

RTE-6@TAC2010

The Sixth Recognizing Textual Entailment Challenge

Luisa Bentivogli (coordinator, CELCT & FBK-irst)
Danilo Giampiccolo (coordinator, CELCT)
Hoa Trang Dang (NIST)
Ido Dagan (Bar Ilan University)
Peter Clark (Vulcan Inc.)

- The RTE Challenge
- RTE-6 Main Task: RTE within a Corpus
 - RTE-6 Novelty Detection Subtask
 - Knowledge Resources and Tools for RTE
- RTE-6 KBP Validation Pilot Task
- Conclusion and Future Perspectives

- The RTE Challenge
- RTE-6 Main Task: RTE within a Corpus
 - RTE-6 Novelty Detection Subtask
 - Knowledge Resources and Tools for RTE
- RTE-6 KBP Validation Pilot Task
- Conclusion and Future Perspectives

Textual entailment is a directional relation between two text fragments:

- the entailing text, called ***T***(ext)
- the entailed text, called ***H***(ypothesis)

T entails H if, typically, a human reading T would infer that H is most likely true

- **YES**

T: The Christian Science Monitor named a US journalist kidnapped in Iraq as freelancer Jill Carroll.

H: Jill Carroll was abducted in Iraq.

- **NO**

T: The Christian Science Monitor named a US journalist kidnapped in Iraq as freelancer Jill Carroll.

H: Jill Carroll is the daughter of Mary Beth Carroll.

The RTE-6 Challenge



Main goals:

- Apply RTE systems to specific applications
- Move the RTE task towards more realistic scenarios

RTE-6 major innovations:

- New Main Task: Textual Entailment within a Corpus
(Piloted in RTE-5 - *Summarization setting*)
- New Novelty Detection Subtask (based on the Main Task)
- New KBP Validation Pilot
(*Knowledge Base Population setting*)
- Exploratory effort on resource evaluation extended to tools

RTE-6 Participants



- Number of participants: **18**
 - RTE-1: 18, RTE-2: 23, RTE-3: 26, RTE-4: 26, RTE-5: 21
- Provenance
 - ASIA: 7, EU: 6, NORTH AMERICA: 3, SOUTH AMERICA: 1, AUSTRALIA: 1
- Participants per task
 - Main Task: 18 (48 runs)
 - Novelty Detection Subtask: 9 (22 runs)
 - KBP Validation Pilot Task: 3 (10 runs)

- The RTE Challenge
- **RTE-6 Main Task: RTE within a Corpus**
 - RTE-6 Novelty Detection Subtask
- RTE-6 KBP Validation Pilot Task
- Knowledge Resources and Tools for RTE
- Conclusion and Future Perspectives

RTE-6 Main Task Description



- Given
 - a corpus
 - a hypothesis H
 - a set of "candidate" entailing sentences for that H retrieved by Lucene from the corpus
- RTE systems are required
 - to identify all the sentences among the candidate sentences that entail a given Hypothesis

RTE-6 Main Task Example



Topic 918: Betty Friedan

Hs SET

H380 :Betty Friedan is the author of "The Feminine Mystique."
H391 : "The Feminine Mystique" was published in 1963.
H401 : In 1962, Judy Mott was laid off from her job with Sears.

Document 1

S1: Betty Friedan, a founder of the modern feminist movement in the United States, died here Saturday of congestive heart failure, feminist leaders announced.

S2: She was 85.

S3: Friedan achieved prominence in 1963 with the publication of her book "The Feminine Mystique," which detailed the lives of American women who were expected to find fulfillment through the achievements of their husbands and children.

S4: The book sparked a movement for a re-evaluation of women's role in American society and is credited with laying the foundation of modern feminism.

S5: She was a founder of the National Organization for Women and a leading advocate of the Equal Rights Amendment, a proposed amendment to the US constitution banning sex-based discrimination, women's rights activists said.

