Proceedings of the Fourth IASTED International Conference
Advances in Computer Science and Technology (ACST 2008)
April 2-4, 2008 Langkawi, Malaysia
ISBN CD: 978-0-88986-730-7

A CLASSIFICATION AND VISUALIZATION APPROACH FOR KNOWLEDGE MANAGEMENT OF A SPECIAL INTEREST GROUP

Shahida Sulaiman, Aisyah Ismail,
Ahamad Tajudin Khader, Maziani Sabudin
School of Computer Sciences, Universiti Sains Malaysia,
11800 USM, Penang,
Malaysia.
shahida@cs.usm.my, acaizman@gmail.com,
tajudin@cs.usm.my, maziani@cs.usm.my

Sarina Sulaiman
Faculty of Computer Science and Information Systems,
Universiti Teknologi Malaysia,
81310, Skudai, Johor,
Malaysia.
sarina@utm.my

ABSTRACT
Some Web portals are developed to support electronic community or e-community of special interest groups (SIG) that provide the platforms to communicate, share ideas and knowledge among the members. By incorporating the Web portal with a knowledge-based system, knowledge acquisition and sharing will be more efficient. This type of knowledge system is known as knowledge sharing system rather than expert system. Existing SIG portals provide the information by listing the names of the related people in a textual form. Some Web portals reveal the level of expertise of the members but the textual representation does not manage to highlight the different level of expertise among the community explicitly. Thus we have proposed KM-ClaVis approach that consists of a point-based semi automatic classification method to classify the users' expertise levels and the tree view method to visualize the classified expertise of a software engineering interest group (SEIG) e-community. We apply KM-ClaVis approach in an existing SEIG knowledge portal. We anticipate the proposed approach will improve knowledge management among the members.

KEY WORDS
Knowledge management, knowledge portal, expertise classification, tree view, software engineering interest group

1. Introduction

Most special interest groups (SIG) use Web applications either portal based or not in order to communicate, share ideas and knowledge among the community. Some portals provide very basic information to the community for example by listing relevant activities of the community's interest and providing the search utility or links to other related resources. On the other hand, a number of such portals may also provide the utility to communicate via emails, forum or chatting to seek help from the others. By incorporating the Web portal with a knowledge-based system, knowledge acquisition will be more efficient. This type of knowledge-based system categorized as knowledge sharing system rather than an expert system. According to Niwa [1] a knowledge-sharing paradigm perceives knowledge suppliers as the same set of system users who use the knowledge base. This differs from expert systems that require suppliers to be the experts while users are the novices. Our proposed approach produces a knowledge sharing system not an expert system. It focuses on users' classification and the visualization of expertise levels besides search utility.

Web portals request users to register in order to join the e-community. The registration may acquire users to notify their field of interest, level of expertise and experience. The information will be stored in a database as the knowledge of users or community in the portal. Based on the data, the knowledge sharing system can classify the members' level of expertise. Existing expertise classification approaches [2][3][4][5] provide the algorithm to classify users' expertise automatically. Although automatic classification is effective, the results are still questionable particularly in term of the quality of answers posted. Thus users' intervention is still necessary in order to rate posted answers for instance.

Existing special interest groups' Web portals provide the information by listing the names of the e-community members textually such as in ITTutor.net [6] and Computerforum.com [7]. Such textual representation does not highlight the different level of expertise explicitly. We believe the utilization of a visualization technique in order to represent the information of the users' field of interest, expertise and experience will make knowledge acquisition and sharing to be more effective among the knowledge portal community.

Thus we propose KM-ClaVis approach that applies a point-based semi automatic expertise classification for knowledge management and visualize the expertise classification using tree view method in a SIG portal to improve the sharing of expertise and knowledge among a knowledge portal community. We have chosen software engineering community or software engineering interest group (SEIG) as the problem domain. A prototype of the portal is developed to materialize the proposed idea.

