Named and specific entity detection in varied data: the Quaero named entity baseline evaluation

Olivier Galibert, Ludovic Quintard, Sophie Rosset, Pierre Zweigenbaum, Claire Nédellec, Sophie Aubin, Laurent Gillard, Jean Pierre Raysz, Delphine Pois, Xavier Tannier, et al.

Named and specific entity detection in varied data: The Quæro Named Entity baseline evaluation

Olivier Galibert[1], Ludovic Quintard[1], Sophie Rosset[2], Pierre Zweigenbaum[2], Claire Nédellec[3], Sophie Aubin[3], Laurent Gillard[3], Jean-Pierre Raysz[5], Delphine Pois[5], Xavier Tannier[24], Louise Deléger[2], Dominique Laurent[6]

[1] LNE, Trappes, France [2] LIMSI-CNRS, Orsay, France
[3] INRA, Jouy-en-Josas, France [4] Université Paris-Sud 11, Orsay, France
[5] Jouve, Mayenne, France [6] Synapse Développement, Toulouse, France

Abstract

The Quæro program that promotes research and industrial innovation on technologies for automatic analysis and classification of multimedia and multilingual documents. Within its context a set of evaluations of Named Entity recognition systems was held in 2009. Four tasks were defined. The first two concerned traditional named entities in French broadcast news for one (a rerun of ESTER 2) and of OCR-ed old newspapers for the other. The third was a gene and protein name extraction in medical abstracts. The last one was the detection of references in patents. Four different partners participated, giving a total of 16 systems. We provide a synthetic descriptions of all of them classifying them by the main approaches chosen (resource-based, rules-based or statistical), without forgetting the fact that any modern system is at some point hybrid. The metric (the relatively standard Slot Error Rate) and the results are also presented and discussed. Finally, a process is ongoing with preliminary acceptance of the partners to ensure the availability for the community of all the corpora used with the exception of the non-Quæro produced ESTER 2 one.

1. Introduction

Named Entity Detection is a problem that has been studied since the eighties, starting with the MUC conferences in 1987 (see (Grishman and Sundheim, 1996) for an history of these). The notion has been extended to cover any type of mono- or multi-word expression which designates an object or concept of the real world that belongs to a class of potential interest for a given application. Given a series of entity definitions and a corpus of natural language, systems try to extract and categorize all the relevant occurring entities. System output can then be used to feed further systems such as Information Retrieval, Question-Answering, Distillation, Terminology studies, etc.

Quæro[1] is a program that promotes research and industrial innovation on technologies for automatic analysis and classification of multimedia and multilingual documents. One of the requirements of the Quæro project is to organize periodic internal evaluations of the technologies developed by the partners. Within this framework a Named/Specific Entity Detection evaluation was organized in 2009, and will be reconducted in a slightly modified and expanded form in the following years.

2. The Tasks

The general named entity recognition problem consists of two parts: *detecting* words or word sequences corresponding to interesting entities and *categorizing* them into predefined types. Defining a task requires to select a corpus of text to search and a set of target entity types.

Within the framework of the Quæro evaluations, we decided to tackle three different domains. The first is the detection of traditional named entities in news documents. The originality comes from the nature of the documents. Two types of documents have been taken into account:

- Speech transcriptions (French broadcast news);
- Scanned and OCR-ed old newspapers (*Le Temps*, from 1920, in French).

Detecting named entities in speech transcripts is not new. This task was already addressed in 2009 within the ESTER 2 evaluation (Galliano et al., 2009), and we simply reused its dataset. NE detection on scanned newspapers is new for French, and the difficulty is high, especially due to OCR errors. We had to simplify named entity categorization to three types, person, location and organisation, to make it tractable. There seems to have been an experiment in ACE 2002 on OCR-ed English texts, but almost no information is available. Figure 1 shows a sample of the OCR corpus, while Figure 2 shows a sample from ESTER 2.

