Ideograms Representation for Cognitive Systems in Robotics

María D. R-Moreno
Universidad de Alcalá
Escuela Politécnica Superior, Campus Universitario
Ctra. Madrid-Barcelona, km 33.6
Alcalá de Henares, 28871 Madrid
malola.rmoreno@uah.es

Adrian Stoica
Jet Propulsion Laboratory
California Institute of Technology
4800 Oak Grove Drive
Pasadena, CA 91109
adrian.stoica@jpl.nasa.gov

Abstract

The cognitive psychology has focused on understanding how humans think and solve problems. As analogy, the robotics systems have used this research to try to make robots more "intelligent" imitating cognitive models. How things are perceived and transformed in mental structures, and how they are retained, is still an open question, although a lot of progress has been done. For example, some studies have demonstrated that the use of diagrams retains concepts in memory better than other approaches based on text or words, and making the reasoning about them easier.

In this direction, we want to propose an ideogram representation for robotic systems in cooperation with humans that can help to provide a natural link between the determination of perceptive symbols and their manipulation in reasoning.

Introduction

The fields of cognitive psychology and robotics share many goals and interest. For example, the development of robot cognitive architectures is an attempt to apply the results of cognitive modeling to the difficult problems faced by robotics systems in real world applications.

Cognitive psychology focuses on understanding the mind, that is, the way the humans think, perceive, remember, process the information and solve the problems. The term cognition refers to all processes by which the information can be transformed, stored, recovered, and used in order to gain knowledge and understanding. Then, memory (storage of information) occupies an important place inside the cognitive model. Different models have been proposed in cognitive psychology in order to understand how the memory works. We don´t want to get into many details of which one is better and pros and cons, but one of the more referenced one is the one based on three components (names also may vary):

- Sensory memory or "active memory" (SM) stores a small amount of information in mind for a short period of time. Some of the information in this memory moves into long-term memory thanks to the consolidation process while kept active in the working memory.
- Working memory (WM) refers to the structures and processes used for temporarily holding and accessing the information in the short-term memory.
- Long-term memory (LM) stores an unlimited amount of information for the rest of the life, although it cannot always be available since the depth of retained process influences the forgetting process.

Therefore, we need to analyze what makes that certain concepts are better retained than others. Many researchers have studied different meta-cognitive strategies that can aid to recall past events such as text, diagrams or sounds. Some studies have shown the effectiveness of diagrams or draws in the memory process (Lambert, 2007). We want to analyze this option in the design of cognitive robotics. For that, we propose a perceptual representation in memory called perceptograms that can generate large number of propositions depending of what we want to infer. This contrast to the pure predicate-symbolic representations where initial information is translated into predicates having in mind some relevant information that we want to consider. Once this is translated, no more information can be represented. That is, the latter involve qualitative abstractions, casting away metric information, and thus cannot support new relational perceptions (Chandrasekaran, 2006).

Then, we need to save in memory structures easy to handle with minimal but enough information that add symbolic predicates when necessary. The perceptive symbol, "perceptogram" is a minimal representation in a perceptive space. An ideogram is a simple combination (idea) of perceptograms, reflecting a linkage between them. In visual perception a perceptogram is a graphical "sketch" - a minimal visual representation composed of graphical (primitive) components (e.g. straight lines, angle set).

We describe a way to simply transform and uniquely convert visual images into perceptograms. Multiple

978-1-4799-7007-0/14 $31.00 © 2014 IEEE
DOI 10.1109/EST.2014.21

perceptograms can be associated with the same world object (e.g. encoding different viewpoints, etc). Then, storage in memory is much simpler and its translation into a symbolic representation can be as large as needed in each situation. The paper is an attempt to define how this can be done. We have tried to define the process as generic as possible that any robotic architecture could adopt it.
The structure is as follows. Next section reviews the main architectures in robotics. Then, we define the perceptual concept *ideograms*. After, a cognitive architecture proposal, called CIbAr (Cognitive Ideogram-based Architecture) based on that concept is presented. Finally, conclusions are outlined.

Robotic Architectures Review

We can group robotic architectures in four main categories: Cognitives, Behavior-based, 3T and Multiagent. In any of these architectures, the set of tasks or plans to perform should be generated. The plan generation can be manual (no action explicit representation is needed) or automatic. In this last case, state of the art planning techniques are group in two approaches:

- Action-oriented: it uses predicates logic and the world is seen as an entity that can be in different states. The domain specifies actions that can be performed to change the state of the world and only applicable when some particular states are set. States and actions are represented using standard language PDDL (Gerevini and Long, 2006). The objective is to find a sequence of actions that, from an initial world state, through applying successive actions, the system achieves a desired goal state.
- Timelines-based: it represents the world in terms of functions that describes the behavior of the system from a time perspective: a timeline is a logical structure used to represent and reason about the evolution of an attribute over a period of time. Rules must be defined to specify how the timelines can change, in order to obtain a sequence of decisions from the planner that brings the set of temporal functions to a final state in which a set of constraints are satisfied.

Cognitive architectures are those who try to imitate the mental human processes that define the intelligent behavior. This implies problem solving, and, specially, learning capabilities. Many cognitive architectures for robotic systems have been proposed (Langley et al, 2008), trying to cover from high-level to low-level range of capabilities. However, most of them are biased towards one or another end of this spectrum. We can mention SOAR (Laird et al., 1987), ICARUS (Langley and Choi, 2006) or ACT-R/E (Kennedy et al., 2009). They are centered around unified amodal representations and mainly dealing with processing of high-level information with less emphasis on grounding this high-level information in raw sensorial data.
Other architectures are **behavior-based** and they focus on using the world as its own model and react accordingly. They are modeled as finite state machines since the behavior is deterministic. When a detector detects a strange behavior, there is a program based on rules that reacts and defines the behavior to follow. The role of reaction is to control real time behavior while the role of sequencing is to generate series of real time behaviors. Examples of this architecture, known in the literature as Reactive architectures, are the Subsumption architecture (Brooks 1991), TDL (Simmons and Apfelbaum, 1998) or Reactive Action Packages (RAP) (Firby, 1987). Figure 1 shows the basic elements of the RAP architecture extracted from Firby (1987) since it can be considered as the more representative of this category.

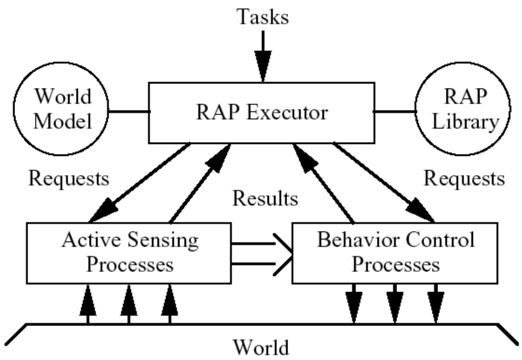

Figure 1: RAP architecture (Firby, 1987)

An intermediate approach between the previous types is 3T (Bonasso et al, 1997) or **3 layer** architectures (see Figure 2). The higher layer corresponds to the deliberative (planning) system, which looks for the activity sequences to satisfy high-level goals, by considering both temporal and resource constraints as well as initial assumptions defined by the domain theory. The low layer is an abstraction of the hardware system that we want to control. The middle layer is the execution system that executes pre-planned actions. That is, dispatches (low-level) action sequences (commands) directly translated from the plans provided by the higher levels into real actions (coupled motor and sensor processes to control the robot hardware) by modifying the world by means of the actuators Each layer uses a different model so there is a lack of integration between components and a clearly separation between deliberation and reaction. Other architectures based on this approach are: ATLANTIS (Gat, 1992), Coupled Layer Architecture for Robotic Autonomy (CLARAty) (Nesnas et al., 2006) Tripodal Schematic Control Architecture

(Kim and Chung, 2006) or Remote Agent (RAX) (Muscettola et al, 98).

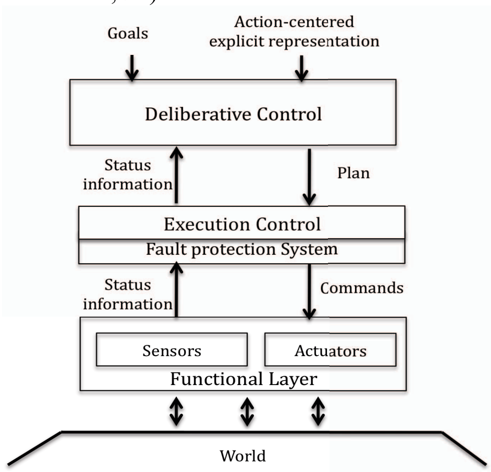

Figure 2: 3T architecture

Some 3T layer architectures have evolved into **Multi-agent** systems. The architectural decomposition into different agents that could involve reactive, deliberative or both capabilities. Combining the capabilities of each agent is possible to solve complex tasks between the cooperation of the different agents. The idea is to divide the problem into subproblems that are interconnected. The architecture is described in a high level, defining the agents that will be part of the system, their roles, the interactions between them, the resources they need under the temporal restrictions imposed by the domain. Another problem that these architectures aim to face is the invalidation of the generated plan before it is completely executed, due to the changes in the real world. To do that, they implement a feedback mechanism to communicate the bottom layers with the (top) deliberative ones (see Figure 3). Thus, when execution incidences are detected, the feedback control could react in two ways: by using fast but short-sighted stored methods; or by notifying the incidence to the (upper) deliberative layers to request a more intelligent response. Some possible recovering actuations are: modify the former plan or replanning having into account the current circumstances, or give up completely and adopt a different goal.

We can mention in this group: IDEA (Intelligent Distributed Execution Architecture) (Aschwanden et al, 2006), VOMAS (Hsu, H.C.H. and Liu, A., 2007), TREX (Teleo-Reactive Executive) (Py et al, 2010) or GOAC (Ceballos et al, 2011). One inconvenient of this approach is how to obtain the right decomposition. Another limitation is that it encapsulates the different techniques into a vertical-communicated layered scheme: the formal and deliberative strategies are located at the highest levels of abstraction. Meanwhile the reactive behavior to unknown events is implemented at the bottom levels. These limitations imply high latency produced by the number of calls done to the planner during the execution process.

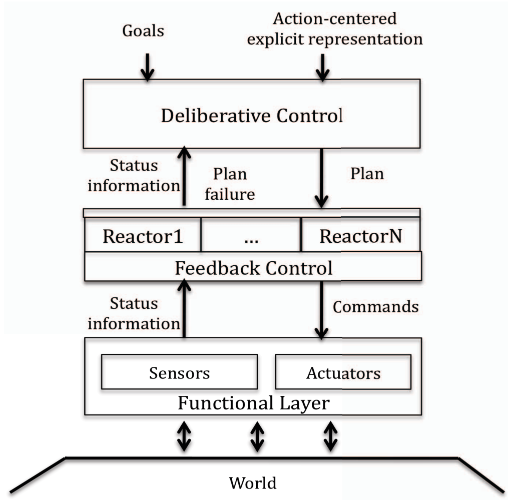

Figure 3: Multiagent architecture with several reactors

What are Ideograms?

The perceptive symbol (ideogram, perceptogram) is a minimal representation in a perceptive domain/field (domain of perception). An ideogram is a simple combination (idea) of perceptograms, reflecting a linkage (bond) between them.

The unique ideogramatic representation facilitates the communication between the determination of perceptive symbols (perception) and their manipulation in reasoning scenarios, 'living' through external manifestation in the form of actions. For example, figure 4 shows the Chinese ideogram for king. The king intermediates the connection between heaven and earth (middle line, half way between top (heaven) and bottom (earth) horizontal lines).

In the visual perception domain a perceptogram is a graphical sketch/schematic (a minimal(istic)/ simplified/distilled visual representation) with a minimum number of graphical components (e.g. straight lines, angle value set, etc).

Figure 4 – Chinese ideogram for king

The determination is made by a simplification (reduction) and projection on – and maximal correlation with axes of the system of representation. For example, from human shape, if we consider lines of equal length, arbitrarily of one unit, l=1, three lines are originated in the center of the figure (say center of the area). Then, the representation of a person could be, for example, *Man = l (Pi, Pi/12, -Pi/12)*

Figure 5a) shows the human shape, then figure 5b) the perceptogram derived from the shape, and figure 5c) the Chinese character 人 (Rén), meaning person.

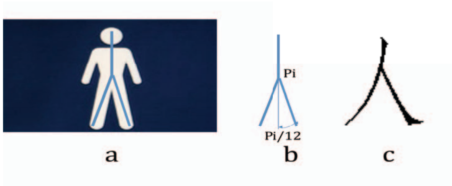

Figure 5 Perceptogram example derived from its shape

A word, e.g. man, is an equivalent simple representation in the space of language characters or sounds, expressing a link between its components (it is what brings together the 'm' the 'a' and the 'n'). The 'vector' base, or the set of primitives, determines the representation, thus Figure 6 shows several representations of increasing level of simplification, with one or two primitives (e.g. line and circle) and 24 variables (segments). Figure 6a) shows the original shape, then Figure 6b) shows the same image in 24 variables system. Figure 6c) uses 6 variables – 2 primitives; figure 6d) also uses 6 variables – 2 primitives; figure 6e) 5 variables, and figure 6f) 3 variables.

Once this mode of perception (as a process) if defined/determined/set, the perceptogram is also uniquely set/determined. So, in this context, perception can be defined as the set of transformations which, in an entropy decreasing process, it minimizes, in a controlled way, the information, aligning it along (correlating it with) a given set of (projective) dimensions (or representation). By minimizing 'redundant' information one realizes a form of compression.

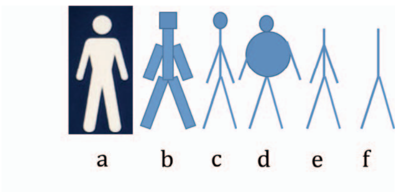

Figure 6 – Example of shape and perceptograms of increasing level of simplification.

Cognition – Perception: discovery of perceptograms

A perceptogram is formed in relation to a system of primitives on which one makes projections.

From an object, n variables, m parameters, results a set of perceptograms (perhaps this can be associated with primary directions of oscillations in the neurons). So in perception-driven cognitive development (intending here to express the process of getting to know, through perception) we assume that the projection system and granularity (discretization level) are known. Perceptograms will not be just arrays of pixels. They will correspond to the stage in perception where figure-ground separation has been performed, i.e., the image has been discretized, reducing the resolution. Figure 7 shows the process of reducing the image of a sheep into a perceptogram. This process, intended not to be computationally expensive, will allow us to save the essential information to perform late reasoning.

The mind (derives projections) correlates with pictograms that can rotate in 3D, but mostly in the plane we learned them, for most things being the horizontal plane and in maximum coverage/view.

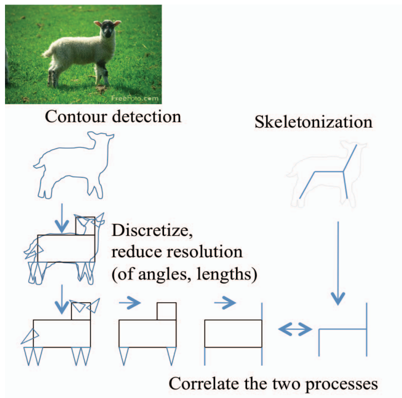

Figure 7 – Percetogram process of a sheep

For example, we recognize the sheep more easily if we watch it from the same plane, in profile, face or half side, rather than from above. The 'preferred' perceptogram is one in the observation with maximal information, e.g. if we see the sheep from side/profile, not from the back. In general they have an elongated dimension along the direction of movement. The main reason may be that in moving towards/from us there are less informative changes in consecutive moment of observation. In this context the observation perpendicular to direction of movement gives maximal changes.

The emotion influences the perception – we have representations of tiger faces, which may have triggered a strong emotion in the viewer, but rarely representations/models of tiger backs/bottoms.

Thus, m multiple perceptual representations (multiple perceptograms) in different perceptive domains, but also in the same perceptive domain (e.g. the sheep from side, ram head, etc) are bonded/tied together and potentially all activated when one is activated.

A Cognitive architecture based on Ideograms

Our proposal of cognitive architecture called CIbAr (Cognitive Ideogram-based Architecture) shares some of the ideas of Chandrasekaran (2006). He believed that a perceptual representation in memory has the additional advantage of answering queries about relations that were defined to the agent after the time of experience. Traditional cognitive architectures generate a symbolic representation of relations, but they cannot cope with all possible quantitative situations at once since it is impossible to foreseen what kind of reasoning the agent will need.

If instead we also save in memory ideograms (or in our proposal, perceptograms), we can generate the quantitative symbolic information when necessary. Then, our architecture is a bi-modal cognitive one. It could be extended to multi-modes if we were also considering other input information such as sounds, texture, etc. Ideograms can also be used for describing what is happening with smaller vocabulary than traditional approaches.

For example, imagine a team of rovers on the Martian surface. Some of them are performing exploration tasks in a defined area looking for places that may contain water. We will for sure save in memory the positions of the interesting areas, but it is unlikely that we want to save if, for example, rover1 was closer to rover2 than rover3 when the water was detected and who detected it. That information that may be seemed irrelevant, can be needed after by the operations team or scientists, and since initially we have not generated a symbolic predicate that can represent that information (i.e. Closer-than(rover1 rover2 rover3)) we will not be able to infer (at least not easily) any relational position among rovers.

Figure 8 shows our cognitive architecture proposal centered on the WM. We will omitted the mechanisms to retrieve the information from the LTM and place in the WM the information relevant to the tasks since we find it out of the scope of the paper. We propose to try to connect both ends of the architecture space. That is, to permit the use of specialized representations that can be closely coupled to sensorial inputs and allow concurrent asynchronous updating of these. But it should also support abstraction from these low-level modality-specific representations to higher-level amodal representations, which can be used for higher-level cognitive tasks, such as planning and communication, while still keeping the links to raw sensorial data.

Then, we will keep the higher levels (i.e. planning) with their representations and connect them to the low-level sensor information through ideogramatic representations (perceptive symbols). That is, we will facilitate the translation between the perceptive symbols and their manipulation in reasoning.

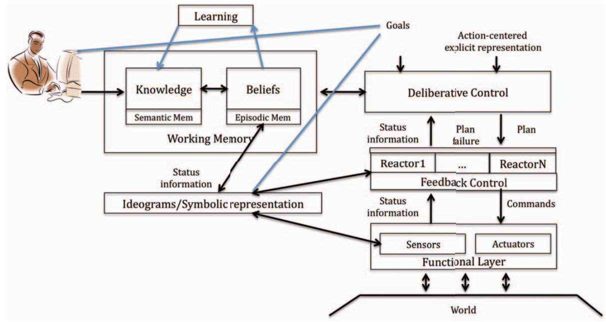

Figure 8: CIbAr architecture

The information processing will be centered around "Knowledge" and "Beliefs". The knowledge will contain all the long term ontological knowledge that the system has either built-in at the design time with the help of the domain experts or subsequently learned using, i.e. Reinforcement Learning techniques. The knowledge is considered as part of the *Semantic Memory* since it contains meanings, understandings, and other concept-based knowledge about the world and the system(s) it wants to control. Instead, the beliefs are part of the *Episodic memory* that will contain the information of events such as times, situation, objects or places as well as other contextual knowledge that can be explicitly stated. These beliefs will be formed using the Knowledge and the interaction/communication with the outside world, the input of which is received through perception capabilities from exploration with humans. The beliefs will be partially represented in an abstracted common representation, i.e. ideograms, which will be understandable by all the elements of the architecture. Then, they will therefore also serve as a main communication medium between them.

At the same time, the current beliefs will represent the current state of the system. It will encode everything the system knows about the particular situation at a given time and in a recent past. Therefore, they will be appropriate for usage by higher-level cognitive components and will be grounded into the low-level sensor data at the same time.

The beliefs will be mainly created by the Perceptual System, that is, the image and space understanding components will both use the information (images) of the different elements available in the scene. These components will therefore provide the symbolic information about the environment (i.e. qualitative and quantitative information about where particular objects are located in space), which will be needed by the planner to produce the plans. Beliefs could also be acquired by the

system through communication with the human. Then, a declarative language, should be provided to facilitate the knowledge acquisition and simple enough and unambiguous to be understandable by the system. That is, a representation that allows the system to recognize the activities even if the human cannot interact with it. The beliefs will also serve as a mean of providing novel information to the system. The new information that will be provided by the human will mainly enter into the system in the form of beliefs (either new ones, or by altering the existing ones), and will be as such later on accessible to all parts of the system.
Finally, our proposal can be integrated in a centralized architecture such as classical 3T-architectures or distributed using multi-agent technology, the final decision will depend on the needs of the final application.

Conclusions

In the last two decades cognitive psychology is been applied to the robotics fields in order to understand the way the humans think, perceive, remember, process the information and solve the problems.
Many cognitive architectures have been proposed based on the research of that field but still a long way to go to fully create machines that think and *feel* like humans.
In this paper we have presented *CIbAr,* a proposal of a cognitive architecture based on: a minimal diagram representation (ideograms), a multi-modal memory and the concepts of knowledge and beliefs.