In Section 2 we discuss the related work and the problems that motivate the work. Section 3 describes knowledge-based system for knowledge management, followed by Section 4 discusses visualization techniques focusing on the graph drawing technique using tree view. Section 5 explains the KM-ClaVis approach and its application in a SEIG portal. Finally we conclude our work in Section 6.

2. The Motivation

In this section we will discuss the related work in expertise classification algorithms including classification and representation problems in existing Web portals for SIG that motivates our research.

2.1 Expertise Classification

The existing approaches include that of Zhang *et al.* [2] who proposed z-score measures and PageRank like algorithm called ExpertiseRank that was based on PageRank algorithm proposed by Page *et al.* [3]. In the work of Zhang *et al.* [2] the proposed algorithms were compared with PageRank [3] and HITS (hypertext induced topic selection) of Kleinberg [4] in a Java forum of an e-community to analyze the relative expertise of different users. The evaluation showed that both ExpertiseRank and z-score performed the best in e-community with different characteristics.

The z-score measures [2] combine both the asking and replying patterns. For example if users ask about the same number of queries and answers, the z-score will be close to 0. If they answer more than asking questions, the z-score will be positive otherwise it will be negative. In addition, ExpertiseRank [2] increases expertise scores using question-answer network. For instance if *User2* is able to answer *User1*'s questions, and *User3* is able to answer *User2*'s questions, then *User3*'s expertise rank should be promoted because *User3* is able to *User2*'s question where *User2* also happen to be someone who has some expertise. Nevertheless, the measures produced are still questionable, as the quality of the answers is not considered in the measures.

On the other hand HITS [4] rate e-community users based on their authority and hub values in the community network nodes. Authority value is the sum of the scaled hubs values that point to the page and hub value is the sum of the scaled authority values of the pages. Users with the highest authority score are experts in the community whilst users with the highest hub values are newbie who have good contact with the experts. Yet the setting of values for authority and hub could be affected if the actual contents of network nodes are of low quality that cause the increased number of authority and hub values when more unnecessary communication occurs.

Another work by Loser and Tempich [5] suggested three semantic overlay layers to give scores to e-community peers using peer monitor based on the frequency to answer a query either as responses to information requests, asking similar questions, providing related documents, asking questions of diverse topics in the past. Peer monitor is a good way that needs users' intervention to rank the peers. However the peers may give unjustified scores that cause discrepancies in the peer monitor.

Hence, we propose a semi-automatic point-based classification that employs z-score of Zhang *et al.* [2] that is mapped to a 5-scale point with the combination of a manual classification towards the answers given by members of a SIG e-community. We will discuss our proposed work more detail in Section 5.

2.2 Web Portals for SIG

The problems exist in the current Web portals cover the following aspects:

(i) Search of member's details:
Most Web portals provide a link where users can view all registered users. For example ITTutor.net [6] provides the list of registered community to be viewed by alphabetical order with brief description of users such as last login and status. Viewers can customize users list by selecting the status (core, normal member, administrator, professional member and member) from a dropdown menu. Another example is Computerforum.com portal [7] that also lists the members by alphabetical order of the portal community. More advanced search should be provided for instance to search users by field of interests and rankings or expertise.

(ii) Viewing textual information:
Existing portals list the members' name textually. Using textual listing requires users to scroll vertically in order to view the list. For instance in Computerforum.com [7] the members are listed textually sorted by rankings. Textual views do not reflect the levels and expertise in a meaningful way compared that of graphical views. However graphical views will be meaningful with textual information.

(iii) Classification of users:
Most portals require users to determine their own level of knowledge in order to classify them into expert, intermediate and beginner. In this case, knowledge may include users' experience, expertise and interest. On the other hand, intelligent-based Web portal has the facility to classify users without asking them directly. System learning capability is one way to develop intelligent-based Web portals. The motivation to achieve this goal is that

users do not always tell the truth. Users may have certain constraint in classifying themselves. They might classify themselves as an expert even though they are not. In Computerforum.com [7] ranks were given based on minimum posts made by the e-community members. Classification based on number of posts can be misleading as it disregards the quality of forums posted. On the other hand manual classification by the administrator will be a tedious task and unjustified, as the administrator might not understand the whole field of interests among the e-community members.

