

The Presence of Beneficial Knowledge in Web Forum: Analysis by Kipling's Framework

Mohd Zaidi Abd Rozan^a and Yoshiki Mikami^b

^aGraduate School of Information Science & Control Engineering

^bProfessor, Department of Management & Information System Science

Nagaoka University of Technology, Nagaoka, Niigata, JAPAN

Tel : +81-258-471-1611 (ext. 3194), Fax : +81-258-479-350

E-mail : zaidi@mis.nagaokaut.ac.jp, mikami@kjs.nagaokaut.ac.jp

ABSTRACT

Words of doubt such as "What benefit can one get from web forums?" or "Web forums are meant just for small-talk" are examples of beliefs that were held by many individuals and researchers for a long time. Based on our observation of a web forum (<http://asamboi.org>), we found that at least quite a number of topics are dealing with deep knowledge contents. Our analysis of a so-called learning zone that contains 35 forum topics and 789 messages, utilizing Kipling's 5W1H framework six "service-men", found that the depth of details were ranked in descending order as 'Siapa' (Who), 'Apa' (What), 'Bagaimana' (How), 'Kenapa' (Why), 'Mana' (Where) and 'Bila' (When). We conclude that web forums are a good web-resource for digital age learning styles, especially for youth.

Keywords

Web Forums, Knowledge Details, Kipling's 5W1H Framework.

1.0 INTRODUCTION

Much has been said about the usage of web forums in cyberspace. In general, many readers will agree that a web forum is an online interaction medium that provides a positive environment for discussion. As it is mostly used for discussions and exchanging of ideas, the degree of knowledge contents involved are still viewed as near to the ground. This situation often leads to the assumption that web forums are full of less useful discourse where the only significant thing is its insignificance.

In this paper, we will draw the reader's attention to our attempt to analyze knowledge details based on the basics of 5W1H. The idea is to gauge the presence of knowledge details by assessing the depth of information that is being discussed plus knowledge that is being generated from it by the interlocutors.

1.1 Importance of this Study

Computer mediated communication (CMC) is evolving at a rapid rate and the only static shape is its dynamism. According to the Eight Malaysia Plan (2001-2005), 53.6% of the total Internet subscribers in Malaysia are concentrated in the Klang Valley, followed by 8.8% in Johor and 7.3% in Pulau Pinang (Jaring Internet, 2006). The high Internet penetration in Malaysia will always be a supporting factor to boost web based online interaction. It is certain that online communication is here to stay. Also based on a study by Mind Share Media Guide 2000 in (Jaring Internet, 2006), the most users are 20-29 years old with a 40% share, followed by user group 15-19 years old and 30-39 years old with 24% and 19%, respectively. The smallest group came from age 40 years and older, with a 16% share.

By analyzing the contents of the sample messages, we found that several discussions lead to certain information related to their age. Although not mentioned explicitly (by putting numbers representing their age), it is possible to figure out that interlocutors are in the range of 15-39 years old. The piles of discourse taken as sample were the results of this age group that carries the biggest portion of online users (83%). From our point of view, it is important to monitor and assess the discussion of this age group by keeping an eye on the knowledge details. This will assist in evaluating the strength of web

forums relative to knowledge quality. In order to observe such discourse, our long-term goal is to produce tools that enable discussion analysis and assessment of knowledge level.

2.0 KNOWLEDGE ESSENTIALS

2.1 Kipling Method, 5W1H Framework Utilization

Various frameworks are being studied to create a set of standard data categories that can be used for the analysis and description of knowledge. In the area of discourse analysis, intensive studies are ongoing under the ISO/TC37/SC4 umbrella (International Organization for Standardization, 2005), whose hope is delivering semantic content representation. In particular, the work under Thematic Domain Group 3 (TDG3) of ISO/TC37/SC4 on “Semantic Content Representation” is expected to produce a useful set of data categories for description of discourse relations, dialogue acts, referential structures and links. These studies, however, are still under development and not available yet.

