KNOWLEDGE ENGINEERING: PRINCIPLES AND TECHNIQUES, KEPT 2009
International Conference on Knowledge Engineering Principles and Techniques
Selected Papers, Cluj-Napoca (Romania), July 2–4, 2009, pp. 61–68

IDENTIFICATION OF TEMPORAL EXPRESSIONS IN THE DOMAIN OF TOURISM

ANDREA VARGA[(1)], GEORGIANA PUŞCAŞU[(2)], AND CONSTANTIN ORĂSAN[(3)]

Abstract. This paper presents how an existing temporal processor was adapted to be used by the English Question Answering system developed part of the EU-funded project QALL-ME. Experiments applying the existing temporal processor to questions from the domain of tourism revealed that the existing temporal processor tackles far too many temporal expressions, and this makes it slower than necessary. In light of this, a simplified temporal processor which identifies only temporal expressions present in user questions was implemented. The two temporal annotators are evaluated on 1,118 randomly selected user questions and an error analysis is presented.

1. Introduction

QALL-ME (Question Answering Learning technologies in a multiLingual and Multimodal Environment)[1] is an EU-funded project that aims to develop a shared infrastructure for multilingual and multimodal question answering in the domain of tourism. The purpose of this system is to be able to answer questions about local events such as movie showtimes, directions to sites (e.g. cinemas, hotels) from structured data sources, such as databases. Investigation of the questions users normally ask in this context revealed that a large number of them contain temporal constraints. This paper presents the temporal annotator employed to process English questions and its evaluation.

Our temporal tagger follows the design and methodology of the temporal tagger described in [6] that is capable of identifying both self-contained temporal expressions (TEs), which get tagged with their value, and indexical/under-specified TEs, which, depending on their semantics, receive a value computed by a temporal function having as argument the time they are relative to. Investigation of the temporal expressions used in this domain revealed that the temporal tagger

Key words and phrases. Temporal Expressions; Question Answering; Temporal Annotator.
This work is supported by the EU-funded project QALL-ME (FP6 IST-033860).
[1]The project's webpage is http://qallme.fbk.eu/

proposed by Puscasu (2004) tackles more cases than necessary making it too slow for our purposes. For this reason a simplified temporal annotator was developed.

The remainder of the paper is structured as following: Section 2 presents related work. Section 3 is focused on the types of temporal expressions one can encounter in natural language texts. An investigation of the temporal expressions that appear in the QALL-ME user questions is presented in Section 4. The temporal annotator used in QALL-ME is presented in Section 5 followed by its evaluation in Section 6.

2. Related work

Many research efforts have focused on the identification and normalisation of temporal expressions, such as [1, 2, 3, 4, 5, 6, 7, 8].

The importance of the proper treatment of TEs is reflected by the relatively large number of NLP evaluation efforts centered on their identification and normalisation, such as the MUC 6 and 7 Named Entity Recognition tasks [9], the ACE-2004 Event Recognition task [10], the Temporal Expression Recognition and Normalisation [11] task.

Machine Learning approaches have been found to work well in detecting the boundaries of the temporal expressions [7, 8], but they are outperformed by rule-based ones at the stage of extracting the TE's temporal meaning.

3. Types of temporal expressions

Temporal expressions (TEs) are natural language phrases that refer directly to time. They are normally signalled by one or more time words called lexical triggers. The terms that indicate a temporal expression can be: nouns (e.g. *day*, *month*, *year*), proper names (e.g. *Friday*, *June*, *Christmas*), time patterns (e.g. *10:15*, *26/03/2009*), adjectives (e.g. *current*, *past*), adverbs (e.g. *weekly*, *then*), time nouns/adverbs (e.g. *today*, *tomorrow*, *now*). The extent of a temporal expression should be either a noun, adjective, adverb or any of the corresponding phrases (noun, adjectival or adverbial phrases).

