SOLR TOTAL

17 000 000  Bibliographic records
3 500 000  Authority records
1 000  MARC fields/subfields
2 000  Indexes
200 Go  Indexes
>1  Levels of structuration of indexes
1,5  Search engine : Solr + SIREN

www.abes.fr

Yann NICOLAS  @071625348

Christophe
We need to mine our MARC data

Bibliographic records

Authority records
We are demanding!

- All the records
- All the fields, subfields, indicators
- All the time ➔ real time updating of indexes
  
  + full text indexing
  + facets
Solution #1 (2011) : vanilla SolR

UNIMARC/XML in DB

⇒ XQuery

⇒ Solr XML Documents (flat lists of fields)

⇒ Indexation by Solr
  • as text
  • as string

Query

• Via API (JAVA) or web service
• Via a gloomy web interface
# Non flat MARC vs Flat indexes

## Search vs Result

<table>
<thead>
<tr>
<th>Search</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Firstname = <strong>Josef</strong></td>
<td>Record with two authors :</td>
</tr>
<tr>
<td>Lastname = <strong>Chvéïk</strong></td>
<td>• <strong>Josef</strong> Colomb</td>
</tr>
<tr>
<td></td>
<td>• Jean-Jacques <strong>Chvéïk</strong></td>
</tr>
</tbody>
</table>

Data: \[[a_1, b_1], [a_2, b_2]\]  

Index: \((a_1, a_2, b_1, b_2)\)

OK, we could enrich the index this way: \((a_1, a_1b_1, a_2, a_2b_2, b_1, b_2)\)
We could... #lazy
Solution #2 (2016) : SolR + SIREN

Non flat indexes for non flat MARC data

Two levels : fields + their own subfields

Three levels : groups of fields

Etc. (in theory)
Solr+SIREN document to index

Structure:
One field and its subfields

<BB700>
  <B700-ind1></B700-ind1>
  <B700-ind2>1</B700-ind2>
  <B700S3>027994007</B700S3>
  <B700Sa>Bensoussan</B700Sa>
  <B700Sb>Alain</B700Sb>
  <B700Sf>1951-....</B700Sf>
  <B700Sc>Avocat</B700Sc>
  <B700S4>070</B700S4>
</BB700>
Solr+SIREN document to index

Structure:
Fields grouped by holding

<EE>
  <B915>
    <B915-ind1></B915-ind1>
    <B915-ind2></B915-ind2>
    <B915S5>914712101:402653661</B915S5>
    <B915Sa>9530403382</B915Sa>
  </B915>
(...)
  <B930>
    <B930-ind1></B930-ind1>
    <B930-ind2></B930-ind2>
    <B930S5>914712101:402653661</B930S5>
    <B930Sb>914712101</B930Sb>
    <B930Sa>346.01 BEN inf</B930Sa>
    <B930Sj>u</B930Sj>
  </B930>
</EE>
{  
  "boolean" : {  
    "clause" : [  
      {  
        "node" : {  
          "attribute" : "B200Sa",  
          "query" : "voltaire"  
        }  
      },  
      {  
        "twig" : {  
          "root" : "B701",  
          "child" : [  
            {  
              "occur" : "MUST",  
              "node" : {  
                "attribute" : "B700Sa",  
                "query" : "Chveïk"  
              }  
            },  
            {  
              "occur" : "MUST",  
              "node" : {  
                "attribute" : "B700Sb",  
                "query" : "Josef"  
              }  
            }  
          ]  
        }  
      }  
    ]  
  }  
}
Siren is 2018 Irish Startup of the Year

Published: Friday June 1st, 2018

Share

SEARCH ARTICLES
Search...

EXPLORE OUR TOPICS
Announcements
Case Study
General
Release

LATEST TWEETS

Vladimir Tichy, Managing Director-Schneider Electric Ireland; John Randles, CEO-Siren & John Campbell, Marketing Manager-Siren

Ireland’s National Tech Excellence awards represent the most prestigious recognition of achievement in the technology sector. Former winners have included the Collison brothers of Stripe