The second task was related to the biomedical domain. Gene name recognition in the biomedical literature is a critical first step in biomedical information extraction (Fukuda et al., 1998). For our first evaluation we decided to detect gene and protein names, without distinguishing them, in English-language biomedical abstracts. Compared to other previous challenges (NLPBA (Kim et al., 2004), BioCreative (Krallinger et al., 2008)), the domain is bacteriology, which has important applications in Health and AgroFood. The focus is on the recognition of the named entities without neighboring words describing properties or types of the named entities (Nédellec et al., 2006), as close as possible to entries from biomedical nomenclatures. The names to be annotated are those that are liable to be recorded as an entry or synonym in the GenBank or SwissProt nomenclatures. Figure 3 shows a sample of the corpus.

Finally, the last task was related to intellectual property. One of the tasks a prior art researcher has to do is to extract a bibliography from the text of the patent, which is currently done by hand. To help in this process we defined a task where systems had to detect citations to other

[1] http://www.quaero.org

patents and to general literature in English-language patent text. Figure 4 shows a sample of these annotations.

3. Data and annotations

The data sets used for the different evaluation tasks were supplied by four Quæro partners: INRA, Jouve, DGA and BNF. Specifically, the annotated biomedical paper abstracts were prepared by INRA (Nédellec, 2009). The annotated patents were prepared by Jouve. The annotated broadcast news transcriptions and annotations were provided by DGA within the ESTER 2 project. The images and the converted text (xml file) for old newspapers were provided by BNF and then annotated by Jouve. All resulting corpora were split into training and development data on the one side and evaluation data on the other.

It is important to note that patent annotation was initially produced for human use, making the presence of an annotation more important than its exact boundaries. That has had an impact at the evaluation stage.

In the case of the ESTER 2 corpus, the entity types were (see Table 1): location, organization, person, position, product, quantity, time and others. In the Oldpress corpus, the entities were: location, organization and person. In the biomedical corpus: gene (no separation between protein and gene). In the patent recognition corpus: patent and non-patent literature.

ESTER 2	Oldpress	Gene	Patent
loc	loc	gene	patcit
org	org		nplcit
pers	pers		
fonct			
prod			
amount			
time			
unk			

Table 1: Entity types names for each task. loc=location, gene=gene/protein, org=organization, patcit=patent citation, nplcit=non patent literature citation, pers=person, fonct=function, prod=production, unk=unknown

4. Metrics

Scoring was performed using SER (Slot Error Rate), recall, precision and F-measure metrics, where type-only or boundary-only errors cost half a point, and complete errors, insertions and deletions cost one full point. The score is then divided by the number of entities present in the reference.

Given the following definitions:

- R: Reference entity count
- H: Hypothesis entity count
- C: Number of correct entities (aka. True Positives)
- T: Number of entities with correct boundaries but incorrect type
- F: Number of entities with correct type but incorrect boundaries
- TF: Number of entities with incorrect type and boundaries
- I: Number of entities inserted (aka. False Positives)
- D: Number of entities forgotten (deletions, aka. False Negatives)

Slot Error Rate (SER) is computed as:

$$SER = \frac{D + I + TF + 0.5 \times (T + F)}{R}$$

The other standard measurements were calculated using the following formulas:

$$Precision = \frac{C + 0.5 \times (T + F)}{H} \quad (1)$$

$$Recall = \frac{C + 0.5 \times (T + F)}{R} \quad (2)$$

$$F\text{-}measure = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (3)$$

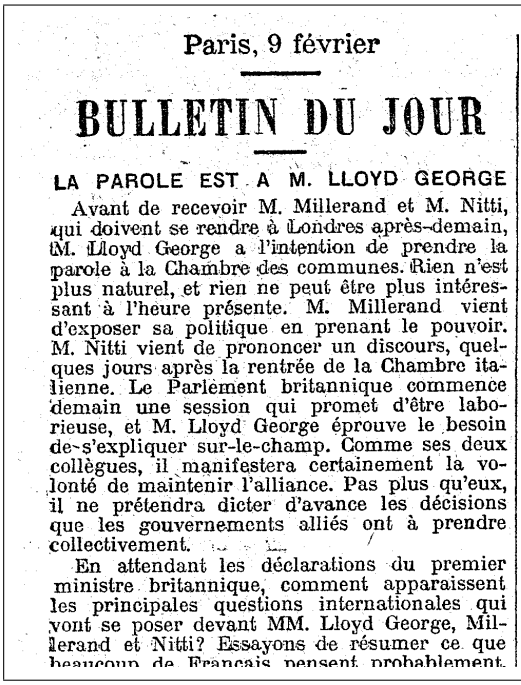
$$= 2 \times \frac{C + 0.5 \times (T + F)}{R + H} \quad (4)$$

For SER, a lower value is better. For all the others, a higher value is better. They are all traditionally given as percentages. It is interesting to note that the "half-point" evaluation of frontier errors may not be pertinent for the gene name extraction task where an incorrect segmentation can easily yield the name of a different object than the referenced object. In contrast, the presence or absence of a determiner does not usually change the meaning of an organization name. As a result, full-point errors are more appropriate for gene name extraction, whereas half-point errors may be better suited for the more traditional entity extraction tasks.

5. Participants and Systems

Four Quæro partners participated in the evaluation: INRA, Jouve, LIMSI and Synapse Développement. Each participant proposed a system for each of the subtasks, for a total of 16 systems. Those systems, as all modern systems, tend to be a mix of linguistic resources, rules and stocastic approaches. However, it is still possible to classify them on what the authors consider central to their approach.

The systems from Synapse Développement are archetypal of the resource-heavy approach. Initially working in the grammatical correction field and since then diversifying into other fields such as question-answering, they developed a huge amount of resources describing the French and now the English language, including such things as a WordNet-equivalent, morphosyntaxic derivation tables, syntactic and semantic compatibility information, tables of expressions, etc. Using undisclosed algorithms leveraging these resources they can produce a complete, in-depth syntactic and semantic analysis of the text. Extracting named entities is then recognizing the appropriate semantic types or contexts. Their approach works extremely well for clean



Paris, 9 février

BULLETIN DU JOUR

LA PAROLE EST A M. LLOYD GEORGE

Avant de recevoir M. Millerand et M. Nitti, qui doivent se rendre à Londres après-demain, M. Lloyd George a l'intention de prendre la parole à la Chambre des communes. Rien n'est plus naturel, et rien ne peut être plus intéressant à l'heure présente. M. Millerand vient d'exposer sa politique en prenant le pouvoir. M. Nitti vient de prononcer un discours, quelques jours après la rentrée de la Chambre italienne. Le Parlement britannique commence demain une session qui promet d'être laborieuse, et M. Lloyd George éprouve le besoin de s'expliquer sur-le-champ. Comme ses deux collègues, il manifestera certainement la volonté de maintenir l'alliance. Pas plus qu'eux, il ne prétendra dicter d'avance les décisions que les gouvernements alliés ont à prendre collectivement.

En attendant les déclarations du premier ministre britannique, comment apparaissent les principales questions internationales qui vont se poser devant MM. Lloyd George, Millerand et Nitti? Essayons de résumer ce que beaucoup de Français pensent probablement.

