

Association for Information Systems

AIS Electronic Library (AISeL)

ICEB 2007 Proceedings

International Conference on Electronic Business
(ICEB)

Winter 12-2-2007

Summary of a Topical Forum FAQ Based on the Chinese Composition Structure

Yu-Hui Tao

Shu-Chu Liu

Chih-Lung Lin

Follow this and additional works at: <https://aisel.aisnet.org/iceb2007>

This material is brought to you by the International Conference on Electronic Business (ICEB) at AIS Electronic Library (AISeL). It has been accepted for inclusion in ICEB 2007 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

SUMMARY OF A TOPICAL FORUM FAQ BASED ON THE CHINESE COMPOSITION STRUCTURE

Yu-Hui Tao, National University of Kaohsiung, Taiwan, ytao@nuk.edu.tw

Shu-Chu Liu, National Pingtung University of Science and Technology, Taiwan, scliu@mail.npust.edu.tw

Chih-Lung Lin, National Pingtung University of Science and Technology, Taiwan, ngo0591@gmail.com

ABSTRACT

An automatic multiple-document summarization system for producing frequently asked questions (FAQ) of a topical forum can save forum Webmasters a great deal of time in theory. This work will address summary composition issue of a previous work by proposing a structured presentation based on a four-part pattern of traditional Chinese articles. The result of the experiment shows that the enhanced system with both domain-terminology corpus methods produced a significantly better summary presentation than the original system. Recall rate and precision rate performance indices and user evaluations are also presented and discussed to show their practical implications.

Keywords: Chinese composition structure, FAQ, presentation, summarization, topical forum.

INTRODUCTION

From the perspective of knowledge management, the long existing Frequently Asked Questions (FAQ) has been a common knowledge sharing format used in Internet newsgroups, bulletin boards, forums, and virtual communities. An FAQ is usually maintained by a Webmaster who manually summarizes the frequently asked topics from related articles into static question-and-answer format for easy access by its members. To take the advantage of the text summarization technologies for implementing useful application, Tao et al. [9] proposed a conceptual model for automatically transforming topical forum articles into FAQ summary, and empirically demonstrated the acceptability of this model via a prototype system. Their experiment implied the time and manpower savings in producing FAQ and illustrated the technical feasibility of such a model.

而這樣的過程既稱為正規化動作	(第11篇文章, 第3句) (過程, 稱為, 動作)
正規化不僅限於關聯式檔案	(第13篇文章, 第4句) (限於, 關聯, 檔案)
也是索引檔案常見的設計方式	(第13篇文章, 第5句) (索引, 常見, 設計, 方式)
正規化既是捨棄重複群組	(第13篇文章, 第9句) (捨棄, 重複)
消除部分相依的組合索引鍵	(第13篇文章, 第11句) (消除, 部分, 相依, 索引)
定義	(第14篇文章, 第8句) (定義)
一方面節省儲存空間	(第14篇文章, 第9句) (節省, 儲存, 空間)
藉由程序方式解決以上可能發生的問題	(第14篇文章, 第13句) (程序, 方式, 解決, 以上, 發生, 問題)
刪除	(第14篇文章, 第18句) (刪除)
在這分配的過程中	(第15篇文章, 第5句) (分配, 過程)
且降低儲存空間的使用量	(第15篇文章, 第12句) (降低, 儲存, 空間, 使用)
以決定關聯表應該擁有那些屬性	(第16篇文章, 第1句) (決定, 關聯, 擁有, 屬性)
其目的是建立良好結構關聯表Well	(第16篇文章, 第2句) (建立, 良好, 結構, 關聯, 其目的)
去除重複性Eliminating Redundancy	(第16篇文章, 第13句) (去除, 重複, 性Eliminating)
這是因為將不應該放在同一個關聯表的屬性拆放或硬置於同一個關聯表	(第16篇文章, 第26句) (屬性, 拆放, 關聯)
稱為非正規化型式Unnormalized Form關聯表	(第16篇文章, 第29句) (稱為, 型式, 關聯)
以期達到提高查詢的速度	(第18篇文章, 第5句) (達到, 提高, 查詢, 速度)
因為存放重複的訊息	(第19篇文章, 第7句) (存放, 重複, 訊息)
利用功能相依性以及主鍵來達到減少儲存空間	(第21篇文章, 第1句) (利用, 功能, 相依, 達到, 減少, 儲存, 空間)
不合正規化的關聯無法通過正規化檢查	(第21篇文章, 第3句) (關聯, 通過, 檢查)
必須設法拆解成更小的關聯	(第21篇文章, 第4句) (設法, 拆解, 關聯)

