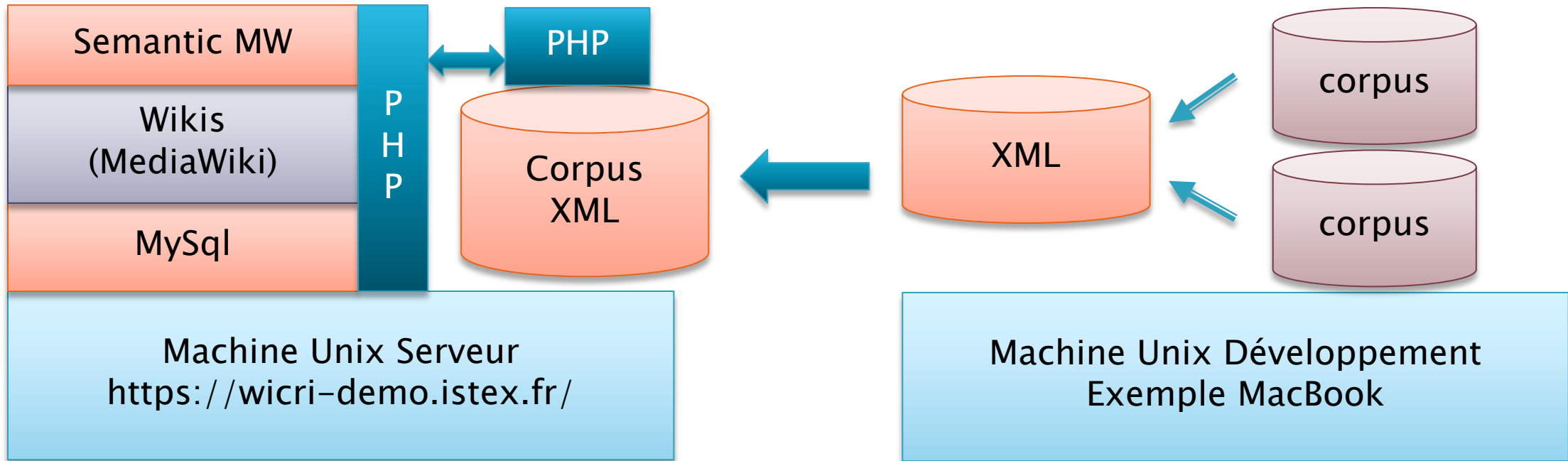


Machine serveur – machine développement

Unix, langage C, PHP, XML, JSON, etc...



Dilib, une boîte à outils Sxml

- ▶ SXML : XML lite (mais JSON+)
 - Compatible avec les outils Unix
 - Un document = Une ligne Unix
- ▶ Origine : Un Lego pour les corpus
 - 1990 : Ilib : ISO 2709 (MARC, Pascal...)
- ▶ Formats MARC



210 \$a Paris \$c Dunod \$d 1988

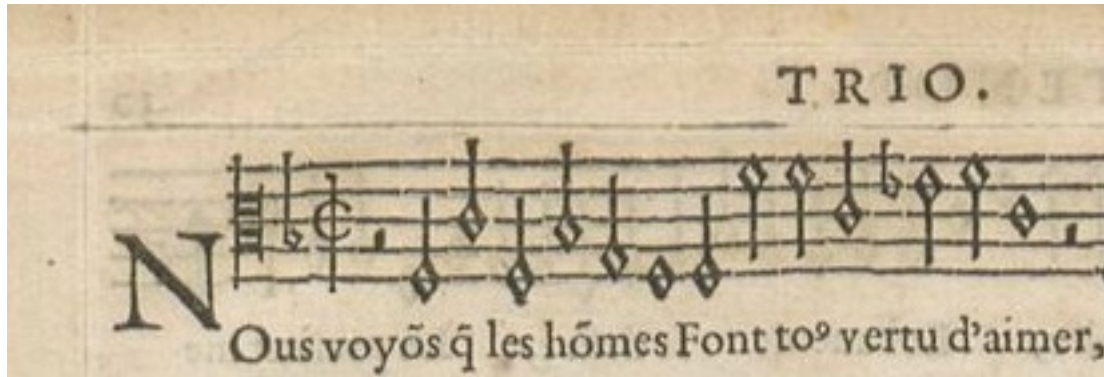
```
<unimarc>...<f210><sa>Paris</sa><sc>Dunod</sc>  
<sd>1988</sd><f210>...</unimarc>
```

- ▶ Fichiers inverses

```
<idx><loc>Paris</loc><f>2</f>  
<l><e>0023</e><e>4123</e></l></idx>
```

