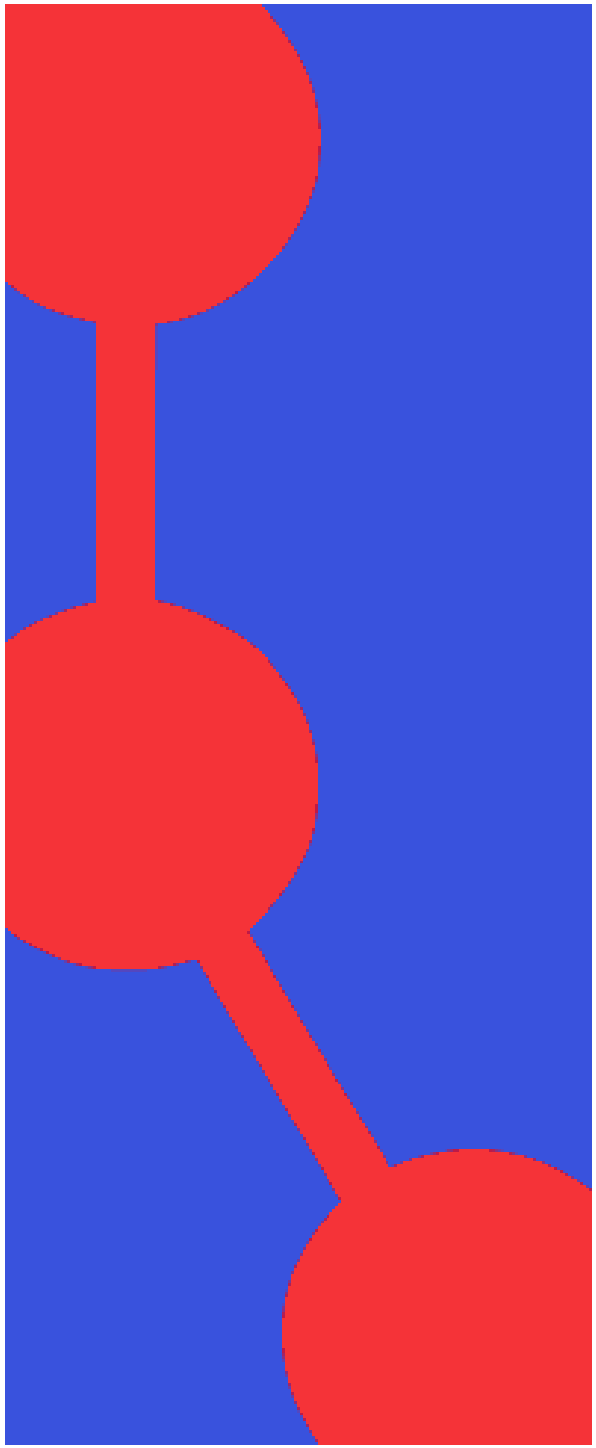


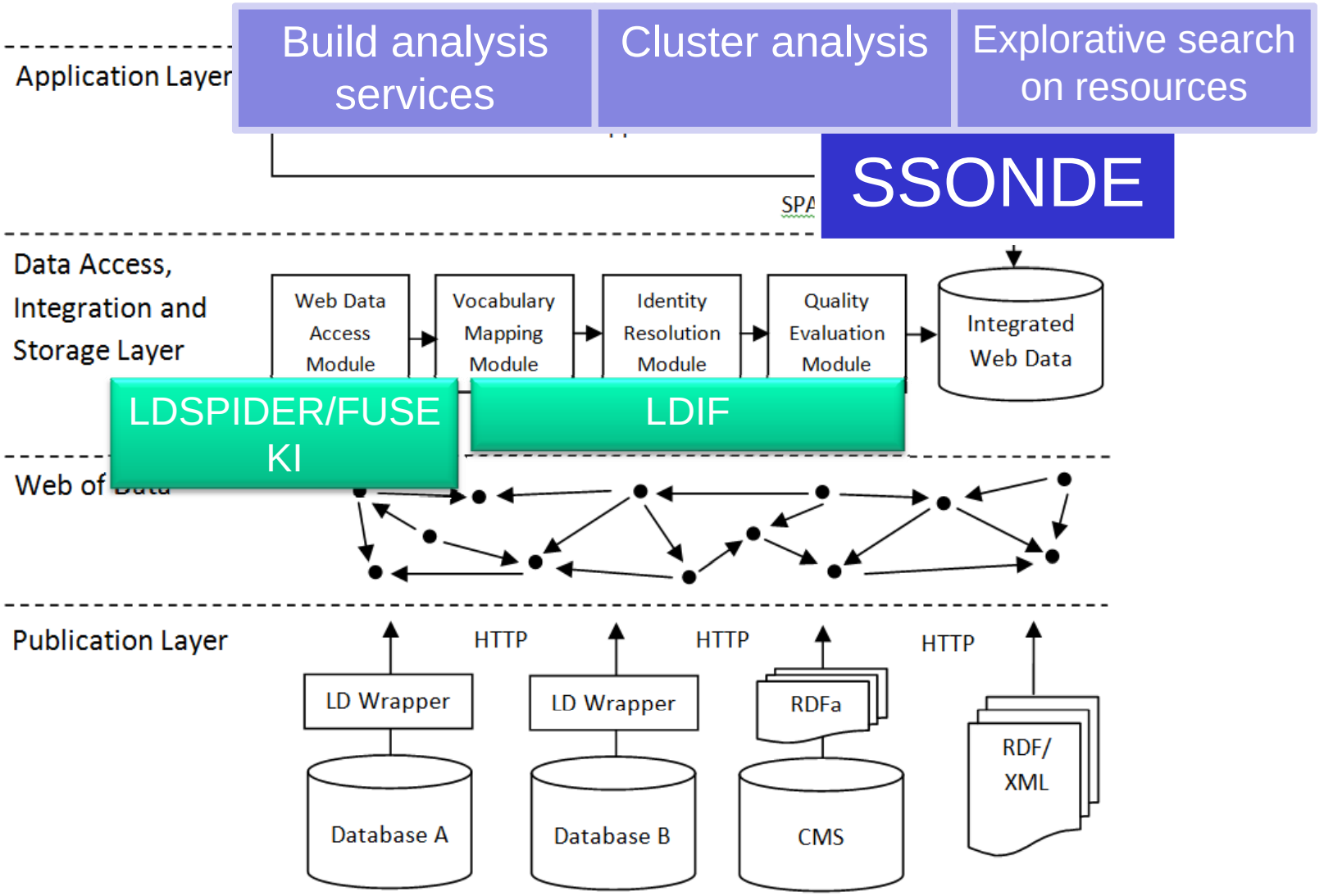
SSONDE:Semantic Similarity On liNked Data Entities (Example/DEMO)

