

# Tratamiento de la implicación textual a través de patrones semánticos

Saul León, Darnes Vilariño, David Pinto, Mireya Tovar, Beatriz Beltrán

Benemérita Universidad Autónoma de Puebla, Facultad de Ciencias de la Computación,  
Puebla, México

{saul.leon, darnes, dpinto, mtovar, bbeltran}@cs.buap.mx

**Resumen.** En el presente trabajo se presenta una nueva metodología para el Problema de Implicación Textual. Esta metodología propone el empare de patrones sintácticos, los cuales están contruidos por la cantidad de términos. Esta aproximación es una variante del modelo de aproximación por empatamiento. Esta metodología fue probada y validada con tres corpus ofrecidos por varias conferencias internacionales (RTE1, CLTE y SICK). Los resultados obtenidos son comparables con los mejores resultados internacionales.

**Palabras clave:** Implicación textual, similitud semántica, taxonomía de WordNet, medidas semánticas.

## 1. Introducción

Encontrar de manera automática cuando dos textos comparten la misma información y en el mismo sentido, es uno de los problemas que en la actualidad está siendo muy estudiado por la comunidad de investigadores en Procesamiento del Lenguaje Natural (PLN). Lograr resolver de manera eficiente este problema beneficiaría a varias tareas de PLN tales como: Recuperación de Información, Traducción Automática, Resúmenes Automáticos, Preguntas y Respuestas, Pruebas de Lectura Comprensiva, entre otras. Anualmente se propone la conferencia International Workshop on Semantic Evaluation (SemEval), en este sentido desde hace cuatro años se ha planteado la tarea de Implicación Textual (IT). Los modelos desarrollados hasta el momento no superan más del 80% de precisión y dependen mucho de las características del corpus de entrenamiento y del corpus de prueba.

El problema a resolver se formula de la siguiente forma: Sean  $H$  y  $T$ , dos fragmentos de textos en un mismo idioma, el problema de implicación textual consiste en determinar si se puede inferir el significado de  $H$  a partir del significado de  $T$ . Cuando esto ocurre se dice que existe una implicación textual entre ambos fragmentos. De la misma forma se puede detectar la dirección de la implicación, es decir, si  $T$  contiene más información que  $H$ , y sostienen implicación textual, se dice que la dirección de la implicación es de  $T \rightarrow H$ , en caso contrario, si  $H$  contiene más información que  $T$  y sostienen implicación textual, la dirección es  $T \leftarrow H$ .

También en muchas situaciones de la vida real es necesario detectar de manera automática cuando dos segmentos de texto, son contradictorios, esta situación es un caso particular de la implicación textual.

Para resolver el problema de la implicación textual, algunos autores han recurrido a modelos que utilizan fundamentalmente conteo estadístico, donde explotan diversas medidas de similitud a partir de un par de oraciones y como resultado confeccionan vectores característicos, los vectores representativos obtenidos alimentan a un algoritmo de aprendizaje supervisado y utilizan algún modelo de clasificación. Uno de los mejores trabajos reportados que emplea este tipo de metodología, es el presentado en [2]. En este trabajo se utilizan tres tipos de características; dentro de las características del primer tipo se emplean métricas basadas en conjuntos, suponiendo que A es el conjunto de palabras que conforman al texto 1 y B el conjunto de palabras del texto 2, entonces se calcula:  $|A|$ ,  $|B|$ ,  $|A - B|$ ,  $|B - A|$ ,  $|A \cap B|$ ,  $|A \cup B|$ ,  $|A / B|$ ,  $|B / A|$ . El segundo tipo de características es calculado a partir de métricas de similitud, algunas de estas son: Coeficiente de Jaccard, Coeficiente Dice, Coeficiente de Sobrelapamiento, Pesos de Coeficientes sobrelapados, Similitud Coseno, Distancia de Manhattan, Distancia Euclideana, Distancia de Edit, Distancia de Jaro-Winker, por mencionar algunas. Como tercer grupo de características, se cuantifican las Relaciones Gramaticales a través de dependencias funcionales, Relaciones Parciales y Número de Entidades en común.