S6: "The movement that Friedan's energy sparked continues to grow, and is bigger today than she could ever have dreamed ...

...

Document 2

S1: Betty Friedan, the visionary, combative feminist who launched a social revolution with her provocative 1963 book, "The Feminine Mystique," died Saturday, which was her 85th birthday.

S2: Friedan died of congestive heart failure at her home in Washington, D.C., according to Emily Bazelon, a cousin who was speaking for the family.

S3: She said Friedan had been in failing health for some time.

S4: Her best-selling book identified "the problem that has no name," the unhappiness of post-World War II American women unfulfilled by traditional notions of female domesticity.

S5: Melding sociology and humanistic psychology, the book became the cornerstone of one of the last century's most profound movements, unleashing the first full flowering of American feminism since the 1800s.

S6: It gave Friedan, an obscure suburban New York housewife and freelance writer, the mantle to...

...

Document 3

S26: What is perhaps most surprising, though, is not that feminists like Hirshman believe homemaking is second-class drudgery, but that so many people still get worked up over the issue.

S27: After all, feminist thinkers have been proclaiming the need to free women from the bondage of housework for a long time..

S28: It is, as Hirshman freely acknowledges, precisely what Friedan argued in "The Feminine Mystique," first published more than 40 years ago.

S29 "The only kind of work which permits an able woman to realize her abilities fully," Friedan wrote, "is the kind that was forbidden by the feminine mystique, the lifelong commitment to an art or science, to politics or profession."

S30: Not homemaking, not motherhood.

S31: In an interview, Hirshman said that in the course of researching a book, she began to wonder when feminism switched from offering a clear blueprint for liberation to choosing from Column A and Column B.

...

RTE-6 Main Task Example



Topic 918: Betty Friedan

H380: Betty Friedan is the author of "The Feminine Mystique."

H380: Betty Friedan is the author of "The Feminine Mystique"

H401: In 1962, Judy Mott was laid off from her job with Sears.

Document 1

S1: Betty Friedan, a founder of the modern feminist movement in the United States, died here Saturday of congestive heart failure, feminist leaders announced.

S2: She was 85.

S3: Friedan achieved prominence in 1963 with the publication of her book "The Feminine Mystique," which detailed the lives of American women who were expected to find fulfillment through the achievements of their husbands and children.

S4: The book sparked a movement for a re-evaluation of women's role in American society and is credited with laying the foundation of modern feminism.

S5: She was a founder of the National Organization for Women and a leading advocate of the Equal Rights Amendment, a proposed amendment to the US constitution banning sex-based discrimination, women's rights activists said.

S6: "The movement that Friedan's energy sparked continues to grow, and is bigger today than she could ever have dreamed ...

...

Document 2

S1: Betty Friedan, the visionary, combative feminist who launched a social revolution with her provocative 1963 book, "The Feminine Mystique," died Saturday, which was her 85th birthday.

S2: Friedan died of congestive heart failure at her home in Washington, D.C., according to Emily Bazelon, a cousin who was speaking for the family.

S3: She said Friedan had been in failing health for some time.

S4: Her best-selling book identified "the problem that has no name," the unhappiness of post-World War II American women unfulfilled by traditional notions of female domesticity.

S5: Melding sociology and humanistic psychology, the book became the cornerstone of one of the last century's most profound movements, unleashing the first full flowering of American feminism since the 1800s.

S6: It gave Friedan, an obscure suburban New York housewife and freelance writer, the mantle to...

...

Document 3

S26: What is perhaps most surprising, though, is not that feminists like Hirshman believe homemaking is second-class drudgery, but that so many people still get worked up over the issue.

S27: After all, feminist thinkers have been proclaiming the need to free women from the bondage of housework for a long time..

S28: It is, as Hirshman freely acknowledges, precisely what Friedan argued in "The Feminine Mystique," first published more than 40 years ago.

