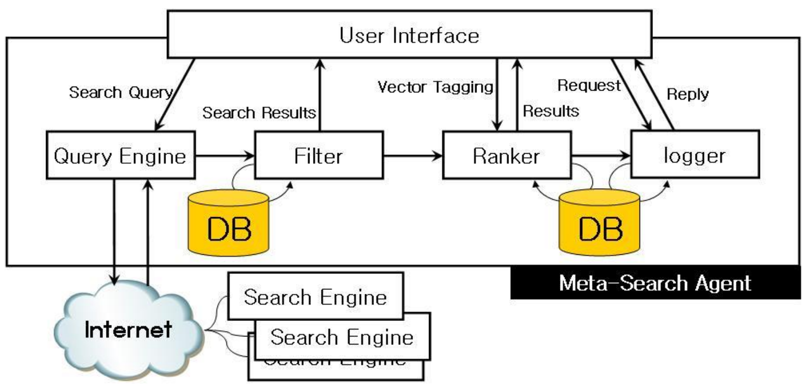

Figure 1: Systems architecture of the proposed meta search agent model

Especially, we are focusing on a ranking scheme using semantized tags and a sharing scheme using ontological approach for accessing relevant information.

3.1 System Architecture

We firstly put a brief description of the meta search task, while a user is trying to exploit the proposed meta-search agent. The major components of the proposed system are as follows; i) user interface, ii) query engine, iii) filter, iv) ranker, and v) logger. We want to discuss each component and how these components are related to each other.

The user interface is primarily applied to input the query (i.e., terms or phrases) for searching task. The query engine receives user queries through the user interface, and determines how query phrases should be considered to be matched with the others. Also, the query engine sends out queries to several selected search engines, and collects the results from the web. The filter is capable of eliminating irrelevant web pages as duplicated web documents from the set of documents returned from the underlying source search engines.

In order to approve primary search results, we propose a ranking scheme with tags introduced by a principle of position vector. Precedence users annotate each document with semantized tags, and then they can get search results in ascending or descending rank order. The ranker re-rankes the documents with the relevance vector value, and then through the user interface, users can browse the results returned from the ranker. All of the browsing results on the user interface are kept at the logger. The logger is a database for sharing search histories of similar users interests. For effective sharing among users, the ontology of user is deployed to categorize users histories in the logger.

3.2 Mechanism of Pattern Filtering

This task of the filter proceeds on pattern matching the URL address of each document. The architecture of the filter (URL pattern filter) is shown in Fig. 2.

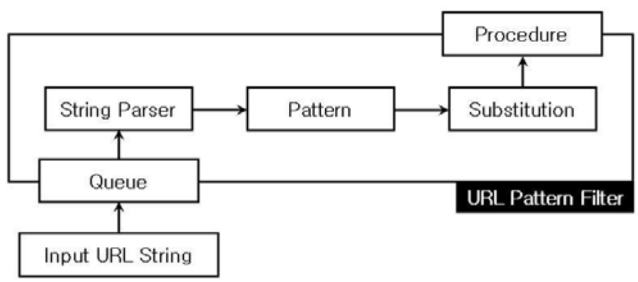

Figure 2: URL Filter architecture of the proposed system

The filter eliminates duplicate documents from same source or with similar address from the returned search results via the query engine. URLs pattern elements are extracted from incoming URLs, parsed by the string parser, and matched among these extracted patterns in the subcomponent of pattern and substitution. The mechanism of URLs pattern filtering is shown in Fig. 3. In order to explain how filtering works, we assume that a filtering amount sets “Filter: Amount ≥ 0.35” as shown in Fig. 3.

