

How Watson reads a clue

Nguyễn Tuấn Linh



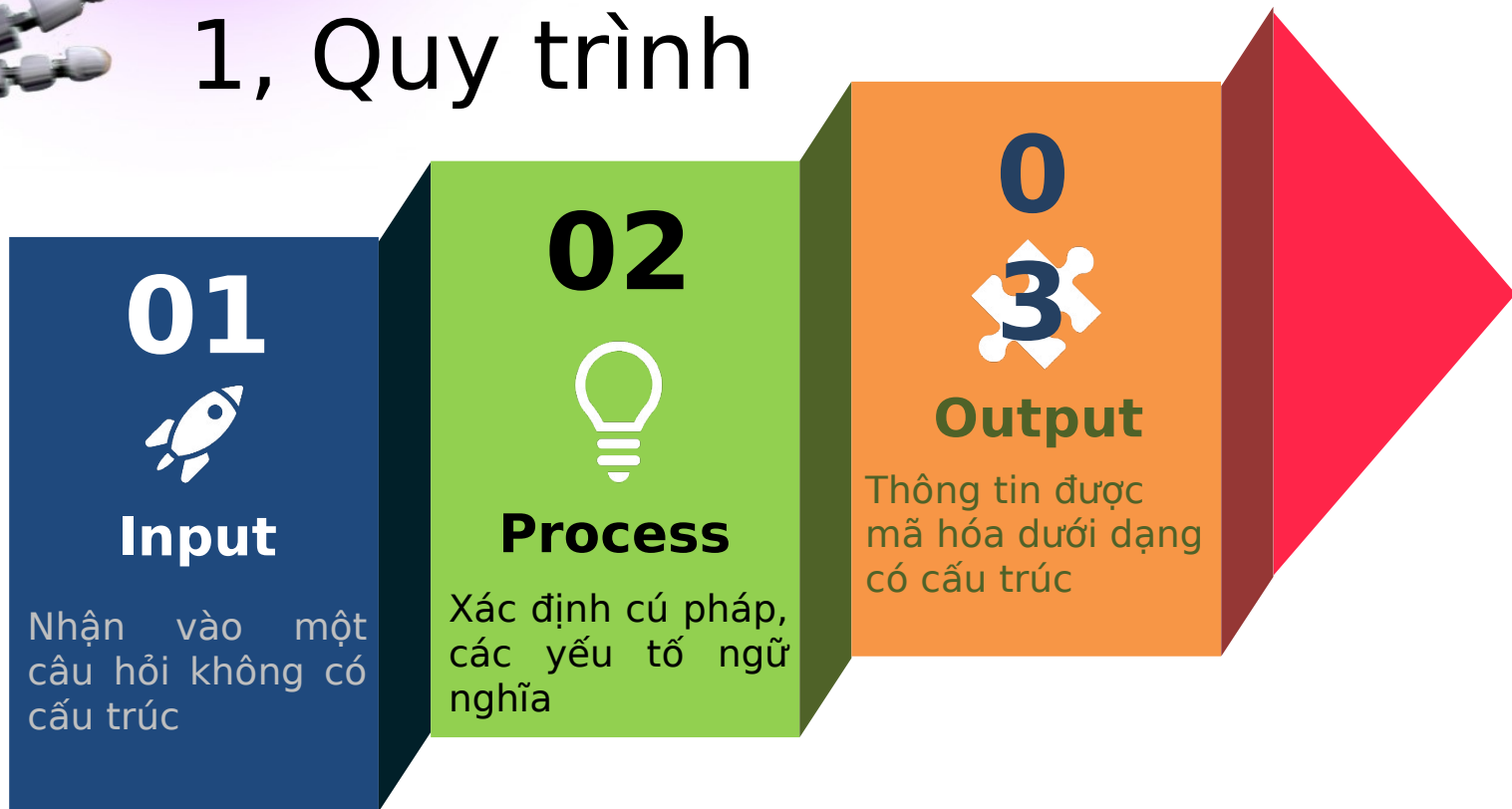
Đặt vấn đề

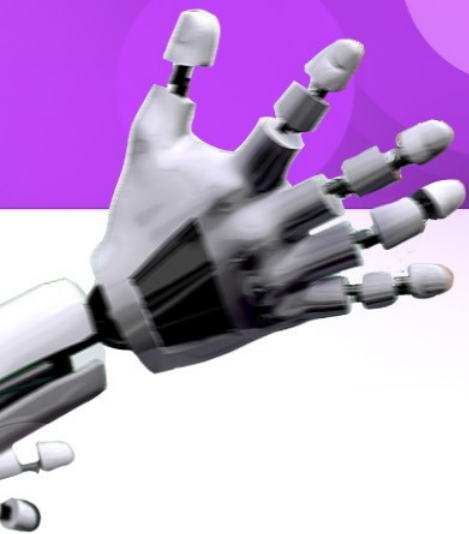
- Xác định xem câu hỏi yêu cầu gì
- Tìm cách tốt nhất để tiếp cận trả lời câu hỏi
- Tùy biến cần thiết để tham dự chương trình *“Jeopardy!”*



I, Giới thiệu chung

1, Quy trình

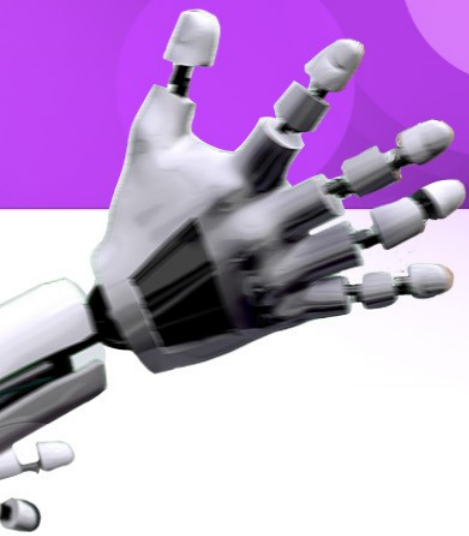




I, Giới thiệu chung

2, Các yếu tố quan trọng

- Focus
- Lexical answer types (LATs)
- Question classification
- Question sections



I, Giới thiệu chung

2, Các yếu tố quan trọng

Ví dụ:

POETS & POETRY: He was a bank clerk in the Yukon before he published “Songs of a Sourdough” in 1907

(*Tạm dịch:* CHỦ ĐỀ THI SĨ VÀ THƠ: Anh ấy là nhân viên ngân hàng ở Yukon trước khi anh xuất bản “Songs of a Sourdough” năm 1907)



1, Giới thiệu chung

2, Các yếu tố quan trọng

Trọng tâm là phần của câu hỏi mà có
tham chiếu đến câu trả lời

Trong ví dụ trên, trọng tâm là **“he”**



