



EDITS: An Open Source Framework for Recognizing Textual Entailment

<http://edits.fbk.eu>

Milen Kouylekov

with:

Yashar Mehdad, Matteo Negri, Elena Cabrio, and Bernardo Magnini

Fondazione Bruno Kessler - Human Language Technology Unit

What is EDITS

- * A general purpose Recognizing Textual Entailment (RTE) framework.
- * Main Features:
 - * Open Source Distribution (JAVA) - LGPL
 - * Language Independent
 - * Configurability - configuration file & shell options
 - * Extendability - interfaces for all modules (plugins)
 - * Reads and Outputs the RTE entailment corpus format
 - * Task adaptability (optimize on different dimensions)

Edit Distance Approach

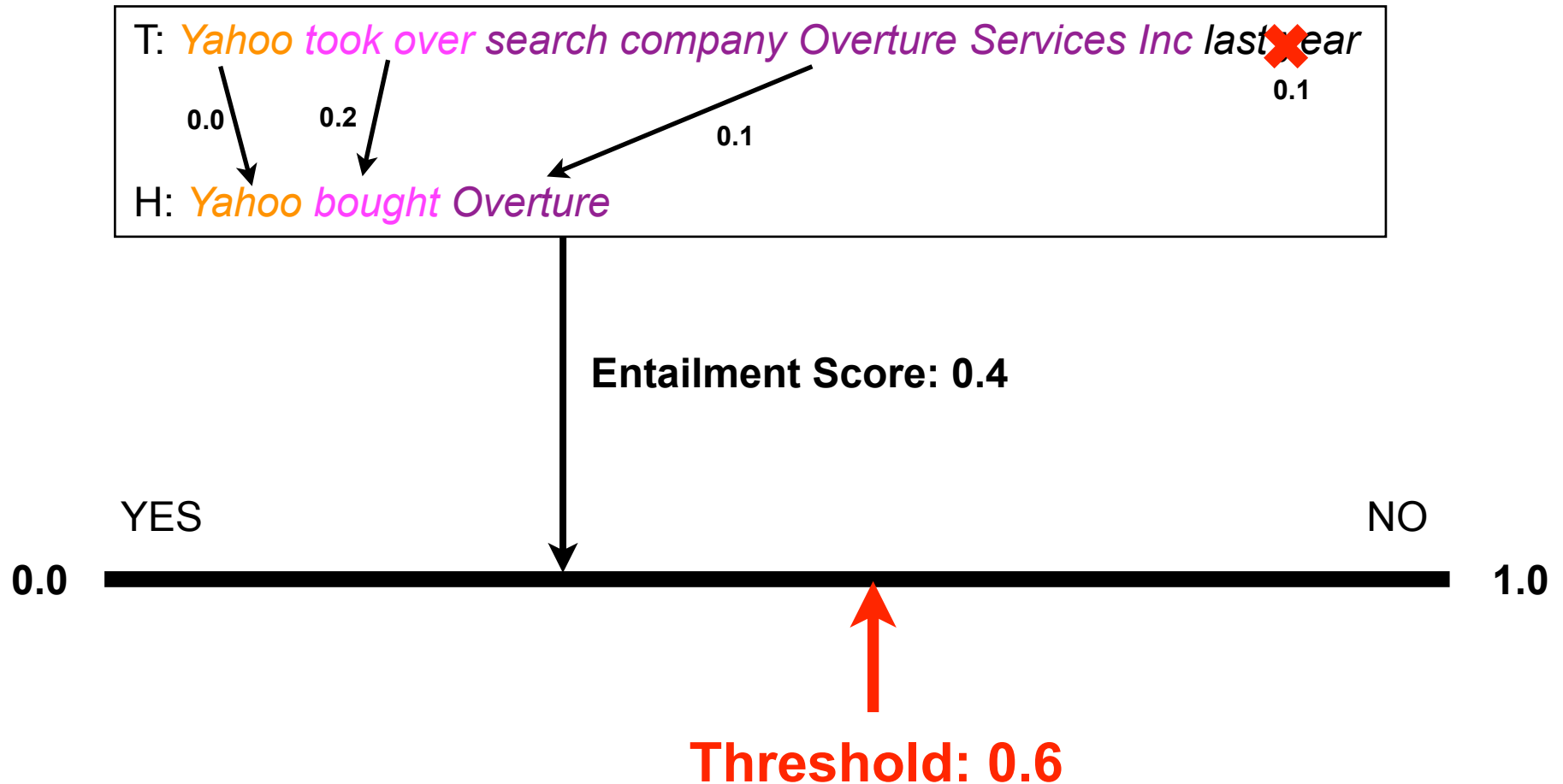
T: *Yahoo took over search company Overture Services Inc last year*

