05151 Abstracts Collection

Annotating, Extracting and Reasoning about Time and Events
— Dagstuhl Seminar —

Graham Katz[1], James Pustejovsky[2] and Frank Schilder[3]

[1] Univ. Osnabrück, DE
`gkatz@uos.de`
[2] Brandeis Univ. Waltham, US
`jamesp@cs.brandeis.edu`
[3] Thomson Legal & Regulatory, Eagan MN, US
`Frank.Schilder@thomson.com`

Abstract. From 10.04.05 to 15.04.05, the Dagstuhl Seminar 05151 "Annotating, Extracting and Reasoning about Time and Events" was held in the International Conference and Research Center (IBFI), Schloss Dagstuhl. During the seminar, several participants presented their current research, and ongoing work and open problems were discussed. Abstracts of the presentations given during the seminar as well as abstracts of seminar results and ideas are put together in this paper. The first section describes the seminar topics and goals in general. Links to extended abstracts or full papers are provided, if available.

Keywords. Text annotation, information extraction and retrieval, summarization, question answering, temporal reasoning

05151 Summary—Annotating, Extracting and Reasoning about Time and Events

The main focus of the seminar was on TimeML-based temporal annotation and reasoning. We were concerned with three main points: determining how effectively one can use the TimeML language for consistent annotation, determining how useful such annotation is for further processing, and determining what modifications should be applied to the standard to improve its usefulness in applications such as question-answering and information retrieval.

Keywords: Temporal information extraction, annotation, temporal reasoning, events

Joint work of: Katz, Graham; Pustejovsky, James; Schilder, Frank

Full Paper: http://drops.dagstuhl.de/opus/volltexte/2005/354

Towards task-based temporal extraction and recognition

David Ahn (University of Amsterdam, NL)

We seek to improve the robustness and portability of temporal information extraction systems by incorporating data-driven techniques. We present two sets of experiments pointing us in this direction. The first shows that machine-learning-based recognition of temporal expressions not only achieves high accuracy on its own but can also improve rule-based normalization. The second makes use of a staged normalization architecture to experiment with machine learned classifiers for certain disambiguation sub-tasks within the normalization task.

Keywords: Information extraction, natural language, temporal reasoning, text mining

Joint work of: Ahn, David; Fissaha Adafre, Sisay; de Rijke, Maarten

Full Paper: http://drops.dagstuhl.de/opus/volltexte/2005/315

Approximate Qualitative Temporal Reasoning

Tom Bittner (Univ. des Saarlandes, D)

We partition the time-line in different ways, for example, into minutes, hours, days, etc. When reasoning about relations between events and processes we often reason about their location within such partitions. For example, x happened yesterday and y happened today, consequently x and y are disjoint. Reasoning about these temporal granularities so far has focussed on temporal units (relations between minute, hour slots). I shall argue in that in our representations and reasoning procedures we need into account that events and processes often lie skew to the cells of our partitions (For example, 'happened yesterday' does not mean that x started at 12 a. m. and ended 0 p. m.) This has the consequence that our descriptions of temporal location of events and processes are often approximate and rough in nature rather than exact and crisp. In this talk I describe representation and reasoning methods that take the approximate character of our descriptions and the resulting limits (granularity) of our knowledge explicitly into account.

Keywords: Approximate Reasoning, Qualitative Reasoning, Temporal Relations, Granularity, Ontology

TimeBank-Driven TimeML Analysis

Branimir Boguraev (IBM TJ Watson Research Center - Hawthorne, USA)

The design of TimeML as an expressive language for temporal information brings promises, and challenges; in particular, its representational properties raise the bar for traditional information extraction methods applied to the task of text-to-TimeML analysis. A reference corpus, such as TimeBank, is an invaluable asset in this situation; however, certain characteristics of TimeBank—size and consistency, primarily—present challenges of their own. We discuss the design, implementation, and performance of an automatic TimeML-compliant annotator, trained on TimeBank, and deploying a hybrid analytical strategy of mixing aggressive finite-state processing over linguistic annotations with a state-of-the-art machine learning technique capable of leveraging large amounts of unannotated data. The results we report are encouraging in the light of a close analysis of TimeBank; at the same time they are indicative of the need for more infrastructure work, especially in the direction of creating a larger and more robust reference corpus.

Joint work of: Boguraev, Branimir; Ando, Rie Kubota

Anchoring Temporal Expressions in Scheduling-related Emails

Benjamin Han (CMU - Pittsburgh, USA)

In this paper we adopt a constraint-based representation of time, Time Calculus (TC), for anchoring temporal expressions in a novel genre, emails. Email is sufficiently different from the most studied genre - newswire texts, and its highly under-specified nature fits well with our representation. The evaluation of our anchoring system shows that it performs significantly better than the baseline, and the result compares favorably with some of the closest related work.

