

Fourth Workshop on Building and Evaluating Resources for Health and Biomedical Text Processing

Sophia Ananiadou
National Centre for Text Mining
School of Computer Science
University of Manchester

Workshop Motivation

- Exponential growth in biomedical and health information in digital form
 - 23 million references to biomedical literature in PubMed
 - Clinical records
 - Specialised social media groups
 - Extensive information on languages other than English

Workshop Motivation (2)

- Domain experts need to find nuggets of information regardless of:
 - Information source
 - Different topics, language styles, terminology differences, etc.
 - Mother tongue
 - Language barriers should not result in information overlook

Resources (1)

- Applications are reliant on high quality resources
 - **Databases and ontologies**
 - **Lexica**
 - **Annotated Corpora**

Resources (2)

- To ensure that resources are of maximum utility of applications, they must:
 - Be up-to-date
 - Account for frequent changes in biomedical terminology
 - Be adapted to cover multiple languages
 - Have wide coverage
 - Shared tasks (e.g. BioNLP, BioCreative) helping to increase subdomains covered by annotated corpora
 - Be reusable and interoperable
 - BioCreative Interoperability Initiative (BioC)
 - UIMA
 - Workbenches (e.g., Argo, U-Compare)

Workshop Statistics

- 16 Submissions
 - 6 accepted oral presentations
 - 8 accepted poster presentations
- International representation of authors :
 - Austria, Denmark, Finland, France, Germany, Hungary, Ireland, Japan, Norway, Spain, UK, USA

Paper Themes (1)

- Text Types
 - Clinical corpora
 - Medical notes
 - Biomedical literature
 - DrugBank
 - Social media
 - Climate science, marine science and environmental science

Paper Themes (2)

- Languages addressed
 - Danish
 - English
 - French
 - German
 - Spanish

Paper Themes (3)

- New/enhanced resources and schemes
 - Knowledge base for patient-centred question answering
 - Disease knowledge base
 - French biomedical annotated corpus of entities and concepts
 - Corpus of tweets annotated with adverse drug reactions
 - Enrichment of Drug-Drug Interaction (DDI) corpus with negation cues and scopes
 - Annotation scheme for entities and events climate science, marine science and environmental science
 - BioC interoperable annotation format and BioC-PMC dataset

Paper Themes (4)

- Systems
 - Information extraction
 - Concept extraction
 - Clinical text
 - Social media
 - Abbreviation resolution
 - Question answering
 - Generation of queries from medical reports
 - Classification of medical question types
 - QA system to answer Danish patient queries
 - Multi-lingual biomedical QA system

Workshop Schedule

- **9:30 – 10:30** - Invited Talk
 - Georgios Paliouras, Institute of Informatics and Telecommunications, Greece
The BioASQ Project
- **10:30 – 11:00** - Coffee break
- **11:00 – 12:20** - 4 Oral presentations
- **12:20 – 14:00** - Lunch
- **14:00 – 14:40** - 2 Oral presentations
- **14:40 – 15:10** - Short poster presentations
- **15:10 – 16:30** - Poster session and coffee break
- **16:30 – 18:00** - Structured discussion
 - Which new resources are most needed?
 - How will they be used?
 - Why are they not currently available

Thank you!

- We would like to thank:
 - Our invited speaker, George Paliouras, for agreeing to give a very interesting talk
 - Oral and poster presenters for their excellent work
 - Our programme committee, for their efforts in reviewing submissions and providing helpful feedback to authors
 - All workshop attendees, for their contributions towards the fruitful discussions
- We hope that everybody enjoyed the workshop!