(iv) Personalization:
Personalization works when the system is able to identify type of users, interest, knowledge and other required details. The Web portal needs to recognize its users in order to provide presentation of the portal that is suitable for users in terms of its content as well as interfaces without manual setting from the users. Hence knowledge portal may produce a better personalization of each registered member based on the information or knowledge archived in the knowledge sharing system.

(v) User interface
Users might take more time to find desired information from Web portals, as there are too many links that they need to choose. The process becomes more difficult for new users who are not familiar with the portal. Some users may have no problem to view each link as they have time, interest and experience in searching the related knowledge via the SIG portal. However some new or busy users will get frustrated easily and abandon the portal. Thus presentation of the information in Web portals must be attractive, effective and user-friendly.

Web portals' interfaces play important roles in obtaining users understanding to navigate a site [8]. For example a search should be a type-in field and not a link. Portals should reduce time and cost in searching information and provide the flexibility to cope with any applications and operating systems [9]. Most users do not know how to use advanced search or Boolean query syntax. Ability to handle single-word queries or very short multi-word queries but still produce high-quality results is needed. One way to implement effective query is by using knowledge-based system.

The problems previously discussed motivate us to propose KM-ClaVis approach that can manage to classify expertise and represent the members and their expertise in a graphical method using tree view. The aspects to be improved in the proposed approach include providing searching of members by field of interests and level of expertise, integrating graphical view to represent the community members besides textual information using tree views, classifying users' expertise both automatic and based on users' intervention, and finally providing better personalization and user interface.

3. Knowledge Management

Knowledge management is a discipline that has the main goal to enhance the performance of individuals and organizations by managing the past, present and future knowledge assets [10]. There are two important components in a knowledge-based system: knowledge acquisition and inference engine [11]. Knowledge-based system can be an expert system or a knowledge sharing system. Based on Niwa's definition [1] our work is not considered as an expert system but as a knowledge sharing system where by experts and users are from the same set of system users. We focus on users' classification and visualization of expertise level to support knowledge sharing among SEIG e-community.

Three ways to represent the knowledge include [11]:

(i) Rules: Easy to implement, ordering of the rules is important.
(ii) Semantic net: Provide a simple, economical, and relatively intuitive representation form. Semantic networks are easy to implement and to manipulate due to its flexibility to cluster related knowledge.
(iii) Frames: Easier to understand; allow unrestrained alteration or cancellation of slots. Any slot can be changed hence the properties a frame inherits can be altered or cancelled anywhere in the hierarchy.

We have chosen semantic net to represent the SIG knowledge because it is easy to be implemented and suitable to represent different field of interests in software engineering field. In our study we apply the semantic net to represent the fields of interest based on software engineering body of knowledge (SWEBOK) [12] for a SEIG e-community.

4. Visualization Techniques

Visualization is transformation of data or information into pictures. It connect human sensory, which is vision and processing power of human mind. Thus, to human the result of visualization is simple and effective medium for representing complex information [13]. There are various techniques to implement visualization such as tree-map, icon-based, daisy chart and graph drawing [13]. We use the graph drawing technique in a tree view method.

A graph $G = (V, E)$ consists of a set of vertices V and a set of edges E, such that each edge in E is a connection between a pair of vertices in V. A research survey by Koschke [14] depicted that graph is the most often used kind of visualization because a graph represents a generic way to represent information, which is probably the reason why it is so popular. Graph drawing requires a graph layout algorithm.

Buchsbaum, *et al.* [15] outlines the objective properties for graph layout algorithms as:

(i) Easy to recognize and read individual objects for example having labelled nodes,
(ii) Avoid aliases including edge crossings, sharp bends and intersection of unrelated objects,
(iii) Reveal patterns by emphasizing symmetry, parallelism and regularity.

