I UNL Panel

December 15th, 2012
Coling 2012, Mumbai, India
www.unlweb.net/presentations/coling2012.pdf

Program

09:30 – Opening
10:00 – Introduction
10:30 – Tea break
11:00 – First session

- Nicoletta Calzolari (Istituto di Linguistica Computazionale Antonio Zampolli, Pisa)
- Mike Dillinger (eBay)
- Eric Wehrli (Université de Genève)

13:00 – Lunch break
14:30 – Second session

- Igor Boguslavsky (Universidad Politécnica de Madrid, Russian Academy of Sciences)
- Sameh Alansary (University of Alexandria, Library of Alexandria)
- Pushpak Bhattacharyya (IIT-Bombay)

16:30 – Tea break
17:00 – Discussion session
18:00 – Closing

UNDL FOUNDATION

Introduction

Outline

- History
- Commitments
- Properties
- Structure
- Road map
- Issues concerning UW's
- Goals

UNDL FOUNDATION

The Universal Networking Language (UNL)

UNDL FOUNDATION

History

1996-2001

- Extensive financial support
- Network topology (language servers)
- UNL Society
- Restricted access
- Instability

UNITED NATIONS UNIVERSITY

2002-2008

- Scarce financial support
- Schisms (U++, CDL, CWL)
- Stability
- Paradigm shift

Universal Networking Language

2009-

- UNL+3, UNL^{2015}
- Centralized repositories (UNL servers)
- UNL Community (www.unlweb.net)
- Transparency and openness (CC-BY-SA)
- Revision of the Specs

UNDL FOUNDATION

UNDL FOUNDATION

Commitments

1. The UNL must represent information

   The UNL must represent "what was meant" (and not "what was said").

2. The UNL must be a language for computers

   The UNL must be computable.

3. The UNL must be self-sufficient

   The UNL representation must not depend on any implicit knowledge.

4. The UNL must be general-purpose

   The UNL must not be bound to translation.

5. The UNL must be independent from any particular natural language

   As a language of the UN, the UNL must be neutral.

UNDL FOUNDATION

Properties

- Non-Ambiguity
  - the boys saw the girl with the telescope
  - [[the boys] [[[saw(icl>perceive) [the girl]] [with the telescope]]]]
- Non-Redundancy
  - Peter killed Mary ≅ Mary was killed by Peter ≅ Peter caused Mary to die
- Compositionality
  - John devoured thousands of books = John read many books
- Declarativeness
  - Can you pass me the salt? = (you pass the salt to me).@request.@polite
- Completeness
  - The monkey took the banana and ate it
  - The monkey_i took the banana_j and the monkey_i ate the banana_j

UNDL FOUNDATION

Structure

Information can be represented by semantic networks made of three different types of discrete semantic entities:

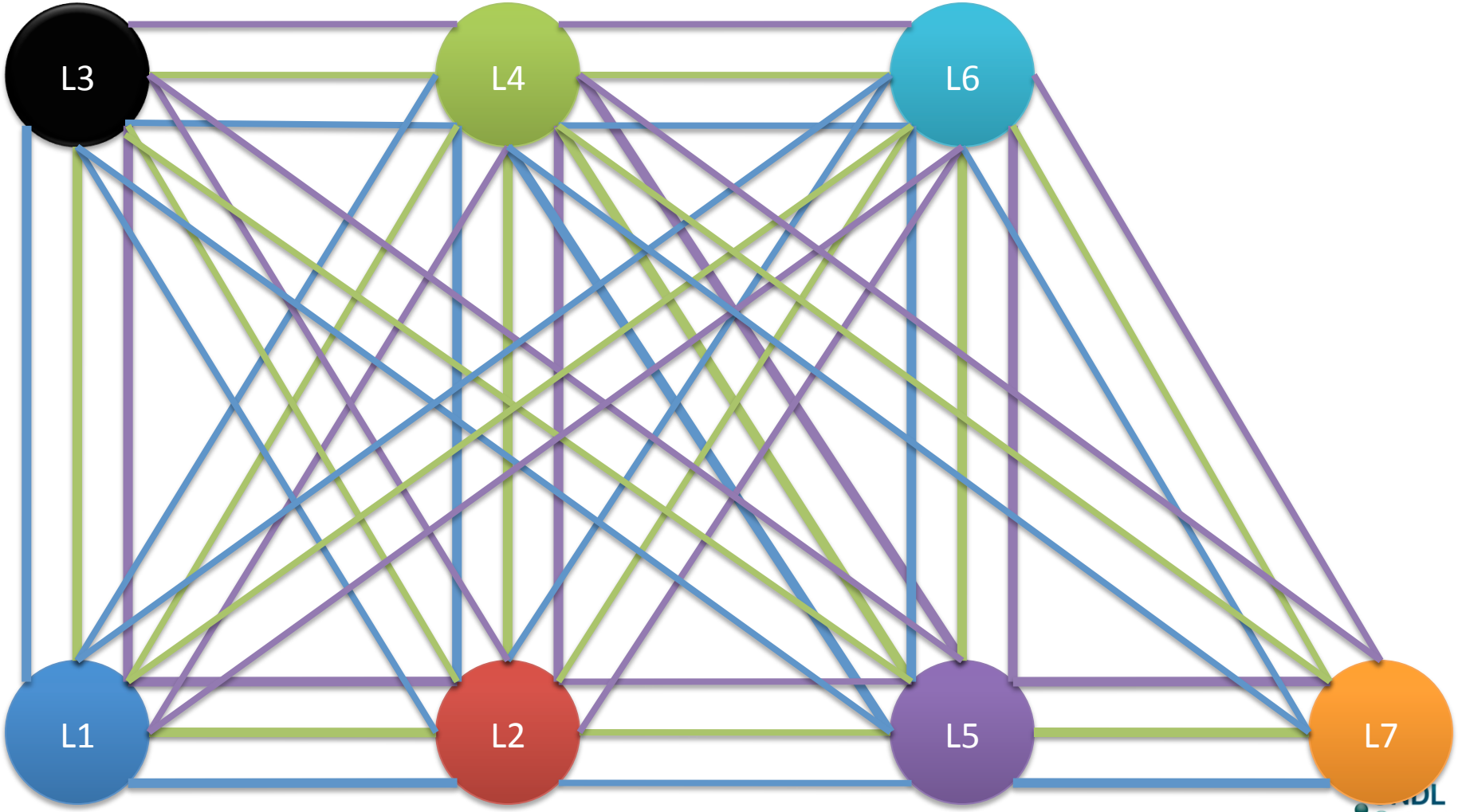
CONCEPTS = Universal Words (UWs)

