Conceptual Retrieval of Chinese Frequently Asked Healthcare Questions

Rey-Long Liu*, Shu-Ling Lin**

ARTICLE INFO

ABSTRACT

Given a query (a health question), retrieval of relevant frequently asked questions (FAQs) is essential as the FAQs provide both reliable and readable information to healthcare consumers. The retrieval requires the estimation of the semantic similarity between the query and each FAQ. The similarity estimation is challenging as semantic structures of Chinese healthcare FAQs are quite different from those of the FAQs in other domains. In this paper, we propose a conceptual model for Chinese healthcare FAQs, and based on the conceptual model, present a technique ECA that estimates conceptual similarities between FAQs. Empirical evaluation shows that ECA can help various kinds of retrievers to rank relevant FAQs significantly higher. We also make ECA online to provide services for FAQ retrievers.

1. Introduction

When compared with general information, healthcare information needs to be more *reliable* and *readable* as the information is actually used for disease management and health promotion. Frequently asked questions (FAQs) in healthcare provide such information as they are often written and compiled by healthcare professionals in response to specific questions of healthcare consumers. The healthcare consumers thus tend to read those web pages that have healthcare FAQs (Sillence et al., 2004), and many healthcare information providers have collected and maintained a large number of healthcare FAQs for healthcare consumers.

Therefore, given a healthcare question as a query, the retrieval of relevant FAQs is essential for the utility of the reliable and readable healthcare information in the FAQs. Given a database of FAQs and a query, a FAQ retriever ranks the FAQs based on the relevancy of the FAQs to the query. The FAQ retrieval task is challenging as both the query and the FAQs are often quite short, making it difficult to collect helpful evidences to identify relevant FAQs.

 * Professor, Department of Medical Informatics, Tzu Chi University Hualien, Taiwan, R.O.C. (rlliutcu@mail.tcu.edu.tw)
** Department of Medical Informatics, Tzu Chi University Hualien, Taiwan, R.O.C. (splend126@gmail.com)

1.1. Problem Definition

In this paper, we analyze the conceptual structure of Chinese healthcare FAQs, and present a conceptual scoring technique to enhance retrievers of Chinese healthcare FAQs. More specifically, we extend a preliminary work (Liu & Lin, 2012) and develop an automatic concept recognition and scoring technique to estimate the conceptual similarities between healthcare queries and FAQs. Three types of essential concepts in Chinese healthcare FAQs are identified: *event*, *condition*, and *aspect*, as a Chinese healthcare FAQ often cares about some aspects (e.g., cause) of some events (e.g., cardiovascular disease) under some condition (e.g., patients of the periodontal disease). For example, a Chinese healthcare FAQ "兒童常吃山藥會不會引發性早熟?" (For children, will frequently eating yams cause precocious puberty?) has two event concepts "山藥" (yams) and "性早熟" (precocious puberty); a condition concept "兒童" (children); and an aspect concept "引發" (cause). Obviously, to identify relevant FAQs for a query, the FAQ retriever should consider the similarities on the three types of essential concepts.

Therefore, we develop a conceptual scoring technique ECA (Event, Condition, and Aspect) to automatically estimate the conceptual similarities between healthcare queries and FAQs. The similarity scores can be integrated with the scores provided by other FAQ retrievers so that relevant FAQs with respect to Chinese healthcare queries can be ranked significantly higher, facilitating the sharing of reliable and readable healthcare information.

1.2. Contribution and Organization of the Paper

Main contributions of ECA include (1) practically, retrieval of healthcare FAQs is a key to share reliable and readable healthcare information; (2) technically, previous studies have developed many FAQ retrievers but none of them have considered the three types of essential concepts in Chinese healthcare FAQs, and hence the collaboration between ECA and the previous retrievers can further enhance the retrievers by the conceptual similarity evidences provided by ECA. We also make ECA online (http://203.64.84.94:126/) to provide the similarity estimation service for various kinds of FAQ retrievers, facilitating the utility of reliable and readable healthcare information.

In the next section, we discuss related work and accordingly identify the technical contributions of ECA. In section 3, we present a conceptual model for Chinese healthcare FAQs. Based on the model, in Section 4 we present how ECA recognizes the essential concepts in healthcare FAQs and estimates the conceptual similarity between a query and each FAQ. An empirical evaluation on thousands of Chinese healthcare FAQs is reported in Section 5. We implement and test several FAQ retrieval techniques as well as their integrations using SVM (Support Vector Machine). The results show that, ECA can be integrated with each of them to produce better performance in ranking relevant healthcare FAQs. Performance of ECA is also robust under different settings of the knowledge (terms and patterns) for concept recognition.

2. related work

A FAQ consists of a question part and an answer part. Several previous techniques considered both parts for FAQ retrieval (e.g., Wu, Yeh, & Chen, 2005; Wu, Yeh, & Lai, 2006; Xue, Jeon, & Croft, 2008), and in some cases other information concerning the FAQ was employed as well (e.g., the content of the web page containing the FAQ, Jijkoun & de Rijke, 2005). In this paper, ECA focuses on the question part, which is the most important part in FAQ retrieval (Jeon, Croft, & Lee, 2005). It provides the conceptual similarity scores between the question parts of queries and FAQs. ECA can thus be integrated with the previous techniques by integrating the conceptual similarity produced by ECA with the similarity estimated based on other parts of FAQs (e.g., the answer parts of the FAQs and the webpage containing the FAQs).

Given q as a query (question) and f as the question part of a FAQ, previous FAQ retrievers estimated the similarity between q and f by two typical methodologies: *term matching* and *semantic analysis*. Term matching methods often considered four types of information for the similarity estimation: (1) overlap of the words in q and f (i.e., q and f may be more similar to each other if many words co-occur in them, Bernhard and Gurevych, 2008), (2) cosine similarity based on the vectors of q and f (Bernhard & Gurevych, 2008; Burke et al., 1997; Wu et al., 2005), (3) relatedness of the words in q and f (e.g., measured by the distance of the words on an ontology, Burke et al., 1997; Wu et al., 2005; Wu et al., 2006, or the handling of the spelling errors of words in q, Bernhard & Gurevych, 2008; Contractor et al., 2010; Kothari et al., 2009), and (4) mapping or translation between the words in q and f (to tackle the problem of word mismatch between q and f, Jeon et al., 2005; Lee et al., 2008b; Riezler et al., 2007; Xue et al., 2008). However, none of the term matching methods considered essential concepts (i.e., *event*, *condition*, and *aspect*) in Chinese healthcare FAQs. ECA considers the essential concepts, which actually indicate a main part of semantics of the healthcare FAQs.

Semantic analysis was thus noted as a methodology for FAQ retrieval as well. However, previous semantic analysis methods were not aimed at the retrieval of Chinese healthcare FAQs. They often aimed at typical kinds of semantic information, including (1) types of questions (e.g., typical types include "what," "how," and "where," Wu et al., 2005), (2) semantic categories of questions (Mishra, Mishra, & Sharma, 2013; Pan et al., 2008), and (3) syntactic or semantic structures of questions (by deeper analysis such as parsing, Casellas et al., 2007; Wang, Ming, & Chua, 2009; Wu et al., 2006; Winiwarter, 2000).

