

Full Text Searches in E-books. Crossing the Borders of Publishing Houses

René [Deplanque](#)
FIZ CHEMIE Berlin, Germany



ICIC
The International Conference on Trends for Scientific Information Professionals
Nice 19-22 October 2008

Definition of an E- Book?

- General Definition: Books in digital form
- Solves Problems like: Transfer of data between book and database, retro-digitalisation etc.
- Name similar for hardware (readers) and for software (reading material)

The development of the E-Books-market

- First books required specific e-book reader
- Starting in the late 1990ies e-book providers offered on-line books (e.g. NetLibrary starting from 1998, Ciando 2001)
- Today's e-books market: amounts to approx. 1% of the total German book market (USA: approx. 9%) *
- The e-books market comprises: Publishing houses, providers/aggregators, agencies, booksellers, libraries, Google etc.

•Association of trade of the German book trade after Peter Just: E-Books for libraries. An inventory analysis. Berlin 2006

Distributor of E-Books

- Publishing houses:
 - Parallel offer of printed books or original e-books/e-book collections, complete publishing house programs or technical packages
- Provider/Aggregator:
 - Database producers (e.g. Ovid, Gale) or special providers for e-books (e.g. Ciando, NetLibrary, ebrary). Platforms for individual e-books of different publishing houses
- On-line bookstores (e.g. Libri, Amazon)
- Distributers (e.g. MISSING Link) and agencies (e.g. Swets, EBSCO)

The problems of libraries with the E-Books-market

- Additional costs for print copies, because of unsolved long-term archiving
- No reduced tax rate
- Partially obscure price and discount structure
- New purchase of new editions often without discounts
- Partly no purchase of the books possible, but only licensing without archive data
- Favorable consortia model is missing (content sharing)

Suggestions of improvement from a library point of view

- Flexible publishing house offers to favorable conditions (packages and single titles; purchase and license models; simultaneous accesses, download and check-out counter)
- Co-operation of the providers, agencies and booksellers to help the libraries in the acquisition of E-Books
- Consortia models, which have flexible and progressive discount rates, content sharing and CROSS ACCESS
- Own platforms and archiving tools for libraries/library groups (at the same time entering of titles into the library catalogs)

Full text search engine for eBooks Library Questions

**Linking of OPAC directly to the book,
quotation ability of links and
searching of links in external search
engines**

Possible with special software tools

**Do you try to initiate a listing of
available books/books in print for e-
books? How well does this cover the
existing offer?**

Yes and if possible complete

**No restrictions temporal or regarding
the number of simultaneous users**

Yes

**Storing on PC of the user; insert of
notes to the own (downloaded) copy;
also passing comments to third
parties; copy for working groups**

Yes

**But in dependence of the licenses
with the publishing house**

Full text search engine for eBooks Library Questions

What can FCH offer compared to agencies or aggregator within the e-book business?

Background

FCH is working for approx. 10 years with search engines

Libraries wish a better use of licensed e-books.

Users wish a problem oriented usage of e-books

Solution

Full text search engine - data base over all e-books

publishing house-independent

with keyword in context

combination of full texts and meta data

Interface to the search only in licensed books

Full text search engine for eBooks Library Questions

What does your business model look like; do you have a variable acquisition model??

Publishing house pays FCH:

Fee per e-book (eg. 2\$/year)

FCH receives commission for book sales

Library pays FCH: Nothing but:

For using a customized interface for industry and libraries a commission has to be paid.

Price in dependence of the number of licensed books (1\$/book/year)

Advertizing can be placed only by publishing houses involved

Full text search engine for eBooks Library Questions

Which interfaces are made available for (possibly) search, linking, supply and resources management (e.g. SFX, Verde)

Searches are generally via all publishing houses.

Linking and resources management can be made available individually

Full text search engine for eBooks Library Questions

Questions concerning the use and the functionalities:
No different user interfaces for different publishing houses
(independent of publishing house)

- Full text data base system of all e-books with keyword in context display.
- Combination of full text and meta data.
- Special interface e.g. for libraries to search in licensed books only
- E-books remain with the publishing house, index files with FCH.