S29: "The only kind of work which permits an able woman to realize her abilities fully," Friedan wrote, "is the kind that was forbidden by the feminine mystique, the lifelong commitment to an art or science, to politics or profession."

S30: Not homemaking, not motherhood.

S31: In an interview, Hirshman said that in the course of researching a book, she began to wonder when feminism switched from offering a clear blueprint for liberation to choosing from Column A and Column B.

...

RTE-6 Main Task Example



Topic 918: Betty Friedan

H380: Betty Friedan is the author of "The Feminine Mystique."

H380: Betty Friedan is the author of "The Feminine Mystique"

H401: In 1962, Judy Mott was laid off from her job with Sears.

Document 1

S1: Betty Friedan, a founder of the modern feminist movement in the United States, died here Saturday of congestive heart failure, feminist leaders announced.

S2: She was 85.

S3: Friedan achieved prominence in 1963 with the publication of her book "The Feminine Mystique," which detailed the lives of American women who were expected to find fulfillment through the achievements of their husbands and children.

S4: The book sparked a movement for a re-evaluation of women's role in American society and is credited with laying the foundation of modern feminism.

S5: She was a founder of the National Organization for Women and a leading advocate of the Equal Rights Amendment, a proposed amendment to the US constitution banning sex-based discrimination, women's rights activists said.

S6: "The movement that Friedan's energy sparked continues to grow, and is bigger today than she could ever have dreamed ...

...

Document 2

S1: Betty Friedan, the visionary, combative feminist who launched a social revolution with her provocative 1963 book, "The Feminine Mystique," died Saturday, which was her 85th birthday.

S2: Friedan died of congestive heart failure at her home in Washington, D.C., according to Emily Bazelon, a cousin who was speaking for the family.

S3: She said Friedan had been in failing health for some time.

S4: Her best-selling book identified "the problem that has no name," the unhappiness of post-World War II American women unfulfilled by traditional notions of female domesticity.

S5: Melding sociology and humanistic psychology, the book became the cornerstone of one of the last century's most profound movements, unleashing the first full flowering of American feminism since the 1800s.

S6: It gave Friedan, an obscure suburban New York housewife and freelance writer, the mantle to...

...

Document 3

S26: What is perhaps most surprising, though, is not that feminists like Hirshman believe homemaking is second-class drudgery, but that so many people still get worked up over the issue.

S27: After all, feminist thinkers have been proclaiming the need to free women from the bondage of housework for a long time..

S28: It is, as Hirshman freely acknowledges, precisely what Friedan argued in "The Feminine Mystique," first published more than 40 years ago.

S29: "The only kind of work which permits an able woman to realize her abilities fully," Friedan wrote, "is the kind that was forbidden by the feminine mystique, the lifelong commitment to an art or science, to politics or profession."

S30: Not homemaking, not motherhood.

S31: In an interview, Hirshman said that in the course of researching a book, she began to wonder when feminism switched from offering a clear blueprint for liberation to choosing from Column A and Column B.

...

RTE-6 Main Task Example



Topic 918: Betty Friedan

H380: Betty Friedan is the author of "The Feminine Mystique"

H380: Betty Friedan is the author of "The Feminine Mystique"

H401: In 1962, Judy Mott was laid off from her job with Sears.

Document 1

S1: Betty Friedan, a founder of the modern feminist movement in the United States, died here Saturday of congestive heart failure, feminist leaders announced.

S2: She was 85.

S3: Friedan achieved prominence in 1963 with the publication of her book "The Feminine Mystique," which detailed the lives of American women ...

have dreamed ...

...

Document 2

S1: Betty Friedan, the visionary, combative feminist who launched a social revolution with her provocative 1963 book, "The Feminine Mystique," died ...

1800s.

S6: It gave Friedan, an obscure suburban New York housewife and freelance writer, the mantle to...

...

Document 3

S26: What is perhaps most surprising, though, is not that feminists like Hirshman believe homemaking is second-class drudgery, but that so many people still get worked up over the issue.

