



Murdoch
UNIVERSITY

MURDOCH RESEARCH REPOSITORY

<http://researchrepository.murdoch.edu.au/644/>

Goh, O.S., Depickere, A., Fung, C.C. and Wong, K.W. (2006) Top-down natural language query approach for embodied conversational agent. In: International MultiConference of Engineers and Computer Scientists (IMECS 2006), 27-29 June 2006, Hong Kong.

It is posted here for your personal use. No further distribution is permitted.

Top-down Natural Language Query Approach for Embodied Conversational Agent

¹Ong Sing Goh, ²Arnold Depickere, ³Chun Che Fung, ⁴Kok Wai Wong
School of Information Technology, Murdoch University, Western Australia
{¹os.goh, ²a.depickere, ³l.fung, ⁴k.wong} @murdoch.edu.au

Abstract— This paper describes research work in implementing a conversational intelligent agent on the web focusing on a top-down natural language query approach. While the present World-Wide Web provides a distributed hypermedia interface to the vast amount of information on the Internet, there is a lack of appropriate metadata to that content. Instead of being a giant library as intended, increasing sections of the Web are looking like a giant dump. A multi-level natural language query system is described in this paper for the representation of knowledge in specific and open domains. The six layers system includes spell check, Natural Language Understanding and Reasoning, FAQChat, Metadata Index Search, Pattern matching and case-based reasoning, and, semi-automated matching approach. Extracts from queries on the field of pandemic crisis, Bird Flu H5N1 is demonstrated.

Index Terms—Natural Language Processing (NLP), Artificial Intelligence (AI), Question answering (QA) System, Embodied Conversational Agent (ECA)

I. INTRODUCTION

The ability for computers to converse with human users using natural language would arguably increase their usefulness. Research in practical dialogue systems has gained much attention in recent years [1]. Most of the dialogue systems today typically focus on helping users to complete a specific task, such as information search, planning, event management, or diagnosis. Recent advances in Natural Language Processing (NLP) and Artificial Intelligence (AI) in general have advanced this field in realizing the vision of a more humanoid interactive system. Several known futurists believe that computers will reach the capabilities comparable to human reasoning and understanding of languages by 2020 [2].

In this paper, we present our novel conversational human-computer interface with embodied conversation agent. The amalgamation of the computer graphics and AI has lead to the possibility of creating believable avatar or embodied conversation intelligent agent. Furthermore, we attempts to present an architecture that move one big step forward in this field, and includes discussion of practical application. Our architecture is aimed specifically for the interaction on Web. It involves

a talking virtual embodied character that is capable of having a meaningful conversation with the user who types in texts. Our system architecture represents a speech on demand system use Text-to-Speech (TTS) technology with lip sync encoder.

II. CONVERSATION ENGINE

This research project involves the establishment of a Crisis Communication Network (CCNet) portal. The objective is to use the embodied conversational agent, called Artificial Intelligent Neural-network Identity (AINI) [3] as the basic architecture. Our real-time prototype relies on distributed agent architecture designed specifically for the Web. A software agent, such as the conversation engine, multidomain knowledge model, multimodal human-computer communication interface and multilevel natural language query, communicates with one another via TCP/IP can be used. AINI is a conversation agent designed by the authors that is capable of having a meaningful conversation with users who interact with her. AINI is a software conversation robot, which uses human-computer communication system. This is a combination of natural language processing and multimodal communication. A human user can communicate with the developed system using typed natural language conversation. The embodied conversation agent system will reply text-prompts or Text-to-Speech Synthesis together with appropriate facial-expressions.

For the purposes of this research, the application area chosen for designing the conversation agent is primarily grounded in an ability to communicate based upon scripting and/or artificial intelligence programming in the field of pandemic crisis, Bird Flu. A sample of the communication interface between a user and AINI in the CCNet portal is depicted in Fig. 4.

