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,mdr}@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.

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[3] 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

---

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

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 more reliable than information on the web in general. However, for some events, only a starting date, an ending date, or an underspecified date is given. Furthermore, for most large–scale events, no structured temporal information is given at all in Wikipedia. We cope with this lack of information by searching the web for beginning dates and ending dates using a simple pattern–based approach. If there is sufficient agreement among different web pages about the beginning and ending date of an event, we represent the time span of this event as an interval; if not, we use the techniques described in [11] to construct a suitable fuzzy set, which we call a fuzzy (time) interval in this context.

To increase the coverage of our knowledge base, we again make use of a pattern–based approach. For example, to find events that happened during an event e, we may send the query "happened during e" to Google[4]. The search results returned by Google are then analysed to find noun phrases that match this pattern. Finally, heuristics are used to decide if such a noun phrase constitutes a good, umambiguous description of a unique event. For some events, it is not possible to find any starting dates or ending dates on the web. Using these patterns, we can link these events to other events in the knowledge base, using chains of before and during relations. The construction and expansion of the knowledge base is discussed in more detail in a separate paper [1].

In this section, we discuss how the temporal information in our knowledge base can be represented. After providing the necessary background on fuzzy set theory, we show how the definitions of Allen's qualitative temporal relations [2] can be generalized to cope with fuzzy time intervals, allowing to effectively calculate which qualitative temporal relations hold between two grounded events, and to what degree. Next, we show how qualitative, as well as underspecified temporal information, can be represented in this framework. Finally, we introduce an algebra of temporal relations, which will serve as the basis for qualitative reasoning in Section 3.

2.1   Grounded events

In defining a traditional set in a universe U we draw a sharp boundary between those objects of U that satisfy a certain property and those objects that do not. Therefore, classical sets are sometimes referred to as crisp sets in this context. Natural language, on the other hand, is pervaded with ill–defined concepts and properties for which such a sharp boundary may be difficult, if not impossible, to define. Fuzzy set theory [12] provides an alternative where this boundary between objects that satisfy a given property and objects that do not can be gradual. Formally, a fuzzy set A in a universe U is a mapping from U to the unit interval $[0, 1]$. For u in U, $A(u)$ is called the membership degree of u in A, where $A(u) = 1$ means u fully belongs to A and $A(u) = 0$ means u does not belong to

---

[4] http://www.google.com

A at all. If $A(u) = 1$ for some u in U, A is called normalised. A fuzzy relation in U is defined as a fuzzy set in $U \times U$.

In this paper, normalised fuzzy sets in $\mathbb{R}$ are used to represent the time span of vague events such as the Cold War, or the Great Depression. To adequately generalize the notion of a closed interval, we require that these fuzzy sets be convex and upper semicontinuous. Recall that a fuzzy set A is convex and upper semicontinuous iff for each α in $]0,1]$, the α–level set $\{x|A(x) \geq \alpha\}$ is a closed interval.

Note that, for crisp intervals $E_1 = [p^-, p^+]$ and $E_2 = [q^-, q^+]$, the following equivalence holds:

$$p^- \leq q^- \Leftrightarrow (\forall y \in \mathbb{R})(y \in E_2 \Rightarrow (\exists x \in \mathbb{R})(x \in E_1 \wedge x \leq y)) \qquad (1)$$

To express that the beginning of a fuzzy interval E_1 is before the beginning of a fuzzy interval E_2, we generalize the expression in the right hand side of (1). In particular, to generalize the logical conjunction and implication to the unit interval $[0,1]$, we use the Łukasiewicz t–norm T_W and Łukasiewicz implicator I_W, defined by:

$$T_W(a,b) = \max(0, a+b-1)$$
$$I_W(a,b) = \min(1, 1-a+b)$$

for all a and b in $[0,1]$. It can be shown that this choice of fuzzy logic connectives leads to a generalization that satisfies many important properties [10]. To generalize the universal and existential quantification, the infimum and supremum are used. Finally to generalize the ordering relation $\leq$, we use the fuzzy relation $L^{\preccurlyeq}$ in $\mathbb{R}$ defined by $L^{\preccurlyeq}(x,y) = 1$ if $x \leq y$ and $L^{\preccurlyeq}(x,y) = 0$ otherwise[5]. Thus, we obtain the following formulation of $bb^{\preccurlyeq}(E_1, E_2)$, the degree to which the beginning of the fuzzy interval E_1 is before the beginning of the fuzzy interval E_2 [9]:

$$bb^{\preccurlyeq}(E_1, E_2) = \inf_{y \in \mathbb{R}} I_W(E_2(y), \sup_{x \in \mathbb{R}} T_W(E_1(x), L^{\preccurlyeq}(x,y))) \qquad (2)$$

