Recognizing Textual Entailment

Group 8
Maunik Shah
Hemant Adil
Akanksha Patel

Given	John Smith spent six years in jail for his role in a number of violent armed robberies.
Is it true?	John Smith was charged with two or more violent crimes.

The Question is:

- Given a text fragment is true, can we predict the truth value of another text fragment?

- This relationship among texts is textual entailment.

Order of Presentation

- What is Textual Entailment
- Motivation
- Basic Process of RTE
- PASCAL RTE Challenges
- RTE Approaches
- ML based approach
- Applications
- Conclusions

Textual Entailment

- A text hypothesis (h) is said to entail a text (t) if, a human reading t would infer that h is most likely true. [1]
- “h entails y” represented as h=>y

Probilistic Interpretation

- *t probabilistically entails h* if:
  P(*h is true* | *t*) > P(*h is true*)
  - *t* increases the likelihood of *h* being true
- P(*h is true* | *t*): *entailment confidence*

From: Textual Entailment, Ido Dagan, Dan Roth, Fabio Zanzotto, ACL 2007

Role of knowledge

- For textual entailment to hold we require:
  - *text* AND *knowledge* $\Rightarrow$ *h*

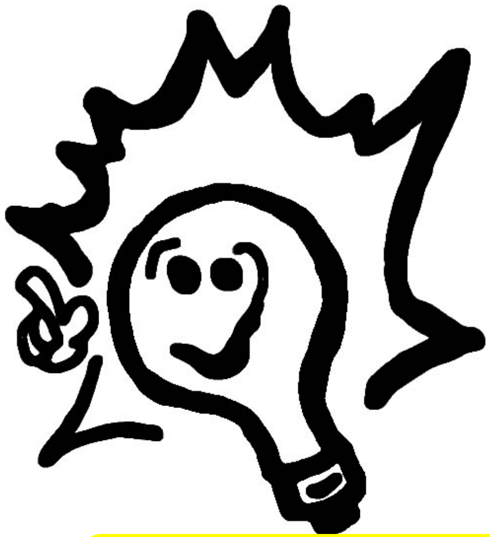
  but
  - *knowledge* should not entail *h* alone

- Systems are not supposed to validate *h*'s truth regardless of *t* (e.g. by searching *h* on the web)

From: Textual Entailment, Ido Dagan, Dan Roth, Fabio Zanzotto, ACL 2007

Motivation

- Text applications require *semantic* inference
- A common framework for applied semantics is needed, b missing
- Textual entailment may provide such framework



Motivation

From: Textual Entailment, Ido Dagan, Dan Roth, Fabio Zanzotto, ACL 2007

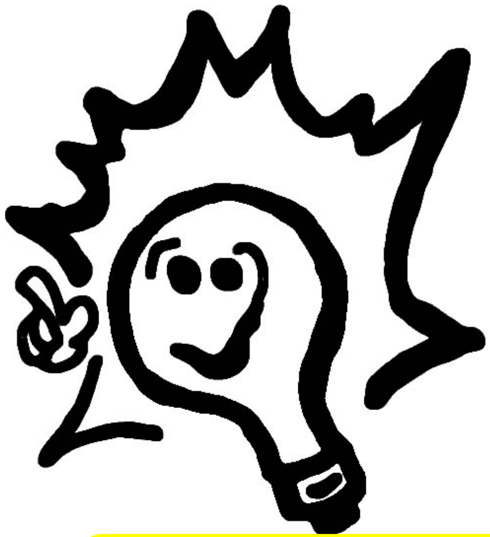
Why textual entailment?

- <u>Variability of semantic Expression</u>

"Same meaning can be inferred from different texts."

- <u>Ambiguity in meaning of words</u>

"Different meanings can be inferred from same text"

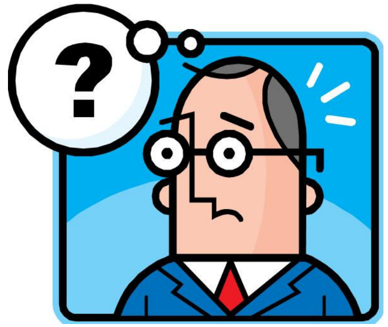


Motivation

Need of common solution for modeling language variability in NLP tasks…

Recognizing Textual Entailment

Why Textual Entailment a Challenge?



Two main underlying problems:

- Paraphrasing
- Strict Entailment

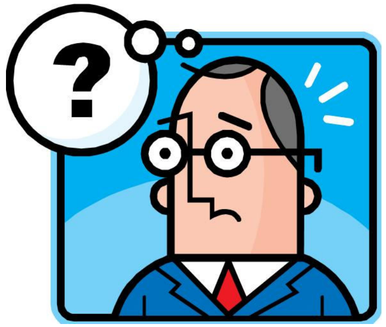
Why Textual Entailment a Challenge?

- Paraphrasing:

The hypothesis h carries a fact fh that is also in the target text t but is expressed with different words.

“the cat devours the mouse"
is a paraphrase of
"the cat consumes the mouse"

Why Textual Entailment a Challenge?



- Strict entailment:

Target sentences carry different fact, can be inferred from the other.

There is strict entailment between

"the cat devours the mouse" → "the cat eats the mouse"

Basic Process of Textual Entailment

Entails

Subsumed by

$\subseteq$

Eyeing the huge market potential, currently led by Google, Yahoo took over ~~search company~~ Overture Services Inc. last year

Yahoo acquired Overture

Overture is a search company

Google is a search company

Google owns Overture

....

