Koha...
International challenges

Henri-Damien LAURENT

BibLibre

23 février 2009

1. Introduction
   - A little bit of geography
1. Technical challenges
   - Diacritics Encoding
   - Different MARC
1. Human Challenges
   - Academic/Public use
   - User skills
   - Coder skills or habits
1. History Challenges
1. Conclusion

Some figures

Portugal

Population

- 2008 estimate **10,676,910** (77^{th})
- 2001 census 10,355,824

Area

- Total 92,345 km^2 (110^{th})

Some figures

Portugal

Population
- 2008 estimate **10,676,910** (77^{th})
- 2001 census 10,355,824

Area
- Total 92,345 km^2 (110^{th})

Michigan

Population Ranked 8^{th} in the US
- Total **10,003,422** (2008 est.)

Area
- Total 97,990 sq mi (253,793 km^2)

A little drawing

Texas

France

A little drawing

Texas

France

A little drawing

Texas

696,241 km^2

France

674,843 km^2

Some global figures

Federation	Countries	Population (Millions)	Area(km^2)	Official Languages
United States	52	306	9,826,630	7
European Union	25	494.8	4,422,773	23

American Context

No diacritics

American Context

No diacritics

European Context

many languages use diacritics

American Context

No diacritics

European Context

many languages use diacritics

But utf8 has more than one way to code it

For instance é can be

American Context

No diacritics

European Context

many languages use diacritics
But utf8 has more than one way to code it
For instance é can be

- coded on one byte 0xe9

American Context

No diacritics

European Context

many languages use diacritics
But utf8 has more than one way to code it
For instance é can be

- coded on one byte 0xe9
- coded on two bytes using NFC 0xc3a9

American Context

No diacritics

European Context

many languages use diacritics
But utf8 has more than one way to code it
For instance é can be

- coded on one byte 0xe9
- coded on two bytes using NFC 0xc3a9
- coded on two bytes using NFD 0xa9c3

American Context

No diacritics

European Context

many languages use diacritics
But utf8 has more than one way to code it
For instance é can be

- coded on one byte 0xe9
- coded on two bytes using NFC 0xc3a9
- coded on two bytes using NFD 0xa9c3

But .... Searching should return all that matches ...

American Context

No diacritics

European Context

many languages use diacritics
But utf8 has more than one way to code it
For instance é can be

- coded on one byte 0xe9
- coded on two bytes using NFC 0xc3a9
- coded on two bytes using NFD 0xa9c3

But .... Searching should return all that matches ... 0x65

American Context

No diacritics

European Context

many languages use diacritics
But utf8 has more than one way to code it
For instance é can be

- coded on one byte 0xe9
- coded on two bytes using NFC 0xc3a9
- coded on two bytes using NFD 0xa9c3

But .... Searching should return all that matches ... 0x65
Jesus

Different MARC

American Context : MARC21

```
LDR 00892namaa2200265 a 4500
001     4004663
003     OSt
005     19960603133300.5
008     960403s1995    fr b          001 0 fre
010    a   96133130
020    a2738429319
245 00 aAnthologie d'introduction à la poésie
comorienne d'expression française /
       céditée et annotée par Carole Beckett.
300     a251 p. :
       bmaps ;
       c22 cm.
490 1  aCollection Lettres de l'océan Indien
650  0 aComorian poetry (French)
700 1  aBeckett, Carole.
```

European Context : UNIMARC

```
LDR 00806cam a2200229    4500
010     a2738429319
       bbr.
100     a20090222              frey50
200 1  aAnthologie d'introduction à la poésie
comorienne d'expression française
       bMONO
       féd. et annot. par Carole Beckett
215     a251 p.
       d22 cm
410  0 034277200
       tLettres de l'Océan indien.
       x1242-7640
       d1995
702     3123141
       aBeckett
       bCarole
       4340
```

Different MARC

American Context : MARC21

```
LDR 00892namaa2200265 a 4500
001     4004663
003     OSt
005     19960603133300.5
008     960403s1995    fr b          001 0 fre
010    a   96133130
020    a2738429319
245 00 aAnthologie d'introduction à la poésie
comorienne d'expression française /
       céditée et annotée par Carole Beckett.
300    a251 p. :
       bmaps ;
       c22 cm.
490 1  aCollection Lettres de l'océan Indien
650  0 aComorian poetry (French)
700 1  aBeckett, Carole.
```

European Context : UNIMARC

```
LDR 00806cam a2200229    4500
010    a2738429319
       bbr.
100    a20090222              frey50
200 1  aAnthologie d'introduction à la poésie
comorienne d'expression française
       bMONO
       féd. et annot. par Carole Beckett
215    a251 p.
       d22 cm
410  0 034277200
       tLettres de l'Océan indien.
       x1242-7640
       d1995
702    3123141
       aBeckett
       bCarole
       4340
```

