

PUBLISHED BY

INTECH

open science | open minds

World's largest Science,
Technology & Medicine
Open Access book publisher



3,100+
OPEN ACCESS BOOKS



103,000+
INTERNATIONAL
AUTHORS AND EDITORS



102+ MILLION
DOWNLOADS



BOOKS
DELIVERED TO
151 COUNTRIES



AUTHORS AMONG
TOP 1%
MOST CITED SCIENTIST

12.2%
AUTHORS AND EDITORS
FROM TOP 500 UNIVERSITIES



Selection of our books indexed in the
Book Citation Index in Web of Science™
Core Collection (BKCI)

WEB OF SCIENCE™

Chapter from the book

Downloaded from: <http://www.intechopen.com/books/>

Interested in publishing with InTechOpen?
Contact us at book.department@intechopen.com

Humanoid Introspection: A Practical Approach

Regular Paper

Ignazio Infantino^{1,*}, Giovanni Pilato¹, Riccardo Rizzo¹ and Filippo Vella¹¹ Istituto di Calcolo e Reti ad Alte Prestazioni, Consiglio Nazionale delle Ricerche, ICAR-CNR, Palermo (PA), Italy

* Corresponding author E-mail: ignazio.infantino@cnr.it

Received 4 May 2012; Accepted 21 Jan 2013

DOI: 10.5772/55919

© 2013 Infantino et al.; licensee InTech. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract We describe an approach to robot introspection based on self observation and communication. Self observation is what the robot should do in order to build, represent and understand its internal state. It is necessary to translate the state representation in order to build a suitable input to an ontology that supplies the meaning of the internal state. The ontology supports the linguistic level that is used to communicate information about the robot state to the human user.

Keywords Humanoid, Introspection, Neural Networks

1. Introduction

Interaction between man and robot can benefit from models of the human mind and its cognitive abilities. The reproduction, even if simplified, of known cognitive processes could lead to a reduction in the differences in behaviour between the robotic and human mind. Surely, cognitive architectures [2, 29] are a useful tool in this direction of research, although many of them were focused primarily on aspects of reasoning and planning. We definitely need to improve models of the cognitive mechanisms that integrate the knowledge of sensory data arising from the environment in which the artificial agent

acts [19, 22] and bind them to its particular physical configuration.

For example, humanoid robots [32], which are designed to be companions to humans, and to be an integral part of their socio-ecological system [10, 16], may have a closer appearance and behaviour if they could have a unified model to represent themselves and the humans with whom they interact.

One aspect that has not received enough attention is the mechanism that allows the robot to have a kind of self-awareness [8]. Some of the concepts associated with it have been widely discussed when it comes to consciousness [3], the robot's intentionality and perceived intentions [23, 24], emotions [33, 34, 43] and so on, but it still lacks integration into a cognitive architecture.

Self-awareness, including the body's physical parts and the management functions related to them, is often referred to as an a priori knowledge and it is usually implicitly integrated into the architecture.

Recent studies on the functioning of the human brain, obtained using fMRI, assume the existence of resonance models [6], which make it possible to perceive the

external entities (objects, actions, environment, events, emotions, etc.) through entities that represent the agent itself and its internal states. For example, the mechanism of mirror neurons is used not only for the recognition of actions, but also to explain empathy, imitation, and so on [5, 15, 17, 18].

This paper focuses on introspection, one of the basic mechanisms of the human mind, aiming to have an internal model, which can however easily be related to external entities. The inspiration for the work presented is summarized by the following sentence from Sloman [39] *“non-trivial introspection involves: An architecture with self-observation subsystems running concurrently with others and using a meta-semantic ontology that refers to relatively high level (e.g. representational) states, events and processes in the system, expressed in enduring multi-purpose forms of representation, as opposed to transient, low-level contents of conditionals and selection procedures”*.

We will illustrate a possible approach for the realization of an introspective capacity (at present mainly oriented to the physical components of the robot) and we will show the potential that could be realized with some practical examples of uses of this module: it is easy even to reach a linguistic level consistently with the internal representation of the knowledge of the robot.

Based on an empirical approach, our idea of introspection of the robot starts from the analysis of information obtained automatically by embedded software and its related documentation, in particular regarding the relationship (direct or explicit) between the physical components and software modules. This documentation is used to construct a representation of the hardware and software of the robot on a map similar to the human somatosensory map.

Using a Self Organizing Map (SOM) [25] it is possible to organize documents [20]. This map structure can be used to quickly retrieve semantically related information [21]. Such an approach could be merged with the Latent Semantic Analysis (LSA) paradigm [27][28], which can be used to successfully simulate several psycholinguistic phenomena. From this map we can use different approaches to ascend to a higher level of abstraction. In particular, the approach used in this paper involves a simple association between labels arising from data and ontology entities, trying to achieve expressiveness with an enormous knowledge based on common sense using Cyc [11].

The presented work thus aims to automatically obtain significant semantic links between entities linked to the running processes, software libraries and physical components of the robot. The approach is more interesting if we consider that the robot, motivated by

curiosity, tries to fill the gaps of “self-knowledge” by making new hypotheses to be tested on its internal representation, with the aim of discovering new possible relationships. For example, the robot could find answers through introspection to questions such as:

- Is my hand visible from my visual system? How do I distinguish my right hand from the left one?
- What is the physical extension of my visual field with respect to the movements of my hands, my legs and the rest of my body?

