We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

12

Classifying Expertise in a Special Interest Group Knowledge Portal Using a Point-Based Semi-Automatic Expertise (PBASE) Method

Aisyah Ismail[1], Shahida Sulaiman[1], Maziani Sabudin[1],
Rosni Abdullah[1] and Sarina Sulaiman[2]
Universiti Sains Malaysia[1], Universiti Teknologi Malaysia[2]
Malaysia

1. Introduction

Knowledge is information and skills acquired through experience or education. We live in the knowledge era where knowledge is available almost everywhere in abundance. Therefore, knowledge should not be neglected; it needs to be shared and exchanged. Based on Newman and Conrad (1999), knowledge management is a discipline that seeks to improve the performance of individuals and organizations by maintaining and leveraging the present and future value of knowledge assets.

Knowledge portal is an enhancement of the ordinary web portal. While the web portal focuses on offering users a broad array of resources and services, the knowledge portal does not only offer the resources and services, it also acts as a knowledge repository where it will extract and analyze knowledge submitted among its community members. According to Niwa (1990), a knowledge sharing paradigm perceives knowledge supplier as the same set of system users who use the knowledge base system. Hence, knowledge portal is one of the means for knowledge sharing.

Based on Giarratano and Riley (1998), there are three ways to represent knowledge: rules, frames and semantic nets. Rules are the most common type of knowledge representation. Rules are easy to implement due to its straightforward structure. However, ordering of the rules is important. Frames represent related knowledge about an object. Frames are easy to understand and they allow unrestrained alteration or cancellation of slots. Frames are suitable to describe a mechanical device. Semantic nets are simple, economical and relatively intuitive representation form. The structure of semantic nets is denoted by nodes and arcs as shown in Fig. 1. This research will use semantic nets to represent its knowledge because it is easy to be implemented and manipulated due to its flexibility to cluster related knowledge in our problem domain.

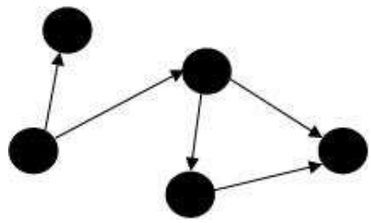


Fig. 1. Semantic nets consist of nodes and arcs to represent knowledge

In a Special Interest Group (SIG) knowledge portal, people from various backgrounds gather for several reasons. For instance, students join a SIG to derive some guidance from people who are already in the industry. They can also be experts in certain fields who are willing to answer questions from anyone and share their expertise. On the other hand, there are also some people who join the portal simply to make new friends with others who have the same interest. Likewise, these people posses knowledge and they are willing and eager to share their knowledge with each other through this online community.

Having people with various backgrounds in the community, we find the need to classify the users' expertise. Knowledge can be organized by classifying expertise of the user. In other words, users' expertise is the knowledge in the portal. When users join the portal for the first time, they may want to find other users' who share the same interests and problems. They may also want to seek help with their problems by looking for someone in the portal who is an expert in a certain field. Classification of the users' expertise is a very crucial task. Hence, we anticipate the expertise classification of the SIG knowledge portal will ensure the convenience of the members to exchange their knowledge and seek help among various expertise levels.

In Section 2 we will discuss the related work and the problems that motivate this study. Section 3 will describe the proposed method, followed by Section 4 that which explains will explain the implementation of the proposed solution. Section 5 will explain the qualitative evaluation of the proposed method. Finally, we will conclude our work in Section 6.

2. The Motivation

Online communities are not much different from other real world communities. Both communities consist of people who are tied together by their interests. In an online community, a group of people from different backgrounds are strangers to each other and this makes them become keen to get some information about the people in their community. Knowing one's level of expertise will make knowledge sharing and discussion more meaningful. Usually, the portal will state users' level of expertise for all community members to view. This section will discuss the existing classification methods in Web portals and the related work in classifying expertise in SIG portals.

