BioNLP 2010 Poster

Ontology-Based Extraction and Summarization of Protein Mutation Impact Information

Concordia

Nona Naderi and René Witte Semantic Software Lab, http://www.semanticsoftware.info Department of Computer Science and Software Engineering, Concordia University, Montréal

Our Mutation Impact Miner extracts mutation impacts from full-text papers and populates an OWL-DL ontology with impact information for guerying and summarization.

Impacts are described through other concepts, in particular, changes to physical quantities such as pH and temperature:

Impact word of at 60°C by up to 85 and the transition erature by 2.5°C to 5.8°C.

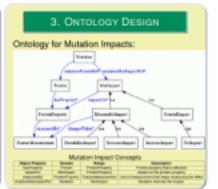
Unit of measurement Protein property

2. MUTATION-IMPACT RELATIONS

Proximity heuristic to associate mutations with their impacts. For example:

HMAEFICO, reduced the specific activity by 32-74%; whereas mutation HOW increased it by 18%.

The impact, increased, is caused by "H86W", which is nearer to the impact word than "HB5A/E/F/K/Q". And in the same way, reduced is associated with "H86A/E/F/K/Q" rather than "H86W"



5. Design & Implementation

Mutation impacts are detected by an NLP pipeline developed based on the General Architecture for Text Engineering (GATE) [1]:

Preprocessing: Tokenizing and sentence splitting

Entity recognition: Ontogazetteering and JAPE grammars to capture impact words, units of measurement and protein properties

Relation extraction: Using rules and grammars mutations are associated with their impacts and sentences containing impacts are

Ontology population: Populating the OWL-DL ontology with the detected entities and relations



6. EVALUATION

The performance of the system was evaluated on the abstracts of four different mutation corpora. Altogether, 58 documents were manually annotated with their impacts. The results are shown in the table below; here, #C, #P, #M, and #S correspond to the correct, partially correct, missing, and spurious impact sentences, respectively; and P, R, F are the precision, recall, and F-measure:

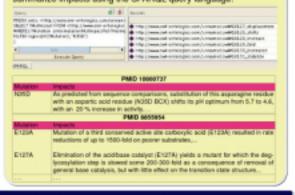
Impact detection evaluation results on four corpora Dioxygenase 77 0 13 14 0.846 0.855

The evaluation of associating the mutations with their impacts on the "Xylanase" corpus:

Relation detection evaluation results			
	Precision	Recall	F-Measure
Lenient (Partial matches included)	88%	80%	91%
Average (of Lenient and Strict)	86%	76%	80%
Strict (Partial matches not counted)	51.8%	46.6%	49.06%

6. APPLICATION

The OWL-DL ontology is populated with the entities and their relations to create a knowledge base, which can then be queried for a certain protein, mutation or protein property. For example, to summarize impacts using the SPARQL query language:



Maynard, K. Bontcheva, and W. Tablan. GATE: an Architecture for Development of Robust HLT Applications. Proceedings of the 48th Anniversary Meeting of the Association for Comp., 2010. http://gate.ac.uk/.

BioNLP 2010 Poster

Published on semanticsoftware.info (https://www.semanticsoftware.info)

Except where otherwise noted, all original content on this site is copyright by its author and licensed under a <u>Creative Commons Attribution-Share Alike 2.5 Canada License</u>.

Source URL (retrieved on 2025-12-02 13:51): https://www.semanticsoftware.info/image/bionlp-2010-poster