





Programmation et projet encadré Boîte à outils Série 3 : extraction de patrons









Extraction de patrons

Boîte à outils : série 3





Objectif

- Extraire des patrons dans les sorties produites à l'issue de la *BàO série* 2
- 2 méthodes
 - Un programme qui prend en entrée les sorties «brutes » de l'étiquetage
 - En particulier la sortie obtenue avec Cordial
 - Requête XPATH sur le fichier XML construit
 - En particulier avec *treetagger* et le programme treetagger2xml





Travail personnel [*série* 3]:

extraction de patrons dans les contenus des fils

- Objectif :
 - Vous devez construire une liste de patrons à partir des contenus textuels des fils traités
 - Liste des patrons à extraire :
 - NOM ADJ
 - NOM NOM
 - NOM PREP NOM
 - En sortie, une liste de patrons au format TXT
 - Cette liste sera le point d'entrée de la BàO série 4





Phase 1

- Extraction des patrons via XPath
- En entrée, la <u>sortie XML</u> produite dans la *BàO série 2*
 - Celle construite avec *treetagger* et reformatée par le programme fourni treetagger2xml
- Écriture d'une requête pour extraire les éléments correspondant au patron...





| xml version="1.0" encoding="iso-8859-1" ? |
|--|
| xml-stylesheet type="text/xsl" href="SORTIE-etiquetage-xsl.xsl"? |
| <parcours></parcours> |
| <nom>Votre nom</nom> |
| <etiquetage><file></file></etiquetage> |
| <name>arbo-fils/2005/Nov/4/01-00-00/0,2-3208,1-0,0.xml</name> |
| <element><data type="type">VER:pres</data><data type="lemma">retrouver</data><data type="string">Retrouvez</data></element> |
| <e]ement><data type="type">NUM</data><data type="]emma">L</data><data type="string">l</data></e]ement> |
| <element><data type="type">PUN</data><data type="lemma">'</data><data type="string">'</data></element> |
| <element><data type="type">ADV</data><data type="lemma">ensemble</data><data type="string">ensemble</data></element> |
| <e]ement><data type="type">PRP:det</data><data type="lemma">du</data><data type="string">des</data></e]ement> |
| <element><data type="type">NOM</data><data type="lemma">dépêche</data><data type="string">dépêches</data></element> |
| <element><data type="type">PRP</data><data type="lemma">sur</data><data type="string">sur</data></element> |
| <element><data type="type">NOM</data><data type="lemma">http</data><data type="string">http</data></element> |
| <element><data type="type">PUN</data><data type="lemma">:</data><data type="string">:</data></element> |
| <element><data type="type">PUN</data><data type="lemma">/</data><data type="string">/</data></element> |
| <element><data type="type">PUN</data><data type="lemma">/</data><data type="string">/</data></element> |
| <element><data type="type">NOM</data><data type="lemma">www</data><data type="string">www</data></element> |
| <element><data type="type">SENT</data><data type="lemma">.</data><data type="string">.</data></element> |
| <pre><element><data type="type">NOM</data><data type="lemma">lemonde</data><data type="string">lemonde</data></element></pre> |
| <pre><element><data type="type">SENT</data><data type="lemma">.</data><data type="string">.</data></element></pre> |
| <pre><clement><data type="type">NOM</data><data data="" ptr<="" type="lemma"><data type="string">tr</data></data></clement></pre> |
| |
| |
| <pre>chame>arbn=t1ls//UU5/NDV/4/UL=UU=UU/U./=3/LU.L=U.U.Xml</pre> |
| colomosta (data) tumo "tumo" DET: ADT (data) tumo "loc (data) (data) (data) (data) (data) (data) (data) (data) |
| <pre>kelement> {data type "type">DET:ART {/data> {data type "lemma">le {data type "string">Le {/element> celement><data type="string">Le</data> {/element></pre> |
| <pre>kelement> {data type "type">DET:ART {/data> {data type "lemma">le {/data> {data type "string">Le {/data> {/element></pre> |
| <pre>celement> {data type "type">DET:ANT {/data> {data type "lemma">le {/data> {data type "string">Le {/data> {/element> <element><data type="type">NOM</data><data type="lemma">désastre</data><data type="string">désastre</data></element> <element><data type="type">ADJ</data><data type="lemma">écologique</data><data type="string">désastre</data></element> <element><data type="type">PRF</data><data type="lemma">écologique</data><data type="string">désastre</data></element></pre> |
| <pre>celement> {data type "type">DET:ART {/data> {data type "lemma">le {/data> {data type "string">Le {/data> {/element> celement><data type="type">NOM</data><data type="lemma">désastre</data><data type="string">désastre</data> celement><data type="type">ADJ</data><data type="lemma">écologique</data><data type="string">écologique</data> celement><data type="type">ADJ</data><data type="lemma">écologique</data><data type="string">écologique</data> celement> <data type="type">PRP </data> <data type="lemma">de </data> <data type="string">de </data> celement> <data type="type">DET:ART </data> <data type="lemma">de </data> <data type="string">de </data> </pre> |
| <pre>celement> {data type "type">DET:ADT {/data> {data type "lemma">le {/data> {data type "string">Le {/data> {/element> <element><data type="type">NOM</data><data type="lemma">désastre</data><data type="string">désastre</data></element> <element><data type="type">ADJ</data><data type="lemma">écologique</data><data type="string">écologique</data></element> <element><data type="type">ADJ</data><data type="lemma">écologique</data><data type="string">écologique</data></element> <element><data type="type">PRP </data><data type="lemma">de </data> <data type="string">de </data> </element> <element><data type="type">DET:ADT</data> <data type="lemma">le </data> <data type="string">data> </data></element> <element><data type="type">DET:ADT</data> <data type="lemma">le </data> <data type="string">la </data> </element> <element><data type="type">NOM</data> </element></pre> |
| <pre>celement> {data type "type">DET:ADT {/data> {data type "lemma">le {/data> {data type "string">Le {/data> {/element> <element><data type="type">NOM</data><data type="lemma">désastre</data><data type="string">désastre</data> </element> <element><data type="type">ADJ</data><data type="lemma">desastre</data><data type="string">désastre</data></element> <element><data type="type">ADJ</data><data type="lemma">desastre</data><data type="string">desastre</data></element> <element><data type="type">ADJ</data><data type="lemma">des/data> <data type="string">des/data> </data></data></element> <element><data type="type">ADJ</data></element> <element><data type="type">ADJ</data> <data type="lemma">des/data> <data type="string">des/data> </data></data></element> <element><data type="type">ADJ</data> <data type="lemma">lo</data> <data type="string">data> </data></element> <element><data type="type">ADJ</data> <data type="lemma">lo</data> <data type="string">adata type="string">la</data> </element> <element><data type="type">ADJ</data> <data type="lemma">lemma">Songhua</data> <data type="string">string">string">Songhua</data> </element></pre> |
| <pre>celement> {data type "type">DET:ADT {/data> {data type "lemma">le {/data> {data type "string">Le {/data> {/element> <element><data type="type">NOM</data><data type="lemma">désastre</data><data type="string">désastre</data> </element> <element><data type="type">ADJ</data><data type="lemma">désastre</data><data type="string">désastre</data></element> <element><data type="type">ADJ</data><data type="lemma">de </data><data type="string">de </data> </element> <element><data type="type">DET:ADT</data><data type="lemma">de </data> <data type="string">de </data> </element> <element><data type="type">DET:ADT</data> <data type="lemma">le</data> <data type="string">data> </data></element> <element><data type="type">DET:ADT</data> </element> <element><data type="type">DET:ADT</data> <data type="lemma">le</data> <data type="string">data> </data></element> <element><data type="type">DET:ADT</data> </element> <element><data type="type">DET:ADT</data> </element> <element><data type="type">DET:ADT</data> <data type="lemma">le</data> </element> <element><data type="type">DET:ADT</data> </element> <element> </element></pre> |
| <pre>celement> {data type "type">DET:ADT {/data> {data type "lemma">le {/data> {data type "string">Le {/data> {/element> <element><data type="type">NOM</data><data type="lemma">désastre</data><data type="string">Le {/data> {/element> <element><data type="type">ADJ</data><data type="lemma">désastre</data><data type="string">désastre</data></element> <element><data type="type">ADJ</data><data type="lemma">de {/data> <data type="string">de {/data> </data> </data></element> <element><data type="type">DET:ADT</data> <data type="lemma">de {/data> <data type="string">de {/data> </data></data></element> <element><data type="type">DET:ADT</data> <data type="lemma">lo</data> <data type="string">data> </data></element> <element><data type="type">DET:ADT</data> <data type="lemma">lo</data> <data type="string">data> </data></element> <element><data type="type">NOM</data> </element> <element><data type="type">NOM</data> <data type="lemma">lo</data> <data type="string">lo</data> </element> <element><data type="type">NOM</data> </element> <element><data type="type">NOM</data> </element> <element><data type="type">NOM</data> <data type="lemma">lo</data> <data type="string">string</data></element></data></element></pre> |
| <pre>celement> <data "type"="" type="">DET:ANT </data> <data "lemma"="" type="">le</data> <data "string"="" type="">Le</data> <element><data type="type">NOM</data><data type="lemma">désastre</data><data type="string">Le</data> </element> <element><data type="type">ADJ</data> <data type="lemma">désastre</data> <data type="string">désastre</data> </element> <element> <data type="type">ADJ</data> <data type="lemma">de</data> <data type="string">de</data> </element> <element> <data type="type">NOM</data> </element> <element> <data type="type">NOM</data> <data type="lemma">de</data> <data type="string">de</data> </element> <element> <data type="type">NOM</data> </element> <element> <data type="type">NOM</data> </element> <element> <data type="type">NOM</data> </element> <element> <data type="type">String">Songhua</data> </element> <element> <data type="type">Songhua</data> </element> <element> <data type="type">Songhua</data> </element> </pre> |
| <pre>celement> data type "type">DET:ANT data type "lemma">le data type "string">Le celement><data type="type">NOM</data><data type="lemma">désastre</data><data type="string">désastre</data> celement><data type="type">ADJ</data><data type="lemma">désastre</data><data type="string">désastre</data> celement><data type="type">NOM</data><data type="lemma">des/data><data type="string">des/data> celement><data type="type">NOM</data><data type="lemma">des/data><data type="string">des/data> celement><data type="type">NOM</data><data type="lemma">des/data><data type="string">la</data> celement><data type="type">NOM</data><data type="lemma">des/data><data type="string">la</data> celement><data type="type">NOM</data></data><data type="lemma">lo</data><data type="string">la</data> celement><data type="type">NOM</data> celement><data type="type">NOM</data></data>lo</data><data type="string">string">la</data> celement><data type="type">Songhua</data> celement><data type="type">Songhua</data> celement><data type="type">Songhua</data> celement><data type="type">DET:ANT</data> celement><data type="type">Songhua</data> celement><data type="type">DET:ANT</data> celement><data type="type">Songhua</data> celement><data type="type">DET:ANT</data> celement><data type="type">Songhua</data> celement><data type="type">DET:ANT</data> celement><data type="type">DET:ANT</data> celement><data type="type">Songhua</data> celement><data type="type">DET:ANT</data> celement><data type="type">DET:ANT</data> celement><data type="type">DET:ANT</data> celement><data type="type">DET:ANT</data> celement><data type="type">DET:ANT</data> celement><data type="type">DET:ANT</data> celement><data type="type">DET:ANT</data> celement></data></data></data></pre> |
| <pre>celement><data "type"="" type="">DET:Ant</data><data "lemma"="" type="">le</data><data "string"="" type="">Le</data> <<lerement><data type="type">NOM</data><data type="lemma">desastre</data><data type="string">desastre</data> <<lerement><data type="type">NOM</data><data type="lemma">desastre</data><data type="string">desastre</data> <<lerement><data type="type">NOM</data>des/data><data type="string">des/data><data type="string">des/data> <<lerement><data type="type">NOM</data></lerement></data> <<lerement><data type="type">NOM</data></lerement></data></lerement></lerement></lerement></pre> |
