The Golem Project

A Laboratory for the Construction of Service Robots

Luis A. Pineda & the Golem Group
Department of Computer Science
Institute for Applied Mathematics and Systems
Universidad Nacional Autónoma de México
lpineda@unam.mx

Content

Service Robots

Service Robots

- Intuitive notion
- Previous robots in the Golem Project:
  - Golem
  - Golem in Universum

Intuitive Notion of Service Robots

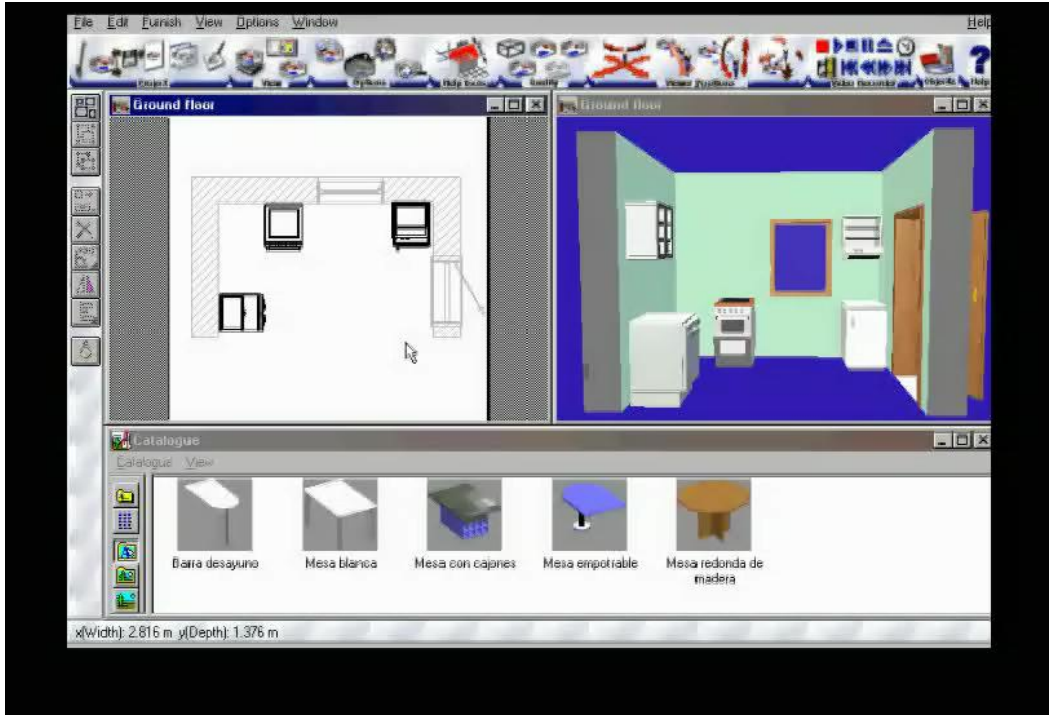
Golem

Golem in Universum

History and Philosophy of the Golem Project

The DIME Corpus

example


File Edit Furnish View Options Window Help
Ground floor
Ground floor
Catalogue
Catalogue View
Barra desayuno
Mesa blanca
Mesa con cajones
Mesa empotrable
Mesa redonda de madera
x(Width): 2.816 m y(Depth): 1.376 m

Conversational Structures

utterance	obliga-tions		common ground				dialogue acts
			agree		unders		
	ch	cr	ch	cr	ch	cr	
1. u: after that <sil> can you put <sil> the extractor on top of the <sil> of the stove	1		1				action-dir
2. s: okay	2			1			commit, accept
3. s: <move-obj>		2 1	3				move-obj
4. is this okay?	4		4				info-request
5. u: yes, it's okay		4		4 3			answer, accept

Pineda et al, 2007

A conservation principle

- the obligations and common ground structures can be thought of as independent planes of expression
- dialogue acts can be thought of as having a charge/credit import on these two planes
- successful transactions should be complete and balanced in each plane!

Service robots should assist people and support structured conversations!

Two main concerns:

- The focus should be in the interpretation of speech acts
- The common ground structure is essential for reliable communication and interaction: Recovery protocols belong to the common ground structure!

Speech acts:

- Direct:
  - Please, move the extractor to the left
- Indirect:
  - Could you please move extractor to the left
  - The extractor is too much to the right!

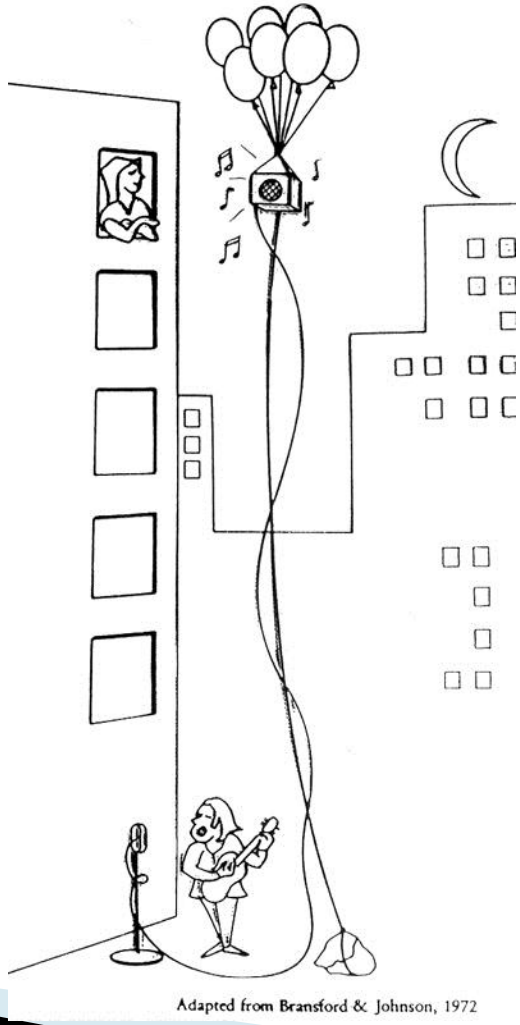
Inference is in service of communication

Interpretation:

If the balloon popped the sound wouldn't be able to carry since everything would be too far away from the correct floor. A closed window would also prevent the sound from carrying, since most buildings tend to be well insulated. Since the whole operation depends on steady flow of electricity, a break in the middle of the wire would also cause problems. Of course, the fellow could shout, but the human voice is not loud enough to carry that far. An additional problem is that a string could break on the instrument. Then there could not be accompaniment to the message. It is clear that the best situation would involve less distance. Then there would be fewer potential problems. With face to face contact, the least number of things could go wrong.

