Named Entity Recognition in Questions:
Towards a Golden Collection

Ana Cristina Mendes, Luísa Coheur, Paula Vaz Lobo

Spoken Language Systems Laboratory - L^{2}F/INESC-ID
Instituto Superior Técnico, Technical University of Lisbon
R. Alves Redol, 9 - 2º– 1000-029 Lisboa, Portugal
{ana.mendes,luisa.coheur,paula.vaz}@l2f.inesc-id.pt

Abstract

Named Entity Recognition (NER) plays a relevant role in several Natural Language Processing tasks. Question-Answering (QA) is an example of such, since answers are frequently named entities in agreement with the semantic category expected by a given question. In this context, the recognition of named entities is usually applied in free text data. NER in natural language questions can also aid QA and, thus, should not be disregarded. Nevertheless, it has not yet been given the necessary importance.

In this paper, we approach the identification and classification of named entities in natural language questions. We hypothesize that NER results can benefit with the inclusion of previously labeled questions in the training corpus. We present a broad study addressing that hypothesis and focusing, among others, on the balance to be achieved between the amount of free text and questions in order to build a suitable training corpus. This work also contributes by providing a set of nearly 5,500 annotated questions with their named entities, freely available for research purposes.

1. Introduction

The task of identifying and classifying entities in natural language texts, denoted Named Entity Recognition (NER), has been attributed increasing importance for some years now. Approaches to NER can roughly be split in two lines of study: hand-crafted rule-based and machine learning-based. Despite the impressive marks machine learning-based systems can achieve, their high reliance on the training phase is an important drawback: changing the format of the test corpus while maintaining the training corpus can lead to unpleasant results.

Aiming to create resources that can be employed in training machine learning-based NER systems, substantial efforts have been put on labeling free text data, often collected from newspaper articles. Nevertheless, a corpus composed merely of questions greatly differs from a free text corpus, due to the particular characteristics of questions: they tend to be shorter than sentences found, for instance, on newspaper articles, reducing considerably the context needed for disambiguation. Given the aforementioned, the guess is that recognizing named entities in questions will not work well if using solely free text as training data, since both have distinct characteristics.

NER is of crucial importance in several Natural Language Processing tasks, like Question-Answering (QA) (Toral et al., 2005). Nevertheless, till now and to our knowledge, no work has focused the NER task applied uniquely to natural language questions. Partly similar this subject, a recent work (Guo et al., 2009) presents a weakly supervised learning model for recognizing and classifying named entities in queries, using query log data. Queries are considered as very short ill-formed sentences (2-3 words on average), and the proposed method is able to assign a set of probable categories to named entities in queries (only to one named entity per query); questions, as they are seen in our work, are short well-formed sentences, typically of type interrogative,

and we assume that only one category is to be assigned per named entity (multiple named entities can exist in a single question).

In this work we approach the identification and classification of named entities in questions, through machine learning techniques. We hypothesize that NER results can benefit with the inclusion of previously labeled questions in the training corpus.

This work contributes by providing nearly 5,500 annotated questions to be used as training corpus in machine learning-based NER systems, available in `https://qa.l2f.inesc-id.pt/wiki/index.php/Resources`.

The named entities in these questions were identified and classified according to the categories: PERSON, LOCATION and ORGANIZATION. For that purpose, we extended the guidelines of the shared task of the Conference on Computational Natural Language Learning (CoNLL) 2003 on NER to face the demands presented by questions.

Also, this work presents a broad study on the creation of language models with the concerns: *1)* which machine learning technique better applies; and, *2)* how to balance the amount of free text and questions in order to build a suitable training corpus.[1]

The remainder of this paper is organized as follows: Section 2. presents the resources used in this work; Section 3. describes the directives that guided the annotation of questions; Section 4. describes the experiments we conducted to assess the impact training parameters on NER on questions and Section 5. presents a brief discussion on the NER task applied to questions. The paper finishes in Section 6., where conclusions are drawn and future work directions are presented.

---

[1] We consider a training corpus to be suitable if it generates a language model which attains an F-measure greater than 75, when evaluated against a golden test collection.

2. Resources

This section presents the resources (corpora and software) used in this work for the automatic discovery and classification of named entities.

