

## Overview

The isseek 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 isseek 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>

The screenshot shows the Cochrane Library website's 'Browse by Topic' page. The header includes the Cochrane Library logo, the tagline 'Trusted evidence. Informed decisions. Better health.', and a search bar. Below the header is a navigation bar with links to 'Cochrane Reviews', 'Trials', 'More Resources', 'About', and 'Help'. The main content area is titled 'Browse by Topic' and includes the subtitle 'Browse the Cochrane Database of Systematic Reviews by Topic...'. The topics are organized into three columns, each headed by a letter. The first column (A) includes 'Allergy & intolerance' (highlighted with a red box), 'Blood disorders', 'Cancer', 'Child health', 'Complementary & alternative medicine', 'Consumer & communication strategies', 'Dentistry & oral health', 'Developmental, psychosocial & learning problems', 'Diagnosis', 'Ear, nose & throat', 'Effective practice & health systems', 'Endocrine & metabolic', and 'Eyes & vision'. The second column (G) includes 'Gastroenterology & hepatology', 'Genetic disorders', 'Gynaecology', 'Health & safety at work', 'Health professional education', 'Heart & circulation', 'Infectious disease', 'Kidney disease', 'Lungs & airways', 'Mental health', 'Methodology', 'Neonatal care', 'Neurology', and 'Orthopaedics & trauma'. The third column (P) includes 'Pain & anaesthesia', 'Pregnancy & childbirth', 'Public health', 'Rheumatology', 'Skin disorders', 'Tobacco, drugs & alcohol', 'Urology', and 'Wounds'.

A	G	P
Allergy & intolerance	Gastroenterology & hepatology	Pain & anaesthesia
Blood disorders	Genetic disorders	Pregnancy & childbirth
Cancer	Gynaecology	Public health
Child health	Health & safety at work	Rheumatology
Complementary & alternative medicine	Health professional education	Skin disorders
Consumer & communication strategies	Heart & circulation	Tobacco, drugs & alcohol
Dentistry & oral health	Infectious disease	Urology
Developmental, psychosocial & learning problems	Kidney disease	Wounds
Diagnosis	Lungs & airways	
Ear, nose & throat	Mental health	
Effective practice & health systems	Methodology	
Endocrine & metabolic	Neonatal care	
Eyes & vision	Neurology	
	Orthopaedics & trauma	

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

The screenshot shows the Cochrane Library website interface. At the top, there is a search bar with the text 'Search title, abstract, keyword' and buttons for 'Browse' and 'Advanced Search'. Below the search bar is a navigation menu with 'Cochrane Reviews', 'Trials', 'More Resources', 'About', and 'Help'. The main content area is titled 'Allergy & intolerance'. On the left, there is a sidebar with 'Narrow your results' and 'Refine your results' sections. The 'Narrow your results' section shows a list of topics under 'Allergy & intolerance', including 'Allergic aspergillosis (3)', 'Allergic asthma (12)', 'Allergic conjunctivitis (3)', 'Allergic dermatitis/eczema (7)', 'Allergic rhinitis (hay fever) (10)', 'Allergy prevention (9)', and 'Allergy to stings & bites (2)'. The 'Refine your results' section shows 'Date' filters: 'All dates (62)', '1 year (2)', and '2 years (6)'. The 'Cochrane Review Group' section shows 'Airways Group (15)', 'Anaesthesia, Critical and Emergency Care Group (4)', 'Cystic Fibrosis and Genetic Disorders Group (2)', and 'show more (4)'. The main content area displays 'Highlighted Cochrane Review' results. The first result is 'Polyunsaturated fatty acid supplementation in infancy for the prevention of allergy' by Tim Schindler, John KH Sinn, David A Osborn. Below this, it says '62 items matching Allergy & intolerance'. There are pagination controls showing 'Page: 1 2 3 Next' and '25 per page'. A 'Sort by:' dropdown menu is set to 'Date: Newest First'. The second result is 'Antifungal therapies for allergic bronchopulmonary aspergillosis in people with cystic fibrosis' by Heather E Elphick, Kevin W Southern, with an 'Online Publication Date: November 2016'. This result is highlighted with a red box. Below it are buttons for 'New search', 'Review', and 'Intervention'. The third result is 'Polyunsaturated fatty acid supplementation in infancy for the prevention of allergy' by Tim Schindler, John KH Sinn, David A Osborn, with an 'Online Publication Date: October 2016'. Below it are buttons for 'Review' and 'Intervention'. The fourth result is 'Specific allergen immunotherapy for the treatment of atopic eczema' by Herman Tam, Moises A Calderon, Logan Manikam, Helen Nankervis, Ignacio García Núñez, Hywel C Williams, Stephen Durham, Robert J Boyle, with an 'Online Publication Date: February 2016'. Below it are buttons for 'Review' and 'Intervention'.

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.