TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA CÔNG NGHỆ THÔNG TIN
BỘ MÔN CÔNG NGHỆ TRI THỨC

BÙI TRUNG HẢI – PHẠM NGỌC TUẤN

XÂY DỰNG ỨNG DỤNG TRỢ LÝ ẢO CHO MÁY TÍNH SỬ DỤNG GOOGLE SPEECH API

KHÓA LUẬN TỐT NGHIỆP CỬ NHÂN CNTT

TP. HCM, 2017

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA CÔNG NGHỆ THÔNG TIN
BỘ MÔN CÔNG NGHỆ TRI THỨC

BÙI TRUNG HẢI – 1312165
PHẠM NGỌC TUẤN – 1312669

XÂY DỰNG ỨNG DỤNG TRỢ LÝ ẢO CHO MÁY TÍNH SỬ DỤNG GOOGLE SPEECH API

KHÓA LUẬN TỐT NGHIỆP CỬ NHÂN CNTT

GIÁO VIÊN HƯỚNG DẪN

TS. NGÔ MINH NHỰT

KHÓA 2013 - 2017

NHẬN XÉT CỦA GIÁO VIÊN HƯỚNG DẪN

....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....

TpHCM, ngày ….. tháng …… năm ……

Giáo viên hướng dẫn

[Ký tên và ghi rõ họ tên]

NHẬN XÉT CỦA GIÁO VIÊN PHẢN BIỆN

....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....
....

Khóa luận đáp ứng yêu cầu của Khóa luận cử nhân CNTT.

TpHCM, ngày ….. tháng …… năm ……

Giáo viên phản biện

[Ký tên và ghi rõ họ tên]

LỜI CẢM ƠN

Với lòng biết ơn sâu sắc, trước hết chúng em xin chân thành cảm ơn quý Thầy Cô khoa Công Nghệ Thông Tin - trường đại học Khoa Học Tự Nhiên, những người đã ân cần giảng dạy, xây dựng cho em một nền tảng kiến thức vững chắc để chúng em có thể thực hiện khóa luận này.

Đặc biệt, chúng em xin gửi lời tri ân sâu sắc đến Thầy Ngô Minh Nhựt. Thầy đã rất tận tâm, nhiệt tình hướng dẫn và chỉ bảo chúng em trong suốt quá trình thực hiện luận văn. Nếu không có sự giúp đỡ tận tình của thầy, chúng em chắc chắn không thể hoàn thành luận văn.

Cuối cùng, chúng con xin cảm ơn ba mẹ đã sinh thành, nuôi dưỡng, và dạy dỗ để chúng con có được thành quả như ngày hôm nay. Ba mẹ luôn là nguồn động viên, nguồn sức mạnh hết sức lớn lao mỗi khi chúng con gặp khó khăn trong cuộc sống.

Để hoàn thành luận văn này là tất cả những cố gắng, nỗ lực của chúng em. Tuy nhiên, sẽ không thể tránh khỏi những thiếu sót, kính mong nhận được sự cảm thông và giúp đỡ của quý Thầy Cô và các bạn.

TP. Hồ Chí Minh, 7/2017

Bùi Trung Hải

Phạm Ngọc Tuấn

ĐỀ CƯƠNG CHI TIẾT

Tên Đề Tài: Xây dựng ứng dụng trợ lý ảo cho máy tính sử dụng Google Speech API
Giáo viên hướng dẫn: ThS. Ngô Minh Nhựt
Thời gian thực hiện: từ ngày 15/12/2016 đến ngày 15/07/2017
Sinh viên thực hiện: Bùi Trung Hải – 1312165, Phạm Ngọc Tuấn - 1312669
Loại đề tài: Phát triển hệ thống, nghiên cứu thuật toán

Nội Dung Đề Tài: Tìm hiểu các phương pháp xử lý và tương tác với tín hiệu âm thanh, tiếng nói. Nghiên cứu hệ thống Natural Language Understanding đơn giản. Ứng dụng vào xây dựng ứng dụng trợ lý ảo cho máy tính.