Riccardo Albertoni
ralbertoni@delicias.dia.fi.upm.es

Ontology Engineering Group
Departamento de Inteligencia Artificial
Facultad de Informática
Universidad Politécnica de Madrid

Acknowledgment: SSONDE in its current (pre)release results mainly from the research activity I did at IMATI-CNR in collaboration with M. De Martino

Linked data Crawling architectural pattern



Getting started

SSONDE's code

- is hosted as an Google Code project, http://code.google.com/p/ssonde/
- licenced as open source code (GNU GPL v3)

Info about how getting started are available at http://code.google.com/p/ssonde/wiki/GettingStarted

Applying SSONDE

Let's apply SSONDE

- on Linked data resources exposed by third parties
  - Info about CNR provided by Gangemi et all, Data.cnr.it
- To work out similarities among researchers according to their research interests

What to download/install for this Example

- SSONDE
  - svn checkout http://ssonde.googlecode.com/svn/SSONDEv1 SSONDEV1DEMO

For crawling data

- LDSpider
  - code.google.com/p/ldspider/
- Fuseski
  - http://jena.apache.org/documentation/serving_data/index.html

Crawling RDF fragments from data.cnr.it

Which resources are you interested on?

1. Identify a set of seeds entities

- E.g. researchers, http://www.cnr.it/ontology/cnr/individuo/unitaDiPersonaleEsterno/ID226 (Riccardo Albertoni)

2. Figure out which entity features (i.e., Object properties and Data properties) are interesting

- Analyse schemas which are deployed to characterize that kind of resources, e.g.,
  - http://www.cnr.it/ontology/cnr/pubblicazioni.owl (prefix:pub)
    - **pub:autoreCNRDi** (the papers written by the authors)
    - **dc:subject** (authors scientific interest)

Crawling RDF fragments from data.cnr.it

3. Crawl metadata of resources to be analyzed creating a seedDATACNRIT.txt file

- java -jar ./LDSpider/ldspider-1.1e.jar -s seedDATACNRIT.txt -b **5** -1 -oe http://localhost:3030/XXX/update

3 bis. if you have a more precise idea of which feature you need, the list of properties to be traversed can be specified by –f option

