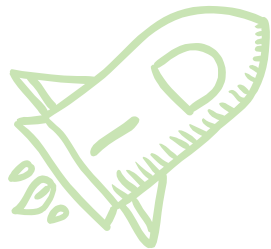
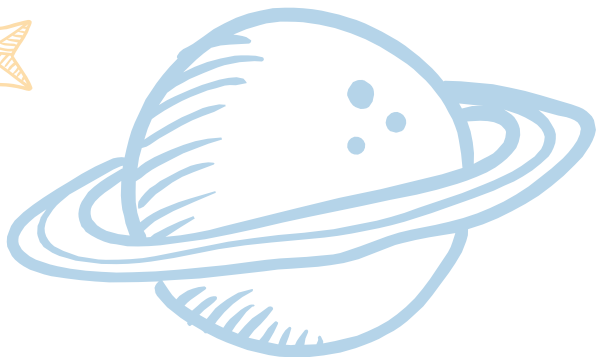


NHẬP MÔN TRÍ TUỆ NHÂN TẠO

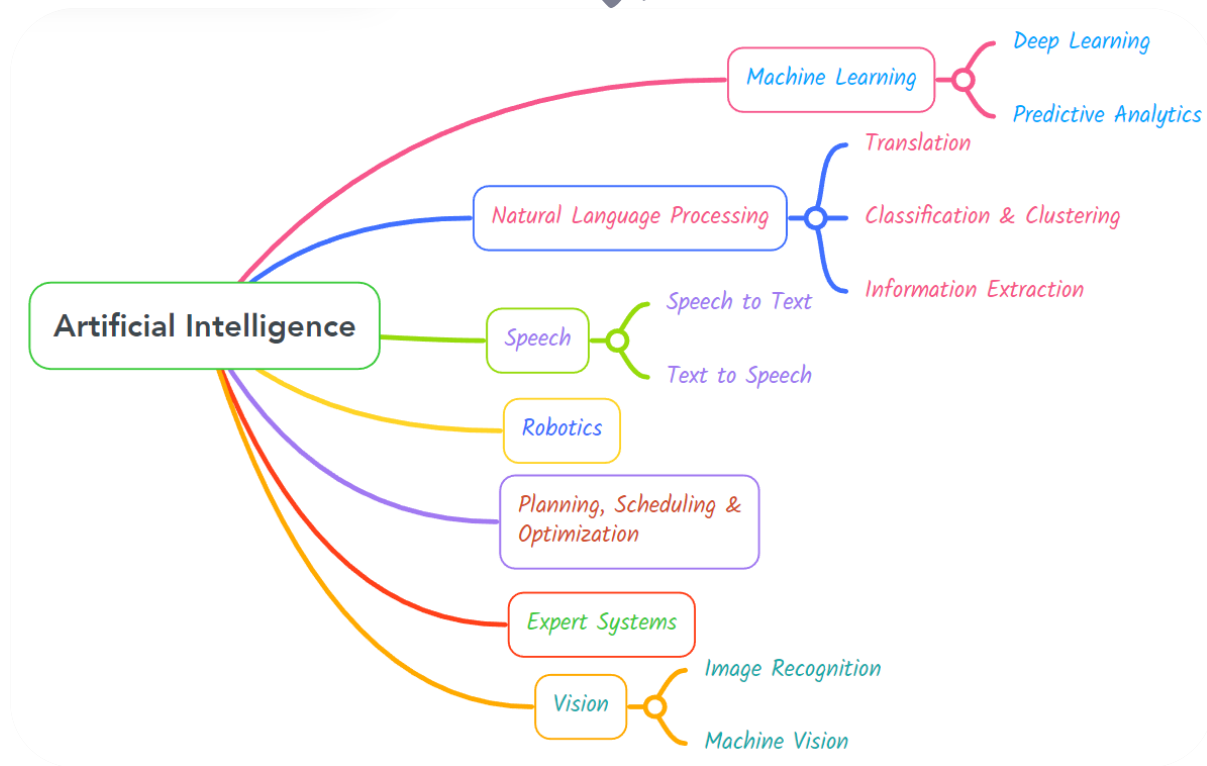




THÀNH VIÊN NHÓM

1. Cao Thị Cẩm Ngân
2. Trần Thu Trang
3. Lý Đại Phát

CÁC NHÁNH TRÍ TUỆ NHÂN TẠO



ROADMAP

1. Giới thiệu sơ lược

3. Một số ứng dụng của NLP

5. Demo Chatbot



2. Các hướng tiếp cận NLP

