BLENDER
Cross-lingual Cross-document IE Lab

COGNITIVE COMPUTATION GROUP UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Temporal Information Extraction and Shallow Temporal Reasoning

Dan Roth*, Heng Ji†, Taylor Cassidy†, Quang Do*

*Computer Science Department
University of Illinois at Urbana-Champaign

†Computer Science Department and Linguistics Department,
Queens College and the Graduate Center
City University of New York

December 8, 2012


SPeak
ers

- Slides are available at

http://nlp.cs.qc.cuny.edu/tietutorial.pptx

"Time is a sort of river of passing events, and strong is its current; no sooner is a thing brought to sight than it is swept by and another takes its place, and this too will be swept away."

- *Marcus Aurelius*

Outline

1. Background: Motivations and Goals 9:30
2. Temporal Information Representation Theories 9:35
3. Temporal Expression Extraction and Normalization 10:10
4. Temporal Slot Filling 10:30

5. Tea Break 11:30

6. Event Timelining and Temporal Reasoning 12:00
7. Resources and Demos 12:55
8. Conclusions 1:00

- Background

Why Extracting Temporal Information?

- Many relations and events are temporally bounded
  - a person's place of residence or employer
  - an organization's members
  - the duration of a war between two countries
  - the precise time at which a plane landed
  - …
- Temporal Information Distribution
  - One of every fifty lines of database application code involves a date or time value (Snodgrass,1998)
  - Each news document in PropBank (Kingsbury and Palmer, 2002) includes eight temporal arguments

Why Extracting Temporal Information?

- Important to many NLP applications
  - Textual inference (Baral et al., 2005)
  - Multi-document text summarization (Barzilay e al., 2002),
  - Temporal event tracking (e.g. Chambers et al., 2009; Ji and Chen, 2009)
  - Temporal grounding for semantic relations (Do et al., 2012)
  - Template based question answering (Ahn et al., 2006, Schockaert et al., 2006)
  - Knowledge Base Population (Ji et al., 2011)

Outline

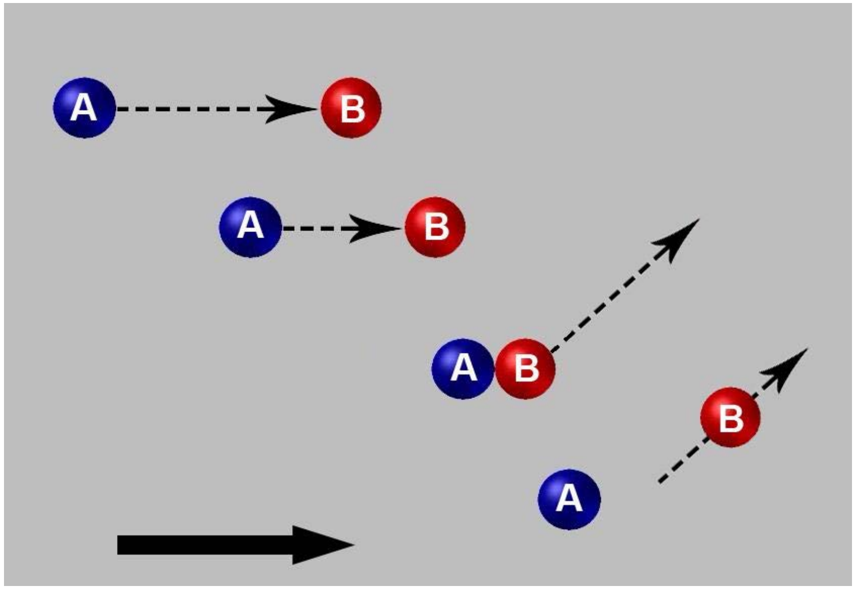
1. Background: Motivations and Goals
2. Temporal Information Representation Theories
3. Temporal Expression Extraction and Normalization
4. Temporal Slot Filling

5. Tea Break

6. Event Timelining and Temporal Reasoning
7. Resources and Demos
8. Conclusions

9:35

- Temporal Information Representation Theories



Temporal Information Representation Theories

- General Goal
- Semantics of Events
- Grounding Eventualities in Time
- Temporal Representation Formalisms
  - Allen Relation (Allen, 1983)
  - TimeML (Pustejovsky et al., 2003)
  - Temporal Closure (Verhagen, 2005)
  - Fuzzy Intervals (Schockaert et al., 2008)
  - 4-Tuple Temporal Representation (Ji et al., 2011)
  - Timeline Representation (Do et al., 2012)

Temporal information

- Concept of time
  - Communicate and reason about **change** and **action**
- Temporal Representation Scheme
  - Determines the order of and distance between **events** and **states**, i.e., ***eventualities****
  - Admits of varying granularities and levels of certainty
  - Consists of scalar quantities and relations that hold between them
- Provides a target to which a system maps linguistic objects
  - Cognitive/Human or Computational/NLP

* We often use ***event*** to refer to events and states from here on

Time expressions in language

- Temporal Expressions (**TE**)
  - Denote intervals and points at varying granularity and (under-) specificity
  - Can be combined with temporal functors
- Calendar based TE
  - *in* 1992; 1990's; *on* Dec. 8th, 1992; *at* 8:00 am
  - *From* Monday *to* Friday; *between* 9 *and* 10 am
  - *This/next* Century; year; day; hour; minute; second, …
  - *Before* 1992; *until next* year; *for* (about) one year; tomorrow; *after* 8:00;
  - Tuesdays; first of the month; several times per week; often

Under-specification and granularity

- **Now** → December 8th, 2012, 3:00pm.
  - 2012-12-08-T15:00
- "December 8th, 2012
  - 2012-12-08-TXX:XX
  - (2012-12-08T00:00, 2012-12-08-T11:59)
- "December 8th"
  - xxxx-12-08-Txx:xx
  - Set of all December 8th's
  - Each is an interval
    - { (xxxx-12-08:T00:00, xxxx-12-08:T11:59) }

Granularity during composition

- One day
- “Tomorrow”
- “within twenty-four hours”
- **Now ⚪** Tomorrow
  - → 2012-12-**09**
  - +0000-00+01 (Dale and Mazur, 2006)
- **Now ⚪** “within twenty-four hours”
  - → (2012-12-08-T15:00, 2012-12-09-T15:00)
- Finish the assignment <u>tomorrow</u>
- Finish the assignment <u>within twenty-four hours</u>

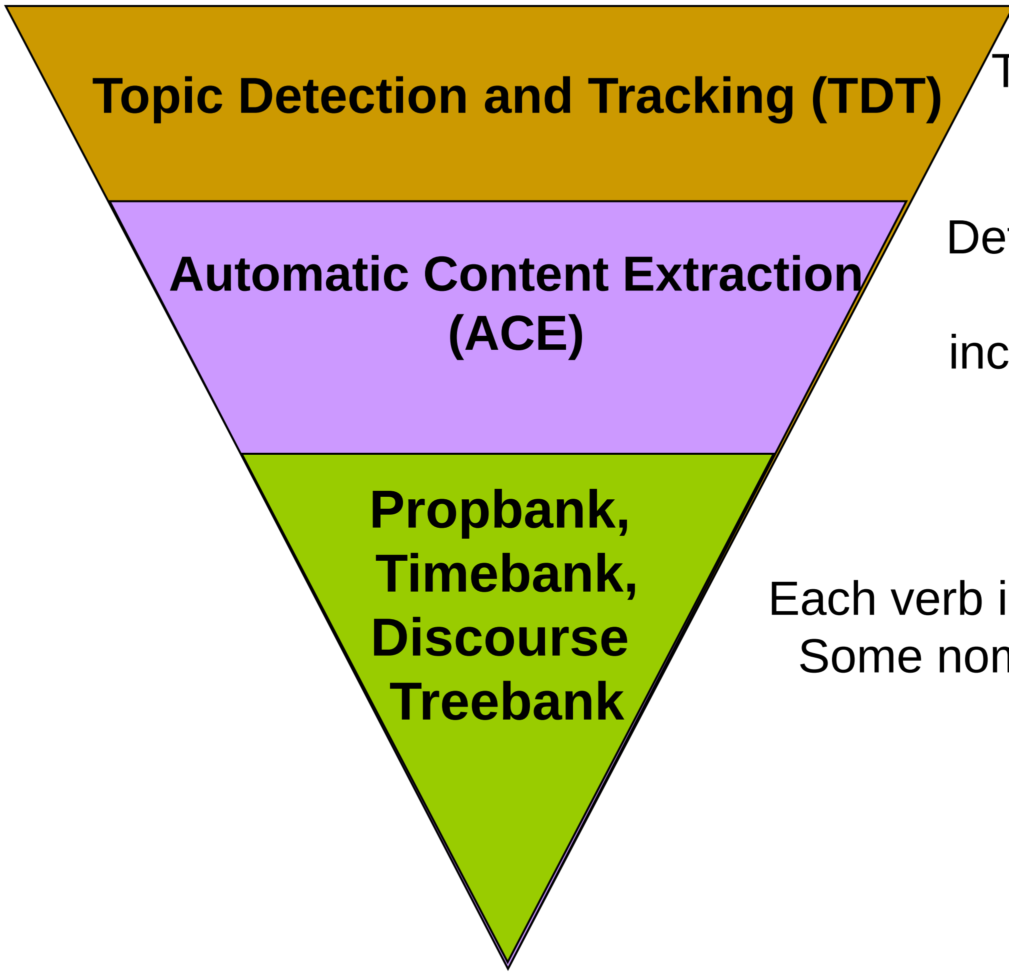
Event structure & classification

- Event reification & subatomic semantics
  - Quantification over and predication of: **events** (Davidson; 1967)
  - ∃**e**[*Stabbing*(**e**)∧*AG*(**B**, **e**)∧*TH*(**C**, **e**)∧*Violently*(**e**)∧ with(*K*, e)]
  - …on Friday → […∧on(Friday, e)] (Parsons, 1990)
  - See also: Tenny & Pustejovksy (2000)
- Verb phrase classes/Aktionsarten/Lexical Aspect (**LA**)
  - State, activity, accomplishment, achievement (Vendler, 1957)
  - Coercion by adv modification (Moens and Steedman, 1998)
  - Generalized coercion with fine-grained categories (Dölling, 2011)
  - Structural analogy with nouns (Bach 1986; Krifka, 1998)
  - **Telicity:** Telic events require result/goal to occur
  - **Dynamicity**: Distinguishes states from other events
  - **Durativity**: Is event conceptualized as instantaneous?

Granularity during anchoring

- Now → 2012-12-08-T15:00
- “Finish the assignment **tomorrow**”
  - Achievement
  - Concerns a single TBD point within 2012-12-09
- “Do the assignment **tomorrow**”
  - Accomplishment
  - Concerns a TBD subinterval of 2012-12-09
- “Be good **tomorrow**”
  - State
  - Event is co-extensional within 2012-12-09

What is an event?



Temporal order of coarse-grained groups of events ("topics")

Defined 33 types of events, each event mention includes a trigger word and arguments with roles

Each verb is an event type, no arguments
Some nominals and adjectives allowed

Types of eventualities

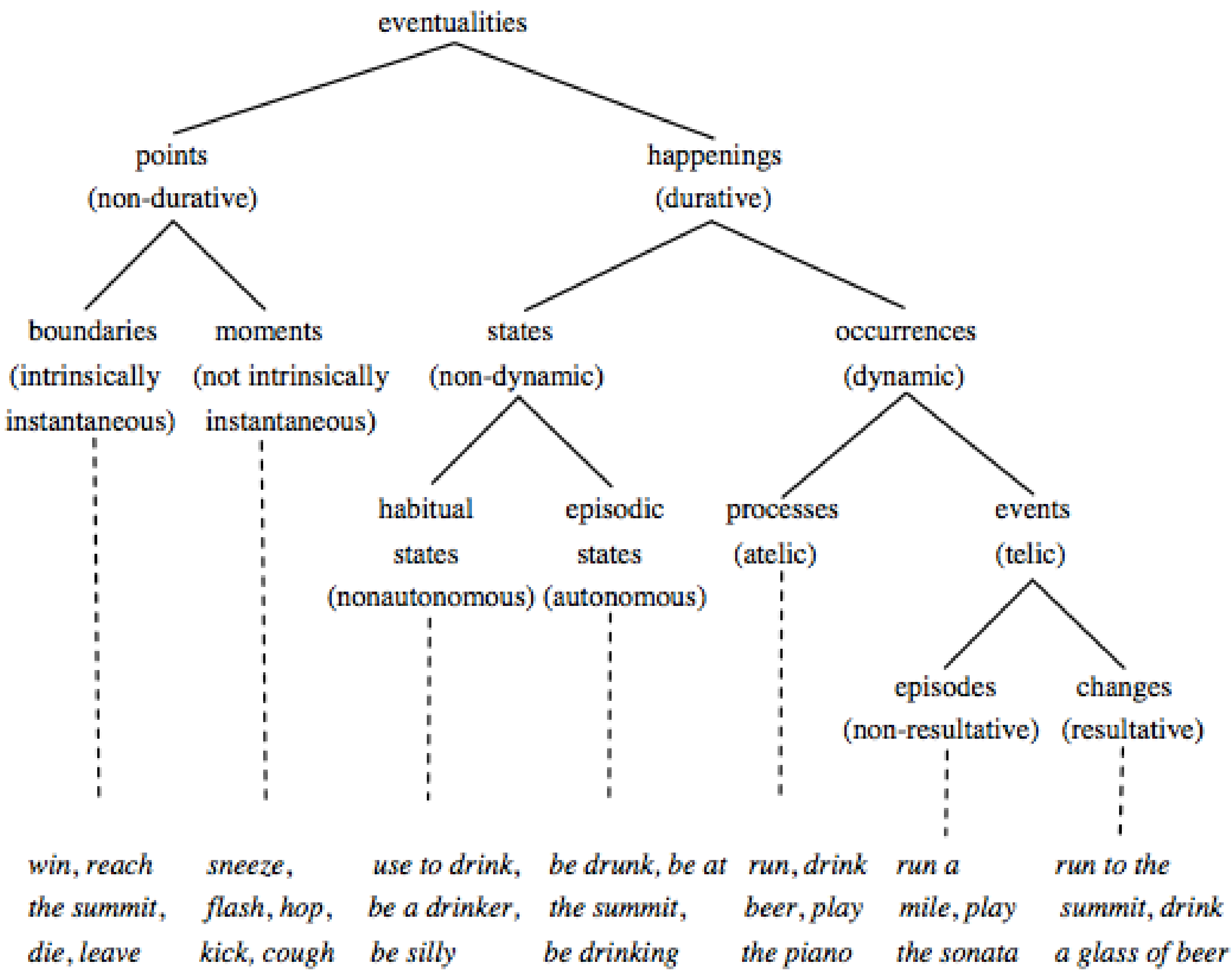


Chart from (Dölling, 2011)

Inter-eventuality relations

- A **boundary** *begins/ends* a **happening**
- A **boundary** *culminates* an **event**
- A **moment** is the *reduction of* an **episode**
- A **state** is the *result of* a **change**
- A **habitual state** is *realized by* a class of **occurrences**
- A **Processes** is made of **event** *constituents* …

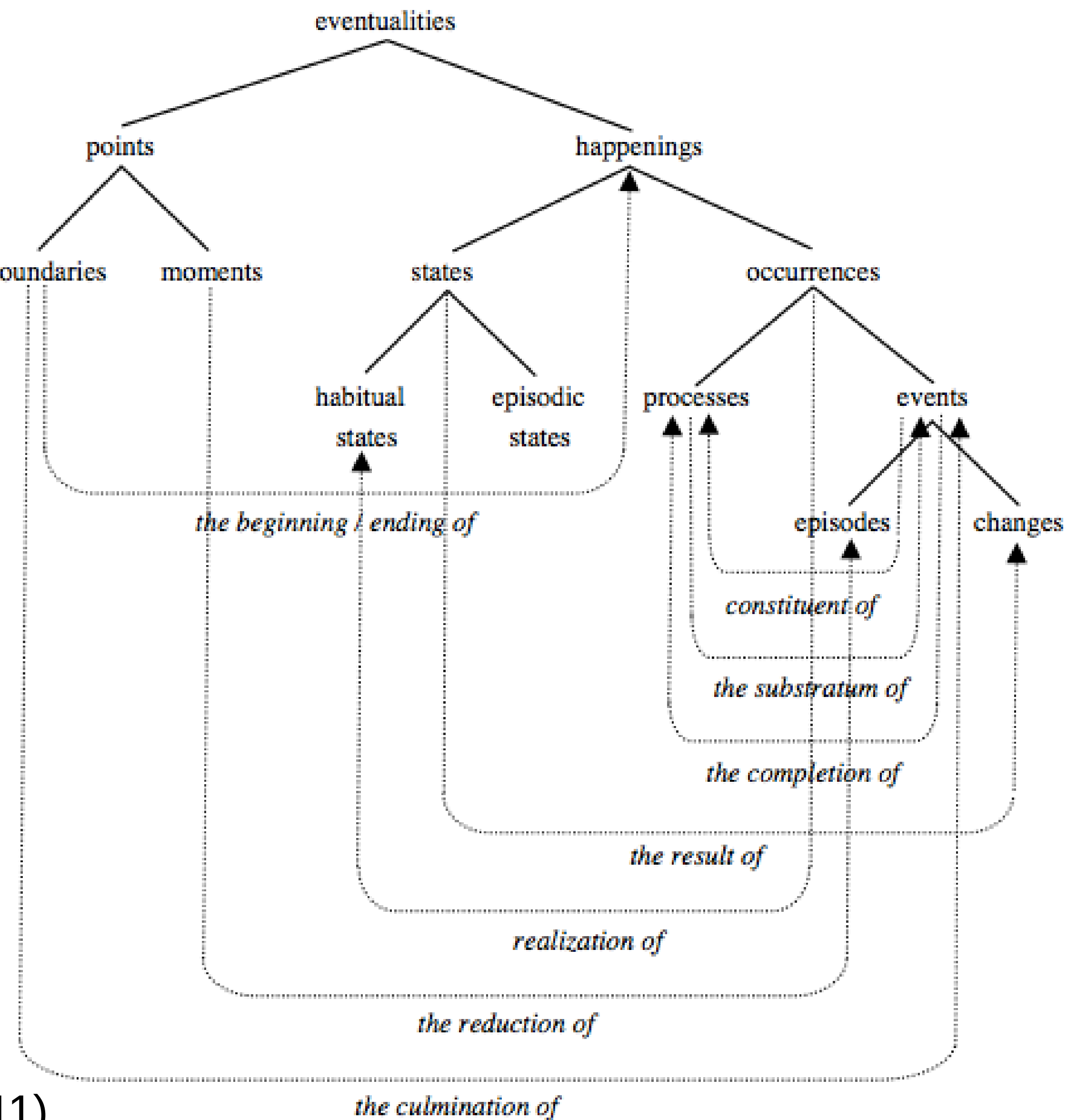


Chart from (Dölling, 2011)

Aspectual coercion

Event Predicate	Win:= λe. **win**(e)
Type restrictions	$\forall$e[**win**(e) $\rightarrow$ BOUNDARY(e)]
	$\forall$e$\forall$t [**for**(e, t) $\rightarrow$ Time_Int(t) $\wedge$ (STATE(e) $\vee$ PROCESS(e)]
Inter-eventuality relations	$\forall$e [HAPPENING(e) $\rightarrow\exists$e'$\exists$e''[BEG(e',e)]$\wedge$END(e'',e)]] $\forall$e$\forall$e' [END(e, e') $\rightarrow$BOUNDARY(e)$\wedge$HAPPENING(e')] $\forall$e [EVENT(e) $\rightarrow\exists$e'[CULM(e', e)] $\forall$e$\forall$e' [CULM(e, e') $\rightarrow$ EVENT(e') $\wedge$END(e, e') $\forall$e [BOUNDARY(e) $\rightarrow$ BEG(e)$\vee$END(e) $\forall$e [PROC(e) $\rightarrow$ $\exists$e'$\exists$e''[CONST(e',e)$\wedge$CONST(e'',e) $\wedge$ADJ(e',e'')]] $\forall$e$\forall$e' [CONST(e, e') $\rightarrow$ EVENT(e)$\wedge$PROCESS(e')]
Additive coercion	λPλe. $\exists$e':CULM(e',e)[P(e')]
Iterative coercion	λPλe. $\forall$e':CONST(e',e)[P(e')]

Generalized Coercion λPλe. Qe':R(e',e)[P(e')]

Aspectual coercion

Event Predicate	A win is a win
Type restrictions	Wins are of type BOUNDARY Only a STATE or PROCESS can happen **for** a time
Inter-eventuality relations	HAPPENINGS have a beginning and end Each end is a BOUNDARY of a HAPPENING EVENTS must culminate Culminations are the ends of EVENTS Any BOUNDARY is a beginning or an end A PROCESS consists of two or more temporally adjecent Constituents. Any constituent is an EVENT that makes up part of a PROCESS
Additive coercion	Apply the property to the EVENT the BOUNDARY culminates
Iterative coercion	Apply the property to the PROCESS of which the EVENT is a constituent

Chris won for three hours

A win is a BOUNDARY, but one can only engage in a STATE or PROCESS for three hours.

A PROCESS is made of two or more temporally adjacent EVENTS. An EVENT is a HAPPENING that must end due to some BOUNDARY occurring. Since every BOUNDARY is a beginning or an end of some happening, the natural interpretation of the proposition is that there was a PROCESS consisting of two or more EVENTS, each of which culminated with a BOUNDARY of type win.

Thus, the proposition will be true just in case there are two or more EVENTS whose boundaries are wins that make up such a PROCESS, lasting three hours.

win: $\lambda e.\ \mathbf{win}(e)$

win: $\lambda e.\ \forall e'{:}CONST(e',e)[\exists e'':CULM(e'',e')[\mathbf{win}(e'')]]$

$\exists e[AG(\mathbf{chris}, e) \wedge \forall e'{:}CONST(e',e)\ [\exists e'' : CULM(e'',e')\ [\mathbf{win}(e'')] \wedge for(e, \mathbf{3hours})]$

Ordering of events in time

- Order of events conveyed and understood
  - Event to time interval mapping
  - In terms of inherent ordering on time intervals
  - In terms of events to event relations
- Tense and Grammatical Aspect (**T**)
  - Expressed morpho-syntactically
  - Past, Present, Future (-*ed*; *will* + V)
  - Perfective, Imperfective, Unmarked (*Has* + V_{part}; V-*ing*)
- TE, LA, & T, guided by commonsense knowledge interact to **anchor events in time**

Ordering events in time

- Speech (S), Event (E), & Reference (R) time (Reichenbach, 1947)

Sentence	Tense	Order
John wins the game	Present	E,R,S
John won the game	Simple Past	E,R<S
John had won the game	Perfective Past	E<R<S
John has won the game	Present Perfect	E<S,R
John will win the game	Future	S<E,R
Etc…	Etc…	Etc…

- **Tense:** relates R and S; **Gr. Aspect:** relates R and E
- R associated with *temporal anaphora* (Partee 1984)
- Order events by comparing R across sentences
- By the time Boris noticed his blunder, John had (already) won the game

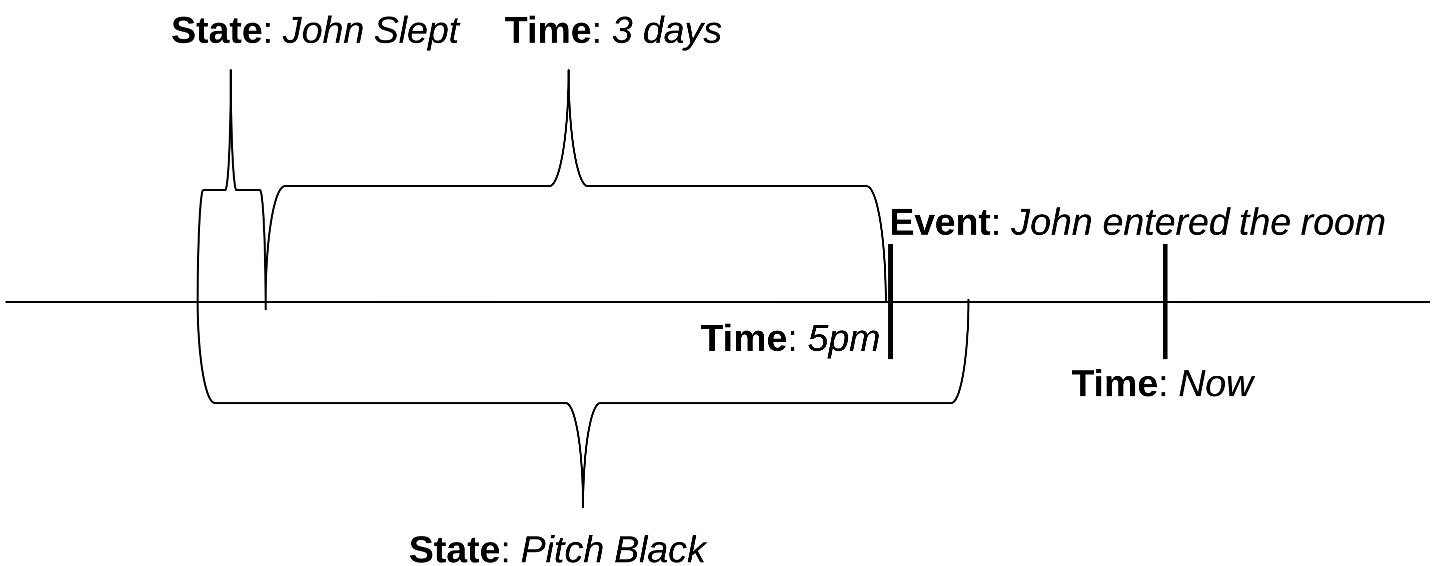
See Michaelis (2006) for a good explanation of tense and grammatical aspect

Ordering events in a discourse

- Incorporation into Discourse Representation Theory (e.g. Hinrichs, 1986)
  - Default assumption: Eventualities in consecutive clauses may not overlap unless one or both are stative
- Temporal Discourse Interpretation Principle (Dowty 1986)
  - Particulars of reference time movement and aspectual class of verbs interact, but both are significantly informed by world knowledge
- “It is crucial … that semantic theory determine what options may be left open by the information given, so that other modules of information may provide additional constraints that the central logic “service” of the system may exploit in generating conclusions” (Ter Meulen, 1991)
- Syntax, semantics, causal and linguistic knowledge accounted for in single logic without reference time (Lascarides and Asher; 1992)
- **Semantics Literature does not fully address:**
  - **How are pragmatic/world knowledge constraints on meaning represented and how might they be learned?**
  - **Representation and reasoning over temporal information of widely varying granularity and scope**
  - ***How exactly* do we associate events with their temporal arguments?**

Ordering events in discourse

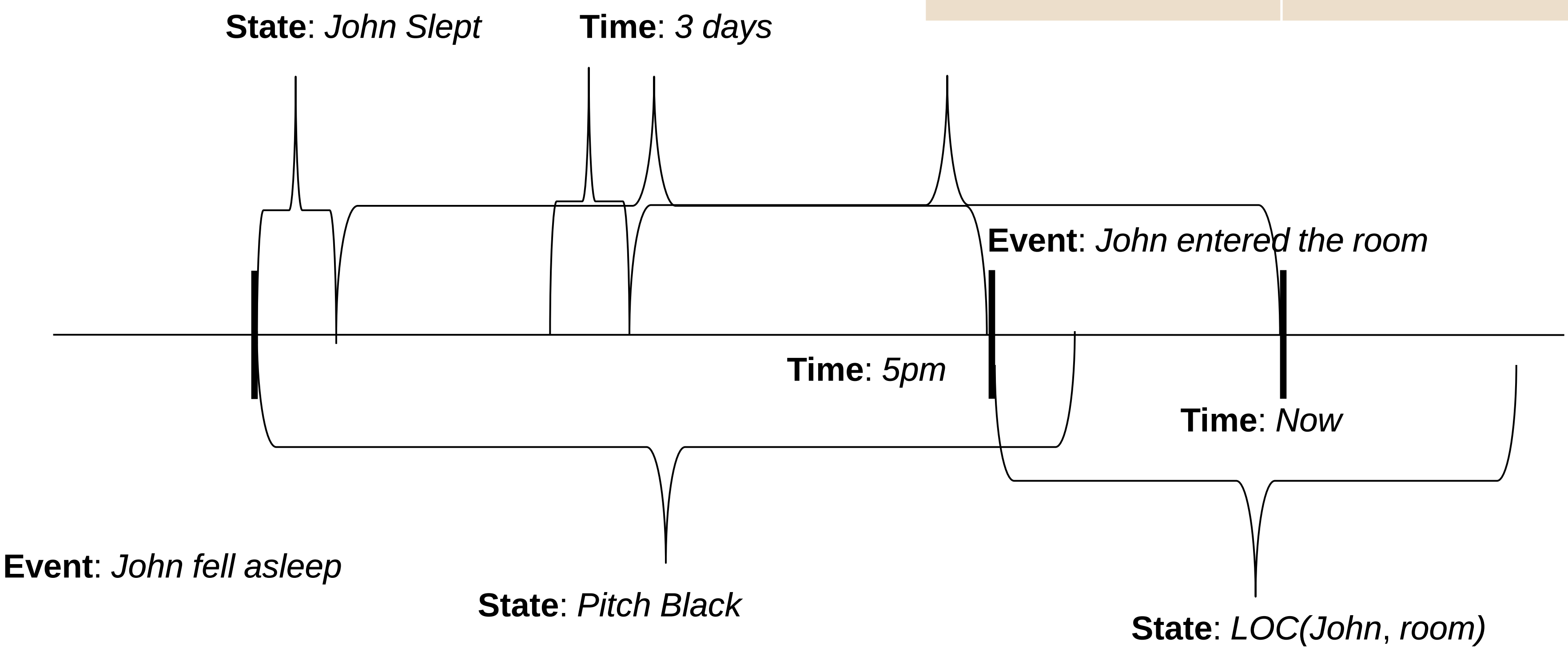
- (1) John entered the room at 5:00pm.
- (2) It was pitch black.
- (3) It had been three days since he'd slept.



Grounding eventualities in time

- (1) John entered the room at 5:00pm.
- (2) It was pitch black.
- (3) It had been three days since he'd slept.

Perfective Past	E—R—S



Role of knowledge

- Recognize events and temporal expressions
- Determine which are related
- Determine type of relation
- Draw Inferences about implicit events and relations
  - Cause/effect, contingency, etc…

Linguistic knowledge

John **worked** for company A in **1999**

[S John **[VP** ***t*** **[V'[V''[V** **worked** **][PP** for company A **]][PP**in **1999** **]]]]**

Activity **d∧di∧=**

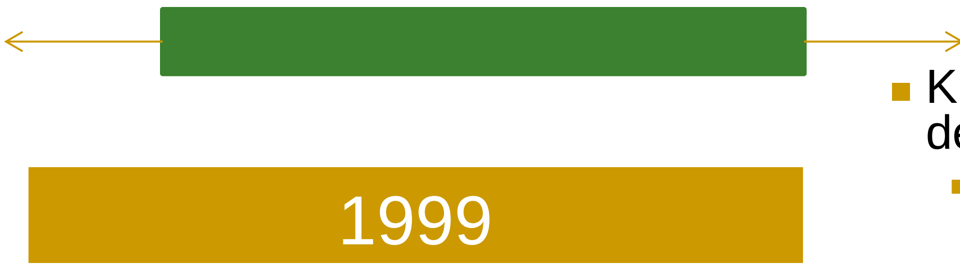
Lexical knowledge needed to identify event triggers and temporal expressions.

