Call for Papers 2nd workshop on MultilingualBIO: Multilingual Biomedical Text Processing
LREC 2020, Marseille (France), Saturday, May 16 2020 (afternoon)

As in other NLP areas, we are currently witnessing fast developments, with improved access, analysis and integration of healthcare-relevant information from heterogeneous content types, including electronic health records, medical literature, clinical trials, medical agency reports or patient-reported information available form social media and forums. There is an increasing automation of tasks in many critical areas, such as detecting interactions or supporting clinical decision. However, progress is very uneven depending on the language. Main achievements in processing biomedical text are almost restricted to English, with most other languages lagging behind in this respect, due to lack of annotated resources, incomplete vocabularies and insufficient in-domain corpora. More effort from the research community is needed to endow these languages with the necessary resources.
Also, machine translation in the biomedical domain is an important field of application. The need to translate biomedical texts occurs in many situations. Increasing cross-border mobility may require specific translation of medical records and discharge reports. In addition, internationalization of the pharmaceutical industry demands that technical specifications and package leaflets of medicines be translated to the language of the customer in several countries. Other areas of interest are translation of medical patents, laboratory reports, clinical trials or scientific publications.

MultilingualBIO at LREC2020 is a unique opportunity to promote the development of biomedical text processing resources and components in languages beyond English, exploring the use of novel methodological advances, e.g. transfer-learning techniques such as contextual embeddings, for a diversity of tasks in the domain, including machine translation.

In this workshop, we plan to address issues, such as the following (but not restricted to):

- Building of MT systems adapted to the biomedical domain.
- Production of multilingual corpora in the biomedical domain.
- Creation of multilingual biomedical glossaries, ontologies and terminological resources
- Application of transfer-learning techniques across tasks in the biomedical domain
- Extension of the coverage of the normative terminologies to languages other than English (e.g. ontologies from the Open Biomedical Ontology repository like HPO, LOINC, MEDRA, UMLS, SNOMED-CT, RxNorm etc).
- Dealing with localization issues, including adaptation to local varieties of international languages (UK vs USA English, Spanish from Spain and Latin America or USA, etc.).
- NLP and text mining applied to health, biomedicine and related domains, including also food safety.
- Medical named entity recognition and grounding systems beyond English.

Submissions
Three types of submissions are invited:
- Research papers, describing original research; these can be either long (6-8 pages, not including references) or short (3-4 pages, not including references);
- Project notes, describing recent, ongoing or planned projects (2-4 pages including references);
- Demonstration notes, accompanying demonstration of software, tools, or systems (2-4 pages including references).
Papers should be in compliance with the style sheet adopted for the LREC Proceedings. The MultilingualBIO proceedings will be published in the LREC 2020 proceedings, with a specific ISBN.

Identify, Describe and Share your LRs!
Describing your LRs in the LRE Map is now a normal practice in the submission procedure of LREC (introduced in 2010 and adopted by other conferences). To continue the efforts initiated at LREC 2014 about "Sharing LRs" (data, tools, web-services, etc.), authors will have the possibility, when submitting a paper, to upload LRs in a special LREC repository. This effort of sharing LRs, linked to the LRE Map for their description, may become a new "regular" feature for conferences in our field, thus contributing to creating a common repository where everyone can deposit and share data.
As scientific work requires accurate citations of referenced work so as to allow the community to understand the whole context and also replicate the experiments conducted by other researchers, LREC 2020 endorses the need to uniquely Identify LRs through the use of the International Standard Language Resource Number (ISLRN, www.islrn.org), a Persistent Unique Identifier to be assigned to each Language Resource. The assignment of ISLRNs to LRs cited in LREC papers will be offered at submission time.

Organising Committee:
- Maite Melero, Barcelona Supercomputing Center, Spain.
- Martin Krallinger, Barcelona Supercomputing Center, Spain.
- Marta Villegas, Barcelona Supercomputing Center, Spain.
- Jordi Armengol, Barcelona Supercomputing Center, Spain.
- Aitor Gonzalez-Agirre, Barcelona Supercomputing Center, Spain.

Scientific Committee (if available):
- Sophia Ananiadou (NaCTeM), UK.
- Nigel Collier (EMBL-EBI), UK
- Marta R. Costa-Jussá (Universitat Politècnica de Catalunya), Spain
- Hercules Dalianis (Stockholm University), Sweden
- Cristina España-Bonet (DFKI), Germany
- Jin-Dong Kim (DBCLS / ROIS), Japan
- Anália Lourenço (Universidad de Vigo), Spain

- Paloma Martínez (Universidad Carlos III de Madrid)
- Raquel Martínez Unanue (UNED), Spain
- Roser Morante (Vrije Universiteit Amsterdam), Holland
- Mariana Neves (German Federal Institute for Risk Assessment), Germany
- Patrick Ruch (University of Applied Sciences), Geneva
- Isabel Segura-Bedmar (Universidad Carlos III), Madrid
- Hua Xu (UTHealth School of Biomedical Informatics), US
- Pierre Zweigenbaum (CNRS), France