4. Chatbot

6. End

1. GIỚI THIỆU SƠ LƯỢC

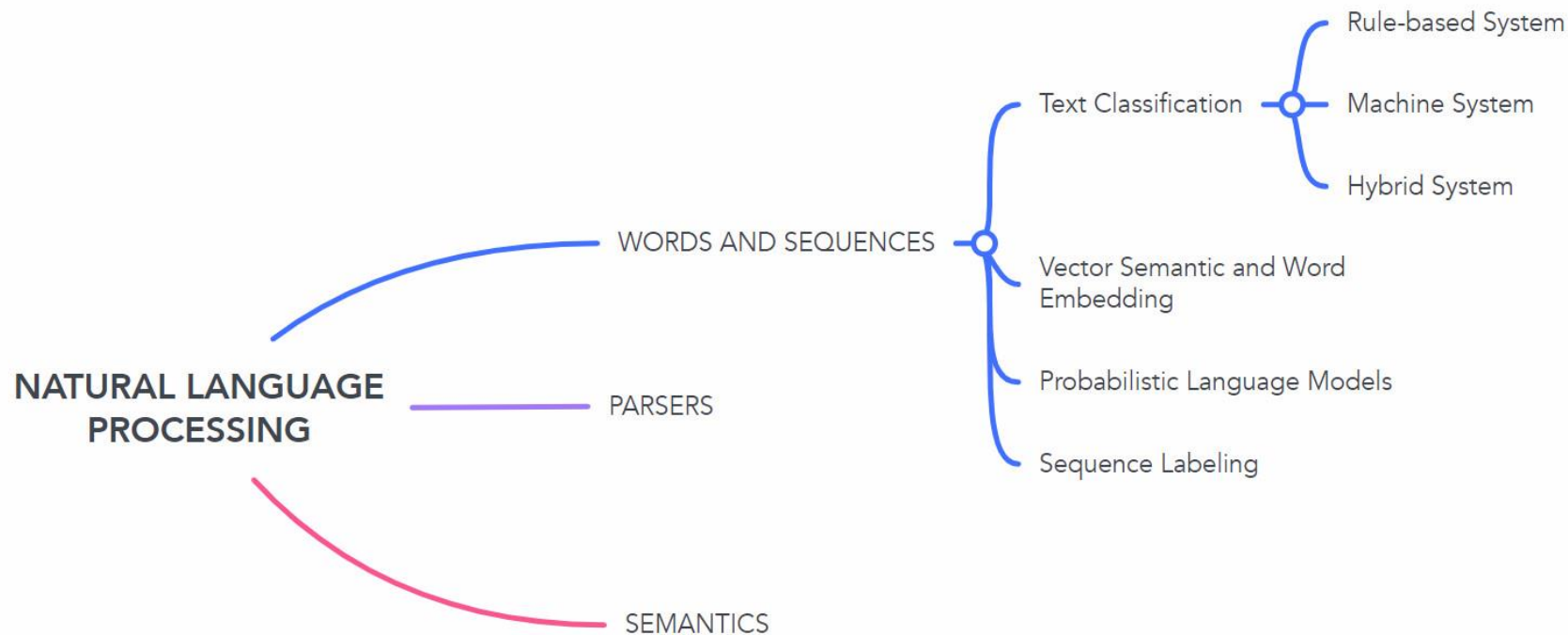


1. GIỚI THIỆU SƠ LƯỢC

- Xử lý ngôn ngữ tự nhiên (natural language processing), một nhánh nghiên cứu của trí tuệ nhân tạo, thông qua việc phát triển các thuật toán, để xây dựng các chương trình máy tính có khả năng phân tích, xử lý, và hiểu ngôn ngữ của con người.