Phrasal verb paraphrasing

Entity matching

Alignment

Semantic Role Labeling

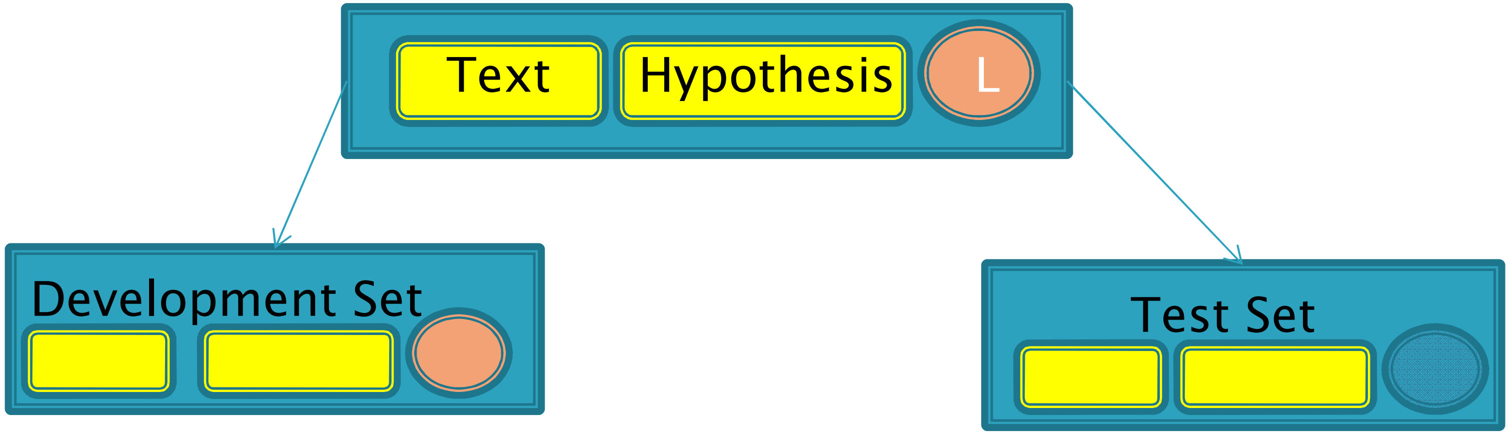
Integration

From: Textual Entailment, Ido Dagan, Dan Roth, Fabio Zanzotto, ACL 2007

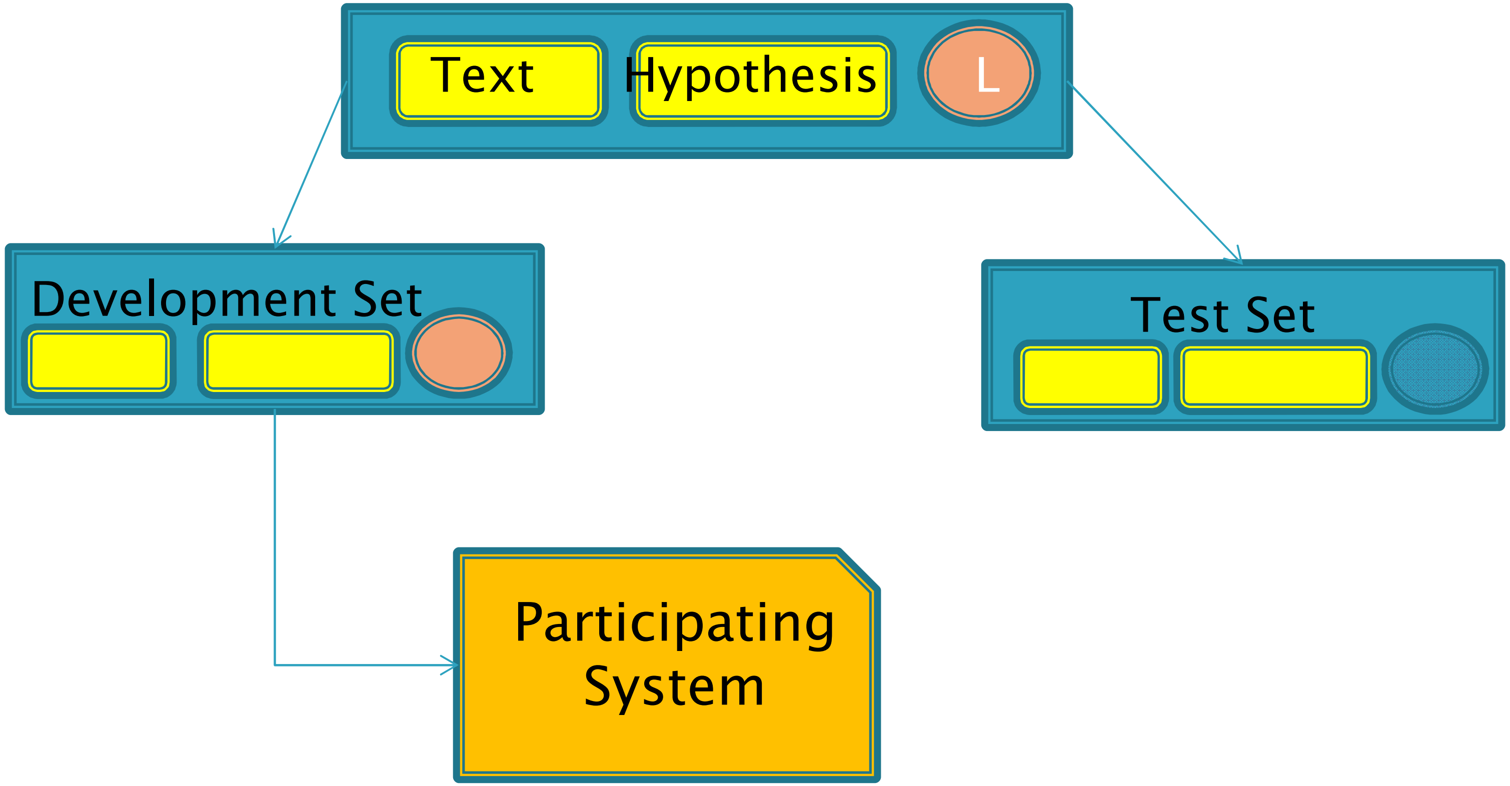
PASCAL RTE Challenges

- [Goal]
- to provide opportunity for presenting and comparing possible approaches for modeling *textual entailment.*