TIMEX2 [12] is the worldwide adopted standard for the annotation of temporal expressions in text, due to its coverage and level of detail. The QALL-ME consortium has also adopted the TIMEX2 standard for the purposes of temporal expression annotation in QALL-ME. Therefore, each TE present in the user questions is supposed to be annotated with the TIMEX2 tag that captures the meaning of the TE.

TEs normally denote position in time, duration or time frequency. The following classes and subclasses of TEs can be distinguished:

1. **TEs indicating time position**
   - 1.1 **Precise TEs:** are the ones for which one can confidently determine their position on the time axis.

- 1.1.1 **Calendar dates:** include TEs denoting specific years, months, dates, decades, centuries, millennia (e.g. <TIMEX2 VAL="2009-07-02">**2nd of July**</TIMEX2>).
- 1.1.2 **Times of day:** such expressions can specify times of the day at the level of hour, minute, second or millisecond (e.g. <TIMEX2 VAL="2009-07-02T10:00">**10 o'clock**</TIMEX2>, <TIMEX2 VAL="2009-07-02T07:53">**7:53 am**</TIMEX2>). They can include references to the time zone where that specific time of day is applicable to.
- 1.1.3 **Week references:** These expressions refer to periods of time having the granularity at the week level (e.g. <TIMEX2 VAL="2009-W27">**next week**</TIMEX2>).

1.2 **Fuzzy TEs:** are vague or have imprecise boundaries.

- 1.2.1 **Generic references to the past, present or future:** TEs that refer in general terms to the past, present or future (e.g. <TIMEX2 VAL="PRESENT_REF" ANCHOR_DIR="AS_OF" ANCHOR_VAL ="2009-07-02">**now**</TIMEX2>).
- 1.2.2 **Seasons, parts of the year (quarters and halves):** denote certain parts of a year (e.g. <TIMEX2 VAL="2009-SU">**this summer**</TIMEX2>).
- 1.2.3 **Weekends:** TEs referring to weekends (e.g. <TIMEX2 VAL="2009-W26-WE">**this weekend**</TIMEX2>).
- 1.2.4 **Fuzzy day parts:** denote parts of the day (e.g. <TIMEX2 VAL= "2009-07-02TAF">**afternoon**</TIMEX2>).

1.3 **Non-specific TEs referring to time position:** are time position TEs mentioned in generic contexts (e.g. <TIMEX2 VAL="XXXX-SU">**summers**</TIMEX2>).

2. **TEs capturing durations:** indicate periods of time by specifying how long something lasted.
   - 2.1 **Precise durations:** specify exactly how long something lasted (e.g. <TIMEX2 ANCHOR_DIR="WITHIN" ANCHOR_VAL="2009-07-02" VAL="PT24H">**24 hours**</TIMEX2>).
   - 2.2 **Fuzzy durations:** denote an unspecified number of temporal units included in a period of time (e.g. <TIMEX2 ANCHOR_DIR="BEFORE" ANCHOR_VAL="2009-W26" VAL="PXW">**preceding weeks**</TIMEX2>).
   - 2.3 **Non-specific durations:** are durations occurring in sentences that state generalisations (e.g. <TIMEX2 VAL="P1D">**all day**</TIMEX2>).
3. **Set-denoting time expressions:** give information about the frequency of a certain event.

3.1 **Precise frequency TEs:** indicate sets of times telling precisely how often something happens (e.g. <TIMEX2 VAL="XXXX-XX-XX" SET="YES"> **every day**</TIMEX2>).

3.2 **Non-specific frequencies:** are set-denoting TEs used in generic contexts (e.g. <TIMEX2 VAL="TNI" SET="YES">**some nights**</TIMEX2>).