CONCEPT SPECIFIERS = Universal Attributes

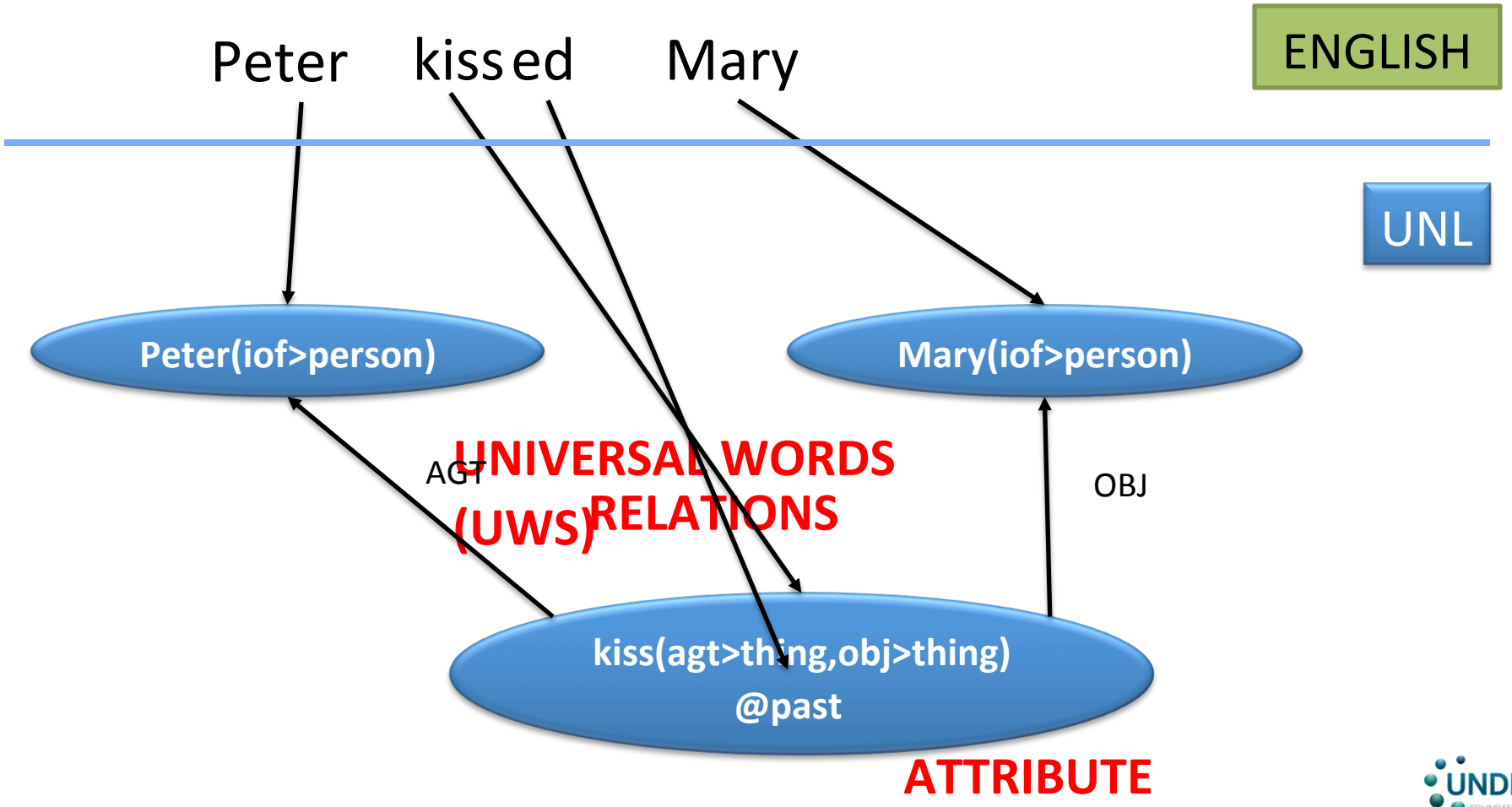
RELATIONS BETWEEN CONCEPTS = Universal Relations

UNDL FOUNDATION

The Universal NETWORKING Language



Natural Language-to-UNL (UNL-ization)



UNL document

[D]

[S:01]

{org:en}

Peter kissed Mary.

{/org}

{unl}

agt(kiss(agt>thing,obj>thing).@entry.@past, Peter(iof>person))

obj(kiss(agt>thing,obj>thing).@entry.@past, Mary(iof>person))

{/unl}

[/S]

[/D]

UNDL FOUNDATION

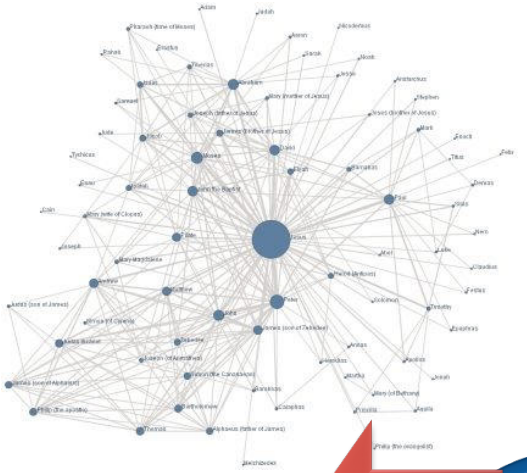
The UNL System

C2 cotidiano TERÇA-FEIRA, 23 DE JANEIRO DE 2007 FOLHA DE S.PAULO

RUBEM ALVES

O aluno perfeito

Ele se chamava Memorioso, pois seus pais julgavam que a memória perfeita é essencial para uma boa educação

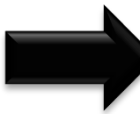


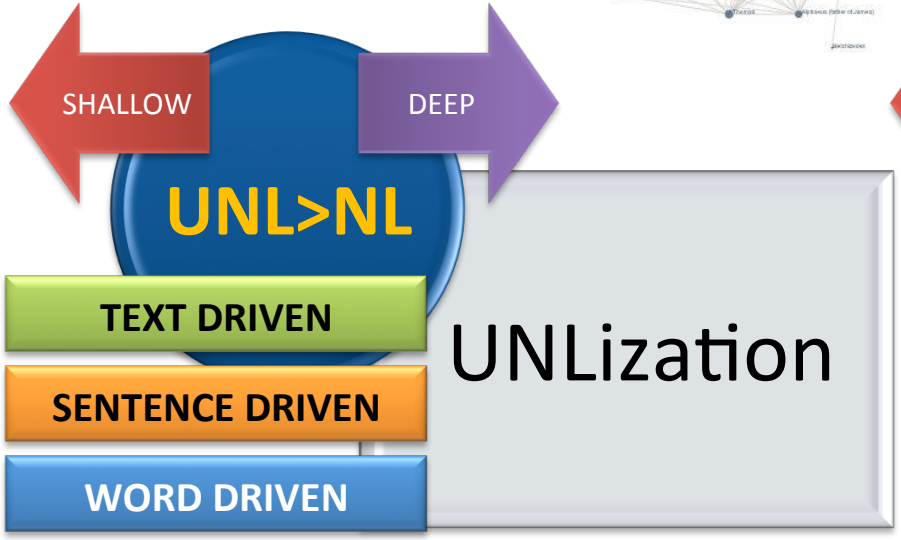
C2 cotidiano TERÇA-FEIRA, 23 DE JANEIRO DE 2007 FOLHA DE S.PAULO