H: *Yahoo bought Overture*

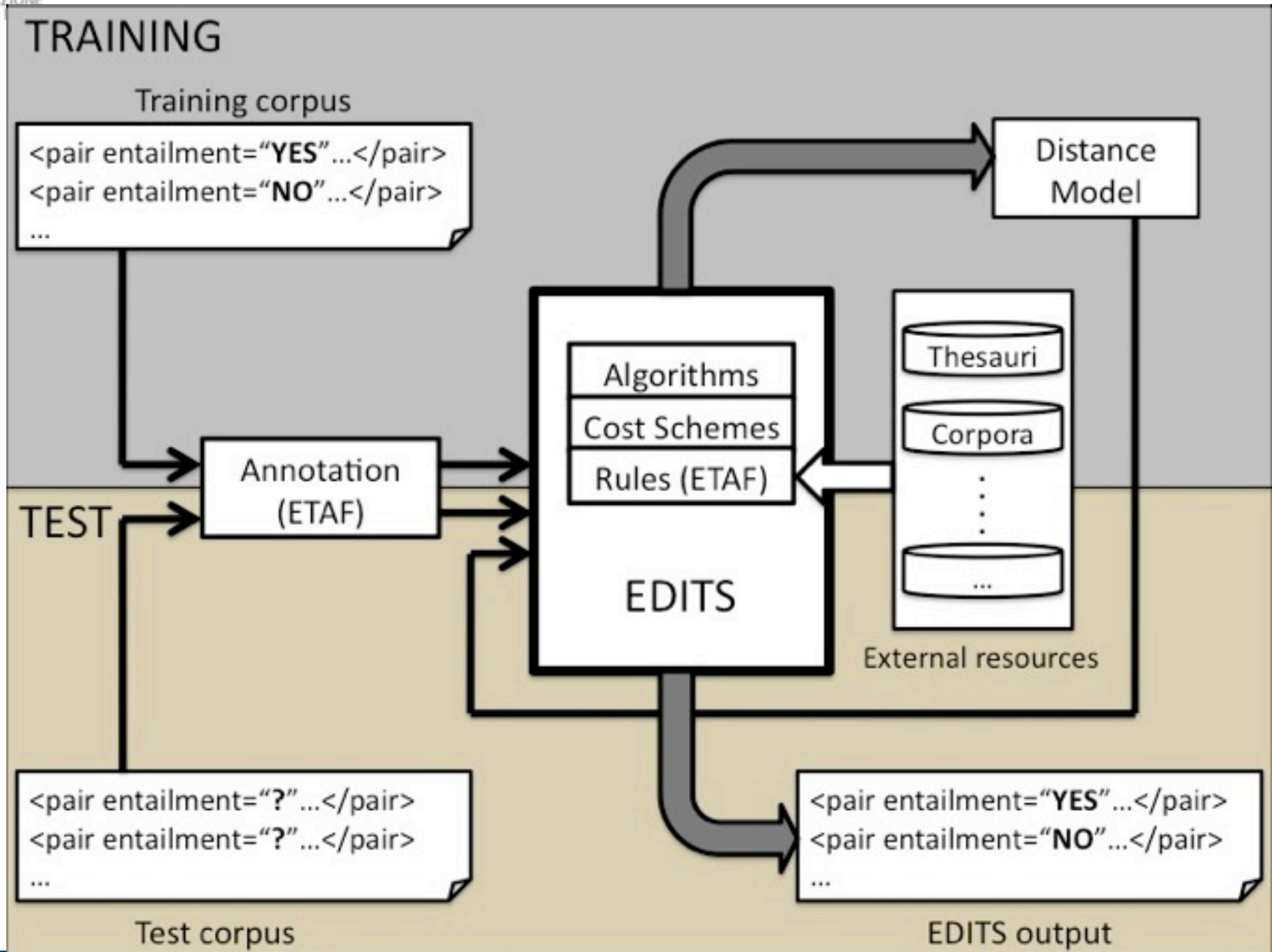


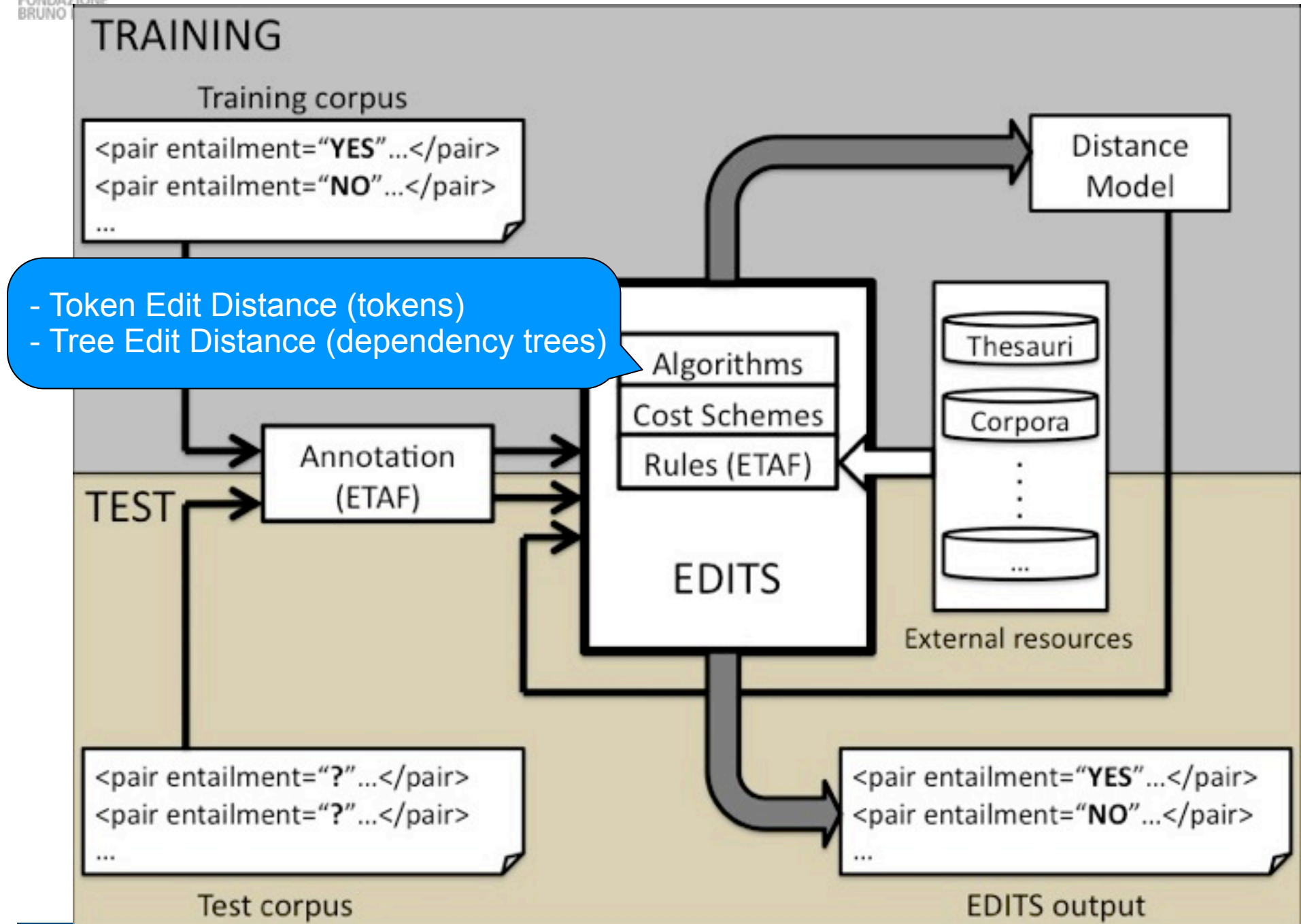
Threshold: 0.6

Edit Distance Approach

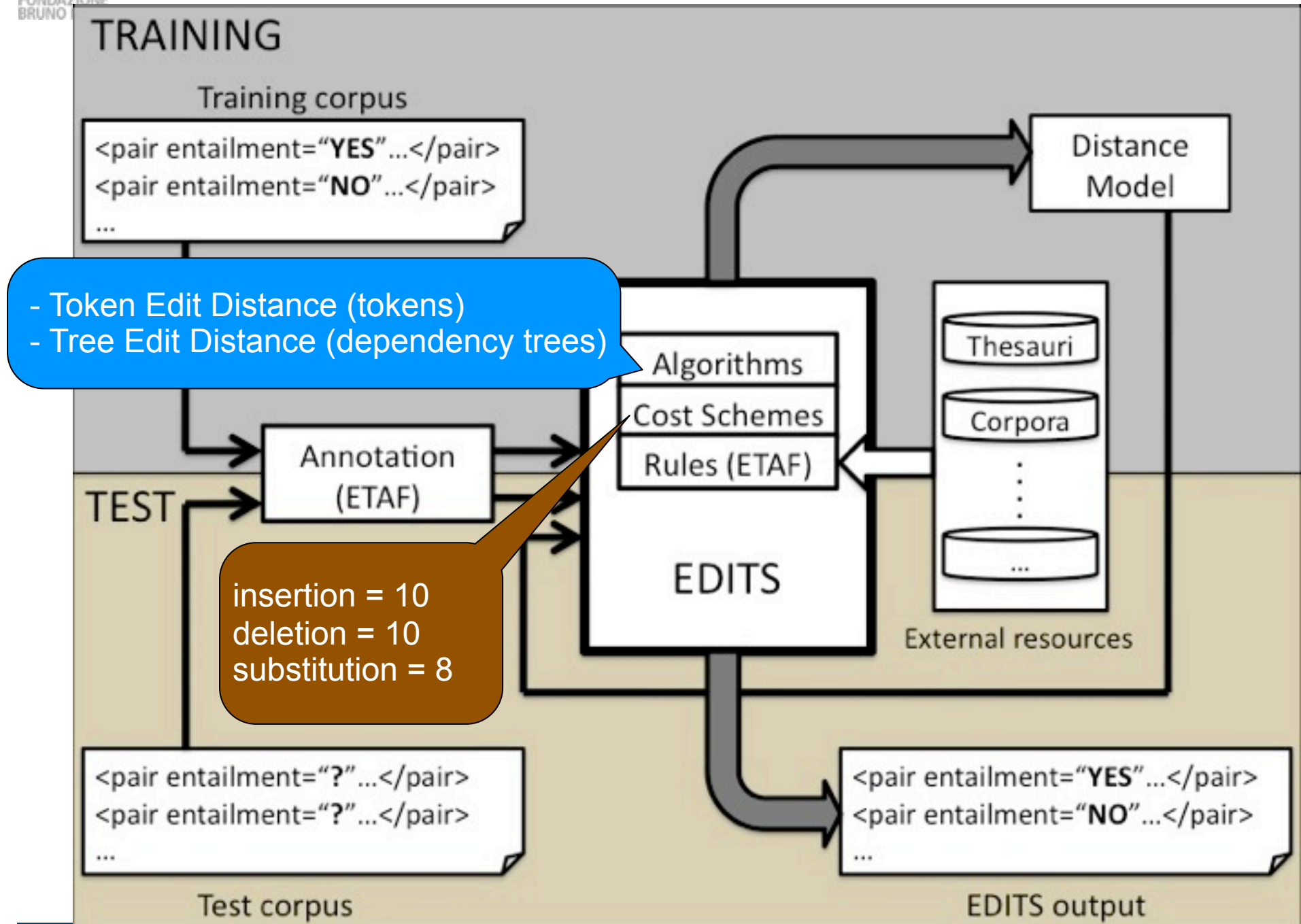


EDITS Architecture

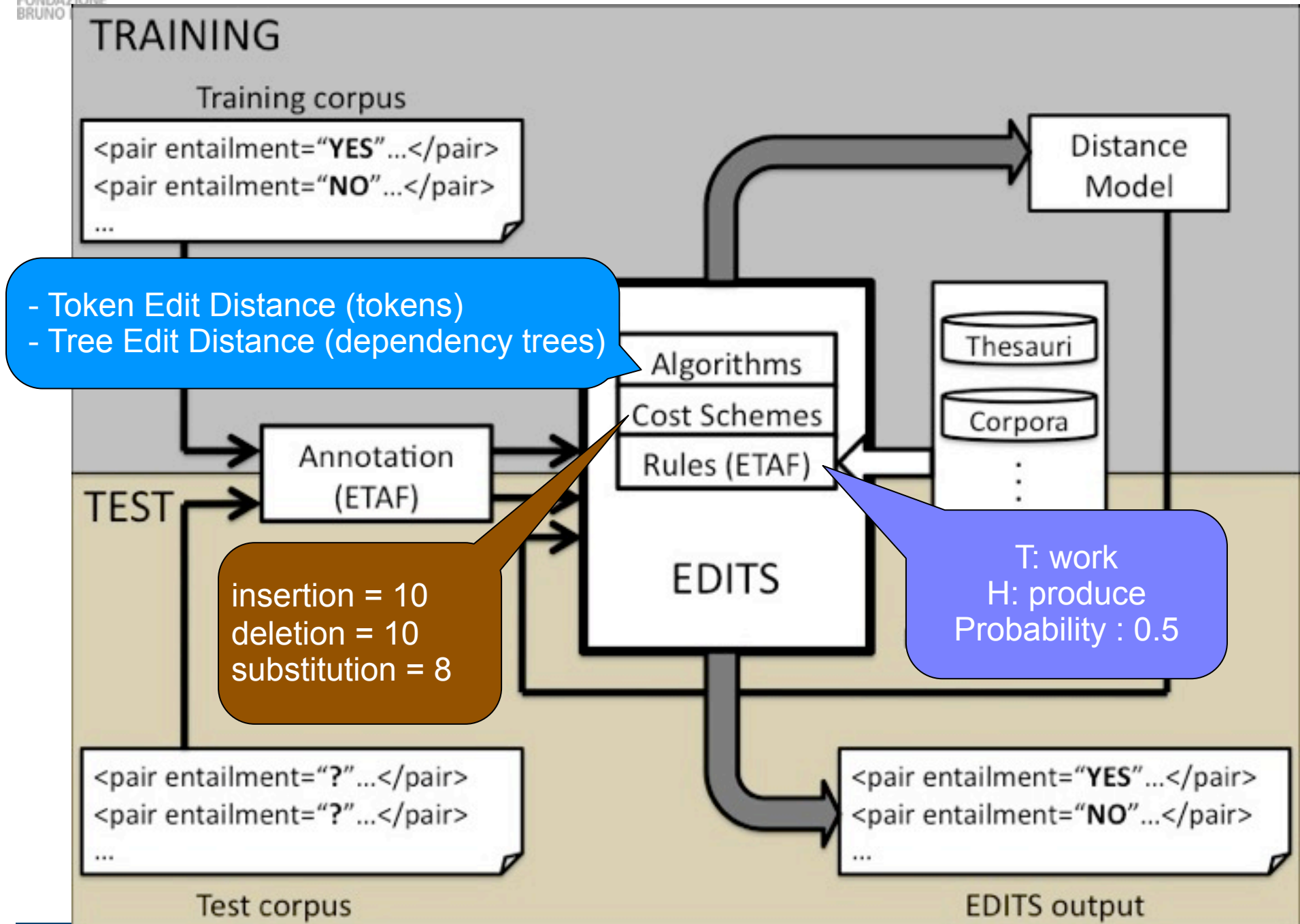




EDITS Architecture



EDITS Architecture



EDITS Text Annotation Format (ETAF)

Common representation format for RTE data and entailment/contradiction rules.

```
<entailment-corpus>
  <pair id="1" task="QA" entailment="YES" benchmark="Entailment" score="0.4" confidence="0.9">
