

Computing Seminar

23rd May 2007

- The CERN Search Service

Andreas Wagner
(CERN IT/IS)

- The Future of Search

Torstein Thorsen,
(VP, Tech. Lead, FAST, Norway)

IS

Internet Services

CERN IT
Department



The CERN Search Service

Computing Seminar
23rd May 2007

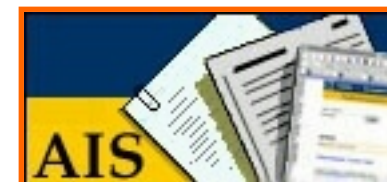
Andreas Wagner – CERN IT/IS



- Search Technology at CERN
- The CERN Search Project
- A CERN Search Solution using FAST ESP

- A wide range information and documentation is available at CERN and more is produced every day:
 - Scientific documents
 - Engineering documents
 - Administrative documents
 - Web-pages
 - Minutes of meetings.
 - etc.
- Depending on their nature, they are stored in different document collection systems such as:
 - CERN Document Server (CDS),
 - Electronic Document Handling (EDH),
 - Engineering Data Management Service (EDMS),
 - websites and databases.
- Being able to easily locate and access this information can benefit every CERN user.

- AIS - Administrative Documents:
 - **What:** Index of administrative documents such as: purchase orders, Leave requests, contracts etc.
 - **Size:** ~2.5 million documents
 - **How:** *Various application specific solutions used (Oracle Intermedia, "ht://Dig") foreseen*
 - **Who:** Database and Engineering Services (IT-DES) group



- CDS – CERN Document Server
 - **What:** Covers preprints, articles, books, journals, photos, videos, exhibition objects, and more.
 - **Size:** ~900,000 bibliographic records in ~500 collections including ~350,000 full-text documents
 - **How:** Search engine provided by CDSware.
 - **Who:** CDS team
- Indico/Agenda
 - **What:** Conference contributions and talks, minutes etc.
 - **Size:** ~120,000 documents
 - **How:** Search engine provided by CDSware
 - **Who:** CDS team



- CERN Web Intranet Search:
 - **What:** Search collection of CERN intranet Web pages.
 - **How:** Search with Verity Ultraseek search engine software.
 - **Who:** CERN Web Services
 - **Size:** ~ 1,000,000 Web pages/documents
- EDMS:
 - **What:** CERN engineering document collection.
 - **How:** Oracle search on metadata, limited/no full text search
 - **Who:** EDMS team in TS department
 - **Size:** ~ 1,000,000 documents



- Xfind – User Support Knowledgebase:
 - *What:* Access to QAs, CERN Computer Newsletter, Technical Writeups (cernlib), Service descriptions, computer book shop catalogue
 - *Size:* <50,000 documents
 - *How:* Search with xfind search engine
 - *Who:* User and Document Services Group
- and others: e.g. CERN Library, Buildings, etc

IS

Scale: Number of Documents

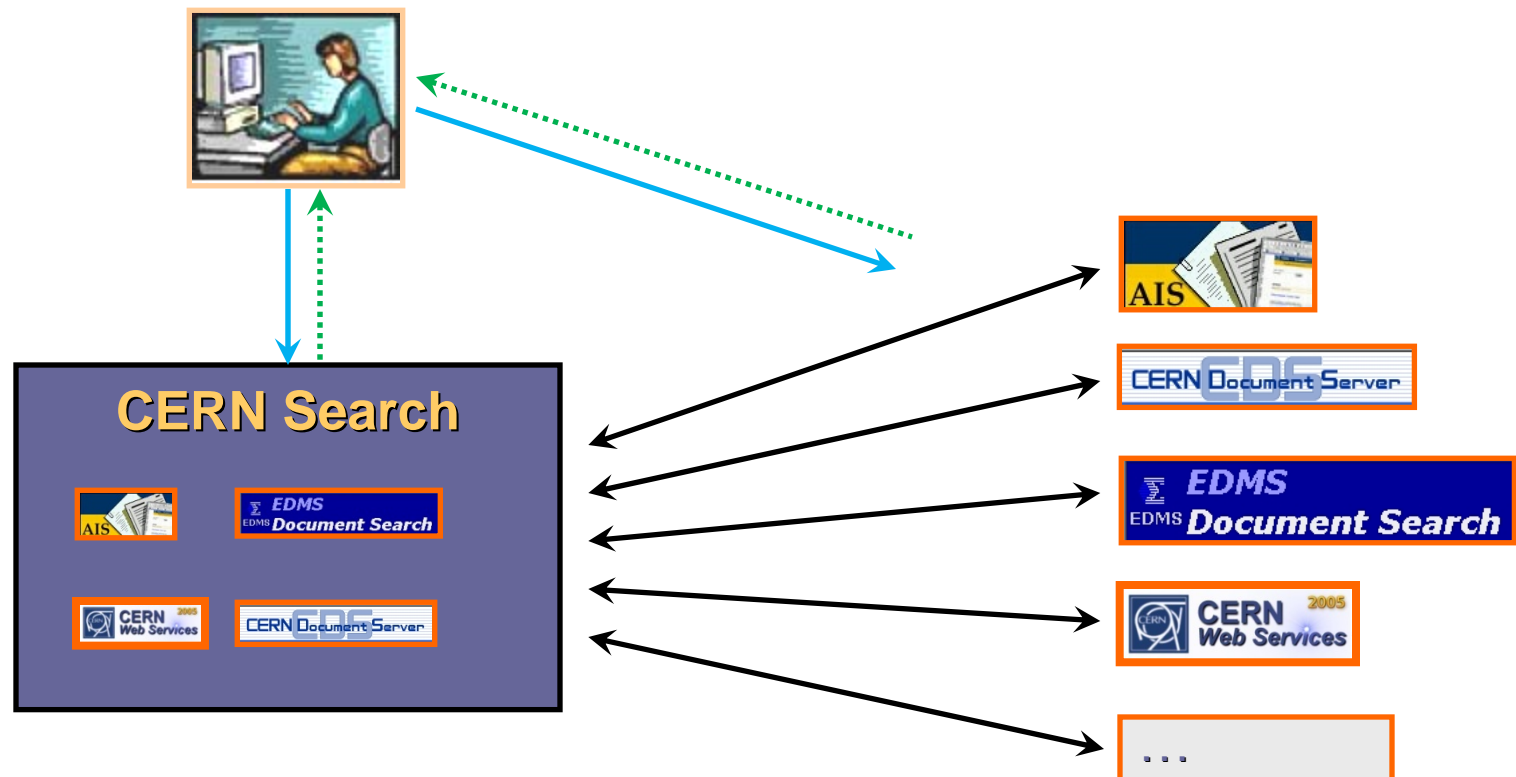
Collection	Number of Documents
Admin documents	~ 2,500,000
CERN Web Intranet	~ 900,000
EDMS	~ 1,000,000
CDS	~ 1,000,000
Agenda/Indico	~120,000
Xfind (knowledgebase)	< 50,000
Total	~ 5 - 6 million documents



