

# **Semantic Search With NLP And Elasticsearch**

---

**Planning Team**

**2020/04**

# **AGENDA**

1. Philosophy of Semantic Search
2. Semantic Search For Recruiting
3. NLP (Natural Language Processing)
4. Elasticsearch
5. Q&A



# **1. Philosophy of Sematic Search**

Why Semantic Search?



# 1. Philosophy of Sematic Search

Keyword: “harry potter”

×
🔍

🔍 All
🖼 Images
📺 Videos
📰 News
📖 Books
⋮ More
Settings
Tools

About 478,000,000 results (1.03 seconds)

www.wizardingworld.com ▾

**Wizarding World - the official home of Harry Potter**

Official home of **Harry Potter** & Fantastic Beasts. Discover your Hogwarts House, create a Wizarding Passport & unlock more magic with Wizarding World Gold.

[Introducing Harry Potter At Home](#) · [Harry Potter at Home](#) · [Features](#) · [Books](#)

www.wizardingworld.com › collections › harry-potter-... ▾

**Harry Potter at Home - Wizarding World**

Welcome to the **Harry Potter** At Home hub where you'll find all the latest magical treats to keep you occupied - including special contributions from Bloomsbury ...

en.wikipedia.org › wiki › Harry\_Potter ▾

**Harry Potter - Wikipedia**

**Harry Potter** is a series of fantasy novels written by British author J. K. Rowling. The novels chronicle the lives of a young wizard, **Harry Potter**, and his friends Hermione Granger and Ron Weasley, all of whom are students at Hogwarts School of Witchcraft and Wizardry.


**No. of books:** 7 **Publisher:** Bloomsbury Publishing (UK); Potter...

**Published:** 26 June 1997 – 21 July 2007 (initial ... **Country:** United Kingdom

[Harry Potter \(film series\)](#) · [Harry Potter \(character\)](#) · [Harry Potter \(disambiguation\)](#)

People also ask

Did Harry Potter win any Oscars? ▾



**Harry Potter**

Novel series

Harry Potter is a series of fantasy novels written by British author J. K. Rowling. The novels chronicle the lives of a young wizard, Harry Potter, and his friends Hermione Granger and Ron Weasley, all of whom are students at Hogwarts School of Witchcraft and Wizardry.

[Wikipedia](#)

**Author:** [J. K. Rowling](#)

**First book:** [Harry Potter and the Philosopher's Stone](#)

**Characters:** [Hermione Granger](#), [Harry Potter](#), [Lord Voldemort](#), [MORE](#)

**Genres:** Fantasy Fiction, Drama, Young adult fiction, Mystery, Thriller, Bildungsroman

**Sport:** [Quidditch](#)





# 1. Philosophy of Sematic Search

Keyword: “harry potter characters”

harry potter characters

All Images Videos News Shopping More Settings Tools

Harry Potter > Characters

Hermione Granger	Ron Weasley	Luna Lovegood	Dolores Umbridge	Gellert Grindelwald	Hedwig
Harry Potter	Professor Severus Snape	Neville Longbottom	Bellatrix Lestrange	Ginny Weasley	Sorting H
Lord Voldemort	Professor Albus Dumbledore	Dobby the House Elf	Sirius Black	Peter Pettigrew	Newt Sca
Draco Malfoy	Rubeus Hagrid	Professor Minerva McGonagall	Remus Lupin	Cho Chang	Nagini

en.wikipedia.org > wiki > List\_of\_Harry\_Potter\_charact...

List of Harry Potter characters - Wikipedia

This is a list of **characters** in the **Harry Potter** series. They are all **characters** who have appeared ... **Hermione Granger** – Muggle-born Gryffindor student in Harry's year and one of his best friends. ... **Draco Malfoy** – Slytherin student in Harry's year. ... Predicted the prophecy that prompted **Lord Voldemort** to go after the Potters.

Characters by surname · P · W · Characters with no or ...


People also ask

How many characters are in Harry Potter?

Who are the three main characters in Harry Potter?

What are all the characters names in Harry Potter?

Who are the bad characters in Harry Potter?



Harry Potter

Novel series

Harry Potter is a series of fantasy novels written by British author J. K. Rowling. The novels chronicle the lives of a young wizard, Harry Potter, and his friends Hermione Granger and Ron Weasley, all of whom are students at Hogwarts School of Witchcraft and Wizardry.