- java -jar ./LDSpider/ldspider-1.1e.jar -b **5** -1

-f **http://www.cnr.it/ontology/cnr/pubblicazioni.owl#autoreCNRDi**

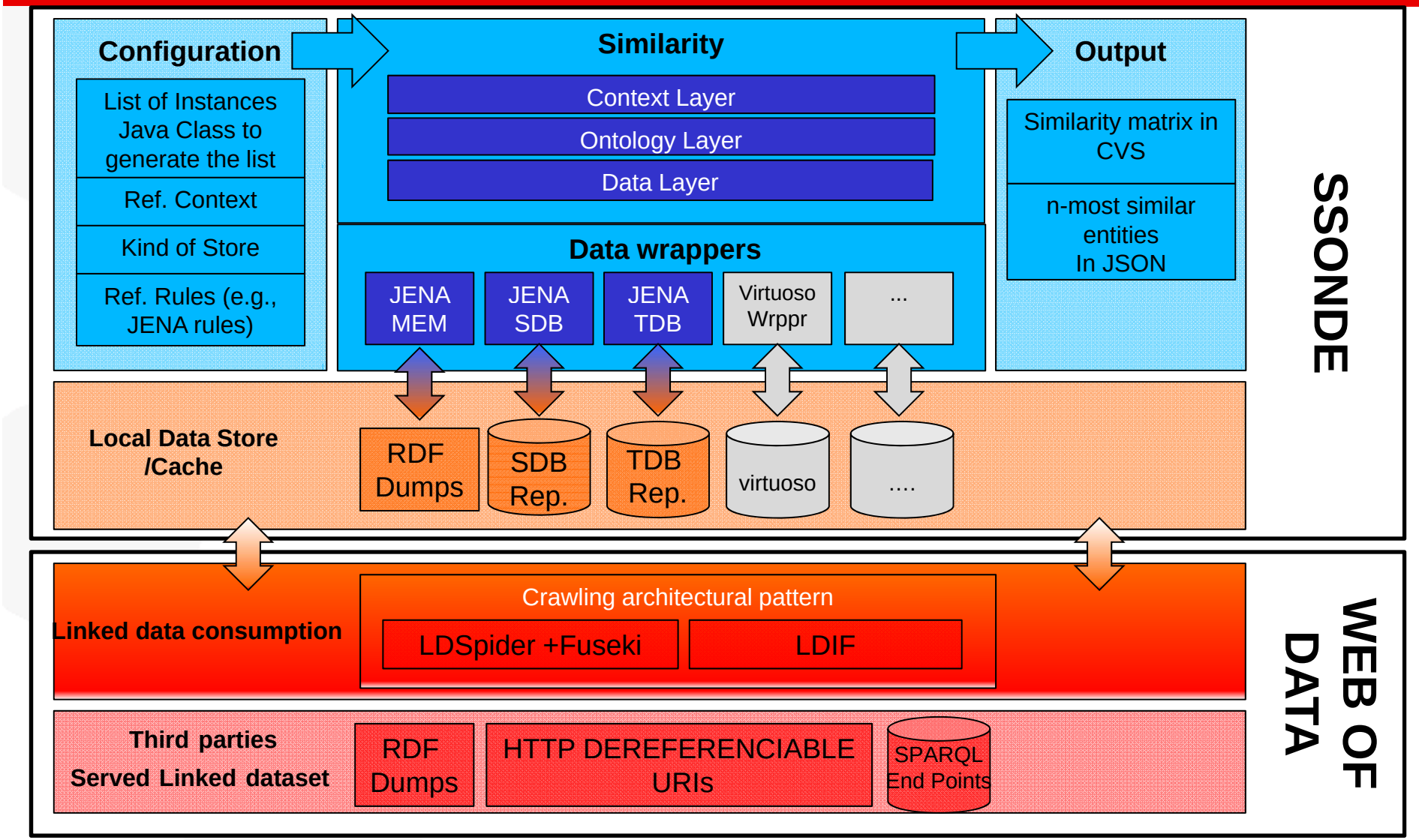
-f **http://www.cnr.it/ontology/cnr/pubblicazioni.owl#coautore**

-f **http://purl.org/dc/terms/subject** -f **http://www.w3.org/2004/02/skos/core#broader** -s seedDATACNRIT.txt -oe "http://localhost:3030/**XXX**/update"

generally,. not in this case,

If you are considering different linked data source you must perform

- Vocabulary mapping
- Identity resolution
- Quality filtering
- …
- …


Configuration
List of Instances Java Class to generate the list
Ref. Context
Kind of Store
Ref. Rules (e.g., JENA rules)
Similarity
Context Layer
Ontology Layer
Data Layer
Data wrappers
JENA MEM
JENA SDB
JENA TDB
Virtuoso Wrppr
...
Output
Similarity matrix in CVS
n-most similar entities In JSON
SSONDE
Local Data Store /Cache
RDF Dumps
SDB Rep.
TDB Rep.
virtuoso
....
Linked data consumption
Crawling architectural pattern
LDSpider +Fuseki
LDIF
Third parties Served Linked dataset
RDF Dumps
HTTP DEREFERENCIABLE URIs
SPARQL End Points
WEB OF DATA

