

AIDA: Artificial Intelligent Dialogue Agent

Rafael E. Banchs, Ridong Jiang, Seokhwan Kim, Arthur Niswar, Kheng Hui Yeo

Natural Language Understanding Lab, Human Language Technology Department

Institute for Infocomm Research, Singapore 138632

{rembanchs,rjiang,kims,aniswar,yeokh}@i2r.a-star.edu.sg

Abstract

This demo paper describes our Artificial Intelligent Dialogue Agent (AIDA), a dialogue management and orchestration platform under development at the Institute for Infocomm Research. Among other features, it integrates different human-computer interaction engines across multiple domains and communication styles such as command, question answering, task-oriented dialogue and chat-oriented dialogue. The platform accepts both speech and text as input modalities by either direct microphone/keyboard connections or by means of mobile device wireless connection. The output interface, which is supported by a talking avatar, integrates speech and text along with other visual aids.

1 Introduction

Some recent efforts towards the development of a more comprehensive framework for dialogue supported applications include research on multi-domain or multi-task dialogue agents (Komatani et. al 2006, Lee et. al 2009, Nakano et. al 2011, Lee et. al 2012). With this direction in mind, our Artificial Intelligent Dialogue Agent (AIDA) has been created aiming the following two objectives: (1) serving as a demonstrator platform for showcasing different dialogue systems and related technologies, and (2) providing an experimental framework for conducting research in the area of dialogue management and orchestration.

The main objective of this paper is to present and describe the main characteristics of AIDA. The rest of the paper is structured as follows. First, in section 2, a description of APOLLO, the software integration platform supporting AIDA is presented. Then, in section 3, the main features of AIDA as a dialogue management and orchestration platform are described, and a real example of human interaction with AIDA is reported. Finally, in section 4, our conclusions and future work plans are presented.

2 The APOLLO Integration Platform

APOLLO (Jiang et al. 2012) is a component pluggable dialogue framework, which allows for the interconnection and control of the different components required for the implementation of dialogue systems. This framework allows for the interoperability of four different classes of components: dialogue (ASR, NLU, NLG, TTS, etc.), managers (vertical domain-dependent task managers), input/output (speech, text, image and video devices), and backend (databases, web crawlers and indexes, rules and inference engines).

The different components can be connected to APOLLO either by means of specifically created plug-ins or by using TCP-IP based socket communications. All component interactions are controlled by using XML scripts. Figure 1 presents a general overview of the APOLLO framework.

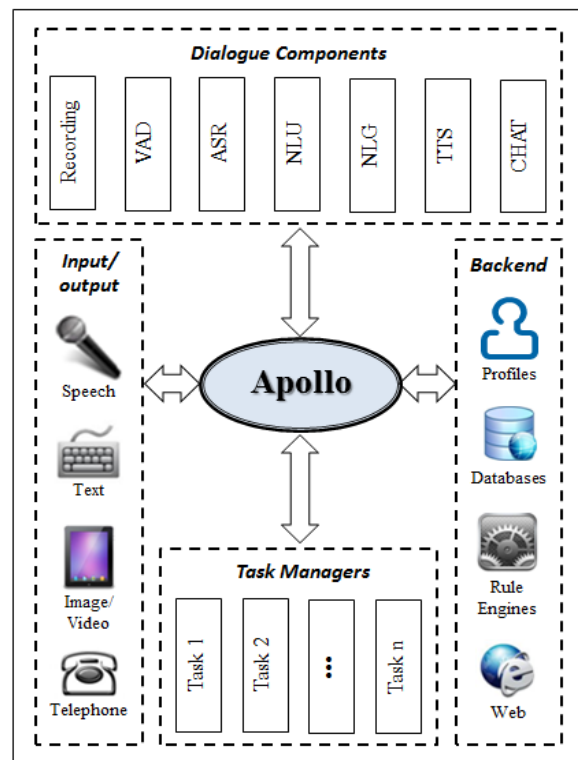


Figure 1: The APOLLO framework

3 Main Features of AIDA

AIDA (Artificial Intelligent Dialogue Agent) is a dialogue management and orchestration platform, which is implemented over the APOLLO framework. In AIDA, different communication task styles (command, question answering, task-oriented dialogue and chatting) are hierarchically organized according to their atomicity; i.e. more atomic (less interruptible) tasks are given preference over less atomic (more interruptible) tasks.

In the case of the chatting engine, as it is the least atomic task of all, it is located in the bottom of the hierarchy. This engine also behaves as a back-off system, which is responsible for taking care of all the user interactions that other engines fail to resolve properly.

In AIDA, a dialogue orchestration mechanism is used to simultaneously address the problems of domain switching and task selection. One of the main components of this mechanism is the user intention inference module, which makes informed decisions for selecting and assigning turns across the different individual engines in the platform.

Domain and task selection decisions are made based on three different sources of information: the current user utterance, which includes standard semantic and pragmatic features extracted from the user utterance; engine information states, which takes into account individual information states from all active engines in the platform; and system expectations, which is constructed based on the most recent history of user-system interactions, the task hierarchy previously described and the archived profile of the current user interacting with the system.

Our current implementation of AIDA integrates six different dialogue engines: **(BC)** a basic command application, which is responsible for serving basic requests such as accessing calendar and clock applications, interfacing with search engines, displaying maps, etc.; **(RA)** a receptionist application, which consists of a question answering system for providing information about the Fusionopolis Complex; **(IR)** I²R information system, which implements as question answering system about our institute; **(FR)** a flight reservation system, which consists of a frame-based dialogue engine that uses statistical natural language understanding; **(RR)** a restaurant recommendation system, which implements a three-stage frame-based dialogue system that uses rule-based natural language understanding, and **(CH)** our IRIS chatting agent (Banchs and Li, 2012).

