OGER: OntoGene's Entity Recogniser in the BeCalm TIPS Task
Lenz Furrer, Fabio Rinaldi

Input Formats
- PubMed abstracts
- PubMed Central full-text
- plain text (through JSON)

Output Formats
- BioC XML
- BeCalm TSV
- BeCalm JSON

Annotations
- task: NER + linking
- approach: knowledge-based
- fuzzy-matching strategy: SRC-1 = SRC 1 = SRC1
- entity types: flexible
  - Cellosaurus cell lines
  - CTD chemicals
  - CTD diseases
  - GO cellular components
  - NCBI taxonomy

Processing Time
- Contributions to total processing time:
  - Depending on document volume
  - fixed overhead
  - does not contribute

Results
- seconds/document: 1.06943 s
- seconds/Byte: 0.00086 s
- seconds/annotation: 0.07227 s
- annotations/document: 14.7923
- time between failures: [no failure] (shared)
- time to repair: [no failure] (shared)

Bio Term Hub
- Gene Ontology
- Cell Ontology
- ChEBI
- NCBI

Log-based Analysis
- Bar chart showing request times for different servers:
  - PubMed
  - Abstract server
  - Patent server
  - all

Institute of Computational Linguistics
Andreasstr. 15, CH-8050 Zürich
{furrer,rinaldi}@cl.uzh.ch

OntoGene
www.ontogene.org