Different MARC

American Context : MARC21

```
LDR 00892namaa2200265 a 4500
001     4004663
003     OSt
005     19960603133300.5
008     960403s1995    fr b          001 0 fre
010    a   96133130
020    a2738429319
245 00 aAnthologie d'introduction à la poésie
comorienne d'expression française /
       céditée et annotée par Carole Beckett.
300     a251 p. :
       bmaps ;
       c22 cm.
490 1  aCollection Lettres de l'océan Indien
650  0 aComorian poetry (French)
700 1  aBeckett, Carole.
```

European Context : UNIMARC

```
LDR 00806cam a2200229    4500
010    a2738429319
       bbr.
100    a20090222              frey50
200 1  aAnthologie d'introduction à la poésie
comorienne d'expression française
       bMONO
       féd. et annot. par Carole Beckett
215    a251 p.
       d22 cm
410  0 034277200
       tLettres de l'Océan indien.
       x1242-7640
       d1995
702    3123141
       aBeckett
       bCarole
       4340
```

Different MARC

American Context : MARC21

```
LDR 00892namaa2200265 a 4500
001     4004663
003     OSt
005     19960603133300.5
008     960403s1995    fr b          001 0 fre
010    a   96133130
020    a2738429319
245 00 aAnthologie d'introduction à la poésie
comorienne d'expression française /
       céditée et annotée par Carole Beckett.
300     a251 p. :
       bmaps ;
       c22 cm.
490 1  aCollection Lettres de l'océan Indien
650  0 aComorian poetry (French)
700 1  aBeckett, Carole.
```

European Context : UNIMARC

```
LDR 00806cam a2200229    4500
010     a2738429319
       bbr.
100     a20090222              frey50
200 1  aAnthologie d'introduction à la poésie
comorienne d'expression française
       bMONO
       féd. et annot. par Carole Beckett
215     a251 p.
       d22 cm
410  0 034277200
       tLettres de l'Océan indien.
       x1242-7640
       d1995
702     3123141
       aBeckett
       bCarole
       4340
```

Different MARC

American Context : MARC21

```
LDR 00892namaa2200265 a 4500
001     4004663
003     OSt
005     19960603133300.5
008     960403s1995    fr b          001 0 fre
010    a   96133130
020    a2738429319
245 00 aAnthologie d'introduction à la poésie
comorienne d'expression française /
       céditée et annotée par Carole Beckett.
300     a251 p. :
       bmaps ;
       c22 cm.
490 1  aCollection Lettres de l'océan Indien
650  0 aComorian poetry (French)
700 1  aBeckett, Carole.
```

European Context : UNIMARC

```
LDR 00806cam a2200229    4500
010     a2738429319
       bbr.
100     a20090222              frey50
200 1  aAnthologie d'introduction à la poésie
comorienne d'expression française
       bMONO
       féd. et annot. par Carole Beckett
215     a251 p.
       d22 cm
410  0 034277200
       tLettres de l'Océan indien.
       x1242-7640
       d1995
702     3123141
       aBeckett
       bCarole
       4340
```

Different MARC

American Context : MARC21

LoC has released many tools to manipulate data :

- MARC21slim Basic stylesheets
- MARC conversions to MODS DC RDF

Different MARC

American Context : MARC21

LoC has released many tools to manipulate data :

- MARC21slim Basic stylesheets
- MARC conversions to MODS DC RDF

European Context : UNIMARC

- UNIMARC : very little tools

Different MARC

American Context : MARC21

LoC has released many tools to manipulate data :

- MARC21slim Basic stylesheets
- MARC conversions to MODS DC RDF

European Context : UNIMARC

- UNIMARC : very little tools
- UKMARC...

Different MARC

American Context : MARC21

LoC has released many tools to manipulate data :

- MARC21slim Basic stylesheets
- MARC conversions to MODS DC RDF

European Context : UNIMARC

- UNIMARC : very little tools
- UKMARC...
- FINMARC...

Different MARC

American Context : MARC21

LoC has released many tools to manipulate data :

- MARC21slim Basic stylesheets
- MARC conversions to MODS DC RDF

European Context : UNIMARC

- UNIMARC : very little tools
- UKMARC...
- FINMARC...
- IBERMARC....

Different usage requirements

Google box effect

Users are quite used to single box queries and like it,

Different usage requirements

Google box effect

Users are quite used to single box queries and like it,

provided they have results to their queries.

