

GENERAREA UNEI ONTOLOGII PENTRU LIMBA ROMÂNĂ

PORNIND DE LA DEXONLINE

SCOICĂ ADRIAN, GRUPA 322CC

CUPRINS

- SCOPUL PROIECTULUI
- ARHITECTURA SOLUȚIEI
- FACTORI DE LIMITARE
- DEZVOLTARE VIITOARE

SCOPUL PROIECTULUI

GENERAREA UNEI ONTOLOGII LEXICALE PENTRU
LIMBA ROMÂNĂ

POSIBILE APLICAȚII:

SCOPUL PROIECTULUI

GENERAREA UNEI ONTOLOGII LEXICALE PENTRU
LIMBA ROMÂNĂ

POSIBILE APLICAȚII:
– LANȚURI LEXICALE

SCOPUL PROIECTULUI

GENERAREA UNEI ONTOLOGII LEXICALE PENTRU LIMBA ROMÂNĂ

POSIBILE APLICAȚII:

- LANȚURI LEXICALE
- DELIMITAREA CONTEXTELOR

SCOPUL PROIECTULUI

GENERAREA UNEI ONTOLOGII LEXICALE PENTRU LIMBA ROMÂNĂ

POSIBILE APLICAȚII:

- LANȚURI LEXICALE
- DELIMITAREA CONTEXTELOR
- DEZAMBIGUIZĂRI

SCOPUL PROIECTULUI

GENERAREA UNEI ONTOLOGII LEXICALE PENTRU LIMBA ROMÂNĂ

POSIBILE APLICAȚII:

- LANȚURI LEXICALE
- DELIMITAREA CONTEXTELOR
- DEZAMBIGUIZĂRI
- GENERAL KR

CERINȚE SPECIFICE

NE DORIM:

CERINȚE SPECIFICE

NE DORIM:

- MĂSURAREA APROPIERII

CERINȚE SPECIFICE

NE DORIM:

- MĂSURAREA APROPIERII
- CLASIFICĂRI (HOLONIMIE, MERONIMIE, SINONIMIE, etc.)

CERINȚE SPECIFICE

NE DORIM:

- MĂSURAREA APROPIERII
- CLASIFICĂRI (HOLONIMIE, MERONIMIE, SINONIMIE, etc.)
- LANGUAGE INDEPENDENCE (CONCEPT - ORIENTED)

CERINȚE SPECIFICE

NE DORIM:

- MĂSURAREA APROPIERII
- CLASIFICĂRI (HOLONIMIE, MERONIMIE, SINONIMIE, etc.)
- LANGUAGE INDEPENDENCE (CONCEPT - ORIENTED)
- EXTINDERE AUTOMATĂ

ARHITECTURA SOLUȚIEI

Tipul soluției: *MultiNet*

ARHITECTURA SOLUȚIEI

Tipul soluției: *MultiNet*

– Propusă inițial de Prof. Dr. Hermann Helbig

ARHITECTURA SOLUȚIEI

Tipul soluției: *MultiNet*

- Propusă inițial de Prof. Dr. Hermann Helbig
- Structură complexă, dar omogenă

ARHITECTURA SOLUȚIEI

Tipul soluției: *MultiNet*

- Propusă inițial de Prof. Dr. Hermann Helbig
- Structură complexă, dar omogenă
- Nu impune metrice

ARHITECTURA SOLUȚIEI

Tipul soluției: *MultiNet*

- Propusă inițial de Prof. Dr. Hermann Helbig
- Structură complexă, dar omogenă
- Nu impune metrici
- Modelare conceptuală

ARHITECTURA SOLUȚIEI

Tipul soluției: *MultiNet*

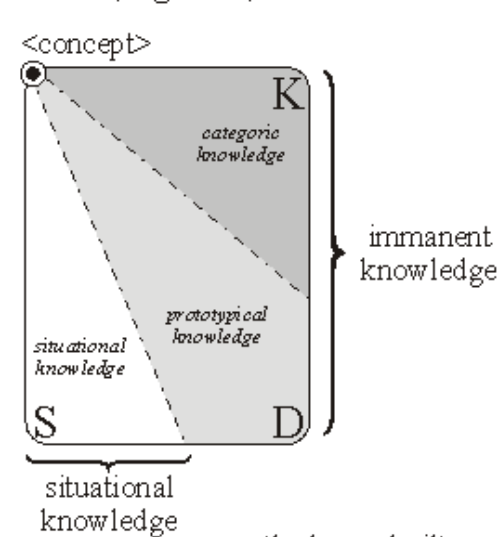
- Propusă inițial de Prof. Dr. Hermann Helbig
- Structură complexă, dar omogenă
- Nu impune metrici
- Modelare conceptuală
- Greu generabilă automat

ARHITECTURA SOLUȚIEI

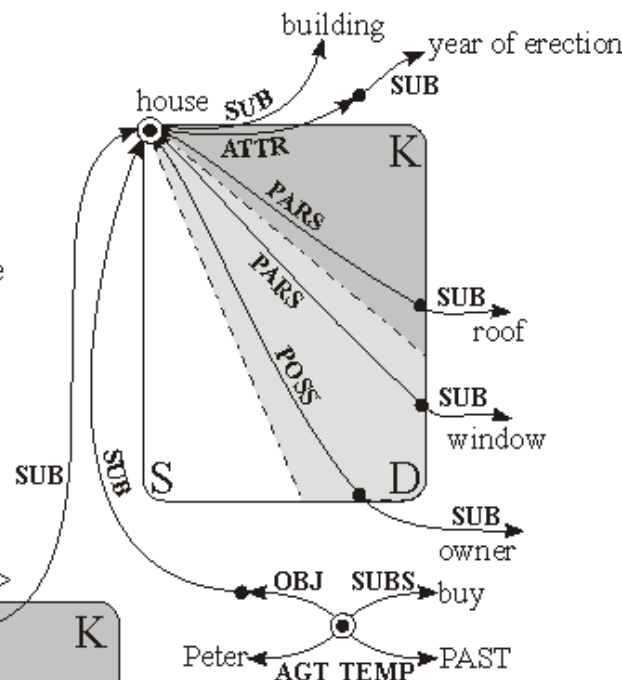
Tipul soluției: *MultiNet*

- Propusă inițial de Prof. Dr. Hermann Helbig
- Structură complexă, dar omogenă
- Nu impune metrici
- Modelare conceptuală
- Greu generabilă automat
- Nu are specificații operaționale

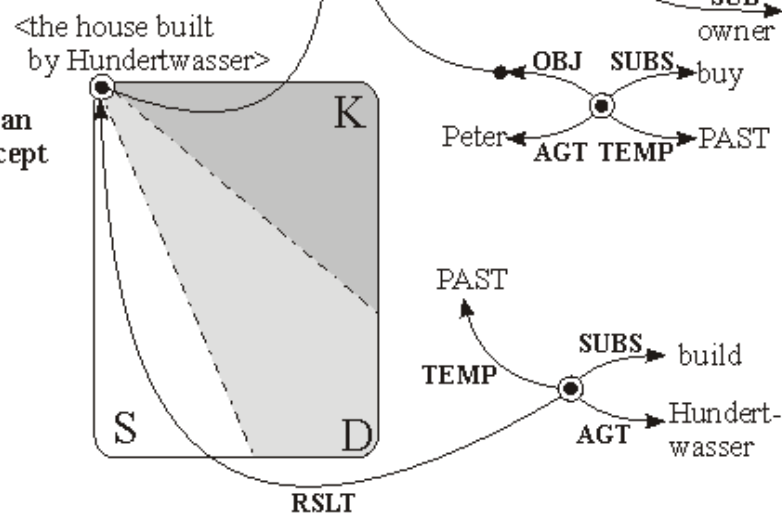
a) Components of a concept (in general)



b) Example for a generic concept



c) Example for an individual concept



EXEMPLE MULTINET

<casa construită de Hundertwasser>

<Peter a fost lovit de mașină pentru că a trecut strada.>

K immanent (categorical):

A house is a building; it always has a roof and is characterized by its year of erection.