Syntactic and Semantic knowledge needed to determine whether, and if so how, an event and temporal expression are related

John **was hired** in 1999

- **Temporal Representation framework** needed to map temporal expressions to their extents

John **worked** for company A in 1999



- Knowledge of **verb class** determines type of interval
  - Punctuated or Persistent?

John **worked** for company A **until** 1999

- Knowledge of **the mode** in which prepositions map an event to its time argument
  - Requires interval-based reasoning

J ompany A **until** 1999

…

Immediately after leaving comp

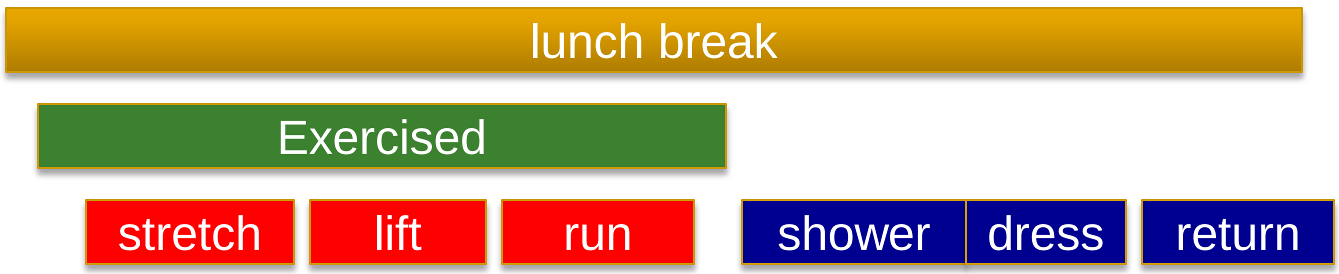
1999

- Relations **between events** interact with relations **between events and time intervals**, and vice versa
- Controlled by:
  - 1. Mapping from events & temporal expressions, to intervals
  - 2. Interval relations

<u>Goal:</u> Enrich text with information that enables machines to (learn to) extract necessary temporal information

Commonsense Knowledge

- (1) John [**exercised**]$_{e1a}$ during his [lunch break]$_{e1b}$.
- (2) He [stretched,]$_{e2a}$ [lifted weights,]$_{e2b}$ and [ran]$_{e2c}$.
- (3) He [showered,]$_{e3a}$ [got dressed,]$_{e3b}$ and [returned to work.]$_{e3c}$



- **<u>Explicit:</u>** $e_{1a} \subseteq e_{1b}$
- **<u>Implicit:</u>** $e_{2a}, e_{2b}, e_{2c,} \subseteq e_{1a} \wedge e_{3a} < e_{3b} < e_{3c} \wedge e_{3a}, e_{3b}, e_{3c} \subseteq e_{1b}$
  - Requires an appeal to the “normal course of events”

Temporal representation formalisms

- Representation required to specify ordering and extension of eventualities in time
- TR in general
  - Eventualities and Temporal Expressions map to **intervals**
  - Reasoning
    - Relations among intervals, world knowledge, some notion of uncertainty and under-specification
    - Determine order of events to some degree
- Ultimate test: **Question Answering** with temporal awareness

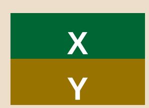
Temporal relations – motivation (Pustejovsky et al., 2005)

- Answering *temporally relevant queries* requires **reasoning**
- Reasoning requires the ability to represent eventualities, temporal extents, and relations between them
  - **E-T E-E T-T**
    - Who **won** the Turing award in **1966**?
    - Who **died** during **The Clinton Administration**?
    - On **what day** was **Dec 25th** in **2004**?
- **Temporal extents** can be referred to explicitly
  - Date, indirect reference, WH-word
- Or implicitly, via **events**

Allen interval relations (Allen, 1983)

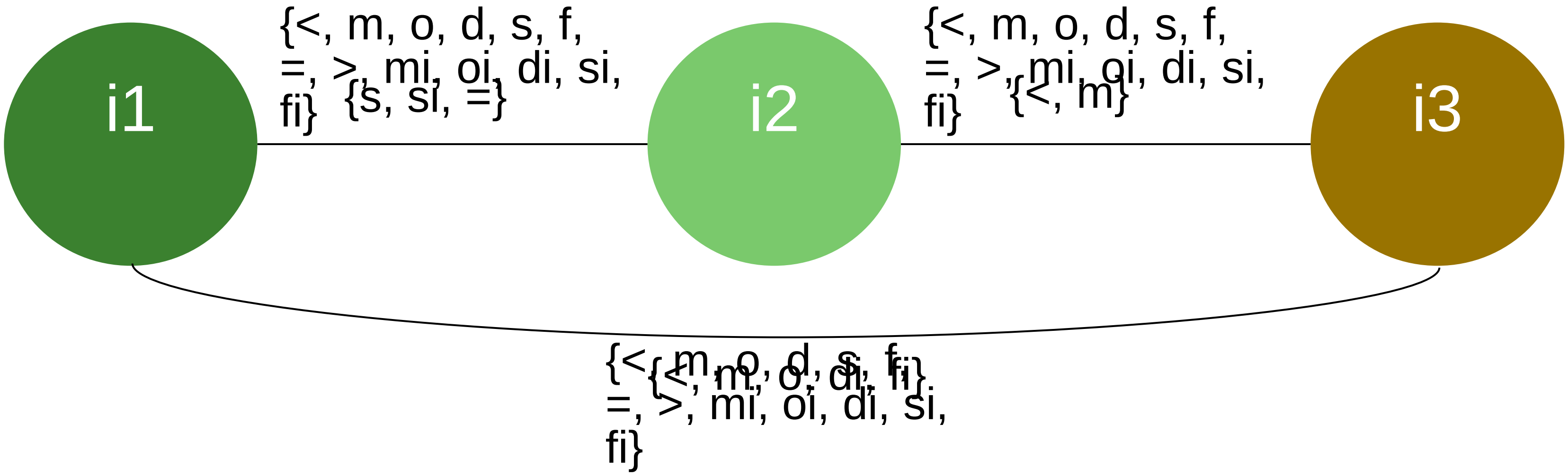
- Ideal Framework will allow
  - Imprecision
  - Uncertainty
  - Varying granularity
  - Persistence
- Proposed Framework
  - 13 Interval relations
  - Inference algorithm to characterize understanding of implicit temporal relation information in terms of what's explicitly provided

Allen interval relations (Allen, 1983)

Relation	Symbol	Inverse	Example
X before Y	<	>	
X meets Y	m	mi	
X overlaps Y	o	oi	
X during Y	d	di	
X starts Y	s	si	
X finishes Y	f	fi	
X equals Y	=	=	

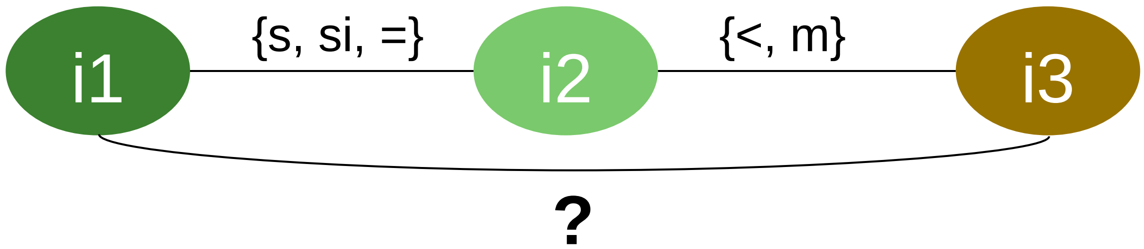
- 13 total relations
- One or more relations hold between any pair of eventualities
- Reasoning done via propagation, using **transitivity table**
  - (see Allen, 1983)
- Transitivity table facilitates compositions of sets of interval relations

Propagation (Allen, 1983)



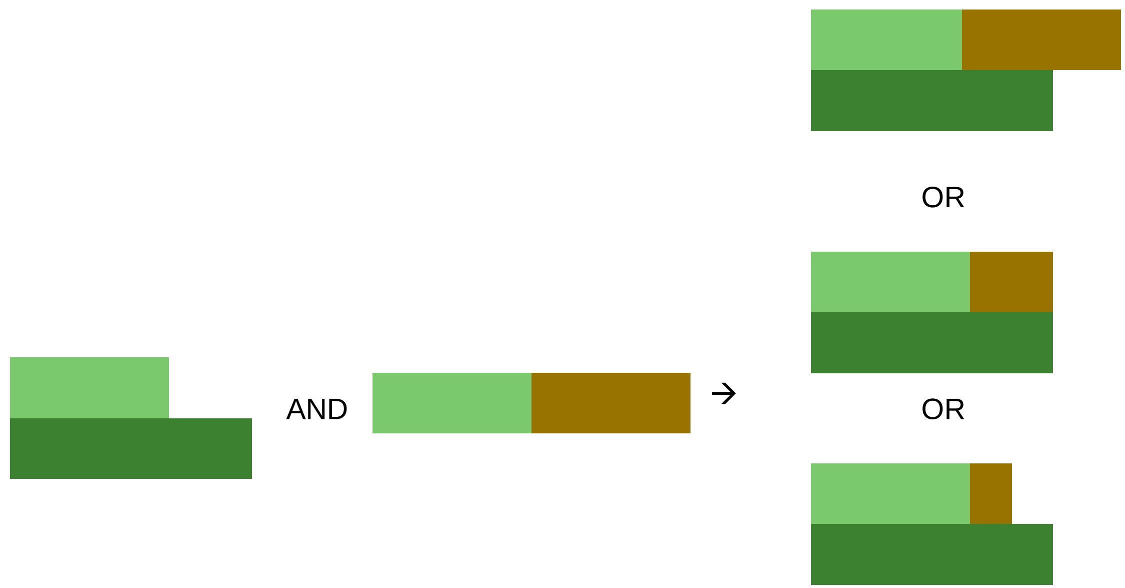
- **A priori:** no information is known about how any two intervals are related
- Partial information is acquired…
- Reasoning is performed…

Propagation (Allen, 1983)

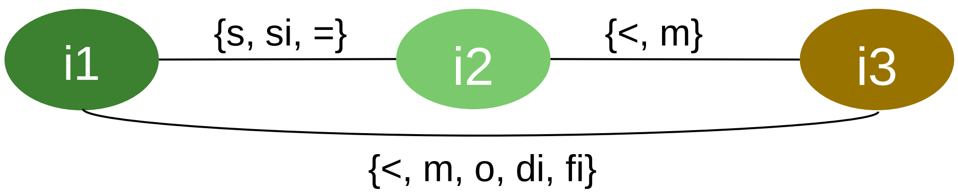


{s, si, =}X{<, m} → { (s, <), (s, m), (si, <), (si, m), (=, <), (=, m)}

i1 - i2	i2 - i3	i1 - i3
s	<	<
s	m	<
si	<	<, o, m, fi, di
si	**m**	**o, fi, di**
=	<	<
=	m	m



Potential to augment IE?



John was **hired** as a **technician** by company A in **June of 2006**, and he **began training** shortly thereafter. After **that**, He **worked** for company B in **2010**.

- J worked for A, some time passed, J started working for B. **(<)**
- J worked for A. This ended, and J started working for B simultaneously. **(m)**
- J worked for A, at some point J began working for B, his employment at A ended during this time. **(o)**
- J worked for A, then stopped. Later on he was working for B. In fact, he started working for B first, and still was upon leaving A. **(di)**
- J was working for B, started working for A, and stopped working for A and B at the exact same time. **(fi)**

TimeML: mark-up language for time (Pustejovsky et al., 2005)

- Annotation language for marking *temporal* and *event expressions*, as well as *links* between them
  - Components: `TIMEX3, EVENT, SIGNAL, LINK`
- Broad community effort of design & specification
  - Annotation guidelines (Saurí et al, 2006)
  - Resource instantiation:
    - annotation ‘standard’; appeals both to annotation and analysis
    - TimeBank corpus (Sundheim and Radev, 2002)
- Evolution of design
  - TERQAS / TANGO workshops : 2002 — 2003
- Builds upon TIDES’ `Timex2,` Sheffield STAG

TimeML : problems addressed

- Dependencies among events and timex
  - Specifying an event's temporal extent
  - Ordering of events, relative to one another
- Reasoning
  - with underspecified temporal expressions : "*last week*", "*eight years before*", *"October (of 2003)"*
  - About duration of events & outcomes of events
- Signals used to form complex temporal expressions
  - Signals: for, during, at, before, after, while, …
- Complex event structures
  - Aspectual phrase: initiation (e.g. <u>started</u> <u>working</u>), continuation, …
  - Subordinating relations: counterfactive (<u>prevented</u> the <u>attack</u>), reporting...
  - Polarity indicators: not, no, none, …
- Temporal quantification:
  - twice, three times, everyday…

Time ML example

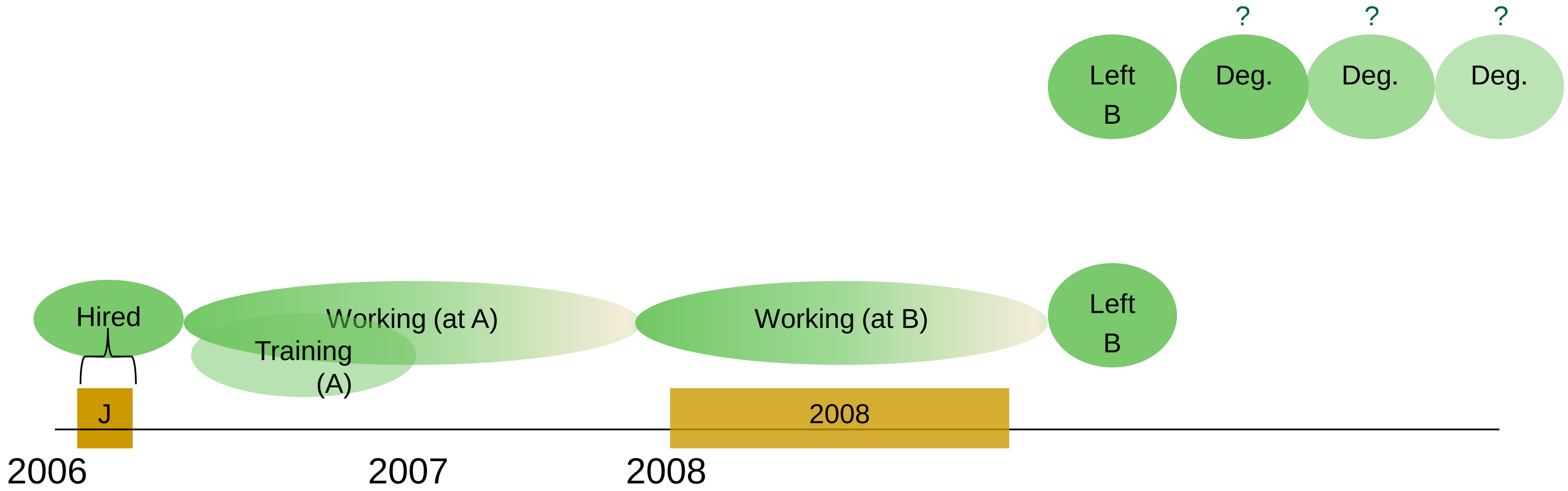
John was **hired** as a **technician** by company A in **June of 2006**. He **began training** shortly thereafter. He **worked** for company B in **2008**.

Time ML annotation

John was **hired** as a **technician** by company A **in June of 2006**. He **began training shortly thereafter**. He **worked** for company B **in 2008**.

…

After leaving Company B, John **earned** a degree in engineering.



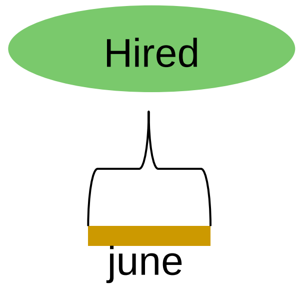
Time ML annotation

John was
<EVENT eid=e1 class= “OCCURRENCE” >
Hired </EVENT>
as a
<EVENT eid=e2 class= “STATE” >
technician </EVENT>
by company A
<SIGNAL sid=s1>
in </SIGNAL>
<TIMEX3 tid=t1 type= “DATE” value= “200606XX” temporalFunction= “true” >
June of 2006. </TIMEX3>
…
He
<EVENT eid=e4 class= “OCCURRENCE” > worked </EVENT>
for company B
<SIGNAL sid=s2> in </SIGNAL>
<TIMEX3 tid=t2 type= “DATE” value= “2008XXXX” temporalFunction= “true” > 2008 </TIMEX3>
…
<SIGNAL sid=s3> After </SIGNAL>
<EVENT eid=e3 class= “OCCURRENCE” > leaving </EVENT>
the company, John <EVENT eid=e5 class= “OCCURRENCE” > earned a degree in engineering</EVENT>

Time ML annotation

<MAKEINSTANCE eiid= “ei1” eventID= “e1” pos= “VERB” tense= “PAST” aspect= “NONE” …>

<TLINK eventInstanceId= “ei1” signalID=s1 relatedToTime= “t1” relType= “IS_INCLUDED” >



Limit one relType!

<TLINK eventInstanceId= “ei2” signalId=s2 relatedToTime= “t2” relType= “DURING” >

<TLINK eventInstanceId= “ei2” signalId=s2 relatedToTime= “t2”

re

Working at B

Working at B

2008

2008

<TLINK eventI

Earned a degree in engineering

=s3 relatedToEv

After

Left B

d= “ei3”

relType= “AFT

Tempeval

- Tempeval-1
  - Relate an event and a timex in the same sentence
  - Relate an event and the document creation time
  - Relate the main events of two consecutive sentences
- Tempeval-2 added
  - Find timex and events and their attributes
  - Restrict event/time relations to where event syntactically dominates timex, or both appear in the same NP
  - Relate events where one syntactically dominates the other
  - Italian, Chinese, Spanish, & Korean
- Tempeval-3: Bottom-up TIE
  - Find timex, events, & attributes (including event class)
  - Determine which temporal entities need to be related, provide any TimeML relation type

Time ML annotation

Would you annotate this?

Earned Degree in Engineering (ei5) **After** Training (ei2)

<TLINK eventInstanceId= “ei5” relatedToEventInstanceId= “ei2” relType= “AFTER” >

Given this?

John was **hired** as a **technician** by company A in **June of 2006**. He **began training** shortly thereafter. He **worked** for company B in **2010**.

…

After leaving Company B, John **earned** a degree in engineering.

John was **hired** as a **technician** by company A in **June of 2006**.
He **began training** shortly thereafter. He **worked** for company B in **2008**.

…

After leaving Company B, John **met** Mary.

How about **all of these?**

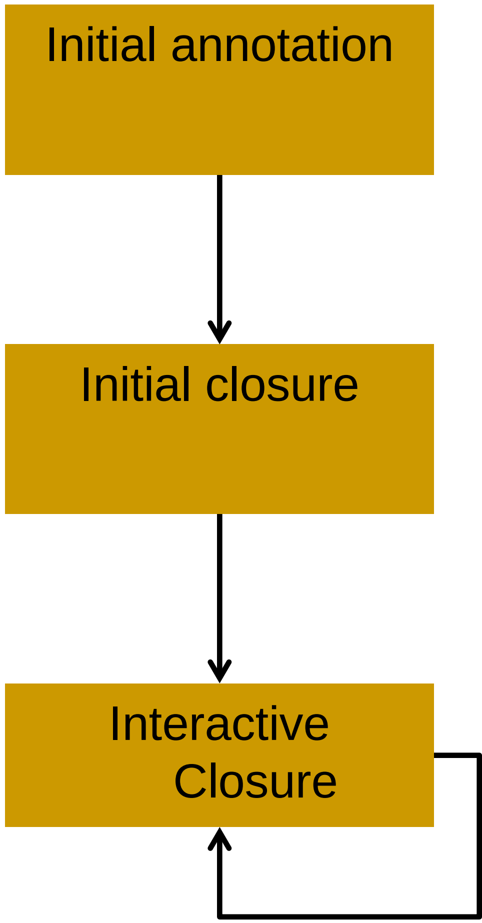
Temporal closure (Verhagen, 2005)

- Can we annotate all temporal relations in a text?
- Few temporal relations are expressed explicitly
- *N(N-1)/2* relations given N events & timex
  - Annotators cover about 1-5%
  - Annotators disagree on what to annotate
- System annotation is unable to capture complex cases
- **Solution: harness strengths of both types**

Temporal closure

SputLink (Verhagen, 2005)

- TimeML mapped into reduced Allen relations in terms of interval endpoints
- Human annotates; Machine propagates
  - Annotation task linear in document size
  - No need to compare distant items
- **Drastically improves number of annotated pairs (density) while minimizing inconsistencies (UzZaman and Allen, 2011; Setzer et al, 2003)**
- Use closure to aid temporal IE evaluation



Fuzzy intervals and reasoning (Schockaert et al., 2008)

- Allen relations are not 100% realistic
- Real life is fuzzier...
- We saw incomplete knowledge
  - Disjunction of interval relations
- But relations and events may be inherently vague
  - “Roosevelt died *just* before the **cold war**”
- **Approach:** Allen relations are a special case of a more general framework

“Long Before”

$$L^{\ll}_{(\alpha,\beta)}(a,b) = \begin{cases} 1, & \text{if } b-a > \alpha+\beta \\ 0, & \text{if } b-a \leq \alpha \\ \frac{b-a-\alpha}{\beta}, & \text{otherwise} \end{cases}$$

Let $\alpha \in \mathbb{R}$ and $\beta \in [0,+\infty[$.

$\alpha = 0$

$\beta = 0$

$L^{\ll}_{(\alpha,\beta)}$ reduces to $<$

"Long Before"

$$L^{\ll}_{(\alpha,\beta)}(a,b) = \begin{cases} 1, & \text{if } b-a > \alpha+\beta \\ 0, & \text{if } b-a \le \alpha \\ \frac{b-a-\alpha}{\beta}, & \text{otherwise} \end{cases}$$

Let $\alpha \in \mathbb{R}$ and $\beta \in [0,+\infty[$.

α = 0

β = 0

$\mathbf{a}\, L^{\ll}_{(\alpha,\beta)}\, \mathbf{b} = 1$

“Long Before”

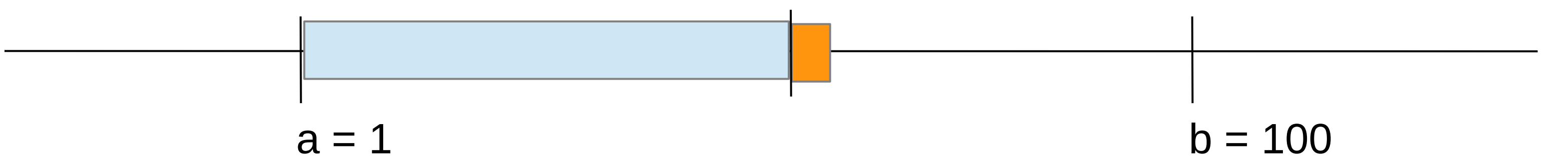
$$L^{\ll}_{(\alpha,\beta)}(a,b) = \begin{cases} 1, & \text{if } b-a > \alpha+\beta \\ 0, & \text{if } b-a \leq \alpha \\ \frac{b-a-\alpha}{\beta}, & \text{otherwise} \end{cases}$$

Let $\alpha \in \mathbb{R}$ and $\beta \in [0, +\infty[$.

α = 50

β = 10

100-1 > 50 + 10
99 > 60

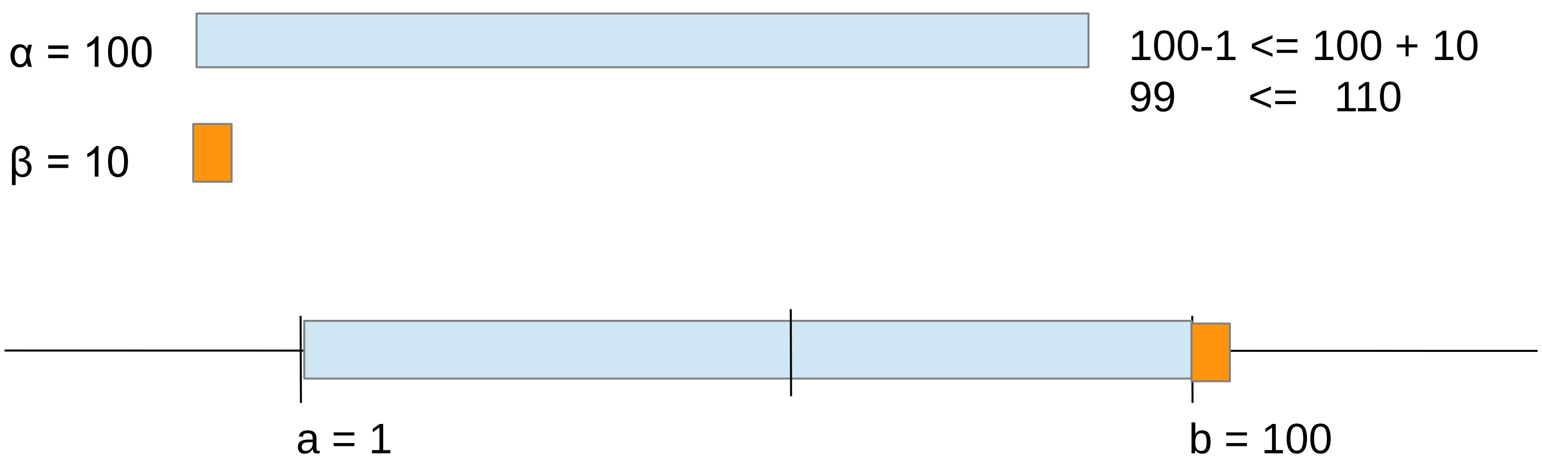


$$\mathbf{a}\; L^{\ll}_{(\alpha,\beta)}\; \mathbf{b} = 1$$

“Long Before”

$$L^{\ll}_{(\alpha,\beta)}(a,b) = \begin{cases} 1, & \text{if } b - a > \alpha + \beta \\ 0, & \text{if } b - a \leq \alpha \\ \frac{b-a-\alpha}{\beta}, & \text{otherwise} \end{cases}$$

Let $\alpha \in \mathbb{R}$ and $\beta \in [0, +\infty[$.



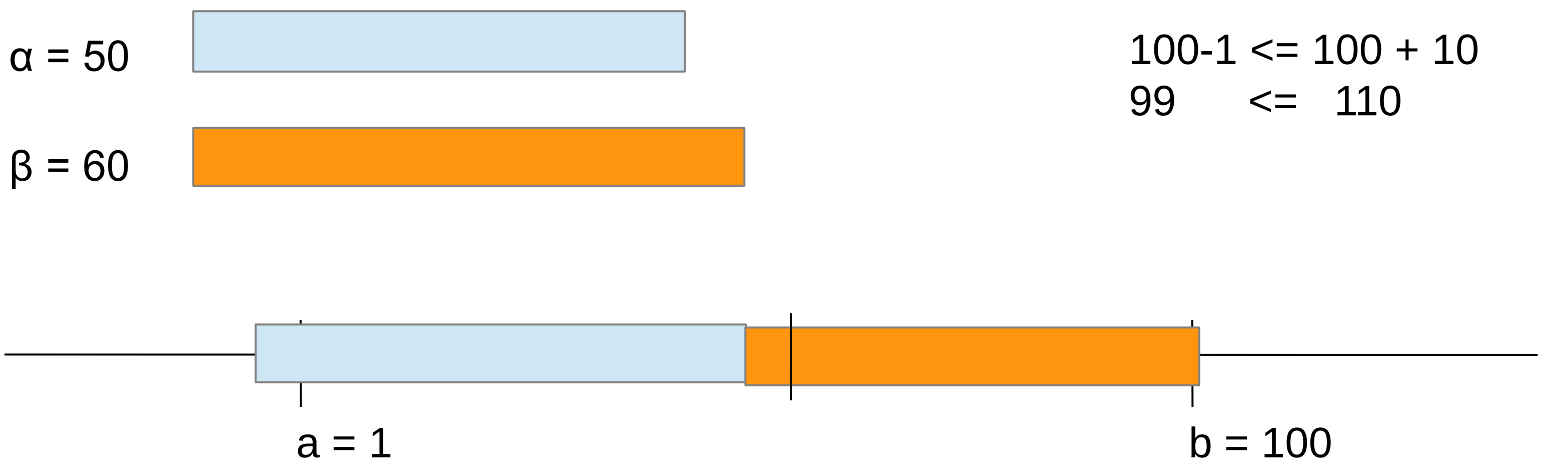
100-1 <= 100 + 10
99 <= 110

$$\mathbf{a}\ L^{\ll}_{(\alpha,\beta)}\ \mathbf{b} = 0$$

"Long Before"

$$L^{\ll}_{(\alpha,\beta)}(a,b) = \begin{cases} 1, & \text{if } b - a > \alpha + \beta \\ 0, & \text{if } b - a \leq \alpha \\ \frac{b-a-\alpha}{\beta}, & \text{otherwise} \end{cases}$$

Let $\alpha \in \mathbb{R}$ and $\beta \in [0, +\infty[$.



100-1 <= 100 + 10
99 <= 110

$$\mathbf{a}\, L^{\ll}_{(\alpha,\beta)}\, \mathbf{b} = .8167$$

Fuzzy time periods

- Also defined:
  - **a** occurs before or at approximately the same time as **b**
  - **a** occurs at approximately the same time as **b**
  - **a** occurs just before **b**
- Intuitive notions follow from fuzzy logic
  - E.g. "if **b** is *long before* **a**, **a** and **b** cannot be at *approximately the same time*, and **b** cannot be *before* **a**"
- Relations among fuzzy time periods, fuzzy transitivity table, defined in terms of fuzzy time point orderings
- Conclusions drawn about degree to which two fuzzy time periods stand in a fuzzy relation to one another:
- E.g. The degree to which **A** *is during* **B** and the degree to which **B** *more or less meets* C can be used to compute a lower bound for the degree to which **A** *is long before* **C**

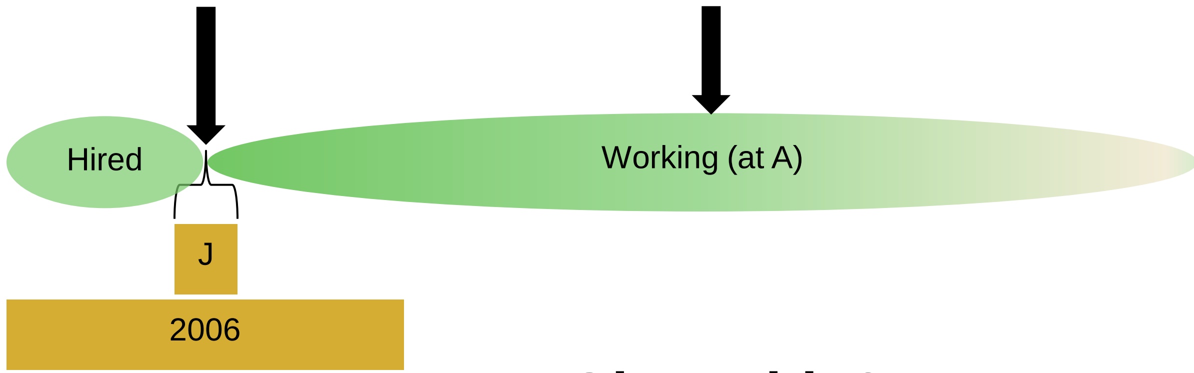
4-tuple representation (Ji et al., 2011)

- Challenges:
  - Be consistent with 'data base' approach of Knowledge Base Population (KBP)
  - Accommodate incomplete information
  - Accommodate uncertainty
  - Accommodate different granularities
- Solution:
  - express constraints on start and end times for slot value
  - 4-tuple $<t_1, t_2, t_3, t_4>$: $t_1 < t_{start} < t_2$ $t_3 < t_{end} < t_4$

Document text (2001-01-01)	T1	T2	T3	T4
Chairman Smith	-infinite	20010101	20010101	+infinite
Smith, who has been chairman for two years	-infinite	19990101	20010101	+infinite
Smith, who was named chairman two years ago	19990101	19990101	19990101	+infinite
Smith, who resigned last October	-infinite	20001001	20001001	20001031
Smith served as chairman for 7 years before leaving in 1991	19840101	19841231	19910101	19911231
Smith was named chairman in 1980	19800101	19801231	19800101	+infinite

Time ML annotation

Wouldn't you want to annotate this?