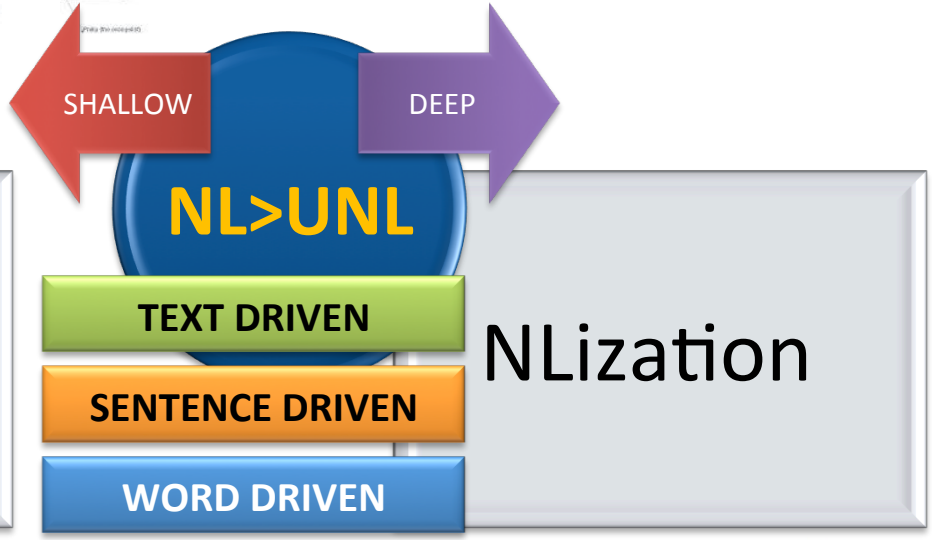
RUBEM ALVES

O aluno perfeito

Ele se chamava Memorioso, pois seus pais julgavam que a memória perfeita é essencial para uma boa educação

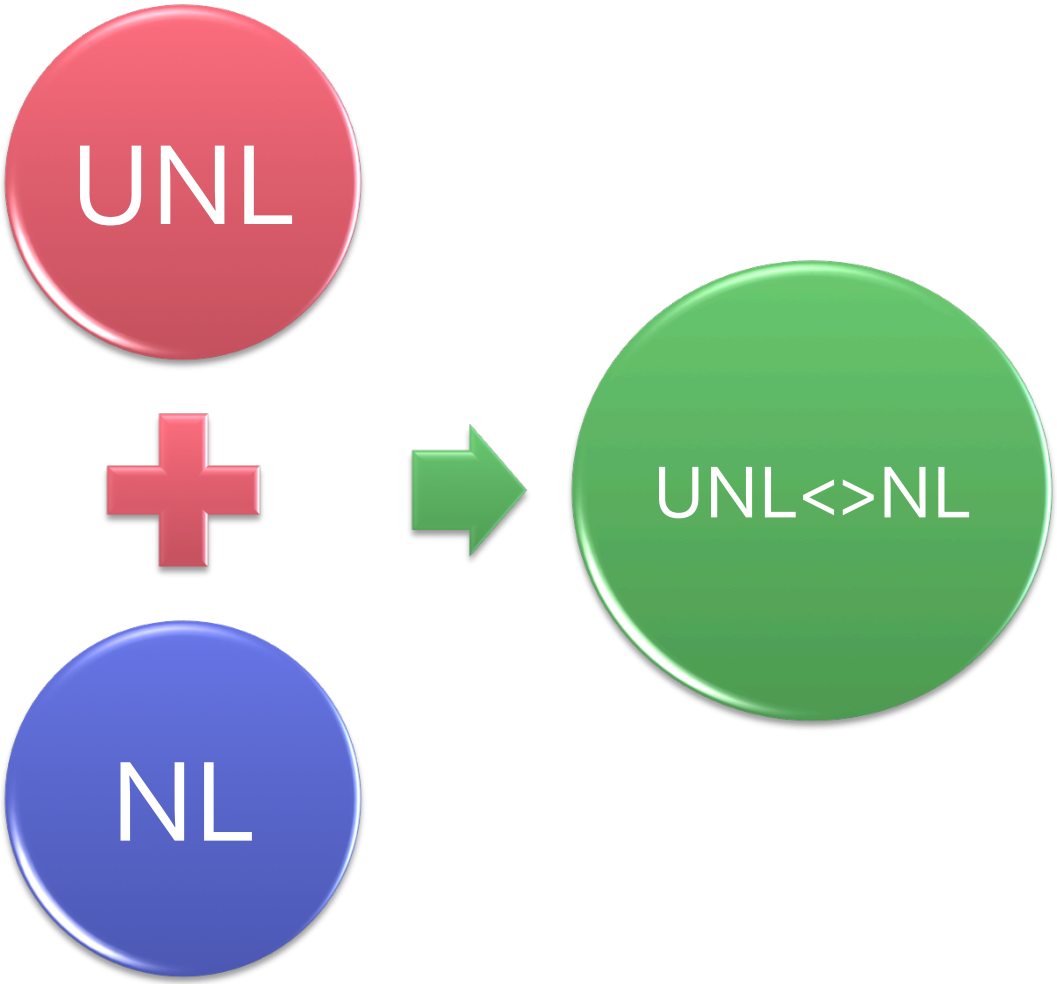






UNDL FOUNDATION

Resources



LEXICAL RESOURCES

GRAMMATICAL RESOURCES

COMPUTATIONAL RESOURCES

UNDL FOUNDATION

Universal Words

- nodes in the UNL graph
- temporary (not to be included in the UNL Dictionary)
  - 3.14159
  - www.undlfoundation.org
  - H_2O
- permanent (to be included in the UNL Dictionary)

TYPE	GRANULARITY	UNL (simplified)	ENGLISH
SIMPLE	NODE	big	big
COMPOUND	NODE + ATTRIBUTE	big.@more	bigger
COMPLEX	HYPER-NODE	obj(affix,stamp)	to stamp