Nội dung chi tiết của đề tài bao gồm:

- Xây dựng ứng dụng trợ lí ảo tương tác bằng giọng nói tiếng Anh cho máy tính
- Các vấn đề quan tâm:
    - Xử lý tín hiệu số (âm thanh, tiếng nói), hệ thống chuyển đổi tiếng nói thành văn bản.
    - Tương tác audio I/O, hệ thống chuyển đổi văn bản thành tiếng nói.
    - Giao thức truyền nhận dữ liệu REST API,…
    - Hệ thống Natural Language Understanding hiệu quả để xác định ý muốn của người dùng
    - Hệ thống chạy đa nhiệm
    - Tối ưu hóa hệ thống

- Các thành phần cơ bản của hệ thống:
    - Module thu âm từ microphone
    - Module nhận dạng từ khóa wake up
    - Module chuyển đổi tiếng nói thành văn bản
    - Module chuyển đổi văn bản thành hành động.
    - Module chuyển đổi văn bản thành tiếng nói
- Ứng dụng thử nghiệm sẽ hỗ trợ các tính năng:
    - Thông báo giờ hiện tại
    - Dự báo thời tiết trong ngày
    - Phát nhạc
    - Trả lời các câu hỏi Wh-question
    - Trả lời các thông tin cơ bản của ứng dụng: tên, tuổi,…

Kế Hoạch Thực Hiện:

- 15/12/2016 – 14/01/2017: Khảo sát, tìm hiểu về các thư viện python phục vụ cho việc tương tác và xử lý tín hiệu âm thanh.

- 15/01/2017 – 14/02/2017: Thiết kế các thành phần của hệ thống.

- 15/02/2017 – 14/03/2017: Cài đặt và thử nghiệm các thành phần: thu âm, nhận dạng từ khóa wake up, chuyển giọng nói thành văn bản, chuyển văn bản thành giọng nói.

- 15/03/2017 – 14/04/2017: Tìm hiểu và xây dựng hệ thống Natural Language Understanding.

- 15/04/2017 – 14/05/2017: Cài đặt các chức năng mà ứng dụng hỗ trợ.

- 15/05/2017 – 31/05/2017: Ráp nối tất cả các thành phần, tiến hành thử nghiện và hoàn thiện hệ thống.

- 01/06/2017 – 28/06/2017: Viết và hoàn thiện luận văn.

Xác nhận của GVHD	**Ngày 08 tháng 07 năm 2013** **SV Thực hiện**

MỤC LỤC

DANH MỤC HÌNH ẢNH

DANH MỤC BẢNG

TÓM TẮT KHÓA LUẬN

Trong xu hướng công nghệ hiện nay, vai trò của các trợ lý ảo ngày càng trở nên quan trọng. Các hãng công nghệ lớn thay nhau tung ra những trợ lý ảo của riêng mình tích hợp trên các thiết bị di động: Siri của Apple, Cortana của Microsoft, Google Assistant của Google, Alexa của Amazon,... Chức năng của các trợ lý ảo này ngày càng được mở rộng, từ những chức năng đơn giản như tra cứu, hỏi đáp, đến những chức năng cao hơn như quản lý lịch, gọi điện thoại, dẫn đường, điều khiển các thiết bị khác,... Khóa luận này có mục đích tạo ra một trợ lý ảo có khả năng chạy được trên nhiều nền tảng hệ điều hành khác nhau trên máy tính cá nhân.

Nhận diện giọng nói là một trong những thành phần quan trọng nhất của một trợ lý ảo. Nhiều công ty và nhóm nghiên cứu lớn nhỏ đã nghiên cứu và đưa ra các bộ toolkit cũng như API cho việc nhận diện giọng nói, trong đó một trong những API có chất lượng được đánh giá tốt nhất là Google Speech API của gã khổng lồ công nghệ Google. Do đó, chúng tôi muốn tận dụng chất lượng của Google Speech API để tạo nên một trợ lý ảo có độ chính xác cao về nhận diện giọng nói.

Kết quả sơ bộ mà khóa luận đạt được là tạo ra một trợ lý ảo có thể chạy trên các hệ điều hành phổ biến trên máy tính cá nhân như Windows, Linux, Mac. Trợ lý ảo có những chức năng cơ bản của một trợ lý ảo như hỏi đáp, tra cứu thông tin, trả lời các câu hỏi về thời gian, thời tiết, ngoài ra còn có thể phát nhạc theo yêu cầu và chào hỏi ở mức độ đơn giản.