D immanent (prototypical):

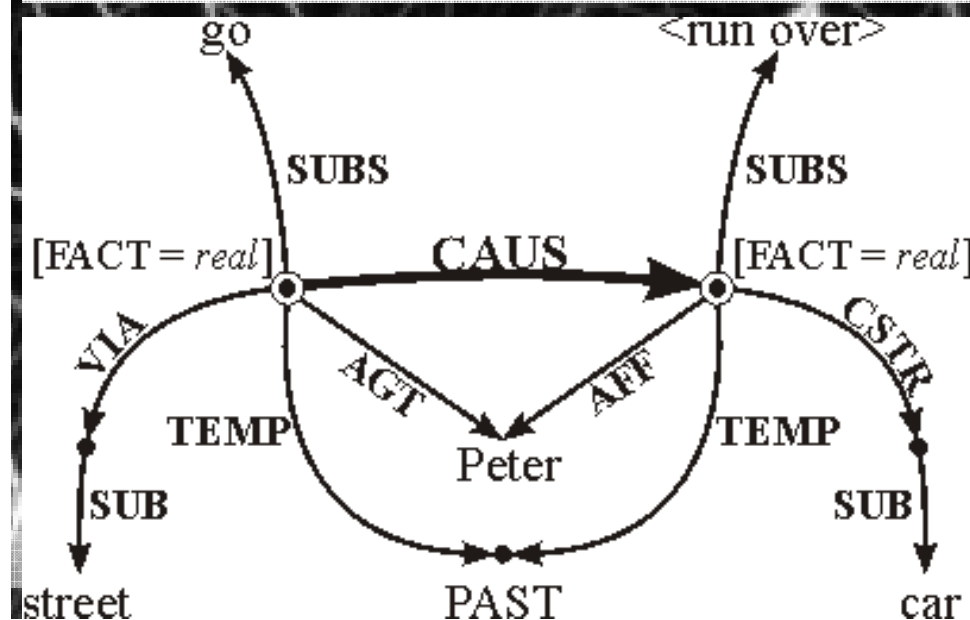
A house (usually) has windows and (in general) an owner.

S situational & assertional:

Peter has bought a house.
[OBJ+SUB outside upper capsule b)]

situational & definitional:

The house built by Hundertwasser
[RSLT inside lower capsule c)]