Connecting the modern semantic tools and the ability of introspection [39], the robot can acquire not only a knowledge of its physical structure and functionality (see Figure 1), but it could also communicate this information to a human: it can talk about its perception and it could be able to describe its status to its human interlocutor (and partner) in natural language [9, 36, 40]. Moreover, this information could be integrated in a suitable cognitive architecture.

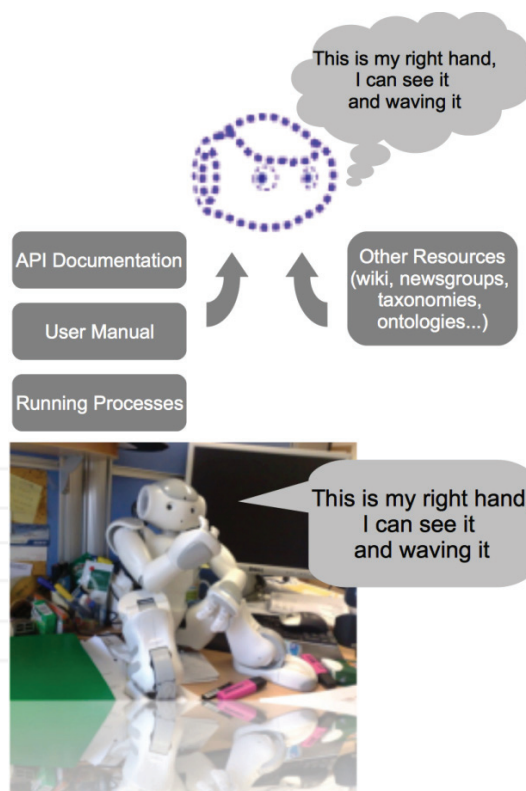


Figure 1. Humanoid robot NAO is the experimental platform used to validate the presented approach to introspection. On the upper side, the diagram that shows the components involved.

Our analysis is based on the assumption that software modules, processes or hardware robot, can contribute to the self-knowledge of the robot directly or less (for example, when derived through some correlated entities). This hypothesis is based on an analogy with a human being, for which the internal organs and their characteristics are not directly accessible through his

experience, but through specific sensations (pain, comfort, discomfort etc.), which can be deduced.

There are several works that indicate introspection as a key component for the relationship between self-awareness and consciousness, either implicitly or explicitly (see for example [35]). There are also other aspects such as self-reflection, sense-of-self, self-recognition etc., which definitely should be taken into account for a complete discussion of this research topic. In this work we have chosen a practical approach that allows us to acquire an implementation of an introspection system to perform experiments in real situations, using a specific humanoid robotics platform. The structure of the paper is as follows. Section 2 deals with the general architecture of the system introspection, Section 3 describes the components that make the self-observation of the robot and Section 4 describes how the observation is projected to the semantic level discussed in the previous section. Section 5 describes the linguistic level and the use of Cyc.

2. The Introspection Architecture System

Our approach to introspection is based on self observation and communication.

Figure 2 shows the proposed architecture for introspection. Considering the definition by Sloman [39], self observation is what the robot should do in order to build, represent and understand its internal state. In particular it is necessary to have a set of sub-systems dedicated to making a snapshot of the Nao robot state. Some systems are supplied by Nao system software, while others are developed ad-hoc.

The data obtained are used to build a rich state representation that is, in some way, biologically inspired.

This state representation should be supplied to an ontology that associates a meaning with the internal state. Our approach integrates static and dynamic information with robot operation.

Static information is related mainly to robot hardware parts and software modules, like hardware drivers, or modules that supply some services like face tracking. But these parts can be active or not during robot operation and their state is part of the robot state itself.

A simple list of these parts can be difficult to manage, considering that each part comes with a list of functioning parameters (for example, each arm movement has an attached angle, movement velocity etc.). In order to obtain a manageable state representation we decided to develop a map that collects all the robot parts and to highlight on this map all the parts that are involved in

any robot operation. This is inspired by the somatosensory map in the human brain.

Such a map can be obtained using the information contained in the robot documentation that is rich in hardware and software details.

Dynamic information is also represented on the map but comes from robot operation. When the robot is operating, moving a hand for example, it is possible to highlight on the map the units corresponding to the active modules and the unit corresponding to the hand. This is a part of the robot state representation.

According to Figure 2 this Nao State Representation is supplied to a Semantic Bridge that analyses the representation and gives as an output a set of information (semantic labels) that is used to activate the right concepts in the ontology.

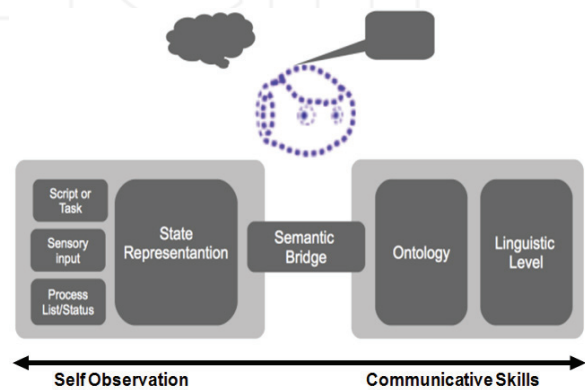


Figure 2. The proposed architecture for introspection

The Linguistic Level exploits these activations in order to perform a verbal interaction with the human user. In the present implementation the Semantic Bridge is constituted by a set of labels. Figure 3 illustrates in more detail how the modules provided achieve the introspection, when the robot is moving his hand:

- the introspective task module accounts for labels assigned by the programmer (e.g., `slow_movement_rigth_hand.py`) and allows queries on the documentation map to select units
- the perceptual module (`visual_recognition_visual_tracking_moving_object.py`) identifies the salient points of the map in the recognition and tracking by the vision of a moving object on a static scene
- the inner state monitoring module contains the updated list of active modules at regular intervals of time (or even real-time).