Figure 1. Summary ordered by article and sentence sequences

Unlike news or technical papers, Internet forum articles often suffer from short, unstructured, and even incomplete article content; incomplete format of sentences or paragraphs; and divergent concepts across multiple response articles of a specific original posting. Therefore, Tao et al.'s [9] model and prototype system adopted the simple extraction approach to produce a generic summary. Two issues are worth of concern. The first is that the summary was not presented as a typical article structure familiar to its readers. Specifically, the sentences were redundant and presented based only on the posting order of their corresponding articles without any further treatment for better readability. Another issue is that domain terminologies were not extracted appropriately. This is due to the prototype system which extracted important sentences from multiple articles by

comparing the Chinese character combinations to the Chinese Electronic Dictionary (CED) developed by the Computational Linguistics Society of R.O.C. Although CED contains 80 thousand Chinese word combinations, special domain jargons or terminologies are not covered. This issue has been addressed in the work of Tao et al. [10] and must be integrated into the FAQ summarization model and prototype system.

In order to improve the readability of FAQ summary presentation, this study intends to enhance the summary presentation module which currently arranges the extracted sentences in the posting order of articles and then their corresponding sentence as shown in Figure 1. Extracted sentences are in the first column, while the second column lists the article and the sentence numbers with the keywords. Notice that the sentences are grouped by article and are separated by blank spaces. Ideally speaking, the summary should have a fixed structure like an ordinary article in addition to the hyperlinks of keywords and their source articles as seen in commercial software [2]. As shown in Figure 2, the summary would be more readable with a conventional article structure, which will in turn efficiently save the reader's time in understanding the summary and linking it to the original article if needed. Although the article will not be like a human written article, in particular, the sentences look more like phrases which according to Ueda et al. [11] were effective, Figure 2 did reduce the redundant sentences by clustering sentences and mimicking a human writing as much as possible. As compared to Figure 1, Figure 2 does have a potential to achieve what this study is aiming for.

主題討論區FAQ摘要	
討論主題	為何要正規化，其目的、重要性為何？
關鍵字	正規化、目的、重要性
摘要句子內容 (多文件摘要)	
forum topic	
keywords	
「起」關鍵字	
既稱為正規化 (Normalization)	(正規化, 稱為)
正規化的定義	(定義, 正規化)
正規化是必經過程	(經過, 正規化)
正規化的重要性	(重要性, 正規化)
正規化的目的	(目的, 正規化)
因為沒有使用正規化	(正規化, 使用)
透過正規化	(正規化, 透過)
不要因為過度正規化	(正規化)
重要性	(重要性)
目的	(目的)
body and sub-theme	
「承轉」關鍵字	
概念 1	
正規化原因與目的	(目的, 原因, 正規化, 何在)
正規化的目的何在	(目的)
keywords	
概念 2	
而正規化的目的就是要去避免這些問題	(問題, 這些, 避免, 目的, 正規化)
概念 3	
重要性資料正規化可以看成關聯網要的分析的過程	(過程, 分析, 網要, 關聯網要, 看成, 正規化, 資料, 重要性)
概念 4	
資料透過正規化後	(正規化, 透過, 資料)

Figure 2. Summary ordered by intended article structure

To address the problem described above, this research aims to enhance Tao et al.'s [9] FAQ presentation model together with the domain terminology extraction module addressed in the work of Tao et al. [10]. With the domain terminology extraction module, those keywords not covered by the CED will be retained. Most importantly, the final FAQ summary generation will be structured more toward human-writing logic. Ideally, the informative level and readability of the FAQ summary will be significantly improved, which can be judged by subjective and objective indices in a comparative experiment. Due to limitation in page length, the background information is not included, and the suggested FAQ presentation module is instead presented next.

FAQ PRESENTATION MODULE

Basic Principle

News and academic papers have specific structures. For instance, the typical academic paper has an introduction, literature review, research methods, data analysis, and conclusions [5]. When generating the summary, the heading can be used to differentiate the segment to which a sentence belongs, which makes the summary presentation apt with the format of the original article's structure. However, forum articles may not have a formal structure due to their chatting nature. Therefore, this study attempts to introduce a formal article structure for the FAQ summary to match.