Bransford and Johnson, 1972.

The Context



Adapted from Bransford & Johnson, 1972

The result of interpretation is a representation!

Interpretation is relative to the context!

The Context

- A set of agents
- A spatial and temporal situation (Indexical)
- Discourse Information (Anaphoric)
- Domain knowledge
- The dialogue and task structure!

But… we could simply input the command!

put(stove, x,y)

Yes, but this would be the result of thinking (on the side of the human user)!

The command and the parameters is just what is negotiated along the dialogue!

The dilema:

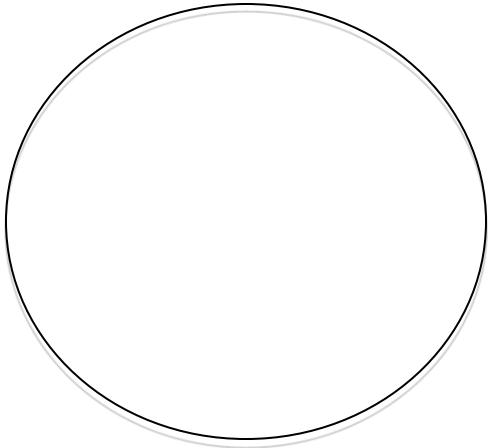
To think (AI) or to talk (HCI)?

What is the balance?

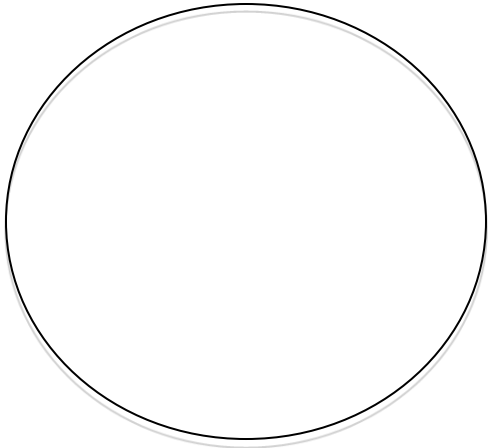
Dialogue Models and Cognitive Architecture

Service Robots are Situated!

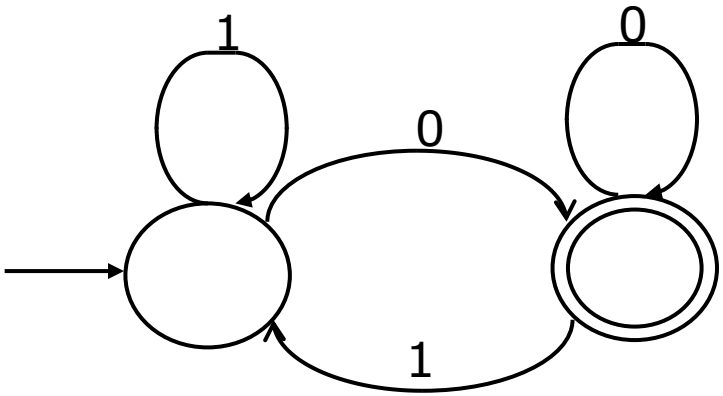
A situation: An information State



How much information?

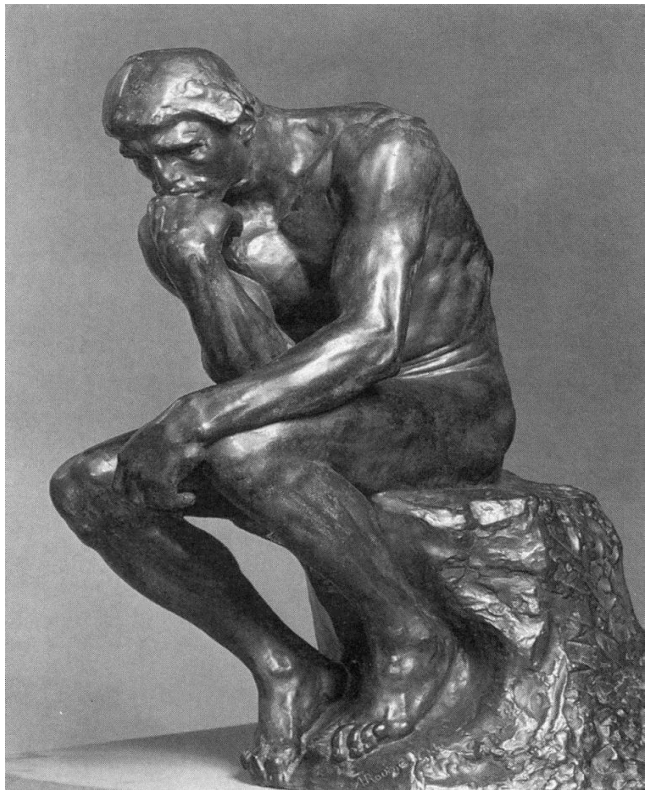


A finite state machine?



Too little: almost empty!

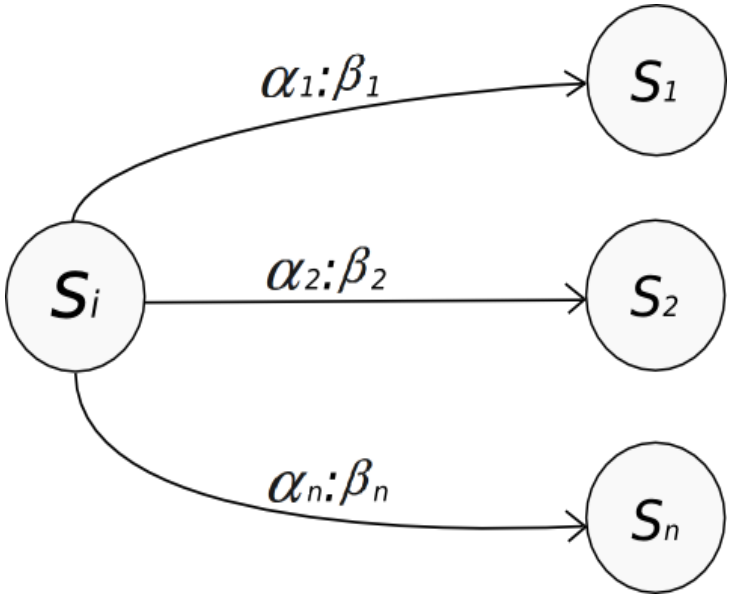
The full mental state?