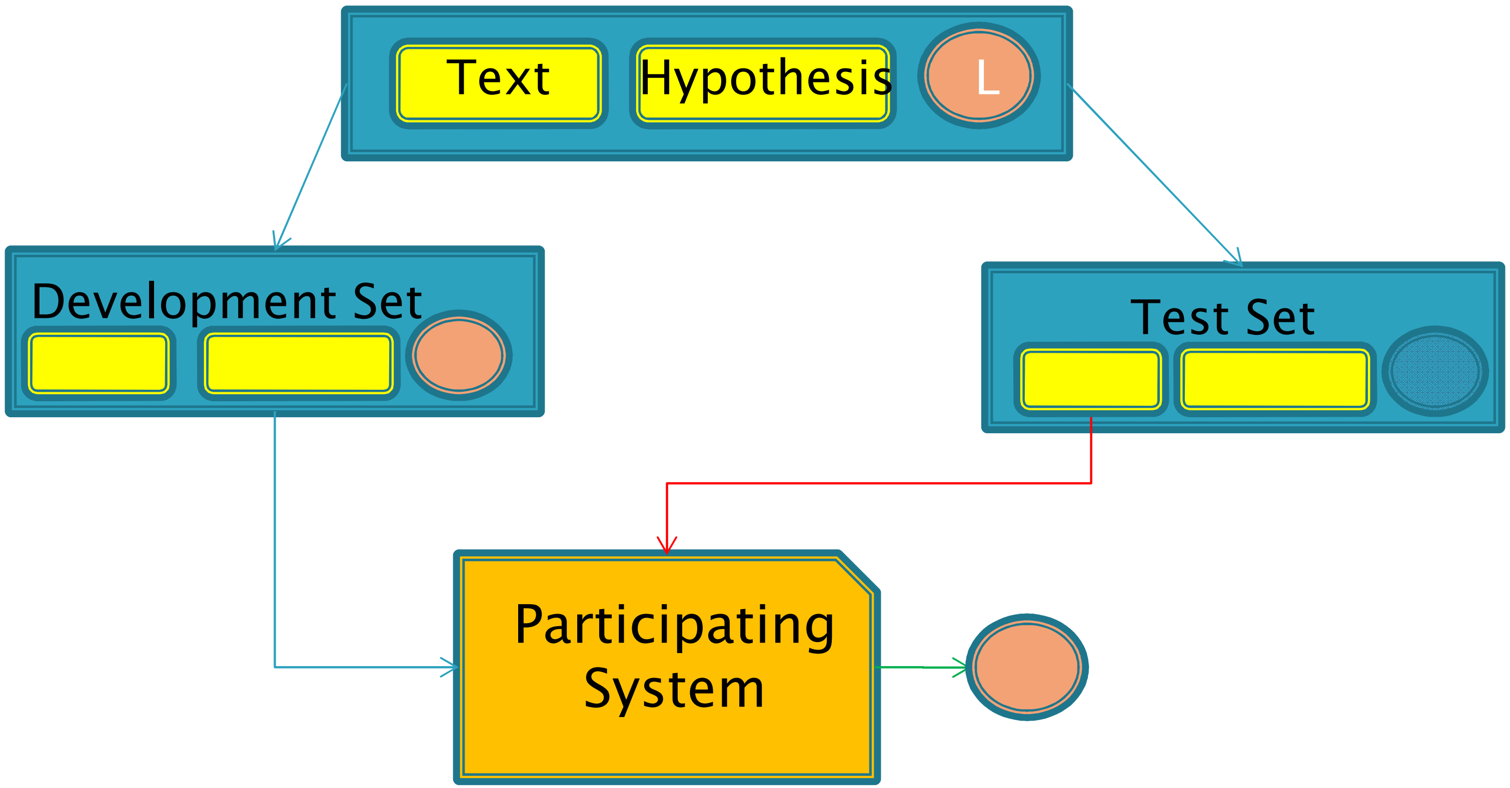
PASCAL RTE Challenges



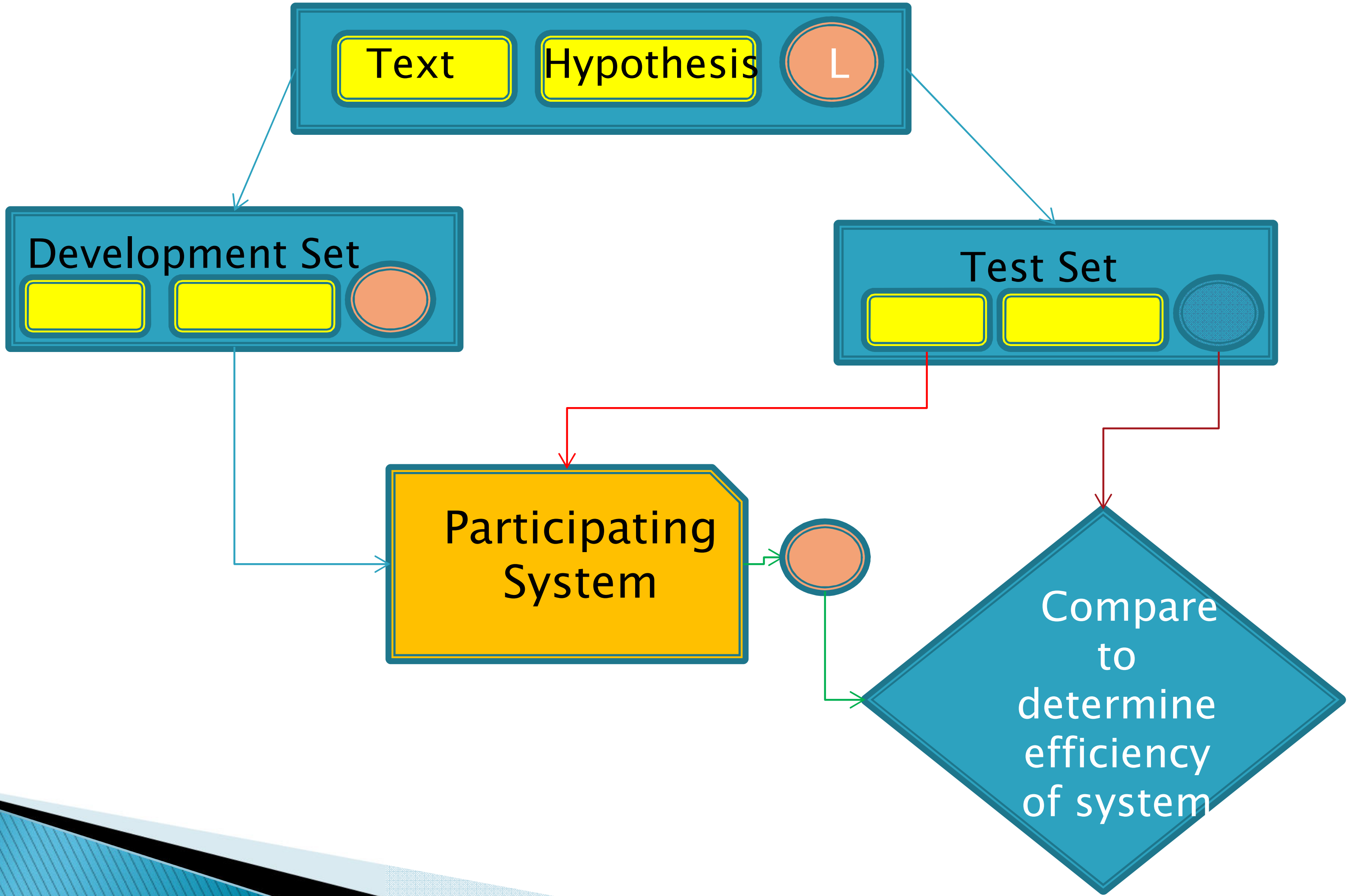
PASCAL RTE Challenges



PASCAL RTE Challenges



PASCAL RTE Challenges



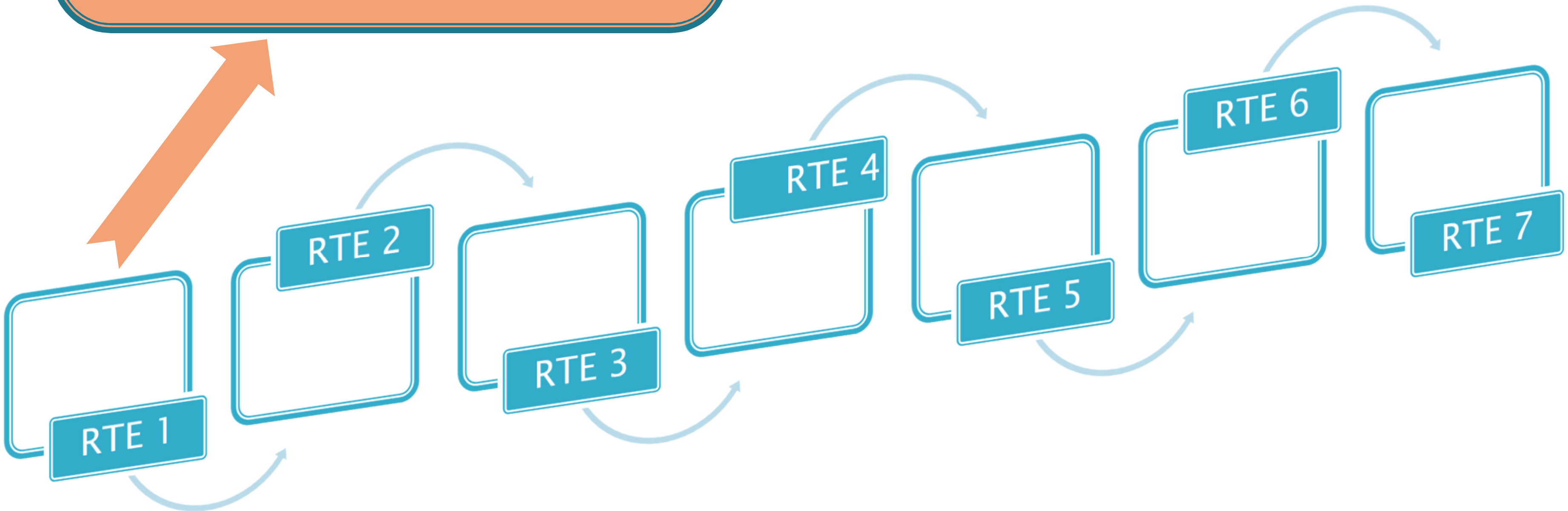
Main Task

Recognizing Entailment

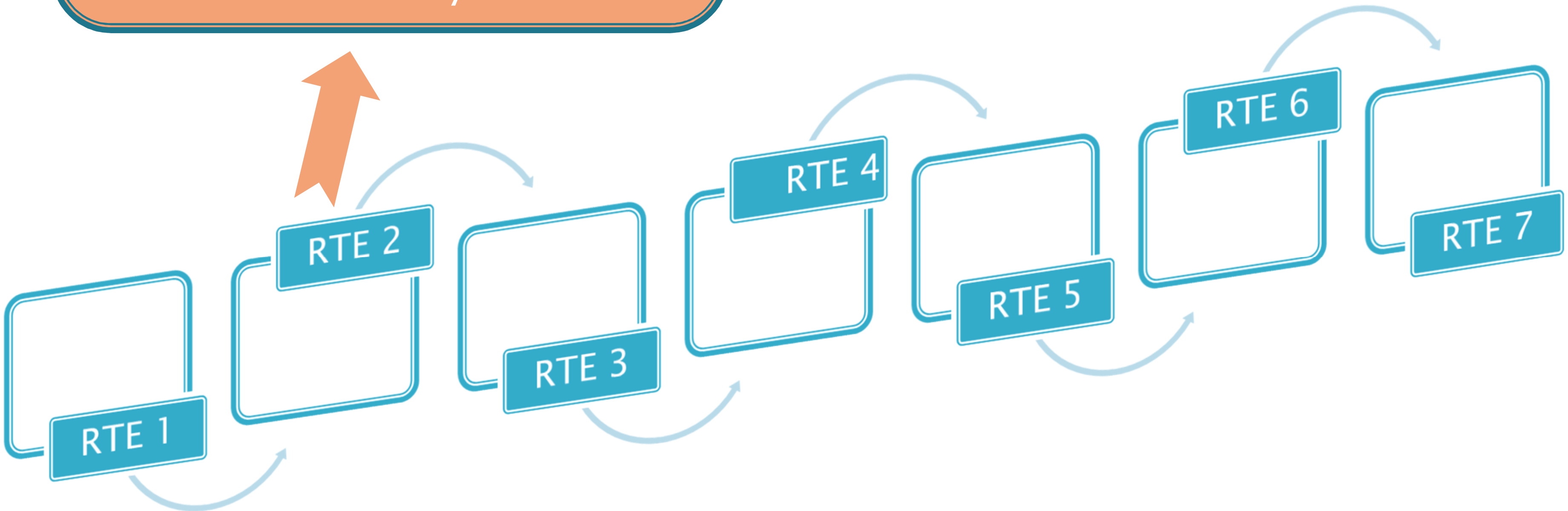
2 way entailment

Best Accuracy : 70%

Average Accuracy: 50 to 60%



Main Task
Recognizing Entailment
(on more realistic examples
from real systems)
2 way entailment
Best Accuracy : 75%