UNDL FOUNDATION

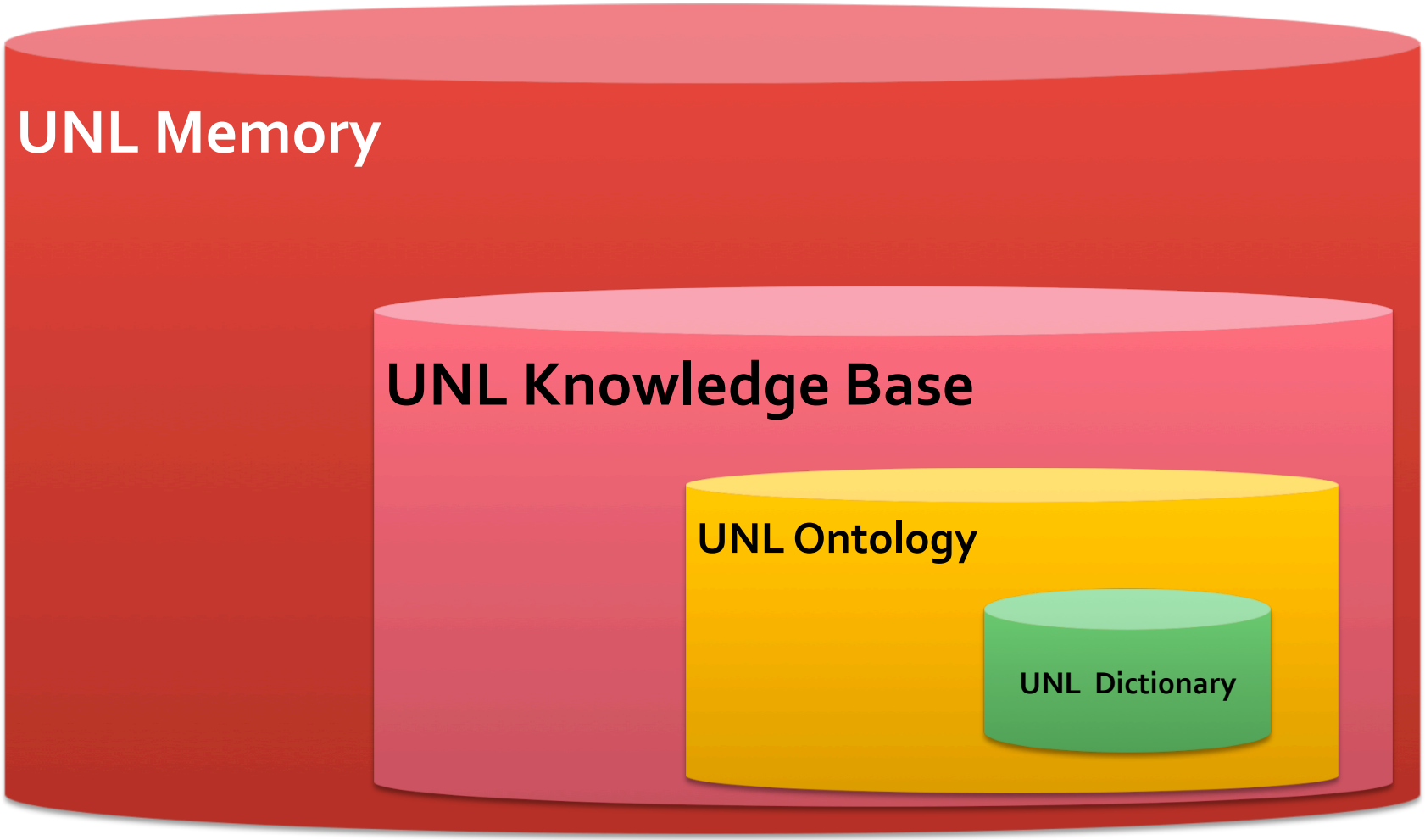
Semantic Accessibility

UCI (Uniform Concept Identifier)

- UCL (Uniform Concept Locator)
  - ucl://<AUTHORITY>/<ID>
    - ucl://unlkb.unlweb.net/104379964
- UCN (Uniform Concept Name)
  - ucn:<LID>:<NSS>
    - ucn:eng:table(icl>furniture)
    - ucn:fra:table(icl>mobilier)
    - ucn:esp:mesa(icl>mobiliario)
    - ucn:deu:Tisch(icl>Möbel)
    - ucn:rus:стол(icl>мебель)

