:acoli dc:isPartOf <http://uni-frankfurt.de>.

Datathon Miniproject
Linking XML corpora

Tutors: Maxim Ionov, Christian Chiarcos

Participants: Sabine Tittel, ...

HEIDELBERGER AKADEMIE
DER WISSENSCHAFTEN

GOETHE
UNIVERSITÄT
FRANKFURT AM MAIN

Corpora and edited works in the Humanities

- TEI/XML as a de facto standard
- Linked Data in Digital Philology
  - prosopography ([social] network analysis)
  - managing cross references
- Little to no application of LOD for resource modeling
  - the lack of convincing prototypes
  - *in this field,* XML is likely to stay

TEI and LOD

- the relevance of LOD has also been recognized in the TEI community
  - yet, there is no convenient formalization of semantically typed URI references in the TEI
    - beyond special-purpose applications (prosopography)
- here, experiment with using RDFa for the purpose
  - preserve TEI spirit and formalization
  - add LOD

Sample Data: Edition (printed book)

[14v°a] Ou nom de Dieu misericord. Cy commence le premier traictier de ceste oevre qui parle de l'anathomie et contient deux doctrines: la premiere doctrine parle de l'anathomie des membres *com*muns, universelz et simples. La seconde sera des me*m*bres propres, particuliers [14v°b] et compost. La premiere doctrine co*n*tient .v. chappitres: le premier, c'est ung chappitre universel qui parle de l'anathomie et de la nature des me*m*bres du corps.

[14v°a] POUR CE QUE, selon Galien, lumiere des medicins, ou .xvij.^{e} [14v°b] livre qui se intitule ‹De utilité des parties›, ou penultime chappitre, y sont quatre utilités de la science de anathomie: l'une, qui est la tres grande, pour amiracion de la puissance de Dieu; la seconde, po*ur* cognoistre les parties des paciens; la tierce, pour pronostiquer des [15r°a] disposicions du corps qui doivent avenir; la quarte si est pour curer les maladies. Et pour ce, c'est chose necessaire et prouffitable a ung ch*acu*m medicin de savoir la anathomie. Et c'est ce q*ue* disoit Galien ou livre qui se intitule ‹Liber scienciarum sive interiorum medico*rum*›, ou il dit ainsi: les jeunes clers, et les anciens aussi, estudient a cognoistre les parties et les passions d'icelles, car selon la difference d'icelles. Et ja soit ce que les parties qui appare*n*t ou sens soient cogneues appertemant, toutesvoies, celles qui sont en p*ar*font occultes, elles ont mestier de ho*m*me qui soit excercités en l'anathomie et es accions et utilités d'icelles. Et de ce lieu ci est prins le principes de tout le continent. Et dit qu'il est escript ou premier du livre des me*m*bres que le medicin hardi doit estre sage en la cognoissance des me*m*bres qui viennent en ch*acu*m lieu. Et se c'est chose prouffitable aux phisiciens, elle est plus necessaire aux

Sample Data: Edition (LaTeX source)

```
\ProvidesFile{EDITION.tex}[2017/06/13]

\begin{edition}{Kritischer Text}
\markboth{Die \emph{Anathomie}}{Kritischer Text}

\beginnumbering
\pstart
%
[14v\hoch{o}a] Ou nom\wdx{nom}{m. `mot
servant à désigner les êtres, les choses
qui appartiennent à une même catégorie
logique'}{\textbf{au nom de} \emph{`en vertu de'}}
de Dieu misericord\wdx{misericort}{adj. `qui a de la
miséricorde; miséricordieux'}{misericord
\emph{m.sg.}}. Cy
co\emph{m}mence\wdx{*comencier}{v.intr. `entrer
dans son commencement'}{commence
\emph{3.p.sg.
ind.prés.}} le premier\wdx{premier}{adj. `qui vient
avant les autres, dans un ordre; premier'}{}
traictier\wdx{*traitié}{m. `ouvrage
didactique, où est exposé d'une manière
systématique un sujet ou un ensemble de
```

