A Chat-bot based Multimodal Virtual Guide for Cultural Heritage Tours.

Conference Paper · January 2006

Source: DBLP

CITATIONS
17

READS
2,140

5 authors, including:

Antonella Santangelo
Università degli Studi di Palermo
23 PUBLICATIONS **159** CITATIONS
SEE PROFILE

Agnese Augello
Italian National Research Council
141 PUBLICATIONS **956** CITATIONS
SEE PROFILE

Antonio Gentile
Università degli Studi di Palermo
119 PUBLICATIONS **1,021** CITATIONS
SEE PROFILE

Giovanni Pilato
Italian National Research Council
217 PUBLICATIONS **1,556** CITATIONS
SEE PROFILE

A Chat-Bot based Multimodal Virtual Guide for Cultural Heritage Tours

A. Santangelo[1], A. Augello[1], A. Gentile[1]

[1]DINFO - Dipartimento di ingegneria INFOrmatica
University of Palermo
Viale delle Scienze - 90128 Palermo – Italy
Tel: +39 0917028528 Fax: +39 0916529124
{augello,santangelo}@csai.unipa.it
{gentile,gaglio}@unipa.it

G. Pilato[2], S. Gaglio[1,2]

[2]ICAR - Istituto di CAlcolo e Reti ad alte prestazioni
Italian National Research Council
Viale delle Scienze - 90128 Palermo - Italy
Tel: +39 091238269 Fax: +39 0916529124
pilato@pa.icar.cnr.it

Abstract *- The use of a PDA with ad-hoc built-in information retrieval functionalities can help people in visiting an historical site in a natural manner instead of traditional audio/visual pre-recorded guides. The aim of this work is to build a versatile virtual-guide system adaptable to the user needs of mobility and therefore usable on different devices (e.g. PDAs, Smartphones). An information retrieval service is included and is easily accessible through a spoken language interaction. The system takes the advantages of chat-bot and speech recognition technologies, allowing a natural interaction with the user. The system has been implemented on a Qtek PDA with Windows Mobile 2003 with the aim of assisting an user during a visit to a cultural heritage institution.*

Keywords: *Multimodal Interfaces, Chat-bots, PDAs, Information Retrieval, Cultural Heritage.*

1 Introduction

In this work it is presented a multimodal, user friendly, virtual guide for specific contexts. The aim is to build a system adaptable to the user needs of mobility and therefore usable on different devices (i.e. PDAs, Smartphone). Besides information retrieval service is easily accessible through a spoken language interaction. In recent years, the Internet access availability trough mobile and handheld devices, the advent of third-generation wireless communication and Bluetooth have made pervasive computing a reality. Pervasive, or ubiquitous, computing aims to totally modify the way people interact with computers. Nowadays computing and networking technologies are more an more powerful and reliable. People can deal with a wide set of computing devices in their living and working spaces. Often we can find computing devices of various size and capabilities in an unusual context, such as car console, cultural heritage institutions or other public environments. Therefore, the goal is to allow an easy access to such technologies and enable the user to focus on their activities and not on the device [9]. However, the constraints inherent to pervasive devices such as PDAs, imply many problems. In [12] it has been shown how the complexity of information searching on pervasive devices can be further reduced if the browsing is limited on a specific topic. This focused search allows useful topic-specific query formulation assistance, including word completion as well as phrase completion and expansion. Human speech language usage, in human-machine communication, enables a user fulfilling interaction. In particular this kind of interaction shows an alternative way for traditional PDA interfaces. In [8] we can see the difference between using speech for input compared to traditional pen input on a PDA. In this manner computing device can replace human assistants like secretaries or museum guides. It is possible to build ad hoc applications in according to context and user needs obtaining customize interaction [1]. In [2] an example of a set of personal assistants, each specialized in a different task such as email or calendar management can be seen. This Smart Personal Assistants open the possibility of vocal interaction with the user. They are like intelligent virtual entities with which users may interact like with an human being. Speech recognition technology has been used in chat-bot systems. As an example, AliceTalker is a [4] Swing-based Java client application which enables the user to interact via spoken words and synthesized speech with an AliceBot server running on the same or a remote machine. It uses Cloudgarden's implementation of Sun's Java Speech API. The goal of this work is to implement a simple human-computer dialogue system that can interact with users through a multimodal interface. The dialogue is achieved by means of chat-bots technology, and in particular it is implemented through the Alicebot (Artificial Linguistic Computer Entity) project [3]. This chat-bot uses a natural language processing (NLP) module that provides a natural language sentences interpretation mechanism that follows a pattern-matching scheme. The dialogue is managed exploiting the chat-bot knowledge base described by question-answer modules. The interaction system is equipped with a framework that allows the processing and the comprehension of the speech, performing a vocal user interface (VUI). The multimodal interface has been developed with XHTML+Voice language. This language combines XHTML and Voice and is called X+V for short. The Voice part is a subset of VoiceXML 2.0. The developed environment is the Multimodal Tools 4.1.2.2 for

