Conceptual Analysis, Terminology, Ontology

Jun'ichi Tsujii

University of Tokyo, Tokyo, Japan

tsujii@is.s.u-tokyo.ac.jp

Challenges

- Tools for terminologies, supported by large corpora (Medline, BioMed Central) and NLP techniques
- Feasibility of the tools for real world applications
  - Bioinformatics, Clinical bioinformatics, Health-care
  - e-Science, e-Learning, e-Education
  - From successful seeds (eg: IR), from real needs (eg: Bioinformatics)
- Linking terminology with other technologies
  - Semantic web, Grid (from sharing computational power to sharing knowledge and data)

Possible contribution of NLP

(1) Recognizing terms (70% with semantic class recognition)
-- from techniques dependent on subject domains to those independent ones
-- integration of larger structures

(2) Gathering related terms and term variants
-- Machine Learning (semi-unsupervised)

(3) Gathering semantic similar terms
-- Knowledge discovery from Web to that from specialized subject domains
(4) NER for gathering new terms

(5) Large discrepancy between the concept domain and the language domain

(6) Expressions in context

Experiment (Gathering Terms)

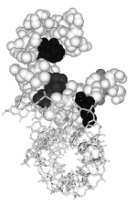
Automatic learning of rules of spelling variations
[Tsuruoka, SIGIR 03]

- Corpus
  - MEDLINE: the largest collection of abstracts in the biomedical domain
- Rule learning
  - 83,142 abstracts
  - Obtained rules: 14,158
- Evaluation
  - 18,930 abstracts
  - Count the occurrences of each generated variant.

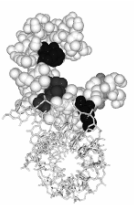
Automatic learning of rules of term variations
[Tsuruoka, Applied bioinformatics 04]

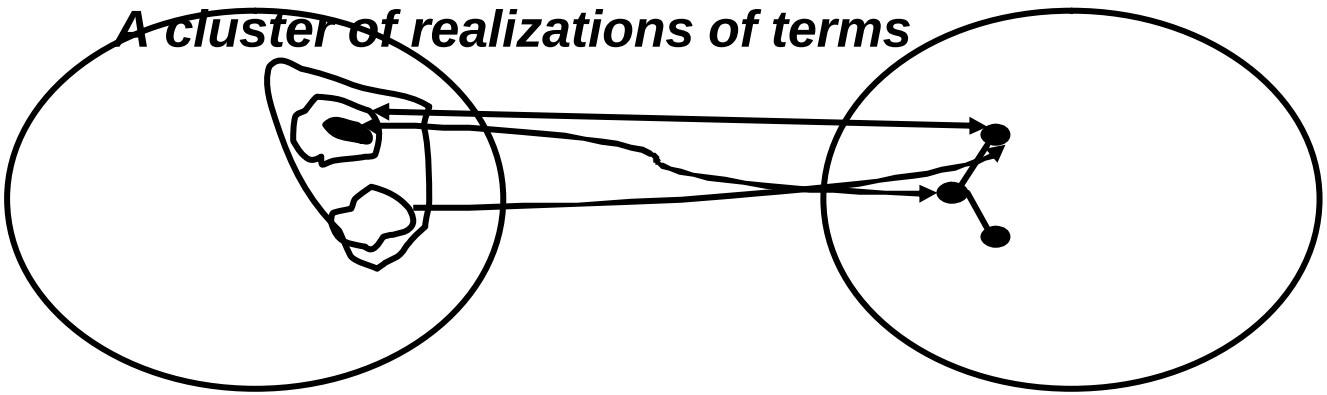
- Training Data
  - Meta-thesaurus
  - Variant pairs with the same concept IDs
  - Under “Amino acid or protein”,
  - 36,112 variant pairs
- Rule induced
  - Rules: 4,780,793 rules
- Evaluation
  - Matching against running texts

1.000	NF kappa B	128
0.500	Transcription Factor NF kappa B	0
0.429	NF-kappa B	912
0.286	NF kB, Transcription Factor	0
0.286	NF kB	0
0.286	Immunoglobulin Enhancer-Binding Protein	0
0.286	Immunoglobulin Enhancer Binding Protein	0
0.286	Enhancer-Binding Protein, Immunoglobulin	0
0.286	kappa B Enhancer Binding Protein	0
0.286	Transcription Factor NF-kB	0
0.286	Transcription Factor NF kB	0
0.286	Factor NF-kB, Transcription	0
0.286	nuclear factor kappa beta	2
0.286	NF kappaB	1
0.273	NF kappa B chain	0
0.273	NF kappa B subunit	0
0.214	Transcription Factor NF-kappa B	0
0.214	NF-kB, Transcription Factor	0
0.214	NF-kB	67
0.200	Neurofibromatosis Type kappa B	0