The first kind of semantic information (i.e., the question type of a FAQ) cannot indicate the essential concepts of the FAQ, and more importantly many Chinese healthcare FAQs cannot fall into any specific question types. For example, among thousands of healthcare FAQs in KingNet[1] (a provider of Chinese healthcare information) we find that it is quite difficult to determine the question types of many FAQs, such as "失眠的原因與診斷" (the cause and diagnosis of insomnia). Some FAQs, such as "安樂死" (euthanasia), even consist of a single health topic without any question words.

---

1) KingNet is available at http://www.kingnet.com.tw.

On the other hand, the second and the third kinds of semantic information (semantic categories and syntactic and semantic structures of FAQs) are quite difficult to recognize from Chinese healthcare FAQs. To recognize the two kinds of semantic information, the previous studies often employed parsing. However, a good parser for Chinese healthcare queries is often unavailable since (1) parsing Chinese questions is still a challenging task (Lee et al., 2008a), and (2) Chinese healthcare questions are not always well-formed for parsing: they may even consist of a single term (e.g., a disease name or a treatment), multiple sentences or multiple fragments.[2] It is thus difficult to get the semantic structures of healthcare FAQs by parsing. ECA is concerned with the existence of essential concepts in a question rather than the semantic structure of the question, making it able to estimate the conceptual similarities without relying on parsing.

3. A Conceptual model for Chinese Healthcare FAQs

A conceptual model for the retrieval of Chinese healthcare FAQs should be both *expressive* (able to indicate the core intention of a healthcare question) and *realizable* (able to be used for similarity estimation without assuming the well-formedness of the question). To develop such an expressive and realizable conceptual model, we analyze a large number of Chinese healthcare FAQs and find that most of the FAQs have the core intentions about health promotion and disease management. This is because healthcare questions are asked by general healthcare consumers, who mainly care about the way to keep them healthy. This is also a reason why healthcare information providers for the consumers mainly focus on the topics about diseases and health for healthcare consumers. We thus identify three types of essential concepts in Chinese healthcare FAQs: (1) *event*: the target event under discussion, (2) *condition*: the condition of the discussion, and (3) *aspect*: the information aspect of the discussion. The core intention of a Chinese healthcare question is to ask for some aspect of information (e.g., cause) about target events (e.g., cardiovascular disease) under a certain condition (e.g., patients of the periodontal disease).

Table 1. Three types of essential concepts in a Chinese healthcare FAQ

Type	Concept	Definition
Event	E1	The first target event (in the FAQ) under discussion
	E2	The second target event (in the FAQ) under discussion (may be none if the question has only one target event)
Condition	C	The condition or the context (in the FAQ) of the discussion
Aspect	A	A_{cause}: The causal aspect of the discussion (e.g., risk factors and prevention of diseases)
		$A_{process}$: The processing aspect of the discussion (e.g., treatment and management of diseases)
		$A_{diagnosis}$: The diagnosis aspect of the discussion (e.g., symptoms and diagnosis of diseases)

---

2) We randomly selected 38 Chinese healthcare FAQs from KingNet and invoked a parsing system (http://parser.iis.sinica.edu.tw/) to parse them. We found that 34.2% of the FAQs cannot have a single and correct parse tree.

More specifically, Table 1 defines the three types of essential concepts. As healthcare questions are often quite short and specific to one to two target events, we consider at most two events *E1* and *E2* in the conceptual model. All concepts governing the condition of the discussion are the condition concept (*C*). To facilitate the recognition of the aspect concept (*A*), we define three high-level categories of aspects: A_{cause}, $A_{process}$, and $A_{diagnosis}$ that are about the *causal*, *processing*, and *diagnosis* aspects of the health topics respectively.[3] The three aspects were noted as the key concepts to retrieve documents for clinical questions posted by healthcare professionals (Lin and Demner-Fushman, 2006). ECA is the first framework aiming at the conceptual retrieval of FAQs (rather than documents) for healthcare consumers (rather than professionals).

As an example, consider a FAQ from KingNet[4]: "兒童常吃山藥會不會引發性早熟" (For <u>children</u>, will frequently eating <u>yams</u> <u>cause</u> <u>precocious puberty</u>?). There are two *event* concepts: *E1*=山藥 (yams) and *E2* = 性早熟 (precocious puberty), a *condition* concept: *C* = 兒童 (children), and an *aspect* concept: *A* = {A_{cause}}, which is determined based on the term 引發 (cause). It is interesting to note that a healthcare FAQ does not necessarily have all types of concepts, but it should have at least one concept, which is *E1* and in that case the FAQ may be simply a term about a health topic (e.g., the name of a disease) and hence its intention is asking for all kinds of information about the topic (e.g., all information about a disease).

4. ECA: A technique to estimate conceptual similarity between Healthcare FAQs

Based on the conceptual model, ECA is developed to provide the conceptual similarity between each query and each FAQ. As illustrated in Figure 1, ECA has two phases of tasks: (1) the offline phase for recognizing the essential concepts (*event*, *condition*, and *aspect*) in each FAQ, and (2) the online phase for receiving a query *q* and estimating the similarity between *q* and each FAQ. The similarity scores of *q* with respect to the FAQs can be used to improve the ranking performance of various kinds of FAQ retrievers.

### 4.1. Offline Tasks: Recognition of Essential Concepts in FAQs

Among the three types of essential concepts, event concepts are the most *unrestricted*, since they can be whatever health topics. Aspect concepts are the most *restricted*, since they can only be three categories of interest: causal, processing, and diagnosis. Therefore, the three types of essential concepts in a FAQ should be recognized by following the sequence: aspect concepts → condition concepts → event concepts. More specifically, given *f* as the question part of a FAQ, ECA first recognizes the aspect concepts in *f* (A1 in Figure 1), and then recognizes the condition concepts from *f* with the aspect strings removed (A2 in Figure 1). Finally ECA recognizes the event concepts from *f* with both the aspect strings and the condition strings removed (A3 in Figure 1).

---

3) The terms corresponding to each aspect should consist of at least two Chinese characters.
4) Available at http://www.kingnet.com.tw.
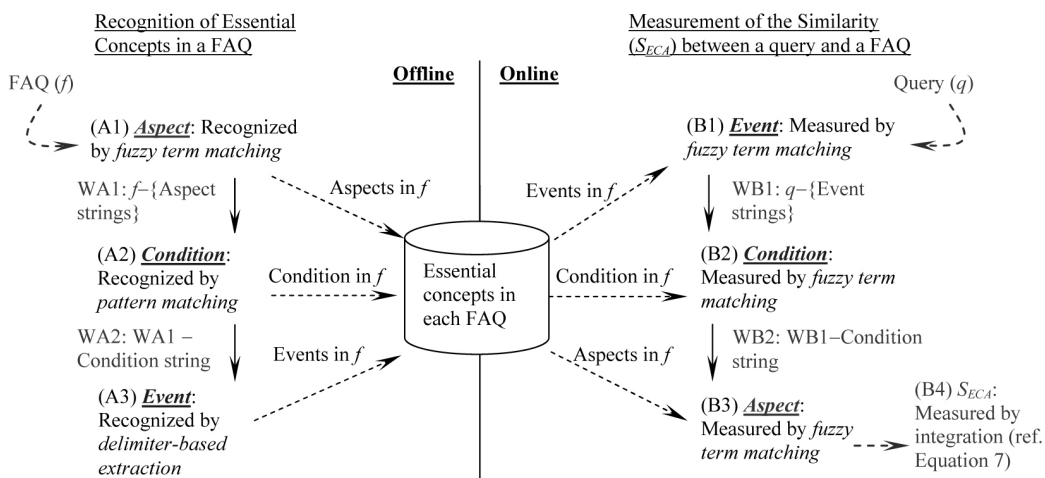
Fig. 1. ECA consists of two phases of tasks: the offline phase and the online phase. The former recognizes the concepts in each FAQ (*f*), and given a query (*q*) the latter measures the similarity between *q* and each FAQ.

