Overview

The iseek medical product provides a searchable index of a medical school's curriculum content (Files, Videos, Events, etc.).

For some queries it is helpful to return resources beyond those of the school alone. Iseek maintains a separate index of medical journals to allow for this broader search.

Customers have expressed an interest in iseek adding the Cochrane Library as a new source to the medical journal index.
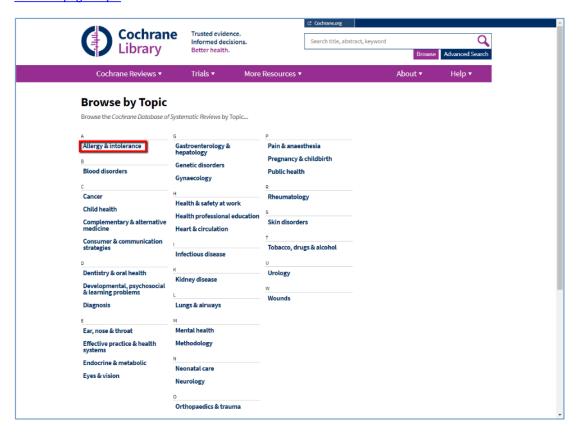
Project

We must build a Java command line application to gather and store URLs and associated metadata from the Cochrane Library review collection (https://www.cochranelibrary.com/cdsr/reviews/topics).
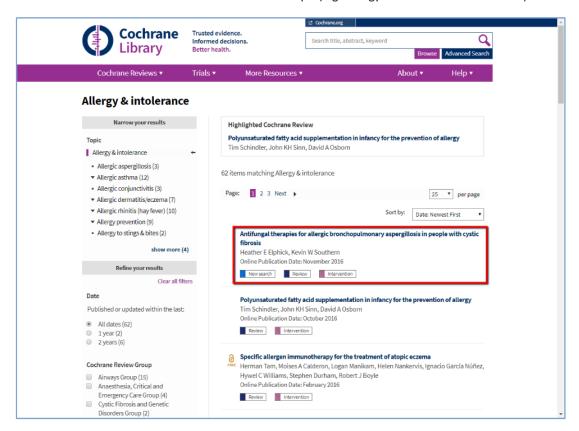
Tools

- Required: JDK 1.7+
- Required: Apache HttpClient 4.5+
- Optional (dependency management): Maven 3.3+
- Add any supporting libraries you deem necessary (not crawlers, of course).

Input

The library may be browsed by Topic: http://www.cochranelibrary.com/home/topic-and-review-group-list.html?page=topic

You must discover ***all*** of the reviews for at least one topic (e.g. Allergy & intolerance – **62** reviews).

Programmatically visit each page of the topic browse results and collect the following data for each review:

URL - http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD002204.pub4/full

Topic - Allergy & intolerance

Title - Antifungal therapies for allergic bronchopulmonary aspergillosis in people with cystic fibrosis

Author - Heather E Elphick, Kevin W Southern

Date - 2016-11-08

Output

Write the collected data to a single text file (e.g. cochrane_reviews.txt). This file must use pipes "|" to delimit the data. The reviews are separated by newlines. For example, the first two reviews:

http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD002204.pub4/full|Allergy & intolerance|Antifungal therapies for allergic bronchopulmonary aspergillosis in people with cystic fibrosis|Heather E Elphick, Kevin W Southern|2016-11-08

http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD010112.pub2/full|Allergy & intolerance|Polyunsaturated fatty acid supplementation in infancy for the prevention of allergy|Tim Schindler, John KH Sinn, David A Osborn| 2016-10-28

Documentation

Please write a one or two page document that describes your solution to the technical lead of a potential customer. The reader should understand:

- The reasoning behind your design.
- High-level flow of application logic.
- Any additional features you found on the Cochrane Library site that might be interesting to include in the injector file.

If images or other graphics are included, the page limit may be increased beyond two pages to accommodate.