2.1 Existing SIG portals

Both ITTutor.net (2009) and Computer Forum (2009) is popular web portals with registered users more than 40,000 and the number of members keep increasing. The portals rank users based on the number of posts they make in the portal in which the more forums posted, the higher users' rank will be. By doing so, even when users post query on a certain topic or post something irrelevant to the topic, users' rank will increase. Given a scenario where *A*,

who is a total beginner, posts a lot of queries in the forum without really contributing anything. Then there is *B*, who on contrary answers other users query in the forum. However, the number of *A*'s posts are larger than *B*'s posts. Based on the existing ranking approach, *A* will be ranked higher than *B*, which is inappropriate and misleading.
In ITTutor.net (2009), there are three ways to identify users' position in the portal (See Fig. 2). They are users' status, military-based ranks and rating from other users in the portal. Users' status will be assigned *Core*, *Ahli Biasa* (Normal Member), *Pengendali* (Administrator), *Ahli Professional* (Professional Member) or *Ahli* (Member). However, the users' status is not used to classify the users' expertise.

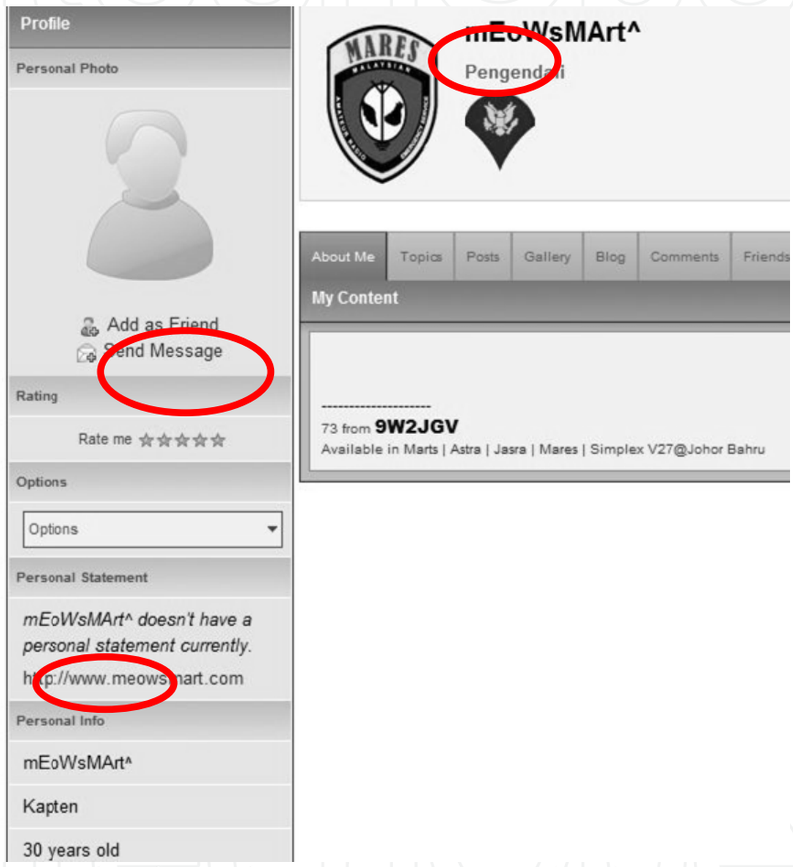


Fig. 2. Three ways to identify users' position in ITTutor.net (2009)

Instead military-based ranks as listed in Table 1 are used to rank the users in the portal. Users are ranked based on points they collected in the portal. The rating function will collect points given by other users in the portal and will be presented using 5-star rating. The three ways (users' status, military-based ranks and rating from other users in the portal) to identify users' position in the portal, will lead to users' confusion of the actual users' level of expertise.