In several words it represents the monitoring of the "brain" by activating on the NAO map, which he observes as a part of himself. The collection of these

states becomes a model to characterize the situation (slow movement, moving a physical part such as the right hand, visual recognition, visual tracking). The map derived from the available documentation of the NAO, could also be generated with additional documentation on the web.

3. Self observation

The Nao state representation is obtained by using the information in documentation, the list of processes running on Nao system and the information from the sensor input. In the following subsections we show the structure module involved in this process (see Figure 3).

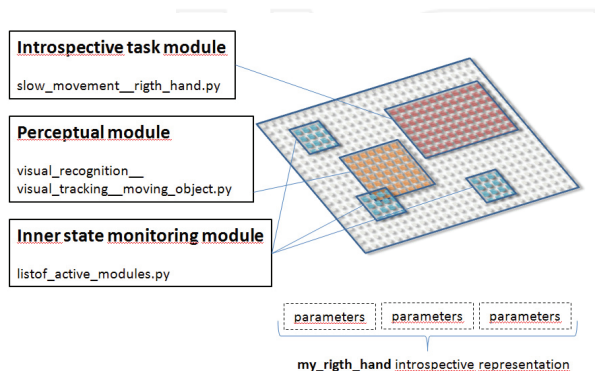


Figure 3. Modules involved in introspection.

3.1 Offline Processing Data

The main goal of this part of the work is to build a representation of the internal state of the robot, mixing together the information related to the robot hardware and the information available about software modules running during robot operation.

This representation should be rich enough to represent the perception of the robot body from sensors, to distinguish between different kinds of sensors and also to describe the list of active processes that control the robot movements and any other useful state component.

The robot documentation collects information about the robot hardware and software, so it is a good starting point for this kind of representation. Nao documentation is based on a corpus of HTML pages and a few introductory pdf documents. The HTML pages represent the core documents and are divided into 3 sections: Green Docs, Blue Docs and Red Docs. The Green Docs are introductory documents and describe the software installation, robot connection to the computer and hardware devices. There are also the descriptions of the Nao simulator and an introduction to Choreographer, the graphical programming environment. The Blue Docs are the documentation of Nao APIs: each page presents a short module description with the list of available methods and their descriptions. The described modules

are related to standard Nao behaviours, such as finding a marker and face detection, or related to hardware drivers for LEDs, lasers, infrared and Bluetooth. There is also documentation for other services, such as file manager, robot pose and text to speech. The Red Docs are related to *Naoqi*, the programming framework and the SDK, together with a set of documents, mostly images, related to the hardware details of the robot parts. The whole document collection is made by 422 HTML pages and three pdf documents.

In order to obtain a suitable document representation, which is useful for clustering, a sequence of standard pre-processing techniques was applied. Firstly, all the graphical details were stripped from the HTML pages, including details such as sidebars and footers that do not convey any information, then the stopwords were cancelled and the remaining words were stemmed using the Porter algorithm [37]. A vocabulary of about 4000 words was obtained after these processing steps. Several words were also deleted because they do not convey any information in this context (e.g., robot, Nao, or Aldebaran, which is the robot producer). The remaining words were ordered using their frequency and the 400 most common words were used to represent the documents using the *tfidf* (term frequency inverse document frequency) representation technique [4]. The output of this document vector representation is a set of vectors, one for each document, in a 400-dimension space where the distance between vectors is a function of the semantic distance among the documents. Each component of these vectors corresponds to one of the 400 keywords used to represent the documents. These vectors can be clustered using any suitable algorithm.

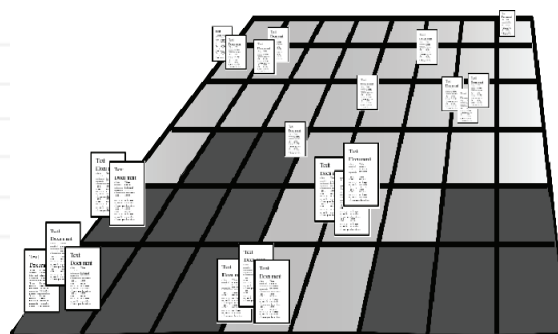


Figure 4. A representation of a section of the document SOM Documents semantically related are associated to the same neural unit.

Self Organizing Maps (SOMs) [25] are able to build a topographic map of vector distribution in a multidimensional space. In these maps objects are grouped according to their similarity and groups of similar objects are close on the map. SOM networks are already used to cluster documents [26, 30] and we employ such an approach to organize the NAO document set. In particular, the SOM used is a square map made

Also, among them regions with little dimension are detected as objects, while they don't have the characteristics of a relevant items in the scene. Filtering the areas through morphology functions with suitable structuring elements can filter the irrelevant sparse areas, preserving the larger homogeneous areas. The portion of the images that are captured by cameras are processed as previously described and are characterized with low level features to represent them in a suitable form for downstream processing.

