**HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG**

**KHOA CÔNG NGHỆ THÔNG TIN 1**

**\_\_\_o0o\_\_\_**

**ĐỒ ÁN**

**TỐT NGHIỆP ĐẠI HỌC**

***Đề tài***:

**XÂY DỰNG ỨNG DỤNG HỌC TIẾNG ANH**

**TRÊN NỀN TẢNG IOS**

|  |  |
| --- | --- |
| **Giảng viên hướng dẫn:** | **ThS. TRỊNH THỊ VÂN ANH** |
| **Sinh viên thực hiện:** | **ĐỖ TIẾN HƯNG** |
| **Lớp:** | **E14CN** |
| **Khoá:** | **2014 - 2019** |
| **Hệ:** | **ĐẠI HỌC CHÍNH QUY** |

**Hà Nội, Tháng 12/2018**

**LỜI CẢM ƠN**

Để có được kết quả học tập đến ngày hôm nay, em xin chân thành gửi đến các thầy cô giáo Học Viện Công Nghệ Bưu Chính Viễn Thông, đặc biệt là quý thầy cô trong khoa Công nghệ thông tin lời cảm ơn chân thành nhất. Các thầy cô luôn luôn tận tình truyền đạt cho em những kiến thức về bộ môn cũng như kinh nghiệm thực tế trong suốt thời gian học tập tại trường.

Em xin chân thành cảm ơn cô Trịnh Thị Vân Anh đã tận tình động viên, giúp đỡ em trong suốt thời gian làm đồ án tốt nghiệp. Được cô hướng dẫn dìu dắt là sự may mắn và vinh dự to lớn của em. Một lần nữa em xin chân thành cảm ơn cô.

Em xin gửi lời chúc tốt đẹp nhất tới các thầy cô và các bạn tham gia đợt bảo vệ tốt nghiệp của khóa này, chúc mọi người luôn mạnh khỏe, thành công trong cuộc sống và đặc biệt bảo vệ tốt nghiệp thành công.

Với vốn kiến thức và kinh nghiệm thực tiễn còn ít nên đồ án không tránh khỏi những thiếu sót, em rất mong nhận được những ý kiến đóng góp của các thầy các cô và các bạn để đồ án được hoàn thiện hơn.

Em xin chân thành cảm ơn!

Hà Nội, ngày tháng 12 năm 2018

Đỗ Tiến Hưng

**NHẬN XÉT, ĐÁNH GIÁ, CHO ĐIỂM**

**(Của giảng viên hướng dẫn)**

…………………………………………………………………………………………..………..………..

…………………………………………………………………………………………..………..………..

…………………………………………………………………………………………..………..………..

…………………………………………………………………………………………..………..………..

…………………………………………………………………………………………..………..………..

…………………………………………………………………………………………..………..………..

…………………………………………………………………………………………..………..………..

…………………………………………………………………………………………..………..………..

…………………………………………………………………………………………..………..………..

…………………………………………………………………………………………..………..………..

…………………………………………………………………………………………..………..………..

…………………………………………………………………………………………..………..………..

…………………………………………………………………………………………..………..………..

…………………………………………………………………………………………..………..………..

…………………………………………………………………………………………..………..………..

…………………………………………………………………………………………..………..………..

…………………………………………………………………………………………..………..………..

…………………………………………………………………………………………..………..………..

…………………………………………………………………………………………..………..………..

…………………………………………………………………………………………..………..………..

…………………………………………………………………………………………..………..………..

**Điểm:** …………………….……… (bằng chữ: …………………………………)

**Đồng ý/Không đồng ý** cho sinh viên bảo vệ trước hội đồng chấm đồ án tốt nghiệp**?**

*Hà Nội, ngày tháng năm 2018*

**CÁN BỘ - GIẢNG VIÊN HƯỚNG DẪN**

**NHẬN XÉT, ĐÁNH GIÁ, CHO ĐIỂM**

**(Của giảng viên phản biện)**

…………………………………………………………………………………………..………..………..

…………………………………………………………………………………………..………..………..

…………………………………………………………………………………………..………..………..

…………………………………………………………………………………………..………..………..

…………………………………………………………………………………………..………..………..

…………………………………………………………………………………………..………..………..

…………………………………………………………………………………………..………..………..