4. Temporal expressions in the QALL-ME user questions

The corpus used for experiments in this work is the QALL-ME benchmark - the set of user questions collected in the context of the QALL-ME project. The QALL-ME benchmark is a collection of several thousand spoken questions in the four languages involved in the project: Italian, English, Spanish and German [13]. The benchmark was created for two purposes: to allow development of applications based on machine-learning for QA and to enable testing their performance in a controlled laboratory setting. To date, the benchmark contains 15,479 questions related to cultural events and tourism, such as accommodation, gastro, cinemas, movies, exhibitions, etc., associated with the relevant information necessary to train and test the core components of a QA system. In this paper only the English part of the QALL-ME benchmark was used with a total of 4,501 questions.

In order to identify the most frequent types of TEs encountered in the user questions, and with a view towards evaluating a temporal expression identifier on questions in the domain of tourism, a set of 1,118 randomly selected user questions from the QALL-ME benchmark has been manually annotated according to the TIMEX2 standard.

On the basis of benchmark investigation it was noticed that QALL-ME user questions do not contain the full range of possibly existing TEs defined in the TIMEX2 annotation guidelines. Therefore, not all the classes of TEs need to be tackled by a real-time QA system aiming at answering questions in the domain of tourism. The distribution of types of temporal expressions in the user questions is captured in Table 1.

	Time Position		**Duration**	**Frequency**
Precise	Calendar dates	86	19	13
	Times of day	81		
	Week	12		
Fuzzy	Past, Present, Future	0	1	N/A
	Seasons and parts of year	1		
	Weekends	16		
	Day parts	20		
Non-specific	43		10	0

Table 1. Distribution of TEs in the QALL-ME user questions

5. Temporal Annotator for QALL-ME

The Question Answering system developed as part of the QALL-ME project requires as part of the Question Processing stage a module that adds temporal expression annotations to a user question. As stated before, the TIMEX2 standard was adopted as temporal annotation schema for this module. Besides the issue of a shared common annotation schema among all QALL-ME partners and the issue concerning the usability of the QALL-ME benchmark outside the QALL-ME project, the idea behind adopting the TIMEX2 standard in QALL-ME was also to re-use existing annotation tools capable of annotating according to the TIMEX2 standard. Such a tool is available at the University of Wolverhampton, and a brief description of the tool can be found in [6]. Since it is able to annotate all types of existing TEs, this tool is very complex, and despite its high performance, a real-time QA system would compromise performance against a faster runtime. Also, by investigating the user questions, one can easily notice that only a few types of temporal expressions are more frequently encountered in the data, and therefore by covering only a few TE classes, a simpler TE annotator would be enough for the QALL-ME system. Henceforth, a simplified version of the existing TE identifier was implemented following the design and methodology employed in the initial TE annotator. The simpler TE annotator covers only the most frequent types of temporal expressions present in the user questions, such as certain precise TEs denoting time position (e.g. some ways of expressing calendar dates, times of the day), as well as certain expressions of precise durations. It also covers some cases of fuzzy TEs, like certain expressions referring to weekends and day parts. The set-referring TEs, as well as all non-specific TEs are not covered by the simplified TE annotator.

The initial temporal expression identifier takes a rule-based approach to the identification and normalisation of TEs. The two stages are separated but they are not independent, as the information gathered at the identification stage is essential for normalisation. The identification stage involves two modules: one that applies patterns to raw text that did not undergo any pre-processing (a short example of a pattern that identifies certain date expressions can be found in Figure 1), and another module that checks the syntactic correctness of the expressions recognised by the first module. Whenever a pattern is matched, it does not only identify the sequence of words representing a TE, but it also generates a semantic representation for the identified expression. The semantic representation takes the form of a typed feature structure that depends on the semantic class of the TE, with features such as the temporal unit and value for durations, or the year, month, day of the month for calendar points that specify explicitly these values. For certain underspecified TEs, the semantic representation includes the function to be used in computing their final value.

```
1to9      =  [1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9]
0to9      =  [0 | 1to9]
Punct     =  [, | -]
Day       =  [ Monday | Tuesday | ... | Sunday ]
Month     =  [ January | February | ... | December ]
Date      =  [ 1to9 | [ 1 | 2 ] 0to9 | 3 [ 0 | 1 ] ]
Year      =  [1 | 2] 0to9 0to9 0to9
DateExpr  =  Day (Punct )? Month Date ( Punct Year )?
```