Motivation for a CERN wide Search

- “Search solutions” exist in several units
 - **try to find synergies**
- Information dispersed:
“You have to know already where to find the document of interest before you start looking”
 - **create single search entry point for users**

- A common search infrastructure

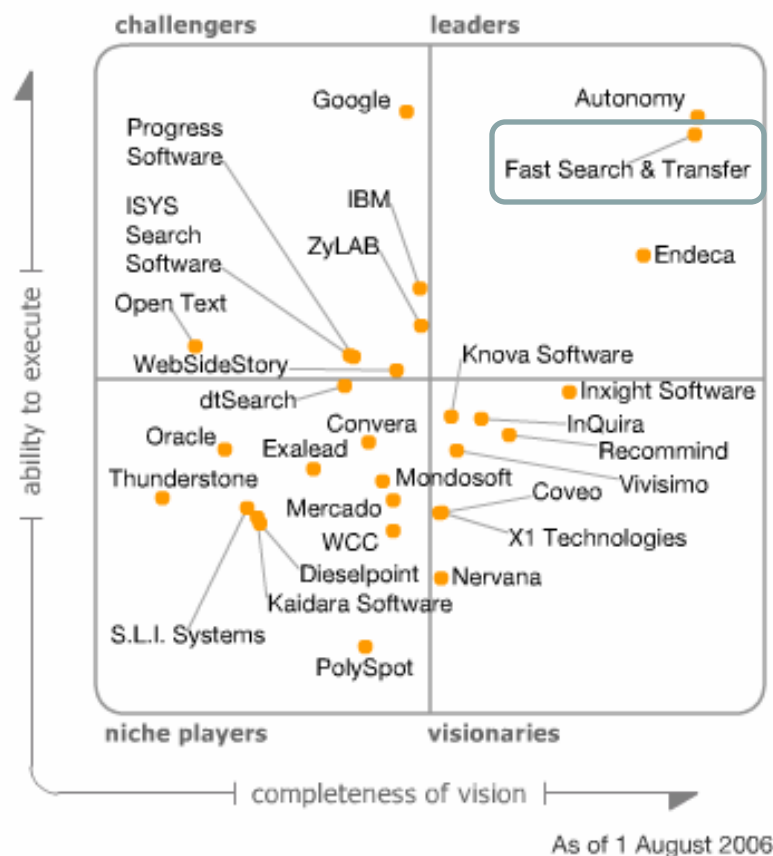
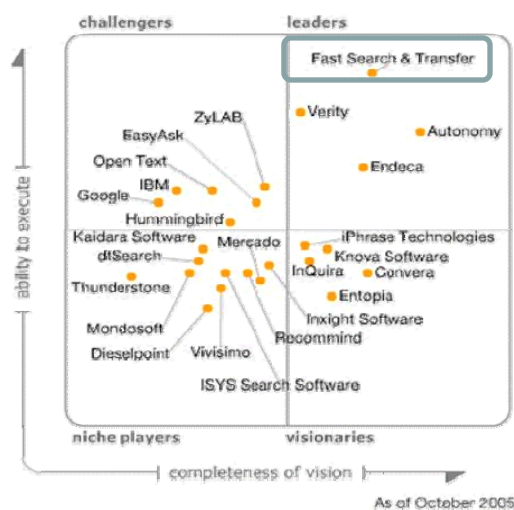
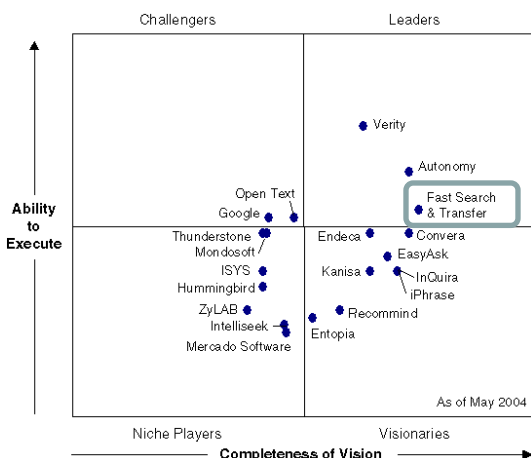