2.1. Corpora

Different resources were used as corpora[2]:

TrainFT A set of news wire articles from the Reuters Corpus, Volume 1 (Lewis et al., 2004), manually annotated in the University of Antwerp for the shared task of the 2003 edition of the CoNLL (CoNLL-2003) focusing Language-Independent Named Entity Recognition (Tjong Kim Sang and De Meulder, 2003);

DevQ A collection of 5,500 questions gathered by Li&Roth (Li and Roth, 2002), used in several learning question classification experiments;

EvalFT A set of news wire articles from the Reuters Corpus, Volume 1 (TrainFT and EvalFT are disjoint sets);

EvalQ A collection of 500 questions from the TREC 10 Question-Answering task (also used in the work of Li&Roth, but not belonging to DevQ).

Table 1 presents some statistics about the used corpora.

	Type	Size		
		Articles	**Sentences**	**Tokens**
TrainFT	free text	946	14,987	203,621
DevQ	questions	n.a.	5,452	55,201
EvalFT	free text	231	3,684	46,435
EvalQ	questions	n.a.	500	3,758

Table 1: Corpora sources.

2.2. Tools

All the experiments related with the automatic identification and classification of named entities were conducted using E-txt2db (Simões, 2009), a framework for specifying and executing Information Extraction tasks. This tool offers a simple and straightforward Java API to an engine responsible for the creation, execution and evaluation of language models. Specifically, it allows the creation of models based on several different techniques, from regular expressions or dictionary-based, to Hidden Markov Models (HMM), Conditional Random Fields (CRF) or Support Vector Machines (SVM). E-txt2db bases its implementation of the HMM technique from the Lingpipe[3] framework. When it comes to E-txt2db implementation of the CRF and the SVM techniques, both were adapted from the Minorthird (Cohen, 2004) framework.

---

[2]For simplicity reasons, throughout this paper we use the following notation when referring to corpora: "<purpose><type>", where <purpose> is **Train**ing, **Dev**elopment or **Eval**uation, and <type> is free text (**FT**) or questions (**Q**).

[3]http://alias-i.com/lingpipe/

The general functioning of E-txt2db goes as follows: it takes as training corpus a text file with XML tags classifying segments. In the NER task, segments correspond to named entities, and tags to categories. A model is created based on one of the pre-implemented techniques. After the application of the model to a text, its segments are marked. E-txt2db allows to evaluate the model against a testing corpus, through recall, precision and the F-measure.

In this work, we utilized as learning features the default provided by the E-txt2db framework: the lower-case version of the segment; lexical properties of the segment; the length of the segment; features for tokens in a window of dimension 3 to either side of the segment; and features for the first and last tokens of the segment. Indeed it was not our main goal to improve the precision of the recognition task, rather than to test the impact of the training corpora in the results.

3. Annotation of Questions

The main goal of this work is to annotate a corpus of natural language questions with their named entities.

Broadly used by the Question-Answering community, we annotated the question corpus from Li&Roth. For each question, its named entities are marked between two equal XML tags, corresponding to a single category. Tags can be PER, which stands for the category PERSON, LOC standing for LOCATION and ORG referring to category ORGANIZATION. Only one category is attributed to each named entity and there are no nested named entities. As previously mentioned, the result of the annotation is freely available for research purposes.

3.1. Annotation Directives

The directives of CoNLL2003 were considered in the manual annotation of questions, and further extended to deal with our train questions:

- Named entities are labeled when their category is explicit, either in the context where it occurs, within the entity or from world knowledge. In case of high ambiguity, they are not labeled. For instance, in *What is the abbreviated expression for the <ORG>National Bureau of Investigation</ORG> ?*, there is an indication ("bureau") that the entity is an ORGANIZATION;

- A named entity will only be labeled LOCATION when there is a clear evidence that it refers to a concrete geographical place. Notice the difference between: *What is the largest city on the <LOC>Great Lakes</LOC>?* and *How many Great Lakes are there?*: in the former, the Great Lakes refer to a specific entity (a region), while in the latter they better refer to a group of entities, similar to a question like *How many oceans are there?* This applies both to entities in the earth, and in the outer space: *How can I find out my Moon sign ?* and *How many astronauts have been on the <LOC>moon</LOC> ?*;

