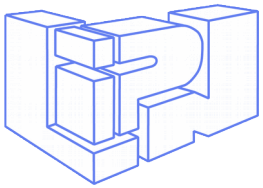
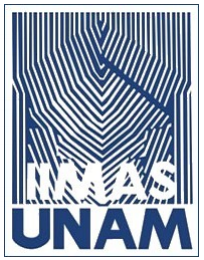



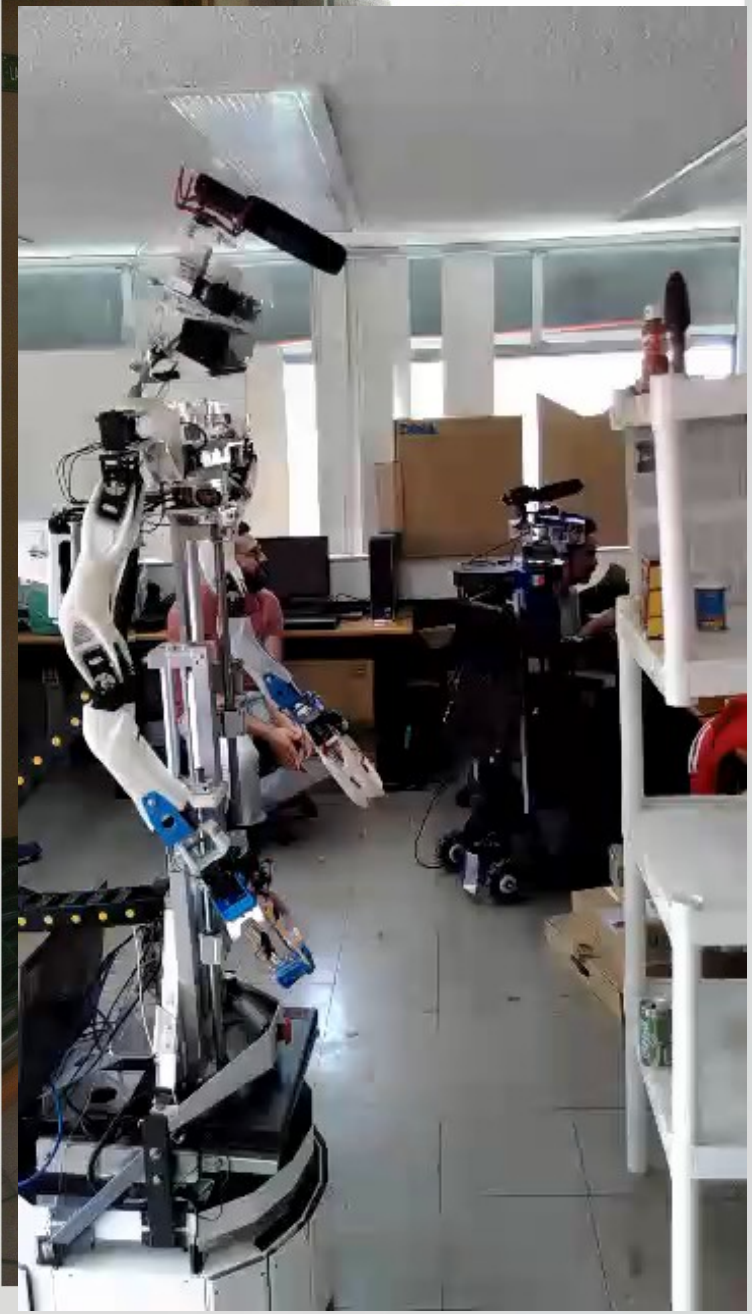

Golfred : Vers la génération de récits d'expérience pour des robots de service

Jorge Garcia Flores
LIPN – CNRS – Paris 13

Claire Gardent
LORIA - CNRS

