

Universal  
Networking  
Language



通用网络语言

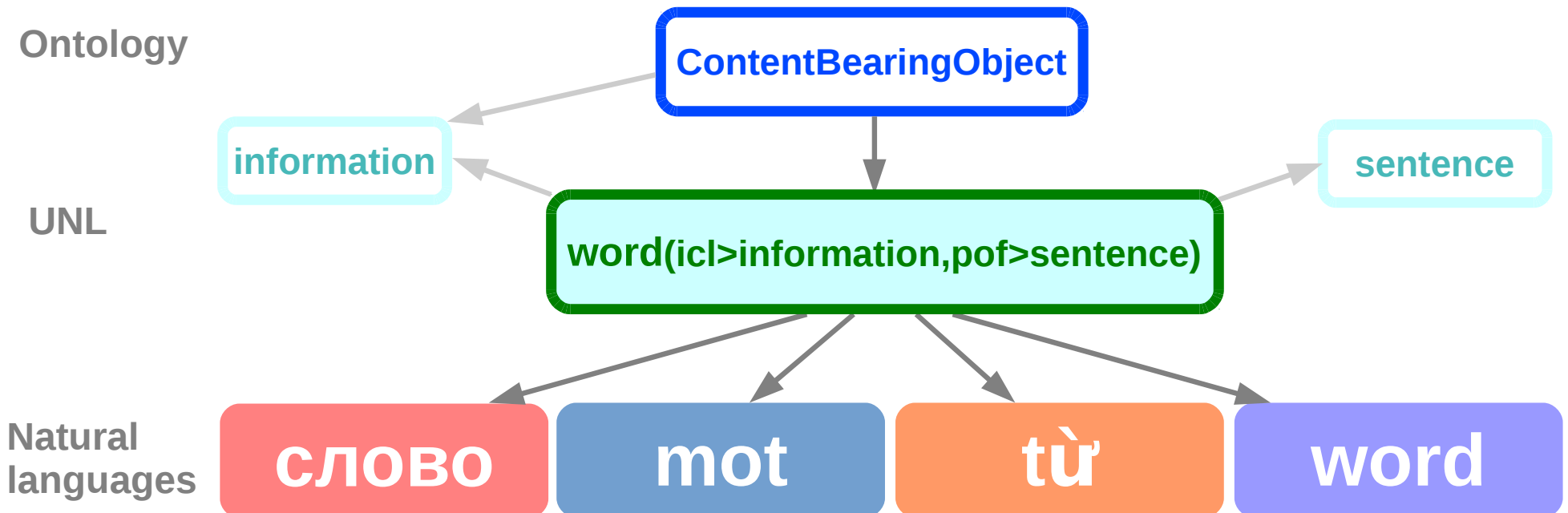
универсальный  
сетевой  
язык

## Development of lexical basis for the Universal Dictionary of UNL Concepts

Viacheslav Dikonov  
Dialog 2013

# What is it?

- Dictionary of the artificial interlingua UNL «Universal Networking Language».
- Basic lexical units of UNL are so-called concepts equal to lexical senses. Each concept has a unique identifier called "Universal Word" (UW)
  - UWs are linked to words and expressions of natural languages
  - UWs have semantic links to other UWs
  - UWs are linked to ontology classes.



## Current status

Part	Number of UW	Status
General lexics	82804	Downloadable
Terminology	688617	Under development
Named Entities	2109240	Soon to be released

- Universal Words are sorted into 3 groups: general lexics, terminology, named entities.
- Some of the data are already available for download.

# Natural Languages

Language	General	Terms	Names	Total	Quality
English	<b>82804</b> (40894 words)	688617	2109240	<b>2 880 661</b>	<b>*****</b>
Russian	<b>48555</b> (30818 words)	<b>688613</b>	<b>226595</b>	<b>963 763</b>	<b>**** Proofreading</b>
French	<b>36324</b> (25068 words)	103060	367888	<b>507 272</b>	<b>*** Autoranking</b>
Hindi	<b>27815</b> (30220 words)	0	10823	<b>38 638</b>	<b>*** Auto</b>
Spanish	<b>11758</b> (6983 words)	21990	298674	<b>332 422</b>	<b>** Experimental</b>
Malay	<b>21861</b> (17457 words)	0	46044	<b>67 905</b>	<b>** Experimental</b>
Vietnamese	<b>5927</b> (6456 words)	0	171367	<b>177 294</b>	<b>*** Experimental</b>