En general, realizar un conteo estadístico de características que comparten ambos textos, ofrece un buen comportamiento siempre y cuando en los datos de prueba aparezca información que se encuentra en los datos de entrenamiento, esto se ve reflejado en que estas propuestas no logran detectar de manera correcta el juicio de implicación con una exactitud mayor al 60%.

En la actualidad se está trabajando en la construcción de modelos no supervisados, es decir que no dependan de los datos de entrenamiento. Por ejemplo, en el trabajo desarrollado en [3] busca la construcción de lo que se conoce como patrones de empatamiento. Un patrón de empatamiento se basa en la aparición de términos iguales o similares que comparten un par de oraciones, por ejemplo:

Oración 1: Leonardo Da Vinci *Painted the* Mona Lisa.

Oración 2: Mona Lisa *is the work of* Leonardo da Vinci.

Plantilla 1: X *paint* Y.

Plantilla 2: Y *is the work of* X.

Patrón de empatamiento: X *paint* Y  $\rightarrow$  Y *is the work of* X.

El uso de plantillas garantiza que la implicación textual siempre es descubierta. En la actualidad se ha propuesto una metodología no supervisada para generar patrones de empatamiento sobre un large-scale-corpus [4], sin embargo construir dichos patrones de empatamiento es una tarea compleja y nuevamente depende de las características que tengan los textos con los que se está trabajando.

Una vertiente nueva intenta generalizar los patrones extraídos aplicando lógica proposicional [5], a continuación se muestra un ejemplo:

Oración 1: Leonardo Da Vinci *Painted the* Mona Lisa.

Oración 2: Mona Lisa *is the work of* Leonardo da Vinci.

Equivalencia Lógica 1: *isPainterOf*(Da Vinci, Mona Lisa).

Equivalencia Lógica 2: *isWorkOf*(Mona Lisa, Da Vinci).

Regla:  $\forall x \forall y \text{ isPainterOf}(x,y) \rightarrow \text{isWorkOf}(y,x)$ .

La complejidad de esta metodología al igual que la de las plantillas en que es necesario construir todos los axiomas que se presentan en el texto que se está estudiando, la generación de reglas de manera automática aún no está totalmente desarrollada.

La detección de la implicación textual utilizando estas dos últimas metodologías, solamente depende de los patrones y las reglas que puedan ser descubiertas en la fase de construcción, sin embargo es conocido que los seres humanos cuando escriben, no emplean reglas ni patrones rígidos, la riqueza de los textos depende mucho del conocimiento propio de la lengua de cada persona, automatizar este proceso es prácticamente imposible, la riqueza del lenguaje y la forma en que se establecen las relaciones en el mismo es infinita, de modo que aplicar plantillas a textos de la vida real, es similar a estudiar una muestra muy pequeña de textos textualmente implicados.

La metodología de patrones de empatamiento, a pesar de su rigidez, puede ser mejorada, con vistas a detectar juicios de implicación textual en textos que no empaten con los patrones construidos, en este sentido está enfocada nuestra investigación.

Este trabajo está organizado de la siguiente forma: en la Sección 2 se presenta el modelo propuesto para abordar la IT. En la Sección 3 se presentan las características de las colecciones de datos y el comportamiento del modelo desarrollado. En la Sección 4 se analizan los resultados obtenidos para tres colecciones de datos. Finalmente las conclusiones y el trabajo a futuro se encuentran en la Sección 5.

## 2. Modelo basado en plantillas para resolver la IT

Para introducir el modelo propuesto, se considera el siguiente ejemplo:

Oración 1: *La becerra de Manuel está en la calle.*

Oración 2: *Esta es la calle de Manuel Becerra.*

Ambas oraciones comparten la misma cantidad de información, pero no transmiten la misma idea, ante esta evidencia, es necesario proponer un modelo que para descubrir la implicación textual sea capaz de detectar los *tokens* que comparten un par de oraciones, y descubrir si estos *tokens* transmiten la misma idea en ambas oraciones, en este sentido se retoma el concepto de términos anclas de la metodología de empatamiento por plantillas [15].