| <pre>celement><data "type"="" type="">DET:AnT</data><data "lemma"="" type="">le</data><data "string"="" type="">Le</data></pre> |
| <pre>celement><data "type"="" type="">DET:ATT/data><data "lemma"="" type="">le</data><data "string"="" type="">Le</data> <element><data type="type">ADJ</data><data type="lemma">désastre</data><data type="string">désastre</data></element> <element><data type="type">ADJ</data><data type="lemma">de</data><data type="string">de</data></element></data> <element><data type="type">DET:ATT</data><data type="lemma">de</data><data type="string">data></data></element> <element><data type="type">DET</data></element> <element><data type="type">NOM</data><data type="lemma">de</data><data type="string">data></data></element> <element><data type="type">DET</data></element> <element><data type="type">NOM</data><data type="lemma">de</data><data type="string">data></data></element> <element><data type="type">NOM</data><data type="lemma">data type="string">initiere</data></element> <element><data type="type">NOM</data><data type="lemma">data type="string">initiere</data></element> <element><data type="type">NOM</data><data type="lemma">data type="string">initiere</data></element> <element><data type="type">NOM</data>data type="string">initiere</element> <element><data type="type">NOM</data>data type="string">initiere</element> <element><data type="type">NOM</data></element> <element><data type="type">NOM</data>data type="lemma">initiere</element> <element><data type="type">NOM</data>data type="lemma">initiere</element> <element><data type="type">NOM</data>data type="lemma">initiereinitiere</element> <element><data type="type">NOM</data>data type="lemma">initiereinitiere</element> <element><data type="type">NOM</data>initiereinitiere</element> <element><data type="type">NOM</data>initiereininitiere</element></pre> |
| <pre>celement><data "type"="" type="">DET:ANT</data><data "lemma"="" type="">de</data><data "type"="" type="">Let/data></data> <element><data type="type">NOM</data><data type="lemma">désastre</data><data type="string">Let/data></data></element> <element><data type="type">NOM</data><data type="lemma">désastre</data><data type="string">désastre</data></element> <clement><data type="type">NOM</data><data type="lemma">désastre</data><data type="string">désastre</data> <clement><data type="type">NOM</data><data type="lemma">désastre</data><data type="string">des/data></data> <clement><data type="type">NOM</data><data type="lemma">des/data><data type="string">data> <clement><data type="type">NOM</data><data type="lemma">des/data><data type="string">le</data> <clement><data type="type">NOM</data><data type="lemma">data><data type="string">string">string">string">le</data> <clement><data type="type">NOM</data><data type="lemma">data><data type="string">string">string">string">le</data> <clement><data type="type">NOM</data><data type="lemma">data></data><data type="string"></data></clement></data></clement></data></clement></data></clement></data></data></clement></clement></clement></pre> |
| <pre>celement><data "type"="" type="">DTIANT//dtta><data "lemma"="" type="">le//data><data "string"="" type="">Le//data></data> <celement><data type="type">NOM</data><data type="lemma">désastre</data><data type="string">désastre</data> <celement><data type="type">NDM</data></celement></celement></data>desastre</data><data type="string">désastre</data> <celement><data type="type">NDM</data>desastre<data type="string">desastre</data> <celement><data type="type">NDM</data>des/data><data type="string">des/data></data> <celement><data type="type">NDM</data><data type="lemma">data><data type="string">data><data type="string">data></data> <celement><data type="type">NDM</data><data type="lemma">data><data type="string">data><data type="string">data></data> <celement><data type="type">NDM</data><data type="lemma">data><data type="string">data><data type="string">data> <celement><data type="type">NDM</data><data type="lemma">data><data type="string">data><data< element=""> <celement><data type="type">NDM</data><data type="lemma">data><data type="string">data><data type="string">data> <celement><data type="type">NDM</data><data type="lemma">data><data type="string">data> <celement><data type="type">NDM</data><data type="lemma">data><data type="string">data> <celement><data type="type">NDM</data><data type="lemma">data><data type="string">data> <celement><data type="type">NDM</data><data type="lemma">data><data type="string">data> <celement><data type="type">NDM</data></celement></data></data></celement></data></data></celement></data></data></celement></data></data></celement></data></data></data></celement></data<> <celement><data type="type">NDM</data></celement></data></data>data></celement></data></data></data></celement></data> <celement><data type="type">NDM</data></celement></data></celement></data>data></data> <celement><data type="type">NDM</data><data></celement> <celement><data type="t</td></tr><tr><td><pre>celement><data type " type"="">DET+ART//data</data>le<data "string"="" type="">Le</data> celement><data type="type">NOM</data><data type="lemma">desastre</data><data type="string">desastre</data> celement><data type="type">ADJ</data><data type="lemma">desastre</data><data type="string">desastre</data> celement><data type="type">ADJ</data><data type="lemma">des/data><data type="string">desastre</data></data>des/data>des/data></celement></celement></celement></celement></pre> |
| <pre>celement><data type="type">DETIANT/data<data type="lemma">leLe</data></data></pre> |
| <pre>valement> data type "type">DET:ATT/data>data type "lemma">le/data> data type "string">Le/data> /element> celement><data type="type">NOM</data><data type="lemma">desastre</data><data type="string">desastre</data> celement><data type="type">Comparison of type "lemma">desastre</data><data type="string">desastre</data> celement><data type="type">Comparison of type">Comparison of type"</data></pre> |
| <pre>velement><data "type"="" type="">NOM</data><data type="lemma">lex/data><data type="type">tring">lex/data> celement><data type="type">NOM</data><data type="lemma">desastre</data><data type="string">desastre</data> <celement><data type="type">NOM</data><data type="lemma">desastre</data><data type="string">desastre</data> <celement><data type="type">NOM</data><data type="lemma">des/data><data type="string">des/data> <celement><data type="type">NOM</data><data type="lemma">des/data><data type="string">das/data> <celement><data type="type">NOM</data><data type="lemma">des/data><data type="string">das/data> <celement><data type="type">NOM</data><data type="lemma">data></data><data type="string">data> <celement><data type="type">NOM</data><data type="lemma">data></data><data type="string">data> <celemen< td=""></celemen<></data></celement></data></celement></data></celement></data></celement></data></celement></data></celement></data></celement></data></celement></data></data></celement></data></data></celement></data></data></celement></data></data></celement></data></data></celement></data></data></celement></data></data></celement></data></data></celement></celement></data></data></pre> |
| <pre>celement><data type="type">DTFINIT/dtaskdata type="lemma">le/datakdata type="string">le/datak/element> celement><data type="type">NOMde/datakdata type="string">de/datak/element> celement><data type="type">NOMde/datakdata type="string">de/datak/element> celement><data type="type">NOMde/datakdata type="string">de/datak/element> celement><data type="type">NOMde/datakdata type="string">de/datak/element> celement><data type="type">NOMde/datakdata type="string">rivière celement><data type="type">NOMde/datakdata type="string">rivière celement><data type="type">NOMdans/datakdata type="string">rivière celement><data type="type">NOMdans/datakdata type="string">rivière celement><data type="type">NOMdans/datakdata type="string">datak/element> celement><data type="type">NOMdans/datakdata type="string">datak/element> celement><data type="type">NOMdans/datakdata type="string">datak/element> celement><data type="type">NOMdans/datakdata type="string">datak/element> celement><data type="type">NOMdans/datakdata type="string">datakdata type="string">datak/element> celement><data type="type">NOMdans/datakdata type="string">datakdata string">les/datakdans/datakdata type="string">datakdatak celement><data type="type">NOMdans/datakdata type="string">datakdatakkdatakkdatakkdatakkdatakkdata type="string" celement><data type="type">NOMdatakdata type="string">datakdatakkdatakkdatakkdatakkdatakkdatakkdatakkdatakdatakkdatakkdatakkdatakkdatakkdatakkdatakkdatakkdatakk</data></data></data></data></data></data></data></data></data></data></data></data></data></data></data></data></data></pre> |
| <pre>celement><data type="type">DETINET</data><data type="lemma">lex/data><data type="string">Lex/data> celement><data type="type">NDM</data><data type="lemma">dex/data><data type="string">dex/data></data><data type="string">dex/data></data></data> celement><data type="type">NDM</data><data type="lemma">dex/data><data type="string">dex/data></data> celement><data type="type">NDM</data><data type="lemma">dex/data><data type="string">dex/data></data> celement><data type="type">NDM</data><data type="lemma">dex/data><data type="string">dex/data></data> celement><data type="type">NDM</data><data type="lemma">dex/data><data type="string">Songhua</data> celement><data type="type">NDM</data><data type="lemma">data><data type="string">Songhua</data><data type="string">Songhua</data> celement><data type="type">NDM</data><data type="lemma">data><data type="string">Songhua</data><data type="string">Cata> celement><data type="type">NDM</data><data type="lemma">data></data><data type="string">Cata> celement><data type="type">NDM</data><data type="lemma">Heinongiang</data><data type="string">string">Songhua</data> celement><data type="type">NDM</data><data type="lemma">Heinongiang</data><data type="string">string">string">Songhua</data></data> celement><data type="type">NDM</data><data type="lemma">te</data><data type="string">string</data></data></data></data></data></data></data></data></data></data></pre> celement> <data type="type">Songhua</data> celement> <data type="type">Songhua</data> celement> <data type="type">spent spent</data> |
| <pre>celement><data type="type">NOM</data><data type="lemma">desastre</data><data type="string">desastre</data> celement><data type="type">NOM</data><data type="lemma">desastre</data><data type="string">desastre</data> celement><data type="type">NOM</data><data type="lemma">desastre</data><data type="string">desastre</data> celement><data type="type">NOM</data><data type="lemma">des/data><data type="string">des/data></data> celement><data type="type">NOM</data><data type="lemma">des/data><data type="string">des/data></data> celement><data type="type">NOM</data><data type="lemma">des/data><data type="string">des/data></data> celement><data type="type">NOM</data><data type="lemma">des/data><data type="string">des/data></data> celement><data type="type">NOM</data><data type="lemma">das</data><data type="string">das/data></data></data> celement><data type="type">NOM</data><data type="lemma">das</data><data type="string">das</data></data> celement><data type="type">DET:ART</data><data type="lemma">das</data><data type="string">das</data> celement><data type="type">DET:ART</data><data type="lemma">das</data><data type="string">das</data> celement><data type="type">NOM</data><data type="lemma">data type="lemma">das</data></data></data> celement><data type="type">NOM</data><data type="lemma">data></data><data type="string">data></data> celement><data type="type">NOM</data><data type="lemma">data></data><data type="string">data> celement><data type="type">NOM</data><data type="lemma">data></data><data type="string">data> celement><data type="type">NOM</data><data type="lemma">data></data></data><data type="string">data> celement><data type="type">NOM</data></data><data type="lemma">data type="string">data></data> celement><data type="type">NOM</data></data><data type="lemma">data type="string">data></data> celement><data type="type">NOM</data></pre> |