As shown in Fig. 1, AINI adopts a hybrid architecture that combines the utility of multidomain knowledge bases, multimodal interface and multilevel natural language query. Given a question, AINI first performs question analysis by extracting pertinent information to be used in query formulation, such as the NPs and VPs. AINI employs an Internet three-tier, thin-client architecture that may be configured to work with

any web application. It comprises of a data server, application and client layers. This Internet specific architecture offers a flexible solution to the unique implementation requirements of the AINI system. The

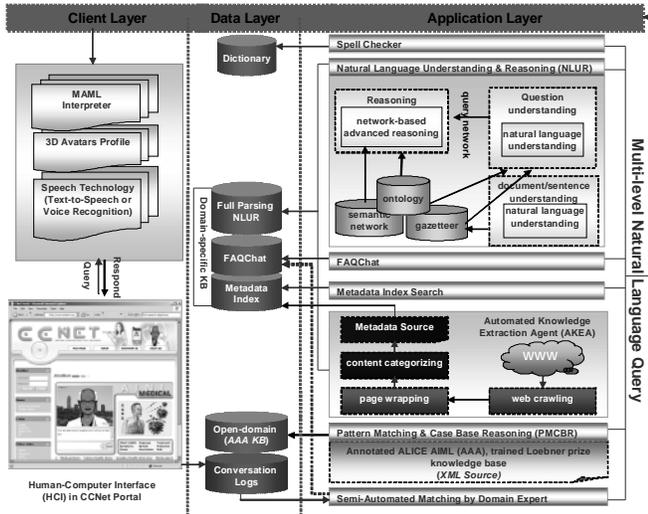


Fig. 1 AINI's Architecture in the CCNet Portal

data server layer serves as storage for permanent data required by the system, where the pandemic knowledge bases are stored. These databases are Dictionary, Domain-Specific, Open Domain and conversation logs. The dictionary is an ispell which is the first time run on TOPS-20 systems at MIT-AI lab¹. Domain-Specific database is extracted by the Automated Knowledge Extraction Agent (AKEA) [4] which consist of Full Parsing NLUR, FAQChat and Metadata Index. The Open-Domain database is taken from the existing award winning Turing Test. This trained Knowledge Base is also called Annotated ALICE Artificial Intelligence Markup Language (AAA) [5] and the conversation logs reside. These web-enabled databases are accessible via the SQL query standard for database connectivity using MySQL database.

The application server layer handles the processing of logic and information requests. Here, one or more application server is configured to compute the dialogue logic through the hybrid approach multilevel natural language query algorithm as shown in the Fig 2. The user interface resides in the thin-client layer and is completely browser based, employing Multimodal Agent Markup Language (MAML) interpreter or Microsoft SAPI to handle the users interface. MAML is a prototype multimodal markup language based on XML that enables animated presentation agents or avatars. It involves a talking virtual lifelike 3D agent character that is capable of involvement in a fairly meaningful conversation. The conversation engine is based on the Web-based and

towards an architectural open-source practice by employing PHP, Perl scripting language, Apache Server and knowledge base stored in a MySQL server.

III. NL-QUERY - TOP-DOWN APPROACH

The human brain has often times been known as the world's most extraordinary computing processor. It is capable of quickly and efficiently processed huge amounts of information. The most quoted example for this is the natural language processing. Recently in the field of AI, researchers are debating whether bottom-up or top-down approach can be best used to model human brain. Mentalese or 'language of thought' and conceptual representation support the ideas of a top-down approach [6]. However the MIT Cog Robot Team fervently supports the bottom-up approach when modeling the human brain [7].

In this paper, we simulated goal-driven or top-down NL-Query approach as human's process their language. Human examine a sentence or phrase as whole and if they do not recognize it, they will break it down into its component parts until they can recognize the parts. If a sentence or phrase is used often, it will become as down approaches are more often found in generation schemas [8], rhetorical structure theory [9] and plan-based approaches, [10] are examples of top-down approaches, where the schema or plan specifies the kind of information to include in a generated text. In early work, schemas were used to generate definitions, but the information for the definitional text was found in a knowledge base. In more recent work, information extraction is used to create a top-down approach for summarization [11] by searching for specific types of information which can be extracted from the input texts (e.g., perpetrator in a news article on terrorism).