Chương 1

Mở Đầu

Nội dung của chương 1 giới thiệu tổng quan về đề tài, nêu ra mục tiêu của khóa luận, và cấu trúc nội dung của luận văn.

1.1 Tổng quan về đề tài

Trợ lý ảo là một phần mềm trên máy tính hoặc thiết bị di động có khả năng hỗ trợ người dùng thực hiện nhiều loại công việc, nhận lệnh từ người dùng dưới dạng ngôn ngữ tự nhiên, thường là giọng nói. Nhờ khả năng nhận lệnh và phản hồi qua giọng nói, người dùng có thể ra lệnh cho trợ lý ảo mà không cần phải thao tác bằng tay trên thiết bị.

Trong xu hướng công nghệ ngày càng tiên tiến, việc sở hữu một trợ lý ảo sẽ giúp cho người dùng có những trải nghiệm mới mẻ và thú vị hơn khi sử dụng các thiết bị công nghệ nhờ vào sự tiện dụng, mạnh mẽ với nhiều chức năng đa dạng, cũng như tính tự nhiên trong giao tiếp giữa người và máy. Khi sử dụng các trợ lý ảo tiên tiến nhất hiện nay, người dùng sẽ có cảm giác được giao tiếp với một người trợ lý thực sự chứ không phải là một cái máy. Số lượng chức năng của các trợ lý ảo ngày càng tăng, từ những chức năng cơ bản như hỏi đáp, tra cứu, tìm kiếm thông tin, đến những chức năng nâng cao hơn như quản lý lịch, quản lý email, thực hiện cuộc gọi, gửi tin nhắn, điều khiển các thiết bị trong nhà, và thậm chí là đặt chỗ nhà hàng!

...

1.2 Mục tiêu của khóa luận

...

1.3 Nội dung luận văn

...

Chương 2

Tín hiệu âm thanh, tiếng nói. Thư viện PyAudio

2.1 Tổng quan về âm thanh, tiếng nói

Âm thanh đóng vai trò quan trọng trọng cuộc sống con người
Âm thanh có thể được cảm nhận bởi con người thông qua thính giác
Một trong những dạng của âm thanh là tiếng nói
Tiếng nói đóng vai trò quan trọng trong hoạt động giao tiếp
Giao tiếp bằng tiếng nói là hoạt động giao tiếp nhanh, phổ biến, tiện lợi nhất
Âm thanh được nghiên cứu và ứng dụng trong nhiều lĩnh vực

2.2 Các khái niệm cơ bản của âm thanh, tiếng nói

wiki: https://en.wikipedia.org/wiki/Sound

2.3 Cách lưu trữ âm thanh trong máy tính

Âm thanh trong tự nhiên có dạng liên tục
Máy tính lưu trữ dạng rời rạc -> cần lấy mẫu âm thanh theo 1 tần số.

2.3.1 Các thông số của âm thanh khi lưu trữ trên máy tính

sample rate

bitdepth

chanel ...

2.3.2 Lưu trữ không nén

file wav

2.3.3 Lưu trữ nén

file mp3

2.4 Ứng dụng

Âm thanh là công cụ tương tác giữa người sử dụng và ứng dụng.

Người sử dụng sử dụng tiếng nói để ra lệnh

Ứng dụng dùng tiếng nói để phản hồi

2.5 Thư viện PyAudio

2.5.1 Tổng quan

Là thư viện viết bằng Python hỗ trợ tất cả các Hđh

Hỗ trợ người dùng tương tác với âm thanh trên máy tính dễ dàng.

https://people.csail.mit.edu/hubert/pyaudio/docs/

2.5.2 Chức năng

Thu âm từ microphone của máy tính dưới dạng dữ liệu thô

Phát âm thanh ra loa của máy tính từ dữ liệu thô

2.5.3 Cài đặt

pip
build từ source

2.5.4 Cách sử dụng

2 cách sử dụng: blocking, non-blocking
các thông số của ứng dụng: samplerate, bitdeptg, chanel, ...