I, Giới thiệu chung

2, Các yếu tố quan trọng

LATs là thuật ngữ cho biết loại thực thể nào đang được yêu cầu trong câu hỏi

- Từ đầu tiên của trọng tâm thường là một LAT
- Trong ví dụ trên, LATs là **“he”**, **“clerk”**, và **“poet”**



1, Giới thiệu chung

2, Các yếu tố quan trọng

Question Classification (phân loại câu hỏi) xác định câu hỏi là thuộc một hoặc nhiều loại

- ***Defination*** (xác định), ***Multiple- Choice*** (nhiều lựa chọn), ***Puzzle*** (câu đố), ***Fill-in-the-Blanks*** (điền từ vào chỗ trống), và ***Abbreviation*** (viết tắt),...
- Câu hỏi trong ví dụ trên là loại câu hỏi “Defination”



I, Giới thiệu chung

2, Các yếu tố quan trọng

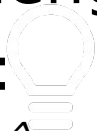
QSections (mảnh câu hỏi) là các phần nhỏ của câu hỏi có cách diễn giải yêu cầu xử lý đặc biệt

- Xác định các ràng buộc từ vựng trên câu trả lời (ví dụ: “4-letter” hoặc “3-word”)
- Phân tích câu hỏi thành nhiều câu hỏi con



