**Speech to text conversion**

The first requirement of the application is to understand speech and convert the speech to text. For example, a plumber (user) may say ask questions, e.g.

1. What is the closest store to Toorak?

2. What is the closest Plumbing store to me?

3. Call the Burwood branch

The first task for the application will be to recognize the voice and convert it to text, so that it can be processed by the application.

**Options**

There are three major options in the marketplace for speech to text recognition:

* Google Cloud Speech API (Google Cloud SDK)
  + https://cloud.google.com/speech-to-text/
* IBM Watson API (IBM Watson SDK)
  + https://www.ibm.com/watson/services/speech-to-text/
* Microsoft Speech Bing API (HTTP API)
  + https://azure.microsoft.com/en-gb/services/cognitive-services/speech/

Based on the research completed, the Google Cloud Speech API is the most error resistant, accurate and provides good response time, specifically:

* Highest percentage of exact text matches
* Lowest percentage of errors in word recognition
* Good average response time
* Good support for English (supports other languages if future extensions required, 120 other languages supported)
* Cloud Based
* Most expensive of the given solutions; however given the usage levels, may be free for the client, at least until the usage becomes popular and many plumbers are using the service



The Microsoft API scores in the middle and IBM Watson Speech to Text fares at the third place.

The recommendation is to use Google API for our project.

**Natural Language Classification**

Once the speech has been converted to text, the next task will be to build the context for the speech. For example, these are the sample questions provided by the client:

1. What is the closest store to Toorak?

2. What is the closest Plumbing store to me?

3. Call the Burwood branch

4. Who is the manager of Burwood?

5. Smart locations … if I’m in Vic, then Burwood means 3125 not 2134

6. What is the address of the Burwood outlet?

7. Is Burwood open right now?

8. When will the Burwood store open?

9. What are the trading hours for Burwood?

In these questions, there are two components to be extracted from each one:

* Store name

Question Context (what is being requested)

* + Location of the closest store to another location
  + Call a particular store
  + Location of the closest store to the current location
  + Find manager details
  + Find opening hours
    - Opening hours
    - Store open right now or not

This is called Natural Language Classification. In other words, given the natural language spoken by someone, the application will need to build the context (i.e. understand the question which is being asked).

**IBM Watson Natural Language Classification Service**

The IBM Watson Natural Language Classification can be used to build context from natural language.

This is the wesite

<https://github.com/watson-developer-cloud/natural-language-classifier-nodejs/blob/master/training/weather_data_train.csv>

There is a demonstration here:

<https://github.com/watson-developer-cloud/natural-language-classifier-nodejs/blob/master/training/weather_data_train.csv>

The IBM Watson’s Natural Language Classifier is a cloud based API which works in the following way:

1. A range of training questions are developed along with what context they represent, e.g. in our application, the questions will be developed in the following way:

|  |  |
| --- | --- |
| **Question** | **Context** |
| What is the closest suburb to Burwood? | <Closest Store> <Burwood> |
| Call the Chadstone store manager. | <Call Manager> <Chadstone> |
| Phone Chadstone store. | <Call Manager> <Chadstone> |
| Dial Chadstone store manager. | <Call Manager> <Chadstone> |
| Direct me to Chadstone store. | <Directions> <Chadstone> |
| Drive to Chadstone. | <Directions> <Chadstone> |

***Request example***

import json

from watson\_developer\_cloud import NaturalLanguageClassifierV1

natural\_language\_classifier = NaturalLanguageClassifierV1(

username='YOUR SERVICE USERNAME',

password='YOUR SERVICE PASSWORD')

classes = natural\_language\_classifier.classify('10D41B-nlc-1', 'Drive me to Chadstone store')

print(json.dumps(classes, indent=2))

***Response example***

{

"classifier\_id" : "10D41B-nlc-1",

"url" : "https://gateway.watsonplatform.net/natural-language-classifier/api/v1/classifiers/10D41B-nlc-1/classify?text=How%20hot%20wil/10D41B-nlc-1",

"text" : "Drive me to Chadstone store",

"top\_class" : "directions<chadstone>",

"classes" : [ {

"class\_name" : "directions",

"confidence" : 0.9998201258549781

}, {

"class\_name" : "call store",

"confidence" : 1.7987414502176904E-4

} ]

}

**User interface proposal**

