

*Golfred :*  
Vers la génération de récits d'expérience pour des  
robots de service

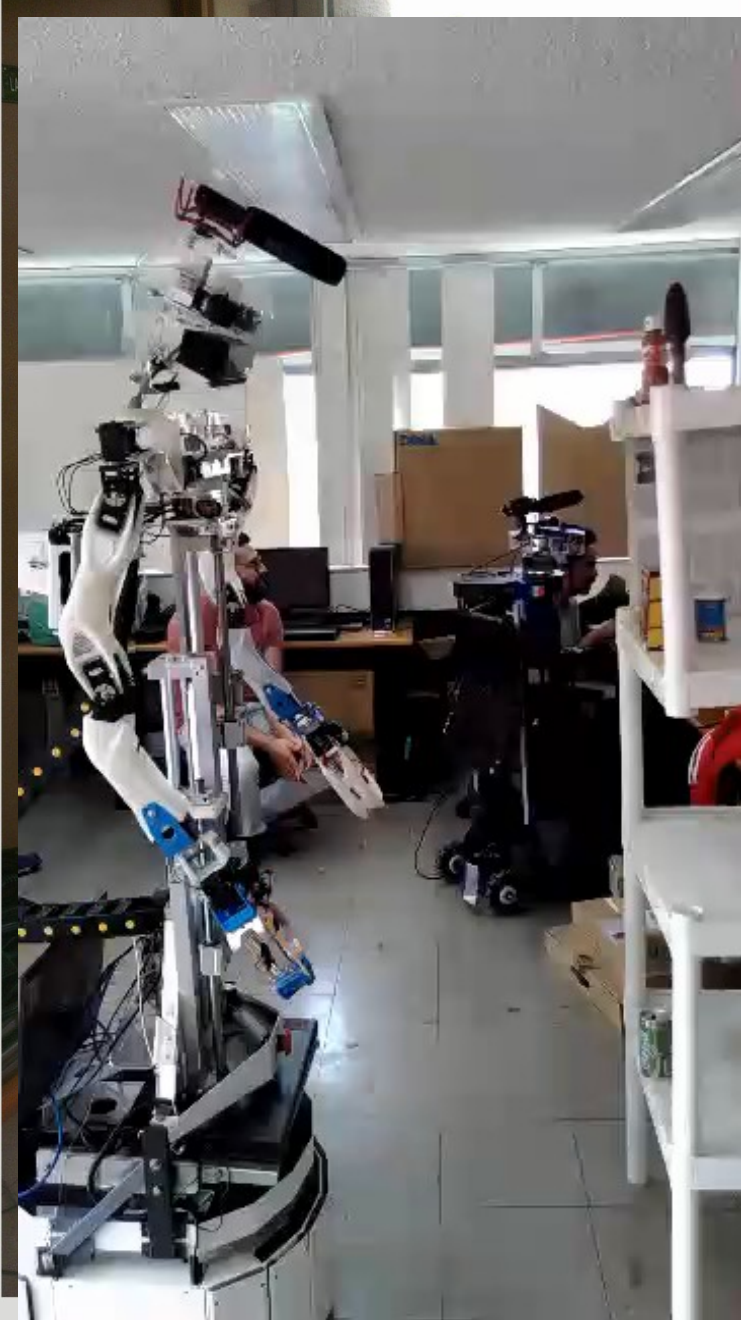
**Jorge Garcia Flores**  
**LIPN – CNRS – Paris 13**

**Claire Gardent**  
**LORIA - CNRS**

**Aldo Gangemi**  
**LIPN – CNRS – Paris 13**

**Iván Meza, Luis Pineda**  
**IIMAS - UNAM**





- Golem III: a service robot
  - Dialogue system
  - Cognitive architecture
  - Command interpreter
- 3d place in Robocup@home 2012