Low level visual features describe the image content with a sequence of values that can be interpreted as the projection of the image in a particular feature space. The most used features are based on colour representation with histograms in different colour spaces and a description of image texture with statistical descriptors.

The feature values extracted from a large set of images are not evenly distributed but tend to have different densities in the vector space. The centroids of the regions, shaped by feature distribution, can be considered as forming a base for data representation and any image can be represented as a function of these points called *visual terms* [13]. These terms act as tokens that can be used to map the single feature values. The *K*-means algorithm has been widely used in automatic image annotation [13, 17]. An alternative is to extract the *visual terms* by applying Vector Quantization to the entire set of the characteristic vectors. As an alternative approach, the codebooks are produced by the LBG algorithm [31], ensuring less computational cost and a limited quantization error [41] [42].

Since the extraction of the *visual terms* is data driven, it allows the representation to emerge from the data set and build generic sets of symbols with a representation power that is limited only by the coverage of the chosen training set. The *visual terms* calculated from multiple features from a training set can be combined allowing particular information aspects of the visual information to be captured according to multiple characteristics. Feature statistics in the image are dependent from the feature itself and are a function of its statistical occurrence in the image.

A more powerful representation can be adopted to exploit the spatial distribution of the *visual terms* in the visual input. In this case not only the presence of a *visual term* is used to represent an image but also the presence in a spatial window of a pair of *visual terms*. This representation is analogue to the representation of bigrams where the words are given by the found *visual terms* [41].

The increased expressivity of the bigrams allows over performing of the results achieved with the simple *visual*

terms (unigrams), although at the cost of an increased dimensionality for image representation.

The representation with *visual terms* and visual bigrams allows the relevant visual information to be represented by a vector that is suitable for the downstream cognitive elaborations.

3.3 Process list state

The Nao software platform supplies a list of processes available on the Nao system.

Figure 8 shows the web page that contains the active processes.

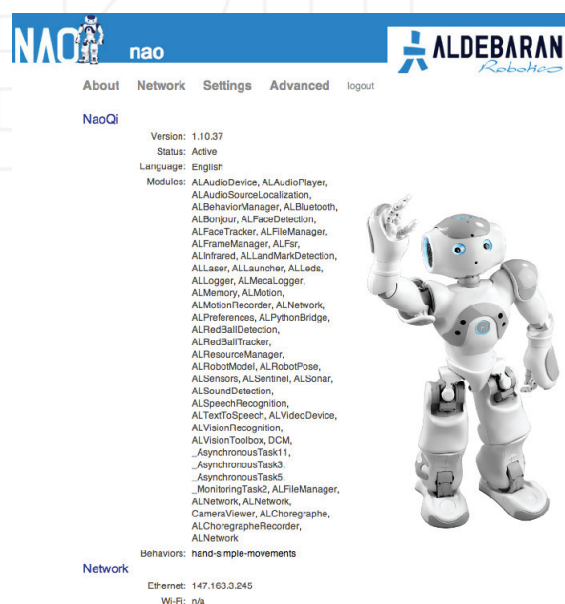


Figure 8. The web page containing the active processes

This page is composed of many sections: the first one contains the information related to the Modules loaded in the memory; each module name is directly linked to the corresponding documentation page. This makes it easy to obtain a representation of the available NAO modules on the document SOM, a representation that is integrated with the others in the State Generation block in Figure 9.

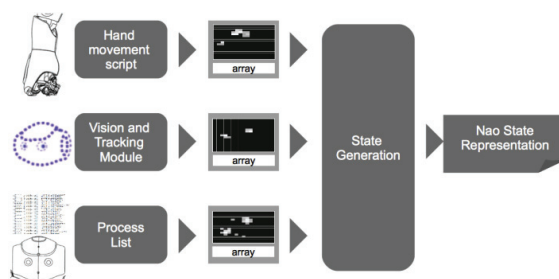


Figure 9. The procedure that calculates a Nao State Representation. Each part of the Nao generates a state representation on the Map together with an array of physical configuration parameters; these are combined in order to obtain the Nao State Representation.

This background information is merged to the behaviours section of the same page, where all the scripts that are running on the Nao system are listed. These scripts are also mapped in the document SOM using the available set of keywords (see Section 3.1).

Finally a third part of the Nao state is obtained using some custom scripts that output the list of processes running on the Nao system. This list depends on the behaviours that are required by the robot.

4. The Semantic Bridge

The trained map can be used as a search tool: if a query text can be represented in the same space as the training documents (i.e., contains at least one of the 400 keywords used to represent the documents) it is possible to calculate the activation of the SOM neural units as a normalized distance from the query text; the nearest units are the most active, while other units have a lower activation according to their distance. If we set an activation threshold and we convert the activation in a level of grey scale, it is possible to obtain an image as a representation of the response to the input.

Figure 10 shows the activation pattern for the queries "hand", "arm", "joint" and "leg". As Figure 4 shows, it is possible to identify the position of a single file on the document map. Using a tool that queries the robot system and retrieves the list of running modules it is possible to represent on the map the set of running processes in the robot, since each process has a documentation page. The obtained representation is similar to the one illustrated in Figure 10.

