Rough Set Based Information Retrieval from Argumentative Data Points in Weblogs

Sea Woo Kim*, Chin-Wan Chung†

**Division of Information and Communication Engineering, Korea Advanced Institute of Science and Technology, 373-1, Kusong-Dong, Yusong-Gu, Taejon 305-701, Korea.*

E-mail: seawoo@islab.kaist.ac.kr, * Corresponding author

† *Department of Computer Science, Korea Advanced Institute of Science and Technology, 373-1, Kusong-Dong, Yusong-Gu, Taejon 305-701, Korea.*

E-mail: chungcw@cs.kaist.ac.kr

Abstract

This paper describes a decision tree model and 3-dimensional representation of information retrieved from various weblogs in relation to argumentative logics. The weblogs are considered as datasets that show significant correlations between the queries applied to them. We have extracted a compact set of rules to support the dataset with the queries and employed effective evaluation metrics to evaluate the weighted average of the weblogs categorized into different types. The opinions from the weblogs are retrieved and represented as an object oriented 3-Dimensional system. The goal of our approach is to generate rules from rough sets and to represent them in a 3-dimensional interactive program, Blog Cosmos. We used rough set theory as a candidate framework for query refinement.

Keywords: *Information search and retrieval, Information filtering, Rough sets, Decision model*.

1. Introduction

Argumentation is a ubiquitous tool that permeates all fields of study. Argumentative essays and research plays an important role in linguistics. The study of argumentation can be further divided into formal and informal logic based on the methodology involved. Formal logic arguments are decontextualized sets of sentences or symbols viewed in terms of their syntactic or semantic relationships. On the other hand, informal logic arguments are pragmatic, i.e., their meaning is a function of their purposive context [3].

The importance of argumentation technology has increased due to the fact that decision making is required by various databases on the web. In this paper, we analyze decision making based on argumentation technology to extract results from various weblogs. The evaluation of weblogs is a new field and the extraction of decisions from these weblogs has various advantages, such as their use as a predictive tool in ascertaining public opinion on various topics of interest.

A blog (abbreviation of weblog) is a tool that enables people to publish their views or comments on the web. It is basically updated using software, which allows people to update or maintain the blog. This, activity is called blogging. The term "blog" was first used by Peter Merholz in April 1999, at a time when only 50 blogs were estimated to exist. In recent years, blogs have increased in popularity, as seen in such examples as Blogger, which was purchased in 2003 by Google [1]. Blogs have become an important tool for online surveys, for both corporate and private policies. Data obtained from weblogs may be utilized in decision making.

Decision making using argumentation over a group of data will be an approximation. The argumentation tree of a topic is manifold in that it requires its result to be achieved by dividing the data into sets and subsets. The result of this argumentation over this group of sets is analyzed using the theory of rough sets. The use of rough sets is suitable for such indiscernible data sets, allowing us to evaluate the weblogs more optimally.

A rough set is a formal approximation of a crisp set (i.e., conventional set) in terms of a pair of sets that give the lower and upper approximations of the original set. The lower and upper approximation sets themselves are crisp sets in the standard version of rough set theory. The main purpose of this theory is the “automated transformation of data into knowledge” [6].

The opinions from different blogs can be obtained using effective evaluation metrics and subsequently could be displayed as a 3-dimensional representation of the results, as a ‘3-D Blog Cosmos.’ We suggest that a dynamic group of experts / opinion leaders for a certain weblog be automatically created among users to evaluate domain specific weblogs. The experts would have dynamic authority weights based on their performance in the ranking evaluation. In addition, we suggest developing evaluation effectiveness metrics for the ranking processes. Furthermore, a dynamic change of authority weight would increase the credibility of the evaluation effectiveness of a given expert. Our search engine is domain specific, which improves the performance of search results and displays the top ranked weblogs. Also, a meta-search engine structure is very useful in collecting relevant information in specialized domains. We have clustered top rank documents by their rank order, and applied various expert evaluation effectiveness measures to those clusters to test whether our ranking list is reliable. Finally, we present a survey collection system and have shown them in a format called 3-D (Three Dimensional) BLOG COSMOS. The decision and investigation of a relevant survey can be enriched by this method. User performance tests verified the helpfulness of our system as a tool for the analysis of opinions in the Blog Survey.