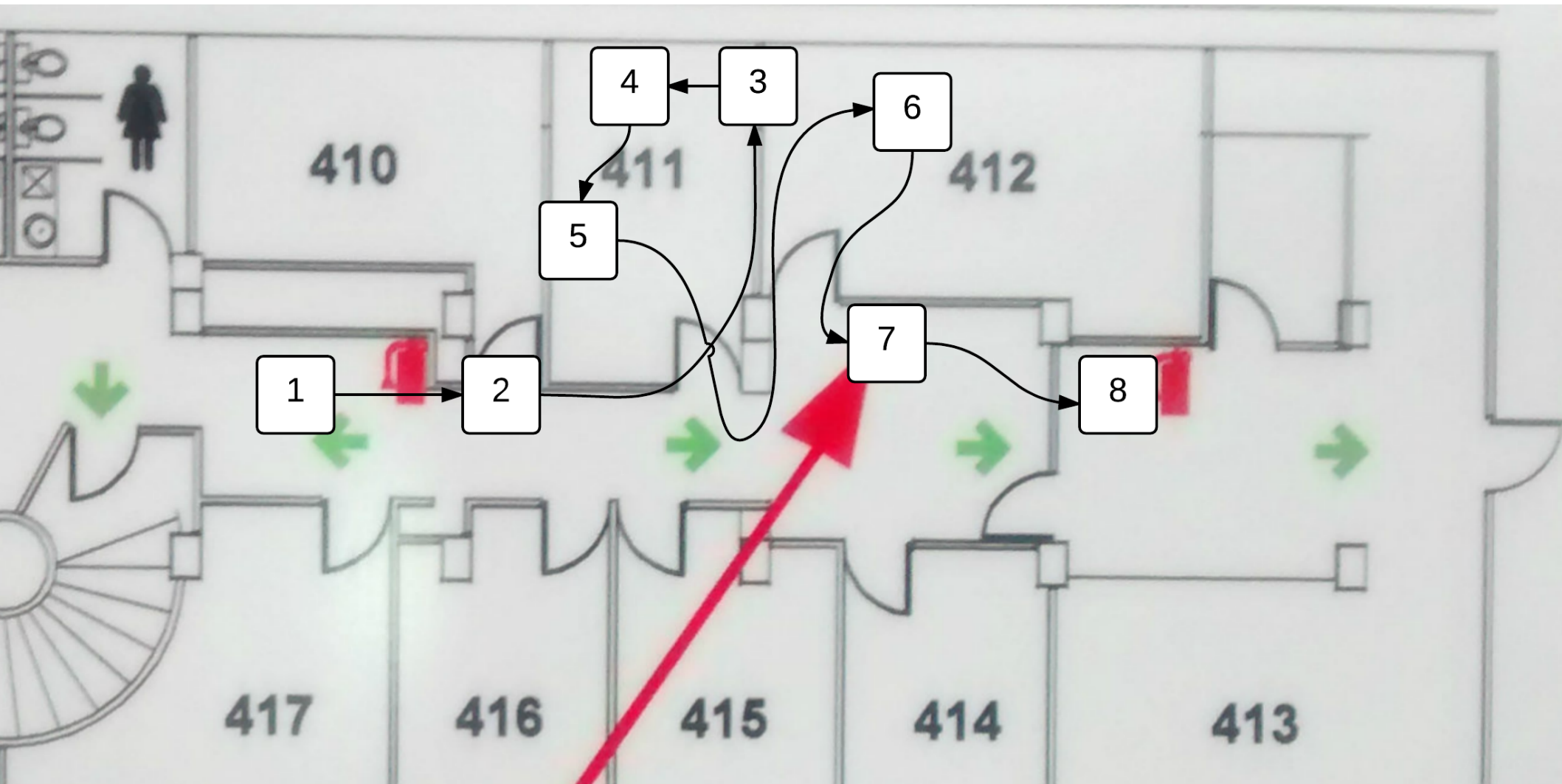
2015

2012

2006

- Luis A. Pineda, Arturo Rodríguez, Gibran Fuentes, Caleb Rascon and Ivan V. Meza. *Concept and Functional Structure of a Service Robot*. *International Journal of Advanced Robot Systems*, 2015, 12:6.
- Luis Pineda, *The Golem Group: The Golem Team, RoboCup@Home 2012*. *Proceedings of Robocup 2012*. vol , pp . 2012.

# Récits d'expérience



## Récits d'expérience



**"I was at Iván V. Meza Ruiz office. Iván is a researcher at IIMAS, UNAM working at the Department of Computer Science with the Golem group...."**



# Récits d'expérience

"In his office, I read on the wall the phrase **Semantic Textual Similarity**. Semantic similarity is a metric defined over a set of documents or terms, where the idea of distance between them is based on the likeness of their meaning or semantic content as opposed to similarity which can be estimated regarding their syntactical representation."

**SOPA: Random Forests Regression for the Semantic Textual Similarity task**  
Davide Buscaldi<sup>1</sup>, Jorge J. García Flores<sup>1</sup>, Ivan V. Meza<sup>2</sup> and Isaac Rodríguez<sup>2</sup>  
<sup>1</sup> Laboratoire d'Informatique de Paris Nord, CNRS (UMR 7030), Université Paris 13, Sorbonne Paris Cité, Villetaneuse, France  
<sup>2</sup> Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas (IIMAS), Universidad Nacional Autónoma de México (UNAM), Ciudad Universitaria, DF,

**Results**

	Wikipedia	News	Overall
LIPN-RF	0,5637	0,5655	0,5649
LIPN-MLP	0,25257	0,5342	0,4401
LIPN-SVM	0,4194	0,4007	0,4069

Spanish results

	Wikipedia	News	Overall
LIPN-RF	0,7243	0,8123	0,8414
LIPN-MLP	0,6886	0,8121	0,8184
LIPN-SVM	0,7028	0,7985	0,8104

English results

	Wikipedia	News	Overall
LIPN-RF	0,6178	0,5864	0,7175
LIPN-MLP	0,5918	0,5864	0,7175
LIPN-SVM	0,5918	0,5864	0,7175

Comparison of the results obtained with corpus selection and using the full corpus.

**Learning models**

We used v-Support Vector Regression (v-SVR) (Schölkopf et al., 1999), Multilayer Perceptron (Bishop et al., 1995) and Random Forests (Breiman, 2001). In our runs, we selected the training set according to a similarity measure between the test and the training set based on a 1- to 3-grams language model and average sentence length. The idea behind this selection process is that learning sentence similarities on a specific type of text will increase the accuracy of predictions on text with similar characteristics: image descriptions are usually written in a very different form than word definitions or forum answers. The complete English training set was composed by the data from SemEval STS 2012, 2013 and 2014. In Spanish, we used our 2014 training set and a corpus we made from RAE (Royal Spanish Academy dictionary) definitions, and the 2014 Spanish test set.