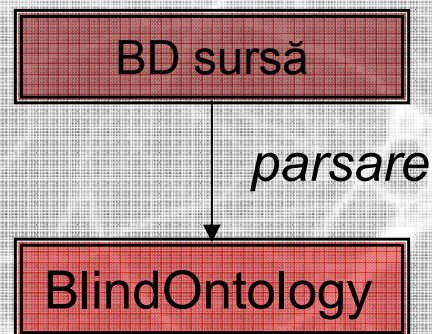
ARHITECTURA SOLUȚIEI

Abordare: *construcție incrementală*

BD sursă

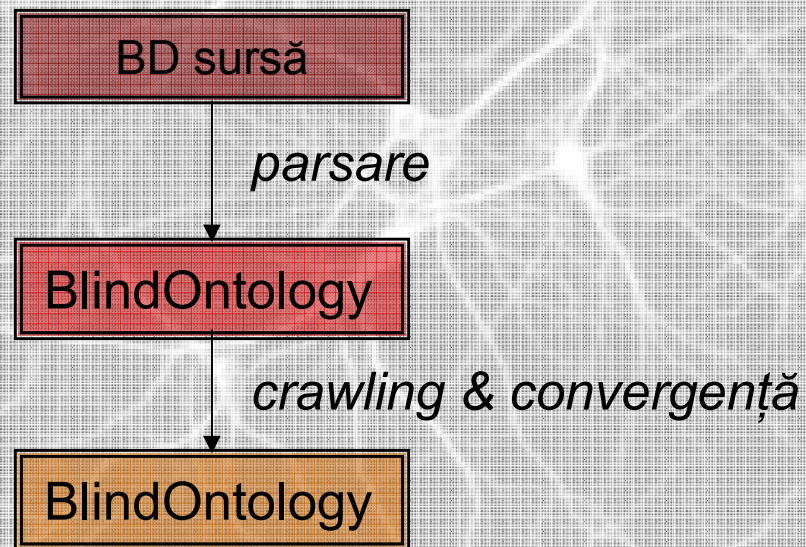
ARHITECTURA SOLUȚIEI

Abordare: *construcție incrementală*



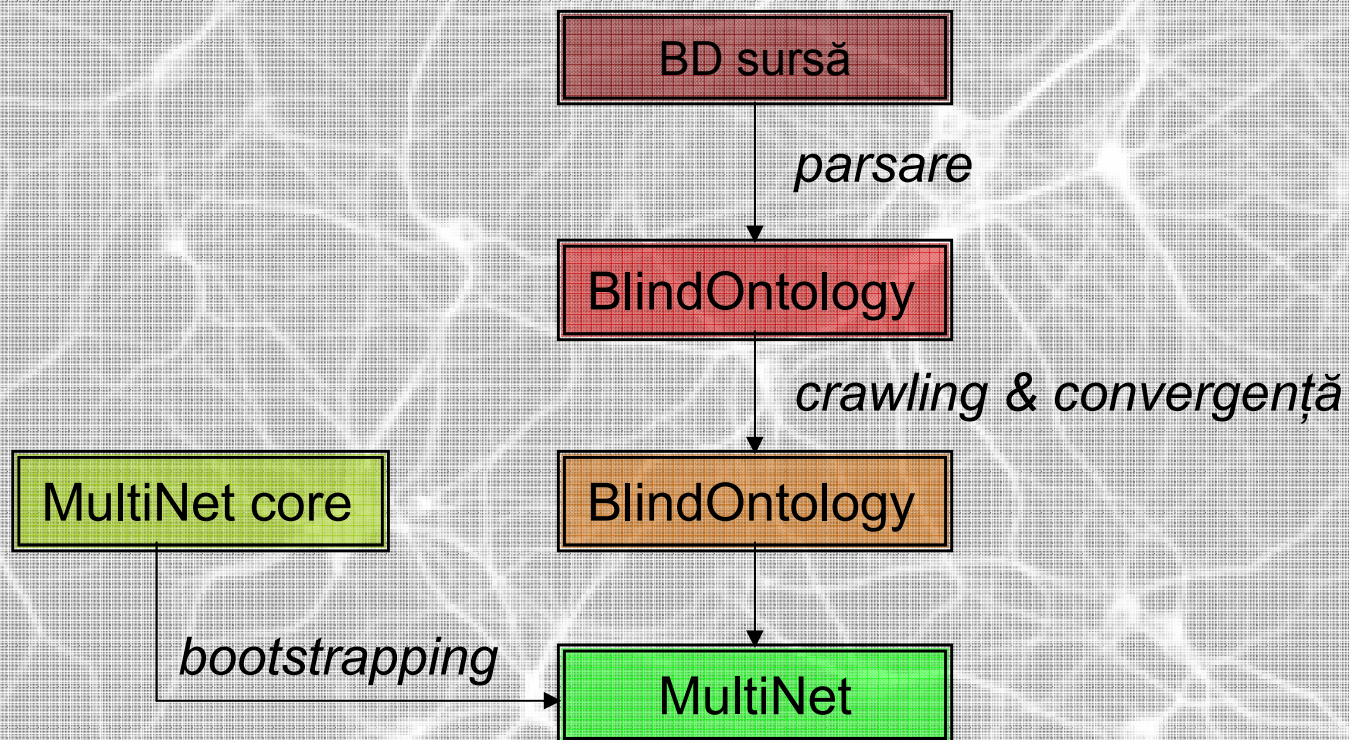
ARHITECTURA SOLUȚIEI

Abordare: *construcție incrementală*



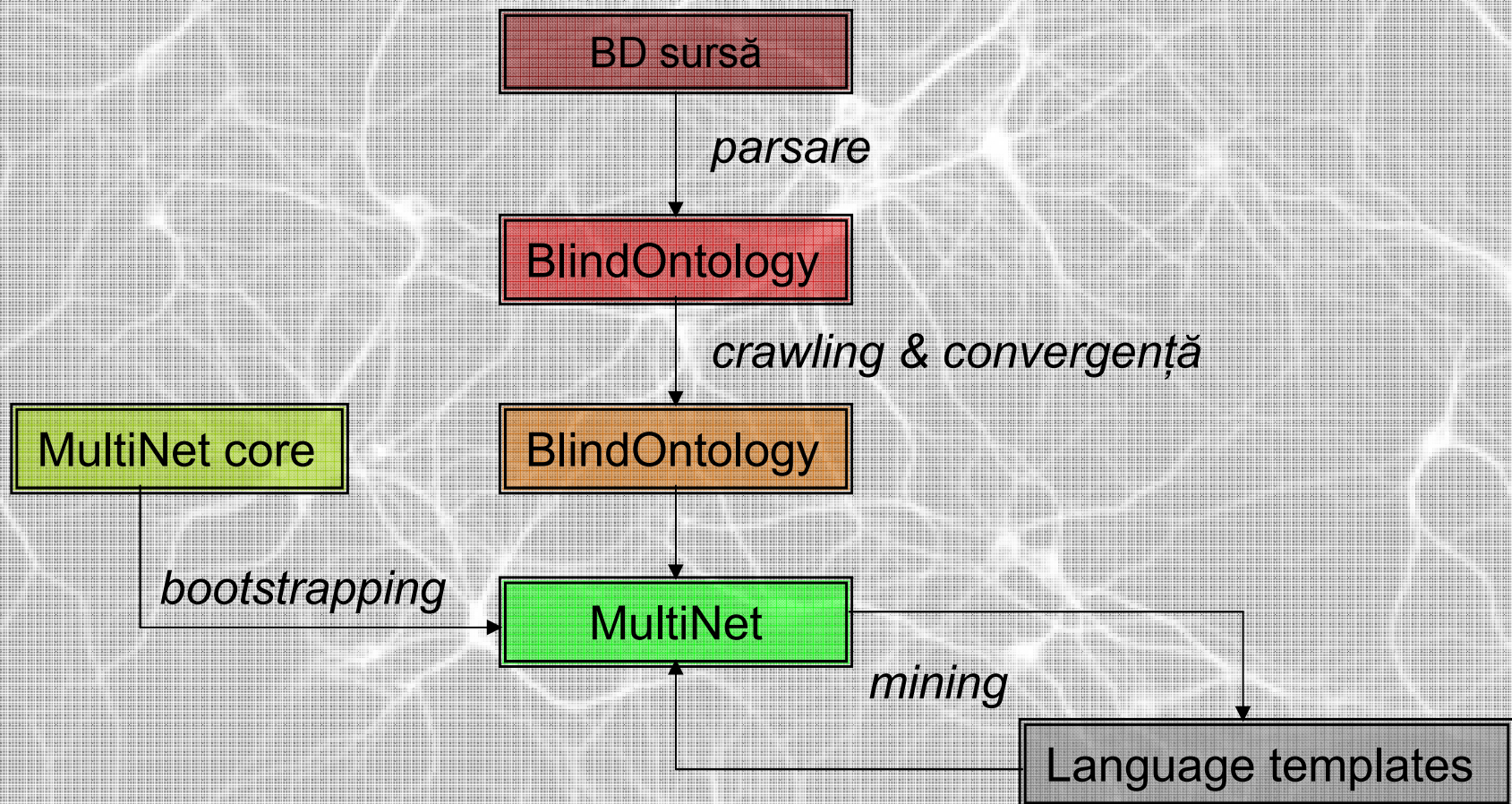
ARHITECTURA SOLUȚIEI

Abordare: *construcție incrementală*



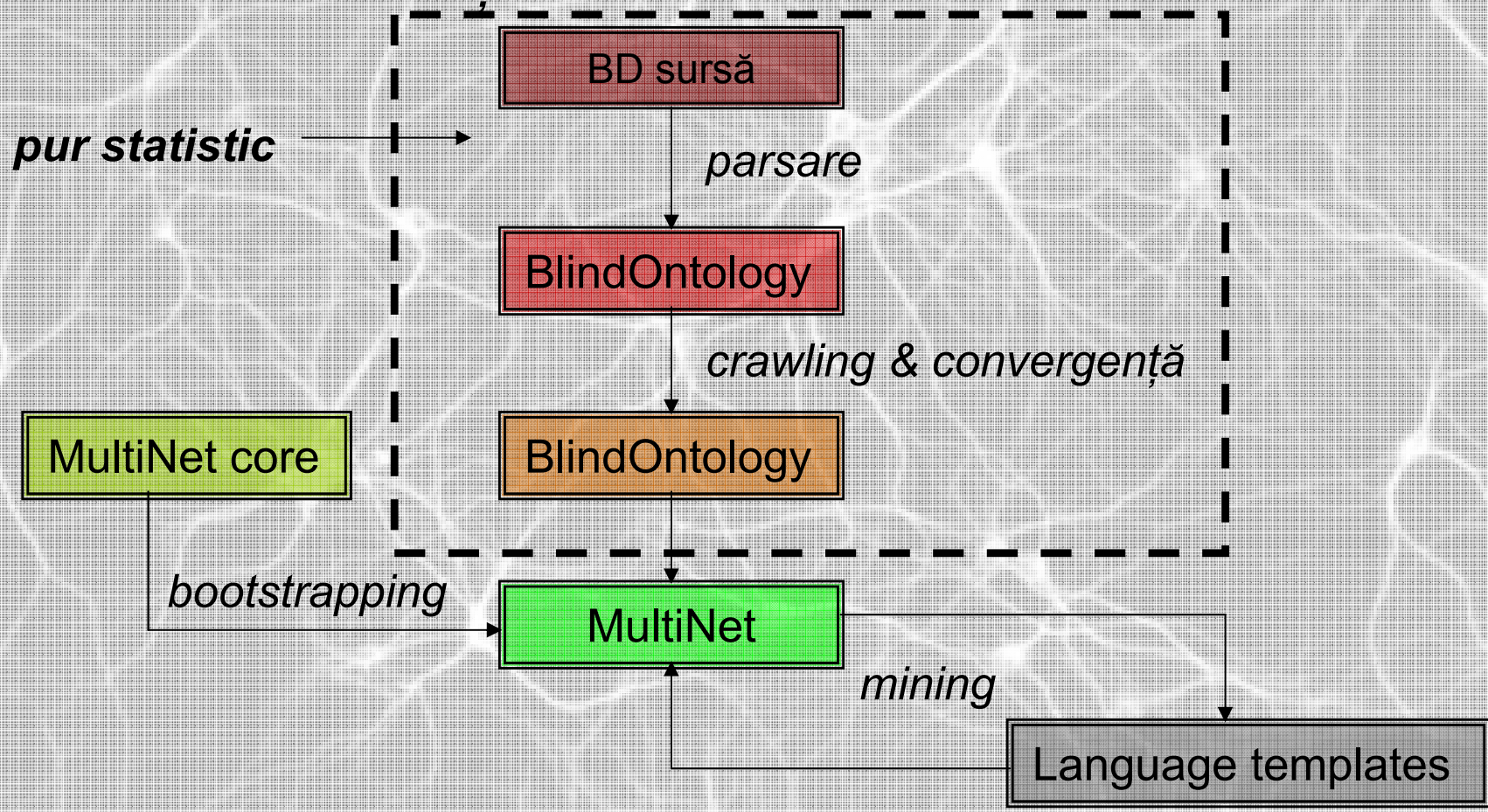