2. Argumentation in weblogs

There are different models involved in the argumentation framework of a system. Some of the well established models are the Zeno model, GeoMed, Logic of Argumentation etc. [7]. Among the various models, the Logic of Argumentation or LA model comes closer to real world practice. The logic of argumentation, introduced by [4] is a model wherein large numbers of arguments are aggregated. This argument model uses an analysis of a proposition and the obtaining of points that either support or detract from it. But there is a disadvantage in their model as to how to combine the arguments and get an evaluation with respect to the proposition. Our work in this paper is related to the evaluation of the results from the argumentation in the weblogs. The positive or negative opinions of the candidates in the weblogs are derived using the effective evaluation metrics.

A blog thread is a set of entries connected to each other via reply links and referring to a common web page via a source link. An opinion leader often stimulates the discussion in a blog thread so that it becomes more active. Thus, we may be able to predict whether a blog thread grows by watching an influential leader’s entries. If the system statistically judges that threads often grow just after a particular blogger has published an entry, then that blogger is judged to be an influential leader.

This influential leader can be of two types. One is a positive influence and the other is a negative influence [2]. Both influence the activities on blogs, but we value positive leaders more, since they tend to lead the general leader’s opinion. We evaluated each article based on the positive and negative words found in the page and accumulated these. The evaluation of the page is as follows: Points were added for positive words and were subtracted for negative words. The weights were given as follows: strong positive words carried more positive points and strong negative words carried more negative points. The overall weight of a page is calculated from the summation of positive and negative points.

3. Decision making rules based on rough set theory

The discovery of decision rules and the recognition of patterns from data examples is one of the most challenging problems in machine learning. The decision tree needs a discretization process for numerical attributes. Data classification requires the discretization of data by decision trees. Approaches based on decision trees involve making the continuous-valued attributes discrete in input space, creating many rectangular divisions.

An information system is a 4-tuple [9],

$$S = \langle U, Q, v, \rho \rangle$$

Where, U is a finite set of objects, Q is a finite set of features,

$$V = \bigcup_{q \in Q} V_q$$

V_q is a domain of feature *q*, and ρ:U → is a function, such that $\rho(x,q) \in V_q$ for every $q \in Q, x \in U$ called information function. If $S = \langle U, Q, v, \rho \rangle$ is an information system, $Q = C \cup D$, $C \cap D = \Phi$, C is a condition attribute set, D is a decision-making attribute set. An information system with a disjoint set of condition attributes and decision-making attributes is called a decision-making table.

The importance of various attributes is different in a decision making table. More attention should be given to the attributes that are important to decision-making. In the rough set theory, the importance to decision-making is defined by the support degree of a decision-making attribute.

4. Query refinement and Rough sets

The theory of rough sets takes objects, attributes, and decision values and creates rules for upper, lower, and boundary approximations of the set. With these rules, a new object can easily be classified into one of the regions. Query refinement has found its way to popular web search engines, and is even becoming one of those features in which search engines aim to differentiate in their attempts to create their own identity [5].

We use rough sets for evaluating web documents and weblogs. There are two groups to evaluate web documents and weblogs:

- Expert group
- User group

A group of people with great authority are considered an expert group. This expert group is automatically promoted from active users. These groups evaluate the web documents and weblogs through rough sets. Let us consider the web documents or weblogs, and how the two groups, namely the expert group and user group, rank the web documents. Say for example that the user group ranks the document as 32 and the expert group ranks it as 33. It is then in a positive class (positive decision class). If the user group ranks the document as 30 and the expert ranks it as 24 then it is in a boundary (BND) decision class. If the user group ranks the document as 22 and the expert ranks it as zero, then it is in a negative class in rough sets (negative decision class).

In global document analysis [5], the whole corpus of searchable documents is preprocessed and transformed into an automatically generated thesaurus. On the other hand, local document analysis only considers the top ranked documents for the initial query. In its most naïve form, terms that appear most frequently in these top ranked documents are added to the query. Local document analysis is referred to as a pseudo-relevance feedback approach, because it tacitly assumes that the highest ranked documents are indeed relevant to the query. A true relevance feedback approach takes into account the documents marked as relevant by the user. Finally, correlations between terms are computed based on their co-occurrences in query logs, instead of in documents.