Aldo Gangemi
LIPN – CNRS – Paris 13

Iván Meza, Luis Pineda
IIMAS - UNAM

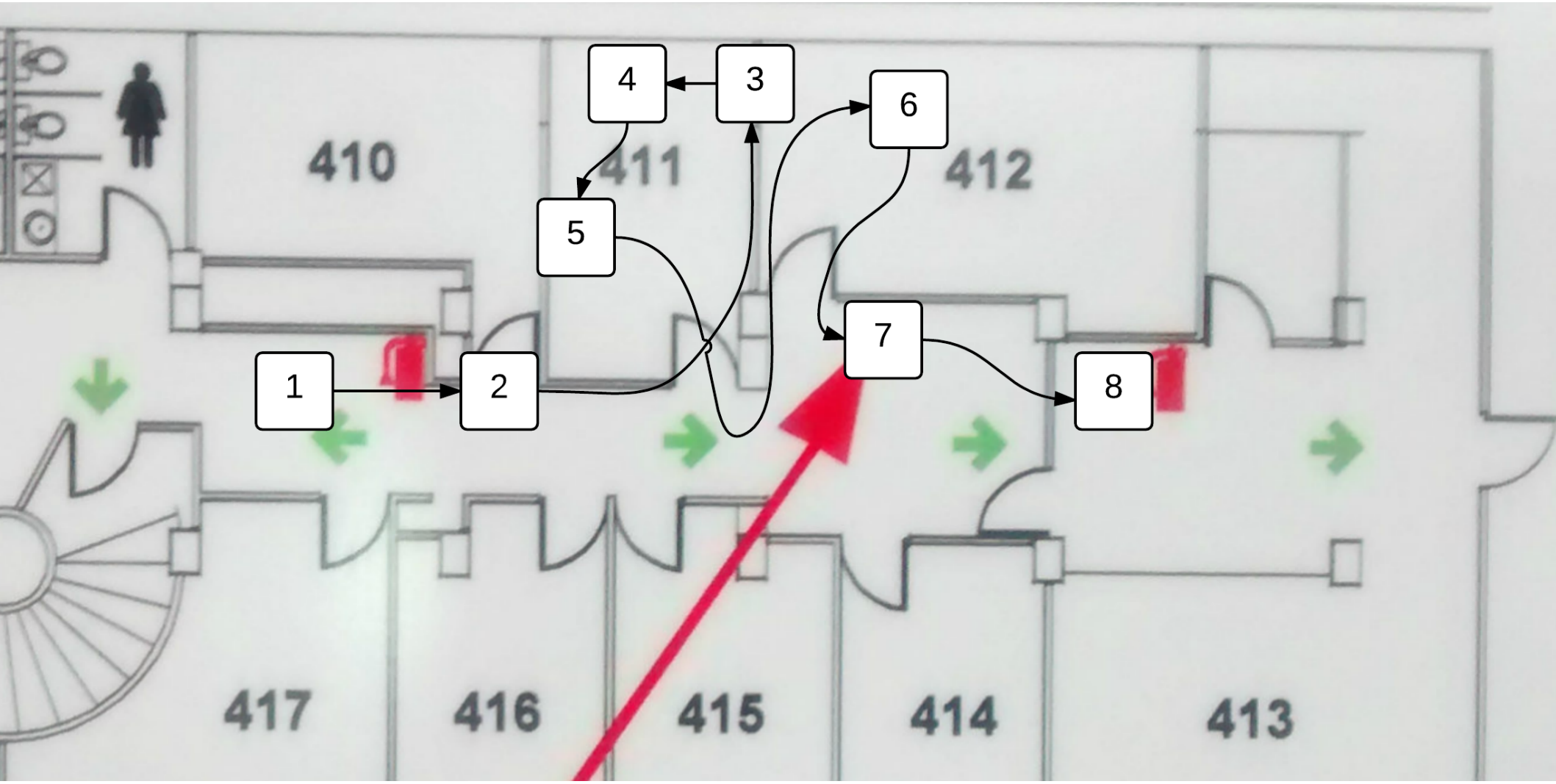


- **Golem III: a service robot**
  - **Dialogue system**
  - **Cognitive architecture**
  - **Command interpreter**
- **3d place in Robocup@home 2012**

