

# The Effect of Entity Recognition in the Answer Validation

Álvaro Rodrigo, Anselmo Peñas, Felisa Verdejo  
Departamento de Lenguajes y Sistemas Informáticos  
Universidad Nacional de Educación a Distancia  
Madrid, Spain  
{alvaroroy, anselmo, felisa}@lsi.uned.es

## Abstract

The Answer Validation Exercise (AVE) 2006 is aimed at developing systems able to decide whether the answer of a Question Answering (QA) system is correct or not using textual entailment. The most answers to be validated are given from questions that need entities as responses. The paper presents a system that has only used entities to participate in the AVE 2006. The results of the propose system are better than the ones of a baseline system that always accepts all answers, therefore the use of entities can improve the results of an answer validation system.

## Keywords

Question Answering, Answer Validation, Textual Entailment, Entity Recognition

## 1 Motivation

The Answer Validation Exercise (AVE) [6] of the Cross Language Evaluation Forum (CLEF) 2006 is aimed at developing systems able to decide whether the answer of a Question Answering (QA) system is correct or not using textual entailment [2]. The test corpus has hypothesis-text pairs where the hypothesis contains the response from a QA system and the text is the snippet given by the system to support its answer. Participant systems must return YES or NO for each hypothesis-text pair to indicate if the text entails the hypothesis or not (i.e. the answer is correct according to the text).

The paper shows a participant system that only uses information about entities (numeric expressions, temporal expressions and named entities) in order to study the importance of entities in answer validation.

The motivation for this experiment comes from the study of the QA task in the CLEF 2005, where 75% of the questions in Spanish were factoids [7]. The responses to factoid questions contain entities (e.g. person names, locations, numbers, dates...). In this way and following the methodology to build the test corpus in the AVE, the most of the pairs will contain entities. For this reason, we think it is important the study of entities in this task.

In this experiment, we have used the hypothesis that in the recognition of textual entailment all the elements in the hypothesis must be entailed by elements of the text. We consider this hypothesis mainly must happen with the entities due to entities contain the main information about the hypothesis-text pairs.

## 2 System description

The propose system receives hypothesis-text pairs and decides, only using information about entities, if the text entails the hypothesis. Next subchapters describe in detail the components of the system.

### 2.1 Entity recognition

First, the system recognises the entities of the hypothesis-text pairs. We have used the Freeling [1] Name Entity Recogniser (NER) to tag numeric expressions (NUMEX), named entities (NE) and time expressions (TIMEX), using the same label (ENTITY) to tag these elements. We decided to use only one label for two reasons:

- To reduce errors due to a wrong recognition. The recogniser sometimes tags an entity with a wrong label. For example, the expression 1990 is sometimes recognized as a numeric expression although it usually appears in texts as a temporal expression. With this kind of errors, an entity of a certain type in the hypothesis could not find an entity of the same type in the text that entails it. An example is shown in figure 1. In the example of the figure the expression 1990 is a year but it is recognized as a numeric expression in the hypothesis, however the same expression is recognized as a temporal expression in the text, therefore there is no numeric expression in the text that entails the numeric expression in the hypothesis.
- We think in the answer validation it is important to decide if there is or no entailment between the entities without consider the type of these entities.

```
<t>...Irak invadió Kuwait en <TIMEX>agosto_de_1990</TIMEX>...</t>  
<h>Irak invadió Kuwait en <NUMEX>1990</NUMEX></h>
```

Figure 1: An error detecting the type of the entity.

### 2.2 Entailment between entities

Once entities of the hypothesis-text pairs are tagged, the next step is to find if the entities in the hypothesis are entailed by the entities in the text. In this process, for each entity in the hypothesis the system searches an entity in the text that entails it. If there is an entity in the hypothesis that is not entailed by one or more entities in the text, then the system responses that there is not entailment between entities in the pair.