Thus the expert and user groups will evaluate the web documents using rough sets. In turn rough sets will take the decision rules of the expert and user groups and create rules for upper, lower, and boundary approximations of the set. Hence, using rough sets we can evaluate the weblogs and web documents

4.1. Approximation of rough sets

As a simple example, consider a set of objects 'O_w' (weblogs) a set of object attributes 'A_r' (responses) a set of values 'V_R' (Rankings) and a function to relate the above parameters.

f: O_w x $A_r \rightarrow V_R$

Therefore each object is described by the values of its attributes. We define an equivalence relation $R(A_s)$, where A_s is a subset of A_r with two objects Ow1, Ow2 such that,

$O_w1\ R(A_s)\ O_w2 \Leftrightarrow f(O_w1.a) = f(O_w2,a)$,
for all a in A_s

Here O_w1 and O_w2 are indiscernible. The indiscernibility relation defines a partition in U. Let $P \subseteq A_r$,U/Ind(P) denotes a family of all equivalence classes of the relation *Ind(P)* called elementary sets. *Ind(P)* is an equivalence relation.

Let $O_w \in A_r, P \subseteq A_r$, the indiscernibility relation Ind(P) is defined as follows :

$$Ind(P) = \{(x,y) \in U \times U : for\ all\ O_w \in P, O_w(x) = O_w(y)\}$$

Now we utilize this relation to partition the collection of data or the universe into equivalence classes,

$\{e_0, e1, e2, \ldots\ldots e_n\} = R(A_s)^*$.

The pair (O_w, R) forms an approximation space with which we approximate arbitrary subsets of O, referred to as concepts. Given O_s, an arbitrary subset of O_w, we can approximate O_s by a union of equivalence classes**:**

The lower approximation of O_s (also known as the positive region):

Lower (O_s) = POS(O_s) = Union $\{e_i \subseteq O_w\} e_i$

The upper approximation of O_s **:**

Upper (O_s) = Union $\{e_i \cap O_w\} e_i$

NEG (O) = O_w – POS (O)

BND (O) = UPPER (O) – LOWER (O)

Therefore the model of rough sets that is used in our experiments is one defined by its lower and upper approximations.

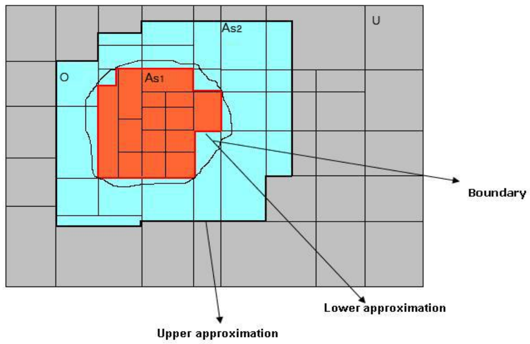

Figure 1. Approximation of sets

In Figure 1, U is a non-empty set of finite objects (the universe) that represents the collection of weblogs. O is the object and As_1, As_2 are the responses [8]. The object (O) we have chosen in our paper is "Euthanasia pros and cons" and As_1, As_2 are the responses over the object.

The concept that has been used in our work is based on the approximation space in rough sets. The approximation space is a classification of the domain interest into disjoint categories. Lower approximation is a description of the domain objects that are known with certainty to belong to the subset of interest, whereas the upper approximation is a description of the objects that possibly belong to the subset. Any subset defined through its lower and upper approximations is called a rough set.

5. Evaluation Metrics

A meta-search engine collects the addresses of cited blogs from conventional search engines. A web crawler automatically stores the list of web documents in the Document Database. At the time of a query, the ranking from the number of citations from the website and the expert ranking are combined. The combined ranks of the blogs are shown to the users, and the users visit the website they decide to explore. These search engines are recorded and monitored. The method of employing an expert group/opinion leader is based on the idea that in a given decision-making task that requires expert knowledge, many experts may be better than one if their individual judgments are properly combined. The explanation of the selection of group members and the determining authority, i.e. effectiveness, for each expert is as follows.