SƠ ĐỒ HƯỚNG TIẾP CẬP NLP



2. CÁC HƯỚNG TIẾP CẬN NLP

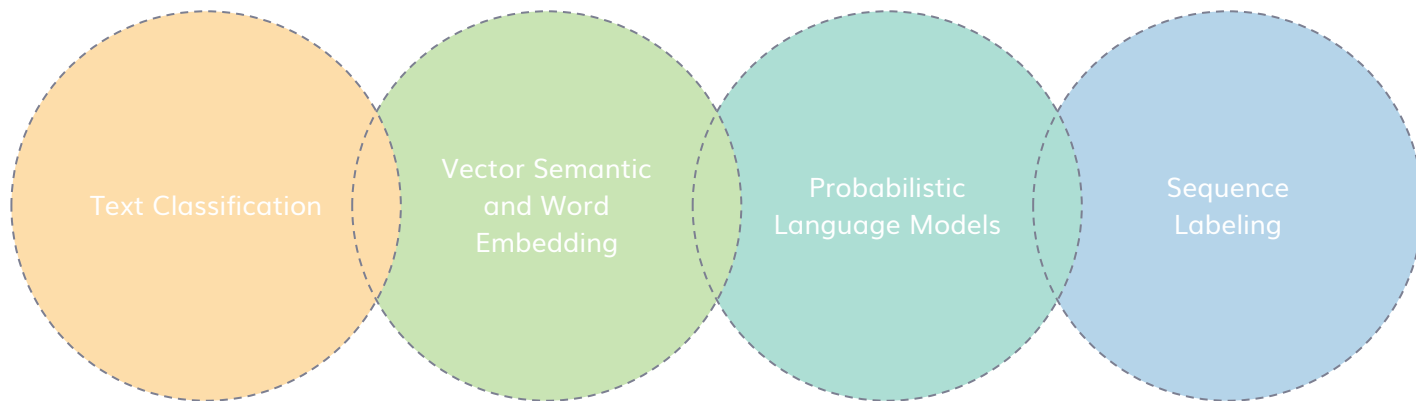
2.1. Word and Sequences

- Text Classification
- Vector Semantic and Word Embedding
- Probabilistic Language Models
- Sequen labeling

2.2. Parsers

2.3. Semantic

2.1 WORDS AND SEQUENCES



2.1.1. TEXT CLASSIFICATION

Là quá trình phân loại văn bản thành các nhóm từ, các nhóm này sẽ được gán các tag (hoặc nhãn) đã được định nghĩa trước dựa trên ngữ cảnh của văn bản.

Một số kỹ thuật thường dùng trong bài toán phân loại: Naïve Bayes, Decision Tree, Perceptron, **SVM**

Có 3 hướng chính để phân loại là:

Rule-based System

Các văn bản được tách thành một nhóm có tổ chức bằng cách sử dụng các bộ quy tắc về ngôn ngữ.

VD: Donald Trump & Boris Jhonson sẽ được phân loại vào nhóm chính trị, LeBron James & Messi sẽ được xếp vào nhóm thể thao.

Machine System

Hệ thống sẽ được huấn luyện từ tập dữ liệu sẵn có. Dữ liệu đã được gán nhãn sẵn và hệ thống sẽ được học liên tục thông qua các dữ liệu này, dữ liệu càng nhiều thì tỷ lệ hệ thống phân loại đúng càng cao.

Việc phân loại văn bản sẽ dựa trên các quan sát và kinh nghiệm từ các lần huấn luyện.

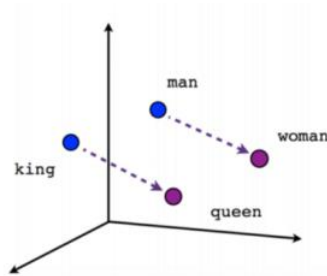
Hybird System

Kết hợp cả 2 cách tiếp cận trên.