Consideremos el par de plantillas *X es obra de Y* y *X escribió Y*, se observa que en ambas plantillas se repiten las variables X e Y, estas variables pueden ser sustituidas por cualquier par de *tokens*, como por ejemplo *Tokio-Blues* es obra de *Murakami*, *Murakami* escribió *Tokio-Blues*; o incluso *Cien-años-de-soledad* es obra de *García-Márquez*, *García-Márquez* escribió *Cien-años-de-soledad*. Los tokens *Cien-años-de-soledad* y *García-Márquez* son denominados términos anclas, porque están presentes en ambas oraciones.

Una vez detectada las posiciones que ocupan en ambas oraciones los términos ancla, se puede extraer por cada oración la relación que sostienen, esta relación corresponde a la información que está contenida entre las anclas, siguiendo el ejemplo, la relación para las oraciones serían: *es-obra-de* y *escribió*, la manera tradicional de validar si un par de relaciones sostiene o no implicación textual, es usar un repositorio de plantillas. Si en el repositorio de plantillas no se encuentran estas relaciones se dice que no existe implicación textual. Es importante destacar que a pesar de que no se encuentren ese par de relaciones en el repositorio, puede existir una implicación textual.

En este trabajo de investigación se propone emplear medidas de similitud semántica, para establecer si dos relaciones significan lo mismo, es decir, se propone un mecanismo que determine si *es-obra-de* y *escribió* son similares semánticamente. Se proponen utilizar seis medidas de similitud semántica entre palabra y palabra y dos medidas de similitud estadística, las cuales se explican a continuación:

## 2.1 Medidas de similitud semántica basadas en la taxonomía de WordNet

Este tipo de medidas utiliza la taxonomía de WordNet [20] para encontrar la relación semántica que existe entre pares de términos, de acuerdo a la jerarquía establecida. La diferencia entre cada una de las medidas radica en la forma en que se recorre esta estructura jerárquica. Las medidas utilizadas son:

- Similitud de Path [8],
- Similitud de Leacock-Chodorow [10],
- Similitud de Wu-Palmer [11],
- Similitud de Resnik [12],
- Similitud de Jiang-Conrath [13],
- Similitud de Lin [14].

## 2.2 Medidas de similitud estadísticas

Para detectar si dos términos se encuentran relacionados se puede aplicar el algoritmo de Análisis Semántico Latente (LSA por sus siglas en inglés) [16] la idea de esta técnica es partir de una colección de documentos y si dentro de los documentos los términos aparecen relacionados es porque se refieren al mismo concepto. La limitante de esta medida está precisamente en la colección de documentos para calcular las relaciones entre los términos. El corpus debe estar conformado por un dominio de información amplio.

En los últimos años se ha utilizado con mucha frecuencia el concepto de Información Mutua (*Pointwise Mutual Information*, PMI [8]) en diversas tareas de procesamiento de lenguaje natural. Esta técnica busca detectar si dos términos son colocaciones, esto quiere decir que la probabilidad de aparición de ellos juntos, en el mismo contexto, supera a la aparición de los términos por separado. Aplicado el PMI a nuestra investigación nos permite detectar la probabilidad de que dos términos estén cerca, partiendo del hecho que si están muy cercanos reflejan el mismo sentido.

La obtención adecuada de los valores PMI dependen totalmente del tamaño del corpus de donde se extraen, de la misma manera que en LSA, se recomienda un corpus de dominio amplio.

