UNICOMP: identification of enterprise competencies to build collaborative networks

PRO-VE 2010

Kafil Hajlaoui, Xavier Boucher and Omar Boussaid

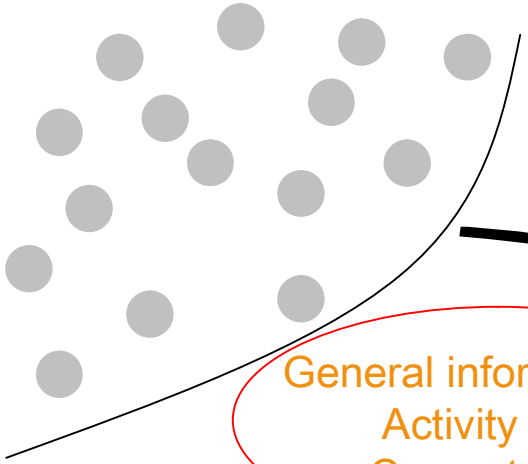
Kafil.hajlaoui@univ-lyon2.fr

Plan

1. Introduction: competence identification
2. Ontology and lexical patterns
3. UNICOMP
4. Conclusion

Introduction: competence identification

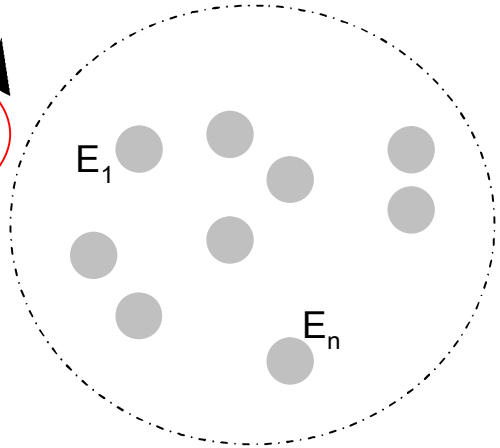
Hypothesis :
Open universe provided by the web
Public information : web sites



General information on:
Activity fields
Competencies
+
Clustering method

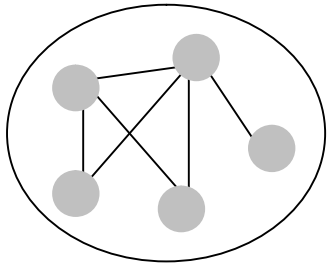
Semi-automatic extraction of information from the web sites

Virtual Breeding Environment



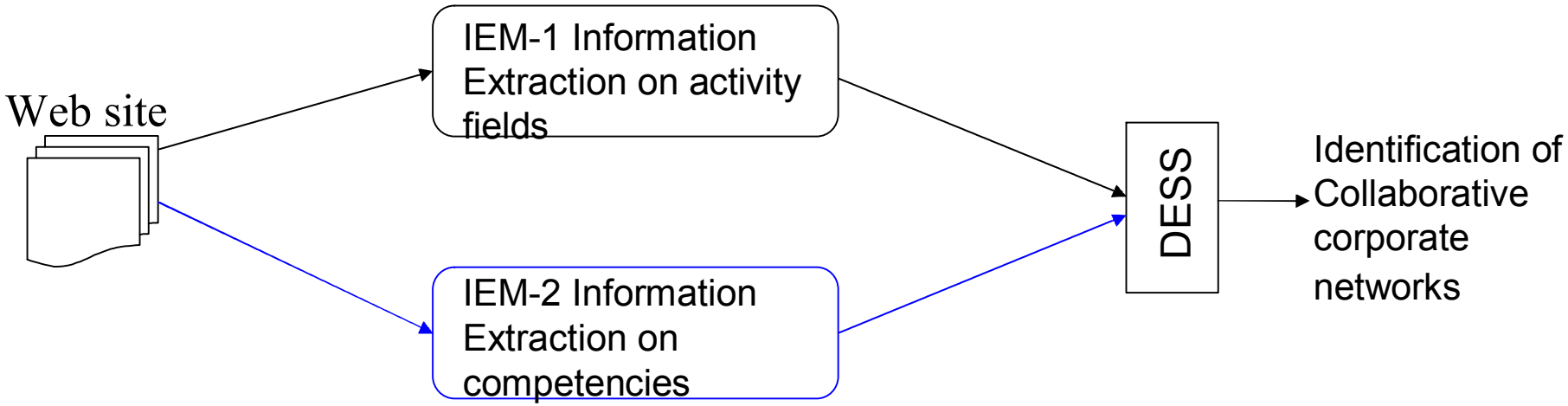
Market opportunity
+
Data base information
Shared by the partners

Virtual organization



Introduction: competence identification

Co-operation between firms = Complementary activities
&
Similar competencies



Informational context :