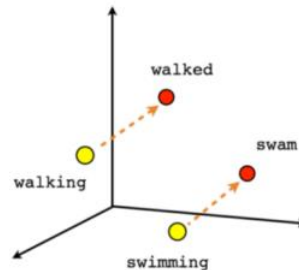
Trong đó Rule-based system sẽ có nhiệm vụ tạo ra các tag và sử dụng Machine System để đào tạo và tạo ra các quy tắc, sau đó các quy tắc được tạo ra này sẽ được so sánh với các bộ quy tắc chuẩn, nếu có sai sót thì con người sẽ cải thiện lại các quy tắc này.

2.1.2 VECTOR SEMANTIC & WORD EMBEDDING

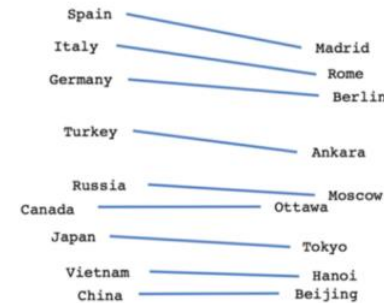
Vector semantic: là phương pháp phân chia các từ trong không gian vector nhiều chiều, được dùng để xác định và giải thích ý nghĩa của từ, ý tưởng chính của vector semantic là hai từ được xem là tương đồng nhau nếu chúng được sử dụng trong ngữ cảnh giống nhau.



Male-Female



Verb tense



Country-Capital

2.1.2 VECTOR SEMANTIC & WORD EMBEDDING

Word Embedding: là phương pháp ánh xạ (map) những từ ngữ vào các vector hoặc số thực, còn được gọi là phương pháp mô hình hóa ngôn ngữ/dữ liệu. Một phép nhúng từ tốt sẽ mang lại nhiều lợi ích cho việc tính toán lẫn minh họa dữ liệu.

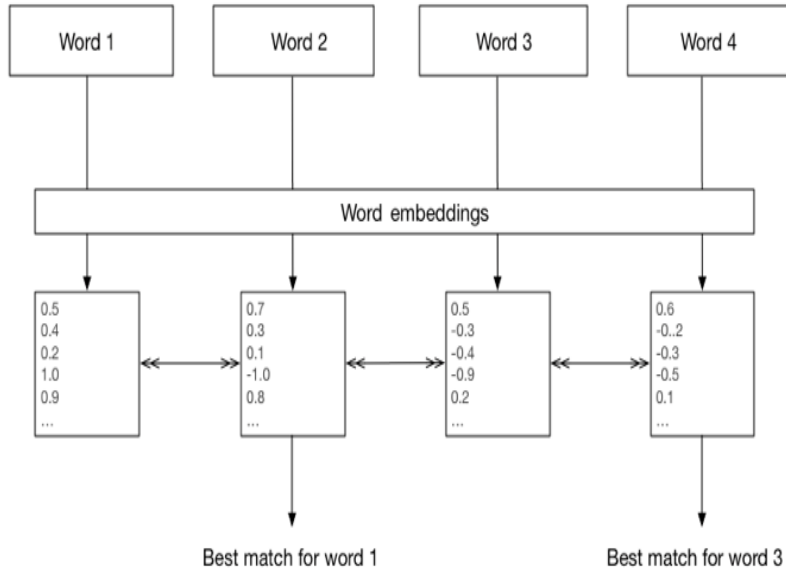
Gồm 2 loại: Word2Vec và Doc2Vec

Ý tưởng cơ bản của Word2Vec:

- Hai từ xuất hiện trong những văn cảnh giống nhau thường có ý nghĩa gần với nhau.
- Ta có thể đoán được một từ nếu biết các từ xung quanh nó trong câu. Ví dụ, với câu “Hà Nội là ... của Việt Nam” thì từ trong dấu ba chấm khả năng cao là “thủ đô”. Với câu hoàn chỉnh “Hà Nội là thủ đô của Việt Nam”, thì mô hình Vector Semantic & Word embedding sẽ xây dựng ra embedding của các từ sao cho xác suất để từ trong dấu ba chấm là “thủ đô” là cao nhất.