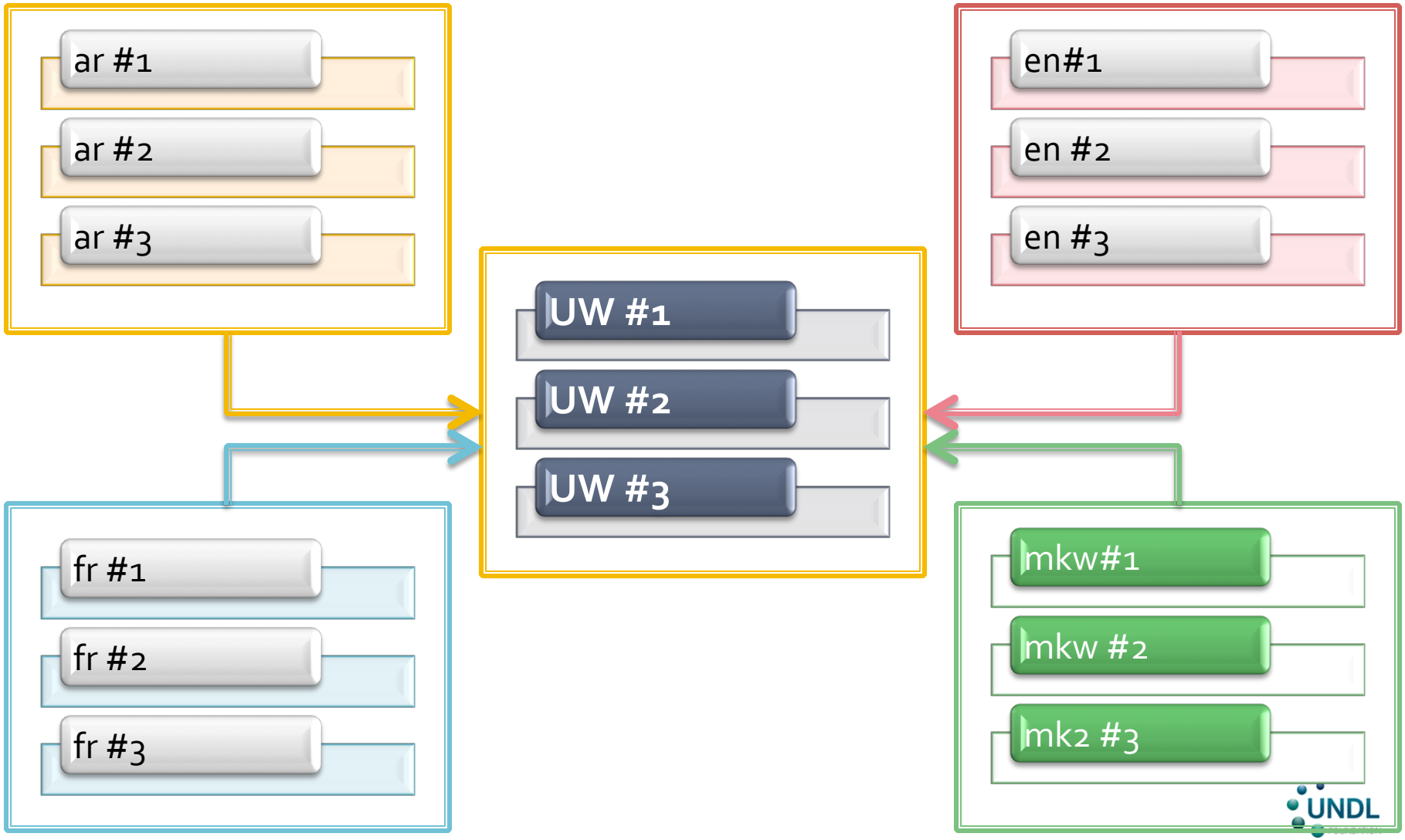
UNDL FOUNDATION

UNL Lexical Resources

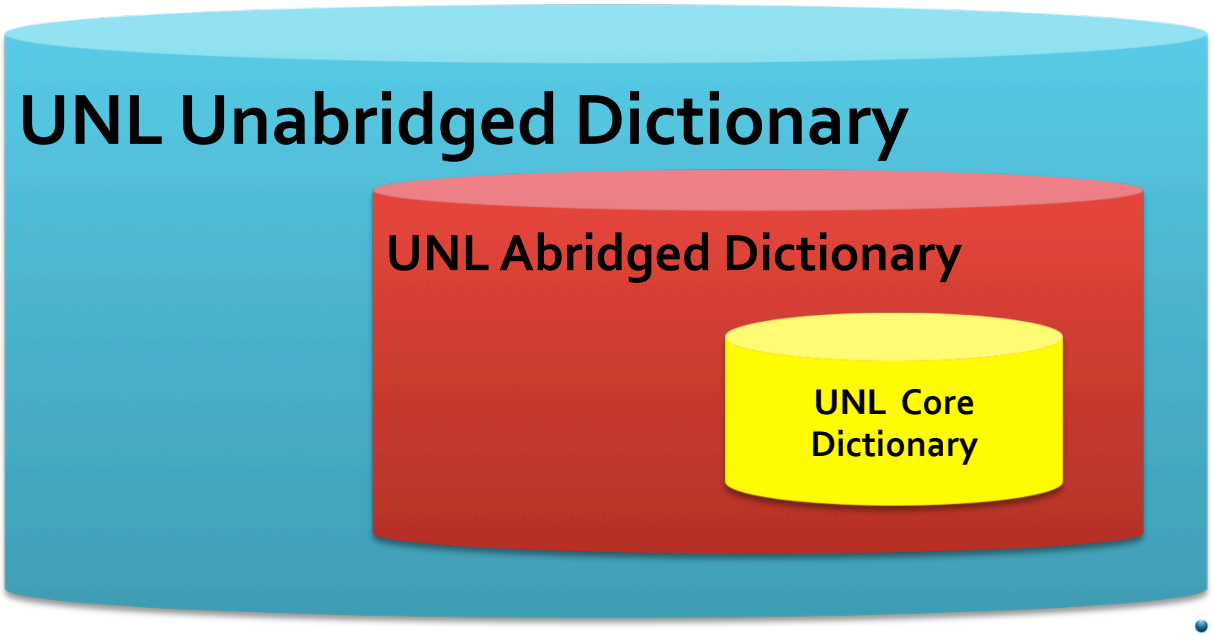


UNDL FOUNDATION

UNL Dictionary (I)

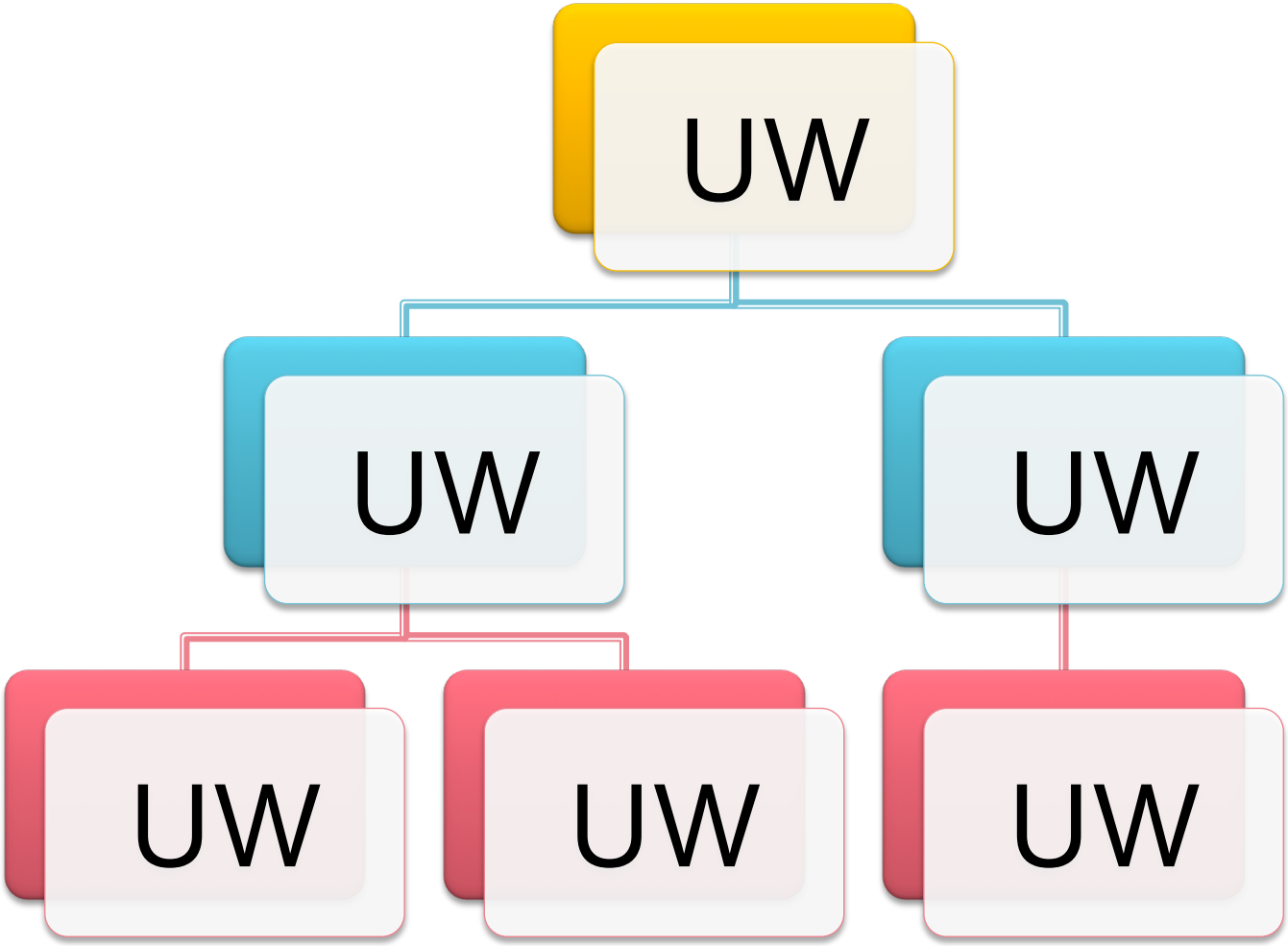


UNL Dictionary (II)