Joint work of: Han, Benjamin; Gates, Donna; Levin, Lori

Full Paper: http://drops.dagstuhl.de/opus/volltexte/2005/316

Text Type and the Position of a Temporal Adverbial within the Sentence

Janet Hitzeman (MITRE - Bedford, USA)

A sentence with a certain type of temporal adverbial is ambiguous, and one reading is lost when the adverbial appears in sentence-initial position. Sentence (1a), for example, has a reading in which there was some three-hear period in the past during which Mary lived in Amsterdam and a reading in which Mary has lived in Amsterdam for the three years preceding speech time:

(1) a. Mary has lived in Amsterdam for three years. b. For three years Mary has lived in Amsterdam.

Sentence (1b) has only the reading in which Mary lives in Amsterdam at speech time and has done so for the preceding three years. The reading that remains when the adverbial is in sentence-initial position is more specific about the time at which the event occurs, and therefore one would expect to see more initial-position adverbials in a narrative text, where the order of events is important. In testing this hypothesis on the ECI corpus, it was found that it is not the narrative/non-narrative distinction that results in a significant difference in initial-position adverbial usage; Instead, narratives with a large amount of flashback material have significantly more initial position adverbials, indicating that in order to accurately predict adverbial position a subclassification of the category "narrative" based on the amount of flashback material is needed.

Keywords: Temporal adverbials, narrative, flashbacks

Full Paper: http://drops.dagstuhl.de/opus/volltexte/2005/317

A Temporal Ontology for the Semantic Web

Jerry Hobbs (USC/ISI - Marina del Rey, USA) vspace-0,2cm

I will first describe the OWL-Time ontology that was developed in conjunction with the DARPA Agent Markup Language (DAML) program. It covers the topological properties of time, including Allen's interval relations, measures of duration, and the clock and calendar.

Predicates are "declared" in OWL and axiomatized in first-order predicate calculus. I will then describe two more recent efforts.

The first is to axiomatize temporal aggregates, such as "every third Monday of every other month". In developing this theory, we have attempted to cover a wide range of natural language constructions as well as subsume the coverage of existing calendar systems. The second is an effort to annotate events in news with the range in which their duration is likely to fall. Here we have developed annotation guidelines to disambiguate the most common uncertain cases, and we have examined issues of inter-annotator agreement.

Keywords: Temporal ontology, temporal aggregates, event durations

Veridicity

Lauri Karttunen (PARC - Palo Alto, USA)

This paper addresses the problem of assessing the veridicity of textual content. Has an event mentioned in the text really occurred? Who is the source of the information? What is the stance of the author of the text? Does the author indicate whether he believes the source? We will survey some of linguistic conventions that indicate the author's commitment, or the lack thereof, to the propositions contained in her text. In particular we discuss phenomena that have been studied as presuppositions or conventional implicatures in previous literature. Some of those, such as factive and non-factive verbs, have received extensive attention in the past. Some others, such as supplemental expressions (e.g. appositives, parentheticals), have not received much previous attention, although they are very common and a rich source of textual inferences. A recent study by Christopher Potts classifies supplemental expressions as conventional implicatures. We agree with Potts on the label but not on what it means. In contrast to Potts, we claim that supplemental expressions cannot always be treated as the author's direct commitments and argue that they do not constitute a basis for a distinction between presuppositions and conventional implicatures. We illustrate some cases of conventional implicature and show how they indicate an author's commitment to the truth of his statements and briefly state the importance of these distinctions for Information Extraction (IE).

Toward a denotational semantics for TimeML

Graham Katz (Universität Osnabrück, D)

The annotation language TimeML represents temporal relationships among times and events. A model-theoretic semantics for this annotation language is presented, problems with the most straightforward suggestions raised, and an alternative for a solution provided. Chief among the difficulties is that the language of TimeML provides no way of specifying semantic scope. This raises familiar problems for the interaction with negation and other operators.

Chronoscopes: A theory of underspecified temporal representations

Inderjeet Mani (Georgetown Univ. - Washington, USA)

Representation and reasoning about time and events is a fundamental aspect of our cognitive abilities and intrinsic to our construal of the structure of our personal and historical lives and recall of past experiences. This talk describes an abstract device called a Chronoscope, that allows a temporal representation (a set of events and their temporal relations) to be viewed based on temporal abstractions. The temporal representation is augmented with abstract events called episodes that stand for discourse segments. The temporal abstractions allow one to collapse temporal relations, or view the representation at different time granularities (hour, day, month, year, etc.), with corresponding changes in event characterization and temporal relations at those granularities. A temporal representation can also be filtered to specify temporal trajectories of particular participants. Trajectories, in turn, can be intersected at various levels of granularity. Chronoscopes can be used to compare temporal representations (e.g., for aggregation, summarization, or evaluation purposes), as well as help in the visualization of temporal narratives

Keywords: Temporal abstraction, granularity, event structure

Full Paper: http://drops.dagstuhl.de/opus/volltexte/2005/336

Computational Treatment of Temporal Notions – The CTTN-System

Hans-Jürgen Ohlbach (Universität München, D)

The CTTN-system is a computer program which provides advanced processing or temporal notions.

The basic data structures of the CTTN-system are time points, crisp and fuzzy time intervals, labelled partitionings of the time line, durations, and calendar systems. The labelled partitionings are used to model periodic temporal notions, quite regular ones like years, months etc., partially regular ones like timetables, but also very irregular ones like, for example, dates of a conference series.