Sample Data: Edition

- (snippet of the) edition of the Middle French *Grande Chirurgie* by *Gui de Chauliac* (15th c.)
  - text with glossary (originally typeset in LaTeX)

ryr et engendrer le corps. Les non naturelles sont devisees: en sont envoyés ens
1060 certains [32r°b] lieux pour aucuns aides, ou elles sont expellees du corps. Et sa-
ches que la cole est envoyee ou cestim du fiel, et la merancolie a l'esplein, et le
fleume aux joinctures, et la superfluité aigouse aux reins et a la vecie. Celles qui
sont expellees du corps, elles vont avec le sang et se putreffient aucunes fois et

edited text

cestim du fiel m. terme d'anat. 'vésicule bilaire' 1061
Isolierte Form. Cp. AmphYpL^2 f° 150a *la ciste ou bourse du fiel*. Cp. DEAF H 676,40ss. *huche du fiel* mit gleicher Bedeutung und den Kommentar ib. 673,42–674,3. Cp. *cistis fellis*.

sample gloss (\index)

Text edition (Mid. French, 15th c.)

```
Et
saches
que
la
<wdx>
    <orth>cole</orth>
    <lemma>cole</lemma>
    <gloss>f. terme de méd. `bile jaune (l'une des quat
</wdx>
est
envoyee
ou
cestim
du
<wdx>
    <orth>fiel</orth>
    <lemma>cestim du fiel</lemma>
    <gloss>m. terme d'anat. `vésicule bilaire'</gloss>
</wdx>
,
et
la
<wdx>
    <orth>merancolie</orth>
    <lemma>melancolie</lemma>
    <gloss>f. terme de méd. `bile noire (l'une des quat
    <var>merancolie</var>
</wdx>
a
<wdx>
    <orth>l'esplein</orth>
    <lemma>esplein</lemma>
    <gloss>m. terme d'anat. `organe lymphoïde situé sou
</wdx>
```

- philological edition
  - primary text
  - plus in-line glossary
- XML
  - converted from LaTeX sources
  - one-word-per-line formatting for viewing convenience only

Sample Data: Dictionary

- http://deaf-server.adw.uni-heidelberg.de/lemme/fiel

FIEL m.

[Étymologie]

(*fiel*, *fel* ca. 1000, *feel*, *fele*, *feil*, *fieil*, *fius*)

◆ **1°** t. de méd. "liquide verdâtre et amer qui est contenu dans la vésicule biliaire, bile" (dep. ca. 1160, EneasS^{2} 8221 [*el cors m'as mis une amartume Peor que sui*

apelez arc ou ciel. Couleur avra semb

Fevres f°25r°7(3); f°97v°10; f°165v°2; e

TLF 8,844b ['vieilli']; FEW 3,445a)

◆ dans des collocations *huche de*

foie et qui emmagasine la bile, vésicu

fel]; MoamT II 48,17 [*burse dou fiel*

dedens, et sont .6.: l'estomach, les bo

l'esplain, la ciste ou bourse du fiel];

Hu 2,160a [CERTEAU]; Li 1,1666b; DM

```
<part type="mainpart">
    <title>
        <lemma developed="false" language="afr.">fiel</lemma>
        <pos>m.</pos>
    </title>
    <etymology>
    <variants>
    <senses>
        <sense>
            <description>&#x2060;<m:terminology type="medecine">t. de
                m&#xE9;d.</m:terminology>
            <m:definition>liquide verd&#xE2;tre et amer qui est contenu dans la
                v&#xE9;sicule biliaire,
            bile</m:definition>&#x2060;</description>
            <datings>
                <dating type="since">
                    <date numeric="1160">ca.1160</date>
                </dating>
            </datings>
            <references type="primary">
                <reference id="10698">
                    <m:date numeric="1160">ca.1160</m:date>
                    <m:siglum>EneasS<m:sup>2</m:sup></m:siglum>
                    <m:text-reference>8221</m:text-reference>
                    <context>
                        <m:quotation>el cors m&#x2019;as mis une amartume Peor que suie
                            ne que fiel</m:quotation>
                    </context>
                </reference>
```

Miniproject

Data:

- edition
  - text
  - glosses
- dictionary

Ultimate goal:

- dictionary and glossary as lemon
- edited text as TEI/XML
- cross-references between edited text (XML) and lemon dictionary

Here:

- model for referencing lemon data from the edited text
- RDFa+XML

for this and similar datasets provided by participants