VisDic – Wordnet Browsing and Editing Tool

Aleš Horák

Faculty of Informatics, Masaryk University
Botanická 68a, CZ-602 00 Brno, Czech Republic
E-mail: hales@fi.muni.cz

Euromasters Summer School 2005
Outline

1 Motivation and History
   - Motivation
   - History
   - Design Goals

2 VisDic Functionality
   - Basic Functionality
   - Advanced Features

3 Conclusions and Future Directions
   - Conclusions
   - Future Directions
Outline

1 Motivation and History
   - Motivation
   - History
   - Design Goals

2 VisDic Functionality
   - Basic Functionality
   - Advanced Features

3 Conclusions and Future Directions
   - Conclusions
   - Future Directions
1 Motivation and History
   - Motivation
   - History
   - Design Goals

2 VisDic Functionality
   - Basic Functionality
   - Advanced Features

3 Conclusions and Future Directions
   - Conclusions
   - Future Directions
Motivations for VisDic Tool

- the need for Wordnet Editing Tool
- need for multilingual approach in EuroWordNet and Balkanet projects
- EuroWordNet tools – Polaris (editor) and Periscope (browser)

Polaris:
- licensed software
- for MS Windows only
- development is closed
Motivations for VisDic Tool

- the need for Wordnet Editing Tool
- need for multilingual approach in EuroWordNet and Balkanet projects
- EuroWordNet tools – Polaris (editor) and Periscope (browser)
- Polaris:
  - licensed software
  - for MS Windows only
  - development is closed
Motivations for VisDic Tool

- the need for Wordnet Editing Tool
- need for multilingual approach in EuroWordNet and Balkanet projects
- EuroWordNet tools – Polaris (editor) and Periscope (browser)
  - Polaris:
    - licensed software
    - for MS Windows only
    - development is closed
Motivations for VisDic Tool

- the need for Wordnet Editing Tool
- need for multilingual approach in EuroWordNet and Balkanet projects
- EuroWordNet tools – Polaris (editor) and Periscope (browser)
- Polaris:
  - licensed software
  - for MS Windows only
  - development is closed
History of VisDic

- NLP laboratory at FI MU (Karel Pala and Tomáš Pavelek) started to work on VisDic
- current team at NLP lab: Aleš Horák, Pavel Rychlá, Karel Pala and others (contact visdic@aurora.fi.muni.cz)
History of VisDic

- NLP laboratory at FI MU (Karel Pala and Tomáš Pavelek) started to work on VisDic
- current team at NLP lab: Aleš Horák, Pavel Rychlí, Karel Pala and others (contact visdic@aurora.fi.muni.cz)
History of VisDic

- NLP laboratory at FI MU (Karel Pala and Tomáš Pavelek) started to work on VisDic
- current team at NLP lab: Aleš Horák, Pavel Rychlý, Karel Pala and others (contact visdic@aurora.fi.muni.cz)
VisDic Design Goals

- tool for presentation and editing of XML databases
- highly configurable
- multiplatform (Linux and Windows)
- primarily for WordNets, but adoptable to various dictionary types (monolingual, translational, thesaurus or generally linked wordnet lexicon)
# VisDic Design Goals

- tool for presentation and editing of XML databases
- highly configurable
  - multiplatform (Linux and Windows)
- primarily for WordNets, but adoptable to various dictionary types (monolingual, translational, thesaurus or generally linked wordnet lexicon)
### VisDic Design Goals

- tool for presentation and editing of XML databases
- highly configurable
- multiplatform (Linux and Windows)
- primarily for WordNets, but adoptable to various dictionary types (monolingual, translational, thesaurus or generally linked wordnet lexicon)
VisDic Design Goals

- tool for presentation and editing of XML databases
- highly configurable
- multiplatform (Linux and Windows)
- primarily for WordNets, but adoptable to various dictionary types (monolingual, translational, thesaurus or generally linked wordnet lexicon)
Basic Functions of VisDic