Rank	Minimum Points
Kadet	0
Korporal	50
Sarjan	100
Staf Sarjan	150
Sarjan Mejar	200
Pegawai Waran 1	300
Pegawai Waran 2	400
Leftenan Muda	500
Leftenan	1000
Kapten	1500
Mejar	2500
Leftenan Kolonel	3000
Kolonel	3500
Certified ITTutor Professional	10000

Table 1. Military-based rank used in ITTutor.net (2009)

On the other hand, Computer Forum (2009) ranks its users based on the minimum posts made by users in the portal as listed in Table 2. Besides, there are also special ranks given by the administrator of the portal to selected users. Administrator also has the right to ban users who violate the rules and regulations of the portal.

Rank	Minimum Posts
New Member	0
Bronze Member	25
Silver Member	100
Gold Member	250
Platinum Member	500
Diamond Member	1000
Unspecified by computerforum.com	2000
Unspecified by computerforum.com	4000
Unspecified by computerforum.com	6000
Unspecified by computerforum.com	8000
Unspecified by computerforum.com	10000

Table 2. Ranks in Computer Forum (2009)

2.2 Expertise classification methods

The existing methods include that of Zhang *et al.* (2007) who proposed z-score measures, and ExpertiseRank that was based on PageRank algorithm proposed by Page *et al.* (1998). In the work of Zhang *et al.* (2007), the proposed algorithms were compared with Hypertext

Induced Topic Selection (HITS) of Kleinberg (1999) and simple statistical measures 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 (Zhang *et al.*, 2007) combine both the asking and replying patterns. For example if users ask 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 (Zhang *et al.*, 2007) increases expertise scores using question-answer network. For instance if *A* is able to answer *B*'s questions, and *C* is able to answer *B*'s questions, then *C*'s expertise rank should be promoted because *C* is able to answer *B*'s question where *B* also happens 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 (Kleinberg, 1999) 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 user and hub value is the sum of the scaled authority values of the user. Users with the highest authority score are experts in the community whilst users with the highest hub values are beginners 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 Löser and Tempich (2005) 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 and 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, this research proposes a point-based semi-automatic expertise classification that employs z-score of Zhang *et al.* (2007). The score is mapped to a 5-scale point with the combination of a manual classification towards the answers given by the members of a SIG e-community.

3. Point-Based Semi-Automatic Expertise (PBASE) classification method

The proposed work is called Point-Based Semi-Automatic Expertise (PBASE) classification method. This is a two-way classification method in which the knowledge portal will automatically classify users' expertise level based on users' interaction in the portal and users' rating. PBASE method consists of two parts; automatic classification using z-score measures of Zhang *et al*. (2007) and manual classification using users' rating. PBASE method takes the average of the two parts as the users' level of expertise. Users are classified as beginner, intermediate and expert based on the accumulated points.

There are two types of post in the portal. They are 'query' and 'answer' posts. The 'query' post is made by a user to ask questions under a certain topic. On the other hand, 'answer' post is a post that answers questions to the 'query' post. Logically, users who make more 'answer' posts are experts and users who make more 'query' posts are beginners in the

portal. This research will be using the z-score measures as introduced by Zhang *et al*. (2007) to classify users in the portal.

$$\sum_{i=1}^{n} Z_i = \frac{a_i - q_i}{\sqrt{a_i + q_i}} \tag{1}$$

Let Z_i be the z-score for user *i*, *i* =1 until *n* where *n* is the number of users, *a* is the number of 'answer' post made by a user and *q* is the number of 'query' post made by a user.
Once the z-score for all users are calculated, the value will be sorted in ascending order and will be mapped to an appropriate point as illustrated in Fig. 3. The top 20% of the users will be given 5 points. The last 20% of the users will be given 1 point. The other users will be given points as shown in Fig. 3. The top 20% of the users (based on an ascending order of calculated z-score) are the active contributors in the portal and will be given 5 points each.
The rationale behind this mapping system is that the experts are always the active contributors of the portal. This means, even when a user is an expert but if the user stops contributing to the portal, the user's level of expertise may drop if there are other users who contribute more. If there is a tie for the highest contributor, both users will be given 5 points.