**4. Information Retrieval-based Similarity**

We consider an IR system as a document collection and a query. The similarity measure is based on the assumption that  $p$  and  $q$  are similar if the documents retrieved by  $S$  for the two texts, used as input queries, are similar. Let  $h_p = \{d_1, \dots, d_n\}$  and  $h_q = \{d_1, \dots, d_m\}$  be the top  $n$  and  $m$  documents retrieved by  $S$  for the query  $p$  and  $q$ , respectively. Let us define  $f_p$  and  $f_q$  the scores assigned by  $S$  to a document  $d$  for the query  $p$  and  $q$ , respectively. Then, the similarity score is calculated as:

$$\text{sim}_{IR}(p, q) = 1 - \frac{\sum_{d \in h_p \cap h_q} f_p(d) \cdot f_q(d)}{\sum_{d \in h_p \cup h_q} f_p(d) \cdot f_q(d)}$$

**5. N-gram Based Similarity**

This measure tries to capture the fact that similar sentences have similar n-grams, even if they are not placed in the same post.

$$\text{sim}_{NG}(p, q) = \frac{\sum_{n=1}^N \sum_{i=1}^{|p|} \sum_{j=1}^{|q|} \mathbb{1}(p[i:i+n] = q[j:j+n])}{\sum_{n=1}^N \sum_{i=1}^{|p|} \sum_{j=1}^{|q|} \mathbb{1}(p[i:i+n] \neq q[j:j+n])}$$

Where  $P$  is the set of the n-grams.

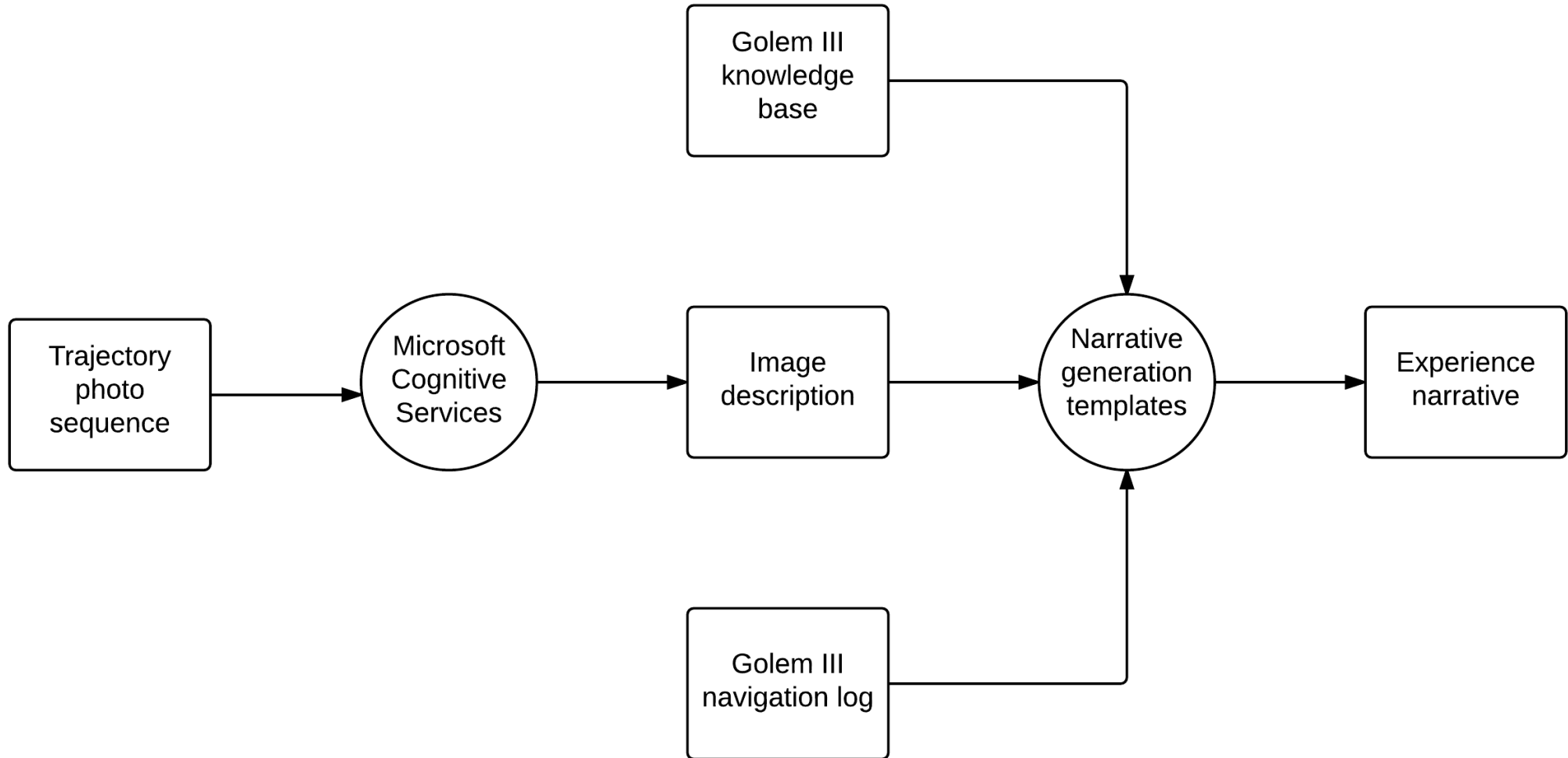
**7. Word Alignment Similarity**

This similarity metric is based on the word alignment between the two sentences. We built a word alignment model using the word co-occurrence matrix and the word alignment algorithm proposed by (García Flores et al., 2014). The word alignment similarity is calculated as follows:

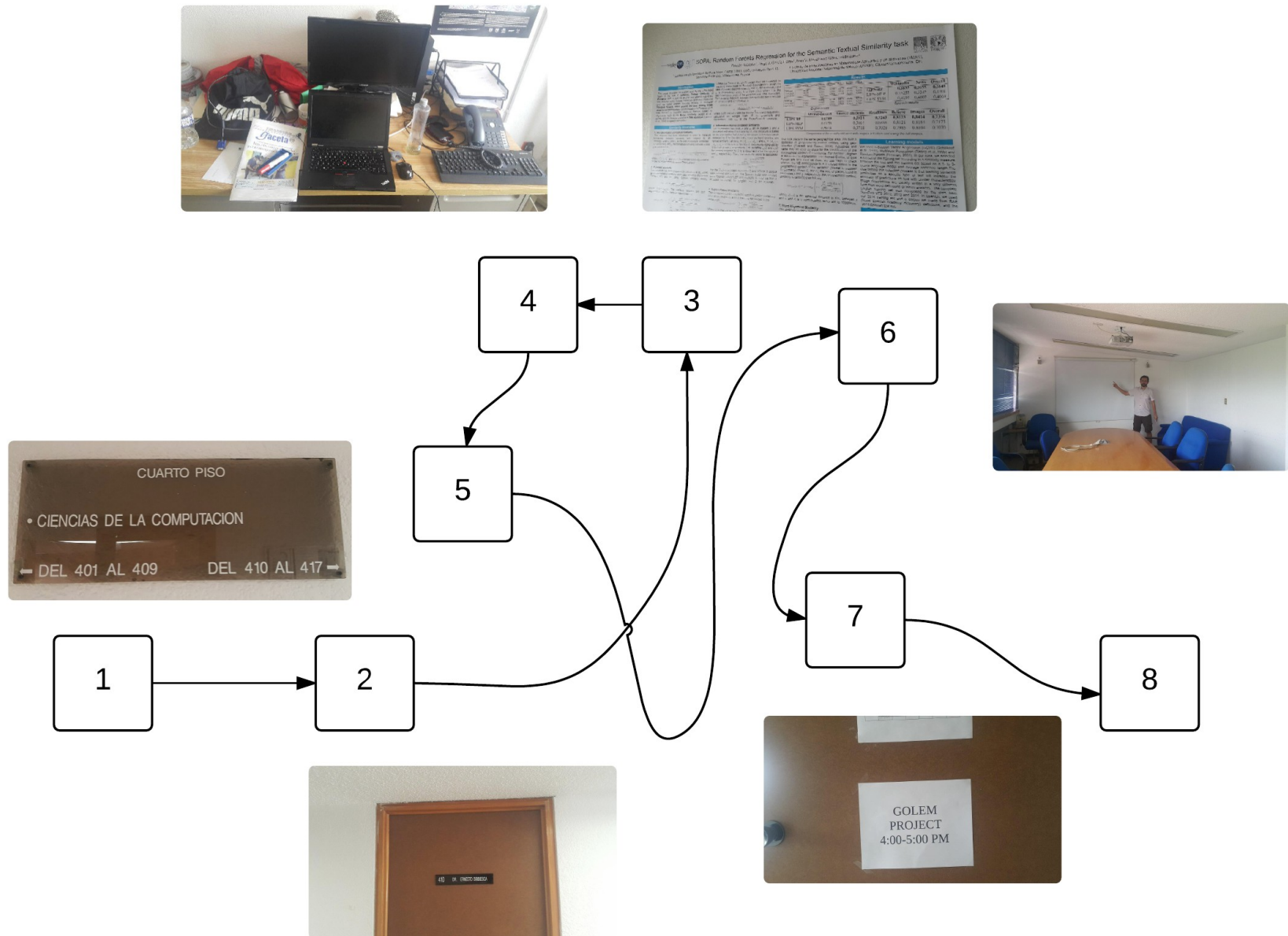
$$\text{sim}_{WA}(p, q) = 1 - \log_K \left( 1 + \frac{\sum_{i=1}^{|p|} \sum_{j=1}^{|q|} d(x_i, y_j)}{\max(|G_p|, |G_q|)} \right)$$

where  $d(x, y)$  is the spherical distance in Km. between  $x$  and  $y$ , and  $K$  is a normalization factor set to 10000Km.

# Architecture (itération 0)



# Trajectoire en images




# Description des images

Memories - Golfred - Chromium

Memories - Golfred x


← → ↻ golem.iimas.unam.mx/golfred/edit/experience/8ace1242-1ef



analysis read

Read ↻

Analysis ↻




```
{"metadata": {"width": 5312, "format": "Jpeg", "height": 2988},
"requestId": "5c849008-ec47-4044-ac87-a12eb6e91ab7",
"description": {"captions": [{"text": "a cluttered desk with a laptop", "confidence": 0.9065554787673039}], "tags": ["indoor", "table", "computer", "laptop", "coffee", "items", "desk", "sitting", "small", "topped", "cluttered", "top", "wooden", "keyboard", "covered", "phone", "bed", "white", "room"]}, "categories": [{"score": 0.0625, "name": "others_"}, {"score": 0.00390625, "name": "outdoor_"}], "tags": [{"confidence": 0.9680249691009521, "name": "indoor"}, {"confidence": 0.7372207641601562, "name": "computer"}, {"confidence": 0.5437814593315125, "name": "items"}, {"confidence": 0.2664460241794586, "name": "cluttered"}, {"confidence": 0.1823427826166153, "name": "desk"}]}
```



Memories - Golfred

golem.iimas.unam.mx/golfred/edit/experience/8ace1242-1ef



analysis

Read


Analysis

```
{"metadata": {"width": 5312, "format": "Jpeg", "height": 2988}, "requestId": "2b968932-1108-492b-a0ad-f7394c35fb64", "description": {"captions": [{"text": "a man standing in a room", "confidence": 0.7833076051181184}], "tags": ["indoor", "ceiling", "man", "standing", "person", "room", "woman", "bed", "young", "holding", "bedroom", "shirt", "posing", "wearing", "top", "table", "air", "suitcase", "hat", "living", "luggage", "blue"]}, "categories": [{"score": 0.015625, "name": "abstract_"}, {"score": 0.015625, "name": "others_"}], "tags": [{"confidence": 0.9999456405639648, "name": "wall"}, {"confidence": 0.9988609552383423, "name": "indoor"}, {"confidence": 0.9841926693916321, "name": "ceiling"}, {"confidence": 0.9046129584312439, "name": "man"}, {"confidence": 0.891359269618988, "name": "standing"}, {"confidence": 0.8507639169692993, "name": "person"}, {"confidence": 0.49494460225105286, "name": "room"}]}
```