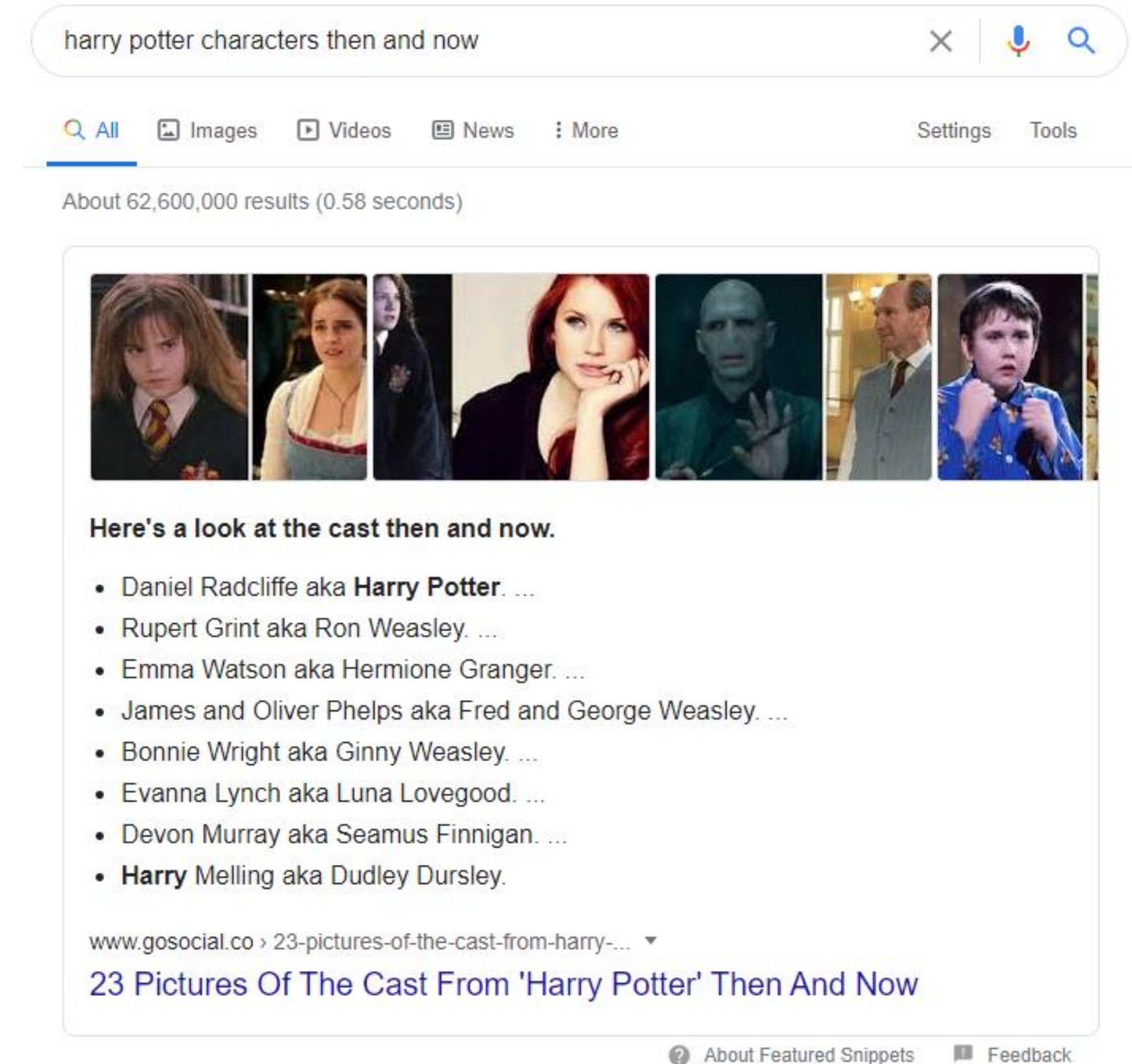
Wikipedia





# 1. Philosophy of Sematic Search

Keyword: “harry potter characters  
then and now”



harry potter characters then and now

All Images Videos News More Settings Tools

About 62,600,000 results (0.58 seconds)

Here's a look at the cast then and now.

- Daniel Radcliffe aka **Harry Potter**. ...
- Rupert Grint aka Ron Weasley. ...
- Emma Watson aka Hermione Granger. ...
- James and Oliver Phelps aka Fred and George Weasley. ...
- Bonnie Wright aka Ginny Weasley. ...
- Evanna Lynch aka Luna Lovegood. ...
- Devon Murray aka Seamus Finnigan. ...
- **Harry** Melling aka Dudley Dursley.

www.gosocial.co > 23-pictures-of-the-cast-from-harry-...  
23 Pictures Of The Cast From 'Harry Potter' Then And Now

About Featured Snippets Feedback

## Videos



Harry Potter Cast Then and Now

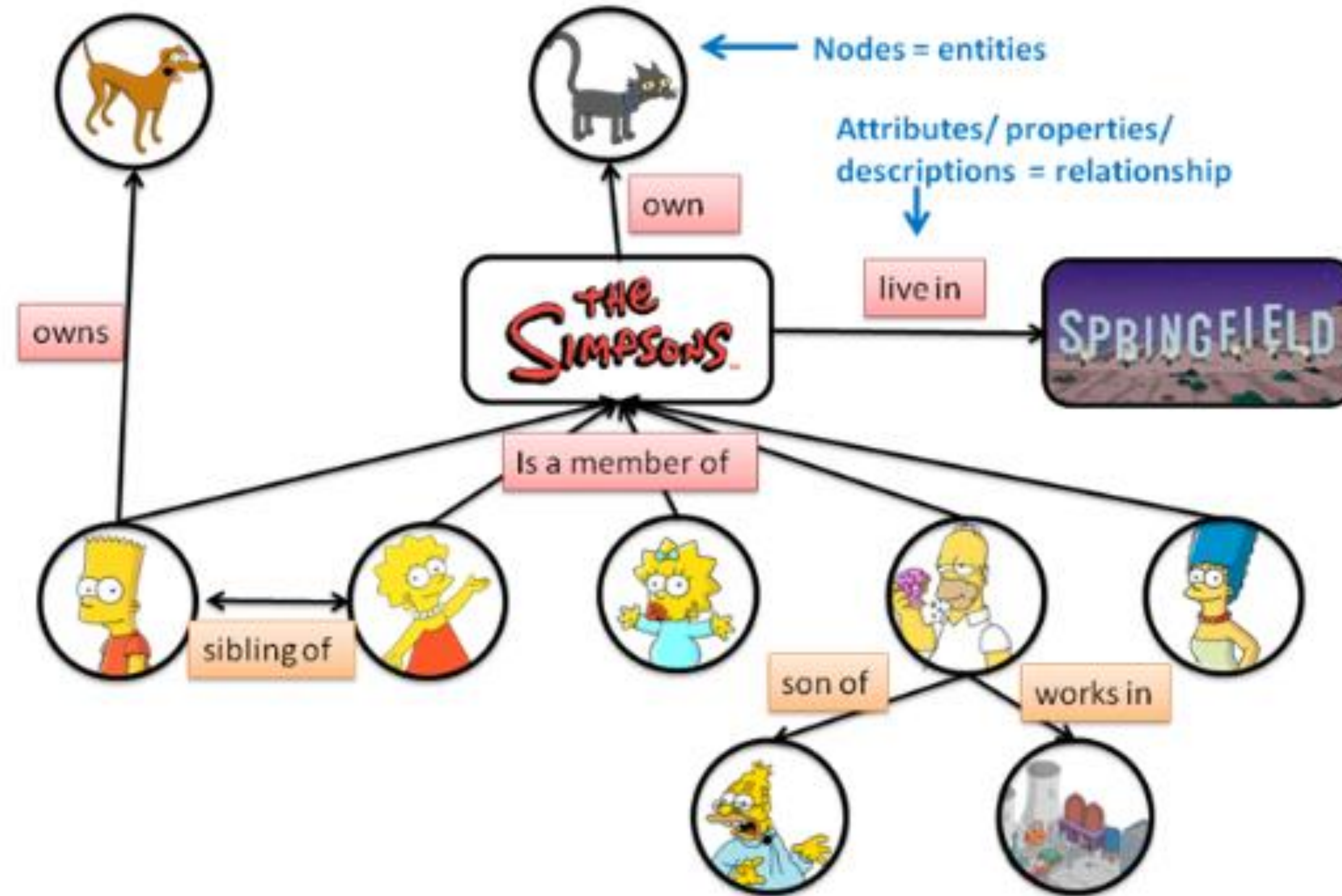
Harry Potter (2001) Cast Then And Now 2020