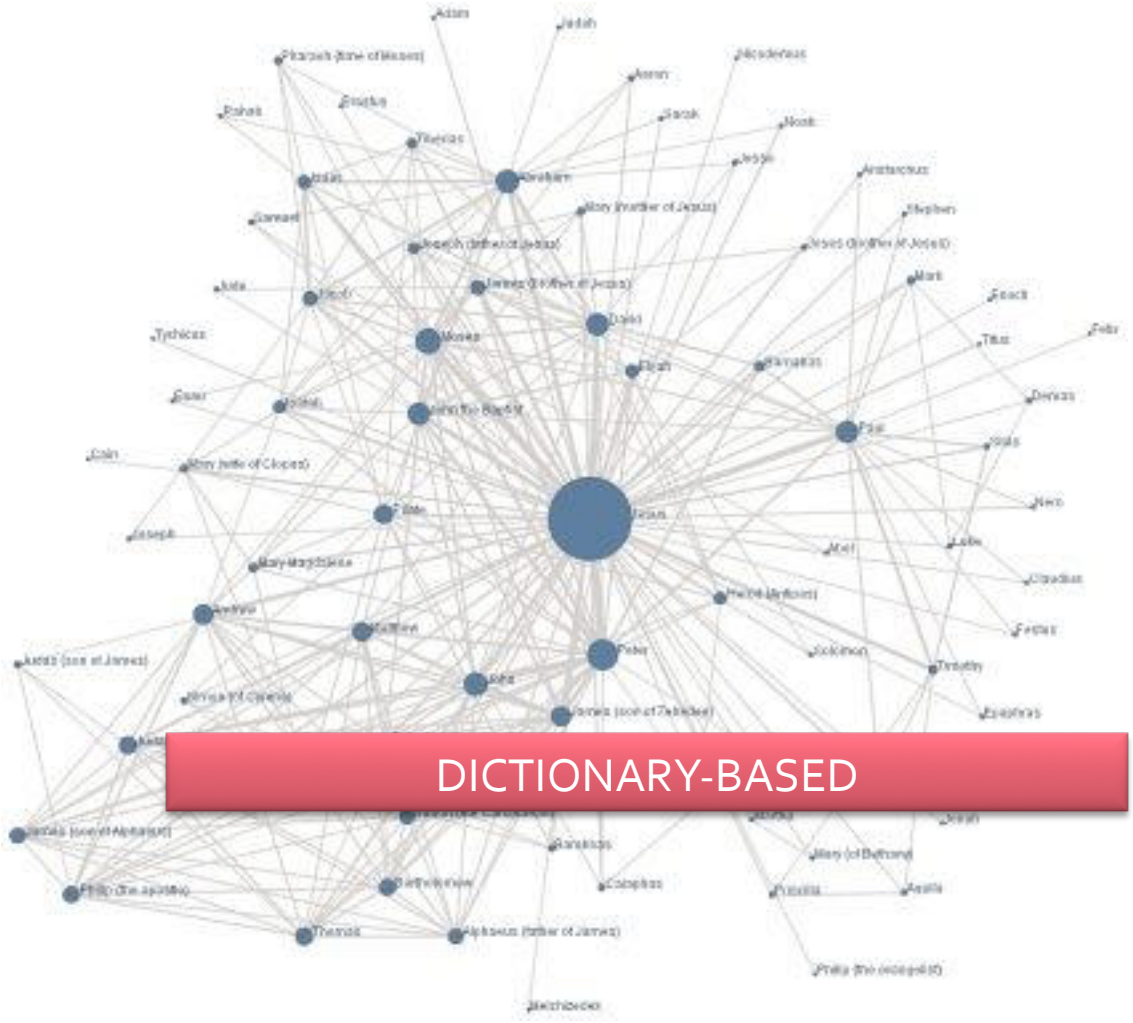
UNDL FOUNDATION

UNL Ontology



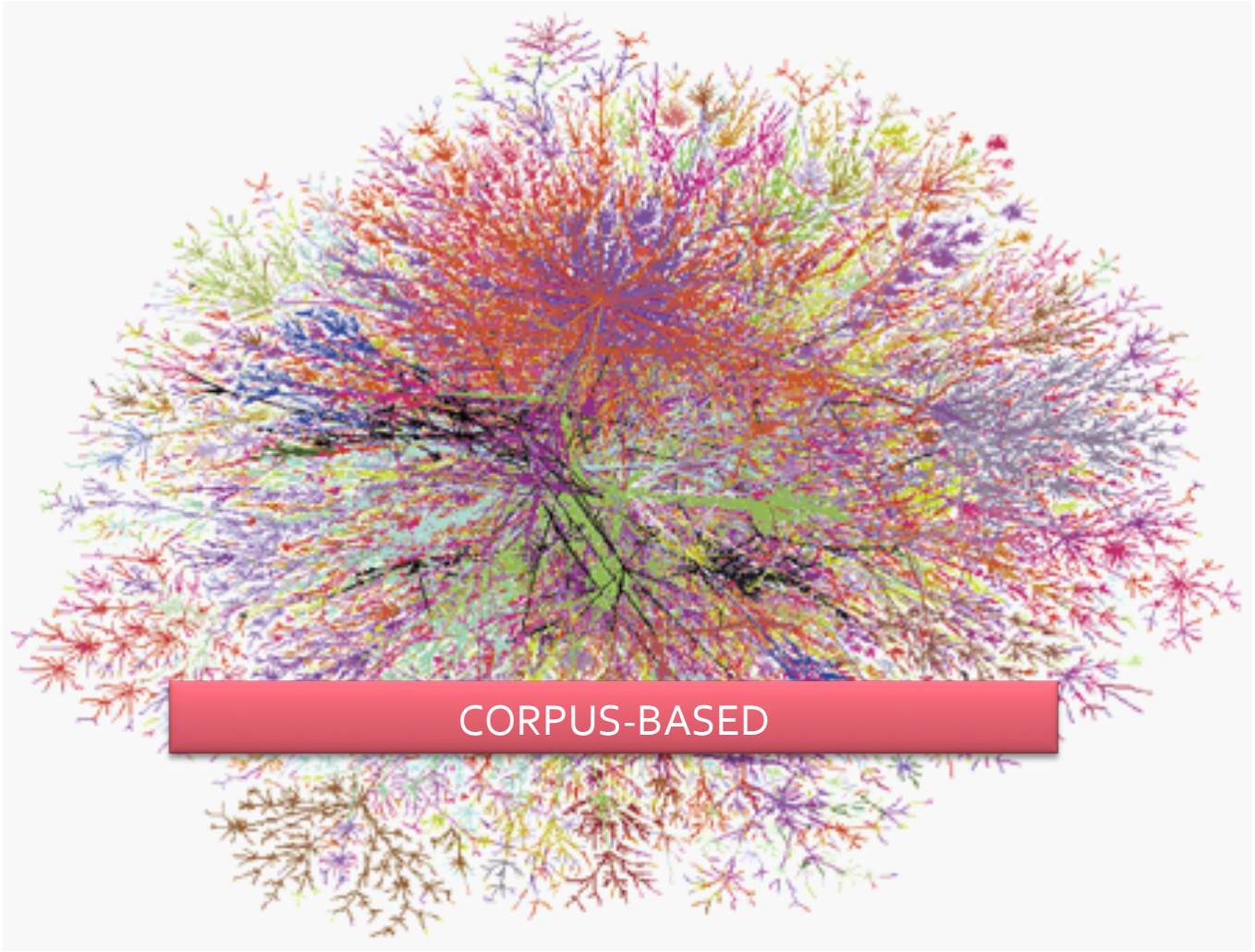
UNDL FOUNDATION

UNL Knowledge Base



UNDL FOUNDATION

UNL Memory

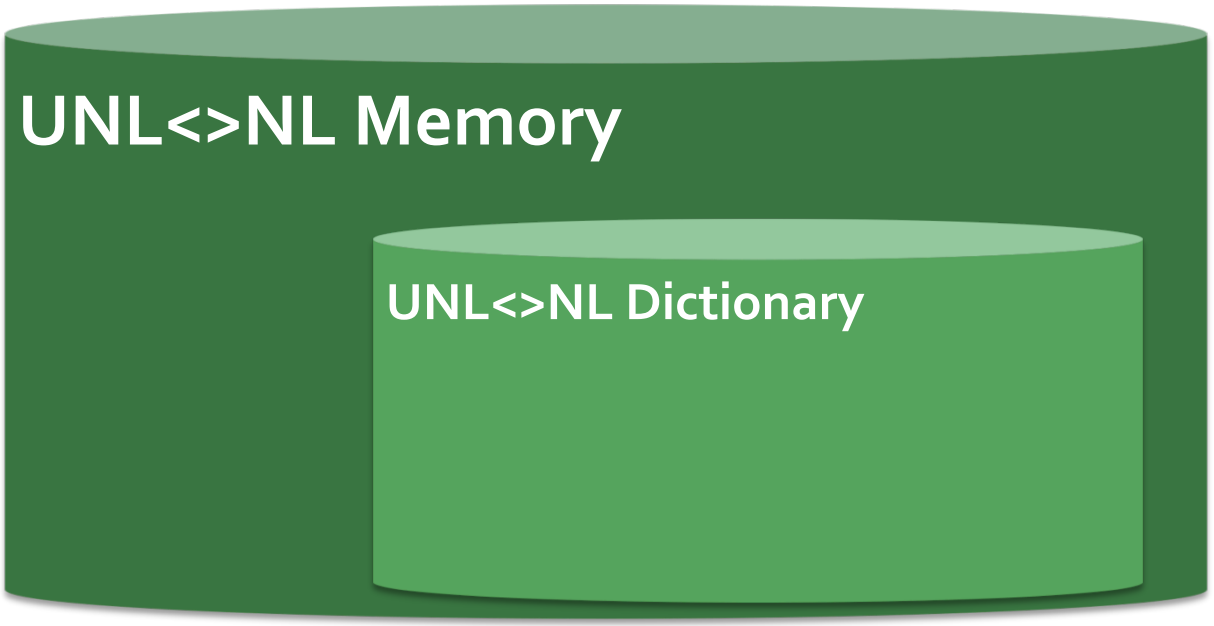


UNDL FOUNDATION

NL Lexical Resources

UNDL FOUNDATION

UNL<>NL Lexical Resources