In the same way, we can express the degree $ee^{\preccurlyeq}(E_1, E_2)$ to which the end of E_1 is before the end of E_2, the degree $eb^{\preccurlyeq}(E_1, E_2)$ to which the end of E_1 is before the beginning of E_2, and the degree $be^{\preccurlyeq}(E_1, E_2)$ to which the beginning of E_1 is before the end of E_2 [9]:

$$ee^{\preccurlyeq}(E_1, E_2) = \inf_{x \in \mathbb{R}} I_W(E_1(x), \sup_{y \in \mathbb{R}} T_W(E_2(y), L^{\preccurlyeq}(x,y))) \qquad (3)$$

$$eb^{\preccurlyeq}(E_1, E_2) = \inf_{x \in \mathbb{R}} I_W(E_1(x), \inf_{y \in \mathbb{R}} I_W(E_2(y), L^{\preccurlyeq}(x,y))) \qquad (4)$$

$$be^{\preccurlyeq}(E_1, E_2) = \sup_{x \in \mathbb{R}} T_W(E_1(x), \sup_{y \in \mathbb{R}} T_W(E_2(y), L^{\preccurlyeq}(x,y))) \qquad (5)$$

---

[5] In [9], a more general approach is taken, where $L^{\preccurlyeq}(x,y)$ may express the degree to which x is before or at *approximately* the same time as y. This is useful to model imprecise qualitative relations such as "E_1 took place *more or less* during E_2".

Note that $be^{\preccurlyeq}$, $bb^{\preccurlyeq}$, $ee^{\preccurlyeq}$ and $eb^{\preccurlyeq}$ are fuzzy relations in the universe of fuzzy time intervals. These fuzzy relations generalize some of the constraints between boundary points that are used in Allen's algebra [2] to define qualitative relations between crisp intervals.

2.2   Qualitative relations

A lot of temporal information on the web is qualitative by nature, stating for example that event e_1 happened before e_2. Even if we have no groundings for those events, i.e., even if we can not run formulas (2)–(5), we know that the ending of e_1 is before the beginning of e_2. Hence, we can interpret this as $eb^{\preccurlyeq}(e_1, e_2) = 1$, where, for convenience, we use e_1 and e_2 both to refer to the events and to the unknown fuzzy intervals corresponding to their time spans[6]. In the same way, the fact that e_1 happened during e_2 can be interpreted as $bb^{\preccurlyeq}(e_2, e_1) = 1 \wedge ee^{\preccurlyeq}(e_1, e_2) = 1$. As another example, if e_1 and e_2 represent the life spans of two persons, $be^{\preccurlyeq}(e_1, e_2) = 1 \wedge be^{\preccurlyeq}(e_2, e_1) = 1$ expresses the fact that these persons were contemporaries.

In a similar way, we can represent underspecified information. Assume, for example, that we only know that event e began in September, 1939. In this case, e is added to the knowledge base as an ungrounded event, and a new event e' which is grounded with a fuzzy interval representing September 1939 (which will correspond to a crisp interval in this case) is also added to the knowledge base. The knowledge that e began in September 1939 can now be expressed as $bb^{\preccurlyeq}(e', e) = 1 \wedge be^{\preccurlyeq}(e, e') = 1$.

Finally, note that while we interpret, for example, "e_1 happened before e_2" as $eb^{\preccurlyeq}(e_1, e_2) = 1$, this might be too strong in some cases, i.e., it may only be that $eb^{\preccurlyeq}(e_1, e_2) \geq \delta$ (with $\delta \in]0, 1[$) holds. To make our knowledge base as informative as possible, we always use the strongest interpretations that do not lead to inconsistencies. In Section 3.2 we discuss how these initial interpretations can be weakened when the knowledge base is inconsistent.

