Recognizing Textual Entailment for Italian

EDITS @ EVALITA 2009

Matteo Negri,

Elena Cabrio, Yashar Mehdad, Milen Kouylekov, and Bernardo Magnini

FBK-irst, Human Language Technology group

EDITS (Edit Distance Textual Entailment Suite)

- A general purpose RTE package.

- Main Features:
  - Distance-based approach (edit distance)
  - Allows for language dependent/independent configurations
  - Configurability - XML configuration file & shell options
  - Extendibility - interfaces for all modules (plugins)
  - Task adaptability (optimizable on different dimensions)
  - Reads and outputs the RTE entailment corpus format
  - **…Open Source Distribution (JAVA) – LGPL**

Edit Distance Approach

- Assumption: the *distance* between T and H is a characteristic that separates positive from negative pairs.

- Ingredients:
  - A ***function***, with range from 0 to K, that calculates an *entailment score S* for a T-H pair, based on the *edit distance* between T and H
    - If T and H are the same, then T entails H (S=0)
    - If T and H are completely different then, T does not entail H (S=K)
  - A ***threshold*** *Z*, 0<Z<K, that separates positive from negative examples

Edit Distance Approach

Goal: transform T into H through <u>*edit operations*</u> (*insertion/deletion/substitution* of characters, tokens, nodes in a syntactic tree)

T: *Yahoo took over search company Overture Services Inc last year*

H: *Yahoo bought Overture*

YES **NO**

Threshold (Z)=0.6

Edit Distance Approach

Goal: transform T into H through *edit operations* (*insertion/deletion/substitution* of characters, tokens, nodes in a syntactic tree)