Given this?

John was **hired** as a **technician** by company A in **June of 2006**.
He **began training** shortly thereafter. He **worked** for company B in **2010**.

…

After leaving Company B, John **earned** a degree in engineering.

And this? Hiring → Employment

Employee_of(J, A) → <20060601, 20060630, 20060601, ∞>

Limitations of TimeML

- Redundancy: no need to have temporal relations between all pairs of events and time points and events and events
- Normalization: Hard to construct a timeline of events across documents
- Inference: Does not support well global inference for timelines

An Interval based Representation (Do et al., 2012)

- An interval-based representation
- Each temporal expression is normalized to an absolute interval and put on a universal timeline
- Each event is associated with an interval and thus is in partial order relation with other events on the timeline
- The interval-based representation allows one to construct an absolute timeline of events, so it’s easy to construct timeline of events across document
- This representation supports a concise inference model

Timeline Relation Representation & Mapping

- Each event is represented by a time interval, denoted by (e^-, e^+):
  - e^- and e^+ are two time endpoints.
  - represent the lower and upper bounds of the time interval of an event.
- Example:
  - *The election was held in September, 2008.*
  - e^- = '2008-09-01', e^+ = '2008-09-30'
- 3 base relations on endpoints
  - Before ($\prec$), After ($\succ$) Equal (=)
    (Denis and Muller, 2011)
- Hard constraint: $e^- \preceq e^+$
- Transitivity constraints of endpoints (supports inference):

	$\prec$	$\preceq$	$\succ$	$\succeq$	$=$
$\prec$	$\prec$	$\prec$			$\prec$
$\preceq$	$\prec$	$\preceq$			$\preceq$
$\succ$			$\succ$	$\succ$	$\succ$
$\succeq$			$\succ$	$\succeq$	$\succeq$
$=$	$\prec$	$\preceq$	$\succ$	$\succeq$	$=$

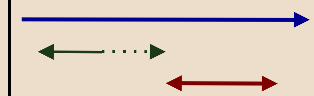
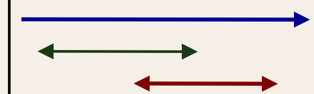
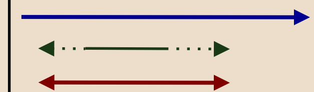
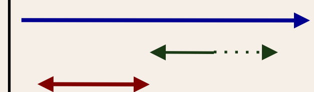
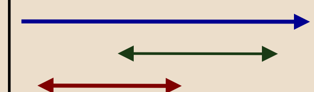
Comments:

- The timeline relations are used to represent both event-event and event-temporal expression relations.
- There is no explicit equal relation, however, we define two events to be equal *iff* they occur during each other.
- The relations can apply both to ttime intervals and time points.
- Transitivity constraints of timeline relations:

	b	*o*	*d*	*bi*	*oi*	*di*
b	*b*	*b*	*b/o/d*		*b/o/d*	*b*
o	*b*	*b/o*	*o/d*	*bi/oi/di*	*oi/di*	*b/o/di*
d	*b*	*b/o/d*	*d*	*bi*	*d/bi/oi*	
bi		*d/bi/oi*	*d/bi/oi*	*bi*	*bi*	*d/bi/oi*
oi	*b/o/di*	*o/d/oi*	*d/oi*	*bi*	*bi/oi*	*bi/oi/di*
di	*b/o/di*	*o/di*		*bi/oi/di*	*oi/di*	*di*

- An application can selectively enforce the constraints. In our work, we used a slightly different sets of relations and constraints.

val Representation

TL relation	Endpoint	Graphical illustration
b	(≺, ≤, ≺, ≺)	
o	(≺, ≻, ≺, ≺)	
d	(≥, ≥, ≤, ≤)	
bi	(≻, ≻, ≥, ≻)	
oi	(≻, ≻, ≺, ≻)	
di	(≤, ≥, ≥, ≤)	

Interval Representation: Summary

- The Interval based formulation provides an interval based representation of time along with:
  - Calculus for reasoning about end points
  - Transitivity reasoning for end points
  - Calculus for reasoning about intervals and this also events)
  - Transitivity reasoning for intervals (and thus events)
- As we will see, these properties give rise to natural way to reason about events and time, resulting in inference for time lining of event.

Outline

1. Background: Motivations and Goals
2. Temporal Information Representation Theories
3. Temporal Expression Extraction and Normalization
4. Temporal Slot Filling

5. Tea Break

6. Event Timelining and Temporal Reasoning
7. Resources and Demos
8. Conclusions

10:10

- Temporal Expression Extraction and Normalization

Temporal Expression Examples

Expression	Value in Timex Format
December 8, 2012	2012-12-08
Friday	2012-12-07
today	2012-12-08
1993	1993
the 1990's	199X
midnight, December 8, 2012	2012-12-08T00:00:00
5pm	2012-12-08T17:00
the previous day	2012-12-07
last October	2011-10
last autumn	2011-FA
last week	2012-W48
Thursday evening	2012-12-06TEV
three months ago	2012:09

Reference Date = December 8, 2012

Timex Value Attribute

the early 1990's	: `value="199X" MOD="START"`
the past 10 years	: `value="P10Y" anchor_val="2012" anchor_dir="BEFORE"`
the next week	: `value="P1W" anchor_val="2012-W49" anchor_dir="AFTER"`
the previous day	: [cf. `point` above]
recent	: `value="PAST_REF" anchor_val="2012-12-08T09:00" anchor_dir="BEFORE"`

Temporal Expression Extraction

- Rule-based (Strtotgen and Gertz, 2010; Chang and Manning, 2012; Do et al., 2012)

- Machine Learning
  - Risk Minimization Model (Boguraev and Ando, 2005)
  - Conditional Random Fields (Ahn et al., 2005; UzZaman and Allen, 2010)

- State-of-the-art: about 95% F-measure for extraction and 85% F-measure for normalization

A Grammar for Date Expressions (Boguraev and Ando, 2005)

1to9 = [1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9]
0to9 = [%0 | 1to9]
SP = [“, “]
Day = [Monday | Tuesday | ... | Sunday]
Month = [January | February | ... | December]
Date = [1to9 | [1 | 2] 0to9 | 3 [%0 | 1]]
Year = 1t09 (0t09 (0to9 0to9)))
Date Expression =

Day | (Day Sp) Month “ “ Date (SP Year)

Example Rules (Chang and Manning, 2012)

- Token Patterns
  - "4 to 5 years":

    { ruleType: "tokens",
      pattern: (($NUM) /to|-/ ($NUM) ["-"]? ($TEUNITS_NODE)),
      result: Duration($1, $2, $3) }
- String Patterns
  - "3-years":

    { ruleType: "text",
      pattern: /(\d+)[-\s]($TEUnits)(s)?([-\s]old)?/ ,
      result: Duration($1, $2) }
- Time Patterns
  - "Date at Time":

    { ruleType: "composite",
      pattern: (([{ temporal::IS_TIMEX_DATE }]) /at/
    ([{   temporal::IS_TIMEX_TIME }])),
     result: TemporalCompose(INTERSECT, $0[0].temporal,
     $0[-1].temporal) }

Machine Learning: Sequential Labeling

- Output tags: B-Timex2, I-Timex2, O
- IOB2 encoding (Sang & Veenstra, 1999)
- Lexical features include word, shape, is year, is date of week, is month, is number, is time, is day, is quarter, is punctuation, if belong to word-list like init-list7, follow-list8

O	O	O	B-Timex2	I-Timex2	O
|	|	|	|	|	|
Elections	**are**	**on**	**November**	**2nd**	**.**

Elections are on <TIMEX2> November 2nd </TIMEX2> .

- Several other approaches have been attempted but, so far, the extraction step seems simple enough and rule based systems perform besetfollow-list8

Rule-based Normalization (Ahn et al., 2005)

- Lexical lookup: mapping names to numbers, units to ISO values, etc.
- Context-independent composition: combining the values of the lexical tokens within a timex to produce a context-independent semantic representation
- Context-dependent classification: determining whether a timex is a point or duration, looks forward or backward, makes specific or generic reference, etc.
- Reference time, or temporal focus, tracking: for anaphoric timexes, whose values must be computed with respect to a reference time
- Final computation: combining the results of all of these steps to produce a final normalized value

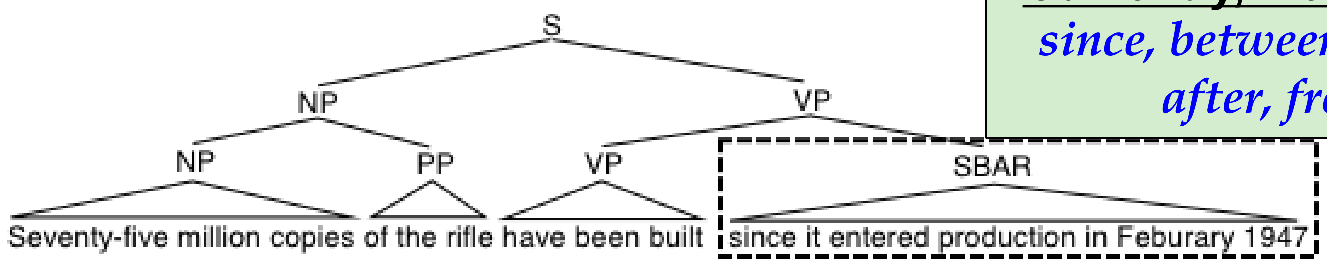
Rules for “today”

- “today” has a possessive inflection?
- “today” is inside of a quotation?
- “said”/”will”/”even”/”most”/… in sentence?
- “year” in same sentence?
- CCYY (4-digit year)/DOW in same sentence?
- POS_{before} “today” POS_{after}
- ………

An Extended Approach Extraction (Do et al., 2012)

- Built on top of a state-of-the-art temporal expression extractor[1] to extract *basic expressions*.
  - For example: ***February 1947***
- Extends the basic extractor to capture *complex expressions* by using full syntactic parse tree.
  - For example: ***since […] February 1947***

Currently, we capture:
since, between, before, after, from

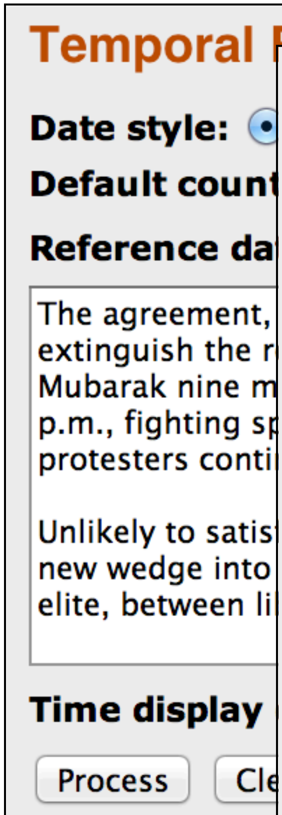


- Normalized to canonical *absolute time intervals* [*start point*, *end point*]
- Compared the normalized intervals by directly comparing their endpoints: *before, before-n-overlap, contain, equal, after, after-n-overlap*

---

[1] We used the **HeidelTime** package: http://code.google.com/p/heideltime/
This system achieved the best performance in the extraction task in TempEval-2 (2010)

An Extended Approach Extraction (Do et al., 2012)



Temporal Reasoning System

Text: The agreement, which centered on a presidential election by next June, appeared unlikely to extinguish the resurgent protest movement the largest since the ouster of President Hosni Mubarak nine months ago. The crowd roared its disapproval when the deal was announced at 8 p.m., fighting spiked on the avenue leading to the Interior Ministry, and the number of protesters continued to swell. Unlikely to satisfy the public demands for the military to leave power, the deal may have driven a new wedge into the opposition, reopening a divide between the seething public and the political elite, between liberals and Islamists and, as events unfolded, among the Islamists themselves.

Reference Date (aka Document Creation Time -- DCT): 2006-05-24

Extracted Expression	Normalized Interval	Relation to DCT
next June	2006-06-01T00:00:00.000/2006-06-30T23:59:59.000	after
since the ouster of President Hosni Mubarak nine months ago	2005-08-01T00:00:00.000/2006-05-24T00:00:00.001	before and overlap
nine months ago	2005-08-01T00:00:00.000/2005-08-31T23:59:59.000	before
8 p.m.	2006-05-24T20:00:00.000/2006-05-24T20:00:00.000	inside

Demo URL: http://cogcomp.cs.illinois.edu/demo/tempdemo

Evaluation

HeidelTime: *High Qualitiy Rule-based Extraction and Normalization of Temporal Expressions*, [Jannik Strötgen and Michael Gertz: HeidelTime, SemEval'10]

- Data: TempEval'10 data sets derived from TimeBank.
- Precision: 90%
- Recall: 82%
- F_1: 86%

IllinoisTime: *A Robust Shallow Temporal Reasoning System*, [Ran Zhao, Quang Do and Dan Roth, NAACL'12 Demo]

- 486 sentences from 183 articles in TimeBank 1.2, which contain at least one of the five temporal connectives *since*, *between*, *from*, *before*, *after*

Extractor

Connective	# sent.	# appear.	Prec.	Rec.	F_1
since	31	31	1.0	1.0	1.0
between	32	33	1.0	1.0	1.0
from	340	366	0.8	1.0	0.89
before	33	33	0.8	1.0	0.89
after	78	81	0.72	1.0	0.84
Average			0.86	1.0	0.92

Normalizer & Comparator

Module	Correct	Incorrect	Accuracy
Normalizer	191	16	0.92
Comparator	191	0	1.0

Outline

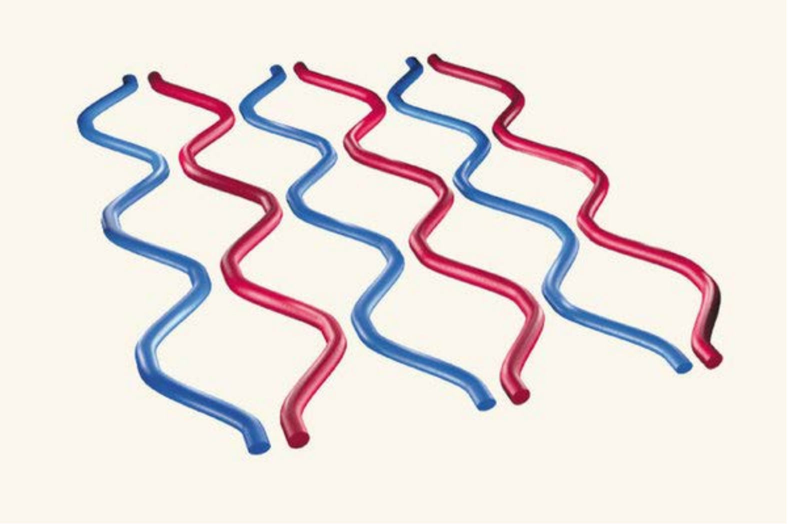
1. Background: Motivations and Goals
2. Temporal Information Representation Theories
3. Temporal Expression Extraction and Normalization
4. Temporal Slot Filling

5. Tea Break

6. Event Timelining and Temporal Reasoning
7. Resources and Demos
8. Conclusions

10:30

- Temporal Slot Filling



Temporal Slot Filling

- Task Definition
- Approach Overview
- Annotation Challenges and Solutions
  - Distant Supervision and its Problems
  - Multi-layer Annotations
  - Global Time Discovery
  - Feature Reduction and Instance Re-labeling
  - Multi-instance Multi-class Learning
  - Pattern Re-weighting
- Temporal Classification Challenges and Solutions
  - Capturing Long Contexts
  - Flat Approach
  - Structured Approach

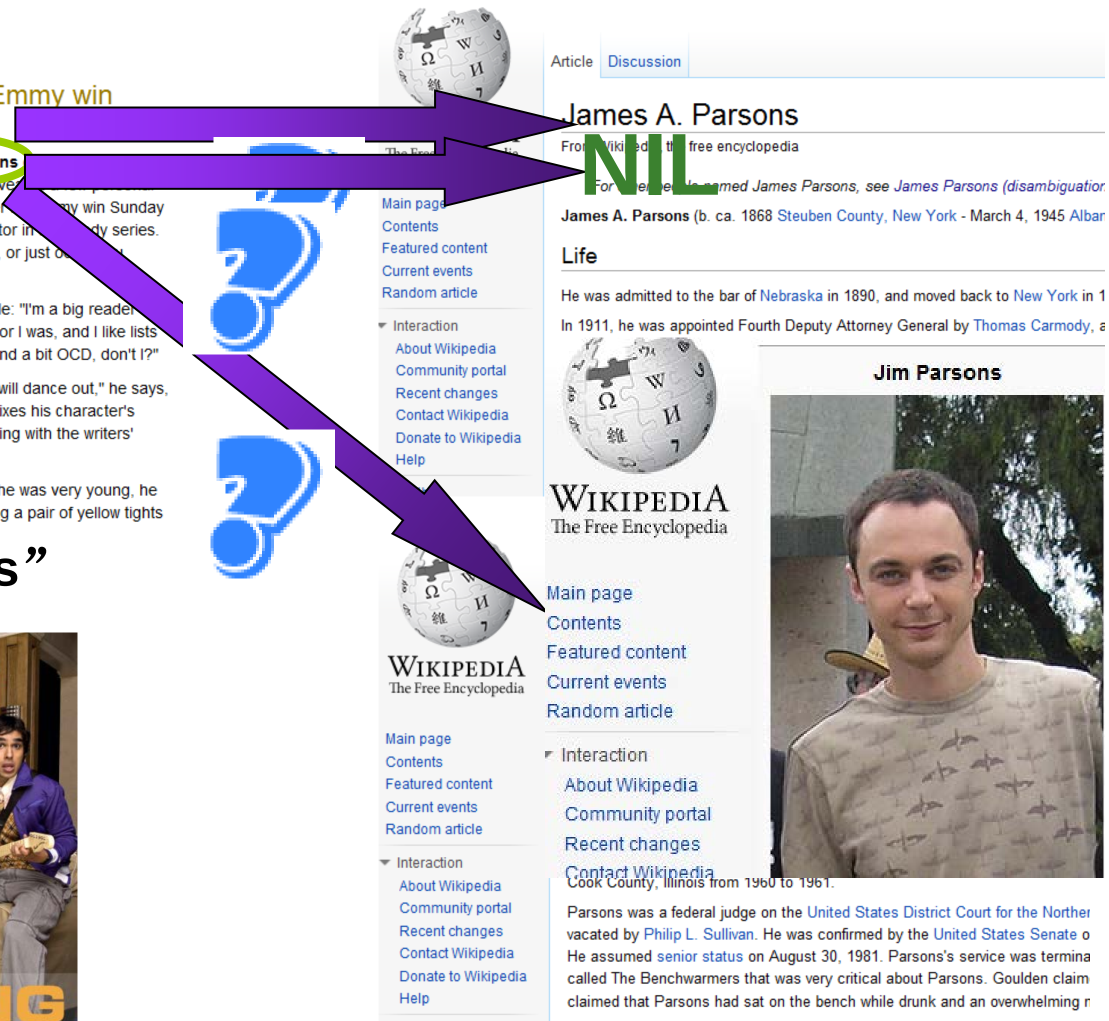
Common Information Extraction (IE) Bottleneck

- One of the initial goals for IE was to create a knowledge base (KB) from the entire input corpus, such as a profile or a series of activities about any entity, and allow further logical reasoning on the KB
- Such information may be scattered among a variety of sources (large-scale documents, languages, genres and data modalities)
- Problem: the KB constructed from a typical IE pipeline often contains lots of erroneous and conflicting facts
  - Single-document event extraction < 70%; Cross-document slot filling < 30%; worse for non-newswire genres, languages, multimedia data
- Improve Quality of IE: Identify topically-related documents and to integrate facts, possibly redundant, possibly complementary, possibly in conflict, coming from these documents

Knowledge Base Population (KBP)

- General Goal
  - Promote research in discovering facts about entities to create and expand a knowledge source automatically
- What's New
  - Extraction at large scale (> 3 million documents)
  - Using a representative collection (not selected for relevance)
  - Cross-document entity resolution (extending the limited effort in ACE)
  - Linking the facts in text to a knowledge base
  - Support multi-lingual information fusion
  - Capture temporal information – Temporal Slot Filling

Entity Linking



Query = "James Parsons"

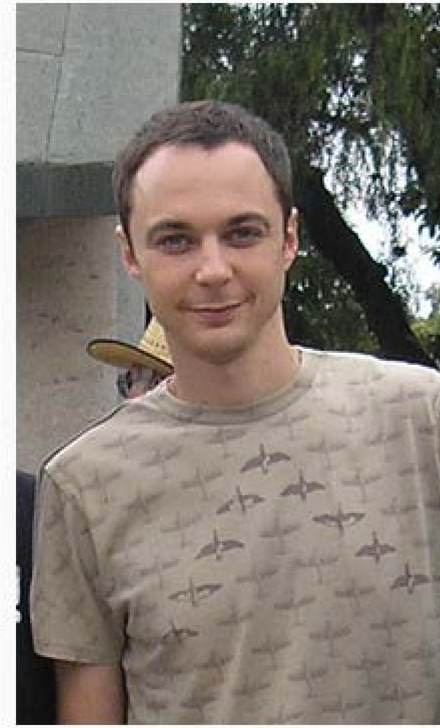
(Chen and Ji, EMNLP2011)

Slot Filling

Jim Parsons, a graduate of the University of Houston School of Theater and Dance, won the Emmy on Sunday for Lead Actor in a Comedy Series for his work on The Big Bang Theory.

School Attended: University of Houston

Jim Parsons



Parsons in 2008

Born	James Joseph Parsons March 24, 1973 (age 37) Houston, Texas, U.S.
Occupation	Actor
Years active	2000–present

Time-intensive Slot Types

Person		Organization
per:alternate_names	**per:title**	org:alternate_names
per:date_of_birth	per:member_of	org:political/religious_affiliation
per:age	**per:employee_of**	**org:top_members/employees**
per:country_of_birth	per:religion	org:number_of_employees/members
per:stateorprovince_of_birth	**per:spouse**	org:members
per:city_of_birth	per:children	org:member_of
per:origin	per:parents	org:subsidiaries
per:date_of_death	per:siblings	org:parents
per:country_of_death	per:other_family	org:founded_by
per:stateorprovince_of_death	per:charges	org:founded
per:city_of_death		org:dissolved
per:cause_of_death		org:country_of_headquarters
per:countries_of_residence		org:stateorprovince_of_headquarters
per:stateorprovinces_of_residence		org:city_of_headquarters
per:cities_of_residence		org:shareholders
per:schools_attended		org:website

Temporal Slot Filling (TSF)

- Given a query entity, a knowledge base (KB) and a source corpus, a system must return slot fills and temporal information must be gathered across the entire corpus
- Query Example

<query id="SFT201">
<name>Angela Merkel</name>
<docid>NYT_ENG_20071015.0123.LDC2009T13</docid>
<enttype>PER</enttype>
<nodeid>E0288830</nodeid>
</query>

- Output Example

SFT201 per:countries_of_residence T2 20051231 AFP_ENG_20081022.0383 Germany
SFT201 per:countries_of_residence T3 20081022 AFP_ENG_20081022.0383 Germany
SFT201 per:spouse T1 19980101 APW_ENG_20051122.0372.LDC2007T07 Joachim Sauer
SFT201 per:spouse T2 19981231 APW_ENG_20051122.0372.LDC2007T07 Joachim Sauer
SFT201 per:spouse T3 20051122 APW_ENG_20051122.0372.LDC2007T07 Joachim Sauer

Scoring Metric

- Temporal Quality
  - Let <t_1, t_2, t_3, t_4> be system output, <g_1, g_2, g_3, g_4> be gold standard

$$Q(S) = \frac{1}{4} \sum_i \frac{c}{c + |t_i - g_i|}$$

  - An error of *c* time units produces a 0.5 score; scores produced with c = 1 year
  - Each element in tuple is scored independently
  - For temporal SF task, a correct slot fill with temporal information ***t*** gets credit Q(S) (instead of 1)
- Overall Metric

$$P = \frac{\sum_{S^i \in C(S)} Q(S^i)}{M} \qquad R = \frac{\sum_{S^i \in C(S)} Q(S^i)}{N}$$

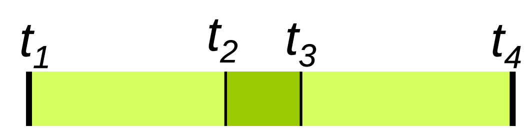
  - *M*: the number of system output tuples
  - *N*: the number of gold standard tuples
  - C(S): the number of instances that have correct slot fills

Evaluation Metric and Formal Constraints (Cont')

- Parameterization Constraint
  - A **parameter should determine if a certain amount of *vagueness*** is worse/better than a certain amount of *over-constraining*

$$c_i = \begin{cases} c_{vag}, if\ (i \in \{1,3\} \wedge t_i \le g_i) \vee (i \in \{2,4\} \wedge t_i \ge g_i) \\ c_{cons}, otherwise \end{cases}$$

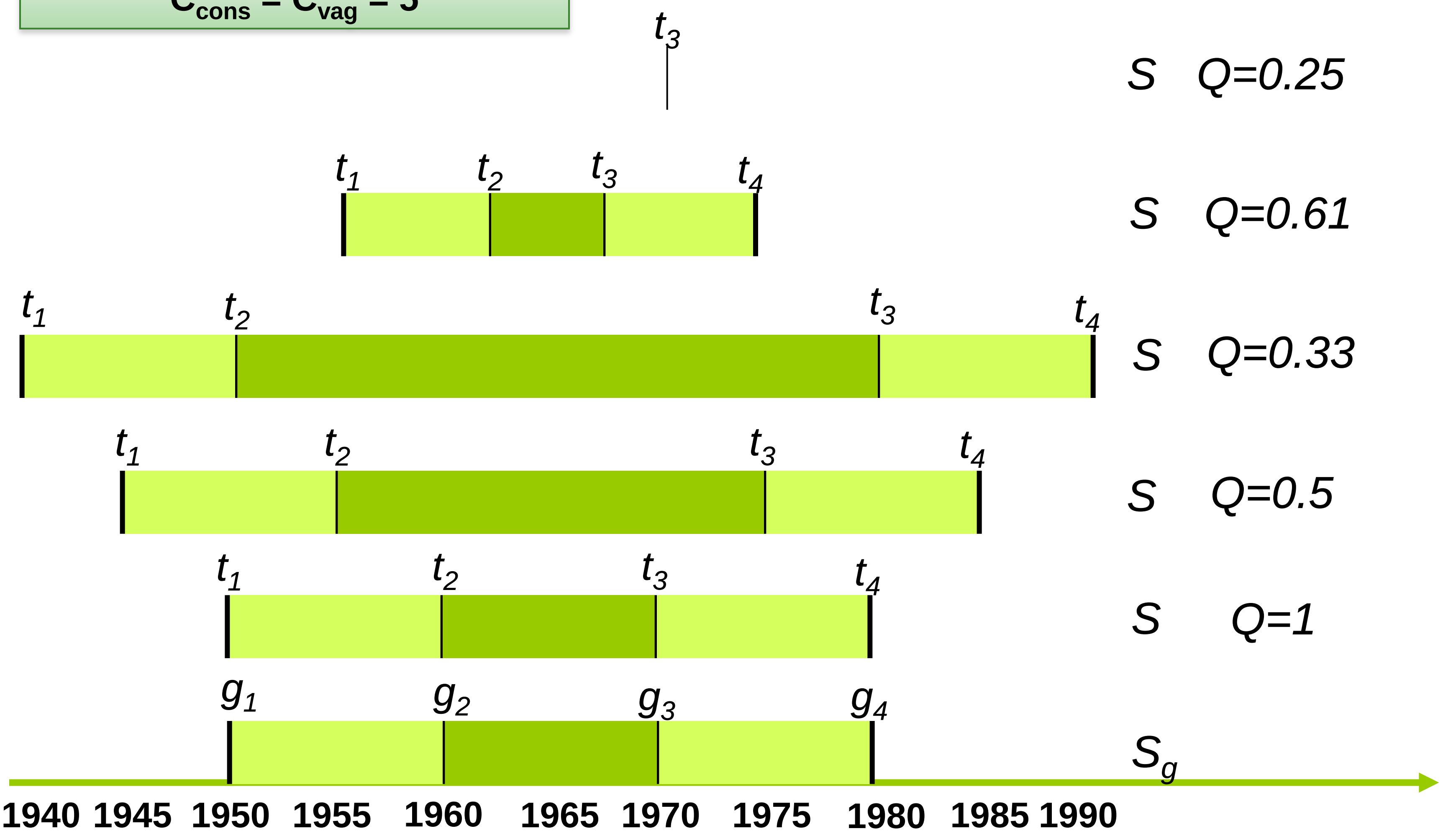
- If $S \subset S_g$: $Q(S) = \frac{1}{4}\sum_i \frac{c_{cons}}{c_{cons} + |t_i - g_i|}$



- If $S_g \subset S'$: $Q(S') = \frac{1}{4}\sum_i \frac{c_{vag}}{c_{vag} + |t_i - g_i|}$

