Question Answering with Imperfect Temporal Information

Steven Schockaert[1], David Ahn[2], Martine De Cock[1], and Etienne E. Kerre[1]

[1] Department of Applied Mathematics and Computer Science
Ghent University, Krijgslaan 281 (S9), B-9000 Gent, Belgium
{Steven.Schockaert, Martine.DeCock, Etienne.Kerre}@UGent.be
[2] ISLA, University of Amsterdam
Kruislaan 403, 1098 SJ Amsterdam, The Netherlands
ahn@science.uva.nl

Abstract. A temporal question answering system must be able to deduce which qualitative temporal relation holds between two events, a reasoning task that is complicated by the fact that historical events tend to have a gradual beginning and ending. In this paper, we introduce an algebra of temporal relations that is well–suited to represent the qualitative temporal information we have at our disposal. We provide a practical algorithm for deducing new temporal knowledge, and show how this can be used to answer questions that require several pieces of qualitative and quantitative temporal information to be combined. Finally, we propose a heuristic technique to cope with inconsistencies that may arise when integrating qualitative and quantitative information.

1 Introduction

Question answering systems (QA–systems) are information retrieval systems that differ from traditional search engines in two ways: users can express their information need as natural language questions, and the result of the system is an answer to a question instead of a ranked list of possibly relevant documents.

In this paper we focus on (complex) temporal questions, such as *Which battles were fought in Belgium between D-Day and the unconditional surrender of Germany*. Not only are temporal questions interesting in their own right, a thorough understanding of temporal question answering is also indispensable to answer, for example, definition questions about events or even about persons. Furthermore, we believe that a temporal QA–system can provide a first step towards causal question answering, as for example an event B can only be a consequence of A, if A happened before B.

Temporal question answering [1, 5, 8] offers a lot of interesting challenges. For some events we may be able to extract an accurate time span from, for example, the web. For other events, however, we will only be able to find qualitative temporal information (e.g., A happened before B, A happened during B,...); hence, qualitative temporal reasoning is sometimes necessary to determine if an event is relevant for a particular question. Moreover, this qualitative temporal information may conflict with some of the time spans we have at our disposal.

H.Legind Larsen et al. (Eds.): FQAS 2006, LNAI 4027, pp. 647–658, 2006.

Temporal reasoning is further complicated by the fact that many historical events are vague, i.e., their time span cannot be accurately captured by an interval with well–defined boundaries. This vagueness can be due to the fact that an event is characterized by a gradual beginning or ending (e.g., the Cold War, the Great Depression, ...). Another important cause for vagueness is that many large–scale historical events are in fact ill–defined aggregations of small–scale events. For example, World War II is a name that has been coined to refer to a number of battles and military operations around the first half of the 1940s. Some of these battles and military operations are clearly a part of World War II (e.g., the battle of the Bulge in 1944), while for others it may be hard to say whether or not this holds (e.g., the Japanese invasion of China in 1937).

In order to support efficient temporal question answering, we have (automatically) constructed a large knowledge base consisting of tens of thousands of events [1]. For some of these events, we have been able to extract an accurate time span, while for other events we only have qualitative temporal information at our disposal. To cope with vague events, we represent time spans of events as fuzzy sets, and model qualitative temporal relations using fuzzy relations. However, many temporal questions require reasoning to obtain an answer, i.e., several pieces of information, possibly coming from different sources, may have to be combined. Although there already exist some approaches to qualitative temporal reasoning that effectively deal with possibilistic uncertainty (e.g., [3, 4]), to our knowledge, the problem of qualitative temporal reasoning with vague events has not yet been considered.

In the next section, we explain how temporal information extracted from Wikipedia[1] and from the web is represented in the temporal relation algebra underlying the knowledge base of our system. It encompasses grounded events, i.e., dated events from Wikipedia and events for which we were able to construct a reliable (fuzzy) time interval, as well as ungrounded events for which we have only qualitative information at our disposal. In Section 3 we present an algebraic closure algorithm to derive new knowledge from the qualitative information in our initial knowledge base. At this point, because of space and time requirements, the available quantitative information about the grounded events is used only for inconsistency repairing. Finally, in Section 4 we explain how at question answering time both the initial and the newly derived qualitative information, as well as the quantitative information in our knowledge base are used to provide the answer.

2 Representing Temporal Information

To efficiently support temporal question answering, we have constructed (automatically) a large knowledge base by extracting relevant information from Wikipedia and from the web in general. Wikipedia is a freely available, online encyclopedia with broad coverage. It contains large lists of dated events which are relatively easy to extract. Moreover, the information in Wikipedia is much

[1] `http://www.wikipedia.org`