The 2nd IEEE International Workshop on Advanced Information Systems for Enterprises

IWAISE'12

SYNTACTICO-SEMANTIC INTERPRETATION OF NATURAL LANGUAGE QUERIES ON A MEDICAL ONTOLOGY

Bouhalika Chouaib, Boufaida Zizette
LIRE laboratory
Mentouri University
Constantine, Algeria

Constantine, November 10- 11, 2012

PLAN

1. Introduction
2. Related works
3. The proposed approach
4. System Architecture
5. Conclusion & perspectives.

1. INTRODUCTION

- The interrogation of knowledge representation formalisms such as ontologies is a central requirement of the Semantic Web.

- Providing a natural language interface to ontologies will not only offer ordinary users the convenience of acquiring needed information from ontologies, but also expand the influence of ontologies and of the semantic web.

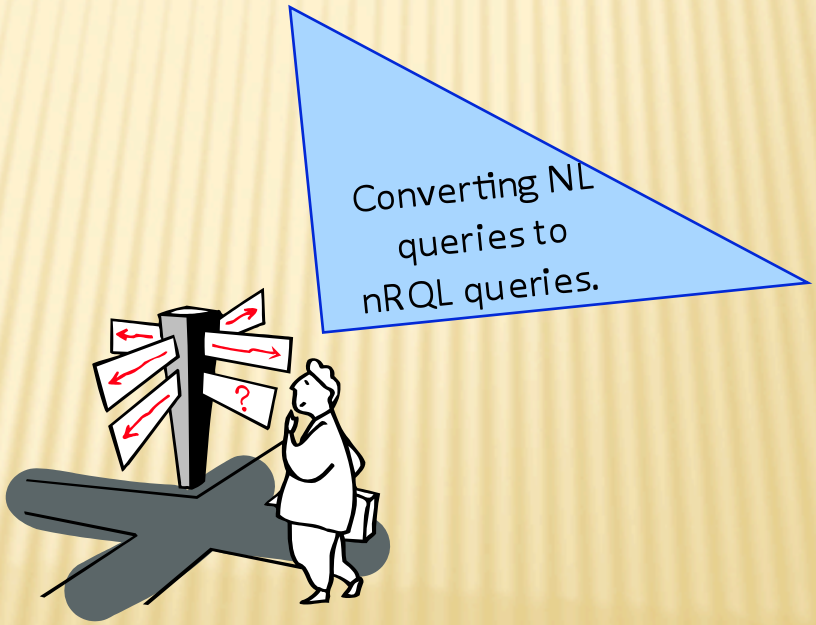
1. INTRODUCTION

Problems to overcome:

- The vocabulary gap;
- The interpretation of natural language queries into the nRQL query language;
- A faithful matching ;
- Ranking the different nRQL queries generated by the system.