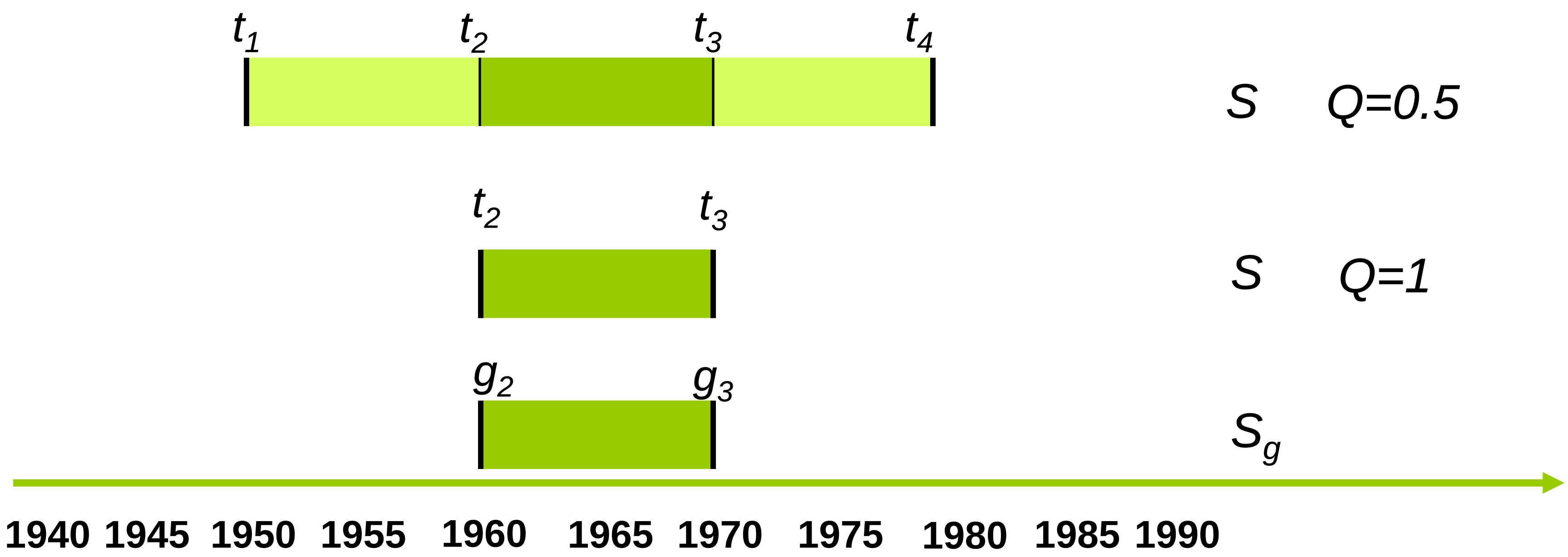
Evaluation Examples

Infinite = 10000 -Infinite = 0
$C_{cons} = C_{vag} = 5$

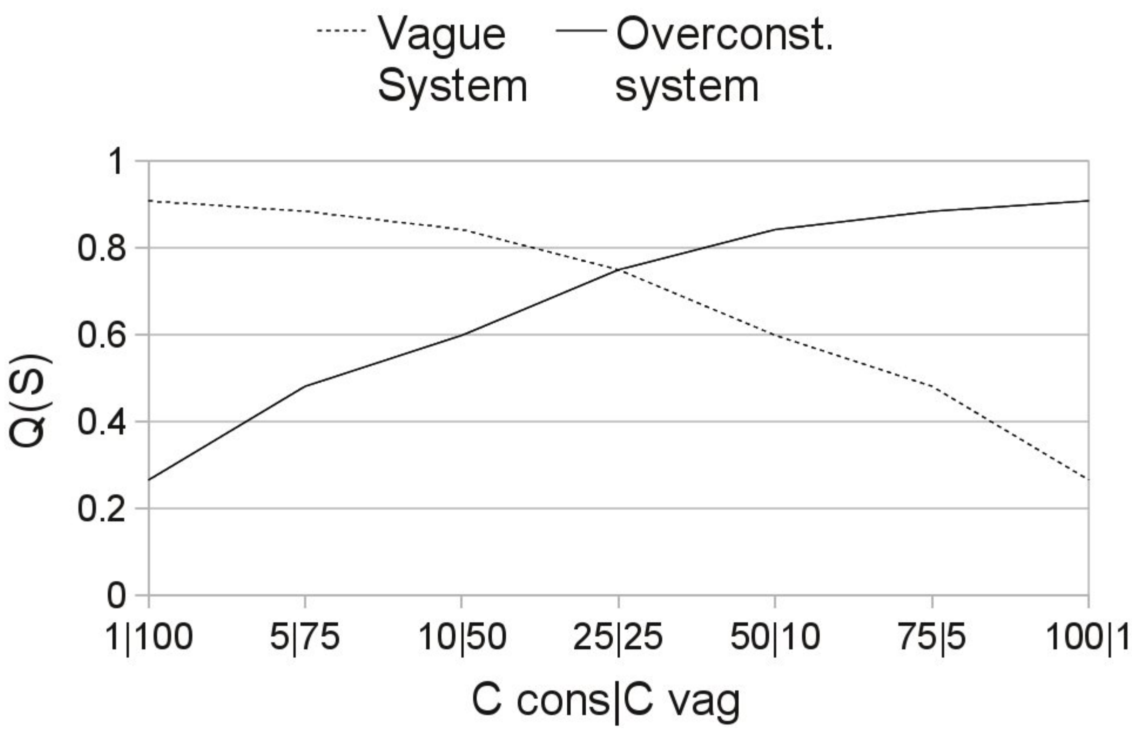


Evaluation Examples

Infinite = 10000 -Infinite = 0
$C_{cons} = C_{vag} = 5$

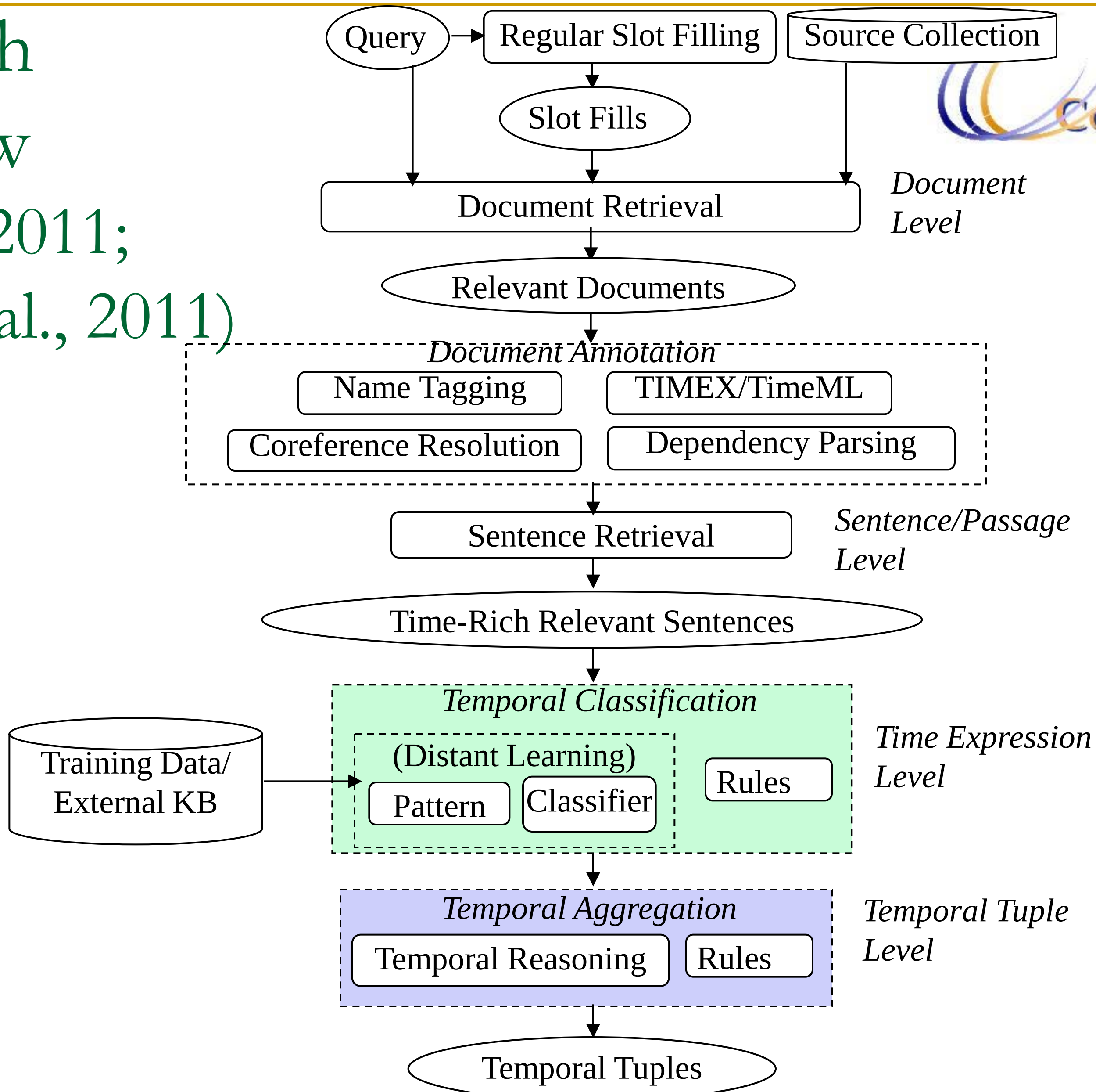


Parameterization



- Randomly generate 100 gold standard tuples
- Randomly generate *Over-cons* tuple and Vague tuple for each gold standard tuple by adding the same offsets to each element
- Our metric is able to differentiate *Over-cons* and Vagueness by using different $\mathbf{C}_{cons}$ and $\mathbf{C}_{vag}$

Approach Overview (Li et al., 2011; Artiles et al., 2011)



Regular Slot Filling

- (Chen et al., 2010; Tamang and Ji, 2011)
- Query expansion based on templates and Wikipedia links
- Pattern Learning
  - Selection of query-answer pairs from Wikipedia Infobox
    - split into two sets
  - Pattern extraction
    - For each {*q*,*a*} pair, generalize patterns by entity tagging and regular expressions e.g. <*q*> died at the age of <*a*>
  - Pattern assessment
    - Evaluate and filter based on matching rate
  - Pattern matching
    - Combine with coreference resolution
  - Answer filtering based on entity type checking, dictionary checking and dependency parsing constraint filtering

Regular Slot Filling (Cont')

- Automatic Content Extraction (ACE) Information Extraction
  - Apply ACE Cross-document IE (Ji et al., 2009)
  - Mapping ACE to KBP
- Question Answering
  - Apply open domain QA system, OpenEphyra (Schlaefer et al., 2007)
  - Relevance metric related to PMI and CCP

    P (q, a) = P (q NEAR a): NEAR within the same sentence boundary

    $$R(q,a) = \frac{freq(q\ NEAR\ a)}{freq(q) \times freq(a)} \times \#sentences$$
  - Heuristic rules for Answer Filtering

Regular Slot Filling (Cont')

- Maximum Entropy (MaxEnt) based supervised re-ranking model to re-rank candidate answers for the same slot
- Low-Transparency Features
  - System and Slot Type: identifies the system of origin and the slot type
  - Number of Tokens and Slot Type: the number of tokens in the answer by the slot type
  - Answer Frequency
- High-Transparency Features
  - Answer Name Type: the name type of the candidate answer
  - Dependence Parse and its length
  - Trigger Words: if a slot type related trigger word is in the system provided context sentence
  - Comma Delimited List: if the context sentence is a long comma delimited list
  - Query Subset of Answer: if the query is a subset of the answer
  - Invalid Answer: if an answer is listed in set of predefined invalid answers (e.g., \the" or \city")
  - Date/Age/Number Validation
  - Country, City, Nationality and Title Validation with gazetteer
  - …

Temporal Classification

- *In **1975**, after being fired from Columbia amid allegations that he used company funds to pay for his son's bar mitzvah, **Davis** founded **Arista***
  - Is ‘1975’ related to the employee_of relation between Davis and Arista?
  - If so, does it indicate START, END, HOLDS… ?
- Each classification instance represents a temporal expression in the context of the entity and slot value.
- We consider the following classes
  - START *Rob joined Microsoft in 1999.*
  - END *Rob left Microsoft in 1999.*
  - HOLDS *In 1999 Rob was still working for Microsoft.*
  - RANGE *Rob has worked for Microsoft for the last ten years.*
  - NONE *Last Sunday Rob’s friend joined Microsoft.*

Temporal Aggregation

- What is the best way to combine a set of classified temporal expressions in a 4-tuple?
  - Individual classifications can be in conflict
  - Temporal classifier makes mistakes
- A temporal expression is normalized, classified and mapped to a 4-tuple
  - START $<T_a, T_b, T_a, +INF>$
  - END $<-INF, T_b, T_a, T_b>$
  - HOLDS $<-INF, T_a, T_b, +INF>$
  - RANGE $<T_a, T_b, T_a, T_b>$
  - NONE $<-INF, INF, -INF, INF>$
- Iterative aggregation (Li et al., 2012)

$$T \wedge T' = < \max(t_1, t_1'), \min(t_2, t_2'), \max(t_3, t_3'), \min(t_4, t_4') >$$

- Aggregation with global constraints (McClosky and Manning, 2012)

Annotation Challenges

- Manually annotated data is not enough (1,172 instances for 8 slot types)

	Total	Start	End	Holds	Range	Others
Spouse	28	10	3	15	0	9
Title	461	69	42	318	2	30
Employee	592	111	67	272	6	146
Residence	91	2	9	79	0	1

- Moderate inter-annotator agreement (pairwise Cohen's Kappa of 0.57)
- Many simple but useless sentences/patterns:
  - ``Tom LaSorda, president and CEO, Sept. 2005-Aug. 2007
  - Dieter Zetsche, president and CEO, Nov. 2000- Sept. 2005
  - …
  - Eugene A. Cafiero, president, Oct. 1975-Nov. 1978"

Annotation Challenges (Cont')

- Explicit temporal information is very sparse and scattered across documents
  - 35 of the 107 KBP queries with "employee_of" answers have documents including temporal arguments
  - 1/3 queries could be reliably associated with either a start or end date
  - On average 518 relevant documents returned for <entity, slot>, but only 21 sentences returned for <entity, slot, temporal expression>

Solution: Distant Supervision (Mintz et al., 2009)

Retrieve events from a Knowledge Base (Freebase)

Slot type: per_spouse
<John, Mary, 1997, 2003>

Collect top web search results for
"Entity A" "Entity B"

www.mary-and-john.com
www.mymarriage.com/john33/
www.wedding-photos.com/joma/
...

Collect all sentences that mention the two entities.

On ***1997****,* ***John*** *and* ***Mary*** *renewed their vows in Florida.*

Label to each temporal expression.
We compare with the corresponding information in the Knowledge Base

	START---	HOLDS---	END
KB.:	1997		2003
Sent.:	*1997*		

Distant Supervision Results

Category	Type	Total	Start	End	Holds	Range	Others
Spouse	Manual	28	10	3	15	0	9
	Automatic	**10,196**	**2,463**	**716**	**1,705**	**182**	**5,130**
Title	Manual	461	69	42	318	2	30
	Automatic	**14,983**	**2,229**	**501**	**7,989**	**275**	**3,989**
Employment	Manual	592	111	67	272	6	146
	Automatic	**17,315**	**3,888**	**965**	**5,833**	**403**	**6,226**
Residence	Manual	91	2	9	79	0	1
	Automatic	**4,168**	**930**	**240**	**727**	**18**	**2,253**

- We obtained over 50k training instances with no human intervention

Advantages of Distant Supervision

- Diverse contexts that can be captured
  - Common patterns

    Alexander and Susan married on Jan. of 2005.

  - Less common patterns

    On September 2002 Mary Jones and John Smith eloped on the SF bay.

  - Implied information…

    After John's death in 2003, Mary fought with his children over the distribution of the Senator’s state.

More Annotation Challenges

- Each knowledge base entry includes temporal information in addition to an entity, slot fill pair
- Noise can be introduced when the unlabeled data is collected from the Web:
  - Coreference errors yield incorrect name matching
  - Temporal expressions are normalized incorrectly
  - Temporal information with different granularities

    "John married Mary in 1997" → "September 3, 1997" as a START?
  - Knowledge base may contain incorrect or contradictory information with Web documents
- Over 100,000 features are required to generalize the complicated contexts for each slot type, data sparsity → Learning supervised models becomes unfeasible

Wrong Assumptions

- One sense per query

  Raul Castro set the date for local (city and town) *general* elections as *October 21* with a second round *October 28*.

- One query per context

  Slow Club 's *Chris Kronner* faced similar challenges taking on his second executive chef position at Serpentine , which opened in *December*.

- One sentence per query
  - Applied at the sentence-level, e.g. assume three elements in the tuple should exist in the same sentence after entity coreference resolution
  - Invalid when a document is typically talking about a centroid entity (e.g. the employment history of a person or an organization)

Solutions 1: Multi-layer Annotations (Artiles et al., 2011)

- Document segmentation for fine-grained reference date extraction
  - **"Aug. 6, 2007**: Bob Nardelli appointed Chrysler chairman and CEO. "
- Multi-layer annotations to expand relevant sentence matching
  - name tagging
  - **co-reference chains**
  - dependency parses

Solution 2: Global Time Discovery

- Half of the event instances don’t include explicit time arguments
- Prior work of detecting implicit time arguments
  - Filatova and Hovy, 2001; Mani et al., 2003; Lapata and Lascarides, 2006; Eidelman, 2008
  - Most work focused on sentence level
  - Linguistic evidence such as verb tense was used for inference
- Cross-event Inference (Gupta and Ji, 2009)
  - More fine-grained events
  - An event mention and all of its coreferential event mentions do not include any explicit or implicit time expressions

Observations about Events in News

- Based on series of events
  - Various situations are evolving, updated, repeated and corrected in different event mentions
- Events occur as chains
  - Conflict→Life-Die/Life-Injure
  - Justice-Convict → Justice-Charge-Indict/Justice-Trial-Hearing
- Writer won't mention time repeatedly
  - To avoid redundancy, rarely provide time arguments for all of the related events
- Reader is expected to use inference
  - On Aug 4 there is fantastic food in Suntec…Millions of people came to attend the IE session. → the IE session is on Aug 4

Background Knowledge Reasoning

- Time Search from Related Documents

  [*Test Sentence*]

  <entity>Al-Douri</entity> said in the <entity>AP</entity> interview he would love to return to teaching but for now he plans to remain at the United Nations.

  [*Sentences from Related Documents*]

  In an interview with <entity>The Associated Press</entity>

  <time>Wednesday<time> night, <entity>Al-Douri</entity> said he will continue to work at the United Nations and had no intention of defecting.

- Time Search from Wikipedia

  [*Test Sentence*]

  <person>Diller</person> started his entertainment career at <entity>ABC</entity>, where he is credited with creating the ``movie of the week'' concept.

  [*Sentences from Wikipedia*]

  <person>Diller</person> was hired by <entity>ABC</entity> in <time>1966</time> and was soon placed in charge of negotiating broadcast rights to feature films.

Time Propagation between Events

Event Mention with time

Injured Russian diplomats and a convoy of America's Kurdish comrades in arms were among unintended victims caught in crossfire and friendly fire [Sunday]

Event Mention without time

Kurds said 18 of their own died in the mistaken U.S. air strike.

Event Mention with time

A state security court suspended a newspaper critical of the government [Saturday] after convicting it of publishing religiously inflammatory material.

Event Mention without time

The sentence was the latest in a series of state actions against the Monitor, the only English language daily in Sudan and a leading critic of conditions in the south of the country, where a civil war has been waged for 20 years.

(Gupta and Ji, 2009)

Inference Rules

- Same-Sentence Propagation
  - EM_i and EM_j are in the same sentence and only one time expression exists in the sentence
- Relevant-Type Propagation
  - $type_i$= "Conflict", $type_j$= "Life-Die/Life-Injure"
  - arg_i is coreferential with arg_j
  - $role_i$="Target" and $role_j$="Victim", or $role_i$=$role_j$= "Instrument"
- Same–Type Propagation
  - arg_i is coreferential with arg_j, $type_i$= $type_j$, $role_i$= $role_j$, and match ***time-cue roles***

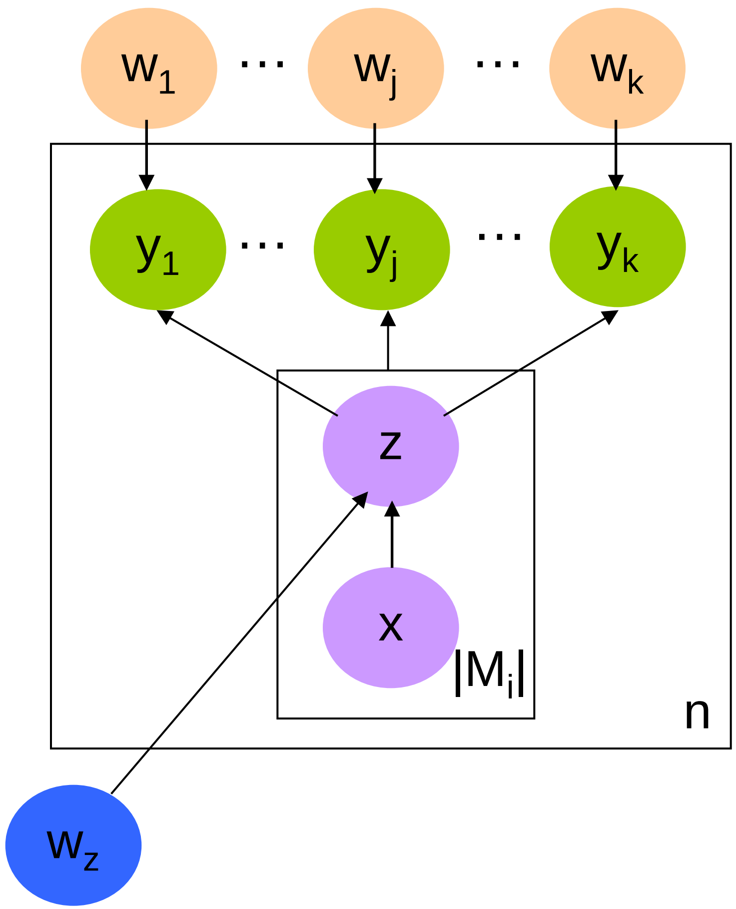
$Type_i$	$Role_i$	$Type_i$	$Role_i$
Conflict	Target/Attacker/Crime	Movement-Transport	Destination/Origin
Justice	Defendant/Crime/Plaintiff	Transaction	Buyer/Seller/Giver/Recipient
Life-Die/Life-Injure	Victim	Contact	Person/Entity
Life-Be-Born/Life-Marry/Life-Divorce	Person/Entity	Personnel	Person/Entity
		Business	Organization/Entity

- Results: 72.2% F-measure

Solutions 3: Feature Reduction and Instance Re-labeling (Tamang and Ji, 2012)

- Feature Reduction
  - Test the statistical significance of each feature as a predictor for a temporal class label using multi-class *logistic regression*
  - Create a minimal feature set
- Relabeling
  - Approximately 0.01% of the distant supervision data for each slot was labeled
  - *Lasso regression* was used to classify the unlabeled instances using self-training techniques

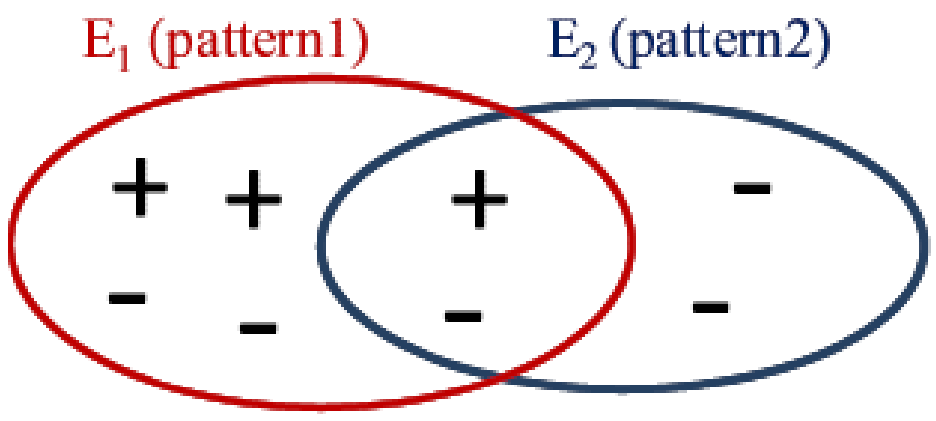
Solutions 4: Multi-instance Multi-label Learning (Surdeanu et al., 2012)



- n: the number of distinct tuples in knowledge base
- M_i: the set of mentions for the i^{th} entity pair
- x: a sentence
- z: latent relation classification for x
- w_z: weight for muli-class mention-level classifier
- y_j: top-level classification decision as to whether the j^{th} relation holds
- w_j: weight vector for binary top-level classifier for the j^{th} relation
- Training based on Expectation Maximization (EM)

Solutions 5: Pattern Re-weighting (Takamatsu et al., 2012)

- A generative model to predict whether each pattern expresses each relation via hidden variables
- Remove wrong labels using the negative pattern list



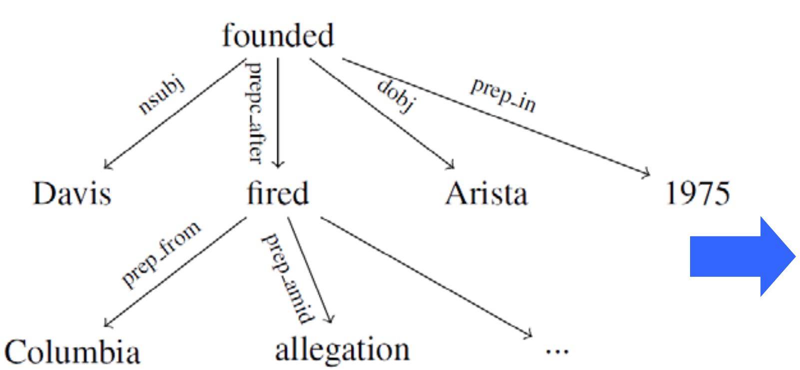
- Probability of instances in E1 expressing pattern 1 = 3/6=0.5
- Probability of instances in E2 expressing pattern 2 = 3/6*2/4=0.25

Temporal Classification Challenges

- Problems
  - Capture long contexts
- Solutions
  - Use parsing structures to *compress* long contexts
  - Core NLP annotation tools (e.g. dependency parsing, coreference) are far from perfect, not robust enough
  - Tradeoff between flat representation and structured representation

Structured Representation Approach

- Representation based on three shortest dependency paths
  In ***1975****, after being fired from Columbia amid allegations that he used company funds to pay for his son's bar mitzvah,* ***Davis*** *founded* ***Arista.***
- Surface sentence:
  - Long distance between *1975* and *Davis founded Arista*
  - Some words in between cause ambiguity: fired
- Dependency paths:
  - Help remove irrelevant information
  - Build syntactic and semantic links from long distance



$$\text{Davis} \xleftarrow{\text{nsubj}} \text{founded/found/VBD/Start-Position} \xrightarrow{\text{prep\_in}} 1975$$

$$\text{Arista} \xleftarrow{\text{dobj}} \text{founded/found/VBD/Start-Position} \xrightarrow{\text{prep\_in}} 1975$$

$$\text{Davis} \xleftarrow{\text{nsubj}} \text{founded/found/VBD/Start-Position} \xrightarrow{\text{dobj}} \text{Arista}$$

Structured Representation Approach

- Kernel function on two paths: enumerate all sub-patterns in two paths

$$K_p(P_x, P_y) = \sum_{k=1}^{Min(l(P_x),l(P_y))} \sum_{a \in P_x[k], b \in P_y[k]} \prod_{i=1}^{k} c(a_i, b_i)$$

  - Count number of common substrings
  - *a* is any substring of P_x with length *k*
  - $c(a_i, b_i)$ is inner product of feature vector of nodes a_i and b_i
- Kernel function on two sentences: combine kernel values of three paths

$$K_s(x, y) = \sum_{i=1}^{3} K_p(x.P_i, y.P_i)$$

- Normalization: avoid bias towards long paths

$$K_p'(P_x, P_y) = \frac{K_p(P_x, P_y)}{\sqrt{K_p(P_x, P_x) \cdot K_p(P_y, P_y)}}$$

Flat Representation Approach

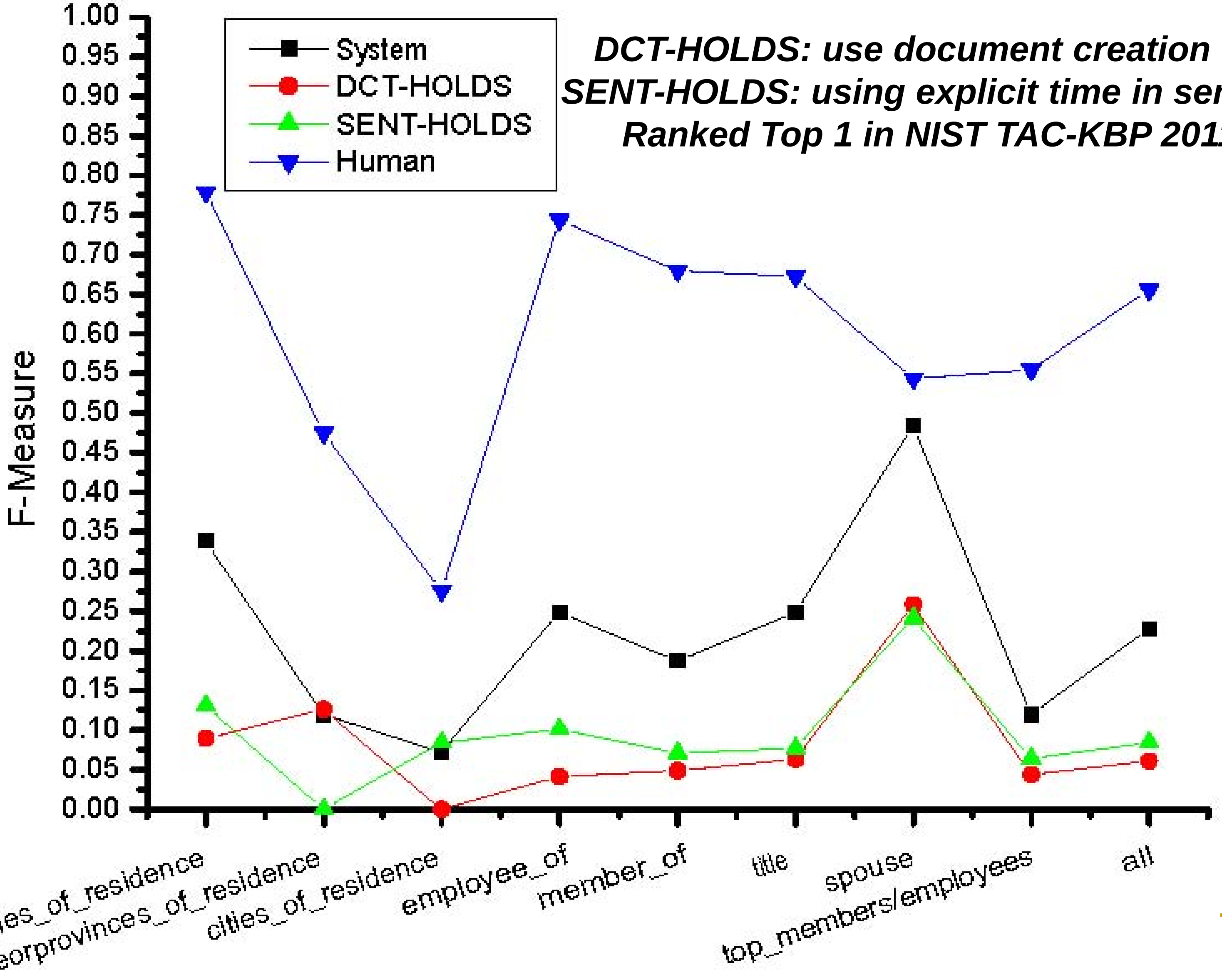
- Window of words around TARGET_DATE, TARGET_ENTITY and TARGET_ATTRIBUTE.
- Shallow Dependency Features: governor and dependent words of the target entity, date and attribute
- Sentences are normalized for the specific query (entity and attribute) and target date.
  - *In **1975**, after being fired from Columbia amid allegations that he used company funds to pay for his son's bar mitzvah, **Davis** founded **Arista.***
  - *In **TARGET_DATE**, after being fired from **ORGANIZATION** amid allegations that **TARGET_ENTITY** used company funds to pay for **TARGET_ENTITY** son's bar mitzvah, **TARGET_ENTITY** founded **TARGET_ATTRIBUTE**.*