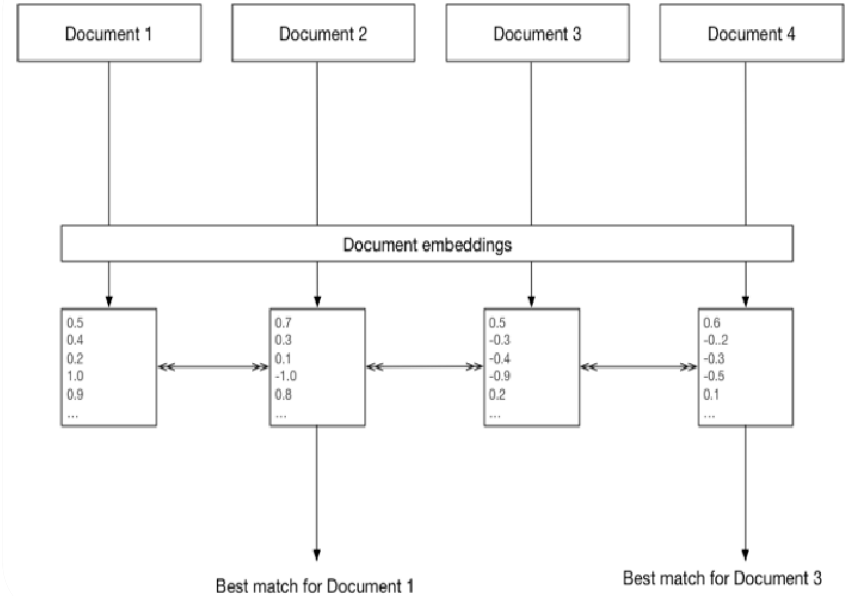
2.1.2 VECTOR SEMANTIC & WORD EMBEDDING

Word similarity based on word embeddings.



Word2Vec

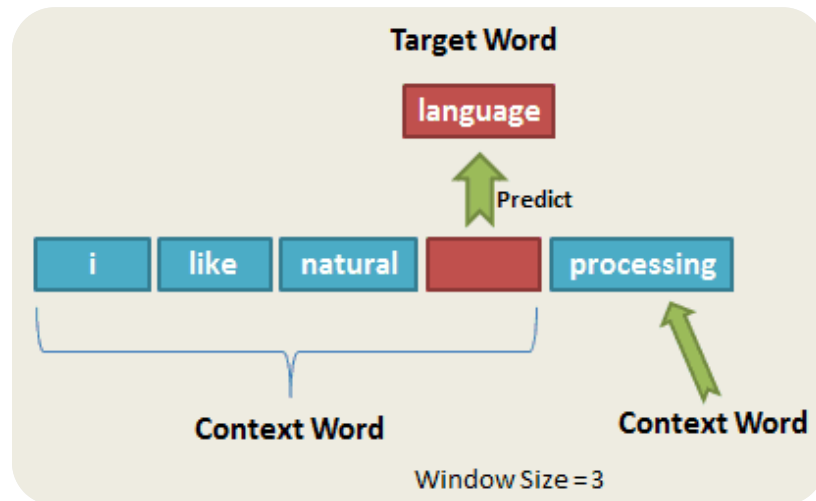
Document similarity based on document embeddings.



Doc2Vec

2.1.3. PROBABILISTIC LANGUAGE MODEL

Một cách tiếp cận khác để phân tích từ và trình tự là mô hình ngôn ngữ xác suất. Mục tiêu của mô hình ngôn ngữ xác suất là tính toán xác suất của một câu của một chuỗi từ. Ví dụ: xác suất của từ “a” xuất hiện trong một từ “to” đã cho là 0,00013131 phần trăm.