Acknowledgment

This work was supported by a mobility scholarship funded by the Universidad de Alcalá.

References

A. Ceballos, S. Bensalem, A. Cesta, L. D. Silva, S. Fratini, F. Ingrand, J. Ocón, A. Orlandini, F. Py, K. Rajan, R. Rasconi, and M. V. Winnendael. 2011. "A goal-oriented autonomous controller for space exploration," in Procs. of the 11th ASTRA Symposium, Noordwijk, the Netherlands.

Aschwanden, P., Baskaran, V., Bernardini, S., Fry, C., R-Moreno, M.D., Muscettola, N., Plaunt, C., Rijsman, D., and Tompkins, P. 2006. Model-Unified Planning and Execution for Distributed Autonomous System Control. In AAAI 2006 Fall Symposia. Washington DC (USA).

Bonasso, R., Firby, R., Gat, E., Kortenkamp, D., Miller, D., and Slack, M. 1997. "Experiences with an architecture for intelligent, reactive agents". JETAI, 9(2-3): 237–256.

Brooks, R.A. 1991. Intelligence without representation. AI, (47): 139–159.

Chandrasekaran B. 2006. Multimodal Cognitive Architecture: Making Perception More Central to Intelligent Behavior. In Procs of the AAAI, pp: 1508-1512, Boston, MA, USA.

Claus C. and Boutilier C. 1998. The dynamics of reinforcement learning in cooperative multi-agent systems. In Procs of the 15th International Conference on AI, pp: 746-752, Madison, WI.

Firby, R.J. 1987. "An investigation into reactive planning in complex domains. In procs. of AAAI'87, Seattle, WA, USA.

Gat, E. 1992. Integrating Planning and Reaction in a Heterogeneous Asynchronous Architecture for Controlling Real-World Mobile Robots. In Procs. of AAAI-92, San Jose, CA.

Gerevini, A. and Long, D. 2006. "Preferences and Soft Constraints in PDDL3". In Procs of the ICAPS06 Workshop on Preferences and Soft Constraints in Planning, pp: 46–54.

Hsu, H.C.H. and Liu, A. 2007. A Flexible Architecture for Navigation Control of a Mobile Robot," IEEE Transactions on Systems, Men, and Cybernetics – Part A, 37 (3): 310-318.

Kennedy, W.G., Bugajska, M.D., Harrison, A. M., Trafton., J. G. 2009. "Like-Me" Simulation as an Effective and Cognitively Plausible Basis for Social Robotics. International Journal of Social Robotics, 1: 181–194.

Kim, G. and Chung, W. 2006. Tripodal schematic control architecture for integration of multi-functional indoor service robots. In Procs. of the IEEE Transactions on Industrial Electronics (SCI), 53(5): 1723-1736.

Laird, J. E., Newell, A., and Rosenbloom, P. S. 1987. SOAR: An architecture for general intelligence. AI, 33(3): 1–64.

Lambert, E B. 2007. Diagrammatic representation and event memory in preschoolers. An International Journal of Research and Development, 27 (1): 65-76.

Langley P., Laird J. E. and Rogers, S. 2008. Cognitive architectures: Research issues and challenges. Cognitive Systems Research, 10: 141–160.

Langley, P., and Choi, D. 2006. A unified cognitive architecture for physical agents. In Procs of the AAAI'06, Boston, MA, USA.

Muscettola, N., Nayak, P., Pell, B., and Williams, B.C. 1998. Remote Agent: To Boldly Go Where No AI System Has Gone Before. AI, 103(1-2): 5–48.

Nesnas, I., Simmons, R., Gaines, D., Kunz, C., Diaz-Calderon, A., Estlin, T., Madison, R., Guineau, J., McHenry, M., Shu, I.H., and Apfelbaum, D. 2006. "CLARAty: Challenges and steps toward reusable robotic software". Advanced Robotic Systems, 3(1): 23--30.

Py, F., Rajan, K. and McGann, C. 2010. A Systematic Agent Framework for Situated Autonomous Systems. In Procs. of AAMAS, Toronto, Canada.

Simmons, R. and Apfelbaum, D. 1998. A Task Description Language for Robot Control". In Procs of IROS, Vancouver, Canada.