<TextBlock HEIGHT="758" HPOS="99" ID="PAG_1_TB000018" STYLEREFS="TXT_5" VPOS="1351" WIDTH="899"> LA PAROLE EST A M. LLOYD (3E0RGE Avant de recevoir <person> M. Millerand </person> et <person> M. Nitti </person>, qui doivent se 'rendre à i <location> Londres </location> + après-demain, i t <person> M. iLiloyd George </person> a l'intention de prendre la l parole à la <organisation> Chambre des communes </organisation> . Rien n'est plus naturel, et rien ne peut être plus intéres- sant'à l'heure présente. <person> M. Millerand </person> vient d'exposer sa politique en prenant le pouvoir. <person> M. Nitti </person> vient de prononcer un discours, quel- ques jours après la rentrée de la <organisation> Chambre ita- lienne </organisation> . Le <organisation> Parlement britannique </organisation> commence demain une session qui promet d'être labo- rieuse, et <person> M. Lloyd George </person> éprouve le besoin de-s'expliquer sur-le-champ. Comme ses deux collègues, il manifestera certainement là vo- .lonté de maintenir l'alliance. Pas plus qu'eux, il ne prétendra dicter d'avance les décisions que les gouvernements alliés ont à prendre collectivement. i. I </TextBlock>

Figure 1: Example of scanned then transcribed and annotated old newspaper (oldpress corpus)

20070710_1900_1920_inter 1 Mickaël_Thébault 113.379 118.384 <o,f3,male> et si c'était un Français , socialiste de surcroît qui prenait la tête du [org.non-profit FMI] ?
20070710_1900_1920_inter 1 Mickaël_Thébault 118.384 123.559 <o,f3,male> [pers.hum Dominique Strauss-Kahn] semble en tout cas très bien parti pour décrocher la direction du [org.non-profit Fonds monétaire international] .
20070710_1900_1920_inter 1 Mickaël_Thébault 123.559 131.365 <o,f3,male> il y avait déjà le soutien du [fonc.pol chef de l'État] , l' [org.pol Union européenne] lui dit aussi oui . concert de louanges avec de très rares voix discordantes .
20070710_1900_1920_inter 1 Mickaël_Thébault 131.365 138.684 <o,f3,male> [pers.hum Dominique de Villepin] , lui , devant les juges . c'est [time.date.rel pour la fin du mois] . affaire [org.com Clearstream] et cette fois la mise en examen semble proche .

Figure 2: Example of transcribed and annotated broadcast news speech (ESTER 2 corpus)

text as are the ESTER 2 manual transcriptions and reasonably well, given their lack of specific resources for the medical domain, on the gene task. The results degrade on lower quality text such as ASR or OCR outputs, because of the importance of each individual word. Still, these resources make it possible to use smart error-correction methodologies, reducing the expected damage.

Given the limited amount of annotated data available in general for named entities related tasks, rules-based approaches are still very popular by leveraging the human capabilities of generalization. The INRA systems (Bossy et al., 2009) use resources under the form of dictionaries of relevant names, such as gene and protein names for the medical domain, people, places, organisations names for the general named entities and country codes for the citations. The rules applied to gene name recognition have been automatically learned. In addition to the belonging to dictionaries, a number of attributes have been defined for describing the candidate entities. They include the description of the candidate context and morphology. For instance, regular expressions-based rules computes certain useful words morpheme such as *starting with desoxy*. A machine learning setup using Weka's implementation of the Induction of Decision Tree algorithm then computes a classifier from the description of the training exemples. The named entity recognition rules applied to the other tasks take the form of automatas written under Unitex (Paumier, 2008).

Two of the LIMSI systems also belong to the rules-based category. Their oldpress system is a pure rules system implemented using their internal *wmatch* (Galibert, 2009) engine. That engine allows the creation of rules using regular expressions on words or characters within an incremental parsing methodology. They tried to build the rules on the words with the highest a-priori probability of being recognized correctly, and also included series of possible variants for the important words (titles, countries, important names of that time, etc). Their ESTER 2 system is an experiment in system adaptation. They started from their internal analyzer for QA systems (Rosset et al., 2009), which includes among other things a named entities extractor for their specific definition of named entities. They then created mapping rules between their output and what was expected by aligning their analysis of the training data with the associated reference.