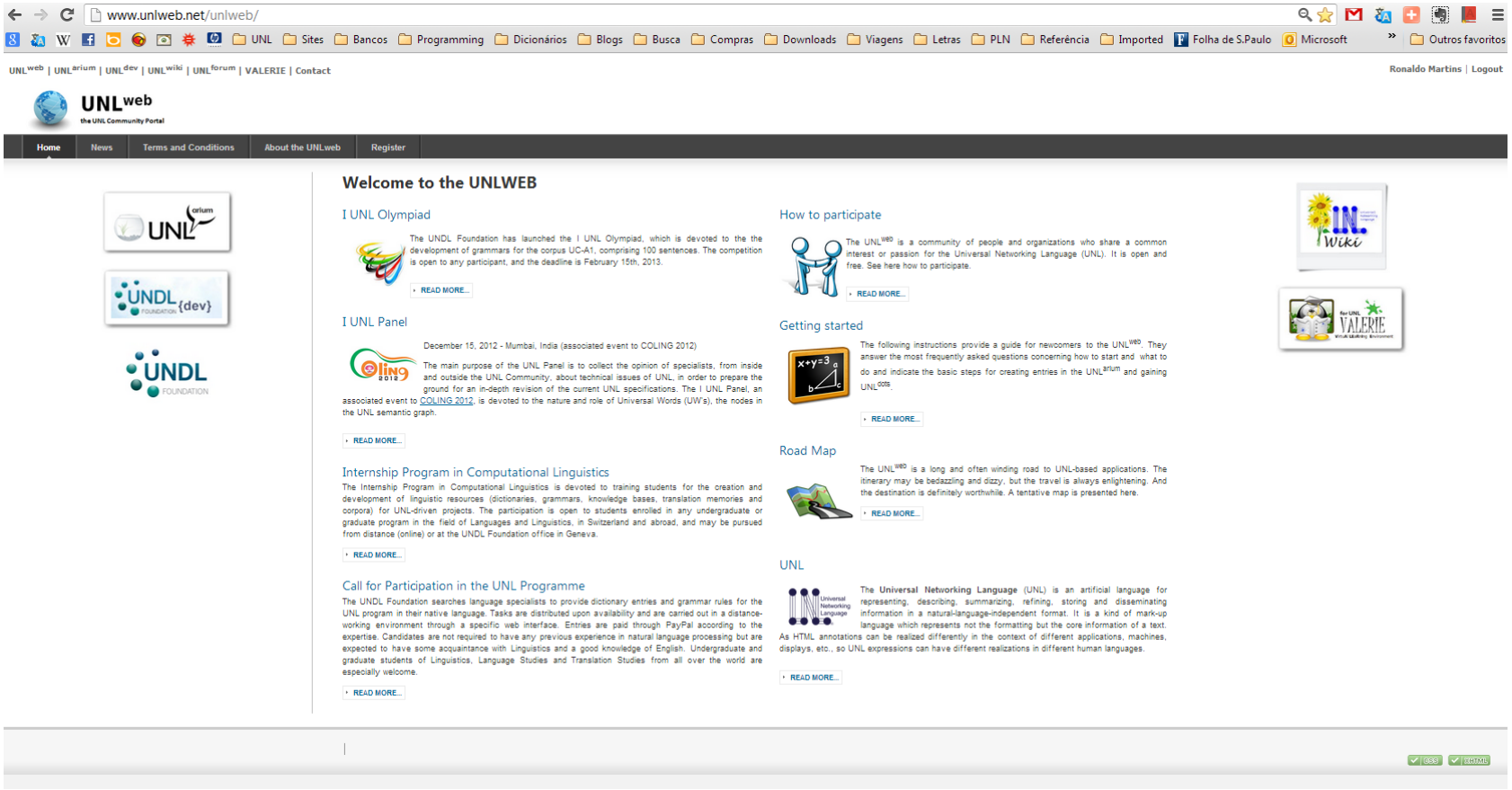


FoR-UNL

LEVEL	UNL-NL DIC	NL-UNL DIC	UNL-NL Grammar	NL-UNL Grammar
A1	MIR-A1	BRUNO-A1	UC-A1	NC-A1
A2	MIR-A2	BRUNO-A2	UC-A2	NC-A2
B1	MIR-B1	BRUNO-B1	UC-B1	NC-B1
B2	MIR-B2	BRUNO-B2	UC-B2	NC-B2
C1	MIR-C1	BRUNO-C1	UC-C1	NC-C1
C2	MIR-C2	BRUNO-C2	UC-C2	NC-C2