The value of ≥ 0.35 means that the filter allows remaining search lists to be below 35% matching. With this filter, any withdrawn URLs that have an amount of pattern

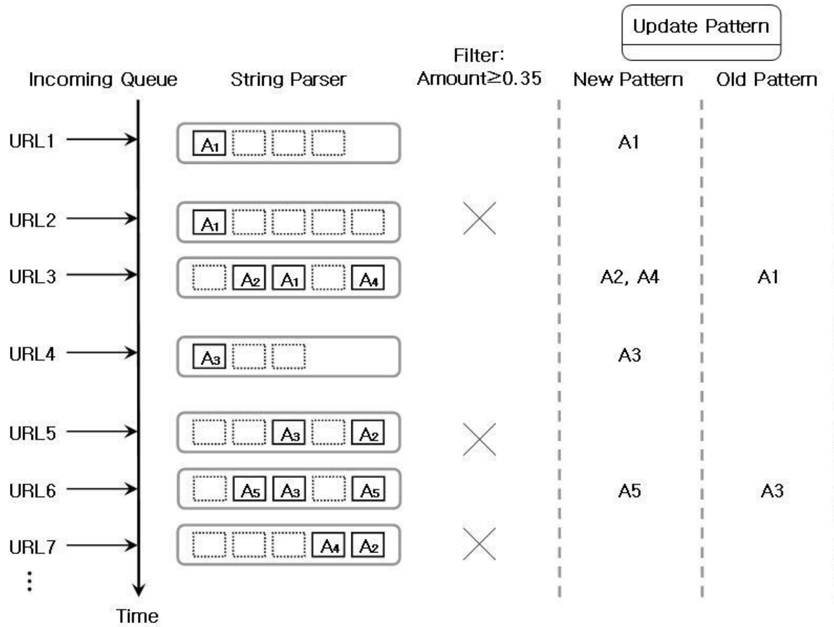

Figure 3: Mechanism of URL filtering under "Filter: Amount $\geq$ 0.35"

matching more than 35% do not enter the search lists and are therefore not passed to update patterns. Users can see the search lists of the remaining good pages based on the relevance pattern provided by URLs filtering results.

3.3 Principle of Rank Using Proposed semantized tag

After filtering, the user interface shows search results returned by the underlying filter to all users. For approving of these primary search results, we suggest a semantized tag method as a good ranking scheme. The proposed semantized tag consists of both a keyword and a vector component. The syntax for the semantized tag is as follows:

$$KEYWORD\,[X, Y, Z] \quad (1)$$

The $KEYWORD$, in the above syntax, expresses users aim and becomes a criterion of categorization for classification in the logger. The vector component $[X, Y, Z]$ represents the position of an object in space in relation to an arbitrary inertial frame of reference, and referred to as a location point that exist in a 3-dimensional space (see Fig. 5). It is also used as a means of deriving displacement by the spatial comparison of distances from the keyword which is an origin of vector and represents user intention (as shown in Fig. 4 and Fig. 5).

