NATerm - authoring tool for terminologies

J.João Almeida   Alberto Simões   Alvaro Iriarte

Dep. de Informática, Univ. do Minho

IPCA

ELACH, Univ. do Minho

December 2021

> **Claim**
> - Crucial problem in translation
> - Not taken as seriously as it should

Different goals:
- glossaries – simple list of (term lang-source , term lang-target)
- terminologies – (multilingual) complex ontology of (Technical) concepts and terms:
  - information fields about the concepts, definition, about each term
  - scope notes

Chalange: Takes time and practice to master terminologies

- concept-oriented vs. word-oriented
- conceptual relations
- attributes, terms, information fields, ...
- relation properties, inference
- micro-structure
- macro-structure

NATerm = Natura Ascii-based Terminology toolkit

to...

- (imediatly) build your own terminologies
- discuss and play with core central concepts
- discuss the best way to express your models

- Multiterm (terminologies, Trados, multiterm-xml)
- glossaries of SmartCAT
- TEI dictionaries
- TBX
- XDXF
- stardict, stardict-editor
- goldendict
- FixMe : ... help me to add some more
- LaTeX + (lua)naterm.sty
- projects excel-based (ex: http://yourterm.eu)

Difficulties

- hundreds of mouse clicks to create an entry
- difficult do cooperate, versions
- installation
- ...some times Low-level

```
%title Dictionary EN - PT
%author JJoão
%lang EN PT ES
%inv hpr hyp

PT: gato
+G: m
EN: cat
hpr: animal

PT: cão
+G: m
EN: dog
ES: perro
def: ....
!img: dog.jpg
hpr: animal
```

- Textual input (just a text editor):
  - similar approach to markdown or wikis
  - %... - directives, metadata
  - entries separated by empty-lines
- Command line
  naterm -html f.naterm
  ... creates "f.html"
- Web interface (no installation; rudimentary)
  ... uploads "f.naterm" and "MEDIA.zip"
  ... downloads "t.pdf", "f.xdxf", etc

Variety of output formats

- HTML (search, links)
- PDF dictionary
- XDXF (XML format for dictionaries) – goldendict, ...
- Stardict – goldendict, startdict, ...
- ... other formats planned

Micro structure (fields, attributes)

- Concept oriented
- 2 level field:
    - Language term (EN: cat):
        - sub-fields of term (+Gender: masculine)
    - Concept relations (HPR: animal)
    - data propertie (def: is the second month of the year):
        - sub-fields (+src: wikipedia)
        - attribute values – text, number, multimedia,
- relations and properties (ex: InverseOf)

- Macro-structure (domain trees)
- Advanced... – Entries from CSV (tables)

a1.naterm monolingual glossary

```
EN: train
def: a very interesting invention used to ....
```

naterm -html a1.naterm
naterm -lang=EN -html a1.naterm

a2.naterm (EN – PT) dictionary

```
EN: train
PT: comboio
PT: trem

EN: car
PT: carro
PT: autmóvel
```

naterm -html a2.naterm

- a1:
  - language (EN)
  - textual information fiels (def)
  - 1 entry
- a2:
  - language (EN, PT)
  - 2 entries (concept)
- Command line
  naterm -html f.naterm
  naterm -lang=EN -html f.naterm
- Next : add metadata

a3.naterm monolingual glossary

```
%lang EN PT
%tit Dictionary of Transportes
%author JJoão; Alberto; Alvaro
%date Dec 2021
%pre {In this fantastic dicionary we
  ...
  Contributions Welcome.}

%inline var

EN: train
PT: comboio
+var: pt
PT: trem
+var: br

EN: car
PT: carro
PT: automóvel
```

naterm -html a3.naterm

- Metadados:
    - title, author, date
- Introduction, Appendices, (%pre   %pos)
- Visual config
  (%inline, %ignore, %rename)
- Language (%lang   %rellang)
- Relation properties (%inv)
- Next : more on Micro structure

a4.naterm [→]

```
%lang EN PT
%rellang EN
%inv hpn hpr
%inv hpr hpn
%inline var

EN: train
+ex: If you miss the train I'm on, ...
PT: comboio
+var: pt
+ex: O comboio chegou à tabela
PT: trem
+var: br
!img: train.jpg
hpn: TGV
hpr: transports

EN: guitar
PT: viola
hpr: musical instrument
!snd: guitar.mp3
```

- Images (!img: ....)
- Sound (!snd: ....)
- subfields (+ex: ....)
- several terms otained by inference (transportes, TGV, ...)
- Next : more on Macro structure

Macro-structure Domain trees

a5.naterm Animals P[→] H[→] N[→]

```
%tit Polish:English
%author Sylwia Foryńska
%lang ENG PL PT
%rellang ENG
%inv dom voc
%inline G


==  animals

===  birds
====  bird of prey

EN: eagle


===  mammals
====  canines
====  felines


EN: lion

===  amphibians
===  reptiles
===  insects
```

Domain tree (navigation, ...):

- top-down:
    - TOP subdom animal
    - animal subdom mammal
    - mammal subdom feline
    - feline voc {lion, cat, ...}
- bottom-up:
    - {lion, cat, ...} dom feline
    - feline supdom mammal
    - mammal supdom animal
    - animal supdom TOP

InverseOf:

- subdom ←→ supdom
- dom ←→ voc

a6.naterm Animals P[→] H[→]

```
%tit  Csv table example
%lang PT EN
%inv hpr hpn
%inv hpn hpr

PT: $1
EN: $2
hpn: $3
*tab{
árvore :: tree :: bananeira | palmeira | carvalho
fruto :: fruit :: maça | banana | uvas
bebida :: drink :: vinho | água | chá | café
flor :: flower :: rosa | girassol | lírio
}
```

- For each line in a csv-table :
    - split by '::'
    - define a new concept-entry
    - field 1 — $1
- tables:
    - external: * tab(food.csv)
    - inline: * tab{ line1 ....}

Dicionário de áreas vocabulares

Tabelas
Inererence
naterm → (PDF, HTML, ...)

Importance of:

- textual syntax
- build an output in a short time
- consequence of conceptual relations (simple inference)
- multimedia
- micro-structure
- macro-structure (domain, trees)
- good-looking output

- Improve HTML output:
  - options for:
    - multi page, multi lingual navigation, ... site
    - search
- webservice
- options for selecting single-sense terms (def. is multi-sense)
- better error messages, and documentation
- improve TBX output
- produce microstruture atribute-tree for documentation