

ĐẠI HỌC QUỐC GIA TP HỒ CHÍ MINH
ĐẠI HỌC BÁCH KHOA
KHOA KHOA HỌC VÀ KỸ THUẬT MÁY TÍNH



NHẬP MÔN ĐIỆN TOÁN (CO1005)

BÁO CÁO BÀI TẬP LỚN

XỬ LÝ NGÔN NGỮ TỰ NHIÊN (NLP)

Giảng viên: **Quản Thành Thơ**
Huỳnh Nam

Thành viên: **Đoàn Nguyễn Thành Lương - 1611947**
Nguyễn Trương Ngọc Bách - 1610155
Lê Thị Thúy Hằng - 1610956
Võ Khắc Trí - 1613754

TP Hồ Chí Minh, 12/2016

Mục lục

1	Xử lý ngôn ngữ tự nhiên (NLP)	2
1.1	Khái niệm	2
1.2	Lịch sử	2
1.3	Các bước chính trong NLP	5
1.4	Những bài toán	5
1.4.1	Hiểu ngôn ngữ tự nhiên (Natural language understanding)	5
1.4.2	Sinh ngôn ngữ tự nhiên (Natural language generation) . .	6
1.5	Ứng dụng của NLP	7
1.5.1	Tóm tắt văn bản tự động (Automatic summarization) . .	7
1.5.2	Chấm điểm tự động (Automated essay scoring)	7
1.5.3	Dịch tự động (Machine translation)	8
1.5.4	Phát hiện thư rác (Spam detection)	8
1.5.5	Nhận dạng giọng nói (Speech recognition)	9
2	Khó khăn trong lĩnh vực NLP	10
2.1	Sự nhập nhằng ngữ nghĩa (semantic ambiguity)	10
2.2	Sự đa dạng của ngôn ngữ (language variability)	11
2.3	Kiến trúc phần cứng (hardware architecture)	11
3	Tài liệu tham khảo	11
4	Đóng góp các thành viên	12

1 Xử lý ngôn ngữ tự nhiên (NLP)

1.1 Khái niệm

- Xử lý ngôn ngữ tự nhiên (Natural language processing - NLP) là một nhánh của trí tuệ nhân tạo liên quan đến vấn đề phân tích, sự hiểu và tạo ra những ngôn ngữ mà con người có thể sử dụng một cách tự nhiên để giao tiếp với máy tính cả trong nói và viết bằng ngôn ngữ (tự nhiên) của con người thay vì ngôn ngữ máy tính.



1.2 Lịch sử

- **1940s-1950s:** Đây là giai đoạn của những viên gạch đầu tiên về NLP với nền tảng: máy tự động và mô hình về xác suất hay thông tin lý thuyết. Máy tự động nổi lên trong những năm 50 bắt nguồn từ mô hình tính toán thuật toán của Alan Turing, được nhiều người cho là nền tảng của khoa học máy tính hiện đại. Shannon (1948) áp dụng mô hình xác suất của các quá trình rời rạc Markov vào máy tự động cho ngôn ngữ. Năm 1950 Alan Turing xuất bản bài viết "Computing Machinery and Intelligence" để đề xuất một tiêu chuẩn đánh giá trí thông minh của máy tính, ngày nay gọi là "Turing test". Giai đoạn này đã xuất hiện máy dịch thuật (MT)- máy tính đầu tiên dựa trên ứng dụng liên quan đến xử lý ngôn ngữ tự nhiên. Công việc ban đầu của MT chỉ đơn giản là chuyển đổi giữa các ngôn ngữ khác nhau dựa trên từ vựng - với quan điểm rất đơn giản là tra cứu từ điển và sắp xếp lại cho phù hợp với quy tắc về thứ tự của ngôn ngữ đó - mà bỏ qua nhiều vấn đề như sự nhập nhằng của ngôn ngữ,.. Kết quả không tốt đã khiến các nhà nghiên cứu nhận ra những thiếu sót này, và họ thấy nhiệm vụ thực sự rất khó khăn, và họ cần một cơ sở lý thuyết đầy đủ hơn cho xử lý ngôn ngữ. Và đến giai đoạn tiếp theo, những công trình của Chomsky đã giúp cải thiện phần nào điểm yếu đó của MT.