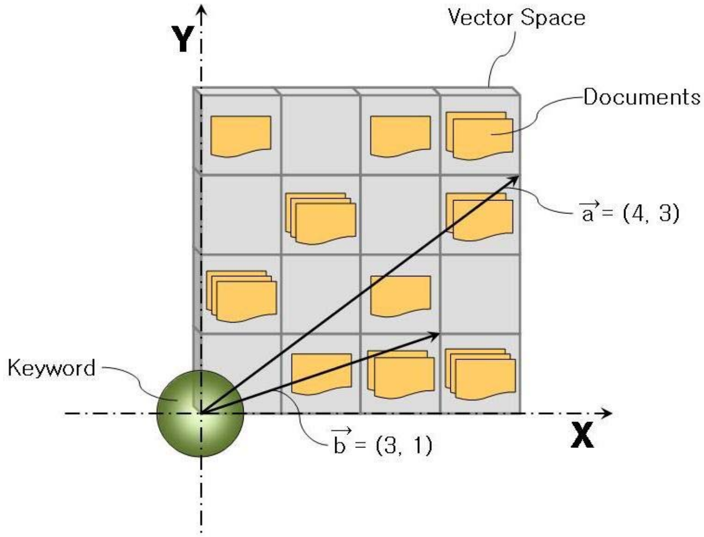

Figure 4: Conceptual structure to define with 2-dimensional semantized tag

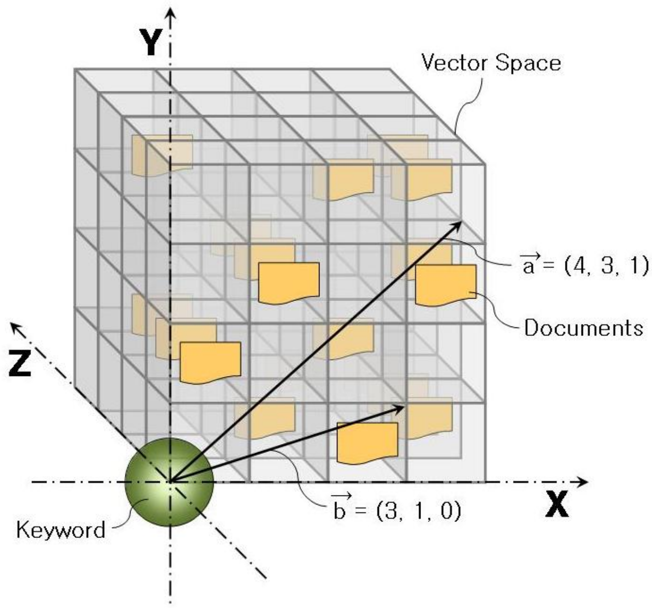

Figure 5: 3-dimensional semantized tag

The vector space of Fig. 4 explained by a Cartesian coordinates system with X and Y axis represents a position vector — X axis is rank of importance based on recommendation, citation, etc; Y axis is time of information that is an indicator for creation time. In advanced, we extend the 2-dimensional vector space to Z axis direction as Fig. 5 — Z axis is reference value introduced by others evaluation for the same web content. We give that the 3-dimensioanl semantized tag (Fig. 5) is therefore more effective and detailed expression for informativeness than the 2-dimensional one (Fig. 4). A semantized tag based on position vector can be used to represent any quantity that has both a magnitude and a direction. Vectors are usually shown in graphs or other diagrams as arrows, as illustrated Fig. 4 and Fig. 5. The length of the arrow makes known the vectors magnitude which means a conceptual distance between search intention and search results. So a short arrow expresses near distance with search intention, and stands for more usefulness than a long one. In Fig. 5, for example, the vector $\overrightarrow{a}$ displays more valuable information than the vector $\overrightarrow{b}$.

The position vector of tag is made in vector space as Fig. 5. Vectors magnitude (length) R for quantitative comparison is defined with Equ. 2, and nearer information to search intent (origin of vector represented by keyword) has small cost. Also, weight value according to bias of α, β, and γ can be controlled.

$$R = \sqrt{\alpha x^2 + \beta y^2 + \gamma z^2} \tag{2}$$

If several users tag same search result, one document has several semantized tags. Each Vectors magnitude R of these tags has to be accumulated, then informative value (V) can be written as Equ. 3 according to estimate cost of individuals $[X, Y, Z]$.

$$V = \frac{\sum_{i \in T_N} \sqrt{X_t^2 + Y_t^2 + Z_t^2}}{N_{total}} \tag{3}$$

In Equ. 3, where T_N shows total tags which can be classified by the same category, i is one of total tags in a category. A tag has an eigenvalue with X_t, Y_t, and Z_t . X_t has some x of components value defined at some point of importance value; Y_t has some y of components value defined in concurrence of time value; Z_t has some z of components value defined by other users evaluation. The magnitude of position vector that is worked by x, y, and z displays informative value (V). The summation of value in Equ. 3 divides by whole tags (N_{total}) to represent mean value being accumulated.

3.4 Ontological Approach for Sharing Scheme of the Logger

The logger, a kind of database, maintains all of the browsing results on the user interface for sharing histories of web search with similar users. For effective sharing, a request to detect similar user induce a practical suggestion as ontological approach. Especially, in order that user information could be understood and processed by machine automatically, a linkage of both the logger and not existing databases but more semantic descriptions is needed. Ontology has become a promising technology to express semantics [McGuinness and van Harmelen 2004].