Event

move

```
{"command": {"args": ["p7"], "name": "move", "type": "move"}}
```



analysis

read

Read

Analysis

```
{"metadata": {"width": 5312, "format": "Jpeg", "height": 2988}, "requestId": "67ca810c-3085-45bd-81d2-06f1c2df752b", "description": {"captions": [{"text": "two giraffes in a fenced in area", "confidence": 0.004919304795390664}], "tags": []}, "categories": [{"score": 0.00545685, "name": "abstract_"}, {"score": 0.00545685, "name": "others_"}], "tags": [{"confidence": 0.00545685, "name": "abstract_"}, {"confidence": 0.00545685, "name": "others_"}]}
```

Memories - Golfred

seminaire.lipn.jun

ijcai-2016 (9).pdf

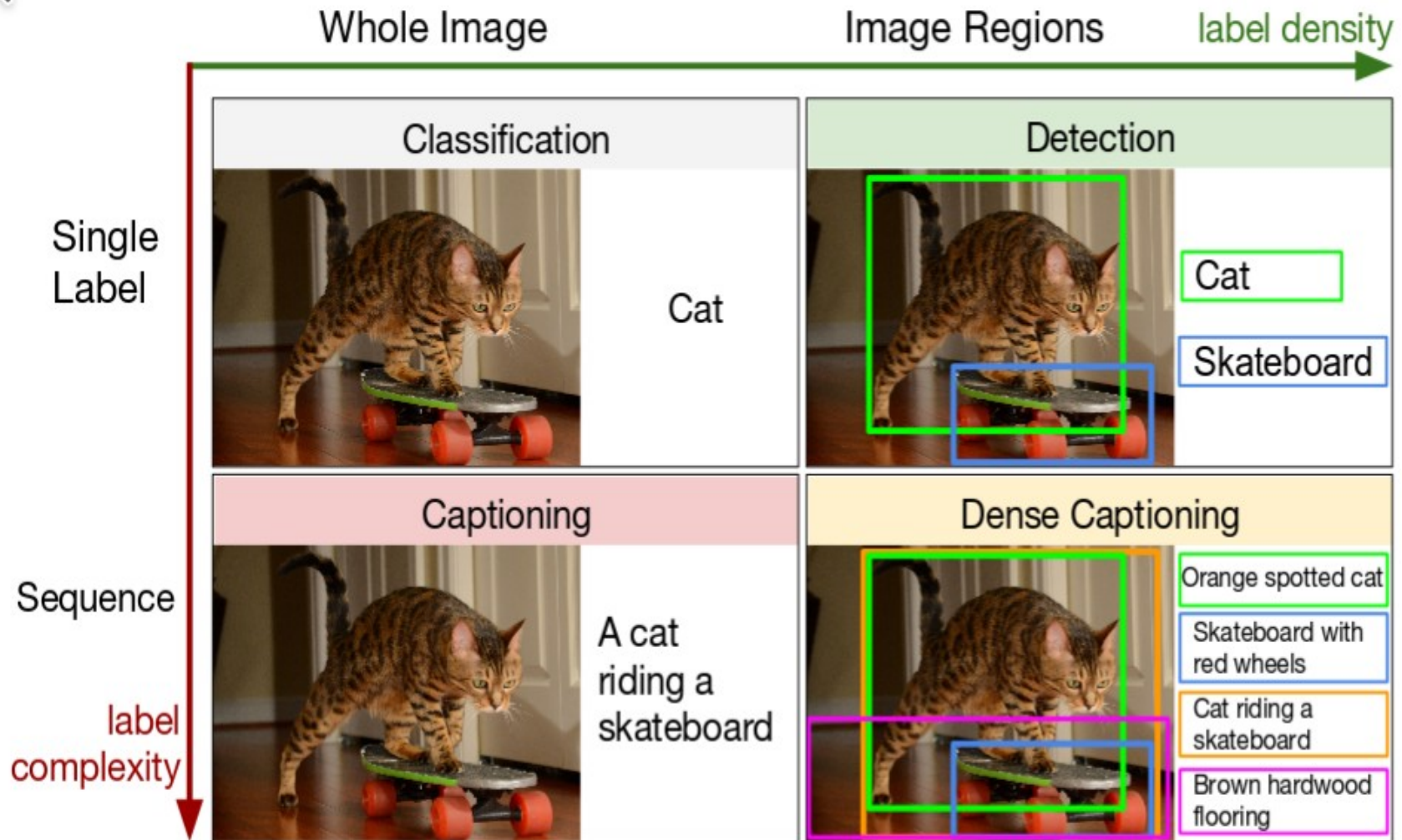
Telegram (27)

Courrier entrant

Préférences de Th

Éditeur d'image

# Description des images



**DenseCap: Fully Convolutional Localization Networks for Dense Captioning.** Johnson, Justin, Karpathy, Andrej and Fei-Fei, Li (2016), In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