- **1957-1970s:** Đến cuối những năm 50 và đầu những năm 60, xử lý lời nói

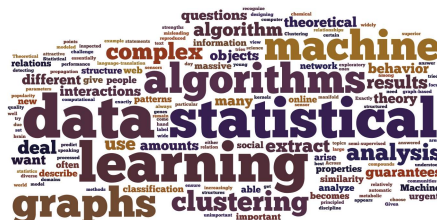
và ngôn ngữ đã chia ra rất rõ ràng thành hai loại: symbolic (biểu tượng) và stochastic (ngẫu nhiên)...Symbolic được bắt đầu từ hai hướng nghiên cứu. Đầu tiên là công trình của Chomsky và những người khác về lý luận về cú pháp ngôn ngữ hình thức và "cú pháp biến tạo"(generative syntax) trong cuối những năm 50 đến giữa thập niên 60, và công trình của những nhà ngôn ngữ học và khoa học máy tính về phân tích thuật toán. Hướng nghiên cứu còn lại là một lĩnh vực mới: trí tuệ nhân tạo. Vào mùa hè năm 1956 John McCarthy, Marvin Minsky, Claude Shannon và Nathaniel Rochester cùng một nhóm các nhà nghiên cứu đã tổ chức một hội thảo kéo dài hai tháng về thứ mà họ đã quyết định gọi là trí tuệ nhân tạo (AI). Stochastic được sử dụng chủ yếu trong ngành thống kê và kỹ thuật điện (electrical engineering) .Vào cuối thập kỷ 50, phương pháp Bayesian đã bắt đầu được áp dụng cho nhận dạng ký tự.

Symbolic vs. Stochastic

- **Symbolic**
 - Formal grammars as basis for NLP and learning systems. (Chomsky, Harris)
 - Logic and logic-based programming for characterizing syntactic/semantic inference (Kaplan, Kay, Pereira)
 - Toy natural-language understanding/generation systems (Woods, Minsky, Schank, Winograd, Colmerauer)
 - Discourse processing: Role of intention, focus (Grosz, Sidner, Hobbs)
- **Stochastic Modeling**
 - Probabilistic methods for early speech recognition, OCR (Bledsoe and Browning, Jelinek, Black, Mercer)
- Những bất cập tồn tại trong hệ thống dẫn đến ALPAC (Automatic Language Processing Advisory Committee của Viện Hàn lâm khoa học – Hội đồng nghiên cứu quốc gia) năm 1966. Báo cáo kết luận rằng MT đã không thể đáp ứng những dự tính/ dự định đặt ra từ trước, dẫn đến những dự án MT chỉ còn được tài trợ rất ít. Những nghiên cứu về MT chỉ được tiếp tục rất lâu sau đó (cuối những năm 80) khi hệ thống dịch máy thống kê đầu tiên (statistical machine translation systems) được tạo ra. Các công trình nghiên cứu lý thuyết từ cuối 1960 đến đầu 1970 tập trung vào vấn đề diễn giải nghĩa và tạo ra những hệ thống tính toán với nhiều phương pháp khác nhau. Cùng với phát triển lý thuyết, nhiều mẫu thử nghiệm hệ thống được phát triển để chứng minh hiệu quả của các nguyên tắc cụ thể. Tiêu biểu nhất là Eliza - một trong những chatterbot đầu tiên, và cũng là phần mềm đầu tiên có thể vượt qua được Turing test.



- **1970s:** Nhiều lập trình viên bắt đầu viết những "conceptual ontologies", thứ cấu trúc lại những thông tin thực tế thành những dữ liệu mà máy tính có thể hiểu được. Cũng trong khoảng thời gian này, nhiều chatterbots được viết bao gồm PARRY, Racter, và Jabberwacky.
- **1980s-1993:** Cho đến thập niên 80, hầu hết các hệ thống NLP được dựa trên những bộ quy tắc phức tạp được viết tay. Bắt đầu từ cuối những năm 80 nhưng thật sự đã có một cuộc cách mạng trong NLP với sự ra đời của thuật toán Machine learning cho xử lý ngôn ngữ. Điều này có được là nhờ cả sự tăng ổn định về năng lực tính toán của máy tính và sự thống trị bị giảm dần của các lý thuyết về ngôn ngữ học của Chomsky - thứ mà nền tảng chỉ là lý thuyết suông chống lại Corpus Linguistics - nền tảng cho mối quan hệ của Machine learning và xử lý ngôn ngữ - cùng những gì liên quan. Những thuật toán ML đầu tiên được sử dụng, như Decision Tree, tạo ra những hệ thống với những quy tắc về if-then hết sức cứng nhắc - tương tự như những quy định về viết tay cùng tồn tại lúc đó.