Evaluation Data Set

- KB: October 2008 dump of English Wikipedia, 818,741 nodes
- Source Corpus: includes 1,286,609 newswire documents, 490,596 web documents and hundreds of transcribed spoken documents
- 100 queries, 80 person entities and 20 organization entities
- Gold-standard creation: pooled the responses from all the systems and human annotators; human assessors judged the responses

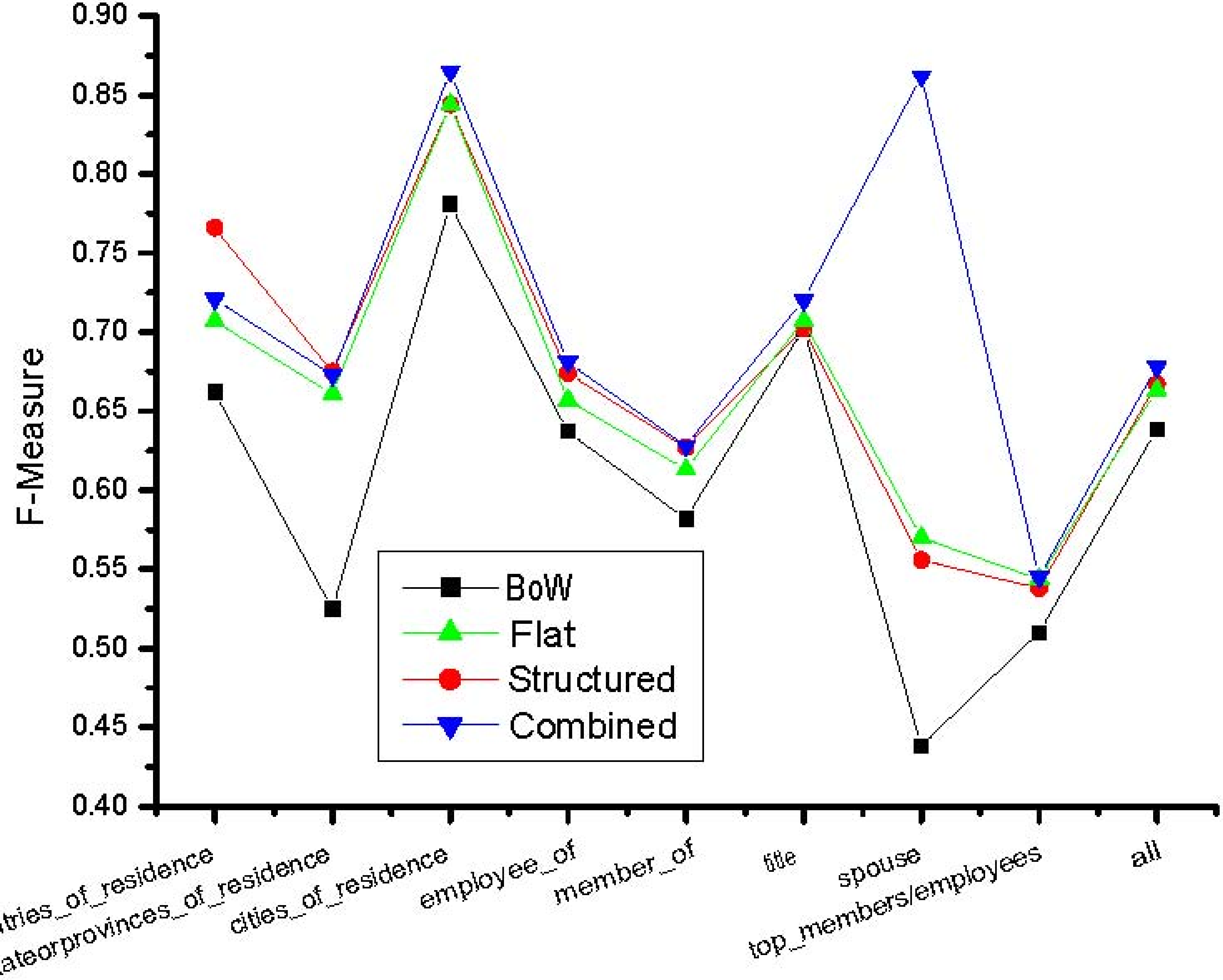
Slot Type	# of Tuples
per:countries_of_residence	**287**
per:statesorprovinces_of_residence	**44**
per:cities_of_residence	**109**
per:member_of	**86**
per:employee_of	**20**
per:title	**89**
per:spouse	**52**
org:top_members/employees	**24**
total	**711**

Overall Performance



DCT-HOLDS: use document creation time as HOLD
SENT-HOLDS: using explicit time in sentence as HOLD
Ranked Top 1 in NIST TAC-KBP 2011 Evaluation

Comparison on Classification Approaches



Flat works but Structured doesn't

- Structured fails to distill informative pattern from parsed tree
  - Avi Dichter per_employee_of Brookings Institute Sep 2005
    In September 2005, Dichter left office and **became** a research fellow at the Brookings Institute in Washington , D.C.
- Flat classifier: Start

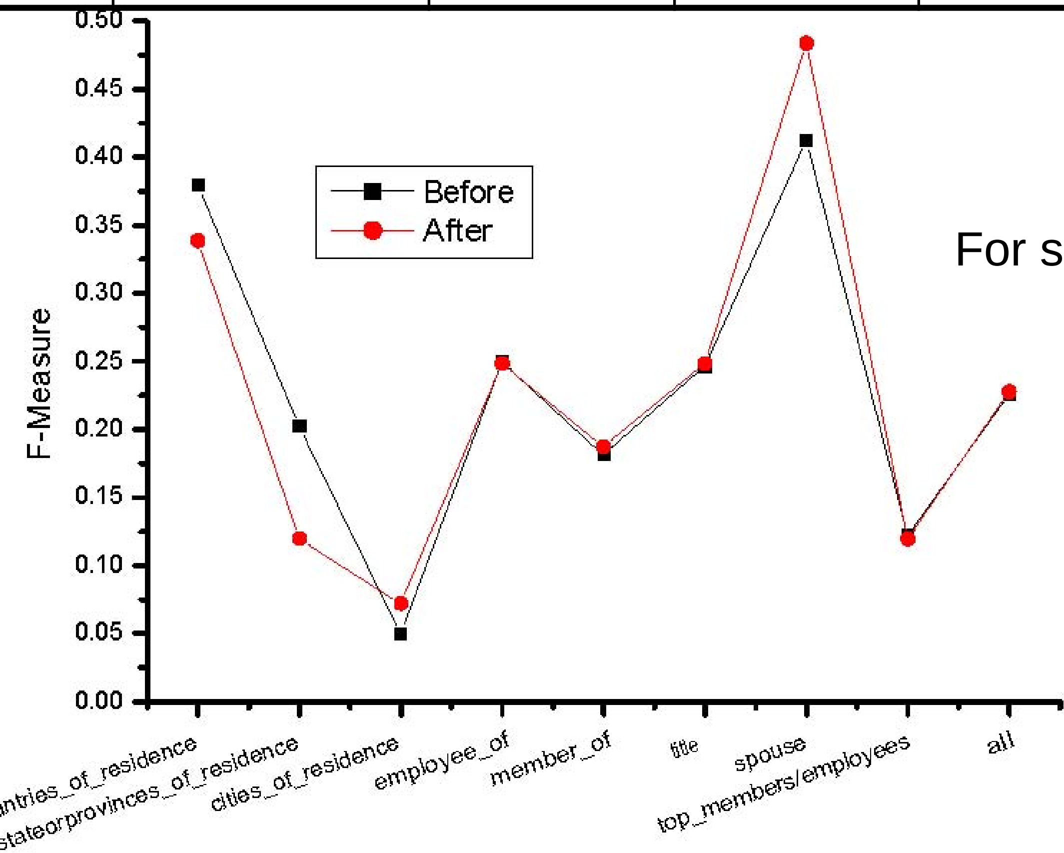
  Structured classifier: None
- Dependency paths are inappropriate:
  Dichter (nsubj)← left →(prep_in) September
  Institute (prep_at)← fellow (conj_and)← left →(prep_in)September
  Dichter (nsubj)← left →(conj_and) fellow →(prep_at)Institute

Structured works but Flat doesn't

- Central Theoretical Council org_top_members_employees Nguyen Phu 2001
  - Trong became secretary of the Hanoi Party Committee in January 2000, chairman of the Central Theoretical Council in 2001, member of the CPVCC in April 2001, and member of the Political Bureau in April 2006.
  - TARGET_ENTITY became TITLE_GAZ of the ORGANIZATION in DATE, TITLE_GAZ of the TARGET_ATTRIBUTE in TARGET_DATE, member of the cpvcc in DATE , and member of the ORGANIZATION in DATE
- Flat classifier : None
- Structured classifier: Start
- Dependency Paths:
  Trong (*nsubj*)← became→(*dobj*) secretary→(*conj_and*) chairman→(*prep_of*) Council→(*prep_in*) 2001
  Council→(*prep_in*) 2001
  Trong (*nsubj*)← became→(*dobj*) secretary→(*conj_and*) chairman→(*prep_of*)Council

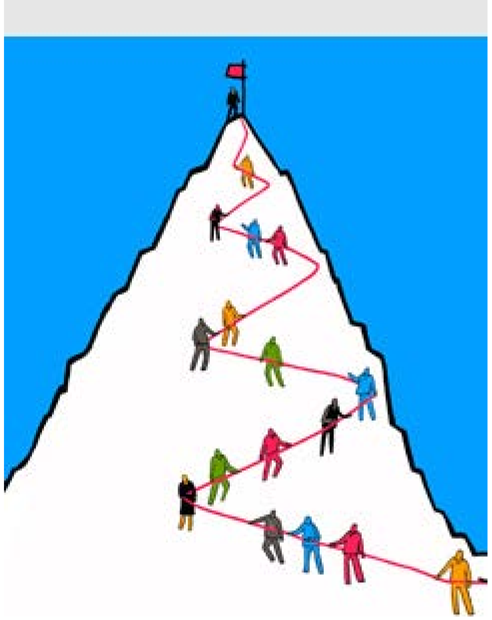
Re-labeling Results

Features	residence	title	spouse	employment
Initial set	10757	31947	40979	51399
Final set	451	2024	1247	2151
Reduction (%)	95.81	93.67	96.96	95.82



For structured approach

- Remaining Challenges



Capture Implicit and Wide Context

- ***Sutil**, a trained pianist, **tested** for **Midland** in 2006 and raced for Spyker in 2007 where he scored one point in the Japanese Grand Prix.*

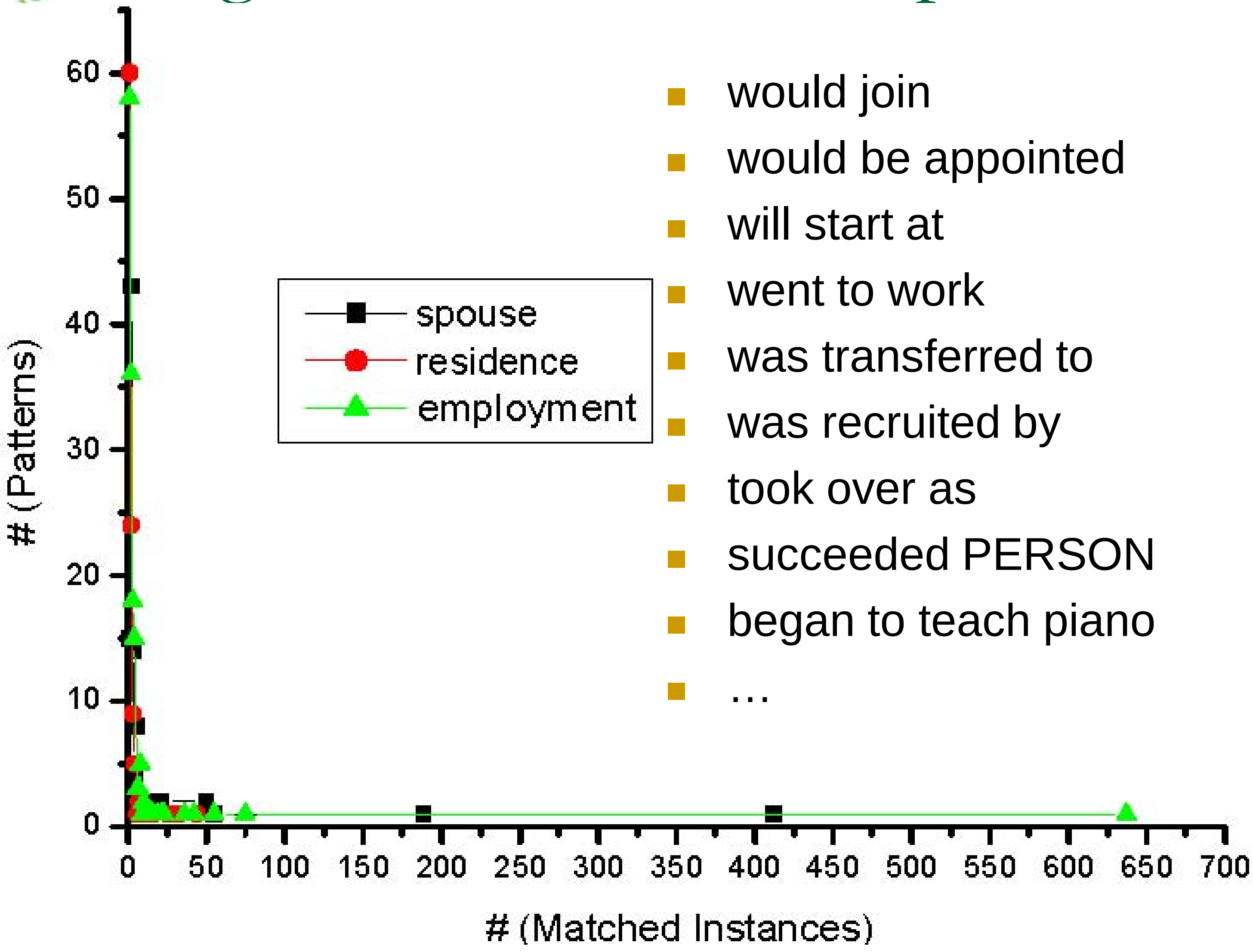
- ***Daimler Chrysler** reports 2004 profits of $3.3 billion; **Chrysler earns** $1.9 billion.*

- *"**Daimler Chrysler** is not yet where we want it to be, but we are headed precisely in the right direction", **Schrempp says**.*

Coreference Resolution Errors

- Name Coreference: "R" = "Republican Party", "Brooklyn Dodgers" = "Brooklyn"
- Nominal Coreference
  - Almost overnight, he became fabulously rich, with a $3-million book deal, a $100,000 speech making fee, and a lucrative multifaceted consulting business, **Giuliani Partners**. As a celebrity rainmaker and lawyer, his income last year exceeded $17 million. His consulting partners included seven of those who were with him on 9/11, and in 2002 Alan Placa, his boyhood pal, went to work at **the firm**.
  - After successful karting career in Europe, Perera became part of the **Toyota** F1 Young Drivers Development Program and was a Formula One test driver for **the Japanese company** in 2006.
  - "Alexandra Burke is out with the video for her second single … taken from the British artist's debut album"
  - "a woman charged with running a prostitution ring … her business, Pamela Martin and Associates"
- Pronoun Coreference
  - Meteorologist **Kelly Cass** became an On-Camera Meteorologist at The Weather Channel, after ***David Kenny*** was named the chairman and chief executive. **She** first appeared on air at The Weather Channel in January 2000.

Long-tail distribution of patterns



- would join
- would be appointed
- will start at
- went to work
- was transferred to
- was recruited by
- took over as
- succeeded PERSON
- began to teach piano
- …

Toward Temporal Reasoning

- ***Sheetrit*** *is a* ***Knesset*** *(parliament) member.*
- ***Sheetrit*** *was born in* ***Morocco*** *and immigrated to* ***Israel*** *in* ***1957****.*
- Reference date = December 8, 2012
- Without Reasoning

Query	Slot Type	Slot Fill	T1	T2	T3	T4
Sheetrit	Member_of	***Knesset***	$-\infty$	2012-12-08	2012-12-08	$+\infty$
Sheetrit	Resident	***Morocco***	$-\infty$	$-\infty$	1957-01-01	1957-12-31
Sheetrit	Resident	***Israel***	1957-01-01	1957-12-31	$+\infty$	$+\infty$

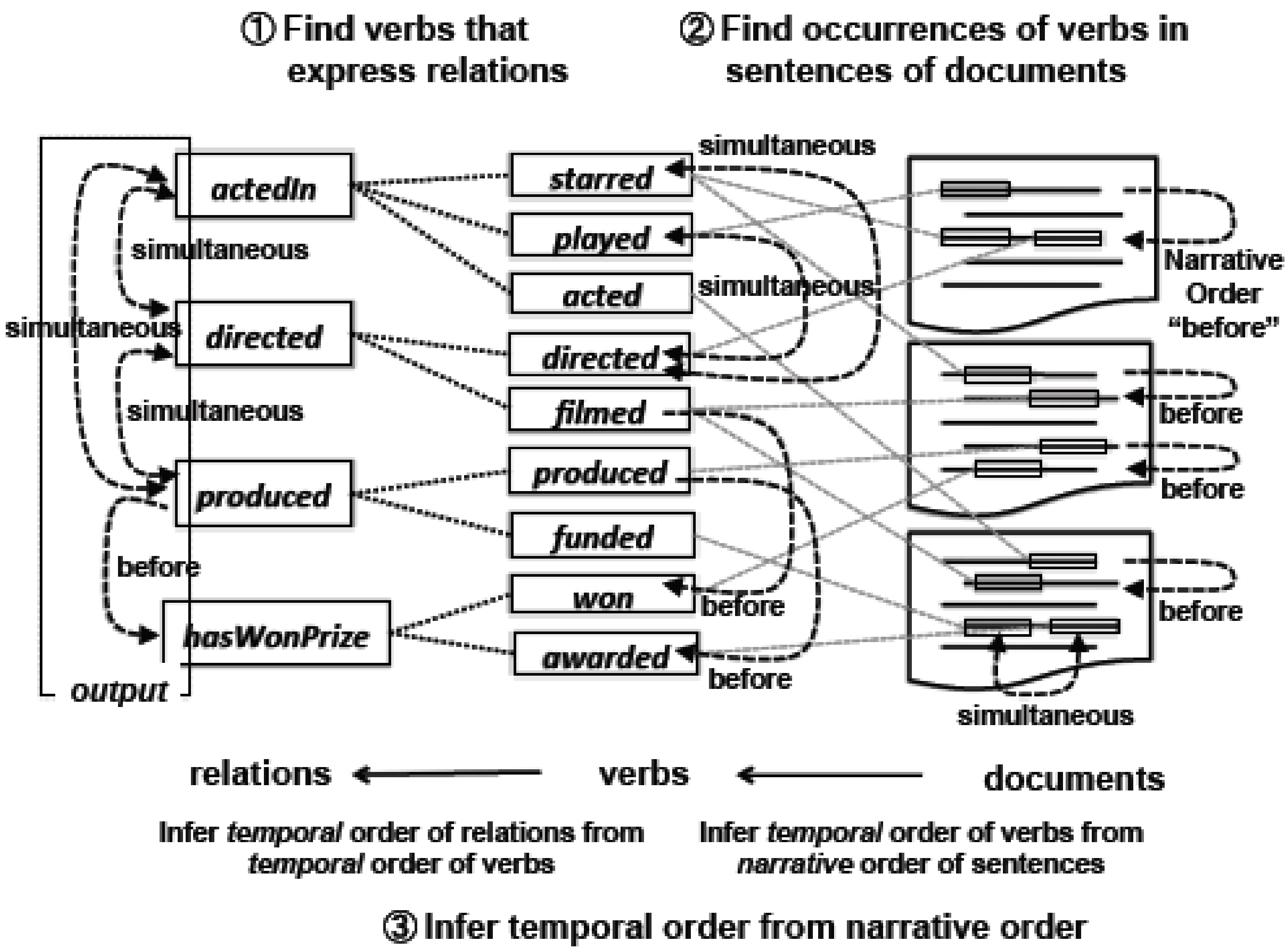
Facts are often Inter-dependent

- ***Sheetrit*** *is a* ***Knesset*** *(parliament) member.*
- ***Sheetrit*** *was born in* ***Morocco*** *and immigrated to* ***Israel*** *in* ***1957****.*
- Reference date = December 8, 2012
- With Reasoning

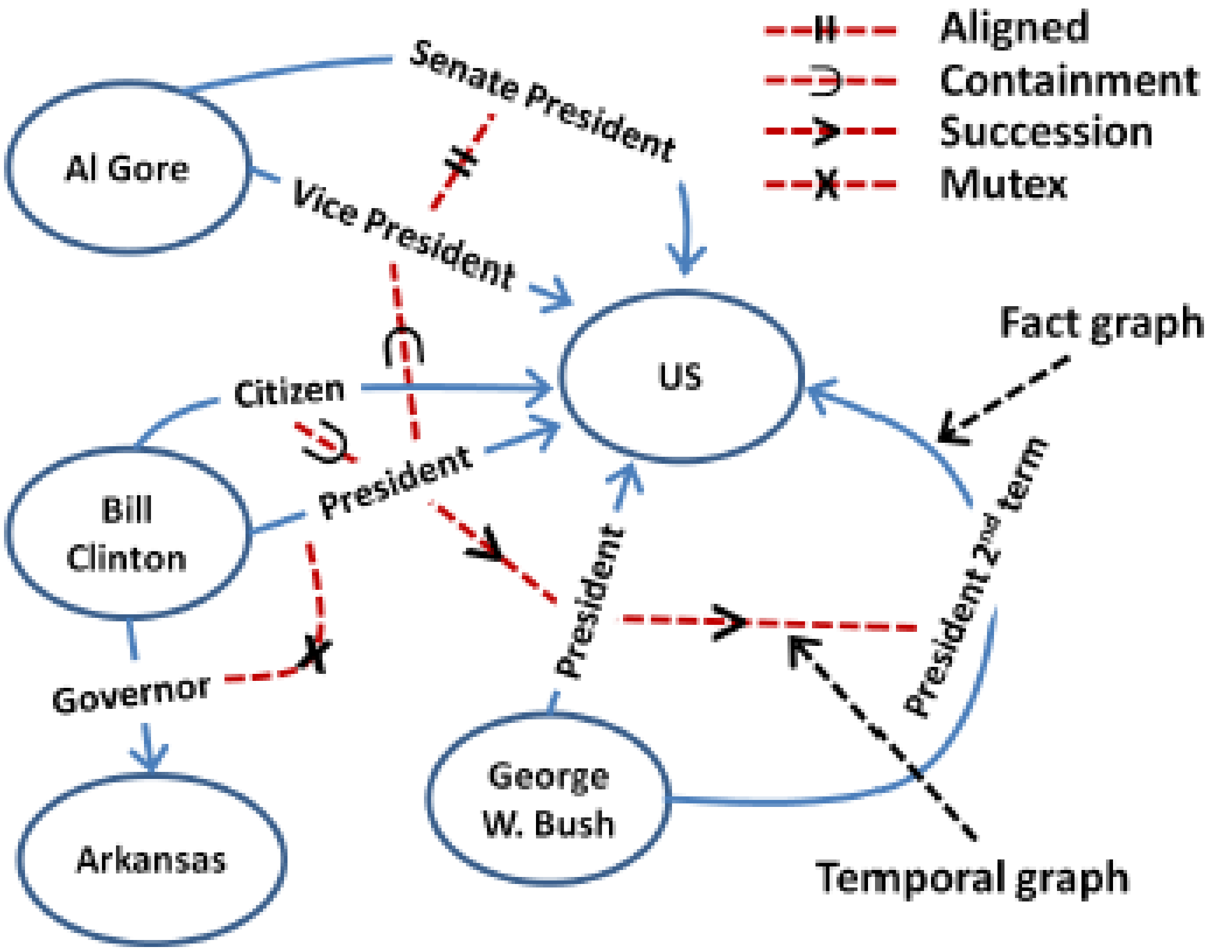
Query	Slot Type	Slot Fill	T1	T2	T3	T4
Sheetrit	Member_of	***Knesset***	-∞	2012-12-08	2012-12-08	+∞
Sheetrit	Resident	***Morocco***	-∞	-∞	1957-01-01	1957-12-31
Sheetrit	Resident	***Israel***	1957-01-01	1957-12-31	+∞	+∞

- *Member (S, K) ^ Located (K, I) → Resident (S, I)*
- *Member (S, K) cannot be earlier than Resident (S, I)*

Learning Temporal Constraints among Relations (Talukdar et al., 2012)



Joint Inference for TSF (Talukdar et al., 2012)



- Solved by an Integer Linear Programming framework

Outline

1. Background: Motivations and Goals
2. Temporal Information Representation Theories
3. Temporal Expression Extraction and Normalization
4. Temporal Slot Filling
5. Tea Break
6. Event Timelining and Temporal Reasoning
7. Resources and Demos
8. Conclusions

11:30

Outline

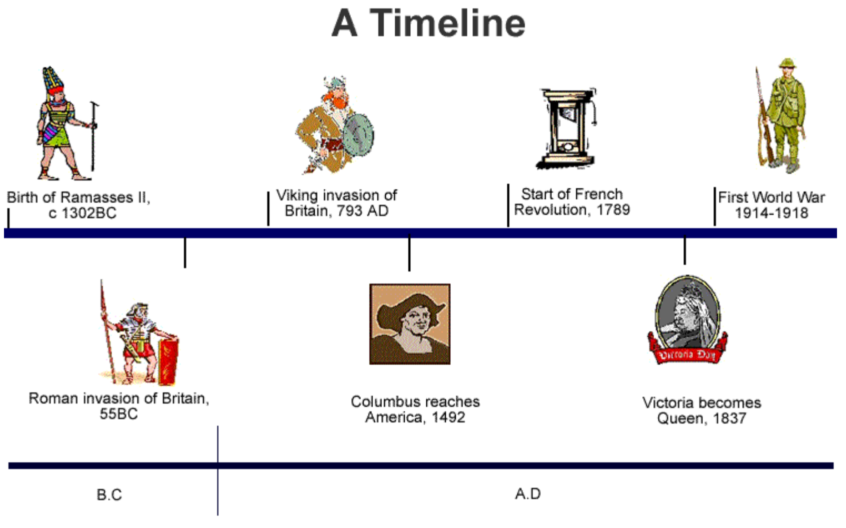
1. Background: Motivations and Goals
2. Temporal Information Representation Theories
3. Temporal Expression Extraction and Normalization
4. Temporal Slot Filling

5. Tea Break

6. Event Timelining and Temporal Reasoning
7. Resources and Demos
8. Conclusions

12:00

- Event Timelining and Shallow Temporal Reasoning



An Example

Seventy-five million copies of the rifle have been built since it entered production in February 1947.

(Publishing date: Feb. 27^{th}, 1998)

basic

complex

Extraction	Normalization	
	Start point	End point
February 1947	1947-02-01 00:00:00	1947-02-28 23:59:59
since […] February 1947	1947-02-01 00:00:00	1998-02-27 23:59:59
Feb. 27^{th}, 1998	1998-02-27 00:00:00	1998-02-27 23:59:59

- **Comparison examples:**

Interval 1	Interval 2	I1 Vs. I2
February 1947	***since […] February 1947***	*inside*
February 1947	Feb. 27^{th}, 1998	*before*

Event Timeline Construction

Wed., May 24th, 2006

[…] The Iraqi insurgents attacked a police station in T[…] on Tuesday killing 6 policemen and injuring 8 other people. This action is allegedly to respond to the bombing action by the coalition armies three days earlier in Baghdad. The police is now sketching a plan to arrest the insurgents in a campaign next week. […]

Motivation

Event Timeline Construction

Wed., May 24th, 2006

Motivation

[…] The Iraqi insurgents *attacked* a police station in Tal Afar on Tuesday *killing* 6 policemen and *injuring* 8 other people. This action is allegedly to respond to the *bombing* action by the coalition armies three days earlier in Baghdad. The police is now sketching a plan to *arrest* the insurgents in a campaign next week. […]

Event	Type	Trigger	Arguments
e_1	Attack	*attacked*	Iraqi insurgents, police station, Tal Afar
e_2	Kill	*killing*	Iraqi insurgents, 6 policemen, Tal Afar
e_3	Injuring	*injuring*	Iraqi insurgents, 8 other people, Tal Afar
e_4	Bombing	*bombing*	coalition armies, Baghdad
e_5	Arrest	*arrest*	police, insurgents

Event Timeline Construction

Document creation time: Wed., May 24th, 2006

Motivation

[…] The Iraqi insurgents *attacked* a police station in [illegible] on Tuesday *killing* 6 policemen and *injuring* 8 other people. This action is allegedly to respond to the *bombing* action by the coalition armies three days earlier in Baghdad. The police is now sketching a plan to *arrest* the insurgents in a campaign next week. […]

Event	Type
e_1	Attack
e_2	Kill
e_3	Injuring
e_4	Bombing
e_5	Arrest

Time	Temporal Expression	Normalized Time
DCT	Wed., May 24th, 2006	2006-05-24
t_1	Tuesday	2006-05-23
t_2	three days earlier	2006-05-21
t_3	now	2006-05-24
t_4	next week	2006-05-29

DCT *= Document Creation Time*

Event Timeline Construction

Wed., May 24th, 2006

Motivation

[…] The Iraqi insurgents *attacked* a police station in on Tuesday *killing* 6 policemen and *injuring* 8 other people. This action is allegedly to respond to the *bombing* action by the coalition armies three days earlier in Baghdad. The police is now sketching a plan to *arrest* the insurgents in a campaign next week. […]

Event	Type
e_1	Attack
e_2	Kill
e_3	Injuring
e_4	Bombing
e_5	Arrest

Time	Normalized Time
DCT	2006-05-24
t_1	2006-05-23
t_2	2006-05-21
t_3	2006-05-24
t_4	2006-05-29

Temp. Relation
Overlaps(e_1, t_1)
Overlaps(e_2, t_1)
Overlaps(e_3, t_1)
Overlaps(e_4, t_2)
After(e_5, t_4)

Applications:

- Event timeline supports discourse understanding, question answering, and news summarization.

- Event timeline allows us to visualize the order of events' occurrence and thus could support better data & knowledge acquisition.