In this article, we adopt the widely known 5W1H framework for this purpose. The 5W1H method has a long history of more than 50 years (Creatingminds.org, 2005) and is also known as the Kipling Method because of its originator, the Nobel Laureate of Literature in 1906, Rudyard Kipling. This method deals with six keywords that are easily attached with any possible words to create question expression. The keywords or also known as the six “service-men” are ‘who’, ‘what’, ‘where’, ‘when’, ‘why’ and ‘how’ (The Kipling Society, 2005). More explanation will follow in the next smaller subsections.

The major advantage of this approach is its ease of application. Even for such simplistic approach, the notion of question expressions are accepted as sophisticated classes of questions especially when involving how, why and what-if (Maybury, 2004). Specifically here, its capacity to facilitate the understanding of details in written text or spoken dialogue has proved the significance of applying these types of question expressions. Again, as suggested by Kipling, the pinpointing of essential information may lead to the comprehension of knowledge within it. In our analysis, the relatedness between the topic of discussion and its answers is also taken into account. This

contributes to the high quality of answers as it leads to a meaningful discourse.

As web forum messages are made in discourse mode, it is expected that various background knowledge is omitted as in the case of oral conversations. Also expected is that messages tend to be subjective and not to contain much objective knowledge. Hence, our challenge is to investigate whether these messages contain clearly stated meaningful and objective knowledge or not.

The details of 5W1H in the Malay language will be presented in the following sections. The sample phrases are created to simplify the explanation. The original samples of the messages mostly contained several language inaccuracies such as incorrect spelling, utilizing slang, mix with foreign language words and lingo and even the usage of Malay *sebang* language. An explanation of Malay *sebang* language can be found in (Mohd Zaidi & Mikami, 2006). For the sake of this paper’s objectives and simplicity, we treat all these inaccuracies by substituting them with standard Malay words. Another thing to keep in mind is the flexibility of the Malay language in utilizing 5W1H. Malay language tends to be very flexible, with a variety of forms that can be utilized for asking information. Certain elements of questions do not have one-to-one correspondence to English 5W1H. The reasons for will be explained in the next smaller subsections.

2.1.1 *Siapa* (Who) Questions

There are several possibilities for using ‘*siapa*’. For example, ‘*orang itu siapa?*’, or ‘*siapakah orang itu?*’ carry the same meaning as ‘who is that person?’ The first question is purely a question sentence but for the second question, there is an addition between the question sentence and question words (‘*-kah*’ or ‘*-lah*’). This is very much influenced by the position of the word ‘*siapa*’. Question words are not required if the word ‘*siapa*’ is at the end of a sentence (Mahat, 2003).

2.1.2 *Apa* (What) questions

The word ‘*apa*’ is used to retrieve information related to name, type and property. For example, ‘*apakah yang dia makan?*’ is translated as ‘what is he/she eating?’ A simple phrase ‘*apa khabar?*’ means ‘how are you?’ (literally means, what’s

news?). In this case, it is different from English as 'apa' actually means 'how' for this specific question.

2.1.3 *Mana* (Where) questions

'*Mana*' is used to ask questions regarding location or to acquire answers related to a 'where' question. '*Ke mana engkau hendak pergi?*' means 'where do you want to go?'. The sentence '*macam mana awak selesaikan masalah itu?*' represents 'how did you solve that problem?'. This is different from the previous sentence that only uses the word '*mana*' and not '*macam mana*'. The function of the word '*mana*' has been changed from 'where' to 'how' just by adding the word '*macam*' before it.

2.1.4 *Bila* (When) questions

For the most part, '*bila*' is straightforward and easy to use. It has a similar meaning to the word 'when'. A sentence like '*bila perlu saya hantar laporan ini?*' translates as 'when is it necessary for me to submit this report?'. The usage is very simple as the intended request is only a time contextual answer.

2.1.5 *Kenapa* (Why) questions

The reason for such questions is to get information on the purpose of having or doing something. '*Kenapa awak mengikut jalan ini?*' means 'why do you go along this road?'. The answer must be an explanation phrase that usually provides some particular reason.