Full text search engine for eBooks Procedure

Will there be different offers for end users, libraries and agencies? (regarding price, in addition to functionalities)

Publishing house transfers books to FCH in pdf format incl. meta data

FCH converts the data into formats needed and operates the data base

Application is arranged in cooperation with the publishing houses

Search engine is independent of publishing house, it can be made to order

Full text search engine for e-books

Search options

Boolean operators

Basic proximity, e.g. "organic chemistry"

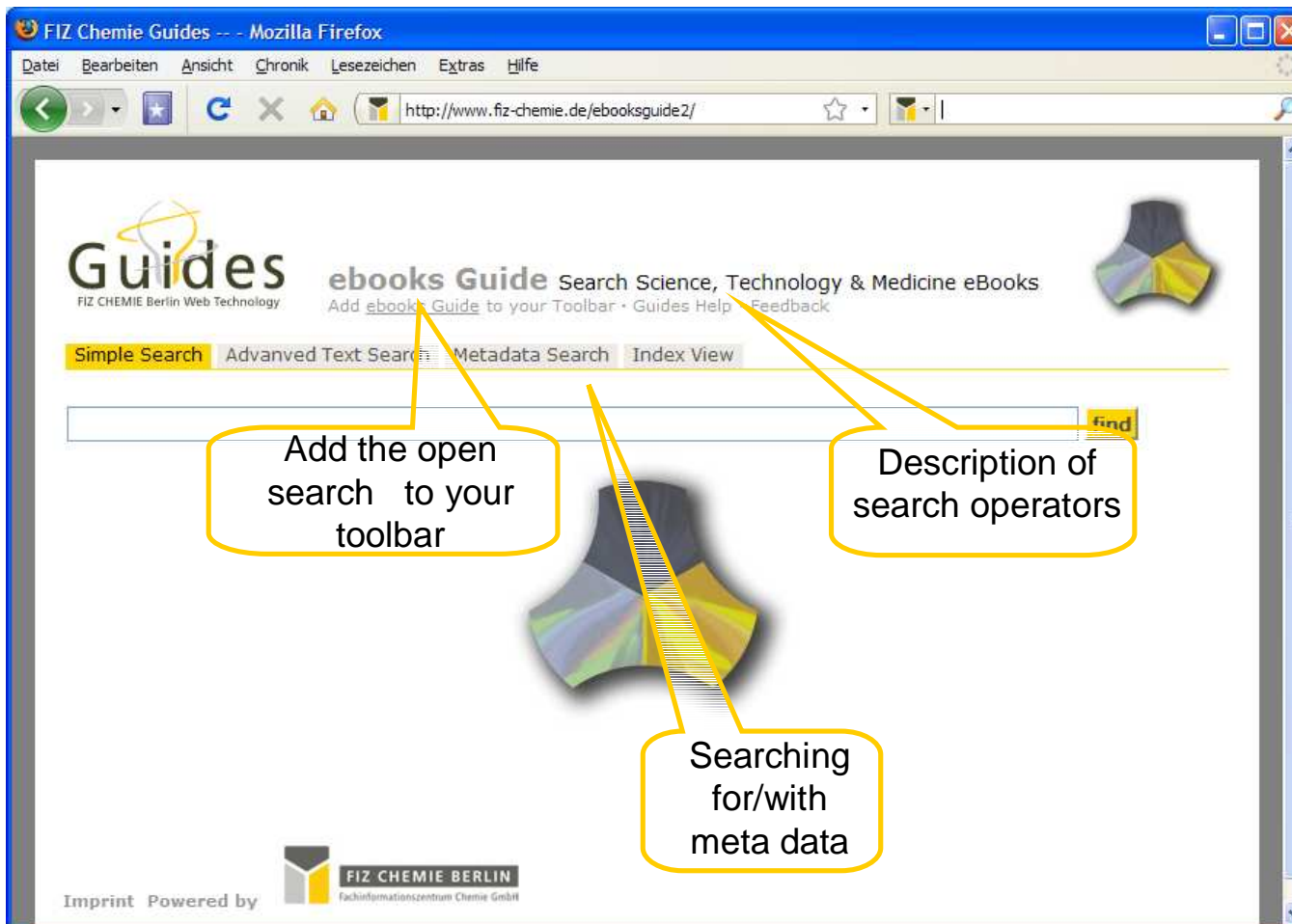
Advanced proximity search

Truncation (right and **left**)

Masking, e.g. sulf **!!** yl will find sulf**en**yl as well as sulf**oxy**l

Searching for meta data (title, author, year of publication, language etc.)