2.3   An algebra of temporal relations

We represent temporal relations as quadruples $[\alpha, \beta, \gamma, \delta]$, where $\alpha, \beta, \gamma, \delta \in [0, 1]$, with the following interpretation:

$$e_1[\alpha, \beta, \gamma, \delta]e_2$$
$$\Leftrightarrow be^{\preccurlyeq}(e_1, e_2) \geq \alpha \wedge bb^{\preccurlyeq}(e_1, e_2) \geq \beta \wedge ee^{\preccurlyeq}(e_1, e_2) \geq \gamma \wedge eb^{\preccurlyeq}(e_1, e_2) \geq \delta$$

i.e., temporal relations are defined as crisp relations expressing a lower bound for the fuzzy relations $be^{\preccurlyeq}$, $bb^{\preccurlyeq}$, $ee^{\preccurlyeq}$ and $eb^{\preccurlyeq}$. In the following, let $\mathcal{R}$ be the set

---

[6] Allen [2] differentiates between on one hand "e_1 happened strictly before e_2", i.e., the ending of e_1 is strictly before the beginning of e_2, and on the other hand "e_1 meets e_2", i.e., the ending of e_1 coincides with the beginning of e_2. Due to the high ambiguity of natural language, such a fine–grained distinction is not useful in the context of our question answering system.

of all temporal relations, i.e., $\mathcal{R} = \{[\alpha, \beta, \gamma, \delta] | \alpha, \beta, \gamma, \delta \in [0,1]\}$. Our qualitative knowledge about the temporal relationship of two events e_1 and e_2 can then be completely described by a statement of the form:

$$e_1[\alpha_1, \beta_1, \gamma_1, \delta_1]e_2 \wedge e_2[\alpha_2, \beta_2, \gamma_2, \delta_2]e_1$$

Intersection, union, equivalence, inclusion, and strict inclusion of temporal relations are defined in the usual way, i.e., for two crisp relations R_1 and R_2 in the universe of all fuzzy time intervals $\mathcal{F}$, we have:

$$(\forall(e_1, e_2) \in \mathcal{F}^2)(e_1(R_1 \cap R_2)e_2 \Leftrightarrow e_1 R_1 e_2 \wedge e_1 R_2 e_2)$$
$$(\forall(e_1, e_2) \in \mathcal{F}^2)(e_1(R_1 \cup R_2)e_2 \Leftrightarrow e_1 R_1 e_2 \vee e_1 R_2 e_2)$$
$$R_1 = R_2 \Leftrightarrow (\forall(e_1, e_2) \in \mathcal{F}^2)(e_1 R_1 e_2 \Leftrightarrow e_1 R_2 e_2)$$
$$R_1 \subseteq R_2 \Leftrightarrow R_1 \cap R_2 = R_1$$
$$R_1 \subset R_2 \Leftrightarrow R_1 \subseteq R_2 \wedge R_2 \not\subseteq R_1$$

If $R_1 \subseteq R_2$, we say that R_1 is a stronger relation than R_2.

Lemma 1. *If E_1 and E_2 are fuzzy time intervals, it holds that:*

$$eb^{\preccurlyeq}(E_1, E_2) \leq bb^{\preccurlyeq}(E_1, E_2) \leq be^{\preccurlyeq}(E_1, E_2) \tag{6}$$
$$eb^{\preccurlyeq}(E_1, E_2) \leq ee^{\preccurlyeq}(E_1, E_2) \leq be^{\preccurlyeq}(E_1, E_2) \tag{7}$$

This lemma shows that the four components of a temporal relation are not independent of each other. For example, if we know that $e_1[0.3, 0.5, 0.1, 0.2]e_2$ holds, we also know that $e_1[0.5, 0.5, 0.1, 0.2]e_2$ and $e_1[0.5, 0.5, 0.2, 0.2]e_2$ must hold. In other words, $[0.3, 0.5, 0.1, 0.2]$, $[0.5, 0.5, 0.1, 0.2]$, and $[0.5, 0.5, 0.2, 0.2]$ denote the same temporal relation.