The last approach used was machine learning. Conditional Random Fields are very popular nowadays for, among others, chunk extraction and typing tasks of which named entities is one. The systems by Jouve are all built around a CRF approach, specifically the Stanford-NER (Finkel et al., 2005) implementation. They used a number of features including POS-tagging and lookups from gazetteers coming in particular from DBpedia and BNF's authority files. In addition, for the ESTER 2 and oldpress tasks an alterna-

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE document SYSTEM "quaero-gene-challenge-2009.dtd">
<document pmid="9826499">
<title>
The kinase activity of the antisigma factor <gene>SpoIIAB</gene> is required for activation as well as inhibition of transcription factor <gene>sigmaF</gene> during sporulation in Bacillus subtilis.
</title>
<abstract>
The activity of the developmental transcription factor <gene>sigmaF</gene> in the spore-forming bacterium Bacillus subtilis is controlled by <gene>SpoIIAB</gene>, which sequesters <gene>sigmaF</gene> in an inactive complex. [...]
```

Figure 3: Excerpt of an annotated medical abstract (biomedical corpus)

```
<p id="p0014" num="0014">"<nplcit>Noncatalytic electrodes for solid-electrolyte oxygen sensors"; Haaland D M; April 1980; Journal of the Electrochemical Society, VOL 127, NR 4, pages 796 - 804</nplcit> discloses, in the context of catalytically active electrodes which can perturb the measurement of oxygen in non-equilibrium mixtures of oxygen and combustible gases such as methane, the technique of poisoning the Pt electrodes with silver, lead (pages 798-799).</p>
<p id="p0008" num="0008">Although usually not as efficient as the general space-filling curves disclosed in the present invention, other well-known geometries such as meandering and zigzag curves can also be used in a novel configuration according to the spirit and scope of the present invention. Some descriptions of using zigzag or meandering curves in antennas can be found for instance in patent publication <patcit>WO96/27219</patcit>, but it should be noticed that in the prior-art such geometries were used mainly in the design of the radiating element rather than in the design of the ground-plane as it is the purpose and basis of several embodiments in the present invention.</p>
```

Figure 4: Two paragraphs extracted from patents with annotated citations (patent corpus)

tive gazetteer-based *distant search* detection module was implemented. The results of both modules (CRF and distant search) were merged together using the confidence levels computed by each. The LIMSI gene system was also CRF-based, this time using the JNET (Hahn et al., 2008) implementation. The JNET models were retrained on the Quaero data, and its performance was enhanced by adding a preprocessing pass with tokenization and POS-tagging, and a post-processing pass fixing the results in some specific cases, like the presence of the words *gene* or *protein* in the extracted chunks which were supposed to be removed as per the evaluation guidelines.

Finally the LIMSI patents system (Galibert et al., 2010) was an example of a fully hybrid system. The non-patent literature citations were detected with a CRF approach using the CRF++ implementation (Kudoh, 2007) with as features the words and the POS as inferred by the TreeTagger(Schmid, 1994). On the other side, the patent citations were detected using a rule-based system detecting boundaries and important features followed by an algorithmic building of the chunks. Both outputs were then merged together. They report that the number of conflicting (overlapping) extractions was low enough (less that 1% of the entities) to make the conflict resolution methodology unimportant.

6. Evaluation

Table 2: Results for Gene tasks for all participants

	Gene			
	SER	Precision	Recall	F-measure
INRA	29,1	93,1	75,3	83,2
Jouve	27,4	93,8	77,3	84,7
LIMSI	26,3	88,6	80,4	84,3
Synapse	51,9	69,7	82,6	75,6

