

Collexis: Comparison between the Search Options

Overview: Conceptual vs. Boolean vs. Last Name Searching

Below is an overview of the difference between thesauri searching via the Concept search box and general keyword searching via the Boolean search box.

Concept Search- If you wanted to see who is studying breast cancer at UM, you can search with those terms in the Concept search box. Collexis will suggest the official term from the MeSH thesaurus for breast cancer, i.e. breast neoplasms. PubMed articles, NIH-funded projects in CRISP, and CV information added by a faculty member are automatically indexed by terms in the MeSH thesaurus, so using the Concept search box is the best choice for locating potential collaborators whose grants and publications are in PubMed or CRISP.

In addition to searching by MeSH terms, the Concept search understands the context of your search and will automatically include synonyms and discriminate between the different meanings of homonyms. For example, a search for the word "oil" will include the words "gas," "petroleum," "petrol," and "diesel." Since Collexis discriminates between the different meanings of homonyms, searching for "mass" will get results about nuclear weapons rather than the time of a church ceremony.

A thesaurus is a collection of terms that are defined as either synonyms (e.g. "doctor" and "physician") or as having broader or narrower meanings relative to other terms (e.g. "dog" is narrower than "pet" but broader than "Golden Retriever").

Collexis makes vigorous use of thesauri to address both the synonym and homonym problems, significantly reducing false negatives and false positives. By looking up related terms in thesauri and searching for these as well, we find terms similar in meaning, even when our queries contain different but related terms. Thus a search for "car" can also find documents with "auto," "automobile," "station wagon," "SUV," and "van." This improves search recall by finding more relevant documents.

Lookup in the thesaurus also provides semantic clues to disambiguate terms. For example, if Collexis finds the word "church" in the same sentence as the word "mass," it concludes that it has a religious meaning. If it finds "weapon" or "kill" near the word "mass," then it concludes that the document is about a means of attack. Disambiguation increases the precision of our search results by returning a higher percentage of documents with relevant meanings.

Terms in the thesaurus are also defined hierarchically (broader/narrower), as well as synonymously. Collexis further improves search precision by *ranking documents with more specific terms higher*. For example, if you were a medical professional researching cancer-causing substances, which document would you rank higher: One that mentions "cancer" eight times but never uses more specific terms; or another that uses "cancer" three times and also mentions "sarcoma," "Hodgkin's," and "metastasis"?

Collexis: Comparison between the Search Options

Boolean Search- If Collexis tells you there are no thesauri terms for your topic, or if the suggested thesauri terms are not relevant, you can do a keyword search on your topic by using the Boolean search box. For example, “Miami-Dade county” is not in the MeSH Thesaurus, so you will get zero hits in the Concept search box. However, searching with this topic in the Boolean search box will retrieve UM investigators and their works that mention this topic.

You can further refine your Boolean search box hits to results from the PubMed, CRISP, or CV sources. UM created the “CV Tab” for users to enter their professional and personal achievements not reflected in PubMed publications or CRISP grants. The information in the CV is indexed with MeSH terms and keywords, so the graphical display, “the fingerprint,” shows terms from the MeSH thesaurus and keywords. (The fingerprint for the faculty member’s PubMed and CRISP sections illustrate MeSH terms only.) Using the Boolean search box will allow you to search the curriculum vitae of UM faculty and thus help you to identify those faculty that are doing work in this area. Lastly, the Boolean search box will allow you to join together concepts with AND, OR, or NOT.

Last Name Search- With conceptual searching, Collexis is able disambiguate between synonyms, acronyms, and homonyms. Names are the same. Collexis disambiguates the names by taking the concepts around those names into consideration, similar to looking for the concept “jaguar.” How does one differentiate when discussing the animal or the car? If the concepts around the term are discussing V8 engines and leather seats, one is most likely discussing the car. If the concepts around the term are discussing cats native to South and Central America, one can identify that term with the animal.

With John S. Smith and John M. Smith, Collexis can identify that they are two separate people based on their areas of research. UM supplies Collexis with the names of UM faculty members and the Collexis software automatically adds the PubMed publications and NIH-funded grants to their Collexis profiles. It’s not a complete list of publications, since not all articles are indexed in PubMed and it doesn’t include non-NIH-funded grants. However, faculty members can add in that information to make their profiles more robust by inputting curriculum vitae information. There are two options for making or requesting corrections or additions to your Collexis Researcher Profile:

<http://www.researchprofiles.collexis.com/miami/feedback.asp> The profile will be updated by Collexis staff in about a week.