Lemma 2. *If E_1 and E_2 are fuzzy time intervals, it holds that:*

$$be^{\preccurlyeq}(E_1, E_2) = 1 \vee eb^{\preccurlyeq}(E_1, E_2) = 0$$

Hence, we also have that $[0.3, 0.5, 0.1, 0.2]$, and $[1, 0.5, 0.1, 0.2]$ denote the same temporal relation. Therefore, we introduce an operator *norm* that transforms a temporal relation into a canonical form:

$$norm([\alpha, \beta, \gamma, \delta]) = \begin{cases} [1, \max(\beta, \delta), \max(\gamma, \delta), \delta] & \text{if } \delta > 0 \\ [\max(\alpha, \beta, \gamma), \beta, \gamma, 0] & \text{otherwise} \end{cases}$$

If $norm(R) = R$, then R is called normalised. Obviously $norm(R)$ is normalised for every temporal relation R. We can prove the following proposition.

Proposition 1. *If $R = [\alpha, \beta, \gamma, \delta]$ is a normalised temporal relation, there always exist fuzzy intervals A and B such that:*

$$be^{\preccurlyeq}(A, B) = \alpha \wedge bb^{\preccurlyeq}(A, B) = \beta \wedge ee^{\preccurlyeq}(A, B) = \gamma \wedge eb^{\preccurlyeq}(A, B) = \delta$$

Corollary 1. *Let $R_1 = [\alpha_1, \beta_1, \gamma_1, \delta_1]$ and $R_2 = [\alpha_2, \beta_2, \gamma_2, \delta_2]$ be normalised temporal relations. It holds that:*

$$R_1 = R_2 \Leftrightarrow \alpha_1 = \alpha_2 \wedge \beta_1 = \beta_2 \wedge \gamma_1 = \gamma_2 \wedge \delta_1 = \delta_2$$

This corollary expresses that the operator *norm* indeed transforms temporal relations into a canonical form; hence, our normalization operator is well–defined. It is easy to see that if $R_1 = [\alpha_1, \beta_1, \gamma_1, \delta_1]$ and $R_2 = [\alpha_2, \beta_2, \gamma_2, \delta_2]$, it holds that:

$$R_1 \cap R_2 = [\max(\alpha_1, \alpha_2), \max(\beta_1, \beta_2), \max(\gamma_1, \gamma_2), \max(\delta_1, \delta_2)] \qquad (8)$$

Moreover, if R_1 and R_2 are normalised, we have that:

$$R_1 \subseteq R_2 \Leftrightarrow \alpha_1 \geq \alpha_2 \wedge \beta_1 \geq \beta_2 \wedge \gamma_1 \geq \gamma_2 \wedge \delta_1 \geq \delta_2 \qquad (9)$$

3   Reasoning with imperfect temporal information

3.1   Deduction algorithm

The composition of R_1 and R_2 is defined as:

$$e_1(R_1 \circ R_2)e_2 \Leftrightarrow (\exists e \in \mathcal{F})(e_1 R_1 e \wedge e R_2 e_2) \qquad (10)$$

i.e., $R_1 \circ R_2$ is the strongest relation that is assured to hold between e_1 and e_2 using only the knowledge that $e_1 R_1 e$ and $e R_2 e_2$ for some fuzzy time interval e. As in many algebras for qualitative temporal or spatial reasoning, our algebra is not closed under composition. Therefore, we use the notion of weak composition [7], which is defined for relations R_1 and R_2 as the strongest relation $R_1 \diamond R_2$ in $\mathcal{R}$ that contains $R_1 \circ R_2$, i.e.:

$$R_1 \circ R_2 \subseteq R_1 \diamond R_2 \wedge (\forall R \in \mathcal{R})(R_1 \circ R_2 \subseteq R \Rightarrow R_1 \diamond R_2 \subseteq R) \qquad (11)$$

i.e., $R_1 \diamond R_2$ is the strongest relation in $\mathcal{R}$ that is assured to hold between e_1 and e_2 using only the knowledge that $e_1 R_1 e$ and $e R_2 e_2$ for some fuzzy time interval e. Intuitively, the weak composition operator can be seen as an optimal operator for propagating temporal knowledge in a particular algebra. Unfortunately, (11) does not tell us how to compute the weak composition of two relations in practice. The following important proposition gives us a practical characterization of weak composition of temporal relations.