Improvement in accuracy
as compared to RTE1

Main Task

Recognizing Entailment

Pilot Task

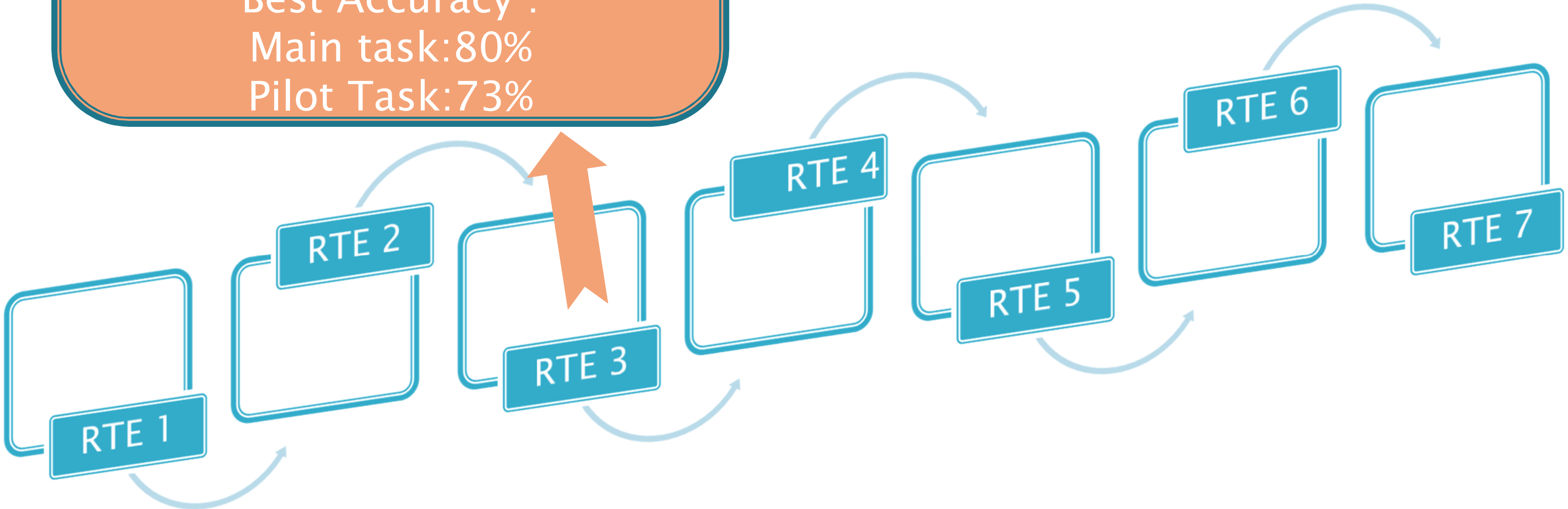
Extending Evaluation of Inference Text

2 way entailment

Best Accuracy :

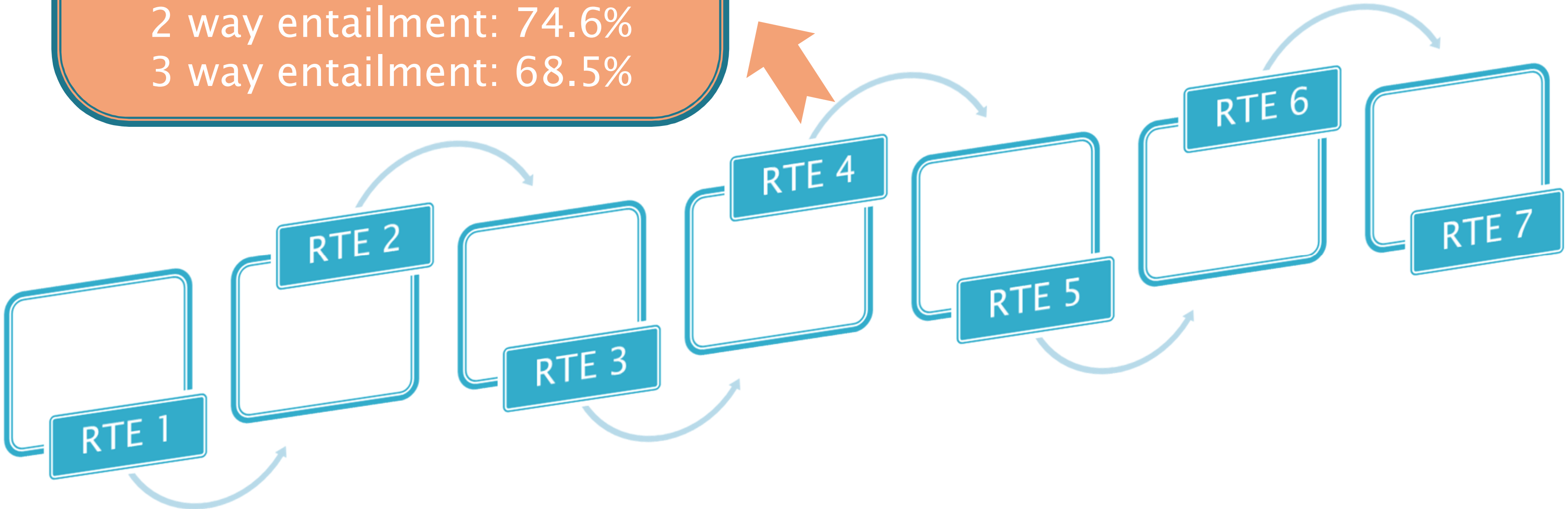
Main task:80%

Pilot Task:73%



Improvement in accuracy as compared to RTE1 and RTE2

Main Task
Recognizing Entailment
(Development Set not given before-hand)
2 way and 3 way entailment
Best Accuracy :
2 way entailment: 74.6%
3 way entailment: 68.5%



Reduction in accuracy as compared to previous campaigns

Main Task

Recognizing Entailment (length of text increased)

Pilot Task:

solving TE in summarization and Knowledge Base Population (KBP) Validation

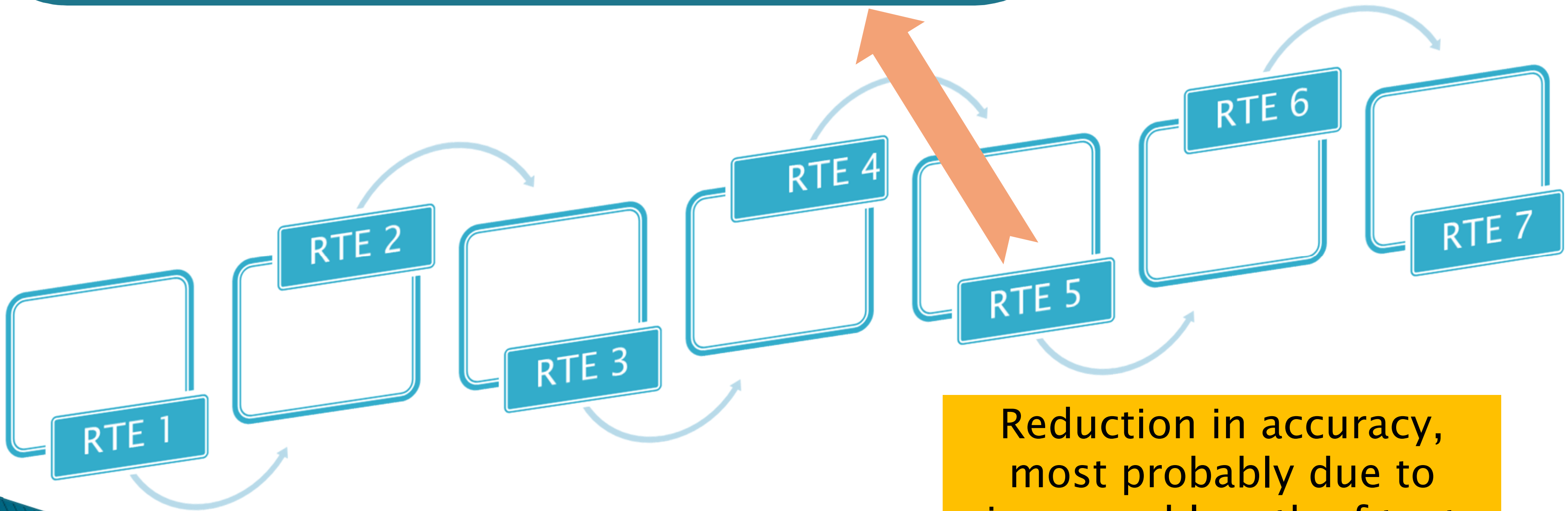
Best Accuracy :

Main Task:

2 way entailment: 68.3%

3 way entailment: 73.5%

Pilot Task: Precision=0.4098,Recall=0.5138, F-measure=0.4559



Reduction in accuracy, most probably due to increased length of text as compared to previous challenges

Main task: *Text Entailment in Corpus*

Main Subtask: *Novelty Detection*

Pilot Task: *Knowledge Base Population (KBP) Validation*

Best Accuracy :

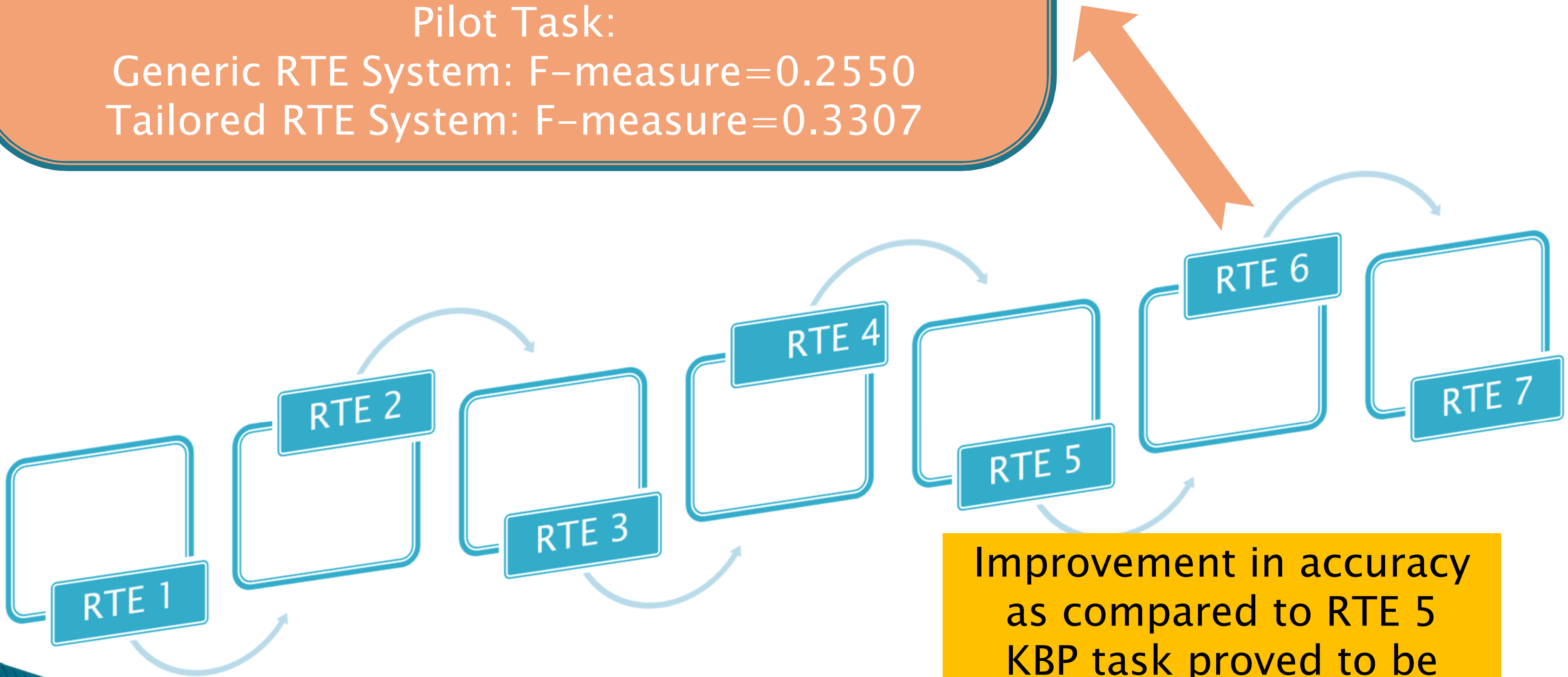
Main Task: F-measure=0.4801

Main Subtask: F-measure=0.8291

Pilot Task:

Generic RTE System: F-measure=0.2550

Tailored RTE System: F-measure=0.3307



Improvement in accuracy as compared to RTE 5 KBP task proved to be very challenging due to difference in Development and Test sets

Main task: *Text Entailment in Corpus*

Subtask: *Novelty Detection and Knowledge Base Population (KBP) Validation*

Best Accuracy :

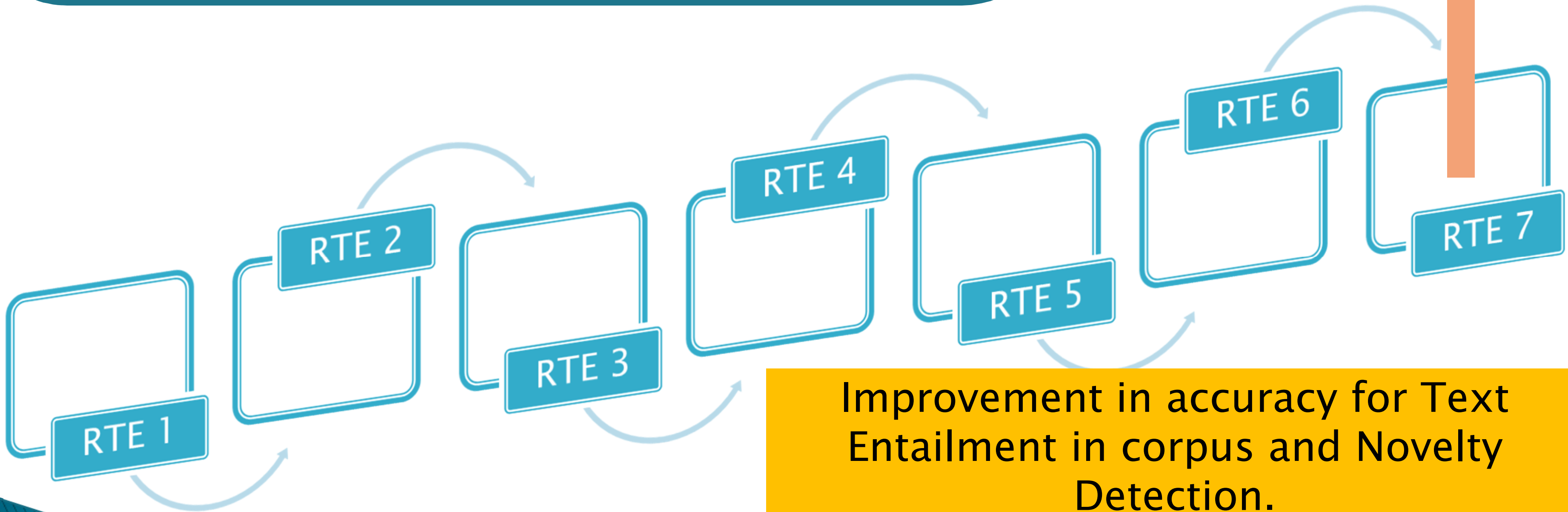
Main Task: F-measure=0.4800

Sub Task: Novelty Detection: F-measure=0.9095

KBP Validation:

Generic RTE System: F-measure=0.1902

Tailored RTE System: F-measure=0.1834



Improvement in accuracy for Text Entailment in corpus and Novelty Detection.