Harry Potter Cast Then and Now | 2019 |





# 1. Philosophy of Sematic Search




# **AGENDA**

1. Philosophy of Semantic Search
- 2. Semantic Search For Recruiting**
3. NLP (Natural Language Processing)
4. Elasticsearch
5. Q&A






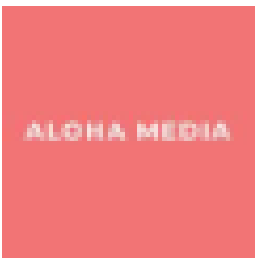
## 2. Semantic Search For Recruiting




**Nhân viên Quan hệ khách hàng doanh nghiệp**  
 Công ty Cho thuê tài chính TNHH MTV ...  
 💰 Thỏa thuận 📍 Hà Nội, Hồ Chí Minh




**[Toàn quốc] Chuyên viên Tư vấn Tài chính**  
 Công Ty TNHH Bảo Hiểm Nhân Thọ AIG ...  
 💰 Trên 6 triệu 📍 Hà Nội, Hồ Chí Minh




**Nhân Viên Trục Page Online (Part time)**  
 CÔNG TY CỔ PHẦN TRUYỀN THÔNG ...  
 💰 7-16 triệu 📍 Hà Nội



**Nhân Viên Tư vấn Tuyển sinh (Lương cao)**  
 Công ty cổ phần đào tạo Công nghệ Trẻ ...  
 💰 7-15 triệu 📍 Hà Nội, Hồ Chí Minh




**Senior Social Content**  
 CÔNG TY CỔ PHẦN TV HUB  
 💰 13-18 triệu 📍 Hồ Chí Minh



**Nhân Viên Kinh Doanh Máy Móc Kỹ Thuật**  
 Công ty TNHH TMDV DTC  
 💰 10-20 triệu 📍 Hồ Chí Minh



**Lê Phương Anh** Đang tìm việc  
 🏢 Department of Accountance - FPT Software  
 🏢 Sale of portable cosmetics - Online Business  
 🎓 - Academy of Finance  
 📍 Hà Nội



**Khúc Thị Hồng Ngọc** Đang tìm việc  
 🏢 Nhân viên bán hàng ( part-time ) - Công ty TNHH An Nhiên  
 🏢 Nhân viên PG, hỗ trợ sự kiện ( Part-time ) - Công ty M-Tech  
 🎓 Chuyên ngành: Khối D (Toán, Văn, Anh) - Trung học phổ thông Kim Liên  
 🎓 Chuyên ngành: Quản lý - Khoa quốc tế - Đại học Quốc Gia Hà Nội  
 📍 Hà Nội



**Nguyễn Thị Vân Anh**  
 🏢 Nhân viên bán hàng - Công Ty Cổ Phần Dịch Vụ Tổng Hợp Thái Minh : Kiến  
 Hương Tích (Xã Thiên Lộc...  
 🏢 Lễ tân - Viện Thẩm Mỹ Bác Sĩ Hoàng  
 🎓 Chuyên ngành : Quản trị Dịch vụ Du lịch và Lữ hành - ĐẠI HỌC HÀ TĨNH  
 📍 Hà Tĩnh



# **AGENDA**

1. Philosophy of Semantic Search
2. Semantic Search For Recruiting
- 3. NLP (Natural Language Processing)**
4. Elasticsearch
5. Q&A



# **3. Natural Language Processing**

## 3.1 Mô tả mạng kỹ năng

## 3.2 Thu thập và xử lý dữ liệu

- ❖ Dữ liệu huấn luyện
- ❖ Dữ liệu ứng viên

## 3.3 Xác định các kỹ năng tương đương

## 3.4 Phương pháp xếp hạng ứng viên

- ❖ Xếp hạng ứng viên theo kỹ năng
- ❖ Xếp hạng ứng viên theo kỹ năng kết hợp thêm các yếu tố khác





# 3. Natural Language Processing

## 3.1 Mô tả mạng kỹ năng

- ❖ Distributional Semantics cố gắng tìm ra ý nghĩa của một từ thông qua các ngữ cảnh khác nhau mà từ ấy xuất hiện. Ý tưởng cơ bản nhất của Distributional Semantics là dựa trên một giả thiết (Distributional Hypothesis): "Các từ có ngữ cảnh tương tự với nhau sẽ có ý nghĩa tương tự nhau".