We have a visiting access matrix C= [Cij] that is in the citation ranking between users and web documents, which is given by,

$$C_{ij} = \begin{cases} 1 & \text{If a user } u_i \text{ visits a document } d_j \\ 0 & \end{cases}$$

C_{ij} is binary to prevent spam effects. Since it is binary, the frequent votes of a single user do not influence the visiting access matrix. From this information, the activity metric $I_a(u_i)$ for a user u_i and a visiting access metric $I_v(\beta d_j)$ for a Weblog are defined in Eq. (1) and Eq. (2) respectively. Here β is a constant and the value is 1 for our experiments. It can be 1/2 or 2 for different conditions in other experiments.

$$I_a(u_i) = \sum_{j=1}^{N_d} C_{ij} \qquad (1)$$

$$I_v(d_j) = \sum_{i=1}^{N_u} C_{ij} \qquad (2)$$

Where N_d is the number of blogs and N_u is the number of users registered in the meta-search engine. Our meta-search engine extracts weblogs from N_s - existing search engines, which are denoted as S_j, j=1,…,N_s. Then, the frequency authority metric $I_f(d_j)$, based on the search engines over a blog d_j, is defined in Eq. (3).

$$I_f(d_i) = \sum_{j=1}^{N_s} \delta_j m_{ij} \qquad (3)$$

$$m_{ij} = \begin{cases} 1 & \text{If a blog } d_j \text{ is seen in } s_j \text{ ,} \\ 0 & \end{cases}$$

Where δ_j is a weight for each search engine, initially set to 1.

The frequency authority metric represents the frequency of a blog in a meta-search engine. If this metric $I_f(d_j)$ is larger than a threshold, then we can assume that the blog d_j is of good quality and importance. Using the visiting access matrix $C=[c_{ij}]$ and the weblogs matrix $M=[m_{ij}]$ we can calculate the popularity and performance of each search engine used in the meta-search engine. The search engine frequency matrix $Y=[y_{ij}]$ is defined as $Y=C.M^T$, and then the weight δ_κ for each search engine can be updated as in Eq. (4).

$$\delta_k = \sum_{i=1}^{N_u} y_{ik} \Big/ \sum_{j=1}^{N_s} \sum_{i=1}^{N_u} y_{ij} \qquad (4)$$

The candidate users who are promoted to experts for a category can be determined by checking the metric I_a calculated during a given period. Every week or so, this "activity" metric is updated. Candidates are required to pass a test to become an expert. The weblogs of good quality or importance are determined by checking the frequency authority metric I_f and visiting access metric I_v over weblogs. Thus, we can select a candidate blog from among general blogs using Eq. (5).

$$I_a = s_a I_f + s_b I_v + s_c \tag{5}$$

Where S_a,S_b,S_c are scaling factors. If I_a is larger than a threshold τ , the corresponding blog will be accepted as a candidate document. The selected blogs will then be evaluated by experts for a given category. For each candidate blog, experts are required to evaluate and score the blog. An evaluation score matrix is defined as $X=[x_{ij}]$ when the i[th] expert evaluates a weblog d_j with a voting score x_{ij}. We have a weighted importance or authority over experts for each category. The evaluation score matrix shows a relation between experts and candidate blogs. The weight is determined by the experts' activity metric I_a, test scores, and other factors. Experts are assigned the same weight at the initial stage, as they are not differentiated at this point. The weight is dynamically changed by their activity and feedback from online users about recommended blogs through voting results.

This weighted metric is useful even when the number of expert members is not fixed. Thus, for each blog d_j listed as a candidate document the weighted authority voting score is defined in Eq. (6).

$$V(d_j) = \sum_{k=1}^{N_e} r_k \chi_{kj}$$

$$= \sum_{k=1}^{N_e} (w_k \Big/ \sum\nolimits_{i=1}^{N_e} w_i) \chi_{kj} \tag{6}$$

$$w_k = w_a I_a(u_k) + w_b \tag{7}$$

where N_e is the number of experts for a given category, r_k is the relative authority for the k-th expert in the expert pool, and w_k (Eq. (7)) is the weight calculated using the activity metric, test scores, and other career factors for the k-th expert member. The scaling factors are w_a, w_b. We consider that w_k should be positive at all times.