Unlike the linear development of English writing, traditional Chinese writing presents a spiral pattern [4]. In parallel to the typical five-paragraph organization in English articles, the most common Chinese article organization is a four-part pattern of introduction-body-related or contrasting sub theme-conclusion, where contrasting is often treated as irrelevant to the subject [7]. Simply put, the introduction (I) breaks the topic into first sentences to attract the readers' interest, the body (B) supports and elaborates the theme, the related or contrasting sub theme (S) makes a logical turn using positive or negative examples, and conclusion (C) closes the article. Accordingly, this study intends to match this traditional IBSC structure to represent the

summary using the extracted sentences. As shown in Figure 3, the left-hand side is the desirable article structure which links those article sentences on its right.

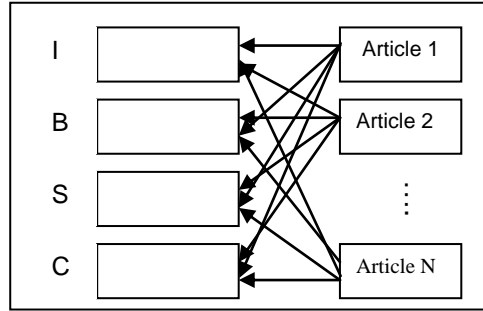


Figure 3. Summary structure with matched sentences

Observations of Forum Articles

To realize the article sentences matching the proposed Chinese four-part article structure, this study randomly selected three forums from representative forums on the Internet. These are the Blue ship, Java technology, and Smart creative teacher forums. Two situations were found from observing the articles on these three forums. First, majority of the articles have three categories of content, including topical content, extended topics from the topical content, and new topical content with no correlations to the topical content. We only considered the first two categories since the extended topical contents provide the depth of the summary related to the topical content. Second, 100 articles from three forums were selected for matching the pattern of IBSC. Not every forum article has the complete structure as we expected based on statistics: 12.66% have IBSC, 30.66% have IBS, 51% have BS, and 5.66% have C.

Summary Structure

Based on the observed types of article structure from representative forums, we know that over 50% have the B&S-part, which are difficult to distinguish and simultaneously coexist at the ratio of 94.34%. Meanwhile, as compared to the technical article structure [5] of previous knowledge, content, method and results, I-part is equivalent to content, B&S-part is equivalent to method, and C-part is equivalent to the results. Forum articles have a unique characteristic of divergence in which the replied articles of a topic may extend to new topic that usually occur while explaining the main subject. This happens in the B&S-part, and all the sentences in a new topic express the same concept [1]. Therefore, we need to group sentences into homogeneous topics for obtaining the B&S-part in the summary. Meanwhile, the sentences within each group need to be ranked based on their similarities to their common topic. In relation to this, Song et al. [8] indicated two summary generation approaches using either a set of main sentences from each cluster or using a representative cluster. In this study, the summary generation for fitting to I, BS, and C parts is taking one sentence from each concept first, and then repeating the process until the limit of compression rate restriction is reached. The C-part can select one sentence from the B&S-part as its conclusion. These C-part sentences are ordered according to their pre-determined topics. As for the I-part, it should contain sentences that are closely related to the topical sentences in the posting articles for matching the idea of breaking the article.

Process Design

To formalize the process of matching the sentences to the proposed summary structure, this study intuitively adopts sentence similarity as the basis to describe the representation algorithm as see. According to Zhang et al. [12], issues such as anti-redundancy and cohesion and coherence become critical in MDS, and current MDS systems often apply a two-phase process, namely, topic identification and summary generation. Therefore, we need to first cluster the sentences with similar concepts for the purpose of sentence grouping. At the same time, there is a need to eliminate sentences with similar semantics in the same group:

Calculate similarity between sentences

If the similarity value is larger than α_1 , merge two sentences into the same group. The calculation is as expressed in Formula 1, where the α_1 value ranges between 0 and 1 and is given by its users.

$$\text{Sim}(S_i, S_j) = \frac{S_i \cap S_j}{S_i \cup S_j} \quad i \neq j \quad (1)$$

In Formula 1, S_i is the i th sentence among all sentences, and $i=1, \dots, N$; S_j is similar to S_i ; N is the number of total sentences; $S_i \cap S_j$ contains the number of the same keywords in S_i and S_j ; $S_i \cup S_j$ contains the number of all keywords within S_i and S_j .

Delete similar sentences within the same group

When the similarity value between two sentences is larger than α_2 , delete the shorter sentence. α_2 is a value between 0 and 1 and is given by its users.