FIGURE 1. Building a pattern for capturing date expressions of the type *Saturday, November 28, 2009*

For many TEs, the normalisation stage involves only a direct translation of these semantic representations into normalised values of the attributes. In the case of underspecified expressions (either expressions denoting a value that is not fully specified up to the year level, or expressions that convey a function through their semantics), a temporal anchor is required for computing their final normalised value. For such expressions, the normalisation process continues by finding the anchor that will, in the end, fill the unknown slots of their values. The anchor should be an already resolved and fully specified temporal expression.

The simplified temporal annotator employed in QALL-ME follows the methodology described above, but targets only certain types of TEs. As opposed to the initial annotator, the simplified version does not include a module that checks the syntactic correctness of the expressions it identifies.

6. Evaluation

The set of 1,118 user questions that were manually annotated according to the TIMEX2 guidelines was used as gold standard in our evaluation. Both the initial TE identifier, as well as the simplified version employed in QALL-ME were evaluated against this gold standard both in terms of identifying parts of an annotated TE, as well as at the level of identifying correctly the entire extent of the annotated TEs. If the evaluation takes into consideration both complete matches, as well as partial matches, the original TE annotator has an F-measure of 95.6%, while the simplified version achieves an F-measure of 85.2%. When looking only at the TEs for which the full extent was correctly identified, the F-measure obtained by the initial TE identifier is 88.5% before checking syntactic correctness and 91.5% afterwards, and the one achieved by the simplified TE annotator is 72.6%. At the normalisation stage the TIMEX2 VAL attribute is filled in by the initial annotator with an accurary of 90.7%, and by the simplified annotator with an accuracy of 81.7%. The evaluation results are captured in more detail in Table 2.

	Initial annotator			**Simplified annotator**		
	Precision	Recall	F-measure	Precision	Recall	F-measure
TIMEX2	96.6%	94.6%	**95.6%**	96.6%	76.2%	**85.2%**
TEXT	92.4%	90.6%	**91.5%**	82.4%	64.9%	**72.6%**
TIMEX2:ANCHOR_DIR	72.7%	100%	**84.2%**	48.1%	92.9%	**63.4%**
TIMEX2:ANCHOR_VAL	72.7%	100%	**84.2%**	51.9%	100%	**68.3%**
TIMEX2:MOD	100%	100%	**100%**	100%	71.4%	**83.3%**
TIMEX2:SET	100%	100%	**100%**	20.0%	100%	**33.3%**
TIMEX2:VAL	90.3%	91.1%	**90.7%**	81.4%	82.1%	**81.7%**

TABLE 2. Detailed evaluation results on the QALL-ME user questions

An analysis of the errors revealed that many errors were due to unimplemented patterns. Expressions that are identified correctly using the initial annotator can not be identified by the simplified annotator due to the fact that many patterns are not implemented (e.g. *the check-out time*, *the opening hours*, *70s*, *weekday daytime*). For the same reason, certain expressions are only partially recognised (e.g. in the case of the expression *12 midnight*, only *midnight* is annotated).

Another important source of errors that appears at the normalisation stage is due to language ambiguity. In the case of the two examples below, both containing the temporal expression *1*, it is impossible for an automatic temporal annotator to correctly normalise the two expressions. Only commonsense knowledge can distinguish between the typical closing time of a bank that is normally in the afternoon, and the closing time of a restaurant that is normally at night, not mentioning that if the time passes midnight, the date must be changed as well. A temporal annotator can at its best normalise both expressions to *"2007-09-24T01"*, given the temporal anchor *24 September 2007*, but it is obvious that this annotation is wrong in both cases.

Can you please tell me the address of a Bank in Birmingham which closes at <TIMEX2 val="2007-09-24T13">1</TIMEX2>?

I'm looking for a modern European restaurant close to the Symphony Hall which closes at <TIMEX2 val="2007-09-25T01">1</TIMEX2>.