- powerful query language
- multiple views of multiple wordnets
- freely defined text views
- editing
- Tree and RevTree
- query result and external entry lists
- plain XML view

and Advanced Features ...
Basic Functions of VisDic

- powerful query language
- multiple views of multiple wordnets
- freely defined text views
- editing
- Tree and RevTree
- query result and external entry lists
- plain XML view

and Advanced Features ...
Basic Functions of VisDic

- powerful query language
- multiple views of multiple wordnets
- freely defined text views
- editing
- Tree and RevTree
- query result and external entry lists
- plain XML view

and Advanced Features...
Basic Functions of VisDic

- powerful query language
- multiple views of multiple wordnets
- freely defined text views
- editing
- Tree and RevTree
- query result and external entry lists
- plain XML view

and Advanced Features …
Basic Functions of VisDic

- powerful query language ▶ screenshots
- multiple views of multiple wordnets ▶ screenshots
- freely defined text views ▶ screenshots
- editing ▶ screenshots
- Tree and RevTree ▶ screenshots
- query result and external entry lists ▶ screenshots
- plain XML view ▶ screenshots

and Advanced Features . . .
Basic Functions of VisDic

- powerful query language
- multiple views of multiple wordnets
- freely defined text views
- editing
- Tree and RevTree
- query result and external entry lists
- plain XML view

and Advanced Features …
Basic Functions of VisDic

- powerful query language
- multiple views of multiple wordnets
- freely defined text views
- editing
- Tree and RevTree
- query result and external entry lists
- plain XML view

and Advanced Features ...
Basic Functions of VisDic

- powerful query language
- multiple views of multiple wordnets
- freely defined text views
- editing
- Tree and RevTree
- query result and external entry lists
- plain XML view

and Advanced Features...
Advanced Features

Advanced Functions of VisDic

- synchronization – *Show (by)* and *AutoLookUp (by)*
- editing support – inter-dictionary copying with *Take key from* or *Copy entry to*
- tree browsing – *Topmost entries* and *Full expansion*
- consistency checks – user defined duplicity checks
- journaling
- XML configuration
Advanced Features

Advanced Functions of VisDic

- synchronization – *Show (by)* and *AutoLookUp (by)*
- editing support – inter-dictionary copying with *Take key from* or *Copy entry to*
- tree browsing – *Topmost entries* and *Full expansion*
- consistency checks – user defined duplicity checks
- journaling
- XML configuration
Advanced Functions of VisDic

- synchronization – *Show (by)* and *AutoLookUp (by)*
- editing support – inter-dictionary copying with *Take key from* or *Copy entry to*
- tree browsing – *Topmost entries* and *Full expansion*
- consistency checks – user defined duplicity checks
- journaling
- XML configuration
Advanced Functions of VisDic

- synchronization – *Show (by)* and *AutoLookUp (by)*
- editing support – inter-dictionary copying with *Take key from* or *Copy entry to*
- tree browsing – *Topmost entries* and *Full expansion*
- consistency checks – user defined duplicity checks
- journaling
- XML configuration
Advanced Features

Advanced Functions of VisDic

- synchronization – *Show (by)* and *AutoLookUp (by)*  
  - [screenshots]
- editing support – inter-dictionary copying with *Take key from* or *Copy entry to*  
  - [screenshots]
- tree browsing – *Topmost entries* and *Full expansion*  
  - [screenshots]
- consistency checks – user defined duplicity checks  
  - [details]
- journaling  
  - [details]
- XML configuration  
  - [details]
Advanced Functions of VisDic

- synchronization – *Show (by)* and *AutoLookUp (by)*
- editing support – inter-dictionary copying with *Take key from* or *Copy entry to*
- tree browsing – *Topmost entries* and *Full expansion*
- consistency checks – user defined duplicity checks
- journaling
- XML configuration
Outline

1 Motivation and History
   - Motivation
   - History
   - Design Goals

2 VisDic Functionality
   - Basic Functionality
   - Advanced Features

3 Conclusions and Future Directions
   - Conclusions
   - Future Directions
Conclusions and Future Directions

VisDic usability:

- VisDic proved its usability in Balkanet
- it has achieved the leading role in many wordnet editing projects

Future of VisDic:

- specific support for linguists
- user interface and team cooperation
- VisDic successor – client-server lexical database editor DEB
Conclusions and Future Directions

VisDic usability:
- VisDic proved its usability in Balkanet
- it has achieved the leading role in many wordnet editing projects

Future of VisDic:
- specific support for linguists
- user interface and team cooperation
- VisDic successor – client-server lexical database editor DEB
Conclusions and Future Directions

VisDic usability:

- VisDic proved its usability in Balkanet
- it has achieved the leading role in many wordnet editing projects

Future of VisDic:

- specific support for linguists
- user interface and team cooperation
- VisDic successor – client-server lexical database editor DEB
Conclusions and Future Directions

VisDic usability:

- VisDic proved its usability in Balkanet
- It has achieved the leading role in many wordnet editing projects

Future of VisDic:

- Specific support for linguists
- User interface and team cooperation
- VisDic successor – client-server lexical database editor DEB
Conclusions and Future Directions

VisDic usability:
- VisDic proved its usability in Balkanet
- it has achieved the leading role in many wordnet editing projects

Future of VisDic:
- specific support for linguists
- user interface and team cooperation
- VisDic successor – client-server lexical database editor DEB
query language expression is constructed with rules:

Expr: \[ Eval \ Oper \ Expr \]
Expr: \[ Eval \mid (Expr) \]
Eval: \[ Tag \ Match \ STR \]
Eval: \[ Literal \]
Tag: \[ ID.Tag \]
Tag: \[ ID \]
Match: \[ == \mid = \]
Oper: \[ && \mid || \]
Literal: \[ STR \]
Literal: \[ STR:STR \]
Query language – examples

dog
  ... searches for all literals containing word 'dog'
dog:1
  ... searches for literal 'dog' with sense '1'
"hunting dog"
  ... searches for all literals containing string 'hunting dog'
/butt.*/
  ... searches for all literals starting with 'butt' (butter, button, butterfly, ...)
SYNSET.ID=ENG20-00003009-n
  ... searches for synsets with id 'ENG20-00003009-n'
SYNSET.POS=/[an]/
  ... searches for synsets with POS 'a' or 'n'
POS: n  ID: ENG20-02001223-n  BCS: 3
Synonyms: dog:1, domestic dog:1, Canis familiaris:1

Definition: a member of the genus Canis (probably descended from the common wolf) that has been domesticated by man since prehistoric times; occurs in many breeds
Usage: the dog barked all night
Domain: zoology
SUMO/MILO: + Canine

--- [hyponym] *[n] canine:2, canid:1
--- [holo_member] +[n] Canis:1, genus Canis:1
--- [holo_member] +[n] pack:6
<<- [hyponym] [n] pooch:1, doggie:1, doggy:1, barker:2, bow-wow:2
<<- [hyponym] +[n] cur:1, mongrel:2, mutt:1
<<- [hyponym] [n] lapdog:1
<<- [hyponym] +[n] toy dog:1, toy:4
<<- [hyponym] +[n] hunting dog:1
<<- [hyponym] +[n] working dog:1
<<- [hyponym] [n] dalmatian:2, coach dog:1, carriage dog:1
<<- [hyponym] [n] basenji:1
<<- [hyponym] +[n] pug:1, pug-dog:1
<<- [hyponym] [n] Leonberg:1
<<- [hyponym] [n] Newfoundland:1
<<- [hyponym] [n] Great Pyrenees:1
<<- [hyponym] +[n] spitz:1
přístroj na měření teploty

Celsiusův, Réamurův, Fahrenheitův; rtuťový, lihový, plynový t.

sač1

n

teoměř

hyponym
- H*[n] organism:1, being:2
  - HHM*[n] person:1, individual:1, someone:1, somebody:1, mortal:1, h
  - H*[n] leader:1
- H*[n] head:4, chief:1, top dog:1