4.1.1. Recognition of the aspect concepts in a FAQ (A1 in Figure 1)

As information aspects are actually categories of interest, an aspect may be indicated by many different terms. We thus analyze the typical terms for A_{cause}, $A_{process}$, and $A_{diagnosis}$, and accordingly construct TA_{cause}, $TA_{process}$, and $TA_{diagnosis}$ as their sets of corresponding terms respectively. The three sets have 74, 124, and 35 terms respectively. Example terms in TA_{cause} include "危險因子" (risk factor) and "引發" (incur). Example terms in $TA_{process}$ include "用藥" (medication) and "抑制" (inhibit). Example terms in $TA_{diagnosis}$ include "症狀" (symptom) and "檢驗" (examination and test). Obviously, the development of the sets of aspect terms is a knowledge engineering task. In empirical evaluation, we will show that the performance of ECA is robust under different settings for the sets of aspect terms (ref. Section 5.4.2).

With the three sets of typical terms for the three aspects, ECA recognizes the aspects of *f* (the question part of a FAQ) by checking whether *f* mentions the terms. More specifically, Equation 1 defines a fuzzy term matching method to estimate the similarity (*StrSim*) between two strings t_1 and t_2, where $idf(w)$ is the inverse document frequency[5]) of the Chinese character *w*. *StrSim* is higher if there are more matched characters with higher *idf* values.

$$StrSim(t_1, t_2) = \frac{\sum_{y \in t_1 \ and \ y \in t_2} idf(y)}{\sum_{z \in t_1 \ or \ z \in t_2} idf(z)} \times \log_2 \left| \{ y \mid y \in t_1 \ and \ y \in t_2 \} \right| \qquad (1)$$

Based on Equation 1, the strengths of classifying a FAQ *f* into A_{cause}, $A_{process}$, and $A_{diagnosis}$ are

---

5) The *idf* value of a Chinese character *w* is calculated by treating each FAQ as a document.

estimated by Equation 2 ~ Equation 4, respectively.

$$ASF(f, A_{cause}) = \max_{t \in TA_{cause}} \max_{c \in f} \{StrSim(t,c)\} \tag{2}$$

$$ASF(f, A_{process}) = \max_{t \in TA_{process}} \max_{c \in f} \{StrSim(t,c)\} \tag{3}$$

$$ASF(f, A_{diagnosis}) = \max_{t \in TA_{diagnosis}} \max_{c \in f} \{StrSim(t,c)\} \tag{4}$$

We employ 0.5 as the threshold for the strength: f is classified into an aspect only if its strength corresponding to the aspect is higher than 0.5. It is interesting to note that, if f is not classified into any aspect (e.g., f may simply consist of a single term about a health topic), a '*don't-care*' is assigned to each aspect for f, indicating that f does not ask for any specific aspect and hence *all* aspects might be related to the intention of f.

4.1.2. Recognition of the condition concepts in a FAQ (A2 in Figure 1)

By analyzing a large number of Chinese healthcare FAQs, we find that the FAQs often employ the description of *time* and *people* to indicate the context of discussion. Therefore, ECA employs pattern matching to extract the string that indicates the condition concepts in f. We define 21 patterns for specific concepts for time and people:

> "春天"(spring) ‖ "冬天"(winter) ‖ "夏天"(summer) ‖
> "秋天"(autumn) ‖ "春季"(spring) ‖ "冬季"(winter) ‖ "夏季"(summer) ‖
> "秋季"(autumn) ‖ "男人"(male) ‖ "女人"(female) ‖ "小孩"(child) ‖
> "老人"(old people) ‖ "兒童"(child) ‖ "嬰兒"(infant) ‖
> "幼兒"(little child) ‖ "孕婦"(pregnant woman) ‖ "小朋友"(child) ‖
> "寶寶"(baby) ‖ "男性"(male) ‖"女性"(female) ‖ "父母"(parent) (5)

The strings (in f) that match the patterns are extracted as the strings that indicate the condition concepts in f. After applying the above patterns to f, the following 7 patterns with wildcards are applied to extract other possible condition concepts, ('*' denotes any string in f):

> "*期"(period) ‖ "*時"(when) ‖ "*後(after) ‖ "*前"(before) ‖
> "*族"(some group) ‖ "*員"(somebody) ‖ "*者"(somebody) (6)

Obviously, the development of the condition patterns is a knowledge engineering task. In empirical evaluation, we will show that the performance of ECA is robust under different settings for the patterns (ref. Section 5.4.2).

4.1.3. Recognition of the event concept in a FAQ (A3 in Figure 1)

After removing from f the strings about the aspects and conditions in f, ECA gets a set of separate strings, which can be treated as "string islands" as they are not consecutive strings. Punctuations

are then removed to further separate the strings, and each of the resulting strings may be a candidate string that indicates an event concept in *f*. To further make the candidate strings more precise, 693 terms (including stop words) that are unlikely to be in event strings are defined[6]. The terms are removed from the strings, and each of the resulting strings is treated as possible strings of event concepts in *f*. In empirical evaluation, we will also show that the performance of ECA is robust under different settings for the terms (ref. Section 5.4.2).

4.2. Online Tasks: Measurement of Similarity between Queries and FAQs

Given *q* as a query and *f* as the question of a FAQ whose essential concepts have been recognized, ECA estimates the conceptual similarity (S_{ECA}) between *q* and *f* by Equation 7, where $S_{E1}(q,f)$, $S_{E2}(q,f)$, $S_C(q,f)$, and $S_A(q,f)$ are the similarity values on E1, E2, C, and A, respectively.

$$S_{ECA}(q,f) = \begin{cases} 0, & \text{if } f \text{ has no } E2, \text{ and } S_{E1} = 0; \\ 0, & \text{if } f \text{ has both } E1 \text{ and } E2, \text{ and } S_{E1} = S_{E2} = 0; \\ \text{average } \{S_x(q,f) \mid x \in \{E1, E2, C, A\}\}, & \text{otherwise} . \end{cases} \quad (7)$$

ECA assigns 0 to the similarity between *q* and *f* if they talk about totally different events (i.e., E1 and E2); otherwise the conceptual similarity is the average of the similarity values on those concepts in *f* (recall that a FAQ does not necessarily have E2, A, and C).