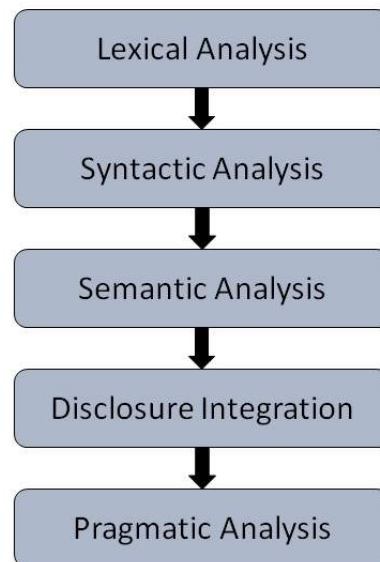
- **1994 cho tới nay:** Trong 5 năm cuối của thiên niên kỉ, có thể thấy sự thay đổi rõ rệt của NLP. Thứ nhất là những mô hình thống kê và điều khiển bằng dữ liệu (probabilistic and data-driven models) đã trở thành những tiêu chuẩn của NLP. Thứ 2 là sự phát triển về tốc độ cũng như bộ nhớ của máy tính đã mở ra sự khai thác về mặt kinh tế những subarea của NLP như nhận dạng ngôn ngữ hay kiểm tra chính tả và ngữ pháp. Có 3 xu hướng chính thúc đẩy sự phát triển của NLP: (1) Một lượng lớn dữ liệu về tài liệu nói và viết trở nên phổ biến dưới sự bảo trợ của Linguistic Data Consortium (LDC) và những tổ chức tương tự khác. Những nguồn tài nguyên này đã góp phần thúc đẩy xu hướng xử lí những vấn đề truyền thống nhưng phức tạp, như phân tích ngữ pháp hay ngữ nghĩa, cũng như những vấn đề về supervised machine learning; (2) Sự chú ý vào nghiên cứu được gia tăng dẫn đến những sự tác động qua lại ngày càng mật thiết với **cộng đồng Học máy và thống kê**. (3) Cuối cùng là sự phổ biến của những hệ thống điện toán với hiệu suất cao (high-performance computing

●



•

-



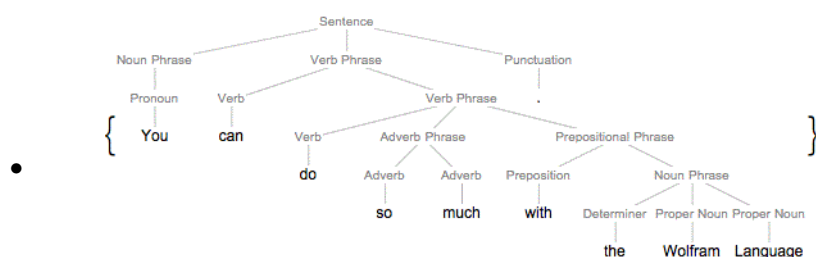
Hình 1: Sơ đồ thực hiện các bước trong NLP

•

•

- Hiểu ngôn ngữ tự nhiên (Natural language understanding - NLU) là một bài toán trong lĩnh vực Xử lý ngôn ngữ tự nhiên. NLU có liên hệ mật thiết tới công nghệ Đọc hiểu của máy (Machine reading comprehension).
- NLU là một bài toán AI-khó.