Proposition 2. *Let $R_1 = [\alpha_1, \beta_1, \gamma_1, \delta_1]$ and $R_2 = [\alpha_2, \beta_2, \gamma_2, \delta_2]$ be normalised temporal relations. It holds that:*

$$
\begin{aligned}
R_1 \diamond R_2 = [\alpha', \\
& \max(T_W(\beta_1, \beta_2), \min(\alpha_1 + T_W(\delta_2, \gamma_1), \delta_2, \beta_1 + T_W(\delta_2, \alpha_1))), \\
& \max(T_W(\gamma_1, \gamma_2), \min(\alpha_2 + T_W(\delta_1, \beta_2), \delta_1, \gamma_2 + T_W(\delta_1, \alpha_2))), \\
& \max(T_W(\delta_1, \beta_2), T_W(\gamma_1, \delta_2), \min(\delta_1, \delta_2))]
\end{aligned}
$$

where

$$\alpha' = \begin{cases} \max(T_W(\beta_1,\alpha_2), T_W(\alpha_1,\gamma_2)) & \textit{if } T_W(\alpha_1,\delta_2) = 0 \textit{ and } T_W(\delta_1,\alpha_2) = 0 \\ \min(\alpha_1,\alpha_2) & \textit{otherwise} \end{cases}$$

Let the qualitative temporal information we have at our disposal be represented as temporal relations $R_{ij} = [\alpha_{ij}, \beta_{ij}, \gamma_{ij}, \delta_{ij}]$ such that $e_i R_{ij} e_j$ holds for all $1 \le i, j \le n$, $i \ne j$. We apply the weak composition to derive as much new information as possible from the qualitative information in our knowledge base, which was initially gathered as described in Section 2.2.

Example 1. Let the temporal relations between e_1, e_2 and e_3 be given by $R_{12} = [0.8, 0.5, 0.3, 0]$, $R_{23} = [1, 0.6, 0.9, 0]$, and $R_{13} = [0.6, 0.3, 0.3, 0]$. Using Proposition 2 we can deduce that also $R_{12} \diamond R_{23} = [0.7, 0.1, 0.2, 0]$ holds between e_1 and e_3. Thus we obtain that R_{13} should be replaced by the stronger temporal relation $[0.6, 0.3, 0.3, 0] \cap [0.7, 0.1, 0.2, 0] = [0.7, 0.3, 0.3, 0]$.

To deduce stronger temporal relations in the way of Example 1, we use Algorithm 1, which is similar in spirit to Allen's path–consistency algorithm [2]. It can be shown that this algorithm has a worst–case complexity of $O(n^4)$. We stress once more that only temporal relations based on qualitative information gathered as described in Section 2.2 serve as input to Algorithm 1. Hence, newly derived temporal relations are also based solely on qualitative information. However, we do check if these new relations are consistent with the known groundings, i.e., with the available quantitative information, and only in this case they are used in further deductions.

It can be shown that the set of qualitative temporal relations in our knowledge base is always consistent, i.e., for every event e_i we can always find a fuzzy interval E_i such that all the relations R_{jk} are satisfied $(1 \le j, k \le n)$. In other words, inconsistencies can only arise when at least some of the events are grounded. In the following discussion, let $\mathcal{E}$ be the set of events and $\mathcal{G}$ the set of grounded events in our knowledge base $(\mathcal{G} \subseteq \mathcal{E})$. If, for example, $e_1 \in \mathcal{G}$ and $e_2 \in \mathcal{G}$ then we know the exact degree $be^{\prec}(E_1, E_2)$ to which the beginning of e_1 is before the end of e_2. On the other hand, suppose that we have deduced, using the available qualitative information, that α_{12} is a lower bound for this degree. Hence, if $be^{\prec}(E_1, E_2) < \alpha_{12}$, the lower bound α_{12} is too strong, and the temporal relation R_{12} is inconsistent with the groundings of e_1 and e_2. In general, we define the predicate *consistent* for e_1 and e_2 in $\mathcal{G}$ as:

$$consistent(e_1, e_2) \Leftrightarrow \alpha_{12} \le be^{\prec}(E_1, E_2) \vee \beta_{12} \le bb^{\prec}(E_1, E_2)$$
$$\vee\, \gamma_{12} \le ee^{\prec}(E_1, E_2) \vee \delta_{12} \le eb^{\prec}(E_1, E_2)$$

In other words, $consistent(e_1, e_2)$ expresses that the available qualitative information about events e_1 and e_2, namely $R_{12} = [\alpha_{12}, \beta_{12}, \gamma_{12}, \delta_{12}]$, does not contradict the values that can be computed from quantitative information, i.e., from the groundings E_1 and E_2. For $e_1 \in \mathcal{E} \setminus \mathcal{G}$ or $e_2 \in \mathcal{E} \setminus \mathcal{G}$, $consistent(e_1, e_2)$ always holds.