II, Nền tảng của phân tích câu hỏi

1, Phân tích cú pháp và ngữ nghĩa

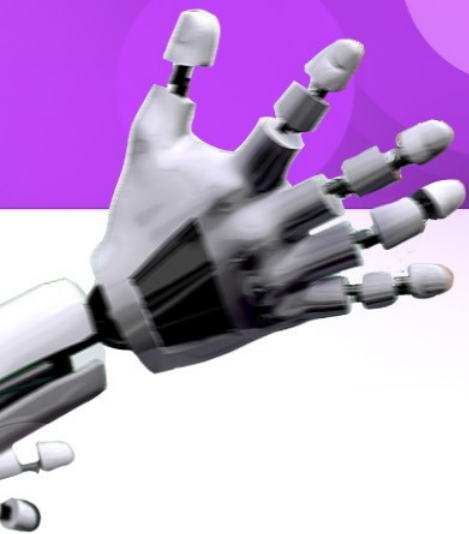
- Được tạo ra từ trình phân tích cú pháp ngữ pháp Tiếng Anh
- Kết hợp với:
 1. Thành phần xây dựng cấu trúc đối số vị từ liên quan (PAS)
 2. Trình nhận dạng các đối tượng đã được đặt tên (named entity recognizer)
 3. Thành phần đồng tham chiếu (co-reference resolution component)



II, Nền tảng của phân tích câu hỏi

1, Phân tích cú pháp và ngữ nghĩa

- Được sử dụng để phân tích từng câu trong một câu hỏi thành một cây
- Mỗi nút của cây chứa:
 1. Một từ hoặc một thuật ngữ đa từ với một vị từ liên quan và logic lập luận của nó
 2. Một danh sách các tính năng, một số hình thái, những ngữ nghĩa khác
 3. Các bổ nghĩa bên trái và bên phải của nút, cũng như vị trí mà nó được điền



II, Nền tảng của phân tích câu hỏi

2, Tùy biến phân tích câu hỏi

Các câu hỏi trong chương trình “Jeopardy!” khác với các câu hỏi trong đời sống thường ngày



=> Cần chỉnh sửa cách tiếp cận, xử lý



II, Nền tảng của phân tích câu hỏi

2, Tùy biến phân tích câu hỏi

a) Toàn bộ các câu hỏi được in hoa -> Khó phát hiện tên riêng

=> Áp dụng thành phần thống kê true-caser



b) Không sử dụng đại từ “wh” mà dùng “this/these” và “he/she/it”

=> Sửa đổi trình phân tích cú pháp



II, Nền tảng của phân tích câu hỏi

2, Tùy biến phân tích câu hỏi

c) các câu hỏi trong chương trình
“Jeopardy!” thường bao gồm một đại từ



=> **Chỉnh sửa thành phần đồng tham chiếu**



II, Nền tảng của phân tích câu hỏi

2, Tùy biến phân tích câu hỏi

Ví dụ: “Astronaut Dave Bowman is brought back to life in his recent novel 3001: Final Odyssey”



* **“his”** đề cập đến câu trả lời (Arthur C. Clarke), không phải Dave Bowman

=> tìm kiếm trọng tâm trước khi chạy thành phần đồng tham chiếu



III, Phát hiện trọng tâm và LAT

1, Các quy tắc cơ bản

- Phát hiện trọng tâm (focus):
 - ❖ Một cụm từ danh từ được quyết định với “this” hoặc “these”
 - ❖ “This” hoặc “these” như một đại từ
 - ❖ Khi câu hỏi là một cụm từ danh từ, dán nhãn toàn bộ câu hỏi làm trọng tâm
 - ❖ Một trong những đại từ “he/she/his/her/him/hers”
 - ❖ Một trong những đại từ “it/they/them/its/their”
 - ❖ Đại từ “one”

**** Khi không có trường hợp nào đúng, câu hỏi có thể không có trọng tâm**

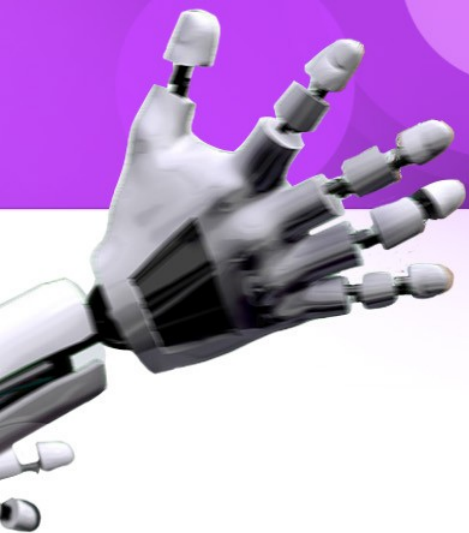


III, Phát hiện trọng tâm và LAT

1, Các quy tắc cơ bản

- Phát hiện LAT: Phương pháp phát hiện LAT cơ bản thường chọn trọng tâm là LAT duy nhất, với các ngoại lệ sau