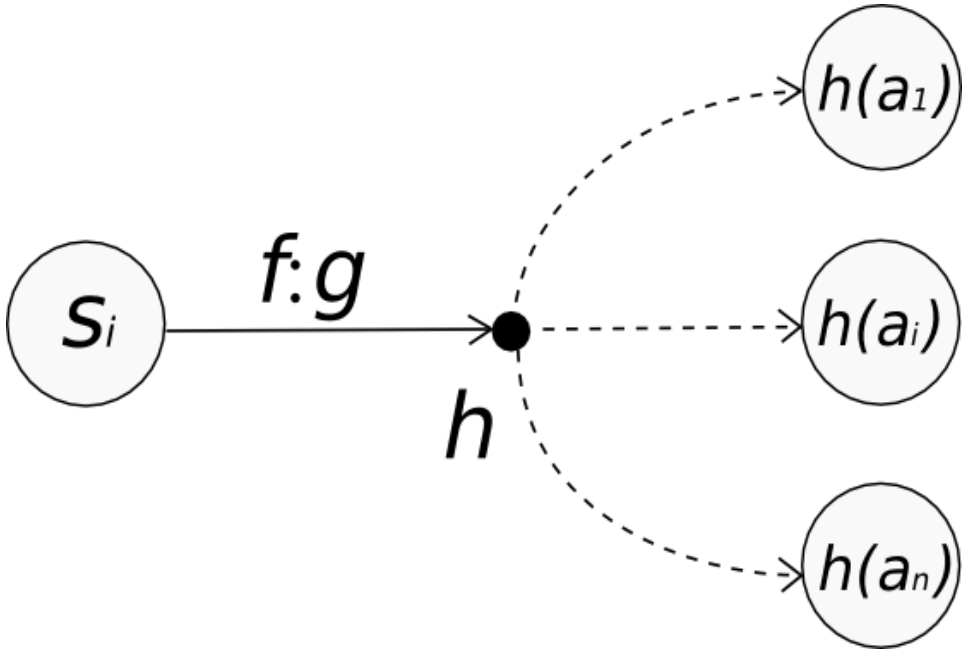
A Situation

- A state within the task structure
- The set of local expectations and dispositions to act at the particular state: speech acts and expected events
- A situation can contain a full task: stack structure (Recursive Transition Network)
- Functional Structure: Speech Act types and content can be parametric (F-RTN)
- Dialogue history: anaphoric information accessed through functions
- Memory access and reasoning embedded within speech act interpretation

Dialogue Models

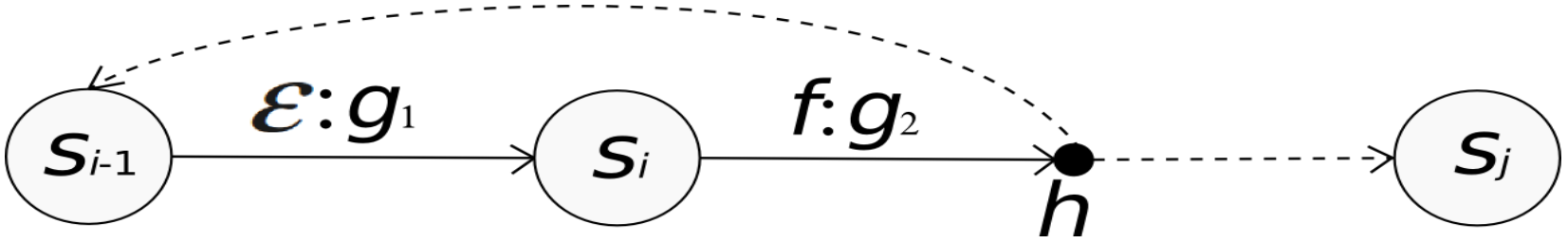


A Functional Specification



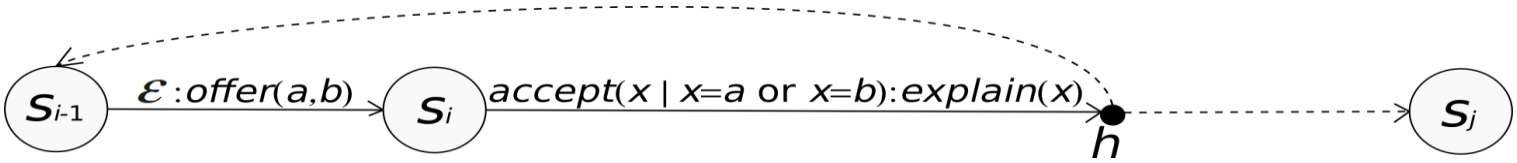
A Dialogue Model…

Functional specification



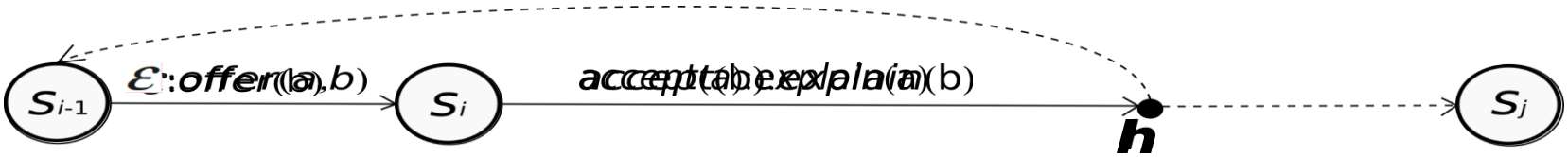
Intentional Protocols

The function values are Speech Acts



Discourse Context

History of concrete interpretations and actions!

