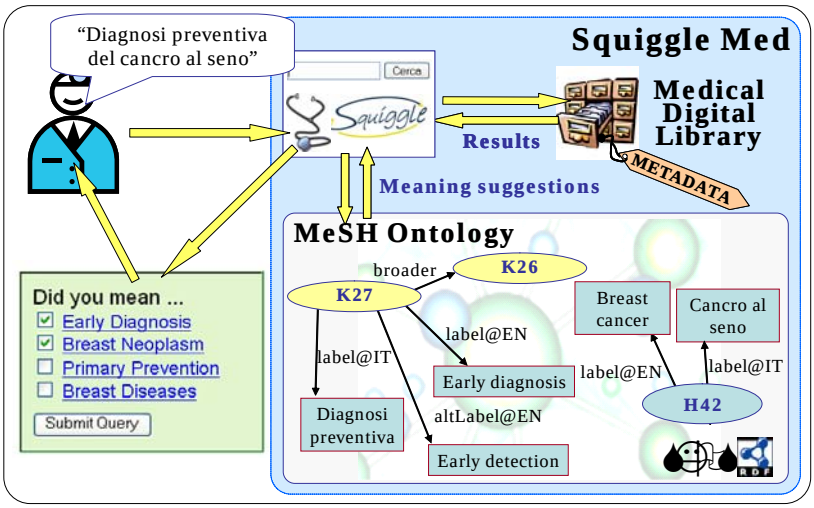
Squiggle Med: Semantic Search for Medical Digital Library

I. Celino, A. Turati, E. Della Valle, D. Cerizza
CEFRIEL – Politecnico di Milano, via Fucini, 2 20133 Milano, Italy
{irene.celino, andrea.turati, emanuele.dellavalle, dario.cerizza}@cefriel.it

I. Introduction

Searching everything everywhere is becoming our habit when we need to find something. However, *finding* what we need is often a hard job. Current search engine technology is very good in finding complete Web pages, but it lacks the desired precision and recall when searching for multi-lingual resources such as within pan-european Medical Digital Libraries. For instance, searching "Diagnosi preventiva del cancro al seno" (which in Italian means "early diagnosis of breast cancer") in a syntactic search engine results in a subset of the relevant document, those written in Italian.



Squiggle [1] is an extensible semantic search framework for the development of semantic search engines. By adding a conceptual flavour to the crawling and the indexing of resources, Squiggle can exploit ontological elements to improve and enrich searching time, without undermining the user experience. These features, together with the employment of SKOS [2] model, make Squiggle a powerful and reusable framework to build engines with both syntactic and semantic functionalities.

II. Searching with Squiggle Med

The interaction of a final user with Squiggle is intuitive and very similar to the use of a traditional search engine; however, the results are better and more meaningful. In the following we provide examples of searches with Squiggle Med (a demonstrative semantic search engine for medical digital library) that uses the SKOS version of MeSH [3]

1st step: Syntactic search and query analysis

The user is presented with a simple search form in which he can insert some keywords. Squiggle, like syntactic search engines, can immediately retrieve all results containing those keywords. For example, a General Practitioner could search for "Diagnosi preventiva del cancro al seno"; and the user not only obtains syntactically-matching results (documents containing the searched keywords), but his query is also analyzed in order to identify its **meaning**. This is possible because of Squiggle *disambiguation* capabilities: the search engine can access an ontology of the domain (i.e., the SKOS version of MeSH) and try to identify the concepts that could have some connections with the query (in the previous example: early diagnosis, breast neoplasm, primary prevention, breast diseases).

2nd step: Semantic search

The results of the previous *disambiguation* phase are displayed, together with the syntactic results, in a lateral box under the heading "Did you mean...?". Therefore, the user can manually choose the *meaning* of his query among the proposed ones. In response, he obtains more precise and numerous results; this is possible because, during the contents indexing, Squiggle is able to *semantically annotate* those resources with regards to the domain ontology. In the previous example, during the *conceptual indexing* phase, all the articles in the Medical Digital Library whose syntactic annotations contained both "breast neoplasm" (the concept's `skos:prefLabel`) and his Italian version "cancro al seno" (represented with a `skos:altLabel` relation) were annotated with MeSH *concept* "breast neoplasm".

3rd step: Semantic suggestions

But there is more: accessing the domain ontology, Squiggle is able to exploit all its content, i.e. not only the alternative labels of its concepts, but also their relations. This capability lets Squiggle expand the user query, by following the relations between the concepts identified in it and other ontological elements, and propose to the user possible searches of his interest. In the previous example, the General Practitioner, could be presented with a lateral box suggesting an expansion of his query to include results related to "complication" (which is in relation with "breast neoplasm" by a `mesh:qualifier` which is an RDF sub-property of `skos:broader` in MeSH). This *meaning suggestion*, as well as the disambiguation phase described before, are possible because Squiggle has access to a domain ontology.

III. Conclusions

We believe that the employment of Semantic Web technologies in the development of search engines provides real benefits to end users. In the examples the contents are those in PubMed [4], but a semantic search engine like Squiggle Med appears at a glace to be more usable, in that users are supported with semantic "suggestions".

References

[1] CEFRIEL Semantic Web Activities group. Squiggle home. http://squiggle.cefriel.it/ , 2005.
[2] A. Miles and D. Brickley. SKOS Core Guide, W3C Working Draft. http://www.w3.org/TR/swbp-skos-core-guide, 2 November 2005.
[3] Mark van Assem, Véronique Malaisé, Alistair Miles, Guus Schreiber: A Method to Convert Thesauri to SKOS. ESWC 2006: 95-109
[4] PubMed, a service of National Library of Medicine and the National Institute of Health. http://www.pubmed.gov