Simple Search



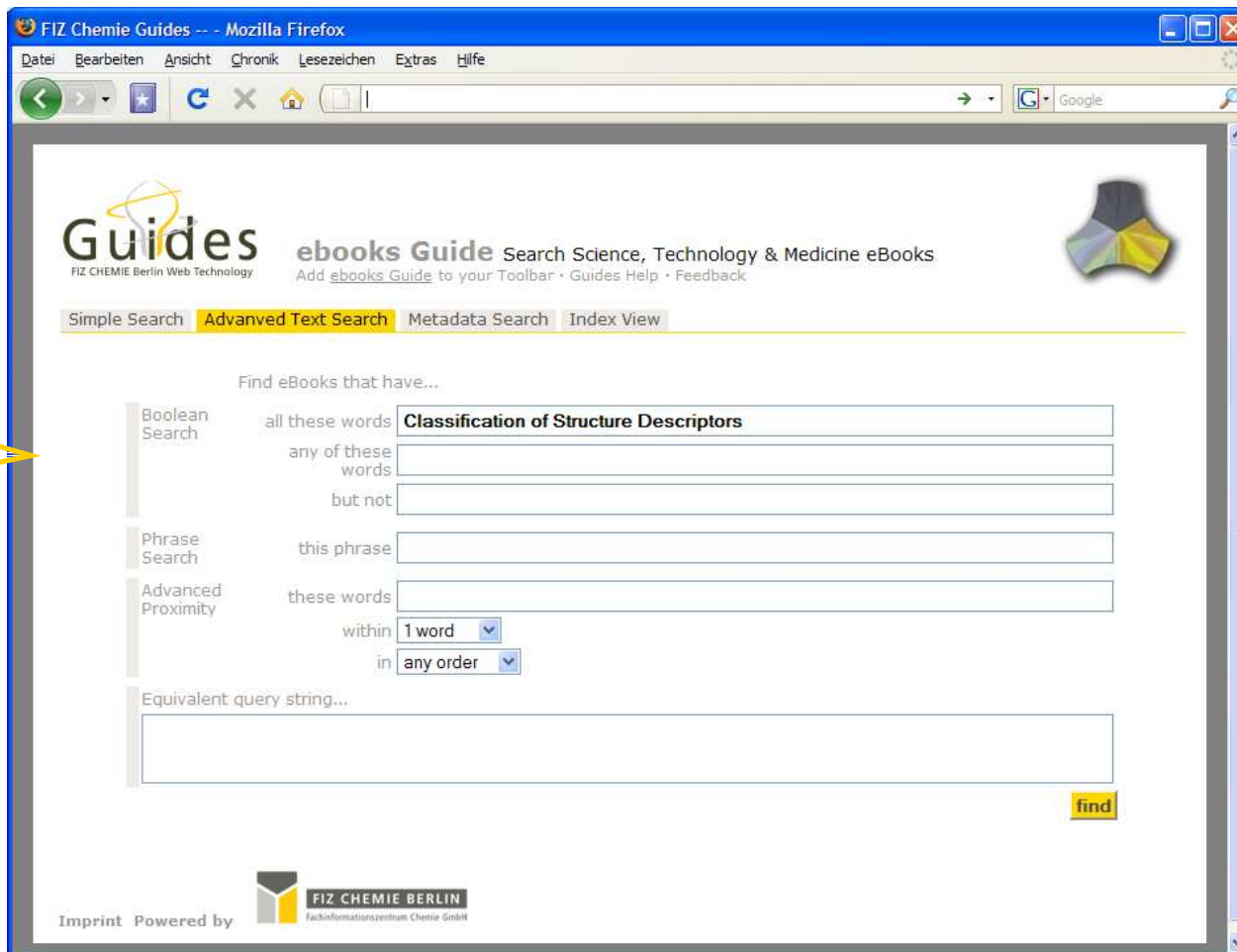
The screenshot shows a Mozilla Firefox browser window displaying the FIZ Chemie Guides website. The browser's address bar shows the URL <http://www.fiz-chemie.de/ebooksguide2/>. The website header includes the logo "Guides" and the text "ebooks Guide Search Science, Technology & Medicine eBooks". Below the header, there are search options: "Simple Search", "Advanced Text Search", "Metadata Search", and "Index View". A search input field is present with a "find" button. Three callout boxes with yellow borders and black text provide instructions:

- "Add the open search to your toolbar" (pointing to the search input field)
- "Description of search operators" (pointing to the search input field)
- "Searching for/with meta data" (pointing to the "Metadata Search" tab)

The footer of the website includes the text "Imprint Powered by" and the FIZ CHEMIE BERLIN logo.

Advanced Text Search

Full text
search with
operators



FIZ Chemie Guides --- Mozilla Firefox

Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe

Google

Guides ebooks Guide Search Science, Technology & Medicine eBooks
FIZ CHEMIE Berlin Web Technology Add [ebooks_Guide](#) to your Toolbar · Guides Help · Feedback

Simple Search **Advanced Text Search** Metadata Search Index View

Find eBooks that have...


Boolean Search all these words
any of these words
but not

Phrase Search this phrase

Advanced Proximity these words
within
in

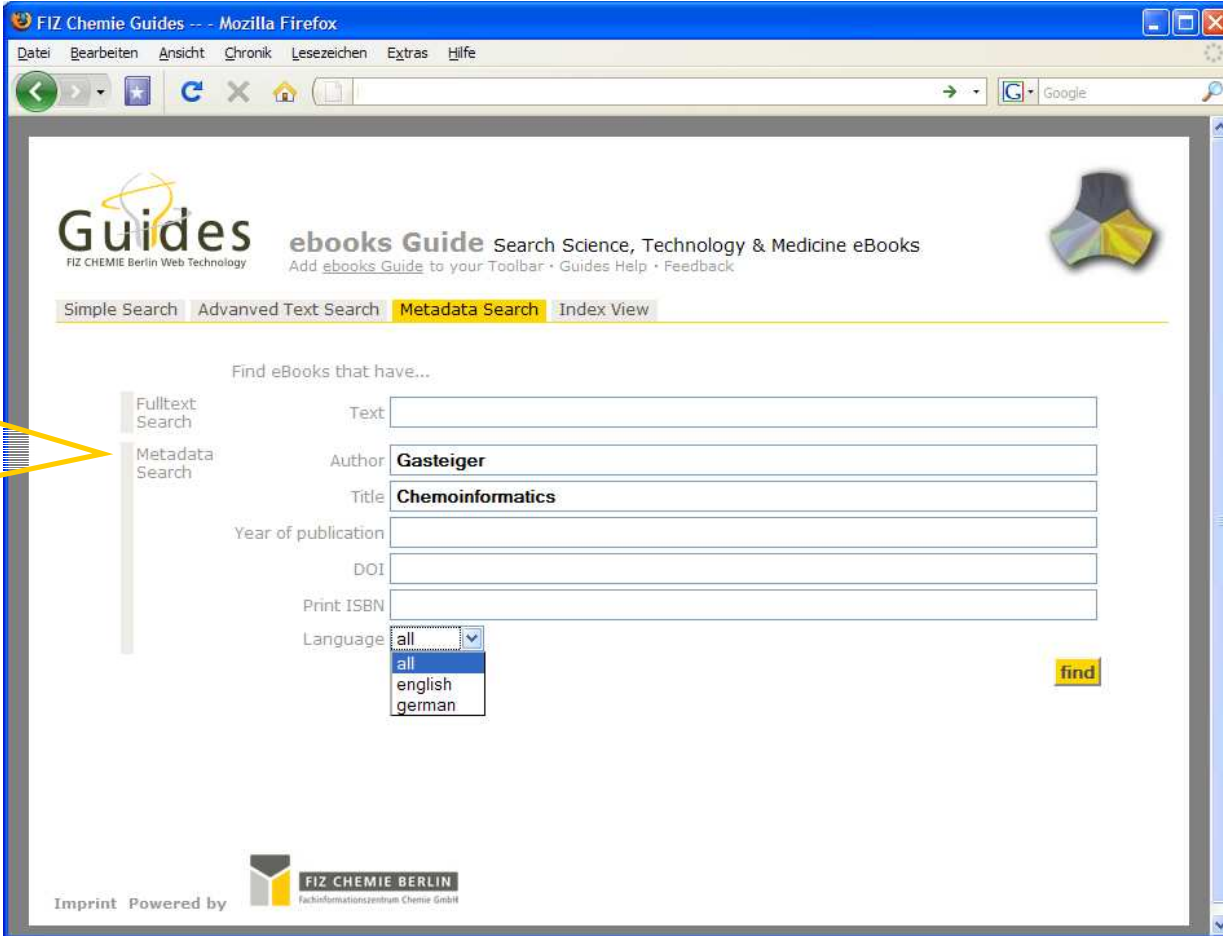
Equivalent query string...

