

Transys Ontology design task

1. Specify the domain and the role of the ontology

Domain: *Membrane proteins, specifically insertion, assembling and organization into complexes.*

Role:

- *Have better results in a search engine using expansion request mechanisms.*
- *Share common representation and vocabulary of the project domains.*

2. Acquire the knowledge from biologist expertise, scientific papers, databanks, thesaurus

Biological literature corpora:

- secretome: 3 998 Medline abstracts from request
("bacillus subtilis" OR "escherichia coli") AND ("protein secretion" OR secretome OR sec OR TAT OR "protein export" OR "signal peptide" OR "signal peptides")
- promemb: 1032 Medline abstracts from request
("bacillus subtilis" OR "escherichia coli") AND ("membrane protein structure" OR "membrane protein structures" OR "transmembrane protein" OR "transmembrane proteins" OR "membrane protein complex" OR "membrane protein complexes" OR "membrane protein insertion" OR "membrane protein assembly" OR "membrane protein assemblies" OR "membrane protein topology" OR "TM helices" OR "TM helix" OR "transmembrane helices" OR "transmembrane helix")

Automatic extraction of list of term candidates:

- secretome: 75,685 term candidates
574 preselected by a biologist
- promemb: 23,296 term candidates
183 (specific to structure) preselected by a biologist
284 preselected by a biologist

Biologists expertise

3. Conceptualize

- identify the **key concept** of the ontology, their properties and relations

Main concepts

- *biological process*
- *cellular component*
- *physico-chemical property*
- *protein region*
- *protein structure*
- *transport system*

Relations

- ISA
- part of
- *codes for*
- *is part of complex*
- identify the terms to be used as **concept labels**
- build the structures

4. Integrate / align existing thesaurus / ontology (GO, MeSH, BactriOnto)