- A CERN Search solution to provide a service
 - **for users**
 - **for service providers**
- Design/build a solution based on one (commercial) product to fulfill the requirements of all described activities above and possibly also other units
- Start of project in early 2006

- Indexing of a variety of document collections
 - Web pages, DBs, administrative/engineering/scientific documents, mailing lists, ...
 - Various document formats
 - Document meta data / full text documents
- Authentication/Authorization
 - Access Restrictions
 - Public / Collections / Documents
 - Confidentiality of documents
 - User Roles

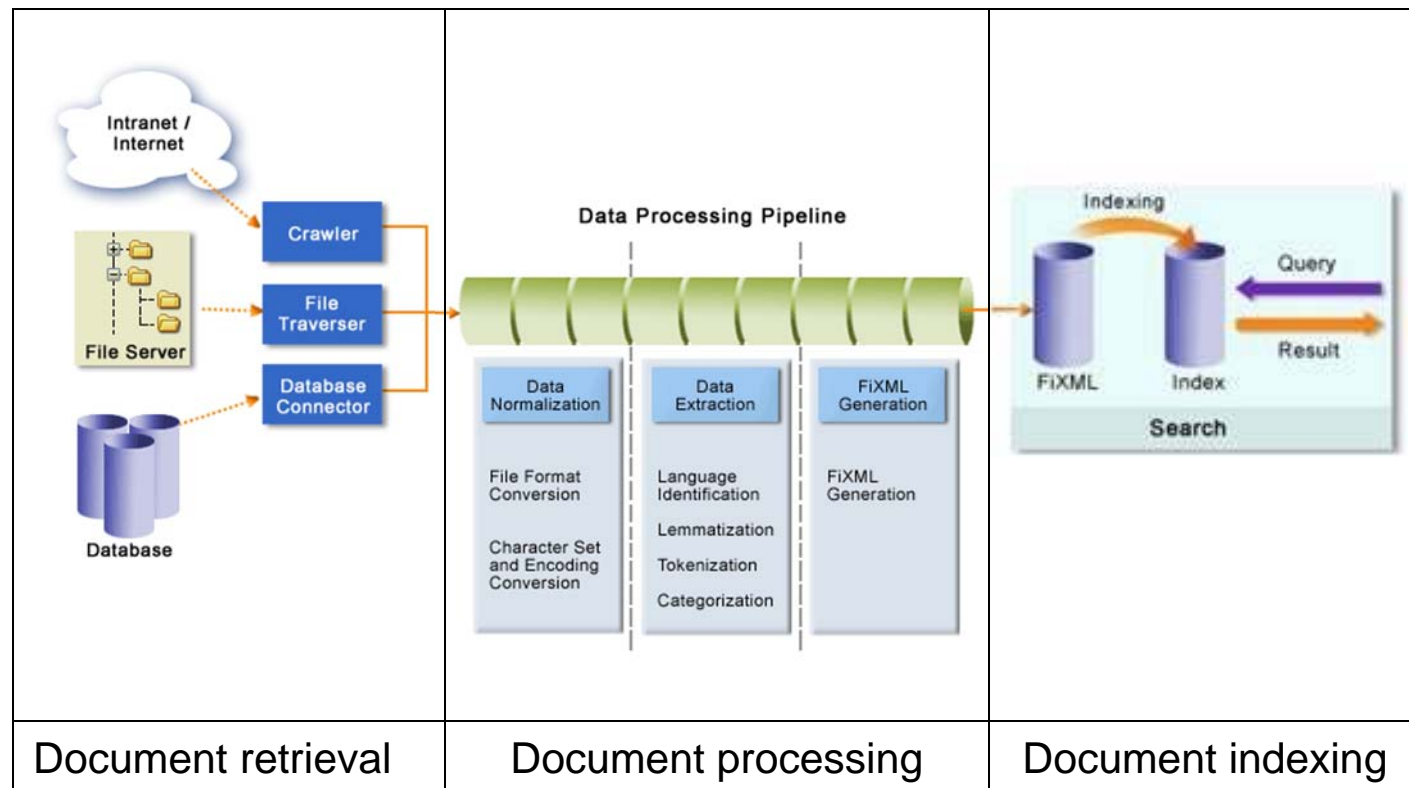
- Are CERN requirements unique ?
 - No, Enterprise Search is commodity at the same time:
 - Products are evolving & not easy to compare
 - “out-of-the-box” solution does not exist
 - Implementing suitable solution challenging
- Enterprise search products:
 - ...