Regarding input/output modalities, speech and text can be used as input channels for user utterances. Direct connections via microphone and keyboard are supported, as well as remote connections via mobile devices.

Additionally, audio and video inputs are used to provide AIDA with user identification and tracking capabilities. In the first case, speaker identification techniques are used to compare the voice profile of the current speaker with a set of users already known by the system. In the second case, face detection and tracking are used in combination with sound localization to determine what the current speaker's location is when dealing with multi-party dialogue scenarios.

The main output of AIDA is composed of a browser interface in which several frames, including different visual elements, are presented along with a talking avatar. The different visual elements include a banner, a text display and a general purpose frame for displaying different html files, images or URL's as required.

For avatar and text-to-speech, AIDA relies on the SitePal API (SitePal 2013). This API allows for controlling both the avatar speech and gaze. Speech functionalities are used for generating a spoken form of the system's output (which is also displayed as text in one of the frames of the interface) and gaze functionalities are used in combination with face tracking to direct the avatar attention to the user who is speaking.

Figure 2 illustrates the main output interface of AIDA.



Figure 2: AIDA's output interface.

Finally, an actual example of a user interaction with the AIDA platform is presented in Table 1. In the table, the following information is provided for each of the turns composing the dialogue: the user utterance, the corresponding system response, the engine that was responsible for generating the response, and the action taken by the system (if any action was taken).

Turn	Speaker	Utterance/Action
1	USER	Good morning, how are you?
	AIDA-CH	How can I help you?
2	USER	Can you help me to book a flight?
	AIDA-FR	Flight reservation, how can I help you?
	(action)	<i>Displays flight reservation main screen</i>
3	USER	I want to fly from Singapore to Tokyo
	AIDA-FR	What day do you want to leave?
4	USER	Next Monday in the evening
	AIDA-FR	What day do you want to return?
5	USER	The second of April in the morning
	AIDA-FR	Here is your itinerary
	(action)	<i>Displays flight schedule on the screen</i>
6	USER	Is there any coffee shop nearby?
	AIDA-RA	There are three coffee shops in this building: Jass Café, Starbucks and Ya Kun
	(action)	<i>Shows logos of retrieved coffee shops</i>
7	USER	Do you drink coffee?
	AIDA-CH	Coffee?
8	USER	Yes, coffee
	AIDA-CH	No
9	USER	Why not?
	AIDA-CH	Well, for one thing you can't spell.
10	USER	Please suggest a good Italian restaurant
	AIDA-RR	What area of the city do you want me to search for?
	(action)	<i>Displays restaurant system main screen</i>
11	USER	Near Sentosa
	AIDA-RR	Ok, I have found one restaurant for you: Alkaff Mansion Ristorante. Do you want more information or do you want to make a reservation?
	(action)	<i>Displays http://www.alkaff.com.sg/</i>
12	USER	Yes please
	AIDA-RR	Ok, I will do a reservation for you! And, how many guests is this reservation for?
13	USER	Only two persons
	AIDA-RR	And, will you be interested in dinner or lunch?
14	USER	Please make it for tonight
	AIDA-RR	Your reservation has been made.
	(action)	<i>Displays reservation details</i>
15	USER	Great, thank you
	AIDA-CH	You are welcome

Table 2: Example of a user interaction with AIDA

4 Conclusions and Future Work

In this demo paper we have presented AIDA (Artificial Intelligent Dialogue Agent), the dialogue management and orchestration platform currently under development at the Institute for Infocomm Research in Singapore.

AIDA integrates different communication task styles (command, question answering, dialogue and chat) into a single conversational agent. It also combines speech recognition and synthesis, image processing and avatar technologies to improve the user-system interaction experience.

Our future plans include the incorporation of more vertical applications into the platform, as well as the integration of machine translation engines to develop multilingual capabilities.

Demonstration Plan

During the SIGDIAL demo presentation, the following functionalities will be demonstrated: text and speech input; dialogue orchestration among receptionist, flight reservation, I²R information system, restaurant booking and chatting agent; and avatar-supported speech and visual output interface. For the case of speech input and avatar-supported output, the use of these technologies is subject to the availability of internet connection at the location of the demo.

References

- R. E. Banchs and H. Li. 2012. IRIS: a chat-oriented dialogue system based on the vector space model, in Demo Session of Association of Computational Linguistics, pp. 37–42.
- R. Jiang, Y. K. Tan, D. K. Limbu and H. Li. 2012. Component pluggable dialogue framework and its application to social robots. In *Proc. Int'l Workshop on Spoken Language Dialog Systems*.
- K. Komatani, N. Kanda, M. Nakano, K. Nakadai, H. Tsujino, T. Ogata and H. G. Okuno. 2006. Multi-domain spoken dialogue system with extensibility and robustness against speech recognition errors. In *Proc. SIGdial Workshop on Discourse and Dialogue*, pp. 9–17.
- C. Lee, S. Jung, S. Kim and G. G. Lee. 2009. Example-based dialog modeling for practical multi-domain dialog system. *Speech Communication*, 51, pp. 466–484.
- I. Lee, S. Kim, K. Kim, D. Lee, J. Choi, S. Ryu and G. G. Lee. 2012. A two step approach for efficient domain selection in multi-domain dialog systems. In *Proc. Int'l Workshop on Spoken Dialogue Systems*.
- M. Nakano, S. Sato, K. Komatani, K. Matsutama, K. Funakoshi and H. G. Okuno. 2011. A two stage domain selection framework for extensible multi-domain spoken dialogue systems. In *Proc. SIGdial Workshop on Discourse and Dialogue*.
- SitePal API & Programmer Information, accessed on June 27th, 2013 <http://www.sitepal.com/support/>