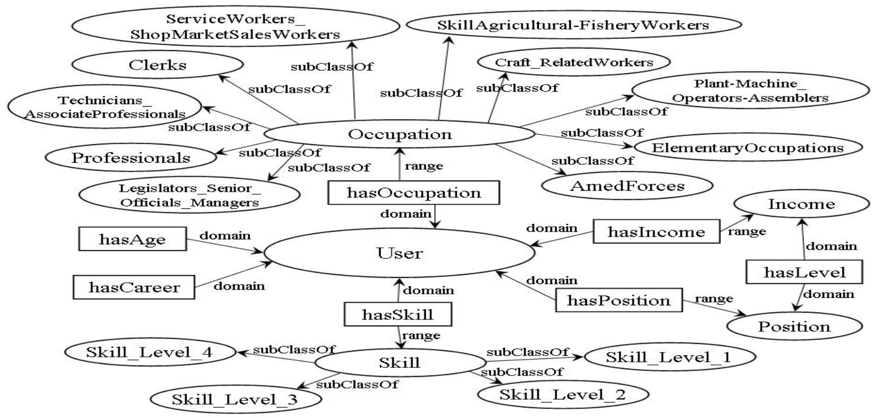

Figure 6: An ontology for the proposed sharing scheme

In building domain ontology people should explore the rationality in expressing domain concepts and relations between the concepts, considering the capabilities of OWL (Ontology Web Language). In Fig. 6, we put forward a framework of User ontology. For the framework of User ontology, we have considered capabilities to express semantic relations provided by OWL [McGuinness and van Harmelen 2004]. OWL classes are interpreted as sets that contain individuals. They are described using formal descriptions that state precisely the requirements for membership of the class. Properties are binary relations on individuals. Individuals represent objects in the domain that we are interested in.

In Fig. 6, the classes are properties are illustrated in ellipses and rectangles, respectively. OWL is property centricity, and each property has two constraints, e.g., domain and range. User is defined as a class. The domain of the six properties (hasAge, hasCareer, hasSkill, hasPosition, hasIncome, hasOccupation) is the class User. The range of the property hasOccupation is the class Occupation. The range of the property hasSkill is the class Skill. The range of the property hasPosition is the class Position. The range of the property hasIncome is the class Income. The domain of the properties hasLevel is the class Income and the class Position. The class Occupation has ten subclasses (e.g., Professionals, Clerks, and so on).

Also, the class Skill has four subclasses (e.g., Skill_Level). A particular class may have a restriction on a property that at least one value for that property is of a certain type

[McGuinness and van Harmelen 2004]. In the environment of Protege, we have built OWL ontology of user based on Fig. 6. OWL provides additional vocabulary along with a formal semantics as disjointWith, intersectionOf, unionOf, complementOf, oneOf, and relationOf.

4 Performance Evaluation

4.1 Effectiveness of Pattern Filtering

We conducted experiments to evaluate the Filter implemented in the component of the proposed meta-search agent. For accomplishing our evaluation of filtering effectiveness, we compare search results (with filtering) of our system with that (without filtering) of Google.

No.	U R L	Remark
1	http://www.terms.co.kr/ontology.htm	
2	http://www.dal.kr/blog/archives/001010.html	
3	http://humbroll.springnote.com/pages/344340	
4	http://complingone.georgetown.edu/~prot/	
5	http://bklab.snu.ac.kr/blog/kwangsub/45	
6	http://europa.snu.ac.kr/index.php/Some_issues_of_ontology_building_and_collaboration	
7	http://bric.postech.ac.kr/topic/63.htm	
8	http://osdir.com/ml/misc.ontology.protege.general/2002-10/msg00025.html	
9	http://iems.net/votal/index.php?cat=215	
10	http://osdir.com/ml/misc.ontology.protege.general/2002-10/msg00024.html	*
11	http://bklab.snu.ac.kr/~kskim/wp/?cat=5	
12	http://shannon.springnote.com/tags/ontology	
13	http://shannon.springnote.com/pages/501678	*
14	http://libterm.springnote.com/pages/541582	
15	http://chord.snu.ac.kr/blog/kwangsub/44	
16	http://mar.gar.in/brandon/ontology&mp=b&msort=datedesc	
17	http://www.blogweb.co.kr/?p=31	
18	http://www.blogweb.co.kr/?cat=3	*
19	http://report.empas.com/search/index.hcam?type=total&qu=ontology	
20	http://gamedev.springnote.com/pages/467242	

Figure 7: Search results by Google

Table 7 shows top 20 of search lists for query "ontology" provided by Google at Oct. 28th 2007, and Table 8 shows top 20 of search lists for the same query with our system at the same date. In Table 8, we can see #18, #19, and #20 instead of #10, #13, and #18 of Table 7. The documents of #10, #13, and #18 in Table 7 have similar URLs patterns of #8, #12, and #17, respectively. So our system filters these documents for exclusion in the collection of the search results.

