



národní
úložiště
šedé
literatury

Solr Total

Nicolas, Yann
2018

Dostupný z <http://www.nusl.cz/ntk/nusl-380932>

Dílo je chráněno podle autorského zákona č. 121/2000 Sb.

Licence Creative Commons Uveďte původ-Neužívejte komerčně-Nezpracovávejte 4.0

Tento dokument byl stažen z Národního úložiště šedé literatury (NUŠL).

Datum stažení: 27.04.2024

Další dokumenty můžete najít prostřednictvím vyhledávacího rozhraní nusl.cz .

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

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

SOLIR TOTAL
REQUETE SUR LE FORMAT D'EXPORT (UNX)

RECHERCHER

- Exemple
- Phrase
- Phrase

AND

Plus...

NOTICE DE

0

50

TRI

asc

SUR

RESULTAT

- EXPORTER EN EXCEL
- AFFICHER EN XML

Enregistrer la requete

Charger une requete

id

id

- B200SE
- B200SK
- B200SL
- B200SP
- B200ST
- B200SU
- B200Sa**
- B200Sb
- B200Sc
- B200Sd
- B200Se
- B200Sf
- B200Sg
- B200Sh
- B200Si
- B200Sk
- B200Sl
- B200Sm
- B200So
- B200Sr

ETTES

AFFICHER LES ZONES

- id
- ppn_z
- A001
- A003
- A004

FACETTES



Non flat MARC vs Flat indexes

Search	Result
Firstname = Josef Lastname = Chvéik	Record with two authors : <ul style="list-style-type: none">• Josef Colomb• Jean-Jacques Chvéik

Data [[a1, b1], [a2, b2]] #repetableFields

Index (a1, a2, b1, b2) #flatList

OK, we could enrich the index this way : (a1, a1b1, a2, a2b2, b1, b2)

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>
```

QUERY

```
{
  "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"
          }
        }
      }
    ]
  }
}
```

Field

Subfield in the same field

Subfield in the same field



2018
Tech
EXCELLENCE AWARDS
WINNER

Siren is 2018 Irish Startup of the Year

Published: Friday, June 1st, 2018

Share



Startup Award 2018
Siren



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 Collision brothers of

SEARCH ARTICLES

EXPLORE OUR TOPICS

[Announcements](#)
[Case Study](#)
[General](#)
[Release](#)

LATEST TWEETS

 @SirenResearch · Wed Jun 06
 Join us for Siren 10 launch Webinar on June 12th, 4pm BST. We'll talk about #InvestigativeIntelligence, Siren 10 Federation technology, Siren Distributed Join Technology and more. Register here: <https://t.co/hajCYn1rNam> #Siren10 #BigDataAnalytics #KnowledgeGraphs <https://t.co/zlIQn6C3mt>