The weight w_k is a dynamic factor, and it differentiates bad experts from good experts in terms of their activity and users' voting results. When some experts show little participation in voting or evaluate incorrectly, their authority weight w_k becomes smaller. An error measure E as a squared sum of differences between desired voting scores and actual voting scores is defined in Eq. (8).

$$E = 1/2 \sum_{j=1}^{n} (V(d_j) - V'(d_j))^2$$

$$= 1/2 \sum_{j=1}^{n} (\sum_{k=1}^{N_e} (w_k \Big/ \sum\nolimits_{i=1}^{N_e} w_i) \chi_{kj} - V'(d_j)^2) \tag{8}$$

Where n is the number of blogs evaluated by users, *V'(d_j)* is the desired voting score for an expert-voting document d_j. We assume that *V'(d_j)* is the average score evaluated by all users, but in reality it is rarely possible to receive feedback from all users. After each feedback, the authority weight for each expert changes and at the same time *V'(d_j)* can be obtained by calculating the average of user validations during the given session. We chose the coefficient 1/2 to make its gradient formula simpler, which will be shown later. We assumed that this value could be determined by the feedback from general on-line users. The voting scores of the experts should reflect the common ideas of users about ranking and satisfy the desire of many users to find relevant and appropriate information, because an expert is a representative of the general users and has extensive expert knowledge in a specific domain.

If we update the experts' ranking weights by feedback from users about a blog *dj*, the weight is changed by the dynamic equation (Eq. (9)).

$$\omega_i(t+1) = \omega_i(t) - \eta[\chi_{ij} - V(d_j)]\Delta_j \big/ S \tag{9}$$

According to the equation, weight change involves a correlation between a voting score difference among experts and the error difference. For example, when an expert-voted score is larger than the weighted average voting score and the average score is smaller than the desirable rank score, the expert gets rewards. The expert is penalized when the average score is larger than the desirable rank score. Some experts will have rewards and others receive penalties based on the weighted average voting score of their expert group. In instances where experts have too many penalties, they will be removed from the expert group and new experts will be added to the group.

6. System Architecture of blog cosmos

The system architecture of the blog survey system "**BLOG COSMOS**" is represented in Figure 2. **BLOG COSMOS** consists of 4 stages. Initially, we have to create a web crawler to find the related blogs and bloggers. Then the User Interface (UI) will be used to get the specific query from the users.

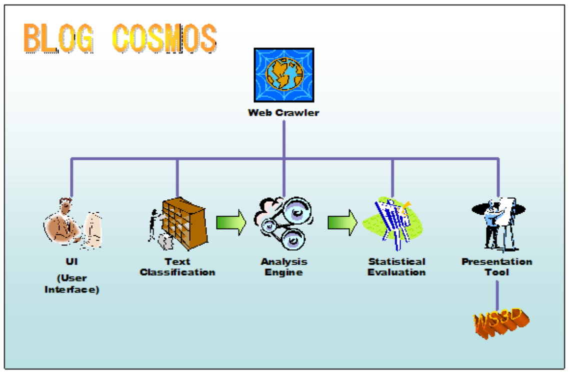

Figure 2. System architecture of blog cosmos

After the query is fetched, the text categorization classifies the websites and analyzes the blog postings. An analysis engine will do an analysis for classification of the text and give the result. We then use a statistical method to evaluate the data. The final result will be shown by a presentation tool that we call WS3D (Web Survey 3D).

7. Experiment

We carried out an experiment based on the methodology. We took into account 15 major blogs and also the topic "Euthanasia pros and cons debate" for the experiment. To ascertain the leading opinion, using evaluation metrics, we selected the 60 opinion leaders from the 15 major blogs using natural language processing and manual confirmation. Figure 3 shows the major leading blogs. Figure 5 shows the foremost opinion leader in the naver blog and the responses to that opinion leader's opinion.

Eng Site Name	Kor Site Name	Hit Count	No Of Opinions	Approval (%)	Opposition (%)	Undecided (%)	Filter Out (%)
Empas	엠파스	0	16	37.50	50.00	12.50	0.00
Hani	한토마	0	1	0.00	100.00	0.00	0.00
Chosun	조선일보	0	18	22.22	50.00	27.78	0.00
Naver	네이버	0	60	26.67	21.67	1.67	50.00
Total			95	21.60	55.42	10.49	12.50

Figure 3. Major leading blogs

The controversy regarding the practice of euthanasia is essentially a controversy about ethics and morality. The debate about euthanasia is a value debate among people who weigh values differently. People have voted in different sites as given below. In Naver there were 26.83% approvals and 12.20% undecided responses, while 60.98% are against the approval. In Empas 66.67% were against the approval and 60.98% were undecided responses. In Hani 22.16% were approvals and 45.41% were undecided responses, while 32.43% were against the approval.