After clustering the selected sentences into groups, induced rules are applied to each group for classifying the sentences into the

three-part pattern as follows:

I-part

Since it is highly correlated to the topic article, the rule considers the similarity between the sentences and the topic sentence of the posting forum article. The similarity formula is as shown in Formula 2. When a sentence has a similarity value greater than β , classify it into the I-part set and order them by similarity value. The β value is given by the users and is a real number between 0 and 1.

$$\text{Sim}(\text{Topic}, C_{kj}) = \frac{\text{Topic} \cap C_{kj}}{\text{Topic} \cup C_{kj}} \quad (2)$$

In Formula 2, C_{kj} is the j th sentence in the k th topical forum article, k and j are positive integers; $\text{Topic} \cap C_{kj}$ is the number of identical keywords within the *Topical sentence* and C_{kj} ; $\text{Topic} \cup C_{kj}$ is the number of all keywords within the *Topical sentence* and C_{kj} .

B&S-part

Calculate the similarity level between all sentences in one group and the sentences in the posting article. When the similarity level between a group and the topical article is larger than γ , retain this group. Only the retained groups are used for composing the B&S-part. Again, these groups appear in the summary according to their similarity levels from high to low [1], and so are the sentences within each group. The group similarity is calculated according to Formula 3.

$$\sum_{i=1}^n \sum_{j=1}^m \text{Sim}(O_i, C_{kj}) = \sum_{i=1}^n \sum_{j=1}^m \frac{O_i \cap C_{kj}}{O_i \cup C_{kj}} \quad (3)$$

In Formula 3, O_i is the i th sentence in the original posting article; C_{kj} is the j th sentence in group k ; $O_i \cap C_{kj}$ is the number of identical keywords in O_i , and C_{kj} ; $O_i \cup C_{kj}$ is the number of keywords in O_i and C_{kj} .

The similarity within a group is calculated as seen in Formula 4.

$$\frac{\sum_{b=1}^r \text{Sim}(C_{ka}, C_{kb})}{r}, a \neq b \quad (4)$$

In Formula 4, k is the number of topics; r is the number of sentences in group k ; C_{ka} is the a th sentence in the k th group; C_{kb} is the b th sentence in the k th group; $\text{Sim}(C_{ka}, C_{kb})$ is the similarity value of sentences C_{ka} and C_{kb} .

C-part

One sentence from each group is selected as a conclusion drawn from the B&S-part in the C-part. That is, the keywords must coexist within both the B&S-part and the C-part. Therefore, the sentence with the highest similarity value is usually selected to be included in the conclusion. However, to avoid high similarity due to a long sentence with more keywords than shorter ones, the reciprocal of the sentence length is also used together with the similarity value to select the sentence to be included in the C-part, as seen in Formula 5. As in the previous one, the order is the same as that which appeared in the B&S-part for consistency.

$$\text{Score}(C_{ka}) = W_1 \times \frac{1}{\text{Length}(C_{ka})} + W_2 \times \frac{\sum_{b=1}^r \text{Sim}(C_{ka}, C_{kb})}{r} \quad (5)$$

In Formula 5, $a \neq b$; W_1 is the weight of the sentence length; W_2 is the weight of the similarity value; $\text{Score}(C_{ka})$ is between 0 and 1 when $W_1=0.5$ and $W_2=0.5$; $\text{Length}(C_{ka})$ is the length of C_{ka} ; K is the number of topics; r is the number of total sentences in group k ; C_{ka} is the a th sentence in group k ; C_{kb} is the b th sentence in group k ; $\text{Sim}(C_{ka}, C_{kb})$ is the similarity between sentences C_{ka} and C_{kb} .

RESEARCH DESIGN

Sample Source

The forum articles were collected from the Internet forums related to “Database.” Among the 390 articles, 300 articles were randomly taken as the training corpus base, and 90 articles were used in the experiment. The 90 articles were divided into three topics, each with 30 articles. The topics are “The purpose and importance of normalization”, “How to map the ER model to the relational model”, and “How to build the ER model and its requirements.”

Summary Presentation

From an application-oriented perspective, the goal is to evaluate and compare the summary presentations between the raw method used in the work of Tao et al. [9] and the proposed method of the four-part pattern approach without comparing them with different advanced summarization presentation methods. This study also applies the domain-terminology module that

includes two approaches based on statistics and genetic algorithm (GA), respectively [10]. To avoid mistakes and increase effectiveness, a pretest was conducted before the formal experiment. The pretest was conducted by three student evaluators. They manually clustered the sentences of 10 articles from the forum topic of “How to update databases in two IP addresses.” There were two sets of hurdle values for deleting sentences. The first parameter value was 1 that indicated only deleting a similarity level of 1, while the second value was 0.65 for deleting sentences with a similarity level greater than 0.65. However, only when at least two out of the three evaluators agreed on the assigned similarity level were the sentences deleted. Three evaluators unanimously agreed that the summary was better when the 0.65 value was used as the hurdle value for deleting sentences.