No.	U R L	Remark
1	http://www.terms.co.kr/ontology.htm	
2	http://www.dal.kr/blog/archives/001010.html	
3	http://humbroll.springnote.com/pages/344340	
4	http://complingone.georgetown.edu/~prot/	
5	http://bklab.snu.ac.kr/blog/kwangsub/45	
6	http://europa.snu.ac.kr/index.php/Some_issues_of_ontology_building_and_collaboration	
7	http://bric.postech.ac.kr/topic/63.htm	
8	http://osdir.com/ml/misc.ontology.protege.general/2002-10/msg00025.html	
9	http://iems.net/votal/index.php?cat=215	
10	http://bklab.snu.ac.kr/~kskim/wp/?cat=5	
11	http://shannon.springnote.com/tags/ontology	
12	http://libterm.springnote.com/pages/541582	
13	http://chord.snu.ac.kr/blog/kwangsub/44	
14	http://mar.gar.in/brandon/ontology&mp=b&msort=datedesc	
15	http://www.blogweb.co.kr/?p=31	
16	http://report.empas.com/search/index.hcam?type=total&qu=ontology	
17	http://gamedev.springnote.com/pages/467242	
18	http://mar.gar.in/ironpark/ontology&mp=b&msort=datedesc00	*
19	http://coreonto.kaist.ac.kr/	*
20	http://ontology.etnews.co.kr/	*

Figure 8: Search results by the proposal approach

We performed experiments of filtering effectiveness for five example queries "ontology", "tag", "meta", "search", and "agent" on web. Fig. 9 shows individual filtering precision for each 1000 crawling documents using example queries, where we use the filtering mechanism of URLs pattern matching.

The X axis is the five queries with five pairs of columns which represent the percentage of filtering beside the right-contrast column (without filtering returned by Google). From this figure, we can see that the performance depends on the search results returned by Google and the proposed system. The average of filtering performance is 9.2% in the returning results for all queries. For some queries such as ontology and agent, the filtering results are heavily shown to 17.7% (remaining 82.3%) and 12.1% (remaining 87.9%) in comparing the both systems. As the results, our system seems to be effective in applying small browsers of handheld devices to achieve for fairly good filtering.

4.2 Appropriate Ranking and Time Complexity

We have chosen precision as a good measure of reasonable ranking in evaluating the appropriate performance for the document set as shown in Table 10 relating to molecular works.

All the ranking work on the same set of results and try to get the most relevant ones to the top. Hence, good information that has a higher precision at the top can be rated