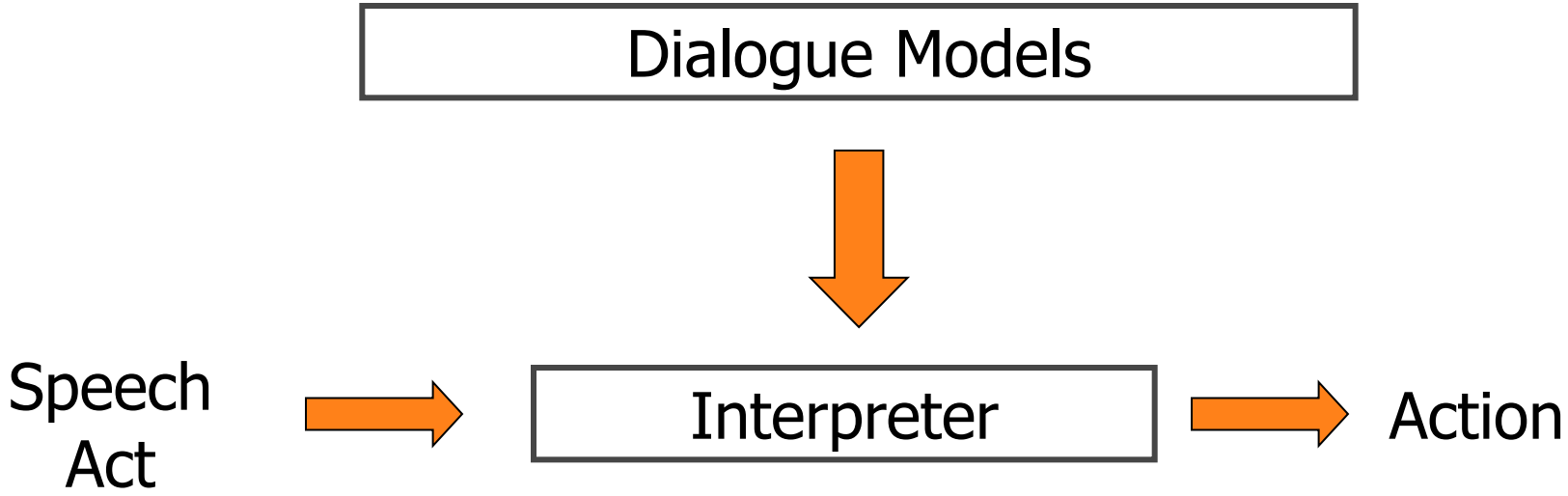


Interaction Context:

Top

If no expectation is met in the situation a recovery dialogue model is executed in order to reestablish the common ground