4.2.1. Similarity Measurement for Event (B1 in Figure 1)

Let E_f be the set of strings in *f* that are recognized as the *event* concepts. ECA employs Equation 8 to find the string $e1_{q,f}$ in *q* and $e1_f$ in E_f that have the largest string similarity (using *StrSim* defined in Equation 1), and accordingly sets $S_{E1}(q,f)$ to the similarity between $e1_{q,f}$ and $e1_f$ by Equation 9.

$$<e1_{q,f}, e1_f> = \arg\max_{e_q \in q; e_f \in E_f} \{StrSim(e_q, e_f)\} \quad (8)$$

$$S_{E1}(q,f) = StrSim(e1_{q,f}, e1_f) \quad (9)$$

Moreover if there are multiple strings in *f* recognized as the *event* concept (i.e., E_f - $\{e1_f\}$ is not empty), ECA employs Equation 10 to find the string $e2_{q,f}$ in q-$\{e1_{q,f}\}$ and $e2_f$ in E_f - $\{e1_f\}$ that have the largest string similarity, and accordingly sets $S_{E2}(q,f)$ to the similarity between $e2_{q,f}$ and $e2_f$ by Equation 11.

$$<e2_{q,f}, e2_f> = \arg\max_{e_q \in q - \{e1_{q,f}\}; e_f \in E_f - \{e1_f\}} \{StrSim(e_q, e_f)\} \quad (10)$$

---

6) Example terms that are unlikely to be healthcare events include the terms about quantity (e.g., "一些些" (a little)), negation (e.g., "不可以" (cannot)), possibility (e.g., "可能" (possible)), question words (e.g., "請問" (could I ask)), inquiry (e.g., "煩請" (please), stopwords in Chinese (e.g., "關於" (about)), and other miscellaneous terms not related to healthcare.

$$S_{E2}(q,f) = StrSim(e2_{q,f}, e2_f) \tag{11}$$

4.2.2. Similarity Measurement for Condition (B2 in Figure 1)

Let C_f be the set of strings in f that are recognized as the *condition* concepts. ECA employs Equation 12 to find the string c_{qf} in q - $\{e1_{qf}\}$ - $\{e2_{qf}\}$ and c_f in C_f that have the largest string similarity, and accordingly sets $S_C(q,f)$ to the similarity between $c_{q,f}$ and c_f by Equation 13.

$$< c_{q,f}, c_f > = \underset{c_q \in q - e1_{q,f} - e2_{q,f};\, t_f \in C_f}{\arg\max} \{ StrSim(c_q, t_f) \} \tag{12}$$

$$S_C(q,f) = StrSim(c_{q,f}, c_f) \tag{13}$$

Table 2. Estimation of $S_a(q,f)$[a]: $S_a(q,f)$ is not zero when the query (q) or the FAQ (f) is not classified into any aspect (i.e., its aspect is '*don't care*'), since in this case all the aspects are related to its intention

Whether query (q) is classified into aspect a	Whether FAQ (f) is classified into aspect a	Similarity ($S_a(q,f)$)
O	O	1
O	X	0
O	?	1/2
?	O	1/2
?	X	1/2
?	?	1
X	O	0
X	X	1
X	?	1/2

[a]$a \in \{A_{cause}, A_{process}, A_{diagnosis}\}$ is an aspect; 'O' denotes 'classified into a';
'X' denotes 'not classified into a'; '?' denotes '*don't-care* on aspect a'

4.2.3 Similarity Measurement for Aspect (B3 in Figure 1)

ECA recognizes the aspect concepts in q by the same way employed to recognize the aspect concepts in FAQs (recall Section 4.1.1). The strengths of correlating q to A_{cause}, $A_{process}$, and $A_{diagnosis}$ are estimated by Equation 14 ~ Equation 16, respectively.

$$ASQ(q, A_{cause}) = \max_{t \in TA_{cause}} \max_{c \in q - e1_{q,f} - e2_{q,f} - c_q} \{ StrSim(t,c) \} \tag{14}$$

$$ASQ(q, A_{process}) = \max_{t \in TA_{process}} \max_{c \in q - e1_{q,f} - e2_{q,f} - c_q} \{ StrSim(t,c) \} \tag{15}$$

$$ASQ(q, A_{diagnosis}) = \max_{t \in TA_{diagnosis}} \max_{c \in q - e1_{q,f} - e2_{q,f} - c_q} \{ StrSim(t,c) \} \tag{16}$$

We employ 0.5 as the threshold for the strength: q is classified into an aspect if its strength corresponding to the aspect is higher than 0.5. Note that if q is not classified into any aspect, a '*don't-care*' is assigned to each aspect for q, indicating that *all* aspects may be related to the intention of q. Given that both q and f can have three labels for each aspect a (i.e., 'classified into a', 'not classified into a', and '*don't-care* on a'), Table 2 defines the way to estimate the similarity value on an aspect. Finally $S_A(q,f)$ is estimated by Equation 17.

$$S_A(q,f) = \frac{\sum_{a \in \{A_{cause}, A_{process}, A_{diagnosis}\}} S_a(q,f)}{3} \qquad (17)$$

5. Empirical Evaluation

ECA was empirically evaluated on thousands of Chinese healthcare FAQs and queries. As the conceptual similarity (i.e., S_{ECA}) can indicate a kind of semantic information not considered by previous FAQ retrievers, we aimed at measuring the contribution of ECA to several kinds of the FAQ retrievers in ranking relevant healthcare FAQs.

5.1. Collection of Chinese healthcare FAQs and Queries

The FAQs were from KingNet[7], which is a Chinese healthcare information provider. All FAQs in KingNet were collected, and we thus got 3517 FAQs. As users tend to employ their own queries (questions) to find relevant FAQs from KingNet, we collected test queries from other healthcare information providers (not from KingNet). We totally got 200 test queries from five healthcare information providers[8]: (1) from the first provider[9], 90 test queries were collected by selecting the top-5 most popular FAQs in each category of FAQs; (2) from the second provider[10], 22 test queries were collected by selecting all FAQs of the categories about physical fitness and nutrition; (3) from the third provider[11], 25 test queries were collected by selecting five FAQs from each category; (4) from the fourth provider[12], 60 test queries were collected by selecting five FAQs from each category; and (5) from the fifth provider[13], 3 test queries were collected by selecting top-3 FAQs.

Each of the 200 test queries was manually checked to identify relevant FAQs from the 3517 FAQs. For each pair of a query and a FAQ, a relevancy level was tagged based on the question

---

7) All FAQs on http://www.kingnet.com.tw were collected in February 2012.

8) The test queries were collected in June 2012.

9) Available at http://www.healthcare.com.tw/healthcare-front/.

10) Available at http://www.ch.com.tw/index.asp?title=1.

11) Available at http://www.tmn.idv.tw/.

12) Available at http://olddoc.tmu.edu.tw/pinging/index.htm.

13) Available at http://cisc.twbbs.org/index.php.

parts of the query and the FAQ: *definitely relevant*, *partially relevant*, and *non-relevant*. Among the 200 queries, 129 queries had relevant (definitely relevant or partially relevant) FAQs and the average number of relevant FAQs of a query was 3.87.

5.2. Underlying FAQ Retrievers

Given that no previous retrievers considered conceptual similarity as ECA (ref. Section 2), we aimed at investigating the extent to which the conceptual similarity provided by ECA was helpful for different kinds of FAQ retrievers to have significantly better performance in identifying relevant FAQs for healthcare queries. Therefore, ECA collaborated with several popular FAQ retrievers, including FAQFinder, Lucene, BM25, query likelihood language model (LM), and their integration by RankingSVM, which a popular technique used to integrate multiple scorers for better ranking.

FAQFinder (Burke et al., 1997) combined three parts of similarities between q (a query) and f (a FAQ): (1) cosine-based term-vector similarity, (2) relatedness of terms measured by the distance of the terms in an ontology, and (3) percentage of the terms in q that matched the terms in f. FAQFinder served as a baseline in many previous studies as well (e.g., Wu et al., 2005). The similarity measures employed by FAQFinder were employed by many previous FAQ retrievers as well. For example, the cosine similarity based on the vectors of input queries and FAQs (Bernhard & Gurevych, 2008; Jeon et al., 2005; Wu et al., 2005) was shown to be one of the best in FAQ retrieval (Bernhard & Gurevych, 2008), and an ontology was employed to measure the relatedness of the words in input queries and FAQs (Wu et al., 2006; Wu et al., 2005).