find

Imprint Powered by  **FIZ CHEMIE BERLIN**
Fachinformationszentrum Chemie GmbH

Metadata Search

Full text or meta data search or combination of both



The screenshot shows a web browser window titled "FIZ Chemie Guides --- Mozilla Firefox". The page header includes the "Guides" logo and the text "ebooks Guide Search Science, Technology & Medicine eBooks". Below the header, there are navigation tabs: "Simple Search", "Advanced Text Search", "Metadata Search" (which is highlighted), and "Index View". The main content area is titled "Find eBooks that have..." and contains a search form with the following fields: "Text", "Author" (filled with "Gasteiger"), "Title" (filled with "Chemoinformatics"), "Year of publication", "DOI", "Print ISBN", and "Language" (a dropdown menu with "all" selected). A yellow "find" button is located at the bottom right of the form. The footer of the page includes the text "Imprint Powered by FIZ CHEMIE BERLIN Fachinformationszentrum Chemie GmbH".

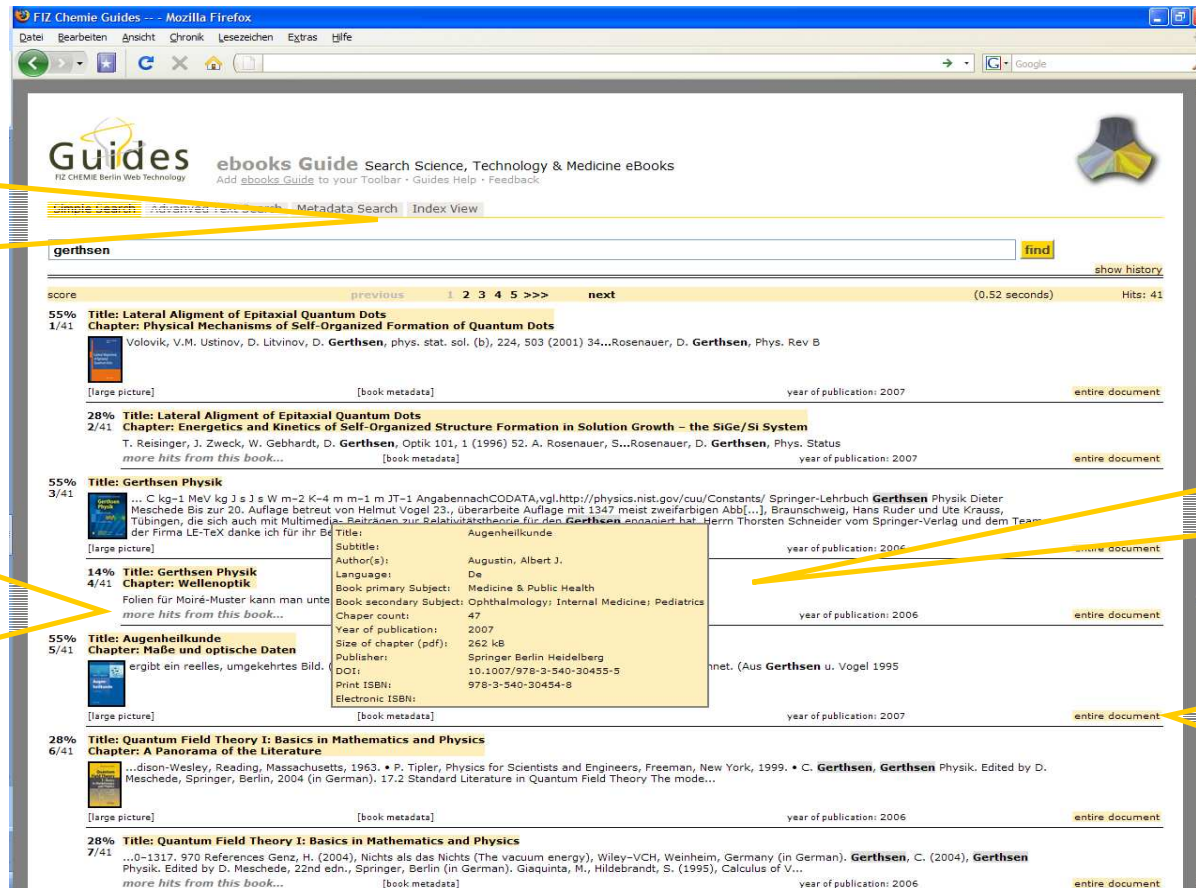
Hitlist

Lookup the search history and index

Multiple chapters of one book will be displayed consolidated

Display the meta data with mouseover

Link to the publisher by DOI



The screenshot shows a search results page in a Mozilla Firefox browser. The search term 'gerthsen' is entered in the search bar. The results are sorted by score, with the top result having a score of 55%. The results list several entries, including chapters and books. A mouseover tooltip is visible over the 5th result, displaying detailed metadata such as the author (Augustin, Albert J.), language (De), and publisher (Springer Berlin Heidelberg). The page also includes navigation links like 'previous', 'next', and 'show history'.

Full text search engine for e-books

Advanced features

- Combination of hits
- Spell checker
- Display of search history
- Index lookup
- Alphabetical listing of all e-books

Full text search engine for e-books

Weighting and relevance

- Relative word frequency
the relationship of the relative sum of the search terms within the total number of the words in a given document.
- Inverse document frequency