- When a named entity, commonly labeled as LOCATION, is related with a verb associated with human beings, it is labeled as PERSON. For example: *What*

does the <PER>*Statue of Liberty*</PER> *wear on her feet?*;

- Only unambiguous wrongly spelled named entities are labeled, like *Where is the* <LOC>*Lourve*</LOC> *?* or *How tall is* <LOC>*kilamanjaro*</LOC> *?*;

- Fictional characters, locations or organizations will be tagged as PERSON, LOCATION or ORGANIZATION unless it is explicit that they refer to books, tv series, movies. Contrast *Where does* <PER>*Barney Rubble*</PER> *go to work after he drops* <PER>*Fred*</PER> *off in the " Flintstones " cartoon series ?* and *What city did the* <PER>*Flintstones*</PER> *live in ?*: on the former, the "Flintstones" refer unambiguously to the cartoon series; on the latter, this decision is debatable, as it can also refer to the "Flintstones" as the fictional family staring on the series with the same name. Thus, in this case, and similar others, we opt to label the name entity with category PERSON.

- Given that entities are denoted and identified by their name, if the question explicitly refers to the name for which the entity is known, instead of the entity itself, the name is labeled: *What is the abbreviation of the company name '* <ORG>*General Motors*</ORG>*' ?*. Note that this does not imply annotating the question: *What is the origin of the name Katie ?*, in which "Katie" is not an entity PERSON;

- The directive of CoNLL2003, which states that "titles such as Mr. and role names such as President are not considered part of a person name" was followed. The directive does not apply, however, when the entity explicitly refers to someone's nickname. Contrast: *What did Mr.* <PER>*Magoo*</PER> *flog on TV for* <ORG>*General Electric*</ORG> *?* and *What comedienne calls her sister-in-law* <PER>*Captain Bligh*</PER> *and her mother-in-law* <PER>*Moby Dick*</PER>. Like in the previous item, the name attributed to a person identifies the entity, regardless of it being the given name, a surname or a nickname. And in "Captain Bligh", the nickname includes the role Captain;

- Hollywood and Broadway are industries and, therefore, are not labeled as ORGANIZATION. When it is explicit that they refer to respective locations, they are labeled as LOCATION. For instance, *What book was* <PER>*F. Scott Fitzgerald*</PER> *working on when he died in* <LOC>*Hollywood*</LOC> *in 194 ?*

- Acronyms in *What is ACRONYM?*-like questions are not labeled, even if they refer to potentially unambiguous entities: *What is FBI?*. Tagging of acronyms occurs when the context disambiguates its category: *What year was the* <ORG>*NACCP*</ORG> *founded?*. With this, we avoid limiting or misguiding further interpretation tasks on the question: for instance, labeling *What is* <ORG>*FBI*</ORG>*?* can

imply and restrict the answer to being the organization Federal Bureau of Investigation, when its motto: Fidelity, Bravery, Integrity, would be also applicable.

- Named entities contained in citations, lines of poems or songs, ..., are not labeled.

Finally, and given the existence of a category MISCELLANEOUS in the CoNLL 2003 guidelines covering events, languages, wars..., and "words of which one part is a location, organization, miscellaneous, or person", complex noun phrases (like "battle of Waterloo", "Valentine 's Day" or "the Andy Griffith show") are not labeled.

We consider that a set of annotated questions is a valuable resource to be used by QA systems, among others, particularly for the interpretation of the input question. Being so, our decisions on what and how to label were mainly directed to that purpose. Namely, we avoided making assumptions on highly ambiguous entities that can compromise and eventually misguide the subsequent tasks to find an answer, after the question has been interpreted.

4. Experiments

This work allowed us to study the impact of the training parameters on the process of automatically identifying and classifying named entities in questions, namely whether there is a compromise to be attained between the size of free text data and the amount of questions to be used.

4.1. Strategy

Our strategy starts by using free text as the unique training corpus for building the language model. We gather part of the development corpus (questions) and use the generated model to automatically annotate it with the named entities. Afterwards, the resulting classified corpus is manually corrected. This corresponds to one annotation cycle. The classified and corrected corpus is then added to the training corpus that will again generate a new model. The process continues until the development corpus is fully annotated.

