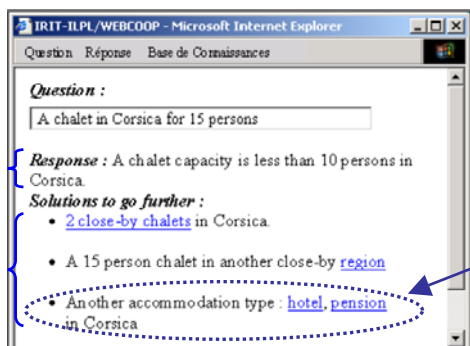


# An Approach for Evaluating Cooperative Question Answering Systems

Farida Aouladomar and Farah Benamara,  
 Institut de Recherches en Informatique de Toulouse, IRIT, France.  
 aouladom@irit.fr, benamara@irit.fr

## I – FRAMEWORK : THE WEBCOOP SYSTEM

WEBCOOP integrates knowledge representation and advanced reasoning procedures to generate Cooperative Responses (CR) to NL queries on the web, in French (no user model).



NL generation of the know-how part is based on under-specified templates T:

$T = \langle \text{Specified terms, lexicalisation Function, logical formulas representing the Remainder of the question} \rangle$

**Template 1:**

*un autre type de lexicalisation (mother\_node) : lexicalisations(sister\_nodes) R*

**Answer 3 :**

*another accommodation type : hotel, pension in Corsica*

## II – EVALUATION GRID OF WEBCOOP OUTPUTS

### I. Within a domain : tourism

#### Intra-Templates evaluation

- (1) adequacy of the responses
- (2) justification and explanation mechanisms
- (3) specified terms and different lexicalisation choices within each template
- (4) adequacy of the hyperlink generation strategy.

#### Inter-Templates evaluation

- (5) display order strategy in responses.
- (6) general fluency (syntactical regularities of responses generated by each template)
- (7) visual aspect of the responses: enumeration vs. paragraph.

### II. Template Portability : over other large public domains such as health or education

## III – EVALUATING RESPONSE ORDER (point 5 above)

**Ordering strategy :** the answer presentation is guided by the inverse reading order of the question (the first constraint is always the last to be relaxed)

**Aim :** identifying the different ordering rules that users would employ

**Task 1 : Subjects have to order answers and to justify their choice (5 QA)**

**Question :**  
 Je voudrais réserver un hôtel 3 étoiles à Monaco

**Réponse du système :**  
 Il n'y a pas d'hôtel trois étoiles à Monaco.

**On vous propose :**

- 1-
- 2-
- 3-

**Responses to order :**

**R -** des hôtels à Monaco dans une autre catégorie d'hôtel : 2 étoiles, 4 étoiles

**F -** un autre type d'hôtellerie 3 étoiles à Monaco : résidence hôtelière, résidence de tourisme

**G -** des hôtels 3 étoiles dans une autre ville proche de Monaco : Nice, Menton

What are the aspects of the question that led you to this order rather than another?  
 .....

**30 Subjects**



**Task 2 : Subjects have to answer the question below**

*In your opinion, is it important to order answers in QA systems ?*

Yes  No

**Explain :** .....

#### Results for task 2:

Question asked in task 2	Percent of subjects choices
Responses have to be ordered	53 %
Responses do not have to be ordered	47%

#### Results for task 1:

The different ordering rules extracted

Response order	Percent of justifications
depends on syntactic criteria	23%
depends on pragmatic criteria	64%
independent from any constraint order	13%

## IV – PERSPECTIVES

- identify a threshold beyond which choices ordering is needed
- characterize the pragmatic rules (depending on the nature of the constraints)