III, Phát hiện trọng tâm và LAT

1, Các quy tắc cơ bản

- ❖ Nếu trọng tâm là một sự kết hợp, trích xuất các liên kết
- ❖ Trích xuất LAT khi trọng tâm là one/name/type/kind
- ❖ Trích xuất LAT khi trọng tâm là name/word/term
- ❖ Nếu không có trọng tâm nào được phát hiện và danh mục là một giai đoạn danh từ, lấy tiêu đề của danh mục là LAT

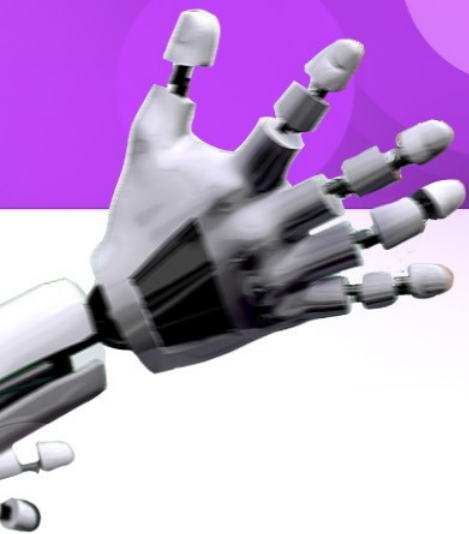


III, Phát hiện trọng tâm và LAT

2, Một số cải tiến

Các quy tắc cơ bản khá chính xác, nhưng cũng có một số sai sót

=> Có rất nhiều LAT hữu ích mà không được phát hiện



III, Phát hiện trọng tâm và LAT

2, Một số cải tiến

Ví dụ 1:

PAIRS: An April 1997 auction of Clyde Barrow's belongings raised money to fund moving his grave next to ***hers***.

* Quy tắc cơ sở tập trung vào “**his**” đầu tiên, nhưng không chính xác.

=> chọn đại từ không ràng buộc với bất kỳ thực thể nào làm trọng tâm, sau đó sử dụng thành phần giải quyết trùng lặp



III, Phát hiện trọng tâm và LAT

2, Một số cải tiến

Ví dụ 2:

FAMOUS AMERICANS: Although he made no campaign speeches, *he* was elected **president** in 1868 by a wide electoral margin

* Quy tắc cơ sở sẽ chọn “**FAMOUS AMERICANS**”

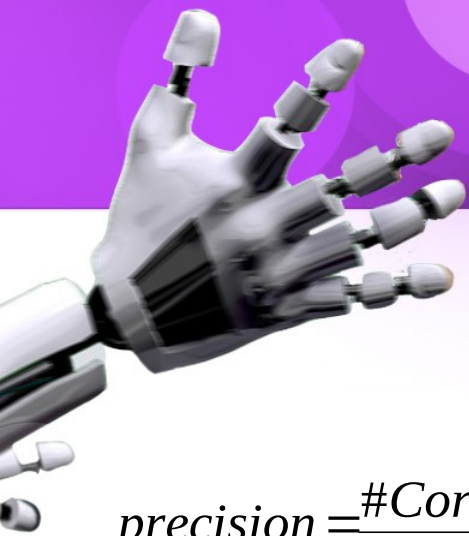
=> chọn các vai trò của tiêu điểm



III, Phát hiện trọng tâm và LAT

3, Đánh giá

- Sử dụng 9128 câu hỏi được tập hợp theo cách thủ công
- Xác thực chéo 10 lần để đào tạo và đánh giá phân loại LAT