Exemple de sortie issue de *BàO série 2*



La requête

(//element[./data[contains(text() , 'NOM')]]/followingsibling::element[1][./data[conta ins(text(), 'ADJ')]/data[3]) (//element[./data[contains(text(), 'ADJ')]]/precedingsibling::element[1][./data[conta ins(text(), 'NOM')]]/data[3]))

En rouge : on recherche parmi les éléments contenant la chaîne NOM, ceux qui sont suivis juste après par un élément contenant la chaîne ADJ, et pour ceux là on affiche le 3^{ème} nœud fils data (le mot)

En bleu : on recherche parmi les éléments contenant la chaîne ADJ, ceux qui sont précédés juste avant par un élément contenant la chaîne NOM, et pour ceux là on affiche le 3^{ème} nœud fils data (le mot)

L'opérateur | permet de rechercher les 2 types de nœuds visés...





Affichage partiel du résultat avec Cooktop

| noc | des: (/ | //eleme | nt[./ | data[o | ontain | s(text() | ,'NOM')]] | /foll | <pre>iowing-sibling::element[1][./data[contains(text(),'ADJ')]]) (//element[./data[contains(text(),'ADJ'</pre> |
|--|---------|---|--------|--------|---|--|-----------|--------|--|
| -</td <td>nodes</td> <td>; xpath</td> <td>- (//e</td> <td>lement</td> <td>[./dat</td> <td>a[contai</td> <td>ns(text()</td> <td>),'NOM</td> <td>1')]]/following-sibling::element[1][./data[contains(text(),'ADJ')]]) (//element[./data[contains(tex</td> | nodes | ; xpath | - (//e | lement | [./dat | a[contai | ns(text() |),'NOM | 1')]]/following-sibling::element[1][./data[contains(text(),'ADJ')]]) (//element[./data[contains(tex |
| <e]< td=""><td>ement:</td><td><data :<="" td=""><td>type=</td><td>"type'</td><td>>NOM<!--</td--><td>data><da< td=""><td>ta type='</td><td>'lemma</td><td>۱">désastre</td></da<></td></td></data><datā type="string">désastre</datā></td></e]<> | ement: | <data :<="" td=""><td>type=</td><td>"type'</td><td>>NOM<!--</td--><td>data><da< td=""><td>ta type='</td><td>'lemma</td><td>۱">désastre</td></da<></td></td></data> <datā type="string">désastre</datā> | type= | "type' | >NOM </td <td>data><da< td=""><td>ta type='</td><td>'lemma</td><td>۱">désastre</td></da<></td> | data> <da< td=""><td>ta type='</td><td>'lemma</td><td>۱">désastre</td></da<> | ta type=' | 'lemma | ۱">désastre |
| <e1< td=""><td>ement:</td><td>-<data :<="" td=""><td>type=</td><td>"type'</td><td>>ADJ<!--</td--><td>data><da< td=""><td>ta type='</td><td>'lemma</td><td>۱">écologique</td></da<></td></td></data><data`type="string">écologique</data`type="string"></td></e1<> | ement: | - <data :<="" td=""><td>type=</td><td>"type'</td><td>>ADJ<!--</td--><td>data><da< td=""><td>ta type='</td><td>'lemma</td><td>۱">écologique</td></da<></td></td></data> <data`type="string">écologique</data`type="string"> | type= | "type' | >ADJ </td <td>data><da< td=""><td>ta type='</td><td>'lemma</td><td>۱">écologique</td></da<></td> | data> <da< td=""><td>ta type='</td><td>'lemma</td><td>۱">écologique</td></da<> | ta type=' | 'lemma | ۱">écologique |
| <e1< td=""><td>ement:</td><td>⊷<data :<="" td=""><td>type=</td><td>"type'</td><td>>NOM<!--</td--><td>data><da< td=""><td>ta type='</td><td>'lemma</td><td>י">rivière</td></da<></td></td></data><data type="string">rivière</data></td></e1<> | ement: | ⊷ <data :<="" td=""><td>type=</td><td>"type'</td><td>>NOM<!--</td--><td>data><da< td=""><td>ta type='</td><td>'lemma</td><td>י">rivière</td></da<></td></td></data> <data type="string">rivière</data> | type= | "type' | >NOM </td <td>data><da< td=""><td>ta type='</td><td>'lemma</td><td>י">rivière</td></da<></td> | data> <da< td=""><td>ta type='</td><td>'lemma</td><td>י">rivière</td></da<> | ta type=' | 'lemma | י">rivière |
| <e1< td=""><td>ement:</td><td></td><td>type=</td><td>"type'</td><td>>ADJ<!--</td--><td>data><da< td=""><td>ta type='</td><td>'lemma</td><td>י>Songhua<data type="string">Songhua</data></td></da<></td></td></e1<> | ement: | | type= | "type' | >ADJ </td <td>data><da< td=""><td>ta type='</td><td>'lemma</td><td>י>Songhua<data type="string">Songhua</data></td></da<></td> | data> <da< td=""><td>ta type='</td><td>'lemma</td><td>י>Songhua<data type="string">Songhua</data></td></da<> | ta type=' | 'lemma | י>Songhua <data type="string">Songhua</data> |
| <e1< td=""><td>ement:</td><td>⊷data :</td><td>type=</td><td>"type'</td><td>>NOM<!--</td--><td>data><da< td=""><td>ta type='</td><td>'lemma</td><td>וי>nivēau<data type="string">nivēau</data></td></da<></td></td></e1<> | ement: | ⊷data : | type= | "type' | >NOM </td <td>data><da< td=""><td>ta type='</td><td>'lemma</td><td>וי>nivēau<data type="string">nivēau</data></td></da<></td> | data> <da< td=""><td>ta type='</td><td>'lemma</td><td>וי>nivēau<data type="string">nivēau</data></td></da<> | ta type=' | 'lemma | וי>nivēau <data type="string">nivēau</data> |
| <e1< td=""><td>ement:</td><td>⊷<data :<="" td=""><td>type=</td><td>"type'</td><td>>ADJ<!--</td--><td>data><da< td=""><td>ta type='</td><td>'lemma</td><td>י">local</td></da<></td></td></data><data type="string">local</data></td></e1<> | ement: | ⊷ <data :<="" td=""><td>type=</td><td>"type'</td><td>>ADJ<!--</td--><td>data><da< td=""><td>ta type='</td><td>'lemma</td><td>י">local</td></da<></td></td></data> <data type="string">local</data> | type= | "type' | >ADJ </td <td>data><da< td=""><td>ta type='</td><td>'lemma</td><td>י">local</td></da<></td> | data> <da< td=""><td>ta type='</td><td>'lemma</td><td>י">local</td></da<> | ta type=' | 'lemma | י">local |
| <e]< td=""><td>ement:</td><td>-<data :<="" td=""><td>type=</td><td>"type'</td><td>>NOM<!--</td--><td>data><da< td=""><td>ta type='</td><td>'lemma</td><td>יאie</td></da<></td></td></data><data type="string">xie</data></td></e]<> | ement: | - <data :<="" td=""><td>type=</td><td>"type'</td><td>>NOM<!--</td--><td>data><da< td=""><td>ta type='</td><td>'lemma</td><td>יאie</td></da<></td></td></data> <data type="string">xie</data> | type= | "type' | >NOM </td <td>data><da< td=""><td>ta type='</td><td>'lemma</td><td>יאie</td></da<></td> | data> <da< td=""><td>ta type='</td><td>'lemma</td><td>יאie</td></da<> | ta type=' | 'lemma | יאie |
| <e1< td=""><td>ement:</td><td>⊷<data :<="" td=""><td>type=</td><td>"type'</td><td>>ADJ<!--</td--><td>data><da< td=""><td>ta type='</td><td>'lemma</td><td>י>zhenhua</td></da<></td></td></data><data type="string">zhenhua</data></td></e1<> | ement: | ⊷ <data :<="" td=""><td>type=</td><td>"type'</td><td>>ADJ<!--</td--><td>data><da< td=""><td>ta type='</td><td>'lemma</td><td>י>zhenhua</td></da<></td></td></data> <data type="string">zhenhua</data> | type= | "type' | >ADJ </td <td>data><da< td=""><td>ta type='</td><td>'lemma</td><td>י>zhenhua</td></da<></td> | data> <da< td=""><td>ta type='</td><td>'lemma</td><td>י>zhenhua</td></da<> | ta type=' | 'lemma | י>zhenhua |
| <e1< td=""><td>ement:</td><td>⊷<data :<="" td=""><td>type=</td><td>"type'</td><td>>NOM<!--</td--><td>data><da< td=""><td>ta type='</td><td>'lemma</td><td>י>agence</td></da<></td></td></data><data type="string">Agence</data></td></e1<> | ement: | ⊷ <data :<="" td=""><td>type=</td><td>"type'</td><td>>NOM<!--</td--><td>data><da< td=""><td>ta type='</td><td>'lemma</td><td>י>agence</td></da<></td></td></data> <data type="string">Agence</data> | type= | "type' | >NOM </td <td>data><da< td=""><td>ta type='</td><td>'lemma</td><td>י>agence</td></da<></td> | data> <da< td=""><td>ta type='</td><td>'lemma</td><td>י>agence</td></da<> | ta type=' | 'lemma | י>agence |
