

Biomedical Data Explosion

- Huge growth in online biological and biomedical data
 - Large databases, ontologies
 - KEGG, ChEBI, UniProt, AraCyc, GO, GOLD, Gramene,...
- Majority of material published as unstructured text
 - 18,000,000+ Medline documents
 - 40,000 new added per month
 - Heavily searched by researchers

Data Integration in ONDEX

- Systems Biology research requires integration
- ONDEX SABR project:
 - Supports research in systems biology through data integration via the ONDEX e-tool, providing:
 - Graph-based data integration & visualization
 - Workflows in Taverna
 - Graph-based analysis of biological networks, and
 - Text mining – NaCTeM

ONDEX Biological Use Cases

ONDEX Use Case 1

- Organism-Habitat Association
- Link protein families, org, habitat
- Newcastle University (NU)
 - Hirt, Nakjang, Wipat

ONDEX Use Case 2

- Yeast metabolic reactions, kinetics
- Enrich, refine metabolic models
- University of Manchester (UM)
 - Dobson, Hull, Mendes
- EML-R: Muller, Rojas, Weidemann

ONDEX Use Case 3

- Link plant traits, genes, enzymes
- Associate with increased biomass
- Rothamsted Research (RR)
 - Canevet, Hanley, Hassani-Pak,...

Text Mining Solutions

- Automatically extract new information from literature[1]
- Recognize key concepts, entities in text
 - Ambiguity: 'CAT': animal, enzyme, procedure,...
 - Variability: 'plague', 'Yersinia pestis', 'Y.pestis',...
- Recognize relations, events involving these concepts
 - Linguistic analysis enables general extraction
- NaCTeM Systems: Facta[2], Medie[3], Kleio[4], Nemine[5]
 - Build on Controlled Vocab., Relations in DBs
 - E.g.: KEGG, UMLS Meta-thesaurus, ChEBI,...

Data Integration for Text Mining in ONDEX

- Integrate databases, ontologies to support recognition and disambiguation of key entities
- Key data resources in ONDEX use cases
 - NU: GOLD, NCBI Taxonomy
 - UM: YEASTNET, HMDB, IUBMB
 - <http://www.comp-sys-bio.org/yeastnet/>
 - <http://www.chem.qmul.ac.uk/iubmb/enzyme/>
 - RR: Gramene, Arabidopsis Hormone DB
 - <http://www.gramene.org/>, <http://ahd.cbi.pku.edu.cn/>
- General NaCTeM Text Mining tools employ:
 - UMLS meta-thesaurus, KEGG, ChEBI, UniProt, HMDB, DrugBank, Biothesaurus, PubChem, NCIDB
 - GENA: FlyBase, HUGO, OMIM, SGD, etc.
- Combining DBs, ontologies with statistical language processing enables high quality entity recognition
 - Up to 92% F-score (best: organisms, tagged sents)
 - Metabolites, habitats more challenging

FACTA

Finds associations between important concepts in biomedical literature. The system can show concepts relevant to the query term(s) by analyzing the concepts mentioned in the documents retrieved by the query.

adrenaline

Gene/Protein Disease Symptom Drug Enzyme Compound

Query: **adrenaline**
15,938 document(s) hit in 18,511,090 MEDLINE articles (0.03 seconds). [Excerpts](#) (click to show).

Concepts found in the documents ranked by [**Frequency** | Pointwise Mutual Information | Freq. * PMI]

Human Gene/Protein	Disease	Symptom	Drug	Enzyme
insulin 982	hypertension 593	anesthesia 943	Norepinephrine 412	renin 401
AMP 466	hypoglycemia 313	pain 374	ICI 143	adenylate cyclase 275
renin 401	essential hypertension 249	analgesia 351	Phentolamine 89	thrombin 199
collagen 393	cold 234	hypothermia 84	caffeine 89	monoamine oxidase 172
glucagon 376	depression 229	exhaustion 72	Clonidine 83	tyrosine hydroxylase 170
vasopressin 344		fatigue 59	NORADRENALINE 58	

References

- [1] Ananiadou, S. & McNaught, J. (Eds) (2006) Text Mining for Biology and Biomedicine, Artech House Books.
- [2] Tsuruoka, Y., Tsujii, J. and Ananiadou, S. 2008. FACTA: a text search engine for finding associated biomedical concepts, Bioinformatics, Vol. 24, No. 21, pp. 2559-2560
- [3] Miyao, Y., Ohta, T., Masuda, K., Tsuruoka, Y., Yoshida, K., Ninomiya, T. and Tsujii, J. (2006) Semantic Retrieval for the Accurate Identification of Relational Concepts in Massive Textbases. In the Proceedings of COLING-ACL 2006. Sydney, Australia, pp. 1017--1024.
- [4] Nobata, C., Cotter, P., Okazaki, N., Rea, B., Sasaki, Y., Tsuruoka, Y., Tsujii, J. and Ananiadou, S. Kleio: a knowledge-enriched information retrieval system for biology. In Proc. of the 31st Annual International ACM SIGIR Conference, pp. 787--788, 2008.
- [5] Sasaki, Y., Tsuruoka, Y., McNaught, J. and Ananiadou, S., How to Make the Most of NE Dictionaries in Statistical NER, ACL-2008 Workshop on Current Trends in Biomedical Natural Language Processing (BioNLP-08), Columbus, pp. 63-70, 2008..

Contributors: Gina-Anne Levow, Sophia Ananiadou, Catherine Canevet, Paul Dobson, Steve Hanley, Keywan Hassani-Pak, Robert Hirt, Duncan Hull, Pedro Mendes, Wolfgang Muller, Sirintra Nakjang, Chikashi Nobata, Yutaka Sasaki, Yoshimasa Tsuruoka, Jun'ichi Tsujii, Andreas Weidemann, Anil Wipat

MEDIE

A semantic extraction engine identifies sentences containing biomedical relations of queried terms from MEDLINE abstracts.

Updated data from Laboratory

subject: interleukin 2 verb: activate object:

Results 1-80 for interleukin 2 activate 1.06 seconds (searched 3.30% of Medline)

Sort by Rank Date Sort

show/50 results subject verb object gene disease

show next >

1. [IL-21 promotes survival and maintains a naive phenotype in human CD4+ T lymphocytes.](#)
Sylvie Ferrari-Lacraz, Rachel Chocheportche, Gregory Schreier, Nicolas Mohari, Jean Villard, Jean-Michel Dayer, pp. 1008-18, Volume 20, Issue 8, International Immunology, 2008 [PMID:18556671]
These effects contrast with those of IL-2, which induces the marked proliferation of CD4(+) T lymphocytes, leading to an activated-memory phenotype.
2. [Serum vascular endothelial growth factor and fibronectin predict clinical response to high-dose interleukin-2 therapy.](#)
Marianna Sabatino, Seunghye Kim-Schulze, Monica C Panelli, David Stroncki, Ewa Wang, Bret Taback, Dae Won Kim, Gal Derafelle, Zoltan Pos, Francesco M Marnicola, Howard L Kaufman, pp. 2645-52, Volume 27, Issue 16, Journal of clinical oncology: official journal of the American Society of Clinical Oncology, 2009 [PMID:19364969]
PURPOSE: High-dose interleukin-2 (IL-2) induces durable therapeutic responses in a small subset of patients with metastatic melanoma and renal cell carcinoma, but simple pretreatment predictors of response have not been identified.