Chapter 1

Question Answering and Generation

Arthur C. Graesser
The University of Memphis, USA

Vasile Rus
The University of Memphis, USA

Zhiqiang Cai
The University of Memphis, USA

Xiangen Hu
The University of Memphis, USA

ABSTRACT

Automated Question Answering and Asking are two active areas of Natural Language Processing with the former dominating the past decade and the latter most likely to dominate the next one. Due to the vast amounts of information available electronically in the Internet era, automated Question Answering is needed to fulfill information needs in an efficient and effective manner. Automated Question Answering is the task of providing answers automatically to questions asked in natural language. Typically, the answers are retrieved from large collections of documents. While answering any question is difficult, successful automated solutions to answer some type of questions, so-called factoid questions, have been developed recently, culminating with the just announced Watson Question Answering system developed by I.B.M. to compete in Jeopardy-like games. The flip process, automated Question Asking or Generation, is about generating questions from some form of input such as a text, meaning representation, or database. Question Asking/Generation is an important component in the full gamut of learning technologies, from conventional computer-based training to tutoring systems. Advances in Question Asking/Generation are projected to revolutionize learning and dialogue systems. This chapter presents an overview of recent developments in Question Answering and Generation starting with the landscape of questions that people ask.

DOI: 10.4018/978-1-60960-741-8.ch001
INTRODUCTION

For the first time in history, a person can ask a question on the web and receive answers in a few seconds. Twenty years ago, it would take hours or weeks to receive answers to the same questions as a person hunted through documents in a library. In the future, electronic textbooks and information sources will be mainstream and they will be accompanied by sophisticated question answering and generation facilities. As a result, we believe that the Google generation is destined to have a much more inquisitive mind than the generations who relied on passive reading and libraries. The new technologies will radically transform how we think and behave.

BACKGROUND

Automatic Question Answering is the task of providing meaningful answers to questions in natural language. There are two features that make automatic Question Answering attractive: (1) it keeps the user-system interface natural because users can ask questions the way they ask other humans, thereby eliminating the need to train users on specific query languages and (2) it can provide effective access for everyone to the huge online repository of knowledge stored on the Internet. Advanced online Question Answering services can provide effective access to information to everyone, computer-savvy or not, as interface barriers are eliminated.

Early explorations of automated Question Answering have been attempted since the beginning of the computing era, but the advent of the Internet in the 1990s greatly stimulated research on Question Answering in order to provide effective access to the vast repositories of information available on the web. In particular, research during the last decade has focused on building Question Answering technologies that can successfully answer one type of questions, namely factoid questions, which have well-defined, relatively short answers. This chapter emphasizes these recent developments on factoid Question Answering.

The reverse process of Question Generation (QG) or asking is a fundamental human capacity that is present in childhood as a primary form of learning, curiosity, and discovery. Students in K12, college, and adult populations are known to improve their learning after they learn how to acquire improved skills of QG. QG is an essential component of learning environments, help systems, information seeking systems, and a myriad of other applications. Mechanisms of QG have been less explored in the Computational Linguistics and Text Retrieval community, the two communities that led the recent efforts on Question Answering processes. We know that language generation is a very difficult task to take on, as all natural language generation tasks are, but do not believe that the inherent difficulty should prevent the exploration of automated QG. Recent efforts by Rus and colleagues (2007, 2009a, 2009b) launched the creation of a coherent and strong QG research community that has grand research plans for the next decade.

QUESTION QUALITY, COMPLEXITY, AND TAXONOMIES

An important initial step in a Question Answering or Generation project is to take stock of the landscape of question categories so that researchers can specify what types of questions they have in mind, as well as the educational context (Rus, Cai, & Graesser, 2007). This section identifies some QG categories, taxonomies, and dimensions that might be considered. The complexity and quality of the questions systematically vary across the broad landscape of questions. Finding the relevant criteria of question quality is a key requirement for good performance of QG systems. What we present in this section is merely the tip of the iceberg.