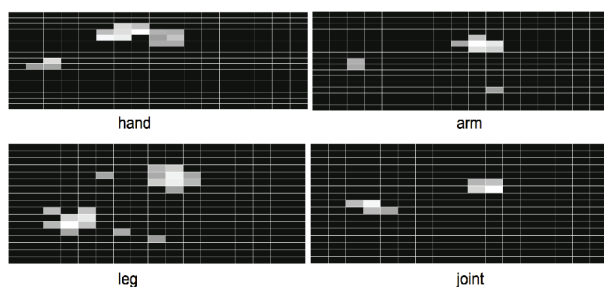


Figure 10. Activation Map for four queries

These representations are mixed with the representation of the scripts used to move the robot parts and with the activation of the sensory inputs traced on the map using suitable keywords like "vision" or "eye" or "camera". Figure 10 shows how the system combines this information to build the NAO State Representation.

Moreover, each one of the fingerprint images is also linked to a vector of parameter values used to quantify some details (for example, the list of the angle values of joints during a movement).

The semantic bridge subsystem should obtain a set of suitable labels from the Nao internal state representation that can be considered as a set of images, as previously described. In the system presented here the implementation of the semantic bridge uses a look-up table that connects the map state images to the labels in the ontology concepts.

This is an easy shortcut that exploits the labels in the document map shown in Figure 5. A particular Nao internal state highlights some units (see Figure) together with the corresponding keywords.

The look-up table is a preliminary implementation that builds rigid mapping between the two subsystems. We plan to study a more refined solution for this multidimensional mapping problem. One suitable replacement can be a multilayer perceptron, but we are still investigating the performances of this solution.

5. Communicative Skills

The Nao robot has an internal knowledge of its physical structure and functionalities. This knowledge is derived from the state representation and can be exploited to support direct communication on the perception capabilities of the robot and to describe his state to a human interlocutor by using natural language. Moreover, modern semantic tools and introspection capability can be exploited together to improve and support direct communication to the robot perception mechanisms.

5.1 Ontology by Cyc Knowledge Base

The Cyc knowledge base (KB) at present appears to be an effective tool for automated logical inference aimed at supporting knowledge-based reasoning applications [11]. Besides, it is the largest and most complete general KB equipped with a performing inference engine.

The Cyc knowledge base is composed of particular collections of concepts and facts concerning a specific domain. These collections are called Microtheories (Mt).

A set of Java APIs facilitates the interaction with Cyc. It supports interoperability among software applications, it is extensible, it provides a common vocabulary and it is well-suited for mapping to/from other ontologies.

The Cyc technology has been chosen to be embedded into the Nao introspective system because of its suitability for common sense representation. Such a feature, together with a more abstract ontology, makes it possible to extend inferences to more generic facts.

All the information regarding the structure of the Nao robot, has therefore been coded into an ontology that we

have implemented into the ResearchCyc Commonsense Knowledge Base.

We have coded relations, concepts, constraints and rules regarding the Nao robot domain. These concepts have been organized in order to fulfil the self-observation task.

The produced ontology involves concepts and relations related to processes running in the robot, its physical parts and its capabilities. We have tried to exploit as the existing concepts and predicates of ResearchCyc as much as possible in carrying out this task.

The domain complexity requires the creation of suitable relations between the concepts in order to implement a high-level introspection mechanism.

As an example we report in the following some of the most common predicates and concepts, among the ResearchCyc predefined items that we have used:

- `#$genls` predicate: relates a given collection to those collections that subsume it
- `#$isa` predicate: relates things of any kind to the collections of which they are instances. It is the most commonly used predicate in the Cyc Knowledge Base and is one of the relations most fundamental to the Cyc ontology
- `#$Robot`: a device which operates autonomously, moving about or manipulating physical objects
- `#$RobotHand`: the collection of mechanical devices that can manipulate objects and/or materials automatically.

In the following we report a sample of items (constants and predicates) that we have defined and inserted into the ResearchCyc KB in order to fulfil our goal:

- `#$Nao-TheRobot`: the collection of the Nao robots;
- `#$IcarNao`: an instance of `#$Nao-TheRobot` and is the robot that we have used for the experiments in our lab.

An example of the assertion that we have inserted into the ResearchCyc KB is:

- `(#$managesToPerformActOfType #IcarNao #ComputerActivity) #IcarNao manages to perform an instance of #ComputerActivity;`

A typical query to the Cyc inferential engine is:

- question: `(?X #Robot ?Y);`
- answer: `(#$properPhysicalPartTypes #Robot #RobotFoot) in BaseKB`

```
(#$genls #Robot #ElectronicDevice) in
UniversalVocabularyMt
($comment #Robot "A device that operates
autonomously, moving about or manipulating
physical objects") in UniversalVocabularyMt
($isa #Robot #SpatiallyDisjointObjectType) in
UniversalVocabularyMt
....
```

From the above example we can see that different results for one single query are obtained and each one may reference to different constants and predicates, exploiting the high potential of the Cyc KB.

5.2 Linguistic Level

The linguistic level is aimed at interpreting natural languages query given by the user. This level exploits a classical pattern-matching technique enhanced with Cyc ontology inference capability. This feature is obtained by transforming natural language requests into symbolic queries, expressed in the ontology language. Such commands are forwarded to the ontology engine that computes the appropriate inferences and gives results in a symbolic form. The symbolic answers are then transformed by the linguistic module into natural language sentences that are finally shown to the user.

The linguistic level has been implemented by using A.L.I.C.E. (Artificial Linguistic Internet Computer Entity) technology [1]. The interaction with the user is based on algorithms for automatic detection of patterns in the dialogue data.