To make FAQFinder able to process Chinese healthcare questions more properly, a sequence of preprocessing steps were conducted to q and f: (1) segmenting the question into Chinese terms by the CKIP system[14]; (2) translating the Chinese terms into English terms by the Google translation system[15]; (3) identifying the concept ID for each term on a medical ontology UMLS[16] (Unified Medical Language System) by the MMTx terminology matching system.[17] The concept IDs were used to locate the concepts on the ontology MRREL[18] from UMLS so that the distance between two concepts on an ontology could be measured (for measuring the second part of similarity values considered by FAQFinder). Note that the second part of FAQFinder has a parameter that controls the maximum distance between two terms on the ontology. To tune the parameter, we conducted 4-fold cross validation using the 129 test queries that have relevant FAQs: 3/4 of the queries were used for parameter tuning and the remaining 1/4 of the queries were used for testing, and the experiment repeated four times.

On the other hand, Lucene estimated similarities between queries and documents based on the vector-space model.[19] It is a good FAQ retriever tested in several previous studies (e.g., Bernhard

---

14) Available at http://ckipsvr.iis.sinica.edu.tw/.

15) Available at http://translate.google.com.tw/.

16) Available at http://www.nlm.nih.gov/research/umls/.

17) Available at http://metamap.nlm.nih.gov/.

18) Available at http://www.nlm.nih.gov/research/umls/licensedcontent/umlsknowledgesources.html.

19) Lucene is available at http://lucene.apache.org.

& Gurevych, 2008; Jijkoun & de Rijke, 2005; Kothari et al., 2009). Moreover, both BM25 and the query likelihood language model (LM) were shown to be the best baseline FAQ retrievers as well (Jeon et al., 2005; Xue et al., 2008). LM was also a basis on which new FAQ retrieval techniques were developed (e.g., Lee et al., 2008b). BM25 estimated the similarity (S_{BM25}) between q and f with Equation 18.

$$S_{BM25}(q,f) = \sum_{q_i \in q \cap f} idf(q_i) \cdot \frac{c(q_i,f) \cdot (k_1+1)}{c(q_i,f) + k_1 \cdot (1-b+b \cdot \frac{|f|}{avgfl})} \qquad (18)$$

In Equation 18, $c(q_i,f)$ is the times q_i appearing in f; $idf(q_i)$ is inverse document frequency of q_i measured by $Log_2(N/df(q_i))$ with N being the number of FAQs and $df(q_i)$ the number of FAQs in which q_i appears; $|f|$ is the length of f (i.e., number of terms in f); k_1=2, b=0.75, and $avgfl$ is average length of FAQ questions. On the other hand, LM estimated the similarity (S_{LM}) between q and f with Equation 19.

$$S_{LM}(q,f) = \prod_{q_i \in q} (1-\lambda)(\frac{c(q_i,F)}{|F|}) + \lambda(\frac{c(q_i,f)}{|f|}) \qquad (19)$$

In Equation 19, F is the set of FAQs; $|F|$ is the total number of terms in F; $c(q_i,F)$ is the times q_i appearing in F; and the parameter λ was tuned in the 4-fold experiments. To make Lucene, BM25, and LM able to process Chinese healthcare questions more properly, a sequence of preprocessing steps were conducted to q and f (by the same preprocessing steps for FAQFinder noted above).

We are also interested in the integration of the FAQ retrievers, since integration of multiple techniques often produces better ranking performance than the individual ones. To integrate the FAQ retrievers, we employed RankingSVM (Joachims, 2002), which is one of the best techniques routinely used to integrate multiple scorers to achieve better ranking (e.g., Liu & Huang, 2011; Veloso, Almeida, Goncalves, & Meira, 2008; Zhou, Xue, Zha, & Yu, 2008). We employ SVM^{rank} (Joachims, 2006) to implement RankingSVM.[20] As the retrievers could produce different scales of similarity scores, the similarity scores were normalized to the range of [0, 1] so that the retrieves could be integrated more properly.[21] We implemented two integrations: integration of the three parts of FAQFinder, and integration all 6 retrievers including three parts of FAQFinder, Lucene, BM25, and LM. The former is named FAQFinder, and the latter is named IntALL. As noted above, we conducted 4-fold experiments in which RankingSVM was trained and tested four times and the average result was reported.

Therefore, we actually have 8 underlying FAQ retrievers, including 6 individual retrievers (i.e., three parts of FAQFinder, Lucene, BM25, and LM) and 2 integrated ones (i.e., FAQFinder and

---

20) SVM^{rank} is available at http://www.cs.cornell.edu/People/tj/svm_light/svm_rank.html.

21) The similarity score between q and f is normalized by dividing it with the maximum similarity score between q and all FAQs. Therefore, the normalized similarity score is in the range of [0, 1] without changing the original order of the FAQs with respect to q.

IntALL). The underlying FAQ retrievers can represent state-of-the-art of FAQ retrieval techniques. We are interested in the collaboration between ECA and each of the retrievers, and hence ECA was integrated with the retrievers by RankingSVM as well. By comparing the performance of the retrievers *before* and *after* ECA was used, we could measure the contribution of ECA.

5.3. Evaluation Criteria

We employed *mean average precision* (MAP) as the evaluation criterion to measure how relevant FAQs are ranked higher. MAP is defined in Equation 20.

$$MAP = \frac{\sum_{i=1}^{|Q|} P(i)}{|Q|}, \;\; AP(i) = \frac{\sum_{j=1}^{k} \frac{j}{FAQ_i(j)}}{k} \tag{20}$$
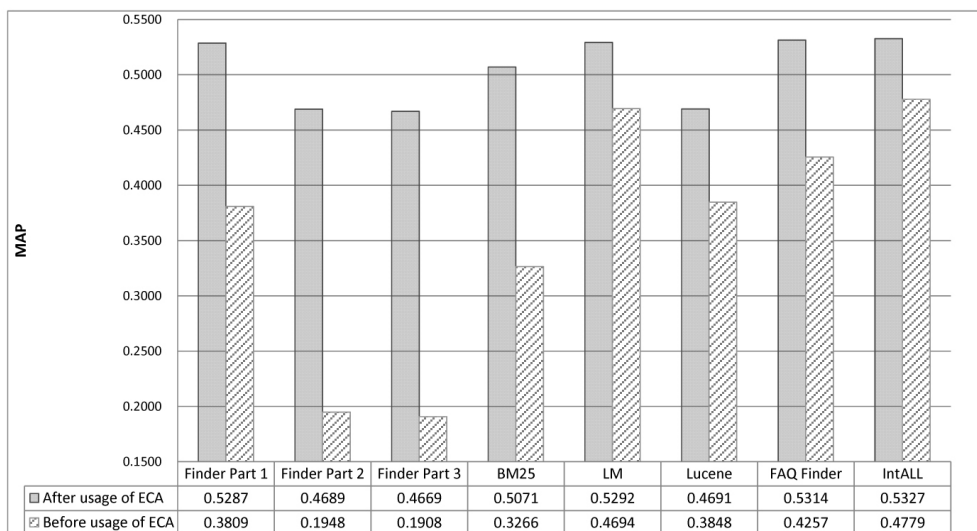
In Equation 20, $|Q|$ is the number of queries, k is number of relevant FAQs for the ith query, and $FAQ_i(j)$ is the number of FAQs whose ranks are higher than or equal to that of the jth relevant FAQ for the ith query. That is, $AP(i)$ is actually the average precision of the ith query, and MAP is the average of the AP values of all queries. When computing AP for a query, those FAQs that are definitely relevant or partially relevant to the query were considered to be relevant to the query.