- Quá trình phân rã và phân tích input của bài toán NLU diễn ra rất phức tạp bởi sự xuất hiện ngoài ý muốn những yếu tố mơ hồ. Sự nhập nhằng đó dẫn tới nhu cầu phải tìm ra chính xác công thức ngữ pháp và ngữ cảnh phù hợp để áp dụng vào việc dự đoán nghĩa của input - yếu tố tiên quyết phải xác định được để cho ra output của bài toán.
- NLU mang lại những lợi ích thương mại rất đáng kể nhờ vào những ứng dụng hữu ích và thiết thực của nó: Thu thập tin tức (news-gathering), Phân loại văn bản (text categorization),...



Hình 2: Biểu diễn quá trình NLU dưới dạng cây

1.4.2 Sinh ngôn ngữ tự nhiên (Natural language generation)

- Sinh ngôn ngữ tự nhiên (Natural language generation - NLG) cũng là một bài toán trong lĩnh vực Xử lý ngôn ngữ tự nhiên. Hệ thống NLG là một công cụ phiên dịch sự biểu diễn bằng ngôn ngữ máy (chẳng hạn như các biểu thức logic) sang sự biểu diễn bằng ngôn ngữ tự nhiên.
- Bài toán NLG có thể được xem là đối nghịch với bài toán **hiểu ngôn ngữ tự nhiên** (NLU): Trong khi NLU nhận vào input là một đoạn văn bản bằng ngôn ngữ tự nhiên và nhiệm vụ nó làm là biểu diễn lại văn bản đó dưới dạng ngôn ngữ máy, thì một hệ thống NLG lại cần phải diễn giải một ý niệm trong máy tính thành lời văn.
- Hệ thống dự báo phần hoa cho Scotland là một ví dụ đơn giản về một hệ thống NLG. Hệ thống này nhận vào input 6 số nguyên dự báo lượng phần hoa cho các vùng khác nhau ở Scotland. Từ những số liệu đó, hệ thống sinh ra output là một văn bản thông báo về lượng phần hoa. Chẳng hạn như vào ngày 01/07/2005, phần mềm đã đưa ra thông báo sau đây: "Lượng phần hoa vào ngày thứ sáu đã tăng từ mức vừa phải lên mức cao với giá trị vào khoảng 6 hoặc 7 tại hầu hết những vùng trên cả nước. Tuy nhiên, tại những vùng ở phía Bắc, lượng phần hoa chỉ ở tầm vừa phải với giá trị là 4."
- Các giai đoạn thực hiện NLG bao gồm: **Xác định nội dung** (Content determination), **Xây dựng cấu trúc văn bản** (Document structing), **Kết hợp** (Aggregation), **Lựa chọn từ ngữ** (Lexical choice), **Sinh ra các biểu thức sở chỉ** (Referring expression generation) và cuối cùng là **Hiện thực hóa** (Realisation).