These properties are also recognized in other studies for example Gansner, *et al.* [16] as the aesthetic criteria or principles. Tree view is one of the techniques for a graph layout. The comparative study of four hierarchy browsers by Andrews and Kasanicka [17] showed that users significantly preferred the tree view browser compared to pyramid, treemap and hyperbolic browsers. Thus we also chose the tree view method in our work.

5. Classification and Visualization Approach for Knowledge Management (KM-ClaVis)

This section will discuss the two main techniques in KM-ClaVis approach that are point-based semi automatic classification method called PBaSE and the classification representation using tree view method.

5.1 Classification of Users Using PBaSE

The proposed classification is called point-based semi-automatic expertise classification (PBaSE) apply z-score measure of Zhang *et al.* [2] that is mapped into corresponding points and the average is computed with users' rating (see Figure 1).

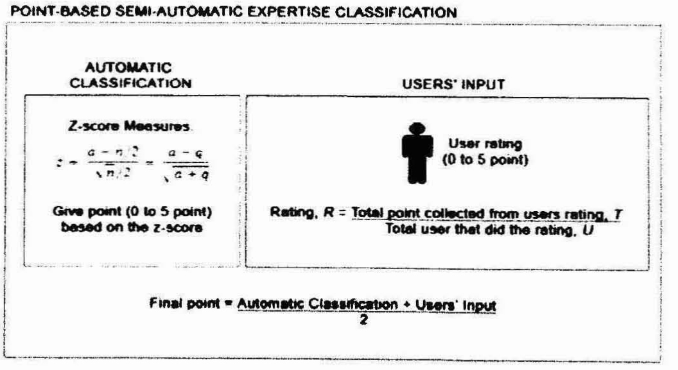

Figure 1: Proposed classification method of PBaSE

Each user in the SIG knowledge portal will be classified as a beginner, an intermediate, or an expert. In order to calculate the z-score, we need to identify the types of posts users create. This requires each user to determine the type of forum upon posting it. When users create a new forum, it will be identified as a query or question q. If users reply to a forum posted, the users will be prompted to indicate whether the forum is a question also or an answer a to the query posted earlier.

We propose the calculation of z-score measure to be mapped to the corresponding points to position the expertise levels as in Figure 2. Based on the z-score measures, if a user posts more questions than answers, the score will be negative. Thus the mapping of z-score measures ensures higher points given to members who post more answers and post less questions to benefit the e-community. We have classified the mapping in which the top 10% of the users will be given 5 points. The last 9% of the users will not be given any point. The other users will be given corresponding points. The users with the highest z-score are the top ten 10% of contributors in the e-community.

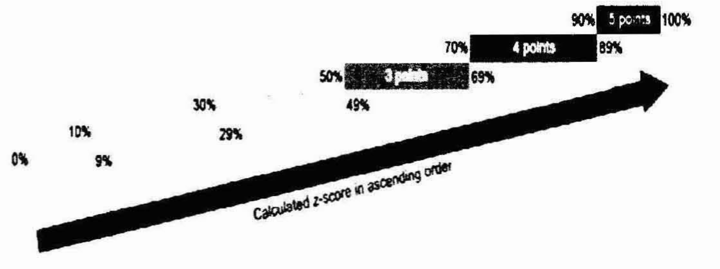

Figure 2: The mapping of z-score measure to points based on overall contribution among the members

The mapping implies that the more active the members of the e-community in posting high quality answers, the more they trigger questions among the new members. Thus the mapping approach ensures a continuous contribution among the members in order to maintain their expertise levels. The final point is automatically calculated and the expertise level is updated once a member has received the ratings for the posted answers. As the mapped z-score values provide the automatic rating of each member, we combine members' ratings to counter check the quality of answers posted. We treat the quality of each question differently, thus we sum the ratings of each question and divide the sum with the number of users who make the ratings.