5.4. Result and Discussion

We discuss and analyze the contribution of ECA to the ranking of relevant healthcare FAQs. Moreover, as the automatic concept recognition is a key task of ECA, we are also concerned with the performance of ECA under different settings of the knowledge for the concept recognition (i.e., patterns and terms, ref. Section 4.1.1 ~ Section 4.1.3). We investigate the performance of ECA under different amounts of the knowledge and then investigate the effect of replacing the automatic concept recognition with manual concept annotation.

5.4.1. Ranking of Relevant Healthcare FAQs
Figure 2 shows the ranking performance of each underlying retriever before and after it is integrated with ECA. The results showed that the conceptual similarity information provided by ECA was helpful for all the retrievers, which have considered different kinds of similarity information employed by state-of-the-art FAQ retrievers. To verify whether the performance improvements were statistically significant, we conducted two-sided and paired t-test with 95% confidence level. The results showed that ECA helped all the retrievers to achieve significantly better performance in MAP. Among the retrievers, LM and IntALL had the best performance, and ECA significantly improved them as well.

	Finder Part 1	Finder Part 2	Finder Part 3	BM25	LM	Lucene	FAQ Finder	IntALL
▣ After usage of ECA	0.5287	0.4689	0.4669	0.5071	0.5292	0.4691	0.5314	0.5327
▨ Before usage of ECA	0.3809	0.1948	0.1908	0.3266	0.4694	0.3848	0.4257	0.4779

Fig. 2. Results on ranking of relevant FAQs: Each retriever with ECA has better MAP performance
than the retriever without ECA, and all the performance differences are statistically significant.

Table 3 shows an example to illustrate the contribution of ECA. The query q asks for the advice about whether seasonings can be added to non-staple food for infants. Among the FAQs, there is a FAQ ($f1$) that talks about the possible problems that should be considered when feeding non-staple food, and hence the FAQ is judged to be relevant to q. Unfortunately, both the best FAQ retrievers (LM and IntALL) make an error of preferring a non-relevant FAQ ($f2$) to $f1$. The non-relevant FAQ talks about the way to deal with the case where the body temperature of an infant is too low. LM and IntALL prefer the non-relevant FAQ because it shares a term "嬰兒" (infant) with q and the term has higher *idf*. ECA successfully gives the non-relevant FAQ a smaller similarity score (S_{ECA}) as $e1_{f2}$ ("體溫太低," body temperature is too low) in the non-relevant FAQ cannot be found in q. Also note that in the example, recognition of essential concepts by ECA is not perfect, especially for the recognition of aspects (see the underlined parts in Table 3). ECA misclassifies $f2$ into the *cause* aspect (i.e., A_{cause} = True) because "怎麼來" (how does it happen) is a term for the cause aspect (i.e., the term is listed in TA_{cause}) and the term happens to match "怎麼" (by Equation 1), which is simply a substring of "怎麼辦" (what to do) in $f2$ (i.e., $f2$ should be classified into process aspect rather than the cause aspect). Even the recognition of concepts cannot always be perfect, ECA can still provide significant contribution to the ranking of relevant healthcare FAQs. We are also exploring the possible ways to further improve the recognition of the concepts in FAQs.
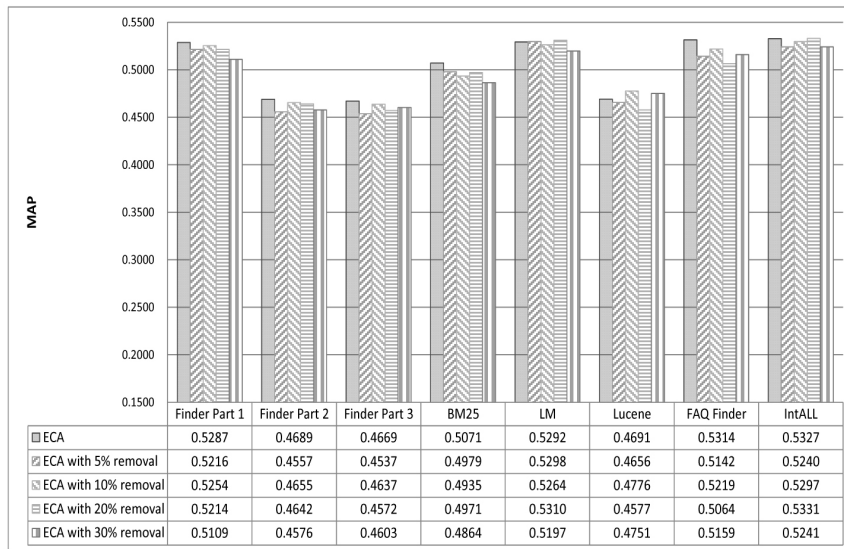
Table 3. An example to illustrate the contribution of ECA: Given the query q, both the best FAQ retrievers (LM and IntALL) make an error of preferring the non-relevant FAQ ($f2$) to the relevant one. Although the recognition of essential concepts by ECA cannot be perfect (see the underlined parts), ECA successfully gives the non-relevant FAQ a smaller similarity score (S_{ECA}) as $e1_{f2}$ ("體溫太低") in the non-relevant FAQ cannot be found in q.

Question of the FAQ	Essential concepts recognized by ECA	Query (q): 嬰兒副食品是否可以添加調味料? (Can seasonings be added to non-staple food for infants?)	
		Aspects in q	Events and conditions in q w.r.t. the FAQ
$f1$ (Relevant FAQ): 餵食副食品可能會遇到的問題? (The problems that should be considered when feeding non-staple food?)	**(E)** $e1_{f1}$ =餵食副食品 (feeding non-staple food); $e2_{f1}$ =題; **(C)** C_{f1}: None; **(A)** A_{cause} = $A_{process}$ = $A_{diagnosis}$ = *don't-care*.	A_{cause} = True (by "可以讓" (can cause)); $A_{process}$ = True (by "調理" (recuperate)); $A_{diagnosis}$ = False.	$e1_{q,f1}$ =副食品 (non-staple food); $e2_{q,f1}$: None; $C_{q,f1}$: None. → S_{ECA} = 0.7243 (= average of the scores on $e1_{q,f1}$ and aspects)
$f2$ (Non-relevant FAQ): 嬰兒體溫太低怎麼辦? (What to do for an infant whose body temperature is too low?)	**(E)** $e1_{f2}$=體溫太低(body temperature is too low); **(C)** C_{f2} =嬰兒(infant); **(A)** A_{cause} = True (by "怎麼來" (how does it happen)); $A_{process}$ = True (by "怎麼辦" (what to do)); $A_{diagnosis}$ = False.		$e1_{q,f2}$: None; $e2_{q,f2}$: None; $C_{q,f2}$ =嬰兒(infant). → S_{ECA}=0 (since $e1_{f2}$ is not found in q)

5.4.2. Effect of Different Knowledge Engineering Strategies for Concept Recognition

As described in Section 4.1.1 ~ Section 4.1.3, several terms and patterns are defined for ECA to recognize essential concepts in FAQs so that conceptual similarity between the FAQs and the queries can be estimated. The development of these patterns and terms is actually a knowledge engineering task, and hence we are concerned with the performance of ECA under different strategies for the knowledge engineering task.