S27: After all, feminist thinkers have been proclaiming the need to free women from the bondage of housework for a long time..

S28: It is, as Hirshman freely acknowledges, precisely what Friedan argued in her book "The Feminine Mystique," first published...

RTE1-5 vs. RTE6 Main Task



RTE1-5

- RTE on isolated T-H pairs
- T-H pairs drawn from multiple applications
- T and H do not contain references to information outside the pair itself
- The distribution of entailment is determined a priori

RTE6

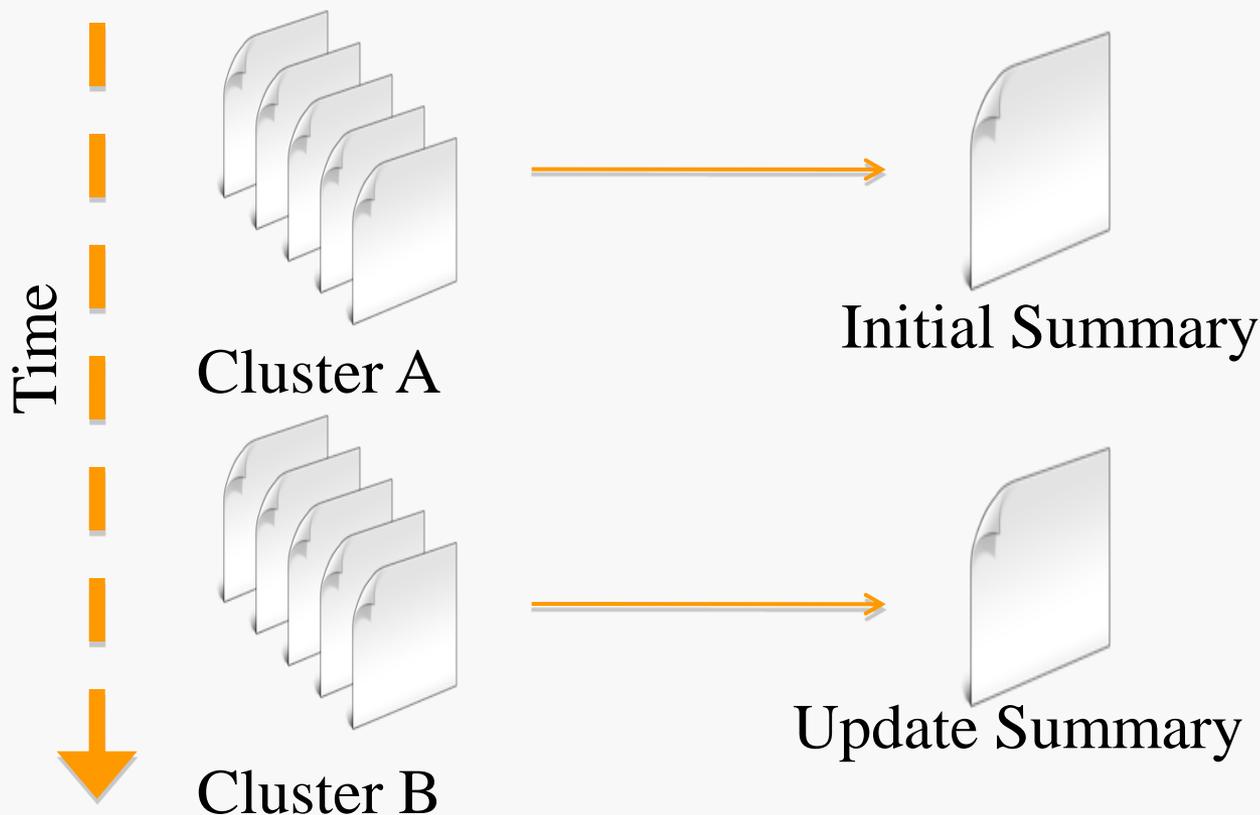
- RTE within a corpus
- Summarization application setting
- Both T and H are to be interpreted within the context of the corpus
- Reflects the natural distribution of entailment in a corpus

