

User manual: EERQI search and query engine

Introduction.....	1
Quick introduction.....	1
List of index fields.....	1
Indexes.....	2
Index overview.....	3
Basic usage.....	3
Query syntax.....	5
Lucene query syntax.....	6

History

Last change of this document

February 11, 2010

Introduction

The EERQI search engine is used for searching for keywords within a selection of the World Wide Web or within the local document corpus from the educational research publishers.

Quick introduction

It is possible to use a keyword search within the document content and combine this with a query within specific fields using boolean operators (AND, OR, NOT).

Example:

content:education AND title:school

searches for documents, containing the term “education” in the content and the term “school” in the title.

List of index fields

For expert queries (see description of Lucene query syntax below), it is necessary to know the names of the fields in the index. For the naming of the fields, current naming conventions within Dublin Core have been used. See for example

<http://www.intute.ac.uk/publications/eprints-uk/simpledc-guidelines.html> for more details.

index field	Corresponding Dublin Core element	Description	Example value
language	dc:language	Language of fulltext document	en for english sv for swedish fr for french de for German
pages		Number of pages, filled with 0 to length 5.	00080
format	dc:format	Mimetype of document	application/pdf, pdf

		format	
url	dc:identifier	URL of document	
content		Entire content of document	
creator	dc:creator	Author name or institution	Sandrine Aeby
creator.surname	eerqi:surname	Author surname	Aeby
title	dc:title	Title of document. If not specified, this is automatically determined, for example from the HTML title and is not always correct.	
subject	dc:subject	Keywords or for example DDC classification	
description	dc:description or dc:abstract	Abstract of document	
publisher	dc:publisher	Name and / or URL of publisher	
journal		Name and / or URL and / or ISSN of journal	
rights	dc:rights	Rights of fulltext document	“Access limited to EERQI members”
date	dc:date	Date of publication, usually just the year.	2004

Indexes

An index is used for querying the document corpus. Specifically, the following indices are now available:

1. ***Content index*** : The *content index* only contains the content of the documents and no additional metadata. The access of the fulltexts is restricted to EERQI members only.
2. ***Metadata index***: This index contains metadata as well as full texts. The access of the fulltexts is restricted to EERQI members only.

The *metadata* contains the content of the documents and additional metadata (for example title, creator etc.). It is possible to access the metadata via the corresponding fields in the index (see description of fields in index above). For index field names the Dublin Core element names have been used, wherever available.

3. ***WWW index***: A list of URLs has been crawled and the corresponding webpages indexed. No metadata has been made available here (except for the title as extracted from the HTML page and the language).
4. ***WWW index (+metadata)***: Fulltext files and corresponding metadata have been indexed wherever available.

Index overview

	Includes metadata	No metadata
Restricted access	<i>Metadata index</i>	<i>Content index</i>
Public access	<i>WWW index (+metadata)</i>	<i>WWW index</i>

Basic usage



The screenshot shows the EERQI search interface. At the top, there is a header with the logo 'EERQI' and the text 'EUROPEAN EDUCATIONAL RESEARCH Quality Indicators'. Below the header is a search bar with a 'Search' button. Underneath the search bar, there are four checkboxes, each with a label and a description in parentheses:

- [content index](#) (restricted documents)
- [metadata index](#) (restricted documents)
- [WWW index](#) (public documents)
- [WWW index \(+metadata\)](#) (public documents)

A red oval highlights the 'metadata index' and 'WWW index (+metadata)' options.

Figure 1: Select one or more indices

EUROPEAN EDUCATIONAL RESEARCH
Quality Indicators
EERQI

Search query

Search button

edu*

[content index](#) [metadata index](#) (restricted documents)

[WWW index](#) [WWW index \(+metadata\)](#) (public documents)

Figure 2: Enter a search query and press the search button.

document granularity page granularity

Figure 3: Choose document or page granularity

The screenshot displays the EERQI search interface. At the top, there is a header with the EERQI logo and the text 'EUROPEAN EDUCATIONAL RESEARCH Quality Indicators'. Below the header, there are logos for the European Union and the Seventh Framework Programme, along with the text 'A PROJECT FUNDED UNDER THE SOCIO-ECONOMIC SCIENCES AND HUMANITIES THEME'.

