



KYOTO: Event Extraction Task

Piek Vossen, VU University Amsterdam

2nd KYOTO Workshop, 25-28th January 2011, Gifu



Overview

- Goal of the task
- Organization of the task
- Triplet format
- Annotation tool
- Evaluation metrics
- Gold standard
- Baseline
- System results

Goal of the task

- Measure precision and recall of open text mining of events in the environment domain
- Events are identified by ranges of word tokens and relations to their participants
- We want to measure both comprehensiveness and precision of event modeling

Organization of the task

- October 1st:
 - Description of the task
 - Example of annotated file in triplet format
 - Annotation tool and documentation of the annotation
 - List of relations
 - Background documents in KAF format
 - Evaluation module
- November 23rd:
 - Release of 3 evaluation files in KAF format
- December 15th:
 - Deadline for results
- December 22nd:
 - Results available

Event representation

- Triplet representation:
 - Relation
 - List of word token identifiers for an event
 - List of word token identifiers for a participant
- Complex event representations are broken down into a list of triplets
- Word token identifiers are taken from the KAF representation of the evaluation documents

Example of KAF tokens

<wf wid="w207" sent="12" page="2">million</wf>

<wf wid="w208" sent="12" page="2">new</wf>

<wf wid="w209" sent="12" page="2">residents</wf>

<wf wid="w210" sent="12" page="2">move</wf>

<wf wid="w211" sent="12" page="2">into</wf>

<wf wid="w212" sent="12" page="2">the</wf>

<wf wid="w213" sent="12" page="2">Chesapeake</wf>

<wf wid="w214" sent="12" page="2">watershed</wf>

<wf wid="w215" sent="12" page="2">.</wf>

Example of triplet structure

```
<triplet id="5" relation="done-by">
  <eventids><event id="w210"/></eventids> <!-- move -->
  <participantids> <!-- million new residents -->
    <participant id="w207"/><participant id="w208"/><participant id="w209"/>
  </participantids>
</triplet>

<triplet id="5" relation="destination-of">
  <eventids><event id="w210"/></eventids> <!-- move -->
  <participantids><!-- into the Chesapeake watershed -->
    <participant id="w211"/><participant id="w212"/><participant id="w213"/>
    <participant id="w214"/>
  </participantids>
</triplet>
```

Relations for endurants (objects and substance)

patient	undergo a change
done-by	do the change
instrument	used as an instrument to make a change
use-of	used as a resource for making a change
has-state	the state expressed as an EVENT applies, for static EVENTS only
motion#path-of	the path along which a motion takes place
motion#destination-of	the target of a motion
motion#source-of	from which a motion departed
generic-location	a location at which an event takes place
product-of	comes into existence through an event
part-of	part of another object or substance

Relations for perdurants (events and states)

simple-cause-of	leads to or causes another process or state
result-of	the result of another process or state
state-of	State that applies to a process
purpose-of	the purpose of an event or object
goal-of	results of events that are intentional

Annotation tool

- Read any KAF file;
- Use any set of tags to code ranges of word tokens at 6 levels at most;
- Special export function to convert annotation to triplets:
 - Event tag
 - Event-scope tag
 - All other tags relate participants to events in the same event-scope

Evaluation module

- Reads a gold-standard triplet file and a system triplet file for the same KAF file;
- Compares each system-triplet with the gold-triplets to see if they cover the same events and the same participants
- Calculation of the scores:
 - Recall $R = \text{Nr. correct} / \text{Nr. Of gold-standard triplets}$
 - Precision $P = \text{Nr. correct} / \text{Nr. Of system triplets}$
 - F-measure $= 2 * (P.R / P+R)$

Evaluation module

- Other features:
 - Partial matches of token identifiers:
 - into Chesapeake watershed
 - Chesapeake watershed
 - watershed
 - Chesapeake
 - Event tokens outside the event tokens of the gold-standard are ignored → relevance is not considered
 - Duplicate matches (overlapping event ids and participant ids) are counted once
 - Average number of tokens per event and participant is calculated to prevent cheating

