

Reasoning Service for the Support of Genome-Wide Association Studies

Zhisheng Huang¹, Gaston Tagni¹, Mark A. Greenwood², Angus Roberts², Raluca Brehar³, and Mattias Johansson⁴

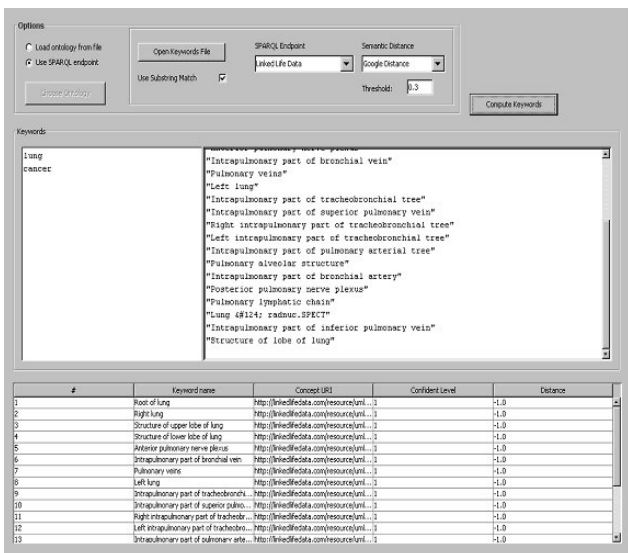
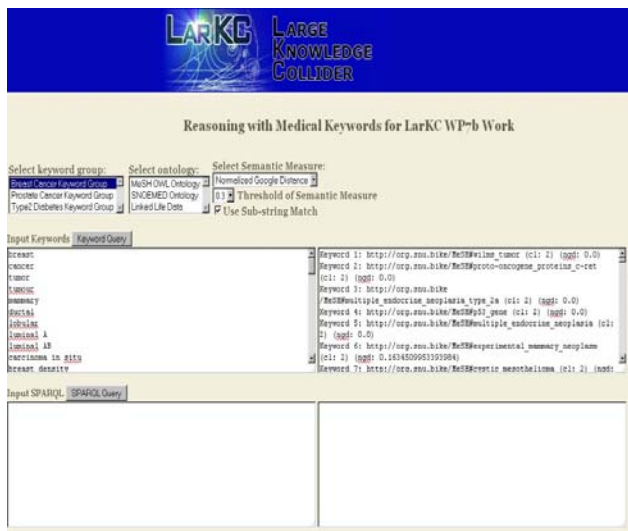
1. Department of Computer Science, Vrije Universiteit Amsterdam, The Netherlands

2. Department of Computer Science, University of Sheffield, UK

3. Technical University of Cluj-Napoca, Romania

4 International Agency for Research on Cancer (IARC), World Health Organization

Genome-Wide Association Studies have been widely used to identify common genetic factors that influence health and disease. We have developed a reasoning service to support the GWAS scientists in obtaining additional information which can be incorporated into the statistical association models to improve ranking techniques. The service is achieved with the support of the LarKC platform for reasoning over massive amounts of biomedical semantic data. Statistical and semantic distance measures, such as Google distance, are used to aid reasoning.



Pseudo Code For Deriving New Keywords

inputKeywords = initial set of keywords (defined by domain expert)

allFlag = true

threshold = the threshold defined by the user (eg. 0.7)

useSubstringMatch = true

FOR EACH keyword K in inputKeywords DO{

 concepts = search for ontology concepts whose name is an exact match with K's string value;

 FOR EACH concept C in concepts DO{

 IF (allFlag){

 outputKeywords += all the descendants of concept C;

 }ELSE{

 outputKeywords += only direct sub concepts of concept C; }

 IF (useSubstringMatch){

 otherConcepts = search for ontology concepts whose name is a super-string of K's string value

 FOR EACH concept D in otherConcepts DO{

 IF (distance(D,K) <= threshold){

 IF (allFlag){

 outputKeywords += all descendants of concept D

 }ELSE{

 outputKeywords += only direct sub concepts of concept D;

 }

 }

 outputKeywords += D

 }

 }

 }

 }

 }

 }

return outputKeywords;