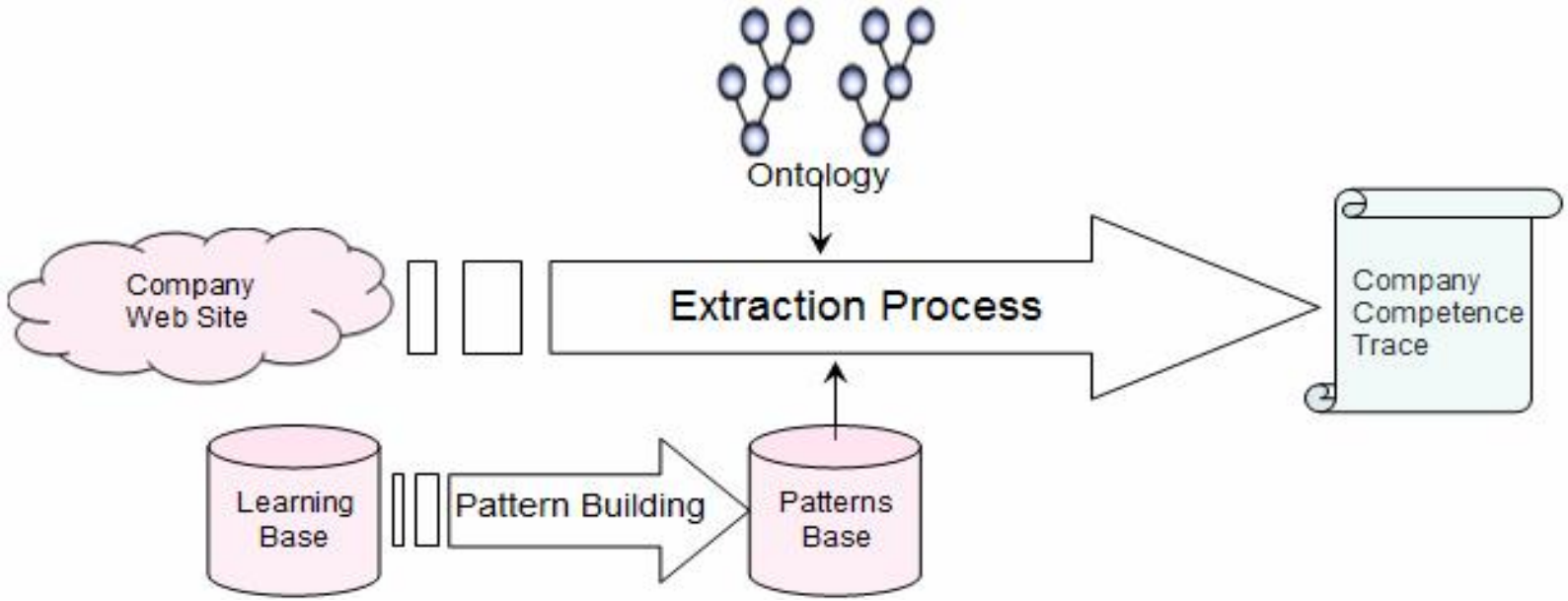
- Complexity of the notion of « Enterprise competence » : linked to technologies, to human resources, to methods and know-how at use in each company
- Necessity of linguistic approach : lots of distinct terms and expressions can bring pieces of information ; semantic ambiguity (context) ; synonymy etc…
- No structured semantic resource available

Plan

1. Introduction: competence identification
2. Ontology and lexical patterns
3. UNICOMP
4. Conclusion

Ontology and lexical patterns

Competence Identification based on ontology and pattern maching



Ontology →

- ✓Conceptual structure of « competence traces »
- ✓Comparison of distinct companies

