

Recognizing Textual Entailment



Estudiante: Ing.Lic. Julio Castillo

Directores :

Ph.D Francisco Tamarit

Ph.D María del Carmen Rojas



Contenido de la charla

1. Introducción
2. Motivación
2. Definición de la tarea de RTE
3. Ejemplos
4. RTE Roadmap
5. Líneas de Investigación



Introducción

- Presentación
- Inicio de este trabajo: Febrero 2008
- Tema de Investigación:
Reconocimiento de Implicación textual
utilizando técnicas de aprendizaje
automático.

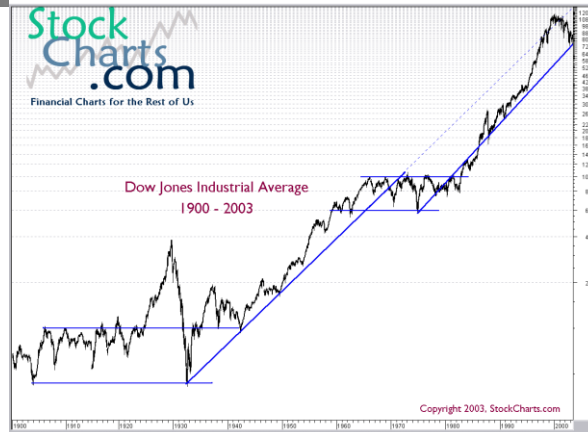


El Lenguaje y el Significado

1. *Relación uno a muchos:* Un significado puede mapearse a N-expresiones del lenguaje.
2. *Relación muchos a uno :* Una expresión del lenguaje se puede mapear a N-significados diferentes, aquí tenemos la “ambigüedad”.

Lenguaje y Significado (2)

Significado =



Fraser 1 = The Dow Jones Industrial Average closed up 255

Fraser 2 = Dow Jones stock market index closed with a rise of 255

Fraser 3 = Stock market hits a record high

Fraser 4 = Dow gains 255 points

Fraser 5 = Dow ends up

Fraser 6 = Dow climbs 255



Tipos de relaciones

Relación entre expresiones de texto:

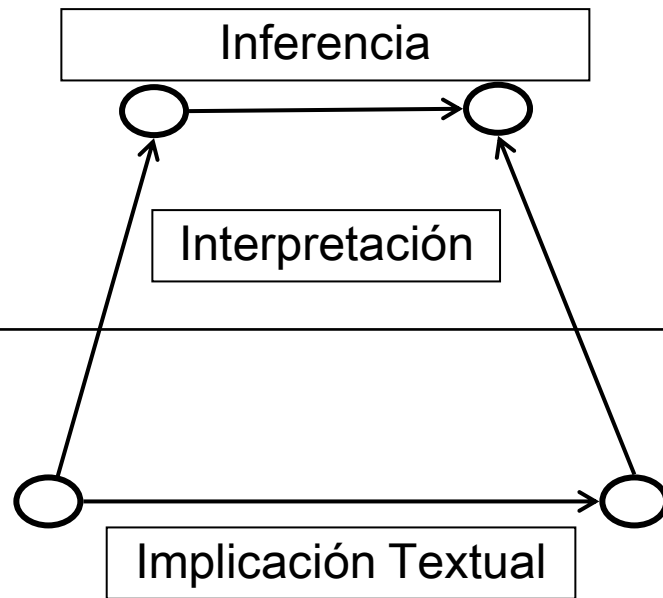
- Equivalencia: $\textit{frase1} \Leftrightarrow \textit{frase2}$
- Implicación textual:
 $\textit{frase1} \Rightarrow \textit{frase2}$



Inferencia

***Representación
del
Conocimiento***

Lenguaje





Definición

- Definición de la tarea:

Dados dos fragmentos de textos llamados T(Texto) e H(Hipótesis), la tarea de reconocimiento de implicación textual consiste en determinar si el significado de H puede ser inferido tomando T como válido o en el contexto de T.