The top-down approach seems to be a good model o explains how humans use their knowledge in conversation. After much literature search, we concluded

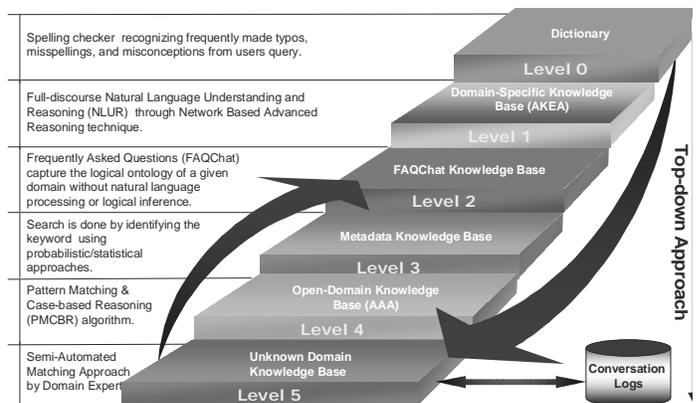


Fig. 2 Top-down Natural Language Query Approach

¹ <http://www.mit.edu/afs/sipb/project/sipb-athena/src/ispell/>

that in the field of NLP, it seems that the top-down approach is far the best approach. Therefore, we use top-down approach as our NL-Query. As shown in the Fig. 2, our top-down NL-query approach consists of 6 level of queries, namely Spell Checker (Level 0), Full-discourse NLUR (Level 1), FAQChat (Level 2), Metadata Index Search (Level 3), PMCBR (Level 4) and Semi-Automated Matching Approach (Level 5). These will discuss in details in the following sub-sections.

A. Spell checker

Level 0 is the most critical level where the system will recognize frequently made typos, spelling mistakes, and misconceptions from users query. It analyzes all terms in the user's query to recognize what is the most likely intention. Its main feature is to suggest any possible replacement for any misspelled word.

The spell check is based on occurrences of all words found on the dictionary, it is able to suggest common spellings for proper nouns (names and places) that might not appear in a standard spell check program or dictionary. The system automatically checks whether you are using the most common spelled word in your query. For example, when you ask for "What is bird flo.", the spelling checker would detected there is a wrongly spelled word in your query and will highlighted the suspected wrongly spelled "flo" which you refers to "flu". Regardless of whether it suggests an alternative spelling, the spell check will returns results that match your query if there are any. After this verification stage, the query will then go to the Level 1.

B. Full-discourse Natural Language Understanding and Reasoning (NLUR)

In the literature, full parsing and other symbolic approaches are commonly called Natural Language Understanding (NLU). Symbolic approaches mean using symbols that have a defined meaning both for humans and machines. The other approaches, e.g. statistical, are often called Natural Language Processing. This use of terms tells us that NLU seeks to do something more than just process the text from one format to another. The end goal is to transform the text into something that the computers can "understand". This means that the computer should be able to understand natural language

(e.g. English) questions from the text, and also be able to reason about facts from different texts. Therefore we have developed a set of tools for extracting data from web sites and transforming it into a structured data format, called AKEA. It was designed to establish the knowledge base for a CCNet global crisis communication system. CCNet was proposed during the height of the SARS epidemic in 2003. As reported in the [12], the AINI architecture can be scale up to be used for any new application domain such as Bird Flu pandemic in the absence of principle approach. Therefore, in this research, Bird Flu pandemic will be our domain-specific knowledge base. AKEA focused on extracting information from the World Wide Web which resolves root domain names such as who.int (World Health Organisation), pandemicflu.gov (United States Department of Health and Human Services), flu.gov.sg (Alerted, Informed, Prepared by Singapore Government), birdflu.org.cn and follows subsequent links that are available on a page until a certain depth as defined by the user.