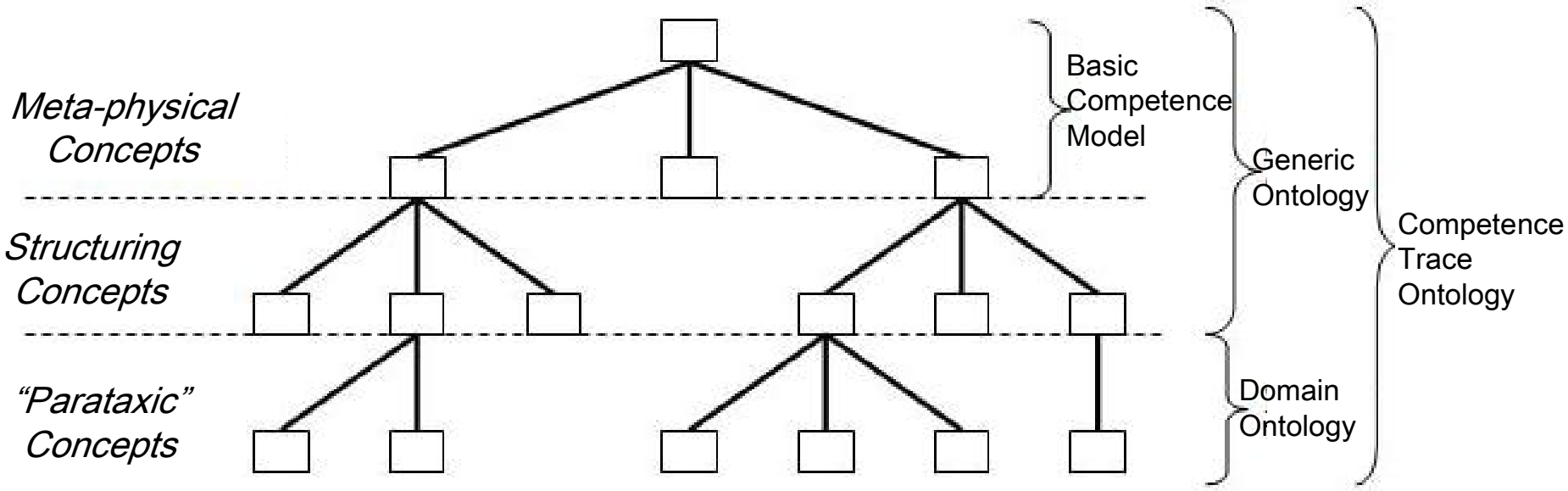
→ Similarity of « competence traces »

Semantic patterns →

- ✓Deal with ambiguïty and other semantic issues during the extraction procedure

Ontology and lexical patterns

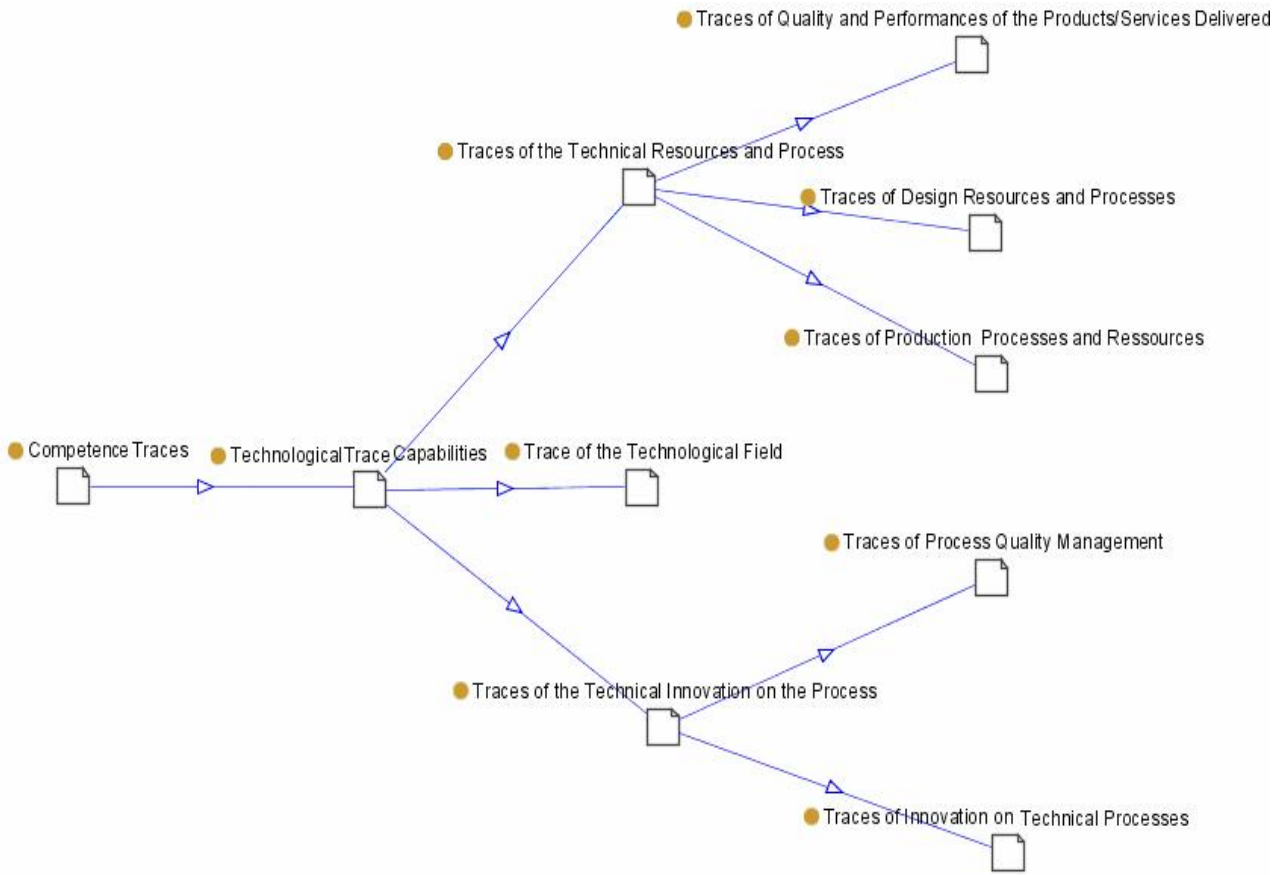
Ontology concepts have been structured on 3 distinct conceptual levels (genericity)