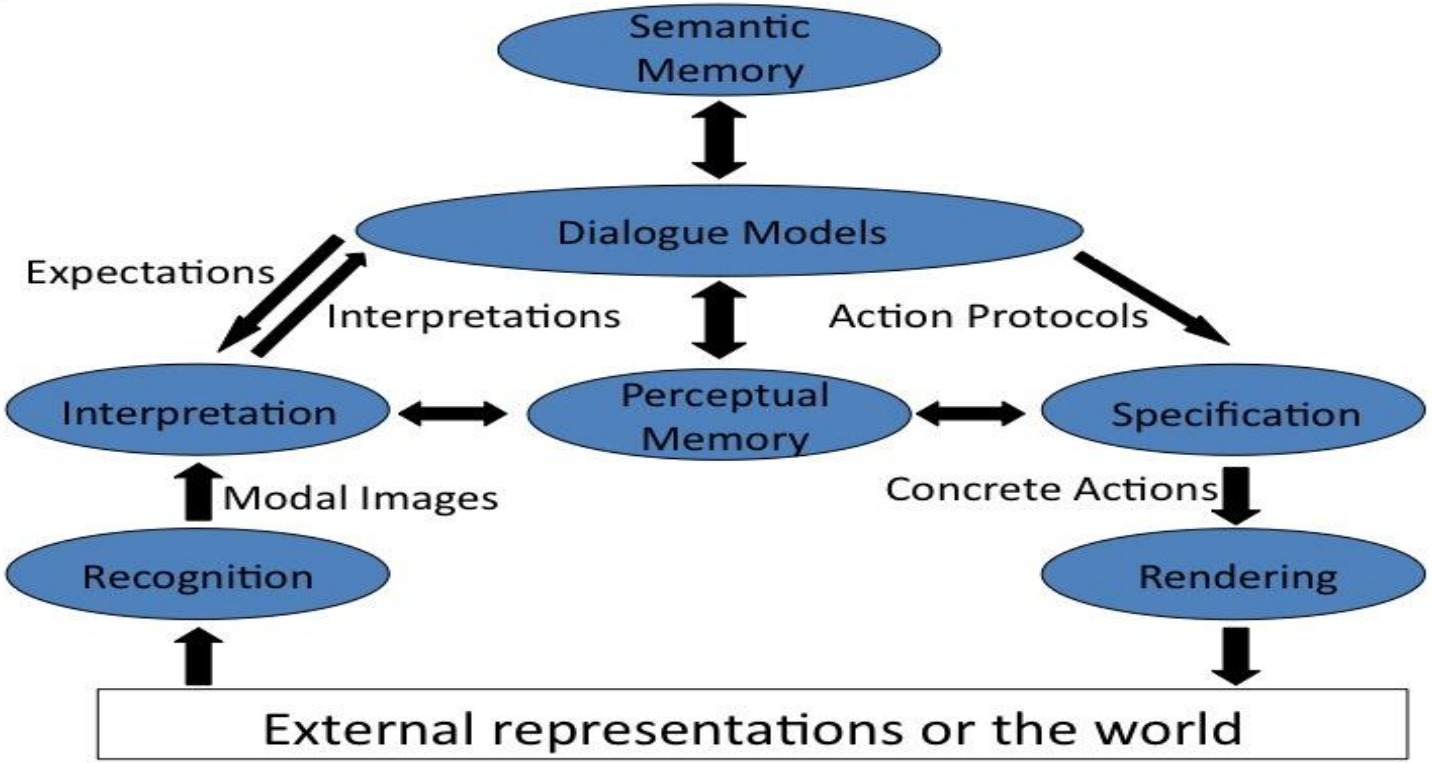
Speech Act Interpretation



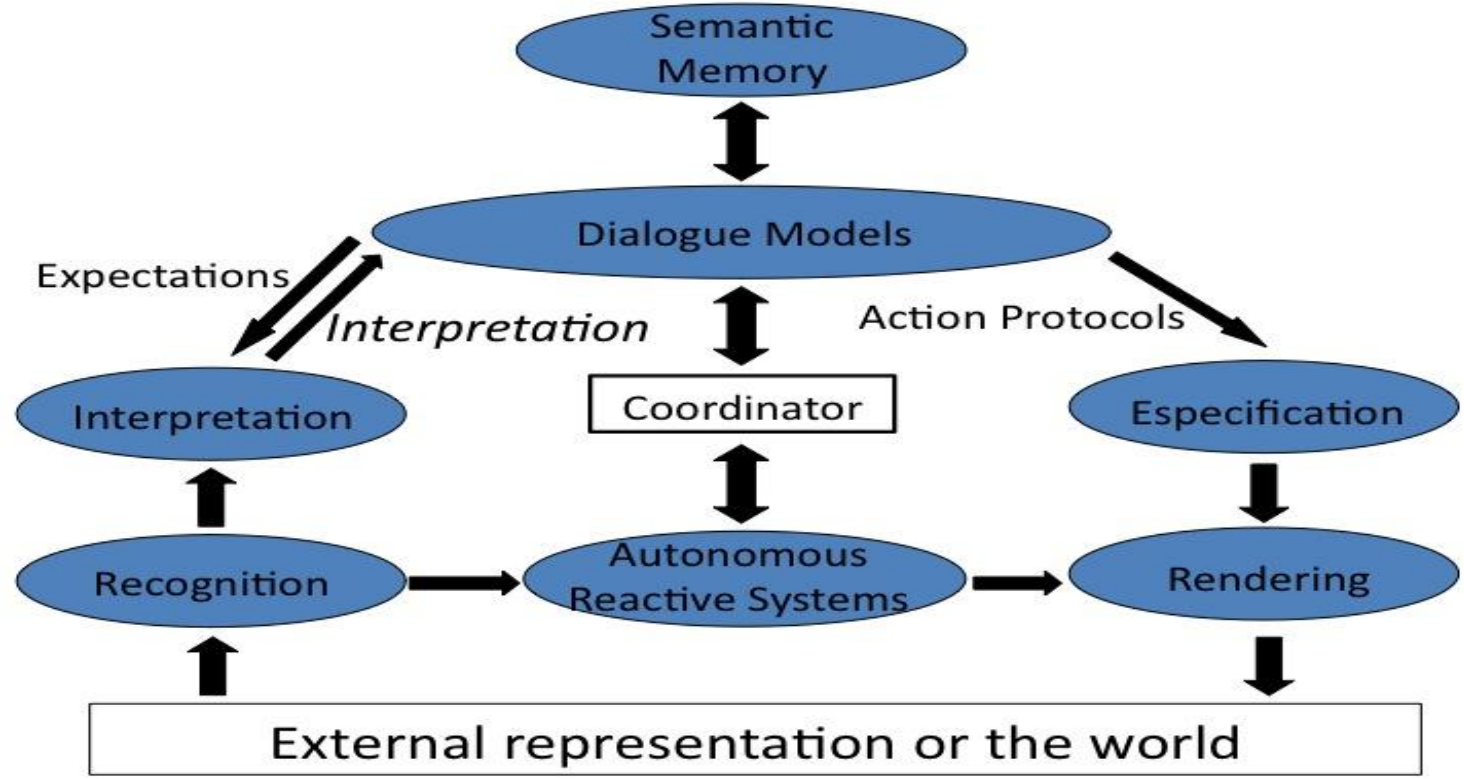
Cognitive Architecture

A modular system in which the main processing disciplines are invariant to information content (independent of task and domain) and local processing mechanisms

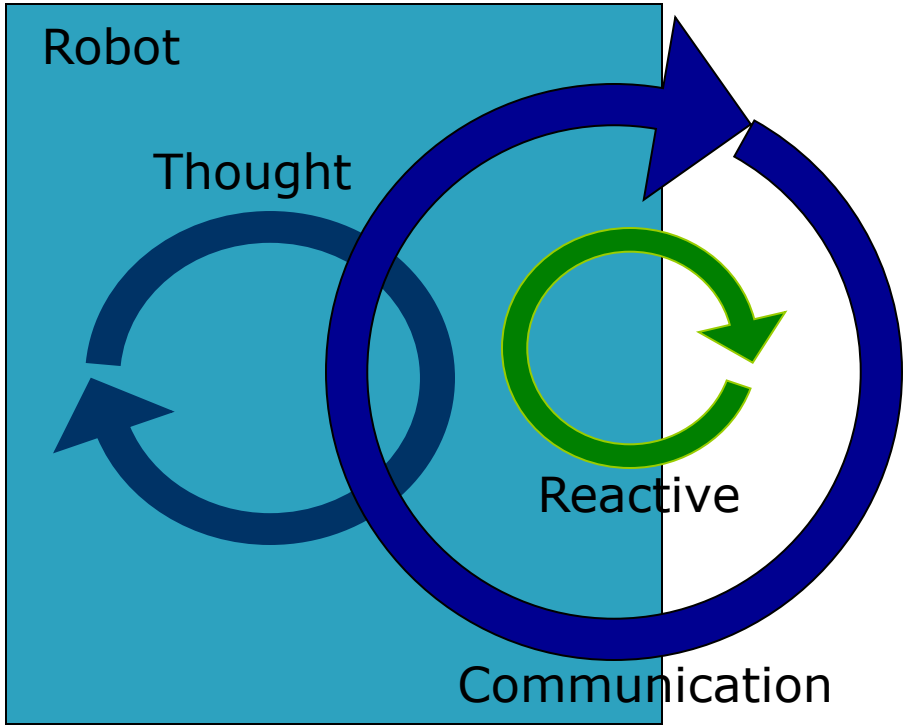
IOCA: Interaction-Oriented Cognitive Architecture



Adding Reactive Behavior



Service Robot's Conceptual Model



Golem-II+

Golem-II+

- Functionalities
- Tasks

Perception and Action Modules (embedded in IOCA)

- Speech and Language (video)
- Navigation (video)
- Vision (video)
- Object Manipulation (video)
- Audio Processing (video)

RoboCup@Home Taks

- Follow me
- General Purpose Service Robots (GPRS)

Speech and Language

Navigation

Vision

Object Manipulation

Audio Processing

Follow me

General Purpose Service Robot

Speech and Language

Speech Recognition

- Live speech recognition
- Context based language models
- Languages:
  - English: WSJ acoustic models and CMU dictionary
  - Spanish: DIMEx100 adults and kids projects
- Coordination with audio-localization
- Recovery and prompting strategies
- Settings:
  - PocketSphinx (Placeway et al, 1997), JACK audio, directional microphone