UNDL FOUNDATION

www.unlweb.net



FRONT-END APPLICATIONS

keys
information retrieval
and extraction

tut
summarization and
simplification

lily
translation

BACK-END APPLICATIONS

UNLdev

UNL editor

IAN

SEAN

EUGENE

NORMA

EDGES

RESOURCES

UNL arium

TRAINING & RESEARCH

VALERIE

UNLversity

UNL forum

Results (as of Nov, 2012)

CODE	Name	For-UNL					Generative Dictionary (# of entries)		Enumerative Dictionary (# of entries)	
		Overall	MIR	BRUNO	UC	NC	Abridged	Unabridged	Abridged	Unabridged
ar	Arabic	A0	A0	A0	A0	A0	131,053	136,209	3,693,033	3,698,189
hy	Armenian	A0	A0	A0	A0	A0	46,254	46,448	2,066,080	2,066,274
as	Assamese	A0	A0	A0	A0	A0	0	232	0	232
bba	Baatonum	A0	A0	A0	A0	A0	0	1,338	0	1,338
bn	Bengali	A0	A0	A0	A0	A0	0	9,183	0	9,183
bg	Bulgarian	A0	A0	A0	A0	A0	2,671	2,971	34,937	35,237
zh	Chinese	A0	A0	A0	A0	A0	0	37,078	0	37,078
hr	Croatian	A0	A0	A0	A0	A0	3,391	3,602	3,391	3,602
nl	Dutch	A0	A0	A0	A0	A0	0	3,075	0	3,075
en	English	A0	A1	A0	A0	A0	206,510	206,516	539,913	539,919
et	Estonian	A0	A0	A0	A0	A0	2,243	5,326	125,198	128,281
fr	French	A0	A0	A0	A0	A0	68,808	71,609	997,533	1,000,334
de	German	A0	A0	A0	A0	A0	902	60,751	13,292	73,141
el	Greek (Modern)	A0	A0	A0	A0	A0	4,199	4,709	8,740	9,250
gu	Gujarati	A0	A0	A0	A0	A0	0	180	0	180
hi	Hindi	A0	A0	A0	A0	A0	0	67	0	67
hu	Hungarian	A0	A0	A0	A0	A0	6	1,143	6	1,143
id	Indonesian	A0	A0	A0	A0	A0	0	9,626	0	9,626
it	Italian	A0	A0	A0	A0	A0	8,156	8,783	222,120	222,747
ja	Japanese	A0	A0	A0	A0	A0	1	1,871	1	1,871
la	Latin	A0	A0	A0	A0	A0	18,179	56,332	959,032	997,185
fa	Persian	A0	A0	A0	A0	A0	0	1,815	0	1,815
pl	Polish	A0	A0	A0	A0	A0	0	125	0	125
pt	Portuguese	A0	A0	A0	A0	A0	39,580	52,967	298,159	311,556
ro	Romanian	A0	A0	A0	A0	A0	92	126	1,090	1,124
ru	Russian	A0	A0	A0	A0	A0	56,465	68,323	1,600,581	1,612,452
sr	Serbian	A0	A0	A0	A0	A0	1,918	4,202	85,026	87,353
sk	Slovak	A0	A0	A0	A0	A0	1,219	3,452	1,219	3,452
sl	Slovenian	A0	A0	A0	A0	A0	6,344	11,380	256,428	261,616
es	Spanish	A0	A0	A0	A0	A0	75,900	80,129	1,388,368	1,392,597
sw	Swahili	A0	A0	A0	A0	A0	0	2,287	0	2,287
sv	Swedish	A0	A0	A0	A0	A0	126	252	539	2,156
ta	Tamil	A0	A0	A0	A0	A0	5	3,477	5	3,477
te	Telugu	A0	A0	A0	A0	A0	0	113	0	113
th	Thai	A0	A0	A0	A0	A0	0	719	0	719
tr	Turkish	A0	A0	A0	A0	A0	2	4,862	2	4,862
uk	Ukrainian	A0	A0	A0	A0	A0	3,042	6,245	59,498	62,889

UNDL FOUNDATION

Issues

1. Concept of UW
2. Named entities
3. Lexical gaps
4. Antonyms
5. Multiword expressions

UNDL FOUNDATION

1. What is to be considered a UW?

How many UW's should be recognized in the sentence below?

« Charles Dickens is generally regarded as the most important English novelist of the Victorian period »?

current answer

UW's must correspond to and only to semantic discrete units conveyed by natural language open lexical categories (noun, verb, adjective and adverb). Any other semantic content (such as the ones conveyed by articles, prepositions, conjunctions etc.) should be represented as attributes or relations.

UNDL FOUNDATION

2. Which UW's should be included in the UNL dictionary?

« Charles Dickens » should be represented as a temporary UW or as as permanent UW?

current answer:

If a proper name is defined as an entry in the Wikipedia, then it should be defined as a permanent UW and included in the UNL Unabridged Dictionary.

UNDL FOUNDATION

3. How to represent derived concepts?

"hunger" (= "a physiological need for food")
"hungry" (= "feeling hunger")
"hungrily" (= "in the manner of someone who is very hungry")
"hunger" (= "to cause to experience hunger")
should be represented as simple, compound or complex UW's?

current answer:

UNDL FOUNDATION

4. How to represent antonyms?

Antonyms such as "mortal" and "immortal", "hot" and "cold", and "son" and "father" should be represented as a single UW (and the corresponding attributes) or as different UW's?

current answer:

UNDL FOUNDATION

5. When a concept can be said compositional or not?

"Farbfernsehgerät" ("color television set", in German) should be represented as a simple or complex UW?

current answer:

UNDL FOUNDATION

Goals

- To listen
- To critize the answers that have been currently given by the UNDL Foundation to these questions (but keeping in mind the commitments and properties of the UNL)
- To prepare the ground for a careful and well-advised in-depth revision of the current specifications

UNDL FOUNDATION

Program

09:30 – Opening
10:00 – Introduction
10:30 – Tea break
11:00 – First session

- Nicoletta Calzolari (Istituto de Linguistica Computazionale Antonio Zampolli, Pisa)
- Mike Dillinger (eBay)
- Eric Wehrli (Université de Genève)

13:00 – Lunch break
14:30 – Second session

- Igor Boguslavsky (Universidad Politecnica de Madrid, Russian Academy of Sciences)
- Sameh Alansary (University of Alexandria, Library of Alexandria)
- Pushpak Bhattacharyya (IIT-Bombay)

16:30 – Tea break
17:00 – Discussion session
18:00 – Closing

UNDL FOUNDATION

Discussion

Closing

www.unlweb.net

r.martins@undlfoundation.org

- UNL OLYMPIAD, Feb 15
- XI UNL School, Mar 4-8, 2013, Macau (Languages of Southeast Asia)

CC BY SA