Configuration file 1

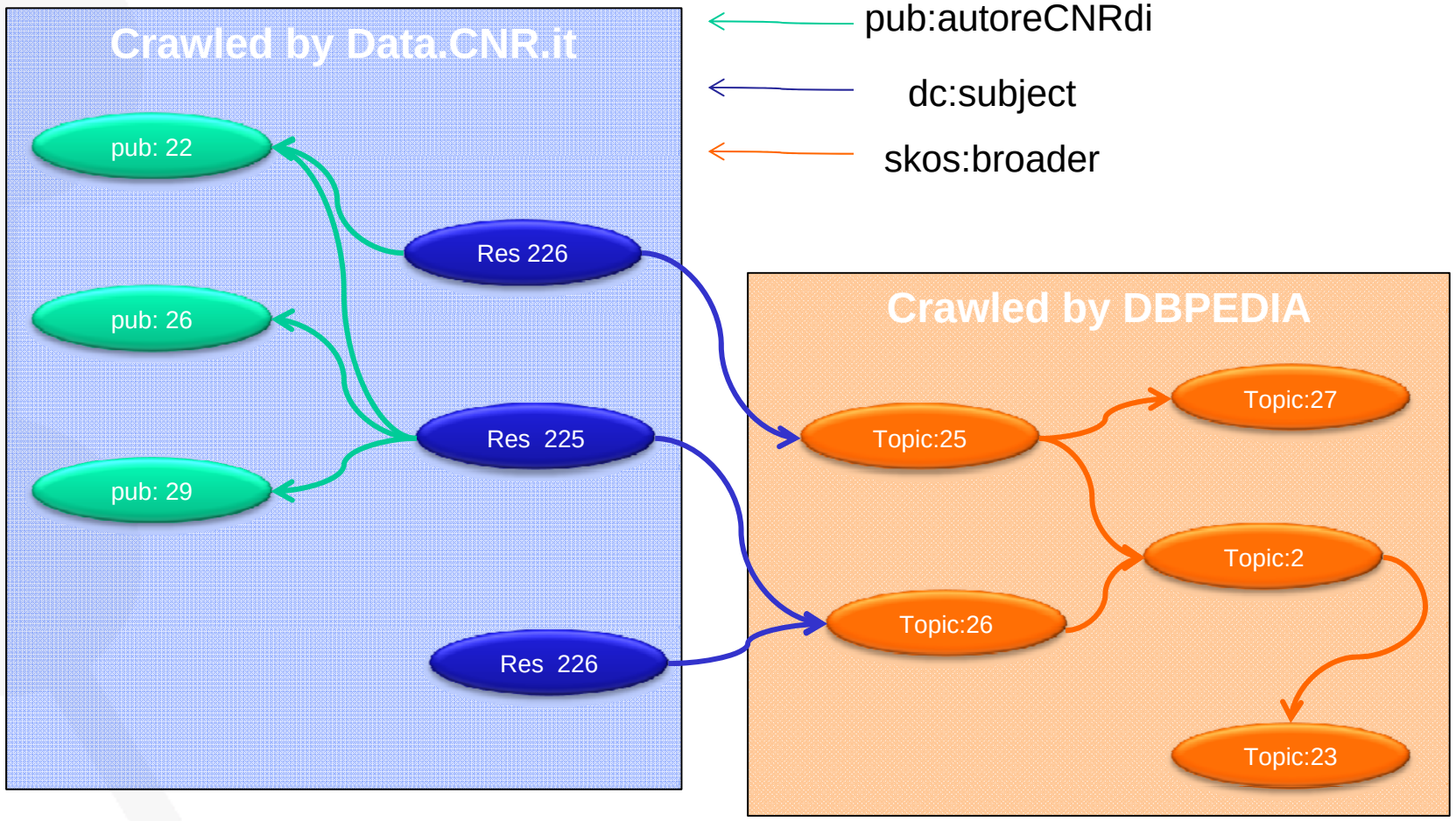
```
{  "StoreConfiguration":{
      "KindOfStore":"JENATDB",
      "RDFDocumentURIs":[  ],
      "TDBDirectory":"data/CNRIT/TDB-0.8.9/CNRR/"
   },
"InstanceConfiguration":{
      "InstanceURIsClass":"application.dataCNRIt.GetResearcherIMATIplusCoauthor"
   },
   "OutputConfiguration":{
      "KindOfOutput":"JSONOr
      "NumberOfOrderedResult":”20",
      "FilePath":"conf/dataCNRIt/ComplexContextResearchInterest/CRRIIntPub.res.json"
   },

 "ContextConfiguration":{
      "ContextFilePath":"conf/dataCNRIt/ComplexContextResearchInterest/CCRIIntPub.ctx"
   }
}
```

List of LOD Entities URI
Java class Implementing ListOfInputInstances

Similarity Matrix CSV - JSON encoding of top n-most similar

Context Encoded in a format in-house text format/ hopefully soon in JSON

Data.cnr.it – defining a context



PREFIX dc: <http://purl.org/dc/terms/>
PREFIX pub: <http://www.cnr.it/ontology/cnr/pubblicazioni.owl#>
[owl:Thing]-> {{}, { (pub:autoreCNRDi, Inter),(dc:subject, Simil)}}
[owl:Thing, dc:subject]-> {{},{(skos:broader, Inter)}}