Ontology and lexical patterns

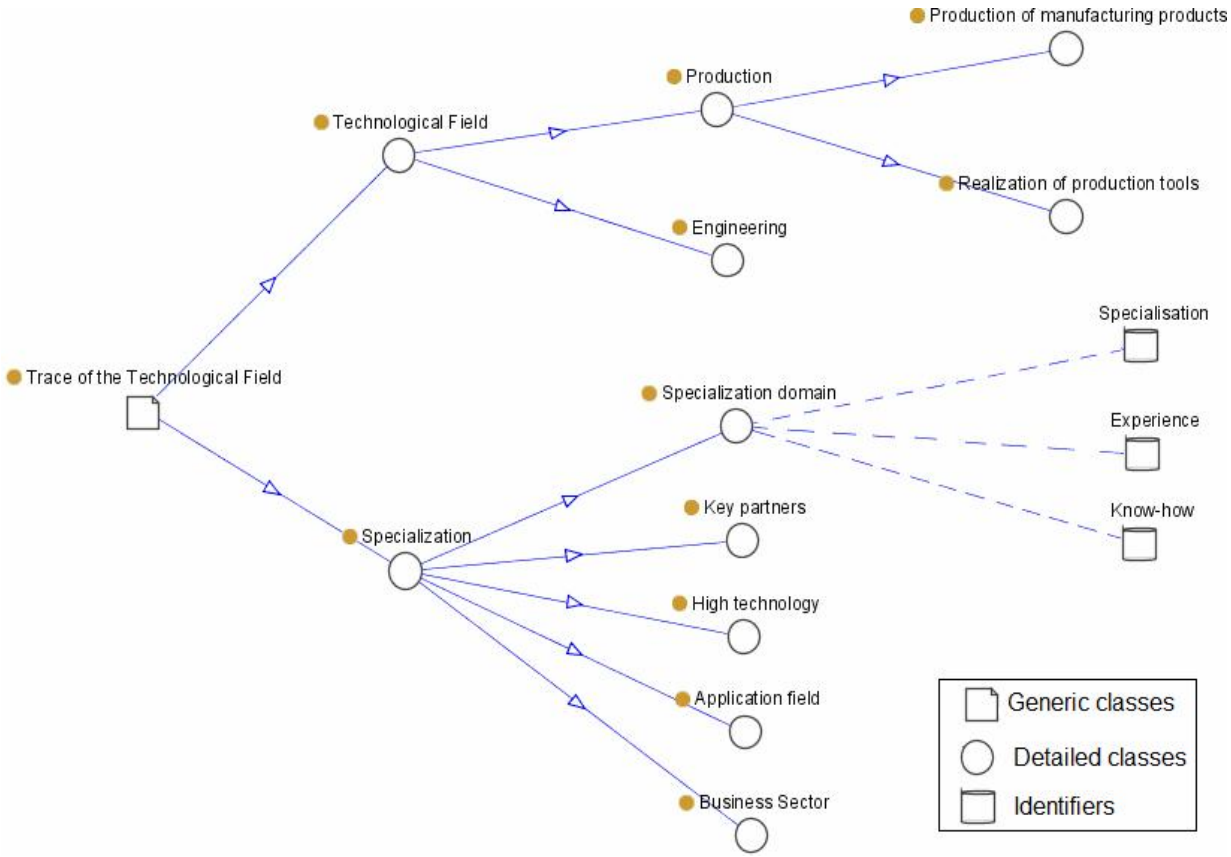
Ontology for competence traces :

Structuring level of the ontology for Technological capabilities



Ontology and lexical patterns

Parataxic level of the ontology (mechanical industry)



A set of identifiers is associated with each conceptual class of the domain ontology for the *activation process*

