

Sub-Symbolic Semantic Layer in Cyc for Intuitive Chat-Bots

G. Pilato
ICAR- Italian National Research Council
Viale delle Scienze, Ed.11
90128, Palermo, Italy
g.pilato@icar.cnr.it

A. Augello, G. Vassallo, S. Gaglio
DINFO - University of Palermo
Viale delle Scienze, Ed. 6
90128 Palermo, Italy
{augello,gvassallo,gaglio}@unipa.it

Abstract

The work presented in this paper aims to combine Latent Semantic Analysis methodology, common sense and traditional knowledge representation in order to improve the dialogue capabilities of a conversational agent. In our approach the agent brain is characterized by two areas: a “rational area”, composed by a structured, rule-based knowledge base, and an “associative area”, obtained through a data-driven semantic space. Concepts are mapped in this space and their mutual geometric distance is related to their conceptual similarity. The geometric distance between concepts implicitly defines a sub-symbolic relationship net, which can be seen as a new “sub-symbolic semantic layer” automatically added to the Cyc ontology. Users queries can also be mapped in the same conceptual space, and evoke similar ontology concepts. As a result the agent can exploit this feature, attempting to retrieve ontological concepts that are not easily reachable by means of the traditional ontology reasoning engine.

1. Introduction

Intelligent user interfaces can help people during the interaction with a system in a natural manner, trying to understand and anticipate user needs[11].

One of the most exploited approaches for interacting with users in natural language is the chat-bot technology. Pattern matching, finite-state-machines and frame-based models are commonly used as methodologies for designing chat-bots. These kind of techniques suffice for simple tasks, since they are based on a static process that assigns in advance all possible types to match [18].

Chat-bots have been used in e-learning systems [17][20]. One of them is AutoTutor[20], which tries to

imitate a human tutor interacting with the student in natural language. The Jabberwacky [14] conversational agent is oriented to “simulate natural human chat in an interesting, entertaining and humorous manner”. In [15] a conversational agent has been proposed as a tool for front-end smart web interaction, named AINI. The system is capable also to gather conversation and domain-specific related knowledge. Other commercial conversational agents, like those developed with Lingubote [16] technology, provide proper design environments with the aim of building intelligent chat-bots having complex and goal driven behaviors.

One of the most widespread conversational agent technologies is ALICE [10], whose knowledge base (KB) is composed of question answer modules, called *categories* and described by means of a mark-up language named AIML. The integration of more sophisticated techniques allows improving this simple approach. As an example, in [12] ALICE-based chat-bots have been provided with advanced reasoning capabilities through the linking of the AIML interpreter with the OpenCyc commonsense ontology[8]. The benefits of ontological resources for a spoken dialog system have also been reported in [19].

In [1] the Latent Semantic Analysis (LSA) technique [4] has been used in order to obtain an associative matching between user questions and chat-bot answers, using the pattern-matching mechanism only as a “default” behaviour.

In this paper we propose to enhance the traditional chat-bots with both common sense and associative reasoning capabilities. For this reason we have implemented two different but interconnected areas in the chat-bots “brain”.

The first one is a *rational reasoning area*, based on the integration of two kinds of structured KBs: the standard AIML KB and a CYC common sense KB.

The second one is an *associative reasoning area* obtained building an LSA-inspired semantic space in

which Cyc concepts are coded as vectors and are connected each other by geometric similarity relationships.

Given a specific Cyc microtheory, that is a collection of concepts and facts concerning a particular domain, the semantic space is inferred from a corpus of texts. The corpus is built using both ad hoc extracted pages from the Wikipedia [9] repository, and the comments on the concepts already present in the specific Cyc microtheory. Each concept is projected in this space. The reciprocal geometric distance between concepts implicitly defines a “sub-symbolic” relationship net, that can be seen as a new “sub-symbolic semantic layer” automatically added to the Cyc ontology. This sub-symbolic layer, which has the same psychological basis claimed by LSA [4] can be exploited by the conversational agent during the dialogue with the user through ad hoc AIML tags, created for trigger an associative behaviour of the chat-bot. As a result, the chat-bot can dialogue with the user exploiting its standard KB and the CYC ontology but it can also make use of the associative reasoning area, attempting to retrieve semantic relations between ontological concepts already stored in the KB that are not easily reachable by means of the traditional ontology rules but that are more easily reachable through associative sub-symbolic paths.

Furthermore, the chat-bot can improve its KB adding unknown concepts introduced by the user in the conversation. In fact every time the chat-bot does not have information about a specific topic, it invites the user to give him a definition of the concept. The description of the new concept is then mapped in the semantic space and is added by the chat-bot into the ontology linking the new concept to the most sub-symbolically conceptually related concepts already present in the ontology.

In the remainder of the paper the chat-bot architecture, divided into the aforementioned reasoning areas is presented. Finally some experimental results regarding a selected Cyc microtheory (the AcademicOrganizationMt) and a dialogue example are reported.

2. Chat-bot Architecture

The chat-bot brain is composed of two main areas, as shown in Figure 1.

The first one is a *rational area*, made of the Cyc ontology and the standard AIML KB of the chat-bot. The second one is an *associative area*, made of a semantic space in which Cyc concepts, AIML categories and user queries are mapped.

2.1. Rational Area

The rational area consists of the standard AIML KB and the CYC ontology. The chat-bot can properly query the ontology in order to better answer the user queries by means of ad-hoc defined AIML tags.

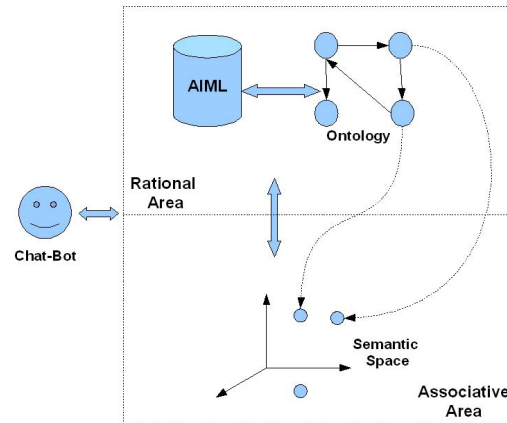


Figure 1: Chat-Bot architecture

2.1.1. AIML KB The chat-bot is based on the ALICE technology [10]. The KB of an ALICE chat-bot is composed of question-answer modules, called categories and described by AIML (Artificial Intelligence Mark-up Language) language. In particular the category is described by the tag *category*. The question is described by the AIML tag *pattern* while the answer by the AIML tag *template*. The dialogue is based on a *pattern matching* algorithm. The user questions are compared by the chat-bot engine with the patterns in its KB. Every time a matching is found, the chat-bot will answer to the user with the template corresponding to the matched pattern.

The presence of special symbols, called wildcards in the pattern allows the chat-bot to answer when in its KB there is not a pattern equal to the user question. Other AIML tags make the dialogue more natural; the tag *srai* allows to recursively call other categories, the tags *set* and *get* allow to set and get the values of variables, the tag *topic* defines a specific topic in the dialogue, the tag *condition* allows to give conditional answers and so on.

The ALICE KB can be constantly increased by the botmaster by means of a “targeting” mechanism. The targeting procedure consists in the analysis of the conversation files in order to detect those user questions, which had an incomplete matching with the AIML pattern. As an example, a user question that matches a pattern with a wildcard is an opportunity to create a new, more specific pattern. As a consequence

the botmaster can write new categories for those questions.

2.1.2. Cyc Ontology The conversational agent can interact with the OpenCyc ontology by means of a module which integrates the AIML interpreter and the OpenCyc inference engine, and which is our Java porting of the CYN project [12]. The template of the chat-bot can contain ad hoc AIML tags which transform it in a meta-answer that must be processed by the OpenCyc inference engine in order to induce and compose the most appropriate response to the user query.

As an example, the tag *cyc-term* allows to translate a natural language term into a Cyc constant. The tag *cyc-system* executes a query into the Cyc KB; if the Cyc query contains a variable, it returns its corresponding value; otherwise it returns the value TRUE or FALSE. The tag *cyc-assert* and *cyc-retract* allow respectively to assert or delete a Cyc formula into or from a microtheory, and so on. The Cyc responses are embedded in a natural language sentence according to the rules of the template. This feature enables the composition of answers that are not present in the traditional AIML KB.

2.2. Associative Area

The associative area is obtained mapping the Cyc concepts as vectors into a semantic space built by means of Latent Semantic Analysis (LSA) methodology. The space is obtained through the statistical analysis of words co-occurrences into a corpus of texts. The texts corpus is built using both ad hoc extracted pages from the Wikipedia [9] repository, and the comments on the concepts already present in a specific CYC microtheory. The chat-bot exploits the associative area, attempting to “guess” semantic relations between ontological concepts evaluating geometric distances computation between the vectors representing the concepts. The chat-bot can also enhance the Cyc KB adding new concepts introduced by the user in the dialogue. The interaction is illustrated in Figure 2.

2.2.1. Building of an LSA-Based Semantic Space

Given N documents of a text corpus let M be the number of unique words occurring in the documents set. Let $\mathbf{A}=\{a_{ij}\}$ be a $M \times N$ matrix whose (i,j) -th entry is the square root of the sample probability of finding the i -th word in the vocabulary in the j -th paragraph (which is a text or a Cyc concept comment). According to the Singular Value Decomposition theorem, \mathbf{A} can be decomposed in the product $\mathbf{A}=\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, where \mathbf{U} is a column-orthonormal $M \times N$ matrix, \mathbf{V} is a column-

orthonormal $N \times N$ matrix and $\mathbf{\Sigma}$ is a $N \times N$ diagonal matrix, whose elements are called singular values of \mathbf{A} .

The matrices \mathbf{U}_R and \mathbf{V}_R obtained after decomposition process reflect a breakdown of the original relationships into linearly independent vectors [2]. These independent R dimensions of the \mathfrak{R}^R space can be tagged in order to interpret this space as a “conceptual” space. Since these vectors are orthogonal, they can be regarded as principal axes representing the “fundamental” concepts residing in the data driven space generated by the LSA.

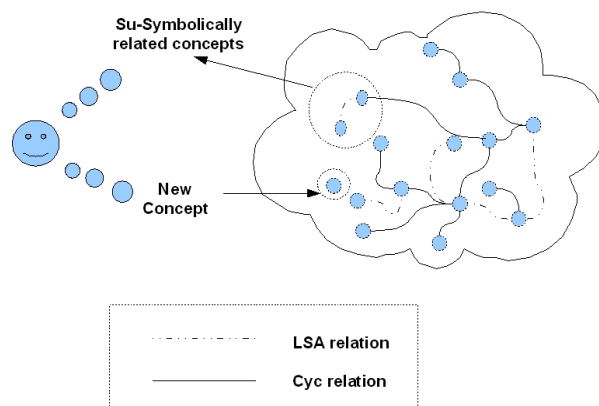


Figure 2. Interaction between the chat-bot and the associative area

To evaluate the distance between two vectors \mathbf{x}_i and \mathbf{x}_j belonging to this space that is coherent with this probabilistic interpretation, a similarity measure is defined as follows:

$$\text{sim}(x_i, x_j) = \begin{cases} \cos^2(x_i, x_j) & \text{if } \cos(x_i, x_j) \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

2.2.2. Evocation of Concepts during the Dialogue

After the creation of the semantic space, each concept of Cyc is encoded as a point in the multi-dimensional semantic space using its Cyc definition and its related documents using the folding-in technique[21]. As a result, each concept is identified by a set of vectors each one related either to the comment already present in the Cyc KB or to a Wikipedia paragraph directly referring to the concept. The geometric similarity measure between two “concepts” as defined in formula 1, establishes a semantic, weighted, sub-symbolic link between them. The net of these semantic connections can be seen as a new *semantic* “layer” superposed to the existing Cyc relations. In particular given a vector \mathbf{u} , associated to the concept c , the set of vectors sub-

symbolically conceptually related to the concept c is given by:

$$CR = \{u_i | sim(u, u_i) \geq T\} \quad (2)$$

where CR is the set of vectors u_i , associated to the concepts c_i whose similarity measure is higher than an experimentally fixed threshold T ($T \in \mathcal{R}; 0 \leq T \leq 1$).

The chat-bot exploits the semantic layer through new specific AIML tags introduced for this interaction. In particular, the *relatedConcept* tag allows the chat-bot to retrieve the concepts conceptually related to a specific ontology concept, while the *sentenceConcept* tag allows the chat-bot to retrieve the concepts related to a sentence introduced by the user.

2.2.3. Ontology Targeting The Ontology Targeting is a mechanism, which allows increasing the Cyc KB during the conversation between the chat-bot and the user. Every time an unknown concept is introduced by the user during the conversation, the chat-bot asks for its definition. The process is analogous to what happens in real life when someone introduces a new term or concept and we ask him further explanation. The verbal definition of the new concept, provided by the user, is mapped as a vector into the semantic space. Its similarity with the concepts of the microtheory already mapped into the space is then computed according to formula 2.

The new concept is added by the chat-bot into the microtheory through the *addConcept* tag, which creates a *conceptuallyRelated* Cyc relation between the new concept and the most related concepts already present in the ontology.

At the end of the dialogue, the ontology/domain expert can furthermore analyze the conversation files. He evaluates the new concepts inserted by the chat-bot and their new relationships, which link them to pre-existing concepts of Cyc. He then decides if the new concepts, together with their associated relations, should be retained or discarded.

3. Sub-Symbolic Mapping of a Cyc Microtheory: Experimental Results

3.1. Analyzed Microtheory

We have selected a small Cyc microtheory in order to validate the proposed technique using its entire set of concepts. The chosen application domain is described by the *AcademicOrganizationMt* microtheory, which describes the American academic structure and the relations among several scholastic institutions by means of a specific set of collections and predicates.

The chosen microtheory represents a good candidate because there are 31 strongly connected elements and, in the worst case, only four links connect each concept with another one belonging to the same microtheory.

3.2. Corpus Building and Creation of a Semantic Space through LSA

The application of the LSA requires a large and meaningful text corpus. The quality of the corpus determines the effectiveness of the semantic space building. For this reason we have searched through the associated *VocabularyMt* all the constants belonging to the microtheory. Results for one generic query can refer to different constants and to predicates that could not belong to the selected microtheory. We have included also these constants which are external, but related to the selected microtheory. Such a choice brought to a higher number of analyzed concepts, which in the particular case of the *AcademicOrganizationMt* has led to a number of 134 analyzed concepts from a starting number of 31 microtheory concepts.

The choice of documents associated to the analyzed concepts is a critical phase of this step. It has been chosen to use the Wikipedia [9] repository, which is one of the most complete semi-structured free documents repositories available today. We have searched documents pertinent to the topic of the microtheory by means of the internal search engine of Wikipedia using the names of the concepts as keywords. A relevance threshold has been experimentally fixed to 50%. Each retrieved document has been filtered in order to retain only informative textual data. Each page has been divided in paragraphs, each one representing a text associated to the concept. Therefore, a variable number of texts is associated to each concept, depending on how much an argument is widespread. The set of documents used to build the semantic space has been expanded using the Cyc comments of each concept in the selected microtheory. A matrix has been constructed analyzing the words-document occurrences in the built corpus. The matrix has been subsequently decomposed into three matrices Σ_R , V_R and U_R according to the TSVD technique with $R=100$, which is also the dimensionality of the generated space.

3.3. An Example of Interaction with the Chat-bot

The following dialogue is a small example of a possible interaction with the chat-bot.

U: *I need information about departments related to the study of plants.*

C: *Do you want information about the BiologyDepartment?*

U: *Can you show me other related Departments?*

C: *I have found a relation with the Agriculture Department.*

U: *Can you give me information about it?*

C: *It is a specialization of AcademicDepartment, instances of which conduct teaching and research in Agriculture.*

U: *Can you give me information about some Private University?*

C: *Sorry, I don't have any information about it. Can you help me in finding a definition of Private University?*

U: *The definition is a university that is run without the control of any government entity.*

C: *Ok, I can add this information In my KB. It is right to relate this concept to the University concept?*

U: *Yes, it is.*

We report below some of the AIML categories, enhanced with ad-hoc defined tags, which allow the chat-bot to sustain the previous dialogue. In particular for each category it is explained the correspondence between the analyzed interaction and the categories belonging to the AIML files.

The following category allows the chat-bot to understand the user information requests about a generic argument matched in the pattern by the wildcard *. Its value, which in the first request of the analyzed interaction is the string “departments related to the study of plants” is recovered in the template through the tag *star* and it is stored in the variable *userSentence*. After this procedure, which is hidden to the user by means of the tag *think*, the tag *srai* starts again the pattern-matching algorithm, searching for the category with pattern “SENTENCE CONCEPT”.

```
<category>
<pattern> I NEED INFORMATION ABOUT *
</pattern>
<template>           <think>           <set
name='userSentence'></star></set></thin
k> <srai> SENTENCE CONCEPT</srai>
</template>
</category>
```

The following category is recursively called by the previous one, by means of the tag *srai*. In its template the Cyc concept most related to the user request is searched through the tag *sentenceConcept*. If a concept is detected, it is stored in the variable *cycConcept*, and

the chat-bot asks to the user if the concept detected is what he was querying for.

```
<category>
<pattern> SENTENCE CONCEPT </pattern>
<template> Do you want information
about           <set name='cycConcept'></
sentenceConcept ></set>?
</template>
</category>
```

If the user asks for other related concepts, the chat-bot searches in the sub-symbolic layer for the concept most similar to the current value stored in the *cycConcept* variable through the tag *relatedConcept* as shown in the following category. The found concept replaces the previous one in the variable *cycConcept*.

```
<category>
<pattern> CAN YOU SHOW ME OTHER RELATED
DEPARTMENTS </pattern>
<template> I have found a relation with
the           <set name='cycConcept'> </
relatedConcept ></set>
</template>
</category>
```

If the user asks for information about the concept currently stored in the *cycConcept* variable, the chat-bot queries the Cyc ontology in order to extract its definition stored through the Cyc predicate *comment*.

```
<category>
<pattern> CAN YOU GIVE ME INFORMATION
ABOUT IT </pattern>
<template>It is <cyssystem>(cyc-query
'(#$comment <cycterm></get name=
'cycConcept'
> </cycterm>
?X))</cyssystem>
</template>
</category>
```

If the user asks for an information related to an unknown concept, the chat-bot asks him for a definition the Cyc ontology in order to extract the its definition stored through the Cyc predicate *comment*.

```
<category>
<pattern> CAN YOU GIVE ME INFORMATION
ABOUT *</pattern>
<template>
<think><set name='cycConcept'>
<cycterm><star/></cycterm>
</set></think>
<condition>
<li name=" cycConcept" value="NULL">
```

Sorry, I don't have any information about it. Can you help me in finding a definition of *star*?
 ... </condition> </template>
 </category>

The user gives the chat-bot the concept definition, and the chat-bot searches for conceptually related concepts to which the new one can be linked.

```
<category>
<pattern> THE DEFINITION IS * </pattern>
<template> <think> <set name='userSentence'><star/></set>
</think>
Ok, I can add this information In my KB. It is right to relate this concept to the <set name='relatedConcept'><sentenceConcept/></set> concept?
</template>
</category>
```

The user gives the confirm and the chat-bot adds the new concept through the Ontology Targeting mechanism by means of the *addConcept* tag, as explained in section 2.2.3.

```
<category>
<pattern>OK YOU CAN ADD IT </pattern>
<template> <think> <addConcept> </get name='cycConcept'> </ addConcept>
</think> The concept has been added
</template>
</category>
```

3.4. Analysis of the Obtained Semantic Layer

In this section some experimental results showing the effectiveness of the relations induced in the semantic space are illustrated. In addition to the constants we have analyzed and stored all the assertions and all the links in the ontology for validation purpose. For the examined microtheory 46 predicates have been found and for each one of them a concept-concept incidence matrix has been created. We have computed the similarity measure given by formula 1 between the vectors related to the microtheory concepts and the vector coding concepts introduced by the user.

In order to validate the quality of the results, some *relevant*, *less relevant* and *not relevant* relations have been estimated. A concept, described by keywords, has been sub-symbolically compared to the corpus of reported documents. The comparison threshold has been fixed to $T=0.5$ (see formula 2).

Table 1 shows some examples of relevant relations obtained for the *MathematicsDepartment* concept, while Table 2 and Table 3 show examples of relevant

relations for two new concepts provided by the user: the *PrivateUniversity* and the *PublicUniversity* concepts.

The relations retrieved for *MathematicsDepartment* concept are appropriate, but some concepts like *HistoryDepartment* and *AnthropologyDepartment* are not found. For *PrivateUniversity* concept pertinent relations with *College*, *University* and *UniversitySystem* are found. It is worthwhile to point out the relation with the *PublicUniversity* concept, which is a new constant inserted by the user. For *PublicUniversity* pertinent relations with the constants *University* and *UniversitySystem* are highlighted.

