Chapter 7
Toward Automatic Answers in User–Interactive Question Answering Systems

Tianyong Hao
Shanghai University of Electric Power, China

Feifei Xu
Shanghai University of Electric Power, China

Jingsheng Lei
Shanghai University of Electric Power, China

Liu Wenyin
City University of Hong Kong, China

Qing Li
City University of Hong Kong, China

ABSTRACT
A strategy of automatic answer retrieval for repeated or similar questions in user-interactive systems by employing semantic question patterns is proposed in this paper. The used semantic question pattern is a generalized representation of a group of questions with both similar structure and relevant semantics. Specifically, it consists of semantic annotations (or constraints) for the variable components in the pattern and hence enhances the semantic representation and greatly reduces the ambiguity of a question instance when asked by a user using such pattern. The proposed method consists of four major steps: structure processing, similar pattern matching and filtering, automatic pattern generation, question similarity evaluation and answer retrieval. Preliminary experiments in a real question answering system show a precision of more than 90% of the method.

1. INTRODUCTION
Though just emerged a couple of years ago, User-Interactive Question Answering (UIQA) (Liu et al., 2009; Buy Answers Consultation Ltd, 2011), aka, Collaborative (or Community) Question Answering (CQA) systems, e.g., Yahoo! Answers, Baidu Knows, and Naver, are becoming popular online information services now. Compared with a search engine, which can automatically return a long list of ranked, possibly relevant but usually long documents, and hence requires a user’s manual, and sometimes, very tedious filtration to find the exact answer, a question answering
system just returns direct and exact answers, and thus, is more preferable. Moreover, only about 80% of desired information can be found by search engines, and even if the user is lucky, it still takes time for the user to manually filter and find the exact answer from the returned long list of documents. With the collaborative efforts from the huge number of users with various knowledge backgrounds, the user may obtain his/her desired information from UIQA/CQA systems with little effort. Hence, we are anticipating a new wave of hot Web information services based on such UIQA/CQA systems since they are natural complements to the automatic search engines; or alternatively, they can be used as manual search engines to meet the ever-increasing information need of users.

Given their Web 2.0 nature, UIQA/CQA systems have accumulated more and more questions and answers with ratings, which in turn become good sources of automatic search engines. For example, Yahoo! Answers alone has acquired an archive of more than 40 million questions and 500 million answers, according to 2008 estimates. However, unaware of previous questions and answers accumulated, or being unwilling to spend time in searching for them, people often ask repeated or very similar questions. As a result, repeated answers need to be provided again and again and cause a huge waste of resources. Hence, it is a new challenge to develop a suitable method that can make reuse of these accumulated questions and answers effectively.

In this paper, we propose a new scheme in our UIQA system (Buy Answers Consultation Ltd, 2011) which can accumulate the questions and answers in a more structured, and even a more semantic-oriented way. We wish such scheme can make it possible to automatically and accurately answer repeated or similar questions based on previously accumulated question-answer pairs (QA pairs). The proposed scheme involves so-called semantic question patterns, each of which is designed to generalize a class of questions with the same sentence structure and relevant semantics (Hao et al., 2007). A question pattern is a generalized question with one or several blank slots, each of which is referred to as a variable component in this paper. Each variable component in a particular pattern is annotated by a semantic label, which is used not only to remind users to fill in correct words when using it, but also let machines know the semantics (from the label) of the filled-in words. Hence, it is called a semantic question pattern.

An initial set of frequently used semantic question patterns are extracted using an automatic question pattern generation method (Hao et al., 2008). These extracted patterns are further revised by human experts and stored in the pattern database in our system (Buy Answers Consultation Ltd, 2011). The system encourages users to ask questions with question patterns. When a user asks a new question in free-text, the system first tries to match it with a suitable pattern in the pattern database. If there is no pattern matched well, the system automatically creates a new pattern and lets the user confirm. Once the user confirms, the created pattern is added into the pattern database which is subject to validation by administrators (or experts). Whether a well-matched pattern is found or a new pattern is created, the question is converted into a pattern-based question, which consists of the pattern ID and the filled-in words for the variable component(s). Such pattern-based questions are accumulated in the question database with certain structures and semantics. Based on such organization, finding similar questions becomes much easier by just checking whether they have the same pattern ID and matching the filled-in words in the variable component(s). Once a similar question is found, its correct answer(s) can be used to reply the new question. Hence, higher efficiency and accuracy are expected than simply matching all text in the questions.

Using semantic question patterns, our automatic approach to answering repeated or similar questions consists of four main steps: (1) structure processing, (2) pattern matching and filtering, (3) automatic pattern generation, (4) question similarity evaluation and answer retrieval. Step