Each member in the e-community may rate the others' answers by the scales: 0 (very very poor), 1 (very poor), 2 (poor), 3 (good), 4 (very good), 5 (excellent) as shown in Figure 3. The scale zero shows the answer posted does not have any contribution at all. Then the average of each member's rating R is calculated by dividing total of points collected from other members' ratings towards all forum posted T with the number of other members who did the rating towards all forum posted U. The final point F will be the average of the two values: the mapped z-score measure M and other users' or members' ratings R. Users' ratings are only made on answers posted not questions, as we perceive anybody can post questions but the quality should be measured based on the ratings of answers posted. We propose the mapping of the final points F to the expertise level L as: expert E ($4 \leq L \leq 5$ points), intermediate I ($2 \leq I < 4$ points), beginner B ($0 \leq B < 2$ point).

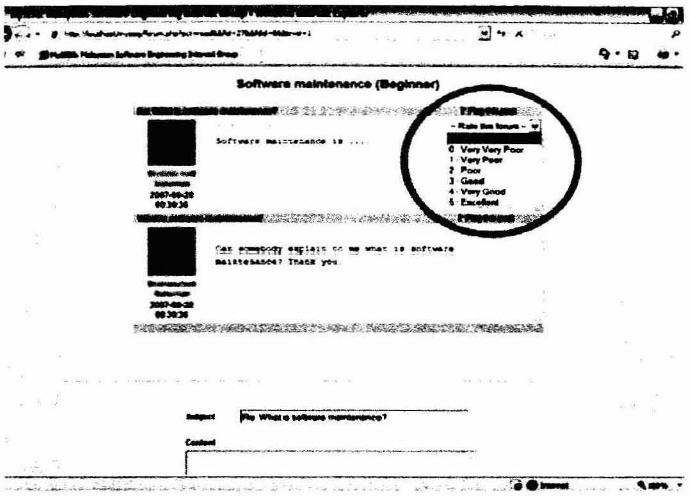

Figure 3: Others users may rate the posted forums

Let U_i is the users where i is number of users, Q is the number of queries or questions posted, A is the number of answers posted, Z is the z-score measures [2], M is the 5-scale point mapping, R is other users' ratings, F is the final rating and L is the level of expertise {B: Beginner, I: Intermediate, E: Expert}. Assume all R rated as 5 (excellent). For R, once there is no answer posted, zero values are set. For M the top 10% of users will be given the 5-point including those with a tie. The rest of the M values are calculated accordingly. An example of the results is illustrated in Table 1.

Table 1: An example of results for ten users

U_i	Q	A	Z	M	R	F	L
U_1	5	0	-2.24	0	0	0	B
U_2	0	5	2.24	3	5	4	E
U_3	5	5	0	0	5	2.5	I
U_4	10	5	-1.29	0	5	2.5	I
U_5	5	10	1.29	2	5	3.5	I
U_6	50	0	-7.07	0	0	0	B
U_7	**0**	**50**	**7.07**	**5**	**5**	**5**	**E**
U_8	50	50	0	0	5	2.5	I
U_9	100	50	-4.08	0	5	2.5	I
U_{10}	50	100	4.08	4	5	4.5	E

The results imply that after the classification algorithms are computed we find that: U_1 and U_6 are beginners, U_3, U_4, U_5, U_8 and U_9 are intermediates, and U_2, U_7 and U_{10} are experts. To counter check the results manually, we find that U_1 and U_6 post only questions thus it is logic that they are beginners. Observe for U_2, U_3, *and U4* although the users post the same number of answers A, the number of questions Q is different. The z-score measures Z will yield negative values if a user posts more questions than answers. Therefore the Z value for U_2 is highly positive because 5 answers posted but no question posted. On the other hand, the Z value for U_3 is zero because U_3 posts the same amount of answers and questions. In addition U_4 posts the number of questions that is double to that of answers thus it yields negative value of Z. The M values are based on Z values. Since the Z value for U_7 is the highest, the M value is assigned with 5 point. In this case, only 1 person (10%) of the community will be granted 5-point as the top 10 ranking. As the R value is given by other users who can only rate on answers posted with the assumption that other users give the same rating 5 for all users, U_7 will get the highest final value of F after taking the mean values of M and R. Notice that the final results F show that not only U_7 is classified as an expert, but also U_2 and U_{10}. Then based on our proposed mapping of the final points F to the expertise level L as: expert E ($4 \leq L \leq 5$ points), intermediate I ($2 \leq I < 4$ points), beginner B ($0 \leq B < 2$ point), we assign the L values correspondingly.