WebSphere Studio V5.1.2. This technology is available at IBM developerWorks site [3] with IBM Scholar Program License. The Multimodal Tools include the IBM WebSphere Multimodal Toolkit and IBM Multimodal Browser, respectively the developed environment and the technology for the application execution. The application has been implemented on a Qtek PDA with Windows Mobile 2003 with the aim of assisting an user during a visit to a cultural heritage institution (Archaeological Sites, Museums, etc.). In fact the use of a PDA with ad-hoc built-in information retrieval functionalities is particularly suitable for a cultural heritage environment and can help people in visiting an historical site in a natural manner instead of traditional audio/visual pre-recorded guides.

2 System architecture

The selected architecture for multimodal application is a client-server paradigm. **Figure 1** shows a representation of such architecture. It runs easily on small handheld devices like cell phones and PDAs that serve many functions and contain sufficient processing power to handle a variety of tasks. The X+V [11] application runs on these handheld devices through the multimodal browser. The multimodal browser is a software application that enables users to display and interact via wireless connection with X+V documents hosted by web servers

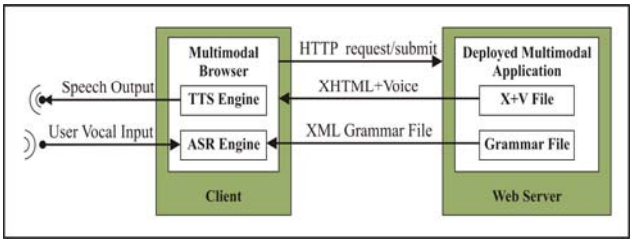

Figure 1: The system architecture.

2.1 A.L.I.C.E. interface

A.L.I.C.E. (Artificial Linguistic Internet Computer Entity) is an artificial intelligence natural language chat robot based on an experiment specified by Alan M. Turing in 1950. A.L.I.C.E has won three times (2000, 2001, 2004) the annual Loebner Prize which declares to "advance AI and serve as a tool to measure the state of the art", which makes it a good candidate tool for human-like natural-language based dialogue systems. The A.L.I.C.E chat-bot's knowledge base is composed of question-answer modules, called categories and structured with AIML (Artificial Intelligence Mark-up Language), an XML-like language designed for creating stimulus-response chat robots. The question, or stimulus, is called "pattern" and consists only of words, spaces, and the wildcard symbols _ and *. The words may consist of letters and numerals, but no other characters. The answer, or response, is called "template". The template is composed of a natural language sentence and sometimes by other AIML tags, that can transform the response in a software that can save data, can start other programs, can give conditional answers and can recursively call other categories. Two optional contexts <that> e <topic> can be used in order to keep trace of the dialogue history (<that>) or switch the dialogue on a specific subject (<topic>). There are three kind of AIML categories: atomic, default and recursive. The former is the simplest kind of category, the pattern does not contain wild-cards and the template is a simple natural language sentence. The default category allows the chat-bot answer according to a partial match with the user question. The latter category is characterized by the <srai> tag presence in the template; this tag performs a recursive call to a given category. Recursive categories allow to implement the synonymy, the symbolic reduction, and the grammar correction, and so on. Besides AIML supports unary predicates with the <set> and <get> tags. To achieve efficient pattern matching time, and a compact memory representation, the AIML software stores all of the categories in a tree managed by an object called the Graphmaster. The graph's branches are the words representing the cultural baggage of the chat-bot, the path from the root to a terminal node represent a specific pattern and the terminal node link the relative template. The Graphmaster merges patterns with common prefixes in the same pathways, achieving considerable compression. The dialogue mechanism is based on a search algorithm that is a special case of backtracking, depth-first search. The dialogue algorithm looks for a match between the user's question and the chat-bot patterns (pattern matching).