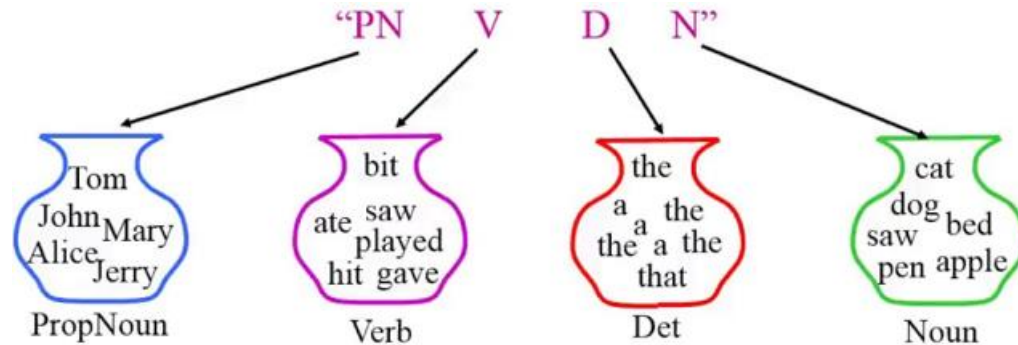


2.1.4. SEQUENCE LABELING

Là quá trình gán một lớp hoặc nhãn cho một chuỗi các quan sát một cách đồng thời.

Ví dụ với câu nói “Hãy xem phim Tom&Jerry”, thì theo thứ tự nhãn sẽ được gán là [xem, phim, Tom&Jerry]. “xem” xác định một hành động, “phim” là một ví dụ của hành động, “Tom&Jerry” là mục đích chính của hành động.

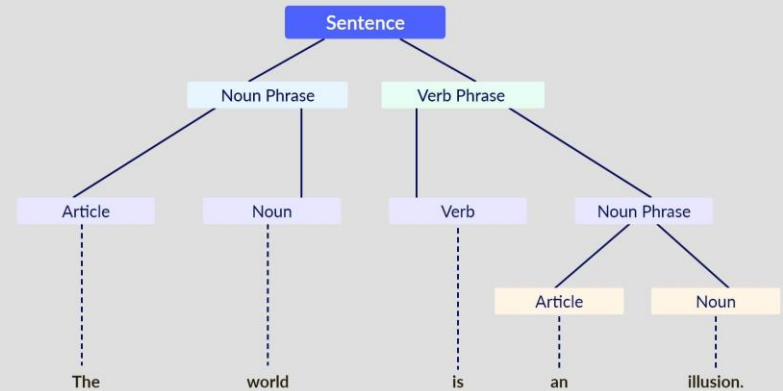
Các kỹ thuật thường dùng có thể kể tới là: Hidden Markov Model, Maximum Entropy Markov Model, **Conditional Random Fields**.



2.2 PARSING

Là phương pháp xác định cấu trúc của văn bản bằng cách phân tích các từ cấu thành của văn bản dựa trên các ngữ pháp cơ bản. Văn bản sẽ được chia thành các thành phần và các từ cấu thành nên các phần đó.

Tùy thuộc vào cây được xây dựng sẽ có những kỹ thuật thích hợp đó là: universal parsing, top-down parsing và bottom-up parsing.



2.3 SEMANTIC

Semantic hay ngữ nghĩa học, là một quá trình tìm cách hiểu ý nghĩa ngôn ngữ bằng cách xây dựng một mô hình nguyên tắc để truyền đạt ý nghĩa. Trong đó có các yếu tố cần quan tâm đến là:

- Ngữ nghĩa theo ngôn ngữ tự nhiên.
- Ngữ cảnh của hành vi truy vấn (các thuật ngữ liên quan).
- Ngữ cảnh của người truy vấn.
- Tham chiếu với thực thể.

Semantic có vai trò quan trọng trong việc loại bỏ các nội dung spam, nội dung có chất lượng kém và nâng cao tính chuẩn xác của kết quả tìm kiếm.