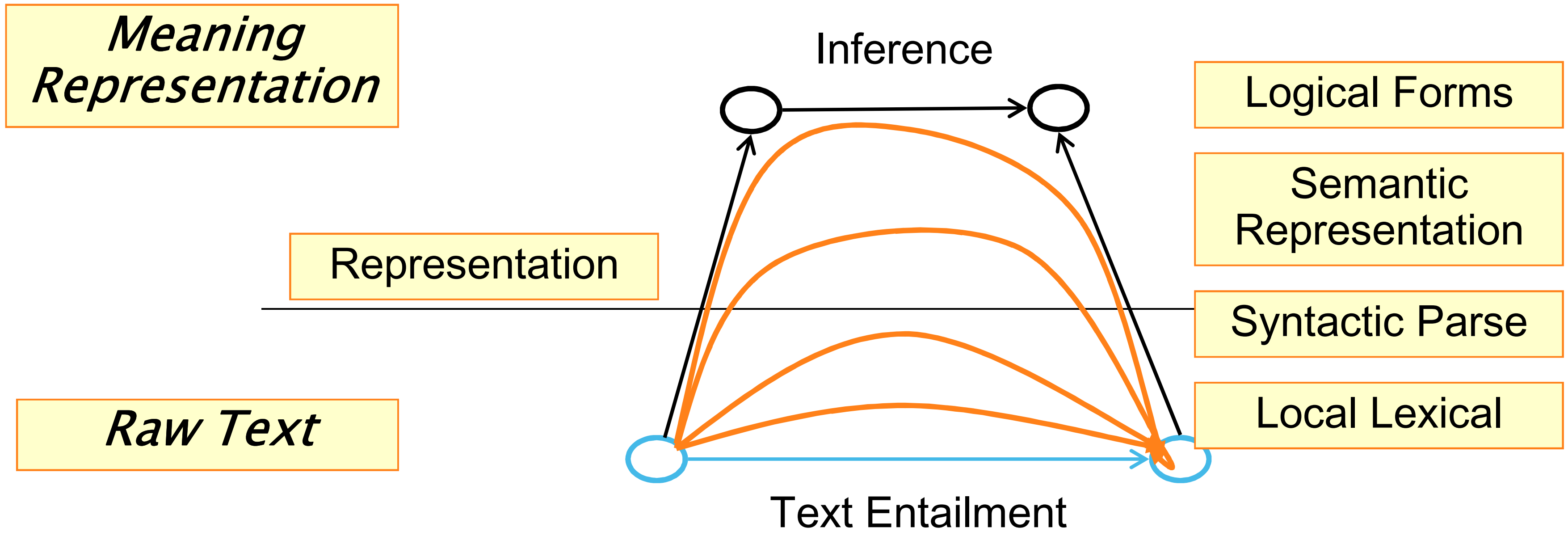
Reduction in performance of KBP task, shows most RTE systems are not robust enough to process large data.

To our surprise...

- RTE 6 Main Task:

Submission	Precision	Recall	F-measure
(Jia et al, 2011)	0.6857	0.3693	0.4801
(Majumdar and Bhattacharyya, 2011)	0.5343	0.4286	0.4756
(Tateishi and Ishikawa, 2011 IKOMA)	0.3971	0.5143	0.4481
(Kouylekov et al., 2011)	0.4346	0.4603	0.4471

Basic Representations



From: Textual Entailment, Ido Dagan, Dan Roth, Fabio Zanzotto, ACL 2007

RTE recognizing approaches

- Lexical only
- Tree similarity
- Predicate-argument structures
- Logical form - BLUE (Boeing)
- cross-pair similarity
- Learning alignment
- Alignment-based + Logic

Our Intuition

- Text :: Everybody loves somebody.
- Hypothesis :: Somebody loves somebody.

- Predicate :: Love(x,y) = x Loves y
- Text :: $\forall x \exists y$ Love (x,y)
- Hypothesis :: $\exists x \exists y$ Love (x,y)

- Here Text =>> Hypothesis...
- So we can say that hypothesis is entailed by Text

Stand alone is challange

- T :: I can lift an elephant with one hand

- H1 :: I can lift very heavy thing.
- H2 :: There exist an elephant with one hand.

- Needs support of parsing and tree structure for finding correct entailment.
- Knowledge is the key to solve text entailment.

- Support also needed from →
  - WSD (Word Sense Disambiguation)
  - NER (Name Entity Recognition)
  - SRL (Statistical Relationship Learning)
  - Parsing
  - Common background Knowledge

Predicate-argument structures

Intuition says that entailment pairs can be solved, in the majority of cases, by examining two types of information,

1) The relation of the verbs in the hypothesis to the ones in the text

2) Each argument or adjunct is an entity, with a set of defined properties

Things we need to know ::

- Levin’s classes
- VerbNet
- A Predication and argument Based Algorithm

Levin’s classes

- “The largest and most widely used classification of English verbs”
- over 3,000 English verbs according to shared meaning and behavior.
- Intuition: a verb's meaning influences its syntactic behavior
- shows how identifying verbs with similar syntactic behavior provides an effective means of distinguishing semantically coherent verb classes, and isolates these classes by examining verb behavior with respect to a wide range of syntactic alternations that reflect verb

VerbNet (VN)

- online verb lexicon for English that provides detailed syntactic and semantic descriptions for *Levin classes* organized into a refined taxonomy.
- hierarchical, domain-independent, broad-coverage verb lexicon.
- has mappings to a number of widely used verb resources, such as FrameNet and WordNet.
- Example of VerbNet *VerbNet Class "eat-39.1"* is given next...

VerbNet Class “eat-39.1

<FRAME><DESCRIPTION descriptionNumber="" primary="NP V NP ADJ"

- secondary="NP-ADJPResultative" xtag=""/><EXAMPLES><EXAMPLE>Cynthia ate herself sick.</EXAMPLE></EXAMPLES><SYNTAX><NP
- value="Agent"><SYNRESTRS/></NP><VERB/><NP
- value="Oblique"><SELRESTRS><SELRESTR Value="+"
- type="refl"/></SELRESTRS></NP><ADJ/></SYNTAX><SEMANTICS><PRED
- value="take_in"><ARGS><ARG type="Event"

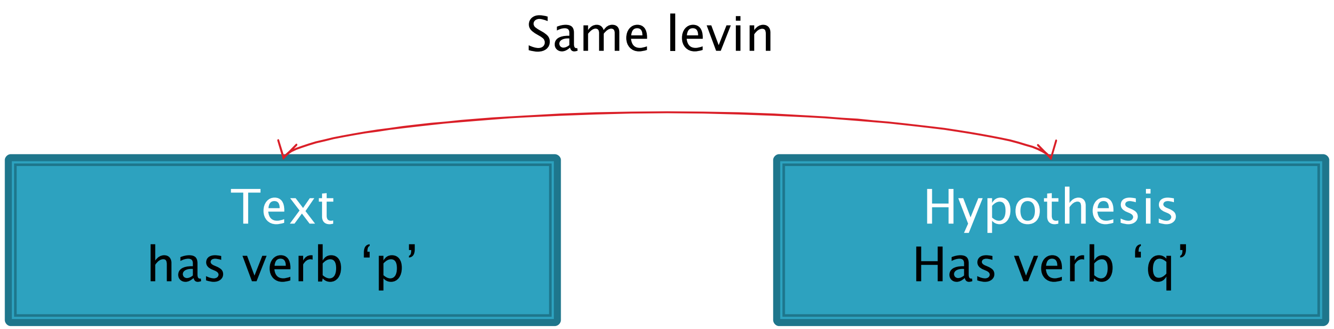
  value="during(E)"/><ARG
- type="ThemRole" value="Agent"/><ARG type="ThemRole"
- value="?Patient"/></ARGS></PRED><PRED value="Pred"><ARGS><ARG
- type="Event" value="result(E)"/><ARG type="ThemRole"
- value="Oblique"/></ARGS></PRED></SEMANTICS></FRAME>

Predicate Argument based Algorithm

- Step 1::

  Extract the Levin class for all the verbs in Text (T) and Hypothesis(H) and attach the appropriate semantic description, on the basis of the Levin class and syntactic analysis.

- Step 2::



- Example –

Step 2 A)

Arguments and adjuncts match, verbs not opposite

T: The cat ate a mouse

Entailment

H: Mouse is eaten by a cat

Step 2 B)

Verb match but arguments and adjuncts opposite

T: The cat ate a large mouse.

Contradiction

H: The cat ate a small mouse.

Step 2 C)

T: The cat ate a mouse.

Unrelated

H: The cat ate in the garden.

