

Search with Invenio

Invenio User Group Workshop 2012

CERN IT-CIS-DLS – Flavio Costa

AGENDA

A quick review (no more than 5') on:

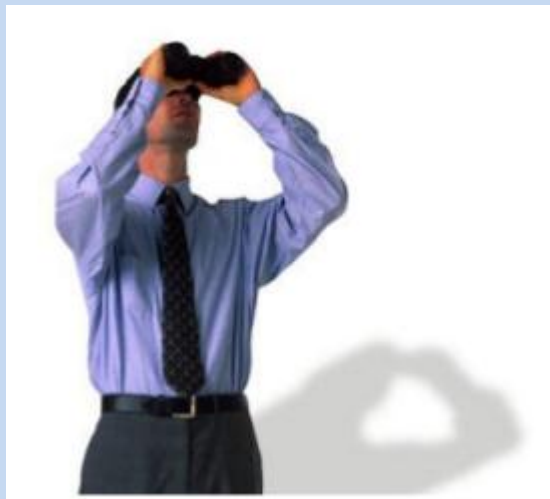
- The characteristics of the Invenio Search interface
- Names and definitions

(present situation, users' point of view)

Default interface is google-like, with some additional options:

Search 1,095,951 records for:

[Search Tips](#) :: [Advanced Search](#)



Search collections:

Sort by:

Display results:

Output format:

How to search?

- Many syntaxes are supported: Google like and SPIRES
- Structured metadata: ex: author:ellis title:muon*, experiment:NA60 year:2001
- Search via MARC: ex: 8564__u: ...
- Boolean queries: AND(+), OR(|), NOT(-)
- Parenthesis: ex: (gravity OR supergravity) AND (ellis OR perelstein)
- Special characters and punctuation: ex: C++, O'Donnell
- International characters: ex: пушкин, Lemaitre (finds Lemaître)
- Not case sensitive

How to search?

- Word truncating: ex: CERN-TH*31, a*, “neutrino mass*”
 - Author-smart: ex "Gys, T*"
 - Stemming: muon → muon, muons, **not** muonic ...
- Search modes
 - Any words: ex: neutrino mass
 - Phrase: ex: 'neutrino mass', "neutrino mass"
 - Regular expression: ex.: /neutrino.*mass/
- See also: <http://cdsweb.cern.ch/help/search-tips>

Advanced Search and searching techniques

- **Ranges:** “->”, ex.1: muon decay year:1983->1992 – ex.2: author:"Ellis, J"->"Ellis, Qqq"
- **Combined searches (metadata/fulltext/ecc.):** All the possible searches types can be combined together in one query.
For example, to find documents that have the word ellis inside author field, that do not contain words like muon, 'muonic' etc in any field, that contain LHC in the title fields, and that were published in year starting by digits '200':
author:ellis -muon* +title:'LHC' year:200*

Advanced Search mode

to combine multiple search terms in different ways

Search 1,144,538 records for:

All of the words: ▾	<input type="text"/>	any field ▾	AND ▾
All of the words:	<input type="text"/>	any field ▾	AND ▾
Any of the words:	<input type="text"/>	any field ▾	
Exact phrase:	<input type="text"/>	any field ▾	<input type="button" value="Search"/> <input type="button" value="Browse"/>
Partial phrase:	<input type="text"/>		
Regular expression:	<input type="text"/>		

[Search Tips](#) :: [Simple Search](#)

Added/modified since: **until:**

Sort by: **Display results:** **Output format:**

Advanced Search

Regular Expressions

```
title:/^E.*s$/
```

Quick Regular Expressions explanations

- ^string matches the beginning (but within [] , exclusion)
- . matching any single character
- Quantification, ? (0 or 1), * (0 or more), + (1 or more)
- string\$ matches the end
- “or”(|), ex: /Gray|Grey/
- Grouping () , ex.: /Gr(a|e)y/
- - Intervals, ex. 0-9
- [:digit:] character class
- { } occurrence count, ex.: {5}

Advanced Search

Regular Expressions

Examples:

- author:/^Ellis, (Jonathan|John)\$/
- title:/dense [^|]* ?matter/ - Title containing “dense” and “matter” separated by at most 1 word not containing an “|”
- collection:PREPRINT -year:/^[0-9]{4}(\?|-|-|[0-9]{4})?\$/ - Preprint collection with not standard format year (YYYY, YYYY?, YYYY-, YYYY-YYYY)
- collection:PREPRINT year:/^[[:digit:]]{4}(\?|-|-|[[:digit:]]{4})?\$/ - alternative way

Advanced sorting

Let's try the following Search:

Search:

[Search Tips](#) :: [Advanced Search](#)

Search collections:

Sort by:

Display results:

Output format:

Check the sorting: on multiple report numbers documents the first is used for sorting.

How to sort by report numbers of the format *CERN-TH...* ?

Advanced sorting

Sort pattern (sp)

sp is a “silent” parameter, not available in the Search interface.

To be added at the end of the URL in the form:

....&sp=<pattern>

Try for example:

...&sp=CERN-TH

THANK YOU!