Reasoning with temporal context in news analysis

Igor Mozetič and **Damjan Bojadžiev**[1]

Abstract. One aspect of implicit, contextual information is its temporal component. Explicating this component in a formal model makes it possible to disambiguate some context-dependent expressions and discover connections between expressions. We have implemented and extended Allen's algebra of temporal intervals in a reasoner that takes into account the linear nature of time and the granularity of temporal expressions (days/weeks/...). If this algebra is used to model the temporal extension of events, the reasoner can track and connect the reference of indexical expressions about them. We intend to use the reasoner for analysing news streams, to help discover connections between news items.

1 INTRODUCTION

This paper expresses the combination of our interests in the subjects of context and ontologies, taken by themselves and in their connections. On the more abstract, logico-philosophical side, there are the questions of definition and significance: what *is* context, what is specific about it, and how does it, its inclusion or its omission, affect cognitive, deductive and computational processes. For example, getting stuck into loops, for humans and for machines, might be conceived as a loss of context. Judging by the many definitions of context in different disciplines, the notion of context is itself context-sensitive, and it is hard to point out the specific characteristic that distinguishes context from background, prior knowledge and/or the multiplicity of implicit facts and assumptions that is simply taken for granted, unnoticed, left out or suppressed as too obvious to mention. This is reflected in the reluctance in some important papers on context to actually define it, such as McCarthy's [8], and in his insistence that "there is no universal context". In connection with ontologies, there is also some context-dependence in the definition of context: ontologies supply context for browsing [5] (which again indicates that context can be practically anything), but mappings between ontologies supply context too, as in C-OWL [6].

On the more computational side, our interest is in ontologies of time, or of the ways we refer to its passage, and in actual implementations of automatic reasoning about temporal information. We have implemented Allen's axiomatization of temporal relations, used eg. in DAML-Time and SUMO [10], in the constraint logic programming system CLP($\mathcal{Q}$) [9]. We plan to use this implementation in automatic news (stream) analysis, for disambiguating context-dependent reference and for news classification. The remainder of this paper gives more information about the axiomatization and its implementation, and some examples of the intended application. More generally, we have a hunch that some of the work done at our Department, eg. on user profiling and on simultaneous ontologies [7], can be formulated as programming context dependency, and we are working on a convincing formulation.

[1] Jozef Stefan Institute, Ljubljana, Slovenia, email: igor.mozetic@ijs.si

In our news analysis system, we are mainly concerned with the temporal aspects of context. The system will take into account the time-stamp part of the metadata about news items, and temporal models of the events reported, to distinguish related news items from unrelated ones. Thus, our working hypothesis about temporal context could be expressed in the equation

temporal context(news) = temporal model + metadata(news)

This definition was originally inspired by [11], which deals with contextual vocabulary acquisition (how to infer the meaning of a new word from textual clues). It identifies two components of context: prior knowledge (which is subject to belief revision) and co-text of the word to be learned. In our case, the task is to find a semantic link between news items. The context of news is prior knowledge in the form of a temporal model, and the metadata that comes with the news. We do not deal with the model revision component, and restrict our system to temporal aspects. However, causal, spatial and other types of models and/or ontologies also represent prior knowledge and thus fit into our definition of context. If the restrictions to temporality and subject matter (news items) are dropped, the equation above generalizes to the form

context(X) = prior knowledge + co-data(X)

2 TEMPORAL ALGEBRA AND ONTOLOGY

Allen [4] proposed an interval algebra to represent relative temporal information, such as the order of events. The representation of events by time intervals rather than points allows the expression of hierarchical, indefinite and incomplete information, at different levels of granularity. The temporal algebra uses the thirteen possible relations between time intervals, such as one interval starting or finishing another interval, or being before or meeting another one.

To represent indefinite and incomplete information, Allen uses disjunction to allow any subset of the basic relations to hold between two time intervals. A set of temporally related events forms a network, with edges corresponding to (possibly disjunctive) relations between events. There are two fundamental queries one can pose about such a network:

- Find the feasible relations between all pairs of events, and
- Determine the consistency of the temporal relations.

When we came accross this algebra, we were not aware of any (complete) reasoner for it. Since its networks of temporal relations express constraints on relations between intervals, we decided to implement the algebra in a constraint logic programming system CLP($\mathcal{Q}$) [9]. The implementation allows automatic reasoning about temporal events, such as:

- "If X precedes Y, and Y overlaps with Z, what are the possible temporal relations between X and Z ?"
- "If X takes longer then Y, can X occur during Y ?"
- "Given a set of temporally related events, what are the possible consistent scenarios on the time line ?"

On top of this basic implementation, we formulated a generic ontology of time which covers everyday concepts such as hours, days, seasons, and the relations between them. Note that there is no fixed underlying time scale. This time ontology is similar to the specifications in SUMO [3], DOLCE [2], and DAML-Time [1], and has the advantage of being executable.