A typical full-discourse NLUR system, as shown in Fig. 3, consists of basically two subsystems namely NLU and network based advanced reasoning system. NLU subsystem is responsible for reading and understanding two things: questions from users and sentences of processed news articles from news repository. The process is carried out in four phases by four natural language processing modules namely sentence parsing, named-entity recognition, relation inference and discourse integration.

However, in network-based advanced reasoning subsystem is responsible for discovering the valid answer and generate unambiguous response or generate an explanation for users' questions [13]. The process is executed in five phases by five modules namely network-to-path reduction, selective path matching, relaxation of event constraint, explanation on failure and template-based response generation.

The network-to-path reduction module collapses the query network into sets of path sequences to reduce the complexity in discovering the answer. The output of network-to-path reduction is two sets of path sequences that will be used by the selective path matching module. This is to discover the answer from the semantic network through a series of conditional path unification. To extend beyond literal matching of path sequence, ontological information is utilized to put into consideration events that are hierarchically equivalent. This process is performed by the module relaxation of event constraint. In case of failure during the discovery for a valid answer by selective path matching, an explanation or justification is dynamically generated by the explanation in failure module as an alternative response. This process is carried out based on the context of the question and the current status of the semantic

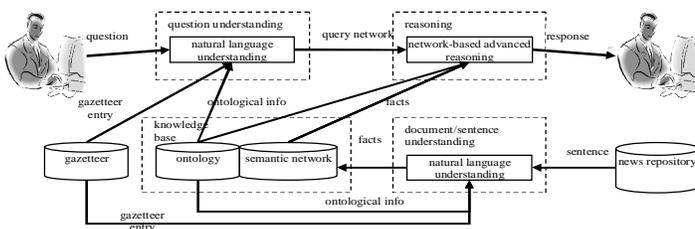


Fig. 3: Architecture of NLUR Knowledge base Query

network. If answers can be validly discovered, then readable natural language responses are generated by the template-based response generation module.

C. Frequently Asked Questions (FAQChat) capture the logical ontology of a given domain

In this Level 2, we are ignoring sophisticated natural language processing or logical inference which has already performed in the Level 1. FAQs are Frequently-Asked Questions documents, designed to capture the logical ontology of a given domain or domain-specific. Any Natural Language interface to an FAQ is constrained to reply with the given Answers, so there is no need for NL generation to recreate well formed answers, or for deep analysis or logical inference to map user input questions onto this logical ontology; a simple (but large) set of pattern-template matching rules will suffice. This simplistic approach works best when the user's conversation with the embodied conversational agent is likely to be constrained to a specific topic (in our case, we restricted to crisis communication on bird flu pandemic). By using this approach, the system have a ability to give direct answers or suggesting related links by using URL push technique [12] while Google only gives links. In addition, since FAQChat used logical ontology of a given specific domain, the pop up page returned by the FAQchat is less than those returned by Google, which saves time browsing/searching [14]).

D. Metadata Index Search

Metadata index is information about information: more precisely, it is structured information about resources indexed. It can be as simple as an author's name or as complex as a geographic code or a controlled-vocabulary subject heading. Library catalogs are remote meta data, as are book reviews, indexes to art collections and summaries. Some document formats allow metadata to be incorporated into documents or records, such as HTML <meta> tags and Dublin Core tags and database keyword fields. It gathers the metadata from pages on the Internet or an Intranet and lets users search the metadata stored in its index.

From a technological perspective, Level 3 rely on the application of a mix of linguistic rules and probabilistic or statistical principles. On one end of the spectrum, solutions apply linguistic rules to "clean" the document of any specific formatting and perform noun-phrase or verb-phrase analyses in the metadata respiratory. At the other end of the spectrum, solutions rely on simple statistics or complex probability models such as Hidden Markov Models to find occurrences and co-occurrences of terms within a document.

Base on the given online documents metadata extracted by AKEA, the query will consider semantic content by processing articles in a manner similar to humans breaking the keyword barrier and achieving

higher performance through hybrid system that incorporates keywords and limited semantic knowledge in the metadata.