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.*

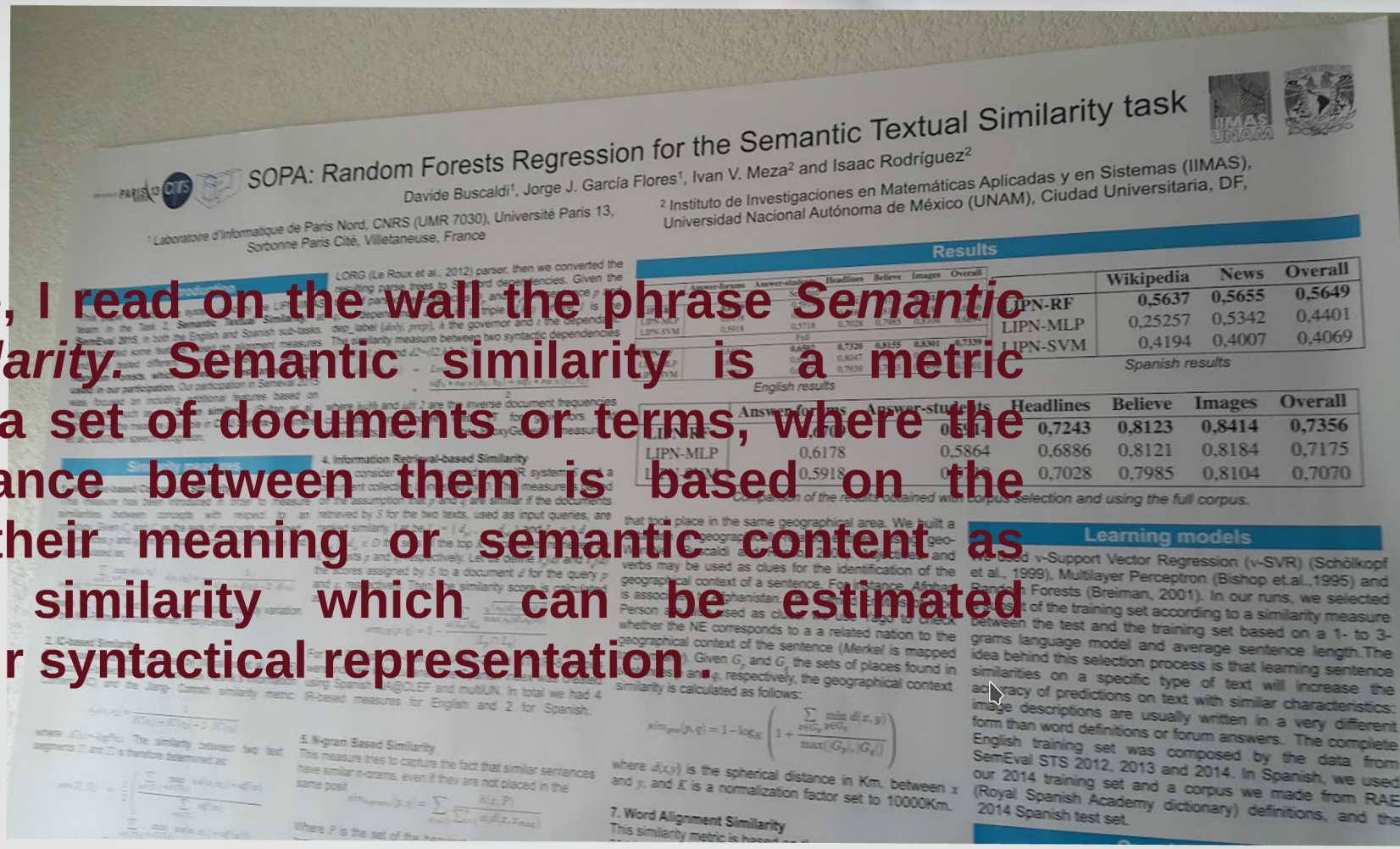
Récits d'expérience



Récits d'expérience

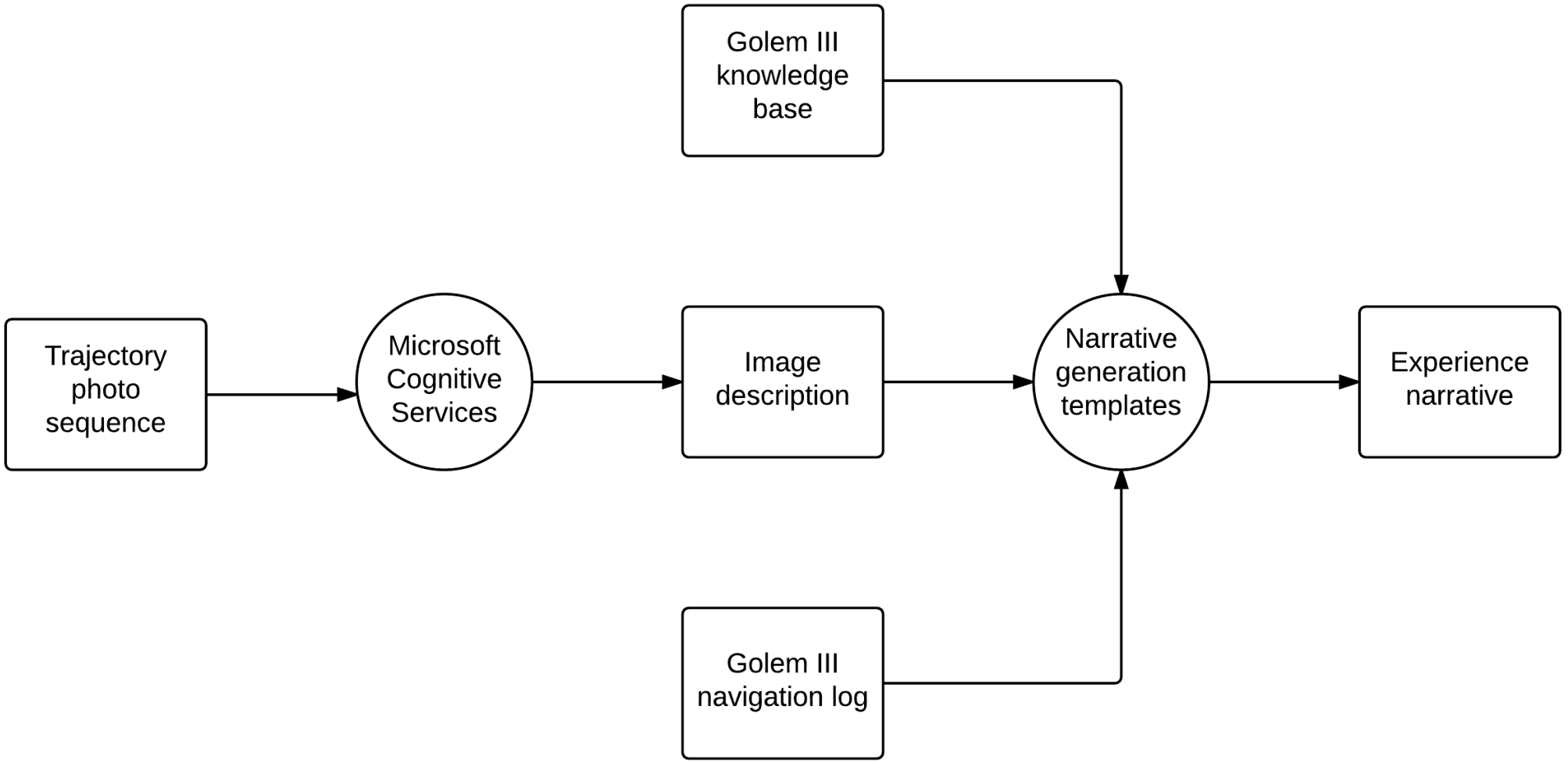
"I was at Iván V. Meza Ruiz office. Iván is a researcher at IIMAS, UNAM working at the Department of Computer Science with the Golem group...."

Récits d'expérience

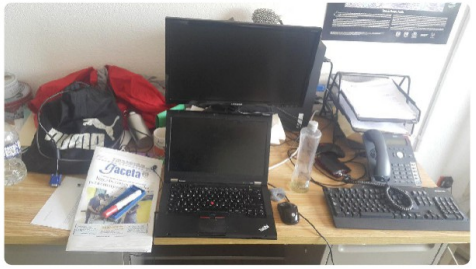


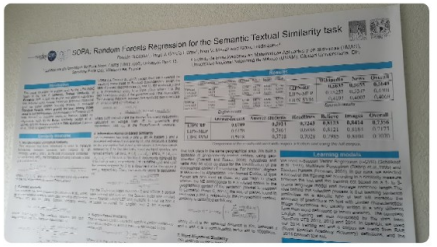
"In his office, I read on the wall the phrase *Semantic Textual Similarity*. Semantic similarity is a metric defined over a set of documents or terms, where the idea of distance between them is based on the likeness of their meaning or semantic content as opposed to similarity which can be estimated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

Architecture (itération 0)