2.2 Multimodal interface

The interface has been developed with the Multimodal Tools 4.1.2.2 for WebSphere Studio V5.1.2. They are available at IBM developerWorks site [3] with IBM Scholar Program License. The Multimodal Tools includes the IBM WebSphere Multimodal Toolkit and IBM Multimodal Browser for PC. The multimodal toolkit includes a developing, debugging and testing environments for multimodal application. Two versions of multimodal browser for PDA are also available: they have been developed in a strategic relationship with Opera Software (based on the Opera Browser V7.55) and ACCESS Systems Company (based on the NetFront Browser V3.1 by ACCESS Systems). Each one of them has been improved with proper extensions that include the IBM ViaVoice speech recognition and text-to-speech technology, allowing user to view and interact with multimodal applications built using X+V on handheld device.

2.3 ALICE-KB induced grammar

In Figure 2 (Alice's knowledge base) it is shown how the Alice-KB categories are stored in the Graphmaster. From this structure it is easy to induce an XML grammar form. This grammar is used by the multimodal system to understand the spoken user utterance. A set of AIML files concerning the cultural heritage field have been written. The list of utterances that the user can pronounce during the interaction with the chat-bot has been extracted from the AIML files. A parser has been created to translate this list of utterances into an XML Form grammar file (see XML grammar in the Figure 2). The process is executed dynamically: the application runs a php file that runtime generates the grammar file allowing the system to adapt itself during the dialogue. If the number of AIML files is too large, it is convenient to generate the grammar offline, in order to reduce the response time of the system. By means of the multimodal browser it is possible to recognize the user utterance and to synthesize the Chat-Bot textual reply. The ASR (Automatic Speech Recognition) is obtained through grammar form. X+V can support the XML Form and the ABNF Form of the W3C Speech Recognition Grammar Specification [10]. For this application we have chosen the XML Form.

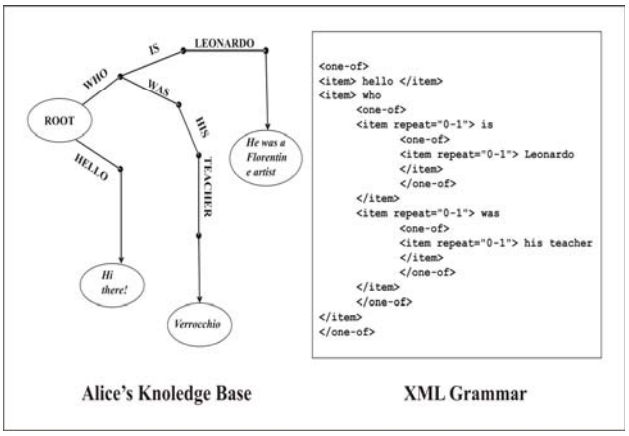

Figure 2: From AIMLTree to XML Grammar

3 System implementation

The interface of the application is a simple web page. This choice allows the reduction of the computational and memory requirements of the PDA. When the user accesses the system from his handheld device, a multimodal browser is open. After a setup procedure where the user profile is settled, the multimodal browser asks the web server for the php file and the dialogue interaction starts. The web server executes the php commands, submits the query (user input) to the chat-bot, waits for response and inserts the reply in a specific VoiceXML form. Finally it provides an X+V document to the PDA browser that makes text-to-speech of the chatbot reply. The multimodal browser, for any user utterance, loads the new page with the response and waits for a new vocal input. As shown in Figure 3 whenever the user pronounces his request, the multimodal browser looks for a match in the grammar file. If a match is found, and the user confirms it, the application answers to the request and the multimodal browser loads the new page, otherwise it asks the user to repeat the question.

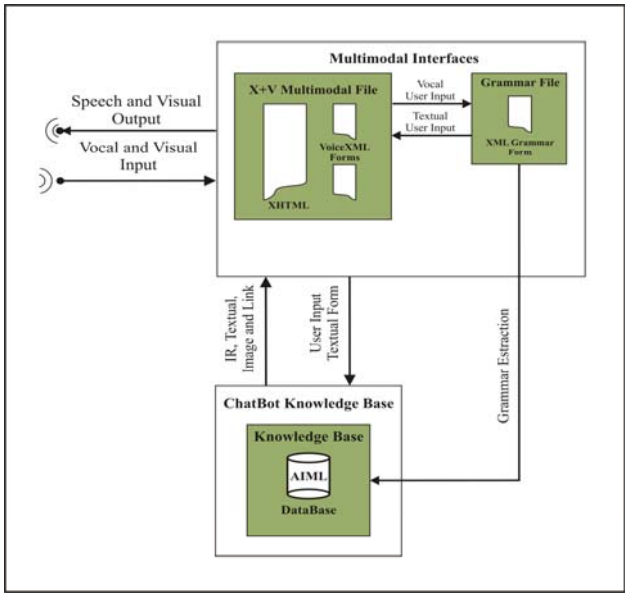

Figure 3: System information flow.