Algorithm 1 Algebraic closure algorithm

1: $todo \leftarrow \{(i,j,k)|1 \le i,j,k \le n, i \ne j \ne k\}$
2: **while** $todo \ne \emptyset$ **do**
3:     select and remove a triplet (i_0, j_0, k_0) from $todo$
4:     **if** $consistent(e_{i_0}, e_{j_0})$ and $consistent(e_{j_0}, e_{k_0})$ **then**
5:         $R \leftarrow R_{i_0 j_0} \diamond R_{j_0 k_0}$
6:         **if** $R \cap R_{i_0 k_0} \subset R_{i_0 k_0}$ **then**
7:             $R_{i_0 k_0} \leftarrow R \cap R_{i_0 k_0}$
8:             $todo \leftarrow todo \cup \{(i_0, k_0, j)|i_0 \ne j \ne k_0\} \cup \{(j, i_0, k_0)|i_0 \ne j \ne k_0\}$
9:         **end if**
10:    **end if**
11: **end while**

3.2 Inconsistency repairing

During the execution of Algorithm 1, inconsistent relations may be deduced. In the following, we make the assumption that all groundings in the knowledge base are correct. This is a reasonably safe assumption because these groundings are extracted either from Wikipedia, or from a large number of webpages. Under this assumption, whenever a temporal relation is deduced that is inconsistent with the groundings in the knowledge base, we know that one or more temporal relations in the original knowledge base are too strong. This may either be because these qualitative relations were obtained by using false information, e.g., some website states that e_1 happened before e_2, while in reality e_1 happened after e_2, because of errors introduced in the information extraction phase, or because our interpretation of the natural language fragment expressing the relation is too strong, e.g., we assume that $eb^{\preccurlyeq}(e_1, e_2) = 1$ holds, while in reality we only have that $eb^{\preccurlyeq}(e_1, e_2) = 0.8$. To make an inconsistent knowledge base consistent, we revise some of these interpretations, and apply Algorithm 1 a second time. This process is repeated until all inconsistencies have been eliminated.

    Let $R_{ij}^{(0)} = [\alpha_{ij}^{(0)}, \beta_{ij}^{(0)}, \gamma_{ij}^{(0)}, \delta_{ij}^{(0)}]$ be the qualitative relation between e_i and e_j in the initial knowledge base, i.e., before the execution of Algorithm 1, and let $R_{ij}^{(1)} = [\alpha_{ij}^{(1)}, \beta_{ij}^{(1)}, \gamma_{ij}^{(1)}, \delta_{ij}^{(1)}]$ be the relation that results after Algorithm 1 is completed. Assume, for example, that $e_1 \in \mathcal{G}$, $e_2 \in \mathcal{G}$, and $\alpha_{12}^{(1)}$ leads to an inconsistency, i.e., $\alpha_{12}^{(1)} > be^{\preccurlyeq}(E_1, E_2)$. During the execution of Algorithm 1, we keep track of the components of the relations $R_{ij}^{(0)}$ that are used to obtain each conclusion. In particular, we have a set V at our disposal containing those values that were used to obtain the value $\alpha_{12}^{(1)}$; assume, for example, $V = \{\alpha_{pq}^{(0)}, \delta_{qr}^{(0)}, \beta_{rs}^{(0)}, \alpha_{st}^{(0)}\}$. Most of the values v in V will also be used in other deduction chains, leading to other conclusions. Some of these conclusions will define temporal relations between ungrounded events; hence, we cannot tell whether they are correct or not. However, other conclusions will define temporal relations between grounded events; we will call these verifiable conclusions. For v in V, let $pos(v)$ be the number of correct verifiable conclusions that were obtained using v, and $neg(v)$ the number of incorrect verifiable conclusions that were obtained

using v; then, $rel(v) = \frac{pos(v)}{pos(v)+neg(v)}$ can be used as an approximation of the reliability of v (provided $pos(v) + neg(v) > 0$). The knowledge base is revised by subtracting a value from each v in V that is proportional to $1 - rel(v)$:

$$v \leftarrow v - (\alpha_{12}^{(1)} - be^{\preccurlyeq}(E_1, E_2))\frac{1 - rel(v)}{\sum_{v' \in V} 1 - rel(v')}$$