…………………………………………………………………………………………..………..………..

…………………………………………………………………………………………..………..………..

…………………………………………………………………………………………..………..………..

…………………………………………………………………………………………..………..………..

…………………………………………………………………………………………..………..………..

…………………………………………………………………………………………..………..………..

…………………………………………………………………………………………..………..………..

…………………………………………………………………………………………..………..………..

…………………………………………………………………………………………..………..………..

…………………………………………………………………………………………..………..………..

…………………………………………………………………………………………..………..………..

…………………………………………………………………………………………..………..………..

…………………………………………………………………………………………..………..………..

…………………………………………………………………………………………..………..………..

**Điểm:** …………………….……… (bằng chữ: …………………………………)

**Đồng ý/Không đồng ý** cho sinh viên bảo vệ trước hội đồng chấm đồ án tốt nghiệp**?**

*Hà Nội, ngày tháng năm 2018*

**CÁN BỘ- GIẢNG VIÊN PHẢN BIỆN**

Table of Contents

[CHƯƠNG 1: IOS VÀ MỘT SỐ VẤN ĐỀ CƠ BẢN 6](#_Toc529517269)

[1. GIỚI THIỆU VỀ HỆ ĐIỀU HÀNH IOS 6](#_Toc529517270)

[1.1. TỔNG QUAN 6](#_Toc529517271)

[1.2. LỊCH SỬ PHÁT TRIỂN 6](#_Toc529517272)

[1.3. ƯU ĐIỂM CỦA IOS 7](#_Toc529517273)

[1.4. NHƯỢC ĐIỂM CỦA IOS 8](#_Toc529517274)

[1.5. PHẦN MỀM VÀ NGÔN NGỮ LẬP TRÌNH 8](#_Toc529517275)

[2. NGÔN NGỮ LẬP TRÌNH SWIFT 8](#_Toc529517276)

[2.1. TỔNG QUAN 8](#_Toc529517277)

[2.2. ƯU ĐIỂM SO VỚI OBJECTIVE-C 9](#_Toc529517278)

[2.3. NHƯỢC ĐIỂM SO VỚI OBJECTIVE-C 10](#_Toc529517279)

[3. KIẾN TRÚC HỆ ĐIỀU HÀNH IOS 10](#_Toc529517280)

[CHƯƠNG 2: MÔ TẢ VỀ ỨNG DỤNG 11](#_Toc529517281)

[1. HIỆN TRẠNG HỌC TIẾNG ANH CỦA TRẺ EM TẠI VIỆT NAM 11](#_Toc529517282)

[2. ĐỐI TƯỢNG VÀ MỤC TIÊU CỦA ỨNG DỤNG 13](#_Toc529517283)

[3. ÁP DỤNG XỬ LÝ NGÔN NGỮ TỰ NHIÊN TRONG ỨNG DỤNG 13](#_Toc529517284)

[3.1. NHỮNG ĐIỀU ĐẶC BIỆT CỦA NGÔN NGỮ 15](#_Toc529517285)

[3.2. CÁC BƯỚC THỰC HIỆN 16](#_Toc529517286)

# **CHƯƠNG 1: IOS VÀ MỘT SỐ VẤN ĐỀ CƠ BẢN**

1. GIỚI THIỆU VỀ HỆ ĐIỀU HÀNH IOS
   1. TỔNG QUAN

iOS (trước đây là iPhone OS) là hệ điều hành trên các thiết bị di động của Apple. Đây là hệ điều hành chạy trên các sản phẩm iPhone, iPad, và iPod Touch và là hệ điều hành phổ biến thứ 2 trên toàn cầu, sau Android.

Ban đầu hệ điều hành này chỉ được phát triển để chạy trên iPhone (gọi là iPhone OS), nhưng sau đó nó đã được mở rộng để chạy trên các thiết bị của Apple như iPod Touch (tháng 9 năm 2007) và iPad (tháng 1 năm 2010). Tính đến tháng 1 năm 2017, App Store của Apple chứa khoảng 2.2 triệu ứng dụng iOS, 1 triệu trong số đó là ứng dụng chỉ cho iPad và được tải về tổng cộng khoảng 130 tỷ lần.

