

Anderson Carlos D. Sanches
Orientadora: Renata Wassermann
MAC 5701 – Tópicos em Ciência da Computação
Prof.: Yoshiharu Kohayakawa

Plano de Estudos

Construção de uma Ontologia de Bioquímica com o uso do PowerLoom.

Proposta de Estudo

Construir uma ontologia [Guarino] de Bioquímica usando o sistema de representação de conhecimento PowerLoom [PowerLoom]. A proposta é testar o sistema numa situação real, tentando ao mesmo tempo construir uma ontologia dos caminhos metabólicos da glicose, fazendo uso de vocabulários já construídos e presentes na rede. Faremos um estudo de caso dos pontos fortes e limitações do sistema PowerLoom quanto a nossa ontologia.

Motivação

Recentemente, com a iniciativa do mapeamento do genoma humano e de várias outras espécies, surgiram vários bancos de conhecimento disponíveis na web [ByoCyC] [Tambis] que tentam produzir vocabulários que podem ser aplicados a todos os organismos, tanto no sentido do conhecimento genético quanto do papel das proteínas nas células [GO]. Além do genoma, outros descrevem o caminho metabólico destes componentes [GK] [Kegg]. Estas vias metabólicas são bastante semelhantes entre as várias espécies, sendo que podemos inferi-las em espécies ainda não estudadas através da presença ou ausência de certos componentes químicos [Lenhof]. Queremos mapear estas vias metabólicas fazendo uso de vocabulários anotados presentes na web, de forma que um especialista no domínio possa entender e pesquisar a base de dados [Hill].

Faremos uso do PowerLoom, ao mesmo tempo escrevendo uma ontologia em uma variante da gramática KIF [Gruber] (que é uma linguagem baseada em lógica) e testando o PowerLoom para ver se ele atende aos propósitos de acomodar nossa base de conhecimento, no sentido de processar regras e fazer inferências automáticas sobre as instâncias ali presentes [GOC].

Bibliografia

[AllGenes] AllGenes. <http://www.allgenes.org/index.html>

- [Ashburner] Ashburner M. *On the representation of gene function in genetic databases*.
- [BioCyC] BioCyC Knowledge Library. <http://biocyc.org/>
- [ENSEMBL] Anotação automática de eucarióticos.
<http://www.ensembl.org/>
- [GK] Genome Knowledge. <http://www.genomeknowledge.org/>
- [GO] The Gene Ontology Consortium. The Gene Ontology:
www.geneontology.org
- [GOC] The Gene Ontology Consortium. *Gene Ontology: tool for the unification of biology*. Nature Genetics 25: 25-29.
- [GOBO]. Global Open Biological Ontologies -
<http://www.geneontology.org/doc/gobo.html>
- [Gruber] Gruber, T. R. *A Translation Approach to Portable Ontology Specifications*. Academic Press.
- [Guarino] Guarino, N. & Giaretta, P. *Ontologies and Knowledge Bases: Towards a Terminological Clarification*
- [Hill] Hill, D.P., Blake, J.A., Richardson, J.E. and Ringwald, M. *Extension and Integration of the Gene Ontology (GO): Combining GO vocabularies with external vocabularies*. Genome Research 12: 1982-1991.
- [Kegg] Kyoto Encyclopedia of Genes and Genomes (KEGG).
<http://www.genome.ad.jp/kegg>
- [Lenhof] Lenhof, H et al. BioMiner – modeling, analyzing, and visualizing biochemical pathways and networks. Oxford University Press.
- [MGI] Mouse Genome Informatics. <http://www.informatics.jax.org/>
- [NAR] Nucleic Acids Research. Oxford Journals online.
<http://www3.oup.co.uk/nar/database/a/>
- [NCBI] National Center for Biotechnology Information.
<http://www.ncbi.nlm.nih.gov/>
- [PowerLoom] Sistema de representação de conhecimento.
<http://www.isi.edu/isd/LOOM/PowerLoom/>
- [Tambis] Tambis System. <http://bagel.cs.man.ac.uk/>

[TIGR] The Institute for Genomic Research. <http://www.tigr.org/>