QUESTION-ANSWERING

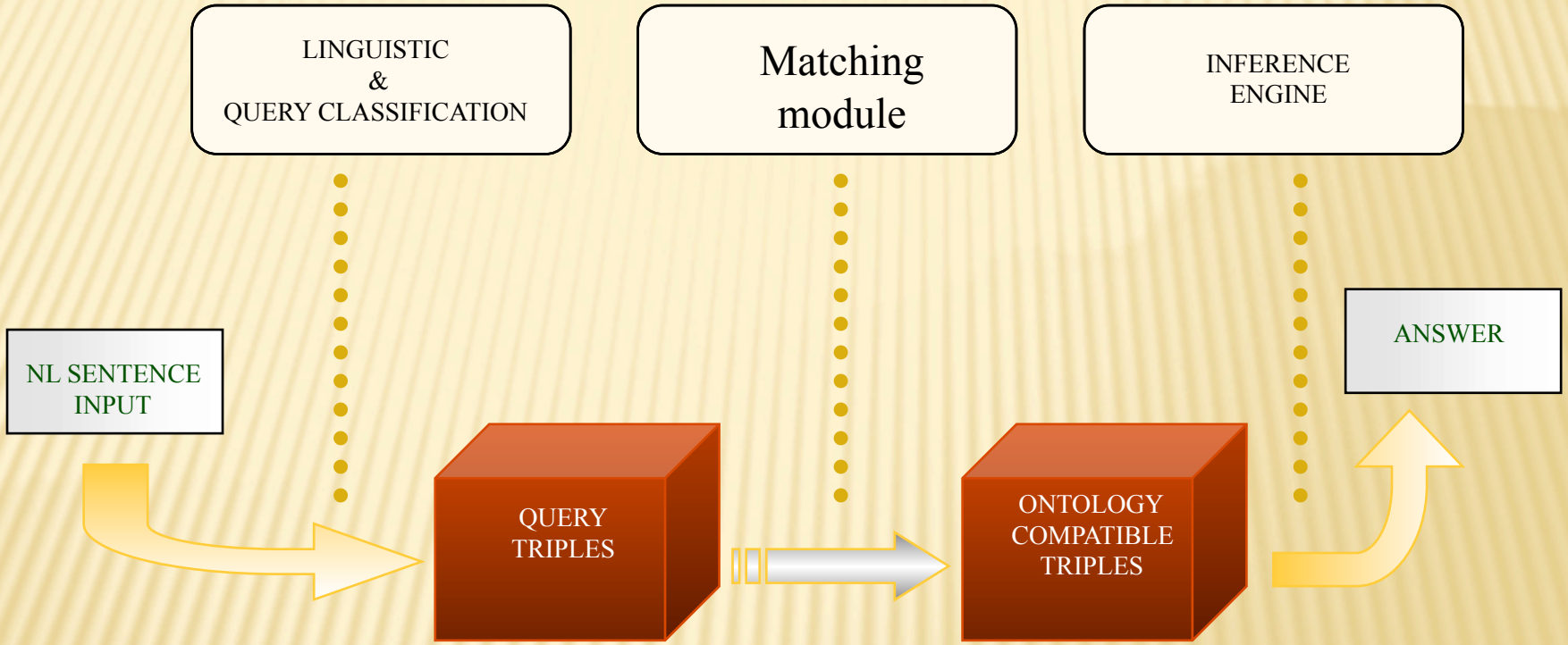
- *The development of a natural language interface for querying the ontology : a front end to RACER(Renamed ABox and Concept Expression Reasoner).*



4.RELATED WORKS

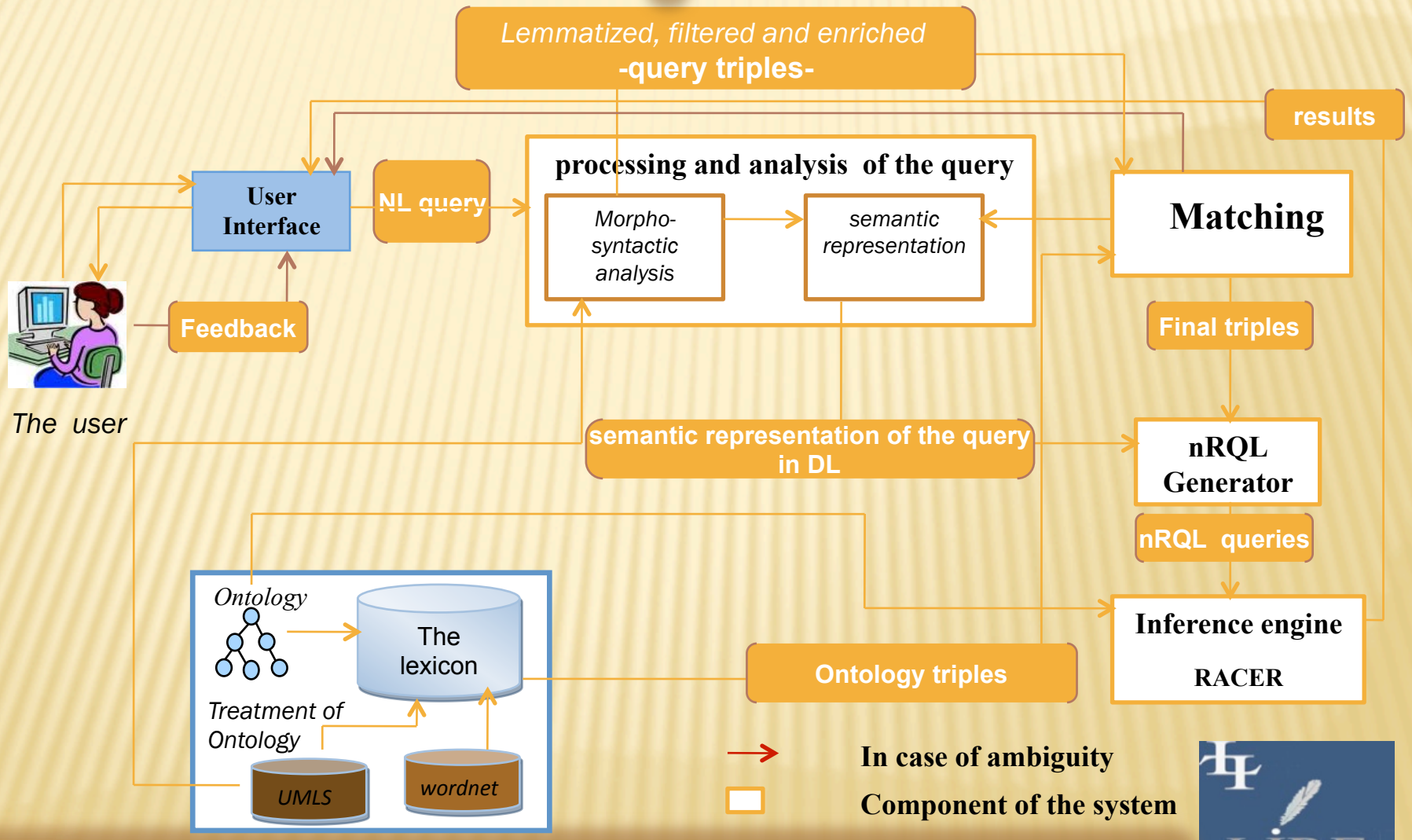
related works	Characteristics of the approach										
	ontology		process		NLP			User query			
	One ontoology	O.D.B	semi	auto	morphologic	syntactic	semantic	Keywords	restricted NL	Free NL	semantic representation
QUESTIO (V. Tablan)	×			×	×	×				×	Triples
PANTO (C. Wang)	×			×	×	×	×		×		Triples
ONLI (L. Kosseim)	×			×	×		×		×		Triples
AquaLog (V.Lopez)	×			×	×	×	×			×	Triples
PowerAqua (V.Lopez)		×		×	×	×	×			×	Triples
SemSearch (Y.lei)		×	×				×	×			Triples