| <e]< td=""><td>ement:</td><td>-<data :<="" td=""><td>type=</td><td>"type'</td><td>>ADJ<!--</td--><td>data><da< td=""><td>ta type='</td><td>'lemma</td><td>a">d</td></da<></td></td></data><data type="string">d</data></td></e]<> | ement: | - <data :<="" td=""><td>type=</td><td>"type'</td><td>>ADJ<!--</td--><td>data><da< td=""><td>ta type='</td><td>'lemma</td><td>a">d</td></da<></td></td></data> <data type="string">d</data> | type= | "type' | >ADJ </td <td>data><da< td=""><td>ta type='</td><td>'lemma</td><td>a">d</td></da<></td> | data> <da< td=""><td>ta type='</td><td>'lemma</td><td>a">d</td></da<> | ta type=' | 'lemma | a">d |
| <e1< td=""><td>ement:</td><td>⊷data :</td><td>type=</td><td>"type'</td><td>>NOM<!--</td--><td>data><da< td=""><td>ta type='</td><td>'lemma</td><td>וי>ministre<data type="string">ministre</data></td></da<></td></td></e1<> | ement: | ⊷data : | type= | "type' | >NOM </td <td>data><da< td=""><td>ta type='</td><td>'lemma</td><td>וי>ministre<data type="string">ministre</data></td></da<></td> | data> <da< td=""><td>ta type='</td><td>'lemma</td><td>וי>ministre<data type="string">ministre</data></td></da<> | ta type=' | 'lemma | וי>ministre <data type="string">ministre</data> |
| <el< td=""><td>ement:</td><td>⊷data :</td><td>type=</td><td>"type'</td><td>>ADJ<!--</td--><td>data><da< td=""><td>ta type='</td><td>'lemma</td><td>י>britannique<data type="string">britannique</data></td></da<></td></td></el<> | ement: | ⊷data : | type= | "type' | >ADJ </td <td>data><da< td=""><td>ta type='</td><td>'lemma</td><td>י>britannique<data type="string">britannique</data></td></da<></td> | data> <da< td=""><td>ta type='</td><td>'lemma</td><td>י>britannique<data type="string">britannique</data></td></da<> | ta type=' | 'lemma | י>britannique <data type="string">britannique</data> |
| <e1< td=""><td>ement:</td><td>⊷data :</td><td>type=</td><td>"type'</td><td>>NOM<!--</td--><td>data><da< td=""><td>ta type='</td><td>'lemma</td><td>י>immigration<data type="string">immigration</data></td></da<></td></td></e1<> | ement: | ⊷data : | type= | "type' | >NOM </td <td>data><da< td=""><td>ta type='</td><td>'lemma</td><td>י>immigration<data type="string">immigration</data></td></da<></td> | data> <da< td=""><td>ta type='</td><td>'lemma</td><td>י>immigration<data type="string">immigration</data></td></da<> | ta type=' | 'lemma | י>immigration <data type="string">immigration</data> |
| <e1< td=""><td>ement:</td><td>⊷≺data ∵</td><td>type=</td><td>"type'</td><td>>ADJ<!--</td--><td>data><da< td=""><td>ta type='</td><td>'lemma</td><td>i">n<data type="string">n</data></td></da<></td></td></e1<> | ement: | ⊷≺data ∵ | type= | "type' | >ADJ </td <td>data><da< td=""><td>ta type='</td><td>'lemma</td><td>i">n<data type="string">n</data></td></da<></td> | data> <da< td=""><td>ta type='</td><td>'lemma</td><td>i">n<data type="string">n</data></td></da<> | ta type=' | 'lemma | i">n <data type="string">n</data> |
| <e1< td=""><td>ement:</td><td>-<data :<="" td=""><td>tÿpe=</td><td>"type'</td><td>>NOM<!--</td--><td>data><da< td=""><td>ta type='</td><td>'lemma</td><td>וי>comité</td></da<></td></td></data><data type="string">Comité</data></td></e1<> | ement: | - <data :<="" td=""><td>tÿpe=</td><td>"type'</td><td>>NOM<!--</td--><td>data><da< td=""><td>ta type='</td><td>'lemma</td><td>וי>comité</td></da<></td></td></data> <data type="string">Comité</data> | tÿpe= | "type' | >NOM </td <td>data><da< td=""><td>ta type='</td><td>'lemma</td><td>וי>comité</td></da<></td> | data> <da< td=""><td>ta type='</td><td>'lemma</td><td>וי>comité</td></da<> | ta type=' | 'lemma | וי>comité |
| <e1< td=""><td>ement:</td><td>⊷data :</td><td>type=</td><td>"type'</td><td>>ADJ<!--</td--><td>data><da< td=""><td>ta type='</td><td>'lemma</td><td>י">interministériel<data type="string">interministériel</data></td></da<></td></td></e1<> | ement: | ⊷data : | type= | "type' | >ADJ </td <td>data><da< td=""><td>ta type='</td><td>'lemma</td><td>י">interministériel<data type="string">interministériel</data></td></da<></td> | data> <da< td=""><td>ta type='</td><td>'lemma</td><td>י">interministériel<data type="string">interministériel</data></td></da<> | ta type=' | 'lemma | י">interministériel <data type="string">interministériel</data> |
| <e1< td=""><td>ement:</td><td>⊷≺data ∵</td><td>type=</td><td>"type'</td><td>>NOM<!--</td--><td>data><da< td=""><td>ta type='</td><td>'lemma</td><td>י>ressource<data type="string">ressources</data></td></da<></td></td></e1<> | ement: | ⊷≺data ∵ | type= | "type' | >NOM </td <td>data><da< td=""><td>ta type='</td><td>'lemma</td><td>י>ressource<data type="string">ressources</data></td></da<></td> | data> <da< td=""><td>ta type='</td><td>'lemma</td><td>י>ressource<data type="string">ressources</data></td></da<> | ta type=' | 'lemma | י>ressource <data type="string">ressources</data> |
| <e1< td=""><td>ement:</td><td>⊷data :</td><td>tÿpe=</td><td>"type'</td><td>>ADJ<!--</td--><td>data><da< td=""><td>ta type='</td><td>'lemma</td><td>וי>minier<data type="string">minières</data></td></da<></td></td></e1<> | ement: | ⊷data : | tÿpe= | "type' | >ADJ </td <td>data><da< td=""><td>ta type='</td><td>'lemma</td><td>וי>minier<data type="string">minières</data></td></da<></td> | data> <da< td=""><td>ta type='</td><td>'lemma</td><td>וי>minier<data type="string">minières</data></td></da<> | ta type=' | 'lemma | וי>minier <data type="string">minières</data> |
| <e1< td=""><td>ement:</td><td>⊷≺data :</td><td>type=</td><td>"type'</td><td>>NOM<!--</td--><td>data><da< td=""><td>ta type='</td><td>'lemma</td><td>י>place<data type="string">place</data></td></da<></td></td></e1<> | ement: | ⊷≺data : | type= | "type' | >NOM </td <td>data><da< td=""><td>ta type='</td><td>'lemma</td><td>י>place<data type="string">place</data></td></da<></td> | data> <da< td=""><td>ta type='</td><td>'lemma</td><td>י>place<data type="string">place</data></td></da<> | ta type=' | 'lemma | י>place <data type="string">place</data> |
| <e1< td=""><td>ement:</td><td>⊷≺data ∵</td><td>type=</td><td>"type'</td><td>>ADJ<!--</td--><td>data><da< td=""><td>ta type='</td><td>'lemma</td><td>יאfinancier<data type="string">financière</data></td></da<></td></td></e1<> | ement: | ⊷≺data ∵ | type= | "type' | >ADJ </td <td>data><da< td=""><td>ta type='</td><td>'lemma</td><td>יאfinancier<data type="string">financière</data></td></da<></td> | data> <da< td=""><td>ta type='</td><td>'lemma</td><td>יאfinancier<data type="string">financière</data></td></da<> | ta type=' | 'lemma | יאfinancier <data type="string">financière</data> |
| <e1< td=""><td>ement:</td><td>⊷data :</td><td>type=</td><td>"type'</td><td>>NOM<!--</td--><td>data><da< td=""><td>ta type='</td><td>'lemma</td><td>וי>place<data type="string">places</data></td></da<></td></td></e1<> | ement: | ⊷data : | type= | "type' | >NOM </td <td>data><da< td=""><td>ta type='</td><td>'lemma</td><td>וי>place<data type="string">places</data></td></da<></td> | data> <da< td=""><td>ta type='</td><td>'lemma</td><td>וי>place<data type="string">places</data></td></da<> | ta type=' | 'lemma | וי>place <data type="string">places</data> |
| <e]< td=""><td>ement:</td><td>-<data :<="" td=""><td>tÿpe=</td><td>"type'</td><td>>ADJ<!--</td--><td>data><da< td=""><td>ta type='</td><td>'lemma</td><td>י>boursier</td></da<></td></td></data><data type="string">boursières</data></td></e]<> | ement: | - <data :<="" td=""><td>tÿpe=</td><td>"type'</td><td>>ADJ<!--</td--><td>data><da< td=""><td>ta type='</td><td>'lemma</td><td>י>boursier</td></da<></td></td></data> <data type="string">boursières</data> | tÿpe= | "type' | >ADJ </td <td>data><da< td=""><td>ta type='</td><td>'lemma</td><td>י>boursier</td></da<></td> | data> <da< td=""><td>ta type='</td><td>'lemma</td><td>י>boursier</td></da<> | ta type=' | 'lemma | י>boursier |