In practice, it may occur that different inconsistencies in the knowledge base require the value v to be updated in different ways. In this case, the final value for v is the minimum of all these different updates.

Finally, Algorithm 1 is executed a second time using this revised knowledge base as a starting point. If there are still inconsistencies detected, this whole process is repeated. In practice, we found that all inconsistencies are typically eliminated after one or two iterations.

4   Question answering

The temporal questions in which we are primarily interested consist of a non–temporal main part asking for an event (e.g., *Which battles took place in Belgium*), the participant of an event (e.g., *Which countries were involved in World War II*), or a time–dependent factoid (e.g., *Who was president of the U.S.*), and a temporal restriction (e.g., *after the invasion of Normandy*). The question analysis module separates the temporal restriction from the main part using a large set of handcrafted patterns. The temporal restriction is further analysed to identify which temporal relation is expressed in it. We use standard question answering techniques to determine, from the main part of the question, the expected semantic type of the answer (e.g., person, battle,...), the question type, i.e., whether the question is asking for an event, a participant, or a time–dependent factoid, and some additional information depending on the question type (see e.g., [6]).

Standard techniques are also used to find events in the knowledge base that satisfy the non–temporal main part of the question. Checking which of these events satisfy the temporal restriction in the question typically boils down to inferring if, and to what degree, a particular qualitative temporal relation holds between these events and the events occurring in the temporal restriction of the question. To support this reasoning task efficiently, Algorithm 1 and the techniques described in Section 3.2 are applied a priori. Hence, at question–answering time R_{ij} is the strongest qualitative relation that we can infer from the available qualitative information, i.e., using groundings of events only to detect inconsistencies. However if e_1 and e_2 are grounded events, with corresponding groundings E_1 and E_2, then $(R_{i1} \diamond R_{12}^*) \diamond R_{2j}$ might be a stronger conclusion, where

$$R_{12}^* = [be^{\preccurlyeq}(E_1, E_2), bb^{\preccurlyeq}(E_1, E_2), ee^{\preccurlyeq}(E_1, E_2), eb^{\preccurlyeq}(E_1, E_2)] \tag{12}$$

In other words, we might obtain a stronger conclusion by also considering the qualitative relation R_{12}^* between e_1 and e_2, computed directly from the groundings of these events. Hence, to obtain a maximal amount of qualitative temporal

information between e_i and e_j, we should calculate:

$$R = R_{ij} \cap \left(\bigcap_{(e_1, e_2) \in \mathcal{G}^2} (R_{i1} \diamond R_{12}^*) \diamond R_{2j} \right)$$

In practice, various techniques can be used to evaluate this efficiently; we omit the details here. Note that, because of space and time requirements, it would not be feasible to add the qualitative relation R_{12}^* to the knowledge base for every pair (e_1, e_2) of grounded events a priori, before the execution of Algorithm 1. Finally, if the knowledge base contains too little information to answer a particular question, we use the web to expand it with relevant information at question answering time.

Example 2. Consider the question *Which battles took place in Asia after V-J Day*, and the knowledge base, after the execution of Algorithm 1, illustrated in Figure 1(a). Assume that only *World War II* and *the Cold War* are grounded. Intuitively, the qualitative information in this knowledge base expresses that *the Battle of Chosin Reservoir* happened during *the Korean War* which happened during *the Cold War*, and that *V–J Day* happened during *World War II*. Assume that, using external knowledge, we were able to establish that *the Battle of Chosin Reservoir* is a battle that took place in Asia. As becomes clear from Figure 1(a), it is not possible to infer that this battle happened after *V–J Day* using qualitative information alone. Using the groundings of *World War II* and *the Cold War* however, we can calculate the strongest temporal relation that holds between these two events. This is illustrated in Figure 1(b). Thus we obtain that *V–J Day* and *the Battle of Chosin Reservoir* satisfy the temporal relation given by

$$([1, 0, 1, 0] \diamond [1, 0.9, 1, 0.7]) \diamond [1, 1, 0, 0] = [1, 0.7, 1, 0.7] \diamond [1, 1, 0, 0]$$
$$= [1, 0.7, 0.7, 0.7]$$