A list of the past dialogue sentences is shown in the left part of the web page. The user can interact with the chat-bot during the museum tour. So he can walk without constraints in the cultural heritage site and access with vocal interaction to information about works that he currently looks, as if the PDA would be a real human guide. An example of interaction is shown as follows:

User: Tell me about the David of Michelangelo.

Chat-bot: David / Apollo 1530. It is kept in the Museo Nazionale del Bargello, at Florence. The statue was commissioned by the hated Papal Governor of Florence…

User: Who are the principal exponents of Mannerism.

Chat-bot: Mannerists include Pontormo, Bronzino, Parmigianino, El Greco ,Tintoretto, Bernini, Brunelleschi, Benvenuto Cellini, Caravaggio and Tiziano.

User: What is the light and shade?

Chat-bot: It is a technique to model and define forms through contrasts of light and shadow.

User: Tell me about Antonello Da Messina life.

Chat-bot: He was trained in Naples, where he studied the work of Flemish artist. His earliest known works are the Crucifixion ad St. Jerome in His Study …

User: Who is Leonardo da Vinci?

Chat-bot: Leonardo da Vinci was a Florentine artist.

User: Who was his teacher?

Chat-bot: In Verrocchio's workshop Leonardo was introduced to many activities.

4 Conclusions

A multimodal pervasive system has been presented to assist an user in visiting cultural heritage sites in a natural manner overcoming the interaction limitation of traditional pre-recorded audio/visual guides. The system takes the advantages of chat-bot and speech recognition technologies embedded in a Personal Digital Assistant, which can be easily carried during the site tour. Future work will regard the enhancement of the system improving the interaction naturalness.

5 References

[1] A. Cavalluzzi, B. De Carolis, S. Pizzutilo and G. Cozzolongo, "Interacting with embodied agents in public environments". In *Proceedings of the Working Conference on Advanced Visual interfaces* (Gallipoli, Italy, May 25 - 28, 2004). AVI '04. ACM Press, New York, NY, 240-243, 2004.

[2] A. Nguyen and W. Wobcke, "An agent-based approach to dialogue management in personal assistants." In *Proceedings of the 10th international Conference on intelligent User interfaces* (San Diego, California, USA, January 10 - 13, 2005). IUI '05. ACM Press, New York, NY, 137-144, 2005.

[3] Alice: http://www.alicebot.org

[4] AliceTalker: http://www.cloudgarden.com/alicetalker/index.html

[5] G. Pilato, G. Vassallo, A. Augello, M. Vasile and S. Gaglio, "Expert Chat-Bots for Cultural Heritage", *IX Convegno della Associazione Italiana Intelligenza Artificiale Proc. of. Workshop Interazione e Comunicazione Visuale nei Beni Culturali*, 15 Sept. 2004, Perugia, Italy.

[6] IBM DeveloperWorks: http://www-128.ibm.com/developerworks/subscription/downloads/index.html

[7] IBM Multimodal Tools: http://www-306.ibm.com/software/pervasive/multimodal/

[8] K. Lyons, C. Skeelsand and T. Starner, "Providing support for mobile calendaring conversations: a wizard of oz evaluation of dual--purpose speech." *In Proceedings of the 7th international Conference on Human Computer interaction with Mobile Devices &Amp; Services* (Salzburg, Austria, September 19 - 22, 2005). MobileHCI '05, vol. 111. ACM Press, New York, NY, 243-246, 2005.

[9] M. WEISER, "The computer for the twenty-first century." *Sci. Amer. 265*, 3 (Sept.), 94–104, 1991.

[10] SRGS Ver1.0: http://www.w3.org/TR/2003/PR-speech-grammar-20031218/

[11] XHTML+Voice Profile 1.2: http://www.voicexml.org/specs/multimodal/x+v/12/

[12] Y. Aridor, D. Carmel, Y. S. Maarek, A. Soffer and R. Lempel, "Knowledge encapsulation for focused search from pervasive devices." *ACM Trans. Inf. Syst.* 20, 1 (Jan. 2002), 25-46, 2002.