ARHITECTURA SOLUȚIEI

Abordare: *construcție incrementală*



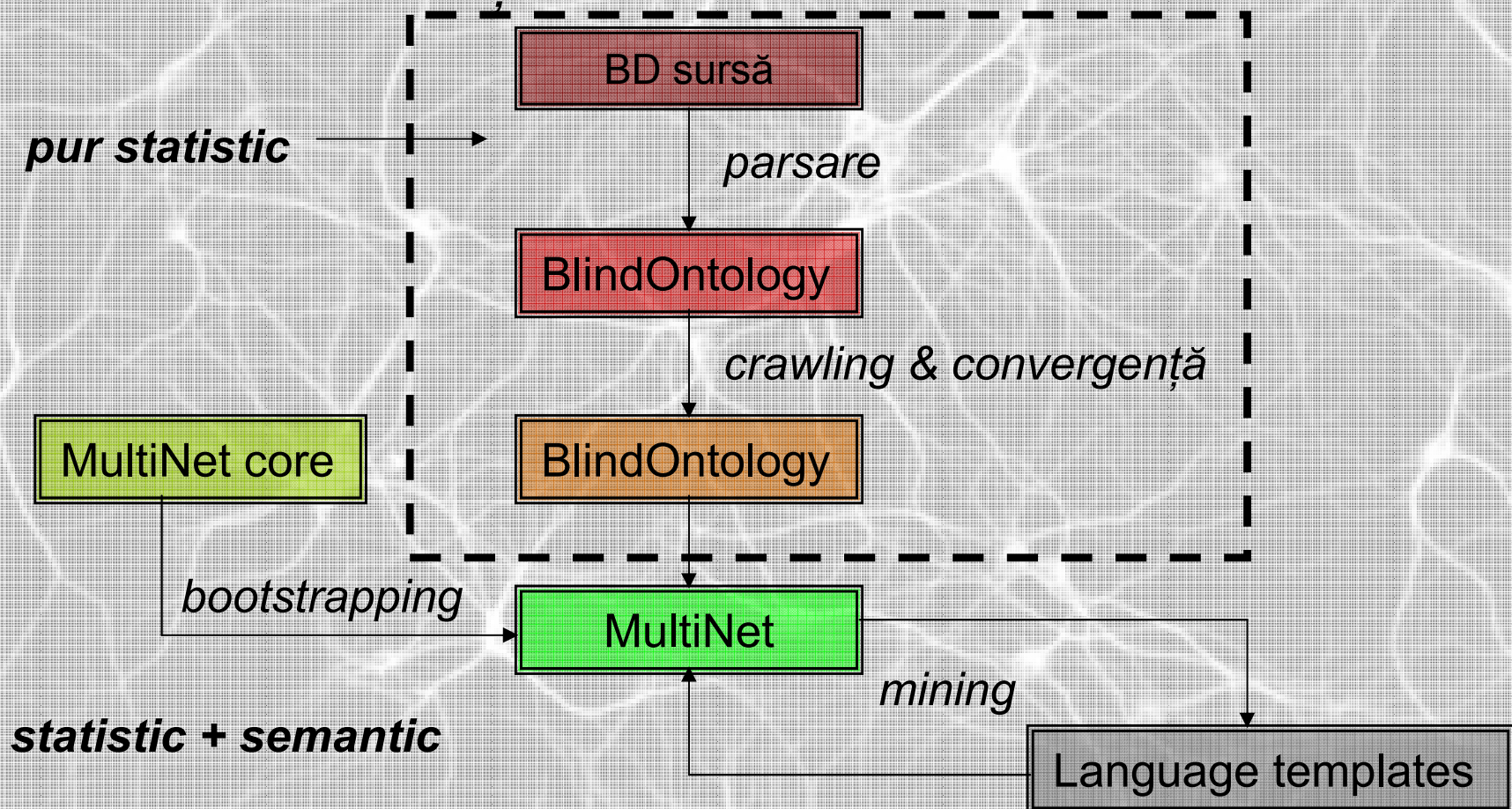
ARHITECTURA SOLUȚIEI

Abordare: *construcție incrementală*



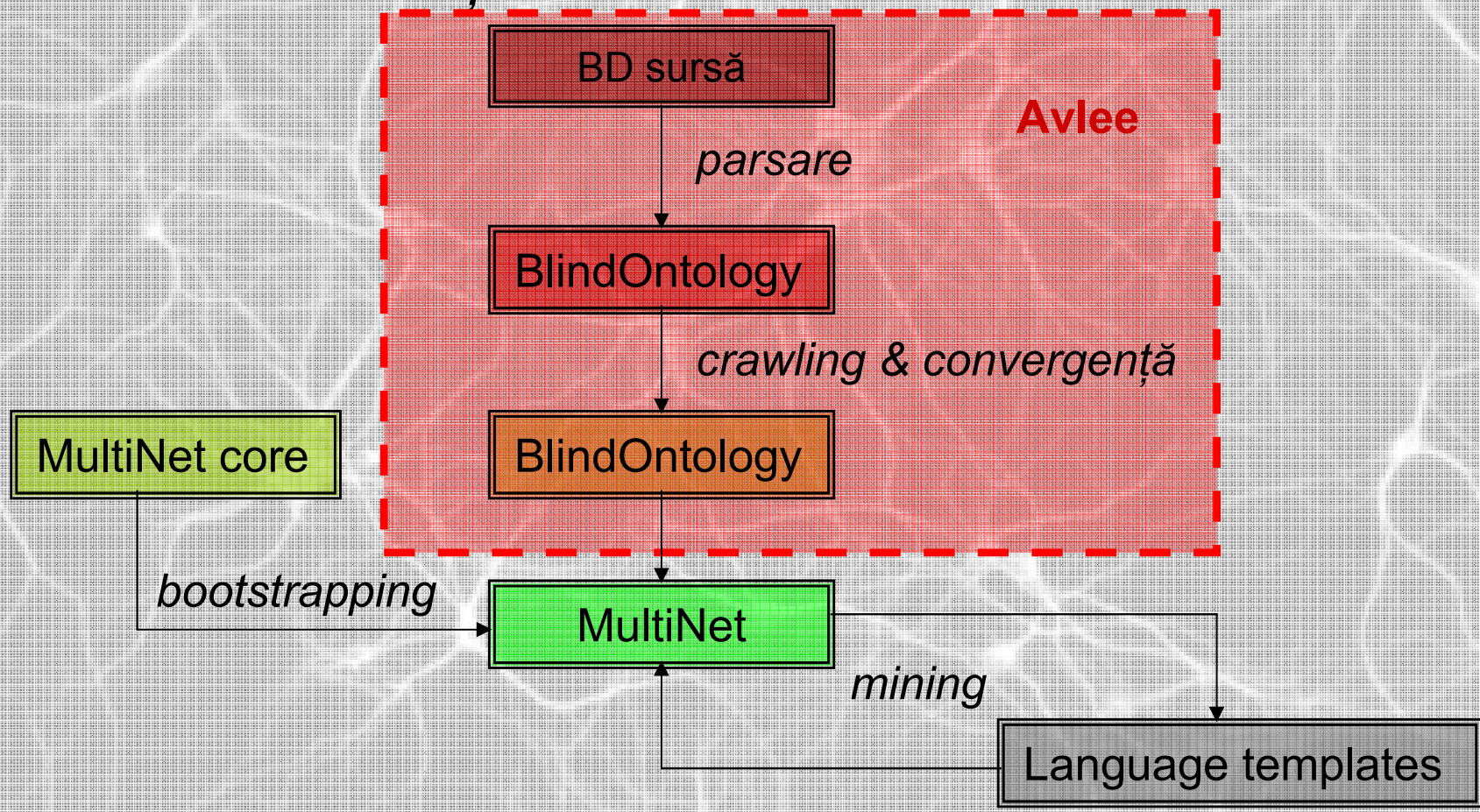
ARHITECTURA SOLUȚIEI

Abordare: *construcție incrementală*



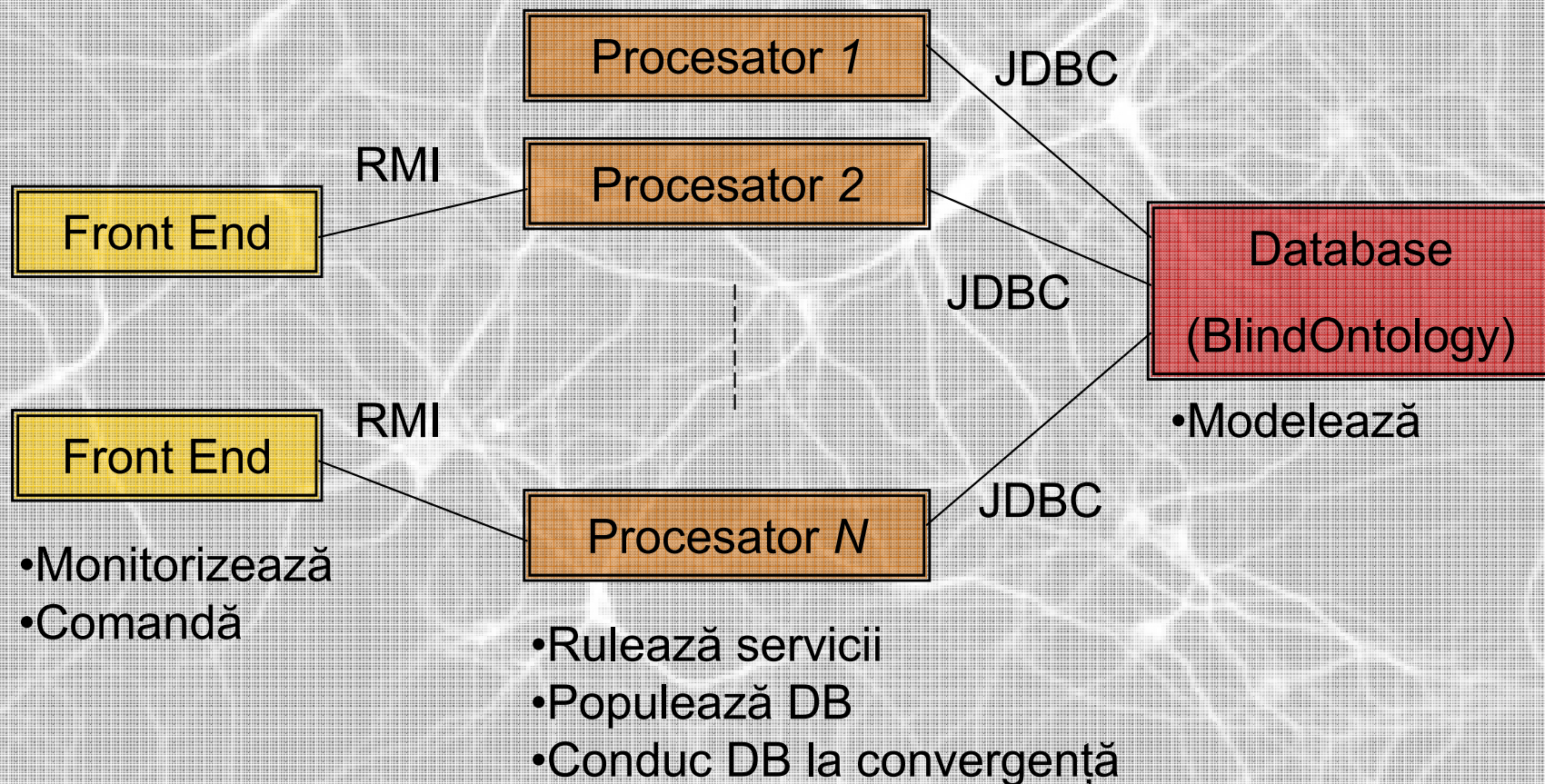
ARHITECTURA SOLUȚIEI

Abordare: *construcție incrementală*

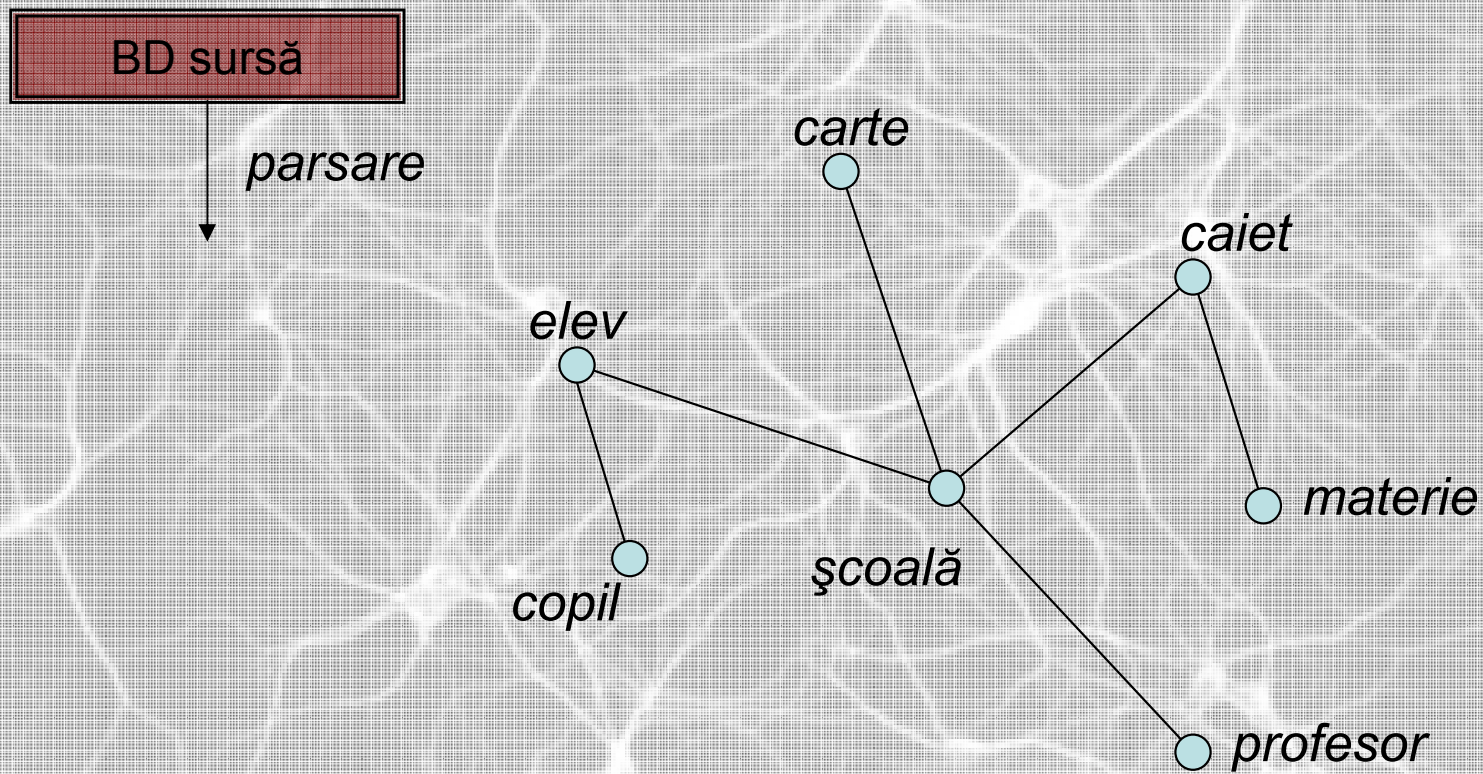


ARHITECTURA SOLUȚIEI

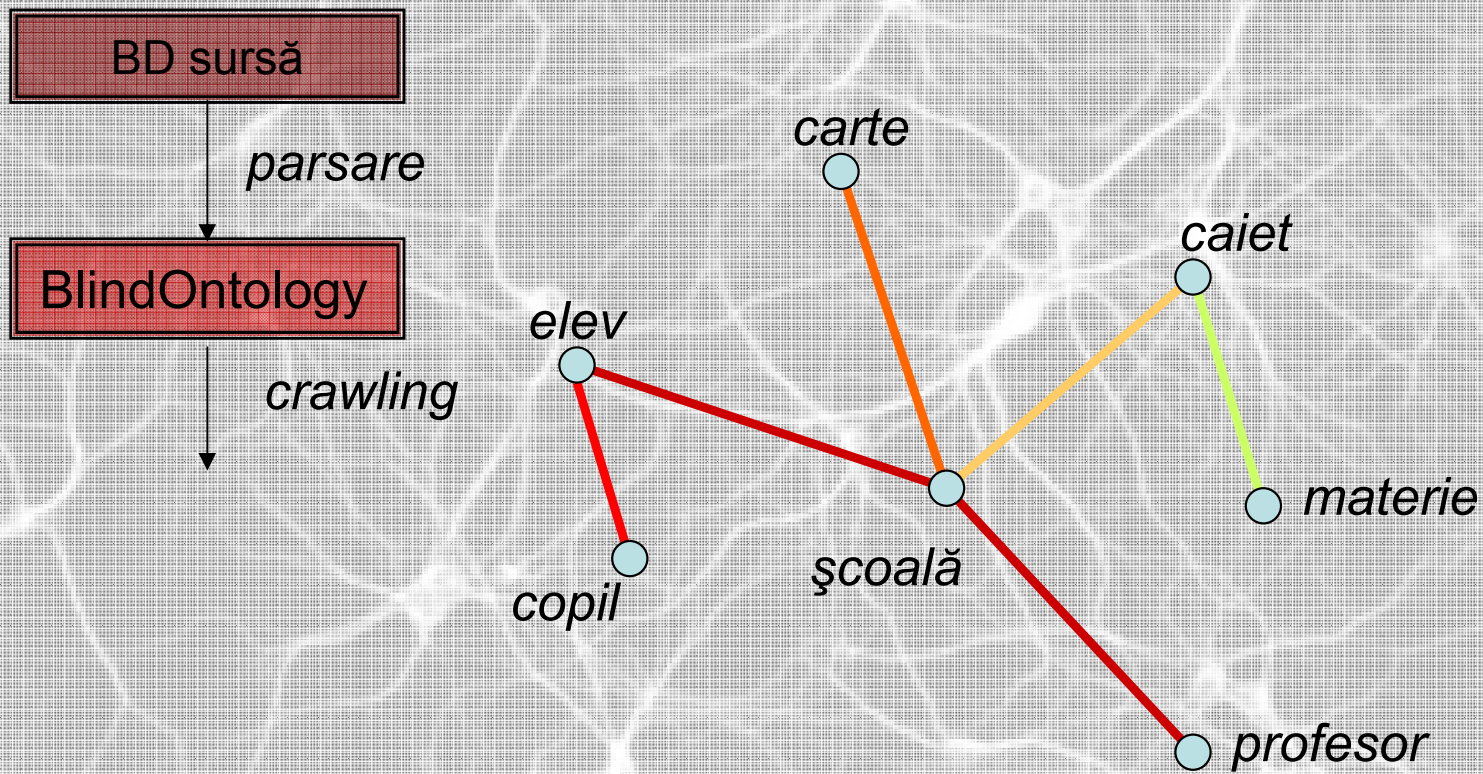