These data structures can be used in the temporal specification language GeTS (GeoTemporal Specifications). GeTS is a functional specification and programming language with a number of built-in constructs for specifying customized temporal notions.

CTTN is implemented as a Web server and as a C++ library.

This paper gives a short overview over the current state of the system and its components.

From TimeML to TPL

Ian Pratt-Hartmann (Manchester University, GB)

This paper describes a subset of the temporal mark-up language TimeML, and explains its relation to various formalisms found in the literature on interval temporal logic. The subset of TimeML we describe can be viewed as an interval temporal logic with a tractable satisfiability problem, but very limited expressive power. Most crucially, that logic does not permit quantification over events. The contribution of this paper is to point out that, by choosing an appropriate interval temporal logic, it is possible to introduce quantification into representations of event-structure without sacrificing decidability.

The Role of Arguments in Event Markup

James Pustejovsky (Brandeis Univ. Waltham, USA)

TimeML is a specification language for the annotation of events and temporal expressions in natural language text. In addition, the language introduces three relational tags, linking temporal objects and events to one another. These links impose both aspectual and temporal ordering over time objects, as well as mark up subordination contexts introduced by modality, evidentiality, and factivity. Given the richness of this specification, the TimeML working group decided not to include the arguments of events within the language specification itself. Full reasoning and inference over natural language texts clearly requires knowledge of events along with their participants. In this paper, I define the appropriate role of argumenthood within event markup and propose that TimeML should make a basic distinction between arguments that are events and those that are entities.

I first review how TimeML treats event arguments in subordinating and aspectual contexts, creating event-event relations between predicate and argument.

As it turns out, these constructions cover a large number of the argument types selected for by event predicates. I suggest that TimeML be enriched slightly to include causal predicates, such as *lead to*, since these also involve event-event relations. All other verbal arguments are ignored by the language, however, and any predicate-argument binding of participants to an event is to be performed by independent means. In fact, except for the event-denoting arguments handled by the extension to TimeML proposed here, almost full temporal ordering of the events in a text can be computed without argument identification.

Keywords: Temporal annotation, event annotation, argument structure, temporal ordering

Temporal information extraction from legal documents

Frank Schilder (Thomson Legal & Regulatory Corp, USA)

The aim of this paper is to analyze what kinds of temporal information can be found in different types of legal documents. In particular, it provides a comparison of different legal document types (case law, statute or transactional document) and it discusses how one can do further reasoning with the extracted temporal information.

Keywords: Extraction of temporal information, temporal reasoning, legal documents

Joint work of: Schilder, Frank; McCulloh, Andrew

Full Paper: http://drops.dagstuhl.de/opus/volltexte/2005/313

TimeML in a Medical Application

Andrea Setzer (University of Sheffield, GB)

We talk about the use of TimeML in a different application - a medical application within the CLEF project. The aim of CLEF is to develop methods for capturing and managing clinical patient data, which comes from two overlapping but complementary sources. One source is the patient notes, unstructured text dicated by doctors. The other source is the structured data, which contains, for example, dates of interventions and investigations.

TimeML is being used to annotate the temporal information of investigations and interventions, the use of which is two-fold.

Information from both sources is inter-related and integrated to build a patient's "chronicle" and to help interpret the structured text. Also, aggregation over multiple patients will allow different research questions to be answered.

Joint work of: Gaizauskas, Rob; Harkema, Henk; Hepple, Mark; Setzer, Andrea

The Calculus of Affordance

Marc Steedman (University of Edinburgh, GB)

This paper analyzes temporal semantics for natural language in terms of a calculus developed for planning and reasoning about action. The calculus depends on two fundamental operation, namely composition of actions into sequences, and type-lifting or the relation of objects to the actions that they afford or make possible. An event calculus with these properties based on Linear Dynamic Logic, and on instantaneous changes rather than intervals, provides a transparent basis for planning of the reactive, forward-chaining kind available to higher animals.

I have argued elsewhere that composition and type-raising offer a universal combinatory basis for natural language syntax. Here I shall argue that the same calculus provides a helpful basis for temporal semantics, analysing phenomena including English tenses and aspects, Navaho de-verbal nouns, and English de-nominal verbs in those terms.

Along the way I shall touch on a very long neuropsychological tradition linking motor planning and the language faculty in evolutionary and developmental terms.

Drawing TimeML Relations with T-BOX

Marc Verhagen (Brandeis Univ. Waltham, USA)

T-BOX is a new way of visualizing the temporal relations in TimeML graphs. Currently, TimeML's temporal relations are usually presented as rows in a table or as directed labeled edges in a graph. I will argue that neither mode of representation scales up nicely when bigger documents are considered and that both make it harder than necessary to get a quick picture of what the temporal structure of a document is. T-BOX is an alternative way of visualizing TimeML graphs that uses left-to-right arrows, box-inclusions and stacking as three distinct ways to visualize precedence, inclusion and simultaneity.