The precision and recall of the annotation is measured, at each annotation cycle, by making use of the evaluation corpus, which was previously labeled with the categories: PERSON, LOCATION and ORGANIZATION. This corpus acts as a golden collection to make a standard comparison between cycles whenever training parameters are modified. Results and statistical reports on the procedure rely on the evaluation made to the annotation on this corpus.

Figure 1 depicts the data flow diagram of the annotation strategy.

4.2. Measures

As previously referred, in order to evaluate the recognition of named entities in questions, we used *Recall*, *Precision* and the *F-measure*.

Precision is the ratio between the amount of entities correctly labeled and all entities labeled:

$$Precision = \frac{C}{C + I + O}$$

in which, C represents the number of entities correctly labeled, I represents the number of entities incorrectly labeled
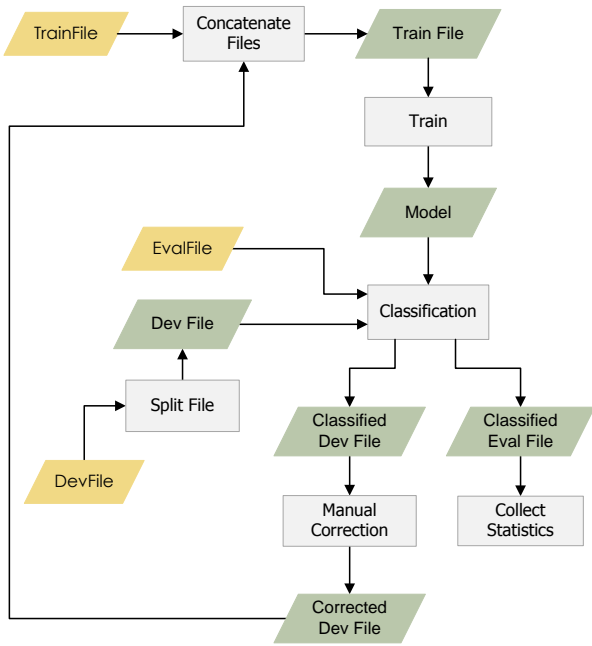
Figure 1: Data flow diagram of the annotation strategy.

and O represents the number of entities that have been labeled and should not have been.

Recall measures the amount of relevant labeled entities:

$$Recall = \frac{C}{T}$$

in which, C represents the number of correctly labeled entities while T represents the total number of entities that should be labeled.

The F-measure is adopted to measure the general performance of the annotation task, balancing the values of recall and precision:

$$F\text{-}measure = \frac{(1+\alpha) \times P \times R}{\alpha \times P + R}$$

in which, R represents the recall, P represents the precision, α allows to define the relative weight of recall and precision. As usually occurs, we took $\alpha = 1$, giving the same weight to recall and precision.

4.3. Results

A first evaluation assessed the performance of different machine learning techniques for the NER task in questions. These represent our baseline results (Table 2), collected with default conditions: only TrainFT is utilized as training corpus.

For all techniques, results show a significantly worse performance when evaluating with questions than with free text, while training uniquely with free text.

Since SVM held the best results, we focused uniquely on this technique to assess if results could be further improved. Being so, different models were generated with different training parameters. Modifications were made on the amount of free text and questions the system should

learn[4]. Results of the evaluation (using uniquely EvalQ) are presented in Table 3.

Train		Results		
TrainFT (# art.)	**Cor.DevQ** (# quest.)	**Recall** (%)	**Precision** (%)	**F-meas.** (%)
946	**0**	**64.96**	**67.56**	**66.23**
946	500	72.65	74.89	73.75
473	500	68.80	74.88	71.71
0	500	53.42	73.10	61.73
946	1000	72.65	79.81	76.06
473	1000	68.80	77.40	72.85
0	1000	60.26	79.66	68.61
946	1000	72.65	81.34	76.75
473	1500	70.09	78.85	74.21
0	1500	62.39	79.78	70.02
946	2000	72.22	79.34	75.62
473	2000	71.37	79.15	75.06
0	2000	67.95	84.13	75.18
946	3000	74.36	81.69	77.85
473	3000	73.50	81.13	77.13
0	3000	67.95	85.00	75.53
946	4000	77.35	84.19	80.62
473	4000	77.35	85.38	81.17
0	4000	74.79	88.38	81.02
946	5000	80.77	85.91	83.26
473	5000	78.63	85.19	81.78
0	5000	76.07	88.12	81.65
946	5500	80.77	85.52	83.08
473	5500	79.06	85.25	82.04
0	5500	74.79	85.37	79.73