- ❖ VD:



# **3. Natural Language Processing**

## 3.1 Mô tả mạng kỹ năng

## 3.2 Thu thập và xử lý dữ liệu

- ❖ Dữ liệu huấn luyện
- ❖ Dữ liệu ứng viên

## 3.3 Xác định các kỹ năng tương đương

## 3.4 Phương pháp xếp hạng ứng viên

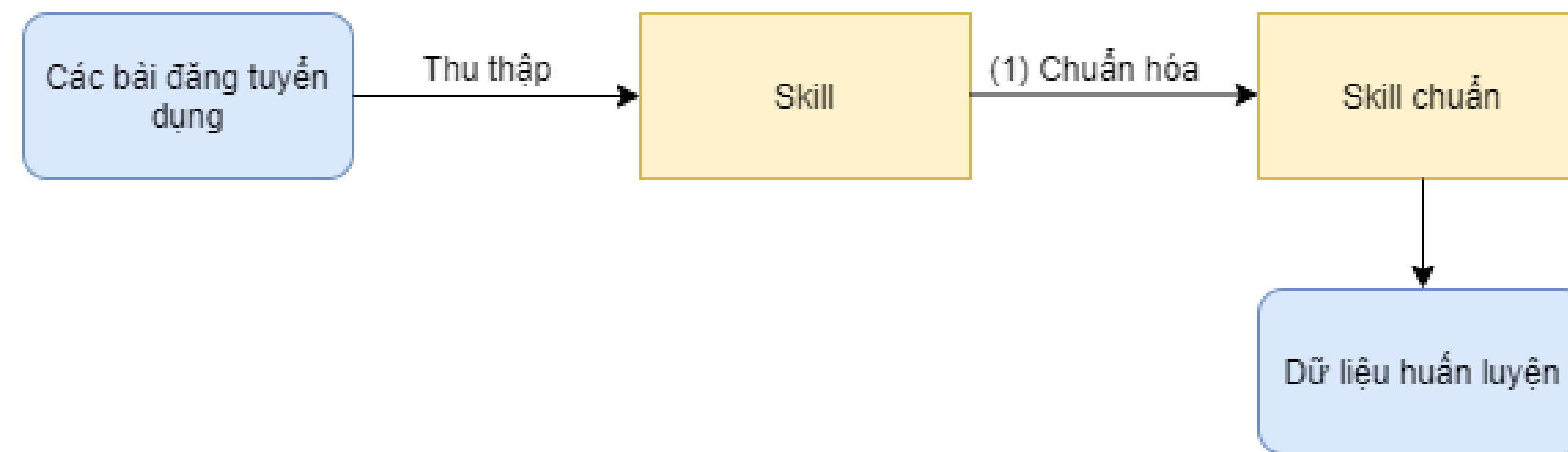
- ❖ Xếp hạng ứng viên theo kỹ năng
- ❖ Xếp hạng ứng viên theo kỹ năng kết hợp thêm các yếu tố khác



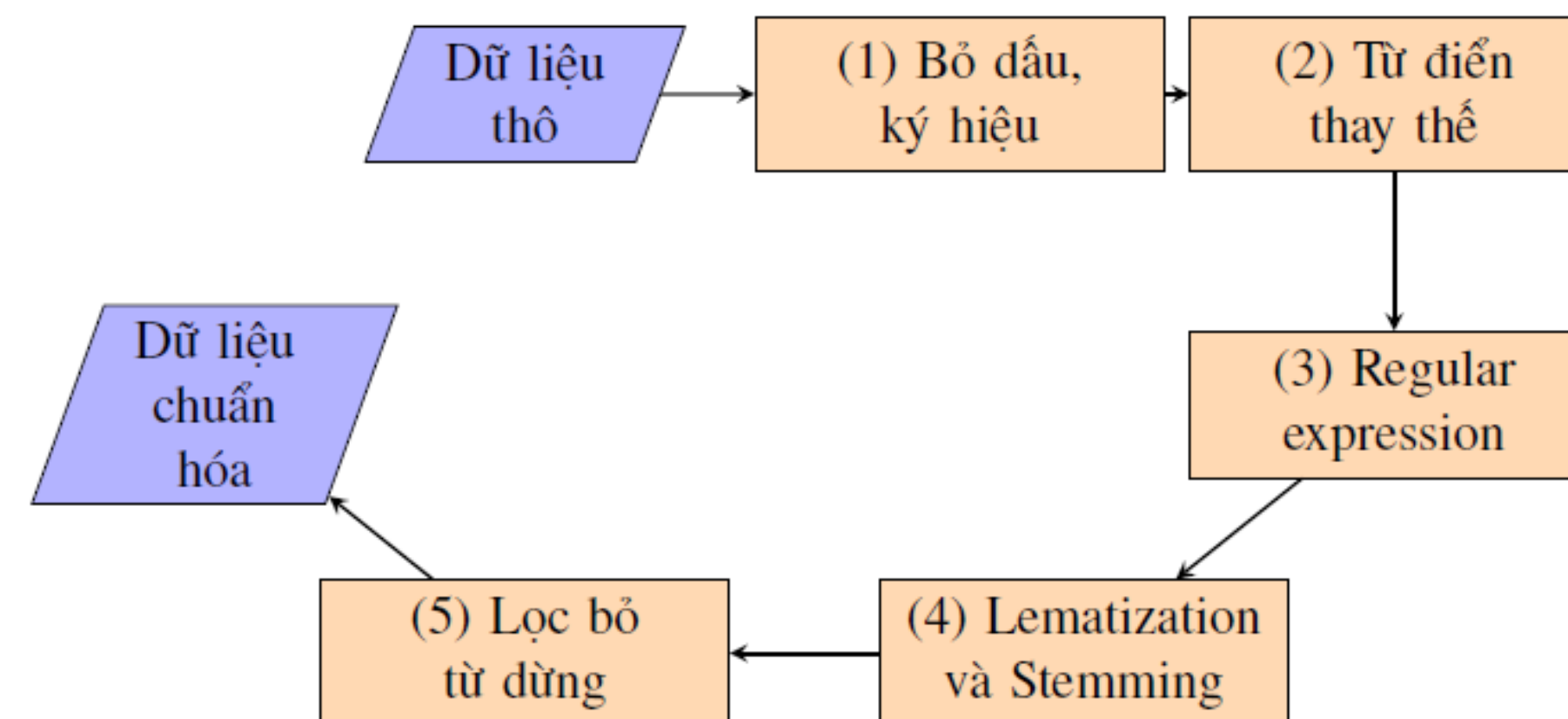
# 3. Natural Language Processing

## 3.2 Thu thập và xử lý dữ liệu

### ❖ Dữ liệu huấn luyện



Quy trình (1) chuẩn hóa





# 3. Natural Language Processing

## 3.2 Thu thập và xử lý dữ liệu

(1) Loại bỏ các ký tự dư thừa, dấu câu, lỗi HTML, mã ký tự đặc biệt HTML(1), chuyển ký tự thành chữ thường (lowercase)