    <t>Yahoo took over search company Overture Services Inc last year</t>
    <h>Yahoo bought Overture</h>
    <tAnnotation></tAnnotation>*
    <hAnnotation></hAnnotation>*
    <log distance="">
      <operation> </operation>
    </log>
  </pair>
</entailment-corpus>
```

EDITS Text Annotation Format (ETAF)

Common representation format for RTE data and entailment/contradiction rules.

```

<entailment-corpus>
  <pair id="1" task="QA" entailment="YES" benchmark="Entailment" score="0.4" confidence="0.9">
    <t>Yahoo took over search company Overture Services Inc last year</t>
    <h>Yahoo bought Overture</h>
    <tAnnotation></tAnnotation>*
    <hAnnotation></hAnnotation>*
    <log distance="">
      <operation> </operation>
    </log>
  </pair>
</entailment-corpus>
  
```

Annotation Levels (Plugins)

- * **Morpho-Syntax** – Text & Hypothesis as sequences of words
 - * TextPro, OpenNLP and TreeTagger
- * **Syntax** – Text & Hypothesis as syntactic trees (graph)
 - * Stanford Parser, OpenNLP, C&C Tools , XIP

EDITS - Configuration

Definition of the basic system components

```
<module alias="distance" type="entailment-engine">
  <module alias="tree-edit-distance" type="distance-algorithm"/>
  <module alias="xml-cost-scheme" type="cost-scheme">
    <option name="scheme-file">scheme.xml</option>
  </module>
  <module alias="file-repository" type="rules-repository">
    <option name="entailment-rules">verbocean.xml</option>
    <option name="entailment-rules">wiki.xml</option>
  </module>
  <module alias="pso" type="scheme-optimizer"/>
</module>
```

Cost Scheme Example

A simple cost scheme ...

Insertion(B) = idf(B)

Deletion(A) = #words T

Substitution(A,B) = $\begin{cases} 0 & A \Rightarrow B \\ 100 & A \neq B \end{cases}$

Cost Scheme Example

... XML Representation

A simple cost scheme ...

Insertion(B) = idf(B)

Deletion(A) = #words T

$$\text{Substitution}(A,B) = \begin{cases} 0 & A \Rightarrow B \\ 100 & A \neq B \end{cases}$$

```

<insertion>
  <cost> (idf B) </cost>
</insertion>
<deletion>
  <cost> (length T) </cost>
</deletion>
<substitution>
  <condition> (entail A B) </condition>
  <cost>0</cost>
</substitution>
<substitution>
  <condition> (not (equals A B)) </condition>
  <cost>10</cost>
</substitution>
  