RTE-6 Main Data Set (1/3)



TAC 2009 SUM *Update* scenario

For each topic:



RTE-6 Main Task scenario

For each topic:

- a corpus
 - Cluster “A” documents
- a set of Hypotheses
 - up to 30 standalone sentences:
 - based on the “B” summary sentences of the 10 best scoring SUM systems
 - based directly on Cluster “B” sentences

For each H of the topic:

- Up to 100 “candidate” entailing sentences
 - Information Retrieval filtering phase:
 - The H is the query
 - The corpus sentences are “the documents” to be retrieved for the query
 - the 100 top-ranked sentences are selected as candidates (80% of all the entailing sentences in the corpus)
 - LUCENE text search engine (v. 2.9.1):
 - *StandardAnalyzer*, Boolean “OR” query, Default Lucene ranking

RTE-5 Pilot

- Generic SUM setting
- Hs are taken from human-authored summaries of the "A" documents.
- All Hs have entailing sentences in the corpus
- Task performed on the whole corpus

RTE-6 Main

- Update SUM setting
- Hs are (mostly) taken from automatic summaries of the "B" documents
- Some Hs have no entailing sentences in the corpus
- Task performed just on the Lucene-retrieved candidates (IR filter)

Data Set Composition



DEVELOPMENT SET		TEST SET	
Topics	10	Topics	10
Hypotheses	211	Hypotheses	243
Entailment: yes no	122 89	Entailment: yes no	143 100
Summaries: yes no	183 28	Summaries: yes no	199 44
Annotations	15,955	Annotations	19,972
“entailment” judg.	810	“entailment” judg.	945

- 3 annotations for the whole data set
- IAA (*Kappa*): 98.83% (Dev), 97.83% (Test)

Main Task Evaluation



18 participants (48 runs)

- Evaluation measures:
 - Precision, Recall, F-measure (micro-averaged)
- IR Baselines:

	Precision	Recall	F1
Lucene_5	30.78	39.58	34.63
Lucene_10	21.87	56.19	31.49
Lucene_15	17.15	66.03	27.23
Lucene_20	14.23	72.70	23.80
Lucene_100	4.73	100.00	9.03

Best Results



Team	Precision	Recall	F-measure
PKUTM2	68.57	36.93	48.01
deb_iitb2	53.43	42.86	47.56
IKOMA1	39.71	51.43	44.81
FBK_irst3	43.46	46.03	44.71
Boeing1	55.1	36.61	43.99
DirRelCond21	38.99	41.8	40.35
DFKI2	55.94	30.9	39.81
SJTU_CIT3	34.35	46.67	39.57
BIU1	37.54	37.46	37.5
JU_CSE_TAC1	38.63	31.64	34.79
Baseline_Lucene5	30.78	39.58	34.63
Baseline_LuceneAll	4.73	100.00	9.03

Results: F-measure statistics



F-measure	Best runs
Highest	48.01
Median	36.14
Average	33.77
Lowest	11.60

WRT RTE-5 Pilot Task:

- overall improvement of results
- best baseline below the Median value
(vs. RTE-5 Pilot: best baseline above the Highest score)

- The RTE Challenge
- RTE-6 Main Task: RTE within a Corpus
 - RTE-6 Novelty Detection Subtask
 - Knowledge Resources and Tools for RTE
- RTE-6 KBP Validation Pilot Task
- Conclusion and Future Perspectives

RTE-6 Novelty Detection Subtask

Motivation:

- Specifically address the needs of the SUM **Update** Task, where it is necessary to distinguish between novel and non novel information
- RTE engines could help summarization systems to filter out non-novel sentences from their summaries

RTE-6 Novelty Detection Subtask

Task:

Judge if the information contained in each H (from Cluster B) is **novel** with respect to the information contained in the set of (Cluster A) candidate entailing sentences