### 2.3 Eliminación de *tokens* para el descubrimiento de la implicación textual

Para detectar el juicio de implicación textual que sostienen dos oraciones, se sigue el siguiente razonamiento: Las oraciones tienen información en común, unas más y otras menos, si se toman estos elementos comunes y se eliminan en ambas oraciones, quedarían términos únicos en cada oración. Los términos sobrantes, al finalizar el proceso de eliminación, se pueden cuantificar en relación a la longitud original de la oración, es decir, que es posible medir el porcentaje de elementos eliminados y así determinar el juicio de implicación textual, por ejemplo si al terminar de eliminar los elementos comunes de un par de oraciones, se obtiene que se ha eliminado más del 80% en ambas oraciones, quiere decir que la cantidad de información compartida es mucha y por lo tanto se dice que existe implicación textual. Como caso particular, se tiene el juicio de implicación textual de contradicción, para detectar éste es necesario introducir elementos de negación propios del idioma con el que se está trabajando, por ejemplo para el idioma inglés se consideran: no, none, nobody, there-\*,-not, \*n't, así como la presencia de antonimia.

Se propone un algoritmo de carácter general que permite obtener el grado de relación entre un par de sentencias, independientemente del tipo de juicio de implicación que se desea detectar, el proceso que aplica es el siguiente: Los términos anclas por simplicidad son los *n-gramas* de palabras más largos presentes en ambas oraciones. Para el procesamiento se tokenizan ambas oraciones y se verifica la distancia a la que se encuentran las anclas, si los términos anclas no exceden un máximo de dos *tokens* entre ellos en ambas oraciones, se verifican las relaciones que sostienen en una y en otra, para ello se emplean las medidas de similitud. Si las relaciones son similares, se eliminan estos términos (criterio de eliminación) y las anclas involucradas en cada oración, ver Algoritmo 1.

El algoritmo propuesto, emplea las siguientes funciones:

- **obtener\_anclas:** Esta función genera los *n-gramas* más largos que comparten ambas oraciones.
- **are\_near:** Esta función regresa verdadero si en una oración, los términos ancla se encuentran a no más de 2 tokens de distancia, en caso contrario el valor de retorno es falso.
- **obtener\_relacion:** Esta función regresa el segmento de texto de una oración, que se encuentra acotado por un par de anclas.
- **are\_similar\_relations:** Es una función que decide si dos segmentos de texto son similares o están relacionados, internamente esta función hace uso de las 8 medidas de similitud. Dentro de la función se ha establecido un umbral (REL), si la medida de similitud semántica supera dicho umbral entonces se dice que la relación es válida, que en otras palabras significa que es un plantilla válida.

```
Funcion obtener_grado_de_relacion(Sentencia S1,
                                Sentencia S2,
                                Vocabulario contradiccion) {
    anclas = obtener_anclas(S1, S2)
```

```
tokens_1 = tokenizar(S1)
tokens_2 = tokenizar(S2)

hay_contradicción = false
for ci in contradicción
    si ci in s1 or ci in s2 or
        existen_antonimos_entre(s1, s2)
        hay_contradicción = true

len_s1_inicial = length(S1)
len_s2_inicial = length(S2)

para cada n-grama ai en anclas hacer
    para cada n-grama aj en anclas hacer
        si ai != aj and are_near(ai, aj, tokens_1) and
            are_near(ai, aj, tokens_2) hacer
                r1 = obtener_relacion(ai, aj, tokens_1)
                r2 = obtener_relacion(ai, aj, tokens_2)

                si are_similar_relations(r1, r2) hacer
                    eliminar ai, r1 y aj de S1
                    eliminar ai, r2 y aj de S2
                    eliminar ai y aj de anclas

len_s1_final = length(S1)
len_s2_final = length(S2)

p1 = 1 - (len_s1_final / len_s1_inicial)
p2 = 1 - (len_s2_final / len_s2_inicial)

regresar p1, p2, hay_contradicción
}
```

**Algoritmo 1.** Algoritmo empleado para obtener el grado de similitud entre un par de sentencias.

### 3. Caso de estudio

Como la metodología propuesta es de carácter general, se propone utilizar tres colecciones de datos, que buscan detectar juicios de implicación textual, pero de diferente tipo. A continuación se describen las tres colecciones de datos, así como la forma de detectar en cada caso el juicio de implicación.

#### 3.1. Conjunto de datos RTE1

Esta colección está diseñada solamente para detectar los juicios de implicación textual: *entailment* y *no\_entailment*. Es una colección que contiene 800 instancias de

las cuales 400 son *entailment* y 400 son *no\_entailment* [17]. Para esta colección a partir de los resultados que obtiene el Algoritmo 1, se hace lo siguiente:

Sea  $C$  el conjunto de elementos comunes entre las oraciones  $S1$  y  $S2$ .