2.1.6 *Bagaimana* (How) questions

The word '*bagaimana*' is used for requesting an answer similar to 'how' or 'in what way' in English. '*Bagaimana perkara itu terjadi?*' has similar meaning to 'how did that happen?'. The answer usually describes a process, way or method with a task. Again, as mentioned in 2.1.3, '*bagaimana*' also has a similar implication to '*macam mana*'.

3.0 METHOD

3.1 Extraction and Preprocessing

Our study is focused on one "Forum Zone" so-called '*Zon Pembelajaran*' or Learning Zone if translated literally into English. This zone contains 35 forum topics and holds a total number of 789 messages (as of 05/01/2005, 1530hrs). Since the analysis is for these 789 messages, it is therefore unnecessary to explain each of the many other available forum zones and forum topics in our selected web forum (<http://asamboi.org>).

Technically, all the data for web forums are commonly kept in its database server and can only be manipulated by the web administrator. However, all the information that is visibly displayed on the internet browser while accessing this web forum can be extracted. Utilizing custom written script is a typical approach. For this reason, we created a script in *Perl* that specifically extracts the poster name, message details (time and date posted) and the attached messages. Such extraction is very efficient, effective and easy to implement as long as the script is written for a specific web forum. This is because most web forums have different schematic arrangement and vary in the usage of tag type and position.

A special challenge faced by our study is to deal with the complicated nature of web forum discourse. The phrases that are used are mostly very simple but created several hurdles when we tried to assess its 5W1H. The extraction of details is challenging, since the phrases normally include different sorts of registers. Linguistically, it means that not only one single writing style is used for communication. Another point to look at is the question and answer pairs or series. In this medium, it is rare to find answers that follow a question without being interrupted by a non-question phrase. For instance, teasing and joke phrases that have no relation to a question exist, intermittently functioning as an interval friendly message. This acts as a fun situation and has no direct meaning as an answer.

Since the 5W1H analyses were conducted wholly by humans, the complicated factors were alleviated. However, the analysis still took several days to finish. On the other hand, several preprocessing operations must be done to the

messages prior to the 5W1H evaluation. These involve removing emoticons, animated gifs, spell checking, substituting foreign words with Malay and utilizing standard Malay words as a conversion to slang and *sembang* words. Figure 1 shows the process flow from web information extraction to message analysis.

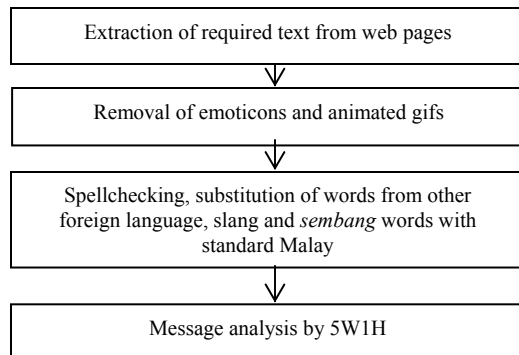


Figure 1: Process Flow for the operations of Extraction, Preprocessing & Analysis

3.2 Knowledge Details Analysis

Every message should hold some elements of knowledge; however, the depths of its detail are not similar. We found out that the first message in a topic would usually act as the central question. Quite a number of these messages do not merely carry question phrases, but are also followed by a series of explanations as well as answers. It is because the presenter put together his or her questions and point of view in one message (2-in-1). Here, such answers are also being assessed for their knowledge details. Furthermore, another observation is that the messages that come after it are either answer phrases or just elaborations of the question. If the central message requests comments or views, answers in affirmative phrases (approving/disapproving) are very likely. In most cases, the messages would contain answers and feedback that could lead to several more answers.

In our analysis, we have found that some messages are not likely to possess any objective knowledge. For example, the act of laughing 'hahahaha' is not considered to have any answer for the 5W1H elements. This is because here it functions only as a friendly gesture together with a message as mentioned before in subsection 3.1.

Here, the explanation of the weighting method for analysis is given. As points are given to the

nominal type attributes of all six variables for 5W1H, some messages may hold answers to more than one specific question type. For example, Message 1 is answering a '*bagaimana*' question. Here the underlined words represent the answer to the question and they occurred three times. Because the analysis looks for the presence of at least one answer, then assigning '1' as a presence is more suitable, rather than allotting '3' for it.