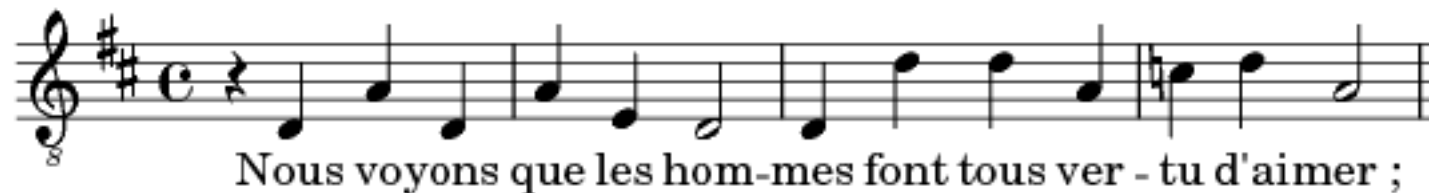
```
<index>  
  <kw>Requiem</kw>  
  <list>  
    <item>004321</item>  
    <item>012345</item>  
  </list>  
  <f>2</f>  
</index>
```

Préfiguration : analyseurs syntaxiques (Lex)



```
\time 4/4 \key f \major  
r4 f4 c' f,4  
c'4 g f2  
f4 f' f c  
ees4 f c2
```

```
%%  
a      printf ("fis");  
bes    printf ("g");  
c      printf ("a");  
d      printf ("b");  
e      printf ("cis");  
ees    printf ("c");  
f      printf ("d");  
g      printf ("e");  
%%  
main()  
{  
    yylex();  
}
```



XML et Unix : variante Sxml

▶ Unix :

- un système pour manipuler des lignes de texte
- Exemple : combien de lignes contiennent le terme Europium dans mon article ?

```
cat monDoc | grep [Ee]uropium | wc
```

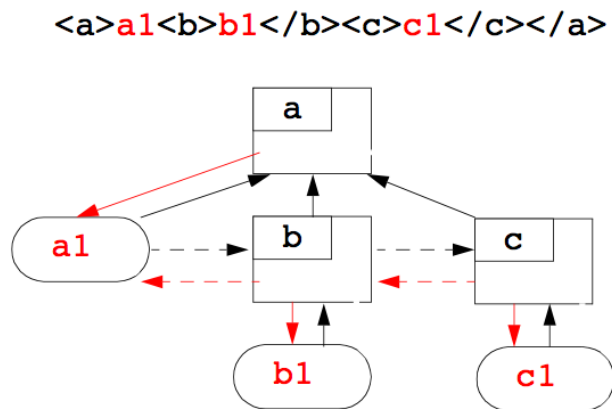
▶ Stream Xml

- Un document XML =
 - une ligne (pas de saut de lignes)
- Combien de documents contiennent le terme Europium dans mon corpus ?

```
cat monCorpus | grep [Ee]uropium | wc
```

DILIB – parser Sxml

▶ Document Object Model (DOM)



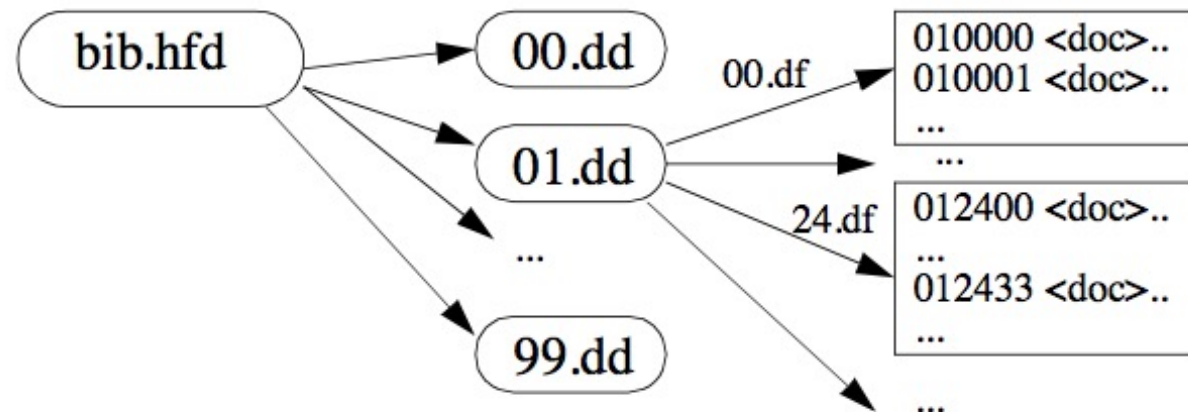
```
/* création de <a><b>b1</b><c>c1</c></a> */  
  
#include "SxmlNode.h"  
main()  
{  
    SxmlNode *root;  
    root =SxmlElementCreate("a");  
    SxmlAppendChild (root, SxmlLeafCreate("b", "b1"));  
    SxmlAppendChild (root, SxmlLeafCreate("c", "c1"));  
    SxmlPrint(root);  
    putchar('\n');  
    exit(0);  
}
```