Dữ liệu thô	Xử lý
Javascript,!	javascript
java&dash;script	java-script
javascript (js)	javascript



# 3. Natural Language Processing

## 3.2 Thu thập và xử lý dữ liệu

(2) Xây dựng và sử dụng từ điển để xử lý, chuẩn hóa thống nhất các skill viết tắt

Dữ liệu thô	Xử lý
JS, Javascript	javascript
object oriented programming, OOP, OO	OOP
user experience, web user experience	UX



# 3. Natural Language Processing

## 3.2 Thu thập và xử lý dữ liệu

(3) Sử dụng biểu thức chính quy (regular expression) để chuẩn hóa các kỹ năng có nhiều cách viết khác nhau.

Dữ liệu thô	Regex	Xử lý
Angularjs, Angular.js, Angular-js	([A-z0-9]+):js	Angular
Reactjs, React.js, React-js	([A-z0-9]+):js	React





# 3. Natural Language Processing

## 3.2 Thu thập và xử lý dữ liệu

(4) Một số phương pháp chuẩn hóa từ trong văn bản:

- Lematization: thay thế họa xóa hậu tố của một từ để đưa từ về nguyên mẫu trong tiếng Anh, lematization thực hiện việc này bằng cách sử dụng các luật từ bổ ngữ hoặc từ điển.
- Stemming: tương tự như Lematization, nhưng stemming sử dụng một thuật toán riêng để chuyển từ về nguyên mẫu, phương pháp này ít được sử dụng hơn.

Dữ liệu thô	Xử lý
Java Programming	Java Program
Java Programmer	Java Program
Java Programmers	Java Program



# 3. Natural Language Processing

## 3.2 Thu thập và xử lý dữ liệu

(5) Lọc bỏ từ dừng (stopwords): có những từ trong kỹ năng thu thập được hay xuất hiện nhưng không mang nhiều ý nghĩa. Ví dụ: application, developer, framework, ...

Dữ liệu thô	Xử lý
Java Program	Java
Web Application	Web
React Developer	React



# 3. Natural Language Processing

## 3.2 Thu thập và xử lý dữ liệu

Bảng mô tả công việc

### Big Data Engineer

📍 Reston, VA   ⌚ Full Time   🎓 Experienced

**Required**

- Bachelor's or Master's degree in computer science or software engineering;
- Experience with object-oriented design, coding and testing patterns as well as experience in engineering (commercial or open source) software platforms and large-scale data infrastructures.
- Ability to architect highly scalable distributed systems, using different open source tools.
- Experience building high-performance algorithms.
- Extensive knowledge of different programming or scripting languages such as Java, Linux, C++, PHP, Ruby, Python and/or R.
- Experience with different NoSQL or RDBMS databases such as MongoDB needed.
- Experience building data processing systems with Hadoop and Hive using Java or Python

Dữ liệu huấn luyện

JD	Kỹ năng
JD1	Hadoop, MapReduce, Java, Hive, SQL
JD2	HTML5, CSS, PHP5, MySQL
JD3	HTML5, PHP5
...	...





# **3. Natural Language Processing**

## 3.1 Mô tả mạng kỹ năng

## 3.2 Thu thập và xử lý dữ liệu

- ❖ Dữ liệu huấn luyện
- ❖ **Dữ liệu ứng viên**

## 3.3 Xác định các kỹ năng tương đương

## 3.4 Phương pháp xếp hạng ứng viên

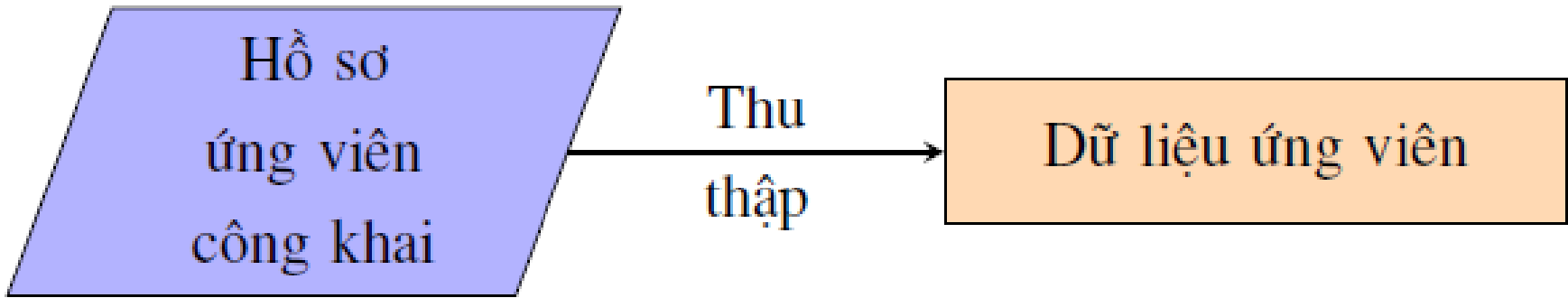
- ❖ Xếp hạng ứng viên theo kỹ năng
- ❖ Xếp hạng ứng viên theo kỹ năng kết hợp thêm các yếu tố khác



# 3. Natural Language Processing

## 3.2 Thu thập và xử lý dữ liệu

❖ Dữ liệu ứng viên



#	Thuộc tính	Mô tả
1	candidate_name	Họ tên
2	url	URL hồ sơ
3	gender	Giới tính
4	address	Địa chỉ
5	skills	Danh sách kỹ năng
6	degree	Bằng cấp
7	experience	Kinh nghiệm (vị trí, thời gian, công ty đã từng làm)