# Description des images

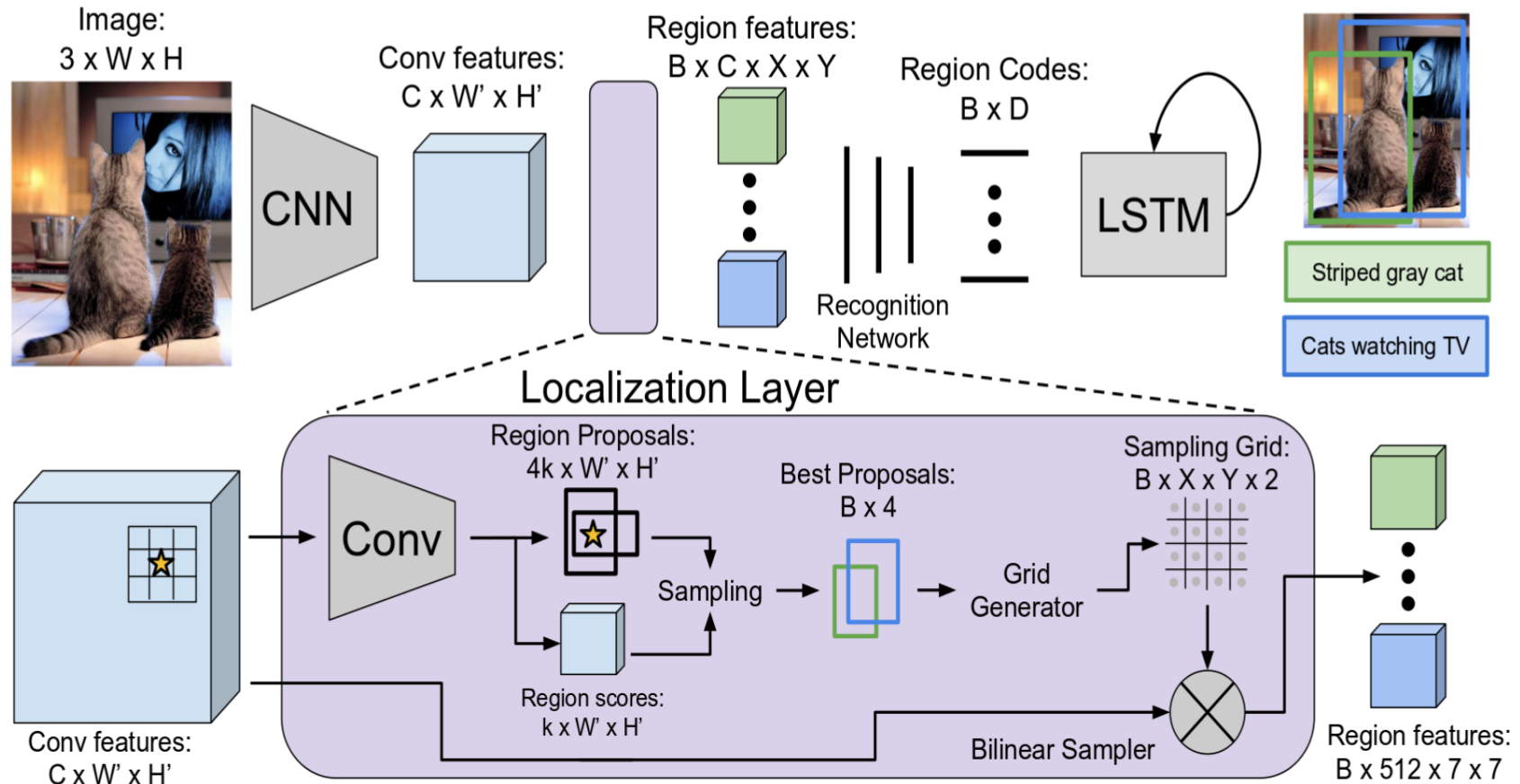
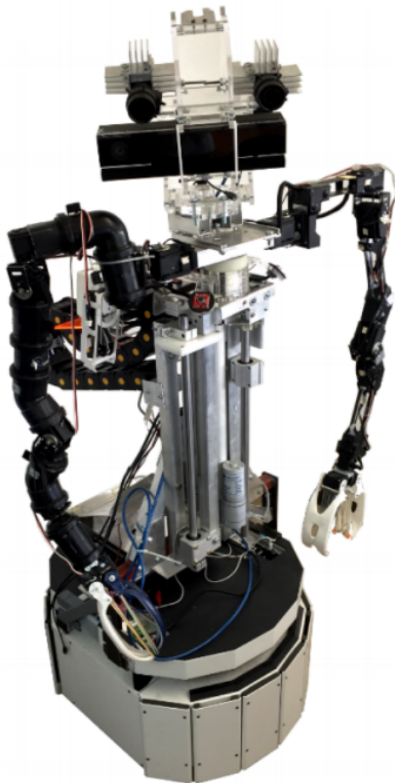
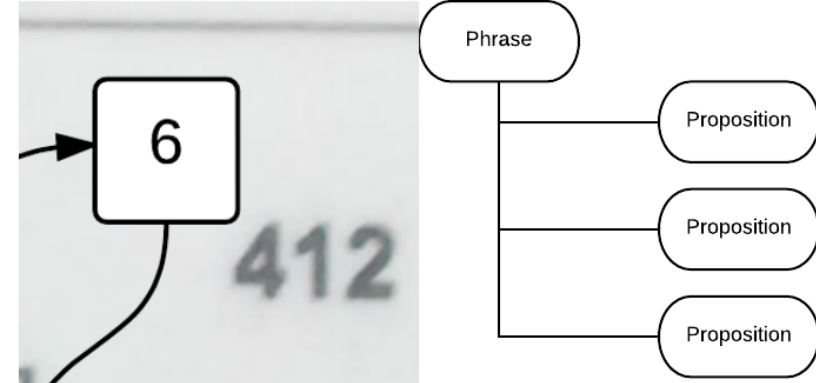
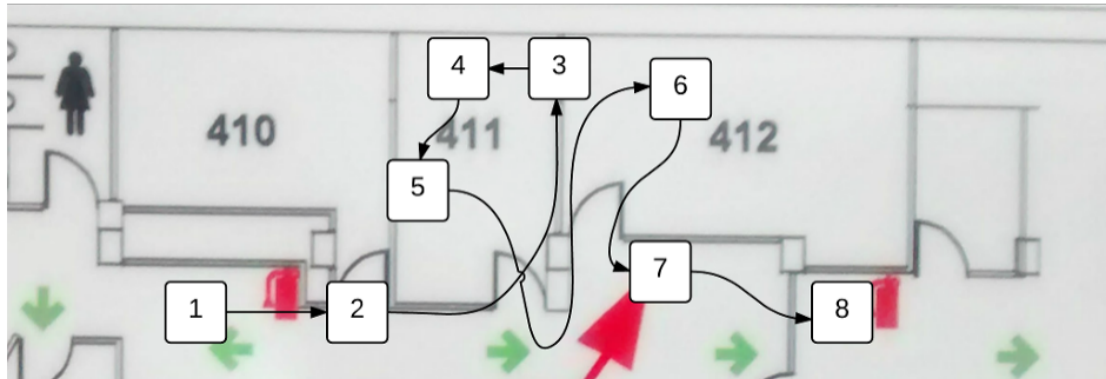