Affichage partiel du résultat avec XMLSpy

| Evaluate XPath | | | | | | |
|---|---|---|---|--|--|--|
| XPath: itains(text(),'ADJ')]]/prece | :ding-sibling::el | ement[1][./data[co | ontains(text(),'NOM')]]/data[3] i Real-time evaluation | | | |
| Allow Complete XPath XML Schema Selector XML Schema Field | From Doc From Sele | From Document Root From Selected Element C Evaluate when ty C Evaluate on butto | | | | |
| name | | value / attributes | ▲ | | | |
| () data | | désastre | | | | |
| () data | | écologique | | | | |
| () data | | rivière | | | | |
| () data | | Songhua | | | | |
| () data | | niveau | | | | |
| () data | | local | | | | |
| () data | | Xie | | | | |
| 🚺 data | | Zhenhua | | | | |
| 🚺 data | | Agence | | | | |
| () data | | d | | | | |
| () data | | ministre | | | | |
| () data | | britannique | | | | |
| () data | | immigration | | | | |





Affichage partiel du résultat avec Cooktop Avec uniquement les contenus des éléments « mot »

| nodes: (//element[| ./data[contains(text(),`NOM`)]]/tollowing_sibling::element[1][./data[contains(text(),`ADJ`)]]/data[3]) |
|--|---|
| nodes xpath (7,</th <th>/element[./data[contains(text(),'NOM')]]/following-sibling::element[1][./data[contains(text(),'ADJ')]]/da</th> | /element[./data[contains(text(),'NOM')]]/following-sibling::element[1][./data[contains(text(),'ADJ')]]/da |
| <data type="string</th><th>">désastre</data> | |
| <pre>kdata type="string</pre> | ">écologique |
| <pre><data type="string</pre></th><th>">rivière</data></pre> | |
| <pre>kdata type="string</pre> | ">Songhua |
| <data type="string</th><th>">nivēau</data> | |
| <data type="string</th><th>">local</data> | |
| <data type="string</th><th>">Xie</data> | |
| <data string<="" th="" type="string</th><th>>Zhenhua</data></th></tr><tr><th><data type="><th>>Agence</th></data> | >Agence |
| <data string<="" th="" type="string</th><th><pre>>d</data></pre></th></tr><tr><th><data type="><th>;>ministre</th></data> | ;>ministre |
| <data type="string</th><th>">britannique</data> | |
| <data string<="" th="" type="string</th><th><pre>>immigration</data></pre></th></tr><tr><th><data type="><th>">n</th></data> | ">n |
| <data type="string</th><th>">Comite</data> | |
| <data type="string</th><th>">interministériel</data> | |
| <data string<="" th="" type="string</th><th>>ressources</data></th></tr><tr><th><data type="><th>">minteres</th></data> | ">minteres |
| <data string<="" th="" type="string</th><th><pre>>place</data></pre></th></tr><tr><th><data type="><th><pre>">tinanclere</pre></th></data> | <pre>">tinanclere</pre> |
| <data type="string</th><th>">places</data> | |
| <data type="string</th><th>">boursteres</data> | |
| <data type="string</th><th>">chasse</data> | |
| <data string<="" th="" type="string</th><th>>s</data></th></tr><tr><th><data type="><th><pre>>tootballeur</pre></th></data> | <pre>>tootballeur</pre> |
| <data string<="" th="" type="string</th><th><pre>>britannique</data></pre></th></tr><tr><th><data type="><th>">ressources</th></data> | ">ressources |
| <data string<="" th="" type="string</th><th><pre>>minieres</data></pre></th></tr><tr><th><data type="><th><pre>">champion</pre></th></data> | <pre>">champion</pre> |
| <data string<="" th="" type="string</th><th>>d</data></th></tr><tr><th><data type="><th><pre>>distance</pre></th></data> | <pre>>distance</pre> |
| <data type="string</th> <th><pre>pectable</pre></th> | <pre>pectable</pre> |
| | |



Une feuille de style XSL pour afficher le résultat

- 1. Après avoir testé la requête XPATH... – Via XMLCooktop, XML Copy Editor...
- 2. On pourra construire une feuille de styles XSLT ou utiliser la <u>feuille de style suivante</u> pour donner à voir le résultat dans un navigateur :
 - <u>les patrons</u> NOM ADJ par exemple

3. On construira une feuille de styles pour construire les listes de patrons recherchées au format TXT



```
<?xml version="1.0" encoding="UTF-8"?>
<xsl:stylesheet xmlns:xsl="http://www.w3.org/1999/XSL/Transform"</pre>
  version="2.0">
                                                         Règle n°1 : préparation du
<rsl:output method="html"/>
                                                          tableau HTML final
                                                         Et lancement de la règle
<rpre><xsl:template match="/"> 
                                                        concernant les éléments visés
<html>
<body bqcolor="#81808E">
<table align="center" width="90%" bqcolor="white" bordercolor="#3300FF"
  border="1">
<font color="white"><h1>Extraction
  de patron <font color="red"><b>NOM</b></font> <font
  color="blue"><b>ADJ</b></font></h1></font>
<tsl:apply-templates select="./PARCOURS/ETIQUETAGE/file/element"/>
   </body>
                                                           Règle n°2 : on filtre
</html>
                                                           les éléments visés et
</xsl:template>
                                                            on applique la
                                                           requête déjà vue...
<rsl:template match="element">
<xsl:choose>
<xsl:when test="(./data[contains(text(),'NOM')]) and (following-</pre>
   sibling::element[1][./data[contains(text(), 'ADJ')]])">
<font color="red"><xsl:value-of select="./data[3]"/></font><xsl:text>
   </xsl:text>
</xsl:when>
<xsl:when test="(./data[contains(text(),'ADJ')]) and (preceding-</pre>
   sibling::element[1][./data[contains(text(), 'NOM')]])">
<font color="blue"><xsl:value-of select="./data[3]"/></font><br/>
</xsl:when>
</xsl:choose>
</xsl:template>
</xsl:stylesheet>
                                                La feuille de styles
```





Extraction de patron <mark>NOM ADJ</mark>

désastre écologique rivière Songhua niveau local Xie Zhenhua Agence d ministre britannique immigration n Comité interministériel ressources minières place financière places boursières chasse s footballeur britannique ressources minières champion d distance respectable président vénézuélien élections législatives ministre britannique finances européens angoisses identitaires crise sociale modèle politique ressources minières chasse s temps glacial footballeur britannique Agence mondiale symposium international dopage génétique Tracy Chapman observatrice lucide champion d distance respectable

Affichage du fichier issu de la *BàO série 2* (dans un navigateur) sur lequel on applique la feuille de styles repérant les patrons visés



Mise en œuvre de XPath

avec XMLCooktop

1.Onglet XML

 Tout se passe dans l'interface proposée par l'outil via l'onglet XPath
 2. Onglet XPath