Based on the number of members in the e-community we can determine the degree of contribution among the members by referring to M values. The sample results show all posted answers are rated as excellent. As we get the final point F by deriving the average of M and R, we can observe that the ratings made by other members of the e-community will be balanced with the degree of contribution made in overall. Thus the classification L derived will be more justified.

5.2 Viewing of Members Using Tree View

A tree view allows users to expand and collapse nodes to view large data sets. By clicking the vertex in the tree view, the members of the fields by their expertise are expanded and displayed. Then clicking the other vertex will automatically expand the concerned vertex and collapse the previous viewed vertex. Details of each member are shown in the lowest vertex (See Figure 4).

The algorithm to implement the tree view is listed below:

(i) Assign the first layer vertex V_1 with the set of field of interest A_n where n is the number of field of interests available.

(ii) For each member of set A_n display the textual information of the field of interest and the total number of members in the set of users U_i where i is the number of users under the same value of A_n.

(iii) Assign the second layer vertex V_2 with the set of rank of users' expertise L_n where n is the number of users classified under the same value of L_n.

(iv) Assign the third level vertex V_3 repeatedly with the member of set of users U_i.

(v) Repeat step (iv) until i times.

(vi) For each vertex V_3 displays the date of the last login for each member of the set of users U_i.

5.3 Searching Users

The e-community users or members can search other members based on names, expertise and field of interests. The results will be viewed in the tree view form. The corresponding tree view is generated using the algorithm in Section 5.2 but it limits to the searched values only. The detail of the search technique is not discussed here.

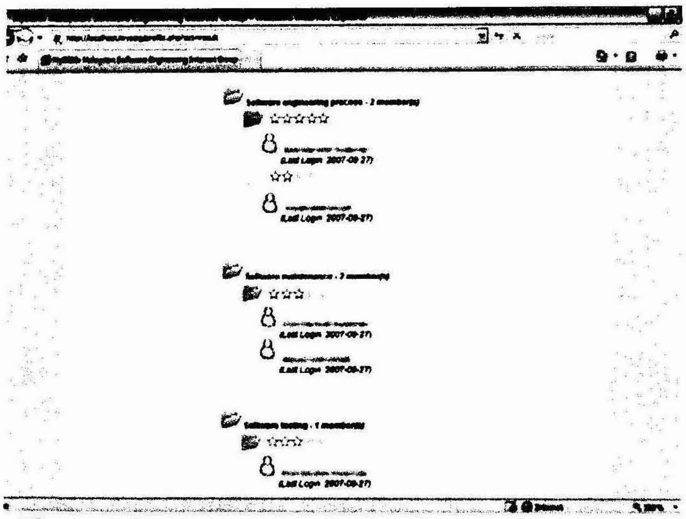

Figure 4: Users represented using the tree view method

6. Conclusion and Future Work

This paper has discussed the proposed KM-ClaVis approach that can classify and visualize the expertise levels of a SEIG portal to manage knowledge sharing. The classification is done using a semi-automatic point-based approach called PBaSE that uses the mapping of z-score measures by Zhang *et al.* [2] to a 5-scale point and we combine the result with the 5-scale ranking made by members of the e-community towards answers given by other members in order to derive the final expertise classification. The members' details can be viewed using interactive navigation of a tree view method by three layers: field of interests, level of expertise and members' details. We tested the approach in an existing SEIG knowledge portal and we found that the KM-ClaVis managed to improve the aspects required as below:

(i) Search of members by names, expertise and field of interests then display search results in a tree view.
(ii) View textual information: the tree view method and the textual information allow the expanding and collapsing of tree views to see members' details.
(iii) Classify users: apply semi-automatic point-based method that combines both automatic and members' classification based on quality of answers posted.
(iv) Provide personalization: members can set their field of interests or limit the views of the e-community members to the selected field of interest only.
(v) Good user interface: Icon-based tree view with short textual information and the member icon has the hyperlink to the details of the concerned member.