We first investigate the robustness of ECA when less knowledge is defined for the concept recognition. We tested different degrees of the removal by randomly removing 5%, 10%, 20%, and 30% of the terms and patterns for concept recognition. The result is shown in Figure 3. It indicated that as more of the terms and patterns were removed (e.g., 30% of them are removed), MAP of ECA tended to drop slightly. However, after conducting significance tests on the performance differences, we found that most of the differences were *not* statistically significant. Significant differences occurred only in three cases: BM25 with 5% removal, FAQFinder with 20% removal, and IntALL with 5% removal. The result shows that ECA is robust in facing the different sets of terms and patterns for concept recognition. A detailed analysis shows that the robustness is contributed by the fuzzy term matching method (i.e., *StrSim* in Equation 1) that is employed by ECA to measure the similarity between each query and FAQ (ref.. Equation 8 ~ Equation 16). When less knowledge is defined to recognize the concepts in each FAQ, more errors might be incurred in concept recognition but the fuzzy string method helps to reduce the effect of the errors by producing robust similarity scores.

	Finder Part 1	Finder Part 2	Finder Part 3	BM25	LM	Lucene	FAQ Finder	IntALL
ECA	0.5287	0.4689	0.4669	0.5071	0.5292	0.4691	0.5314	0.5327
ECA with 5% removal	0.5216	0.4557	0.4537	0.4979	0.5298	0.4656	0.5142	0.5240
ECA with 10% removal	0.5254	0.4655	0.4637	0.4935	0.5264	0.4776	0.5219	0.5297
ECA with 20% removal	0.5214	0.4642	0.4572	0.4971	0.5310	0.4577	0.5064	0.5331
ECA with 30% removal	0.5109	0.4576	0.4603	0.4864	0.5197	0.4751	0.5159	0.5241

Fig. 3. Robustness of ECA under different settings of the knowledge (terms and patterns) for concept recognition: Although MAP of ECA tends to drop when a part of the knowledge is removed, most of the MAP differences are *not* statistically significant.

Moreover, we are also interested in the performance of ECA when the concept recognition is conducted *manually* (i.e., disabling the offline concept recognition tasks of ECA and replacing it with manual annotation). As noted in the example discussed above (ref. Table 3), recognition of essential concepts by ECA cannot always be perfect. Therefore, manual annotation for the essential concepts could be helpful for the ranking of relevant FAQs. Another motivation of considering the manual annotation is that the manual annotation should be both *feasible* and *helpful* for FAQ retrieval, based on two reasons: (1) concept annotation to a FAQ is conducted only once (e.g., conducted when the FAQ is edited and entered to the database), and (2) the annotation can provide higher-quality information to ECA. Therefore, following the definition of the essential concepts, each FAQ was manually annotated with essential concepts.

Figure 4 compares the performance of the two versions of ECA: ECA with automatic concept recognition and ECA with manual concept annotation for FAQs. We surprisingly found that manual annotation did not always help ECA to achieve better MAP. All performance differences between the two versions of ECA were *not* statistically significant. Therefore, the imperfect concept recognition by ECA did not significantly deteriorate its performance in the experiment. The result reconfirms the robustness of ECA under different strategies of concept recognition for FAQs. As the concept annotation task is the offline task of ECA, the result also suggests that future extension of ECA should be directed to the online modules of ECA (i.e., similarity measurement, Section 4.2). We are interested in extending ECA by text classification (for aspects), and otology-based and translation-based term matching (for event and condition concepts).
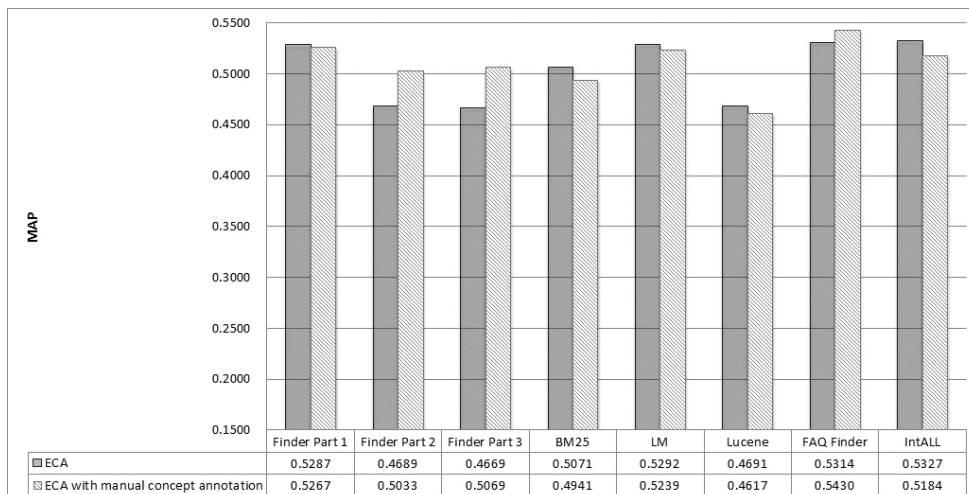
	Finder Part 1	Finder Part 2	Finder Part 3	BM25	LM	Lucene	FAQ Finder	IntALL
☐ ECA	0.5287	0.4689	0.4669	0.5071	0.5292	0.4691	0.5314	0.5327
☒ ECA with manual concept annotation	0.5267	0.5033	0.5069	0.4941	0.5239	0.4617	0.5430	0.5184

Fig. 4. Possible improvement of ECA by manual annotation of the concepts in FAQs: Manual annotation is not always helpful for ECA, and all MAP differences are *not* statistically significant.

6. future research directions

ECA can be extended in two ways. The first way is to expand the conceptual model employed by ECA (ref. Table 1). Currently ECA considers three aspect categories, which are respectively about the cause, diagnosis, and process of disorders. As users (healthcare consumers) are often concerned with the information about different stages of a disorder, it would be interesting to expand the conceptual model by considering four typical stages of a disorder, including: (1) *before* the disorder is diagnosed, (2) *when* the disorder is being diagnosed, (3) *after* the disorder is confirmed, and (4) *after* the disorder is treated. Typical categories may thus include one most-general category (*general description*), four general categories (*prevention*, *diagnosis*, *treatment*, and *prognosis*), and seven specific categories (*risk factors*, *symptoms and signs*, *lab test*, *homecare*, *medicine*, *mortality*, and *recurrence*). It is thus interesting to investigate whether the retrieval of healthcare FAQs can be improved by the expanded conceptual model.

Another interesting way to extend ECA is to develop a machine learning technique to build a classifier to determine the aspect category of a given healthcare FAQ. Currently ECA determines the aspect category by string matching (ref. Section 4.1.1). Although the string matching method performs well in the experiments, it may be costly to construct and maintain a complete set of strings for each aspect category, especially when more aspect categories are considered. Therefore, given a set of healthcare FAQs that are labeled with suitable aspect categories, a machine learning technique can be developed to automatically train a classifier, without needing to manually construct and maintain the strings for each category. It is thus interesting to investigate whether the retrieval of healthcare FAQs can be improved by the machine-learning-based classifier.

