Supplementary Material

## A Functional similarity between GO terms

# Node-based method

Resnik calculates semantic similarity based on the information content (IC) of the terms in ontology. The IC of *t* term is defined by  Equation 1 (Lee et al., 2004):

 (1)

Here is the probability of the occurrence of the *t* term.

Resnik (Resnik, 1995) calculated the similarity as Equation 2:

 (2)

Among them, is the common ancestor of the most abundant IC values of and in ontology.

Lin (Lin, 1998) calculated the similarity as Equation 3:

 (3)

# Edge-based method

Pekar (Pekar and Staab, 2002) calculated the similarity as  Equation 4:

 (4)

Where denotes the longest distance between term and in GO graph, and is Least Common Ancestor (LCA) of , . Three distances are used in the equation.

# Hybrid method

Wang (Wang et al., 2007) calculated the similarity as  Equation 5, 6 and 7:

 (5)

 (6)

 (7)

Factor and the fixed weight assigned by the edge. represents the contribution value of any term *t* to the semantics of term *A*. is the set of all ancestors of GO term *a*, while is the set of corresponding links. represents the term contribution factor of edge *e* connecting term *T* and its sub term *t* in . represents the term contribution factor of edge *e* connecting term *t* and *t's* sub term *t'* in . represents the semantic value of GO term *A*, and is the semantic similarity measure of GO term *A* and *B*.

# References

Lee, S.G., Hur, J.U., and Kim, Y.S. (2004). A graph-theoretic modeling on GO space for biological interpretation of gene clusters. *Bioinformatics* 20(3)**,** 381-388.

Lin, D. (1998). "An information-theoretic definition of similarity", in: *Proceedings of the 15th International Conference on Machine Learning*, 296-304.

Pekar, V., and Staab, S. (2002). "Taxonomy learning-factoring the structure of a taxonomy into a semantic classification decision", in: *COLING 2002: The 19th International Conference on Computational Linguistics*).

Resnik, P. (1995). "Using information content to evaluate semantic similarity in a taxonomy," *In: Proceedings of the International Joint Conference for Artificial Intelligence,* 448–453.

Wang, J.Z., Du, Z., Payattakool, R., Yu, P.S., and Chen, C.-F. (2007). A new method to measure the semantic similarity of GO terms. *Bioinformatics* 23(10)**,** 1274-1281.