Table 3: NER results under different training conditions (SVM as learning technique).

Compared with the baseline, results tend to improve with the inclusion of labeled questions, regardless of the amount free text corpus used for training. Recall increases more significantly when the number of questions is higher than 2000. Precision, however, suffers visible changes regardless of the used questions. Also, precision results are always higher than those of the baseline, which only happens with the recall when the number of questions in the training corpus is 4 times the size of the test corpus. When there is not enough data in the training corpus (TrainFT = 0 and Cor.DevQ $\leq$ 1500), the model marks less entities, but they are accurate.

Furthermore, and contrary to the tendency, when the number of questions included on the training corpus is maximum (5,500), precision and recall results tend to slightly drop or maintain. The exception is when we use TrainFT = 473, and both results augment. Here, probably, increasing the training set with questions would lead to better results; however, on one hand, the increase was small, and on the other, its corresponding F-measure is still around 1.2 less than the best results achieved: 83.26, with TrainFT = 946

---

[4]We used the first half of the articles available in TrainFT, and not a random set. Likewise for the questions: the first 250, 500, 750 and so forth... questions were used from the total of 5500 to be annotated.

| Technique | Train | | Eval | Results | | |
	TrainFT (# articles)	Cor.DevQ (# questions)		Recall (%)	Precision (%)	F-measure
HMM	946 (100%)	0	EvalFT	73.59	76.28	74.91
	946 (100%)	0	EvalQ	53.85	21.36	30.58
SVM	946 (100%)	0	EvalFT	75.60	80.43	77.94
	946 (100%)	0	EvalQ	64.96	67.56	66.23
CRF	946 (100%)	0	EvalFT	73.49	79.59	76.42
	946 (100%)	0	EvalQ	43.16	48.56	45.70

Table 2: Baseline NER results.

and Cor.DevQ = 5000. Graphs in Figure 2 depict the evolution of the precision and recall with the addition of questions to the training corpus, given the different quantities of free text we used in the experiments.
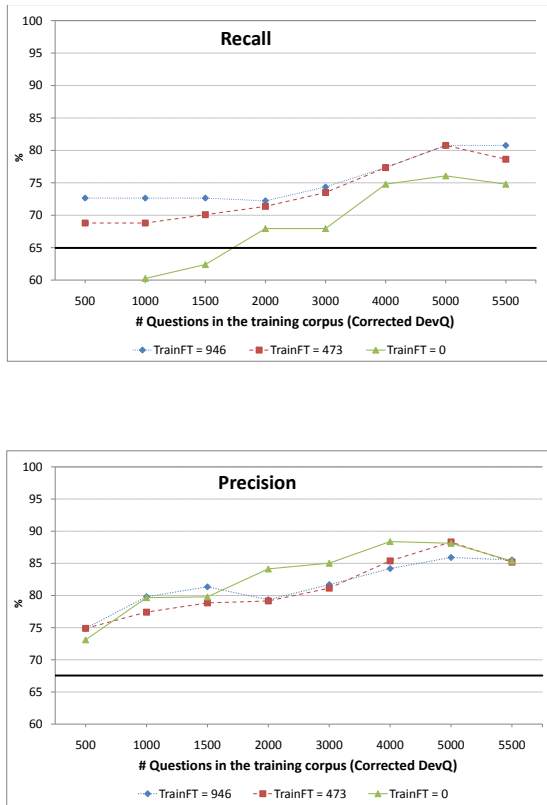

Figure 2: Evolution of recall and precision with the addition of questions to the training corpus, given different amounts of free text. Baseline values are given by the dark horizontal lines.