Ejemplo: (1)

T1: “At the end of the year; all solid companies pay dividends”.

H1 : “At the end of the year; all solid insurance companies pay dividends”.

H2 : “At the end of the year, all solid companies pay cash dividends”.

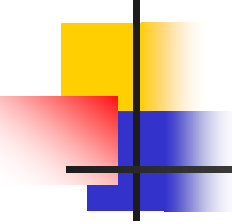
$T1 \Rightarrow H1$ is true.

$T1 \Rightarrow H2$ is false.



Ejemplos de pares

- T1: Internet media company Yahoo Inc. announced Monday it is buying Overture Services Inc. in a \$1.63-billion (U.S.) cash-and-stock deal that will bolster its online search capabilities.
- H1. Overture was acquired by Yahoo.



RTE-3: par id=33, entailment

- T2="As leaders gather in Argentina ahead of this weekends regional talks, Hugo Chávez, Venezuela's populist president, is using an energy windfall to win friends and promote his vision of 21st-century socialism."
- H2="Chávez is a follower of socialism."



RTE-5: par 437, entailment

T3: The Japanese Nikkei 225 has recorded its third biggest drop in history with a massive sell-off in the exchange that has resulted in USD 250 billion being knocked off the index's value. Toyota, which is the second largest carmaker in the world, fell by the largest amount in 21 years, while Elpida Memory, the world's largest manufacturer of computer memory, dropped in value to a record low.

H3: Japan's economy **is not flourishing**.



Implicación textual y el “Conocimiento del mundo”

- En implicación textual se supone :
 - *texto* Y “*world knowledge*” $\Rightarrow h$

Pero:

- “*world knowledge*” por si solo no debe implicar a h
- Se debe tener en cuenta la “información es dependiente del tiempo”. Por ejemplo:
- “Julian Assange es el creador de Wikileaks”.
- “El presidente actual de Ecuador tiene un doctorado en economía”.



Contradicción de pares

- *T: In that aircraft accident, four people were killed: the pilot, who was wearing civilian clothes, and three other people who were wearing military uniforms.*
- *H: Four people were assassinated by the pilot.*



Desconocimiento

- *T: Five people were killed in another suicide bomb blast at a police station in the northern city of Mosul.*
- *H: Five people were killed and 20 others wounded in a car bomb explosion outside an Iraqi police station south of Baghdad.*



RTE

RTE necesita integrar conocimiento de:

- Entidades nombradas
- Correferencias
- Dependencias sintácticas
- Desambiguación de sentidos
- Entre otros...



Aproximaciones

1. Basada en Características
2. Basadas en: Tree-based similarity
3. Basadas en Formulas Lógicas
4. Learning entailment via alignment



Métricas

- Accuracy = cantidad de respuestas correctas / cantidad de respuestas
- Entailment confidence



Aplicaciones

- El reconocimiento de la implicación textual (RTE) es útil para muchas aplicaciones : Question Answering, Information Retrieval, Sumarization, Information Extraction, etc.



Roadmap

- RTE 1,2,3: aprox. 800 dev, 800 test.
 - Decisión booleana: "entailment" y "not entailment" o Yes/No.
 - El RTE1 tuvo lugar en el 2005.
- RTE 4 y 5: Se incrementa la longitud de los textos a 40-100 palabras.

Tareas de 2 y 3 vías, decisión : "entailment", "contradiction", and "unknown"
- En RTE 5 se lanza una tarea piloto.
- RTE 6 (2010): Búsqueda de implicaciones.



Búsqueda de Implicaciones

- La tarea de búsqueda de implicación textual consiste en encontrar todas las frases/oraciones de un conjunto de documentos que impliquen una hipótesis dada.
- Los sistemas deben encontrar todas las frases que impliquen (Textos T's) en un corpus de 10 documentos de agencias de noticias acerca de un tema en común.
- Se diferencia de la tarea principal en que en la Búsqueda de Implicaciones las frases se han de interpretar en el contexto del corpus.