Sea  $P1$  el porcentaje restante de la oración  $S1$  al eliminar los elementos de  $C$ .

Sea  $P2$  el porcentaje restante de la oración  $S2$  al eliminar los elementos de  $C$ .

Sea  $U$  un umbral de porcentaje.

El juicio de implicación se obtiene aplicando las siguientes reglas:

Si  $P1 > U$  y  $P2 > U$  el juicio de implicación es *entailment*, en caso contrario el juicio de implicación es *no\_entailment*.

### 3.2. Conjunto de datos CLTE

Esta colección maneja que  $S1$  se encuentre en un idioma y  $S2$  en otro, sin embargo se puede traducir la oración  $S2$  para que se encuentre en el mismo idioma de  $S1$ , que para esta colección es el idioma inglés [18]. La colección está compuesta por 1000 pares de oraciones, divididas en los siguientes 4 juicios:

- *Bidirectional*: Ambas sentencias comparten la misma información, hablan del mismo tópico y tienen el mismo sentido.
- *Backward*: La segunda oración ( $S2$ ) contiene más información que la primera ( $S1$ ), hablan del mismo tópico y poseen el mismo sentido.
- *Forward*: La primera oración ( $S1$ ) contiene más información que la segunda ( $S2$ ), hablan del mismo tópico y poseen el mismo sentido.
- *No\_entailment*: Las oraciones  $S1$  y  $S2$  no hablan del mismo tópico, ni expresan lo mismo.

Para descubrir el juicio de implicación a partir de los resultados obtenidos del algoritmo 1, se aplican las siguientes reglas.

Sea  $C$  el conjunto de elementos comunes entre las oraciones  $S1$  y  $S2$ .

Sea  $P1$  el porcentaje restante de la oración  $S1$  al eliminar los elementos de  $C$ .

Sea  $P2$  el porcentaje restante de la oración  $S2$  al eliminar los elementos de  $C$ .

Sea  $U$  un umbral de porcentaje.

El juicio de implicación se obtiene aplicando las siguientes reglas:

Si  $P1 > U$  y  $P2 > U$ : El juicio es bidireccional, ya que ambas oraciones comparten mucha información.

Si  $P1 < U$  y  $P2 > U$ : El juicio es *backward*, porque la oración  $S2$  está contenida en la oración  $S1$ , y  $S1$  contiene mucha más información que  $S2$ .

Si  $P1 > U$  y  $P2 < U$ : El juicio es *forward*, porque la oración  $S1$  está contenida en la oración  $S2$ , y  $S2$  contiene mucha más información que  $S1$ .

Si  $P1 < U$  y  $P2 < U$ : El juicio de *no-entailment*, ya que ambas oraciones comparen muy poca información.

### 3.3. Conjunto de Datos SICK

Esta colección fue extraída de los datos ofrecidos en la tarea 1 del SemEval-2014 [9], se propone una colección 4,927 pares de oraciones, donde 2,793 tienen el juicio de implicación neutral, 1,414 tienen juicio de implicación *entailment* y 720 tienen juicio de *contradiction*. El juicio de implicación *neutral* significa que ambas oraciones hablan del mismo tópico. Mientras que el juicio de *contradiction* significa que una es la negación de la otra.

Como el juicio de implicación neutral, puede o no incluir implicación textual, consideramos eliminarlo del conjunto de datos, por lo que finalmente se trabaja con los juicios *entailment* y *contradiction*, haciendo un total de 2,134 pares de oraciones.

Para detectar el juicio de implicación se analiza la salida del Algoritmo 1, si la bandera que indica contradicción está prendida, se dice que ese par de oraciones tiene un juicio de *contradiction*, en caso contrario se aplican las mismas reglas que para el conjunto de datos RTE1.