- Gartner Report: "Magic Quadrant for Information Access Technology, 2004-2006"

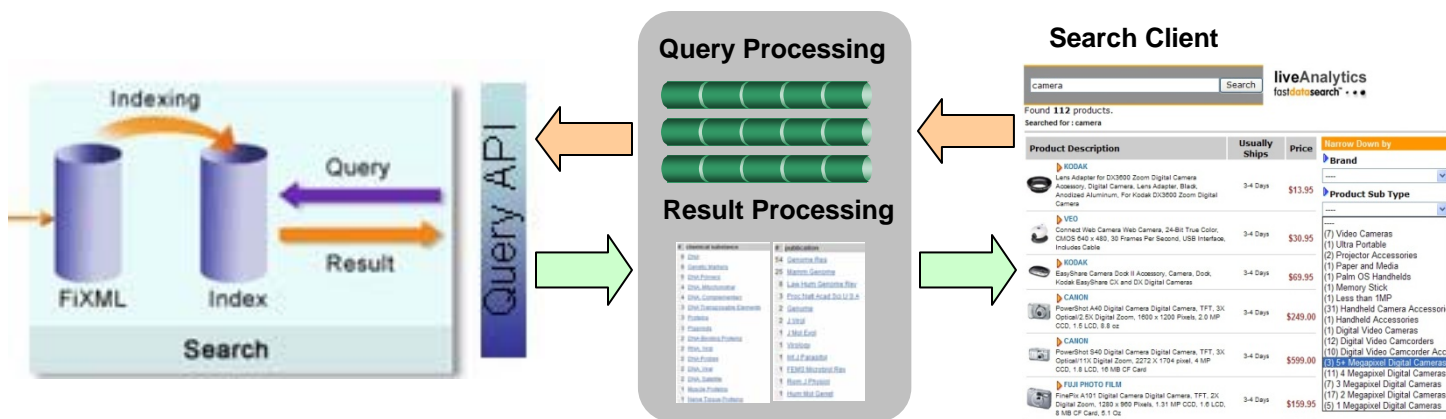


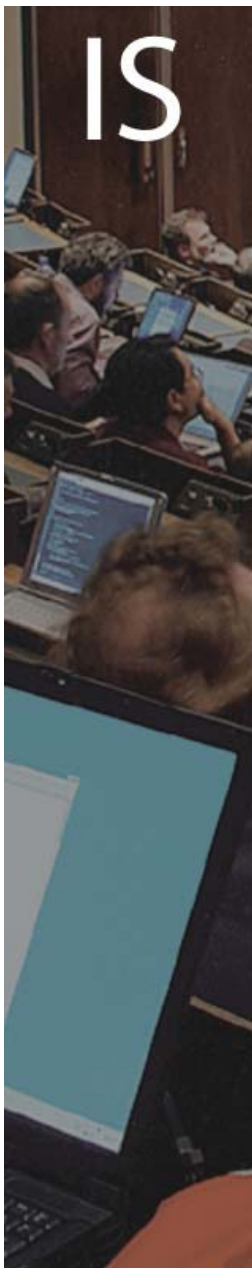
- Dynamic market with a few major players
see Gartner “Magic Quadrant for Information Access Technology”
- Choice was made to base our implementation on FAST ESP:
 - FAST is one of the leaders in Information Access Technology
 - Top ranked in magic quadrant
 - Wide range of clients in many areas
 - FAST ESP: Powerful and flexible product
 - European company