– If a given H:

- has entailing sentences = information is NOT novel
- has not entailing sentences = information is novel

RTE-6 Novelty Detection Subtask



Based on the Main Task:

- Uses a subset of the Main Task data
 - only the Hs taken from the automatic summaries
- Same output format/annotation
 - the novelty detection decision is derived automatically from the number of entailing sentences for each H

Differences:

- Systems are specifically tuned for novelty detection
- Specific scoring metrics designed for assessing novelty detection

Data Set Composition



DEVELOPMENT SET		TEST SET	
Topics	10	Topics	10
Hypotheses Novel: yes no	183 89 94	Hypotheses Novel: yes no	199 100 99
“entailing” judgm.	707	“entailing” judgm.	723

- IAA (*Kappa*): 98.21% (Dev), 97.68% (Test)

9 participants (22 runs)

- **Primary score: Novelty Detection evaluation**
 - Micro Averaged Precision, recall and F-measure computed on the binary novel/non-novel decision
 - derived automatically from the number of entailing sentences provided by the systems
- **Secondary score: Justification evaluation**
 - measures the quality of the justifications provided for non-novel Hs
 - Micro-averaged Precision, Recall and F-measure on the set of all the sentences extracted as entailing the Hs

Best Results – Primary Score



	Novelty Detection Evaluation		
Run	Precision	Recall	F-measure
PKUTM1	72.39	97	82.91
IKOMA2	79.44	85	82.13
JU_CSE_TAC1	80.58	83	81.77
DFKI2	73.5	86	79.26
UAIC20103	73.28	85	78.7
Boeing2	66.43	93	77.5
BIU1	73.53	75	74.26
Baseline_all_new	50.25	100	66.89
SINAI1	65.62	42	51.22
Sagan1	46.15	42	43.98

Best Results – Secondary Score



	Justification Evaluation		
Run	Precision	Recall	F-measure
PKUTM2	67.75	37.48	48.26
IKOMA2	47.63	43.02	45.2
Boeing1	50.62	34.02	40.69
DFKI1	48.85	29.46	36.76
BIU1	34.83	36.38	35.59
JU_CSE_TAC1	40.92	29.6	34.35
UAIC20101	21.91	27.94	24.56
SINAI1	20.72	23.79	22.15
Sagan1	2.15	16.18	3.79

Results: F-measure statistics



	Novelty Detection	Justification (non novel Hs)
F-measure	Best runs	Best runs
Highest	82.91	48.26
Median	78.70	35.59
Average	72.41	32.38
Lowest	43.98	3.79

Assessing RTE impact on SUM



- New Test Set: only Hs contained in the summaries of the SUM Update *best-scoring* system (65 Hs: 30 novel and 35 non-novel)

	P	R	F1
Best-scoring RTE run	74.36	96.67	84.06
“All-novel” baseline	46.15	100	63.15

- RTE has a potential to improve filtering non-novel information

- The RTE Challenge
- RTE-6 Main Task: RTE within a Corpus
 - RTE-6 Novelty Detection Subtask
 - Knowledge Resources and Tools for RTE
- RTE-6 KBP Validation Pilot Task
- Conclusion and Future Perspectives

Knowledge Resources and Tools for RTE



An exploratory effort aimed at studying the relevance of knowledge resources and tools in recognizing TE

- ***Ablation Tests*** for all knowledge resources and tools used in Main Task runs:
 - remove one module at a time from a system, and re-run the system on the test set with the other modules, except the one tested
 - ! Remove only knowledge resources or tools
 - ! Remove one resource or tool at a time

- 78 ablation tests submitted (by 17 teams)
 - 20 tests did not specifically address knowledge resources or tools
 - In 1 test a combination of different resources/components was removed
- 57 ablation tests conformant to the requirements
 - 46 tests for 9 different resources
 - 11 tests for 5 different tools

