

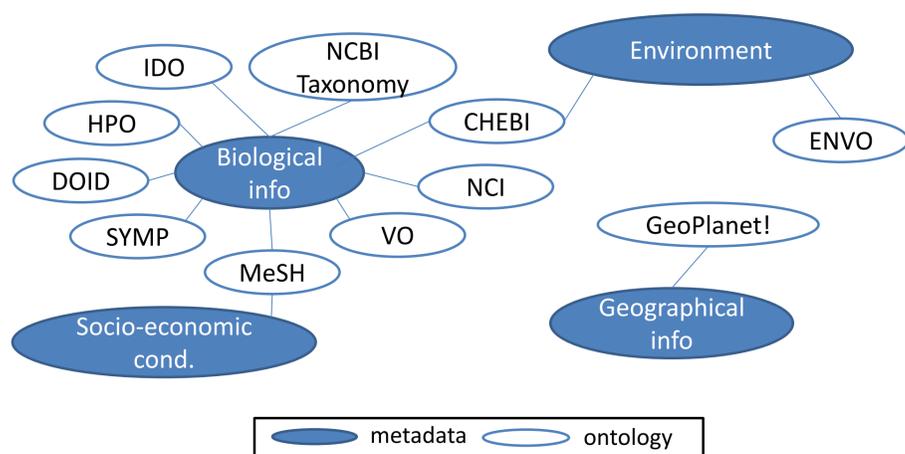
Semi-Automated Annotation of Epidemiological Resources

Catia Pesquita^{1,3}, João D. Ferreira^{1,3}, Francisco M. Couto^{1,3}, Mário J. Silva^{2,3}
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.

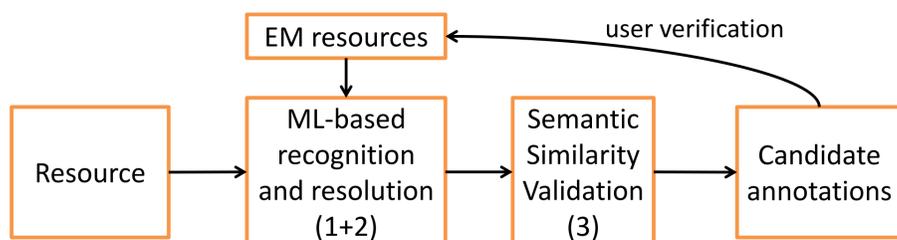
Network of Epidemiology Related Ontologies (NERO)



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] Couto, F. M., Ferreira, J. D., Zamite, J., Santos, C., Posse, T., Graça, P., Domingos, D., & Silva, M. J. (2012). The Epidemic Marketplace Platform: towards semantic characterization of epidemiological resources using biomedical ontologies. In Proc. ICBO
- [2] Ferreira, J. D., Pesquita, C., Couto, F. M., & Silva, M. J. (2012). Bringing epidemiology into the Semantic Web. In Proc. ICBO
- [3] Ferreira, J. D., Paolotti, D., Couto, F. M., & Silva, M. J. (2012). On the usefulness of ontologies in epidemiology research and practice. J EPIDEMIOLOG COMMUN H
- [4] Grego T, Couto FM (2013) Enhancement of Chemical Entity Identification in Text Using Semantic Similarity Validation. PLoS ONE 8(5): e62984.
- [5] Ferreira, J. D., & Couto, F. M. (2011). Generic semantic relatedness measure for biomedical ontologies In Proc. ICBO

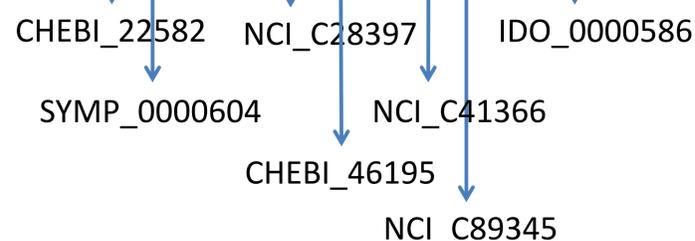
1. ML-based entity recognition

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%).

Legend: ■ disease ■ drug ■ symptom ■ demography

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