# **3. Natural Language Processing**

## 3.1 Mô tả mạng kỹ năng

## 3.2 Thu thập và xử lý dữ liệu

- ❖ Dữ liệu huấn luyện
- ❖ Dữ liệu ứng viên

## 3.3 Xác định các kỹ năng tương đương

## 3.4 Phương pháp xếp hạng ứng viên

- ❖ Xếp hạng ứng viên theo kỹ năng
- ❖ Xếp hạng ứng viên theo kỹ năng kết hợp thêm các yếu tố khác



# 3. Natural Language Processing

## 3.3. Xác định các kỹ năng tương đương

- ❖ Mô hình LSA (Latent Semantic Analysis)
- ❖ Mô hình Word2vec
- ❖ VD: kết quả đánh giá kỹ năng  
thuộc nhóm “Web Developer”

Skill	No	Relevant Skill	Similarity
html5	1	css3	0.9375
	2	ui_developer	0.9245
	3	bootstrap	0.9231
	4	front__end	0.9156
	5	css_3	0.9136
jquery	1	ajax	0.8304
	2	bootstrap	0.8251
	3	html5	0.8221
	4	mvc_architecture	0.813
	5	oops_concept	0.8114
...	...	...	...





# **3. Natural Language Processing**

## 3.1 Mô tả mạng kỹ năng

## 3.2 Thu thập và xử lý dữ liệu

- ❖ Dữ liệu huấn luyện
- ❖ Dữ liệu ứng viên

## 3.3 Xác định các kỹ năng tương đương

## **3.4 Phương pháp xếp hạng ứng viên**

- ❖ **Xếp hạng ứng viên theo kỹ năng**
- ❖ Xếp hạng ứng viên theo kỹ năng kết hợp thêm các yếu tố khác



# **3. Natural Language Processing**

## **3.4. Phương pháp xếp hạng ứng viên**

### **❖ Xếp hạng ứng viên theo kỹ năng**

Xếp hạng ứng viên theo kỹ năng sẽ đánh giá kỹ năng của ứng viên so với kỹ năng được yêu cầu trong bản mô tả công việc.

Kỹ năng trong bản mô tả công việc được chia thành ba loại:

- **Must have:** phải có.
- **Should have:** nên có.
- **Nice to have:** có thêm là một điểm cộng.

Việc đánh giá ứng viên theo 3 tiêu chí trên trước đây phụ thuộc nhiều vào cảm tính chủ quan của nhà tuyển dụng.



# 3. Natural Language Processing

❖ Điểm kỹ năng **must\_have**:

$$\text{score}_{\text{must\_have}} = \frac{|S_{\text{candidate}} \cap S_{\text{must}}|}{|S_{\text{must}}| + 1}$$

Với:

- $S_{\text{candidate}}$  và  $S_{\text{must}}$  lần lượt là tập các kỹ năng của ứng viên và của must\_have yêu cầu.
- $|S_{\text{candidate}} \cap S_{\text{must}}|$  là số kỹ năng giao nhau của ứng viên và kỹ năng must\_have.

Ví dụ: Một bản mô tả công việc có hai kỹ năng cần có là PHP và HTML; ứng viên có các kỹ năng là PHP, JS và CSS:

$$S_{\text{candidate}} = \{\text{PHP, JS, CSS}\}, S_{\text{must}} = \{\text{PHP, HTML}\}$$

$$\text{Vậy } \text{score}_{\text{must\_have}} = \frac{1}{2+1} = 0.33$$



# 3. Natural Language Processing

❖ Điểm kỹ năng **should\_have**:

$$\text{score}_{\text{should\_have}} = \frac{\sum_{i=1}^{|S_{\text{should}}|} \max\{sim(S_{\text{should } i}, S_{\text{candidate } j}) | j = 1, \dots, |S_{\text{candidate}}|\}}{|S_{\text{should}}| + 1}$$

Với:

- $S_{\text{should}}$  và  $S_{\text{candidate}}$  lần lượt là tập những kỹ năng nên có và kỹ năng của ứng viên.
- $sim(a,b)$  là hàm trả về sự tương đồng giữa hai kỹ năng a và b





# 3. Natural Language Processing

## ❖ Điểm kỹ năng **should\_have**:

Ví dụ: JD có ba kỹ năng nên có là PHP, JS, HTML. Một ứng viên có các kỹ năng là PHP, CSS, Angular. Với bảng độ tương đồng các kỹ năng như sau:

		Kỹ năng nên có		
		PHP	JS	HTML
Kỹ năng ứng viên	PHP	1	0.5	0.6
	CSS	0.5	0.7	<b>0.7</b>
	Angular	0.6	<b>0.8</b>	0.5

$$\text{Vậy } \text{score}_{\text{should\_have}} = \frac{1+0.8+0.7}{3+1} = 0.625$$



# 3. Natural Language Processing

❖ Điểm kỹ năng **nice\_to\_have**:

Được tính tương tự như **should\_have**

$$\text{score}_{\text{nice\_to\_have}} = \frac{\sum_{i=1}^{|S_{\text{nice}}|} \max\{sim(S_{\text{nice } i}, S_{\text{candidate } j}) | j = 1, \dots, |S_{\text{candidate}}|\}}{|S_{\text{nice}}| + 1}$$

Với:

- $S_{\text{nice}}$  và  $S_{\text{candidate}}$  lần lượt là tập những kỹ năng nice\_to\_have và kỹ năng của ứng viên.
- $sim(a, b)$  là hàm tính độ tương đồng cosine similarity giữa hai kỹ năng a và b.