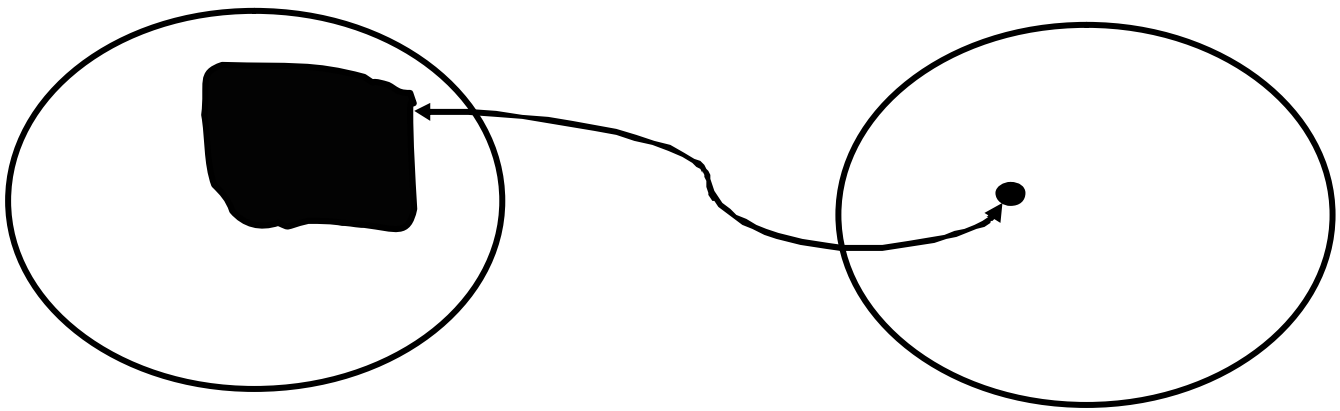
1.000	tumor necrosis factor A	0
0.316	TNF A	1
0.200	tumor necrosis factor	1653
0.158	TNF alpha	358
0.133	TNFA	32
0.133	TNF	2631
0.133	Tumour necrosis factor alpha	14
0.133	Tumor Necrosis Factor alpha	2
0.133	Tumor Necrosis Factor-Alpha	0
0.133	TUMOR NECROSIS FACTOR.ALPHA	0
0.133	Tumor necrosis factor alpha	52
0.133	Tumor Necrosis Factor-alpha	8
0.133	TNF-Alpha	0
0.133	TNF-alpha	6899





Language Domain *Concept Domain*

Homologues/Orthologues



Process of Ribosomal subunit assembly

AmiGO

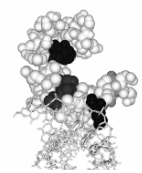
Search GO: Exact Match Subm

Terms Gene Products Advanc

GOst

Top Docs Gene Ontology GO Links GO Summary

- **GO:0003673 : Gene Ontology (146200)**
  - GO:0008150 : biological process (96312)
    - GO:0009987 : cellular process (31905)
      - GO:0050875 : cellular physiological process (25201)
        - GO:0008151 : cell growth and/or maintenance (22177)
          - GO:0016043 : cell organization and biogenesis (6226)
            - GO:0007028 : cytoplasm organization and biogenesis (3607)
              - GO:0042254 : ribosome biogenesis and assembly (736)
                - GO:0042255 : ribosome assembly (107)
                  - GO:0042257 : ribosomal subunit assembly (84)
                    - **GO:0000027 : ribosomal large subunit assembly and maintenanc**
    - GO:0007582 : physiological process (60310)
      - GO:0050875 : cellular physiological process (25201)
        - GO:0008151 : cell growth and/or maintenance (22177)
          - GO:0016043 : cell organization and biogenesis (6226)
            - GO:0007028 : cytoplasm organization and biogenesis (3607)
              - GO:0042254 : ribosome biogenesis and assembly (736)
                - GO:0042255 : ribosome assembly (107)
                  - GO:0042257 : ribosomal subunit assembly (84)
                    - **GO:0000027 : ribosomal large subunit assembly and maintenanc**
    - GO:0005575 : cellular component (79199)
    - GO:0003674 : molecular function (97507)



Get this tree as RDF XML.