Ontology and lexical patterns

Lexical Patterns

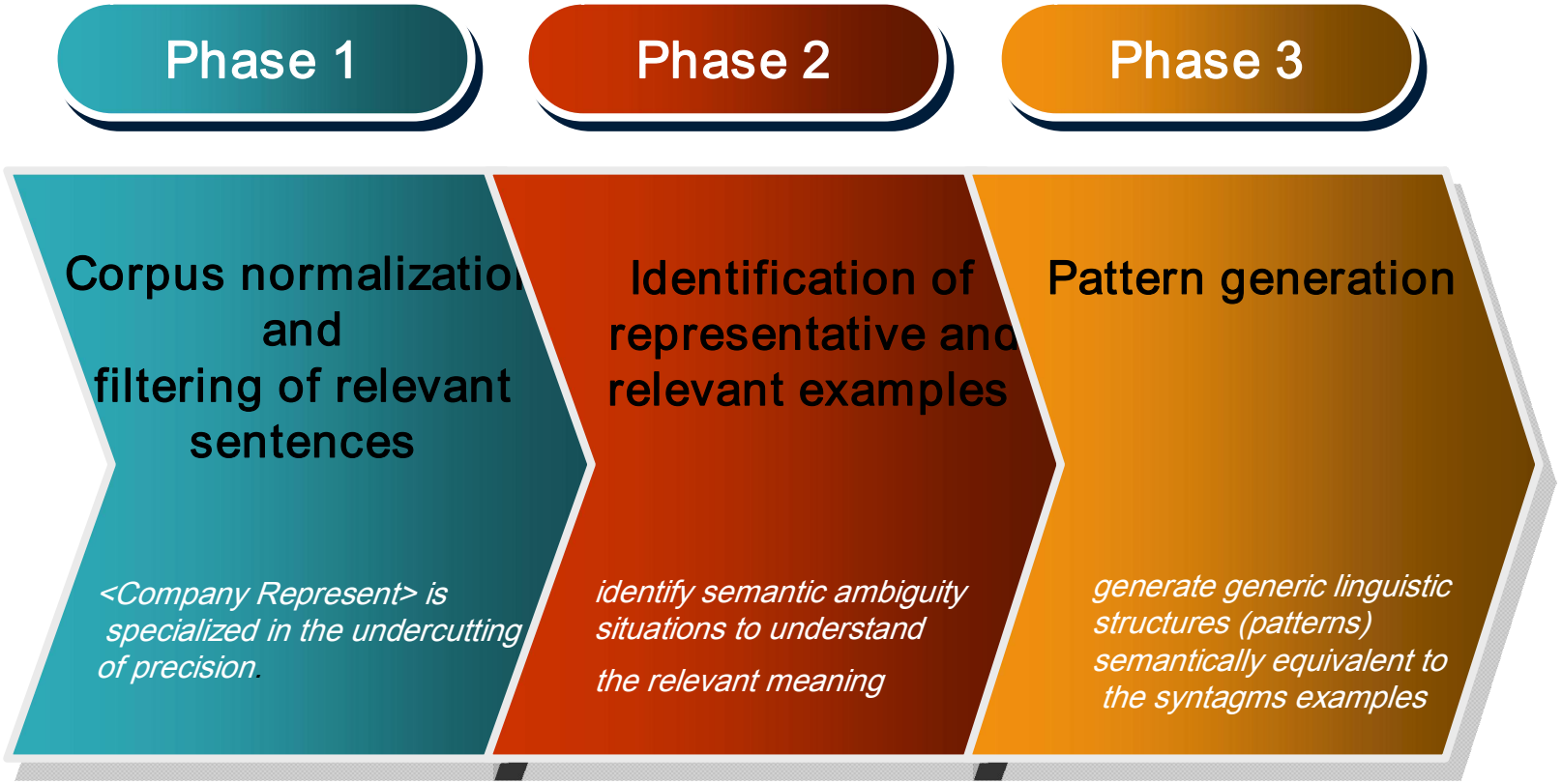
<nom> <verbe> <IDENTIFER> <nom>

- Lexical patterns aim at formalising a contextual signature of identifier.
- Patterns are based on principles of distributive semantics which states that the meaning of a word is strongly correlated to its context.