describes the weighting of a certain term regarding all indexed and/or in the data base existing documents.
- Oracle text ranking algorithm
additional Oracle algorithm after Salton (inverse document frequency)
- Meta data
in addition to this information FCH imports the meta data into the data base, which guarantee a higher weighting of the meta data in connection with the full text.

Full text search engine for e-books *Further developments*

- Developments (Guide-Version 3.0) will be completed beginning 2009
- Subsequent online access for all users
- Interfaces for industry und libraries will be available
- Addition of more scientific publishers

Full text search engine for e-books *Further developments*

- 24 hours availability (from home as well)

however in dependence of the license with the publishing house

- Full text search on single title, package and total content, and/or in the continuous text (no stop at the end of a chapter)

Current Status

- Search engine prototype is already in use.
- Interface to libraries is already used.
- Approx. 7,000 eBooks by Springer with ~ 146,000 chapters are within a daily update routine.
- Test data of Wiley and Walter De Gruyter are processed at present.
- E-books remain with the publishing house, the index with FCH.
- Full texts are transferred to the customers via the publishing house; license information comes from the customer and is verified with the publishing house.
- After the search the customer is transferred to the publishing house for obtaining his licensed text.

Full text search engine for e-books Join our β -test partners

Publisher

Springer (~20.000)
Wiley (~ 16.000)
Thieme (~200)
Walter De Gruyter (~500)
Wolters Kluwer (~200)
(Royal Society (~700))
(Oxford University Press (~2.125))

Industry

Boehringer Ingelheim
Bayer Schering Pharma
(BASF)

Universities Germany/Switzerland

Kooperativer Bibliotheksverbund (KOBV)
Bayrischer Bibliotheksverbund
GoPortis (TIB, ZBMed, ZBW)
ETH Zürich
TU Darmstadt
Humboldt Universität
FU Berlin

Universities USA

Stanford University
University of Chicago
M I T

General release spring 2009

free test account at

www.fiz-chemie.de/ebooksguide2/

User ID : bibotest

Password: irom85ww



FIZ CHEMIE BERLIN