Table 1. Relevant relations obtained for the MathematicsDepartment concept

	MathematicsDepartment
BiologyDepartment	0.80
PhysicsDepartment	0.73
AgricultureDepartment	0.70
University	0.65

Table 2. Relevant relations obtained for the PublicUniversity concept

	PublicUniversity
University	0.75
PrivateUniversity	0.68
UniversitySystem	0.65

Table 3. Relevant relations obtained for the PrivateUniversity concept

	PrivateUniversity
University	0.72
PublicUniversity	0.68
College	0.61
UniversitySystem	0.59

Less relevant relations have been found for the concept *Campus*. This concept has weak connections with the chosen domain because it belongs to the university world but not to the academic structure. One single link has been found with *UniversitySystem* with a score of 0.79, the reason can be found in the fact that in many documents referring to *UniversitySystem* there are names of various university campuses.

Experiments with the not relevant constants *Bedroom* and *Telephone* have been carried out. These constants have been chosen in order to verify two possible situations: the former has been chosen because it appears rarely in the domain documents; while the latter is very frequent in the retrieved text corpus. For

Bedroom no links with other constants have been found, while for *Telephone* a semantic link *UniversitySystem* with a score of 0.79 has been found, but it is not correct. This can be explained by many documents related to the *UniversitySystem* concept: there are references to the telephone numbers of some university.

3.5. A Comparison of Traditional, Common Sense and Intuitive Chat-Bots

In this section we compare three different systems: the traditional ALICE-based, the CYN-based and our (LSA+CYN)- based systems. We consider a possible information request of the user and compare the AIML categories for the three approaches, highlighting drawbacks and benefits. As an example, the user queries the chat-bot about information concerning "Medical Schools".

An ALICE-based chat-bot developer (botmaster) should plan all the possible user expressions for the specific information request. As an example, the botmaster could write patterns like this one:

```
<pattern> DO YOU KNOW SOME MEDICAL
SCHOOL</pattern>
```

It is clear that the botmaster should predict many others expression, and the same should be done for the other main concepts of the analyzed domain the user could ask for. Besides, it is necessary to write a specific answer of each specific concept of the analyzed domain. A template could be the following:

```
<template> I know some medical school.
There are: the Texas A and M University
College of Medicine, the Tufts
University School Of Medicine...
</template>
```

The use of an ontology representing the domain concepts and its relations can lead to a chat-bot having reasoning capabilities. As an example, the botmaster can write categories with generic patterns such as:

```
<pattern> DO YOU KNOW SOME * </pattern>
```

The corresponding template should be:

```
<template>
<think> <set name='cycConcept'> <star/>
</set> </think>
You want information about
<cycsystem>(cyc-query '(&#x27;isa ?X
<cycterm> <star/> </cycterm>
?X))</cycsystem>.
There are the
```

```
<cycsystem>(cyc-query '(&#x27;isa ?X
<cycterm> <star/> </cycterm>
))</cycsystem>
</template>
```

The wildcard * can be read with the tag <star/> and introduced in a subL query which can be analyzed by the Cyc reasoning engine.

In this manner if the user asks: "Do you Know some Medical School?", the string *Medical School* corresponds to the Cyc concept *MedicalSchool*. The chat-bot answer is dynamically composed through the ontology querying and in the specific case is the following: *You want information about Schools that grant medical degrees, whose students usually end up as medical doctors. There are the Texas A and M University College of Medicine, the Tufts University School Of Medicine...*

The disadvantage is that the user is constrained to express his request with the exact name of the ontology concept, or with one of its related "name-strings" (i.e. natural language expression associated to the concepts through the Cyc predicate *nameString*).

As a consequence, if the user asks, "Do you know some university where I can get a medical degree?" the system is not able to understand the request.

In our system we only need to substitute in the previous template the tag *star* with the tag *sentenceConcept*. In this manner the chat-bot will find the Cyc concept *MedicalSchool* which is most similar to the string "University where I can get a medical degree". The retrieved concept will be introduced in the subL queries and the chat-bot will return again the same answer composed by the definition of the concept and by the concepts related to the analyzed concept by means of the *isa* Cyc predicate.

4. Conclusions

In this paper we have presented a chat-bot with associative capabilities. The chat-bot is provided with two reasoning areas. The first one, makes use of the traditional AIML KB and the Cyc ontology and constitutes the "rational reasoning area". The second one makes use of a semantic space automatically induced from the concepts definitions of Cyc and a set of related Wikipedia documents dealing with a specific topic. When the user asks a question, the query is projected in the semantic space and triggers the "closer" Cyc concepts, which are sub-symbolically semantically related. If the user expresses a new concept or says a new "word", the chat-bot is capable to ask more information about its meaning and tries to map it in the Cyc ontology, linking it with the closer concepts in the semantic space. Subsequently a

knowledge engineer can refine the positioning of the new concept into the ontology.