Different usage requirements

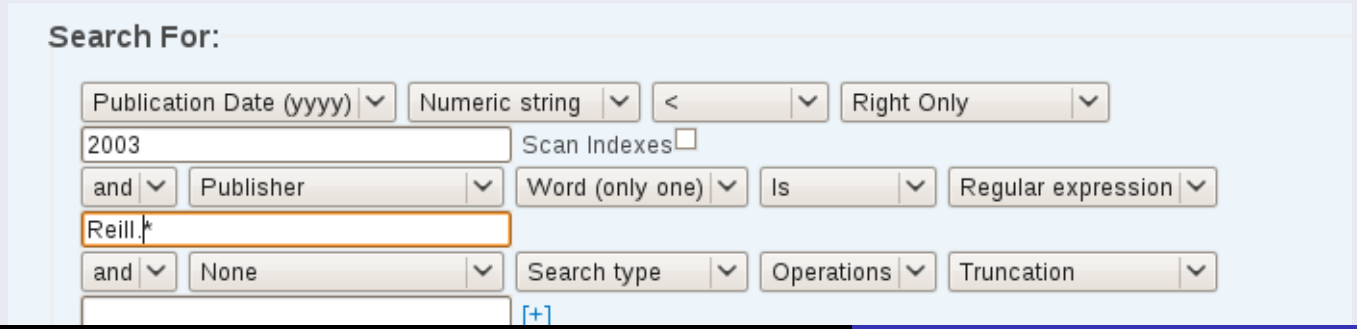
Google box effect

Users are quite used to single box queries and like it,

provided they have results to their queries.

Academic Users : Scholars doing precise queries

Need to search a specific field for information with all the booleans and operators



Different usage requirements

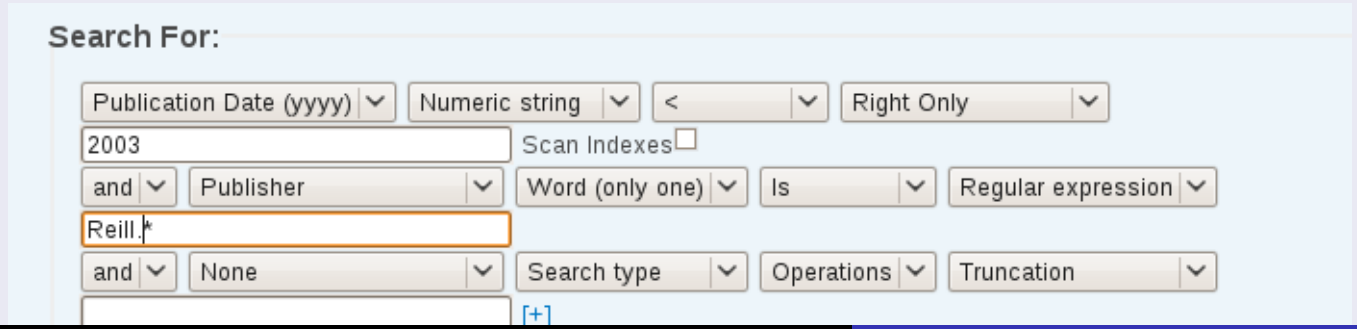
Google box effect

Users are quite used to single box queries and like it,

provided they have results to their queries.

Academic Users : Scholars doing precise queries

Need to search a specific field for information with all the booleans and operators



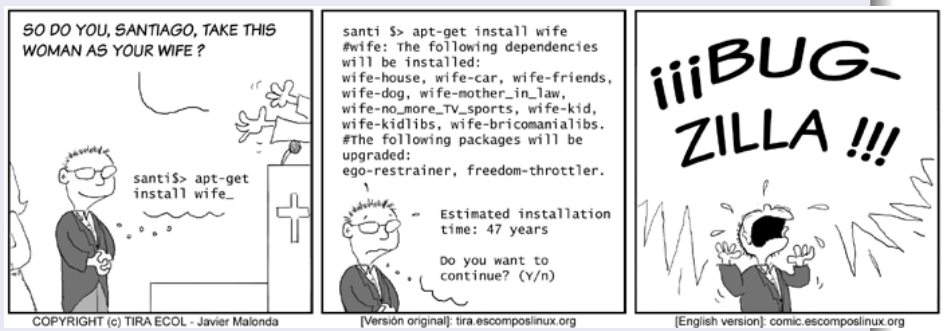
User Skills

Rookie

User Skills

Rookie

Geeks



Coders skills or habits

- procedural
- passing arrays of parameters to functions
- Void return functions
- die soon in web context
- even new coder is welcome

- OOP
- passing hashes to functions
- functions returns at least a status
- error Proof application
- Strong Coding Guidelines

History Challenges

- Some historical features are transformed owed to improved support of standards
- Itemtype can now be on biblio level or on item level
- Some data are synched between marc containers and koha tables
- late written automated tests
- lack of global coordination and mutualisation of coding efforts

Conclusion

Koha Community Has to grow up and get organized.