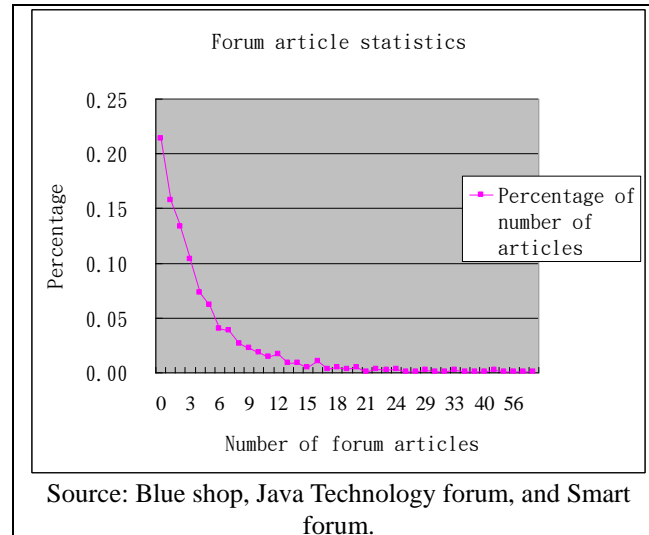


Figure 4. Percentages of number of forum articles

The experiment was divided into three groups, one control group and two experimental groups. The control group (CG) adopted Tao et al.’s [9] FAQ automatic summary system, while the experimental group 1 (EG1) adopted the proposed presentation method with the GA domain-terminology module. The experimental group 2 (EG2) was the same but with the statistical-base domain-terminology module.

Topical articles were randomly selected from the representative forums which again are Blue shop, Java Technology forum, and Smart forum. Based on the articles available on the Internet, some statistics were calculated as shown in Figure 4. Generally speaking, the number of articles can be roughly grouped into low (10), medium (20), and high (over 30). These three levels of article number were adopted in this experiment. Based on three forum topics and three groups of article size, 27 student evaluators were selected from the Graduate Institute in Information Management at a National University of Science and Technology in Taiwan. Like in the pretest, three evaluators worked as a group for one combination of the topic and article size (3 x 3 = 9 groups). The process is briefly described as follows: every evaluator reads through the assigned articles printed on paper, and selects sentences highly related to the topical article. After this paper-based selection step, each evaluator reads the summary automatically generated by the prototype system developed by this study. The evaluators were requested to evaluate the summary, the logical connection of the sentences, length of the summary, its effectiveness, and summary structure. The data collected from these 27 evaluators were used in the analyses.

RESULTS

The objective experimental analysis was conducted based on the criteria of compression rate, recall rate, and precision rate, while the subjective experimental analysis was based on user acceptance.

Objective Experimental Results

This part of the experiment was to compare the recall rate (RR) and precision rate (PR) under different compression rates for three different article sizes. The results are summarized in Table 1. A few observations are discussed as follows:

From the compression-rate perspective, the recall and precision rates of the two experimental groups are very close under the compression rates of 10%, 20%, and 30%. As can be seen in Table 1(a), experiment groups have stable precision rates at different compression rates. For example, EG2 has a precision rate of 32% and a recall rate of 40% under the 10% compression rate, 34% and 40% under the 20% compression rate, and 34% and 30% under the 30% compression rate. On the other hand, the recall and precision rates of the control group vary at different compression rates. As the compression rate increases, the recall rate also increases, and vice versa. On the second topical forum of the control group, the precision and recall rates are 10% and 12% at 10% compression. Similarly, they are 6% and 16% under the 20% compression, and 6% and 24% under the 30% compression. Therefore, these experimental groups demonstrated stable recall and precision rates despite the compression rate. Generally speaking, the recall and precision rates of the two experimental groups were higher than those of the control groups. Based on the data, EG 2 performed the best, then EG1 the second, and CG the last. One exception did occur at topical forum 3. An in-depth

examination of the data revealed that the forum articles presented sentences in either paragraphs or listings. While the first two topics have sentences in paragraph format, the third topic also has sentences in listing format.