No data properties are considered in this context

run SSONDE

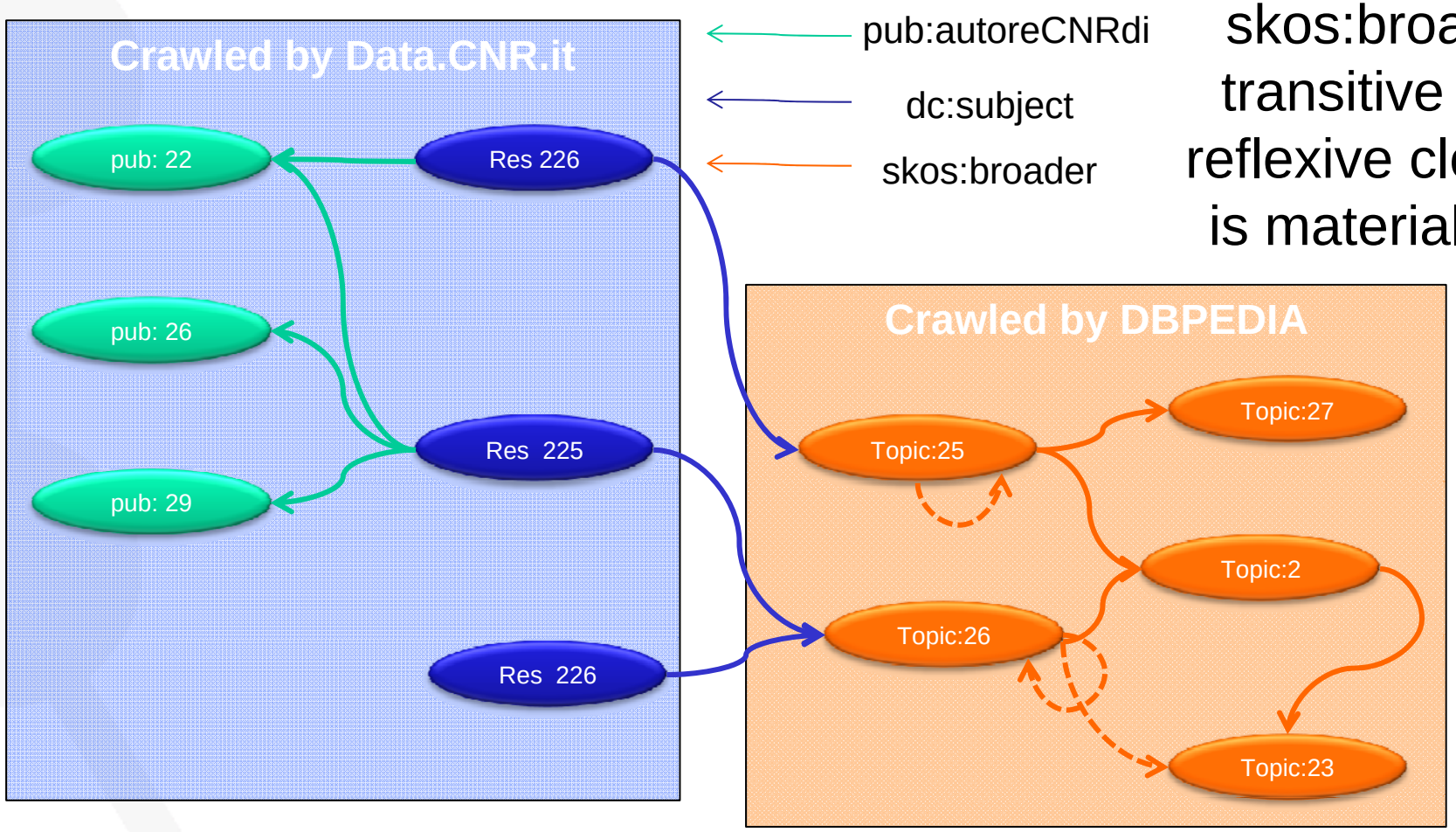
From the directory where you have downloaded SONDE (e.g., SSONDEV1DEMO)

- java -classpath ./lib/*:./bin SSONDEv1.SemSim

–conf JSONconfigurationFile1

- java -classpath ./lib/*:./bin/ SSONDEv1.SemSim -conf conf/dataCNRIt/ComplexContextResearchInterest/**CCRIInt Pub.param.json**

Data.cnr.it – defining a context



skos:broader transitive and reflexive closure is materialized

PREFIX dc: <http://purl.org/dc/terms/>
PREFIX pub: <http://www.cnr.it/ontology/cnr/pubblicazioni.owl#>
[owl:Thing]->{{},{ (pub:autoreCNRDi, Inter),(dc:subject, Simil)}}
[owl:Thing, dc:subject]-> {{},{(skos:broader, Inter)}}

