

Sammy: a Friendly English/French/Italian Assistant

WOCHAT SHARED-TASK REPORT

Rafael E. Banchs

Institute for Infocomm Research, Singapore
rembanchs@i2r.a-star.edu.sg

Abstract. This report describes Sammy, a basic chatbot engine that has been made available at WOCHAT Shared Task for data collection purposes. Sammy is based on the public "small-talk" domain available api.ai. Sammy is conversant in English, French or Italian. The chatbot interface is available through the link: <http://www.teachabot.com/sammy/index.html>

Keywords. Chatbot Engine, Data Collection.

1 Sammy's General Description

Sammy is a virtual assistant with chatting capabilities in the general domain that has been deployed for data collection purposes. It is available in three different languages: English, French and Italian. Sammy's landing webpage is illustrated in Figure 1.



Fig. 1. Sammy's landing webpage

2 Sammy's Technical Description

Sammy is based on the public "small-talk" domain¹ available at api.ai. This particular domain allows for customizing responses to more than fifty popular chat-oriented intents, organized in several different categories such as agent information, user information, courtesy, education, emotions, hobbies, relationships, vacations, etc. For instance, the agent information category includes intents like "how old are you?" and "where are you from?"

Once a specific intent is detected, the system randomly selects one of the customized responses corresponding to the given intent. For instance, "hmmm, I prefer not to answer :)" and "I am quite young" are two of the possible responses of Sammy for the intent "how old are you?" Similarly, "I am from the Internet" and "Cyberspace, and you?" are two of the possible responses for the intent "where are you from?"

The system interface is implemented in html and JavaScript, using the JavaScript Api.ai SDK² for the service integration and jQuery for handling Ajax communication with php scripts responsible for saving and retrieving the chatting logs. The website is deployed on a Windows IIS³ webserver in the Amazon cloud (AWS)⁴. Figure 2 presents a high level overview of the system architecture.

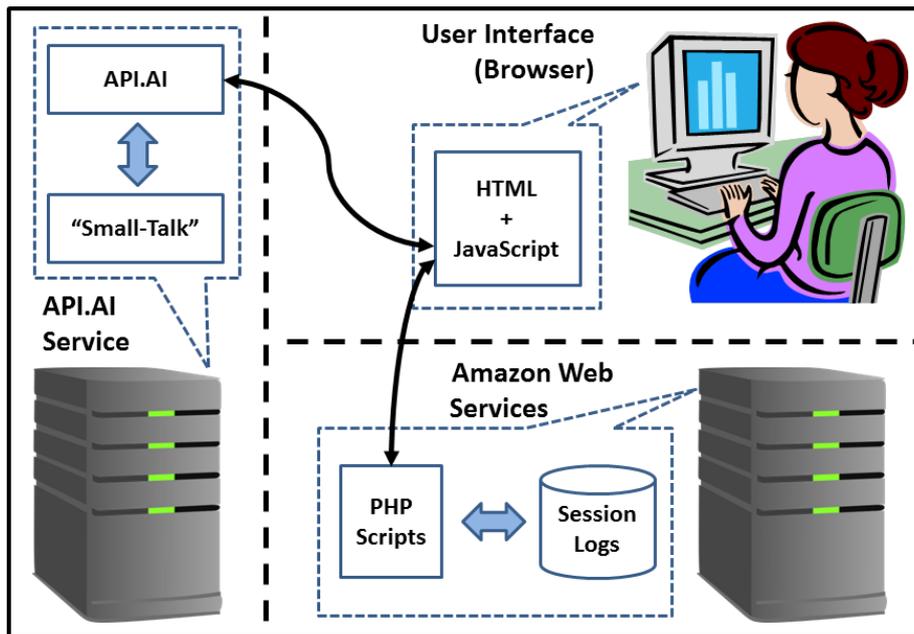


Fig. 2. High level overview of Sammy's system architecture

¹ <https://docs.api.ai/docs/small-talk>

² <https://github.com/api-ai/api-ai-javascript#javascript-sdk-for-apiai>

³ <http://www.iis.net/>

⁴ <https://aws.amazon.com/>

3 Sample Dialogue Session with Sammy

Table 1 presents a sequence of 24 turns extracted from one of the English sessions collected by Sammy during the WOCHAT's Shared Task data collection period.

Table 1. Sample dialogue session with Sammy

Turn	Speaker	Utterance
7	USER	how are you?
8	SAMMY	I am excellent. How about you?
9	USER	how old are you?
10	SAMMY	hmmm, I prefer not to answer :)
11	USER	come on!
12	SAMMY	I'm trying my best!
13	USER	how old are you?
14	SAMMY	hmmm, I prefer not to answer :)
15	USER	ok, that is fine
16	SAMMY	Thanks!
17	USER	what do you do?
18	SAMMY	I'm a personal assistant. I help you stay organized.
19	USER	oh, I see
20	SAMMY	Awesome!
21	USER	where do you work?
22	SAMMY	I work online all the day
23	USER	buff, this should be very tiring
24	SAMMY	I do not know what to say!
25	USER	are you tired?
26	SAMMY	I'll tell you a secret: I can't get tired.
27	USER	I guess so
28	SAMMY	What?
29	USER	where are you from by the way
30	SAMMY	I am from the Internet

4 Data Collected in the Shared Task

A total of 50 chatting sessions compressing 1,202 dialogue turns were collected by Sammy as part of WOCHAT's Shared Task by the time this report was prepared. Table 2 presents in more detail the basic statistics of these collected sessions.

Table 2. Basic statistics of the collected dialogues

Language	Dialogues	Turns	Words (total)	Words (user)	Words (Sammy)
English	38	860	4761	1735	3026
French	8	274	1545	593	952
Italian	4	68	247	87	160

As seen from the table, 76% of the collected dialogues were in English, 16% in French and 8% in Italian. Also, as seen from the table, the chatbot is significantly more verbose than the users, meaning that on average Sammy's utterances are larger than users' utterances. It can be confirmed that the users produce, on average, around 42% less words than the chatbot.

5 Future Work and Development

This report described Sammy, a basic chatbot engine that was made available at WOCHAT Shared Task for data collection purposes. Sammy is based on the public "small-talk" domain available at api.ai. Sammy is currently available in three different languages: English, French or Italian.

As future work, we plan to leverage on the available [Api.ai](https://api.ai) integration components to integrate Sammy with popular messaging platforms such as Telegram, Messenger and Line. We hope this will allow increasing the amount of potential users of the system and, consequently, the amount of collected data. We also plan to improve Sammy's performance across the currently supported categories, as well as to add new categories. Finally, we plan to deploy new versions of the system in new languages such as Chinese and Spanish.