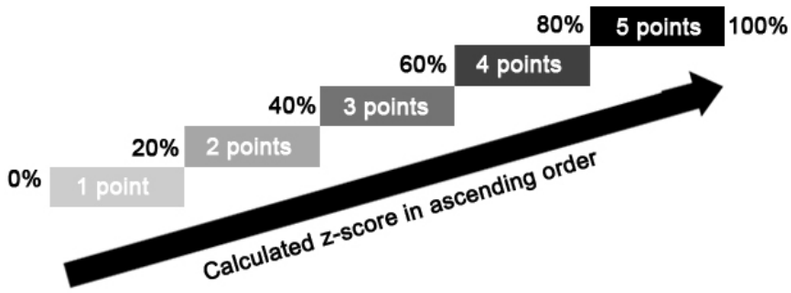


Fig. 3. Mapping of the z-score measures into a five-point scale

Table 3 shows an example of mapping the z-score measures. Let U_i be the users, i = 1 until n where n is the number of users, q is the number of queries posted, a is the number of answers posted, Z is the z-score measures (Zhang *et al.*, 2007) and M is the mapped z-score.
When users view 'answer' posts in the portal, they are required to rate by the scales: 0 (Unacceptable), 1 (Poor), 2 (Fair), 3 (Average), 4 (Very Good) or 5 (Excellent). By doing so, the sender of the post will receive points given by the other users. We treat all post equally, thus the user rating points, R is calculated by dividing the total points collected for each user, T with the numbers of users who make the rating, N. The purpose of user rating, R is to counter check the automatic classification using z-score measure (Zhang *et al.*, 2007).

$$\sum_{i=1}^{n} R_i = \frac{T_i}{N_i} \tag{2}$$

The final points (for each user); F is the average of the sum of mapped z-score, M and users' rating, R. The mapping of the final points, F to the expertise level, L is: expert E (4 or 5 points), intermediate I (2 or 3 points) and beginner B (0 or 1 points).

$$\sum_{i=1}^{n} F_i = \frac{M_i + R_i}{2} \tag{3}$$

				Z-score in ascending order		
U_i	q	a	Z	U_i	Z	M
U_0	0	0	0	U_0	0	0
U_1	5	0	-2.24	U_6	-7.07	1
U_2	0	5	2.24	U_9	-4.08	1
U_3	5	5	0	U_1	-2.24	2
U_4	10	5	-1.29	U_4	-1.29	2
U_5	5	10	1.29	U_3	0	3
U_6	50	0	-7.07	U_8	0	3
U_7	0	50	7.07	U_5	1.29	4
U_8	50	50	0	U_2	2.24	4
U_9	100	50	-4.08	U_{10}	4.08	5
U_{10}	50	100	4.08	U_7	7.07	5

Table 3. An example of mapping the z-score values to a five-point scale

Fig. 4 illustrates an overview of PBASE. Let U_i be the users, i = 1 until n where n is the total number of users, q is the number of queries posted, a is the number of answers posted, Z is the z-score measures (Zhang *et al.*, 2007), M is the mapped z-score, R is the users' rating, F is the final points and L is the level of expertise {B: Beginner, I: Intermediate, E: Expert}.

An example of classification using PBASE is shown in Table 4. Based on PBASE method, the user rating, R played an important role in classifying the users' expertise level. In the case of U_9, although the mapped z-score is the lowest (1 point), the users' expertise can still be promoted through the rating. For U_1 and U_6, the user rating, R will be automatically set to zero since the users do not make any 'answer' post and U_0 is an example of users who do not contribute anything in the SIG.

In addition, users are also allowed to flag posts if they find it inappropriate to the topic. After users flags certain posts, the administrator of the portal will be notified to take further action. Through the rating and flagging process, the users of the community are also contributing in giving point to users. As a result, members of the community also contribute to classify users' level of expertise in the portal.