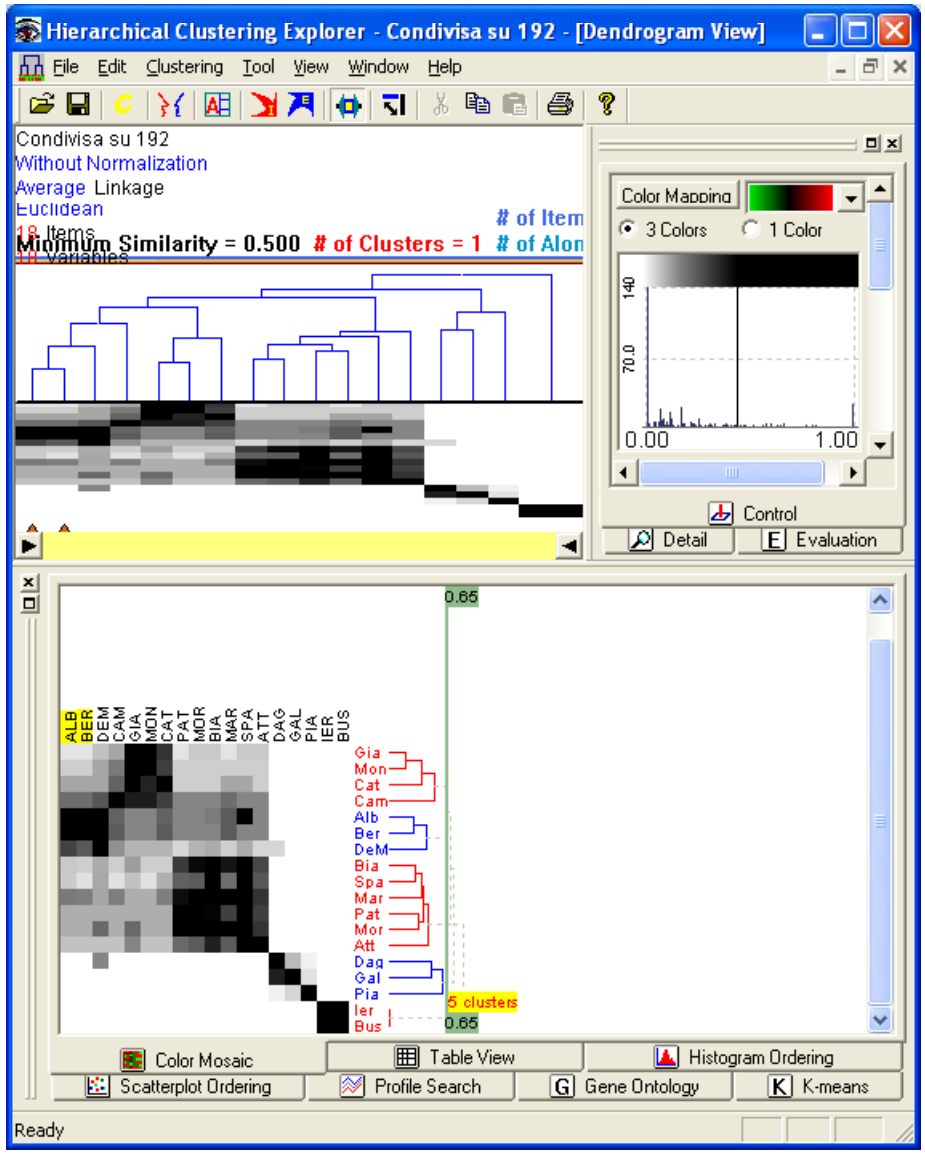
Ex. Of Tool to interpret SSONDE results

SSONDE does not have its own GUI,

results are interpreted by elaborating its similarity matrix via

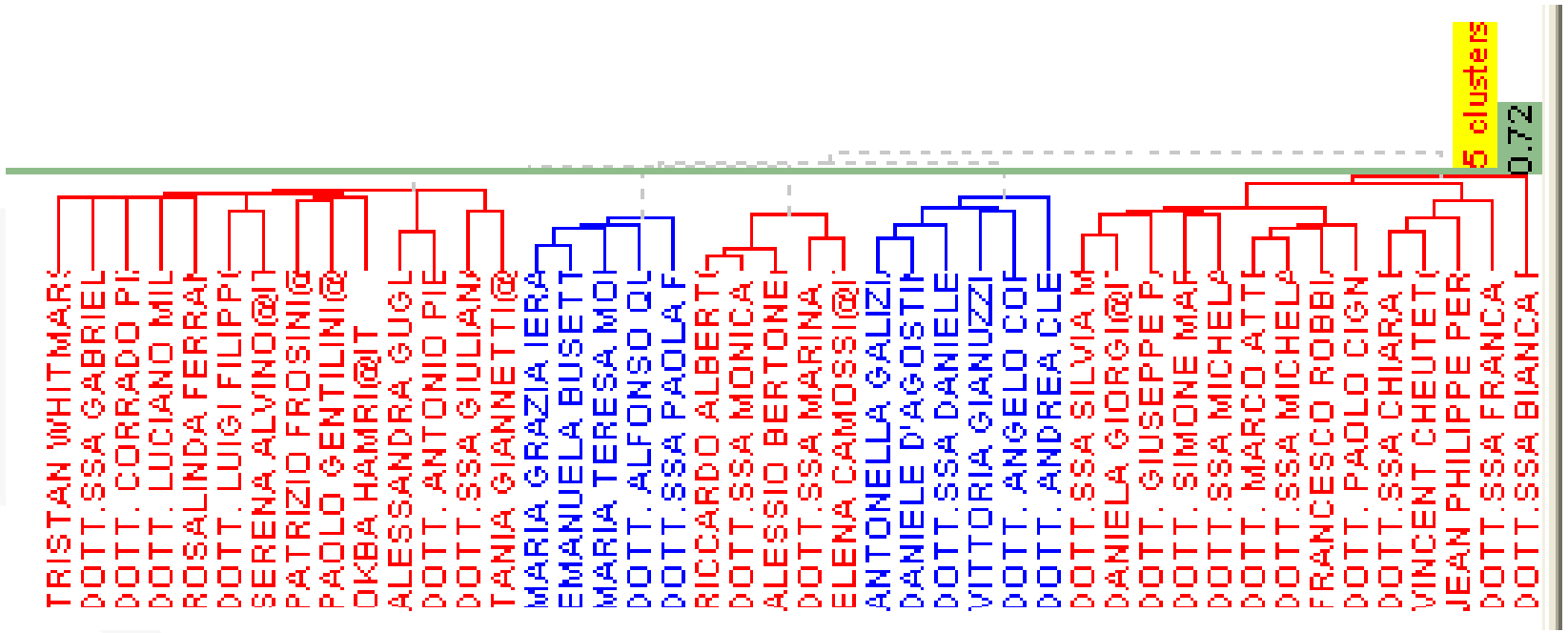
1) Excel files

2) Hierarchical Clustering Explorer, 3.0, Human-Computer Interaction Lab University of Maryland. http://www.cs.umd.edu/hcil/multi-cluster/.