Synthesis

- Template based generation
- Pool of options for the same intention during recovery strategies
- Coordination with Speech Recognizer
- Setting:
  - Festival TTS software (Taylor et al, 2007)

Language Interpretation

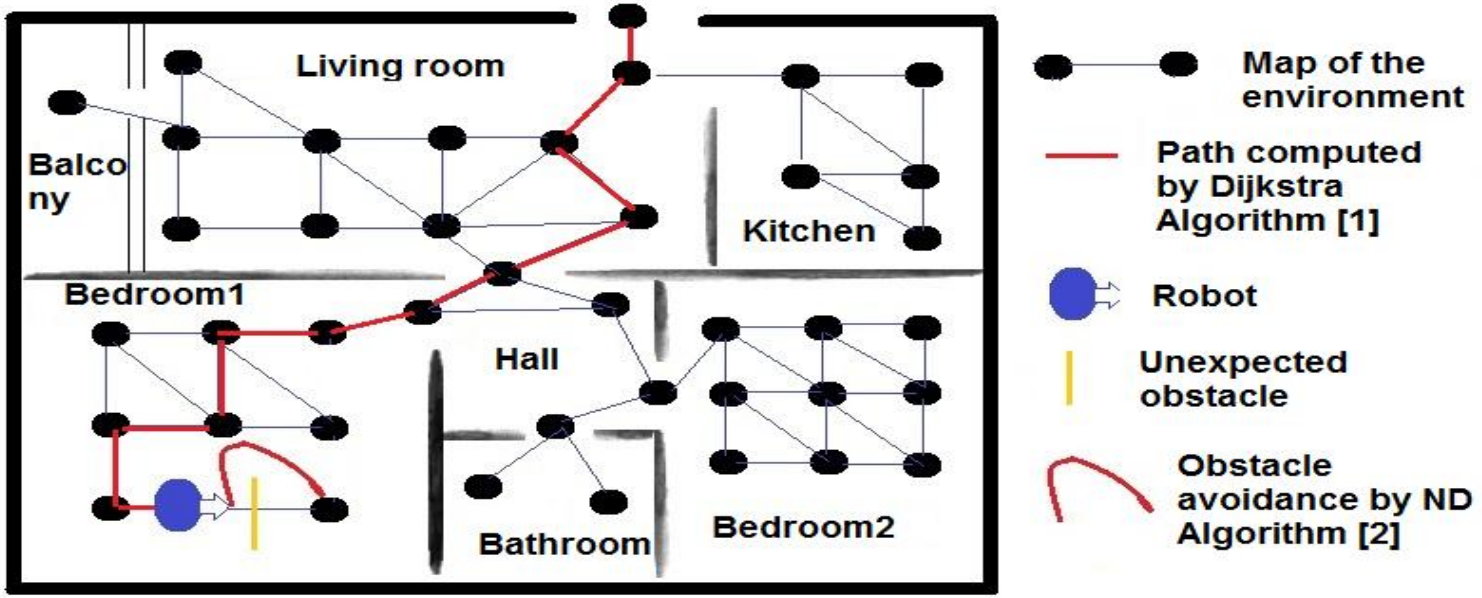
- Context based
- Shallow and deep parsing
- Word spotting
  - Regular expressions
  - Propositional predicates, e.g. *order(X,Y)* (Meza et *al., 2010)*
- Rule based parsing
  - Hand-crafted grammars (GF grammar, Ranta, 2004)
  - Frame based, e. g. *DEST=kitchen, ACTION:take, OBJ: milk*
- Statistical parsing
  - Dependency Parsing and Semantic Role Labeling (LTH system, (LTH system, Johansson and Nugues, 2008))
  - Frame based, e. g. *DEST=kitchen, ACTION:take, OBJ: ??*

Navigation

Navigation

- Planned
  - Environment's map: direct graph
  - Dynamic path finding: Dijkstra's algorithm
- Reactive
  - Object avoidance: ND (Nearest Diagram)
  - Laser Sensor: Hokuyo UTM-30LX

Navigation Example

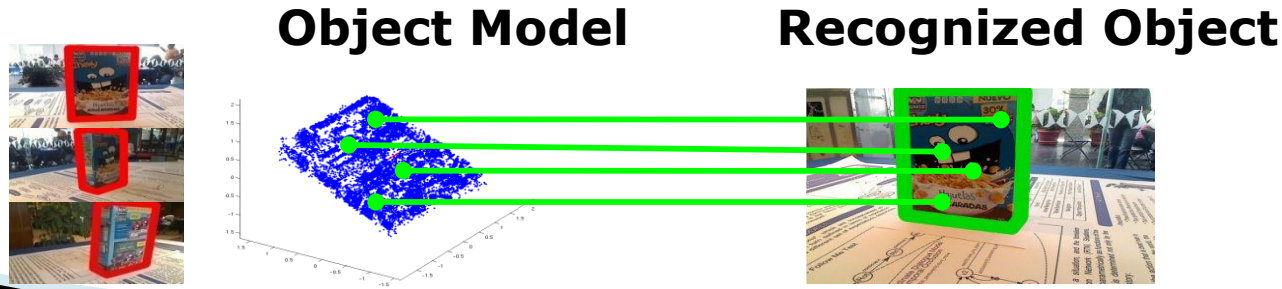


- [1] Briggs A., Detweiler C., Scharstein D., Vandenberg-Rodes A. "Expected Shortest Paths for Landmark-Based Robot Navigation". *International Journal of Robotics Research*, 2004.
- [2] Minguez J., Montano L. "Nearness Diagram Navigation (ND): Collision Avoidance in Troublesome Scenarios". *IEEE Transactions on Robotics and Automation*, pp 154, 2004.