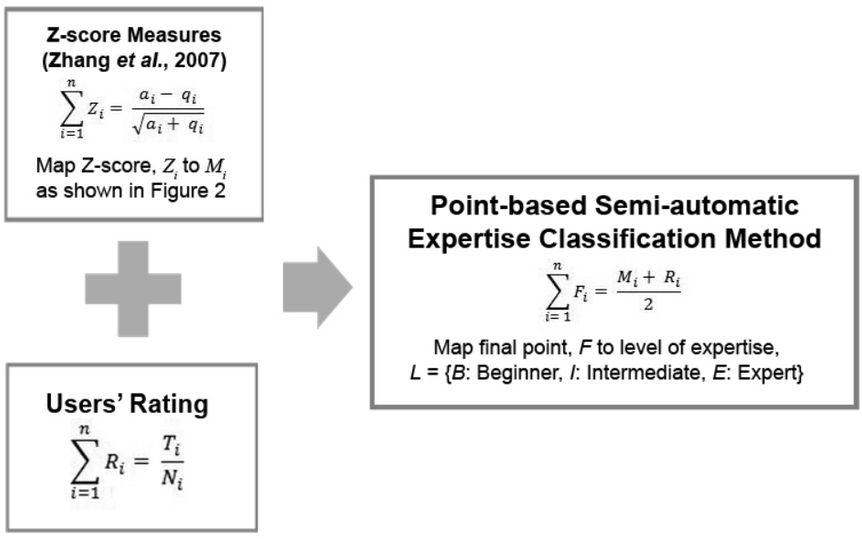


Fig. 4. Measures by PBASE method

4. Implementation and results

This research will use software engineering as the domain problem. We find that software engineering is an interesting domain as it concerns the creation and maintenance of software application by applying technologies and practices from computer sciences, project management, engineering, application domains, and other fields. The proposed PBASE method is applied in an existing web portal for software engineers in Malaysia called Malaysian Software Engineering Interest Group (MySEIG). The online interest group was founded in mid 2005 to provide a platform for software engineers to share knowledge, ideas and experience related to software engineering issues (MySEIG, 2009).

4.1 Knowledge representation

The field topics in MySEIG are based on Software Engineering Body of Knowledge or SWEBOK (Abran *et al.*, 2004) as listed in Table 5. For the convenience of MySEIG users to discuss common knowledge without specific software engineering topic, a general field named 'Others' is added.

Users are allowed to choose one or more field of interest from the listed topic. This means each user have a different set of field of interest. Example of users' set of field of interest includes: User A {Software Design, Software Testing, Software Maintenance, Others}, User B {Software Engineering Process, Software Quality, Others}, and User C {Software Requirement, Others}.

As mentioned previously, users' level of expertise reflects the knowledge in such SIG portals which the knowledge can be presented using a semantic net. Fig. 5 illustrates how semantic net represents users' level of expertise in MySEIG knowledge portal.

U_i	q	a	Z	M	R	F	L
U_0	0	0	0	0	0	0	*B*
U_1	8	0	-2.83	2	0	1	*B*
U_2	0	5	2.24	4	0	2	*I*
					1	2.5	*I*
					2	3	*I*
					3	3.5	*E*
					4	4	*E*
					5	4.5	*E*
U_3	7	7	0	3	0	1.5	*I*
					1	2	*I*
					2	2.5	*I*
					3	3	*I*
					4	3.5	*E*
					5	4	*E*
U_4	20	10	-1.83	2	0	1	*B*
					1	1.5	*I*
					2	2	*I*
					3	2.5	*I*
					4	3	*I*
					5	3.5	*E*
U_5	5	10	1.29	4	0	2	*I*
					1	2.5	*I*
					2	3	*I*
					3	3.5	*E*
					4	4	*E*
					5	4.5	*E*
U_6	40	0	-6.32	1	0	0.5	*B*
U_7	0	60	7.75	5	0	2.5	*I*
					1	3	*I*
					2	3.5	*E*
					3	4	*E*
					4	4.5	*E*
					5	5	*E*
U_8	30	30	0	3	0	1.5	*I*
					1	2	*I*
					2	2.5	*I*
					3	3	*I*
					4	3.5	*E*
					5	4	*E*
U_9	110	40	-5.72	1	0	0.5	*B*
					1	1	*B*
					2	1.5	*I*
					3	2	*I*
					4	2.5	*I*
					5	3	*I*
U_{10}	60	150	6.21	5	0	2.5	*I*
					1	3	*I*
					2	3.5	*E*
					3	4	*E*
					4	4.5	*E*
					5	5	*E*