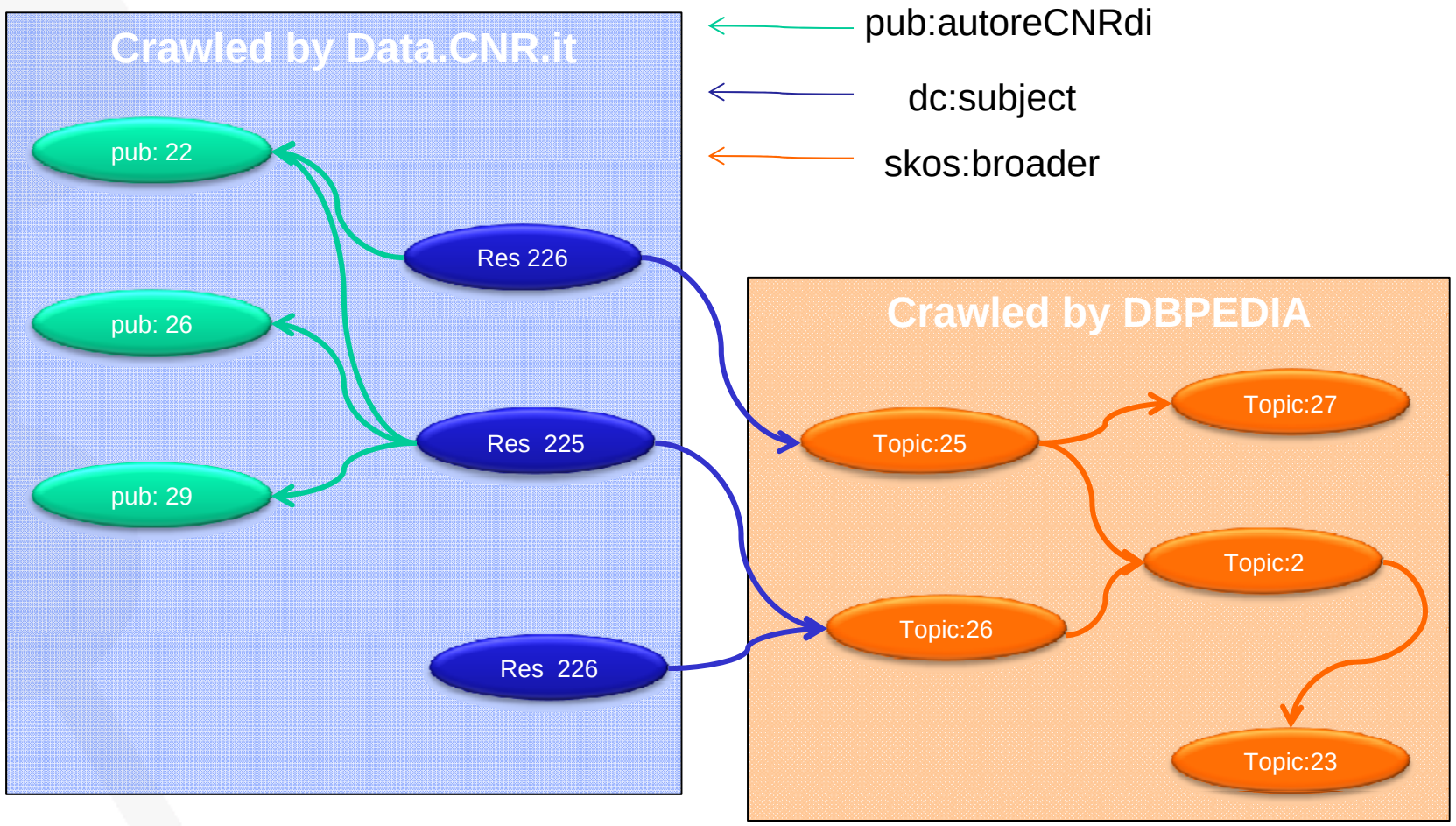


RICCARDO ALBERTONI@IT
DOTT.SSA MARINA MONTI@IT
DOTT.SSA FRANCA GIANNINI@IT
ELENA CAMOSSI@IT
DOTT.SSA MONICA DE MARTINO@I
ALESSIO BERTONE@IT
DOTT. SIMONE MARINI@IT
FRANCESCO ROBBIANO@IT
DOTT.SSA BIANCA FALCIDIENO@IT
DOTT. MARCO ATTENE@IT
DOTT.SSA MICHELA SPAGNUOLO@
DOTT. GIUSEPPE PATANE'@IT
DOTT.SSA MICHELA MORTARA@IT
DOTT.SSA CHIARA EVA CATALANO
DOTT.SSA SILVIA MARIA BIASOTTI
ALESSANDRA GUGLIELMI@IT
DOTT. ANTONIO PIEVATOLO@IT
DANIELA GIORGI@IT
DOTT. CORRADO PIZZI@IT
TRISTAN WHITMARSH@IT
VINCENT CHEUTET@IT
JEAN PHILIPPE PERNOT@IT
OKBA HAMRI@IT
PATRIZIO FROSINI@IT
ROSALINDA FERRANDES@IT
ANTONELLA GALIZIA@IT
DOTT.SSA GABRIELLA SANNITI DI
DOTT. PAOLO CIGNONI@IT
DOTT. ANDREA CLEMATIS@IT
VITTORIA GIANUZZI@IT
DANIELE D'AGOSTINO@IT
DOTT. ALFONSO QUARATI@IT
DOTT.SSA PAOLA FORCHERI@IT
DOTT.SSA DANIELE D'AGOSTINO@
DOTT. ANGELO CORANA@IT
DOTT. LUCIANO MILANESI@IT
MARIA GRAZIA IERARDI@IT
DOTT.SSA GIULIANA DETTORI@IT
PAOLO GENTILINI@IT
EMANUELA BUSETTI@IT
DOTT. LUIGI FILIPPO SARTI@IT
TANIA GIANNETTI@IT
MARIA TERESA MOLFINO@IT
SERENA ALVINO@IT
RICCARDO ALBERT
DOTT.SSA MARINA
DOTT.SSA FRANCA
ELENA CAMOSSI@it
DOTT.SSA MONICA
ALESSIO BERTONE
DOTT. SIMONE MAR
FRANCESCO ROBBI
DOTT.SSA BIANCA
DOTT. MARCO ATTE
DOTT.SSA MICHELA
DOTT. GIUSEPPE P
DOTT.SSA MICHELA
DOTT.SSA CHIARA
DOTT.SSA SILVIA M
ALESSANDRA GUGL
DOTT. ANTONIO PIE
DANIELA GIORGI@it
DOTT. CORRADO PI
TRISTAN WHITMARS
VINCENT CHEUTET
JEAN PHILIPPE PER
OKBA HAMRI@it
PATRIZIO FROSINI@
ROSALINDA FERRAN
ANTONELLA GALIZI
DOTT.SSA GABRIEL
DOTT. PAOLO CIGN
DOTT. ANDREA CLE
VITTORIA GIANUZZI
DANIELE D'AGOSTIN
DOTT. ALFONSO QU
DOTT.SSA PAOLA F
DOTT.SSA DANIELE