Number of UWs linked to NL words and expressions

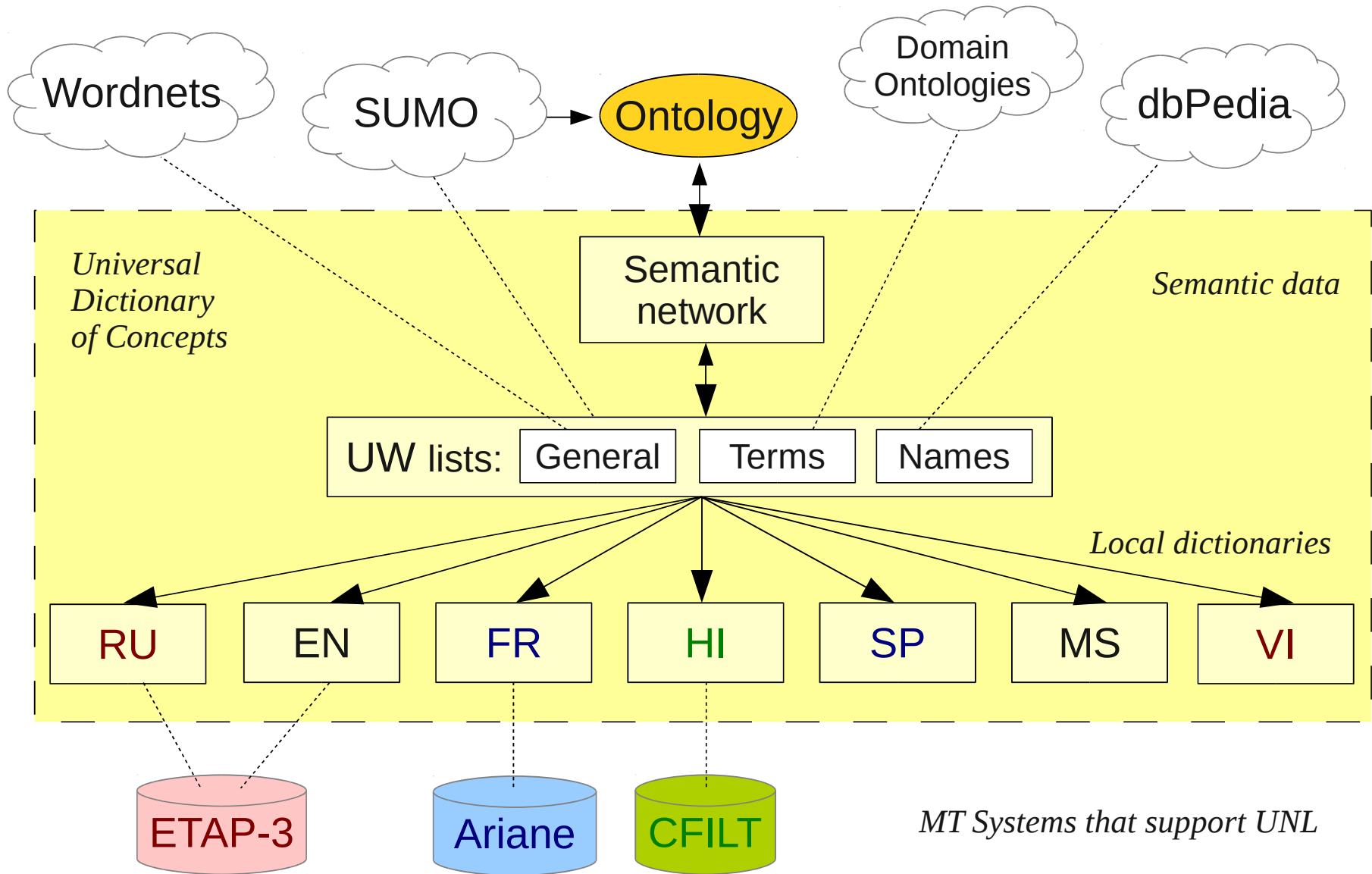
English is based on a subset of Princeton Wordnet

Russian dictionary is being proofread and extended (work in progress)

Links between UWs and other languages are ranked automatically

- Ranking is based on the number of sources confirming translations that can be deduced from the UNL dictionary.