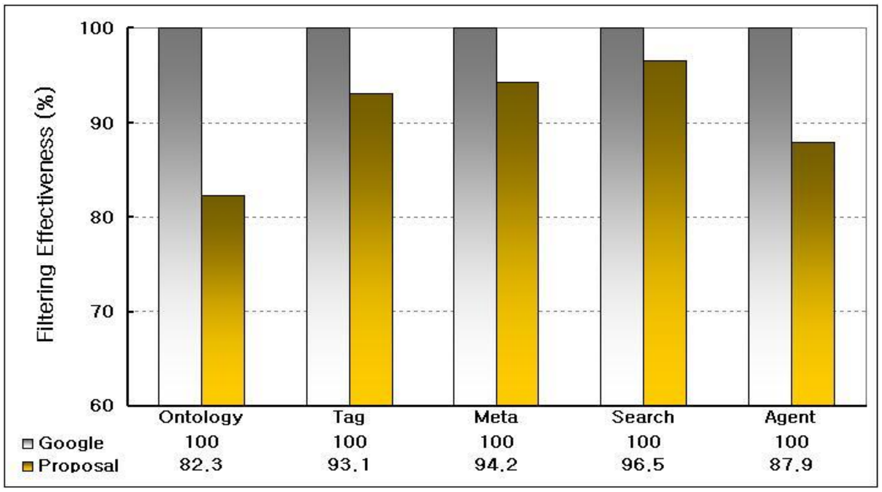

Figure 9: Comparing results with (proposal) and without (Google) filtering

No.	URL
1	http://cat.inist.fr/?aModele=afficheN&cpsidt=4181477
2	http://openwetware.org/wiki/Miniprep/Kit-free_high-throughput_protocol
3	http://people.morehead-st.edu/fs/d.peyton/protocols.html
4	http://sosnick.uchicago.edu/DNA_miniprep.html
5	http://userwww.service.emory.edu/~kressle/protocols/BAC%20miniprep%20protocol.doc
6	http://wolverton.owu.edu/lab/2006/01/eppendorf-miniprep-protocol
7	http://www.bioinformatics.vg/Methods/miniprep.shtml
8	http://www.bio.brandeis.edu/haberlab/jehsite/pdfs/ZymoPrep.pdf
9	http://www.bio.indiana.edu/~chenlab/potocols/qiagenmini.pdf
10	http://www.bio.net/hypermail/chlamydomonas/1993-December/000121.html
11	http://www.genetics.ucla.edu/labs/fan/Protocols_Miniprep.htm
12	http://www.genome.arizona.edu/agi/seq/QIAprep%20Spin%20Miniprep%20Kit%20Protocol.doc
13	http://www.genomed-dna.com/pdf/Quick-PDFs/Protocol%20Quick-Plasmid%20(Vacuum).pdf
14	http://www.gerardbiotech.com/documents/Protocols/GerardBiotech_HurricaneMiniPrep250_gbt010705.pdf
15	http://www.mbi.ufl.edu/~rowland/protocols/minipreps.htm
16	http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=7946306&dopt=Abstract
17	http://www.protocol-online.org/prot/Molecular_Biology/Plasmid/Miniprep/more2.html
18	http://www.tracy.k12.ca.us/thscdna/miniprep.html
19	http://www.umich.edu/~wakil/protocols/miniprep.html
20	http://www1.qiagen.com/HB/QIAprepMiniprep

Figure 10: The list of 20 documents for experimental results

better from the users prospect. Between both of the ranking results obtained by each of the proposed method (based on tag) and PageRank (based on hyperlink), we compare the ranking results processed by the both.

Table 10 shows valid evaluation of comparison with both the proposed method and PageRank. The ranking results of two methods have wide deviations. Our interesting observation is a warning of exclusion in the collection of Table 10 for disputed documents (document # 1, 4, 9, 16, and 19). Carrying out disputed documents seems to give a kind of effectiveness for preventing spam.

In comparison with two results of Table 11, we experimentally clarified the accuracies of ranking by the proposed method. Our ranking of approval rating has 79.8% agreement. And also, the decision-making process for exclusion of disputed documents has 99.4% consensus.

Document #	PageRank	Proposed Rank	Accruacy	Remark
1	16	-	1.00	*
2	17	2	0.83	
3	18	1	0.89	
4	10	-	1.00	*
5	5	14	0.74	
6	11	9	0.77	
7	4	3	0.80	
8	19	4	0.86	
9	2	-	0.97	*
10	7	15	0.71	
11	14	7	0.77	
12	12	12	0.83	
13	15	13	0.83	
14	20	10	0.77	
15	1	8	0.80	
16	13	-	1.00	*
17	3	5	0.80	
18	9	6	0.71	
19	8	-	1.00	*
20	6	10	0.86	

Figure 11: Comparison of ranking evaluation between PageRank and the proposed approach

We performed experiments for the time complexity of PageRank, Topic-Sensitive PageRank, and the proposal as shown in Fig. 12.

