A Conceptual Model for Retrieval of Chinese Frequently Asked Questions in Healthcare

Rey-Long Liu and Shu-Ling Lin
Department of Medical Informatics
Tzu Chi University, Hualien, Taiwan
rlliutcu@mail.tcu.edu.tw

Abstract. Frequently asked questions (FAQs) in healthcare provide general readers with both reliable and readable healthcare information. In this paper, we present a conceptual retrieval technique that serves as a supplement to enhance existing FAQ retrievers to find Chinese healthcare FAQs for each input query. By analyzing the structures and goals of Chinese healthcare FAQs, we identify three types of essential concepts in healthcare FAQs: 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). The proposed conceptual retrieval technique is thus named ECA (Event, Condition, and Aspect). Given healthcare FAQs annotated by the three types of concepts, ECA can measure the conceptual similarities between an input query and the FAQs. Empirical evaluation on real-world Chinese healthcare FAQs shows that the conceptual similarity information provided by ECA is helpful for an FAQ retriever to have significantly better performance in identifying relevant FAQs for input queries.

Keywords: Frequently Asked Questions, Healthcare, Conceptual Retrieval.

1 Introduction

Healthcare information should be both reliable and readable. Frequently asked questions (FAQs) in healthcare provide such reliable and readable healthcare information as they are often written and compiled by healthcare professionals for general reader. Many healthcare information providers thus collect and maintain a large number of healthcare FAQs for general readers. Given an input query in natural language form, the retrieval of relevant FAQs is thus essential for the utility of the readable and reliable healthcare information for health promotion and disease management.

Many FAQ retrieval techniques have been developed in previous studies. Given a database of FAQs and a natural language question as an input query, an FAQ retriever ranks the FAQs in the database 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 much information to identify relevant FAQs.

In this paper, we analyze the conceptual structure of Chinese healthcare FAQs and employ the conceptual structure to provide helpful information to existing FAQ
retrieves so that relevant FAQs can be ranked higher. More specifically, we present a conceptual retrieval technique that serves as a supplement to enhance existing FAQ retrievers to find Chinese healthcare FAQs relevant to input queries. In the conceptual model, we identify three types of essential concepts: 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 an input query, the FAQ retriever should consider the similarities on the three types of concepts. Therefore, our conceptual retrieval technique for healthcare FAQs is named ECA (Event, Condition, and Aspect), which aims at providing such conceptual similarity information to the retriever so that more relevant FAQs may be ranked higher for the input query.

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 above conceptual structures of Chinese healthcare FAQs, and hence the collaboration between ECA and the previous retrievers can further enhance the retrievers by conceptual similarity information. An empirical evaluation on thousands of Chinese healthcare FAQs shows that, by collaborating with ECA, an FAQ retriever can perform significantly better in ranking relevant FAQs higher for input queries.

2 Related Work

An FAQ often consists of two parts: a question part and an answer part. Several previous studies considered both parts for FAQ retrieval (e.g., [6][7][8]). In this paper, the proposed technique ECA focuses on the question part, which has been shown to be the most important part in FAQ retrieval [4]. ECA provides the conceptual similarity on the question part as additional information for FAQ retrievers.

Given \( q \) and \( f \) as an input query (question) and the question part of an FAQ respectively, previous techniques have employed several types of information to estimate the similarity between \( q \) and \( f \): (1) the overlap of the words in \( q \) and \( f \) [1]; (2) the cosine similarity based on the vectors of \( q \) and \( f \) [1–2][7]; (3) the relatedness of the words in \( q \) and \( f \) (e.g., measured by the distance of the words on an ontology [2][6] [7] or the correction of the spelling errors of words in \( q \) [1][3]); (4) the correlation between the question types of \( q \) and \( f \) [7]; (5) the mapping or translation between the words in \( q \) and \( f \) (for tackling the problem of word mismatch between \( q \) and \( f \) [4][8]); and (6) the similarity between the syntactic or semantic structures of \( q \) and \( f \) (by deeper analysis such as parsing [6]).

However, none of the similarity estimation techniques considered essential concepts (event, condition, and aspect) in Chinese healthcare FAQs. ECA is the first framework aiming at the conceptual retrieval of FAQs for healthcare consumers.