## 4. Análisis de resultados

Se realizaron alrededor de 130,000 diferentes experimentos variando el valor de los umbrales de porcentaje  $U$  y  $REL$ , que se movieron en un rango de 0.05 a 0.95, con un incremento de 0.05. En la Tabla 1 se muestran los resultados obtenidos utilizando las 8 medidas de similitud. La medida propuesta por Wu-Palmer mostró un desempeño superior al resto.

Las medidas de WordNet, que son las seis primeras en la tabla, muestran un comportamiento similar para los 3 corpus. Para el corpus SICK, se obtiene una exactitud mayor que para el resto de los corpus. Este resultado es coherente con las características propias de este corpus, ya que ha sido construido con más cuidado, por ejemplo, en el corpus CLTE se ha detectado que una misma oración está asociada a dos clases distintas. Es importante destacar que el corpus SICK sólo se encuentra clasificado en tres categorías, mientras que el corpus CLTE está categorizado en cuatro clases.

Analizando los resultados obtenidos se puede observar que las medidas de similitud PMI y LSA ofrecen un grado de exactitud similar, PMI supera a LSA en una centésima. Se puede concluir que las medidas semánticas que ofrece WordNet, en general, superan a las medidas estadísticas, calculadas a través de PMI y LSA. Se debe destacar que las medidas estadísticas se ven afectadas por las características del corpus de donde se obtienen. Se sabe, que para que PMI ofrezca resultados relevantes, se necesita que el corpus sea muy grande y construido utilizando dominios diferentes. El mismo comportamiento ofrece LSA, ya que los conceptos que va a descubrir tienen que encontrarse en el corpus de referencia.



**Tabla 1.** Resultados utilizando las 8 medidas.

Medidas	Corpus		
	RTE1	CLTE	SICK
Path	0.521	0.574	0.652
Leacock-Chodorow	0.521	0.568	0.661
Wu-Palmer	0.525	0.576	0.679
Resnik	0.520	0.554	0.677
Jiang-Conrath	0.520	0.554	0.677
Lin	0.520	0.554	0.677
LSA	0.521	0.566	0.664
PMI	0.521	0.574	0.672

Por otro lado, la buena inferencia de relaciones, está en función de la calidad de los términos anclas, si estos son consistentes se asegura que la relación entre ellos es válida, mientras que si los términos anclas son ruidosos, las relaciones que se logran inferir son invalidas.

#### 4.1. Patrones válidos

A continuación se muestran algunos patrones válidos detectados por la metodología:

- X **take-captive** Y = X **kidnap** Y
- X **peculiar-anthropomorphic-creature** Y = X **stranger** Y
- X **nickname** Y = X **know-as** Y
- X **arrest** Y = X **take-imprison** Y
- X **together-know** Y = X **collective-refer** Y
- X **call** Y = X **well-know** Y
- X **down-payment** Y = X **discount** Y
- X **initial** Y = X **preliminary** Y
- X **associate** Y = X **link** Y
- X **topple** Y = X **overthrow** Y

La metodología propuesta, también permite determinar si las relaciones extraídas no son válidas, esto se hace posible gracias al umbral REL, a continuación se muestran algunos patrones no válidos.

#### 4.2. Patrones no válidos

- X **place** Y != X **consist** Y
- X **shell** Y != X **bomb** Y

- X **europa** Y != X **ban** Y
- X **jail** Y != X **arrest** Y
- X **justice** Y != X **begin** Y
- X **hindi** Y != X **use** Y
- X **get** Y != X **produce** Y
- X **happen** Y != X **center** Y

Se ha detectado que los patrones no válidos tienen una relación directa con la calidad de las anclas, que éstas están en función de los procesos de tokenización y consolidación de términos, que consiste en la generación de *n-gramas* comunes. Es importante destacar que los patrones son extraídos en tiempo de ejecución, directamente del par de texto de los que se desea descubrir el juicio de implicación.

Los resultados obtenidos no pueden ser comparados de manera directa con la metodología propuesta, y que estas no detectan el juicio de implicación a través del uso de plantillas. Las plantillas se han propuesto de manera general, pero no han sido aplicadas sobre ningún corpus. Para visualizar el comportamiento se ha realizado una comparación de los resultados obtenidos por esta metodología con propuestas de los mejores modelos a nivel internacional y de modelos desarrollados por nuestro grupo de investigación.