Giao diện người dùng của iOS dựa trên cơ sở thao tác bằng tay. Người dùng có thể tương tác với hệ điều hành này thông qua rất nhiều động tác bằng tay trên màn hình cảm ứng của các thiết bị của Apple.  Phiên bản mới nhất của iOS là 12 được ra mắt chính thức vào ngày 4 tháng 6 năm 2018, dành cho các thiết bị từ iPhone 5S trở lên, iPod touch (thế hệ 6), iPad (2017) trở lên, iPad Mini (thế hệ thứ hai) trở lên và iPad Pro.

* 1. LỊCH SỬ PHÁT TRIỂN
* iPhone OS 1

Năm 2007, Steve Jobs đứng trên sân khấu giới thiệu iPhone và iPhone OS cùng nhau. Trong buổi lễ ra mắt, Jobs gọi hệ điều hành này là “OS X” bởi vì nó chia sẻ cùng một nhận Unix với phiên bản dành cho máy tính của Apple. Một năm sau khi Apple công bố iPhone SDK để các nhà phát triển có thể viết app thì công ty mới nhắc đến chữ “iPhone OS”.  Chiếc iPhone đầu tiên là một trong những thiết bị quan trọng nhất trong lịch sử công nghệ vì nó đã thay đổi cách mà smartphone được định hình, thay đổi cách giao tiếp giữa người dùng với điện thoại và thiết bị di động nói chung, cũng như mở đầu cho thời kì bùng nổ của mobile. Kiểu thiết kế của iPhone cũng như nhiều thành phần trong iPhone OS đã đặt nền tảng cho sự phát triển của giao diện người dùng hiện đại – một giao diện thân thiện hơn với ngón tay, đẹp hơn và dễ dùng hơn.

* iPhone OS 2, iPhone OS 3, iOS 4, iOS 5, iOS 6 lần lượt ra đời đem lại các tính năng hữu ích cho người dùng như: App Store, MMS, Spotlight, FaceTime, Notification Center, iMessage, Siri, Apple Maps, …
* iOS 7 là một bản cập nhật lớn khi thay đổi hoàn toàn giao diện của điện thoại iPhone từ skeuomorphism sang thiết kế phẳng.

iOS 7 giống như một làn gió mới thổi vào người dùng iPhone, iPad và iPod Touch bởi vì họ được trải nghiệm một thứ hoàn toàn mới và khác lạ so với những gì họ từng thấy. Ngoài việc thiết kế lại icon, Apple cũng đưa những yếu tố trong mờ vào giao diện của iOS, bổ sung Control Center để truy cập nhanh các kết nối thường dùng và có cả AirDrop để gửi file cho nhau (trước đó AirDrop chỉ dành cho Mac). iOS 7 cũng có app Photos mới, iTunes Radio và cách sắp xếp mới cho giao diện đa nhiệm.

* iOS 8, iOS 9, iOS 10 cũng lần lượt đến tay người sử dụng kèm theo những ứng dụng mới như TestFlight, Apple Music, 3D Touch, …
  1. ƯU ĐIỂM CỦA IOS
* iOS là một hệ điều hành với độ ổn định cao, mượt mà qua thời gian dài.
* Là hệ điều hành có độ bảo mật và tin cậy cao nhất hiện nay
* Hệ điều hành được cập nhật đều đặn hàng năm, bạn được quyền lựa chọn update hoặc không. Ngoài ra khi cập nhật iOS bản mới, tất cả các thiết bị iOS đều được phép cập nhật lên bản cao nhất và không bị giới hạn.
* Kho ứng dụng phong phú và được kiểm duyệt nghiêm ngặt, không có ứng dụng lừa đảo, ứng dụng rác.
* Đơn giản, dễ sử dụng cho người mới
* Có khả năng đồng bộ tốt giữa các thiết bị iOS với nhau
  1. NHƯỢC ĐIỂM CỦA IOS
* Không thể chạy nhiều ứng dụng cùng lúc trên màn hình.
* Trao đổi dữ liệu đều phải qua iTunes gây bất tiện và tốn thời gian vì đây là con đường duy nhất.
* Khả năng tuỳ chỉnh hạn chế.
* iOS chỉ có thể đi kèm với iPhone, trong khi giá các thiết bị Apple thì khá cao
  1. PHẦN MỀM VÀ NGÔN NGỮ LẬP TRÌNH

Các phần mềm (ứng dụng, app) trên iOS thông thường được người dùng tải về từ App Store - kho ứng dụng của Apple tuy vẫn còn vài cách khác không chính thống để cài đặt phần mềm. Các phần mềm này được các nhà phát triển (developer) phát triển bằng các ngôn ngữ lập trình Objective-C và sau này được thay thế bằng Swift.

