Automated Extraction of Protein Mutation Impacts from the Biomedical Literature

Submitted by [nona](#) [1] on Sun, 2011-09-18 07:45

- [Open Mutation Miner](#) [2]
- [OrganismTagger](#) [3]
- [Corpora](#) [4]
- [Summarization](#) [5]
- [Bioinformatics](#) [6]
- [NLP](#) [7]
- [Ontology](#) [8]
- [Text Mining](#) [9]

Title	Automated Extraction of Protein Mutation Impacts from the Biomedical Literature
Publication Type	Thesis
Year of Publication	2011
Authors	[Naderi, N.](#) [10]
Academic Department	Department of Computer Science and Software Engineering
Degree	M. Comp. Sc.
Date Published	09/2011
University	Concordia University
City	Montreal
Thesis Type	Masters

Abstract: Mutations as sources of evolution have long been the focus of attention in the biomedical literature. Accessing the mutational information and their impacts on protein properties facilitates research in various domains, such as enzymology and pharmacology. However, manually reading through the rich and fast growing repository of biomedical literature is expensive and time-consuming. A number of manually curated databases, such as BRENDA (http://www.brenda-enzymes.org [11]), try to index and provide this information; yet the provided data seems to be incomplete. Thus, there is a growing need for automated approaches to extract this information.

In this work, we present a system to automatically extract and summarize impact information from protein mutations. Our system extraction module is split into subtasks: organism analysis, mutation detection, protein property extraction and impact analysis. Organisms, as sources of proteins, are required to be extracted to help disambiguation of genes and proteins. Thus, our system extracts and grounds organisms to NCBI. We detect mutation series to correctly ground our detected impacts. Our system also extracts the affected protein properties as well as the magnitude of the effects.

The output of our system is populated to an OWL-DL ontology, which can then be queried to provide structured information. The performance of the system is evaluated on both external and internal corpora and databases. The results show the reliability of the approaches. Our Organism extraction system achieves a precision and recall of 95% and 94% and a grounding accuracy

of 97.5% on the OT corpus. On the manually annotated corpus of Linneaus-100, the results show a precision and recall of 99% and 97% and grounding with an accuracy of 97.4%.

In the impact detection task, our system achieves a precision and recall of 70.4%-71.8% and 71.2%-71.3% on a manually annotated documents. Our system grounds the detected impacts with an accuracy of 70.1%-71.7% on the manually annotated documents and a precision and recall of 57%-57.5% and 82.5%-84.2% against the BRENDA data.

URL	http://spectrum.library.concordia.ca/35931/ [12]

Attachment	Size
Naderi_MSc_F2011.pdf [13]	3.33 MB

Source URL (retrieved on *2025-12-07 21:48*):
https://www.semanticsoftware.info/biblio/automated-extraction-protein-mutation-impacts-biomedical-literature

Links:
[1] https://www.semanticsoftware.info/users/nona
[2] https://www.semanticsoftware.info/category/project/mutation-miner
[3] https://www.semanticsoftware.info/taxonomy/term/249
[4] https://www.semanticsoftware.info/category/topic/nlp/corpora
[5] https://www.semanticsoftware.info/category/topic/nlp/summarization
[6] https://www.semanticsoftware.info/category/topic/bioinformatics
[7] https://www.semanticsoftware.info/category/topic/nlp
[8] https://www.semanticsoftware.info/category/topic/ontology
[9] https://www.semanticsoftware.info/category/topic/text-mining
[10] https://www.semanticsoftware.info/biblio/author/22
[11] http://www.brenda-enzymes.org
[12] http://spectrum.library.concordia.ca/35931/
[13] https://www.semanticsoftware.info/system/files/Naderi_MSc_F2011.pdf