Vision

Object Recognition with MOPED*

- Modeling
  - Features from different views of an object are matched
  - A 3D model is created from the matched features
- Recognition
  - Finds matches between the scene and the 3D models
  - Generates and refines object hypotheses from clusters matched features, and estimates their pose



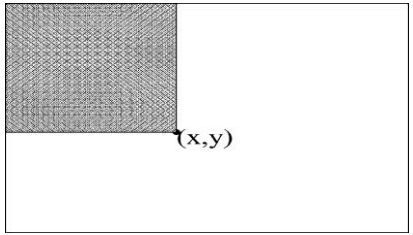
*Collet et al. 2011, The MOPED framework: …

Viola-Jones Face Detector*

- Faces represented by simple rectangular features
  - Fast computation through integral image representation

Rectangular Features

Integral Image



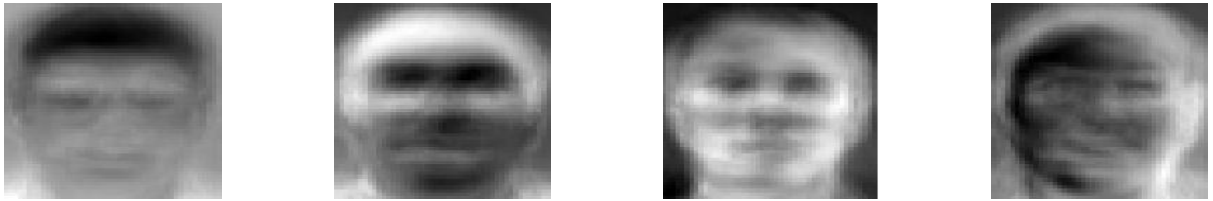
- Fast multi-scale detection by combining multiple weak classifiers
  - Slow training but real time classification

*Viola and Jones 2004, Robust real-time face detection

Eigenfaces Face Classifier*

- PCA on training images
  - Face space with reduced dimensionality
  - The principal components of the distribution of faces are taken as a basis, these are called the eigenfaces

Examples of eigenfaces

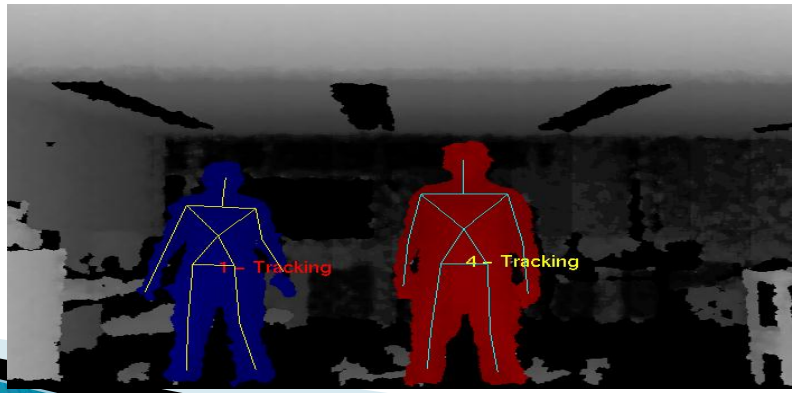


- New face images are projected onto the face space to be recognized
  - The nearest known face is found

*Turk and Pentland 1991, Eigenfaces for recognition

People Tracking and Gesture Recognition

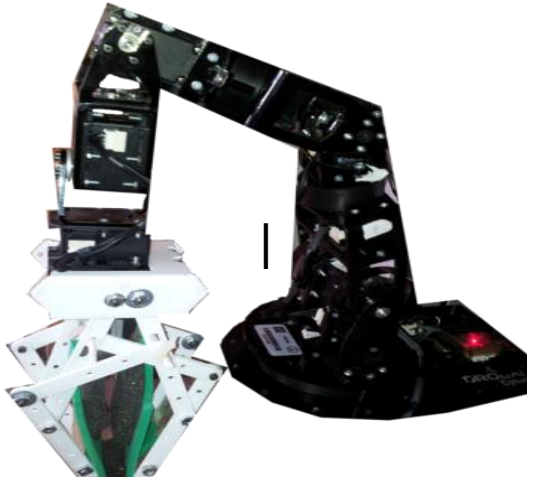
- Microsoft® Kinect sensor
- OpenNI Framework
  - People detection and tracking
  - Skeleton detection and tracking: 15 joints
  - Gesture recognition



Object Manipulation

Manipulation

- Robai Cyton Veta 7DOF
  - Kinematically redundant arm
- Control based on Actin-SE
  - Optimized control system
  - Avoids joint limits, singularities and collisions
  - Minimizes kinetic energy



In-house robot arm

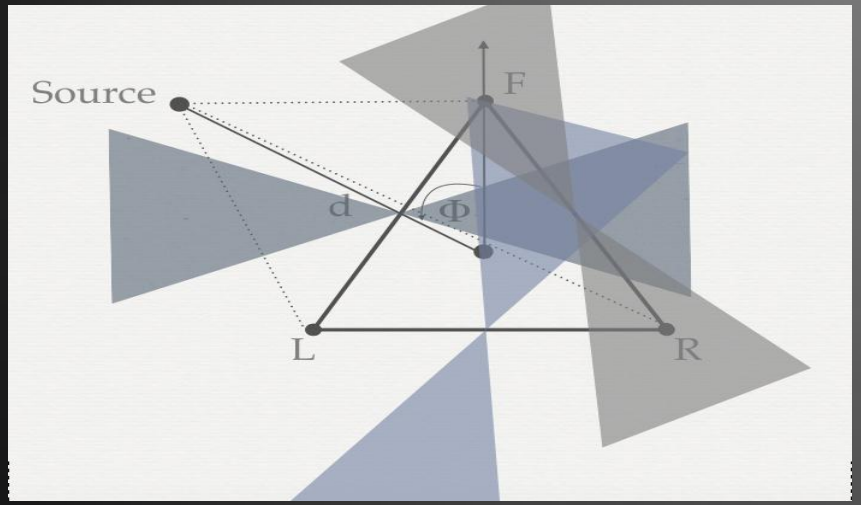
- 4 degrees of freedom
- Reach: max. 90 cm, min. de 10 cm
- 1 hand with IR sensors for object detection, openning: 0 to 30 cm
- Max. load: 500 gr
- Costo: 15,000 pesos aprox.