Table 4. An example of classification

No	Topic
1	Software Configuration Management
2	Software Construction
3	Software Design
4	Software Engineering Management
5	Software Engineering Process
6	Software Engineering Tools and Methods
7	Software Maintenance
8	Software Quality
9	Software Requirement
10	Software Testing
11	Others

Table 5. Field of interest in MySEIG based on SWEBOK (Abran *et al.,* 2004)

4.2 Classification Process Using PBASE

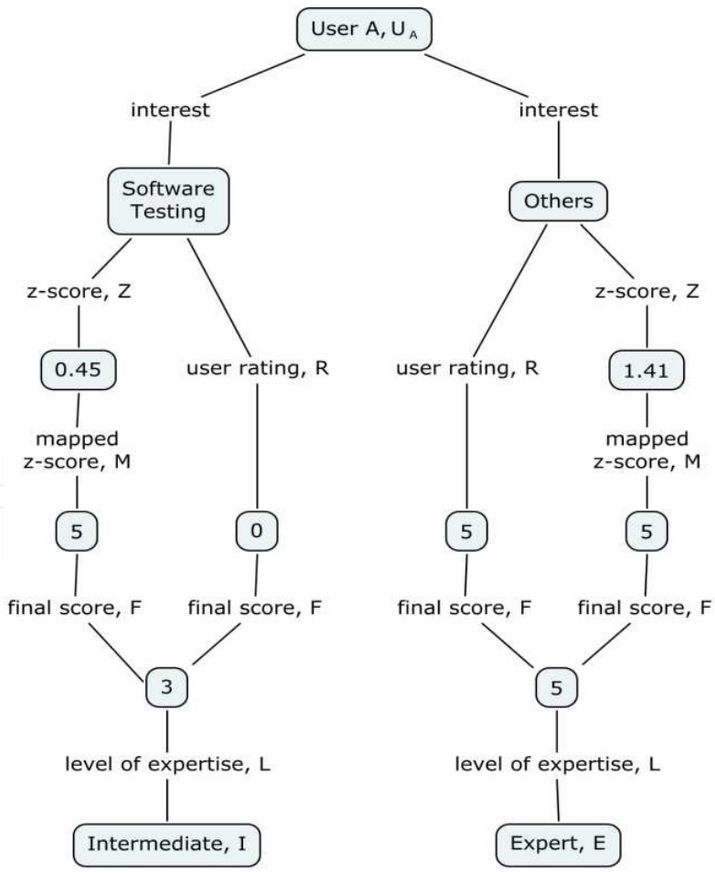


Fig. 5. Example of knowledge representation using semantic net

The first step of PBASE method in MySEIG knowledge portal is to calculate the z-score of each user. In order to calculate the z-score, we have to identify the type of posts or forums created. When users create a new post, they are required to choose the type of post from the dropdown list topic as illustrated in Fig. 6.

There are six types of post the users can choose from. The six types can be categorized into two; 'query' post and 'answer' post. The 'query' post include 'request', 'announcement' and 'question' while 'answer' post could be either 'opinion', 'information' or 'answer'. The default post is 'request'.

Then we can calculate z-score measures (Zhang *et al.*, 2007). After z-score values are calculated for every user under a certain field, we can map the z-score values to the five-point scale as shown in Fig. 3.