Patterns identify and formalize linguistic relationships,
by defining syntactic constraints on the context of the terms

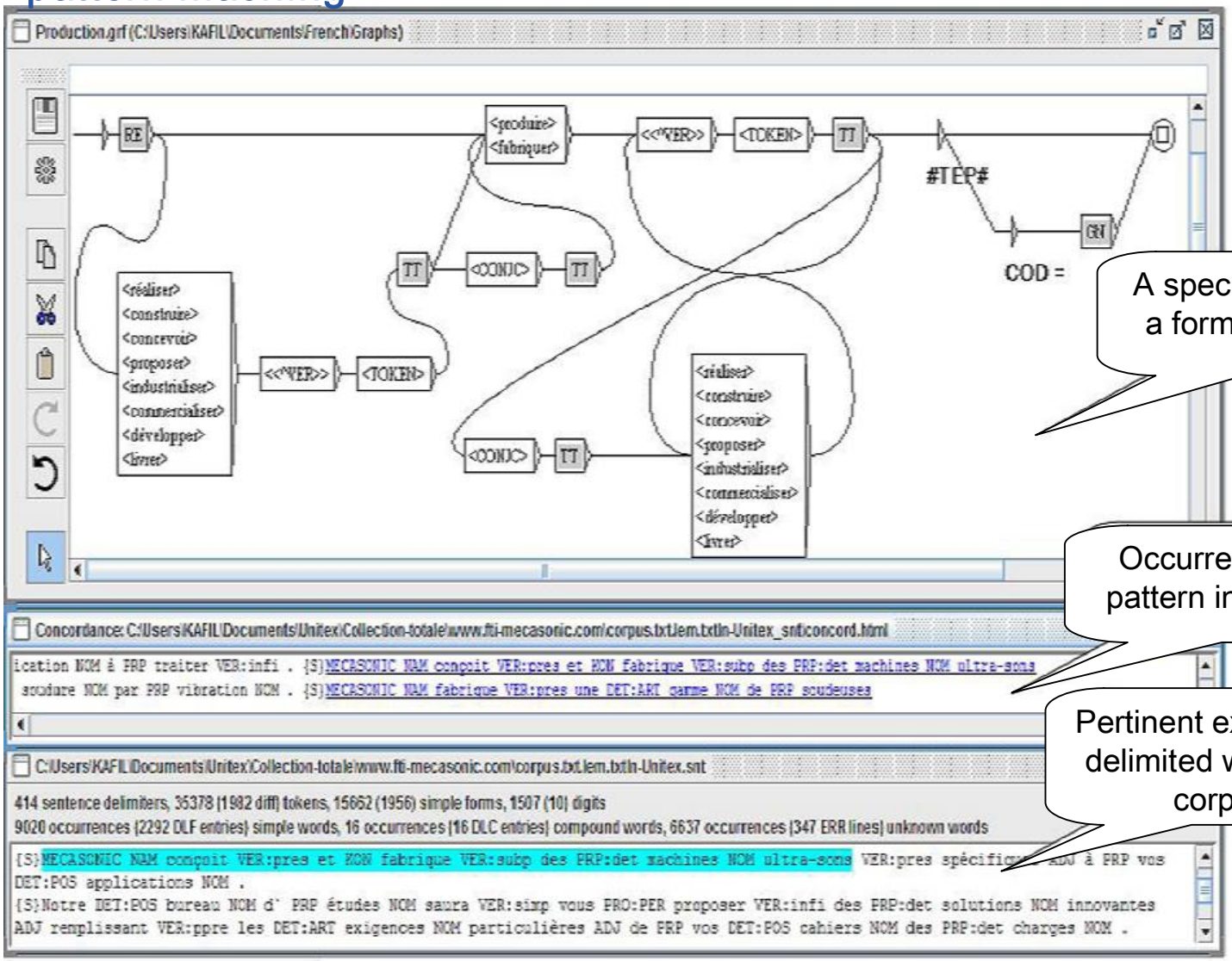
Ontology and lexical patterns

How to identify and formalize patterns ?



Ontology and lexical patterns

pattern maching



A specific pattern in a formal language

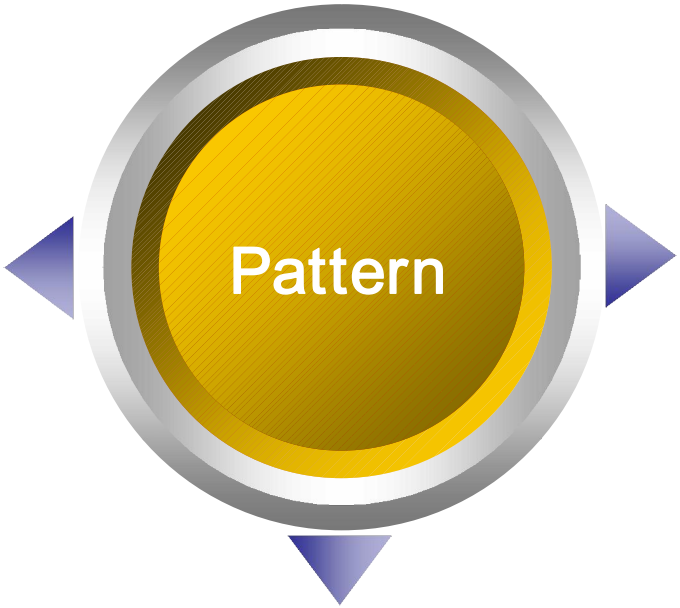
Occurrences of the pattern in a web site

Pertinent expression delimited within the corpus

Ontology and lexical patterns

3 typical use cases of patterns for information extraction:

Simple detection : to confirm the presence of an *identifier*. This pattern category is only applied for *identifiers* without ambiguity.



