

ALL ABOUT REFERENCES

Alessio Deiana
INSPIRE

FROM BIBEDIT

The screenshot shows the INSPIRE Bibedit interface. On the left, there's a sidebar with options like Record, Search, Fields, View, Holding Pen, Undo/Redo, History, and a Ready status indicator. The main area displays a record with fields such as 001, 0247_, 035_, 035_, 037_, 037_, 100_, 245_, 246_, 269_, 300_, 500_, 520_, and 595_. A green arrow points from the top right towards the toolbar icons at the top right of the main window. The toolbar includes icons for Record (with a circled number 854529), Search, Submit, Cancel, Switch to Read-only, Add, Delete selected, View, Holding Pen, Undo/Redo, History, and Special symbols. The status bar at the bottom indicates "record revision: 2012.09.29 14:10:50".

PUBLIC API

- REST and regular html form
- Extract from a URL or upload a PDF

COMMAND LINE

- refextract
 - bibtask
 - deals with records
- docextract
 - standalone
 - deals with pdfs directly

COMMAND LINE

```
$ /opt/invenio/bin/refextract
```

- bibtask
- deals with records
- docextract
- standalone
- deals with pdfs directly

COMMAND LINE

```
$ /opt/invenio/bin/refextract
```

- bibtask
- deals with records

```
$ /opt/invenio/bin/docextract
```

- standalone
- deals with pdfs directly

DOCEXTRACT

```
$ docextract example.pdf --out example.xml
```

```
<record>
  <controlfield tag="001">1091122</controlfield>
  <datafield tag="999" ind1="C" ind2="5">
    <subfield code="o">1</subfield>
    <subfield code="s">PHRVA,D42,3613</subfield>
  </datafield>
  <datafield tag="999" ind1="C" ind2="5">
    <subfield code="o">2</subfield>
    <subfield code="s">APHYE,7,263</subfield>
  </datafield>
```

DOCEXTRACT

```
$ docextract --raw-references refs.txt --out example.xml
```

```
<record>
  <controlfield tag="001">1091122</controlfield>
  <datafield tag="999" ind1="C" ind2="5">
    <subfield code="o">1</subfield>
    <subfield code="s">PHRVA,D42,3613</subfield>
  </datafield>
  <datafield tag="999" ind1="C" ind2="5">
    <subfield code="o">2</subfield>
    <subfield code="s">APHYE,7,263</subfield>
  </datafield>
```

REFEXTRACT

- **refextract --recids 1-100**
- **refextract --recids 2 --create-ticket**
- **refextract --collections 'HEP'**
- **refextract -a (runs on added records)**

KBs

```
ls /opt/invenio/etc/docextract/*.kb  
authors.kb collaborations.kb journal-titles-inspire.kb  
journal-titles-re.kb report-numbers.kb books.kb  
journal-titles.kb publishers.kb special-journals.kb
```

(Can also be stored in the DB)

STEP I: FINDING REFERENCES

References

[1] R. G. Roberts, Nucl. Phys. B615 (2001) 358–384,

[2] hep-ph/0104088

Or

I) or I., etc.

STEP 2: REBUILDING REFERENCES

From

[1] R. G. Roberts,
Nucl. Phys. B615 (2001)
358–384, .

[2] etc

To

[1] R. G. Roberts, Nucl. Phys. B615 (2001) 358–384, .

STEP 3:TAGGING

<cds.AUTHetal>S. Chatrchyan et al.</cds.AUTHetal>Technical Proposal for the Upgrade of the CMS detector through 2020,<cds.JOURNAL>*Nucl.Instrum.Meth.*</cds.JOURNAL> <cds.VOL>A628</cds.VOL> <cds.YR>(2011)</cds.YR><cds.PG>182</cds.PG>

STEP 4: ELEMENTS TRANSFORMATION

line marker [I]

AUTH

```
{'auth_type': 'etal',
 'auth_txt': u'S. Chatrchyan et al.',
 'type': 'AUTH', 'misc_txt': u"}
```

COLLABORATION

```
{'collaboration': u'The CMS Collaboration',
 'type': u'COLLABORATION',
 'misc_txt': u' '}
```

JOURNAL {

```
'volume': u'3',
 'is_ibid': False,
 'title': u'JINST',
 'extra_ibids': [],
```

STEP 5: XML GENERATION

```
<datafield tag="999" ind1="C" ind2="5">
  <subfield code="o">6</subfield>
  <subfield code="h">D.Abbaneo et al.</subfield>
  <subfield code="m">Construction of the first full-size GEM-based Prototype for
the CMS High- $\eta$  Muon System, 2010</subfield>
  <subfield code="s">IEEE Nucl.Sci.Symp.Conf.Rec.,1909,1913</subfield>
  <subfield code="y">2010</subfield>
</datafield>
<datafield tag="999" ind1="C" ind2="5">
  <subfield code="o">6</subfield>
  <subfield code="m">RD51 Note 2010-008</subfield>
  <subfield code="r">arXiv:1012.1524 [physics.insdet]</subfield>
</datafield>
```

TESTING

For each version:

- tested on 5000 pdfs
- we generate the diff between versions

CURATION

- \$9 CURATOR
- TEXT mode for editing MARC
- Tickets creation

BIBRANK INDEXING

Examples of searches

```
search_pattern(p="CMS-UG-TP-1", f="reportnumber")
```

```
search_pattern(p="JINST,3,S08004", f="journal")
```

```
search_pattern(p="10.", f="doi")
```

CITATION DICTIONARY

Dictionary

```
{|:[2,3]}
```

Database

citee	citer	last_updated
796	793	2012-11-12 15:23:57
708	613	2012-11-12 15:23:57
926	927	2012-11-12 15:23:57

Great

- Done