Table 4 presents detailed results for the different categories. Comparing again with the baseline (first row in bold), the recall on the identification of PERSONS and LOCATIONS increases with questions on the training corpus. When it comes to ORGANIZATIONS, and except when no free text is used in the training corpus, recall values are not better; there is actually a degradation with the addition of questions, specially with half of the free text corpus. We believe that this happens because the number of entities belong-

ing to this category is higher on the news articles than on our open domain questions; the organizations present in the question corpus are not representative and not beneficial to the final result.

Precision in PERSONS and LOCATIONS tend to augment with the use of questions in the training corpus, meaning that, besides of identifying more entities of these categories, the technique is able to correctly classify more of them. Precision in ORGANIZATIONS are worse than in the other categories; nevertheless, if the amount of free text diminishes, results improve and largely surpass the baseline with the increase of the number of questions in the training corpus. Precision values are nearly the double of the baseline for parameters TrainFT = 946 and Cor.DevQ = 1000.

Results achieved by this work suggested that the inclusion of questions in the training corpus benefit the annotation of named entities in questions. Indeed, to build a suitable training corpus (which can generate a language model that attains an F-measure greater than 75, when evaluated against the test corpus) we needed merely to add 1000 annotated questions to the full number of articles. When the number of questions included was superior to 2000, the number of articles was not relevant to the suitableness of the training corpus (as the F-measure was always greater than 75). The free text data was decisive, however, to obtain the best evaluation results.

5. A Note on NER on Questions

Named entities are used in almost all Natural Language Processing tasks. In Question-Answering, specifically, they play a proeminent role, being often employed to filter out candidate answers, given a taxonomy of semantic classes that represent the type of information expected to be in the answer, and a function that maps each of them to the respective named entity category (Mollá et al., 2006). For instance, to the question *Who is the President of the U.S.A?*, that asks for the name of a person, the system only considers named entities with category PER as answers. Named entities are, thus, usually searched for in texts that can contain the answer.

The identification and classification of named entities in questions is, however, far from being fully explored. And its importance is considerably. Note that, in the limit, the category of a named entity represents itself the answer to a question. For example, the question *What is "Nine Inch Nails"?* seeks for the definition of *Nine Inch Nails* (*i.e.* a band, a group or, broadly speaking and making the parallel

Train		Results					
TrainFT	**Cor.DevQ**	**Recall**			**Precision**		
(# art.)	**(# quest.)**	PER	LOC	ORG	PER	LOC	ORG
946	**0**	**83.72**	**62.43**	**44.44**	**69.23**	**85.04**	**17.39**
946	500	93.02	70.52	44.44	70.18	89.71	23.53
473	500	88.37	68.21	27.78	67.86	86.76	21.74
0	500	72.09	54.34	0.00	56.36	83.93	0.00
946	1000	88.37	72.25	38.89	73.08	89.93	31.82
473	1000	81.40	69.36	33.33	68.63	85.71	35.29
0	1000	76.74	61.85	5.56	68.75	86.29	20.00
946	1500	88.37	72.25	38.89	80.85	87.41	36.84
473	1500	86.05	69.94	33.33	72.55	85.21	40.00
0	1500	74.42	65.32	5.56	71.11	84.33	25.00
946	2000	90.70	71.10	38.89	76.47	87.23	33.33
473	2000	86.05	72.25	27.78	74.00	85.03	35.71
0	2000	76.74	71.68	11.11	78.57	86.71	50.00
946	3000	95.35	73.41	33.33	82.00	87.59	33.33
473	3000	93.02	73.41	27.78	75.47	88.19	33.33
0	3000	88.37	68.21	16.67	82.61	87.41	50.00
946	4000	93.02	78.03	33.33	80.00	90.60	37.50
473	4000	93.02	78.61	27.78	80.00	90.67	41.67
0	4000	90.70	76.88	16.67	90.70	89.26	50.00
946	5000	93.02	82.66	33.33	90.91	89.94	35.29
473	5000	93.02	83.24	27.78	90.91	90.00	45.45
0	5000	93.02	78.61	11.11	97.56	86.62	50.00
946	5500	93.02	83.24	27.78	86.96	89.44	35.71
473	5500	93.02	80.35	27.78	83.33	89.10	41.67
0	5500	93.02	76.30	16.67	88.89	86.27	42.86

Table 4: NER results under different training conditions, using SVM as learning technique, detailed for the categories PERSON, LOCATION and ORGANIZATION.

with this work, an organization) which is itself the category of the named entity.