Ablation Tests - Resources

Ablated Knowledge Resource	# of Ablation Tests	Impact on Systems	
		Positive	Negative
WordNet	22	14 (+7.54%)	8 (-1.36%)
VerbOcean	7	5 (+1.28%)	2 (-2.58%)
Wikipedia	5	4 (+2.25%)	1 (-3.58%)
FrameNet	3	-	3 (-1.25%)
Dirt	4	1 (+3.97%)	3 (-1.09%)
CatVar	1	1 (+0.63%)	-
Synonym&Acronym Dictionary	1	-	1 (-0.76%)
Dependency-based Thesaurus	1	-	1 (-13.56%)
Proximity -based Thesaurus	1	-	1 (-7.79%)

Ablation Tests - Resources

Ablated Knowledge Resource	# of Ablation Tests	Impact on Systems	
		Positive	Negative
WordNet	22	14 (+7.54%)	8 (-1.36%)
VerbOcean	7	5 (+1.28%)	2 (-2.58%)
Wikipedia	5	4 (+2.25%)	1 (-3.58%)
FrameNet	3	-	3 (-1.25%)
Dirt	4	1 (+3.97%)	3 (-1.09%)
CatVar	1	1 (+0.63%)	-
Synonym&Acronym Dictionary	1	-	1 (-0.76%)
Dependency-based Thesaurus	1	-	1 (-13.56%)
Proximity -based Thesaurus	1	-	1 (-7.79%)

Ablation Tests - Tools

Ablated Tool	# of Ablation Tests	Impact on Systems	
		Positive	Negative
Coreference Resolver	3	1 (+0.17%)	2 (-1.21%)
Named Entities Recognizer	5	4 (10.98%)	1 (-1.23%)
POS Tagger	1	1 (4.99%)	-
Parser	1	-	1 (-1.76%)
Name Normalization	1	1 (0.65%)	-

Remarks on the initiative



- WRT RTE-5 (resources only):
 - More ablation tests (+9)
 - More resources (+2)
 - Higher average impacts
 - Some trends confirmed (WordNet, VerbOcean, Wikipedia)
- Lesson learned
 - the value of a resource is very much dependent on how that resource is used and how it integrates with the rest of the system
 - Need for a deeper comprehension of the usage of the resources and tools

- The RTE Challenge
- RTE-6 Main Task: RTE within a Corpus
 - RTE-6 Novelty Detection Subtask
 - Knowledge Resources and Tools for RTE
- **RTE-6 KBP Validation Pilot Task**
- Conclusion and Future Perspectives

The RTE-6 KBP Validation Pilot Task

Motivations:

- analyze the potential utility of RTE systems in another real NLP application scenario, i.e. the *Knowledge Base Population Slot Filling task*
- use Textual Entailment techniques to validate the output of an NLP system (similar to the AVE experiment in QA)

The KBP Slot Filling Task



Given an *entity* in a knowledge base and an *attribute (slot)* for that entity:

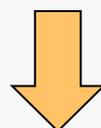
- find in a large corpus the correct value (filler) for that attribute
- return the extracted information together with a corpus document supporting it as a correct slot filler

The RTE-6 KBP Validation Task



- Initial assumption: an extracted slot filler is correct *if and only if* the supporting document entails a hypothesis summarizing the slot filler
- Task : determine whether a candidate slot filler is supported in the associated document using entailment techniques.