In which the X axis stands for the number of results returned from real Web data

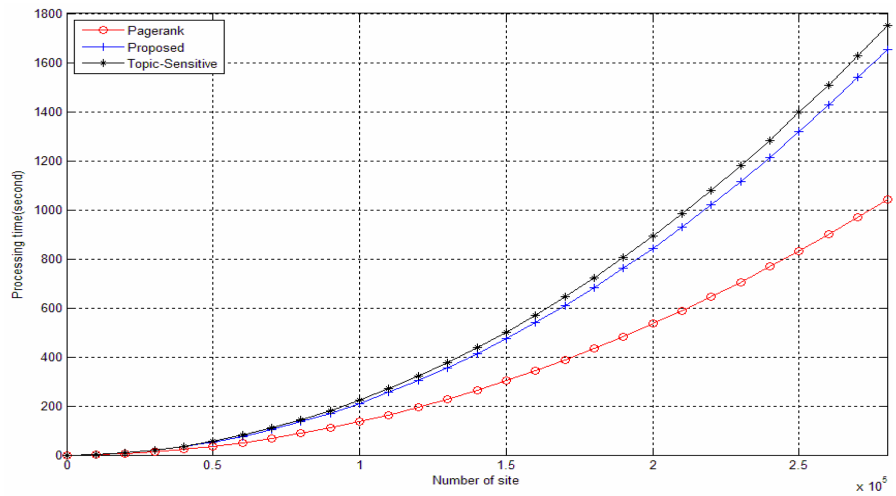

Figure 12: Experiments for the time complexity of three algorithms (PageRank, Topic-Sensitive PageRank, and the proposed approach)

based on Stanfords matrix[3], and the Y axis is the time spent for processing in the whole system. Figure 7 shows that our proposed method slightly outperforms than Topic-Sensitive PageRank. On other side, in comparing with PageRank, the two methods (the proposal and Topic- Sensitive PageRank) are more precipitous than PageRank because of the proposal and Topic-Sensitive ranking requested more additive calculations.

5 Concluding remarks and future work

The semantized tag method and pattern filtering enable Web search to increase performance. Future Web browsers seem to need coordination with a broker, software as a meta-search agent, for more powerful handling of Web contents. Our proposed semantized tag is a core idea for brokers. Page titles, headers, image, and content include the semantized tag into themselves in order to help optimize visibility and ranking in condition of Web browsers within limited devices or manners. It is important not to over use keywords, or hide them within inappropriate alt attributes or invisible text, as this could be perceived as spamming, resulting in exclusion by search engines. We provide a framework of User ontology for effective sharing search histories. Our exploration might not standardize User ontology. We explore with an intention to help Web search better understand the rationality in expressing concepts and relations between

[3] The Stanford Web Matrix, http://nlp.stanford.edu

the concepts in user categorization. In order that proposed User ontology relating to other domain ontology is constantly improved and standardized as possible.

Acknowledgement

This research was financially supported by the Ministry of Commerce, Industry and Energy (MOCIE) and Korea Industrial Technology Foundation (KOTEF) through the Human Resource Training Project for Regional Innovation.

References

[Balabanovic 1998] Balabanovic, M.: An interface for learning multi-topic user profiles from implicit feedback. In *Proceedings of AAAI Workshop on Recommender Systems*, 1998.

[Brin and Page 1998] Brin, S. and Page, L.: The anatomy of a large-scale hypertextual web search engine. *Computer Networks*, 30(1-7):107–117, 1998.

[Chen and Sycara 1998] Chen, L. and Sycara, K.P.: Webmate: A personal agent for browsing and searching. In Michael Wooldridge Katia P. Sycara, editor, *Proceedings of the Second International Conference on Autonomous Agents*, pages 132–139. ACM Press, 1998.

[Chen et al. 2002] Chen, Z., Meng, X., Zhu, B., and Fowler, R.H.: Websail: From on-line learning to web search. *Knowledge and Information Systems*, 4(2):219–227, 2002.

[Convey 1992] Convey, J.: *Online Information Retrieval: An Introductory Manual to Principles and Practice*. Clive Bingley, London., 1992.

[Dwork et al. 2001] Dwork, C., Kumar, R., Naor, M., and Sivakumar, D.: Rank aggregation methods for the web. In *Proceedings of the Tenth International World Wide Web Conference (WWW 10), Hong Kong, China, May 1-5.*, pages 613–622. ACM, 2001.

