The BioCreative V.5 evaluation workshop: tasks, organization, sessions and topics

Martin Krallinger*[1,] Martin Pérez-Pérez[2,3], Gael Pérez-Rodríguez[2,3], Aitor Blanco-Míguez[2,3], Florentino Fdez-Riverola[2,3], Salvador Capella-Gutierrez[4], Anália Lourenço*[2,3,5], Alfonso Valencia[6,7,8,9]

[1]Biological Text Mining Unit, Structural Biology and BioComputing Programme, Spanish National Cancer Research Centre, C/Melchor Fernández Almagro 3, E-28029 Madrid, Spain
[2]ESEI - Department of Computer Science, University of Vigo, Ourense, Spain
[3]CINBIO - Centro de Investigaciones Biomédicas, University of Vigo, Campus Universitario Lagoas-Marcosende, 36310, Vigo, Spain
[4]Spanish National Bioinformatics Institute Unit, Structural Biology and BioComputing Programme, Spanish National Cancer Research Centre (CNIO), Madrid 28029, Spain
[5]CEB - Centre of Biological Engineering, University of Minho, Braga, Portugal
[6]Life Science Department, Barcelona Supercomputing Centre (BSC-CNS). C/Jordi Girona, 29-31, E-08034 Barcelona, Spain
[7]Joint BSC-IRB-CRG Program in Computational Biology. Parc Científic de Barcelona. C/ Baldiri Reixac 10, E-08028 Barcelona, Spain
[8]Institució Catalana de Recerca i Estudis Avançats (ICREA), Passeig de Lluís Companys 23, E-08010 Barcelona, Spain
[9]Spanish Bioinformatics Institute INB-ISCIII ES-ELIXIR, Madrid 28029, Spain

```
{martiperez, gaeperez,
aiblanco,riverola,analia}@.uvigo.es;
alfonso.valencia@bsc.es;
{scapella,mkrallinger}@cnio.es
```

The BioCreative (Critical Assessment of Information Extraction in Biology) is a community-wide effort with the aim of evaluating biomedical text mining and information extraction tools. It is organized in the form of shared tasks or challenges. The evaluation workshop linked to each BioCreative event serves to analyze the results obtained for each track, and to present the used Gold Standard datasets/evaluation settings. For each track, top scoring teams were invited to present a system's description talk to uncover the used methodology and system's implementation. Building on the achievements of previous BioCreative evaluations and workshops (BioCreative I, II, II.5, III, 2012 workshop, IV and V [1–5]) we have organized the BioCreative V.5 challenge evaluation workshop in Barcelona, Spain on April 26th-27th, 2017. To promote synergies, this workshop was co-located with the ELIXIR-EXCELERATE workshop on text mining infrastructure requirements, with a particular emphasis on

the role of Text Mining systems for Data Curation and Knowledge Management in the Life Sciences domain. The goal of BioCreative V.5 was to address some of the major barriers to the adoption and use of text mining tools, related to assessment, accessibility, interoperability, robustness and integration. Two traditional BioCreative tracks focused on monitoring progress on the recognition of relevant bio-entities (chemicals – CEMP track and genes/proteins – GPRO track). A novel track called TIPS (Technical interoperability and performance of annotation servers) focused on the technical aspects of the evaluation of continuous text annotation web services. The topics that were addressed during the workshop included: (1) continuous evaluation/stability of text mining tools, (2) enabling of interoperability of multiple text annotation systems at the technical level (design of compatible annotation schemas), (3) extraction of textual content from heterogeneous document sources, and (4) visualization and comparative assessment of automatic and manual annotations. Through the contribution of invited speakers and panel sessions, we also discussed aspects related to annotation formats, technical integration of text mining components, the experience of text mining techniques for the DARPA Big Mechanism program and the use of text processing web services and workflows. During the collocated workshops, the OpenMinted project and the open call for tenders were presented.

Acknowledgment

REFERENCES

1.   Arighi CN, Lu Z, Krallinger M, et al (2011). Overview of the BioCreative III workshop. BMC bioinformatics, 12(8), S1.
2.   Leitner F, Krallinger M, Rodriguez-Penagos C, et al (2008). Introducing meta-services for biomedical information extraction. Genome biology, 9(2), S6
3.   Pérez-Pérez M, Pérez-Rodríguez G, Rabal O, et al (2016). The Markyt visualisation, prediction and benchmark platform for chemical and gene entity recognition at BioCreative/ CHEMDNER challenge. Database, 2016, baw120.

4. Krallinger M, Rabal O, Lourenço A, et al (2015). Overview of the CHEMDNER patents task. In Proceedings of the fifth BioCreative challenge evaluation workshop (pp. 63-75).
5. Leitner F, Krallinger M, Rodriguez-Penagos C, et al (2008) Introducing meta-services for biomedical information extraction. Genome Biol 9 Suppl 2:S6.