A cross lingual semantic search engine dedicated to medical teaching resources

Griffon N, Darmoni SJ

CISMeF, Rouen University Hospital, Normandy & TIBS, LITIS EA 4108, Institute of Biomedical Research, Rouen University, France

Background: The Catalogue of French-speaking health resources available on the Internet (French acronym: CISMeF) is a portal that gathers French medical Web resources (n≈90,000). Doc'UMVF is a semantic search engine that allows searching in this database but filtering only medical teaching materials (n=9,697). Currently, Doc'UMVF is based on multi-terminology indexing, consequently, it works with the French (n≈500,000) and English terms (n≈1,000,000) of the 32 terminologies included in information system (IS). Using terminologies translated in other language than only French and English would allow cross lingual indexing and information retrieval of teaching material.

Objective: To make Doc'UMVF working in several languages.

Method: Some terminologies integrated in CISMeF IS have been translated in numerous languages (e.g. International classification of disease, tenth revision, is available in 11 languages, Medical Subject Heading in 16 languages, etc). An enhanced version of the IS was necessary to manage multiple translations and Doc'UMVF has to be adapted to this new IS.

Results: HeTOP is a portal that allows cross lingual access to multiple terminologies. A semantic search engine based on such tool could be useful in at least two use-cases. First, for resources that exist in numerous languages, like European Medical Agency (EMA) resources, there is only need for one indexing in one language to allow their retrieval in many European languages. Second, this provides a language independent tool: any teacher, whatever his speaking language was, may index his own teaching material in his own language and any student, whatever his speaking language was, may find it.

Conclusions: Doc'UMVF will soon work in the new IS. Nevertheless, its natural language processing tools work only in French, therefore its use should be reserved to advanced user (those who know terminologies). There is not too much development to achieve an operational language independent search engine dedicated to teaching material.