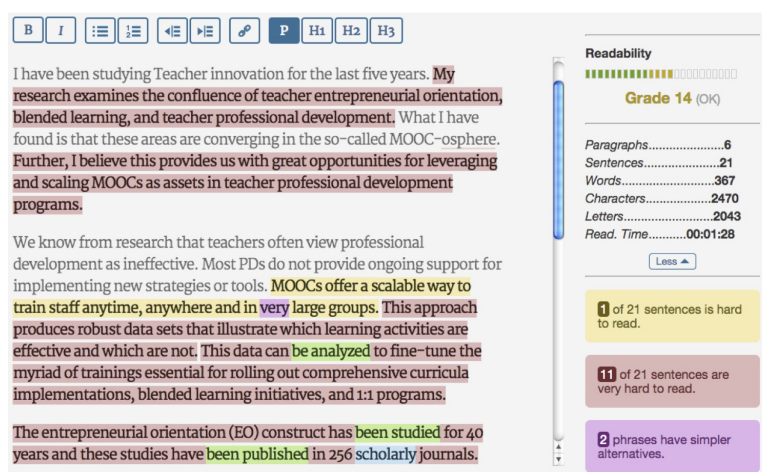
1.5 Ứng dụng của NLP

1.5.1 Tóm tắt văn bản tự động (Automatic summarization)

- Tóm tắt văn bản tự động là quá trình tối giản một tài liệu văn bản bằng một chương trình máy tính để tạo ra một bản tóm tắt ngắn gọn chỉ chứa những điểm cốt lõi, quan trọng trong văn bản gốc. Ý tưởng chính của quá trình tự động tóm tắt là tìm ra một **tập con đại diện** cho tập dữ liệu, sao cho nó có chứa **toàn bộ thông tin** của tập dữ liệu. Ngày nay, công nghệ tóm tắt văn bản tự động đã được áp dụng trên diện rộng trong các lĩnh vực công nghiệp, mà một ví dụ tiêu biểu là trong các máy tìm kiếm như Google.
- Có hai phương pháp chính để tiếp cận bài toán tóm tắt văn bản tự động là **trích xuất** (extraction) và **tóm lược** (abstraction). Phương pháp trích xuất hoạt động bằng cách chọn ra một tập con gồm những từ, cụm từ hoặc câu trong văn bản gốc để xây dựng nên bản tóm tắt. Ngược lại, phương pháp tóm lược đi tìm bản chất ngữ nghĩa của văn bản gốc, sau đó sử dụng công nghệ **sinh ngôn ngữ tự nhiên** để kết xuất nên bản tóm tắt.

1.5.2 Chấm điểm tự động (Automated essay scoring)

- Chấm điểm tự động (Automated essay scoring - AES) là hành động sử dụng một chương trình máy tính được chuyên biệt hóa để tính điểm cho những **bài luận** (essay) được viết ra trong một môi trường giáo dục như trường Trung học hoặc Đại học.
- AES là một ứng dụng của việc Xử lý ngôn ngữ tự nhiên (NLP) và là một phương pháp để tự động hóa quá trình **thẩm định giáo dục** (Educational assessment).



Hình 3: Hemmingway App - Một chương trình AES tiêu biểu để cho điểm một bài luận

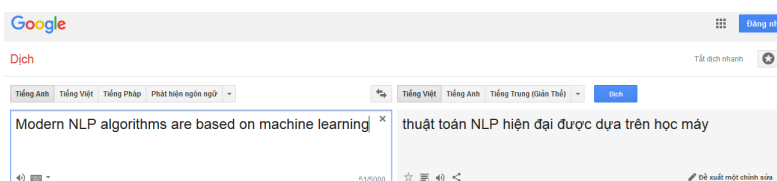
- Cha đẻ của công nghệ AES là Ellis Batten Page. Năm 1966, ông đã lập luận chứng minh cho khả năng chấm điểm bài luận bằng máy tính. Hai

năm sau, Page hoàn thành và công bố một chương trình tên là **Project Essay Grade** (PEG). Nhưng chỉ bằng những công nghệ thời đó, việc chấm điểm tự động kéo theo một chi phí lớn để hoạt động. Do đó ông nhanh chóng nản lòng và trì trệ công việc phát triển của mình suốt hai thập niên tiếp theo.

- Nguyên lý hoạt động của các chương trình AES hiện nay có liên quan mật thiết đến lĩnh vực **Học máy** (Machine learning - ML): Bước đầu AES đọc vào một **tập huấn luyện** (training set - một khái niệm trong ML) chứa những bài luận mẫu đã được cho điểm sẵn (bởi con người), sau đó chương trình tiến hành khái quát hóa những **đặc tính** nổi bật của mỗi bài luận (như số lượng từ ngữ, số mệnh đề phụ thuộc, hoặc tỉ lệ giữa ký tự viết thường và viết hoa...) mà không cần sự chỉ dẫn của con người. Tiếp theo, chương trình AES xây dựng một **model toán học** để liên kết các **đặc tính** trên với điểm số mà bài luận đó nhận được. Và cuối cùng, nó sẽ áp dụng chính **model** đó để tính toán ra số điểm cho các bài luận mới.

1.5.3 Dịch tự động (Machine translation)

- Dịch tự động (Machine translation - MT) là một bài toán trong lĩnh vực NLP, nghiên cứu việc dịch chính xác một văn bản hoặc một bài nói chuyện từ ngôn ngữ này sang một ngôn ngữ khác.
- Ở mức độ sơ đẳng nhất, MT thực hiện việc thay thế từng từ trong văn bản cần dịch bởi từ đồng nghĩa tương ứng ở ngôn ngữ kia (word-by-word). Cách làm này rất đơn giản nhưng nó không bao giờ cho ta một bản dịch tốt được (bởi nó không nhận diện và dịch được các **cụm từ** (phrase) vốn xuất hiện rất nhiều trong mọi thể loại văn bản).
- Những thành tựu nghiên cứu trong lĩnh vực NLP đã được áp dụng để cải thiện chất lượng cho lĩnh vực Machine translation. Một bước đi tiêu biểu là sử dụng kỹ thuật neural (neural technique), một hướng tiếp cận từ công nghệ **Học sâu** (Deep learning).