2.5.5 Các ưu, khuyết điểm

Ưu: hỗ trợ nhiều hđh, cài đặt đơn giản, nhẹ, dễ sử dụng. Nhược: ít tính năng, chưa hỗ trợ phát từ file mp3, chưa hỗ trợ thu từ nhiều micro

2.5.6 Ứng dụng

PyAudio được sử dụng trong module Microphone của ứng dụng. Giúp thu âm và chuyển cho các module khác để xử lý.

Chương 3

Speech to text

3.1 Tổng quan

3.2 Mô hình hoạt động

3.3 Ứng dụng

3.4 Các vấn đề cần giải quyết

3.4.1 Dò tìm keyword

Yêu cầu

Hoạt động liên tục -> nên hoạt động offline Xử lý nhanh từng frame của âm thanh Độ chính xác khá cao.

Giải pháp

sử dụng thư viên pocketsphinx

3.4.2 Chuyển đổi lệnh người dùng thành văn bản

Yêu cầu

Chỉ hoạt động khi người dùng ra lệnh Độ chính xác rất cao.

Giải pháp

sử dụng thư viên google speech to text

3.5 Thư viện pocketsphinx

3.5.1 Tổng quan

3.5.2 Chức năng

3.5.3 Cách cài đặt

3.5.4 Cách sử dụng

các thông số của thư viện

3.5.5 Ưu, nhược điểm

Ưu: hỗ trợ offline, dò keyword Nhược: độ chính xác kém

3.5.6 Ứng dụng

Module wake up, giúp kích hoạt hệ thống khi người dùng gọi wakup word

3.6 Thư viện Google Speech To Text

3.6.1 Tổng quan

3.6.2 Chức năng

3.6.3 Cách cài đặt

3.6.4 Cách sử dụng

Các thông số của thư viện

3.6.5 Ưu, nhược điểm

Ưu: độ chính xác cao Nhược: yêu cầu internet, phải gửi toàn bộ file âm thanh 1 lúc, không stream được

3.6.6 Ứng dụng

Module TTS, giúp chuyển lệnh người dùng thành văn bản.

Chương 4

Text To Speech

4.1 Tổng quan

4.2 Mô hình hoạt động

4.3 Ứng dụng

4.4 googleTTS

4.4.1 Tổng quan

4.4.2 Chức năng

4.4.3 Cách cài đặt

4.4.4 Cách sử dụng

4.4.5 Ưu, nhược điểm

4.4.6 Ứng dụng

4.5 iSpeech

4.5.1 Tổng quan

4.5.2 Chức năng

4.5.3 Cách cài đặt

4.5.4 Cách sử dụng

Chương 5

Phân loại ý định

5.1 Tổng quan

5.2 Mô hình hoạt động

5.3 Ứng dụng

5.4 Thư viện Rasa NLU

5.4.1 Tổng quan

5.4.2 Chức năng

5.4.3 Mô hình hoạt động

5.4.4 Cách cài đặt

5.4.5 Cách sử dụng

5.4.6 Chuẩn bị dữ liệu

5.4.7 Đánh giá model

5.4.8 Ứng dụng

Chương 6

Ứng dụng Alexa

6.1 Tổng quan

6.2 Mô hình hoạt động

6.2.1 Các module chính

Microphone

Chức năng: Các vấn đề và cách giải quyết:

Recorder

Wakeup

Text To Speech

Speech to Text

Intent Classification

Intent Processor

6.2.2 Luồn hoạt động giữa các module

6.3 Các chức năng chính

Thông báo giờ

Chức năng chi tiết: Cách thức hoạt động:

Thông báo thời tiết

Phát nhạc

Giao tiếp cơ bản

Trả lời câu hỏi Wh-question

Chương 7

Kết Luận và Hướng Phát Triển

7.1 Kết quả đạt được

7.1.1 Về mặt lý thuyết

7.1.2 Về mặt thực nghiệm

7.2 Hướng phát triển

Phụ Lục: Các Công Trình Đã Công Bố

Hội nghị quốc tế:

- **K. Tran** and B. Le, "Demystifying Sparse Rectified Auto-Encoders," in *Proceedings of the Fourth Symposium on Information and Communication Technology*, ser. SoICT'13. New York, NY, USA: ACM, 2013, pp. 101–107. [Online]. Available: http://doi.acm.org/10.1145/2542050.2542065