*** Call for Participation BARR track - IberEval 2017***

BARR track - Ibereval 2017

Biomedical Abbreviation Recognition and Resolution (BARR) track

September 19th, 2017: Workshop at SEPLN 2017

http://temu.inab.org/index.html

===============

Overview

===================

The recognition and resolution of abbreviations, acronyms and symbols is a critical step for a number of tasks including named entity recognition (NER), machine translation, information retrieval/indexing and document categorization among others. Therefore the implementation and availability of abbreviation recognition systems is of great practical impact for text mining and language processing.

In case of domains such as biomedicine and clinical research, abbreviations are particularly frequent, often referring to entities and concepts of importance such as genes, diseases, symptoms, drugs/chemicals or treatments. NER, relation extraction and clinical document coding systems usually need to cope with recognizing correctly short forms or abbreviations.

Abbreviations can be regarded as a ShortForm (SF) that denotes a longer word or phrase (LongForm, LF), typically its definition. Different strategies have been tested to detect short forms in English biomedical texts (Torii et al., 2007), using for instance alignment-based approaches, machine learning methods or rule-based strategies and some manually annotated corpora do exist (e.g. MEDSTRACT, Ab3P or BIOADI, see Islamaj Doğan et al., 2014). Far less effort has been made to detect short form- long form pairs in text written in other languages.

There is a growing number of biomedical and clinical documents written in Spanish, such as medical literature, medical agency reports, patents and particularly electronic health records. Moreover, according to some estimates there are over 500 million Spanish speakers worldwide.

As part of the IBEREVAL 2017 (http://cabrillo.lsi.uned.es/nlp/IberEval-2017/index.php) initiative we have proposed the Biomedical Abbreviation Recognition and Resolution (BARR) track with the aim promoting the development and evaluation of biomedical abbreviation identification systems.

For this track, participating teams have to detect mentions of pairs of Short Forms and their corresponding Long Forms from medical article abstracts written in Spanish. BARR track organizers provide a manually labeled training set exhaustively tagged with Short Form-Long Form pairs (in addition to other abbreviations).

This track is particularly interesting as some abstracts were manually transcribed, resembling preprocessing characteristics also found in clinical documents. In addition to the BARR manually annotated Gold Standard corpora, the BARR document collection will be released, consisting in the largest existing unified collection of medical article abstracts written in Spanish distributed through a special agreement with the publisher Elsevier to participating teams.

Additional details, sample sets, FAQ and inscription details can be found at:

BARR track URL: http://temu.inab.org/index.html

Contact e-mail: Martin Krallinger , mkrallinger@cnio.es

Important tentative dates

===================

- ~~**March 28th, 2017:** Release of sample data~~
- ~~**May 19th, 2017:** Release of training data subset1~~
- **May 30th, 2017:** Release of training data (full set)
- **June 19th, 2017:** Release of Testing Set
- **June 24th, 2017:** Submission of participant runs
- **June 26th, 2017:** Working notes submission due (short system description 3-5 pages)
- **June 28th, 2017:** Reviews of Working notes sent out to authors
- **July 1st, 2017:** Deadline to submit Camera ready revised Working notes

- **September 19th, 2017:** Workshop at SEPLN 2017

BARR track organizers

- Martin Krallinger, Biological Text Mining Unit (Bio-TeMUC), CNIO, Spain
- Santiago de la Peña, Biological Text Mining Unit (Bio-TeMUC), CNIO, Spain
- Ander Intxaurrondo, Biological Text Mining Unit (Bio-TeMUC), CNIO, Spain
- Jesús Santamaría, Biological Text Mining Unit (Bio-TeMUC), CNIO, Spain
- Jose A. Lopez-Martin, Medical Oncology, Hospital 12 de Octubre, Spain
- Alfonso Valencia, Life Sciences & Computational Biology, BSC, Spain
- Marta Villegas, Life Sciences & Computational Genomics, BSC, Spain
- Anália Lourenço, Next Generation Computer Systems Group, University of Vigo, Spain
- Gael Pérez, Next Generation Computer Systems Group, University of Vigo, Spain
- Martín Pérez, Next Generation Computer Systems Group, University of Vigo, Spain