Hình 4: Google Translate - một công cụ dịch tự động nổi tiếng

1.5.4 Phát hiện thư rác (Spam detection)

- **Thuật toán Naive Bayes**: là một kĩ thuật thống kê phổ biến của bộ lọc email. Naive Bayes phân loại công việc bằng tương quan trong việc sử

dùng các thẻ với spam hay non-spam emails rồi thì dùng định lý Bayes để tính toán xác suất xem có phải là một spam email hay không:

$$\bullet P(F|E) = \frac{P(E|F)P(F)}{P(E|F)P(F) + P(E|\bar{F})P(\bar{F})}$$

- **Máy vecto hỗ trợ (support vector machine):** Đây là một khái niệm trong thống kê và khoa học máy tính cho một tập hợp các phương pháp liên quan đến nhau để phân loại và phân tích hồi quy. Xét về tổng thể, các lỗi mô hình này ít hơn so với Naive Bayes, các kiểm tra lỗi cũng tốt hơn. Trong khi SVM làm tăng tính phức tạp của mô hình, kết quả cho thấy không có lợi so với thuật toán Bayes.
- **K-nearest neighbor:** Trong phương pháp này, nhân cho một mẫu thử nghiệm được dự đoán dựa trên đa số phiếu của k láng giềng gần nó nhất. Thuật toán K-NN là một trong những thuật toán học máy đơn giản nhất. Các thuật toán K-NN là phương pháp phi tham số được sử dụng để phân loại và hồi quy. Trong phân loại K-NN, đầu ra là một thành phần của lớp. Một đối tượng được phân loại bởi đa số phiếu của các láng giềng. Trong hồi quy, đầu ra là giá trị tài sản cho đối tượng. Giá trị này là trung bình của các giá trị của k láng giềng gần nhất của nó.



1.5.5 Nhận dạng giọng nói (Speech recognition)

- Trước khi máy tính hiểu được những ngôn ngữ con người thì nó cần phải nghe được những gì ta nói trước đã. Điều này liên quan đến một quá trình phức tạp bao gồm lấy mẫu âm thanh, khai thác các tính năng và rồi thì nhận ra âm thanh cá nhân sau đó chuyển chúng thành văn bản.
- Các nhà khoa học học đã phát triển kĩ thuật khai thác các tính năng một cách tương tự như tai người và nhận ra chúng là các âm sắc và âm thanh như của con người. Điều này liên quan tới việc sử dụng các mạng thần kinh nhân tạo và là một phần trong lĩnh vực trí tuệ nhân tạo.
- Mặc dù với những tiến bộ trên, việc nghe được giọng nói của bạn mới chỉ là một nửa của vấn đề. Một khi máy tính xử lý quá trình này, nó chỉ có thể sao chép lại những gì bạn nói. Bước tiếp theo chính là xử lý ngôn ngữ tự nhiên, khi máy tính đã chuyển những gì bạn nói thành văn bản, nó sẽ thật sự hiểu những gì bạn nói. Đây được xem là khó khăn hơn so với quá trình nhận dạng giọng nói bởi vì ngôn ngữ con người đầy bối cảnh và ngữ nghĩa làm cho quá trình xử lý khó khăn hơn.

- **Ứng dụng trong gia đình và văn phòng:** Ứng dụng bao gồm nhận dạng cá nhân để truy cập vào các tòa nhà, máy tính và điều khiển bằng giọng nói của các hệ thống máy tính như cơ sở dữ liệu.
- **Ứng dụng trong quân sự:** Công nghệ này rất hữu ích cho việc kiểm soát “rảnh tay” các hệ thống vũ khí.
- Ngoài ra, trong Ô-tô, các công cụ Speech recognition kiểm soát được các hệ thống ngoại vi như điện thoại và hệ thống âm nhạc, điều khiển bằng giọng nói của thiết bị trợ giúp định hướng.