Extraction of additional information: For instance, with the identifier "specialization", we need to extract the speciality of the company and not only to detect that the company has a speciality.

Semantic clarification : The linguistic scheme makes possible to deal with ambiguity situations by checking contextual information which confirms the semantic interpretation of a text fragment

Plan

1. Introduction: competence identification
2. Ontology and lexical patterns
3. UNICOMP
4. Conclusion

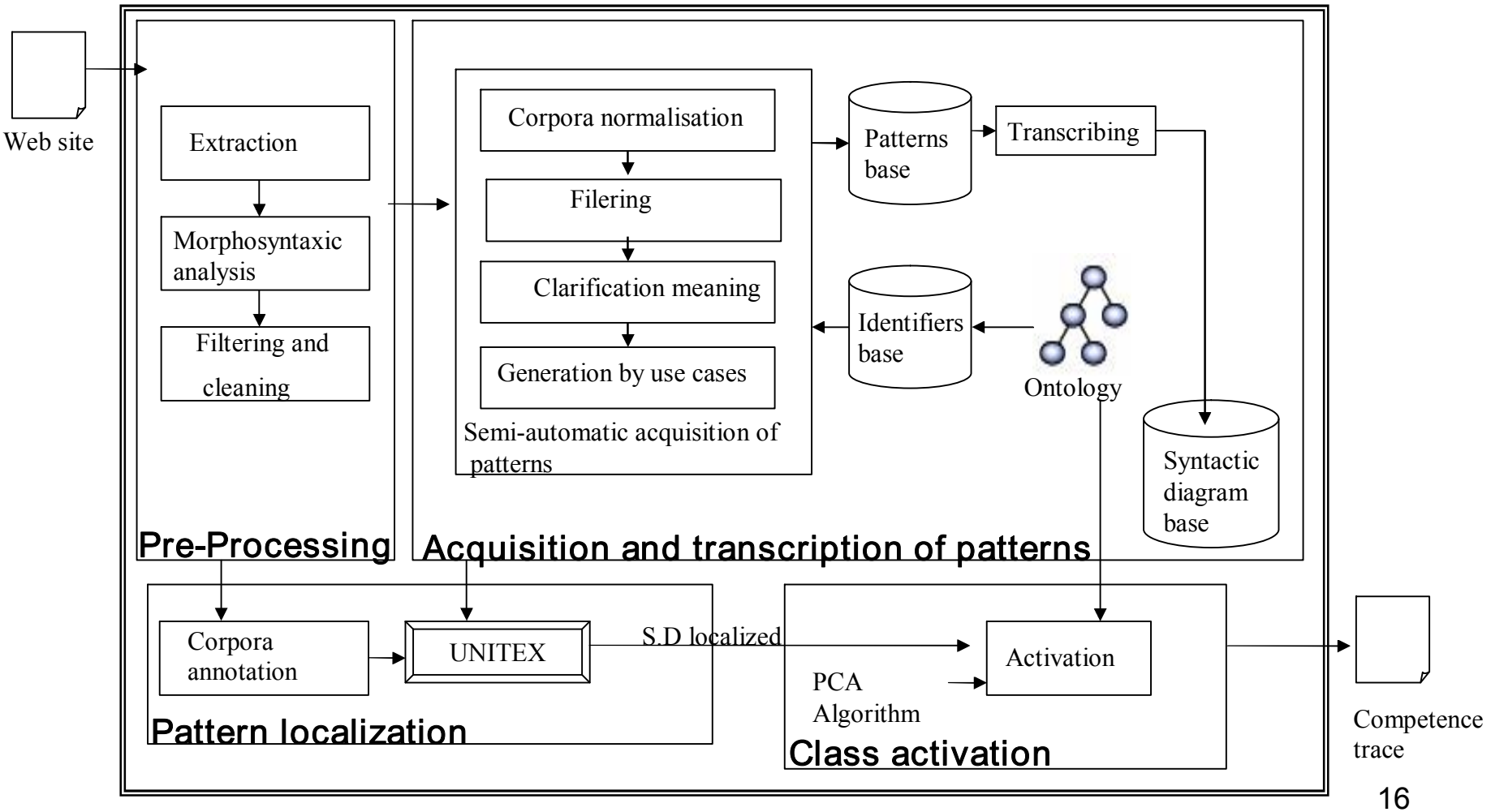
UNICOMP

UNICOMP system is the implementation of our approach which aims at extracting company competence traces, using :