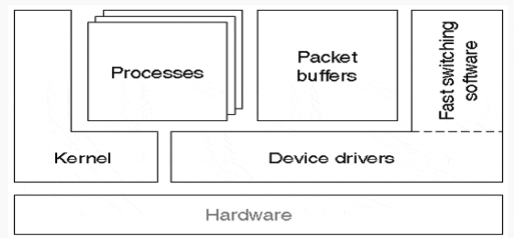
1. NGÔN NGỮ LẬP TRÌNH SWIFT
   1. TỔNG QUAN

Swift là một ngôn ngữ lập trình hướng đối tượng được dùng để phát triển ứng dụng trên iOS và OS X. Trước đây Objective C là ngôn ngữ chính của Apple, việc ra đời Swift có thể giúp các lập trình viên dễ dàng tiếp cận hơn, đặc biệt nó là mã nguồn mở (các tài nguyên về Swift ta có thể tải về từ trang <https://swift.org/)>. Swift được thiết kế là một ngôn ngữ có khả năng phòng chống lỗi cao. Nó được biên dịch với trình biên dịch LLVM. Apple vẫn chủ ý cho chạy song song 2 ngôn ngữ lập trình Objective C và Swift.

Ngày 2 tháng 6 năm 2014, tại hội thảo WWDC (Apple Worldwide Developers Conference), phiên bản Swift đầu tiên đã được công bố.  Vào tháng 3 năm 2017, chưa đầy 3 năm sau khi chính thức ra mắt, Swift đã đứng đầu trong bảng xếp hạng TIOBE hàng tháng về các ngôn ngữ lập trình phổ biến nhất.

* 1. ƯU ĐIỂM SO VỚI OBJECTIVE-C
* Việc code của các lập trình viên iOS giờ đây nhanh và đơn giản hơn rất nhiều do khả năng phát triển các ứng dụng và kiểm tra thành quả của mình theo thời gian thực. Ngoài ra, việc không cần chạy các chương trình Compliler và các chương trình test quá thường xuyên sẽ giúp nhà phát triển tiết kiệm phần cứng. Một điểm mạnh tuyệt với nữa cũng liên quan đến thời gian là các lập trình viên có thể dành nhiều thời gian để thực hiện các bài test ở mức độ tích hợp cao một cách kỹ càng thay vì phải tập trung kiểm tra hay thử nghiệm các tính năng nhỏ lẻ như trước đây.
* Code ngắn gọn, dễ đọc. Theo như Apple, 3 dòng code của Objective-C có thể được gói gọn trong 1 dòng code của Swift, vì vậy, việc viết ứng dụng sẽ tiết kiệm thời gian hơn, lỗi xuất hiện ít hơn và tăng hiệu quả cho quá trình bảo trì hay dễ dàng vá lỗi Swift trong tương lai.
* Giống như một phiên bản đơn giản của Objective-C, Swift có các lớp đối tượng được tối ưu nhất với các dòng mã được sắp xếp một cách hợp lý, ở đó, các nhà phả triển sẽ có một môi trường Runtime nên rất quen thuộc và dễ dàng quản lý.
* Tránh sự cố tràn bộ nhớ vì Xcode Runtime được thực hiện trong thời gian thực, tức là khi lập trình viên viết một dòng mã và nhận kết quả trong thời gian thực, hình ảnh có thể được xem trong IDE, ngoài ra Xcode cũng sẽ dựng các kịch bản thực hiện trong các ứng dụng nên lập trình viên có thể quản lý ứng dụng tốt hơn.
* Dựa trên nền tảng Objective-C, vì vậy các ứng dụng dùng Objective-C có thể được cập nhật bằng Swift, tức là các công cụ hữu ích của Objective-C có thể được sử dụng ngay trên Swift.
* Riêng đối với các ứng dụng Game, người dùng sẽ có những trải nghiệm ấn tượng hơn do các ứng dụng sử dụng Swift sẽ có sức ép phần cứng thấp hơn là những ứng dụng sử dụng Objective-C.
  1. NHƯỢC ĐIỂM SO VỚI OBJECTIVE-C
* Các nhà ứng dụng sẽ phải học lại một ngôn ngữ lập trình hoàn toàn mới, sẽ mất một khoảng thời gian để làm chủ và thích nghi với sự thay đổi này
* Đang còn non trẻ và vẫn đang trong giai đoạn kiểm nghiệm/ đánh giá nên cần phải có thời gian để ngôn ngữ lập trình này đánh bật được Ojective-C với tuổi đời lên đến 30 năm.