5. THE PROPOSED APPROACH:

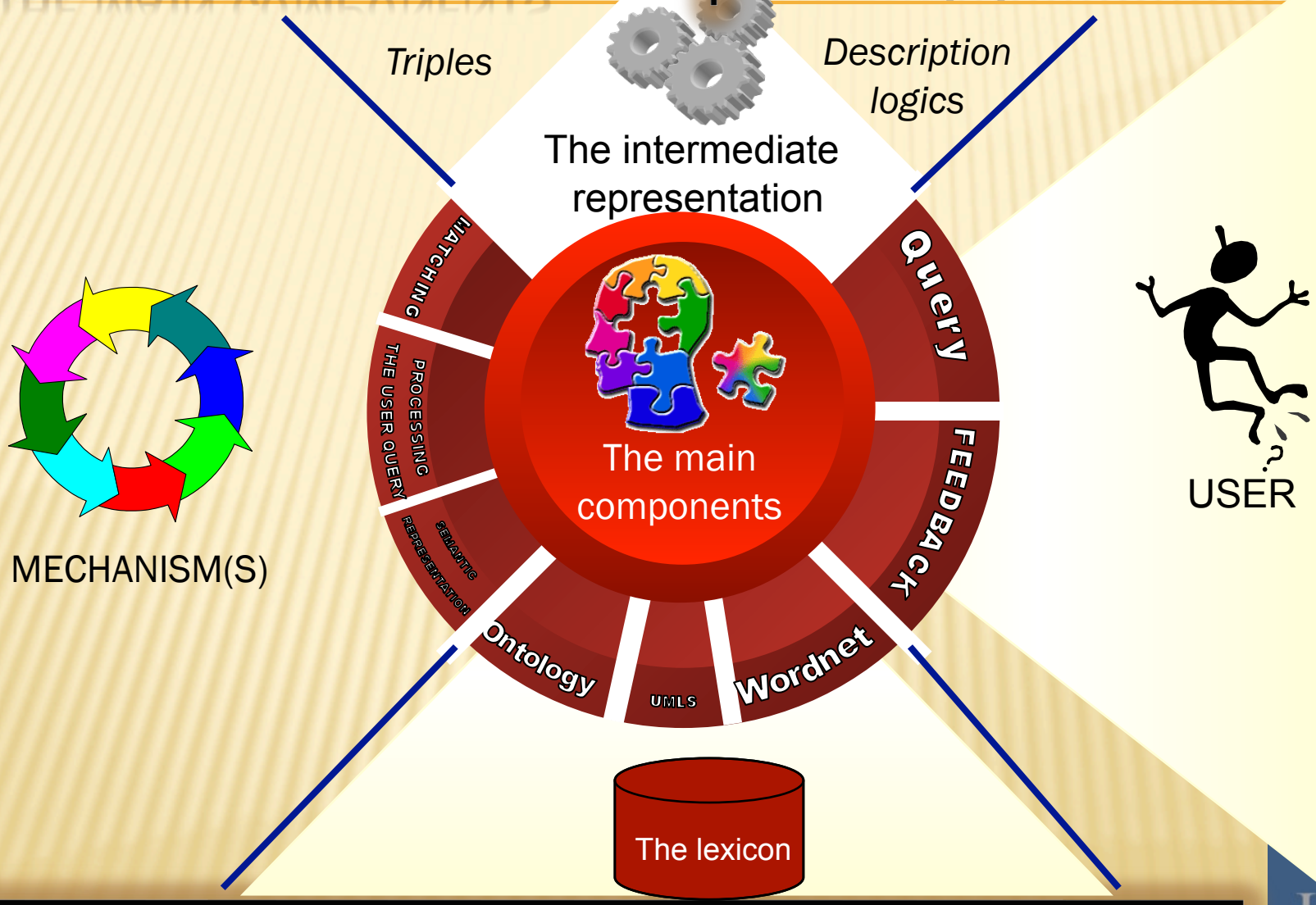


Intermediate triples: <subject, predicate, object>

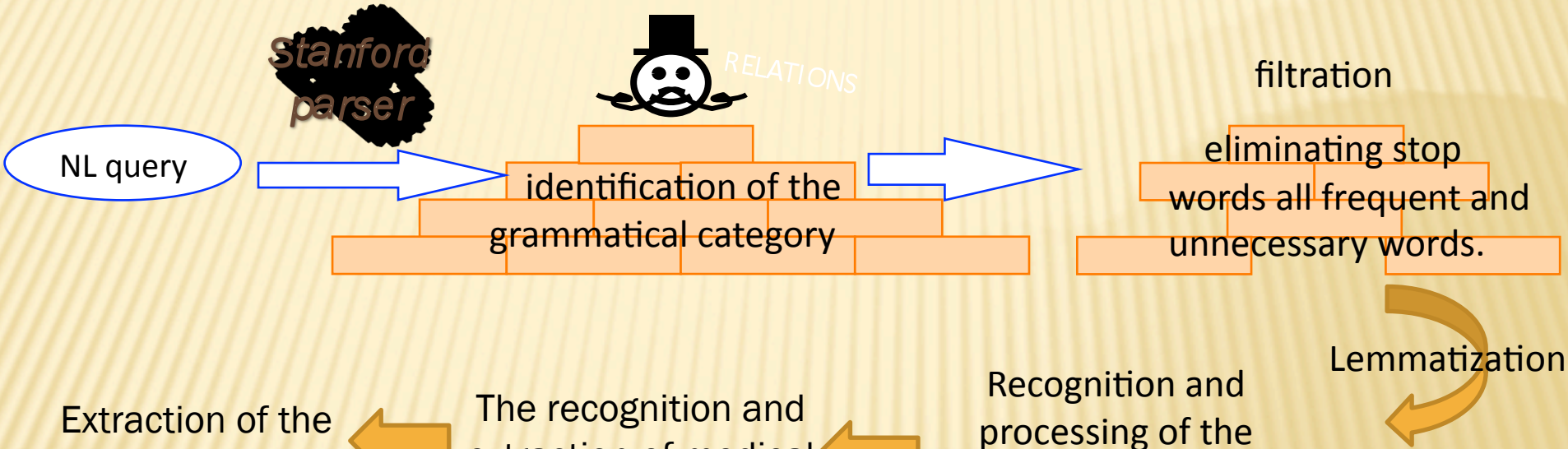
6. ARCHITECTURE OF THE PROPOSED SYSTEM



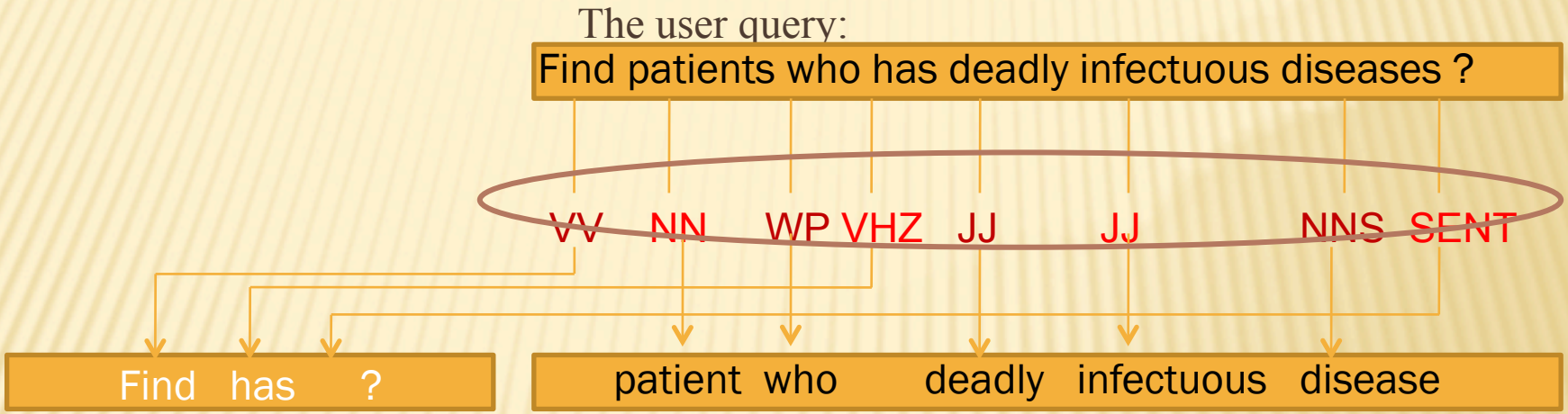
THE MAIN COMPONENTS



PROCESSING & ANALYSING OF THE USER QUERY



EXAMPLE

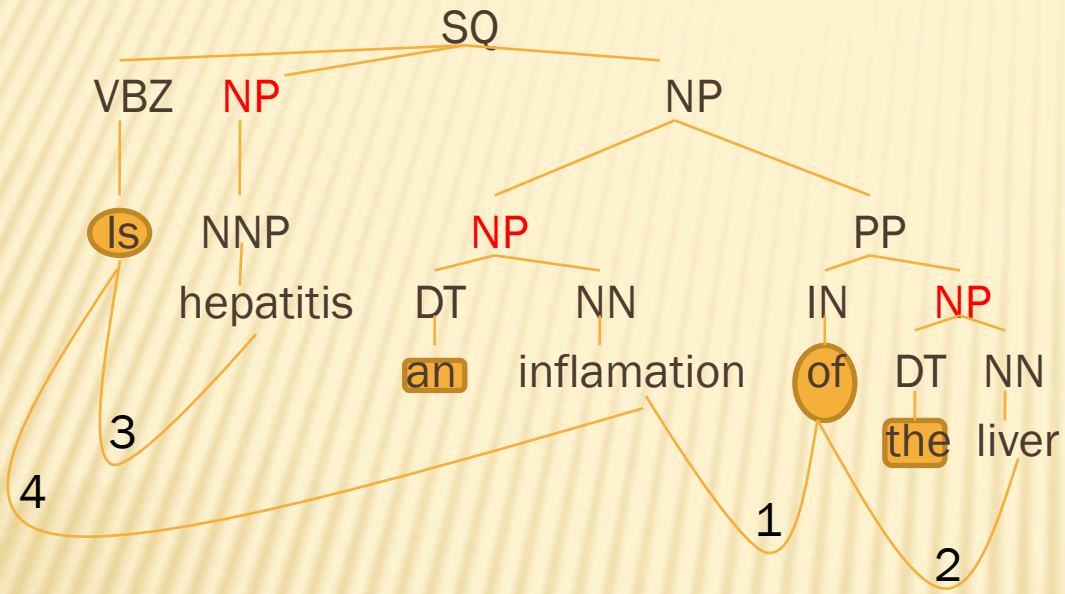


Filtration

Lemmatized words:

VV: verb, basic form
NN: noun, singular or mass
WP: wh-pronoun
VHZ: verb *have, pres 3rd per.sing*
JJ: adjective
NNS: noun plural
SENT: end punctuation

EXTRACTION OF THE INTERMEDIATE REPRESENTATION



Stop word

NP Elementary Noun Phrase

Extraction of the intermediate representation: R (S,O)

Step1: recognition and extraction of prepositional phrases: PN+PP +PN=PP(PN,PN)

Step 2: recognition and extraction of verb phrases: PN+PV+PN=PV(PN,PN)

EXTRACTION OF THE INTERMEDIATE REPRESENTATION

IWAISE'12

Step3: In the case where the relation is a noun phrase;

1- Identify the noun phrases.