The search bar contains the query 'edu*' and a 'Search' button. To the right of the search bar are links for 'Help', 'Lucene Query Syntax', and 'Simple mode'.

Below the search bar, there are two rows of checkboxes for selecting search options:

- [content index](#) [metadata index](#) (restricted documents) Show all metadata
- [WWW index](#) [WWW index \(+metadata\)](#) (public documents)

Below these options, there are two radio buttons for document granularity: document granularity and page granularity.

The current query is displayed as: 'Current Query: creator:edu* title:edu* content:edu* description:edu*'

The search results are shown as '1 - 10 results of 1736: [1](#) [11](#) [21](#) [31](#) [41](#) ... [891](#) [>>](#) [>>>](#)'.

The first result is displayed as follows:

Result 1)

Title: Educating Elites in Democratic Societies: a dialogue
Creator: Joseph Agassi, Ronald Swartz
Journal: pfi
Publisher: [Symposium Journals a division of Wwwwords Ltd](#)
Date: 2007
Description (Abstract): This dialogue centers on the following questions: (1) How can schools help a society select or identify new elites who are hopefully as good as and pe ...
URL: https://eergj.hu-berlin.de/fulltexts/symposium/1478-2103/5/4/pfi_2007_5_4_424.pdf
Language: en (English)
Rights: Access limited to EERQI members

Below the result, there is a snippet of text from the document: 'education?' The **educational** problem of basic learning is very different from the **educational** problem ... scorned by influential **educators** such as school superintendents and mainstream **educational** theorists ... **educational** proposals which clearly reject most of what Dewey had to say about **education**. And after I ... **education** for witch doctors and priests; these early schools were clearly meant to **educate** those who ... Socrates and his failure to **educate** the sophists, who were the **educational** elite establishment during the

Figure 4: Search results.

The current query is highlighted in the displayed results, within appropriate text snippets from the content. In case of the *metadata index*, additional information such as title and creator (author) is displayed. A maximum of 10 results is displayed on the page. If there are more than 10 results, it is possible to skip to one of the next pages:

- [1](#) : results 1 – 10
- [11](#): results 11 – 20
- [21](#): results 21 – 30
- [>>](#) : next page
- [>>>](#) : last page
- [<<](#) : previous page
- [<<<](#) : first page

Query syntax

To query the index, [Lucene query syntax](#) must be used. The query syntax is explained in detail at http://lucene.apache.org/java/2_3_2/queryparsersyntax.html

A short overview is now given:

Lucene query syntax

The words used in the query are not case sensitive. Thus, it makes no difference if you query for **any** or **aNy**. It is possible to search simply for one or more words (e.g. **educational**) or phrases (**"educational research"**) or combine these with AND or OR. By default, if you search, for example for **educational research**, the boolean operator OR is used implicitly. This means all documents are displayed that contain the word educational or research. If you wish to search for all documents, containing both words, the AND operator (or the plus prefix) must be used: **educational AND research**.

Query	Explanation	Example
Phrase query	Search for a specific phrase within ""	educational research (searches for the exact phrase "educational research")
Wildcard query	* - means several characters ? - means one character It is not possible to begin a term with a wildcard.	edu*n (searches for words that begin with edu and end with n)
Field query	Search within a specific field. Specify the field name, followed by a colon (:) and then the search term	author:a* (searches for all authors beginning with A or a)
Range query	Search within a range, can be a date or a string. [] - inclusive range query { } - exclusive range query	author:[a* TO m*] (searches for authors beginning with A to M)
Fuzzy query	Search for similar terms by using the ~ symbol at the end of a term	educatoin~ (will find text containing education, educators etc. in spite of the typo)
Proximity query	When searching for more than one terms, one can specify that they must be within a specific distance of each other.	Content:"educational research"~10 (searches for the words educational research within 10 words of each other)

Table 1: Description of Lucene query syntax.