# 3. Natural Language Processing

## 3.4. Phương pháp xếp hạng ứng viên

### ❖ Xếp hạng ứng viên theo kỹ năng

Tổng hợp các loại kỹ năng lại ta được điểm cho phần kỹ năng được tính theo công thức sau:

$$\text{score}_{\text{skills}} = \frac{\text{score}_{\text{must\_have}} \times 100 + \text{score}_{\text{should\_have}} \times 10 + \text{score}_{\text{nice\_to\_have}} \times 1}{100 + 10 + 1}$$

Kết quả:

	score <sub>must_have</sub>	score <sub>should_have</sub>	score <sub>nice_to_have</sub>	score <sub>skills</sub>
Ứng viên 1	0.5	0.6	0.1	0.505
Ứng viên 2	0.5	0.4	0.8	0.494
Ứng viên 3	0.33	0.8	0.8	0.377



# **3. Natural Language Processing**

## 3.1 Mô tả mạng kỹ năng

## 3.2 Thu thập và xử lý dữ liệu

- ❖ Dữ liệu huấn luyện
- ❖ Dữ liệu ứng viên

## 3.3 Xác định các kỹ năng tương đương

## **3.4 Phương pháp xếp hạng ứng viên**

- ❖ Xếp hạng ứng viên theo kỹ năng
- ❖ **Xếp hạng ứng viên theo kỹ năng kết hợp thêm các yếu tố khác**





# **3. Natural Language Processing**

## **3.4. Phương pháp xếp hạng ứng viên**

❖ Xếp hạng ứng viên theo kỹ năng kết hợp thêm các yếu tố khác

✓ Điểm năm kinh nghiệm

✓ Điểm chức vụ

✓ Điểm công ty ưu tiên

✓ Điểm bằng cấp

...



# **AGENDA**

1. Philosophy of Semantic Search
2. Semantic Search For Recruiting
3. NLP (Natural Language Processing)
- 4. Elasticsearch**
5. Q&A



# **4. Elasticsearch**

## 4.1 Elasticsearch

## 4.2 Tìm kiếm trong Elasticsearch

## 4.3 Kibana

## 4.4 Demo



# 4. Elasticsearch

## 4.1 Elasticsearch

- Tìm kiếm dữ liệu rất nhanh chóng, mạnh mẽ dựa trên Apache Lucene. Tìm kiếm trong elasticsearch gần như là realtime hay còn gọi là near-realtime searching
- Có khả năng phân tích dữ liệu (Analysis data)
- Lưu trữ dữ liệu full-text
- Đánh index cho dữ liệu (near-realtime search/indexing, inverted index)
- Hỗ trợ tìm kiếm mờ (fuzzy), tức là từ khóa tìm kiếm có thể bị sai lỗi chính tả hay không đúng cú pháp thì vẫn có khả năng elasticsearch trả về kết quả tốt.
- Hỗ trợ nhiều Elasticsearch client phổ biến như Java, PHP, Javascript, Ruby, .NET, Python





# 4. Elasticsearch

## 4.1 Elasticsearch

MySQL	Elastic Search
Database	Index
Table	Type
Row	Document
Column	Field
Schema	Mapping
Index	Everything is indexed
SQL	Query DSL
SELECT * FROM table ...	GET http://...
UPDATE table SET ...	PUT http://...



# **4. Elasticsearch**

## 4.1 Elasticsearch

## 4.2 Tìm kiếm trong Elasticsearch

## 4.3 Kibana

## 4.4 Demo



# 4. Elasticsearch

## 4.2 Tìm kiếm trong Elasticsearch

### ❖ Cơ chế tìm kiếm

Giả sử có hai văn bản:

*The quick brown fox jumped over the lazy dog*

*Quick brown foxes leap over lazy dogs in summer*

Term	1	2
Quick		x
The	x	
brown	x	x
dog	x	
dogs		x
fox	x	
foxes		x

Term	1	2
in		x
jumped	x	
lazy	x	x
leap		x
over	x	x
quick	x	
summer		x
the	x	

Khi tìm từ quick brown , ta chỉ cần tìm document mà các term xuất hiện:

Term	1	2
quick	x	x
brown	x	



# **4. Elasticsearch**

## **4.2 Tìm kiếm trong Elasticsearch**

### **❖ Full-text Search**

Bình thường, chúng ta sẽ sử dụng câu truy vấn dạng như sau để tìm kiếm dữ liệu:

**SELECT id,title,description FROM book WHERE title LIKE '%keyword%'**

Nhưng cách truy vấn này có một số hạn chế sau:

- Khi không đánh index thì tốc độ tìm kiếm chậm.
- Kết quả tìm kiếm nhiều nhưng độ nhiễu cao, từ đồng nghĩa nhiều.
- Gặp vấn đề trong tìm kiếm tiếng việt có dấu và không dấu.



# 4. Elasticsearch

## 4.2 Tìm kiếm trong Elasticsearch

### ❖ Full-text Search

#### `intervals` query

A full text query that allows fine-grained control of the ordering and proximity of matching terms.

#### `match` query

The standard query for performing full text queries, including fuzzy matching and phrase or proximity queries.

#### `match_bool_prefix` query

Creates a `bool` query that matches each term as a `term` query, except for the last term, which is matched as a `prefix` query

#### `match_phrase` query

Like the `match` query but used for matching exact phrases or word proximity matches.

#### `match_phrase_prefix` query

Like the `match_phrase` query, but does a wildcard search on the final word.

#### `multi_match` query

The multi-field version of the `match` query.

#### `common terms` query

A more specialized query which gives more preference to uncommon words.

#### `query_string` query

Supports the compact Lucene `query string syntax`, allowing you to specify AND|OR|NOT conditions and multi-field search within a single query string. For expert users only.

#### `simple_query_string` query

A simpler, more robust version of the `query_string` syntax suitable for exposing directly to users.