```

Cost Scheme Optimization

- * Goal: Estimate the optimal cost of each edit operation
- * How: Stochastic method based on Particle Swarm Optimization (**PSO**)

```
<constant value="1" name="OPInsertion"/>
<constant value="1" name="OPDeletion"/>
<constant value="1" name="OPEntailment"/>
<insertion>
  <cost> OPInsertion </cost>
</insertion>
<deletion>
  <cost> OPDeletion </cost>
</deletion>
<substitution>
  <condition> (entail A B) </condition>
  <cost> OPEntailment </cost>
</substitution>
```

Before

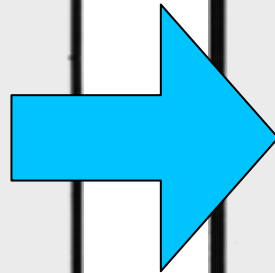
Cost Scheme Optimization

- * Goal: Estimate the optimal cost of each edit operation
- * How: Stochastic method based on Particle Swarm Optimization (**PSO**)

```

<constant value="1" name="OPInsertion"/>
<constant value="1" name="OPDeletion"/>
<constant value="1" name="OPEntailment"/>
<insertion>
  <cost> OPInsertion </cost>
</insertion>
<deletion>
  <cost> OPDeletion </cost>
</deletion>
<substitution>
  <condition> (entail A B) </condition>
  <cost> OPEntailment </cost>
</substitution>
  
```

Before



```

<constant value="5.2" name="OPInsertion"/>
<constant value="17" name="OPDeletion"/>
<constant value="4" name="OPEntailment"/>
<insertion>
  <cost> OPInsertion </cost>
</insertion>
<deletion>
  <cost> OPDeletion </cost>
</deletion>
<substitution>
  <condition> (entail A B) </condition>
  <cost> OPEntailment </cost>
</substitution>
  
```

After

Entailment Rules in RTE5

* Verbocean

- * 18232 entailment rules for all the English verbs
- * Connected by the "stronger-than" relation
Example: "kill [stronger-than] injure"

* Wikipedia

- * Latent Semantic Analysis over Wikipedia for the words inside RTE5 dataset (**jLSI tool**)
- * Relatedness Threshold of 0.7

Main Task Results

| | Main | -Verbocean | -Wikipedia | -PSO |
|----------|------|------------|------------|------|
| Accuracy | 60.2 | 60.3 | 59.16 | 57.3 |

```

<module alias="distance" type="entailment-engine">
  <module alias="tree-edit-distance" type="distance-algorithm"/>
  <module alias="xml-cost-scheme" type="cost-scheme">
    <option name="scheme-file">scheme.xml</option>
  </module>
  <module alias="file-repository" type="rules-repository">
    <option name="entailment-rules">verbocean.xml</option>
    <option name="entailment-rules">wiki.xml</option>
  </module>
  <module alias="pso" type="scheme-optimizer"/>
</module>

```

Pilot Results

- * No system adaptation
- * Paired the Hs with the sentences of the documents in the same topic into entailment pairs.
- * Estimated a threshold optimizing the FMeasure of the YES pairs.

| Micro Average | | | Macro Average (Topic) | | | Macro Average (Hypothesis) | | |
|---------------|-------|--------------|--------------------------|-------|-------|-------------------------------|-------|-------|
| P | R | F | P | R | F | P | R | F |
| 0.225 | 0.648 | 0.334 | 0.274 | 0.635 | 0.383 | 0.333 | 0.680 | 0.447 |

Future Directions

* EDITS

- * Add different entailment algorithms (plugins)
- * Improve Web Interface (<http://qallmedemo.fbk.eu/editsui/>)
- * Add desktop experimental environment (GUI)

* FBK-HLT Group

- * Specialized Engines (Ongoing PhD)
 - * Negation, Coreference etc.
- * Move to 3-way task (separate first the UNKNOWN)
 - * From distance to overlap measures

EDITS - <http://edits.fbk.eu>