Trajectoire en images





Description des images

Memories - Golfred – Chromium

Memories - Golfred

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

analysis read

Read Analisis

{"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"}]}

analysis

Read Analisis

{"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"}

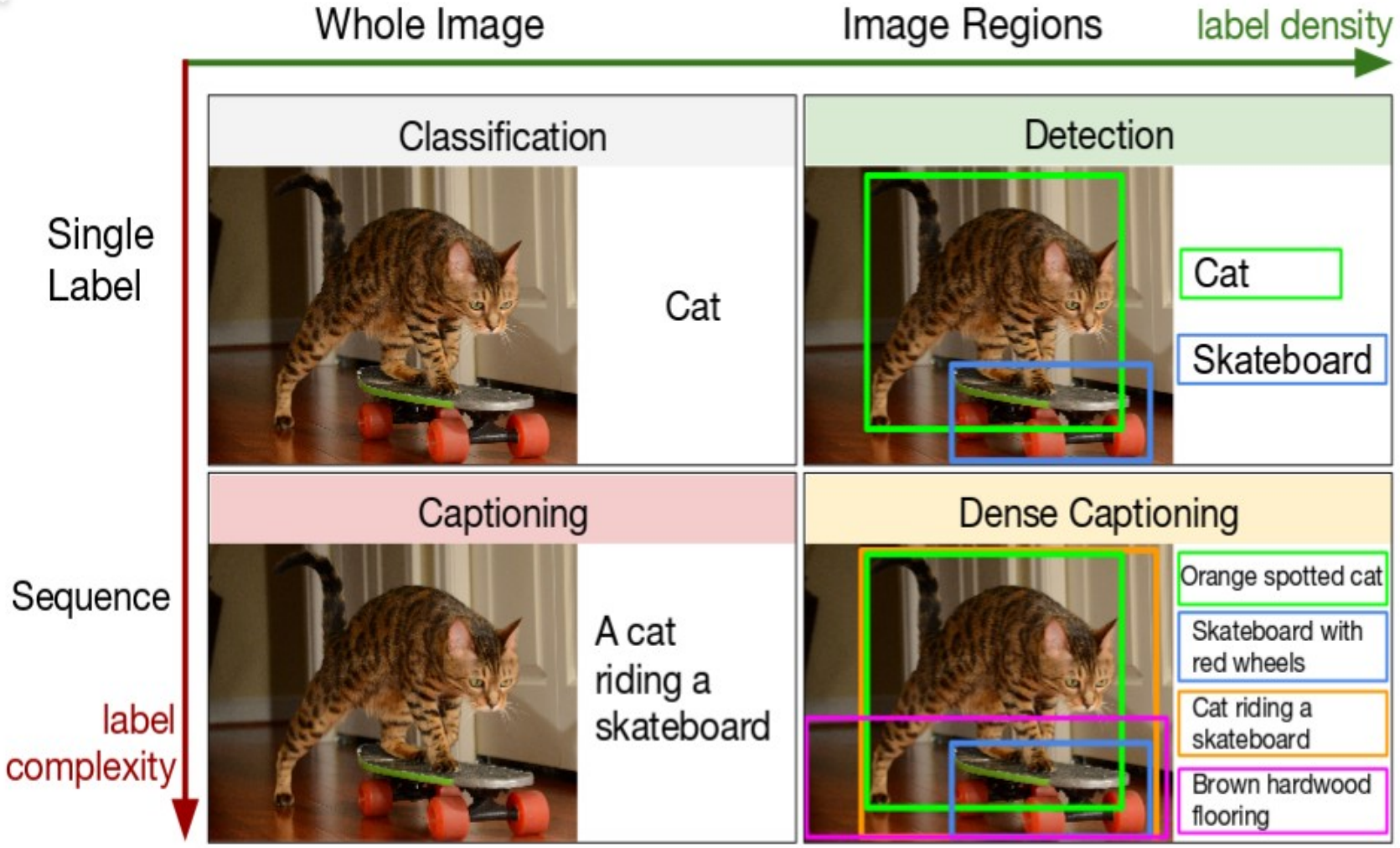
- 

analysis read

Read Analisis

{"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": []},

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

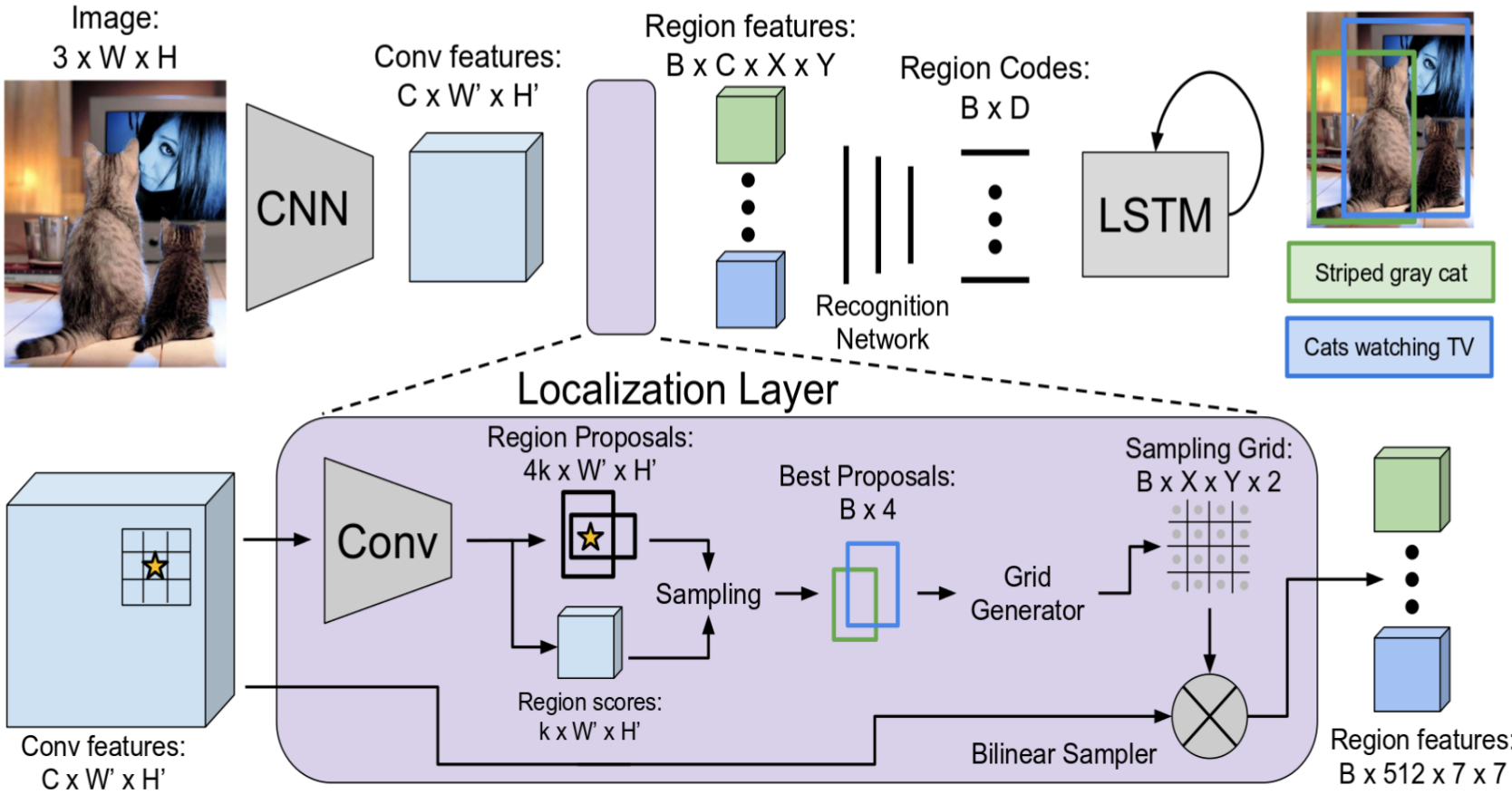
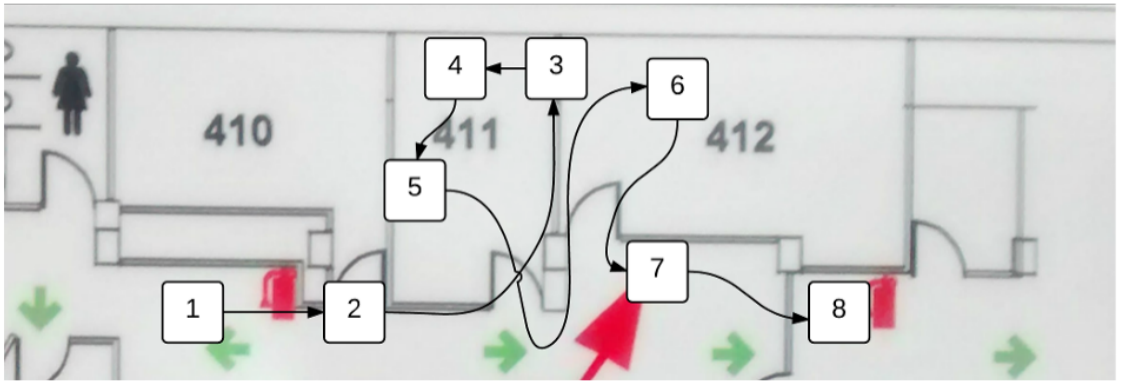
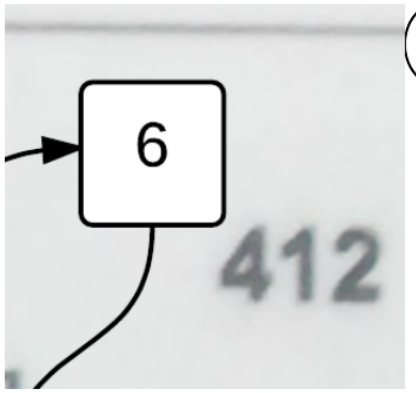


