Test Session

Team 5 - 06.03.2018

# What - tested areas

We are going to test following Amazon Lex intents:

1. Weather Intent
2. Robert’s Coffee Intent
3. Book Finding Intent
4. Help Intent

We could test our solution using Raspberry Pi or Laptop, but we decided to only use Raspberry because hardware/OS stuff was tested very well and we only care about client - AWS server interaction, user experience and learning about target audience - students.

# Why - goal and focus

High level goal of this testing session is to understand how successfully user, who has no prior knowledge about Libby, will interact with it. First of all, main goal of this sprint is refinement of existing intents for actual usage, so we want to hear unbiased opinion from other team on how our Libby-assistant is doing, which things aren’t obvious at all and how to enhance concrete dialogflows for different intents.

This testing session must produce log and statistic of:

* How well did user do without any guidance when trying to use any of the intents?
* How well did Libby recognize the wake word for different voices?
* How well did Libby recognize other intent-specific phrases?
* How satisfied user was with results from intent X (where X is any intent from the first section)? Rate usability from 1 to 5.
* Were replies fast enough? (Actual response times are too network dependent, but if any specific query takes longer than 3 seconds the actual query text must also be logged)
* Would you use it? Why?
* How to enhance the conversation flow?

Testing performed from viewpoint of a regular student.

During this test session we are addressing following product risks:

* If the software skips one of the core features, which users want to see
* If the software’s non-functional attributes (reply-time) are unacceptable
* If the software doesn’t enhance book-searching experience at all

# How - approach

Other team gets Libby and first tries to talk with it without supervision. If supervision is needed (important to log for future user manual), then they have assistance from one of the developers. So, basically, we are interested in a exploratory/blackbox testing.

We don’t want anybody to prepare testing data beforehand, because it would make the experiment senseless.

Tools: voice

# Exploration Log

## Data Files - Any Created Test Data Files

## Test Notes - What, how to reproduce test

## Bugs and Short Descriptions

## Issues, Observations, Feature Requests

**Observations**

**How well did user do without any guidance when trying to use any of the intents?**

User did not know when to talk, the ping-sound was not intuitive enough.

Limited information, e.g. when asking about weather only “-4 degrees” was answered. Weather help is, in its current state, (paraphrasing Tuomas) “fucking useless”.

Robert’s Coffee was unintuitive, the user did not know how he could ask for coffee’s. I.e. Libby only listed everything categories, to which the user did the same query again but with the drink name in. I.e. “How can I get coffee?” straight away.

The book intent is very limited. The user was not able to in several tries ask for books. (Apparently the book intent did not work during the session).

The user often tried to ask “Libby, what can you do?”

The user wants to talk more naturally, instead of having to say only words, e.g. instead of “Libby, help” he would like to say “Libby, what can you do?”

The user did not know what he could ask without help from us. He wanted to only ask one thing and then get help from that, instead of doing multiple queries in a specific order and manner.

**How well did Libby recognize the wake word for different voices?**

Initially Libby was recognizing the wakeword surprisingly well, there were more unanticipated false-positives.

**How well did Libby recognize other intent-specific phrases?**

As stated earlier, the user rephrased his queries often and in unanticipated manners, to which we need to be prepared to answer.

**How satisfied user was with results from intent X (where X is any intent from the first section)? Rate usability from 1 to 5.**

Find book did not work.

Help intent worked, was good to get to know different intents.

Robert’s coffee intent could have been more intuitive. Less specific manners and more ways to ask.

**Were replies fast enough? (Actual response times are too network dependent, but if any specific query takes longer than 3 seconds the actual query text must also be logged)**

We had to change to a laptop during the session, but replies were usually fast, but at a few times when Libby was processing, the user did not know what happened and started talking in the background, which resulted in unanticipated results. E.g. being in the Robert’s coffee intent, the user started asking for help from friends and suddenly Libby replied with weather.

**Would you use it? Why?**

**How to enhance the conversation flow?**