| ource (xml) xpath console stylesheet(xsl) result result(html) |
|--|
| <pre><?xml version="1.0" encoding="iso-8859-1" ?></pre> |
| <pre><?xml-stylesheet type="text/xsl" href="SORTIE-extraction-patron-NOM-ADJ.xsl"?></pre> |
| <parcours></parcours> |
| <nom>Votre nom</nom> |
| <etiquetage><file></file></etiquetage> |
| <pre><name>arbo-fils/2005/Nov/4/01-00-00/0.2-3208.1-0.0.xml</name></pre> |
| <pre><element><data type="type">VER:pres</data><data type="lemma">retrouver</data><data< pre=""></data<></element></pre> |
| <pre><element><data type="type">NUM</data><data type="lemma">L</data><data <="" pre="" type="strind"></data></element></pre> |
| <pre><element><data type="type">PUN</data><data type="lemma"> </data><data <="" pre="" type="strind"></data></element></pre> |
| <pre><element><data type="type">ADV</data><data type="lemma">ensemble</data><data <="" data="" type="lemma"></data>ensemble<data <="" data="" type="lemma"></data>ensemble<data <="" data="" type="l</th></tr><tr><th><pre><element><data type=" type"="">PRP:det</data><data type="lemma">du</data><data type="s</pre></th></tr><tr><th><pre><element><data type=" type"="">NOM</data><data type="lemma">dépêche</data><data type="</pre></th></tr><tr><th><pre><element><data type=" type"="">PRP</data><data type="lemma">sur</data><data type="stri</pre></th></tr><tr><th><pre><element><data type=" type"="">NOM</data><data type="lemma">http</data><data type="stri-</pre></th></tr><tr><th><pre><element><data type=" type"="">PUN</data><data type="lemma">:</data><data <="" pre="" type="string"></data></element></pre> |
| <pre><element><data type="type">PUN</data><data type="lemma">/</data><data <="" pre="" type="strind"></data></element></pre> |
| <pre><element><data type="type">PUN</data><data type="lemma">/</data><data <="" pre="" type="strind"></data></element></pre> |
| <pre><element><data type="type">NOM</data><data type="lemma">www</data><data <="" pre="" type="str"></data></element></pre> |
| <pre><element><data type="type">SENT</data><data type="lemma">.</data><data <="" pre="" type="strir"></data></element></pre> |
| <pre><element><data type="type">NOM</data><data type="lemma">lemonde</data><data type="</pre></th></tr><tr><th><pre><element><data type=" type"="">SENT</data><data type="lemma">.</data><data <="" pre="" type="string"></data></element></pre> |
| <pre><element><data type="type">NOM</data><data type="lemma">fr</data><data <="" pre="" type="string"></data></element></pre> |
| |
| <file></file> |
| <pre><name>arbo-fils/2005/Nov/4/01-00-00/0.2-3210.1-0.0.xml</name></pre> |
| <pre><element><data_type="type">DET:ABT<data_type="lemma">le<data_type="selements.com se<br="" selements.com="">selements.com/selements.com/selements.com/selements.com/selements.com/selements.com/selements.com/selements.com/selements.com/selements.com/selements.com/selements.com/selements.com/selements.com/selements.com/selements.com/selements.c</data_type="selements.com></data_type="lemma"></data_type="type"></element></pre> |
| <pre><element><data type="type">NOM</data><data type="lemma">désastre</data><data <="" pre="" type="lemma"></data></element></pre> |
| <pre><element><data type="type">AD1</data><data type="lemma">écologique</data><data type="lemma">écologique</data></element></pre> |
| <pre><element><data type="type">PRP</data><data type="lemma">de</data><data type="strik</pre></th></tr><tr><th><pre>selement><data type=" type"="">DET:ART</data><data type="lemma">le</data><data <="" pre="" type=""></data></element></pre> |
| <pre><element><data type="type">NOM</data><data type="lemma">rivière</data><data type="</pre></th></tr><tr><th><pre>celement><data type=" type"="">ADI</data><data type="lemma">Songhua</data><data type="</pre></th></tr><tr><th><pre>celement>cdata type=" type"="">EUN</data>cdata type="lemma"> cdata type="string"</element></pre> |
| <pre><clement><data type="type">PRP</data><data type="lemma">dans</data><data <="" pre="" type="structure"></data></clement></pre> |
| <pre>celement><data type="type">ET:ABT</data><data type="lemma">le</data><data type="classical data</pre></th></tr><tr><th><pre>celement><data type=" type"="">NOM</data><data type="lemma">Heiloppijapg</data><data pre="" t<=""></data></pre> |
| celements/data_type="type">PIN//data/cdata_type="lemma">. |
| celements/data_type="type">KONet <data_type="lemma">et<data_type="lemma">et<data_type="lemma">et<data_type="lemma">et<data_type="lemma">et<data_type="lemma">et<data_type="lemma">et<data_type="lemma">et<data_type="lemma">et<data_type="lemma">et<data_type="lemma">et<data_type="lemma">et<data_type="lemma">et<data_type="lemma">et<data_type="lemma">et<data_type="lemma">et<data_type="lemma">et<data_type="lemma">et<data_type="lemma">et<data_type="lemma">et<data_type="lemma">et<data_type="lemma">et<data_type="lemma">et<data_type="lemma">et<data_type="lemma">et<data_type="lemma">et<data_type="lemma">et<data_type="lemma">et<data_type="lemma">et<data_type="lemma">et<data_type="lemma">et<data_type="lemma">et<data_type="lemma">et<data_type="lemma">et<data_type="lemma">et<data_type="lemma">et<data_type="lemma">et<data_type="lemma">et<data_type="lemma">et<data_type="lemma">et<data_type="lemma">et<data_type="lemma">et<data_type="lemma">et<data_type="lemma">et<data_type="lemma">et<data_type="lemma">et<data_type="lemma">et</data_type="lemma"></data_type="lemma"></data_type="lemma"></data_type="lemma"></data_type="lemma"></data_type="lemma"></data_type="lemma"></data_type="lemma"></data_type="lemma"></data_type="lemma"></data_type="lemma"></data_type="lemma"></data_type="lemma"></data_type="lemma"></data_type="lemma"></data_type="lemma"></data_type="lemma"></data_type="lemma"></data_type="lemma"></data_type="lemma"></data_type="lemma"></data_type="lemma"></data_type="lemma"></data_type="lemma"></data_type="lemma"></data_type="lemma"></data_type="lemma"></data_type="lemma"></data_type="lemma"></data_type="lemma"></data_type="lemma"></data_type="lemma"></data_type="lemma"></data_type="lemma"></data_type="lemma"></data_type="lemma"></data_type="lemma"></data_type="lemma"></data_type="lemma"></data_type="lemma"></data_type="lemma"></data_type="lemma"></data_type="lemma"></data_type="lemma"></data_type="lemma"></data_type="lemma"></data_type="lemma"> |
| <pre>celements/data_type="type"splf:/ABT</pre> |
| celements/data type="type">NOM//datas/data type="lemma">tentative//datas/data type=" |
| celements/data_type="type">type='type"/type='type='lemma'/type='lemma'/type='type |
| <pre>celements<data type="type">BPRP:detdata</data></pre> |
| <pre>celement><data type="type">NOM</data><data type="lemma">niveau</data><data <="" pre="" type=""></data></pre> |
| celements/data_type="type="type="type="lemma">local//data>/data_type="ty |
| celements/data_type="type">PRPpourvdata_type="str |
| <pre>celements/data_type="type">vpe://data/vdata_type="temma"/pour/data/vdata_type="type"; celements/data_type="type">vpe:/vpe://data/vdata_type="temma"/pour/data/vdata_type="type="type"; celements/data_type="type";</pre> |
| zelementszdata type-type/vek.ni type/data zydata type-telmia/etourtet/ydata/data |
| valaments/data type="type=ynoms/like/datas/data type="lemma >\//datas/data type="stilling |
| zalamentszdata type-type/srowz/dataszdata type-"lemma/saffairez/dataszdata type-" |
| zelementszdata type="type"splw/data/sdata type="lemma"ariaries/data/sdata type="string |
| valaments/data type= type /rows/data/data type= remma //data/data type= stillio |

| source (xml) xpath console stylesheet(xsl) result result(html) |
|--|
| XPATH Debuger Help:</th |
| The XPATH expression are written on a single line |
| XPATH debugs XPATH expresions against the current XML document. |
| nodes: (or n:) xpath expresion that results in nodes |
| value: (or s:) xpath expression that results in a single value |
| - values. (or v.) xpath expression that results in multiple values |
| Put the cursor on the line that you want |
| to execute and press Ctrl-E or Enter |
| To view all namespaces of the xml document enter namespaces: (or ns:) and press |
| where the state of |
| - to enable XPath to find their elements |
| All namespaces with names like ns1, ns2 etc. were named explicitly by Cooktop - |
| nodes://element |
| nodes xpath //element' |
| zelement> <data type="">VER:pres</data> <data type="lemma">retrouver</data> <data< th=""></data<> |
| <pre>colleget >Kdat / ppe= type >NUM (data type= Temma >L (data> type=""temma >L (data> type=""tempa > type="tempa > type="</pre> |
| celements/data type="type"solv/datas/data type="lemma"soncemble/datas/data type="type"solv/datas/data |
| <pre>celement>cdata type="type">DPD:/data/data_type="lemma">dua/data/data_type="</pre> |
| <pre><element><data type="type">NOM</data><data type="lemma">dénêche</data><data type="</pre></th></tr><tr><th><pre><element><data type=" type"="">PRP</data><data type="lemma">sur</data><data type="stri</pre></th></tr><tr><th><pre><element><data type=" type"="">NOM</data><data type="lemma">http</data><data type="str</pre></th></tr><tr><th><pre><element><data type=" type"="">PUN</data><data type="lemma">:</data><data type="string</pre></th></tr><tr><th><pre><element><data type=" type"="">PUN</data><data type="lemma">/</data><data <="" pre="" type="string"></data></element></pre> |
| <pre><element><data type="type">PUN</data><data type="lemma">/</data><data <="" pre="" type="string"></data></element></pre> |
| <pre><element><data type="type">NOM</data><data type="lemma">www</data><data lemma"="" type="str]</pre></th></tr><tr><th>celements/data type= type >sent/datas/data type= temma >.</data>data type= strin</th></tr><tr><th><pre>celement>cdata type= type >NOM</pre>/data>cdata type= remina >remonder/data>cdata type= remont>cdata type=type>SENT</data>cdata type=">remonder/data>cdata type="type"</data></element></pre> |
| <pre>celement><data type="type">Sentfr</data></pre> |
| <pre><element><data type="type">DET:ABT</data><data type="lemma">le</data><data type="s</pre></th></tr><tr><th><pre><element><data type=" type"="">NOM</data><data type="lemma">désastre</data><data type="</pre"></data></element></pre> |
| <pre><element><data type="type">ADJ</data><data type="lemma">écologique</data><data pre="" typ<=""></data></element></pre> |
| <pre><element><data type="type">PRP</data><data type="lemma">de</data><data <="" pre="" type="strif"></data></element></pre> |
| <pre><element><data type="type">DET:ART</data><data type="lemma">le</data><data type="s</pre></th></tr><tr><th><pre><element><data type=" type"="">NOM</data><data type="lemma">rivière</data><data type="</pre></th></tr><tr><th><pre><element><data type= type >ADJ</data><data type= lemma >Songhua</data><data type=</pre></th></tr><tr><th>celements/data type= type >PUN</data> (data type= lemma >, </data> (data> (data type= string</th></tr><tr><th><pre>celement><uala type= type >PRP</uala><uala type= temma >uals</uala><uala>type= type selement><uala>type=type</pre></th></tr><tr><th>celement x data type=" type"="">Der x type="type"</data></element></pre> |
| <pre>celement><data type="type">PUN</data><data type="lemma"></data><data <="" pre="" type="string"></data></pre> |
| <pre><element><data type="type">KON</data><data type="lemma">et</data><data <="" pre="" type="string"></data></element></pre> |
| <pre><element><data type="type">DET:ART</data><data type="lemma">le</data><data type="s</pre></th></tr><tr><th><pre><element><data type=" type"="">NOM</data><data type="lemma">tentative</data><data pre="" type<=""></data></element></pre> |
| <pre><element><data type="type">VER:pres</data><data type="lemma">faire</data><data pre="" type<=""></data></element></pre> |
| <pre><element><data type="type">PRP:det</data><data type="lemma">au</data><data selement="" type="selement><data type="><data selement"="" type="selement><data type="></data><data <="" data="" selements"<="" type="sele</th></tr><tr><th><pre><element><data type=" type"="">NOM</data><data type="lemma">nlveau</data><data <="" pre="" type=""></data></data></data></data></data></data></data></data></data></data></data></element></pre> |
| <pre>celement><uala lype="type">AUJ</uala><ual type="lemma">local<ual type="st<br">celement><ual type="type">type>type>type>type=st</ual></ual></ual></pre> |
| <pre>keremenc>kuata type= type >pkpk/uata>kuata type= Temma >puurk/uata>kuata type= str</pre> |
| |