Table 1. Comparison of the objective performance

(a) Low volume								(b) Medium volume								(c) High volume							
Topic	Comp	CG		EG1		EG2		Topic	Com.	CR		EG1		EG2		Topic	Com.	CR		EG1		EG2	
		RR	PR	RR	PR	RR	PR			RR	PR	RR	PR	RR	PR			RR	PR	RR	PR	RR	PR
1	10	23	16	23	30	23	36	1	10	11	10	16	21	16	21	1	10	10	9	31	38	31	30
	20	41	14	23	30	23	30		20	16	7	16	21	16	20		20	18	8	31	38	31	30
	30	52	12	23	30	23	30		30	24	8	16	21	16	20		30	22	6	31	38	31	30
2	10	12	10	40	34	40	32	2	10	5	8	13	22	13	22	2	10	16	12	24	19	22	22
	20	16	6	61	27	44	34		20	8	6	23	22	14	23		20	24	9	29	15	29	26
	30	24	6	61	28	44	34		30	14	7	25	24	16	24		30	32	7	29	15	29	25
3	10	27	16	13	36	17	41	3	10	18	16	8	33	8	33	3	10	24	12	17	29	17	27
	20	31	6	13	36	17	41		20	26	15	8	33	8	33		20	29	7	17	28	17	27
	30	34	6	13	36	17	41		30	33	11	8	33	8	33		30	33	5	17	28	17	27

Note: 1. CG=Control Group, EG1=Experiment Group 1, EG2=Experiment Group 2
2. Unit is % for CR, CG, EG1, and EG2.

The evaluators selected partial sentences from the paragraphs, and the whole sentences in the listing. Therefore, the number of sentences in the first two topics had a larger number of sentences. These large numbers also served as the numerator of the recall rate, which therefore made the recall rate a smaller number in the first two topics. Different types of forum articles should have different summary structures. This study suggests that when the articles are in normal paragraph format, the proposed four-part article structure should be used. The listing format of articles may be better with the original summary structure used in the work of Tao et al. [9], which is by the article order and then the sentence order of the original postings.

The recall and precision rates do not change according to the number of articles and sentences. In our data, the numbers of sentences ranged from 15-18 to more than 25 per article in different forum topics. Also, Table 1 shows that the recall and precision rates did not change with the changes in the numbers of articles in either the control group or the experimental groups. Take the control group for example; the recall rates have the highest values at the high volume of articles, while they were the lowest at the medium volume of articles. The precision rate was worst when at the medium volume of all compression rates, while it was best under a high volume at 10% compression, and under a low volume at 20% and 30% compression cases. Therefore, the performance of recall and precision rates will not be affected with the number of articles and sentences. The recall and precision rates are between 10% and 30% for all the cases in this experiment, which are relatively lower compared with the 40% to 50% in the literature. The main reason for such below average performance could be due to the article structure of the Internet forum which is simply not structural and consistent as the Internet news that has a formal recognizable structure shared by writers and readers.

Subjective Experimental Results

The criteria for subjective evaluation were based on the users' acceptance on the summary result as well as the interface support. Therefore, the analyses are divided into the summary results and system interface.

Summary results

The evaluation was done based on users' opinion on the indication, readability, appropriate number of sentences, and structure of the summary. The results are shown on Table 2 for the comparison between the control group and two experimental groups by the number of articles and compression rate. The general observation from the data is that the performance of the experimental groups was better than that of the control group in all the criteria. Among the two experimental groups, EG2 performed better than EG1. For example, under the 30% compression rate, the average scores of indication, readability, appropriate number of sentences, and structure of the summary are 3.11, 2.33, 2.11, and 2.88 for CG, while they are 3.55, 3.33, 4, and 4 for EG1, and 4.22, 3.88, 4.33, and 4.33 for EG2. Under the 10% compression, the averages are again 2.78, 2, 2.11, and 2.78 for CG, while the averages are 3.11, 3.22, 3.11, and 3.11 for the EG1, and 3.11, 3.44, 3.66, and 3.66 for EG2. These two cases confirmed what is concluded earlier that EG2 is better than EG1, which in turn is better than CG. In other words, adding the domain-terminology module and enhanced summary presentation module did prove to be positively perceived by the users. This means that the method of the four-part structure proposed by this study was more acceptable than presenting the summary in the original order of the articles and sentences as in the work of Tao et al. [9]. Furthermore, this experiment also confirmed that statistical-based domain terminology extraction performed better than GA-based terminology extraction in terms of users' perception.