PBASE is a two-way classification method where its users also take part in the classification process. When users view 'answer' post, they are required to rate by the scales: 0 (Unacceptable), 1 (Poor), 2 (Fair), 3 (Average), 4 (Very Good) or 5 (Excellent) as shown in Fig. 7. Users also can contribute in the classification process by flagging post that they find inappropriate to the topic as in Fig. 7.

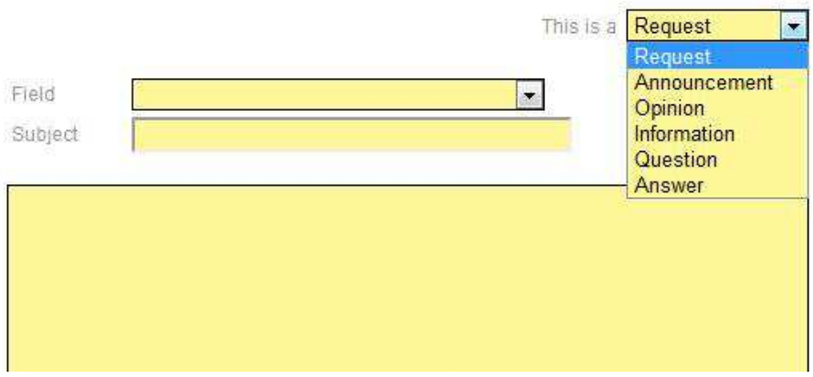


Fig. 6. Types of post to be determined by users before submitting a post

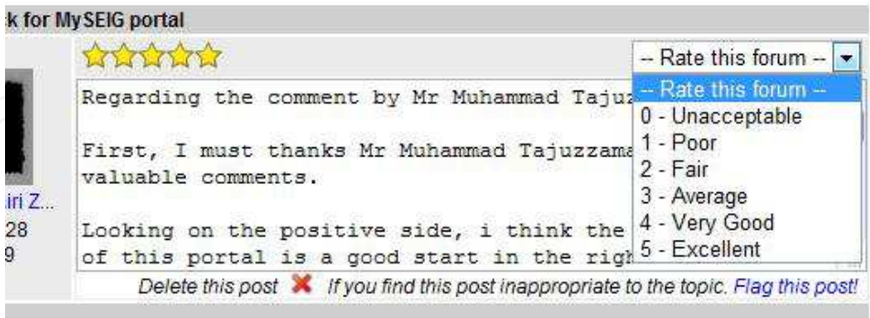


Fig. 7. Users' post that can be rated by other members in MySEIG knowledge portal

The z-score mapping in MySEIG by default will be as in Fig. 3 but the admin can change the z-score mapping as shown in Fig. 8. Admin also can reset the z-score mapping to a default value and the classification for each user will be recalculated.

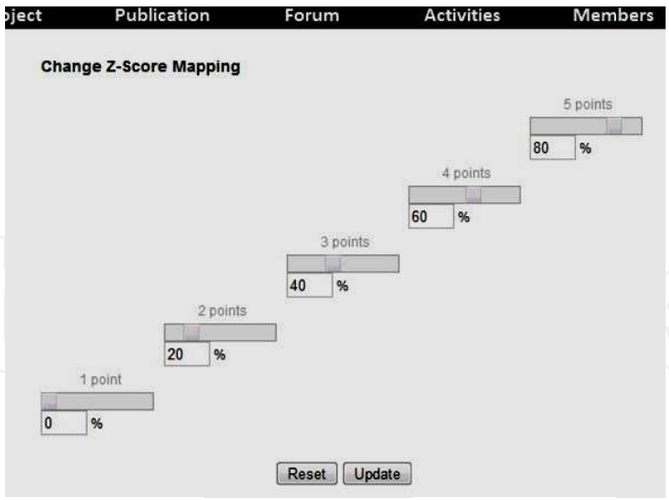


Fig. 8. Administrator can change the percentage of the five-point scale in z-score mapping