In this context, an accurate recognition can aid the task of deciding which semantic class a question is expecting (known as question classification). For instance, *Who is* PER*?* asks for the definition of a person, and *Where is* LOC*?* asks for the name of a geographical place, regardless of what PER and LOC are, respectively. On the other hand, it can make the difference between proceeding to search for the answer or requesting the user to confirm his question[5], if the question is *To which country belong the islands of Madeira?* or *To which country belong the islands of Langerhans?*, keeping in mind that islands of Madeira are a location, and the islands of Langerhans are clusters of cells in the pancreas of most vertebrates.

Similarly to named entities recognition in free text, the task applied to questions presents several challenges, mainly due to ambiguity: multiples categories that can be applied for the same entity. The process frequently obeys to annotation guidelines, that helps to minimize this problem, however not to solve it. For instance, we adopted CoNLL2003 directives, which state that organizations are to be tagged as ORGANIZATION "unless designated only by country names, which are tagged as LOCATION". In Question-Answering this decision can impact the performance of the question classifier. Consider the questions:

Who defeated France in the last final? and *Who defeated Manchester United in the last final?*. It is unarguable that both expect a team name, however on the first it is probably encoded as LOCATION and on the second as ORGANIZATION. This situation is made complicated in the question *Who defeated La Machina Naranja in the last final?*. The nickname *La Machina Naranja* is annotated as ORGANIZATION, but the answer is probably a country name, i.e. a LOCATION. Further experiments on this issue should, thus, be conducted.

6. Conclusions and Future Work

This work, focusing the task of identifying and classifying named entities in questions, resulted in a set of 5,500 labeled questions with categories PERSON, LOCATION and ORGANIZATION, which can be employed by machine learning-based NER systems. In parallel, we studied the impact of changing the training parameters (the amount of free text and questions), and presented a broad study on that topic. The hypothesis was that the introduction of labeled questions in the training corpus of a machine learning-based NER system would benefit NER in questions, which was largely suggested by the achieved results. As future work, we intend to test the combination of several machine learning techniques in order to obtain better results, and to explore the transfer of named entities between languages. Finally, we aim at using the created resource

---

[5]Among other strategies to solve this situation.

in the question analysis module of an in-house Question-Answering system, and study how it can improve its performance.

7. Acknowledgments

8. References

William W. Cohen. 2004. Minorthird: Methods for identifying names and ontological relations in text using heuristics for inducing regularities from data. http://minorthird.sourceforge.net.

Jiafeng Guo, Gu Xu, Xueqi Cheng, and Hang Li. 2009. Named entity recognition in query. In *SIGIR '09: Proc. 32nd Int ACM SIGIR Conf. Research Development in Information Retrieval*. ACM.

David D. Lewis, Yiming Yang, Tony G. Rose, and Fan Li. 2004. RCV1: A New Benchmark Collection for Text Categorization Research. *Journal of Machine Learning Research*, 5:361–397.

Xin Li and Dan Roth. 2002. Learning question classifiers. In *Proc. 19th Int. Conf. Computational Linguistics*. ACL.

Diego Mollá, Menno van Zaanen, and Daniel Smith. 2006. Named Entity Recognition for Question Answering. In *Proc. ALTW 2006*.

Gonçalo Fernandes Simões. 2009. e-txt2db: Giving structure to unstrutured data. Master's thesis, Inst. Sup. Técnico, Technical University Lisbon, Nov.

Erik F. Tjong Kim Sang and Fien De Meulder. 2003. Introduction to the CoNLL-2003 Shared Task: Language-Independent Named Entity Recognition. In *Proc. CoNLL-2003*.

Antonio Toral, Fernando Llopis, Rafael Muñoz, and Elisa Noguera. 2005. Reducing Question Answering input data using Named Entity Recognition. In *Proc. 8th International Conference on Text, Speech & Dialogue*.