Todos los algoritmos del estado del arte emplean aprendizaje supervisado, mientras que esta metodología es no supervisada; los algoritmos del estado del arte emplean módulos de normalización, análisis estadístico, análisis semántico y hasta teoría de grafos, mientras que la propuesta introducida se basa en la eliminación de plantillas. Todo esto podría conducirnos a una mala interpretación de los resultados obtenidos, sin embargo en la tabla 2 se ofrece una comparativa de los resultados obtenidos por esta metodología, empleando como medida de similitud a Wu-Palmer, así mismo se muestran los resultados de otras propuestas de solución empleadas por nuestro grupo de investigación.

**Tabla 2.** Comparación de resultados.

<b>Corpus</b>	<b>Ejecución</b>	<b>Exactitud</b>
RTE1	Manning	0.631
	BUAP-Plantillas	0.525
	HDU-Run2	0.632
CLTE	BUAP-Run2	0.366
	BUAP-Plantillas	0.576
	Illinois-LH_run1	0.845
SICK	BUAP-Run1	0.796
	BUAP-Plantillas	0.679

Analizando la tabla 2, con respecto al corpus CLTE, puede apreciarse que la aproximación utilizando plantillas supera considerablemente a la aproximación enviada al SemEval-2012 (BUAP-Run2), esta aproximación está basada en la selección de características que comparten ambas sentencias [19] y este modelo fue supervisado, es decir se aplicó un modelo de clasificación para detectar el juicio de implicación textual.

A pesar de que esta metodología ha ofrecido mejores resultados no supera a la aproximación HDU-Run2 que ofreció un grado de exactitud del 63%.

La aproximación enviada por nuestro grupo de investigación al SemEval-2014 (BUAP-Run1) aprovecha el tipo de juicio que desea descubrir, ya que construye un conjunto de reglas que le permite prácticamente detectar el juicio CONTRADICTION, y es por ello que el grado de exactitud es prácticamente del 80%. La aproximación utilizando plantillas no incluyó el juicio de implicación neutral, por lo que el resultado obtenido no es comparable.

La metodología propuesta para el Corpus del RTE, no logró detectar más de la mitad de los juicios de implicación *entailment* y *no\_entailment*.

## 5. Conclusiones

Hemos presentado un método para resolver el problema de implicación textual, independientemente del corpus que se está estudiando. La metodología propone inicialmente encontrar anclas entre las dos oraciones, con el objetivo de comparar la similitud entre las relaciones que quedan entre los términos anclas, una vez aplicado el algoritmo de eliminación. Para la búsqueda de la similitud entre las relaciones se propuso utilizar las 6 medidas de WordNet, el Algoritmo LSA y el Algoritmo PMI. Esta propuesta viene dada porque si quedan términos en el medio de las anclas, después de la eliminación, quiere decir que no se pudieron aplicar criterios de sinonimia, pero partimos de la hipótesis que si son términos similares o hablan del mismo contexto, en la taxonomía de WordNet son más cercanos.

Las medidas estadísticas LSA y PMI, no ofrecieron resultados satisfactorios, se considera que esto es producto del corpus que se utilizó para obtenerlas. Se construyó un corpus utilizando EuroParl, Documentación de ayuda de PHP, OpenOffice, novelas extraídas del Proyecto Gutenberg y por último un fragmento filtrado de la Wikipedia. A pesar de que el corpus obtenido es de dominio amplio hay pares de palabras de las que no se pudo detectar el grado de similitud. Es importante destacar que esta metodología no se puede utilizar para el idioma español, pues no se dispone de WordNet en español.

La ventaja de esta propuesta es que propone resolver el problema de detectar el juicio de implicación textual de manera no supervisada y nos permite descubrir patrones de relación, a partir de los términos anclas. Con estos patrones como trabajo a futuro, se puede detectar fácilmente el juicio de implicación.