Each slot filler returned by KBP systems



1 RTE evaluation pair, where:

- T is the entire document supporting the slot filler
- H is a set of synonymous sentences, representing different realizations of the slot filler

Data Set Creation: example

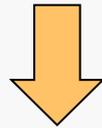


KBP SYSTEM INPUT

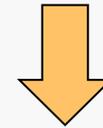
Target Entity: *Chris Simcox*
Slot: Residences
Document collection

KBP SYSTEM OUTPUT

Slot Filler: “*Tucson, Ariz.*”
Supporting Document:
NYT_ENG_20050919.0130
.LDC2007T07



RTE EVALUATION PAIR



T: NYT_ENG_20050919.0130.LDC2007T07

H:	H1:	<i>Chris Simcox lives in Tucson, Ariz.</i>
	H2:	<i>Chris Simcox has residence in Tucson, Ariz.</i>
	H3:	<i>Tucson, Ariz. is the place of residence of Chris Simcox</i>
	H4:	<i>Chris Simcox resides in Tucson, Ariz.</i>
	H5:	<i>Chris Simcox’s home is in Tucson, Ariz.</i>

Hypotheses Creation



Manually created templates

Attribute: *origin*

Target entity: *person*

Instantiated Hs

Slot filler: *Canadian*

Target person entity: *Chris Simcox*

Template 1: *X's origins are in Y*

Template 2: *X comes from Y*

Template 3: *X is from Y*

Template 4: *X origins are Y*

Template 5: *X has Y origins*

Template 6: *X is of Y origin*

H1 *Chris Simcox* origins are in *Canadian*

H2 *Chris Simcox* comes from *Canadian*

H3 *Chris Simcox* is from *Canadian*

H4 *Chris Simcox* origins are *Canadian*

H5 *Chris Simcox* has *Canadian* origin

H6 *Chris Simcox* is of *Canadian* origin

Gold Standard Creation



KBP assessments



(automatically)

RTE gold standard annotations

KBP JUDGMENTS
(4-valued)

ENTAILMENT VALUES
(2-valued)

Correct



YES

Redundant



YES

Wrong



NO

Inexact



(not included)

- RTE evaluation pair
 - T is an entire document
 - H is *a set* of synonymous sentences, possibly ungrammatical
- *(Semi-)automatic* generation
 - Data Set
 - from KBP outputs
 - Gold Standard
 - from KBP output assessments

Data Set Composition



DEVELOPMENT SET		TEST SET	
KBP '09 Slot Filling Task assessments	10,416	KBP '10 Slot Filling Task assessments	24,014
Pairs	9,462	Pairs	23,192
Positive examples	694	Positive examples	2,034
Negative examples	8,768	Negative examples	21,158

Removed pair types:

GPE; “inexact”; “NO_RESPONSE”; duplicates; speech transcriptions; “other_family” slot (only Test Set).

2 TYPES OF SUBMISSIONS:

- generic systems (no adaptation)
- tailored systems (adapted for specific slots)

PARTICIPANTS : 3

SUBMITTED RUNS: 10

- 7 generic
- 3 tailored

EVALUATION MEASURES:

Micro-Averaged Precision, Recall, F-measure

Baseline: All Ts classified as entailing the corresponding H

This baseline:

- reflects the cumulative performance of all KBP Slot Filling Systems
- indicates the percentage of entailing pairs in the Test Set

Results



TYPE	RUN	P	R	F1
Generic	FBK_irst1	20.46	33.82	25.5
	JU_CSE_TAC2	22.4	13.96	17.2
	Baseline	8.77	100	16.13
	BIU2	10.02	39.48	15.98
Tailored	JU_CSE_TAC2	24.32	51.67	33.07

KBP Slot Filling results:

Highest System F1: 65.78

2nd Highest System F1: 29.15

Median F1: 14.12

- The RTE Challenge
- RTE-6 Main Task: RTE within a Corpus
 - RTE-6 Novelty Detection Subtask
 - Knowledge Resources and Tools for RTE
- RTE-6 KBP Validation Pilot Task
- **Conclusion and Future Perspectives**

- New Main Task
 - interaction between the RTE and SUM task
 - textual entailment recognition performed on a real corpus
 - natural distribution of entailment
- Evaluation of Knowledge Resources and Tools
 - very positive response
 - first step towards sharing and reuse
- KBP Validation Pilot Task
 - interaction between the RTE and KBP task

See you all at the RTE Planning Session

Thank you!