Arguments not related

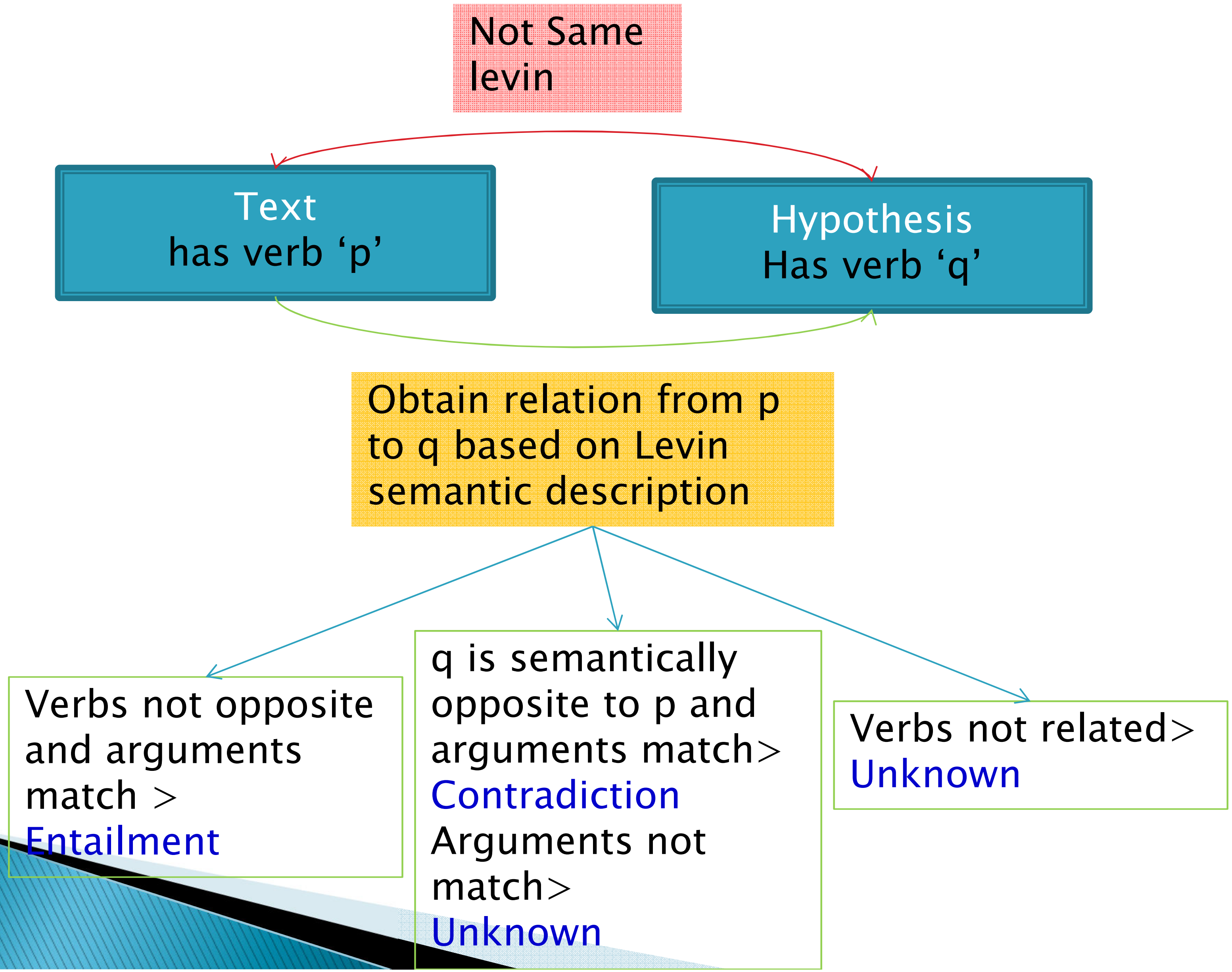
Step 2

A) For all candidates p in T, if the arguments and adjuncts match over p and q, and the verbs are not semantic opposites (e.g. antonyms or negations of one another), return ENTAILMENT

B) Else, (i) if the verbs match, but the arguments and adjuncts are semantic opposites (e.g. antonyms or negations of one another), or the arguments are related but do not match return CONTRADICTION (ii) else if the arguments are not related, return UNKNOWN

C) Else, return UNKNOWN

- Step 3:


Not Same levin
Text has verb ‘p’
Hypothesis Has verb ‘q’
Obtain relation from p to q based on Levin semantic description
Verbs not opposite and arguments match > Entailment
q is semantically opposite to p and arguments match> Contradiction
Arguments not match> Unknown
Verbs not related> Unknown

- Step 3 ::
- For every verb q in H, if there is no verb p in T has the same Levin as q, extract relations between q and p on the basis of Levin semantic descriptions
- A) If the verbs in H are not semantic opposites (e.g. antonyms or negations of one another)of verbs in T, and the arguments match, return ENTAILMENT
- B) Else, (i) if q is semantically opposite to p and the arguments match, or the arguments do not match, return CONTRADICTION (ii) else if the arguments are not related, return UNKNOWN
- C) Else, return UNKNOWN
- Step 4 ::

  Return UNKNOWN

Intuition of this algorithm is taken from structure of VerbNet which has

subset meanings like:

- *give and receive*
- *declared and proclaimed*
- *gain and benefit*
- or synonyms and antonyms and so on..

verb inference like:

- *hungry then eat*
- *thirsty then drink*
- *tired then rest*

Some of pairs solved successfully by this algorithm..

- Example 1: Exact match over VN classes

- *T: MADAGASCAR'S constitutional court declared Andry Rajoelina as the new president of the vast Indian Ocean island today, a day after his arch rival was swept from office by the army. ...*
- *H: Andry Rajoelina was proclaimed president of Madagascar.*

Requirement of backgroud knowledge

Match in terms of verb [Step 3]
(can be verified using VN and Levin classes)

- Example 2: Syntactic description and semantic decomposition

- T: A court in Venezuela has jailed nine former police officers for their role in the deaths of 19 people during demonstrations in 2002. ...
- H: Nine police officers have had a role in the death of 19 people.

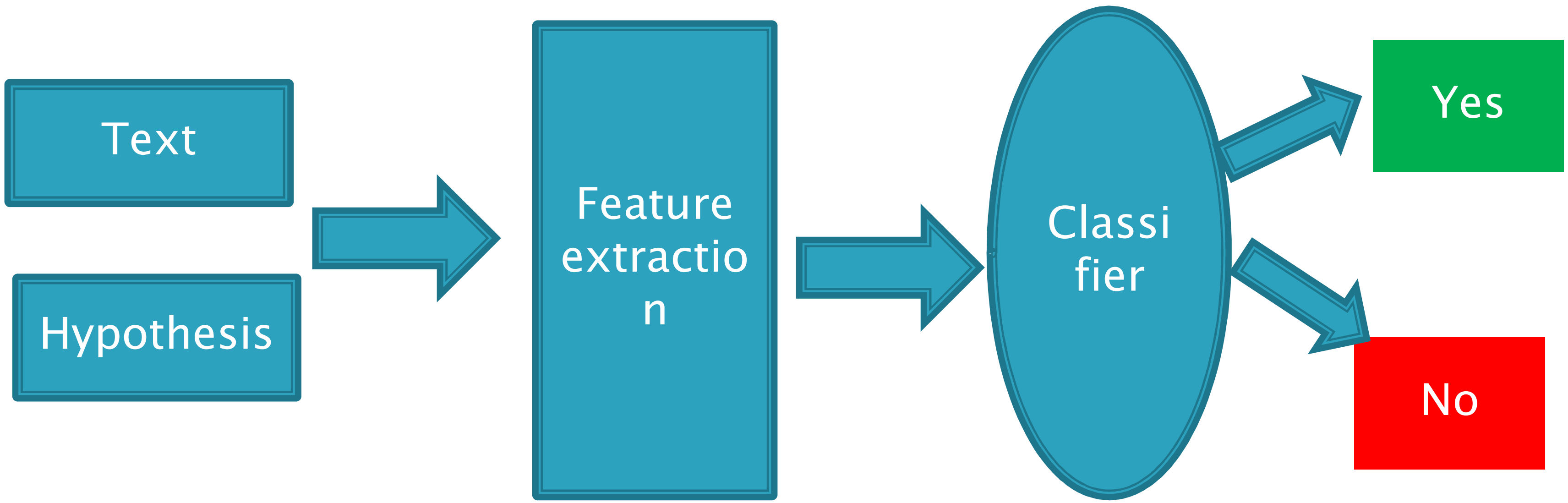
Predicate can be written as
P(theme 1, theme 2)

Result of Predicate Argument Structure

- The results have shown that such an approach solves 38% of the entailment pairs taken into consideration; also, a further 29.5% of the pairs are solved by the use of argument structure matching.
- Even verbs are not attached or one of the key concepts in H is not even existed in T, this method can solve them because of argument structure matching.

Machine Learning Approach

- RTE task can be thought as classification task.
  - Whether hypothesis entails a text or not



What we need

- We have off the shelf classifier tools available. We just need features as input to classifier .

- Possible Features :
  - Distance features
  - Entailment Triggers
  - Pair Feature

Distance Feature

- Numbers of words in common
- Length of Longest common subsequence
- Example
  - T:“All I eat is mangoes.”
  - H: “I eat mangoes.”

  - No. of common words = 3
  - length of lcs = 3

Entailment Triggers

Polarity features

Presence /absence of negative polarity contexts (not , no or few without)

"Dark knight rises" => "Dark knight doesn't fall"

Antonym features

Presence/absence of antonymous word in T and H

"Dark knight is falling" $\not\Rightarrow$ "Dark knight is rising"

Adjunct features

Dropping/adding of syntactic adjunct when moving from T to H

"He is running fast" => "He is running"

Pair feature

Bag of words

Using words in hypothesis and text we can create dictionary and represent it in form vector.