Table 2. Comparison of system interface

(a) Low volume					(b) Medium volume					(c) High volume				
Item	CR	CG	EG1	EG2	Item	CR	CG	EG1	EG2	Item	CR	CG	EG1	EG2
Indication	10	2.78	3.11	3.11	Indication	10	2.77	3.00	3.00	Indication	10	2.00	3.11	3.33
	20	3.10	3.44	3.44		20	2.77	3.33	3.33		20	2.66	3.33	3.55
	30	3.11	3.55	3.22		30	2.55	3.44	3.11		30	3.00	3.55	3.77
Readability	10	2.10	3.22	3.44	Readability	10	2.66	3.00	2.88	Readability	10	2.22	2.88	3.00
	20	1.40	3.66	3.44		20	2.88	3.55	3.44		20	2.88	3.55	3.55
	30	2.33	3.33	3.88		30	3.11	3.33	3.55		30	3.00	3.55	3.55
Appropriate number of articles	10	2.11	3.11	3.66	Appropriate number of articles	10	2.44	3.11	3.22	Appropriate number of articles	10	2.22	3.00	3.11
	20	1.60	3.6	3.88		20	2.88	3.22	3.22		20	2.33	3.11	3.11
	30	2.11	4.00	4.33		30	3.11	3.44	3.33		30	2.77	3.44	3.33
Structure	10	2.78	3.11	3.66	Structure	10	3.00	2.66	3.00	Structure	10	2.22	2.88	3.11
	20	2.40	3.66	3.8		20	3.00	3.22	3.55		20	2.33	3.00	3.11
	30	2.88	4.00	4.33		30	3.00	3.33	3.22		30	2.55	3.22	3.44

Note: 1. CG=Control Group, EG1=Experiment Group 1, EG2=Experiment Group 2
2. Unit is % for CR, CG, CG1, and EG2.

System interface evaluation

This part mainly assessed whether the users perceived better while browsing the FAQ summary via the interface functions such as keywords and hyperlinks to original articles, or whether the users understood the FAQ better as shown in Table 3. The detailed results of the system interface are summarized in Table 4. Within CG in (a), 59.3% of them agree and extremely agree on the “Hyperlink” function, 44.5% of them agree and extremely agree on the “Keyword indication”, 55.6% of them agree and extremely agree on “Keywords highlighted in summary”, and 55.5% of them show medium agreement on “Summary format.” Similarly, within EG1 in (b), 63% of them agree and extremely agree on the “Hyperlink” function, 70.4% of them agree and extremely agree on the “Keyword indication”, 59.3% of them agree and extremely agree on “Keywords highlighted in summary”, and 63% of them agree and extremely agree on “Summary format.” Finally, within CG2 in (c), 70.4% of them agree and extremely agree on the “Hyperlink” function, 77.8% of them agree and extremely agree on the “Keyword indication”, 63% of them agree and extremely agree on “Keywords highlighted in summary”, and 63% of them agree and extremely agree on “Summary format.” Overall, the two experimental groups had higher satisfaction levels than the control group on the four criteria. Again, EG2 had higher levels than EG1.

Table 3. System interface

Item	Question	CG	EG1	EG2
Usability of the Interface	Hyperlinks can effectively assist a user understand the content of the forum topic	3.44*	3.74	3.74
	Keyword indication can effectively assist a user understand the content of the forum topic	3.22	3.7	3.74
	Keywords in the summary can effectively assist a user understand the content of the forum topic	3.44	3.63	3.59
	The format of summary can clearly present the content of the summary	3.33	3.67	3.67

Note: 1. CG=Control Group, EG1=Experiment Group 1, EG2=Experiment Group 2
*. Average scores based on 5-point Likert-scale ranging from 0(very disagree) to 5 (Very agree)

In addition to the above analyses, we compared the experimental groups to the control group using the articles in a forum topic of “The purpose and importance of normalization.” Under 20% compression, the differences are summarized as follows. The number of sentences in EG2 is 39, much less than the 112 in CG. This naturally lowered the readers’ load. The structure of the summary under control group 2 was based on the I-B&S-C order as shown in Figures 3, in which the right-hand side labels these four parts in order for readers to clearly recognize these different concepts in a normal Chinese composition. The summary structure of CG, as seen in Figure 1, was based on the posting order of the articles and their sentences. The user benefit can be illustrated using a simple example: the control group had to locate the sentences with the same concept in different parts of the

summary, such as in Figure 1, that “減少儲存空間 (reducing the storage space)” was separated in the 22nd sentences in the 3rd article and the 9th sentence in the 14th article. On its counterpart in experimental group 2, as seen in Figure 2, “減少儲存空間 (reducing the storage)” was grouped together for users’ convenience. Based on the above comparison, it is clear that the concept-based four-part structure did present a concise and clear summary structure for its readers.