III, Phát hiện trọng tâm và LAT

3, Đánh giá

$$\text{precision} = \frac{\# \text{Correctly Detected LATs}}{\# \text{Detected LATs}}$$

$$\text{recall} = \frac{\# \text{Correctly Detected LATs}}{\text{LATs in Manualy Annotated Set}}$$

$$F_1 = \frac{2(\text{precision})(\text{recall})}{\text{precision} + \text{recall}}$$

$$\text{Per Question Recall} = \frac{\# \text{Questions with at least one correctly detected LAT}}{\# \text{Questions with at least one manually annotated LAT}}$$



III, Phát hiện trọng tâm và LAT

3, Đánh giá

	Baseline	Watson
Precision	0.817	0.829
Recall	0.613	0.766
F_1	0.700	0.796
Per	0.840	0.905
Question		
Recall		

Bảng 1: Đánh giá tỉ lệ phát hiện LAT



IV, Phân loại câu hỏi và phát hiện QSections

1, Phân loại câu hỏi

Trong chương trình “Jeopardy!”, có những câu hỏi khác với câu hỏi thực tế




=> Phân loại để tìm cách trả lời thích hợp

IV, Phân loại câu hỏi và phát hiện QSections

1, Phân loại câu hỏi

Các loại câu hỏi và cách phát hiện

- Puzzle
 - Bond
 - Fill-in-the-blank
 - Multiple-choice
 - Khác
- 
- | | | |
|---|--|--------------------------------------|
| } | | Các biểu thức
chính quy cơ
bản |
| } | | Ký hiệu thông
dụng |
| } | | Được cung cấp chuỗi danh
mục |
| } | | Quy tắc cú pháp PAS |



IV, Phân loại câu hỏi và phát hiện QSections

2, Qsections

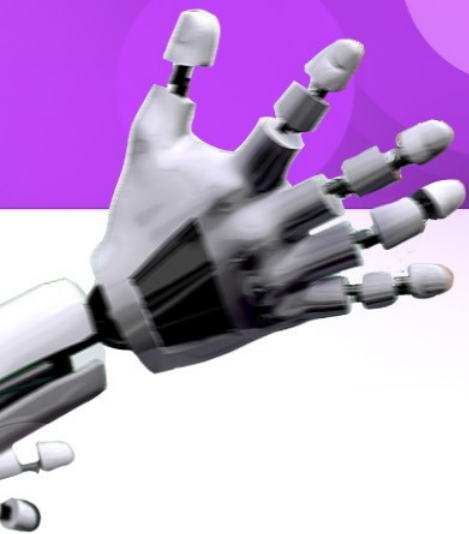
Một QSection là một chú thích được ghi trên một nội dung tiếp giáp với văn bản của câu hỏi (đôi khi, nó nằm trong danh mục)

Ví dụ:



- This **7-letter word** means to presage or forebode
- **12-LETTER WORDS:** Leavenworth, established in 1895, is a federal one.

* Nó rất quan trọng trong việc trả lời câu hỏi



IV, Phân loại câu hỏi và phát hiện QSections

3, Đánh giá

- Đánh giá dựa trên 3500 câu hỏi được phân loại thủ công



- So sánh toàn bộ hệ thống Watson với một hệ thống cơ bản để đo lường tác động tới toàn hệ thống trả lời câu hỏi



IV, Phân loại câu hỏi và phát hiện QSections

3, Đánh giá

QClass	Precision	Recall	F_1
FITB	0.676	0.711	0.693
PUZZLE	0.977	0.525	0.683
BOND	1.000	0.652	0.789
MULTIPLE-CHOICE	0.650	0.684	0.667
DATE	0.692	0.818	0.750
All	0.646	0.629	0.637

Bảng 2: Đánh giá khả năng phân loại câu hỏi



IV, Phân loại câu hỏi và phát hiện QSections

3, Đánh giá

Focus/LAT	QClass/ Qsection	Độ chính xác của hệ thống trả lời
Baseline	Không	65.1
Baseline	Có	67.5
Full	Không	68.1
Full	Có	71.0

Bảng 3: Ảnh hưởng đến hệ thống trả lời câu hỏi



Kết luận

- Trình bày quy trình mà Watson phân tích câu hỏi
- Một số tùy biến để tham gia chương trình “Jeopardy!”
- Đánh giá hiệu suất, độ chính xác của Watson trong trả lời câu hỏi