Alternative interfaces for PubMed searches
Isabelle de Kaevel, Pablo Irriarte
Medical Library, University Hospital; CHUV BH-08, CH-1011 Lausanne, Switzerland
isabelle.de-kaenel@chuv.ch; pablo.irriarte@chuv.ch

Introduction: Over the last 5 years, more than 30 new search interfaces for MEDLINE/PubMed have been made publicly accessible. This wave surged in the early 2000’s with the launch of the PubMed web services, the « Entrez Programming Utilities ». Along with the traditional commercial distributors (Ovid, CSA, Dialog and many others), academic institutions, bioinformatics services, pharmaceutical enterprises and software companies leased the MEDLINE/PubMed data for research purpose and devised alternative interfaces. These applications draw on data mining techniques, linguistic analysis, knowledge management resources, statistical methods, and new web technologies. The new search tools were examined by reference librarians at the Lausanne University Hospital, in order to decide whether they should be used as alternate search engines for specific requests and even be introduced in training sessions for students and clinicians.

NLM Mission: To collect, organise, and disseminate the world’s health-related and biomedical information

Method: To get an overview of these different retrieval systems, literature and web searches were conducted. The sites listed in the « NLM’s registry of licensees who use MEDLINE/PubMed data for research purposes » and in the MEDLINE section of the « Metadatabase for the biological sciences » on Neurotransmitter.net were also systematically explored. The main innovations of these interfaces where then classified along specific criteria and functions.

Information retrieval

• Graphical search interface: PubMed Interact presents slider bars to control the traditional PubMed search limits and parameters.
• Free text / natural language query tool: AskMEDLINE processes clinical questions written in plain English.
• Guided search form: PICO Linguist, a multilingual interface, helps in building a structured clinical query by using the PICO framework.
• Cross-language tool: queries can be submitted to BabelMeSH in different languages. Search terms are then mapped to a multilingual MeSH.
• Similarity quest: the input of at least 8 PMID allows Pubfinder to return a hit list of references in order of relevance.
• Relevance ranking: ReleMed defines & levels of «relevance» to sort the records in a descending order based on the estimated scores.
• Graphical display and concept maps: AllBabes provides a graphical and interactive view of the extracted information.
• Frequency tables: PubReamer generates hyperlinked tables after statistical analysis of the result sets.
• Clusters: MESHPubMed or ClusterMed dynamically create clusters of references classified by topic, author, MeSH headings or date.
• Tag-clouds: iHOP produces tag clouds based on MESH term. The size of each cloud is proportionate to the frequency of occurrence of the keyword.

Knowledge retrieval

• Analysis of word co-occurrence: XolorMed parses abstracts and calculates a degree of relatedness between words.
• Entity recognition: using the Gene Ontology, GoPubMed identifies relevant biological concepts and shows possible hidden correlations between genes, drugs, and diseases.
• Text & data mining: LitMiner annotates key terms in abstracts and predicts relationships between key biomedical terms in 4 categories: genes, chemical compounds, diseases and organs.
• Information extraction: AliBabes scans abstracts for biological objects and creates associations with descriptions of the relationships.
• Semantic networking tools: PubFocus performs terminology extraction and couples the information with external parameters.

Screen

Save & structure

Conclusion: These new interfaces address many different needs and user profiles: clinicians working in private practices in Europe could use the multilingual tools such as BabelMeSH and PICO Linguist; researchers that need to scan vast amount of literature will turn to advanced methods that mine information and bring out hidden relations and associations. For the students, they may be attracted by social tagging and web 2.0 technologies also used in recreational web sites. In short, these new tools have to be introduced in training sessions, at least to appreciate the interest they raise in the different groups of users. These developments threat the transition from « resource centric » to « user centric » systems, and from « information retrieval » to « knowledge retrieval ».