Gold standard

- Selected 3 short news messages from the daily-green website, restricted to the coastal areas: 646, 567, and 790 words respectively.
- Annotated by a PhD student trained in linguistics but not involved in the KYOTO project
- Triplets: 105, 47 and 101 triplets, respectively;
- Not every event or relation is relevant, e.g.: *Sadly, a new report indicates....*

Relations coded in gold-standard

	Nr. relations	Prop.
destination-of	5	1.98%
done-by	33	13.04%
generic-location	11	4.35%
has-state	31	12.25%
instrument	9	3.56%
part-of	23	9.09%
participant	16	6.32%
patient	87	34.39%
product-of	5	1.98%
purpose-of	5	1.98%
result-of	2	0.79%
simple-cause-of	7	2.77%
source-of	3	1.19%
state-of	14	5.53%
use-of	2	0.79%
Total	253	

Baseline

- Triplets between all heads of constituents in KAF in the same sentence, where each head is an event and all the others participants and vice versa;
- Most frequent relation is used as the default: patient relation;

System results

Overall results for all 3 files	Nr. triplets	Nr. of triplets in scope	Nr. of relations	Nr. correct	Recall	Precision	F-measure
GS	256	256	253				
Baseline	14902	1815	1815	50	0.20	0.03	0.05
AST	15	8	8	3	0.01	0.38	0.02
KAIST	165	62	60	34	0.13	0.57	0.22
KYOTO	3461	964	192	58	0.23	0.30	0.26

- Highest F-measure: KYOTO
- Highest precision: KAIST
 - subject/object verb relations only
 - done-by and patient only
- Highest recall: KYOTO
 - all relations and structures

System results per relation

Overall results per relation	GS		Baseline			AST			KAIST				KYOTO-1					
	#		#	C.	R.	P.	#	C.	R.	P.	#	C.	R.	P.	#	C.	R.	P.
destination-of	5	1.98%					2	2	40.00%	100.00%					2	1	20.00%	50.00%
done-by	33	13.04%					2	0	0.00%	0.00%	19	7	21.21%	36.84%	37	8	24.24%	21.62%
generic-location	11	4.35%													13	1	9.09%	7.69%
has-state	31	12.25%													9	4	12.90%	44.44%
instrument	9	3.56%																
part-of	23	9.09%													4	1	4.35%	25.00%
participant	16	6.32%																
patient	87	34.39%	1815	50	57.47%	2.75%	2	1	1.15%	50.00%	41	27	31.03%	65.85%	79	37	42.53%	46.84%
product-of	5	1.98%																
purpose-of	5	1.98%					2	0	0.00%	0.00%					4	0	0.00%	0.00%
result-of	2	0.79%																
simple-cause-of	7	2.77%													33	0	0.00%	0.00%
source-of	3	1.19%													1	1	33.33%	100.00%
state-of	14	5.53%													9	5	35.71%	55.56%
use-of	2	0.79%													1	0	0.00%	0.00%
Total	253		1815	50	19.76%	2.75%	8	3	1.19%	37.50%	60	34	13.44%	56.67%	192	58	22.92%	30.21%

Conclusion and discussion

- Annotation style is important:
 - PhD as trained linguist follows subject-object structure to assign done-by & patient
 - water flows (- control & + affected → patient?)
 - climate change threatens species (subject is event → cause-of?)
- Feature-based definition of relations:

water flows:	SA+, CC-, OA-	SA = self-affected
water intrudes:	SA+, CC-, OA+	CC = controls change
fish eat algae:	SA+, CC+, OA+	OA = other affected
algae kill plant:	SA-, CC-, OA+	
man kills fish:	SA-, CC+, OA+	

Ignoring the relation

Overall results ignoring the relations	Nr. triplets	Nr. of triplets in scope	Nr. of relations	Nr. correct	Recall	Precision	F-measure
GS	256	256	253				
Baseline	50	1815	1815	155	0.61	0.09	0.15
AST	3	8	8	7	0.03	0.88	0.05
KAIST	34	62	60	47	0.18	0.78	0.30
KYOTO	58	964	192	85	0.33	0.44	0.38

- Recall gains: 20%, 2%, 5% and 10%
- Precision gains: 6%, 50%, 21% and 14%
- F-measure gains: 10%, 3%, 8% and 12%