DOTT. ANGELO COR
DOTT. LUCIANO MIL
MARIA GRAZIA IERA
DOTT.SSA GIULIAN
PAOLO GENTILINI@
EMANUELA BUSETT
DOTT. LUIGI FILIPP
TANIA GIANNETTI@
MARIA TERESA MOI
SERENA ALVINO@it

Hierarchical Clustering



Data.cnr.it – defining a context



PREFIX dc: <http://purl.org/dc/terms/>
PREFIX pub: <http://www.cnr.it/ontology/cnr/pubblicazioni.owl#>
[owl:Thing]->{{},{ (pub:autoreCNRDi, Inter),(dc:subject, Simil)}}
[owl:Thing, dc:subject]-> {{},{(skos:broader, Inter)}}

What next?

(i) semantic similarity optimization:

- (i) **the caching of intermediate similarity results**
- (ii) the adoption of MapReduce paradigm to speed up the assessment of semantic similarity;

(ii) domain driven extensions at data layer:

- (i) defining new data layer measures suited for geo-referenced entities
- **(ii) the multilingual similarity recently proposed by Kartic and Jorge could be considered for inclusion in the data layer of SSONDE**

(iii) definition of interfaces sifting entities according to their similarity, e.g., exploiting existing visualization frameworks (e.g., Exibit, Google visualization and JavaScript InfoVis Toolkit).

What I'd like you to think about …

- Can SSONDE be useful in any of your current research activities ?
- Can SSONDE be deployed in some of your future projects (proposal)?
- Are you interested in contributing somehow to SSONDE?

Further details are available in

- R. Albertoni, M. De Martino, SSONDE: Semantic Similarity On liNked Data Entities, 6th Metadata and Semantics Research Conference, 28-30 November 2012 - Cádiz (Spain) [to appear]
- Framework documentation
  - http://code.google.com/p/ssonde/wiki/GettingStarted