Desarrollo del Sistema RTE

- Experimentación con Algoritmos de ML aplicados a RTE.
- Uso de MT para expandir corpus de RTE.
- Reconocimiento de Implicación textual en Español usando ML.

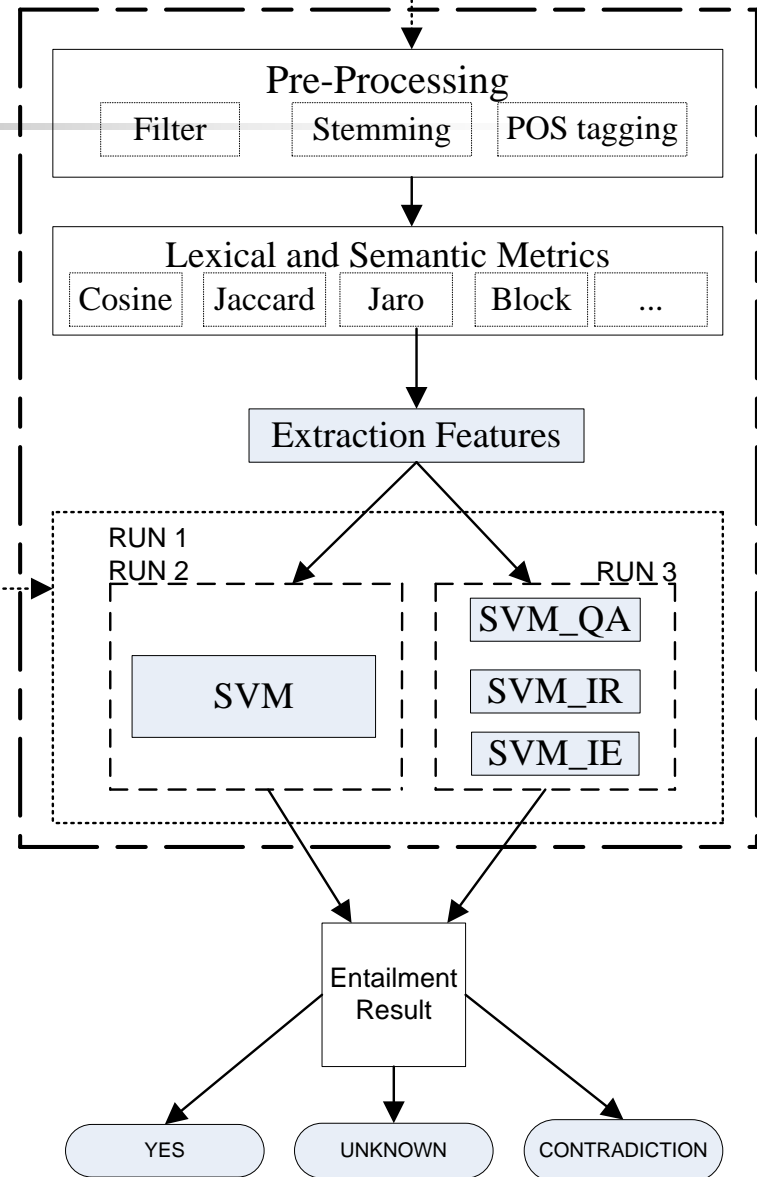
Arquitectura del Sistema en RTE5 (2009)

Características :

- Usa medidas léxicas y semánticas
- Usa varias SVMs para inferir modelos.
- Clasificación en 2 y 3 vías

RTE5 Testset

Training Sets:
RTE3 Devset,
RTE4 Testset and
RTE5 Devset





Líneas de investigación de este trabajo

- Aplicación RTE
- Búsqueda de Implicaciones
- RTE en Español
- Trabajo sobre Corpus RTE



FIN

- Muchas Gracias!
- ¿ Preguntas ? ... en el break ?