Hence, according to our knowledge base, *the Battle of Chosin Reservoir* was after *V–J Day* at least to degree 0.7.

5   Conclusions

We have discussed the problem of reasoning with vague and qualitative temporal information in the context of question answering. We introduced an algebra of temporal relations to express qualitative temporal information between vague events, and provided a practical characterization of weak composition. Furthermore, we introduced an algorithm to compute the algebraic closure of an initial specification, as well as a heuristic technique to eliminate inconsistencies in the knowledge base. Finally, we have shown how qualitative and quantitative temporal information can be combined to effectively support the question answering process.

(a) Using only qualitative information

(b) Using qualitative information and groundings
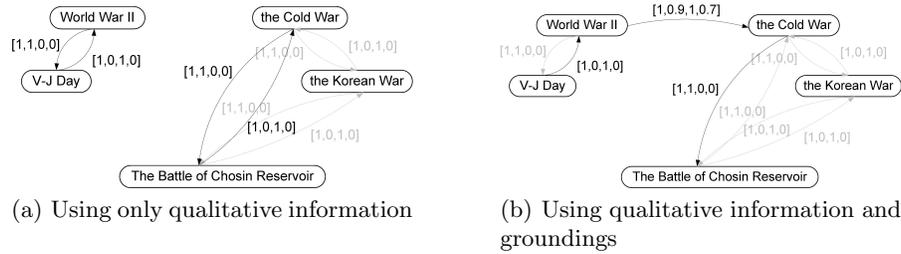
Fig. 1. Verifying whether *the Battle of Chosin Reservoir* happened after *V–J Day*.

Acknowledgments

References

1. Ahn, D., Schockaert, S., De Cock, M., Kerre, E.E.: Supporting Temporal Question Answering: Strategies for Offline Data Collection. International workshop on Inference in Computational Semantics, to appear.
2. Allen, J.F.: Maintaining Knowledge about Temporal Intervals. Communications of the ACM **26** (1983) 832–843
3. Badaloni, S., Giacomin, M.: A Fuzzy Extension of Allen's Interval Algebra. Lecture Notes in Artificial Intelligence **1792** (2000) 155–165
4. Dubois, D., HadjAli, A., Prade, H.: Fuzziness and Uncertainty in Temporal Reasoning. Journal of Universal Computer Science **9** (2003) 1168–1194
5. Harabagiu, S., Bejan, C.A.: Question Answering based on Temporal Inference. AAAI-2005 Workshop on Inference for Textual Question Answering (2005)
6. Jijkoun, V., Tjong Kim Sang, E., Ahn, D., Müller, K., de Rijke, M.: The University of Amsterdam at QA@CLEF 2005. Working Notes for the CLEF 2005 Workshop (2005)
7. Renz, J., Ligozat, G.: Weak Composition for Qualitative Spatial and Temporal Reasoning. Proceedings of the Eleventh International Conference on Principles and Practice of Constraint Programming (2005) 534–548
8. Saquete, E., Martínez–Barco, P., Muñozn, R., Vicedo, J.: Splitting Complex Temporal Questions for Question Answering Systems. ACL04 (2004) 566–573
9. Schockaert, S., De Cock, M., Kerre, E.E.: Imprecise Temporal Interval Relations. Lecture Notes in Computer Science **3849** (2006) 108–113
10. Schockaert, S., De Cock, M., Kerre, E.E.: Fuzzifying Allen's Temporal Interval Relations. Submitted
11. Schockaert, S.: Construction of Membership Functions for Fuzzy Time Periods. Proceedings of the ESSLLI 2005 Student Session (2005) 297–305
12. Zadeh, L.A.: Fuzzy sets. Information and Control **8** (1965) 338–353