2- Group the elementary NP in pairs.

3- Use the semantic network of UMLS, the semantic types source and target for each potential relation.

Step4: Extraction of the relations of exception, (eg, nounnoun, or noun-adj).;

1 - In the case of adjective + noun: be (adjective, noun) as (yellowfever, be, fever)

2 - In the case of noun + noun, if they are not compound nouns, then the triple will be as (macrocytic, anemia)

EXTRACTION OF THE INTERMEDIATE REPRESENTATION

After the identification of the differents triples composing the user query:

Step5 : coordination

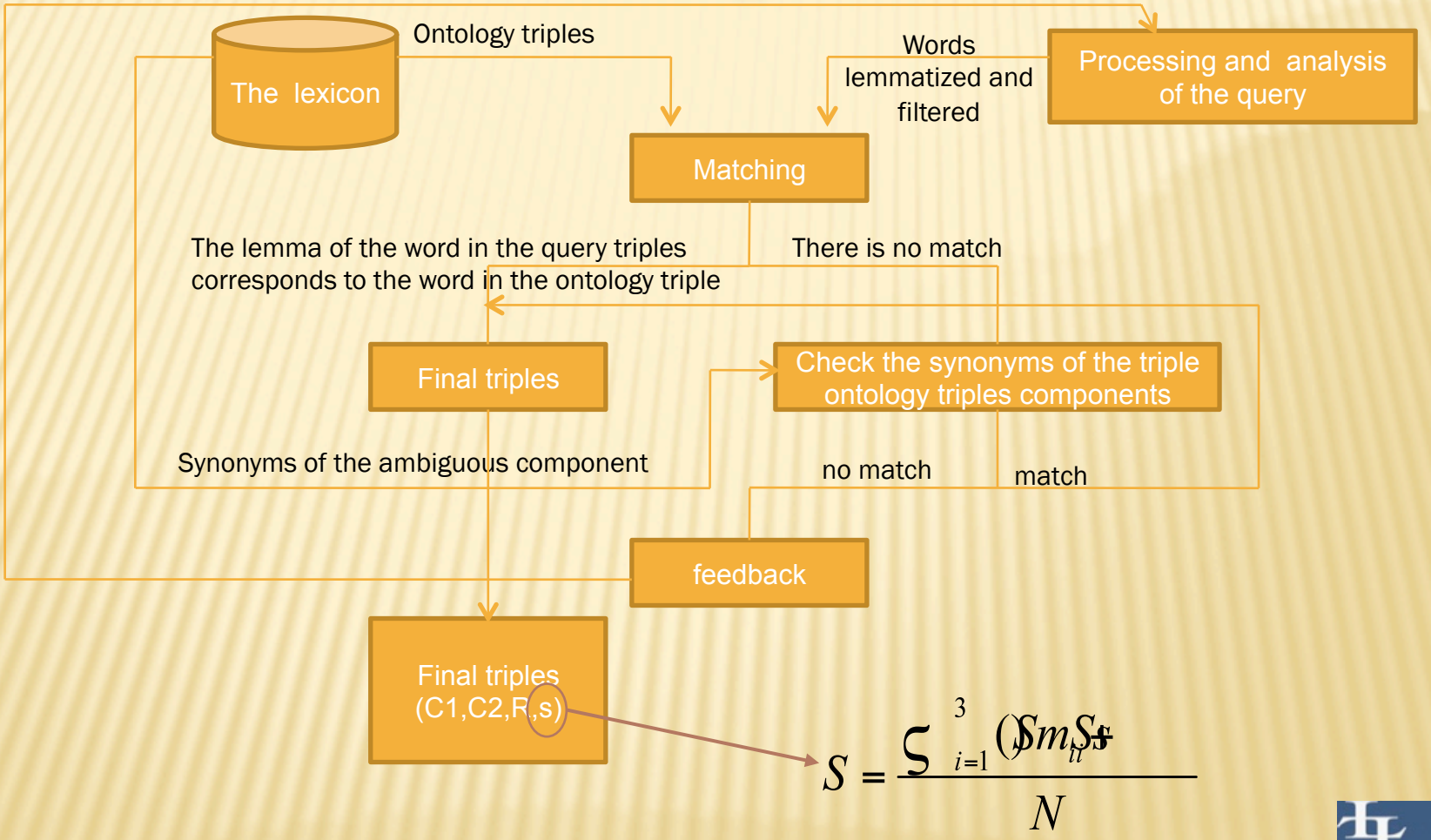
-If the coordination (and / or) is between two separate sentences :
it will be used to connect the triplets they contain.