Document Content Flow

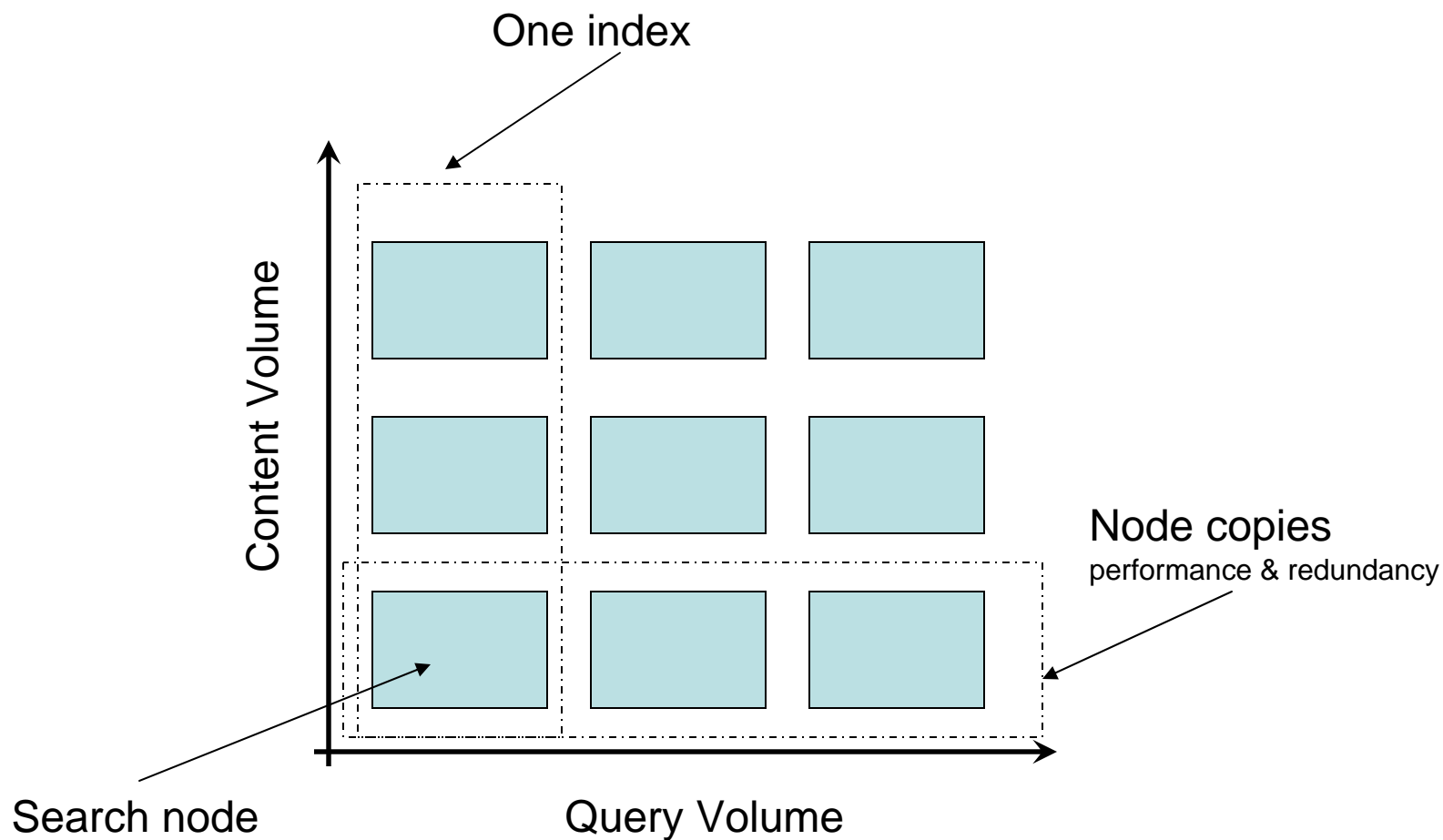


Searching

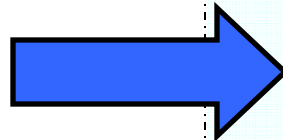




- Highly modular design, very powerful and flexible
- Important Design Considerations for CERN Search Implementation:
 - Scalability
 - Performance
 - Availability, redundancy
 - Authentication/Authorisation model for sensitive content



Web, CDS
EDH, CFU,
EDMS, ...

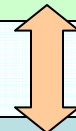
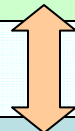


Document Processing & Frontend

search01

search02

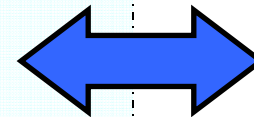
search03



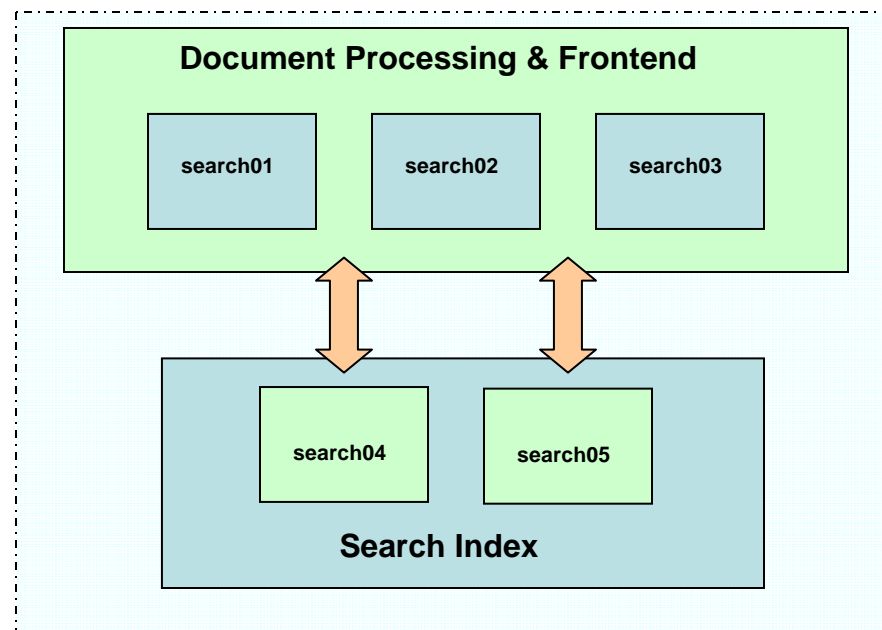
search04

search05

Search Index



- **Search01:**
 - production search frontend
 - anchor server
 - some document processing
- **Search02:**
 - development search frontend
 - document processing
- **Search03:**
 - admin node
 - crawler
 - database connector
- **Search04:**
 - indexer
 - search engine
- **Search05:**
 - search engine




 Administration

[FAST Home](#) | [Logout](#)

Version 5.0

Collection Overview
System Overview

 Document Processing
Logs

 Search View
Data Sources

 System Management
Matching Engines

Collection Overview

 ■ ■ ■ [Create Collection](#) | [Refresh](#) | [Help](#)
Overview - There are 5 Collections

Name	Description	Last input	Docs
CDS	Cern Document Server data	3 hours ago	914616  
CFU	Contract Follow-Up data	N/A	3662  
CernIntranet	CERN Intranet web data	0 seconds ago	1222045  
EDMS	Engineering Data Management Service	N/A	415814  
People	Dump from AD	N/A	21527  

Copyright © 2007 Fast Search & Transfer ASA



- In February 2007:
 - CERN Search went into production replacing the previously used Ultraseek Search Engine.