1. KIẾN TRÚC HỆ ĐIỀU HÀNH IOS



* Các tiến trình (Processes): Là những tuyến riêng lẻ kết hợp với dữ liệu để thực hiện những tác vụ, như duy trì hệ thống, chuyển mạch gói dữ liệu, thực hiện giao thức định tuyến…
* Nhân (Kernel): Cung cấp những dịch vụ cơ bản của hệ thống tùy thuộc vào iOS như: quản lý bộ nhớ, lập lịch các tiến trình…Nó cung cấp quản lý tài nguyên phần cứng (CPU, bộ nhớ) cho các tiến trình.
* Bộ đệm gói (Packet buffer): Cung cấp các bộ đệm toàn cục và kết hợp với chức năng quản lý bộ đệm để lưu trữ gói dữ liệu đang được chuyển mạch.
* Trình điều khiển thiết bị (Device driver): Làm chức năng điều khiển giao tiếp giữa phần cứng và thiết bị ngoại vi, giao tiếp giữa các tiến trình iOS, kernel và phần cứng. Chúng cũng giao tiếp với phần mềm chuyển mạch nhanh (fast switching software).
* Phần mềm chuyển mạch nhanh (Fast switching soft): Chức năng chuyển mạch gói dữ liệu cao.

# **CHƯƠNG 2: MÔ TẢ VỀ ỨNG DỤNG**

1. HIỆN TRẠNG HỌC TIẾNG ANH CỦA TRẺ EM TẠI VIỆT NAM

Tiếng Anh đã trở thành một phương tiện quan trọng trong thời đại hội nhập hiện nay. Kéo theo đó, nhu cầu cho trẻ em sớm tiếp cận tiếng Anh đã trở thành một trào lưu cực phổ biến. Viễn cảnh bé ở nhà nói tiếng Anh líu lo và phát triển thành một người thông thạo tiếng Anh là điều mà ai cũng mong muốn và hướng tới.

Tuy vậy, thực tế thường sẽ không dễ dàng như mong đợi. Cũng giống như việc học bơi sẽ phải sặc nước, học viết sẽ phải lem mực, việc các bé gặp những vấn đề khó khăn khi học tiếng Anh là không thể tránh khỏi.

* Không nhớ từ vựng

Đây là vấn đề cơ bản nhất mà các bé thường gặp phải. Kể cả theo một thời gian dài, việc bé không thể nhớ được từ vựng vẫn sẽ xảy ra thường xuyên. Điều này có liên quan đến phương pháp học cũng như tâm lí học của trẻ.

Về nguyên nhân tâm lí, thường việc không nhớ từ vựng xuất phát từ tâm lí lơ là hoặc chủ quan trong việc học tiếng Anh của bé. Có những bé được học tiếng Anh từ sớm, hoặc được bố mẹ dạy tiếng Anh bằng những phương pháp khác so với việc học tiếng Anh truyền thống tại trường, khiến cho bé không coi trọng cách học hiện tại. Cũng có khả năng, bé không có hứng thú với việc học tiếng Anh , dẫn đến việc học và nhớ từ một cách qua loa, không đưa được từ vựng vào sâu trí nhớ. Đây là những điều mà bố mẹ cần quan tâm và hỗ trợ bé kịp thời, vì nó dễ dẫn đến việc bé chán ghét tiếng Anh.