Co

06-29

Problem Definition:

A notational convention: granularity + axioms

- Input:
  - \+ A temporal ontology O_T
  - \+ An event ontology O_E
  - \+ A set of documents D
  - \+ A list of events E=<e_1, e_2, …, e_n> in D, following O_E

- Output:
  - \+ Order of events in E on a timeline with respect to O_T

Event	Type
e_1	Attack
e_2	Kill
e_3	Injuring
e_4	Bombing
e_5	Arrest

Time	Normalized Time
DCT	2006-05-24
t_1	2006-05-23
t_2	2006-05-21
t_3	2006-05-24
t_4	2006-05-29

Temp. Relation
Overlaps(e_1, t_1)
Overlaps(e_2, t_1)
Overlaps(e_3, t_1)
Overlaps(e_4, t_2)
After(e_5, t_4)

Towards Time Line Construction

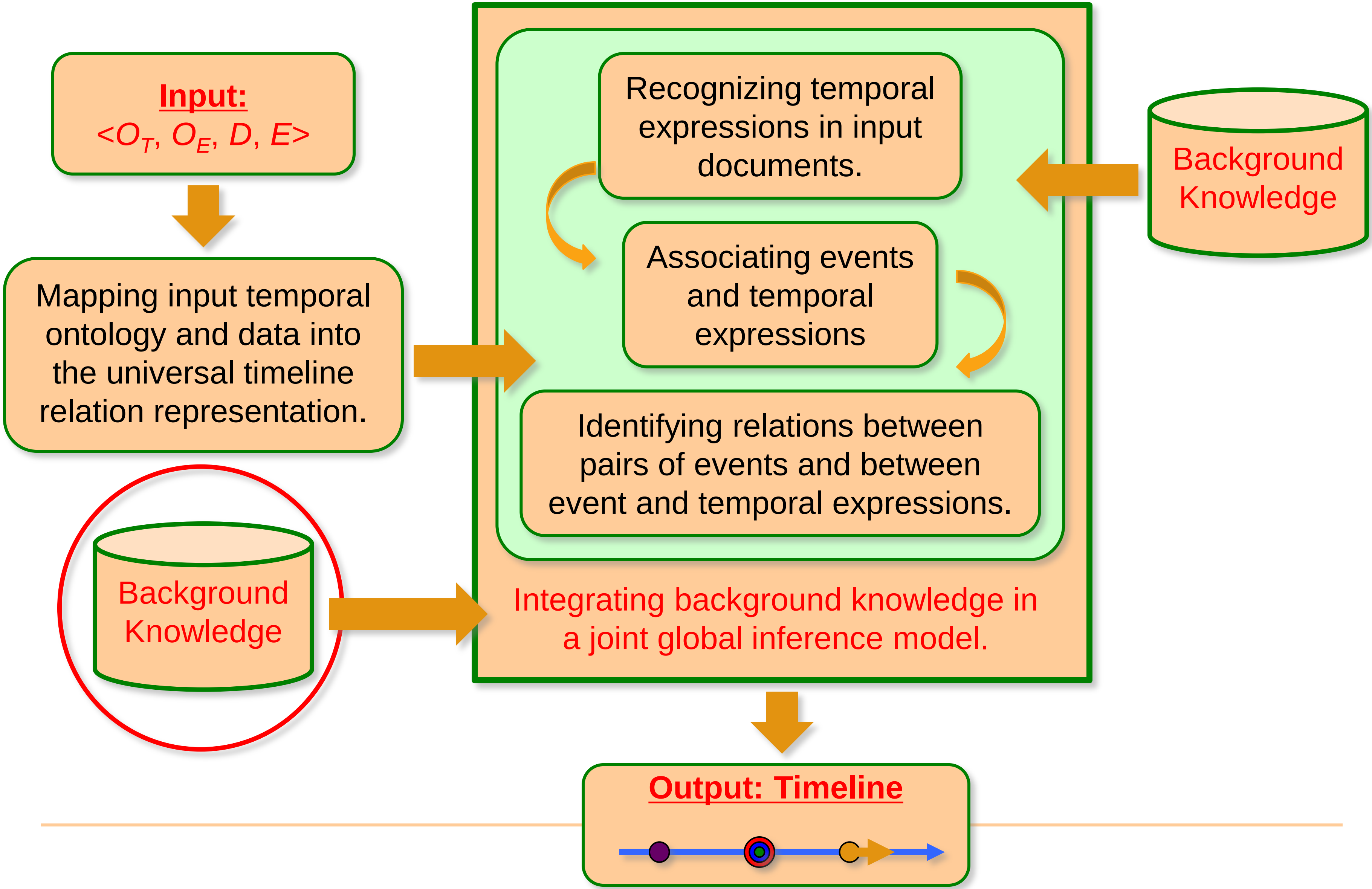
- There has been much work proposing various temporal ontologies and representations (e.g. Allen, 1983; Pustejovsky et al., 2003; Hobbs and Pan, 2004).

- In order to support time line construction there is a need to extend existing representations. We will use the interval based representation described earlier to create a universal timeline relation representation that unifies the efforts of developing many temporal reasoning systems.

- We will then present a timeline construction system that works and performs reasoning on the proposed universal representation.

- Other temporal representations and data can be mapped to the universal representation, thus can be handled by the timeline construction system presented.

An Overview of the Time Lining Approach



Interval Based Event Timeline Construction

Publishing date: Wed., May 24th, 2006

[…] The Iraqi insurgents attacked a police station in Tal Afar on Tuesday killing 6 policemen and injuring 8 other people. This action brings the casualties to over 3000 since the invasion of the coalition armies on 3/20/2003. Police wants to arrest the insurgents in a campaign next week. […]

Interval Based Event Timeline Construction

Publishing date: Wed., May 24th, 2006

[…] The Iraqi insurgents attacked e_1 a police station in Tal Afar on Tuesday killing e_2 6 policemen and injuring e_3 8 other people. This action brings the casualties e_4 to over 3000 since the invasion e_5 of the coalition armies on 3/20/2003. Police wants to arrest e_6 the insurgents in a campaign next week. […]

Interval Based Event Timeline Construction

Publishing date: Wed., May 24th, 2006

[…] The Iraqi insurgents *attacked* e_1 a police station in Tal Afar on Tuesday *killing* e_2 6 policemen and *injuring* e_3 8 other people. This action brings the *casualties* e_4 to over 3000 since the *invasion* e_5 of the coalition armies on 3/20/2003. Police wants to *arrest* e_6 the insurgents in a campaign next week. […]

EVENT: *attacked*
Agent: Iraqi insurgents
Victim: a police station
Location: Tal Afar
Time: Tuesday

killing(…)

arrest(…)

casualties(…)

injuring(…)

invasion(…)

Interval Based Event Timeline Construction

Publishing date: Wed., May 24th, 2006 I_1

[…] The Iraqi insurgents *attacked* a police station in Tal Afar on Tuesday I_2 *killing* 6 policemen and *injuring* 8 other people. This action brings the *casualties* to over 3000 since I_4 the *invasion* of the coalition armies on 3/20/2003 I_3. Police wants to *arrest* the insurgents in a campaign next week I_5. […]

Wed., May 24th, 2006
From: 2006-05-24 00:00:00
To: 2006-05-24 23:59:59

Tuesday
From: 2006-05-23 00:00:00
To: 2006-05-23 23:59:59

since […] 3/20/2003
From: 2003-03-20 00:00:00
To: 2006-05-24 23:59:59

3/20/2003
From: 2003-03-20 00:00:00
To: 2003-03-20 23:59:59

next week
From: 2006-05-29 00:00:00
To: 2006-06-04 23:59:59

Interval Based Event Timeline Construction

Publishing date: Wed., May 24th, 2006

I_1

Comments on timepoint representation (in previous work):

- As annotated in the TimeML corpus.
- A collection of temporal relations between both event-time and event-event.
- Redundant relations
- Hard to construct a timeline of events across documents.
- Inference model becomes cumbersome and less concise (compared to our representation).

Previous work: Timepoint Representation

Interval Based Event Timeline Construction

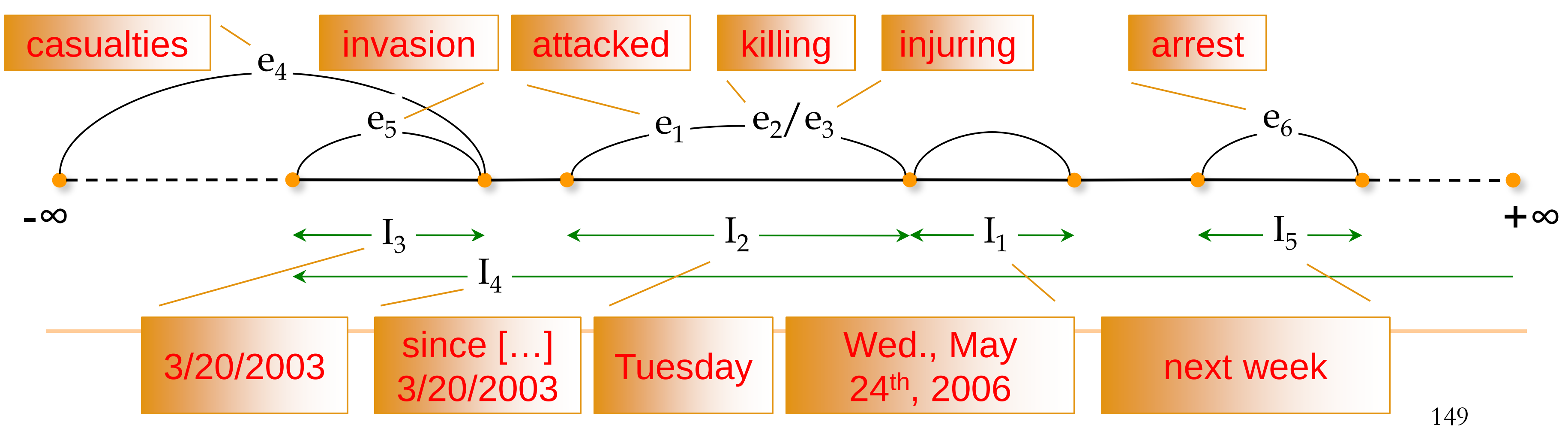
Publishing date: Wed., May 24th, 2006

Comments on our interval-based representation:

- Allows constructing absolute timeline of events.
- Supports constructing timeline of events across documents.
- Concise inference model.

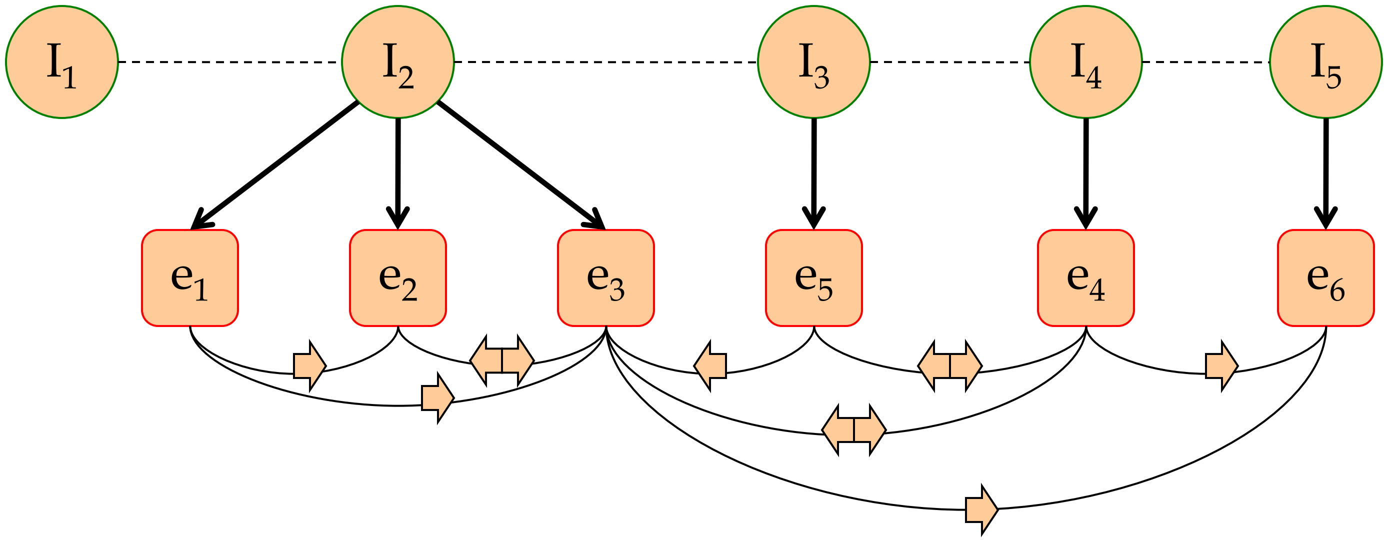
(For more discussion, see Do et. al, EMNLP 2012)

Our interval representation:

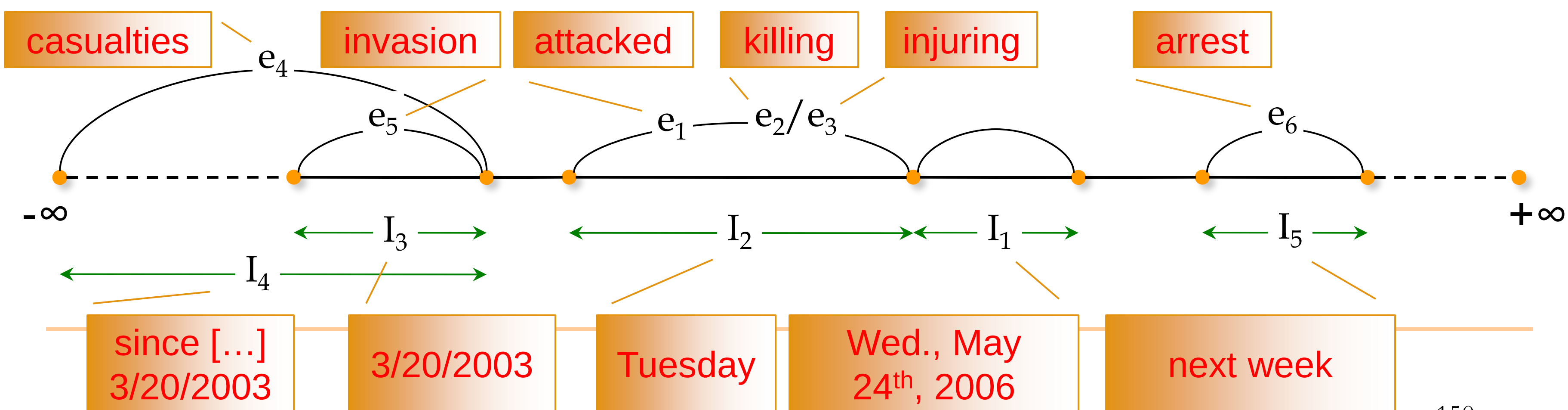


Interval Based Event Timeline Construction

Our proposed document temporal structure:



Our interval representation:



Global Inference for Timeline construction

- Given the interval-based representation we can now reason about relations between events and relations between events and temporal intervals
- We will learn two models:
  - C_{E-E}: Does event A follows event B?
  - C_{T-E}: The relation between event E and time interval T
- We then generate a timeline that attempts to optimize:
  - Respecting the proposals of the two models
  - Respecting common sense constraints

Background Knowledge for Timeline

- Constructing a timeline requires "putting things together": reasoning about temporal intervals & about events and requires incorporating background knowledge
  - Temporal Constraints
    - Enforcing global agreement among the relations between pairs of events and between events and temporal intervals (e.g. reflexivity and transitivity)
  - Statistical Properties
    - Events described in text usually follow a temporal order conveyed via language markers (discourse connectives).
    - Discourse markers and the surrounding context to can be used to time-line temporal entities.

Detour: How to "Put Things Together"

- We will briefly discuss a framework that allows us to
  - incorporate multiple statistical models, along with
  - declarative and statistical background knowledge.

- The knowledge will be modeled as
  - constraints on the outputs of the models and
- The decision problem
  - will be formulated as an Integer Linear Program (ILP)
- The goal is to combine components (models) that have views on parts of the output space in a coherent way—respecting both the models suggestions and the domain/tasks specific background knowledge.

Inference with General Constrain

Improvement over no inference: 2-5%

[Roth&Yih'04,07]

Recognizing Entities and R lations

other	0.05		other	0.10		othe	0.05

Y = argm

= argm

Subject to Constraints

An Objective function that incorporates learned models with knowledge (constraints)

A constrained Conditional Model

spouse_of	**0.45**
born_in	0.50

spouse_of	0.05
born_in	**0.85**

Non Sequential Model

Models could be learned separately; constraints may come up only at decision time.

Constrained Conditional Models

Penalty for violating the constraint.

$$\underset{y}{\operatorname{argmax}}\, \boldsymbol{\lambda} \cdot F(x, y) - \sum_{i=1}^{K} \rho_i d(y, 1_{C_i(x)})$$

(Soft) constraints component

Weight Vector for "local" models

Features, classifiers; log-linear models (HMM, CRF) or a combination

How far y is from a "legal" assignment

How to solve?

This is an Integer Linear Program

Solving using ILP packages gives an exact solution.

Cutting Planes, Dual Decomposition & other search techniques are possible

How to train?

Training is learning the objective function

Decouple? Decompose?

How to exploit the structure to minimize supervision?

Examples: CCM Formulations

$$\operatorname*{argmax}_{y} \boldsymbol{\lambda} \cdot F(x, y) - \sum_{i=1}^{K} \rho_i d(y, 1_{C_i(x)})$$

CCMs can be viewed as a general interface to easily combine declarative domain knowledge with data driven statistical models

Formulate NLP Problems as ILP problems (inference may be done otherwise)

1. Sequence tagging (HMM/CRF + Global constraints)
2. Sentence Compression (Language Model + Global Constraints)
3. SRL (Independent classifiers + Global Constraints)

Sentence Compression/Summarization:

Language Model based:

Argmax $\sum \lambda_{ijk}$ x_{ijk}

Linguistics Constraints

If a modifier chosen, include its head

If verb is chosen, include its arguments

Semantic Role Labeling

I *left* my pearls to my daughter in my will .

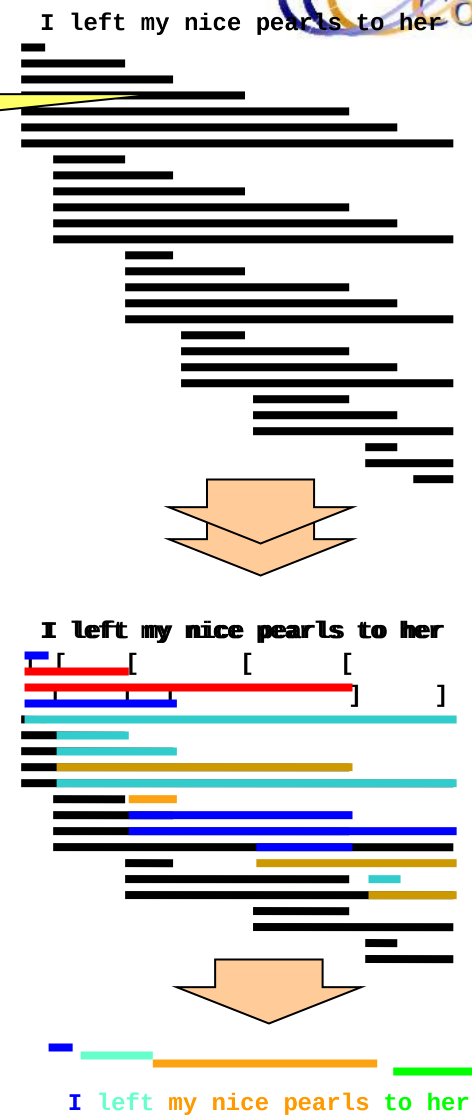
[I]$_{A0}$ *left* [my pearls]$_{A1}$ [to my daughter]$_{A2}$ [in my will]$_{AM\text{-}LOC}$.

- ***A0*** Leaver
- ***A1*** Things left
- ***A2*** Benefactor
- ***AM-LOC*** Location

I *left* my pearls to my daughter in my will .

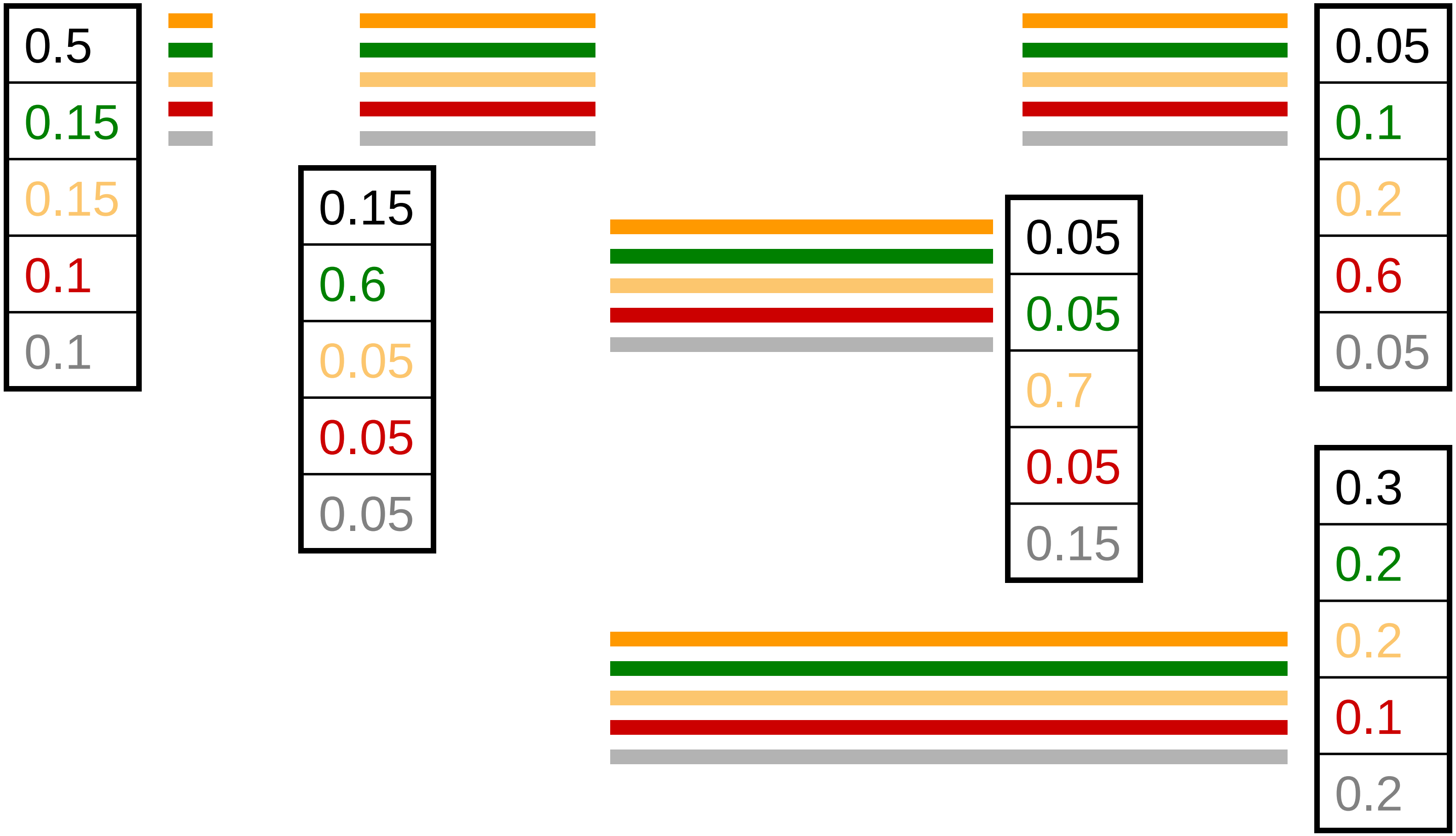
Algorithmic Approach

- Identify argument candidates
  - Pruning [Xue&Palmer, EMNLP'04]
  - Argument Identifier
    - Binary classification
- Classify argument candidates
  - Argument Classifier
    - Multi-class classification
- Inference
  - Use the estimated probability distribution given by the argument classifier
  - Use structural and linguistic constraints
  - Infer the optimal global output



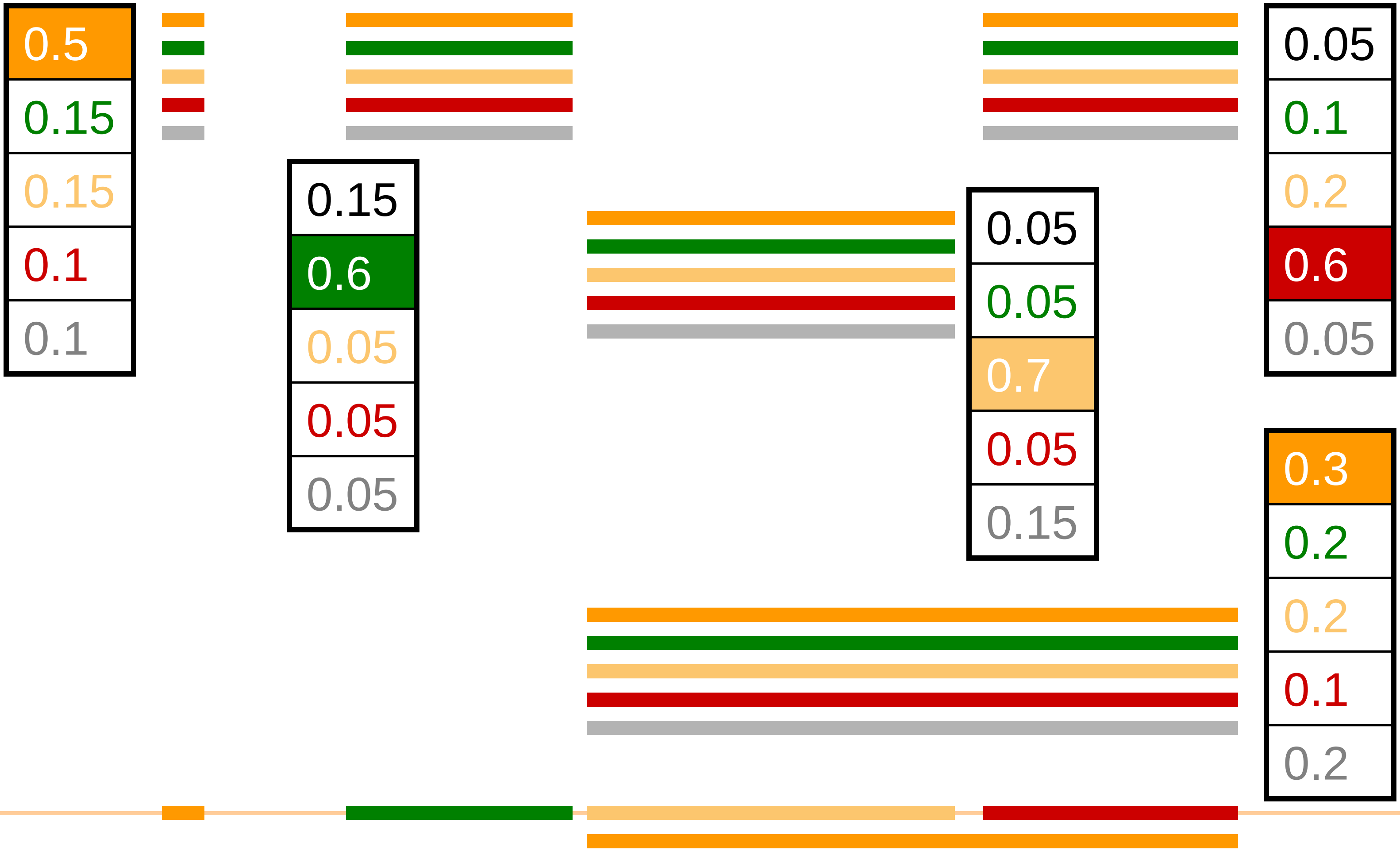
Semantic Role Labeling (SRL)

I *left* my pearls to my daughter in my will .



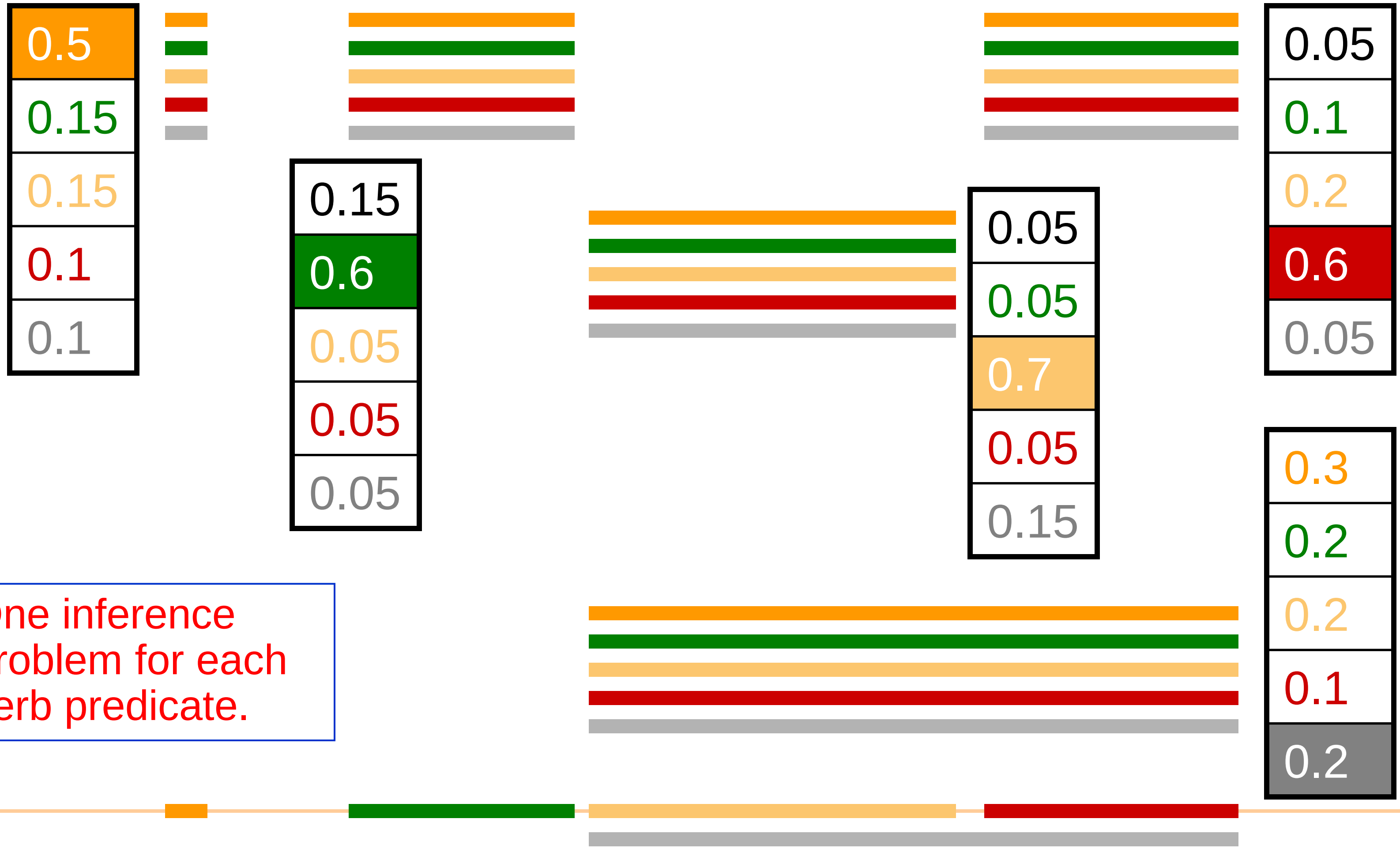
Semantic Role Labeling (SRL)

I *left* my pearls to my daughter in my will .



Semantic Role Labeling (SRL)

I *left* my pearls to my daughter in my will .



One inference problem for each verb predicate.

Constraints

Any Boolean rule can be encoded as a set of linear inequalities.