This kind of “intuitive” answer mechanism allows users a greater freedom of expression than traditional rule-based systems, like ALICE and ALICE+CYC approaches. First experimental results, regarding the "*AcademicOrganizationMt*" Cyc microtheory, are encouraging. Future work will deal on how to enhance the “associative” interaction and updating mechanism of the knowledge base.

5. Acknowledgements

Authors would thank Mr. Mario Scriminaci for his partial contribution to this work.

References

- [1] Agostaro F., Augello A., Pilato G., Vassallo G., Gaglio S.: A Conversational Agent Based on a Conceptual Interpretation of a Data Driven Semantic Space. In Lecture Notes in Artificial Intelligence, Springer-Verlag GmbH, vol. 3673/2005, pp 381-392
- [2] Agostaro F., Pilato G., Vassallo G., Gaglio S.: A Subsymbolic Approach to Word Modelling for Domain Specific Speech Recognition In Proc. of IEEE CAMP05 International Workshop on Computer Architecture for Machine Perception, Terrasini - Palermo, July 4-6 2005, pp. 321-326
- [3] S.T.Dumais T.K.Landauer. A solution to Plato’s problem: the latent semantic analysis theory of acquisition, induction and representation of knowledge. Psychological Review, 1997. 2.
- [4] Landauer, T. K., Foltz, P. W., & Laham, D. (1998). Introduction to Latent Semantic Analysis. *Discourse Processes*, 25, pp.259-284.
- [5] Biemann C.: Ontology learning from text: a survey of methods. In LDV-Forum 2005 – Band 20(2) pp.75-93
- [6] Suguraman V., Storey V.C.: Ontologies for conceptual modeling: their creation, use and management. In *Data & Knowledge Engineering* 42, 2002, pp.251-271
- [7] Chella A., Frixione M., Gaglio S.: An Architecture for Autonomous Agents Exploiting Conceptual Representations. *Robotics and Autonomous Systems* 25, 1998, pp. 231–240
- [8] <http://research.cyc.com>
- [9] <http://www.wikipedia.org>
- [10] Alice: <http://www.alicebot.org>
- [11] Sing, G. O., Wong, K. W., Fung, C. C., Depickere, A. 2006. Towards a more natural and intelligent interface with embodied conversation agent. *ACM Proc. of the 2006 international Conference on Game Research and Development*. vol. 223. pp.177-183.
- [12] Kino Coursey, Daxtron Laboratories, Inc. Living in CYN: mating AIML and CYC together with Program N. 2004.
- [13] Pilato, G., Augello, A., Trecarichi, G., Vassallo, G., Gaglio, S. LSA-Enhanced Ontologies for Information Exploration System on Cultural Heritage. *AI*IA Workshop for Cultural Heritage*, Milano, Italy, 2005.
- [14] Carpenter, R. 1997–2006. www.jabberwacky.com
- [15] Ong Sing Goh, Chun Che Fung, Kok Wai Wong and Arnold Depickere: An Embodied Conversational Agent for Intelligent Web Interaction on Pandemic Crisis Communication, *Proc. of the 2006 IEEE/WIC/ACM international conference on Web Intelligence and Intelligent Agent Technology* Pages: 397-400, 2006
- [16] Creative Virtual. 2004–2006. CreativeVirtual.com. www.creativevirtual.com
- [17] Alice Kerly, Phil Hall, Susan Bull: Bringing chatbots into education: Towards natural language negotiation of open learner models. *Knowledge based systems*, (20) 2007 pp. 177-185
- [18] Kyoung-Min Kim, Jin-Hyuk Hong, Sung-Bae Cho: A semantic Bayesian network approach to retrieving. *Information Processing and Management* 43, 2007, pp. 225–236
- [19] Manish Mehta, Andrea Corradini: Developing a Conversational Agent using Ontologies, 12th International Conference on Human-Computer Interaction, Lul 2007, Beijing, China, pp. 22-27
- [20] Jackson G.T. & Graesser A.C. (2006). Applications of human tutorial dialog in AutoTutor: An intelligent tutoring system. *Revista Signos* 39, pp. 31-48.
- [21] Berry, M., et al. Using Linear Algebra for Intelligent Information Retrieval In: *SIAM Review* 37(4), 1995, pp. 573-595.