Về nguyên nhân phương pháp học, có rất nhiều lí do dẫn đến việc bé không nhớ được từ vựng. Thường vấn đề chủ yếu của các bé là phương pháp học không chặt chẽ, không giúp bé luyện tập và sử dụng được vốn từ vựng đã học. Ví dụ như bé chỉ được học từ vựng trên lớp, nhưng không có bài tập để sử dụng những từ vựng đó, khiến cho số lượng từ vựng có thể nhớ sau giờ học sẽ giảm dần. Đi xa hơn, có thể là bé chỉ được sử dụng tiếng Anh trong giờ học, và sau giờ học thì toàn bộ tiếng Anh được cất qua một bên. Với tình trạng này, không chỉ các bé mà cả người lớn cũng sẽ bị mai một đi vốn từ vựng của mình.

* Phát âm sai

Thực ra thì đây không phải là vấn đề mà chỉ có các bé gặp phải. Bạn có thể bắt gặp những người lớn phát âm tiếng Anh sai hàng ngày. Và thực tế, vấn đề này phần lớn đến từ người dạy hoặc môi trường học tiếng Anh của bé.

Vấn đề phát âm sai trong môi trường học đường ở Việt Nam đã không còn là vấn đề mới lạ nữa. Chuyện người này phát âm sai và sau đó chỉ lại cho người khác cùng phát âm như vậy cũng là tình huống rất phổ biến. Cách tốt nhất và nhanh nhất để giải quyết vấn đề này, là luôn tham khảo kĩ cách phát âm trong từ điển khi học để đối chiếu. Sẽ rất lí tưởng nếu bố mẹ trang bị cho bé các loại Kim từ điển, hoặc cài đặt các phần mềm học tiếng Anh online trên máy, giúp bé có thể nhanh chóng tự tham khảo phát âm bất cứ lúc nào.

* Ít cơ hội được giao tiếp, thực hành

Học ở trường, thầy cô chỉ chú trọng ngữ pháp, từ vựng và thi viết. Học sinh gần như không được thực hành nghe, nói, thảo luận... vì không có phòng thực hành. Kiến thức trong sách giáo khoa đã cũ, không được chỉnh sửa. Chương trình học quá nặng về ngữ pháp, biến các em thành “cuốn ngữ pháp biết đi”.

1. ĐỐI TƯỢNG VÀ MỤC TIÊU CỦA ỨNG DỤNG

Với hiện trạng nêu trên, để cải thiện quá trình học tiếng anh trở nên hiệu quả, chất lượng nhưng cũng không tạo cảm giác chán nản khi học, ứng dụng được xây dựng để hướng đến các em học sinh cấp 1 và các bé mới học Tiếng Anh với hình ảnh sinh động, thú vị. Hi vọng các em sẽ cảm thấy hứng thú với việc học Tiếng Anh để tạo nên thói quen chăm chỉ, xây dựng một nền tảng vững chắc về cả từ vựng, ngữ pháp cũng như giao tiếp để các em có thể dễ dàng phát triển sau này.

1. ÁP DỤNG XỬ LÝ NGÔN NGỮ TỰ NHIÊN TRONG ỨNG DỤNG

Xử lý ngôn ngữ tự nhiên là một lĩnh vực đặc biệt, là sự kết hợp giữa các ngành khoa học máy tính, trí tuệ nhân tạo và ngôn ngữ học.

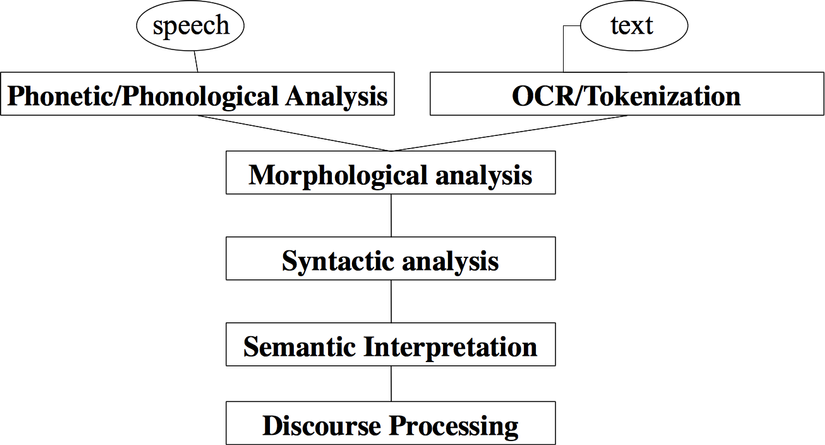
Mục tiêu của việc xử lý ngôn ngữ tự nhiên là để cho máy tính xử lý và hiểu được ngôn ngữ tự nhiên của con người, giúp máy tính có thể thực hiện được một số nhiệm vụ hữu ích thay cho con người như đặt lịch hẹn, mua bán hàng hóa, dịch từ ngôn ngữ này sang ngôn ngữ khác, các hệ tư vấn, hệ hỏi đáp (Ví dụ: Siri, Google Assistant, Facebook M, Cortana, ...).