Hình 5: Trợ lý ảo Siri trên iPhone có thể nhận dạng và xử lý được giọng nói từ người dùng

2 Khó khăn trong lĩnh vực NLP

2.1 Sự nhập nhằng ngữ nghĩa (semantic ambiguity)

Xét ví dụ sau:

- Con sói bông **lồng** lên.
- Con chim bay ra khỏi **lồng**.

Đây là một **sự nhập nhằng về từ loại**, từ **lồng** vừa có là một động từ, vừa có thể là một danh từ. Nếu gặp phải trường hợp này thì sẽ phát sinh nhiều khó khăn trong việc gán nhãn từ loại.

Một ví dụ khác:

- Hổ mang bò lên núi.

Đây là **sự nhập nhằng về cú pháp**. Cùng một câu nói nhưng lại có thể có 2 cách hiểu khác nhau:

- Con rắn hổ mang đang bò lên núi.
- Con hổ mang con bò lên núi.

Sự nhập nhằng về cú pháp khiến hai bước **phân tích cú pháp** và **phân tích ngữ pháp** trong NLP trở nên phức tạp hơn rất nhiều.

2.2 Sự đa dạng của ngôn ngữ (language variability)

Một ví dụ tiêu biểu:

- "Trường Đại học Bách khoa nằm ở đâu vậy?"
- "Bạn có thể chỉ đường cho tôi đến Đại học Bách khoa được không?"

Hai câu hỏi trên rõ ràng là có cùng ý nghĩa. Nhưng để các thuật toán NLP xác định được điều đó thì cần rất nhiều nỗ lực trong việc nghiên cứu thiết kế bởi vì cách viết của chúng không hề giống nhau.

2.3 Kiến trúc phần cứng (hardware architecture)

- Các nhà nghiên cứu đã áp dụng các thành tựu trong lĩnh vực trí tuệ nhân tạo (Artificial Intelligence - AI) để cải thiện lời giải cho một số bài toán NLP. Tuy nhiên, mức độ cao nhất trong sự thấu hiểu ngôn ngữ vẫn phải liên hệ đến giới hạn của não bộ về việc xử lý những câu nói phức tạp. Điều này đòi hỏi khả năng trừu tượng hóa và liên kết thông tin, đồng thời phải truy xuất lại nhiều ký ức, kỷ niệm để dò ra nghĩa. Chỉ một việc đưa ra quyết định cuối cùng đã phải cần tới một tiến trình gồm hàng chục tỉ biến nhớ (và phải được xử lý trong thời gian nhận thức thực tế). Và một thực trạng là các cấu trúc phần cứng máy tính hiện nay của chúng ta chỉ dựa trên những chip vi xử lý thậm chí còn không thể thực hiện những tác vụ đơn giản nhất của bộ não.



Hình 6: Siêu máy tính K-Computer mất đến 40 phút để mô phỏng...1 giây hoạt động của não người

3 Tài liệu tham khảo

- Hutchins, J. (2005). "The history of machine translation in a nutshell".

- "Versatile question answering systems: seeing in synthesis", Int. J. of Intelligent Information Database Systems, 119-142, 2011.
- Kai Wei, Rishabh Iyer, and Jeff Bilmes, Submodularity in Data Subset Selection and Active Learning, To Appear In Proc. International Conference on Machine Learning (ICML), Lille, France, June - 2015
- Knight, Kevin. "Building a large ontology for machine translation (1993)"
- David A. Wheeler, (May 11, 2011) Countering Spam by Using Ham Passwords (Email Passwords)
- "Anti-spam", NZ Department of Internal Affairs
- Juang, B. H.; Rabiner, Lawrence R. "Automatic speech recognition—a brief history of the technology development"
- <https://www.quora.com/What-are-the-major-open-problems-in-natural-language-understanding>

4 Đóng góp của các thành viên

- Lê Thị Thúy Hằng - 1611947 - Đóng góp 25%
- Đoàn Nguyễn Thành Lương - 1611947 - Đóng góp 30%
- Nguyễn Trương Ngọc Bách - 1611947 - Đóng góp 25%
- Võ Khắc Trí - 1611947 - Đóng góp 25%