Knowledge Details in Web Forums: How High or Low above the Ground?

Mohd Zaidi Abd Rozan[1], and Yoshiki Mikami[2]

1 Graduate School of Information Science & Control Engineering
2 Professor, Department of Management & Information System Science
Nagaoka University of Technology, Nagaoka, Niigata, Japan, 940-2188
zaidi@mis.nagaokaut.ac.jp, mikami@kjs.nagaokaut.ac.jp
WWW home page: http://gii.nagaokaut.ac.jp/gii

Abstract. While entertainment web forums provide dynamic medium for interaction, not many researchers feel the need to go deeply into the contents. One of the reasons behind this attitude lays on a widely perceived assumption that web forums does not deal with knowledge matter and has the inclination to take place only as a small talk. In this paper, we will mainly inspect the components of knowledge as are being handled by the *six service-men* of Kipling's 5W1H framework. Based on our observation on a *learning zone* that contains 35 forum topics and 789 messages in a web forum (http://asamboi.org), we found out at least several topics are dealing with deep knowledge contents. From our analysis, we have found that the depth of knowledge details is reasonably significant particularly when responding to specific question demand. On average the depth of details obtained were ranked as 'Who' *Siapa*, 'What' *Apa*, 'How' *Bagaimana*, 'Why' *Kenapa*, 'Where' *Mana* and 'When' *Bila* once sorted descendingly. We conclude that web forums are a good web-resource for digital age learning styles as it provides detailed knowledge for networked learning.

1 Introduction

Web forums serve as means for discussions and exchanging of ideas in a positive environment, however the degree of knowledge contents involved are 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.

Based on www.e-keluarga.com branded as the *Family Information Media* in Malaysian cyberspace, Asamboi.org is positioned first on its ranking [1]. This result was possibly due to the high poll of web users whom admired Asamboi.org very much (well-known *Entertainment Community* web forum). However, from our point

of view it is important to assess the web forums discourse contents relatively to its depth ness of information and not based individually on its popularity. Based on Kipling's 5W1H, our attempt is to analyze knowledge details by exploratory gauging the presence of specific information appearing in such discourse.

2   Kipling Method, the 5W1H Framework Utilization & Interpretation in Malay Language

Various frameworks are under study to create a set of standard data categories that can be used for the analysis and description of knowledge. In area of discourse analysis, intensive studies are going under the International Standard Organization (ISO)/Technical Committee 37 (TC37)/Sub-Committee 4 (SC4) umbrella [2]. Works under the Thematic Domain Group 3 (TDG3) of ISO/TC37/SC4 on "Semantic Content Representation" is expected to produce certain 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. 5W1H is also known as Kipling Method from its originator, the Nobel Laureate of Literature in 1906, Rudyard Kipling [3]. This 5W1H framework deals with six keywords that are also known as the *six service-men*; 'who', 'what', 'where', 'when', 'why' and 'how'. To understand the meaning of 5W1H in Malay language, in Table 1, the interpretations are presented very briefly with its literal meaning and example.

Table 1. 5W1H Interpretation

English	Malay	Example
Who	Siapa	*Orang itu siapa?* similar to 'who is that person?'
What	Apa	*Apa yang awak makan?* similar to 'what do you eat?'
Where	Mana	*Mana awak mahu pergi?* similar to 'where do you want to go?'
When	Bila	*Bila perlu saya pergi?* similar to 'when should I go?'
Why	Kenapa	*Kenapa awak perlu belajar?* similar to 'why do you need to study?'
How	Bagaimana	*Bagaimana perkara itu terjadi?* Similar to 'how did it happen?'

   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 [4] as it requires complicated rules for semantic content analysis. Specifically here, its capacity to facilitate the understanding of details in written text or spoken dialogue has proved the significance of applying this type of question expressions.

   On the other hand, we also found out that this framework could create a very lenient rule as it offers six keywords that can easily be attached to any possible words in order to create question expressions. The tendency to create very general type of questions meant for one specific topic are also high. We believe that this is the limitation of the approach since it leads to very open-ended non-specific discussions. To avoid this, we must make sure that the relatedness between the topic of discussions and its answers must be taken into account. Hence, our challenge is to

investigate whether these messages contain clearly stated and objective knowledge that leads to a meaningful discourse.

3   Knowledge Details Analysis

Our study is focused to one Forum Zone so-called *Zon Pembelajaran* or Learning Zone that contains 35 forum topics and holds a total number of 789 messages (as of 05/01/2005, 1530hrs). This is part of the content for Asamboi.org, an *entertainment community web forum* in Malaysian cyberspace (http://asamboi.org).