The future work may include the improvement of KM-ClaVis approach in the SEIG portal by integrating a method that can analyze the quality of posted answers in order to give a more justified automatic classification of members in the e-community. In addition the rating of answers made by different level of users' expertise may need to be revisited to ensure more realistic classification.

Acknowledgement

This research is partially supported by the research university grant of Universiti Sains Malaysia, 1001/PKOMP/817002.

References

[1] Niwa, K., "Towards Successful Implementation of Knowledge-Based Systems: Expert Systems vs. Knowledge Sharing Systems", *IEEE Transactions on Engineering Management*, 37(4), November 1990.

[2] Zhang, J., Ackerman, M. S., and Adamic, L., "Expertise Networks in Online Communities: Structure and Algorithms", *Proceedings of the International World Wide Web Conference WWW 2007*, ACM Press, 2007, pp. 221-230.

[3] Page, L. Brin, S., Motwani, R. and Winograd, T., The PageRank citation ranking: bringing order to the Web, Stanford Digital Library Technologies Project, 1998.

[4] Kleinberg, J. M., Hubs, authorities, and communities, *ACM Computing Surveys*, 31, U21-U23.

[5] Löser, A. and Tempich, C. "On Ranking Peers in Semantic Overlay Networks", 2005.

[6] ITTutor.net, 2007, http://ittutor.net.

[7] Computer Forum, *Jelsoft Enterprises Ltd.*, 2007, http://www.computerforum.com.

[8] Heck, M., "Diving into Portal", *INFOWORLD*, 2004, pp.37-47.

[9] Knorr, E., "The New Enterprise Portal", *INFOWORLD*, 2004, pp 43-52.

[10] Newman, B. (Bo), and Conrad, K. W., "A Framework for Characterizing Knowledge Management Methods, Practices, and Technologies", The Knowledge Management Forum, 1999.

[11] Giarratano, J. and Riley, G., Expert Systems Principles and Programming: 3rd Edition, PWS Publishing Co., London, 1998.

[12] Abran, A. et al. (eds.), Guide to Software Engineering Body of Knowledge SWEBOK: Version 4, USA, IEEE Computer Society Press, 2004.

[13] Schroeder, W., Martin, K. and Lorensen, B. The Visualization Toolkit – An Object-Oriented Approach to 3D Graphics, 2nd Edition, 2000.

[14] Koschke, R., "Software Visualization in Software Maintenance, Reverse Engineering and Re-engineering: a Research Survey", *Journal of Software Maintenance and Evolution: Research and Practice*, Vol.15, 2003, pp. 87-109.

[15] Buchsbaum, A., Chen, Y-F., Huang, H., Koutsofios, E., Mocenigo, J., Rogers, A., Jankowsky, M. and Mancoridis, S., "Visualizing and Analyzing Software Infrastructures", *IEEE Software*, September/October 2001, pp. 62-70.

[16] Gansner, E. R., Koutsofios, E., North, S. C. and Vo, K-P., "A Technique for Drawing Directed Graphs", *IEEE Transactions on Software Engineering*, 19(3), March 1993, pp. 214-230.

[17] Andrews, K., and Kasanicka, J., "A Comparative Study of Four Hierarchy Browsers Using the Hierarchical Visualization Testing Environment (HVTE)", *Proceedings of 11th International Conference Information Visualization IV'07*, 2007, pp. 81-86.