E. Pattern Matching & Case-based Reasoning (PMCBR) algorithm.

This approach of conversational agents based on empirical language processing techniques is called pattern matching and case based reasoning (PMCBR). These programs descend from the early ELIZA program [15], which demonstrated that an illusion of dialogue could be supported by the recognition of simple topics in the user's discourse and the generation of arbitrary sentences in the same semantic field. They work by matching pre-defined answers to patterns recognized in the user input, such as keywords or a combination of keywords.

Our approach to handling conversations with the embodied intelligent agent is to have case-based rules that run on top of reasoning rules. The case-based rules, which do simple pattern matching, have the advantage of being quick. Therefore, it is able to return near instantaneous responses to the users. This is the ultimate goal in maintaining believability in the interaction. The Conversation Engine handles this reactive component. The PMCBR Conversation Engine is based on the ALICE [16] engine. The ALICE chat engine implements the AIML, which allows dialogs between the user and agent to be easily acknowledged. Judging from the specification based on XML, we selected AIML as our the perfect markup language for our system.

AIML consists of data objects called AIML objects, which are made up of units called topics and categories. The topic is an optional top-level element, it has a name attribute and a set of categories related to that topic. Categories are the basic unit of knowledge in AIML. Each category is a rule for matching an input and converting to an output. Each category also consists of a pattern, which represents the user input, and a template, which implies the AINI robot answer. The AIML pattern is simple, consisting only of words, spaces, and the wildcard symbols _ and *. The words may consist of letters and numerals, but no other characters. Words are separated by a single space, and the wildcard characters function like words. The pattern language is case invariant. The idea of the pattern matching technique is based on finding the best and longest pattern match.

F. Semi-Automated Matching approach by Domain Expert

To date, early systems for extracting "semantic signatures," or conceptual representations, from documents relied strictly on manual processes involving human expert judgment. In essence, the expert would match the analyzed document to a pre-defined taxonomy. The advent of the information overflow phenomenon has forced the gradual automation of this extraction. Today,

several entirely automatic or at least partly automatic solutions have been proposed. These solutions, while not reaching the level of human expertise, have shown that they can efficiently determine the semantic signature within unstructured documents.

It is difficult for a software agent to understand the data representations for different models since the models might differ both semantically and syntactically. A semi-automated matching approach is much more achievable, especially when performed within only one specific domain. Human intervention can improve model matching from two aspects. First, a human expert is able to set up a matching context, by applying domain constraints or configuring heuristic parameters, to speed matching. Second, a domain expert can correct some errors during the matching procedure and follow by training to avoid future errors. Therefore, compared to a fully automated approach, a domain-specific semi-automated approach that utilizes prior matching knowledge and domain knowledge will undoubtedly lead to better performance and accuracy.

In this final stage, the result checking will still be the responsibility of the domain expert, who will be able to correct matching errors and to pick the proper matching result from a list of possible matches from the conversation log which is unanswered by AINI. Finally, the newly generated matching rules subsequently will be stored and upgraded into FAQChat knowledge base in the Level 2 by domain expert. The process of queries from Level 0 to Level 5 will continue until the query can be handled.

IV. INTEGRATION DOMAIN KNOWLEDGE AND NL-QUERY IN CONVERSATIONAL SYSTEM

AINI's domain knowledge model usually incorporates several knowledge domains, thus merging the expertise of one or more experts. A "sales" domain knowledge for instance, would contain expertise on improving sales, but it would also incorporate with an Open-Domain knowledge. Multiple domain knowledge, merged into AINI's single domain knowledge would give the users the best conversation.

We pre-defined the Open-Domain and Domain-Specific in the data layer. Based on the type of input provided by the user, the agent's response state moves smoothly from one domain knowledge base and NL-query Level to another respectively as shown in the Fig. 2. According to K. Mori, et. al. [17], these two intermediate states transition called "Reluctant" and "Concede".