3 NEWS ANALYSIS

Let us first illustrate the desired feature of the analysis system by a simple example. Suppose we receive two news items on two subsequent days:

- **Day1**: "Giant waves hit the shore early today."
- **Day2**: "An ocean floor earthquake was detected yesterday."

One interesting question that a news analyst might then ask is: Are these news items related?

There are various techniques used for news analysis, but all essentially measure the degree of similarity between items. The metrics used can be purely syntactic or increasingly based on semantics. We might roughly distinguish three levels of (semantic) similarity:

1. purely lexical, based only on the presence of keywords
2. weak or lexicographic, taking into account taxonomic meaning
3. strong, using models of word referents

The models in question are formulated in terms of the temporal ontology; in the case of the news items above, a relevant example would be the temporal model of a tsunami, shown in Figure 1.

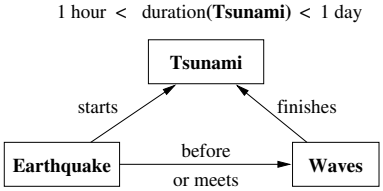

Figure 1. A simple temporal model of a tsunami.

To detect whether the news items above are related or not, we would use the following algorithm:

1) When the first news (waves) arrive, find the temporal terms ("today") and resolve them locally, with respect to the news metadata (time-stamp). The implicit temporal reference can then be transformed into explicit reference in terms of the temporal ontology, resulting in the temporal relation

Waves during **Day1**

2) When the next news arrive (earthquake), the procedure gives

Earthquake during previous(**Day2**)

Here, the reference "yesterday" is expressed by applying the function "previous" to the current **Day2**.

3) Reasoning with the temporal ontology gives (Figure 2):

previous(**Day2**) equals **Day1**

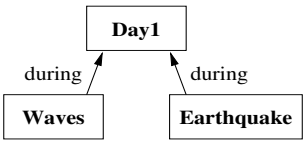

Figure 2. Temporal relations between both news events.

4) Reasoning with the tsunami model then shows that the news are consistent with a tsunami. Therefore, we can formulate a defeasible hypothesis: A tsunami is a possible explanation of the two events, which links the news items in question.

5) However, if the news say that

Waves before **Earthquake**

the tsunami link will be ruled out as a possible explanation of the news sequence.

In this way, the temporal model can provide a stronger measure of semantic similarity and thus increase the quality of the news analysis system.

4 CONCLUSION

Our definition of temporal context seems useful, especially for a news analysis system, because it encompasses both static prior knowledge and dynamic metadata (a sophisticated example of the use of such data in reasoning with abductive constraint logic programs is presented in [12]). If experiments with the news analysis system, augmented by the temporal ontology, the constraint logic program and temporal models such as the tsunami model, prove successful, other semantic models, such as causal and spatial, will be included too. In the tsunami example, these would be needed to capture other relevant relations, such as the fact that the earthquake needs to take place under the see in roughly the same geographic area.

ACKNOWLEDGEMENTS

Supported by the Slovenian Research Agency, and the EU Projects SEKT (IST-2003-506826) and NeOn (IST-2004-27595).

REFERENCES

[1] Daml-time. http://www.cs.rochester.edu/ferguson/daml/.
[2] Dolce. http://www.loa-cnr.it/DOLCE.html.
[3] Sumo. http://www.ontologyportal.org/.
[4] J.F. Allen, 'Maintaining knowledge about temporal intervals', *Communications of the ACM*, **26**, 832–843, (1983).
[5] N. Aussenac-Gilles and J. Mothe, 'Ontologies as background knowledge to explore document collections', in *Proc. RIAO-2004*, pp. 129–142, Avignon, (2004).
[6] P. Bouquet, F. Giunchiglia, F. van Harmelen, L. Serafini, and H. Stuckenschmidt, 'C-owl: Contextualizing ontologies', in *Proc. 2nd ISWC*, pp. 164–179, (2003).
[7] Grobelnik M. Fortuna, B. and D. Mladenič, 'Background knowledge for ontology construction', in *Proc. Contexts and Ontologies*, (2006).
[8] J. McCarthy, 'Notes on formalizing context', in *Proc. 13th IJCAI*, (1993). http://www-formal.stanford.edu/jmc/context3/context3.html.
[9] I. Mozetič, 'Temporal interval reasoning with clp(q)', in *Proc. 8th IS-2005*, pp. 178–181, Ljubljana, (2005).
[10] I. Niles and A. Pease, 'Towards a standard upper ontology', in *Proc. 2nd FOIS-2001*, Ogunquit, Maine, (2001).
[11] W.J. Rapaport, 'What is the 'context' for contextual vocabulary acquisition?', in *Proc. 4th ICCS/ASCS-2003*, pp. 547–552, Sydney.
[12] Madnick S. Zhu, Hongwei and M. Siegel, 'Reasoning about temporal context using ontology and abductive constraint logic programming', Technical Report 15, CISL, MIT, Cambridge, (2004).