- Public information available on websites
- Ontology for competence traces
- Lexical Patterns

UNICOMP

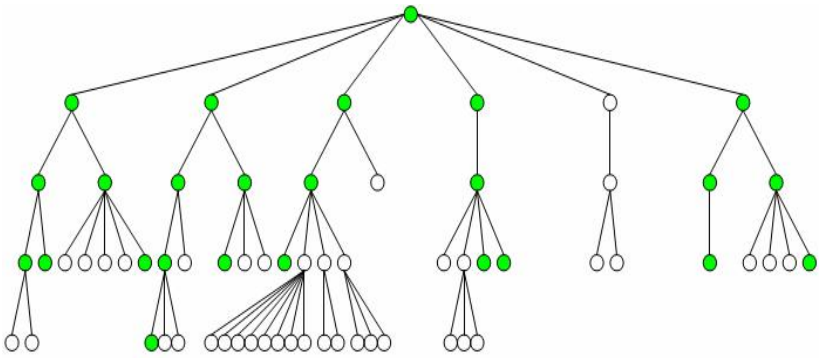
Architecture of the system



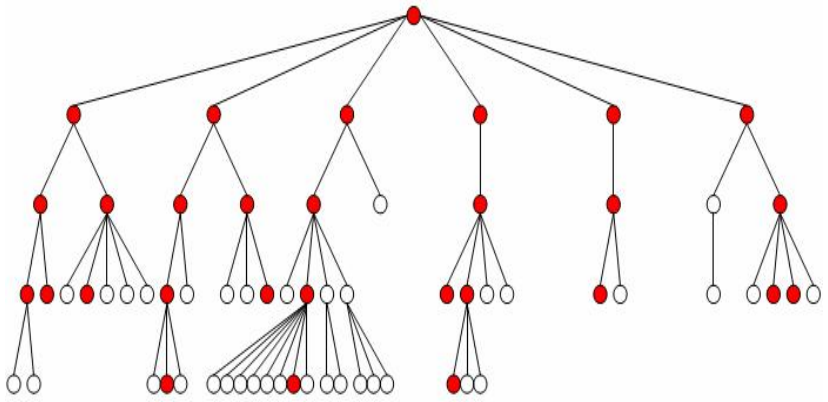
UNICOMP

For each company : which classes of the ontology are considered activated ?

Compagny1: www.boisset-et-cie.fr



E2: www.chambon.com



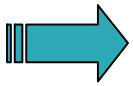
$$\delta(o,o') = 1 - \frac{\sum_{p=1}^{p=4} P\,|L_p(o) \cap L_p(o')|}{\sum_{p=1}^{p=4} P\,|L_p(o) \cup L_p(o')|}$$

UNICOMP

The performance evaluation of UNICOMP system

Precision : the capacity of the system to exclude the non-relevant classes
Recall : the capacity of the system to activate the relevant classes

Entreprise	Precision	Recall
www.boisset-et-cie.fr	0.81	0.56
www.chambon.com	0.92	0.7
www.flip-elec.fr	0.87	0.5
www.martin-joseph.com	1	0.66
www.bargy-decolletage.com	0.75	0.54
www.entechno.fr	1	0.7
www.attax.com	0.76	0.83
www.fti-mecasonic.com	0.8	0.66
www.isojet.com	0.87	0.77
www.sic-marking.com	0.88	0.57
Moyenne	**0.87**	**0.64**



These performances could still be improved both by a more exhaustive ontology and by a larger pattern base

Conclusion

- We have proposed an approach which takes advantages of semantic information from the domain (concerning competence descriptions) …

-… but which remains very adaptable from one business domain to another.

- The association of patterns and ontology, provides to the information extraction system a good ability to identify correctly the classes of a « competence traces ontology », corresponding to a company's web site.

- Further work :

-To measure the similarity among the competences traces from distinct companies.

-An automatic or semi-automatic enrichment of the ontology and the pattern base, using case-learning techniques would be able to increase the overall performance of the system.

Thank You !

Enterprise competence ?

•The overall competence of a firm
•Emerges as a combination of capabilities, notably the technological and methodological capabilities…
• These capabilities results from the utilisation of internal resources : human, technical, informational, organisational resources

Ontolology ?

• Of competence traces
• Built with a methodology which provide a structured approach to control the semantic issues : ARCHONTE was selected

ARCHONTE METHODOLOGY

operationalisation

Terms
of the domain

differential ontology

reference ontology

computable ontology

Ontology Formalisation

Ontology concepts have been structured on 3 distinct conceptual levels (genericity)