Even though the conversation agent allows the user to carry the conversation beyond their domain knowledge, however the conversational agent will continue to remind and recall the user by bringing back to the current topic of the presentation. This is to convey

and direct the users' attention to move back to its original Open-Domain or Domain-Specific state. However, the priority will be Domain-Specific. Therefore, the conversation agent will always give higher priority to Domain-Specific in an attempt to keep the user focused on the topic of the presentation. An example of interaction domain knowledge model and NL-Query conversation between Isabel and AINI on the bird flu pandemic is shown in Fig. 4.

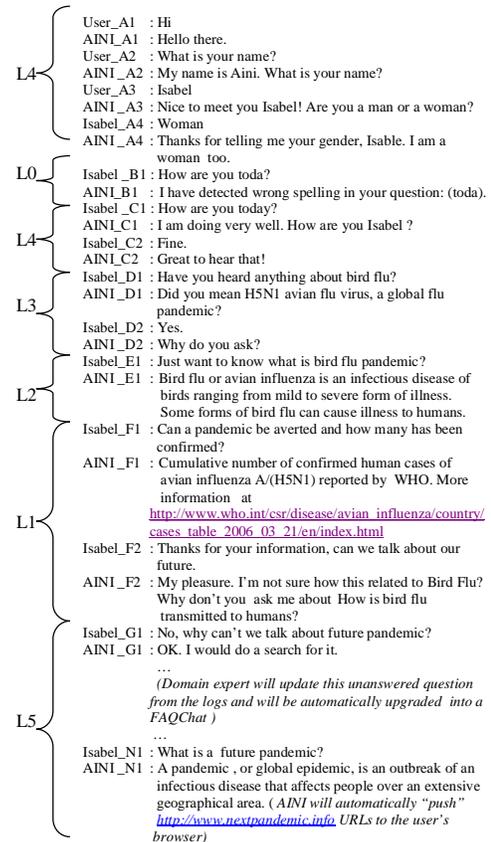


Fig. 4 Conversation logs on H5N1 Bird Flu Pandemic

From the conversation logs, the transition state in dialog pair A1-A4 and C1-C2 used Open-Domain from the NL-Query Level 4 where PMCBR approach has been carried out. In the dialog pair B1, the NL-query can't proceed because the system found misspellings "today" misspelled as [toda] which has been highlighted in the respond. In the dialog pair D1-D2, NL-Query Level 3 has been imposed where the search done by identifying the keyword or phrase using probabilistic or statistical approaches from the metadata index. In the dialog pair E1, FAQChat approach captured the logical ontology of a given domain. In this Level 2, FAQChat is constrained to reply with the given Answers without NL generation to recreate well formed answers. However in dialog pair F1-F2, full-discourse NLUR through Network Based Advanced Reasoning technique with Domain-specific

has been used. In the dialog pair, G1 shows that the AINI is unable to answer user question but she will forward the random statement such as “*I would do a search for it.*”, “*Did I misunderstand your meaning?*”, “*That’s an interesting question.*”, “*I’ll come back to that in a minute*” etc and these statements will be monitored and submitted into the unanswered conversation logs data layer. In this Level 5, domain expert will be responsible to pick up the proper matching result from a list of possible matches. Finally, the newly generated matching rules subsequently will be stored and upgraded into the Domain-specific knowledge set. Another significant result shows that in the dialog pair G2, the user had control of the conversation although the agent reminded the user of the topic of the current presentation in the dialog pair F2. In addition, in the Level 5, the domain expert can also integrate the answer with relevance source from the internet using the “URL Push” technique. This will make the conversation more interesting and the information forwarded to the user is up-to-date.

V. CONCLUSION

In conclusion, the top-down NL-Query approach provides an inexpensive method for reformulating queries based on previously query. In order to exploit the efficiency and the reliability of the algorithm, such a system will be designed to maximize the recall of retrieved candidate answers. Instead of just performing a deep linguistic analysis from domain-specific in Level 1, the system will delegate to the evaluation component from the selection of the right answer from our proposed top-down approach in multilevel NL-Query architecture.