<http://cern.ch/search>



CERN Search

[Advanced search](#) [Help](#)

Presently about **1 million CERN Web pages** are indexed.

We are working on making other document collections such as CDS, EDMS, EDH, etc. available soon.

Search other Internal CERN Sources

CERN Index Pages (main activities): [Alphabetic](#)

CERN Phone Book (people's email, computer accounts, location ...)

Go to the [Phone Book](#) for detailed search

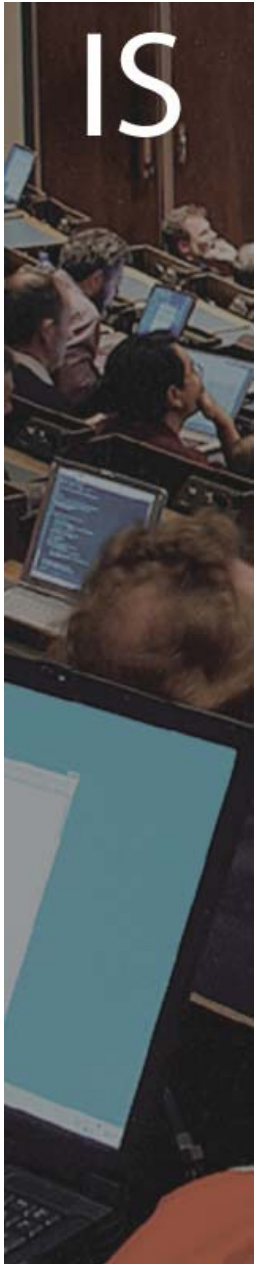
CDS - CERN Document Server (preprints, books, periodicals, reports, photos ...)

Go to the [CERN Document Server](#) for detailed search

EDMS - CERN Engineering Data Management Service (CERN and LHC technical data...)

Go to the [EDMS Service](#) for detailed search

[User feedback and Comments](#) - [About CERN Search](#) - [Help Pages](#)
 Powered by [FAST](#) © CERN 2007



Basic search [Help](#)

Generic search options

Search Type: Query Language:

Correct spelling: Site collapse: Normalize words:

Hits:

Preferences

[Save](#) or [reset](#) generic search options (requires cookies enabled)