T: *Yahoo took over search company Overture Services Inc ~~last~~ ~~year~~*

0.0 0.2 0.1 0.05 0.05

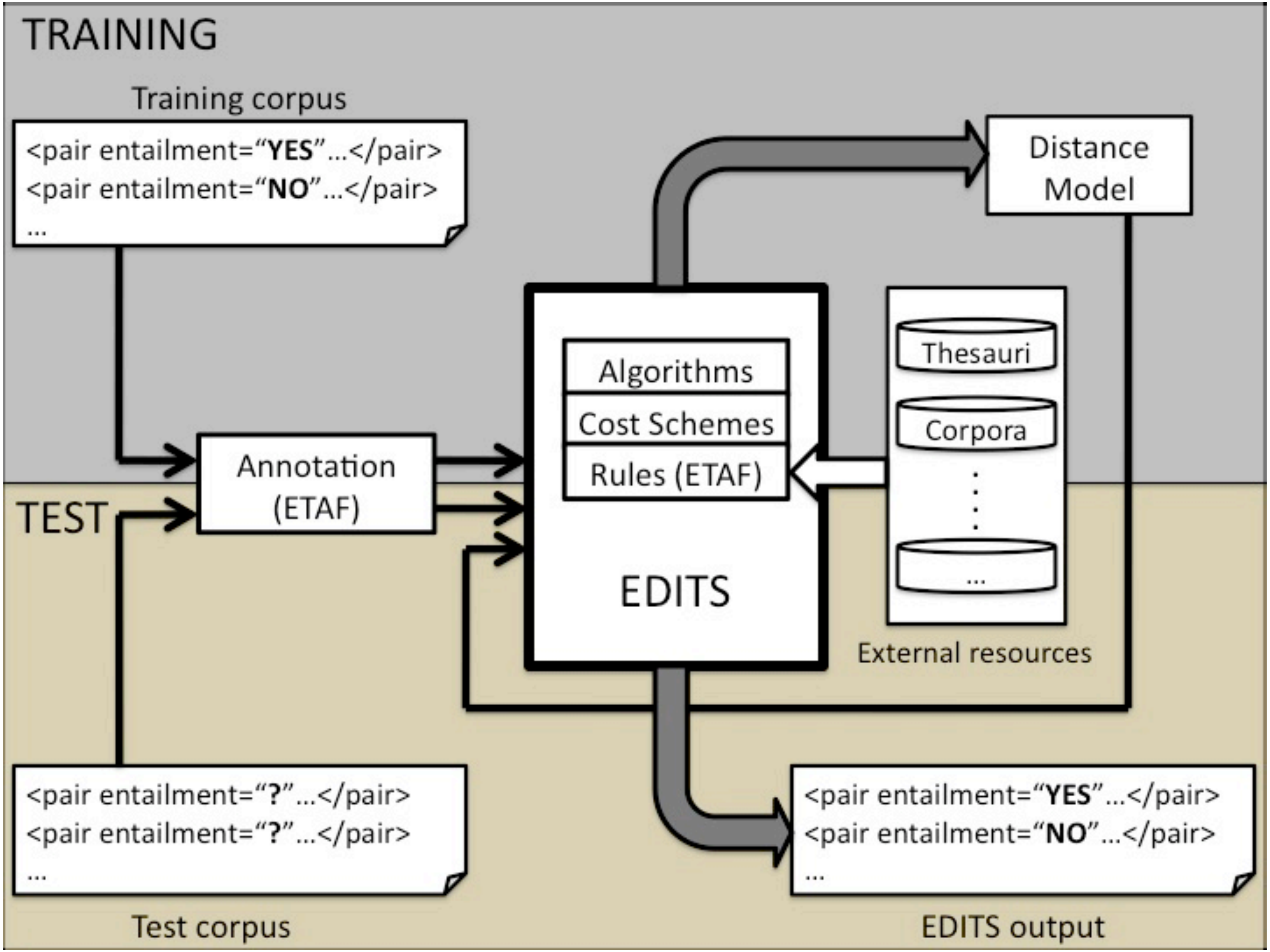
H: *Yahoo bought Overture*

S=0.4 ✔

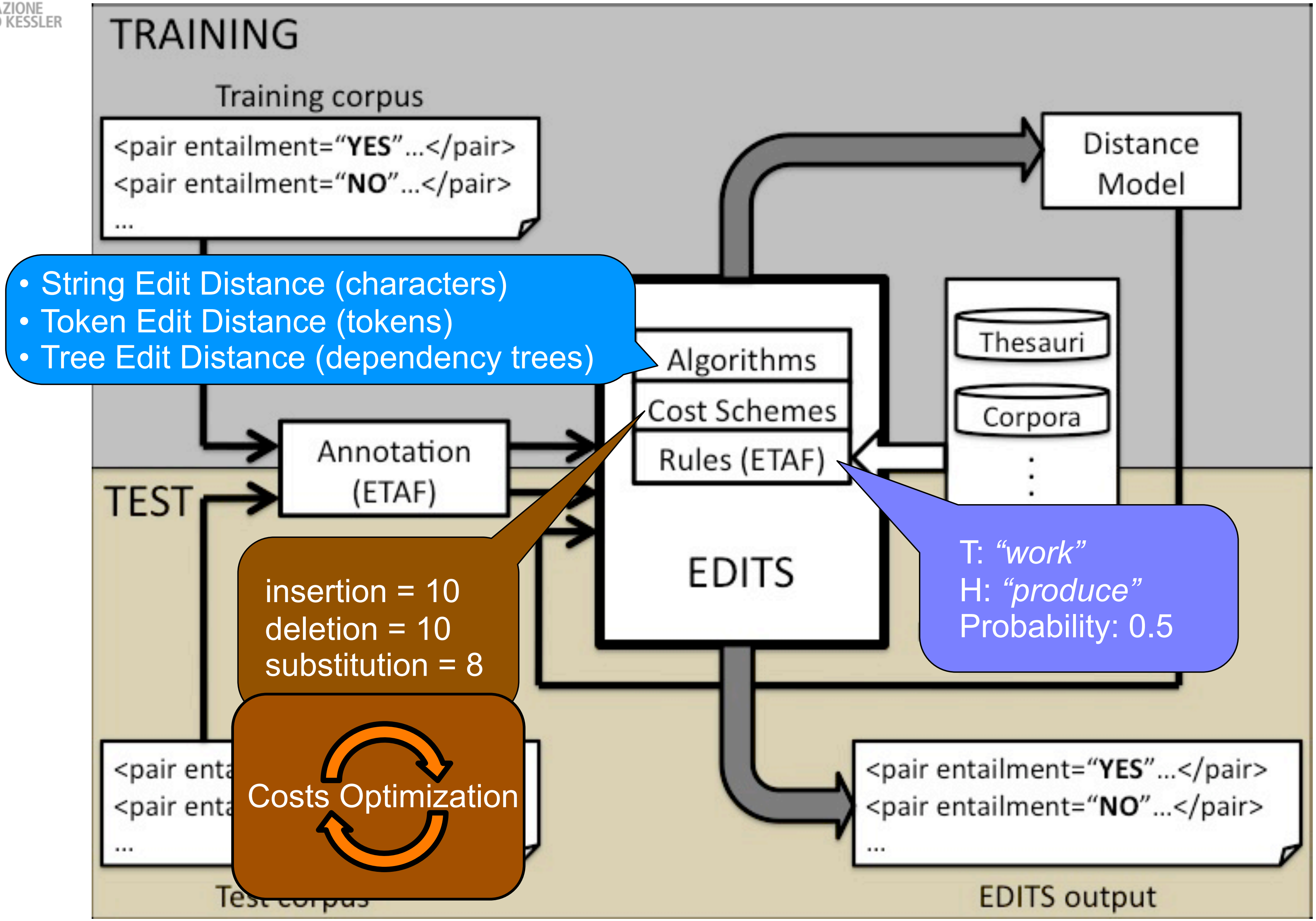
YES NO

Threshold (Z)=0.6

EDITS Architecture



EDITS Architecture



EVALITA submission…

TextPro
Pianta et al.,
LREC 2008

Y. Mehdad
ACL 2009

	Algorithm		Wikipedia Rules (133.500)	Stop Words (150)	Cost Optim.
	Tree-ED	Token-ED			
Run 1		✔		✔	
Run 2		✔	✔	✔	✔
Run 3	✔		✔	✔	✔
Run 4	✔		✔	✔	✔