Để máy tính có thể hiểu được đầy đủ và thể hiện được đúng ý nghĩa của ngôn ngữ là một nhiệm vụ cực kì khó.

Một vài bài toán của xử lý ngôn ngữ tự nhiên như:

* Spell checking: phát hiện và sửa lỗi chính tả
* Finding synonyms: Tìm từ có nghĩa tương đồng
* Extracting information: Trích rút thông tin từ websites như giá sản phẩm, ngày tháng, địa điểm, tên người và tên công ty
* Classifying: Phân loại quan điểm(Tích cực/ Tiêu cực) của một văn bản dài, phân loại tin tức,...
* Machine translation: Dịch từ ngôn ngữ nguồn sang ngôn ngữ đích
* Spoken dialog systems: Các hệ thống hội thoại giữa người và máy(Tư vấn khách hàng tự động, điều khiển thiết bị, đặt hàng,...)
* Question Answering: Các hệ hỏi đáp
* Speech recognition: Nhận dạng giọng nói

Ta chia việc xử lý ngôn ngữ tự nhiên thành các cấp độ khác nhau:



Trong đó, đầu vào thường là hai dạng chính của ngôn ngữ gồm lời nói(speech) và văn bản(text). Sau khi phân tích ngữ âm (đối với dạng speech) hoặc OCR/Tokenization văn bản, chúng ta sẽ trải qua các bước xử lý ngôn ngữ theo các cấp độ:

* Morphological analysis: phân tích hình thái của ngôn ngữ, bao gồm các khâu xử lý:
* Phân đoạn từ vựng (word segmentation): phân giải câu văn được nhập vào thành các từ có thứ tự.
* Phân loại từ (part-of-speech tagging): quyết định từ loại của từ vựng
* Phục hồi thể nguyên dạng của từ (lemmatization) (đối với tiếng anh): làm trở lại nguyên dạng ban đầu các từ vựng bị biến đổi thể (inflection) hoặc được kết hợp (conjugatetion).
* Syntactic analysis: phân tích cú pháp, tìm hiểu cấu trúc của câu (chủ ngữ, động từ chính, ...)
* Semantic interpretaion: diễn dịch ngữ nghĩa, ý nghĩa của câu dựa vào các từ tạo nên câu
* Discourse processing: phân tích ngữ nghĩa dựa trên bối cảnh của câu.
  1. NHỮNG ĐIỀU ĐẶC BIỆT CỦA NGÔN NGỮ

Ngôn ngữ của con người là một hệ thống các tín hiệu/ký hiệu được xây dựng một cách đặc biệt để truyền đạt được thông tin có chủ đích của người viết/người nói. Các tín hiệu/ký hiệu này được con người sử dụng để giao tiếp với nhau theo nhiều cách:

* Âm thanh
* Cử chỉ
* Chữ viết
* Hình ảnh/Tranh vẽ

Mỗi người sẽ có cách mã hóa các tín hiệu/ký hiệu này khác nhau (giọng nói khác nhau, chữ viết khác nhau, nét vẽ khác nhau). Tuy nhiên, ngữ nghĩa là bất biến và thống nhất không phụ thuộc vào đối tượng thể hiện.