All these evaluations are relatively novel on these corpora, preventing us from giving state-of-the-art performance values. Raw results go from 26.3% to 51.9% SER for the gene task (Table 2, 65.3%-74.7% F-measure), 33.1% to 48.7% for the patents task (Table 3), 9.33% to 75.41% for the ESTER 2 task (Table 4) and 56.7% to 80.1% for the Oldpress task (Table 5).

The *gene* evaluation shows different tradeoffs when it

Table 3: Results for the Patents task for all participants

	Patents			
	SER	Precision	Recall	F-measure
INRA	44,9	72,3	65,5	68,8
Jouve	36,7	78,1	69,4	73,5
LIMSI	33,1	78,5	71,2	74,7
Synapse	48,7	63,7	67,0	65,3

Table 4: SER results for the ESTER 2 task. LIMSI and Synapse results come from their participation in the original ESTER 2 task, INRA and Jouve from their participation in the Quaero evaluation of same.

	SER			
	Ref	ASR1	ASR2	ASR3
INRA	43,93	57,62	89,46	90,86
Jouve	75,41	60,43	96,32	96,11
LIMSI	30,88	45,34	55,55	61,16
Synapse	9,93	44,86	60,67	66,22

Table 5: Results for the Oldpress task for all participants

	Oldpress			
	SER	Precision	Recall	F-measure
INRA	80,1	46,6	40,1	43,1
Jouve	61,0	73,0	42,6	53,8
LIMSI	56,7	69,4	50,6	58,5
Synapse	69,3	50,4	48,5	49,4

comes to precision vs. recall. The ratio of correctly detected entities is similar for all four systems (from 73.8% to 80.2%) and the frontier errors are not very numerous (from 1.7% to 9.5%). The real difference comes from insertions and deletions. INRA and Jouve went for high precision systems, trading a low insertion rate (4.3% and 4.7%) against a high deletion rate (23.3% and 21.9%). Synapse Développement went for high recall with 34.4% insertion but still 15.1% deletion. LIMSI struck a globally more efficient median with 6.7% insertion and 14.6% deletion.

For *patents*, the main problems for all systems was a frontier detection issue. The annotated reference data is heavily ambiguous in that area, sometimes including closing parentheses, brackets, or final periods and sometimes not, without any obvious underlying rule. The for-human-use origin of the corpus, current lack of a formal definition of the task, and the sheer size of the test data (around 20,000 annotations, with around 10,000 frontier errors for each system) makes it difficult to correct the references. The systems results are to be considered better than the values announced here in practice. If the task is kept as-is, we will need to study whether systematic corrections are possible (always integrate closing symbols in the entity if the opening symbol is there for instance).

For *ESTER* the highly linguistic-knowledge-based Synapse Développement system obtained impressive results on the clean manual transcriptions, but wasn't robust enough to keep them for the relatively low error rate automatic transcriptions (around 10% Word Error Rate for ASR1). It is interesting to note that the surprising results of Jouve, which are better with the automatic transcription than with the manual ones, are explained by a bug in their results post-processing pass which removed a large number of annotations from the system output. Once that bug corrected the error rate goes sharply down to 29.4%.

7. Data availability

The partners in the Quæro Named Entity subgroup have decided to make all the data and scoring tools associated with that evaluation available to the community at large. That concerns three of the tasks, named entities in OCR, in medical documents and citations in patents. The final packaging is not done nor the terms fully defined, but we expect the effective availability to happen somewhere in 2010.

The ESTER 2 data availability is not under the control of the Quæro project and should be discussed if necessary with the ESTER 2 organizers.

8. Conclusions & Future Work

We presented the 2009 Quæro Named Entity Detection evaluation. Three main entity definitions were used on four different corpora. The results vary widely depending on the task, and tend to show the different strategies system developers can adopt. One of the tasks even shows the difficulties posed by using a corpus initially created for human use for an evaluation.