MaltParser
Lavelli et al.,
EVALITA 2009

XIP
Bolioli et al.,
EVALITA 2009

1. LSA between all T-H terms
2. Relatedness threshold estimated on training data (jLSI tool by C. Giuliano)

EVALITA submission…

	Algorithm		Wikipedia Rules (133.500)	Stop Words (150)	Cost Optim.
	Tree-ED	Token-ED			
Run 1		✔		✔	
Run 2		✔	✔	✔	✔
Run 3	✔		✔	✔	✔
Run 4	✔		✔	✔	✔

…and results

	Run 1	Run 2	Run 3	Run 4
Dev	0.72	0.725	0.645	0.647
Test	0.71	0.71	0.51	0.50

EVALITA submission…

	Algorithm		Wikipedia Rules (133.500)	Stop Words (150)	Cost Optim.
	Tree-ED	Token-ED			
Run 1		✔		✔	
Run 2		✔	✔	✔	✔
Run 3	✔		✔	✔	✔
Run 4	✔		✔	✔	✔

Much higher results (>10%) wrt to English
same configuration used for TAC-RTE 2009

. … d results

	Run 1	Run 2	Run 3	Run 4
Dev	0.72	0.725	0.645	0.647
Test	**0.71**	**0.71**	0.51	0.50

FONDAZIONE BRUNO KESSLER

EVALITA submission…

	Algorithm		Wikipedia Rules (133.500)	Stop Words (150)	Cost Optim.
	Tree-ED	Token-ED			
Run 1		✔		✔	
Run 2		✔	✔	✔	✔
Run 3	✔		✔	✔	✔
Run 4	✔		✔	✔	✔

Much higher results (>10%) wrt to English
same configuration used for TAC-RTE 2009

Better results with the shallow approach.
High lexical similarity between T and H…
=> Easier to handle at the level of tokens?

sults

	Run 1	Run 2	Run 3	Run 4
Dev	0.72	0.725	0.645	0.647
Test	**0.71**	**0.71**	0.51	0.50

EVALITA submission…

	Algorithm		Wikipedia Rules (133.500)	Stop Words (150)	Cost Optim.
	Tree-ED	Token-ED			
Run 1		✔		✔	
Run 2		✔	✔	✔	✔
Run 3	✔		✔	✔	✔
Run 4	✔		✔	✔	✔

Much higher results (>10%) wrt to English same configuration used for TAC-RTE 2009

Better results with the shallow approach.
High lexical similarity between T and H…
=> Easier to handle at the level of tokens?

sults

	Run 1	**Run 2**	**Run 3**	**Run 4**
Dev	0.72	0.725	0.645	0.647
Test	0.71	**0.71**	0.51	0.50

On the test set, no noticeable contribution from Wikipedia and Automatic Cost Optimization

EVALITA submission…

	Algorithm		Wikipedia Rules (133.500)	Stop Words (150)	Cost Optim.
	Tree-ED	Token-ED			
Run 1		✔		✔	
Run 2		✔	✔	✔	✔
Run 3	✔		✔	✔	✔
Run 4	✔		✔	✔	✔

Much higher results (>10%) wrt to English same configuration used for TAC-RTE 2009

Better results with the shallow approach.
High lexical similarity between T and H…
=> Easier to handle at the level of tokens?

sults

	Run 1	Run 2	Run 3	Run 4
Dev	0.72	0.725	0.645	0.647
Test	0.71	0.71	**0.51**	**0.50**

On the test set, no noticeable contribution from Wikipedia and Automatic Cost Optimization

Small performance variations with different parsers

EVALITA submission…

	Algorithm		Wikipedia Rules (133.500)	Stop Words (150)	Cost Optim.
	Tree-ED	Token-ED			
Run 1		✔		✔	
Run 2		✔	✔	✔	✔
Run 3	✔		✔	✔	✔
Run 4	✔		✔	✔	✔

Much higher results (>10%) wrt to English same configuration used for TAC-RTE 2009

Drop from Dev to Test with Tree Edit Distance.
Poor recall on "NO" pairs (10%-15%)!
Higher variability (=higher distance) on training
=> Over-estimated threshold?

Better results with the shallow approach.
High lexical similarity between T and H…
=> Easier to handle at the level of tokens?

sults

	Run 1	Run 2	Run 3	Run 4
Dev	0.72	0.725	**0.645**	**0.647**
Test	0.71	0.71	**0.51**	**0.50**

On the test set, no noticeable contribution from Wikipedia and Automatic Cost Optimization

Small performance variations with different parsers

http://edits.fbk.eu