Figure 2. Model overview. An input image is first processed a CNN. The Localization Layer proposes regions and smoothly extracts a batch of corresponding activations using bilinear interpolation. These regions are processed with a fully-connected recognition network and described with an RNN language model. The model is trained end-to-end with gradient descent.

# Génération spatiale du discours



start(loc1)

Random  
template draw

*"I woke up at the hall..."*

go(loc2)

Random  
template draw

*"afterwards I went to..."*

go(locN)

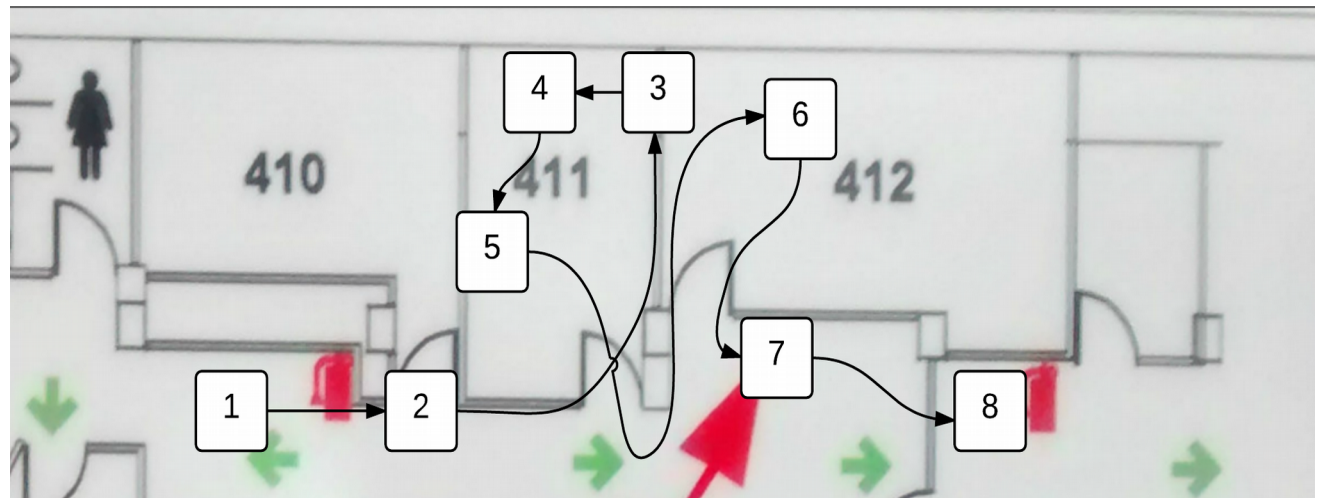
read(text1)

Random  
template draw

*"there was a sign on the  
door, it said..."*



# Expérience *in vitro*



Point	Instruction
1	Start and read the sign
2	Get into Mr. Bribiesca's office and move the water bottle
3	Get into Mr. Meza's office and read the poster
4	Take the water bottle from the desk under the poster
5	Leave the water on the desk behind and describe what's on it
6	Go to the meeting room and describe the scene
7	If the meeting room is busy, look for a sign on the door and read what's going on
8	Go to the lab and stop



## Robot-v0:

I begun at in the hall, I saw a a sign on the door, I read it said COMPUTER SCIENCE DEPARTMENT. Then, I tried to, I walk towards the hall there, but couldn't, I noticed a clock mounted on a wall, it said DR. ERNESTO BRIBIESCA. Then, I went to the office of Ivan there, I saw a a picture of a tv, I read it said SOFA: Random Forests Regression for the Semantic Textual Similarity task, I grabbed a bottle, I moved to the desk there, I dropped the bottle, I noticed a cluttered desk with a laptop, it said sausenc &aceta para consejeros - ASO. When finish, I walk towards the meeting room there, I noticed a man standing in a room, I moved to the door there, I noticed two giraffes in a fenced in area, it said GOLEM. At this point, I went to the laboratory there, there was a a desk with a computer and a chair, I stopped at laboratory.

## Robot-v2:

"I woke up at the hall there was a sign on the door there it said COMPUTER SCIENCE DEPARTMENT. After that I wanted to walk towards the office of Ernesto but couldn't, after that I walked towards the office of Ivan to a desk there it said SOFA: Random Forests Regression for the Semantic Textual Similarity task I took a bottle then I moved to a desk I leaved the bottle I noticed a cluttered desk with a laptop there it said sausenc &aceta para consejos - ASO. Afterwards I moved to the meeting room to a table I saw a man standing in a room then I went to a door there it said GOLEM. After that I moved to the laboratory to a desk I noticed a desk with a computer and a chair afterwards I finished."