Using the approximation space of rough set theory we can determine the decision rules of the various sites as shown in Figure 4. Based on these decision rules, blog cosmos can be designed.

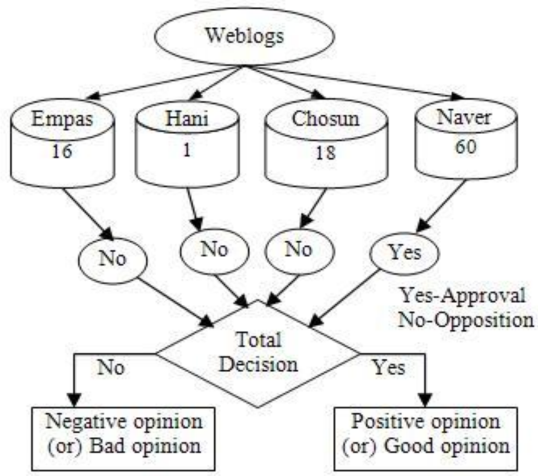

Figure 4. Decision rules using rough sets

The set of positive examples of weblogs with good decisions, O = {Empas, Hani, Chosun}

The set of attributes: A_r = {Euthanasia is necessary?}

The equivalence classes:

R(A) = {{Empas, Chosun},{Naver},{Hani}}

The lower approximation and positive region:

POS(O) = LOWER(O) = {Naver}

The negative region: NEG(O) = {Hani}

The Boundary region: BND(O) = {Empas, Chosun}

The Upper approximation:

UPPER (O) = POS(O) + BND(O) = {Naver, Empas,Chosun}

Decision rules we can derive:

des (POS(O)) → Yes

des (NEG(O)) → No

des (BND(O)) → No

Thus the summary of the decision from the above problem is,

(Euthanasia, Maximum hit counts) = Yes

(Euthanasia, Minimum hit counts) = No
(Euthanasia, Average hit counts) = No
Thus, blog cosmos can be represented in 3-dimensional structure using decision rules derived from rough sets.

Euthanasia pros cons debate

Naver

Article&Reply	Good	Bad	Result(%)
⊞ Euthanasia realted memo	35	-58	-24.73
⊞ Euthanasia muder or right to	39	-33	8.33
⊞ Euthanasia kebonion freed	4	-9	-38.46
⊞ Euthanasia panel talk	61	-98	-23.27
⊞ Euthanasia papers from passive reaction from Buddhism	35	-454	-85.69
⊞ Euthanasia pros - need to know	10	-10	0
⊞ Euthanasia ceasing treatment	1	-6	-71.43
⊞ A story of euthanasia	31	-55	-27.91
⊞ A daily life of euthanasia	14	-38	-46.15
⊞ Cons euthanasia	7	-8	-6.67
⊞ Pressure on euthanasia	4	-15	-57.89
⊞ Euthanasia another opinion	14	-27	-31.71
⊞ The way he lives euthanasia	7	-20	-48.15
⊞ My throught euthanasia	1	-3	-50
⊞ Euthanasia literature on cons position	10	-12	-9.09
⊞ Euthanasia cons standing point	10	-5	33.33
⊞ Euthanasia panel discussion	0	0	0
⊞ Euthanasia pros cons panel	0	0	0
⊞ Euthanasia	0	0	0

Figure 5. Major opinion leader in naver blog

We evaluated each article based on the positive and negative words found on the page. The evaluation of the page is as follows: points were added for positive words and were subtracted for negative words. The weights were given as follows: strong positive words carried more positive points and strong negative words carried more negative points. The overall weight of a page was calculated from the summation of the positive and negative points. A list of all the surveys carried out is shown in Table 1. Through various experiments, users' opinions on certain topics were obtained.