3. ỨNG DỤNG NLP

3. ỨNG DỤNG NLP

Personal Assistant

- Siri, Cortana, Google Assistant



Auto complete

- Tự động điền từ còn thiếu khi bạn tìm kiếm thông tin.



3. ỨNG DỤNG NLP

Spell Checking

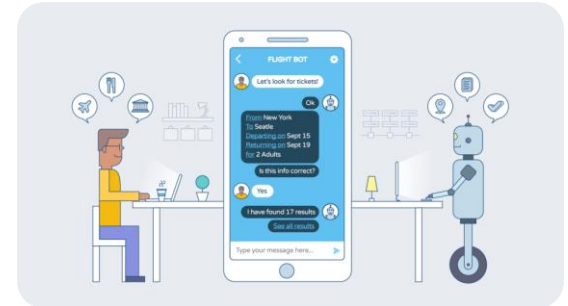
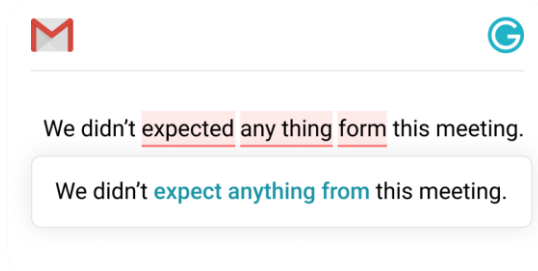
- Google Doc, Microsoft Word, ...

Translation

- Google Translate, TFlat Dictionary, Hanzii Dict, ...

Customer Assistant

- Chatbot





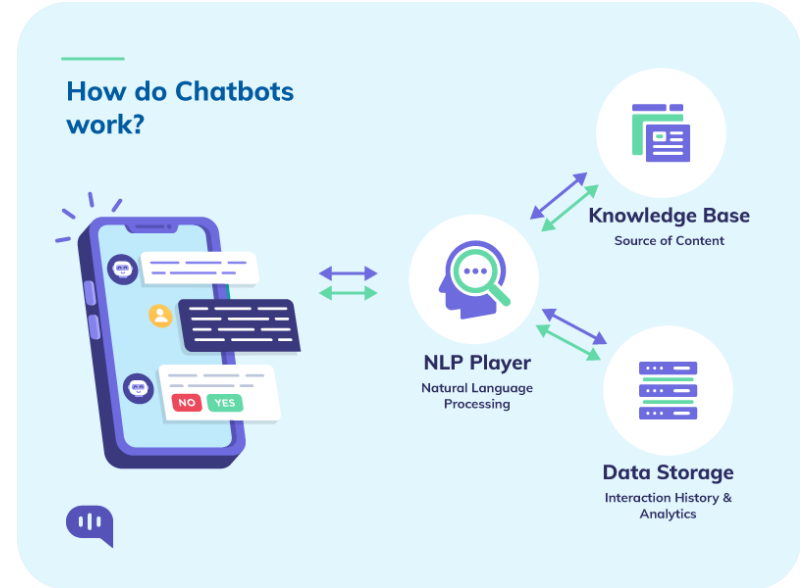
4. CHATBOT!

Hello Human, I'm Smith!

Chatbot là một chương trình kết hợp với trí tuệ nhân tạo (AI) để tương tác với con người. Công cụ này thay thế cho nhân viên để tư vấn trả lời những gì khách hàng thắc mắc.



Chatbot là sự kết hợp của các kịch bản có trước và tự học trong quá trình tương tác. Với các câu hỏi được đặt ra, Chatbot sử dụng nhiều hệ thống quét các từ khóa bên trong đầu vào, sau đó bot khởi động một hành động, kéo một câu trả lời với các từ khóa phù hợp nhất từ một cơ sở dữ liệu (API), hoặc bàn giao cho con người. Nếu tình huống đó chưa có trong cơ sở dữ liệu, Chatbot sẽ bỏ qua nhưng sẽ đồng thời tự học để áp dụng cho các cuộc trò chuyện về sau.





ChatBot Demo



THANK FOR ATTENTION!