Table 4. Comparison of subjective performance

	(a) CG				(b) EG1				(c) EG 2			
Acceptance	Hyperlink	Keyword indication	Summary keywords	Summary format	Hyperlink	Keyword indication	Summary keywords	Summary format	Hyperlink	Keyword indication	Summary keywords	Summary format
Extremely disagree and disagree	14.8	25.9	11.1	7.4	3.7	0	3.7	0	0	3.7	3.7	0
Average	25.9	29.6	33.3	55.5	33.3	29.6	37.0	37.0	29.6	18.5	33.3	37.0
Extremely agree and agree	59.3	44.5	55.6	37.1	63.0	70.4	59.3	63.0	70.4	77.8	63.0	63.0
Total	100	100	100	100	100	100	100	100	100	100	100	100

Note: 1. CG=Control Group, EG1=Experiment Group 1, EG2=Experiment Group 2
2. Unit is % for CR, CG, CG1, and EG2.

CONCLUSIONS

This study proposed using the Chinese composition structure for improving the summary presentation of an automatic Internet forum FAQ system by Tao et al. [9]. Based on the objective evaluations, the comparative experiment results demonstrated that the recall and precision rates increased under different article volumes and compression rates. Moreover, they become stabilized without being affected by the compression rate due to redundant reduction via the clustering of similar concepts of sentences. Also, the statistical-based method for extracting domain terminology seemed to work better than the GA-based method with our sample data. On the counterpart of the subjective evaluation, the users perceived better on the criteria of indication, readability, appropriate number of sentences, and structure in the FAQ system. Based on our experiment, this study suggests the use of the proposed summary presentation module for the topic forum with paragraph-based articles in order to reduce the repeats of the summary content, and the original summary presentation module [9] for listing the style of articles in order to maintain the summary integrity at the cost of higher repeats. To further improve the usability of the FAQ summary presentation, one immediate future work should delve on the use of a complete sentence from the original article instead of partial sentences for better readability. Another one would be to apply this Chinese composition structure to non-Chinese text summarization cases for possible generalized applications. Also, despite the necessary analysis being too immature or computationally intensive [3], other techniques and methods in extraction or abstraction approaches [1] can be used to further enhance the coherence of the multiple-article FAQ summary.

REFERENCES

- [1] Barzilay, R., Elhadad, N. and McKeown, K. (2002) “Inferring strategies for sentence ordering in multidocument news summarization”, *Journal of Artificial Intelligence Research*, Vol. 17, pp. 35-55.
- [2] Hahn, U & Mani, I. (2000) “The challenges of automatic summarization”, *IEEE Computer*, November, pp. 29-36.
- [3] Halteren, H. V. (2003) “New feature sets for summarization by sentence extraction”, *IEEE Intelligent Systems*, Vol. 18, No. 4, pp. 34-42.
- [4] Kaplan, R. B. (1966) “Cultural thought patterns in inter-cultural education”, *Language Learning*, Vol. 16, pp.1-20.
- [5] Lehman, A. (1999) “Text structuration leading to an automatic summary system: RAFI”, *Information Processing & Management*, Vol. 35, No. 2, pp.181-191.
- [6] Mani, I. and Maybury, M.T. (1999) *Advances in Automated Text Summarization*, Cambridge, MA: The MIT Press.
- [7] Pern, “T.-L. A Web classroom assistant NICENET for assisting English writing composition”, last accessed on March 24, http://www.cavesbooks.com.tw/e_magazine/e_magazine_article.aspx?sn=180&language1=0.
- [8] Song, S.-K., Jang, D.-H. and Myaeng, S.-H.(2005) “Text summarization based on sentence clustering with rhetorical structure information”, *International Journal of Computer Processing of Oriental Languages*, Vol. 18, No. 2, pp. 153-170.
- [9] Tao, Y.-H., Huang, C.-C. and Yang, C.-C. (2006) “Designing an automatic FAQ abstraction for internet forum”, *Journal of Information Management (in Chinese)*, Vol. 13, No. 2, pp. 89-112.
- [10] Tao, Y.-H., Lin, C.-L. and Liu, S.-C.(2005) “Comparing genetic algorithm and statistical segmentation in terminology extraction for topical forum, 10th Conference on Artificial Intelligence and Applications, Kaohsiung, Taiwan, December 2-3.
- [11] Ueda, Y., Oka, M., Koyama, T. and Miyauchi, T. (2000) ”Toward the ‘At-a-glance’ summary: Phrase-representation summarization method”, *In Proceedings of COLING*, Saarbrücken, Germany, pp. 878-884.

- [12] Zhang, Y., Zincir-Heywood, N. and Milios, E. (2004) “World wide web site summarization”, *Web Intelligence and Agent Systems: An international journal*, Vol. 2, pp.39–53.