* 1. CÁC BƯỚC THỰC HIỆN
* Phân tích hình thái - Trong bước này từng từ sẽ được phân tích và các ký tự không phải chữ (như các dấu câu) sẽ được tách ra khỏi các từ. Trong tiếng Anh và nhiều ngôn ngữ khác, các từ được phân tách với nhau bằng dấu cách. Tuy nhiên trong tiếng Việt, dấu cách được dùng để phân tách các tiếng (âm tiết) chứ không phải từ. Cùng với các ngôn ngữ như tiếng Trung, tiếng Hàn, tiếng Nhật, phân tách từ trong tiếng Việt là một công việc không hề đơn giản.
* Phân tích cú pháp - Dãy các từ sẽ được biến đổi thành các cấu trúc thể hiện sự liên kết giữa các từ này. Sẽ có những dãy từ bị loại do vi phạm các luật văn phạm.
* Phân tích ngữ nghĩa - Thêm ngữ nghĩa vào các cấu trúc được tạo ra bởi bộ phân tích cú pháp.
* Tích hợp văn bản - Ngữ nghĩa của một câu riêng biệt có thể phụ thuộc vào những câu đứng trước, đồng thời nó cũng có thể ảnh hưởng đến các câu phía sau.
* Phân tích thực nghĩa - Cấu trúc thể hiện điều được phát ngôn sẽ được thông dịch lại để xác định nó thật sự có nghĩa là gì.

Tuy nhiên, ranh giới giữa 5 bước xử lý này cũng rất mong manh. Chúng có thể được tiến hành từng bước một, hoặc tiến hành cùng lúc - tùy thuộc vào giải thuật và ngữ cảnh cụ thể.

* 1. XÂY DỰNG CHATBOT ĐỂ GIAO TIẾP TIẾNG ANH

Chìa khóa để cho một con bot có thể hiểu được, giao tiếp được với con người là nó phải có khả năng hiểu các ý định của con người và đáp lại. Và việc quan trọng cần làm đó là thiết kế các Intent và Entity.

* Intent

Intent đơn giản là ý định của người dùng cuối, những ý định này được chuyển tải bởi người dùng tới con bot của bạn. Bạn có thể chủ yếu đặt ý định vào 2 nhóm chính: casual intents, business intents.

* Casual Intents - Đây là những ý định mở đầu hoặc kết thúc một cuộc hội thoại. Nhưng lời chào như: hi, hello, hallo, bye… là những câu lệnh mở đầu hoặc kết thúc một cuộc hội thoại. Những ý định này nên hướng con bot của bạn phản hồi bằng một câu trả lời gợi ý như: “Hello. How can I help you?” hay “Bye, see you later”.

Các ý định ngẫu nhiên cũng bao gồm các ý định Khẳng định - Affirmative và Phủ định - Negative cho các câu nói như: “OK”, “Yes”, “No”, “Not yet”, …

Bằng các ý định chung chung có thể giúp bạn xử lý tất cả các ý định của người dùng thay vì đưa câu chuyện đã có vào cuộc trò chuyện với bot.

* Business Intents - Đây là các ý định được kết nối trực tiếp với ý đồ ban đầu của các con bot khi được xây dựng, tức là bot cần phải hỗ trợ và xử lý được các ý định của người dùng. Ví dụ người dùng hỏi “What is the weather like today?”, con bot khi đó phải hiểu ý định người dùng sau đó lấy được dữ liệu ở một server nào đó rồi đáp lại người dùng. Tuy nhiên với ứng dụng này, mục đích tạo ra là giao tiếp thông thường nên các intent chỉ cẩn thiết kế để đáp lại một cách hợp lí chứ không cần thực hiện một hành động nào cả.
* Entity

Mục đích chính của giai đoạn này là để nhận dạng thực thể (entity). Nó được giao trách nhiệm phát hiện các từ và cụm từ quan trọng đối với mỗi câu của người dùng.

Ví dụ có 1 câu là: “I have one dog”

one: có thể dễ dàng phân loại thành thực thể số đếm, được gán nhãn “number”

dog: nếu chi tiết có thể xếp vào thực thể loài chó, ví dụ như:



Ta có thể thấy nhãn “dog” gồm các từ: dog, pup, puppy

Khi đó những câu khác như: I have 2 pups, I have 3 puppies sẽ được hiểu nghĩa tương tự với câu ban đầu.

Tuy nhiên nếu tổng quát có thể xếp dog vào thực thể động vật, ví dụ như:



Khi đó, những câu như I have 2 cats sẽ được hiểu tương tự với câu ban đầu.

Tóm lại, mục tiêu của bước này là tìm các từ đồng nghĩa, gán nhãn và xếp chúng vào các nhóm thích hợp.