Audio Processing and orientation through source of sound

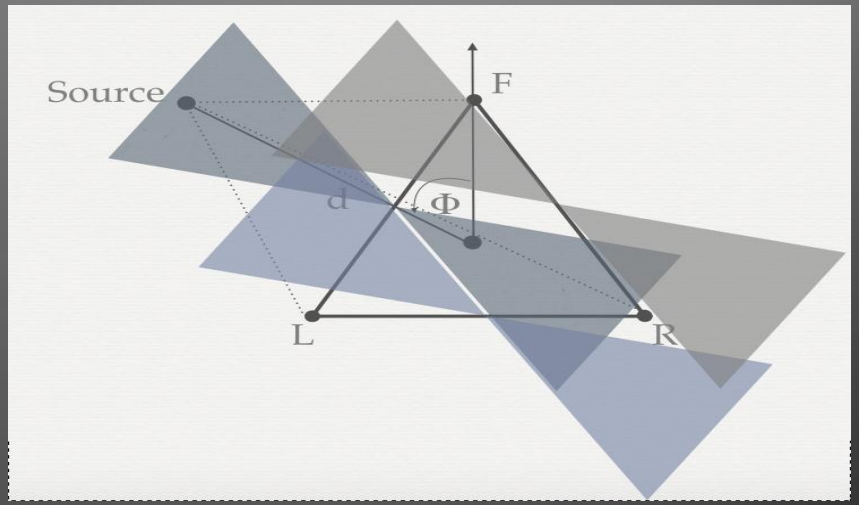
Robotic Orientation towards Speaker

- Motivation:
  - Naturality in HRI (speaker feels as the robot is "putting attention" when it faces him/her)
  - Directional microphone pointed at user automatically
- Using 3 microphones:
  - 360 degree range
  - Close-to-linear response throughout
  - Redundancy

Triangle Array - Redundancy



Sample not acceptable:
will **NOT** be processed

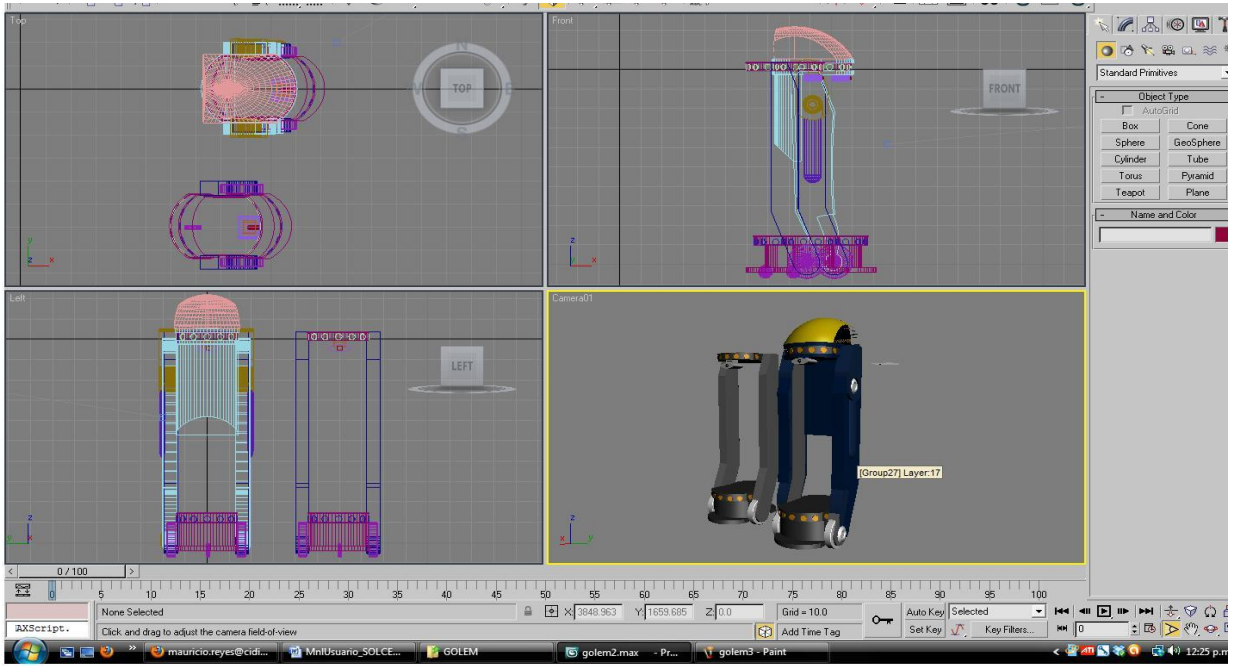


Sample acceptable:
will be processed

Image

A Symbol

- A Golem (as a metaphor of AI)
- From UNAM
- Mexican Technology:
  - Presence
  - Personality (sympathetic!)
  - Thin


Top
Front
Left
Camera01
TOP
FRONT
LEFT
[Group27] Layer:17
Standard Primitives
Object Type
AutoGrid
Box
Cone
Sphere
GeoSphere
Cylinder
Tube
Torus
Pyramid
Teapot
Plane
Name and Color
0 / 100
None Selected
Click and drag to adjust the camera field-of-view
Add Time Tag
Auto Key
Set Key
Key Filters...
MAXScript.
mauricio.reyes@cidi...
MnlUsuario_SOLCE...
GOLEM
golem2.max - Pr...
golem3 - Paint

OPCION 2

Current Image

Future Image

Evaluation

Evaluation

- In-house methodology (based on Paradise)
  - Maximize user satisfaction
  - Task success (quantitative)
  - Objective measures (turns, time, etc.)
  - Qualitative measures (user's subjective judgments)
- RoboCup@Home Competion:

RoboCup Istambul
July 2011

15th Place

RoboCup German Open March 2012

3rd Place

Torneo Mexicano de Robótica
April 2012

1st Place

Concluding Remarks

Perspectives of service robots!

A model for technological development

- The overall product gives a context for particular problems and specialties!
- Research questions are motivated empirically!

The Golem Team

- Dr. Luis A. Pineda (Coordination, Diag. Mod. & Cog. Arch.)
- Dr. Ivan V. Meza (Speech, Language and Diag. Models)
- Dr. Caleb Rascón (Audio Processing, Navigation)
- Dr. Gibrán Fuentes (Vision, manipulation)
- Dr. Mario Peña (Control)
- Dr. Carlos Gershenson (Cog. Architectute)
- M. C. Ivan Sánchez (Navigation)

The Golem Team

- M. C. Arturo Rodríguez (Vision, Navigation)
- M. C. Hernando Ortega (Manipulation)
- M. C. Mauricion Reyes (Image and Design Engineering)
- Ing. Liz Salinas (Dialogue Models)
- Ing. Joel Ortega (Control and Electronics)
- Lic. Esther Venegas (Evaluation)
- Srta. Varinia Estrada (Conversational Structure)

Many Thanks