Message 1: 'masukkan Cd lepas tu restart pc...
tp pastikan boot priority letak
cdrom as first boot device'
(put in CD and restart PC... But make sure the boot priority is set to CDROM as first boot device)

Message 2: 'try dulu .. kalau per per nanti
tanyer a saya pon dah lama tak
buat tanyer kekawan yang biase
buat'
(First try.. if have anything ask it's been a long time I didn't do it ask other friends that normally do it)

As for Message 2, the underlined words are actually a suggestion, semantically they mean 'first try'. These words are not providing answers to 5W1H and together with the rest of the following words in Message 2.

4.0 ANALYSIS & RESULTS

From our analysis, we simplified the results by tabulating the relevant information in Table 1. The forum topic title is presented as it appears on the Web. The meaning in standard Malay and English can be found here in the endnotes¹

For forum topic no. 1, the 'who' and 'what' scored nearly equal. We found out that most answers have both elements that are combined together. We observed this phenomenon since some discussions have messages that exist not simply to answer the question of 'where' but also in need to have 'who' as an element to back up their suggestion. Here, most messages include their friends' or their own experiences that are set-down together as narrative. Narrative or story telling has been accepted as a practical technique for knowledge disclosure, communication and cultural intervention in this new age of uncertainty, as stated here (Snowden, 1999).

Table 1: Forum Messages and Depth of Details

| Forum Topic No | Forum Topic Title | No of messages | Depth of Details & Percentage (%) | | | | | |
|----------------|---|----------------|-----------------------------------|------|-------|------|-----|-----|
| | | | Who | What | Where | When | Why | How |
| 1 | universiti, politeknik, kolej dan lain2? | 26 | 77 | 73 | 65 | 46 | 46 | 35 |
| 2 | 8A...PMR... | 39 | 77 | 31 | 18 | 21 | 54 | 33 |
| 3 | Teknik-teknik menguasai Bahasa Inggeris | 41 | 51 | 22 | 22 | 29 | 27 | 59 |
| 4 | Belajar bahse jepung | 55 | 71 | 85 | 16 | 7 | 24 | 35 |
| 5 | Saper nak belajar bahasa Jepun?? | 101 | 81 | 61 | 14 | 11 | 17 | 22 |
| 6 | UPSR 2005 | 11 | 91 | 91 | 27 | 27 | 36 | 9 |
| 7 | SPM 17A1WAUU!!!! | 25 | 92 | 72 | 16 | 24 | 32 | 36 |
| 8 | mira pon nk blajar gak... | 3 | 100 | 100 | 100 | 67 | 100 | 67 |
| 9 | Magical of Maths..! | 16 | 56 | 56 | 25 | 13 | 44 | 31 |
| 10 | kenali MRSM | 4 | 50 | 100 | 0 | 50 | 50 | 0 |
| 11 | asikom... | 13 | 85 | 54 | 8 | 54 | 15 | 46 |
| 12 | please.....tolon2 | 14 | 50 | 43 | 21 | 14 | 14 | 14 |
| 13 | mata pelajaran sex di sekolah 02 | 5 | 100 | 60 | 60 | 60 | 60 | 60 |
| 14 | Hala Tuju selepas habis sekolah??? | 59 | 71 | 49 | 27 | 19 | 25 | 53 |
| 15 | tips utk pelajar2 ipt | 15 | 57 | 71 | 14 | 29 | 43 | 29 |
| 16 | camne nak kuasai pelajaran matematik??? | 29 | 55 | 10 | 0 | 0 | 17 | 76 |
| 17 | Berbaloi ke study kat OVERSEA? | 41 | 41 | 15 | 20 | 10 | 61 | 39 |
| 18 | tips dan trick utk students... | 11 | 45 | 9 | 9 | 18 | 18 | 82 |
| 19 | utk mereka yang mencari tempat praktikal | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 20 | keyakinan diri | 7 | 57 | 14 | 14 | 14 | 14 | 43 |
| 21 | Doakan Aku Berjaya Dalam PMR | 7 | 71 | 14 | 0 | 15 | 29 | 29 |
| 22 | mata pelajaran seks di sekolah.... | 62 | 48 | 23 | 11 | 5 | 32 | 58 |
| 23 | ::Selamat Hari Guru..Tanpa guru..siapalah kita:: | 9 | 78 | 56 | 33 | 44 | 44 | 11 |
| 24 | STPM students | 10 | 70 | 40 | 10 | 10 | 10 | 60 |
| 25 | Ilmu..... | 9 | 56 | 44 | 22 | 33 | 22 | 67 |
| 26 | Website designer cam mana nak wat nak yg memuaskan hati | 10 | 40 | 20 | 20 | 30 | 10 | 70 |
| 27 | Blaja mana? | 40 | 70 | 5 | 45 | 28 | 15 | 15 |
| 28 | help2 me..... | 7 | 43 | 14 | 71 | 0 | 0 | 14 |
| 29 | PENDIDIKAN ASAS FARDHU AIN (PAFA) | 6 | 67 | 17 | 50 | 100 | 50 | 17 |
| 30 | Need ur idea ..plz. | 6 | 83 | 50 | 33 | 33 | 17 | 100 |
| 31 | Betul ke orang yang gi kelas tambahan nie pemalas ? | 32 | 59 | 34 | 25 | 0 | 78 | 16 |
| 32 | Perlukah pengawasan pelajar diperketatkan? | 29 | 38 | 28 | 14 | 14 | 34 | 34 |
| 33 | matapelajaran vokasional..... | 14 | 43 | 50 | 21 | 7 | 57 | 71 |
| 34 | Form 6 buang masa??? | 9 | 33 | 33 | 11 | 11 | 33 | 22 |
| 35 | cinta:guru ngan murid... | 23 | 91 | 61 | 17 | 43 | 52 | 43 |