5. Qualitative evaluation

Characteristic	ITTutor.net (2009)	Computer Forum (2009)	PBASE
Direction of classification	One-way classification method where the users are not involved in the classification process	One-way classification approach. The users are not involved in the classification process but the administrator has the privilege in awarding selected users.	two-way classification method where users are involved in the classification process
Basis of classification	Classify its users based on the number of posts they created in the forum (See Table 1 and Table 2)		
Differences in type of post	No. All post treated equally and will be included in the classification process		Yes. The z-score measures will calculate the distribution of the questions and answers of each user
Competitiveness to be an expert	Not available because the expertise level of the user will not dropped.	Not available because the expertise level of the user will not dropped but users can be banned by the administrator if they are found violating the rules and regulation of the portal.	Yes because the expertise level of the user can dropped if the user stop contributing in the portal. Experts in the portal are always the current active contributors in the portal.

Table 6. Comparison of classification approaches

Comparison of PBASE with existing expertise classification method used in ITTutor.net (2009) and Computer Forum (2009) is listed in the following aspects:

(a) Direction of classification
(b) Basis of classification
(c) Differences in type of posts
(d) Competitiveness to be an expert

6. Conclusions and future works

Instead of using the conventional way to classify users based on the number of posts, this research proposes a two-way classification method called Point-Based Semi-automatic Expertise (PBASE). By proposing the PBASE method, we hope to maximize the capability of SIG knowledge portal for the convenience of its community members to seek help among the members.

Furthermore, we have identified that there is a limitation in identifying the type of posts. Based on the current approach, users are required to state the type of post. Thus as part of the future work, we plan to integrate Natural Language Processing (NLP) technique with PBASE. Hence, users will no longer need to state the type of post since NLP will automatically analyze and identify the type of posts.

Other future work include that the system should suggest automatically to other members list of people who in the same area or expert. In other word it involves either expert system or decision support system concept.

7. Acknowledgements

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

8. Reference

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

Computer Forum, Jelsoft Enterprises Ltd., 2009, http://www.computerforum.com.

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

ITTutor.net, 2009, http://ittutor.net.

Kleinberg, J. M., "Hubs, Authorities, and Communities", *ACM Computing Surveys*, Volume 31, Article No. 5, 1999.

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

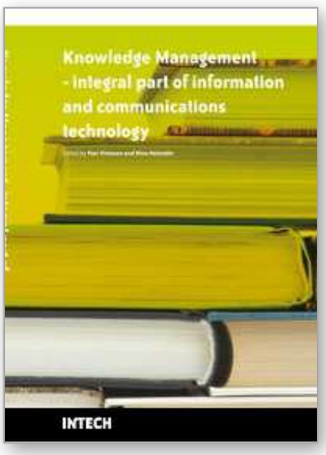
MySEIG, 2009, http://www.myseig.org.

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

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

Page, L., Brin, S., Motwani, R. and Winograd, T., "The PageRank Citation Ranking: Bringing Order to the Web", *Stanford Digital Library Technologies Project*, 1998.

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.



Knowledge Management

Edited by Pasi Virtanen and Nina Helander

ISBN 978-953-7619-94-7
Hard cover, 272 pages
Publisher InTech
Published online 01, March, 2010
Published in print edition March, 2010

This book is a compilation of writings handpicked in esteemed scientific conferences that present the variety of ways to approach this multifaceted phenomenon. In this book, knowledge management is seen as an integral part of information and communications technology (ICT). The topic is first approached from the more general perspective, starting with discussing knowledge management's role as a medium towards increasing productivity in organizations. In the starting chapters of the book, the duality between technology and humans is also taken into account. In the following chapters, one may see the essence and multifaceted nature of knowledge management through branch-specific observations and studies. Towards the end of the book the ontological side of knowledge management is illuminated. The book ends with two special applications of knowledge management.

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen