Studies on Commonsense Server and Dialog Management

Toyoaki Nishida
Graduate School of Information Science Nara Institute of Science and Technology
Takayama 8916-5, Ikoma, Nara 630-01, Japan

1. Introduction

The purpose of this research is to establish a methodology of commonsense knowledge server capable of providing commonsense knowledge for dialog systems. The main contribution is twofold. First, we have developed a computational framework of sharing knowledge based on multi-agent systems. Second, we have developed a methodology and computational supports for collecting and systematizing knowledge from natural information sources.

2. the Knowledgeable Community KC-0

The Knowledgeable Community is our long term research goal of accumulating, systematizing, sharing and creating knowledge for humans and computers. Crucial issues in the Knowledgeable Community are (a) the knowledge media for integrating human- and computer-oriented information media, (b) the ontology for systematically specifying shared concepts, and (c) agent-assisted mediation technology for facilitating content-oriented communication among participating agents.

We have developed the first version of the Knowledgeable Community called KC-0, on top of the federation architecture proposed by the ARPA knowledge sharing effort. KC-0 mainly handles computer-oriented information media.

As an advanced mechanism for facilitating inter-agent communication, we have incorporated an ontology-based mediation mechanism into KC-0. Two kinds of agents are introduced. Facilitators take care of low level communication issues. Mediators handle conceptual-level information, based on a shared ontological information maintained by ontology servers.

We have evaluated the ideas in the context of a prototype travel assistance system called KC-Kansai, in which agents cooperatively help travelers figure out itinerary or execute the plan. We have implemented about thirty KQML speaking agents. Given a request of travel arrangement, a travel planner communicates with other information providing agents to collect necessary information.

3. Building Ontologies from Natural Information Sources

Ontology-development is quite painstaking task. We have investigated a knowledge medium which allows for data-driven ontology development, by accumulating almost raw data and incrementally creating the conceptual structure through human-computer collaboration. Our approach allows for data-driven ontology development, by accumulating almost raw data and incrementally creating the conceptual structure through human-computer collaboration.

Our information base consists of annotated raw material, weakly structured information, and strongly structured information. We assume that incremental conceptualization by human-computer collaboration takes place continually to transform annotated raw materials into weakly and strongly structured information. Strongly structured information is used to synthesize an efficient plan for answering questions, while annotated raw material and weakly structured information are used both to cope with questions that cannot be handled adequately with strongly structured information and to provide human-oriented intelligible information.

Our knowledge medium is based on unstructured means of information representation, called an associative representation. The basic entities of associative representation are (a) a unit which represents either a concept or an external datum, and (b) an association which associates a collection of key concepts with a collection of units which is normally reminded of by the given keys.

We have assumed that people implicitly encode conceptualization into associational representations as a regularity of the usage of conceptual terms, even though they are not given an explicit description tool for conceptualization, and we are able to detect such regularity by some means such as statistical analysis.

We have proposed and investigated a couple of heuristic techniques which will detect inappropriate associations from the information base and suggest a possible way of remedying them. Orthogonal decomposition attempts to decompose a given information base into coherent groups of associations, by analyzing how the user intersects associations. Analogical refinement is a less efficient but more powerful technique for further elaborating information base based on the measurement of similarity.

We have implemented the above idea as a program called CM-2 using Common Lisp and tcl/tk. We are evaluating CM-2 against accumulating various kinds of information such as research memoranda, technical surveys, regional guide, personal diary, and so on. Besides testing against these small examples and the examples described so far, we have made a couple of experiments with a nontrivial scale.

(Experiment 1) We have manually constructed a CM-2 information base for regional guide of Nara, Japan. It contains about 1,850 concepts and 870 associations. As a result of orthogonal decomposition, CM-2 has produced 212 revisions, about 80 of which have been found useful. Others are uninteresting. On the other hand, the analogical refinement heuristic has generated 65 suggestions, 20 of which are found useful.

(Experiment 2) We have gathered WWW pages in the Kansai (west) area of Japan and have generated associational representations using a simple keyword extraction and text analysis program. As a result, we have obtained about 700 concepts and 230 associations. CM-2 has proposed 1130 and 230 revisions using the orthogonal decomposition and analogical refinement heuristics. About 230 and 12 proposals by orthogonal decomposition and analogical refinement heuristics are found useful, respectively.

As an extensive work, we have incorporated an information collection and integration mechanism from Internet. We make use of domain ontology as a shared vocabulary between the user and an information gathering system. Thus, the user uses the ontology to specify the range and scope of information to be collected. In addition, the ontology provides a shared background knowledge between the user and the system, supplementing tacit information referred to during information collection and classification. We have implemented IICA (Intelligent Information Collector and Analyzer). IICA takes a set of keywords and searches for WWW pages which match them. IICA uses a weakly structured ontology.

Given a set of keywords, IICA extends them by taking into account ontological relations. Using the extended set of keywords and an anchor to a WWW page, IICA evaluates the relevance of the anchor by computing the minimal ontological distance between the extended set of keywords and the natural language part of the anchor. IICA also uses heuristic knowledge such as "I-related pages may be anchored at university pages,"to extend the given set of keywords.

IICA classifies documents by combining a keyword vector model with an ontology. IICA creates an initial classification of documents from the ontology and repeats a cycle of computing a representative characteristic vector for each category and classifying documents using the characteristic vectors until a convergence is achieved.

We have tested IICA against several examples.

4. Theory of Multiple Ontologies

We have studied a theory of multiple ontologies, and proposed Aspect Theory. We describe how multiplicity of ontology can be achieved using aspects, each of which represents a particular view of the world. In this formalization, an ontology is defined as a consistent combination of aspects. We allow more than one aspect in the information base, and provide a mechanism for translating information between different aspects.

We have designed a language of aspects called ASPECTOL (Aspect-based Ontology Description Language), which is an extension of Ontolingua-like ontology definition.

Our approach to dealing with the inter-aspect relations is to translate messages among agents with different aspects. Translation formulae are used by a translation agent which converts information from one aspect to another. If there is only a single ontology, there are no ambiguities in interpreting messages. When we allow multiple aspects, it is necessary to specify aspects the message is based on and sometimes to translate messages from an agent to the other. A translation agent is inserted between the mediator and target agents.

Mediation is performed with the almost the same way for a single ontology, except that a step to select aspects is inserted to the procedure. The basic functions of a translation agent are (a) the translation agent receives a message from an agent (source agent) to a target agent via the mediator. (b) It analyzes the message and retrieves translation formulae from the ontology server. (c) It composes a new message written in the aspect of the target agent by consulting the translation formulae, and sends it to the target agent. If the message is a informative one, the procedure will terminate. (d) It waits and receives an answer message from the target agent. (e) It composes an answer message written in the aspect of the source agent by consulting the same translation formulae used to compose the original message and sends it to the source agent.

During the procedure, the translation agent behaves that it can understand a category gent which includes both aspects in the source and the target agents.

Major Publications

[1] Toyoaki Nishida, Hideaki Takeda, and Kenji Iino. Mediation in the knowledgeable community. In Working Notes Software Agents, AAAI Spring Symposium Series, 1994.

[2] Toyoaki Nishida and Hideaki Takeda. Towards the knowledgeable community. In Kazuhiro Fuchi and Toshio Yokoi, editors, Knowledge Building and Knowledge Sharing, pages 155-164. Ohmsha, IOS Press, 1994.

[3] H. Takeda and T. Nishida. Integration of aspects in design processes. In J. S. Gero and F. Sudweeks, editors, Aritificial Intelligence in Design '94, pages 309-326. Kluwer Academic Publishers, 1994.

[4] H. Takeda, K. Iino, and T. Nishida. Agent communication with multiple ontologies. In Proceedings of FGCS'94 Workshop on Heterogeneous Cooperative Knowledge-Bases, pages 111-124, Tokyo, 1994. ICOT.

[5] Toyoaki Nishida, Hideaki Takeda, Kenji Iino, and Masanobu Nishiki. A knowledge media approach to ontology development. In N.J.I. Mars, editor, Towards Very Large Knowledge Bases: Knowledge Building & Knowledge Sharing 1995, pages 84-94. IOS Press, 1995.

[6] Michiaki Iwazume, Hideaki Takeda, and Toyoaki Nishida. Ontology-based approach to information gathering and text categorization. In Proc. ISDL'95, 1995.

[7] Hideaki Takeda, Kenji Iino, and Toyoaki Nishida. Agent organization and communication with multiple ontologies. the International Journal of Cooperative Information Systems (IJCIS), December 1995. (In printing).

[8] Hideaki Takeda, Kenji Iino, and Toyoaki Nishida. Ontology-supported agent communication. In Working notes of 1995 AAAI Spring Symposium on Information Gathering in distributed environments, pages 157-165, 1995.

[9] Toyoaki Nishida, Kazuto Koujitani, and Hideaki Takeda: A Plain Indexing Method for Organizing Conceptually Promiscuous Data, AAAI Fall Symposium Series, Workshop on AI Applications in Knowledge Navigation and Retrieval, 1995.

Keywords: Knowledge media, knowledge sharing, ontology