Fielded search

in the

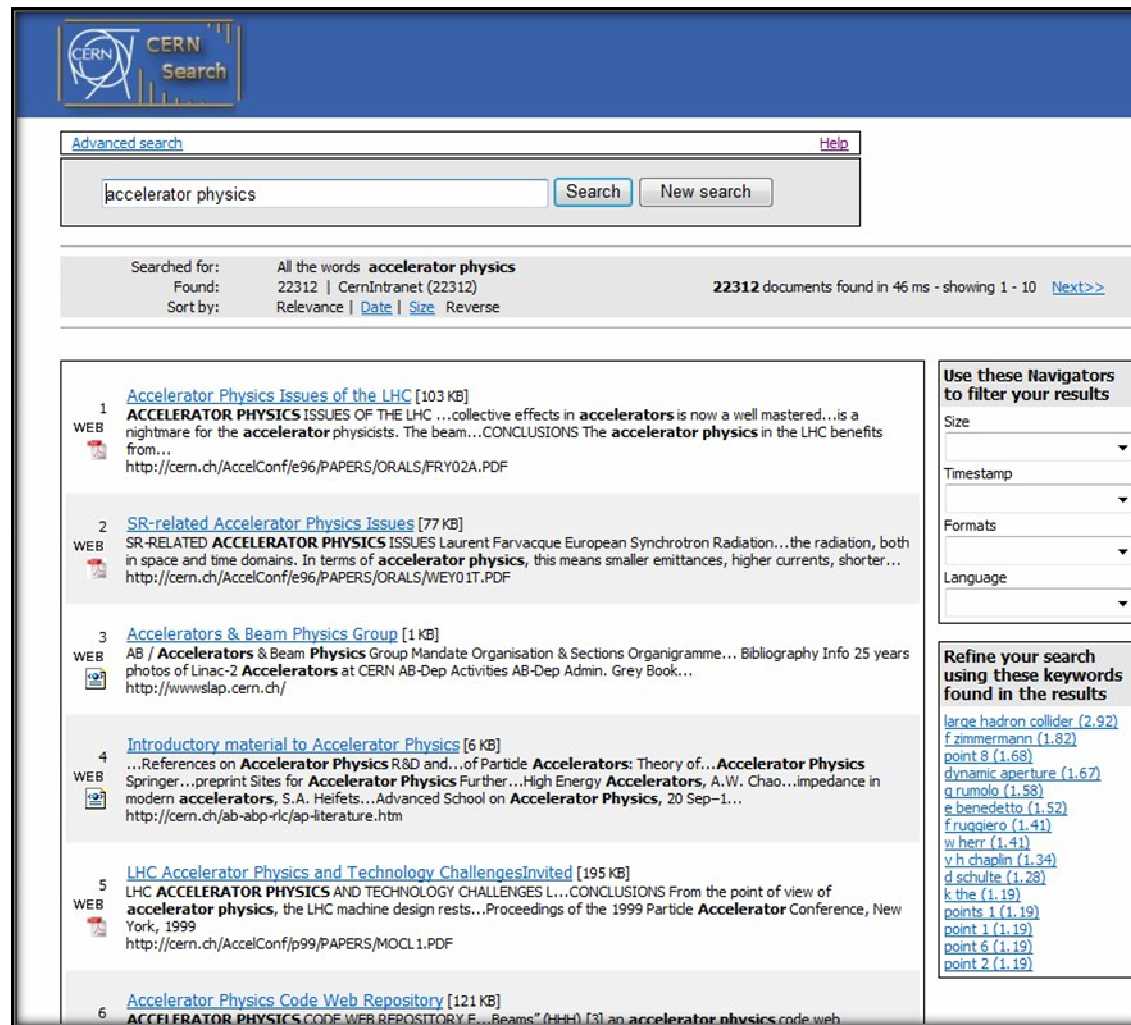
AND

in the

Presently about 1.3 million CERN Web pages are indexed.

[User feedback and Comments](#) - [About CERN Search](#) - [Help Pages](#)
Powered by FAST © CERN 2007





CERN Search

Advanced search [Help](#)

accelerator physics

Searched for: All the words **accelerator physics**
 Found: 22312 | CernIntranet (22312) **22312** documents found in 46 ms - showing 1 - 10 [Next>>](#)
 Sort by: Relevance | [Date](#) | [Size](#) | [Reverse](#)

- 1** [Accelerator Physics Issues of the LHC](#) [103 KB]
 WEB **ACCELERATOR PHYSICS** ISSUES OF THE LHC ...collective effects in **accelerators** is now a well mastered...is a nightmare for the **accelerator** physicists. The beam...CONCLUSIONS The **accelerator physics** in the LHC benefits from...
<http://cern.ch/AccelConf/e96/PAPERS/ORALS/FRY02A.PDF>
- 2** [SR-related Accelerator Physics Issues](#) [77 KB]
 WEB SR-RELATED **ACCELERATOR PHYSICS** ISSUES Laurent Farvacque European Synchrotron Radiation...the radiation, both in space and time domains. In terms of **accelerator physics**, this means smaller emittances, higher currents, shorter...
<http://cern.ch/AccelConf/e96/PAPERS/ORALS/WIY01T.PDF>
- 3** [Accelerators & Beam Physics Group](#) [1 KB]
 WEB AB / **Accelerators & Beam Physics** Group Mandate Organisation & Sections Organigramme... Bibliography Info 25 years photos of Linac-2 **Accelerators** at CERN AB-Dep Activities AB-Dep Admin. Grey Book...
<http://wwwslap.cern.ch/>
- 4** [Introductory material to Accelerator Physics](#) [6 KB]
 WEB ...References on **Accelerator Physics** R&D and...of Particle **Accelerators**: Theory of...**Accelerator Physics** Springer...preprint Sites for **Accelerator Physics** Further...High Energy **Accelerators**, A.W. Chao...impedance in modern **accelerators**, S.A. Heifets...Advanced School on **Accelerator Physics**, 20 Sep-1...
<http://cern.ch/ab-abp-rlc/ap-literature.htm>
- 5** [LHC Accelerator Physics and Technology ChallengesInvited](#) [195 KB]
 WEB LHC **ACCELERATOR PHYSICS** AND TECHNOLOGY CHALLENGES L...CONCLUSIONS From the point of view of **accelerator physics**, the LHC machine design rests...Proceedings of the 1999 Particle **Accelerator** Conference, New York, 1999
<http://cern.ch/AccelConf/p99/PAPERS/MOCL1.PDF>
- 6** [Accelerator Physics Code Web Repository](#) [121 KB]
 WEB **ACCELERATOR PHYSICS** CODE WEB REPOSITORY E...Beams" (HT+I) [3] an **accelerator physics** code web

Use these Navigators to filter your results

Size

Timestamp

Formats

Language

Refine your search using these keywords found in the results

[large hadron collider \(2.92\)](#)
[f zimmermann \(1.82\)](#)
[point 8 \(1.68\)](#)
[dynamic aperture \(1.67\)](#)
[q rumolo \(1.58\)](#)
[e benedetto \(1.52\)](#)
[fruggiero \(1.41\)](#)
[w herr \(1.41\)](#)
[v h chaplin \(1.34\)](#)
[d schulte \(1.28\)](#)
[k the \(1.19\)](#)
[points 1 \(1.19\)](#)
[point 1 \(1.19\)](#)
[point 6 \(1.19\)](#)
[point 2 \(1.19\)](#)

- Search within sub-sets of the Web index:

Home | Search | SiteMap | Contact Us | Credits

Search Results - Résultats de la recherche

Search again:

- [Spotlight on the LHC Olympics](#)
1 ...on... News Ask an Expert The LHC Olympics Coinciding with the kickoff...on in 2007, the Large Hadron Collider (LHC) will be the world's biggest and most...interpreting the data produced from the LHC won't be an easy task. This is because...
<http://cern.ch/public/Content/Chapters/Spotlight/SpotlightOlympics-en.html> - 2 KB - 2007 02 12
- [What is LHC?](#)
2 ...Expert What's next at CERN? What is the LHC? The Large Hadron Collider (LHC) is a particle accelerator which will probe deeper...of motion of a flying mosquito. What makes the LHC so extraordinary is that it squeezes energy into...
<http://cern.ch/public/ACCELERATORS/LHCacc.html> - 2 KB - 2006 11 10
- [Qu'est-ce que le LHC?](#)
3 ...étape au CERN? Qu'est-ce que le LHC? Le Grand collisionneur de hadrons, (Large Hadron Collider, LHC) est un accélérateur de particules qui sondera la matière...du mouvement d'un moustique en vol. Le LHC parviendra à condenser son énergie dans...
<http://cern.ch/public/Content/Chapters/AboutCERN/CERNFuture/WhatLHC/WhatLHC-fr.html> - 2 KB - 2006 11 10
- [LHC milestone triple jump](#)
4 ...Spotlight on... News Ask an Expert LHC milestone triple jump An official ceremony... CERN's major project to build the LHC (Large Hadron Collider), the world's largest...all the main magnets required to build the LHC. This gigantic scientific machine is...
<http://cern.ch/public/Content/Chapters/Spotlight/Spotlight3Jump-en.html> - 3 KB - 2007 02 28
- [Pourquoi le LHC?](#)
5 ...Quelle est la prochaine étape au CERN? Pourquoi le LHC? Le LHC est construit pour compléter notre connaissance de l'Univers...associée, le boson de Higgs. Si celui-ci existe, le LHC devrait permettre de le détecter. Qu'en est-il des quatre forces? Lorsque...
<http://cern.ch/public/Content/Chapters/AboutCERN/CERNFuture/WhyLHC/WhyLHC-fr.html> - 3 KB - 2006 11 10
- [Gros plan sur les Olympiades du LHC](#)
6 ...Interroger un Expert Les Olympiades du LHC Alors que les Jeux Olympiques d'hiver démarraient... 2007, le grand collisionneur de hadrons (LHC) sera le plus grand et le plus puissant des accélérateurs...Cependant, l'interprétation des données du LHC ne sera pas une tâche facile. En effet, les protons...
<http://cern.ch/public/Content/Chapters/Spotlight/SpotlightOlympics-fr.html> - 2 KB - 2007 02 12

71 hit(s) [Sort by date](#) 1 - 10 [Next >>](#) [CERN Search](#)

Copyright CERN 2006, DSU - Communication Group

- CFU data in protected collection
- Indexing of per document ACLs

The screenshot displays the CERN Contract Follow Up (CFU) website. At the top, there is a header with the CERN logo and the text "CERN — European Organization for Nuclear Research". To the right, it shows the user "User AWAGNER" and links for "Help - Logout". Below the header, the main title "CERN Contract Follow Up" is displayed, followed by a search bar containing the text "wagner" and a "Search" button. The main content area is organized into a list of links under various categories:

- [News](#)
- **Main Menu**
 - [Your Purchase Procedures](#)
 - **Divisional Requests**
 - [Divisional Requests](#)
 - [Advance Information Publications](#)
 - [DR Summary](#)
 - **Market Surveys**
 - [Market Surveys](#)
 - [MS Summary](#)
 - **Invitations to Tender and Price Enquiries**
 - [Invitations to Tender & Price Enquiries](#)
 - [IT/PE Summary](#)
 - **Contracts**
 - [Supplies & Other Contracts](#)
 - [Industrial Services Contracts](#)
 - [Protocols, Addenda and Amendments](#)
 - [Contract Financial Milestones](#)
 - [Contracts Summary](#)
 - [Proposals to Award](#)
 - [Interface with SIRIAC](#)
 - [Safety documents and Documents on Completion](#)
 - [Purchase procedure links](#)
 - [Finance Committee Meetings & submitted Documents](#)
 - [Other Committees & Meetings](#)
 - [Referential data](#)
- [CFU for STORES](#)
- **Utilities**
 - [Search CFU](#)
 - [Exit CFU](#)
 - [View your access rights](#)
 - [Report Navigator](#)
- **Useful links**
 - [A Guide to AISmedia](#)
 - [Abstract of CERN Purchase Procedures](#)

At the bottom of the page, there is a footer with contact information: "For problems or comments send email to ais.support@cern.ch" and "New users, please try the [CFU QuickTour](#)". The footer also includes "Copyright CERN".

- Including additional document collections
 - EDMS, Buildings
- Search Backend for new CFU
 - CFU data in private collection with per document ACLs
- Adding administrative documents
 - Database field mapping to index profile done
 - Mapping of Foundation roles to AD groups
- Explore additional information sources:
 - Mailing list archives
 - Sharepoint Web sites
 - etc.

IS

Questions ?

