

Proyecto: Reconocimiento de la Implicación Textual.

Descripción: Un sistema de reconocimiento de la implicación textual (RTE) recibe una lista de pares texto/hipótesis como el siguiente:

```
<pair id="15765" value="YES" task="QA">
  <q lang="ES" src="clef2006" type="ORGANIZATION">¿Qué es
  Lufthansa?</q>
  <t doc="efel9951102-00968">
  A la edad de 66 años falleció en Seeheim, centro de Alemania,
  Reinhardt Abraham, una de las figuras alemanas más destacadas de
  la técnica aeronáutica moderna, informó hoy, jueves, un portavoz
  de las Líneas Aéreas Alemanas (Lufthansa).
  </t>
  <h>Lufthansa es Líneas Aéreas Alemanas.</h>
</pair>
```

Donde:

- Cada par texto/hipótesis aparece dentro de un elemento <pair>
- El elemento <pair> tiene los siguientes atributos:
  - id, un identificador numérico del par texto/hipótesis
  - value, el valor de la implicación, siendo ‘YES’ (si existe la implicación) o ‘NO’ (si no existe)
  - task, el acrónimo de la aplicación del cual fue generado el par. En estos pares siempre es QA (búsqueda de respuestas).
- El elemento <q> (pregunta) contiene la pregunta de donde se generó el par. Este tiene los siguientes atributos:
  - lang, el lenguaje de la pregunta
  - src, la fuente de donde se obtuvo la pregunta
  - type, el tipo de pregunta (PERSON, LOCATION, etc.)
- El elemento <t> (texto) está compuesto de una o más oraciones que soportan la respuesta usada en la hipótesis. Sus atributos son los siguientes:
  - doc, el identificador de un documento que contiene el texto de soporte
- El elemento <h> (hipótesis) no tiene atributos, usualmente contiene sólo una oración

Entonces la salida de un sistema RTE debe ser:

pair\_id [YES|NO] confianza

donde la confianza varía entre 0 (no seguro) – 1 (completamente seguro).

En la evaluación sólo se toman en cuenta las respuestas ‘YES’ del sistema. Las medidas de evaluación son las siguientes:

$$precision = \frac{\#predicted\ as\ YES\ correctly}{\#predicted\ as\ YES} \quad (1)$$

$$recall = \frac{\#predicted\ as\ YES\ correctly}{\#YES\ pairs} \quad (2)$$

**Objetivo:** El objetivo del proyecto es que el alumno proponga y evalúe formas de representación de conocimiento, en este caso textual, para resolver un problema actual como es el reconocimiento de la implicación textual.

Con su estudio el estudiante deberá responder a las siguientes interrogantes (*Reporte*):

- ¿Cuál representación del conocimiento es la adecuada para reconocer la implicación? (*Según el tipo de pares que se le asignaron*)
- Debido a que en muchos pares el conocimiento explícito en la hipótesis no es suficiente:
  - ¿Cómo se puede adquirir conocimiento extra para reconocer la implicación?
  - ¿Cómo se puede incorporar ese conocimiento al razonamiento?
  - ¿Cómo tratar la falta de conocimiento?

Para facilitar el estudio, se proporciona un análisis de dependencias (DA) y una forma lógica (LF) de cada uno de los textos que intervienen en la implicación. Por ejemplo, para la hipótesis del ejemplo anterior:

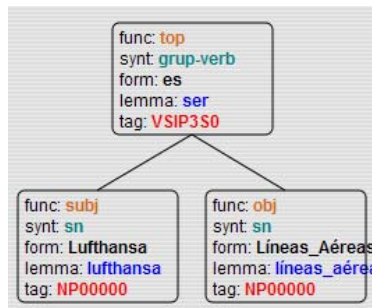
H:

Lufthansa es Líneas Aéreas Alemanas.

DA:

```
grup-verb/top/(es ser VSIP3S0 01775973) [  
  sn/subj/(Lufthansa lufthansa NP00000)  
  sn/obj/(Líneas_Aéreas_Alemanas líneas_aéreas_alemanas NP00V00)  
]
```

El análisis de dependencias refleja la relación sujeto-acción-objeto que hay en las oraciones.



LF:

```
np_lineas_aereas_alemanas(x2) & np_lufthansa(x1) & vs_ser(_, x1, x2).
```

**Nota:** DA y LF pueden contener errores debido a que se generan de forma automática con herramientas que no son robustas. Parte del estudio es evaluar la confianza en tales recursos.

**Referencias:**

Answer Validation Exercise (AVE)

<http://nlp.uned.es/QA/AVE/>

*Las memorias de este congreso se las anexo en el material.*