## Sujet 2 (humain):

First I went past the bathroom, I saw the sign saying that it was the Computer Science Department and the range of cubicles in the department. Then I went to Ernesto Bribiesca's cubicle, it was shut so I couldn't pick up the water nor move the bottle. After that I got into Ivan Meza's cubicle, I read the SOPA poster and I took the water that was on the desk, I left it again and I saw on the desktop a computer with keyboard, mouse, a portrait, speakers. After that I went to the meeting room, where I found one person presenting, the room was busy, there was a Golem's project meeting, and finally I got to the lab.

## Évaluation avec ROUGE:

<b>Narrative</b>	<b>Words</b>	<b>Recall</b>	<b>Precision</b>	<b>F</b>
Robot-v0	164	<b>0.30515</b>	0.45663	<b>0.36583</b>
Robot-v1	139	0.25362	0.45324	0.32524
Robot-v2	139	0.28341	<b>0.50647</b>	0.36344
Subject-1	175	0.43587	0.58409	0.47559
Subject-2	119	<i>0.31863</i>	0.66469	0.42227
Subject-3	211	0.49374	0.50227	0.47204
Subject-4	307	0.60205	0.40819	0.46578
Subject-5	392	0.61060	<i>0.31663</i>	<i>0.40657</i>

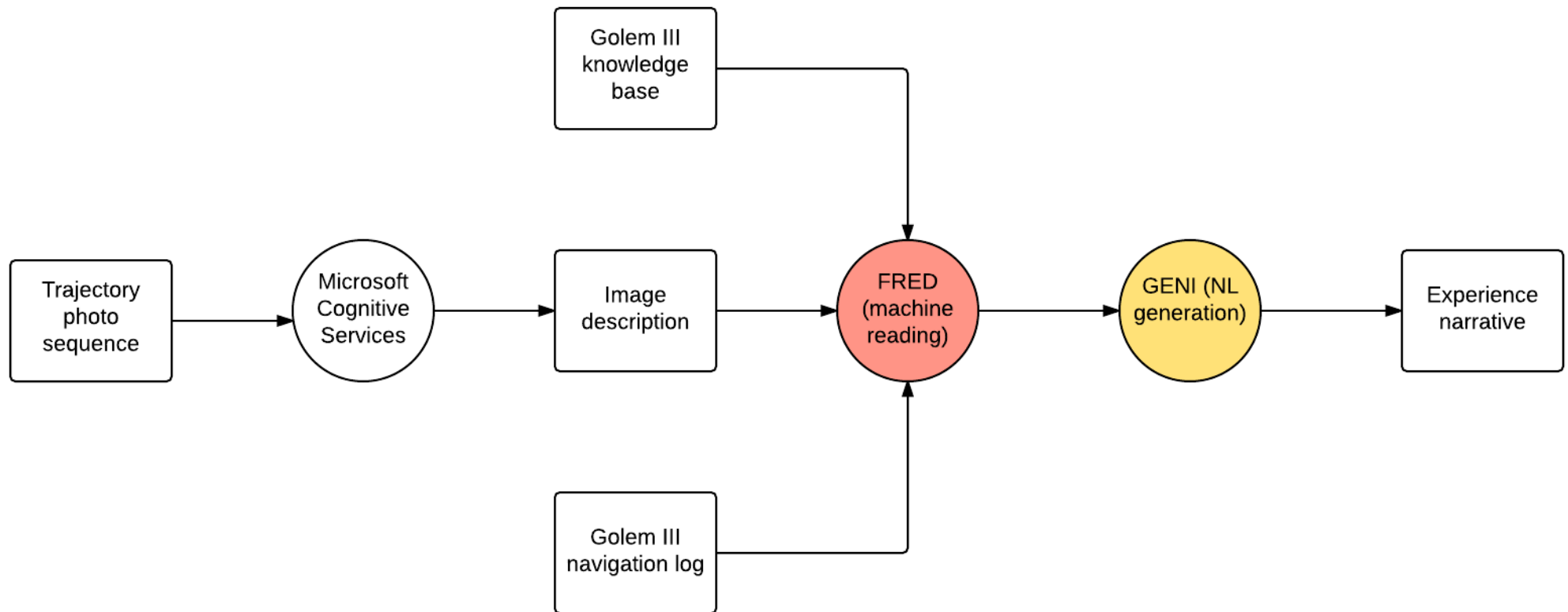
Table 3: ROUGE-L automatic evaluation for robot and human narratives.

## Évaluation par des juges humains:

	<b>Humans avg.</b>	<b>Robot</b>
Precision	4.0	2.8
Completeness	4.2	3.6
Readability	4.0	2.6
Fluency	3.8	2.8

Table 4: Human evaluation (five judges; 1=low quality; 5=high quality)

# Architecture (itération1):



# Perspectives:

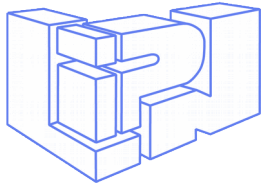
- Générer avec FRED un graphe RDF avec la description des images ainsi que les références au contexte culturel (c'est que le robot lit)
- Produire des templates de génération basés sur le graph RDF de FRED
- Générer le récit final avec l'interface FRED-Geni (outil de génération de langage naturel du LORIA basée sur des grammaires TAG)
- Test des 100 trajectoires
- Générer en langue française

[Gangemi et al. 2014]

[Gardent et al. 2015]

## Démo avec Golem II:





Merci!