AmiGO : GO:0000027 details - Microsoft Internet Explorer

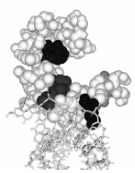
ファイル(F) 編集(E) 表示(V) お気に入り(A) ツール(T) ヘルプ(H)

戻る 検索 お気に入り メディア

アドレス(D) http://godatabase.org/cgi-bin/go.cgi?query=GO:0000027&view=details&search_constraint=terms&depth=0&sessi

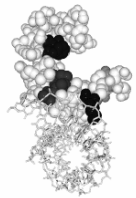
GO Term:

	Gene Symbol:	Datasource:	Evidence:	Full name:
GO:0000027 : ribosomal large subunit assembly and maintenance				
☐	PF02 0173	Pfalciparum	ISS - With	hypothetical protein
☐	RL25 YEAST CCgst	UniProt	IMP	60S ribosomal protein L25
☐	RLA0 YEAST CCgst	UniProt	IMP	60S acidic ribosomal protein P0
☐	RL6B YEAST CCgst	UniProt	IMP	60S ribosomal protein L6-B
☐	RL11 YEAST CCgst	UniProt	IMP	60S ribosomal protein L11
☐	RL3 YEAST CCgst	UniProt	IMP	60S ribosomal protein L3
☐	RL12 YEAST CCgst	UniProt	IMP	60S ribosomal protein L12
☐	DBP3 YEAST CCgst	UniProt	IMP	Probable ATP-dependent RNA helicase
☐	RL5 YEAST CCgst	UniProt	IMP	60S ribosomal protein L
☐	SQT1 YEAST CCgst	UniProt	IGI	Ribosome assembly pro



1: J Biol Chem. 1994 Jun 3;269(22):15689-96.

FREE full text article at www.jbc.org



Ribosomal protein P0, contrary to phosphoproteins P1 and P2, is required for ribosome activity ar Saccharomyces cerevisiae viability.

Santos C, Ballesta JP.

Centro de Biologia Molecular Severo Ochoa, Consejo Superior de Investigaciones Cienticas, Canto Blanco, Madrid, Spain.

Protein P0 in Saccharomyces cerevisiae is found only in the ribosomes and not free in a cytoplasmic pool like the structurally relatec P1 and P2 proteins. Analogously, P0 stays bound to the particles in conditions that release the other P proteins. Attempts to obtain y strains carrying an interrupted P0 gene by direct gene disruption techniques of different yeast strains always resulted in haploid cells one disrupted and one intact copy of the gene. Disruption of the unique P0 genomic copy seems to induce a duplication and occasion chromosomal transposition of the gene. Conditional null mutants of P0 were then constructed carrying the P0 gene under the contrc inducible GAL1 promoter. A 2-3-fold excess of P0 mRNA is found in the conditional mutant when grown in galactose; however, only a increase of the P0 protein is detected in total cell extracts. No P0 protein is detected in the cell supernatant, indicating that, like the standard ribosomal proteins and opposite to the other P proteins, the protein not bound to the ribosomes is degraded. Transfer of the to the restrictive conditions causes, after some generations, a growth stop that finally leads to cell death. The growth decline is paral a reduction in the polysome number and the appearance of half-mer particles as well as by an accumulation of 60 S particles deficient and in the acidic proteins P1 and P2. These results indicate that P0 is required for the interaction of the acidic P1 and P2 proteins w ribosomes, and in its absence, deficient 60 S ribosomes are assembled which are inactive in protein synthesis resulting in cell lethality

PMID: 8195220 [PubMed - indexed for MEDLINE]

Term: Ribosomal large subunit assembly and maintenance

and in its absence, deficient 60 S ribosomes are assembled which are inactive in protein synthesis resulting in cell lethality.

Mutations that completely abolish recognition of 26 S rRNA, however, block the formation of 60S particles, demonstrating that binding of L25 to this rRNA is an essential step in the assembly of the large ribosomal subunit.

Depletion of Saccharmoyces cerevisiae ribosomal protein L16 causes decrease in 60S ribosomal subunits and formation of half-mer polyribosomes.

Without L3, apparent synthesis of several 60 S subunit proteins diminished, and 60S subunit did not assemble. A similar phenomenon occurred, when a second strain, synthesis of ribosomal protein L29 was prevented.