- No duplicate argument classes

$$\forall y \in \mathcal{Y}, \sum_{i=0}^{n-1} 1_{\{y_i = y\}} \leq 1$$

- Reference-Ax

If there is an Reference-Ax phrase, there is an Ax

$$\forall y \in \mathcal{Y}_R, \sum_{i=0}^{n-1} 1_{\{y_i = y = \text{“R-Ax”}\}} \leq \sum_{i=0}^{n-1} 1_{\{y_i = \text{“Ax”}\}}$$

- Continuation-Ax

If there is an Continuation-x phrase, there is an Ax e it

$$\forall j, y \in \mathcal{Y}_C, \; 1_{\{y_j = y = \text{“C-Ax”}\}} \leq \sum_{i=0}^{j} 1_{\{y_i = \text{“Ax”}\}}$$

Universally quantified rules

- Many other possible constraints:
  - Unique labels
  - No overlapping or embedding
  - Relations between number of arguments; order constraints
  - If verb is of type A, no argument of type B

Learning Based Java: allows a developer to encode constraints in First Order Logic; these are compiled into linear inequalities automatically.

SRL: Posing the Problem

$$\boxed{\text{maximize}} \quad \sum_{i=0}^{n-1} \sum_{y \in \mathcal{Y}} \lambda_{\mathbf{x}_i, y} 1_{\{y_i = y\}}$$

$$\text{where} \quad \lambda_{\mathbf{x},y} = \lambda \cdot F(\mathbf{x}, y) = \lambda_y \cdot F(\mathbf{x})$$

$$\boxed{\text{subject to}}$$

$$\forall i, \sum_{y \in \mathcal{Y}} 1_{\{y_i = y\}} = 1$$

$$\forall y \in \mathcal{Y}, \sum_{i=0}^{n-1} 1_{\{y_i = y\}} \leq 1$$

$$\forall y \in \mathcal{Y}_R, \sum_{i=0}^{n-1} 1_{\{y_i = y = \text{“R-Ax”}\}} \leq \sum_{i=0}^{n-1} 1_{\{y_i = \text{“Ax”}\}}$$

$$\forall j, y \in \mathcal{Y}_C, \; 1_{\{y_j = y = \text{“C-Ax”}\}} \leq \sum_{i=0}^{j} 1_{\{y_i = \text{“Ax”}\}}$$

A	bomb [A1]		killer [A0]
car			
bomb			
that	bomb (Reference) [R-A1]		
exploded	V: explode		
outside	location [AM-LOC]		
the			
U.S.			
military	temporal [AM-TMP]		
base			
in	location [AM-LOC]		
Beniji			
killed			V: kill
11			corpse [A1]
Iraqi			
citizens			
.			

Demo:
http://cogcomp.cs.illinois.edu/

Top ranked system in CoNLL' 05 shared task

Key difference is the Inference

Constrained Conditional Models —A Summary

- Constrained Conditional Models – ILP formulations – have been shown useful in the context of many NLP problems
- [Roth&Yih, 04,07: Entities and Relations; Punyakanok et. al: SRL …]
  - Summarization; Co-reference; Information & Relation Extraction; Event Identifications; Transliteration; Textual Entailment; Knowledge Acquisition; Sentiments; Temporal Reasoning, Dependency Parsing,…
- Some theoretical work on training paradigms [Punyakanok et. al., 05 more; Constraints Driven Learning, PR, Constrained EM…]
- Some work on Inference, mostly approximations, bringing back ideas on Lagrangian relaxation, etc.

- **Summary of work & a bibliography: http://L2R.cs.uiuc.edu/tutorials.html**
- **See also: Chang, Ratinov & Roth, Machine Learning Journal 2012.**

Interval Based Event Timeline Construction

Publishing date: Wed., May 24th, 2006

[…] The Iraqi insurgents attacked a police station in Tal Afar on Tuesday killing 6 policemen and injuring 8 other people. This action brings the casualties to over 3000 since the invasion of the coalition armies on 3/20/2003. Police wants to arrest the insurgents in a campaign next week. […]

Interval Based Event Timeline Construction

Publishing date: Wed., May 24th, 2006

[…] The Iraqi insurgents *attacked* $\mathbf{e_1}$ a police station in Tal Afar on Tuesday *killing* $\mathbf{e_2}$ 6 policemen and *injuring* $\mathbf{e_3}$ 8 other people. This action brings the *casualties* $\mathbf{e_4}$ to over 3000 since the *invasion* $\mathbf{e_5}$ of the coalition armies on 3/20/2003. Police wants to *arrest* $\mathbf{e_6}$ the insurgents in a campaign next week. […]

Interval Based Event Timeline Construction

Publishing date: Wed., May 24th, 2006

[…] The Iraqi insurgents *attacked* e_1 a police station in Tal Afar on Tuesday *killing* e_2 6 policemen and *injuring* e_3 8 other people. This action brings the *casualties* e_4 to over 3000 since the *invasion* e_5 of the coalition armies on 3/20/2003. Police wants to *arrest* e_6 the insurgents in a campaign next week. […]

EVENT: ***attacked***
Agent: Iraqi insurgents
Victim: a police station
Location: Tal Afar
Time: Tuesday

killing(…)

arrest(…)

casualties(…)

injuring(…)

invasion(…)

Interval Based Event Timeline Construction

Publishing date: Wed., May 24th, 2006 I_1

[…] The Iraqi insurgents *attacked* a police station in Tal Afar on Tuesday I_2 *killing* 6 policemen and *injuring* 8 other people. This action brings the *casualties* to over 3000 since I_4 the *invasion* of the coalition armies on 3/20/2003 I_3. Police wants to *arrest* the insurgents in a campaign next week I_5. […]

Wed., May 24th, 2006
From: 2006-05-24 00:00:00
To: 2006-05-24 23:59:59

Tuesday
From: 2006-05-23 00:00:00
To: 2006-05-23 23:59:59

since […] 3/20/2003
From: 2003-03-20 00:00:00
To: 2006-05-24 23:59:59

3/20/2003
From: 2003-03-20 00:00:00
To: 2003-03-20 23:59:59

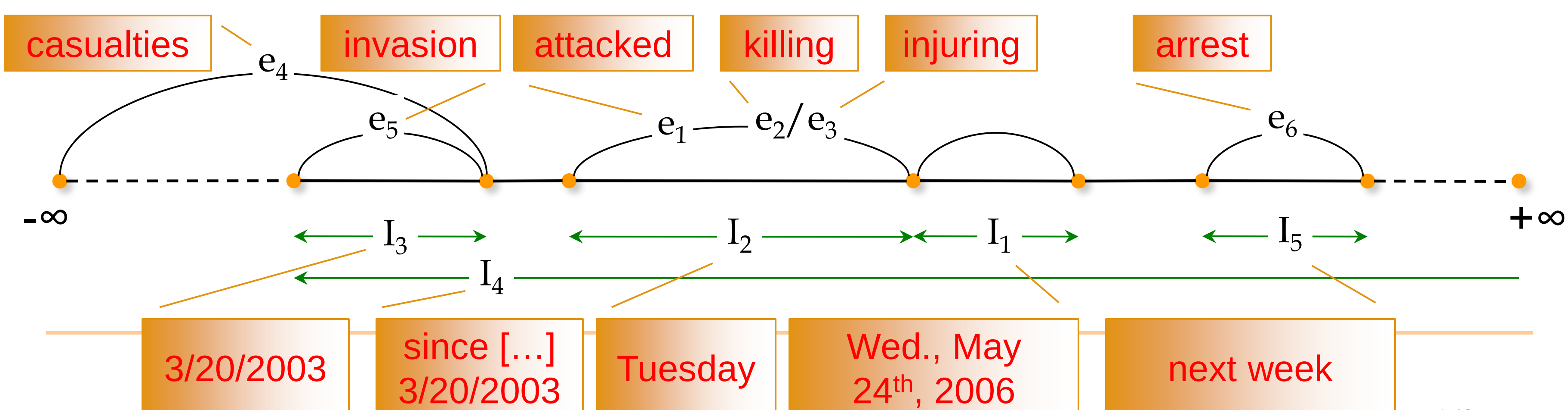
next week
From: 2006-05-29 00:00:00
To: 2006-06-04 23:59:59

Interval Based Event Timeline Construction

Publishing date: Wed., May I_1 24^{th}, 2006

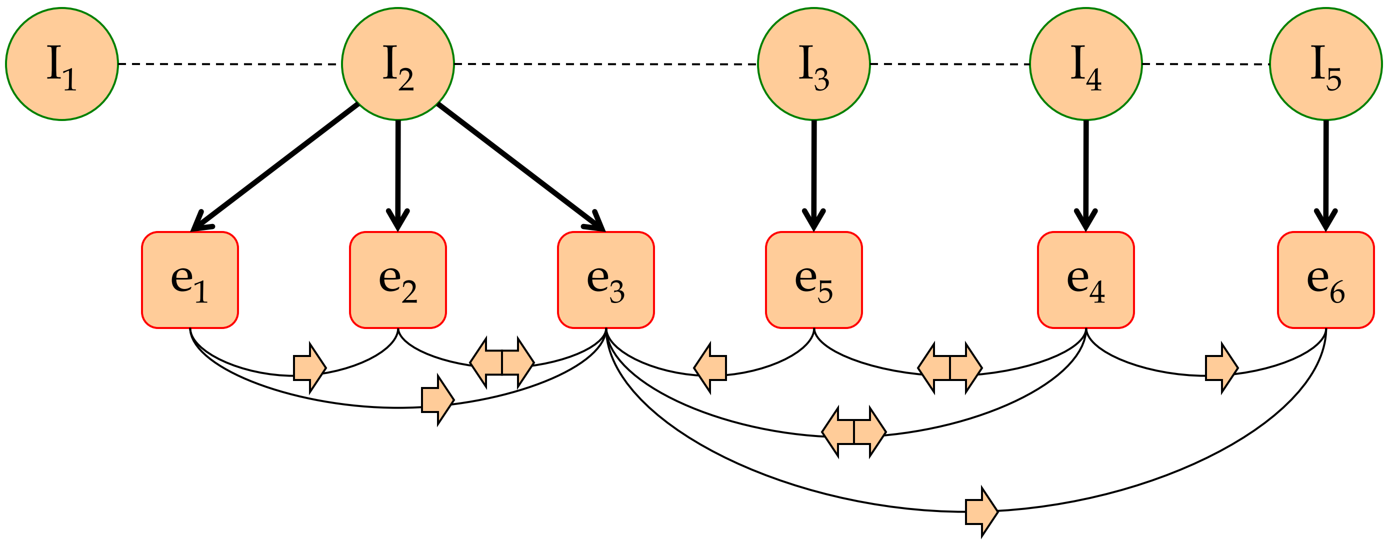
[…] The Iraqi insurgents attacked e_1 a police station in Tal Afar on Tuesday I_2 killing e_2 6 policemen and injuring e_3 8 other people. This action brings the casualties e_4 to over 3000 since I_4 the invasion e_5 of the coalition armies on 3/20/2003 I_3. Police wants to arrest e_6 the insurgents in a campaign next week I_5. […]

Our interval representation:

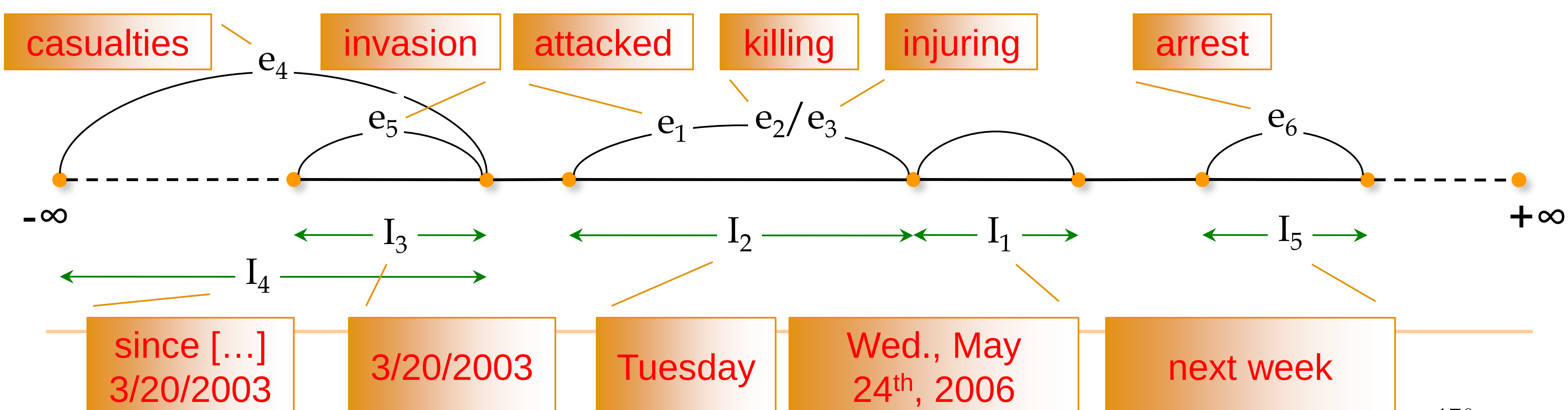


Interval Based Event Timeline Construction

Our proposed document temporal structure:



Our interval representation:



Global Inference for Timeline construction

- Given the interval-based representation we can now reason about relations between events and relations between events and temporal intervals
- We will learn two models:
  - C_{E-E}: Does event A follows event B?
  - C_{T-E}: The relation between event E and time interval T
- We then generate a timeline, via a constrained conditional model that attempts to optimize:
  - Respecting the proposals of the two models
  - Respecting common sense constraints

Background Knowledge for Timeline

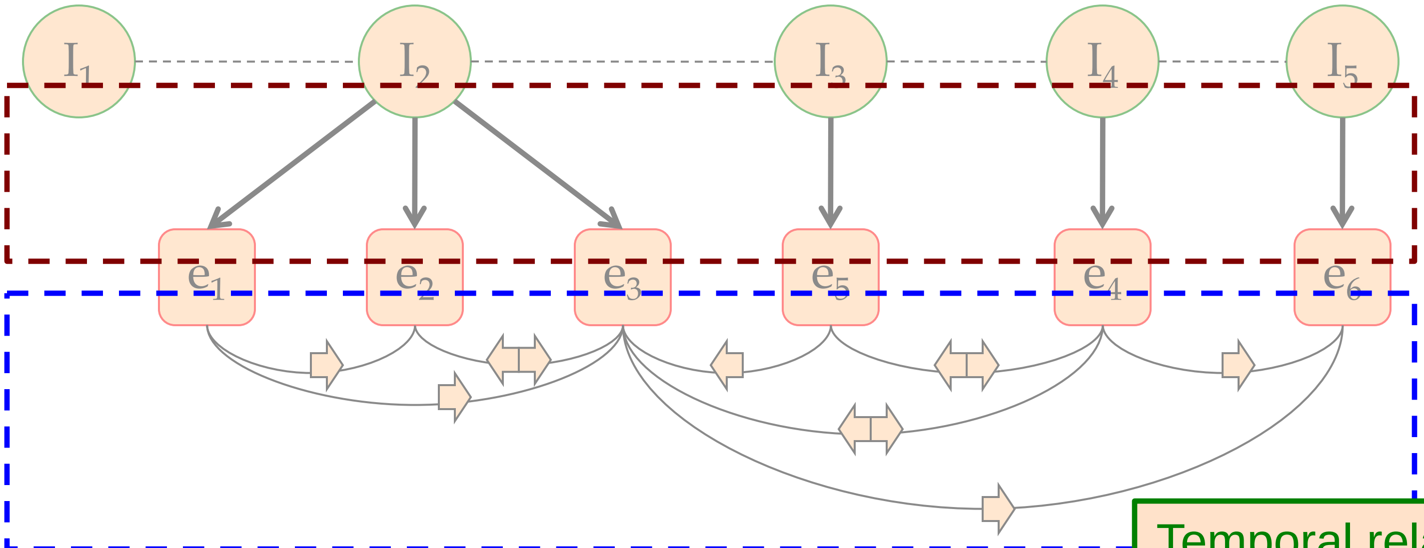
- Constructing a timeline requires "putting things together": reasoning about temporal intervals & about events and requires incorporating background knowledge
  - Temporal Constraints
    - **Enforcing global agreement among the relations between pairs of events and between events and temporal intervals (e.g. reflexivity and transitivity)**
  - Statistical Properties
    - **Events described in text usually follow a temporal order conveyed via language markers (discourse connectives).**
    - **Discourse markers and the surrounding context to can be used to time-line temporal entities.**

A Joint Timeline Model

[1] Event-Time Association: the *E-T* classifier

$$C_{E-T}(e_i, I_j) \rightarrow \{0,1\},$$

$$\forall i, j; 1 \le i \le |E|, 1 \le j \le |I|$$



[2] Event-Event Temporal Order: the *E-E* clas

$$C_{E-E}(e_i, e_j) \rightarrow \{\overline{b}, \overline{a}, \overline{o}, \overline{n}\},$$

$$\forall i, j; 1 \le i, j \le |E|, i \ne j$$

<u>Temporal relations:</u>
before, *after*, *overlap* and *no relation*

A Joint Timeline Model

[1] Event-Time Association: the *E-T* classifier

$$C_{E-T}(e_i, I_j) \rightarrow \{0,1\},$$

$$\forall i, j; 1 \le i \le |E|, 1 \le j \le |I|$$

Supervised Learning

[3] A Global Joint Inference Model with *Common Sense* Knowledge

Constrained Conditional Model

[2] Event-Event Temporal Order: the *E-E* classifier

$$C_{E-E}(e_i, e_j) \rightarrow \{\overline{b}, \overline{a}, \overline{o}, \overline{n}\},$$

$$\forall i, j; 1 \le i, j \le |E|, i \ne j$$

Features of the Temporal Classifiers

<u>Some selective features for a pair of (Event, Time Interval):</u>

The Iraqi insurgents *attacked* a police station in Tal-Afar on *Tuesday.*
(*Publishing date: <u>Wed., May 24th, 2006</u>*)

Temporal entities: attacked (event), Tuesday (time interval)

<u>Syntactic Parse Tree:</u>

```
(ROOT
 (S
  (NP (DT The) (JJ Iraqi) (NNS insurgents))
  (VP (VBD attacked)
   (NP (DT a) (NN police) (NN station))
   (PP (IN in)
    (NP (NNP Tal-Afar)))
   (PP (IN on)
    (NP (NNP Tuesday))))
  (. .)))
```

<u>Syntactic features:</u>

Appearance order: Event_First = True
Same sentence: Same_Sent = True
# of sentences between: Numb_Sent_Diff = None
Prepositional-event phrase: Covered_By_PP = False
Least common non-terminal: LCNT = S

<u>Linguistic features:</u>

Aspect: Event_Aspect = Simple
Tense: Event_Tense = Past

<u>Time interval features:</u>

Explicitness: Explicit_Interval = True
Relative to DCT: Compare_Dct = Before

The Joint Inference Model

- The ILP objective function:

E-T

E-E

$$\arg\max_{x,y}\left[\lambda\sum_{et\in EI} p_{\langle et,1\rangle}\cdot x_{\langle et,1\rangle} + (1-\lambda)\sum_{ee\in EE}\sum_{r\in R} p_{\langle ee,r\rangle}\cdot y_{\langle ee,r\rangle}\right]$$

Probability of event *e* associating with time interval *t*

oral constr

ing basic IL

Probability of two events *e-e* holding temporal relation *r*

E-T indicator variable

E-E indicator variable

Temporal Constraints

1	$\sum_{r \in I}$	y one
2	$\sum_{t \in I}$	only

Reflexivity constraints
- A before B ➔ B after A
- A after B ➔ B before A
- A overlap B ➔ B overlap A
- A has no relation with B ➔ B has no relation with A

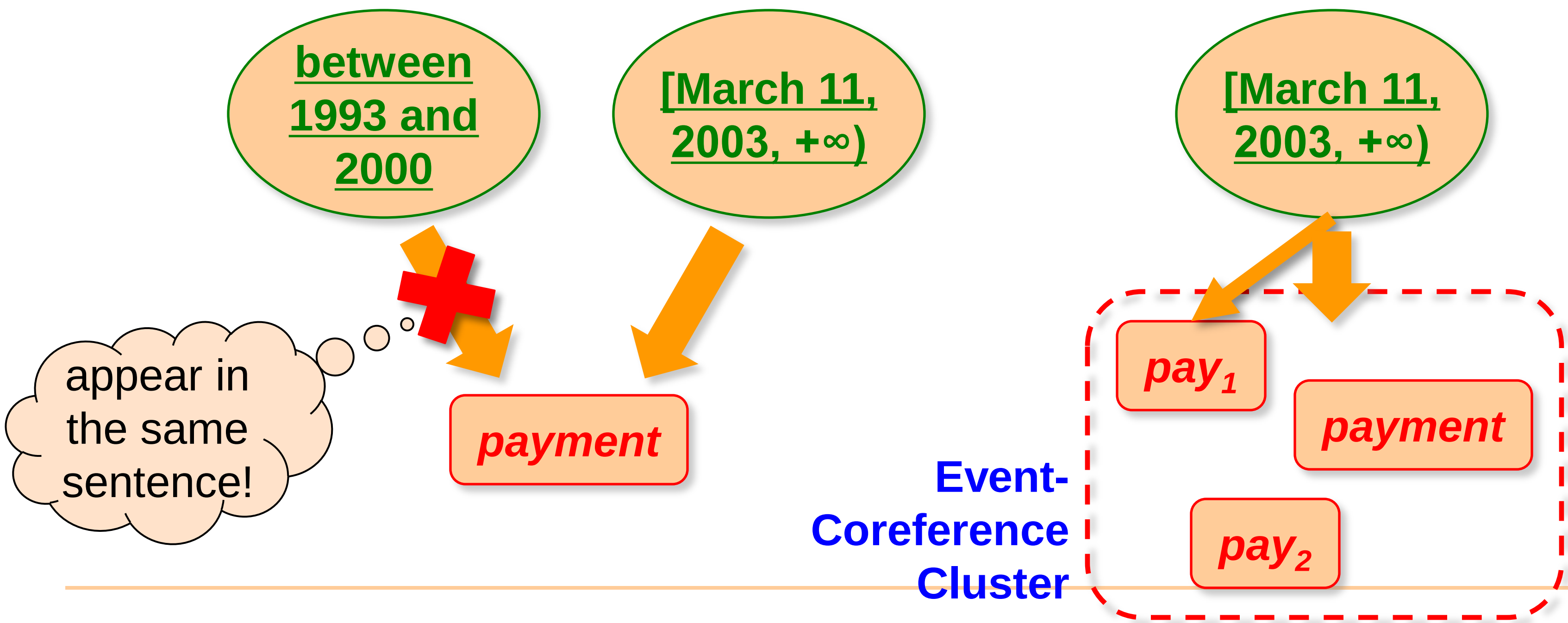
Transitivity constraints
- A before B, B before C ➔ A before C
- A after B, B after C ➔ A after C
- A overlaps B, B overlaps C ➔ A overlaps C
- A before B, B overlaps C ➔ A before C
- A after B, B overlaps C ➔ A after C
- A overlaps B, B before C ➔ A before C
- A overlaps B, B after C ➔ A after C

Knowledge from Eve

Notes:

- We use the knowledge from event-coreference as an additional knowledge source to improve the performance of the timeline construction system.

Sotheby's and Christie's auction houses
dollars to settle an international price-fixing
by the courts, would settle a slew of ***sui***
between 1993 and 2000. Sotheby's and Christie's will each ***pay***$_2$ 20 million dollars.



Knowledge from Event-Coreference

- We propose two principles that make use of event coreference knowledge.

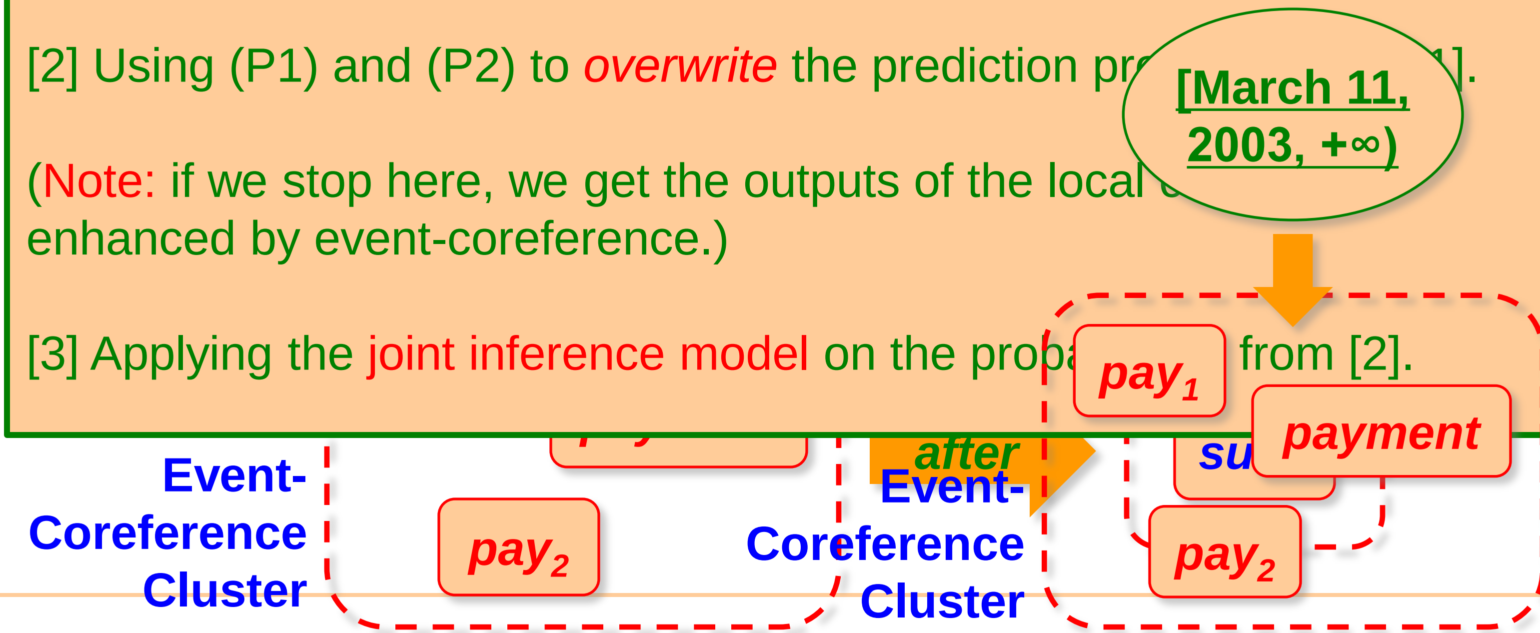
Procedure to Incorporate Event-Coreference Knowledge:

[1] Performing classifications with $C_{E\text{-}T}$ and $C_{E\text{-}E}$.

[2] Using (P1) and (P2) to *overwrite* the prediction pr 1].

(Note: if we stop here, we get the outputs of the local c enhanced by event-coreference.)

[3] Applying the joint inference model on the proba from [2].



Experiments: Data

- 20 newswire articles published in March 2003, from the ACE05 corpus.

valuation

Data	#Intervals	#E-mentions	#E-T	#E-E
Initial	232	324	305	376
Saturated	232	324	324	5940

Notes:

⇰ We do not use the TimeML corpus because:

1. It does not focus on significant events (e.g. Attack, Kill)
2. It does not have event coreference information.

⇰ However, we have some additional experiments on TimeML to compare with previous work (see our paper).

sitivity.

Experiments: Results

All figures are averaged scores from 5-fold cross validation experiments.

	Model	C_{E-T}			C_{E-E}			**Overall**		
		Prec.	*Rec.*	F_1	*Prec.*	*Rec.*	F_1	*Prec.*	*Rec.*	F_1
1	**Baseline**	33.29	33.29	33.29	20.86	32.81	25.03	27.06	33.05	29.16

Baseline:

For E-T:

1. Associates an event mention to the closest interval in the same sentence.

Significance test: Bootstrap Re-sampling (Koehn, 2004)

The overall improvement with the joint inference model is **statistically significant** over the local classifiers (*p ~ 0.01*).

The performance of an event coref system can have significant impact on the task.

An open question: Can event coref benefit from our local classifiers with a joint inference framework?