-If elementary phrases are sharing the same grammatical category:
- then they will share the two remaining components of the same triplet as Ahmed and Ali suffer from cirrhosis (the two triples elements ahmed and ali are sharing the same relation “suffer from” and the same concept”cirrhosis”).

$Q = T1 \wedge T2 \wedge T3 (/) \ldots$

Q: is user the query,
T: the corresponding triple.

MATCHING MODULE



$$S = \frac{\sum_{i=1}^{3} (Sm_{it} S_t}{N}$$

GENERATING THE INTERMEDIATE SEMANTIC REPRESENTATION

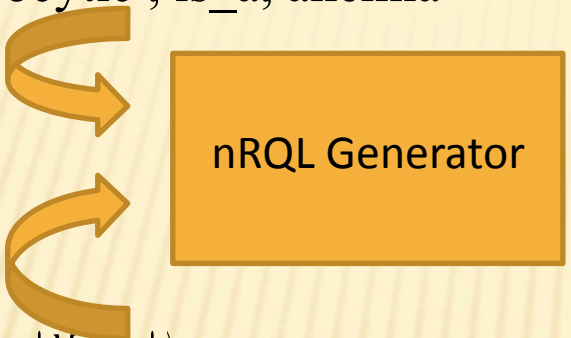
	Production rules	*Semantic association*
1.	P→ SN + SV	$[SN] \sqcap [SV] \neq \phi$
2.	P→ NP + SV	$[NP] \sqsubseteq [SV]$
3.	SN →Art + N	$[SN] = [N]$
4.	SN→ Art + N + Adj	$[SN] = [N] \sqcap [Adj]$
5.	SV→ VT + SN	$[SV] = [VT]:[SN]$

$\top^{\mathcal{I}} = U$
$\bot^{\mathcal{I}} = \phi$
$(\neg A)^{\mathcal{I}} := (A^{\mathcal{I}})'$
$(\neg C)^{\mathcal{I}} = (C^{\mathcal{I}})'$
$(C \sqcap D)^{\mathcal{I}} = C^{\mathcal{I}} \sqcap D^{\mathcal{I}}$
$(C \sqcup D)^{\mathcal{I}} = C^{\mathcal{I}} \sqcup D^{\mathcal{I}}$
$(\exists r.C)^{\mathcal{I}} = r^{\mathcal{I}} : C^{\mathcal{I}}$
$(\forall r.C)^{\mathcal{I}} = (r^{\mathcal{I}} : (C^{\mathcal{I}})')'$

The semantic tree

GENERATING THE NRQL QUERY

<patient,suffer_from,anemia>,
< AnemiaMacrocytic , is_a, anemia>



(Retrieve (?x) (and (?x patient anemia suffer_from) (AnemiaMacrocytic anemia is_a)))

$Patient \sqcap \exists suffer\_from.(anemia \sqcap AnemiaMacrocytic)$

Each query will have as a score the sum of the different scores of triples that compose it:

SQ: Query Score
n: is the number of triplets that compose the user's query.
S: is the score of each triple composing it

$$SQS = \sum_{i=1}^{n} i$$

5.CONCLUSION & PERSPECTIVES

Generally, conventional methods in information retrieval perform limited linguistic processing. So, we propose :

- an approach which takes account, in a better manner, of the linguistic nature of the processed queries, when extracting the triple-based intermediate representation
- a formal interpretation of triples to description logics formalism.

This semantic representation of the query is well adapted to the semantic search needs and provides more significant improvements to the results than question answering systems.

- Overcoming the limitation of our system to a single ontology.
- Enhancing the performance of the proposed system.

QUESTIONS?

THANK YOU

bouhalika.chouaib@yahoo.fr