We will need to study what can be done to make such a corpus more reliable for an evaluation. Systematic corrections may apply. In addition, the tasks will probably be widened in scope, especially in the types of entities to be detected.

Furthermore, an internal reflection is going on to determine on which terms it would be possible to open the next instances of the evaluation to participants external to the Quæro project.

9. Acknowledgement

This work has been partially financed by OSEO under the Quæro program. We want to thank the BNF for providing the old press corpus.

10. References

Robert Bossy, Sophie Aubin, Laurent Gillard, Frédéric Papazian, and Claire Nédellec. 2009. AlvisNER, a combined graph and ML-based approach to NER. Software, INRA, September.

Jenny Rose Finkel, Trond Grenager, and Christopher Manning. 2005. Incorporating non-local information into information extraction systems by gibbs sampling. In *Proceedings of the 43nd Annual Meeting of the Association for Computational Linguistics (ACL 2005)*, pages 363–370.

Ken Fukuda, Tatsuhiko Tsunoda, Akihisa Tamura, and Toshihisa Takagi. 1998. Toward information extraction: Identifying protein names from biological papers. In *Proceedings of the Pacific Symposium on Biocomputing (PSB '98)*, pages 705–716.

Olivier Galibert, Sophie Rosset, Xavier Tannier, and Fanny Grandry. 2010. Hybrid Citation Extraction from Patents. In *LREC'10*.

Olivier Galibert. 2009. *Approches et méthodologies pour la réponse automatique à des questions adaptées à un cadre interactif en domaine ouvert*. Ph.D. thesis, LIMSI-CNRS, June.

Sylvain Galliano, Guillaume Gravier, and Laura Chaubard. 2009. The Ester 2 evaluation campaign for the rich transcription of french radio broadcasts. In *Proceedings of Interspeech 2009*, pages 2583–2586, Brighton, England, September.

Ralph Grishman and Beth Sundheim. 1996. Message Understanding Conference - 6: A brief history. In *Proceedings of the 16th International Conference on Computational Linguistics (COLING)*, pages 466–471, Copenhagen, Denmark, August.

Udo Hahn, Ekaterina Buyko, Rico Landefeld, Matthias Mühlhausen, Michael Poprat, Katrin Tomanek, and Joachim Wermter. 2008. An overview of JCoRe, the Julie Lab UIMA Component Repository. In *Proceedings of the LREC'08 Workshop Towards Enhanced Interoperability for Large HLT Systems: UIMA for NLP*, pages 1–7.

Jin-Dong Kim, Tomoko Ohta, Yoshimasa Tsuruoka, Yuka Tateisi, and Nigel Collier. 2004. Introduction to the Bio-Entity Recognition Task at JNLPBA. In *Proceedings of the International Workshop on Natural Language Processing in Biomedicine and its Applications (JNLPBA-04)*, pages 70–75.

Martin Krallinger, Alexander Morgan, Larry Smith, Florian Leitner, Lorraine Tanabe, John Wilbur, Lynette Hirschman, and Alfonso Valencia. 2008. Evaluation of

text-mining systems for biology: overview of the second biocreative community challenge. *Genome Biology*, 9(Suppl 2).

Taku Kudoh. 2007. Crf++. http://crfpp.sourceforge.net/.

Claire Nédellec, Philippe Bessières, Robert Bossy, Alain Kotoujansky, and Alain-Pierre Manine. 2006. Annotation Guidelines for Machine Learning-Based Named Entity Recognition in Microbiology. In Melanie Hilario and Claire Nédellec, editors, *Proceedings of the ECML/PKDD workshop on Data and Text Mining in Integrative Biology*, pages 40–54.

Claire Nédellec. 2009. Guidelines for the annotation of gene and protein names. Technical report, INRA, June.

Sébastien Paumier. 2008. Unitex 2.0 user manual. `http://www-igm.univ-mlv.fr/~unitex/manuel.html`.

Sophie Rosset, Olivier Galibert, Guillaume Bernard, Eric Bilinski, and Gilles Adda. 2009. The LIMSI Multilingual, Multitask QAst System. *Lecture Notes in Computer Science*, 5706/2009:480–487.

Helmut Schmid. 1994. Probabilistic Part-of-Speech Tagging Using Decision Trees. In *International Conference on New Methods in Language Processing*, pages 44–49, Manchester, UK.