Projet encadré



pluriTAL

http://tal.univ-paris3.fr/plurital/

Avec XML Copy Editor

Menu XML

Evaluate XPath

<?xml-stylesheet type="text/xsl"-href="SORTIE-extraction-patron-NOM-ADJ.xsl"?> <PARCOURS> <NOM>Votre-nom</NOM> <ETIQUETAGE><file> <name>arbo-fils/2005/Nov/4/01-00-00/0,2-3208,1-0,0.xml</name> <element><data type="type">VER:pres</data><data type="lemma"></data><data type="string">Retrouvez</data></element> <element><data type="type">VER:pres</data></element> <element><data type="type">NUM</data><data type="lemma"></data><data type="string"></data></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></element></elem

<element><data type="type">NUM</data><data type="lemma">L</data><data type="string">I</data></element> <element><data type="type">PUN</data><data type="lemma">'</data><data type="string">'</data></element> <element><data type="type">ADV</data><data type="lemma">ensemble</data><data type="string">ensemble</data></element> <element><data type="type">PRP:det</data><data type="lemma">du</data><data type="string">des</data></element> <element><data type="type">NOM</data></data type="lemma">dépêche</data></data type="string">dépêches</data></element> <element><data type="type">PRP</data><data type="lemma">sur</data><data type="string">sur</data></element> <element><data type="type">NOM</data><data type="lemma">http</data><data type="string">http</data></element> <element><data type="type">PUN</data><data type="lemma">:</data><data type="string">:</data></element> <element><data type="type">PUN</data><data type="lemma">/</data><data type="string">/</data></element> <element><data-type="type">PUN</data><data-type="lemma">/</data><data-type="string">/</data></element> <element><data type="type">NOM</data><data type="lemma">www</data><data type="string">www</data></element> <element><data type="type">SENT</data><data type="lemma">.</data><data type="string">.</data></element> <element><data type="type">NOM</data><data type="lemma">lemonde</data><data type="string">lemonde</data></element> <element><data type="type">SENT</data><data type="lemma">.</data><data type="string">.</data></element> <element><data type="type">NOM</data><data type="lemma">fr</data><data type="string">fr</data></element> </file>

<file>

<name>arbo-fils/2005/Nov/4/01-00-00/0,2-3210,1-0,0.xml</name>

?xml·version="1.0"·encoding="iso-8859-1"·?>

<element><data type="type">DET:ART</data><data type="lemma">le</data
<element><data type="type">NOM</data><data type="lemma">désastre</i
<element><data type="type">ADJ</data><data type="lemma">désastre</i
<element><data type="type">ADJ</data><data type="lemma">desidata><data type="lemma">data</data><data type="lemma">desidata><data type="lemma">data</data><data type="lemma">data</data><data type="lemma">data</data><data type="lemma">data><data type="lemma">data><data type="lemma">data</data><data type="lemma">data><data type="lemma">data><data><data type="lemma">data><data type="lemma">data><data type="lemma">data><data type="lemma">data><data type="lemma">data><data type="lemma">data></data><data type="lemma">data<

| valuate XPath | | × |
|---------------|----|--------|
| Enter XPath: | | |
| //element | | |
| | | |
| | OK | Cancel |

<element><data type="type">DET:ART</data><data type="lemma">le</data><data type="string">le</data></element> <element><data type="type">NOM</data><data type="lemma">Heilongjiang</data><data type="string">Heilongjiang</data></element> <element><data type="type">PUN</data><data type="lemma">,</data><data type="string">,</data></element> <element><data type="type">KON</data><data type="lemma">et</data><data type="string">et</data></element> <element><data type="type">DET:ART</data><data type="lemma">le</data><data type="string">les</data></element> <element><data type="type">NOM</data><data type="lemma">tentative</data><data type="string">tentatives</data></element> <element><data type="type">VER:pres</data><data type="lemma">faire</data><data type="string">faites</data></element> <element><data type="type">PRP:det</data><data type="lemma">au</data><data type="string">au</data></element> <element><data type="type">NOM</data><data type="lemma">niveau</data><data type="string">niveau</data></element> <element><data type="type">ADJ</data><data type="lemma">local</data><data type="string">local</data></element> <element><data type="type">PRP</data><data type="lemma">pour</data><data type="string">pour</data></element> <element><data type="type">VER:infi</data><data type="lemma">étouffer</data><data type="string">étouffer</data></element> <element><data type="type">NUM</data><data type="lemma">L</data><data type="string">I</data></element> <element></data-type="type">PUN</data></data-type="lemma">'</data></data-type="string">'</data></element> <element><data-type="type">NOM</data><data-type="lemma">affaire</data><data-type="string">affaire</data></element> <element><data type="type">PUN</data><data type="lemma">,</data><data type="string">,</data></element> <element><data type="type">VER:pres</data><data type="lemma">avoir</data><data type="string">ont</data></element> <element><data type="type">VER:pper</data><data type="lemma">conduire</data><data type="string">conduit</data></element> <element><data type="type">PRP:det</data><data type="lemma">au</data><data type="string">au</data></element> <element><data type="type">NOM</data><data type="lemma">limogeage</data><data type="string">limogeage</data></element> <element><data type="type">PUN</data><data type="lemma">,</data><data type="string">,</data></element> <element><data type="type">NOM</data><data type="lemma">vendredi</data><data type="string">vendredi</data></element> <element><data type="type">NUM</data><data type="lemma">@card@</data><data type="string">2</data></element> <element><data type="type">NOM</data><data type="lemma">décembre</data><data type="string">décembre</data></element> <element><data type="type">PUN</data><data type="lemma">,</data><data type="string">,</data></element> <element><data type="type">PRP</data><data type="lemma">de</data><data type="string">de</data></element> <element><data type="type">NOM</data><data type="lemma">Xie</data><data type="string">Xie</data></element> <element><data-type="type">ADJ</data><data-type="lemma">Zhenhua</data><data-type="string">Zhenhua</data></element> <element><data type="type">PUN</data><data type="lemma">,</data><data type="string">,</data></element> <element><data type="type">DET:ART</data><data type="lemma">le</data><data type="string">le</data></element>





Transformer un document XML (1) *via* une feuille de styles XSLT

• Saxon

- Hypertoile : <u>http://saxon.source</u> <u>forge.net/</u>
- (à récupérer ici :
 <u>programme</u>
 <u>SAXON pour</u>
 <u>windows</u>)

| SAXON 6.4.3 From | Hichael Kay | |
|---|--|--------------|
| Usage: java con. | icl.saxon.StyleSheet [options] source-doc style-doc | {paran=valu |
| e> | | |
| Options: | | |
| | Ose xml-stylesheet r1, not style-doc argument | |
| -01 | Use standard tree data structure | |
| -de Cilleonne | Geod outwith to paped 5510 on dispostory | |
| - Classpann | Hen specified Emittee class for veltessesse output | |
| -re classnane | the specified URIReaniver class | |
| -8 | Display version and tining information | |
| $-\tilde{\mathbf{I}}$ | Set standard TraceListener | |
| -TL classname | Set a specific TraceListener | |
| -u | Manes are URLs not filenanes | |
| | Recover silently from recoverable errors | |
| -w1 | Report recoverable errors and continue (default) | |
| | Treat recoverable errors as fatal | |
| -x classnane | Use specified SRX parser for source file | |
| -y classnane | use specified SAX parser for stylesheet | |
| -1 | Ploplay this nessage | |
| C:\Documents and axon.hat test | Settings\Serge Fleury\Bureau\hao-29822886\Graphm12Fc ./Hawai.xml/GraphML2Pajek.x=1 | ajek\Saxon)s |
| Ave - Classpark - - Classpark - - Classpark - - Vertice: 13 - | The survey and the survey | |
| 12 13 114.8 | | |
| | a set in a set in a set of the | icksSaxon21 |
| Cr.Bocuments and ava -classpath 1: jek.xsl > test.ts | Setting:\Serge Floury\Bureau\bao-2902800\Graphm127v ib/taxon.jar com.icl.taxon.StyleSheet/Hawai.xml tt | ./GraphHL2Pa |
| Cr\Documents and ava -classpath 1: jek.ssl > test.t: Cr\Documents and | Setting:\Serge Floury\Bureau\bao-2982886\Graphm127 ib/taxon.jar com.icl.taxon.StyleSheet/Havai.xml t | -GraphHL2Pa |





Transformer un document XML (2) *via* une feuille de styles XSLT

AltovaXSLT

Processeur
 développé par
 <u>Altova</u>

 DISPONIBLE dans votre archive du cours (à récupérer aussi ici <u>AltoXSLT.zip</u> ou <u>là</u>)

| Cygwin 820 | and a second second | | |
|---|--|---|---|
| hach-2.825 Alt Iltova XSLI Pr Sopyright (c)i Ill rights res | ovaXSLT.exe ocessor 998-2003 Altova erved. | Gaduli | - |
| ise of this so http://www.alt | ftware in subje ova.com/xsltdla | ct to the license agreement at | |
| -sml -sml -out -param -smlstack Coml Hawai.sml un-graph.tst bash-2.825 | [file name] [file name] [file name] [name=expr] [number >= 180 -xs1 GraybME2P | input solt input result stylesheet parameter value I maximum sol template call depth ajek.sol -out un-graph.tst | |





Transformer un document XML (3) *via* une feuille de styles XSLT

AltovaXML

- Hypertoile : <u>http://www.altova.com/altovaxml.html</u>
- Disponible dans votre ARCHIVE du cours
- Mode d'emploi :
 - Dans la ligne de commandes, tapez :