Another interesting point comes from topic titles no.4 and no.5. Both titles were concerned with learning Japanese. It is most expected that forum topic no. 5, has the 'who' element scored highest (81%) because the reason for this message was to look for members who are interested in learning about the Japanese language. In addition, 'what' scored 61% as the contents provide many listings of Japanese to Malay translation. This appears to

correlate with topic title no.4, where 'what' has 85% as its peak: this was also because of such similar Japanese-to-Malay translation listings in most messages found. Topic no. 8 contains a request from a person named 'Mira', in which he mentioned his eagerness to learn Japanese. Anyway, the topic was closed by the moderator due to the availability of topic no. 5 in this

learning zone. Topic no. 4 was later closed as it contains similar topic of discussion as in no. 5.

The forum topic entitled 'mata pelajaran sex di sekolah' or 'sexual education in school' existed as two versions in this *learning zone*. The opening topic was in title no. 22 and the later is title no 13. The later topic was created as an extension of no. 22, in any event, it did not last for very long. Looking at topic no. 22, the depth of details is towards 'how'. It is interesting to see that the discussion started as merely asking for comments on the pros and cons of having sex education in schools. However, more than half of the total messages include 'how', and towards the end of the messages, the forum member decided to compile their suggestions and submit them to the education office (the results of this action is not stated anywhere). The 'how' in implementing this course of study from their own point of view includes how to create a proper name for the subject so that students will feel at ease when learning it and suggesting the role of parents (or close family members) to deal with this specific education topic.

Generally, most of the topics were answered well by the interlocutors, except for title no. 19 which was targeted at students looking for a practical training site. It seems that not even a single reply was recorded. The reason for this is unknown. Since the first message is only a link to a government website offering a practical training spot, most probably it is not that alluring because of a time-specific offer by the government department and very weak interest in it from forum members.

From this brief explanation, all six "service-men" (5W1H) played an important role in 35 forum topics. Every topic has its own distinctive demand on different "service-men" and this has shown that all six "men" serve useful function. On average 'who' took the biggest portion of work (65%) and the smallest was 'when' (19%) for the whole 789 messages. The busiest "service-men" (who) occurred possibly because of the presence of story referencing to a person. They generally know such people by referring to the real person's name or a job function such as "my teacher". For "service-men" (when), it was rare to get him to work regularly because, not many topics have strong concern to him.