- [n] bank commissioner:1
  - H [n] commissioner:1
  - H [n] administrator:3, executive:3
- H *[n] head:4, chief:1, top dog:1
  - H *[n] leader:1
Query: SYNSET.POS=n
Query: SYNSET.BCS=/[12]/
<SYNSET>
  <ID>ENG20-04919813-n</ID>
  <POS>n</POS>
  <SYNONYM>
    <LITERAL>body part</LITERAL>
    <SENSE>1</SENSE>
  </SYNONYM>
  <LITERAL>
  </LITERAL>
  <SYNONYM>
    <ILR>ENG20-08797461-n</ILR>
    <TYPE>hyponym</TYPE>
  </SYNONYM>
  <ILR>ENG20-00003226-n
    <TYPE>holo_part</TYPE>
  </ILR>
  <DEF>any part of an organism such as an organ or extremity</DEF>
  <BCS>1</BCS>
  <DOMAIN>anatomy</DOMAIN>
  <SUMO>BodyPart</SUMO>
  <TYPE>_</TYPE>
</SUMO>
<RILR>ENG20-01814160-n
  <TYPE>hyponym</TYPE>
</RILR>
<RILR>ENG20-01818138-n
  <TYPE>hyponym</TYPE>
</RILR>
<RILR>ENG20-01820900-n
  <TYPE>hyponym</TYPE>
</RILR>
<RILR>ENG20-02074327-n
  <TYPE>hyponym</TYPE>
</RILR>
<RILR>ENG20-02231455-n
  <TYPE>hyponym</TYPE>
</RILR>
</SYNSET>
Number of entries: 1203
Serbian WordNet

Definition: Za zvuk, posebno glas.

Portion of složeným z rožlíných rostlinných společenstev, zejm. z trav a bylin.

Usage: rozkvetlé, horská l.; úrodné louky, posekat louku

Synonyms: piskav:1, tanak:5, tanusxan:1

Domain: geography

SUMO/MILO

O: = Field

-->
Serbian WordNet
Journaling

- all user actions are *logged* in a *journal*
- supports team work – user journals can be *applied* to *central wordnet*
VisDic XML Configuration

- **global configuration** in `visdic.cfg`
  - colors, fonts, dictionaries, open dictionaries, autolookups, ...

- **dictionary specific configuration** in `dictionary.cfg`
  - name, main tag, charset, ...
  - definitions of *visuals*
  - definitions of *views* (panels)
  - definitions of *consistency checks*

- **dictionary structure definition** in `dictionary.def`
  - similar purpose as DTD – Document Type Definition
visdic.cfg

<?xml version="1.0"?>
<CONFIG>Visdic general configuration file
  <COLOR>Colors definition
    <BLACK>0x0000, 0x0000, 0x0000</BLACK>
    <WHITE>0xffffff, 0xffffff, 0xffffff</WHITE>
    ...
  </COLOR>
  <FONT>Fonts definition</FONT>
  <APPL>
    <DICT>/nlp/wn/visdic/data/eng20/wneng20</DICT>
    <DICT>/nlp/wn/visdic/data/eng171/wneng171</DICT>
    <DICT>/nlp/wn/visdic/data/eng15/wneng15</DICT>
    ...
    <AUTOLOOKUP>v2-v1</AUTOLOOKUP>
    <OPEN>1
      <SIZE>43</SIZE>
      <HIST>
        <LINE>trench</LINE>
        <LINE>house</LINE>
        ...
      </HIST>
    </OPEN>
    <OPEN>5
      <SIZE>57</SIZE>
    </OPEN>
  </APPL>
</CONFIG>
wneng20.def

0 SYNSET 1 1  N
  1 ID 1 1  K
  1 POS 1 1  N
  1 SYNONYM 1 1  N
    2 LITERAL 1 -1  N
      3 SENSE 1 1  I  @maxbyparent+1
      3 LNOTE 0 1  N
  1 ILR 0 -1  L
    2 TYPE 1 1  N
  1 RILR 0 -1  R SYNSET.ILR
  1 BCS 0 1  N
  1 DEF 0 1  N
  1 USAGE 0 -1  N
  1 SNOTE 0 -1  N
  1 STAMP 0 1  N
  1 DOMAIN 0 1  N
  1 SUMO 0 -1  N
    2 TYPE 1 1  N