Avlee: organizarea aplicației



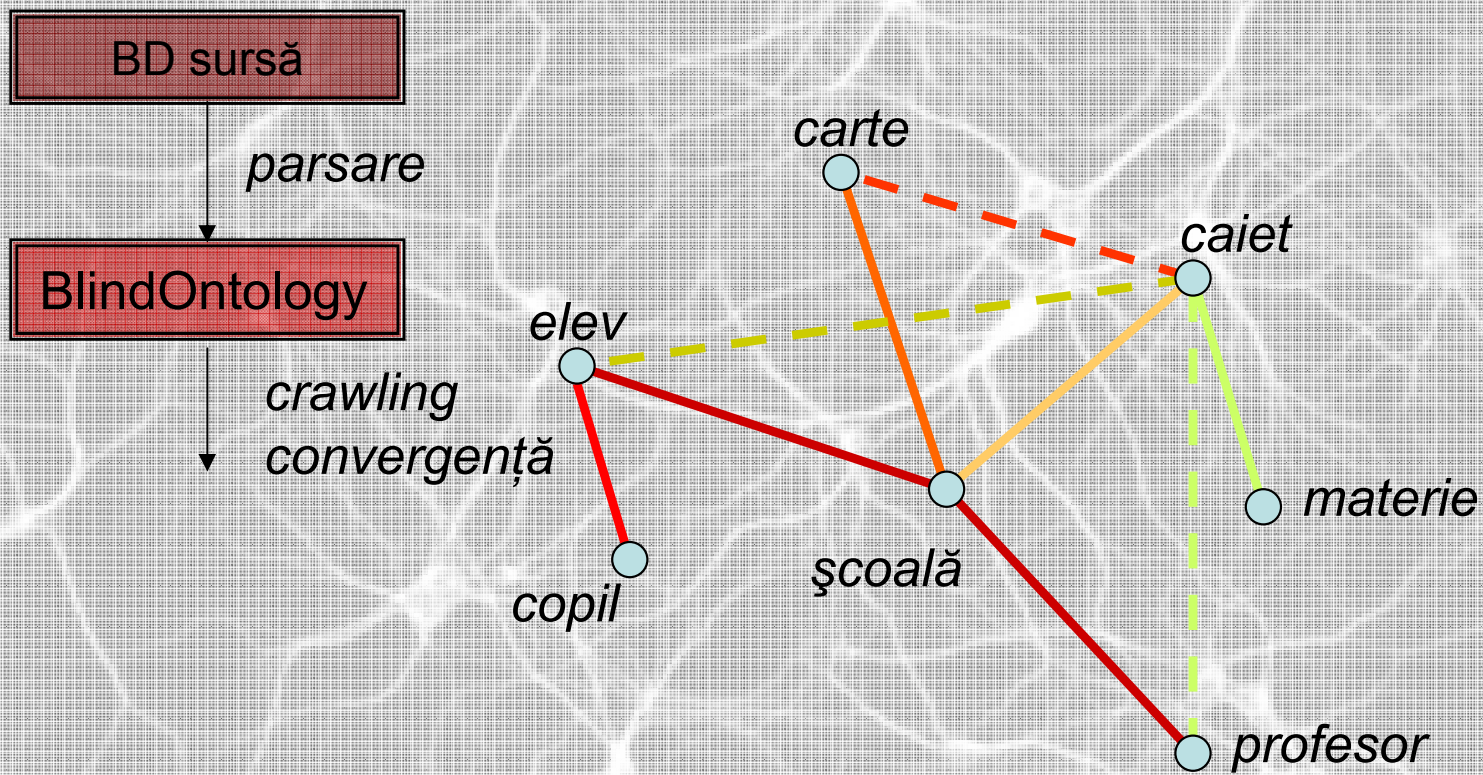
EXEMPLU FUNCȚIONARE



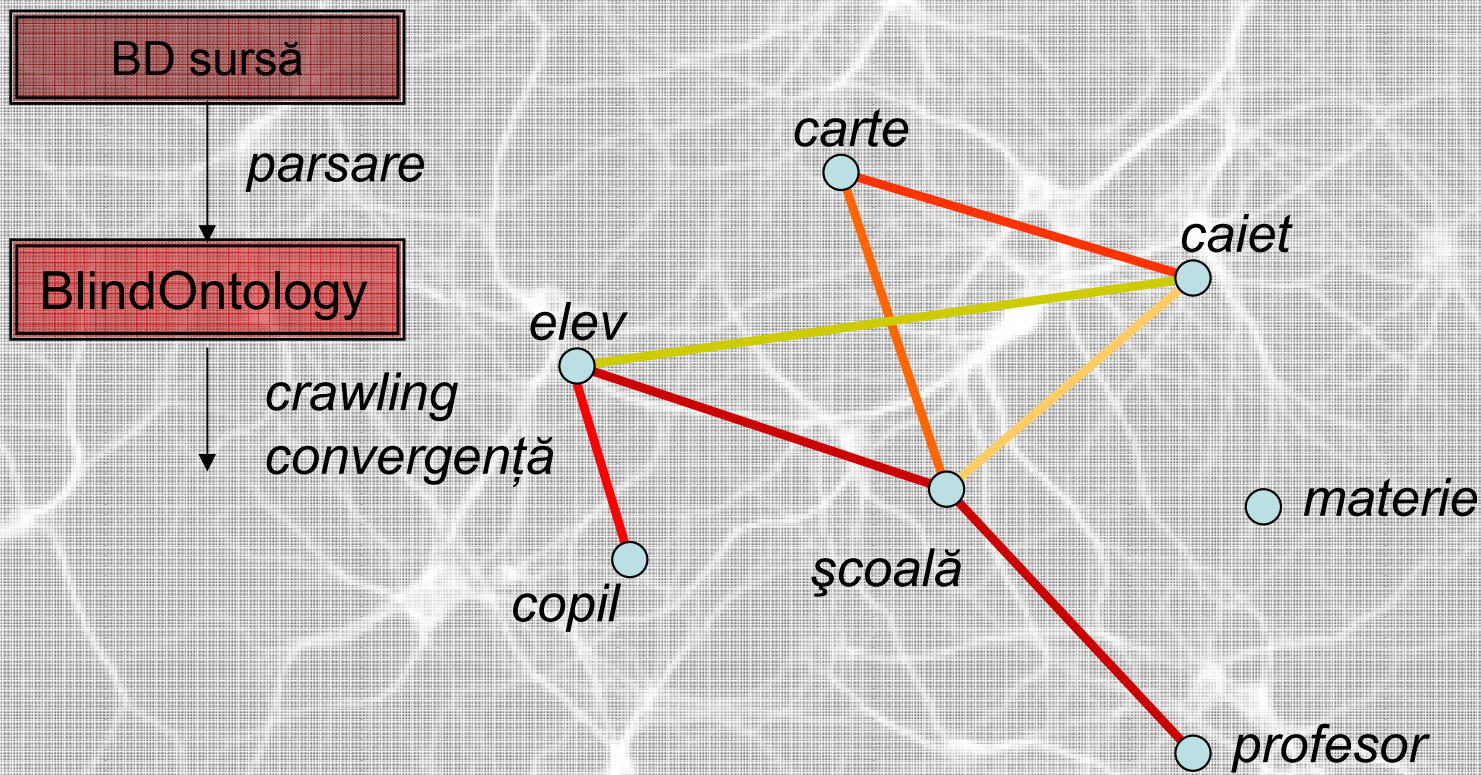
EXEMPLU FUNCȚIONARE



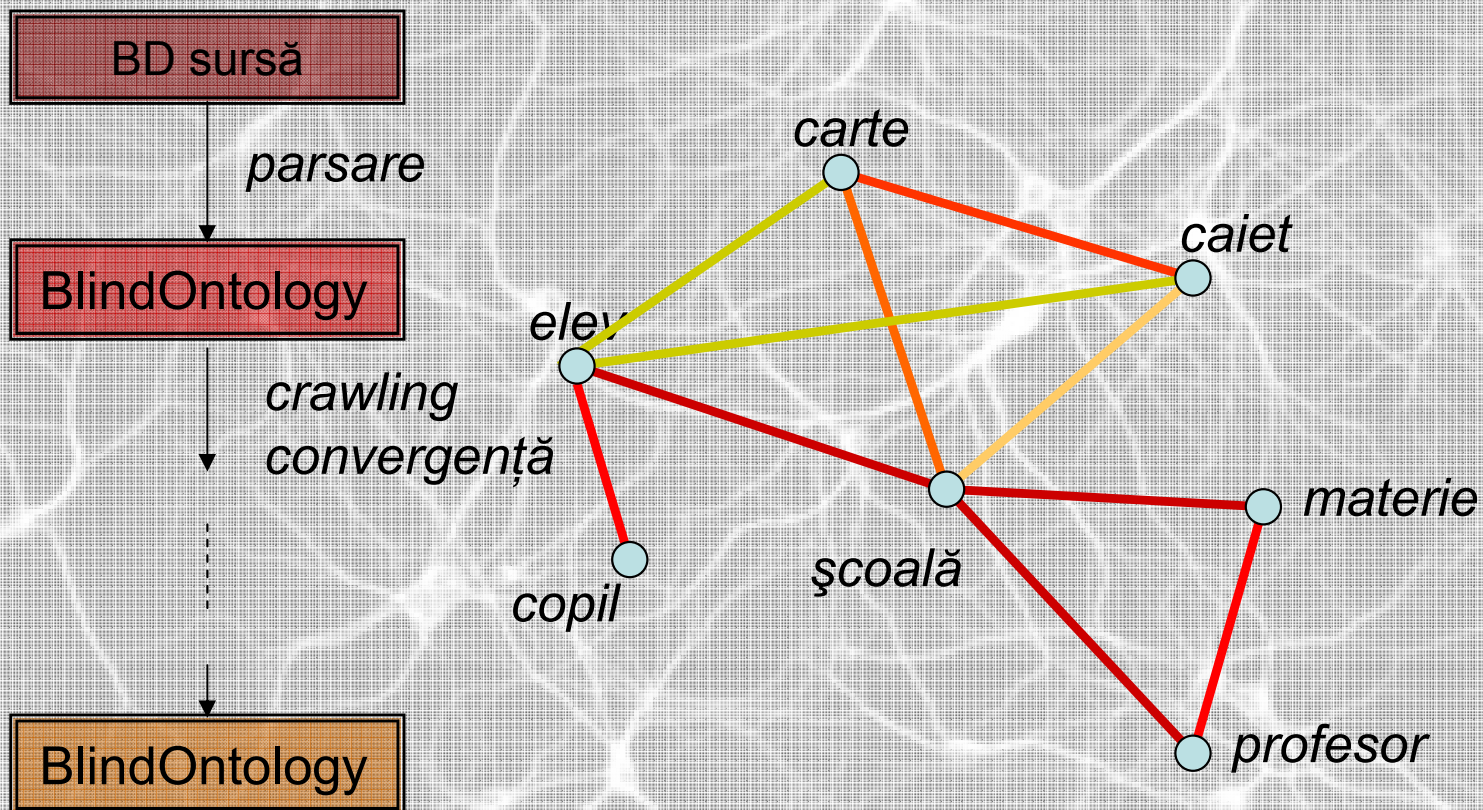
EXEMPLU FUNCȚIONARE

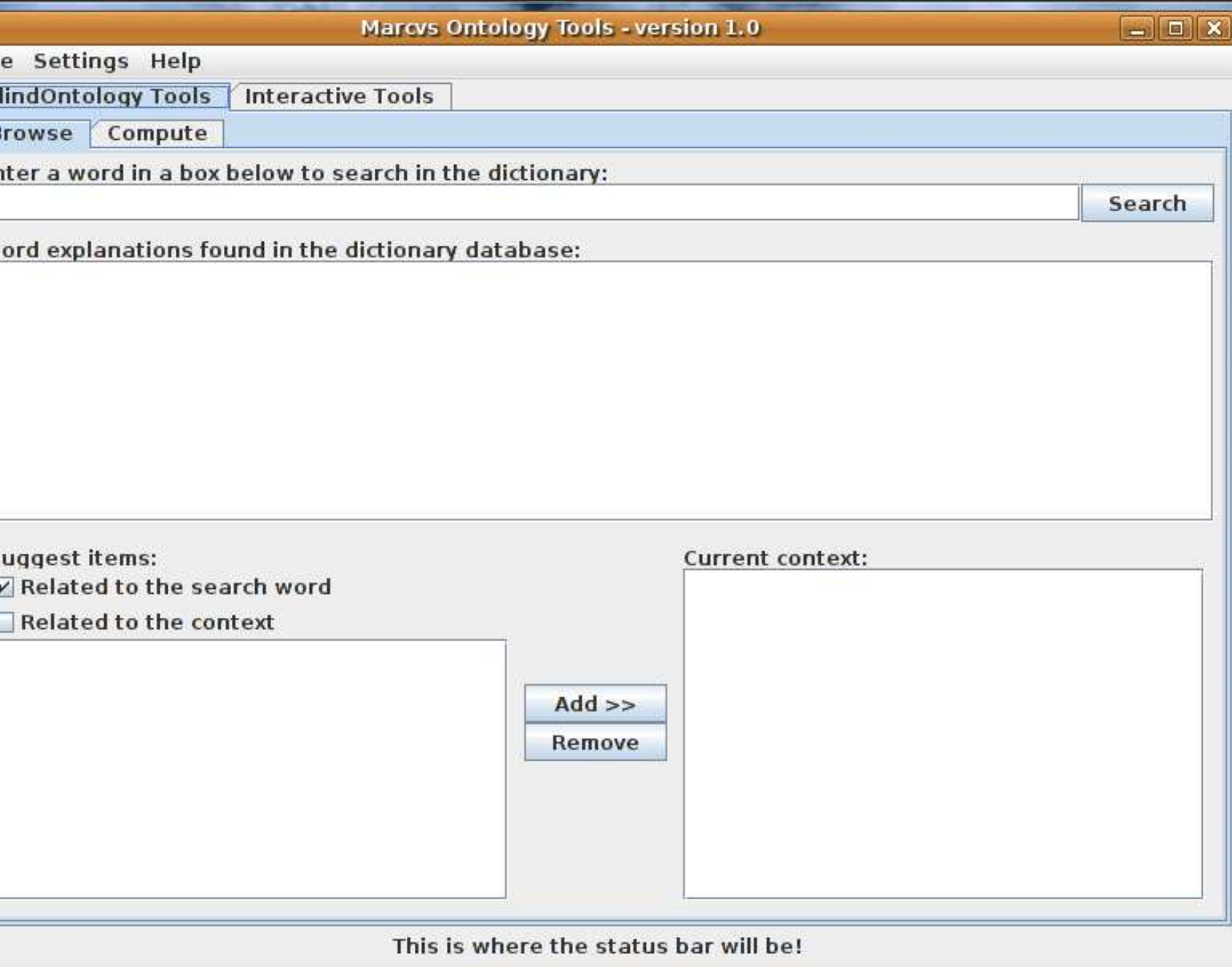


EXEMPLU FUNCȚIONARE



EXEMPLU FUNCȚIONARE





REMINDER.txt REMINDER2.
txt

working
rifoarena
oundations
Statistical
atural Lan...
ictures.zip
Badu
aramakabu
v Cohen R...
MINDER.txt

Marcus Ontology Tools - version 1.0

File Settings Help

BlindOntology Tools Interactive Tools

Browse Compute

Enter a word in a box below to search in the dictionary:

arbore

Search

Word explanations found in the dictionary database:

ARBORE, *arbori*, s. m. 1. Nume generic pentru orice plantă cu trunchi înalt și puternic, lemnos și cu mai multe ramuri cu frunze care formează o coroană: copac; *p. restr.* pom. ◇ Compuse: *arbore-de-cacao* = plantă lemnoasă specifică zonei tropicale, cultivată pentru semințele ei comestibile (*Theobroma cacao*); *arbore-de-cafea* = plantă lemnoasă din regiunile calde ale Americii Latine, Africii și Asiei, cultivată pentru semințele ei comestibile (*Coffea arabica*); *arbore-de-cauciuc* = plantă lemnoasă originară din America de Sud, din scoarța căreia se extrage cauciucul natural (*Hevea brasiliensis*); *arbore-de-pâine* = plantă
ARBORE s.m. ~ 5. (Inform.) Graf orientat, aciclic.