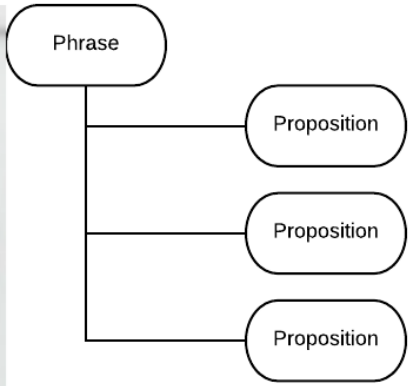
Figure 2. Model overview. An input image is first processed a CNN. The Localization Layer proposes regions and smoothly extracts a batch of corresponding activations using bilinear interpolation. These regions are processed with a fully-connected recognition network and described with an RNN language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

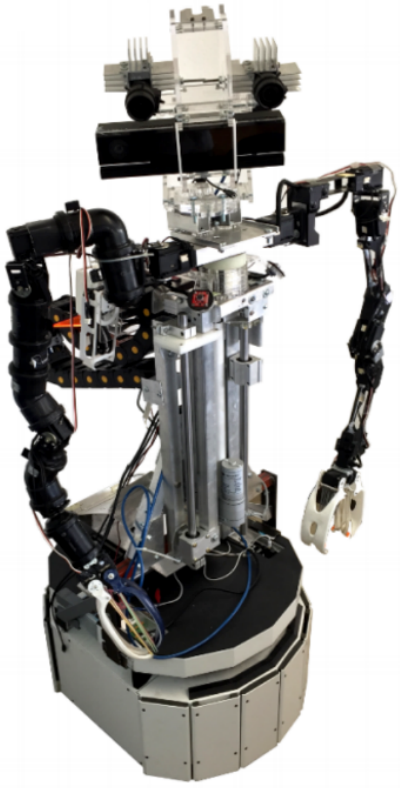
DenseCap: Fully Convolutional Localization Networks for Dense Captioning. Johnson, Justin, Karpathy, Andrej and Fei-Fei, Li (2016), In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

Génération spatiale du discours









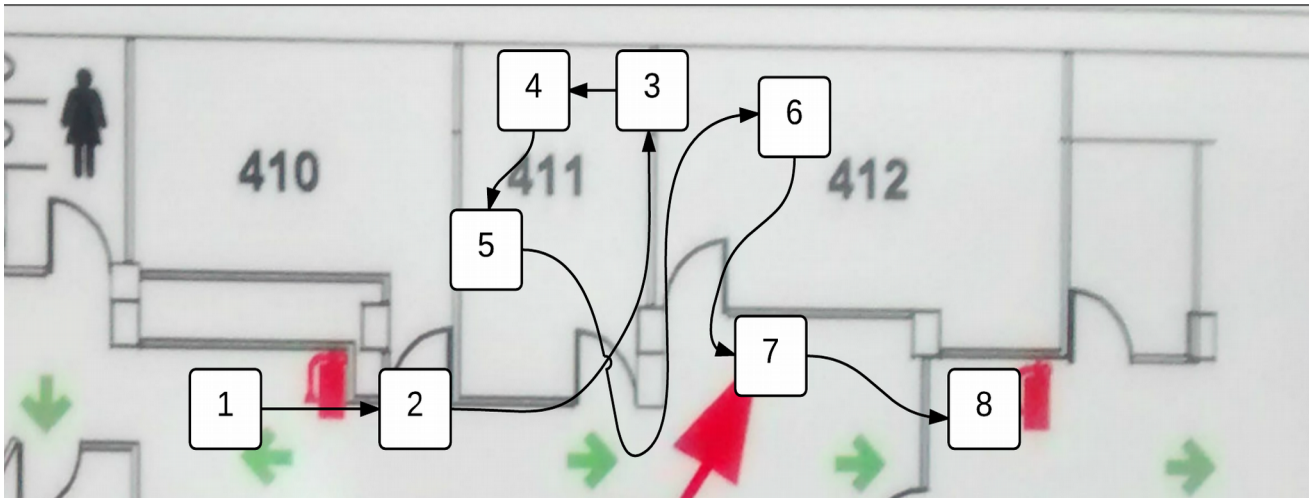
```
start(loc1)
      ⋮
go(loc2)
      ⋮
go(locN)
      ⋮
read(text1)
```