The system responses that an entity E1 entails an entity E2 if the text string of E1 contains the text string of E2. This idea of entailment between entities gives better results than the idea of the two entities must be the same, avoiding errors in pairs as the ones in figure 2. In the text of the first pair in the figure appears the first name and the surname of a person whereas in the hypothesis only appears the surname of the same person. In the second pair of the same figure, a time expression in the text contains day, month and year and an equivalent time expression in the hypothesis only contains month and year. In these examples the entity in the text entails the entity in the hypothesis, but if a system requires that the entity in the text must be the same as the entity in the hypothesis, then the system would return the value NO.

### 2.3 Entailment decision

The system only uses to determine the entailment value of each pair the information about entailment between entities. In this way, we can study the relevance of the use of entities in the automatic answer validation.

<t>...de la OLP, presidida por <ENTITY>Yaser.Arafat</ENTITY>...</t>  
<h><ENTITY>Arafat</ENTITY> preside la OLP.</h>

<t>...durante la fuga de Chernobyl el <ENTITY>26.de.abril.de.1986</ENTITY>...</t>  
<h>La catástrofe de Chernobyl ocurrió en <ENTITY>abril.de.1986</ENTITY></h>

Figure 2: Pairs that justify the process of entailment.

Table 1: Results of the propose system and the baseline system.

System	F-measure	Precision over YES	Recall over YES
Proposed system	0.5315	0.4364	0.6796
100% YES Baseline	0.4538	0.2935	1

The system uses the hypothesis about textual entailment from the section Motivation and adapts it to the use of entities. When an entity in the hypothesis is not entailed by one or more entities in the text, then the system’s output to this pair is NO. However, in pairs where there is entailment between entities, there is not enough information to decide if the value of the pair is YES or NO. To study the importance of the effect of the entities in the answer validation, we decided to compare the system with a baseline system that gives always YES. Then, our system returns the value YES when it detects that there is entailment between entities in the pair.

### 3 Results

The propose system has been checked over the Spanish test set of the AVE 2006. The results of the proposed system and the results of a baseline system that always returns YES are shown in table 1.

The proposed system has obtained better results than the baseline system, achieving an increase of 0.8 in the F-measure (which it is used to rank systems in the AVE).

### 4 Discussion

After submitting the results to the AVE 2006, the errors of the system were studied. First, we studied pairs with value YES in which the system had returned the value NO. The most important errors were:

- Some errors were produced by an incorrect representation of the entities. One of these errors appears in some entities that represent years due to the representation sometimes contains the word *año* (year in Spanish) with the numeric value of the year as in the hypothesis of figure 3. Entities with this error in the representation can not be entailed by entities without this error, as it is shown in figure 3 where the entity in the text does not entail the entity in the hypothesis using the method described in the section Entailment between entities. In a new version of the system, this error does not appear after deleting the word *año* in the representation.
- Another kind of errors comes from the format of the QA systems answers. These answers sometimes are in capital letters and the NER recognised them as named entities and they are not usually named entities, therefore there are sometimes false entities (words that are not entities) in the hypothesis. The system can not find entities in the text that entail these false entities and then the system will return the value NO despite the value of the pair could

Table 2: Results of the new system.

System	F-measure	Precision over YES	Recall over YES
Improved Proposed system	0.5895	0.4579	0.8271

be YES. To solve this problem, QA systems must take into account that their output can be used by answer validation systems and return the answers as they are in the supporting texts.

- In the same way, some QA systems return supporting text snippets in lower case. The NER can not recognise all the named entities in these texts. In these situations, entities of the hypothesis do not find entities in the text that entail them because some entities have not been tagged in the text.
- Some errors were detected due to there are sometimes some characters that change between named entities as, for example, in a proper noun with different wordings (e.g. Yasser Arafat and Yaser Arafat). A new version of the system uses the edit distance of Levenshtein [5] to solve this problem taking the idea of [4]: if two entities differ in less than 20%, then there is entailment between these entities.

```
<t>..cuando Irak invadió Kuwait en <ENTITY>1990</ENTITY></t>
<h>Irak invadió Kuwait en el <ENTITY>año.1990</ENTITY></h>
```

Figure 3: Example of an error in the representation.