It is also interesting to apply ECA to retrieval of healthcare FAQs in languages other than Chinese. ECA is based on a healthcare conceptual model, which has been shown to be helpful for healthcare FAQ retrieval. The conceptual model should be applicable to healthcare FAQs in different natural languages, since healthcare consumers are often concerned with disease management and health promotion, no matter how they express their concerns in different languages. Technically, recognition of the three kinds of concepts (i.e., event, aspect, and condition) in a FAQ should call for different techniques when the FAQ is expressed in different languages.

7. Conclusions

Healthcare FAQs are a valuable source of readable and reliable healthcare information for healthcare consumers. Given a Chinese healthcare question as a query, semantic similarities between the query and FAQs are essential for the ranking of relevant FAQs. To estimate the semantic similarities, semantic analysis on Chinese healthcare FAQs is required, however it is challenging. As healthcare FAQs are actually specific questions about health promotion and disease management, their semantic structure is quite different from those of the FAQs in other domains. Therefore, we identify three types of essential concepts (*event*, *condition*, and *aspect*) as the key semantic elements in Chinese healthcare FAQs, and show that based on the essential concepts, the conceptual similarities between Chinese healthcare queries and FAQs can be estimated by a technique ECA.

The conceptual similarities provided by ECA actually indicate a kind of semantic information that is not considered by previous FAQ retrievers and can serve as a supplement to the FAQ retrievers. An empirical evaluation on thousands of Chinese healthcare FAQs confirms that the similarity scores produced by ECA can be used to significantly enhance several kinds of FAQ retrievers in ranking the FAQs for the input queries. Performance of ECA is also robust under different settings of the knowledge (terms and patterns) for concept recognition. The results thus confirm the expressive power of the conceptual model and the significant contribution of ECA to the ranking of relevant healthcare FAQs. The contribution is of technical significance to the studies of FAQ retrieval in the healthcare domain. It is also of practical significance to the utility of healthcare FAQs, which provide both reliable and readable healthcare information for specific questions of healthcare consumers. We thus make ECA online (http://203.64.84.94:126/). Given two Chinese healthcare questions, the online service returns a conceptual similarity score. It can thus provide similarity estimation for various kinds of FAQ retrievers, facilitating further technical studies and practical applications in the healthcare domain.

Acknowledgement

References

Bernhard, D., & Gurevych, I. (2008). Answering learners' questions by retrieving question paraphrases from social Q&A sites. *Proceedings of the Third Workshop on Innovative Use of NLP for Building Educational Applications*, 44-52.

Burke, R. D., Hammond, K. J., Kulyukin, V., Lytinen, S. L., Tomuro, N., & Schoenberg, S. (1997). Question answering from frequently asked question files: Experiences with the faq finder system. *AI magazine, 18*(2), 57.

Casellas, N., Casanovas, P., Vallbé, J. J., Poblet, M., Blázquez, M., Contreras, J., ... & Benjamins, V. R. (2007). Semantic enhancement for legal information retrieval: Iuriservice performance. *Proceedings of the 11th international conference on Artificial intelligence and law*, 49-57.

Contractor, D., Kothari, G., Faruquie, T. A., Subramaniam, L. V., & Negi, S. (2010). Handling noisy queries in cross language faq retrieval. *Proceedings of the 2010 conference on empirical methods in natural language processing*, 87-96.

Jeon, J., Croft, W. B., & Lee, J. H. (2005). Finding similar questions in large question and answer archives. *Proceedings of the 14th ACM international conference on Information and knowledge management*, 84-90. doi:10.1145/1099554.1099572

Jijkoun, V., & de Rijke, M. (2005). *Retrieving answers from frequently asked questions pages on the web*. Proceedings of the 14th ACM international conference on Information and knowledge management, 76-83.

Joachims, T. (2002). Optimizing search engines using clickthrough data. Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining, 133-142.

Joachims, T. (2006). *Training linear SVMs in linear time.* In Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, 217-226.

Kim, D. S., & Noh, Y. (2014). A study of public library patrons' understanding of library records and data privacy. *International Journal of Knowledge Content Development & Technology, 4*(1), 53-78.

Kothari, G., Negi, S., Faruquie, T. A., Chakaravarthy, V. T., & Subramaniam, L. V. (2009). *SMS based interface for FAQ retrieval.* Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP, 2(7), 852-860.

Lee, C. W., Day, M. Y., Sung, C. L., Lee, Y. H., Jiang, T. J., Wu, C. W., ... & Hsu, W. L. (2008). Boosting Chinese question answering with two lightweight methods: ABSPs and SCO-QAT. *ACM Transactions on Asian Language Information Processing, 7*(4), 12.

Lee, J. T., Kim, S. B., Song, Y. I., & Rim, H. C. (2008). Bridging lexical gaps between queries and questions on large online Q&A collections with compact translation models. Proceedings of the Conference on Empirical Methods in Natural Language Processing, 410-418.

Lin, J., & Demner-Fushman, D. (2006). The role of knowledge in conceptual retrieval: a study in the domain of clinical medicine. Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval, 99-106.

Liu, R.-L., & Huang, Y.-C. (2011). Ranker Enhancement for Proximity-based Ranking of Biomedical

Texts. *Journal of the American Society for Information Science and Technology, 62*(12), 2479-2495.

Liu, R. L., & Lin, S. L. (2012). A Conceptual Model for Retrieval of Chinese Frequently Asked Questions in Healthcare. Information Retrieval Technology, 366-375.

Mishra, M., Mishra, V. K., & Sharma, H. R. (2013). Question Classification using Semantic, Syntactic and Lexical features. *International Journal of Web & Semantic Technology, 4*(3).

Pan, Y., Tang, Y., Lin, L., & Luo, Y. (2008). Question classification with semantic tree kernel. Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval, 837-838.

Riezler, S., Vasserman, A., Tsochantaridis, I., Mittal, V., & Liu, Y. (2007). Statistical machine translation for query expansion in answer retrieval. *Annual Meeting-Association For Computational Linguistics, 45*(1), 464.

Sillence, E., Briggs, P., Fishwick, L., & Harris, P. (2004). Trust and mistrust of online health sites. Proceedings of the SIGCHI conference on Human factors in computing systems, 663-670.

Veloso, A. A., Almeida, H. M., Gonçalves, M. A., & Meira Jr, W. (2008). Learning to rank at query-time using association rules. Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval, 267-274.

Wang, K., Ming, Z., & Chua, T. S. (2009). A syntactic tree matching approach to finding similar questions in community-based qa services. Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval, 187-194.

Winiwarter, W. (2000). Adaptive natural language interfaces to FAQ knowledge bases. *Data & Knowledge Engineering, 35*(2), 181-199.

Wu, C. H., Yeh, J. F., & Chen, M. J. (2005). *Domain-specific FAQ retrieval using independent aspects.* ACM Transactions on Asian Language Information Processing, 4(1), 1-17.

Wu, C. H., Yeh, J. F., & Lai, Y. S. (2006). Semantic segment extraction and matching for internet FAQ retrieval. Knowledge and Data Engineering, *IEEE Transactions on, 18*(7), 930-940.

Xue, X., Jeon, J., & Croft, W. B. (2008). *Retrieval models for question and answer archives.* Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval, 475-482.

Zhou, K., Xue, G. R., Zha, H., & Yu, Y. (2008). *Learning to rank with ties.* Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval (pp. 275-282). ACM.