Dilib – Parser DOM

- ▶ **Parser classique**
 - Libxml, Java, php...
 - Manipule un (ou quelques) éventuellement gros documents.
- ▶ **Parser Dilib Sxml**
 - Manipule de très grandes collections de petits documents,
 - Plus souple d'emploi (espaces de nom)
 - Extensions possibles (types numériques)
 - Peu de restrictions / norme
 - Très peu de restrictions via filtres
- ▶ **La philosophie Dilib est indépendante de Dilib**

Dilib HFD

- ▶ Architecture XML pour construire des systèmes de recherche d'information en kit.
- ▶ Hierarchy File Document
 - $1.000.000 = 100 * 100 * 100$
 - $012433 \rightarrow 01 / 24 / 33$
 - `bib.hfd/01.dd/24.fd`



DILIB – HFD – Fichiers inverses

Un exemple Serveur d'exploration sur l'Europium à partir du wiki Wicri/Terre

[http://ticri.univ-lorraine.fr/Wicri/Terre/
corpus/Europium/EuropiumV1/Data/Main/Exploration/
AffRegion.i.hfd/00.dd/00.df](http://ticri.univ-lorraine.fr/Wicri/Terre/corpus/Europium/EuropiumV1/Data/Main/Exploration/AffRegion.i.hfd/00.dd/00.df)

```
000000 <idx><kw>Alsace</kw><lc>alsace</lc><f>3</f><l><e>000070</e><e>000136
        </e><e>000175</e></l></idx>
000001 <idx><kw>Angleterre</kw><lc>angleterre</lc><f>7</f>
        <l><e>000046</e><e>000091</e><e>000107</e><e>000183</e>
        <e>000221</e><e>000233</e><e>000251</e></l></idx>
000002 <idx><kw>Aquitaine</kw><lc>aquitaine</lc><f>1</f><l><e>000117</e></l></idx>
000003 <idx><kw>Bavière</kw><lc>bavière</lc><f>1</f><l><e>000173</e></l></idx>
000004 <idx><kw>Berlin</kw><lc>berlin</lc><f>2</f>
        <l><e>000005</e><e>000231</e></l></idx>
```


Complémentarité Wiki –Sxml

► Commandes liées aux tables

```
WicriGetPage -l wicri-afrique.fr -p "Wicri:Liste de localisation de pays d'Afrique" \  
MediaWikiCleanTable \  
MediaWikiTable2SxmlRowCol \  
MediaWikiTableTransformCol -l1 -E2 \  
TeiTableRegionFromWiki
```



wicri discussion modifier historique suppri

Jacques Ducloy page

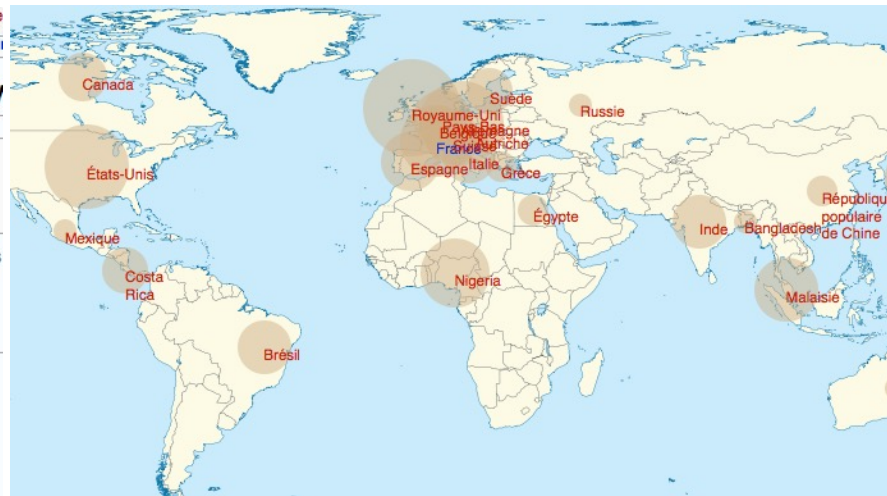
Wicri:Liste de localisation de pay

 page en construction

Cette page introduit une liste de coordonnées géographiques d'Afrique.

La liste

Pays	Barycentre
Afrique du Sud	geo : -30.559482 22.937506
Ghana	geo : 7.68 -0.98
Niger	geo : 17.607789 8.081666
Zimbabwe	geo : -19.015438 29.154857
Nigeria	geo : 9.081999 8.675277
Maroc	geo : 31.791702 -7.09262
Algérie	geo : 28.033886 1.659626
Tunisie	geo : 33.886917 9.537499
Égypte	geo : 26.820553 30.802498



```
<figure>  
<link target="Ghana" />  
<head offset="se">Ghana</head>  
<geo latitude="7.68" longitude="-0.98">7.68 -0.98</geo>  
</figure>
```