Here, the explanation of the weighting method for analysis is given. As the points are of *nominal* type for all six variables of 5W1H, some message may hold answers to more than one specific question type. For example in Message 1, this phrase is answering a *'bagaimana'* or 'how' question. Here the underlined words are representing the answers for the question and it occurs 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: <u>*masukkan Cd*</u> *lepas tu* <u>*restart pc*</u>*… tp* <u>*pastikan boot priority letak cdrom as first boot device*</u>
> (<u>put in CD</u> and <u>restart PC</u>… But <u>make sure the boot priority is set to cdrom as first boot device</u>)

As for Message 2 below, the underlined words are only a suggestion that means 'first try'. These words are not providing answers to 5W1H and together with the rest of the words in this message.

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

4   Analysis & Results

From our analysis, we simplify the results by tabulating the relevant information in Table 2. However, to comply with the maximum page requirement for this paper, only four of 35 forum topics are shown here. The full forum topics and scores can be accessed here http://gii.nagaokaut.ac.jp/~zaidi/IFIPTable2.htm

Forum topic number one relates to the selection of higher-level learning entry points, such as universities, polytechnics, colleges and others. 'Who' and 'what' are found to have relatively higher score compared to the rest. Here, most messages have both elements and this phenomenon occurred since some discussions have messages exist not only to answer the question of 'what' but in need to have 'who' as a function to back-up their suggestion. Here most messages include their friends' or their own experiences, which are set-down together similar to story telling. Story

telling is accepted as a practical technique for knowledge disclosure, communication and cultural intervention as stated here [5]. As the topic also discusses about their preferred destination, timing and specific reason after they have completed their studies in schools, details for 'where', 'when' and 'why' also have relatively higher percentage.

Table 2. Forum Messages and Depth of Details

No	Forum Topic Title*[i]	No of messages	Details Type & Percentage (%)					
			Who	What	Where	When	Why	How
1	universiti,politeknik, kolej dan lain2?	26	77	73	65	46	46	35
2	Berbaloi ke study kat OVERSEA?	41	41	15	20	10	61	39
3	mata pelajaran seks di sekolah....	62	48	23	11	5	32	58
4	Betul ke orang yang gi kelas tambahan nie pemalas ?	32	59	34	25	0	78	16

Titles were taken as it is from http://asamboi.org

Topic number two plays around the matter of whether it is worth studying overseas. The initiator started the forum by exposing the issue of extremely high expenses for studying abroad compared to local as well as the factor that graduates received almost similar income during recruitment. Most messages include the explanations of why it is best to go abroad and why it is good to stay in the country, hence, the responses were mostly on the 'why' for the matter concerned, here the score is found to be 61%.

Forum topic entitled *'mata pelajaran sex di sekolah'* or 'sex subjects in school' has the depth of details '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 members decided to compile their suggestions and submit it to the education office (yet the results of this action is not stated anywhere in the forum). Here, the 'how' in implementing sex education includes, how to create 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.

The fourth topic is about whether people who go for extra classes are lazy. Debate on the 'why' element covered more than 75% of the dialogues. The main rationale is because most forum members are against the statement of the message initiator at the first place. Here in most of the message, the reason why students go to supplementary classes are given (laziness is not a factor why people go for extra classes) and it is frequently opposing to the forum title.

From this brief explanation, all *six service-men* (5W1H) played an important role in 4 forum topics even with diverse scores. 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 (53%) and the smallest was 'when' (12%) carried out for these 161 messages. Highest working *service-men* (who) occurred possibly because of referencing stories to any person (generally

known or unknown by them such as a real person's name). Specifically for this study, it was infrequent to get the lowest working *service-men* (when) to work more because not many topics have strong concern to it.

5   Conclusions

Web Forum discussions 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 are very much influenced by the very friendly interactions among forum members.

A big advantage of a web forum compared to oral discussion is the fact that it can provide link information. Through this, participants can learn from a vast pool of knowledge by participating in the forum. It does not require going to a library or referring to dictionary pages. Assessing the effectiveness of disseminating knowledge via web forums is a point to look at in future and could be one exiting challenge.

In this paper, we have gained understanding for recognizing the discourse properties of messages using the 5W1H approach. This has provided answer to the question as in the title of this paper.

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

1.   E-keluarga.com, Malaysia (March 27, 2006); http://www.e-keluarga.com
2.   International Organization for Standardization (ISO), Technical Committee 37 (TC37), Sub-Committee 4 (SC4), Switzerland (November 25, 2005); http://www.iso.org
3.   The Kipling Society, England (December 11, 2005); http://www.kipling.org.uk/poems_serving.htm
4.   M. T. Maybury, Knowledge and Inference-Enabled Question Answering, New Directions in Question Answering (MIT Press, Massachusetts, 2004).
5.   D. Snowden, Story Telling: An Old Skill in a New Context, Information and Knowledge Management Society of Singapore Conference (IKMS), 1999 (unpublished speakers' notes).

[i] 1. University, polytechnic, college and others; 2. Is it worth studying at overseas?; 3. Sex subjects in school; 4. Is it true, people who go to extra classes are lazy?