7. CONCLUSIONS

This paper presented the temporal processor employed in the English Question Answering system developed as part of the QALL-ME project. As the TEs present in user questions belong only to certain subclasses defined by the TIMEX2 guidelines, an existing temporal processor was simplified in order to provide the speed required by a real-time application. Both the initial and the simplified temporal processors take a rule-based approach to the identification and normalisation of TEs, mainly because rule-based approaches have proved to be more appropriate when targeting not only the identification, but also the normalisation of TEs.

The evaluation of the two temporal annotators shows that, in the case of the QALL-ME user questions, the initial one has an accuracy of 88.5% before checking

syntactic correctness and 91.5% after, and the simplified one an accuracy of 72.6% on complete TEs. The two annotators normalise TEs with accuracies of 90.7% and 81.7%, respectively. After simplifying the initial temporal annotator in order to increase its speed, a lower performance was expected, but since the expressions missed by the simplified annotator were mostly generic (i.e. most of them did not introduce a specific point in time essential for answering the question), it is likely that there would not be a major impact on the entire QA process.

References

[1] Mani, I. and Wilson, G.: Robust temporal processing of news. Proceedings of ACL-2000 (2000)
[2] Setzer, A. and Gaizauskas, R.: Annotating events and temporal information in newswire texts. Proceedings of LREC-2000 (2000)
[3] Filatova, E. and Hovy, E.: Assigning Time-Stamps to Event-Clauses. Proceedings of the 2001 ACL Workshop on Temporal and Spatial Information Processing (2001)
[4] Katz, G. and Arosio, F.: The Annotation of Temporal Information In Natural Language Sentences. Proceedings of the 2001 ACL Workshop on Temporal and Spatial Information Processing (2001)
[5] Schilder, F. and Habel, C.: From Temporal Expressions to Temporal Information: Semantic Tagging of News Messages. Proceedings of the 2001 ACL Workshop on Temporal and Spatial Information Processing (2001)
[6] Puscasu, G.: A Framework for Temporal Resolution. Proceedings of LREC-2004 (2004)
[7] Ahn, D., Adafre, S., F. and de Rijke, M.: Extracting Temporal Information from Open Domain Text. Journal of Digital Information Management (2005)
[8] Ahn, D., van Rantwijk, J. and de Rijke, M.: A Cascaded Machine Learning Approach to Interpreting Temporal Expressions. Proceedings of the 2007 NAACL-HLT (2007)
[9] MUC-6: Named entity task definition. http://www.cs.nyu.edu/cs/faculty/grishman/muc6.html (1995)
[10] ACE: http://www.nist.gov/speech/tests/ace/ace04/index.htm. (2004)
[11] TERN: Time Expression Recognition and Normalization. http://timex2.mitre.org/tern.html (2004)
[12] Ferro, L., Gerber, L., Mani, I., Sundheim, B. and Wilson, G.: TIDES 2005 Standard for the Annotation of Temporal Expressions (2005)
[13] Cabrio, E., Kouylekov, M., Magnini, B., Negri, M., Hasler, L., Orasan, C., Tomas, D., Vicedo, J.L., Neumann, G. and Weber, C: The QALL-ME benchmark: a Multilingual Resource of Annotated Spoken Requests for Question Answering. Proceedings of LREC-2008 (2008)

[1] Research Group in Computational Linguistics, University of Wolverhampton, Stafford Street, Wolverhampton, WV1 1SB, United Kingdom
E-mail address: andrea.varga@wlv.ac.uk

[2] Research Group in Computational Linguistics, University of Wolverhampton, Stafford Street, Wolverhampton, WV1 1SB, United Kingdom
E-mail address: georgie@wlv.ac.uk

[3] Research Group in Computational Linguistics, University of Wolverhampton, Stafford Street, Wolverhampton, WV1 1SB, United Kingdom
E-mail address: c.orasan@wlv.ac.uk