

Developing a Conversational Agent using Ontologies

Manish Mehta¹, Andrea Corradini²

1. Cognitive Computing Lab
Georgia Institute of Technology
Atlanta, GA, USA
mehtama1@cc.gatech.edu

2. Computational Linguistics Department
University of Potsdam
Potsdam, Germany
andrea@ling.uni-potsdam.de

Abstract. We report on the benefits achieved by using ontologies in the context of a fully implemented conversational system that allows for a real-time rich communication between primarily 10 to 18 years old human users and a 3D graphical character through spontaneous speech and gesture. In this paper, we focus on the categorization of ontological resources into domain independent and domain specific components in the effort of both augmenting the agent’s conversational capabilities and enhancing system’s reusability across conversational domains. We also present a novel method of exploiting the existing ontological resources along with Google directory categorization for a semi-automatic understanding of user utterance on general purpose topics like e.g. movies and games.

1 Introduction

A growing research community has worked on developing general purpose ontologies [1, 2]. One of the long term goals of the ontology initiatives is to produce common formalisms so that ontologies can be combined together easily for a new domain [3]. The use of ontologies has also become an important step towards the implementation of actual spoken dialog systems [4, 5]. In this respect, this paper reports on the role of ontological resources in a real time system that we developed to simulate an experientially rich interaction between kids and a conversational agent. To create a context, we deployed a 3D-embodied character personifying the fairy tale author Hans Christian Andersen (HCA) and placed him in a graphical representation of his original study. User interaction with the character occurs through spontaneous speech and 2D gesture. The agent’s communication channels are synthesized speech coupled with the display of 3D gesture and emotions. HCA domains of discourse include: HCA’s fairy tales, his life, the user and his physical appearance and a meta-domain to resolve meta-communication conversational turns during interactive sessions.

Through the paper, we specifically focus on the overall benefits achieved by using ontologies for the development of such a conversational system and present our findings on the following topics:

- Use of ontologies as the underlying common knowledge representation formalism and communication language across the natural language understanding (NLU) and dialog modules.
- Ease and fast development of the dialog system through shared ontological resources across different domains of conversation.
- Ease of adding new conversation domains, such as e.g. movies and games through a combination of existing ontological resources with Google’s directory categorization.

Developing reusable resources is a key challenge for any software application. Ontological reusability across characters and domains depends upon whether the knowledge resources are developed in a generic way. The ability to automatically induce a concept in one character and port it to a new character provides a strong benchmark to test the crafting of ontological resources. For historical characters, where the users are going to address questions regarding their life and physical appearance, concepts for these domains provide a clear case of reuse. We have developed ontological concepts for a characters life and physical appearance that could be used for a new character with little modification. We have also developed properties that are shared across domains. These ontological resources also provide a common representational formalism that simplifies communication across the natural language understanding (NLU) component and dialog modules.

Inside the NLU, the rules that we have defined to detect domain independent dialog acts and properties simplify the addition of new domains to increase the range of discussion topics one could have with the animated agent and not just the topics in its domain of expertise. To the five original domains, we have then added domains like movies and games to provide HCA with the ability to address more general purpose everyday topics. In order to properly capture and understand topics within these new domains during conversation, we utilize Google's directory structure that contains, among other things, updated information and classification of movies and games. For example, if the user asks about a certain computer game that was recently released to the public (and for which we do not have any information in our hand crafted system knowledge base), the NLU uses its internal set of rules to try to classify the question into a dialog act, a property using domain independent rules and an unknown concept which is likely to be the name of the game as it was uttered/typed in by the user. This unknown concept is resolved using Google directory engine. The automatic categorization provided by Google coupled with the domain independent properties and the dialog acts results in an automated representation of the user intent consistent with the current NLU ontological representation formalism.

The rest of the paper is organized as follows. In section 2 we present an overview of the system. We present our natural language understanding module in Section 3. Next, we describe our method of combining existing resources with Google directory categorization to provide a representation for utterances corresponding to general purpose topics. In Section 5, we discuss conversational mover component inside the dialog module that detects the next prospective conversational move of the character. We discuss ontological reusability across different domains in Section 6 and finally conclude with some future directions we are planning to undertake.

2 System Primer

Our framework is a computer game where a player can interact with an embodied character in a 3D world, using spoken conversation as well as 2D gesture. The basic idea behind the game scenario is to have the



Fig. 1. HCA in his study

player communicate with a computer generated representation of the fairytale author Hans Christian Andersen to learn about the writer's life, historical period and fairy tales in an entertaining way that reinforces and support the educational message delivered. There is no visible user avatar as the user perceives the world around him in a first-person perspective. She can explore HCA's study and talk to him, in any order, about any topic within HCA's knowledge domains, using spontaneous speech and mixed-initiative dialog. The user can change the camera view, refer to and talk about objects in the study, and also point at or gesture to them. Typical input gestures are markers like, e.g., lines, points, circles, etc. entered at will via a mouse compatible input device or using a touch-sensitive screen. HCA's domains of discourse are: HCA's fairy tales, his life, his physical presence in his study, the user, HCA's role as gate-keeper for access to the fairy tale world, and the meta-domain of solving problems of meta-communication during speech/gesture conversation. Apart from engaging in conversation on his domain of expertise, the user can discuss everyday topics like movies, games, famous personalities and others using a typed interface.

3 Natural Language Understanding module

3.1 General Overview

The Natural Language Understanding module (Fig. 2) consists of four main components: a key phrase spotter, a semantic analyzer, a concept finder, and a domain spotter. Any user utterance from the speech recognizer is forwarded to the NLU where a key phrase spotter detects multi-word expressions from a stored set of words labeled with semantic and syntactic tags. This first stage of processing usually is helpful

to adjust minor errors due to misrecognized utterances by the Speech recognizer. Key phrases that are domain-related are extracted, and a wider acceptance of utterances is achieved. The processed utterance is sent on to the semantic analyzer. Here, dates, age, and numerals in the user utterance are detected while both the syntactic and semantic categories for single words are retrieved from a lexicon. Relying upon these semantic and syntactic categories, grammar rules are then applied to the utterance to help in performing word sense disambiguation and to create a sequence of semantic and syntactic categories. This higher-level representation of the input is then fed into a set of finite state automata, each associated to a predefined semantic equivalent according to data used to train the automata. Anytime a sequence is able to traverse a given automaton, its associated semantic equivalent is the semantic representation corresponding to the input utterance. At the same time, the NLU calculates a representation of the user utterance in terms of dialog acts. At the next stage, the concept finder relates the representation of the user utterance, in terms of semantic categories, to the domain level ontological representation. Once semantic categories are mapped onto domain level concepts and properties, the relevant domain of the user utterance is extracted. The domain helps in providing a categorization of the character's knowledge set. The final output in form of concept(s)/subconcept(s) pairs, property, dialog act and domain is sent to the dialog module.

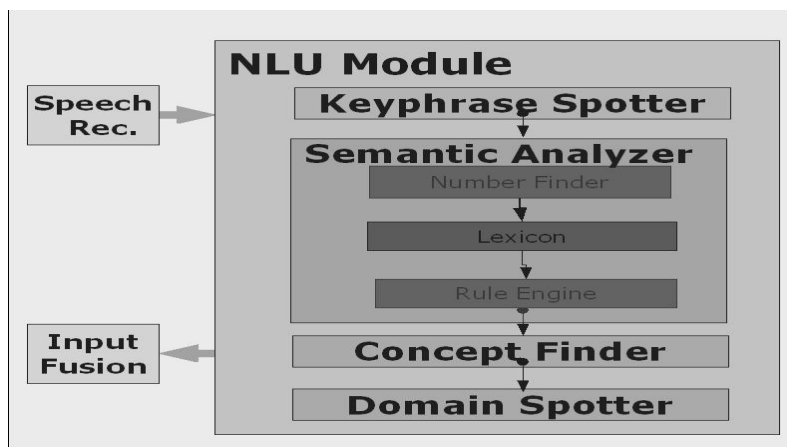


Fig. 2. The main components of the NLU module

3.2 Rule Sharing

Generic rules are defined inside the semantic analyzer for detecting dialog acts (shown in Fig. 3). These dialog acts provide a representation of user intent like types of question asked (e.g., asking about a particular place or a particular reason), opinion statements (like positive, negative or generic comments), greetings (opening, closing) and repairs (clarification, corrections, repeats) [6]. Rules for detecting these dialog acts are defined based on domain independent syntactic knowledge thereby ensuring that once these rules are hand crafted they can be reused across the different domains of conversation. Altogether in our system we have defined about 300 rules out of which approximately a third of them are domain independent. As explained above, a subset of these domain independent rules is used for detecting the dialog acts and is reused across different domains of conversations. The rest of the domain independent rules are used to detect the domain independent properties (e.g., dislike, like, praise, read, write etc). For instance, general wh-questions about a domain are handled by using the domain independent rules to detect dialog acts and properties while simultaneously domain dependent rules are deployed for detecting the concept present in the user utterance.

A typical rule for detecting a dialog act is of the form:

```
<aux:all> <hca>:- <question:yes/no> <hca>
apply_at_position :- [beginning]
Number of conditions :- 0
```

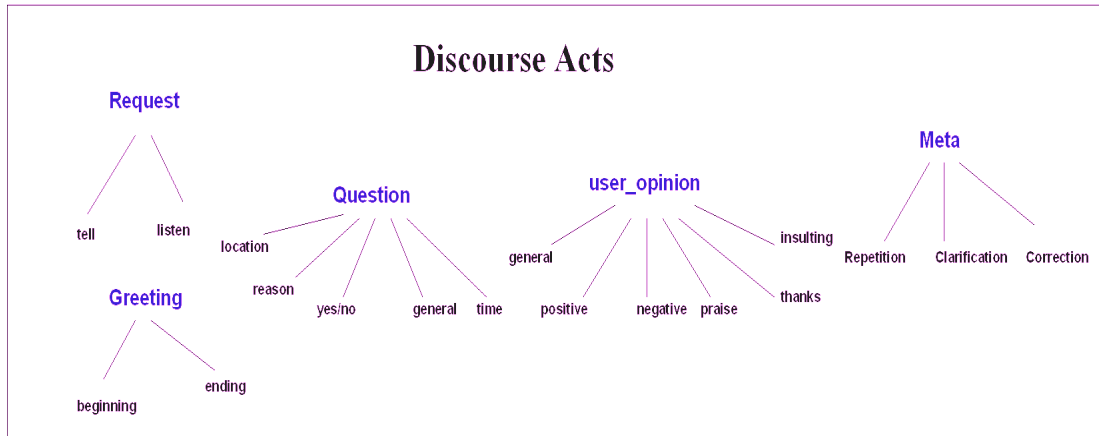


Fig. 3. Examples of different dialog acts that are shared across the domains of conversation

It makes sure that wherever the input sequence “<aux:all> <hca>” (occurring e.g. in sentences starting with “Have you.....”, “Would you.....”) is extracted from the user input at the ‘beginning’ of a sentence, it is then converted into the sequence “<question:yes/no> <hca>”. The rule also states that there is no condition which affects its application. In <aux:all> the field ‘all’ specifies that the rule can be applicable to all the auxiliaries like ‘can’, ‘shall’ etc. This provides a good generalization mechanism so that the rule needs not be created for all the auxiliaries individually. Table 1 shows an example of processing inside the NLU where this rule is applied to detect the dialog act.

One of the rule for detecting the dialog act ‘user opinion’ of type ‘negative’ is

```
<user><aux:all><negative:all> :- <user opinion:negative><subject:user>
apply at position :- [beginning]
Number of conditions :- 0
```

When sentences of the form “I am not...”, “I have not...” are encountered (in this specific case also at the beginning of an input stream), their lexical entries are retrieved from the lexicon and immediately converted into the sequence “<user><aux:am><negative:not>.....”. The above rule is then applied to rewrite this sequence of categories into “<user opinion:negative><subject:user>.....”.

SR output	Do you like your study
Keyphrase Spotter	Do you like your study
Semantic Analyzer	<question:yes/no><verb:like><study:general >
Concept Finder	<dialogact:question><dialogacttype:yes/no> <concept:study><subconcept:general> <property:like>

Table 1. Pipelined processing across different components inside the NLU

3.3 Separation of Ontological and Semantic Representations

In a conversational system, the domain knowledge has to be connected to linguistic system levels of organization such as grammar and lexicon. Domain ontology captures knowledge of one particular domain and serves as a more direct representation of the world. Ontological relationships ‘is-a’ and ‘a-kind-of’ have their lexical counterparts in hyponymy. The part-whole relationships meronymy and holonymy also form hierarchies. This relationship parallelism would suggest that lexical relationships and ontology are the

same but a lexical hierarchy might only serve as a basis for a useful ontology and can at most be called an ersatz ontology [8]. Bateman [9] provides an interesting discussion of the relationship between domain and linguistic ontologies. In our architecture, we use two different sets of representations to support the two contrasting objectives of semantic and domain level representation. The role of the concept finder is to provide a link between the two representations. The semantic categories present in the user utterance in the domain are mapped to the domain level concepts and properties through rules defined on the semantic categories. Table 1 shows an example of this mapping.

Inside the NLU, the concept/property representations along with the dialog acts are combined with Google’s directory categorization for unknown concepts to achieve a semi-automatic categorization for general purpose topics. We explain on this categorization approach in the next section.

4 Understanding General Purpose Topics through Google’s directory

Conversational agent research has focused on effective strategies for developing agents that can co-operate with the human participant to solve a task within a given domain. In the framework of our project, we attempt to move from task oriented agents to life-like virtual agents with model of emotion and personality, a sense of humor and social grace, carrying out a mixed initiative conversation on their life, physical appearance and their domain of expertise. In our research, one of the suggested improvements of our first efforts [11] was to increase the range of discussion topics one could have with the animated agent. This raises an interesting challenge of how to cost-effectively develop agents that have sufficient depth to keep the user involved and are able to handle multiple general purpose topics and not just the topics in their domain of expertise. These topics could range from multiple general-purpose subjects like games, movies, current news, famous personalities, food, and famous places. The task of developing conversational agents with abilities to talk about these open-ended domains would be expensive, requiring significant resources. The simplest way to address these topics would have been to handle them with a standard reply *“I don’t know the answer”*, ignoring them altogether or through shallow pattern matching techniques as illustrated by chat bots like Weizenbaum’s seminal ELIZA [12] and Alice [13].

Web directories represent large databases of hand-selected and human-reviewed sites arranged into a hierarchy of topical categories. Search engines utilize these directories to find high-quality, hand-selected sites to add to their database. Users that are searching for a variety of sites on the same topic also find directories helpful by being able to search in only the category that interests them. Google’s web directory contains, among other things, classification information about names of movies, games, famous personalities, etc. Making entries for these domains manually in the lexicon would be a labor and time intensive effort. Apart from that, these open ended domains evolve over a period of time. As such they need

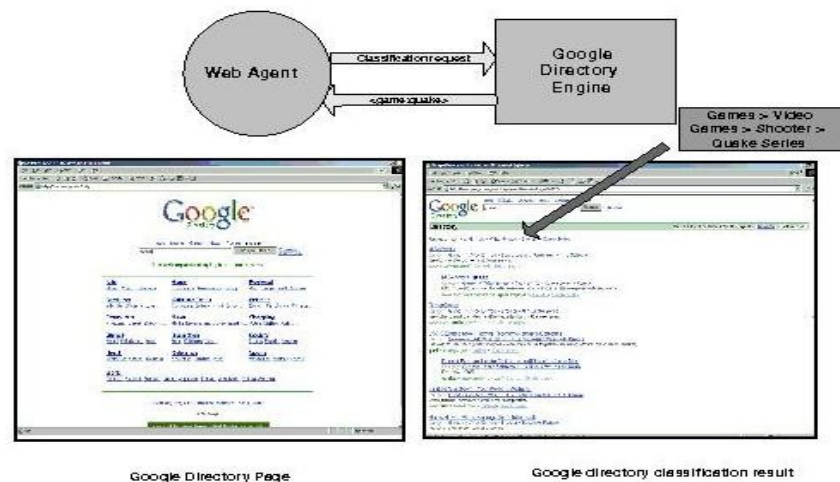


Fig. 4. The unknown category ‘quake’ resolved by Google’s web directory

periodic updates. Thus, using Google’s categorization provides an automatic classification method for terms related to these domains. In our architecture, the NLU categorizes the words without a lexical entry and those that are not detected by the keyphrase spotter, into an *unknown* category. The longest unknown sequence of words is combined into a single phrase. These words are sent to the web agent, which uses Google’s directory structure to find out whether the unknown words refer to a name of a movie, game, or a famous personality and the corresponding category is returned to the NLU. To illustrate the processing let us assume the user asked ”do you like quake?”. In this case, the NLU marks the word quake as an *unknown* category that, as such, needs further resolution. The temporary output of the NLU is thus a *yes/no-question* dialog act, a property of the kind *like* and an *unknown* category. The *unknown* category is resolved by the web agent into the category *game* using Google’s directory engine (see Fig. 4). Using this newly gathered information, the NLU is capable to pass on to the dialog module a complete output which now consists of a *yes/no-question* dialog act, a property of kind *like*, a concept *game* and a sub concept *quake*. The classification provided by Google along with the properties shared across domains and the dialog acts provides a method to build an automated representation consistent with the current output representation generated by the understanding module. Based on this information, the conversational mover inside the dialog module searches for an appropriate conversational move in response to the original sentence as explained in the next session.

5. Conversational Mover

One of the challenges in developing a spoken dialog system for conversational characters is to make system components communicate with each other using a common representation language. This representation language reflects the contradictory ambitions of being rich enough to help encode the personality of the character and general enough so that the formalism doesn't change across characters. At the next stage, inside the dialog module, the output representation from the NLU is used to reason about the next conversational move of the character. This stage of processing is performed inside a module called the conversational mover. For each conversational move of the character, rules are defined using the concept(s)/sub concept(s), property(s)/property type and dialog act/dialog act type pairs delivered by the NLU. This provides a systematic way to connect the user intention to the characters output move. Table 2 shows examples of rules inside the conversational mover. Table 3 provides two examples of processing across different components. Anytime HCA has to produce a response or initiate a new conversational turn on the domain topics, the dialog module selects a contextually appropriate output in accordance with the conversational move produced by this module, the conversational history and the emotional state [7]. The processing after this stage is mainly related to agent’s response generation in terms of behavior display, speech generation and its synthesis but it is outside the scope of the paper.

<pre> define conv move :- movie_opinion { dialog act :- request or question and dialog act type :- listen or general and concept :- movie and sub concept: XX and property :- like or think } define conv move :- famous_personality_knowledge { dialog act :- request or question and dialog act type :- listen or general and concept :- famous_personality and sub_concept: XX property :- know } </pre>
--

Table 2. Example of two rules inside the conversational mover. XX acts as a placeholder for any *sub_concept* type and allows us to reuse the rule for all the sub concepts

6. Ontological Reuse

One of the clear cases of ontological reuse for a historical character is to craft the concepts of his life and physical appearance in a generic way. Figure 5 shows the domain (in)dependent concepts which are currently used in our architecture for HCA life and physical self. The figure also shows general properties which are shared across different domains. To illustrate the reusability of our approach, let us consider some explanatory use cases.

Use case 1: This case represents an utterance which is used by the user to ask about the characters father. The representation from the NLU is independent of a particular character. Similar utterances regarding other family members produce an independent representation with the *sub_concept* slot filled with the corresponding value.

Input: *I want to know a little bit about your father*
 NLU: <dialog_act:question>, <dialog_act_type:general>,
 <concept:family>, <sub_concept:father>

Use case 2: This case represents the use of common properties across different domains. The property *emotion* with property type *scary* is used for the representation across both the utterances belonging to different domain: the first belonging to HCA's fairytales and second one to his physical self.

Input: *Your fairytales are scary*
 Domain: fairytale
 NLU : <dialog_act:user_opinion>, <dialog_act_type:general>
 <concept:fairytale>,<sub_concept:general>,
 <property:emotion>,<property_type:scary>

Input: *You look scary*
 Domain: physical self
 NLU <dialog_act:user_opinion>, <dialog_act_type:general >
 <concept:personal_attribute> <sub_concept:self_identity>
 <property:emotion>, <property_type:scary>

We contend that these reusable portions for the characters life and his physical appearance can save a great deal of development time for a new character.

User:	What do you think about Agatha Christie
NLU:	<dialogact:question><dialogacttype:general> <concept:unknown> <subconcept:agatha_christie> <property:think>
Google Class. :	<dialogact:question><dialogacttype:general> <concept:famous_personality> <subconcept:agatha_christie> > <property:think>
C Mover :	famous_personality_opinion
User:	Do you know about Quake
NLU:	<dialogact:question><dialogacttype:yes/no> <concept:unknown> <subconcept:quake >
Google Class. :	<dialogact:question><dialogacttype:yes/no> <concept:game> <subconcept:quake> <property:know>
C Mover :	famous_personality_opinion

Table 3. Processing of two utterances inside different components

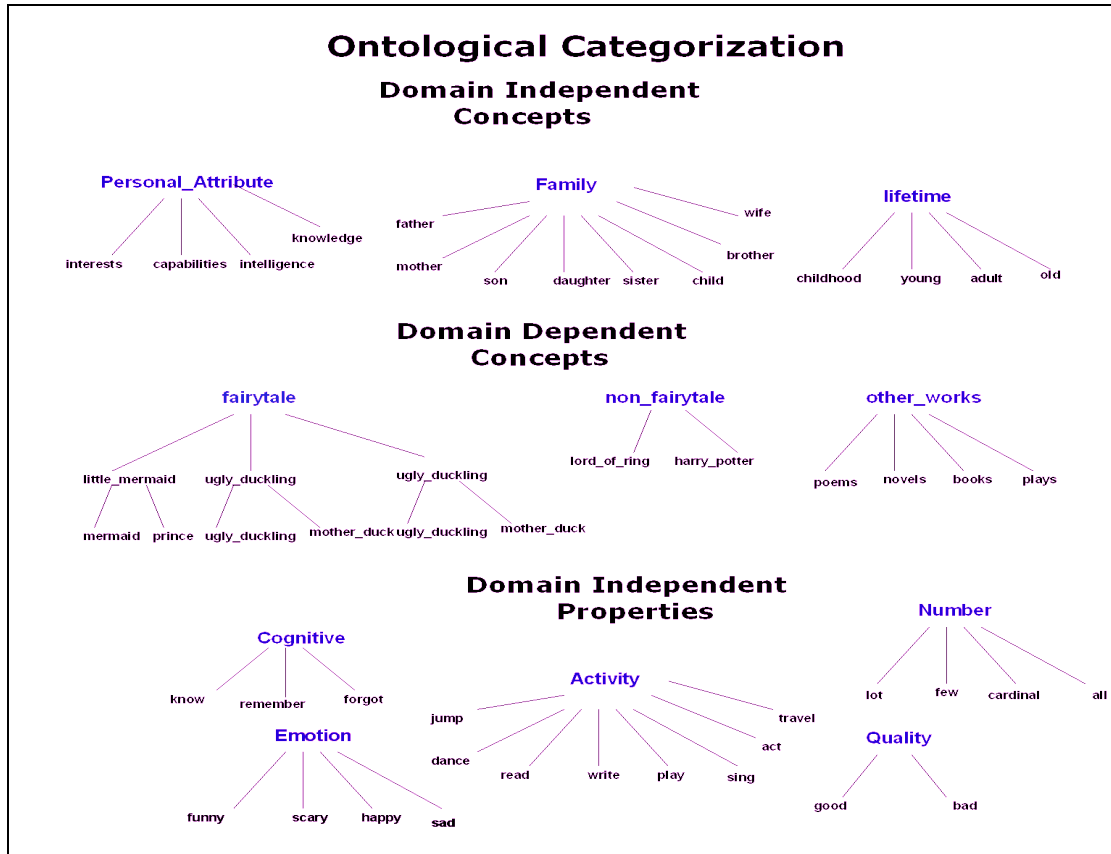


Fig. 5. A set of domain dependent and domain independent concepts and properties

7. Conclusion

In this paper, we discussed the benefits of ontological resources for a spoken dialog system. We reported on the domain independent ontological concepts and properties. These ontological resources have also served as a basis for a common communication language across understanding and dialog modules. We intend to explore what further advantages can be obtained by an ontological based representation and test the reusability of our representation of a characters life and physical appearance through development of a different historical character. For language understanding purposes on topics like movies, games and famous personalities, we have proposed an approach of using web directories along with existing domain independent properties and dialog acts to build a consistent representation with other domain input. This approach helps in providing a semi-automatic understanding of user input for open ended domains.

There have been approaches using Yahoo categories [10] to classify documents using an N-gram classifier but we are not aware of any approaches utilizing directory categorization for language understanding. Our classification approach faces problem when the group of words overlap with the words in the lexicon. For example, when the user says "Do you like Lord of the rings?" where the words 'of' and 'the' have a lexical entry, their category is retrieved from the lexicon and the only unknown words remaining are 'Lord' and 'rings' and the web agent is not able to find the correct category for these individual words. One solution would be to automatically detect the entries, which overlap with the words in the lexicon by parsing the Google directory structure offline and having these entries made in the keyphrase spotter. We plan to solve these issues in the future.

8. Acknowledgments

We gratefully acknowledge the Human Language Technologies Programme of the European Union, contract # IST-2001-35293 that supported both authors in the initial stage of the work presented in this paper. We also thank Abhishek Kaushik at Oracle India Ltd. for programming support.

References

1. Lenat B. D.: 1995, Cyc: A large-scale investment in knowledge infrastructure, *Communication of the ACM*, vol. 38, no. 11, pp. 33–38
2. Philipot A., Hovy E. H., and Pantel P.: 2005, The omega ontology. In: *Proceedings of the ONTOLEX Workshop at the International Conference on Natural Language Processing*, pp. 59–66
3. Kalfoglou, Y., and Schorlemmer, M.: 2003, Ontology mapping: the state of the art, *Knowledge Engineering Review*, vol.18, no.1, pp. 1–31
4. Tsovaltzi, D., and Fiedler, A.: 2003, Enhancement and use of a mathematical ontology in a tutorial dialog system. In: *Proceedings of the IJCAI Workshop on Knowledge Representation and Automated Reasoning for E-Learning Systems, Acapulco (Mexico)*, pp. 23–35
5. Dzikovska M. O., Allen J. F., and Swift D. M.: 2003, Integrating linguistic and domain knowledge for spoken dialogue systems in multiple domains. In: *Proceedings of the IJCAI Workshop on Knowledge and Reasoning in Practical Dialogue Systems, Acapulco (Mexico)*, pp. 25–35.
6. Mehta, M., and Corradini, A.: 2006, Understanding Spoken Language of Children Interacting with an Embodied Conversational Character. In: *Proceedings of the ECAI Workshop on Language-Enabled Educational Technology and Development and Evaluation of Robust Spoken dialog Systems*, pp. 51-58
7. Corradini, A., Mehta, M., Bernsen, N.O., and Charfuelan, M.: 2005, Animating an Interactive Conversational Character for an Educational Game System. In: *Proceedings of the ACM International Conference on Intelligent User Interfaces, San Diego (CA, USA)*, pp. 183-190
8. Hirst, G.: 2004, *Handbook on Ontologies*, Chapter Ontology and the Lexicon, Springer Verlag, pp. 209–230
9. Bateman, J. A.: 1991, The Theoretical Status of Ontologies in Natural Language Processing.. In: *Proceedings of Workshop on Text Representation and Domain Modelling - Ideas from Linguistics and AI*, pp. 50–99
10. Labrou, Y., and Finin, T.: 1999, Yahoo! as an ontology: using Yahoo! Categories to describe documents. In: *Proceedings of the eighth international conference on Information and knowledge management*, pp. 180–187
11. Bernsen, N. O.. and Dybkjær, L.: 2004, Evaluation of spoken multimodal Conversation. In: *Proceedings of the 6th International Conference on Multimodal Interfaces*, pp. 38–45
12. Weizenbaum, J.: 1966, ELIZA: computer program for the study of natural language communication between man and machine, *Communications of the ACM*, Vol. 9, pp. 36–45
13. Wallace, R.: 2002, *The Anatomy of A.L.I.C.E.*