Table 1. Total surveys carried out in blog

Issues	No Of Opinions	Approval (%)	Opposition (%)	Undecided (%)
Does Government's policy have an effect on Immovable property?	196	9.67	60.79	29.54
Is the Korean Government's countermeasure good about North Korean Nuclear Crisis?	70	16.10	59.09	24.81
Is Korea and US FTA good or bad?	255	22.50	46.31	31.17
A Euthanasia pros cons Debate	280	26.10	48.20	25.70

Figure 6 shows a graphical representation of the survey system BLOG COSMOS. Here each planet or Sphere represents a website like yahoo, Naver, Empas, or Daum, which are visited often by numerous users. The size of the sphere will vary based upon the number of user opinions received. When the size of the sphere is large, it means that users have given more opinions on an issue, and vice versa. The opinion of an individual user can be either positive or negative. A positive opinion is considered a good opinion and a negative opinion is considered a bad opinion. The good opinions are shown in blue and the bad opinions are shown in red.

Hence, the total opinions from various sites like yahoo, Naver, and Empas are summarized in the center sphere as shown in Figure 6.

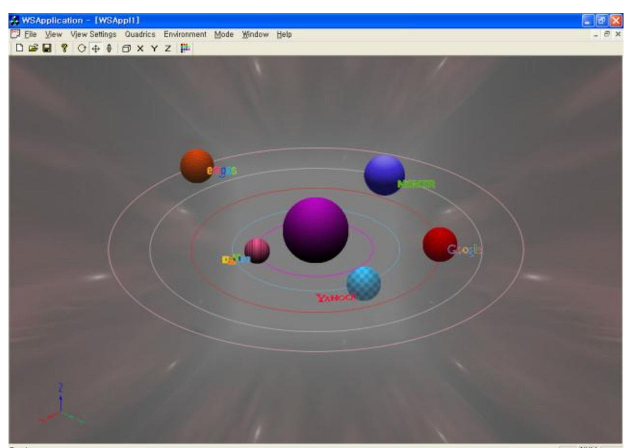

Figure 6. Survey system of blog cosmos

8. Conclusion

Our system's primary task is the retrieval of opinions on arguments from the blogs. We have clustered the top ranked blogs by their rank order, and applied various expert evaluation effectiveness measures to those clusters to test whether our ranking list is reliable. This paper presents the retrieval of data with a hybrid method for constructing a decision tree from rough sets.

This system is more suitable for expert-oriented issues, such as those related to medicine, politics, houses, cars, and banks, rather than general issues like movies or music. Our future work will include improving the web crawler and the redefinition of the parameters in the model. For example, recently posted or popular web entries can be given more weight. Collaborative filtering can be adapted to our system using expertise and trustworthy values.

9. Acknowledgement

This research was supported by the Ministry of Information and Communication, Korea, under the College Information Technology Research Center Support Program, grant number IITA-2006-C1090-0603-0031.

10. References

[1] Matthew Hicks (2004, October 5), Blogger Boss to Leave Google.eWEEK.com, http://www.eweek.com/article2/0,1895,1666646,00.asp

[2] Gilad Mishne, 2005, Experiments in mood classification in Blog Posts. Style2005, Stylistic Analysis of Text for Information Access c 2005 the author/owner.

[3] Thomas F.Gordon, Nikos Karacapilidis (1998), The Zeno Argumentation Framework.

[4] Krause P, Ambler S, Elvang-Gøransson M, Fox J. A (1995), logic of argumentation for reasoning under uncertainty. Comput Intell, 11:113–131.

[5] Martine De Cock and Chris Cornelis (2005), Fuzzy Rough Set Based Web Query Expansion, http://www.fuzzy.ugent.be/martine/papers/volledig/mdecock2005d.pdf

[6] Pawlak, Zdzisław (1991), Rough Sets: Theoretical Aspects of Reasoning about Data. Dordrecht: Kluwer Academic Publishing. ISBN 0-7923-1472-7

[7] G. Brewka, T. Gordon and N. Karacapilidis, (1995), Models of Argumentation, Negotiation and Decision Making , Mediating Systems for Group Decision Making: the Zeno System.

[8] Frida Coaquira (2006), University of Puerto Rico, On Application of Rough sets theory to Knowledge Discovery.

[9] Andrew Kusiak, 2006, Rough Set Theory, Intelligent Systems Laboratory, University of Iowa.