Random template draw → *"I woke up at the hall...*

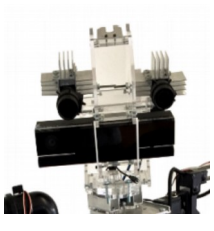
Random template draw → *"afterwards I went to..."*

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 consejoros - 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 consejoros - 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:

Narrative	Words	Recall	Precision	F
Robot-v0	164	**0.30515**	0.45663	**0.36583**
Robot-v1	139	0.25362	0.45324	0.32524
Robot-v2	139	0.28341	**0.50647**	0.36344
Subject-1	175	0.43587	0.58409	0.47559
Subject-2	119	*0.31863*	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	*0.31663*	*0.40657*

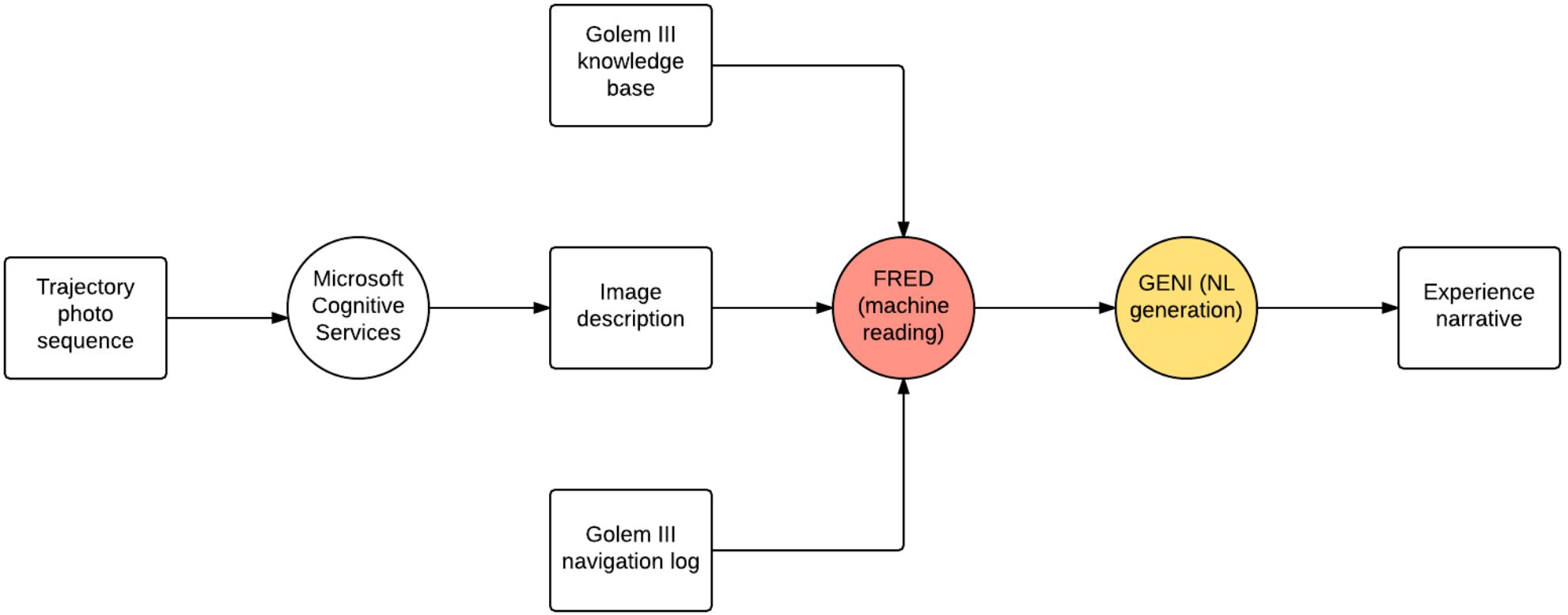
Table 3: ROUGE-L automatic evaluation for robot and human narratives.

Évaluation par des juges humains:

	Humans avg.	**Robot**
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'interace 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!