The A.L.I.C.E chat-bot's knowledge base is composed of question-answer modules, named categories and structured with AIML (Artificial Intelligence Mark-up Language) an XML-like language. The question of the user is described by a pattern tag while the answer by a template tag. Other tags grant the management of more complex answers.

In the following we illustrate the simplest form of category constituting the A.L.I.C.E knowledge base. Every time the user says "Hi", the system answers with "Hello there! What's your name?"

```
<category>
  <pattern>HI</pattern>
  <template>Hello there! What's your name?</template>
</category>
```

To exploit the ontological knowledge base we have used the CYN tool [12] which constitutes a bridge between an AIML interpreter and Cyc. This tool grants an easy interface between AIML language and Cyc in order to answer the questions of the user by querying the Cyc inference engine.

In CYN new tags were created in order to interact with the Cyc KB to achieve maximum flexibility with minimum overheads. As an example “$\\$$\\$$\\$$\\$” translates an English word/phrase into a Cyc symbol, “$\\$$\\$$\\$$\\$” executes a CycL statement and returns the result, “$\\$$\\$$\\$$\\$” executes a CycL query and returns one response at random and so on.

The following example shows a very simple AIML category concerning the interaction with Cyc:

```
<category>
<pattern>
WHAT IS A ROBOT
</pattern>
<template>
A robot is a
<cycsystem>(cyc-query ' (#$isa #Robot ?X) ) </cycsystem>.
</template>
</category>
```

Every time the user queries the Nao Robot with the question “What is a robot?” the ResearchCyc inference engine is activated by using the construct “(“$\\$$\\$$\\$$\\$”)” and the answer is wrapped as a natural language sentence.

5.3 Verbal Interaction examples

The following are two examples of possible dialogue between a human user and the robot Nao. The first, based on the ontology that uses Cyc, has already been implemented and will be improved in the future. The second example is based on a case envisaged in the design of the system of introspection, which is still under refinement, but which is important because it shows some objectives that we intend to achieve.

1. HUMAN: Who are you?
2. NAO: I am Nao-TheRobot, Agent-Generic, Robot, ElectronicDevice and other things.
3. HUMAN: What kind of robots do you know?
4. NAO: I know these kinds of Robots: Robot, Nao-TheRobot, Android.
5. HUMAN: Could you tell me your components that have been the subject of introspection?
6. NAO: The components that have been subject of introspection are:RobotFoot, RoboticArm, RobotHand.
7. HUMAN: What kind of actions can you take?
8. NAO: I can take these actions: RecognizeSomething, Executingstate, ResumingAnActivity, FulfillingAnObligation, ChangingDeviceState, ComputerActivity
9. HUMAN: Could you list your main functional modules?
10. NAO: My main functional modules are: InnerstateMonitoringModule, PerceptualModule, IntrospectiveTaskModule.
11. HUMAN: May you describe a functional module?
12. NAO: A functional module can be ComputerProcessRunning, ComputerActivity, Intangible and many other things.

The information gathered through introspection enables the robot to communicate to a human user important information regarding the modules involved in performing a given task, the relevant data stored and indicate the documents that contain further details.

Three elements compose the information obtained from introspection: the first one is derived from the execution of a movement (configuration of the joints, speed, kinematic chains involved) or motor information, the second one detects the parts of the robot involved (visual bi-grams) through visual information and the third one is derived from the knowledge of the active modules running at any given time or information concerning the internal state.

Similarly, the robot will be able to account for the details of each one of these components, describing the functions involved to a user who wants to learn something about the NAO.

1. HUMAN : How do you *move* your *right hand*?
2. NAO:
 - I have to move the kinematic chain named 'RARM', assigning values to the angles of the following joints: 'RShoulderPitch', 'RShoulderRoll', 'RElbowYaw', 'RElbowRoll', 'RWristYaw', 'rhandir', 'HeadPitch', and deciding what fraction of the speed (fractionMaxSpeed) of execution of the movement must be used.
 - I memorized the characteristic aspects of my right hand through vision and I can recognize it through my camera, which distinguishes it from the left, or from the hand of another similar robot. I can follow it with my cameras while I perform operations.
 - I stored description of the modules that are active during the use of right hand and I can describe them.
3. HUMAN: I would like more details on the *movement*. What are the values of the angle that can take *RWristYaw*?
4. NAO: Look in the documentation RED (Advanced), under the Motion section
5. HUMAN: I would like more details on *perception*. What are the *visual descriptions* of the *right hand*?
6. NAO: The list of bi-grams is in the file named extracted visual bgvisual_RHHand.dat
7. HUMAN: I would like more details on the *internal state*.
8. NAO: Here is a list of active modules: ALAudioDevice, ALAudioPlayer, ALAudioSourceLocalization, ALBehaviorManager, ALBluetooth, ALBonjour, ALFaceDetection, ALFaceTracker, ALFileManager, ALFrameManager, ALFsr, ALInfrared, ALLandMarkDetection, ALLaser, ALLauncher, ALLeds would stay, ALMecaLogger, ALMemory, ALMotion , ALMotionRecorder, ALNetwork, ALPreferences, ALPythonBridge,

ALRedBallDetection, ALRedBallTracker,
 ALResourceManager, ALRobotModel, ALRobotPose,
 ALSensors, ALSentinel, Alson, ALSoundDetection,
 ALSpeechRecognition, ALTextToSpeech,
 ALVideoDevice, ALVisionRecognition,
 ALVisionToolbox, DCM, _AsynchronousTask15,
 _AsynchronousTask3, _AsynchronousTask5,
 _MonitoringTask2, ALFileManager, ALNetwork ,
 ALChoregraphe, ALChoregrapheRecorder,
 ALNetwork

9. HUMAN: What does *ALSpeechRecognition* do?
10. NAO: Find information on this module in the following documents: Red doc - Sound system: *ALSpeechRecognition* (Tutorial), Red Doc - Sound system: *ALSpeechRecognition* (Introduction), Doc Blue - *ALSpeechRecognition*: *ALSpeechRecognition* API.
11.