## Referencias

1. Dagan, I., Glickman, O., Magnini, B.: The PASCAL recognizing textual entailment challenge. *Machine learning challenges: evaluating predictive uncertainty, visual object classification, and recognising textual entailment*, pp. 177–190 (2006)
2. Zhao, J., Lan, M., Niu, Z.: ECNUCS: Recognizing Cross lingual Textual Entailment Using Multiple Text Similarity and Text Difference Measures. Atlanta, Georgia, USA, pp. 118–124 (2013)

3. Zhao, S., Lan, X., Liu, T., Li, S.: Application driven statistical paraphrase generation. In: Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL, Vol. 2, pp. 834–842 (2009)
4. Kouylekov, M., Magnini, B.: Building a large-scale repository of textual entailment rules. In: Proceedings of LREC (2006)
5. Bos, J.: Is there a place for logic in recognizing textual entailment? Linguistic Issues in Language Technology, Vol. 9 (2013)
6. Harabagiu, S., Hickl, A.: Methods for using textual entailment in open-domain question answering. In: Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics, pp. 905–912 (2006)
7. Carrillo, M., Vilariño, D., Pinto, D., Tovar, M., León, S., Castillo, E.: FCC: Three Approaches for Semantic Textual Similarity. In: Proceedings of the First Joint Conference on Lexical and Computational Semantics, Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation, pp. 631–634 (2012)
8. Mihalcea, R., Corley, C., Strapparava, C.: Corpus-based and knowledge-based measures of text semantic similarity. AAAI, Vol. 6, pp. 775–780 (2006)
9. Marelli, M., Menini, S., Baroni, M., Bentivogli, L., Bernardi, R., Zamparelli, R.: A SICK Cure for the Evaluation of Compositional Distributional Semantic Models. In: Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14) (2014)
10. Leacock, C., Chodorow, M.: Combining local context and WordNet similarity for word sense identification. In: Fellbaum, Christiane (ed.), WordNet, pp. 265–283 (1998)
11. Zhibiao, W., Stone, M.: Verb Semantics and Lexical Selection. In: ACL, pp. 133–138 (1994)
12. Resnik, P.: Using Information Content to Evaluate Semantic Similarity in a Taxonomy. In: Proceedings of the 14th International Joint Conference on Artificial Intelligence, pp. 448–453 (1995)
13. Jiang, J., Conrath, D.: Semantic similarity based on corpus statistics and lexical taxonomy. In: Proc of 10th International Conference on Research in Computational Linguistics (ROCLING'97), pp. 19–33 (1997)
14. Lin, D.: An Information-Theoretic Definition of Similarity. In: Proceedings of the Fifteenth International Conference on Machine Learning, pp. 296–304 (1998)
15. Zhao, S., Xiang, L., Ting, L., Shen, L.: Application-driven statistical paraphrase generation. In: 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: Vol. 2, pp. 834–842 (2009)
16. Deerwester, S., Dumais, S., Furnas, G., Landauer, T.: LSA: Indexing by Latent Semantic Analysis. Journal of the American Society for Information Science (1986-1996), pp. 391 (1990)
17. Dagan, I., Glickman, O., Magnini, B.: The PASCAL recognizing textual entailment challenge. Springer Berlin Heidelberg, pp. 177–190 (2006)
18. Negri, M., Marchetti, A., Mehdad, Y., Bentivogli, L., Giampiccolo, D.: Semeval-2013 Task8: Cross-lingual Textual Entailment for Content Synchronization. In: Proceedings of the 7th International Workshop on Semantic Evaluation (SemEval 2013) (2013)
19. Vilariño, D., Pinto, D., León, S., Alemán, Y., Gómez-Adorno, H.: N-gram based Feature Evaluation for the Cross-Lingual Textual Entailment Task. In: Second Joint Conference on Lexical and Computational Semantics (\*SEM), Vol. 2: Seventh International Workshop on Semantic Evaluation (SemEval 2013), pp. 124–127, (2013)
20. <http://wordnetweb.princeton.edu/perl/webwn>