5.0 CONCLUSIONS

Web Forum discussions are organized better than normal oral conversations viewed from the knowledge description framework, probably because non face-to-face conversation requires more explicit statement of background knowledge. On the other hand, such transfer of knowledge is supported by very friendly interactions among forum members.

One advantage of a web forum compared to oral discussion is the fact that it can provide link information. Through this, participants can gain a great deal of knowledge by participating in the forum. It does not require going to a library or referring to dictionary pages. How to assess this advantage, is a matter for another discussion.

In this paper, we have developed further understanding for recognizing the discourse properties of messages using the 5W1H method. We also provide answers to the question posed in the title of this paper, *the presence of beneficial knowledge in web forum*. A study based on a more sophisticated discourse description framework should be the next agenda.

Although the issues discussed in this paper center on the Malay language, still it has demonstrated the importance of building a knowledge extraction and analysis system. Our next step will be developing an agent that will carry out knowledge detail analysis and assessment using Natural Language Processing (NLP). This agent will be developed specifically for web forums environment accessing accordingly to its knowledge depth.

REFERENCES

- Creatingminds.org. (2005). The Kipling Method (5W1H). Retrieved November 30, 2005, from <http://creatingminds.org/tools/kipling.htm>.
- International Organization for Standardization (ISO). (2005). Technical Committee 37 (TC37), Sub-Committee 4 (SC4), Switzerland; Retrieved December 11, 2005, from <http://www.iso.org>.
- Jaring Internet. (2006). Malaysian Readiness Towards K-Society, A Statistic Measurement. Retrieved January 10, 2006, from <http://www.magazine.jaring.my/2001/august/index2.html?content=stay9.html>.

Mahat, A. R. (2003). *Sukatan Pelajaran Bahasa Melayu Pusat Perkembangan Kurikulum*. Kementerian Pendidikan Malaysia, Malaysia.

Maybury, M. T. (2004). *Knowledge and Inference-Enabled Question Answering*. New Directions in Question Answering. MIT Press. Massachusetts.

Mohd Zaidi, A. R & Mikami, Y. (2005). *Bahasa Sembang in Web Forums: Knowledge Management for Piles of Atopian Discourse*, International Conference on Web Information Systems and Technologies (WEBIST2006), Setubal, Portugal. (coming conference on 11-13th April 2006).

Snowden, D. (1999). *Story Telling: An Old Skill in a New Context*. Speakers Notes in Information and Knowledge Management Society of Singapore Conference (IKMS). March 1999.

The Kipling Society. (2005). *The Elephant's Child* poem, "I Keep six honest serving-men: Their names are What and Where and When And How and Why and Who". Retrieved December 11, 2005 from http://www.kipling.org.uk/poems_serving.htm.

tighten the supervision of student?; 33. Vocational subjects; 34. Secondary 6 a waste of time?; 35. Love: Teachers with students.

i

Meaning in English for Forum Topics: 1. University, polytechnic, college and others; 2. 8A, Lower Secondary Evaluation Exam; 3. Techniques to master English language; 4. Learning Japanese Language; 5. Anybody wants to learn Japanese Language?; 6. Primary School Achievement Test 2005; 7. Malaysian Certificate of Education 17A1 Wow!; 8. Mira also wants to learn; 9. Magical of Maths!; 10. Get to know Mara Junior Science College; 11. Peace be upon you; 12. Please help; 13. Sex education at school_02; 14. Destination after finish schooling; 15. Tips for higher institution students; 16. How to master the subject of mathematics?; 17. Worth studying at oversea?; 18. Tips and tricks for student; 19. For those looking for practical training site; 20. Self confidence; 21. Pray for my success in PMR; 22. Sex education in school; 23. Happy teachers day, who are we without teacher; 24. Malaysian Higher Certificate Education Student; 25. Knowledge; 26. Website designer how to do with satisfaction; 27. Where do you study; 28. Help me; 29. Fardhu Ain Basic Education; 30. Need your help...please; 31. Is it true, people goes to extra class are lazy?; 32. Is it necessary to