# Links to other resources

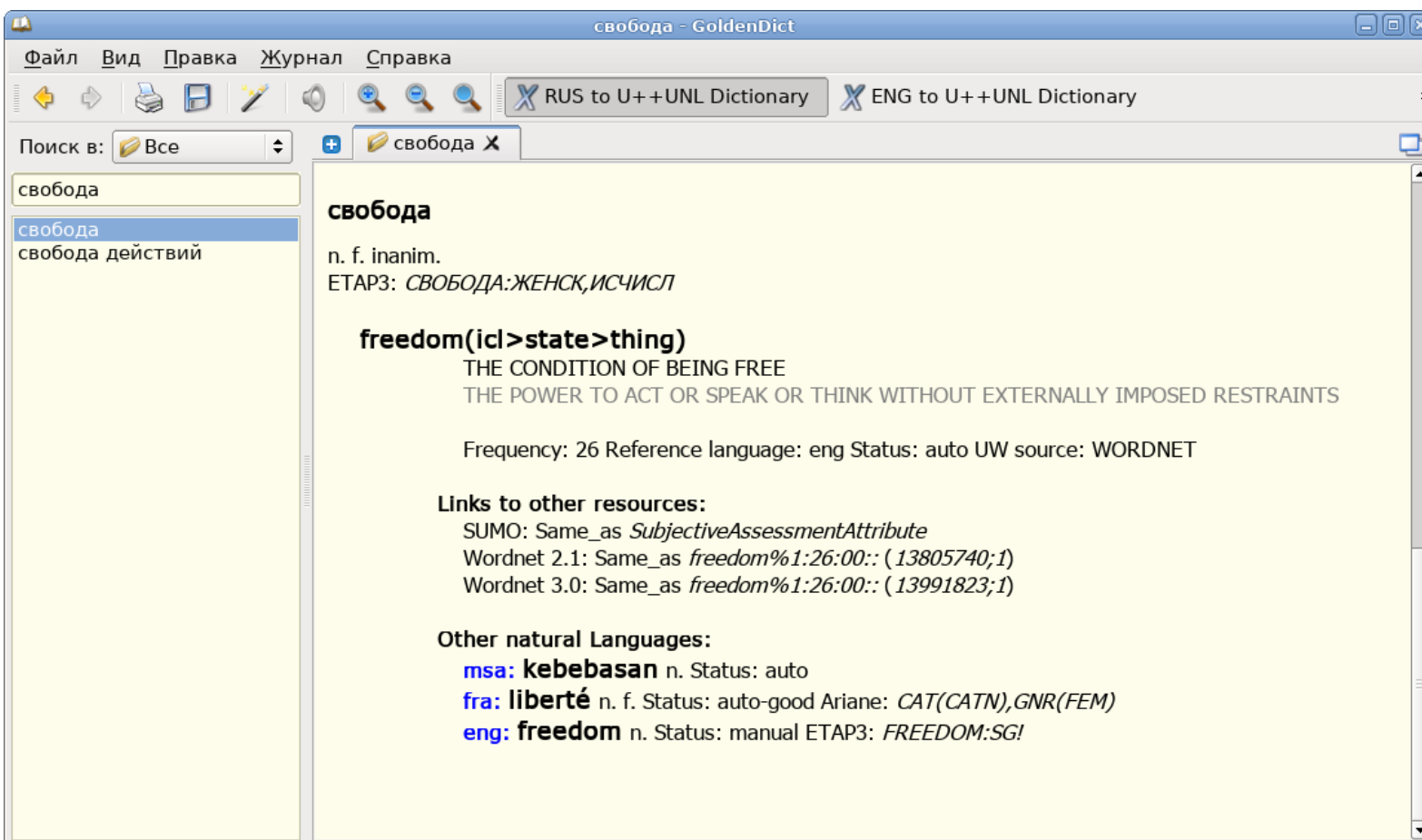


# Data files

- Files available at <http://atoum.imag.fr/geta/User/services/pivax/data/>
- Formats
  - CSV
  - XML (pivax, xdx) for various dictionary shells
  - RDF/Turtle (Open Linked Data) soon to be released
- Free to use under GPLv3+ or Creative Commons CC-BY-SA / CC-BY-SA-(NC)

# Questions?

## Welcome to the poster stand!



A sample search using goldendict dictionary shell...