Suggest items:
☒ Related to the search word
☐ Related to the context

abietacee -- -16.5574806381802
acacia -- -17.0128311386329
acaju -- -16.3223597607493
aciclic -- 0.0
adăposti -- 0.0
afle -- 0.0

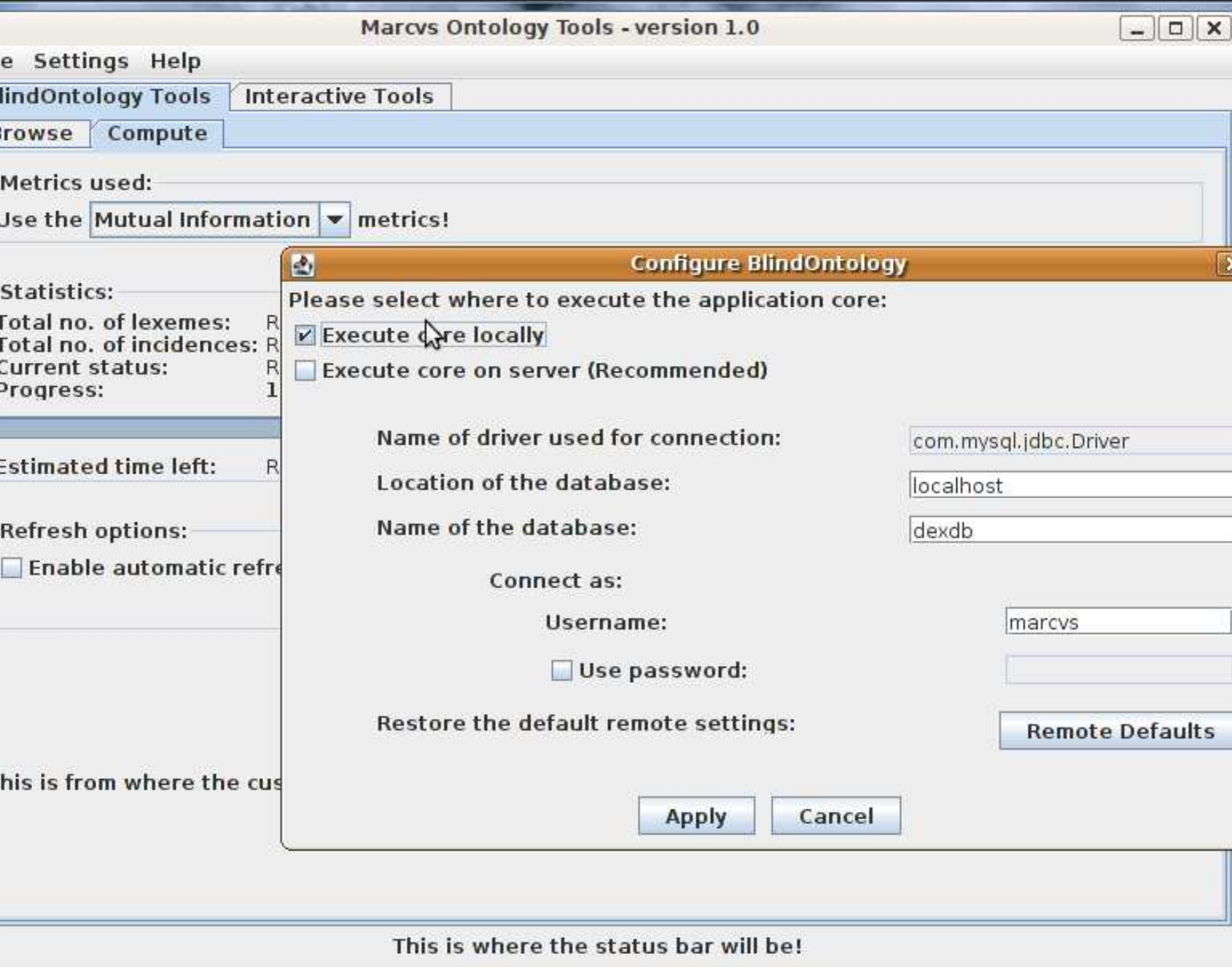
Add >>
Remove

Current context:

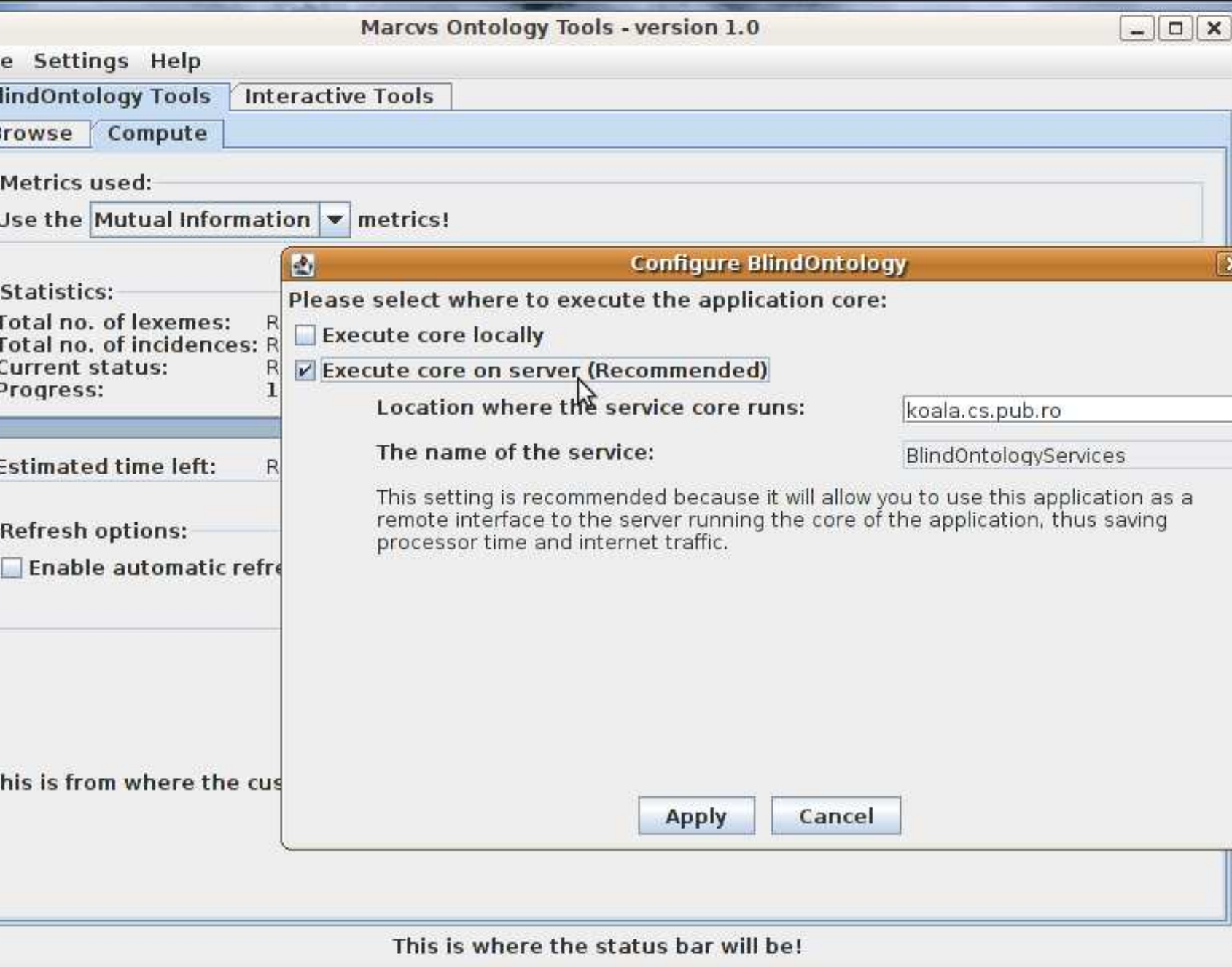
acacia
bonsai
coronament
tui
tulpina

This is where the status bar will be!

www.jeffreyklassen



REMINDEr.txt REMINDER2.txt



REMINDER.txt REMINDER2.txt

Marcvs Ontology Tools - version 1.0

File

Settings

Help

FindOntology Tools

Interactive Tools

Browse

Compute

Metrics used:

Use the

Mutual Information

 metrics!

Statistics:

Total no. of lexemes:

141365

Total no. of incidences:

1398518

Current status:

RUNNING

Progress:

4% (52144/1398518)

Estimated time left:

45h 26min 47sec

Populate

Stop

Refresh options:

☒ Enable automatic refresh

every 10 sec

Refresh Now

This is from where the custom SQL statements will be run!

This is where the status bar will be!

FACTORI DE LIMITARE

Bazele de date au structură plană

- Markup neomogen
- Foarte greu de parsat
- Lipsite de informații semantice

Numărul mare de lexeme (convergență lentă)

Lipsa unui corpus (inaplicabilitatea unor metrici)

Volumul mare de trafic în rețea (mai ales la crawling)

DEZVOLTARE VIITOARE

Suport pentru forme inflexionare

Schițarea unui “Multinet core”

Definirea unui limbaj formal “human readable” pentru interacțiune

Metrici mai adecvate și explorarea folosirii Wikipedia drept corpus

BIBLIOGRAFIE

- “Foundations of Natural Language Processing”, Christopher D. Manning, Heinrich Schütze
- „Speech and Language Processing“, Daniel Jurafsky, James H. Martin
- „Knowledge Representation and The Semantics of Natural Language“, Hermann Helbig

SURSE

Ultimele versiuni ale codului sursă, precum și mai multe detalii despre implementare vor putea fi găsite în curând la [această adresă](#).

VĂ MULȚUMESC!

ÎNTREBĂRI?