In Step (1), the words "move", "hand", and "right", belong to the reference dictionary and are recognized by the introspective system and in this case correspond exactly to the labels of an introspective task already performed. In the rest of the dialogue all the words in italics are items from the dictionary.

In Step (2) the NAO shows everything that can be said about the movement of his right hand, according to the introspective module and relevant knowledge related to it.

6. Conclusions

In this paper we have described an approach for a humanoid robot to understand its internal state. The method is based on self observation and communication with the external world, according to the idea of introspection given by Sloman. Robot introspection arises from information about physical components and software modules. This information is translated in a spatial representation of the hardware and software components of the robot through a SOM.

We have shown that the SOM can be an effective way to realize a semantic bridge that links the state representation of the robot with a high level representation given by an ontology. The ontology is furthermore linked to a linguistic module that makes the interaction with human beings through a conversational agent architecture possible.

We have shown a proof-of-concept procedure that illustrates the effectiveness of the approach. In particular we have demonstrated that it is possible to obtain semantic links among entities related to the running processes, libraries used and physical components of the robot.

Furthermore, we have also shown that the robot is capable of autonomously filling the "gaps of knowledge of himself", to create new hypotheses to be tested with the aim of discovering new possible relationships.

7. Acknowledgments

Financial support of the research is partially given by Regione Sicilia, PO FESR 2007/2013 Linea d'intervento 4.1.1.1. Title of Project: IDS – INNOVATIVE DOCUMENT SHARING.

8. References

- [1] <http://alicebot.blogspot.com/> (Accessed 2012 Dec 8).
- [2] Anderson JR, Bothell D, Byrne MD, Douglass S, Lebiere C and Qin Y (2004) An integrated theory of the mind. *Psychological Review*, 111(4):1036-1060.
- [3] Aydede M and Guzeldere G (2005) Cognitive Architecture, Concepts, and Introspection: An Information-Theoretic Solution to the Problem of Phenomenal Consciousness, *Nous*, 39(2):197-255.
- [4] Baezo-Yates R and Ribeiro-Neto B (1999) *Modern Information Retrieval*. Addison-Wesley.
- [5] Barakova EI and Lourens T (2009) Mirror neuron framework yields representations for robot interaction, *Neurocomputing, Brain Inspired Cognitive Systems (BICS 2006)/Interplay Between Natural and Artificial Computation (IWINAC 2007)*, 72(4-6):895-900. doi: 10.1016/j.neucom.2008.04.057.
- [6] Barsade SG (2002) The Ripple Effect: Emotional Contagion and Its Influence on Group Behavior, *Administrative Science Quarterly*, 47(4):644-675.
- [7] Barnard K, et al (2003) Matching words and pictures, *Journal of Machine Learning Research*, 3:1107-1135
- [8] Birlo M and Tapus A (2011) The crucial role of robot self-awareness in HRI, *Proc. of the 6th Intl. conf. on Human-robot interaction HRI'11*, ACM. doi>10.1145/1957656.1957688
- [9] Cerezo E, Baldassarri S, Hupont I and Seron FJ (2008) *Affective Embodied Conversational Agents for Natural Interaction*, *Affective Computing*, Jimmy Or (Ed.), ISBN: 978-3-902613-23-3, I-Tech Education and Publishing.
- [10] Chaminade T and Cheng G (2009) Social cognitive neuroscience and humanoid robotics, *Journal of Physiology-Paris*, 103:286-295.
- [11] <http://research.cyc.com/> (Accessed 2012 Dec 8).
- [12] http://www.cyc.com/doc/white_papers/Cyn_description.pdf (Accessed 2012 Dec 8).
- [13] Duygulu P et al. (2002), Object recognition as machine translation: Learning a lexicon for a fixed vocabulary, *Proc. of ECCV*.
- [14] Fischler MA and Bolles RC (1981) Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Comm. of the ACM*, 24:381-395.