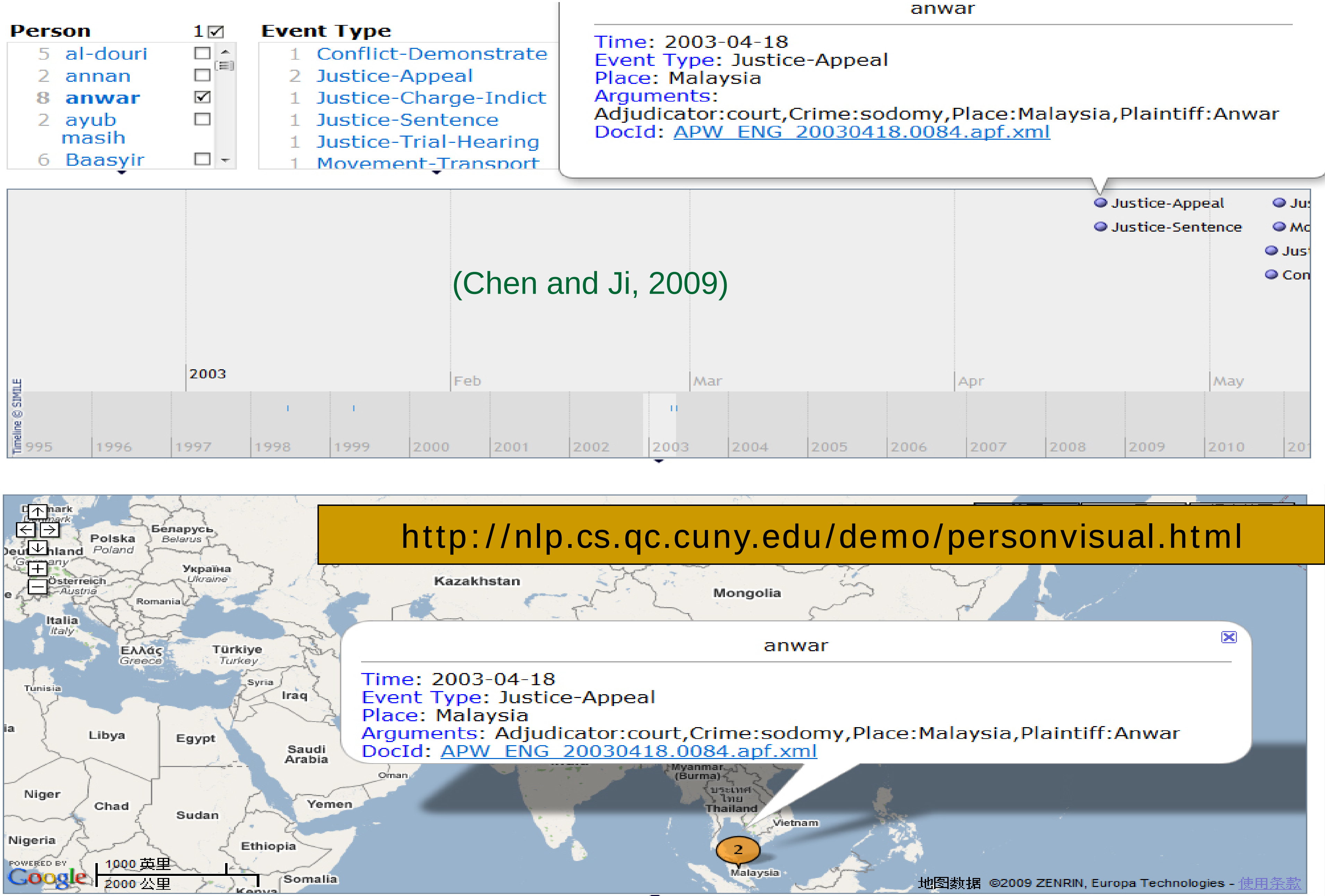
Outline

1. Background: Motivations and Goals
2. Temporal Information Representation Theories
3. Temporal Expression Extraction and Normalization
4. Temporal Slot Filling
5. Tea Break
6. Event Timelining and Temporal Reasoning
7. Resources and Demos
8. Conclusions

12:55

- Resources and Demos

Demo: Temporal Event Tracking



(Chen and Ji, 2009)

http://nlp.cs.qc.cuny.edu/demo/personvisual.html

Resources

- Time Expression Extraction and Normalization
  - Illinois Temporal Extraction and Comparison
    http://cogcomp.cs.illinois.edu/demo/tempdemo/?id=29
  - Stanford SU Time
    http://nlp.stanford.edu/software/sutime.shtml
  - HeidiTime
    http://code.google.com/p/heideltime/
- CUNY Temporal Knowledge Base Population
  - Entity Linking and Regular Slot Filling Programs;
    http://nlp.cs.qc.cuny.edu/kbptoolkit-1.5.0.tar.gz
  - Temporal Slot Filling Programs:
    http://nlp.cs.qc.cuny.edu/Temporal_Slot_Filling_1.0.1.tar.gz
  - Distantly supervised Data:
    http://nlp.cs.qc.cuny.edu/instances.tar.gz
  - Freely available for research purpose

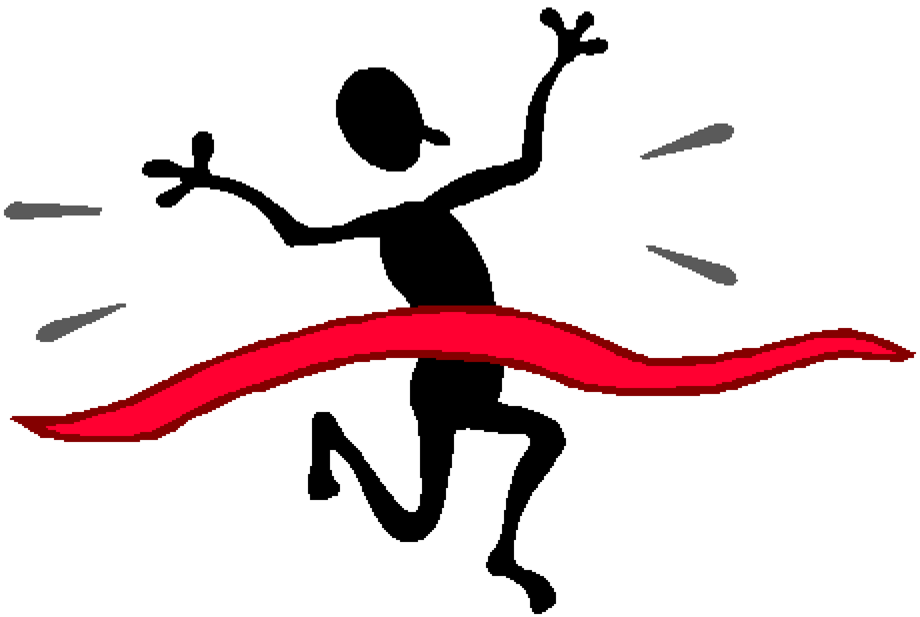
Related Resources

Outline

1. Background: Motivations and Goals
2. Temporal Information Representation Theories
3. Temporal Expression Extraction and Normalization
4. Temporal Slot Filling
5. Tea Break
6. Event Timelining and Temporal Reasoning
7. Resources and Demos
8. Conclusions

1:00

- Conclusions



Other Related Work

- TempEval shared task (Verhagen et al., 2007; Pustejovsky et al., 2009)
  - Single-document, ordering and relation extraction between events and time expressions using TimeBank (Pustejovsky et al., 2003)
  - Flat approaches based on lexical and shallow dependency features (Chambers and Jurafsky, 2008&2009; Yoshikawa et al., 2009; Ling and Weld, 2011)
  - Structured approaches based on syntactic tree and dependency paths (Puscasu, 2007; Cheng et al., 2007; Bethard and Martin, 2007&2008)
- Joint inference for temporal relations (Yoshikawa et al., 2009; Eidelman, 2008; Chambers et al., 2008; Ling, 2010)
  - Focused on single-document and single domain (e.g. medical)

Conclusions

- The "Time" world is fascinating but also challenging
- A lot of problems are open
- Some are easy to define:
  - "Grounding" events and relations
    - Why are we doing so badly?
  - Time lines
    - Why are we doing so badly?
- Some are still ill-defined:
    - I've played Tennis for 10 years vs. I've played Tennis for an hour
- Technical Advances require
  - Cross-document aggregation and grounding
  - Robust Temporal Reasoning
  - Methods that capture long and complex contexts
  - Fast and Accurate ways to obtain training data

Thank you

References

- D Ahn, S.F. Adafre, and M. de Rijke. 2005. Extracting temporal information from open domain text: A comparative exploration. In Digital Information Management.
- Enrique Amigo, Javier Artiles, Qi Li and Heng Ji. 2011. An Evaluation Framework for Aggregated Temporal Information Extraction. Proc. SIGIR2011 Workshop on Entity-Oriented Search.
- J. F. Allen. 1983. Maintaining Knowledge about Temporal Intervals. Communications of the ACM, November 1983, Volume 26, Number 11, volume 26, pages 832-843, New York, NY, USA.
- Silvana Badaloni, Marco Falda and Massimiliano Giacomin. 2004. Integrating Quantitative and Qualitative Fuzzy Temporal Constraints. AI Communications, vol. 17 (4), 187-200, IOS Press.
- Chitta Baral, Gregory Gelfond, Michael Gelfond and Richard B. Scherl. 2005. Textual Inference by Combining Multiple Logic Programming Paradigms. Proc. AAAI 2005 Workshop on Inference for Textual Question Answering.
- Steven Bethard, James H. Martin and Sara Klingenstein. 2007. Finding Temporal Structure in Text: Machine Learning of Syntactic Temporal Relations. International Journal of Semantic Computing (IJSC), 1(4), December 2007.
- Steven Bethard and James H. Martin. 2008. Learning Semantic Links from a Corpus of Parallel Temporal and Causal Relations. Proc. Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT).
- Branimir Bouguraev and Rie Kubota Ando. 2005. TimeBank-Driven TimeML Analysis. Proc. Annotating, Extracting and Reasoning about Time and Events. IBFI, Schloss Dagstuhl, Germany.
- Philip Bramsen, Pawan Deshpande, Yoong Keok Lee, and Regina Barzilay. 2006. Inducing temporal graphs. Proc. Conference on Empirical Methods in Natural Language Processing (EMNLP 2006).
- Nathanael Chambers and Dan Jurafsky. 2008. Jointly Combining Implicit Constraints Improves Temporal Ordering. Proc. Empirical Methods in Natural Language Processing (EMNLP2008).
- Nathanael Chambers and Dan Jurafsky. 2009. Unsupervised Learning of Narrative Schemas and their Participants. Proc. the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing (ACL-IJCNLP 2009).

References (Cont')

- DARPA. 2012. Machine Reading Program. http://www.darpa.mil/Our_Work/I2O/Programs/Machine_Reading.aspx.
- Quang Do, YeeSeng Chan and Dan Roth. 2011. Minimally Supervised Event Causality Identification. Proc. Empirical Methods in Natural Language Processing (EMNLP2012).
- Quang Do, Wei Lu and Dan Roth. 2012. Joint Inference for Event Timeline Construction. Proc. Empirical Methods for Natural Language Processing (EMNLP2012).
- Noemie Elhadad, Regina Barzilay and Kathleen McKeown. 2002. Inferring Strategies for Sentence Ordering in Multidocument Summarization. Journal of Artificial Intelligence Research,17:35-55.
- Prashant Gupta and Heng Ji. 2009. Predicting Unknown Time Arguments based on Cross-event propagation. Proc. the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing (ACL-IJCNLP 2009).
- Heng Ji, Ralph Grishman, Zheng Chen and Prashant Gupta. 2009. Cross-document Event Extraction and Tracking: Task, Evaluation, Techniques and Challenges. Proc. Recent Advances in Natural Language Processing 2009.
- Heng Ji and Ralph Grishman. 2011. Knowledge Base Population: Successful Approaches and Challenges. Proc. The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT 2011).
- Heng Ji, Ralph Grishman and Hoa Trang Dang. 2011. An Overview of the TAC2011 Knowledge Base Population Track. Proc. Text Analysis Conference (TAC2011).
- Paul Kingsbury and Martha Palmer. 2002. From TreeBank to PropBank. Proc. the 3rd International Conference on Language Resources and Evaluation (LREC-2002).
- Qi Li, Javier Artiles, Taylor Cassidy and Heng Ji. 2012. Combining Flat and Structured Approaches for Temporal Slot Filling or: How Much to Compress?. Lecture Notes in Computer Science, 2012, Volume 7182/2012, 194-205.
- Mirella Lapata and Alex Lascarides. 2006. Learning Sentence-internal Temporal Relations. Journal of Artificial Intelligence Research 27. pp. 85-117.

References (Cont')

- Xiao Ling and Daniel S. Weld. Temporal Information Extraction. Proc. the Twenty-Fourth Conference on Artificial Intelligence (AAAI 2010).
- James Pustejovsky, Patrick Hanks, Roser Saurí, Andrew See, Robert Gaizauskas, Andrea Setzer, Dragomir Radev, Beth Sundheim, David Day, Lisa Ferro and Marcia Lazo. 2003. The TIMEBANK Corpus. Proceedings of Corpus Linguistics 2003: 647-656.
- James Pustejovsky, Jos M. Castao, Robert Ingria, Roser Sauri, Robert J. Gaizauskas, Andrea Setzer, Graham Katz, and Dragomir R. Radev. 2003. TimeML: Robust Specification of Event and Temporal Expressions in Text. In Mark T. Maybury, editor, New Directions in Question Answering, pages 28-34. AAAI Press.
- James Pustejovsky and Marc Verhagen. 2009. Semeval-2010 Task 13: Evaluating Events, Time Expressions, and Temporal Relations (Tempeval-2). Proc. the Workshop on Semantic Evaluations: Recent Achievements and Future Directions.
- Steven Schockaert, Martine De Cock, and Etienne E. Kerre. 2008. Fuzzifying Allen's Temporal Interval Relations. IEEE Transactions on Fuzzy Systems, Vol. 16, No. 2.
- Richard Snodgrass. 1998. Of Duplicates and Septuplets. Database Programming and Design.
- Jannik Strtotgen and Michael Gertz. 2010. HeidelTime: High Quality Rule-based Extraction and Normalization of Temporal Expressions. Proc. the 5th International Workshop on Semantic Evaluation.
- Naushad UzZaman and James F. Allen. 2010. TRIPS and TRIOS System for TempEval-2: Extracting Temporal Information from Text. Proc. the 5th International Workshop on Semantic Evaluation.
- Naushad UzZaman and James F. Allen. 2011. Temporal Evaluation. The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT2011).
- Marc Verhagen. 2005. Temporal Closure in an Annotation Environment. Language Resources and Evaluation.
- Marc Verhagen, Roser Sauri, Tommaso Caselli, and James Pustejovsky. 2010. Semeval-2010 task 13: Tempeval 2. In Proceedings of International Workshop on Semantic Evaluations (SemEval 2010), 2010.
- Lluis Vila. 1994. A Survey on Temporal Reasoning in Artificial Intelligence. AI Communications, 7(1):4.
- Katsumasa Yoshikawa, Sebastian Riedel, Masayuki Asahara, and Yuji Matsumoto. 2009. Jointly Identifying Temporal Relations with Markov Logic. Proc. Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP (ACL-IJCNLP2009).

- “ *The events in our lives happen in a sequence in time, but in their significance to ourselves they find their own order: the continuous thread of revelation.*”

- *Eudora Welty*


Grazie
Merci
感謝
Thank you
Gracias
DANKE

- Backup…

- Temporal Event Tracking and Implicit Time Prediction

A Vision (Ji et al., 2009)

…

*Centroid="**Toefting**"*
Rank=26

Time	*2002-01-01*
Event	Attack
Person	***Toefting***
Place	Copenhagen
Target	workers

Time	*2003-03-15*
Event	End-Position
Person	***Toefting***
Entity	Bolton

Time	*2003-03-31*
Event	Sentence
Defendant	***Toefting***
Sentence	four months in prison
Crime	assault

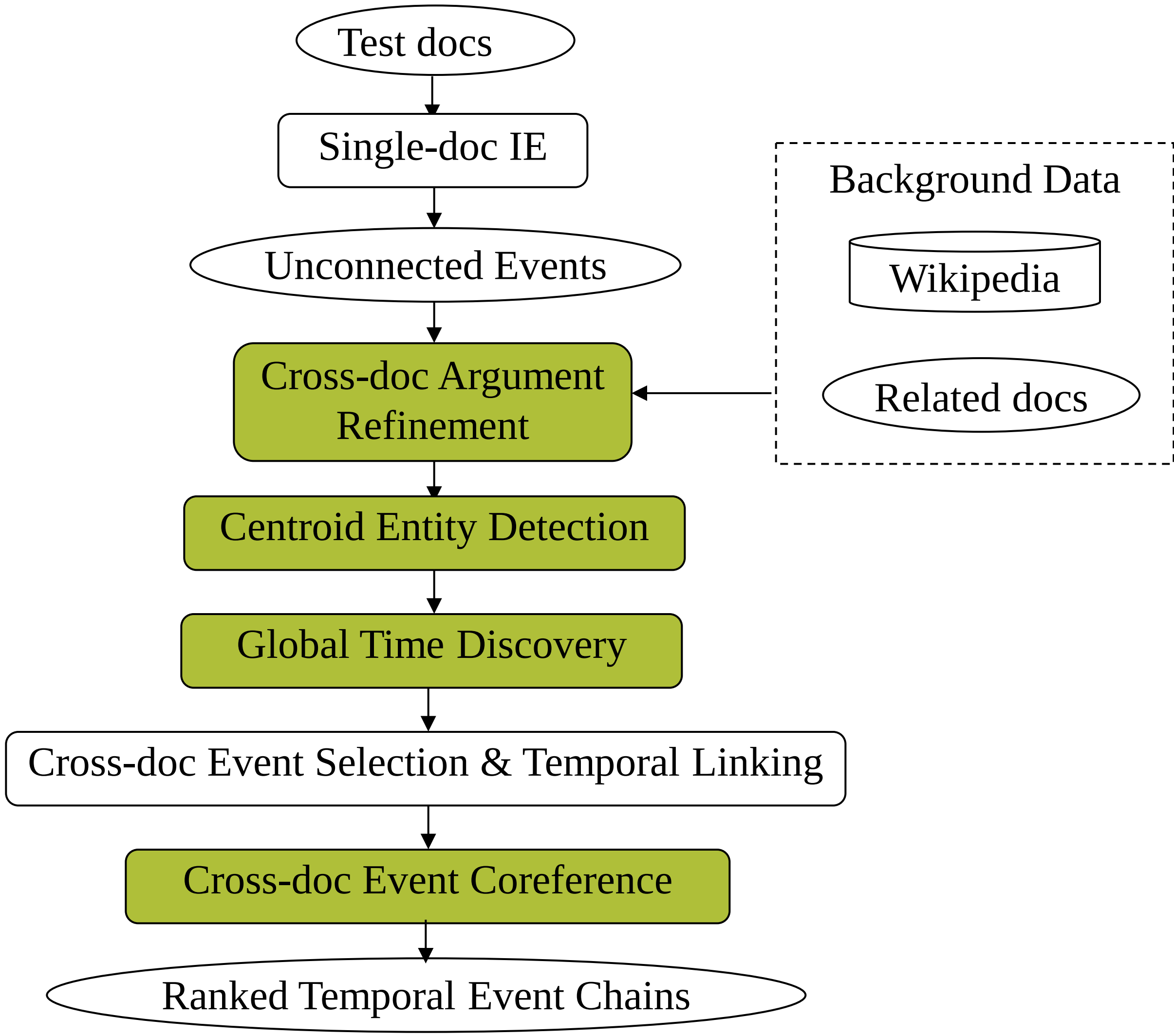
…

- **Input**: A test set of documents
- **Output**: Identify a set of *centroid entities*, and then for each centroid entity, link and order the events centered around it on a time line

Evaluation Metrics

- Centroid Entity Detection
  - ***F-Measure***: *A centroid entity is correctly detected* if its name (and document id) matches the full or partial name of a reference centroid
  - Normalized Kendall tau distance (Centroid entities) = the fraction of correct system centroid entity pairs out of salience order
  - ***Centroid Entity Ranking Accuracy*** = 1- Normalized Kendall tau distance (Centroids)
- Browsing Cost: Incorporate Novelty/Diversity into F-Measure
  - *An argument is correctly extracted* in an event chain if its event type, string and role match any of the reference argument mentions
  - *Two arguments in an event chain are redundant* if their event types, event time, string (the full or partial name) and roles overlap
  - ***Browsing Cost (i)*** = the number of incorrect or redundant event arguments that a user must examine before finding i correct event arguments
- Temporal Correlation: Measure Coherence
  - ***Temporal Correlation*** = the correlation of the temporal order of *argset* in the system output and the answer key
  - ***Argument recall*** = number of unique and correct arguments in response / number of unique arguments in key

Cross-document Temporal Event Tracking



Baseline Single-document IE System

- Includes entity extraction, time expression extraction and normalization, relation extraction and event extraction
- Event Extraction
  - Pattern Matching
    - British and US forces reported gains in the advance on Baghdad
      → PER report gain in advance on LOC
  - Maximum Entropy Models
    - Trigger Labeling: to distinguish event instances from non-events, to classify event instances by type
    - Argument Identification: to distinguish arguments from non-arguments
    - Argument Classification: to classify arguments by argument role
    - Reportable-Event Classifier: to determine whether there is a reportable event instance
    - Each step produces *local confidence*

(Grishman et al., 2005)

What's New: Research Challenges Overview

- More Salient: Detecting centroid entities using global confidence
- More Accurate and Complete: Correcting and enriching arguments from the background data
- More Concise: Conducting cross-document event coreference resolution to remove redundancy

Centroid Entity Detection

- Cross-document Name Coreference
  - Single-doc entity coreference (Ji et al., 2005) + Simple substring matching in the paper
  - Ongoing work: using event chains as feedback (Dogan and Ji, in submission)

- Global Entity Ranking
  - Promote those arguments which are both central to the collection (*high frequency*) and more likely to be accurate (*high confidence*)
  - $\{n_j \mid n_j$ *is a name,* n_j and e_i are coreferential or linked by a relation; and n_j is involved in an event mention$\}$

$$salience(e_i) = \sum_j \sum_k local-confidence(n_j, event_k)$$

Cross-document Argument Refinement

- Hypotheses
  - One Trigger Sense Per Cluster
  - One Argument Role Per Cluster
- Aggregate similar events across related documents and conduct statistical global inference
- Remove triggers and argument annotations with local or cross-doc confidence lower than thresholds
- Propagate highly consistent and frequent arguments with high global confidence to override other, lower confidence, extraction results

(Ji and Grishman, ACL 2008)

Cross-document Event Coreference Resolution

1. An explosion in a cafe at one of the capital's busiest intersections killed one woman and injured another Tuesday

2. Police were investigating the cause of the explosion in the restroom of the multistory Crocodile Cafe in the commercial district of Kizilay during the morning rush hour

3. The blast shattered walls and windows in the building

4. Ankara police chief Ercument Yilmaz visited the site of the morning blast

5. The explosion comes a month after

6. a bomb exploded at a McDonald's restaurant in Istanbul, causing damage but no injuries

7. Radical leftist, Kurdish and Islamic groups are active in the country and have carried out the bombing in the past

(Chen and Ji, 2009)

Spectral Graph Clustering

Trigger	explosion	
Arguments	Role = Place	a cafe
	Role = Time	Tuesday

Trigger	explosion	
Arguments	Role = Place	restroom
	Role = Time	morning rush hour

Trigger	explosion	
Arguments	Role = Place	building

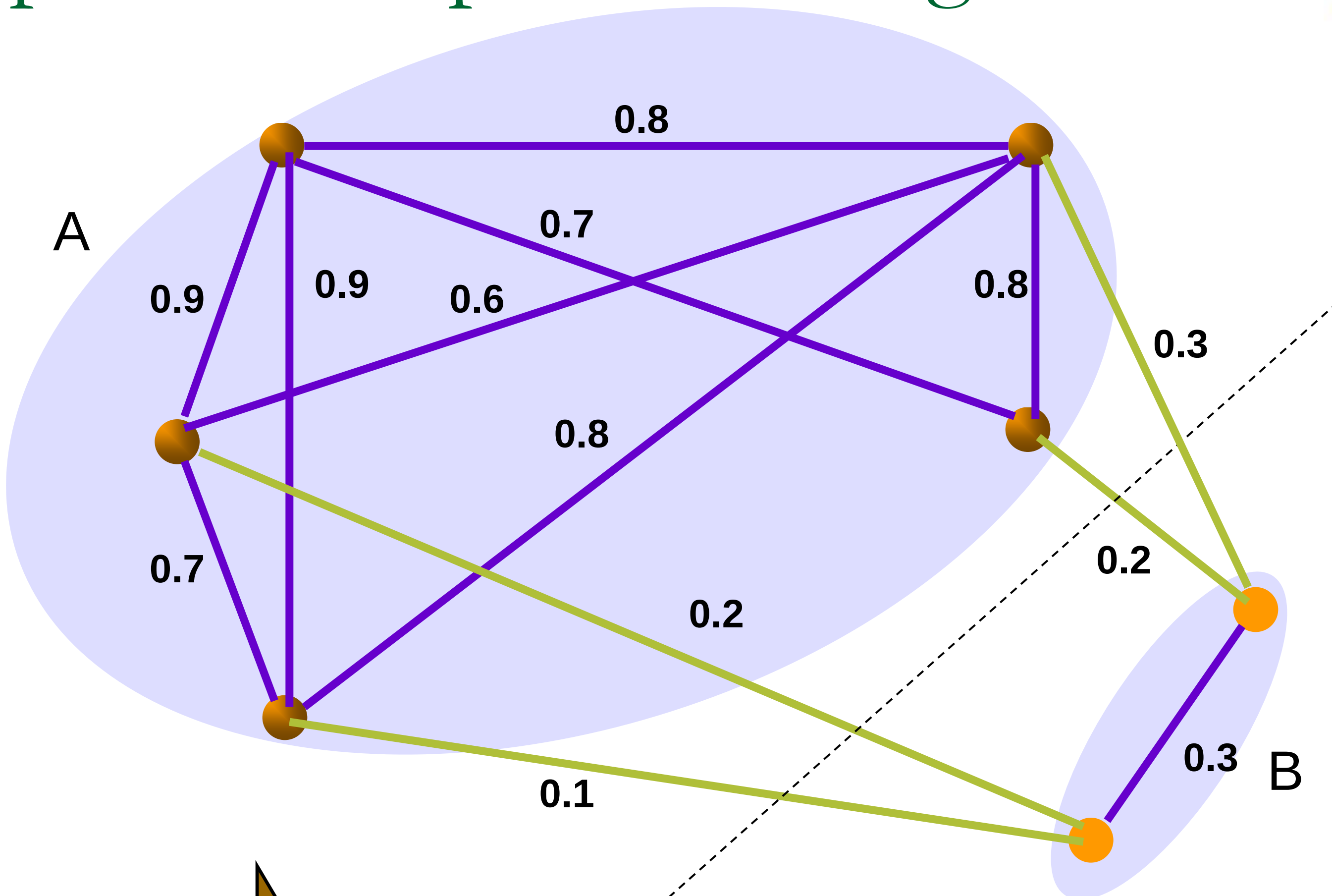
Trigger	blast	
Arguments	Role = Place	site
	Role = Time	morning

Trigger	explosion	
Arguments	Role = Time	a month after

Trigger	exploded	
Arguments	Role = Place	restaurant

Trigger	bombing	
Arguments	Role = Attacker	groups

Spectral Graph Clustering



$$cut(A,B) = 0.1+0.2+0.2+0.3=0.8$$

Experiments: Data

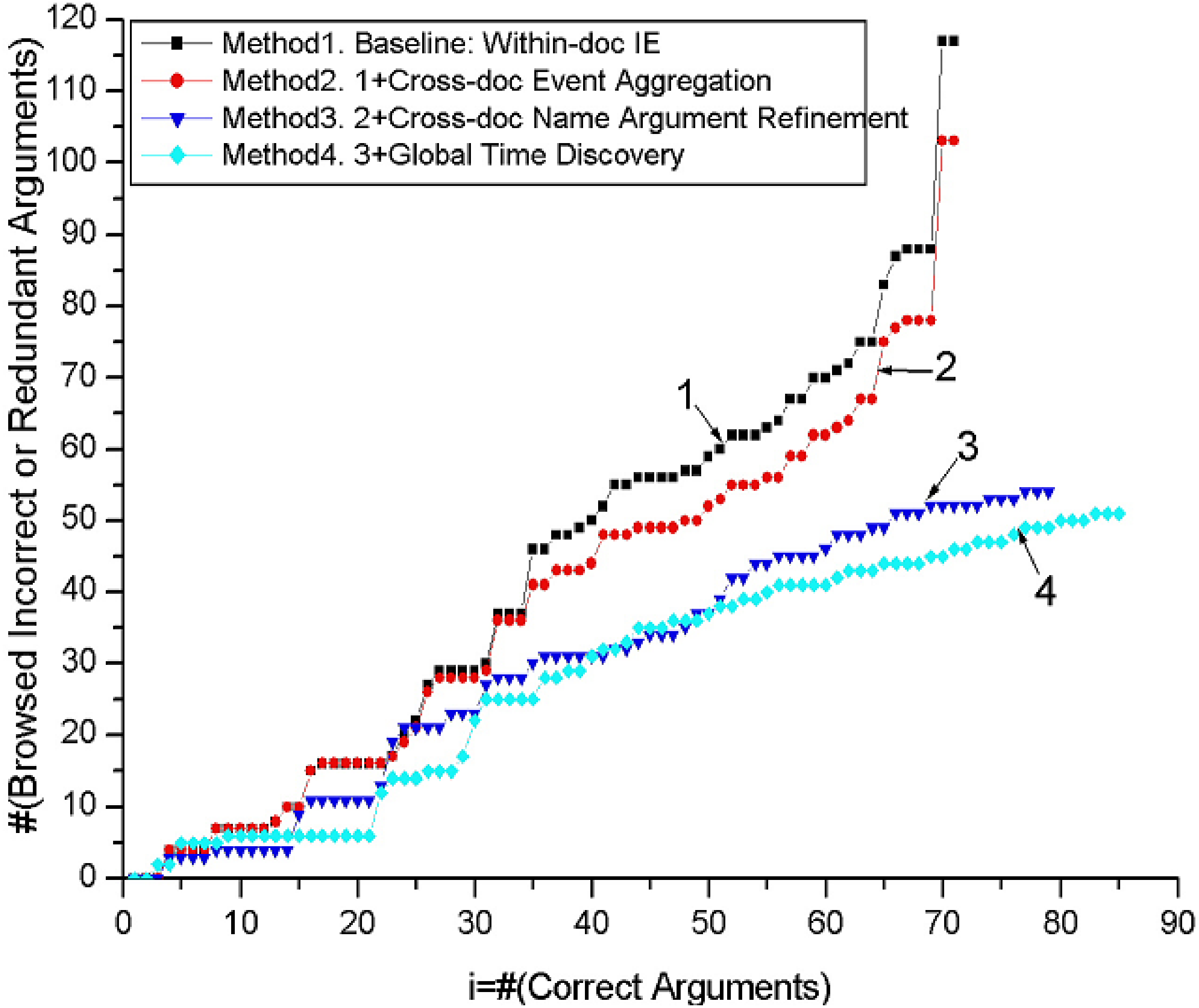
- 106 newswire texts from ACE 2005 training corpora as test set
- extracted the top 40 ranked person names as centroid entities, and manually created temporal event chains by
  - Aggregated reference event mentions (Inter-annotator agreement: ~90%)
  - Filled in the implicit event time arguments from the background data (Inter-annotator agreement: ~82%)
  - Annotated by two annotators independently and adjudicated
- 278,108 texts from English TDT5 corpus and 148 million sentences from Wikipedia as the source for background data
- 140 events with 368 arguments (257 are unique)
- The top ranked centroid entities are "Bush", "Ibrahim", "Putin", "Al-douri", "Blair", etc.

Centroid Entity Detection Performance

- F-Measure
  - Single-document IE: 55% to detect 40 centroid entities
  - Cross-document IE: 67.5% to detect 40 centroid entities, can cover all key centroid entities by using the top 76 system output entities

- Ranking accuracy for 40 centroid entities
  - Cross-document IE: 72.95%
  - Baseline 1 - random ranking: 42%
  - Baseline 2 - ranked by the position where the first mentions of the centroid entities appear as event arguments in the test corpus: 47.31%

Browsing Cost



Temporal Correlation

Method	Temporal Correlation	Argument Recall
Baseline: ordered by event reporting time	3.71%	27.63%
Method1: Single-document IE	44.02%	27.63%
Method2: 1+Cross-doc Event Coreference	46.15%	27.63%
Method3: 2+ Cross-doc Argument Refinement	55.73%	30.74%
Method4: 3 + Global Time Discovery	70.09%	33.07%

KBP Setup

Create/Expand

Source Collection

CNN U.S.

SEARCH

POWERED BY Google

Justice | Entertainment | Tech | Health | Living | Travel | Opinion | iReport | Money | Sports

Steve Jobs, Apple founder, dies

October 05, 2011 | By Brandon Griggs, CNN

Share | Twitter | Email

Recommend 153,356 people recommend this.

Steve Jobs, the visionary in the black turtleneck who co-founded Apple in a Silicon Valley garage, built it into the world's leading tech company and led a mobile-computing revolution with wildly popular devices such as the iPhone, died Wednesday. He was 56.

The hard-driving executive pioneered the concept of the personal computer and of navigating them by clicking onscreen images with a mouse. In more recent years, he introduced the iPod portable music player, the iPhone and the iPad tablet -- all of which changed how we consume content in the digital age.

Reference KB

Steve Jobs

Jobs holding a white iPhone 4 at Worldwide Developers Conference 2010

Born	Steven Paul Jobs February 24, 1955[1][2] San Francisco, California, U.S.[1][2]
Died	October 5, 2011 (aged 56)[2] Palo Alto, California, U.S.
Nationality	American
Occupation	Co-founder, Chairman and CEO, Apple Inc., CEO, Pixar, Co-founder and CEO, NeXT Inc.
Years active	1974–2011

Temporal Slot Filling (TSF)

- Regular temporal task, slot fills and temporal information must be gathered across the entire corpus
- Diagnostic task: the system is given a correct slot fill and must extract the time information for that slot fill from a single document
- Evaluation Metric
  - Let $<t_1, t_2, t_3, t_4>$ be system output, $<g_1, g_2, g_3, g_4>$ be gold standard

$$Q(S) = \frac{1}{4} \sum_i \frac{c}{c + |t_i - g_i|}$$

  - An error of *c* time units produces a 0.5 score
    - scores produced with c = 1 year
  - Each element in tuple is scored independently
  - For temporal SF task, a correct slot fill with temporal information ***t*** gets credit Q(S) (instead of 1)