Embodied Conversational Agent may indeed play an important role in popularizing the concept of conversational characters, paving the way for the more humanoid user interface based on human language technologies, whose deployment can only be envisioned in the mid-term future. Based on this experiment, the top-down NL-query approach shows interesting behavior in the natural conversation agent. The key assumption is that important queries do not necessarily turn up the answers that they can be found in a number of different domains. In this paper, we only worked on selected pandemic crisis websites which are used to perform knowledge extraction for the Domain-Specific database on the server. Although we simulate the proxy conversation log that contains of client requests, there is

a possibility that new simulation result from other traces is different from the result referred in this paper.

REFERENCES

- [1] J. Allen, D. Byron, M. Dzikovska, G. Ferguson, L. Galescu, and A. Stent, "Towards conversational human-computer interaction," *AI Magazine*, vol. 22, 2001.
- [2] R. J. Lempert, S. W. Popper, and S. C. Bankes, *Shaping the next one hundred years: new methods for quantitative, long-term policy analysis*. Santa Monica, CA.: RAND, 2003.
- [3] O. S. Goh, C. C. Fung, and L. M. Ph'ng, "Intelligent Agents for an Internet-based Global Crisis Communication System," *Journal of Technology Management And Entrepreneurship*, vol. 2, 2005.
- [4] O. S. Goh and C. C. Fung, "Automated Knowledge Extraction from Internet for a Crisis Communication Portal," presented at First International Conference on Natural Computation, Changsha, China, 2005.
- [5] H. Loebner, (2003), Loebner Prize Gold Medal, [Online]. Available: <http://www.loebner.net/Prize/loebner-prize.html>
- [6] J. A. Fodor, *Elm and the Expert: An Introduction to Mentalese and Its Semantics*: Cambridge University Press, 1994.
- [7] R. A. Brooks, "The Cog Project: Building a Humanoid Robot," presented at The 1st International Conference on Humanoid Robots and Human friendly Robots, Tsukuba, Japan, 1998.
- [8] K. R. McKeown, *Text Generation*: Cambridge University Press, 1995.
- [9] J. D. Moore and C. L. Paris, "Planning text for advisory dialogues: Capturing intentional and rhetorical information," *Computational Linguistics*, vol. 19, pp. 651-695, 1992.
- [10] E. Reiter and R. Dale, *Building Natural Language Generation Systems*: Cambridge Univ. Press, 2000.
- [11] D. Radev and K. McKeown, "Generating natural language summaries from multiple on-line sources," *Computational Linguistics*, vol. 24, pp. 469-500, 1998.
- [12] O. S. Goh, C. C. Fung, A. Depickere, K. W. Wong, and W. Wilson, "Domain Knowledge Model for Embodied Conversation Agent," presented at the 3rd International Conference on Computational Intelligence, Robotics and Autonomous Systems (CIRAS 2005), Singapore, 2005.
- [13] W. Wong and O. S. Goh, "Evaluation of Response Quality for Heterogeneous Question Answering," presented at IEEE/WIC/ACM International Conference on Web Intelligence (WI 2005), Compiègne, France, 2005.
- [14] B. A. Shawar, E. Atwell, and A. Roberts, "FAQChat as an Information Retrieval System," presented at the 2nd Language & Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics, Poznan, Poland, 2005.
- [15] J. Weizenbaum, "ELIZA - A computer program for the study of natural language communication between man and machine," *Communications of the ACM*, vol. 9, pp. 36-45, 1966.
- [16] Alice, (2005), Artificial Linguistic Internet Computer Entity, [Online]. Available: <http://www.alicebot.org>
- [17] K. Mori, A. Jatowt, and M. Ishizuka, "Enhancing Conversational Flexibility in Multimodal Interactions with Embodied Lifelike Agents," presented at International Conference on Intelligent User Interfaces, Miami, Florida, USA, 2003.