- [15] Gazzola V, Rizzolatti G, Wicker B and Keysers C (2007) The anthropomorphic brain: The mirror neuron system responds to human and robotic actions, *NeuroImage*, 35(4):1674-1684, ISSN 1053-8119, doi: 10.1016/j.neuroimage.2007.02.003.
- [16] Goodrich MA and Schultz A (2007) Human-Robot Interaction : A Survey, *Foundations and Trends in Human-Computer interaction*, 1(3):203-275.
- [17] Fogassi L (2011) The mirror neuron system: How cognitive functions emerge from motor organization, *Journal of Economic Behavior and Organization, Emotions, Natural Selection and Rationality*, 77(1):66-75, doi: 10.1016/j.jebo.2010.04.009.
- [18] Hatfield E, Cacioppo JT and Rapson RL (1994) Emotional contagion. *Current Directions, Psychological Science*, 2:96-99, Cambridge Univ. Press. doi: 10.1111/j.1467-8721.1993.tb00114.x
- [19] Hommel B, Musseler J, Aschersleben G and Prinz W (2001) The Theory of Event Coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences*, 24:849-878.
- [20] Honkela T, Kaski S, Lagus K and Kohonen T (1997) WEBSOM - Self-Organizing Maps of Document Collections, *Neurocomputing*, 21:101-117
- [21] Infantino I, Rizzo R and Gaglio S (2007) A framework for sign language sentence recognition by commonsense context. *IEEE Transactions on Systems, Man, and Cybernetics, Part C*. 4:1034-1039.
- [22] Infantino I, Lodato C, Lopes S and Vella F (2008). Human-humanoid interaction by an intentional system, In *proc. of 8th IEEE-RAS Intl. Conference on Humanoids 2008*, 573-578, 1-3 Dec. 2008.
- [23] Kelley R, Tavakkoli A, King C, Nicolescu M, Nicolescu M and Bebi G (2008) Understanding human intentions via hidden Markov models in autonomous mobile robots. In *Proc. of the 3rd ACM/IEEE Intl. Conf. on Human robot interaction (HRI '08)*. ACM, New York, NY, USA, 367-374.
- [24] Kelley R, Tavakkoli A, King C, Nicolescu M and Nicolescu M. (2010) Understanding Activities and Intentions for Human-Robot Interaction, *Human-Robot Interaction*, Daisuke Chugo (Ed.), ISBN: 978-953-307-051-3, InTech.
- [25] Kohonen T (1995) *Self-Organizing Maps*, Springer-Verlag, Berlin.
- [26] Kohonen T, Kaski S, Lagus K, Salojärvi J, Honkela J, Paatero V and Saarela A (2000) Self-Organization of Massive Document Collections, *IEEE Transactions on Neural Networks*, 11(3):574-585.
- [27] Landauer TK and Dumais TS (1997) A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge, *Psychological Review*, 104(2):211-240.
- [28] Landauer TK, Foltz PW and Laham D (1998) An introduction to latent semantic analysis, *Discourse Processes*, 25:259-284.
- [29] Langley P, Laird JE and Rogers S (2009) Cognitive architectures: Research issues and challenges. *Cognitive Systems Research*, 10:141-160.
- [30] Lagus K, Kaski S and Kohonen T (2004) Mining massive document collections by the WEBSOM method, *Information Sciences*, 163(1-3):135-156.
- [31] Linde Y, Buzo A and Gray R (1980) An Algorithm for Vector Quantizer Design. *IEEE Transaction on Communications*, 28(1):84-94
- [32] MacDorman KF and Ishiguro H (2006) The uncanny advantage of using androids in cognitive and social science research. *Interaction Studies*, 7(3):297-337, doi: 10.1075/is.7.3.03mac.
- [33] Malatesta L, Murray J, Raouzaoui A, Hiolle A, Canamero L and Karpouzis K (2009) Emotion Modelling and Facial Affect Recognition in Human-Computer and Human-Robot Interaction, *State of the Art in Face Recognition*, Julio Ponce and Adem Karahoca (Ed.), ISBN: 978-3-902613-42-4, I-Tech Education and Publishing.
- [34] Marinier RP III, Laird JE and Lewis RL (2009) A computational unification of cognitive behavior and emotion, *Cognitive Systems Research*, vol. 10(1):48-69 doi: 10.1016/j.cogsys.2008.03.004.
- [35] Novianto R and Williams M (2009). The role of attention in robot self-awareness. In *Proc. of the 18th IEEE Intl. Symposium on robot and Human Interactive Communication*, Toyama, Japan, Sept.27-Oct.2, 2009.
- [36] Pantic M, Pentland A, Nijholt A and Huang TS (2006) Human Computing and Machine Understanding of Human Behavior: A Survey, *Proc. of the Eighth ACM Int'l Conf. Multimodal Interfaces (ICMI '06)*, 239-248.
- [37] Porter MF (1980) An algorithm for suffix stripping, *Program*, 14(3):130-137.
- [38] Roy D (2007) A mechanistic Model of Three Facet Meaning. *A Mechanistic Model of Three Facets of Meaning, Symbols, Embodiment, and Meaning*, de Vega, Glenberg and Graesser, eds.
- [39] Sloman A (2010) An Alternative to Working on Machine Consciousness. *International Journal of Machine Consciousness*, 2(1):1-18 doi: 10.1142/S1793843010000400.
- [40] Trafton JG, Cassimatis NL, Bugajska MD, Brock DP, Mintz FE and Schultz AC (2005) Enabling effective human-robot interaction using perspective-taking in robots. *IEEE Transactions on Systems, Man, and Cybernetics*, 35(4):460-470.
- [41] Vella F, Lee CH and Gaglio S (2007) Boosting of maximal figure of merit classifiers for automatic

image annotation. Proc. of IEEE International Conference on Image Processing.

[42] Vella F and Lee CH (2007) Information fusion techniques for automatic image annotation. Proc. of 2nd International Conference on Computer Vision Theory and Applications, 2007.

[43] Zeng Z, Pantic M, Roisman GI and Huang TS (2009) A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions. IEEE Trans. on Pattern Analysis and Machine Intell, 31(2):39-58.

INTECH

INTECH