AltovaXML -xslt1 xsltfile -in xmlfile [-out outputfile] [options] AltovaXML -xslt2 xsltfile -in xmlfile [-out outputfile] [options]





Transformer un document XML (4) *via* une feuille de styles XSLT
Avec XMLCooktop

- *Hypertoile* : <u>http://www.xmlcooktop.com/</u>
- Mode d'emploi :

1. Onglet XML

• Tout se passe dans l'interface proposée par l'outil :



Travail personnel série 3

Projet encadré



pluriTAL

http://tal.univ-paris3.fr/plurital/

🐣 Cooktop - [Hawai.ami - GraphML2Pajek.asl.sle - Hawai.ami | GraphML2Pajek.asl] . OX Bie Edit 1948. Code Bits Tools Options Web Window Heb _ 6 × source (imi) spath console stylesheet(xii) result (result(html) chaml version="1.0" encoding="iso-8859-1">
ckslistylesheet version="1.0" xmlnsixsl="http://www.w3.org/1999/kdL/Transform">
ckslioutput method="text" encoding="iso-8859-1"/> d-- celitero-chelitero ---> collitenplate match="/"> coslitext>*Vertices </xilitext>coslivalue-of select="count(graphm]/graph/node)"/>coslitext> </xsl:text> cxsliapply-templates select="graphml/graph/node"/> <rsl:text>*Edges </kslitexts cxtl:apply-templates select="graphml/graph/edge"/>
</xsl:templates</pre> cxsl:template match="edge"> cxsl:value-of select="./@source"/><xsl:text></xsl:text><xsl:value-of select="./@target"/> </xslitext> </ksl:template> cxsl:template match="node"> coslivalue-of select-"./Bid"/>coslitext> "c/oslitext>coslivalue-of select-"./data"/>coslite <//xsl:text> </xsl:template> </xil:stylesheet> COLT: MOORE Ln I, Col D Ready UNDO

2. Onglet XSL

3. Onglet RESULT

| Cooktop - [Hawaiseni - GraphMt2Pajek.uslade - Hawaiseni | Graphen 20 ajek asl | 1 | | |
|--|---------------------|-------------|--------------|--------|
| Bie Edit 1946. Gode Bits Tools Options Web Window Help | The second second | | | _16 ×1 |
| | 3000 | | | |
| <pre>pource(unt) spath console (sylenteetCot) rem& result() *vertices 13 1 "Hail11" 2 "Wahilawa" 3 "Pearl City" 4 "Laite" 5 "Kaneohe" 6 "Monolulu" 7 "Kahului" 8 "Lahaina" 9 "Keokes" 10 "Mana" 11 "Kamuela" 12 "Kona" 13 "Hilo" *Edges 1 2 15.0 1 3 20.0 2 3 12.0 2 3 12.0 3 6 13.0 4 5 24.0 3 6 13.0 4 5 24.0 3 6 13.0 11 12 31.0 11 12 31.0 11 12 31.0 11 12 31.0</pre> | | | | - |
| 当 | | 1 | | 20 |
| Ready | PC | DOLT: MOONL | Ln 19, Col 9 | READ |





Transformer un document XML (5) via une feuille de styles XSLT

le Edit View Insert XML

Tools

- Avec XML Copy • Editor
 - Hypertoile : <u>http://xml-</u> <u>сору-</u> editor.sourceforge.net/
 - Mode d'emploi :
 - Menu XML
 - XSL
 - XSL transform
 - (ou F8)

| saintex | .xml 🗙 SORTIE-etiquetage-patron.xml 🗙 | - |
|------------|--|---|
| 1 | xml·version="1.0"·encoding="iso-8859-1"·? | 4 |
| 2 | xml-stylesheet type="text/xsl"-href="SORTIE-extraction-patron-NOM-ADJ.xsl"? | |
| ا ک | | |
| 4 5 G | | |
| 6 1 | <a href="https://doi.org/10.100/100/100/100/100/100/100/100/100/</td><td></td></tr><tr><td>7</td><td><element><data type=" type"="">VER:pres<data type="lemma">retrouver</data><data type="string">Retrouvez</data><td></td> | |
| 8 | <element><data type="type">NUM</data><data type="lemma">L</data><data type="string">l</data></element> | |
| 9 | <element><data type="type">PUN</data><data type="lemma">'</data><data type="string">'</data></element> | |
| 10 | <element><data type="type">ADV</data><data type="lemma">ensemble</data><data type="string">ensemble</data></element> | |
| 11 | <element><data type="type">PRP: det</data><data type="lemma">du</data><data type="string">des</data></element> | |
| 12 | <element><datatype="type">NOM<data type="lemma">depeche</data><data type="string">depeches</data></datatype="type"></element> <data type="string">depeches</data> <td></td> | |
| 13 | <element><data type="type">rrk+</data><data type="lemma">sur</data><data type="string">sur</data><delement><data type="string">sur</data><delement><data type="string">sur</data><data type="string">sur</data><data type="string">sur</data><data type="string">sur</data><data type="string">sur</data><data type="string">sur</data><data type="string">sur</data><data type="string">sur</data><data type="string">sur</data><td></td></delement></delement></element> | |
| 14 | <pre><element.scuatarype= snows.cuatarype="enima" snu<="" snups.cuatarype="type" td="" type=""><td></td></element.scuatarype=></pre> | |
| 16 | Solonia in Statia type="type = constraints and a type="initial and a statia" type="type = "type = constraints" in the statia type = type = constraints and a type = "type = "type = type = t | |
| 17 | <pre><pre>celement><data type="type">PLN</data><data <="" pre="" type="type"></data></pre></pre> | |
| 18 | <element><data type="type">NOM</data><data type="lemma">wwww</data><data type="string">www</data></element> | |
| 19 | <element><data type="type">SENT</data><data type="lemma">.</data><data type="string">.</data></element> | |
| 20 | <element><data type="type">NOM</data><data type="lemma">lemonde</data><data type="string">lemonde</data></element> <data type="string">lemonde</data> | |
| 21 | <element><data type="type">SENT</data><data type="lemma">.</data><data type="string">.</data></element> | |
| 22 | <element><data type="type">NOM</data><data type="lemma">fr</data><data type="string">fr</data></element> | |
| 23 | | |
| 24 E |] <1162 | |
| 20 | <pre>stante/atabutas/down/orbit/concounters/atabutas/internance/ colonearbit/atabutas/internance/internan </pre> | |
| 27 | como la solar a y por type "solar solar solar solar y por tempo" a la solar solar a y por solar solar a y por solar solar solar solar | |
| 28 | <pre><lement><data type="type">ADJ</data><data type="lemma">écologique</data><data type="string">écologique</data></lement></pre> | |
| 29 | <element><data type="type">PRP</data><data type="lemma">de</data><data type="string">de</data></element> | |
| 30 | <element><data type="type">DET:ART</data><data type="lemma">le</data><data type="string">la</data></element> | |
| 31 | <element><data type="type">NOM</data><data type="lemma">rivière</data><data type="string">rivière</data></element> | |
| 32 | <element><data type="type">ADJ</data><data type="lemma">Songhua</data><data type="string">Songhua</data></element> <data type="string">Songhua</data> | |
| 33 | <element><data type="type">PUN4:/data><data type="emma">,</data><data type="string">,</data></data></element> | |
| 34 | <element><datatype="type">trV+c/data><datatype="lemma">data><td></td></datatype="lemma"></datatype="type"></element> | |
| 36 | <pre><ceimetrix<datatype= type="">DET.Arti<data<datatype= termina="">textoata<type= string="">textoata</type=></data<datatype=></ceimetrix<datatype=></pre> | |
| 37 | solenoins solata type= type = "ven="ventata= solata type=" emma"> | |
| 38 | <pre></pre> | |
| 39 | <element><data type="type">DET:ART</data><data type="lemma">le</data><data type="string">les</data></element> | |
| 40 | <element><data type="type">NOM</data><data type="lemma">tentative</data><data type="string">tentatives</data><delemma">tentatives<data type="string">tentatives</data><delemma< p=""></delemma<></delemma"></element> | |
| 41 | <element><data type="type">VER:pres</data><data type="lemma">faire</data><data type="string">faites</data></element> | |
| 42 | <element><data type="type">PRP:det</data><data type="lemma">au</data><data type="string">au</data></element> | |
| 43 | <pre><element><data type="type">NOM</data><data type="lemma">niveau</data><data type="string">niveau</data></element></pre> | |
| 44 | <element><datatype="type">ADJ<datatype="lemma">>tocal<datatype="string">>tocal<delement></delement></datatype="string"></datatype="lemma"></datatype="type"></element> | |
| 45 | <pre><element><data type="type">rrr</data><data type="lemma">pour</data><data type="string">pour</data></element></pre> | |
| 40 | selements/sdata/type=type=210/ms/sdata/sdat | |
| 48 | <pre><pre>celement><data type="type">PLN</data><data <="" pre="" type="type"></data></pre></pre> | |
| 49 | <element><data type="type">NOM</data><data type="lemma">affaire</data><data type="string">affaire</data></element> | |
| 50 | <element><data type="type">PUN</data><data type="lemma">,</data><data type="string">,</data></element> | |
| 51 | <element><data type="type">VER:pres</data><data type="lemma">avoir</data><data type="string">ont</data></element> | |
| 52 | <element><data type="type">VER:pper</data><data type="lemma">conduire</data><data type="string">conduit</data></element> <data type="string">conduit</data> <data type="string">conduit</data> | |
| 100 | selene este selete terre ille populate della terre ille ese della terre ille ese della terre ille terre della terre ille terre della terre ille terre | |





• Extraction des patrons via un script PERL à tester

- Présentation générale
- Présentation du script sur le blog pluriTAL
 - <u>http://www.cavi.univ-paris3.fr/ilpga/ilpga/tal/cours/bao3-</u> extractionterminologique.html
 - <u>http://tal-p3.wordpress.com/2006/02/15/boite-a-outils-serie-</u><u>3/</u>
 - Ces ressources sont aussi dans votre archive de travail (dossier bao-extraction-terminologique)





A vous de jouer...

Les listes de patrons construites ici seront réutilisées dans la *BàO série 4* !