Semi-Automated Annotation of Epidemiological Resources

Catia Pesquita1,3, João D. Ferreira1,3, Francisco M. Couto1,3, Mário J. Silva2,3
1LASIGE, 2INESC-ID, 3University of Lisbon

Motivation:
The Epidemic Marketplace (EM) [1] (www.epimarketplace.net) is a platform that enables the sharing of resources and knowledge within the Epidemiology community with a strong focus on the semantic annotation of resources. On uploading their resources to the EM, users can provide an accurate semantic annotation based on metadata and a network of epidemiology related ontologies (NERO)[2,3]. However, this process can be time consuming, since a single epidemiological resource can refer to several diseases, symptoms, locations, etc.

Network of Epidemiology Related Ontologies (NERO)

Proposal:
In order to require minimal human intervention, we will develop a semi-automated annotation module for the EM, which automatically identifies terms of NERO in a given text-based resource and suggests them as a default characterization of the resource. This will be based on structured machine learning [4] and multi-ontology semantic similarity [5] methods, and benefit from the over 100 fully annotated resources already available in the EM. The candidate annotations can later be verified by the users. This will cut the time and effort needed to provide a complete semantic annotation for research papers and other text-based epidemiological resources, effectively encouraging users to contribute with more resources and provide richer annotations.

Semi-Automated Annotation Strategy

1. ML-based entity recognition

The “hygiene hypothesis” postulates that infections during infancy may protect against asthma and atopy. There is also some evidence that antibiotic and/or paracetamol use may increase the risk of asthma. [...] Results: There was little difference in the prevalence of current wheezing between the childhood infections group (prevalence = 23.5%) and the general population group (prevalence = 24.3%).

2. ML-based entity resolution to NERO ontologies

The “hygiene hypothesis” postulates that infections during infancy may protect against asthma and atopy. There is also some evidence that antibiotic and/or paracetamol use may increase the risk of asthma. [...] Results: There was little difference in the prevalence of current wheezing between the childhood infections group (prevalence = 23.5%) and the general population group (prevalence = 24.3%).

3. Multi-Ontology semantic similarity validation