- T: Sachin is in Indian cricket team .
- H: Sachin plays cricket .
- Dictionary [Sachin:1, plays:2, Indian:3, cricket:4, team:5, is:6, in:7]
- Now text and hypothesis can be represented as vector.
- VT=[1,0,1,1,1,1,1]
- VH=[1,1,0,1,0,0,0]
- What we can learn here whether a word is in T than H entails or whether a word is in H or not than T entails H . Too naïve

Pair Feature continued..

Cross-pair similarity

$$K_s((T',H'),(T'',H'')) =$$

$$\max_{c \in C}\Big(K_T(t(H',c),t(H'',i)) + K_T(t(T',c),t(T'',i))\Big)$$

Where

- ***C*** is the set of all the correspondences between anchors of ***(T',H')*** and ***(T'',H'')***
- ***t(S, c)*** returns the parse tree of the text **S** where placeholders of these latter are replaced by means of the substitution ***c***
- ***i*** is the identity substitution
- $\mathbf{K_T(t_1, t_2)}$ is a function that measures the similarity between the two trees $\mathbf{t_1}$ and $\mathbf{t_2}$.(It gives number of subtrees shared by $\mathbf{t_1}$ and $\mathbf{t_2}$.)

Pair Feature continued..

- "All companies file annual reports" => "All Fortune companies file annual reports"

  T1 : (S (NP: 1 (DT All) (NNS: 1 companies)) (VP: 2 (VBP: 2 file) (NP: 3 (JJ: 3 annual) (NNS: 3 reports))))

  H1 : (S (NP: 1 (DT All) (NNP Fortune) (CD 50) (NNS: 1 companies)) (VP: 2 (VBP: 2 file) (NP: 3 (JJ: 3 annual) (NNS: 3 reports))))

- "In autumn all leaves fall" => "in autumn maple leaves fall"

  T2 : (S (PP (IN In) (NP (NN: a autumn))) (, ,) (NP: b (DT all) (NNS: b leaves)) (VP: c (VBP: c fall)))

  H2: (S (PP (IN In) (NP: a (NN: a autumn))) (, ,) (NP: b (DT all) (NN maple) NNS: a leaves)) (VP: c (VBP: c fall)))

- What we can learn

  T3: (S (NP: x (DT all) (NNS: x)) (VP: y (VBP: y)))

  H3: (S (NP: x (DT all) (NN) (NNS: x)) (VP: y (VBP: y)))

Character or number in red are **placeholders** .

Application

NLP applications like following use above phenomenon of variability of semantic expression, and hence phenomenon of *textual entailment*:

i. Question Answering (QA)
ii. Information Extraction (IE)
iii. Information Retrieval (IR)
iv. Comparable Documents (CD)
v. Multi-document Summarization (SUM)
vi. Machine Translation (MT)
vii. Paraphrase Acquisition (PP)

Question Answering (QA)

Zee News, 7th Nov'12

Text: "Barack Obama beats Romney to win re-election as US President"

Hypothesis 1: Barack Obama elected as president.

Entailment

Hypothesis 2: Romney elected as president. Contradiction

Hypothesis 3: Results of presidential election were declared on 14th October in US

Unknown

Information Extraction (IE)

Text: Fab.com, one of the fastest-growing online retail sites in the world, has acquired Pune-based technology venture True Sparrow Systems in a cash-and-stock deal that marks the first time a US-based e-commerce company has bought an Indian technology startup.

Hypothesis: Fab.com is Indian Technology startup

Not Entailment

Information Retrieval (IR)

Text: Fab.com, one of the fastest-growing online retail sites in the world, has acquired Pune-based technology venture True Sparrow Systems in a cash-and-stock deal that marks the first time a US-based e-commerce company has bought an Indian technology startup.

Hypothesis: Fab.com bought True Sparrow Systems

Entailment

Machine Translation (MT)

Hyp: The virus did not infect anybody.

entailment

entailment

Ref: No one was infected by the virus.

Hyp: Virus was infected.

no entailment

no entailment

Ref: No one was infected by the virus.

From: Sebastian Padó, Michel Galley, Dan Jurafsky, and Christopher D. Manning. 2009. Textual Entailment Features for Machine Translation Evaluation. *Proceedings of the Fourth Workshop on Statistical Machine Translation*, pp. 37-41.

Paraphrase Acquisition (PP)

Text: Any trip to Italy should include a visit to Tuscany to sample their exquisite wines.

Hypothesis: Be sure to include a Tuscan wine-tasting experience when visiting Italy.

Entailed

Conclusion

- NLP is all out understanding text and logically deduct that meaning of this sentence is understood by computer and checking that it is the same meaning as human understood or not..
- we need knowledge.. we need data.. but most important we need a framework which has thinking part of his own and has power to find inference using logic.. RTE can be used to develop such a framework..
- RTE can also be used as part of the most important application of NLP, which is summarization..

Future Work

- There is no way that anybody can say that my method of RTE will definitely give correct answer because computer do not have their own thinking and to make them thinking as human is dream of AI people from years
- But still with use of WordNet, Documentations and other resources like wikipedia and so on most of the entailment can be inferred with proper logical inference methods.
- Till now it has no mapping with the proper knowledge resource which is the key for RTE.
- Even to tell a computer that elephant is heavy is a difficult task if u do not have a proper resource and inferring technique.

References

1. The PASCAL Recognizing Textual Entailment Challenge Ido Dagan, Oren Glickman and Bernardo Magnini, 2006
2. The Second PASCAL Recognizing Textual Entailment Challenge Roy Bar-Haim, Ido Dagan, Bill Dolan, Lisa Ferro, Danilo Giampiccolo, Bernardo Magnini and Idan Szpektor,2006
3. The Third PASCAL Recognizing Textual Entailment Challenge Danilo Giampiccolo, Bernardo Magnini, Ido Dagan and Bill Dolan, 2007
4. The Fourth PASCAL Recognizing Textual Entailment Challenge Danilo Giampiccolo, Hoa Trang Dang, Bernardo Magnini, Ido Dagan, Elena Cabrio and Bbill Dolan, 2008
5. The Fifth PASCAL Recognizing Textual Entailment Challenge Luisa Bentivogli, Ido Dagan, Hoa Trang Dang, Danilo Giampiccolo and Bernardo Magnini, 2009
6. The Sixth PASCAL Recognizing Textual Entailment Challenge Ido Dagan, Oren Glickman and Bernardo Magnini, 2010
7. The Seventh PASCAL Recognizing Textual Entailment Challenge Luisa Bentivogli, Peter Clark, Ido Dagan, Hoa Trang Dang and Danilo Giampiccolo, 2011

References

7. Zanzotto, Fabio M., et al. "Learning textual entailment from examples." *Proceedings of the Second PASCAL Challenges Workshop on Recognising Textual Entailment, Venice, Italy*. 2006.
8. Sammons, Mark, V. G. Vydiswaran, and Dan Roth. "Ask not what textual entailment can do for you..." *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 2010.
9. Harabagiu, Sanda, and Andrew Hickl. "Methods for using textual entailment in open-domain question answering." *ANNUAL MEETING-ASSOCIATION FOR COMPUTATIONAL LINGUISTICS*. Vol. 44. No. 2. 2006.

References

- Celikyilmaz, Asli, Marcus Thint, and Zhiheng Huang. "A graph-based semi-supervised learning for question-answering." *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2*. Association for Computational Linguistics, 2009.
- Chambers, Nathanael, et al. "Learning alignments and leveraging natural logic." *Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing*. Association for Computational Linguistics, 2007.

12. Moruz, M. A. *Predication Driven Textual Entailment*. Diss. Ph. D. thesis,"Alexandru Ioan Cuza" University, Faculty of Computer Science, Iasi, 2011.
13. http://art.uniroma2.it/research/te/
14. Padó, Sebastian, et al. "Textual entailment features for machine translation evaluation." *Proceedings of the Fourth Workshop on Statistical Machine Translation*. Association for Computational Linguistics, 2009.
15. Zanzotto, Fabio Massimo, Marco Pennacchiotti, and Alessandro Moschitti. "Shallow semantics in fast textual entailment rule learners." *Proceedings of the ACL-PASCAL workshop on textual entailment and paraphrasing*. Association for Computational Linguistics, 2007.

Questions and Comments