[Filman et al. 2000] Filman, R.E., Korsmeyer, D.J., and Lee, D.D.: A corba extension for intelligent software environments. *Advanced Engineering Software*, 31(8-9):727–732, 2000.

[Godoy and Amandi 2000] Godoy, D. and Amandi, A.: PersonalSearcher: An intelligent agent for searching web pages. In Maria Carolina Monard and Jaime Simão Sichman, editors, *Proceedings of the International Joint Conference, 7th Ibero-American Conference on AI, 15th Brazilian Symposium on AI, IBERAMIA-SBIA 2000, Atibaia, SP, Brazil, November 19-22*, volume 1952 of *Lecture Notes in Computer Science*, pages 43–52. Springer, 2000.

[Goldman et al. 1997] Goldman, C.V., Langer, A., and Rosenschein, J.S.: Musag an agent that learns what you mean. *Applied Artificial Intelligence*, 11(5):413–435, 1997.

[Jansen et al. 1998] Jansen, B.J., Spink, A., Bateman, J., and Saracevic, T.: Real life information retrieval: a study of user queries on the web. *SIGIR Forum*, 32(1):5–17, 1998.

[Jung 2007] Jung, J.J.: Ontological framework based on contextual mediation for collaborative information retrieval. *Information Retrieval*, 10(1):85–109, 2007.

[Jung et al. 2007] Jung, J.J., Juszczyszyn, K., and Nguyen, N.T.: Centrality measurement on semantically multiplex social networks: Divide-and-conquer approach. *International Journal of Intelligent Information and Database Systems*, 1(3-4):277–292, 2007.

[Kleinberg 1999] Kleinberg, J.M.: Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46(5):604–632, 1999.

[Klyne et al. 2004] Klyne, G., Reynolds, F., Woodrow, C., Ohto, H., Hjelm, J., Butler, M.H., and Tran, L.: Composite capability/preference profiles (cc/pp): Structure and vocabularies 1.0. Technical report, W3C, 2004.

[Ligon et al. 2005] Ligon, G.L., Balachandran, B.M., and Sharma, D.: Personalised search on electronic information. In Rajiv Khosla, Robert J. Howlett, and Lakhmi C. Jain, editors, *Proceedings of the 9th International Conference on Knowledge-Based Intelligent Information and Engineering Systems (KES 2005), Melbourne, Australia, September 14-16*, volume 3684 of *Lecture Notes in Computer Science*, pages 35–41. Springer, 2005.

[McGuinness and van Harmelen 2004] McGuinness, D.L. and van Harmelen, F.: Owl web ontology language. Technical report, W3C, 2004.

[Mladenic 2002] Mladenic, D.: Learning word normalization using word suffix and context from unlabeled data. In Claude Sammut and Achim G. Hoffmann, editors, *Machine Learning, Proceedings of the Nineteenth International Conference (ICML 2002), University of New South Wales, Sydney, Australia, July 8-12*, pages 427–434. Morgan Kaufmann, 2002.

[Nguyen 2008] Nguyen, N.T.: *Advanced Methods for Inconsistent Knowledge Management*. Springer-Verlag London, 2008.

[SafetyOnline 2003] SafetyOnline.: The sfs browser: A demonstration system that realizes filtering of inappropriate content accessed with mobile phones. Technical report, 2003.

[Silverstein et al. 1999] Silverstein, C., Marais, H., Henzinger, M., and Moricz, M.: Analysis of a very large web search engine query log. *SIGIR Forum*, 33(1):6–12, 1999.

[Sliwko and Nguyen 2007] Sliwko, L. and Nguyen, N.T.: Using multi-agent systems and consensus methods for information retrieval in internet. *International Journal of Intelligent Information and Database Systems*, 1(2):181–198, 2007.

[Xue et al. 2004] Xue, G.-R., Zeng, H.-J., Chen, Z., Yu, Y., Ma, W.-Y., Xi, W., and Fan, W.: Optimizing web search using web click-through data. In David Grossman, Luis Gravano, ChengXiang Zhai, Otthein Herzog, and David A. Evans, editors, *Proceedings of the 2004 ACM International Conference on Information and Knowledge Management (CIKM)*, pages 118–126. ACM, 2004.