An Instance-based Approach for Identifying Candidate Ontology Relations within a Multi-Agent System

Andrew Brent Williams[1] and **Costas Tsatsoulis**[2]

Abstract. Discovering related concepts in a multi-agent system among agents with diverse ontologies is difficult using existing knowledge representation languages and approaches. We describe an approach for identifying candidate relations between expressive, diverse ontologies using concept cluster integration. We evaluate the feasibility of this approach using lightweight ontologies. These lightweight ontologies are constructed from the Magellan search Web site and consist of Web page categories, or concepts, and their corresponding instances.

1 INTRODUCTION

In order to facilitate knowledge sharing between a group of interacting information agents (i.e. a multi-agent system), a common ontology should be shared. However, we recognize that often agents do not always commit *a priori* to a common, pre-defined global ontology. Our ongoing research investigates approaches for agents with diverse ontologies to share knowledge by automated learning methods and agent communication strategies [15]. When these agents have diverse ontologies there are many challenges for knowledge sharing and communication. One of these challenges is for agents to automatically learn representations for diverse ontologies from categorized Web pages and identify the relationships between two agents' ontologies. In this paper, we demonstrate the feasibility for identifying a candidate 1:N relation between two different agents' ontologies. These ontologies represent natural human categorizations of Web page bookmarks into concepts, or sets of corresponding instances.

Various definitions of an ontology range from simply what is known to exist in an agent's world; the categories in a search engine index; or to more rigorous definitions which lend themselves to constructions of formal ontologies using a description logic. For example, in the classic AI blocks world domain, the ontology only consists of a table surface and three blocks labeled A, B, and C. The Yahoo! search engine has an ontology which consists of a taxonomy of its Web page categories and is referred to as a lightweight ontology [13]. An example of a more extensive formalized ontology is the Cyc Ontology [7].

We recognize that research using formal knowledge representation languages to create formal ontologies for agent knowledge sharing has made significant strides. These approaches, however, must place some limits on the expressiveness of the vocabulary in order to facilitate the use of inference mechanisms for deducing the entailments of a set of sentences in the description logic. MacGregor [8], who worked on the LOOM knowledge representation language, stated the trade off between this type of language and its expressiveness. Also, the essential ontological promiscuity of artificial intelligence states that any agent can create and accommodate an ontology based on its usefulness for the task at hand [4]. Therefore, a group of agents with individualized ontologies may wish to share knowledge but find it difficult to understand the relationships between their concepts.

This situation of agents with diverse ontologies may exist in the World Wide Web domain. Web users may construct simple ontologies while searching and "surfing" the Web. These users search for information using Web browsers and search engines. Once a person finds a Web page of interest, often they will bookmark them for later reference. These bookmarks can be grouped into categories, or concepts, of similar Web pages to form a taxonomy, or a lightweight ontology.

We believe that information agents may benefit from using these ontologies to search for related concepts in a multi-agent system. These agents may understand different concepts that are not exactly the same but may be related. As an example, there may be an agent that understands the concept, "NBA". Another agent may know the concept, "College Hoops". Although these concepts are not exactly the same, they are both clearly related to the concept, "Basketball". Agents that do not know the relationships of their concepts to each other need to be able to teach each other these relationships. If the agents are able to discover these concept relations, this will aid them as a group in sharing knowledge even though they have diverse ontologies. Information agents acting on behalf of a diverse group of users need a way of discovering relationships between the individualized ontologies of users. These agents can use these discovered relationships to help their users find information related to their topic, or concept, of interest. This paper describes a possible approach for discovering these relationships while allowing for maximum expressiveness in the agents' vocabularies.

The rest of this paper first discusses related work in Section 2, describes our approach in Section 3 and then describes our implementation in Section 4. Section 5 describes how we evaluated our system and presents the results. Section 6 presents our conclusions and describes future work.

2 RELATED WORK

Manually constructing ontologies by combining different ontologies on the same or similar subject into one is called merging [11]. Differentiated ontologies having terms that are formally defined as concepts and have local concepts that are shared have been addressed [14]. They use description compatibility measures based on com-

[1] Department of Electrical and Computer Engineering, University of Iowa, Iowa City, IA 52242 email: abwill@eng.uiowa.edu

[2] Information and Telecommunication Center, Department of Electrical Engineering and Computer Science, University of Kansas, Lawrence, KS 66045 email: tsatsoul@ittc.ukans.edu

paring ontology structures represented as graphs and by identifying similarities as mappings between elements of the graphs. The relations they find between concepts are based on the assumption that local concepts inherit from concepts that are shared. Their system was evaluated by generating description logic ontologies in artificial worlds. In our approach, we do not assume that the ontologies share commonly labeled concepts but rather a distributed collective memory of objects that can be selectively categorized into the agent's ontology. Our system also differs in that it uses Web page text as instances that describe examples of the agent's concepts.

Machine learning algorithms have been used to learn how to extract information from other Web pages [3]. Their approach uses manually constructed ontologies with their classes, relations and training data. The objective of this work is to construct a knowledge base from the World Wide Web and not to find relationships between concepts in a multi-agent system.

Several information agent systems attempt to deal with some issues in using ontologies to find information. IICA, or Intelligent Information Collector and Analyzer, is an ontology-based Internet navigation system [5]. IICA gathers, classifies and reorganizes information from the Internet. It uses a common ontology to allow IICA to make inexact matches between users' requests and the candidate locations. They define their ontologies as weakly structured and are built from existing thesauruses and technical books consisting of about 4,500 terms. This system is based on using a common ontology rather than diverse ontologies. For text categorization or classification it uses the information retrieval vector space model. Information agents that can update models of available information sources using inductive, concept learning but applied it to static, relational databases using a formal description logic have been demonstrated [6, 1]. Their system embedded the concept semantics in the initial ontology and in their query reformulation operators. Since they are using a description logic their expressiveness of their vocabulary is limited and would be hampered by the high degree of language expressiveness in the World Wide Web domain. The InfoSleuth Project [2] uses multiple representations of ontologies to help in semantic brokering. Their agents advertise their capabilities in terms of more than one ontology in order to increase the chances of finding a semantic match of concepts in the distributed information system. The InfoSleuth system, however, is not attempting to discover relationships between concepts in the different ontologies.

3 APPROACH

We discuss how our agents represent, learn, share, and interpret concepts using ontologies constructed from Web page bookmark hierarchies. In particular, we show how we use DOGGIE agents to discovery candidate relations between different ontologies. The relations are assumed to be general is-a relations.

3.1 Concept Representation and Learning

A semantic concept comprises a group of semantic objects, i.e. Web pages, that describe that concept. The semantic object representations we use define each token, i.e. word and HTML tag from the Web page, as a boolean feature. The entire collection of Web pages, or semantic objects, that were categorized by a user's bookmark hierarchy is tokenized to find a vocabulary of unique tokens. This vocabulary is used to represent a Web page by a vector of ones and zeroes corresponding to the presence or absence of a token in a Web page. This combination of a unique vocabulary and a vector of corresponding ones and zeroes makes up a concept vector. The concept vector represents a specific Web page and the actual semantic concept is represented by a group of concept vectors judged to be similar by the user.

Our agents use supervised inductive learning to learn their individual ontologies. The output of this ontology learning is semantic concept descriptions (SCD) in the form of interpretation rules. For example, the following is the SCD for the concept in the ontology location Arts/Book/Talk/Reviews using a CLIPS rule representation:

1. (defrule Rule_35 (danny 1)
   =>
   (assert (CONCEPT Arts_Book_Talk_Reviews)))
2. (defrule Rule_34 (ink 1)
   =>
   (assert (CONCEPT Arts_Book_Talk_Reviews)))

Each Web page bookmark folder label represents a semantic concept name. A Web page bookmark folder can contain bookmarks, or URL's, pointing to a semantic concept object, or Web page. A bookmark folder can also contain additional folders. Each set of bookmarks in a folder is used as training instances for the semantic concept learner. The semantic concept learner learns a set of interpretation rules for all of the agent's known semantic concept objects (Figure 1).

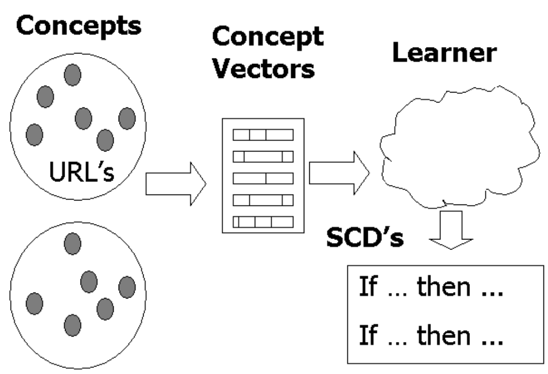

Figure 1. Supervised inductive learning produces ontology rules

For each of these semantic concept description rules, an associated certainty value is determined during the learning process. This certainty value is used later during the interpretation process. It is equal to the percentage these rules were successful in interpreting the training set minus an error prediction rate calculated for our particular semantic concept learner [12].

3.2 Concept-based Queries

DOGGIE agents use concept-based queries (CBQ) to communicate their requests for concepts related to the query. A CBQ occurs when one agent sends example concepts to other neighboring agents, determines by the agents' responses who knows related concepts, and learns new knowledge. This new knowledge can be in the form of learning new semantic concepts or knowledge regarding another agent's ontology. The actual CBQ consists of the concept name, addresses of examples of the concept (i.e. URL's), and flags indicating what type of service the user requests. For this concept-based query scenario, an acquaintance agent is any other agent that the querying agent knows how to locate and communicate with.

3.3 Concept Interpretation and Verification

The querying (Q) agent will send out a CBQ to its acquaintances. The responding (R) agents will use their semantic concept interpreters to determine if they think they know related concepts, and will send their responses back to the Q agent. A semantic concept interpreter is a knowledge-based component that can classify concept instances according to an agent's local ontology. Each Q and R agent have their own local ontologies which represent how they have individually conceptualized their view of the world. The R agent's responses may be a positive (K), neutral (M), or negative (D) interpretation along with the concept name and type. A positive or negative response corresponds to an interpretation value above the positive or negative interpretation threshold, respectively. A neutral response corresponds to the value falling between these two thresholds. A positive interpretation threshold is equal to the percentage accuracy value calculated for a particular concept during the ontology process minus an error prediction rate value for the particular concept interpreter. A negative interpretation threshold is the lower boundary interpretation value that indicates the concept is not known. The concept name corresponds to the bookmark folder the Web pages belong to. The concept type indicates whether the answer to the query is a similar or related concept. If an R agent has a positive response to the CBQ, it will request permission to send examples of its similar or related concept back to the Q agent. The Q agent can then verify whether the R agent actually knows a similar or related concept by using its own concept interpreter on the examples R sends to it.

3.4 Concept Cluster Integration

In this case of multiple M regions in a concept query response, DOGGIE can apply the concept cluster algorithm (CCI) to look for candidate relations between ontologies. As we have previously described, the Q agent sends out a CBQ that will be received by R agents. The R agents will send back the results of its interpretation process. Included in this response are the name of the concept(s) it has interpreted the original concept to be, its interpretation region (K,M, or D), the stored interpretation threshold, the resulting interpretation value, and some examples of the R agent's concept. Since in our multi-agent system, the agents are willing to perform minimal work for each other, the actual concept cluster integration algorithm is performed by the original Q agent instead of an R agent. After the interpretation results have been sent back to the Q agent from the R agent, the Q agent must do several things to complete concept cluster integration. First, it must gather all of the returned examples from each of the returned M region concepts. Then it must combine these into a new directory named after a combination of these concept names. For example, if the M region concepts returned were Sports and NBA, then the new concept cluster directory would be called Sports + NBA. A new ontology category would be created with this label and the returned examples would be combined into this category as the instances that make up this concept. Then the agent must relearn the ontology rules, or semantic concept descriptions (SCD), using its semantic concept learner. Next, the original CBQ sent out by the Q agent will be interpreted according to these new semantic concept descriptions to see if it knows the CBQ concept as the combination of the returned M region concepts. This CCI process integrates the R agents M region concepts into its own ontology since the examples are input into the ontology under a new concept name. This new concept name represents a compound concept.

If there is a match between the original CBQ and the new compound cluster, then new group knowledge which describes a relationship between a Q agent's concept with more than one R agent's concepts can be learned. This group knowledge, or CCI rule, is stored in the form of a concept relation, or compound cluster translation rule.

The Q agent takes the following outlined steps to perform CCI:

1. From the R agent response, determine the names of the concepts to cluster.
2. Create a new compound concept using the above names.
3. Create a new ontology category by combining instances associated with the compound concept.
4. Re-learn the ontology rules.
5. Re-interpret the CBQ using the new ontology rules including the new concept cluster description rules.
6. If the concept is verified, store the new concept relation rule.

4 IMPLEMENTATION

In this section, we describe how we implemented our multi-agent system, the Distributed Ontology Gathering Group Integration Environment (DOGGIE). DOGGIE was used for our investigation into knowledge sharing and learning among agents with diverse ontologies. DOGGIE agents are multithreaded Java applications with a Swing GUI (Figure 2).

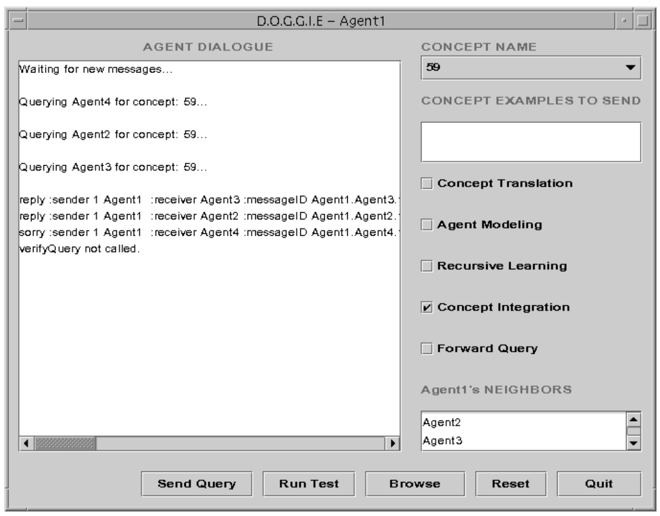

Figure 2. Example DOGGIE Agent GUI with KQML Messages

The underlying multi-agent communication architecture for the DOGGIE system is designed around the Common Object Request Broker Architecture (CORBA). CORBA is used as the underlying communication mechanism between the DOGGIE agents that can be located anywhere on the Internet. The messages between agents are formatted and sent using the Knowledge Query Manipulation Language (KQML). Each agent in DOGGIE is actually composed of both a CORBA client and a CORBA server process running simultaneously so that it can both send and receive queries. A single agent sends its concept-based queries (CBQ) using the CORBA client. The agent receives concept-based queries through its CORBA server component. The CORBA server and CORBA client are the main communication components for the Agent Engine and Agent Control. Each single DOGGIE agent is made up of five major components: Agent Control, Agent Engine, Agent UI, Semantic Concept Learner, Semantic Concept Interpreter (Figure 3).

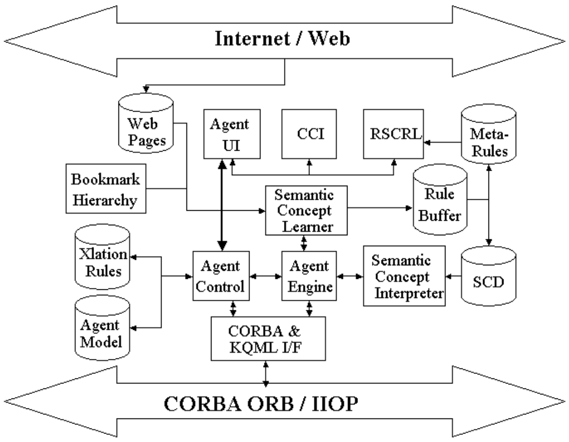

Figure 3. Architecture of a single DOGGIE Agent

5 EVALUATION

We show that it may be feasible for agents to discover relationships between diverse ontologies by testing the concept cluster integration algorithm using ontologies constructed from the Magellan [10, 9] search engine ontology.

5.1 Data Set

A manually constructed subject ontology from the Magellan site was used [16]. This ontology was grabbed from the Magellan site by a spider. The spider started from the Magellan homepage and recursively followed the links to grab both topics and Web pages listed in each Web page topic. This approach assumed that the Web pages listed under a topic were semantically related to the topic. We used an existing open Magellan ontology to objectively measure which Web page instances belonged to particular ontology concepts. The data used consisted of 50 random concept categories taken from the Magellan search Web site. The Magellan ontology consisted of 4,385 nodes, or concept categories. Each of the concept categories used had 20 Web pages in them. Each DOGGIE agent was assigned 5 to 12 concepts from the Magellan ontology. The concept cluster integration experiments were run in single agent to single agent configurations.

The concepts used were each assigned a unique identification (ID) number and some examples of the concepts used are listed in Table 1 below. These were the concepts that we chose from to build our individual agent ontologies. For some of our agent ontologies we randomly chose the concepts used. In others we hand selected the concepts to build "narrow" ontologies that only included closely related concepts.

Table 1. Some example Magellan concepts used for ontology

Number	ID	$Concept$
1	3	Arts/Architecture_Firms
15	170	Business/Companies/Agriculture_and_Fisheries
17	1012	Computing/Hardware/LAN_Hardware
31	2090	Health_and_Medicine/Mental_Health/Resources
33	2120	Hobbies/Arts_and_Crafts/Knitting_and_Stitching
37	3504	Regional/Travel/Travel_Agencies/P_through_Z
39	3535	Science/Astrononmy_and_Space/Resources
46	4030	Shopping/Prized_Possessions/Collectibles
49	4115	Sports/Basketball/NCAA
50	4127	Sports/College/School_Home_Pages

5.2 Experiments

In this section, we describe how we evaluated the feasibility of finding candidate ontology relations using the concept cluster integration algorithm in the context of a multi-agent system.

5.2.1 Hypothesis

It is feasible for agents with diverse ontologies to discover concept relations using concept cluster integration.

5.2.2 Method

We selected a sample of ten queries which produced two M region responses and set up the DOGGIE agents to communicate between the respective Q and R agents. We selected the concept cluster integration option on the DOGGIE agent user interface then sent and processed the queries one at a time.

5.2.3 Prediction

We expected that the CCI algorithm would produce at least some verified concept cluster relations.

5.2.4 Results

The results of our experiments are located in Table 2 below. Only 20% of our queries produced verified concept cluster relations.

Table 2 shows the experiment number, the original queried concept, and the concept responses. It also shows the region type for the responses, the delta, or the difference between the stored interpretation threshold and the actual threshold, the name of the newly created concept cluster, and the results of the concept cluster verification.

5.2.5 Discussion

Our DOGGIE agents discovered two concept relations out of the ten attempts. The resulting concept relation rules are below:

- K(A1, C2090, K(A4, C5+42))
- K(A1, C3504, K(A4, C2024+3504))

Table 2. Concept Cluster Integration Experiments Summary

#	$Query$	$Reply$	□	△	$Cluster$	V
1a	2090	5+42	K	0.04	5+42	Y
1b	2090	2090	M	-0.06	5+42	N
2	3504	2024+3504	K	0.07	2024+3504	Y
3	3562	5+3561	M	-0.26	5+3561	N
4	4030	4030	K	0.24	4004+1014	N
5a	135	3505+1014	M	-0.36	3505+1014	N
5b	135	135	M	-0.06	3505+1014	N
6	59	59	M	-0.08	3504+1014	N
7	170	170	M	-0.50	57+2409	N
8	4002	57+3505	D	-0.46	57+3505	N
9	4002	4002	K	0.15	57+3505	N
10	3561	3561	-	-	57+1027	N

□ region type

The first equation can be read as "Agent 1's concept 2090 is related to Agent 4's concept cluster 5+42". From Table 1 we note that concept 2090 is located in the Magellan ontologies as: Health and Medicine / Mental Health Resources. Concepts 5 and 42 are: Arts /Architecture / Resources and Professional Organizations and Arts / Books / Genres / Non-Fiction. Intuitively, it would be difficult to determine such a relationship between these concepts. However, if we are using the DOGGIE for AI-assisted Web-browsing, this is a relationship that the user may wish to explore.

Similarly, the second equation can be read as "Agent 1's concept 3504 is related to Agent 4's concept cluster 2024+3504". This concept relation rule shows an interesting situation in which one agent's concept 3504 is related to another agent's concept 3504 combined with concept 2024. From Table 2 we see that concept 3504 is Regional/Travel/Travel Agencies/P through Z. Concept 2024 is Health and Medicine/Medicine/Clinics/University Medical Centers. Again, this newly created concept cluster could be worth exploring by the user.

6 CONCLUSION AND FUTURE WORK

Our results have demonstrated that our instance-based approach for discovering candidate relations between ontologies using concept cluster integration is feasible. We believe that further research is required. Our approach does not attempt to identify the specific type of relationship (e.g. part-of) in the ontologies but assumes they consist of general is-a relations. For future experiments, the number of instances included in each concept should be increased to insure that the machine learning algorithm has sufficient training examples. Also, a different experiment design should be used to verify that we can take an existing concept, divide it into two concepts, and determine whether DOGGIE can discover the relations between them. We hope that eventually this multi-agent system approach to finding relations can be used in conjunction with formal ontologies constructed using a description logic.

ACKNOWLEDGEMENTS

We would like to thank the referees for their comments which helped improve this paper.

REFERENCES

[1] J. Ambite and C. Knoblock, 'Reconciling distributed information systems', *AAAI 1995 Spring Symposium on Information Gathering from Distributed, Heterogeneous Environmentss*, (1995).

[2] R. Bayardo, W. Bohrer, R. Brice, A. Cichocki, J. Fowler, A. Helal, V. Kashyap, T. Ksiezyk, G. Martin, M. Nodine, M Rashid, M. Rusinkiewicz, R. Shea, C. Unnikrishnan, A. Unruh, and D. Woelk, *InfoSleuth: Agent-Based Semantic Integration of Information in Open and Dynamic Environments*, 205–216, Morgan Kaufmann, San Francisco, 1998.

[3] M. Craven, D. DiPasquo, D. Freitag, A. McCallum, T. Mitchell, K. Nigam, and S. Slatter, 'Learning to extract symbolic knowledge from the world wide web', in *Proceedings of the 15th National Conference on Artificial Intelligence (AAAI-98)*, (1998).

[4] M. Genesereth and N. Nilsson, *Logical Foundations of Artificial Intelligence*, Morgan Kauffmann, Palo Alto, CA, 1987.

[5] M. Iwasume, K. Shirakami, H. Takeda, and T. Nishida, 'Iica: An ontology-based internet navigation system', *AAAI-96 Workshop on Internet-based Information Systems*, (1996).

[6] C. Knoblock, Y. Arens, and C. Hsu, 'Cooperating agents for information retrieval', in *Proceedings of the 2nd International Conference on Cooperative Information Systems*, Toronto, Canada, (1994). University of Toronto Press.

[7] D. Lenat and R.V. Guha, *Building Large Knowledge-Based Systems*, Addison-Wesley, Reading, Mass., 1990.

[8] R. MacGregor, *The evolving technology of classification-based knowledge representation systems*, Principles of Semantic Networks: Explorations in the Representation of Knowledge, Morgan Kaufmann, 1991.

[9] Magellan. http://www.lib.ua.edu/maghelp.htm, 1998.

[10] Magellan. http://magellan.mckinley.com, 1999.

[11] H.S. Pinto and J.P. Martins, 'Reusing ontologies', *AAAI 2000 Spring Symposium on Bringing Knowledge to Business Processes*, 77–84, (2000).

[12] J.R. Quinlan, *C4.5: Programs for Machine Learning*, Morgan Kaufmann, San Mateo, CA, 1993.

[13] S. Staab, J. Angele, S. Decker, M. Erdmann, A. Hotho, A. Madche, H. Schnurr, R. Studer, and Y. Sure, *Semantic Community Web Portals*, 1–13, Computer Networks (Special Issue: WWW9 - Proceedings of the 9th International World Wide Web Conference), Elsevier, Amsterdam, 2000.

[14] P. Weinstein and W. Birmingham, 'Agent communication with differentiated ontologies: eight new measures of description compatibility', Technical report, Department of Electrical Engineering and Computer Science, University of Michigan, (1999).

[15] A.B. Williams and C. Tsatsoulis, 'Diverse web ontologies: What intelligent agents must teach to each other', *AAAI 1999 Spring Symposium on Intelligent Agents in Cyberspace*, 115–120, (1999).

[16] Xiaolan Zhu, *Incorporating Quality Metrics in Agent-Based Centralized/Decentralized Information Retrieval on the World Wide Web*, Ph.D. dissertation, University of Kansas, 1999.