With a new system that includes a best representation of years and the use of the edit distance of Levenshtein a new experiment was accomplished. The results obtained with this new system over the same test corpus are shown in table 2. The F-measure of the new system is 0.5 better than the one of the previous system.

In the analysis of pairs with value NO in which the system had returned the value YES, a high percentage of errors comes from the evidence that the information of entailment between entities is not enough to determine the entailment value of some pairs. An example is shown in figure 4. In the example the only entity of the hypothesis is Hubble, which is entailed by an exact entity in the text. In the example, the propose system returns YES despite the entailment value is obviously NO. Therefore, the system needs more information to decide about the entailment value of the pair.

```
<t>... el telescopio espacial <ENTITY>Hubble</ENTITY>...</t>
<h>El <ENTITY>Hubble</ENTITY> es una imagen</h>
```

Figure 4: Example of an error in the representation.

## 5 Conclusions and future work

This paper has proposed a system that only uses entities to the answer validation of the AVE 2006. The propose system has obtained better results than a baseline system that always accepts all answers (return YES in all pairs), what means that the entailment between entities can give important information to decide if a text entails a hypothesis and therefore in the answer validation. However, in some pairs an answer validation system needs additional information.

The information about entities that has been proposed can be used by systems in two different ways:

- As a filter to detect pairs with entailment value NO, as in the proposed system, using another system to determine if there is or no entailment in the other pairs.
- As additional information to train a system that uses machine learning.

Future work is oriented in improving the representation of the entities to avoid errors. Another research line is to achieve a more complex process of entailment decision. In this way, we are working in the detection of numeric expressions that represent ranges to adapt the numerical entailment module used in [3] to the proposed system.

## Acknowledgments

This work has been partially supported by the Spanish Ministry of Science and Technology within the project R2D2-SyEMBRA. (TIC-2003-07158-C04-02) and a PhD grant by UNED.

## References

- [1] J. Atserias, B. Casas, E. Comelles, M. González, L. Padró, and M. Padró. FreeLing 1.3: Syntactic and semantic services in an open-source NLP library. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC'06)*. Genoa, Italy. 2006.
- [2] Ido Dagan, Oren Glickman, and Bernardo Magnini. The pascal recognising textual entailment challenge. In *Proceedings of the First PASCAL Recognizing Textual Entailment Workshop*. Southampton, United Kingdom., 2005.
- [3] Jesús Herrera, Anselmo Peñas, Álvaro Rodrigo, and Felisa Verdejo. Uned at pascal rte-2 challenge. In *Proceedings of the Second PASCAL Challenges Workshop on Recognising Textual Entailment*. Venezia, Italy., 2006.
- [4] Jesús Herrera, Anselmo Peñas, and Felisa Verdejo. Textual entailment recognition based on dependency analysis and wordnet. In Springer, editor, *Proceedings of the First PASCAL Recognizing Textual Entailment Workshop*. Southampton, United Kingdom., LNAI, 2005.
- [5] V. I. Levensthein. Binary Codes Capable of Correcting Deletions, Insertions and Reversals. In *Soviet Physics - Doklady*, volume 10, pages 707–710, 1966.
- [6] Anselmo Peñas, Álvaro Rodrigo, Valentín Sama, and Felisa Verdejo. Overview of the answer validation exercise 2006. In *this volume*, 2006.
- [7] A. Vallin, B. Magnini, D. Giampiccolo, L. Aunimo, C. Ayache, P. Osenova, A. Peñas, M. de Rijke, B. Sacaleanu, D. Santos, and R. Sutcliffe. Overview of the clef 2005 multilingual question answering track. In *Proceedings of CLEF 2005*, 2005.