Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

ERC-Project | Universität Hamburg
COBHUNI
Contemporary Bioethics and the History of the Unborn in Islam

POS tagging and conversion of multilingual texts into Annis

Alicia González Martínez

http://gitlab.com/kabikaj/freeling2annis

Outline

1 Presentation

2 Motivation

3 Workflow

    POS tagging: Freeling

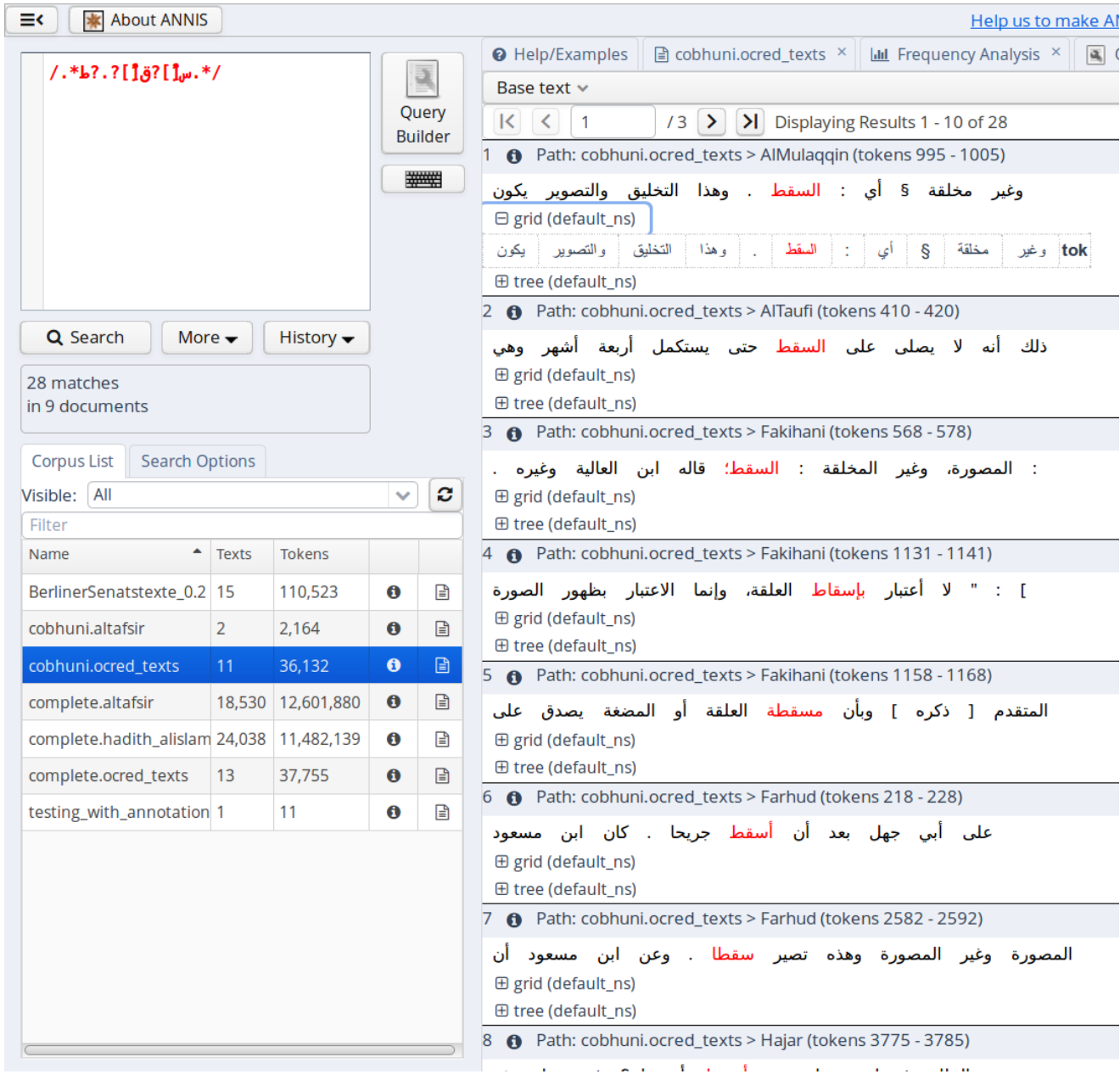
    Conversion: Pepper

    Visualization: Annis

http://gitlab.com/kabikaj/freeling2annis

Motivation

إِنَّمَا الْأَعْمَالُ بِالنِّيَّةِ وَإِنَّمَا لِامْرِئٍ مَا نَوَى فَمَنْ كَانَتْ هِجْرَتُهُ إِلَى اللَّهِ وَرَسُولِهِ فَهِجْرَتُهُ إِلَى اللَّهِ وَرَسُولِهِ وَمَنْ كَانَتْ هِجْرَتُهُ لِدُنْيَا يُصِيبُهَا أَوْ امْرَأَةٍ يَتَزَوَّجُهَا فَهِجْرَتُهُ إِلَى مَا هَاجَرَ إِلَيْهِ



http://gitlab.com/kabikaj/freeling2annis

Workflow

```
Perder la mirada,
distraídamente,

perderla y que nunca la
vuelva a encontrar:

y, figura erguida, entre
cielo y playa,

sentirme el olvido perenne
del mar.
```

FreeLing 4.0

form lemma POS

```
...
cielo cielo NCMS000 1
y y CC 0.999989
playa playa NCFS000 1
, , Fc 1
sentir sentir VMN0000 1
me me PP1CS00 1
el el DA0MS0 1
olvido olvido NCMS000 0.989726
perenne perenne AQ0CS00 1
de de SP 1
el el DA0MS0 1
mar. marzo NCMS000 1
```

freeling2xml.py

```
<?xml version="1.0" encoding="utf-8"?>
<a><b/>
 <content>
  <t POS="VMN0000" lemma="perder">Perder</t>
  <t POS="DA0FS0" lemma="el">la</t>
  <t POS="NCFS000" lemma="mirada">mirada</t>
  <t POS="Fc" lemma=",">,</t>
  <t POS="RG" lemma="distraídamente">distraídamente</t>
  <t POS="Fc" lemma=",">,</t>
  <t POS="VMN0000" lemma="perder">perder</t>
  <t POS="PP3FSA0" lemma="lo">la</t>
  <t POS="CC" lemma="y">y</t>
```

pepper

createPepperWorkflow.py

ANNIS

http://gitlab.com/kabikaj/freeling2annis