# 4. Elasticsearch

## 4.2 Tìm kiếm trong Elasticsearch

### ❖ Match Query

Là truy vấn chuẩn để thực hiện full text query. Bao gồm truy vấn kết hợp và truy vấn cụm từ hoặc gần đúng. Match query chấp nhận văn bản, số, ngày tháng. Match query trả về các document chứa ít nhất 1 trong các từ trong truy vấn.

```
GET /_search
{
  "query": {
    "match": {
      "content": "ruby python"
    }
  }
}
```



# 4. Elasticsearch

## 4.2 Tìm kiếm trong Elasticsearch

### ❖ Match Phrase Prefix Query

Trả về các document khớp với tiền tố trong truy vấn.

```
GET /_search
{
  "query": {
    "match_phrase_prefix": {
      "content": "ruby on rai"
    }
  }
}
```



# 4. Elasticsearch

## 4.2 Tìm kiếm trong Elasticsearch

### ❖ Multi Match Query

Tương tự match query nhưng cho phép tìm kiếm trên nhiều trường.

```
GET /_search
{
  "query": {
    "multi_match" : {
      "query": "ruby",
      "fields": [ "title", "content" ]
    }
  }
}
```



# 4. Elasticsearch

## 4.2 Tìm kiếm trong Elasticsearch

### ❖ Controlling precision

Sẽ làm gì khi người dùng đưa ra 1 truy vấn có 4 từ và cần lấy có các document chứa ít nhất 3 từ trong đó. Elastic hỗ trợ `minimum_should_match` parameter, cho phép chỉ ra số terms sẽ so sánh trong tài liệu chứa các kết quả thích hợp .

```
GET /_search
{
  "query": {
    "multi_match" : {
      "query" : "python ruby php abc",
      "minimum_should_match": "75%"
    }
  }
}
```



# 4. Elasticsearch

## 4.2 Tìm kiếm trong Elasticsearch

### ❖ Combining Queries

Combining queries cho phép thực hiện nhiều điều kiện trong tìm kiếm.

```
GET /_search
{
  "query": {
    "bool": {
      "must": {"match": {"content": "ruby"}},
      "must_not": {"match": {"content":
"rails"}}}
    }
  }
}
```





# **4. Elasticsearch**

## 4.1 Elasticsearch

## 4.2 Tìm kiếm trong Elasticsearch

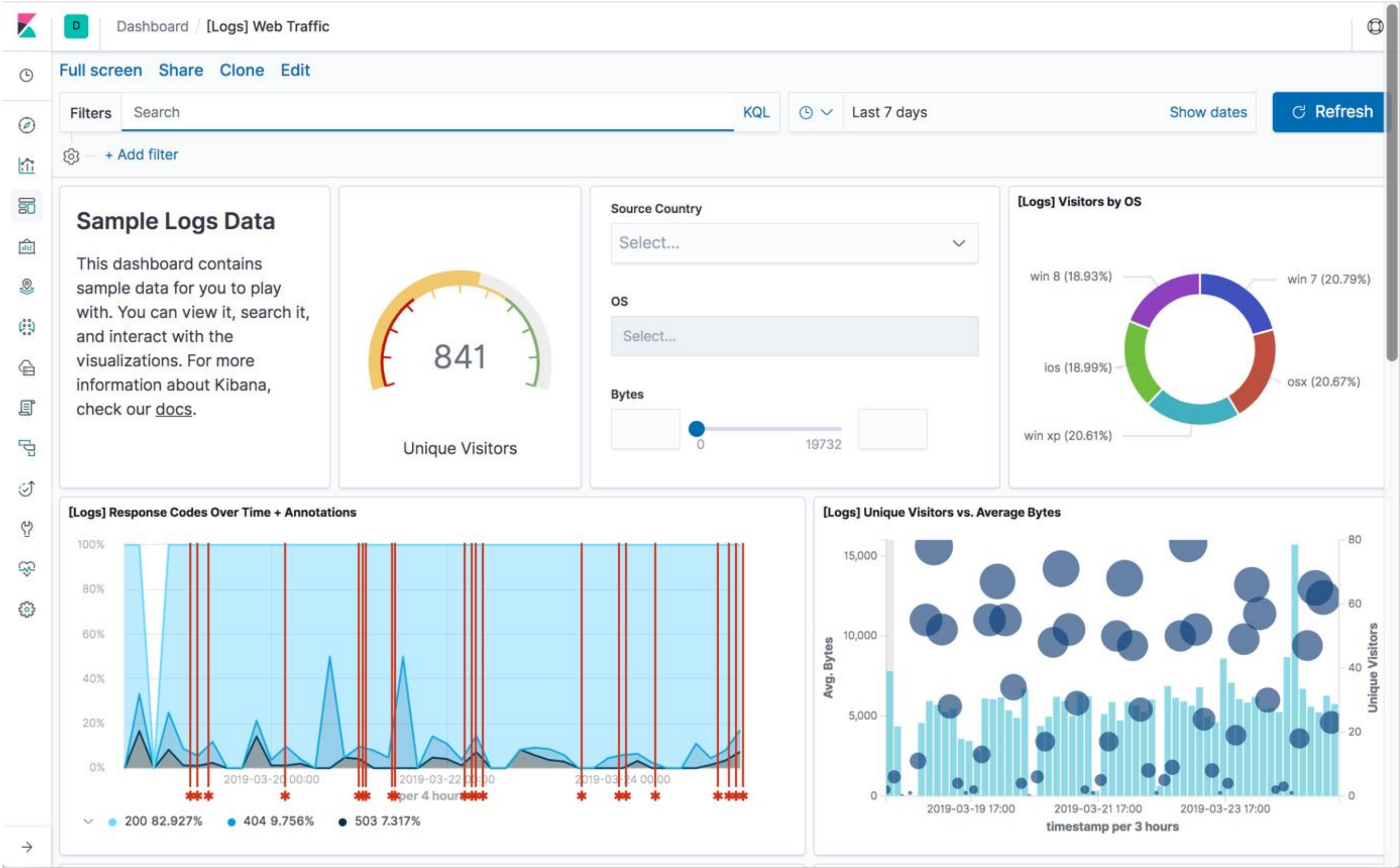
## 4.3 Kibana

## 4.4 Demo



# 4. Elasticsearch

## 4.3 Kibana



# **4. Elasticsearch**

## 4.1 Elasticsearch

## 4.2 Tìm kiếm trong Elasticsearch

## 4.3 Kibana

## 4.4 Demo



# **AGENDA**

1. Philosophy of Semantic Search
2. Semantic Search For Recruiting
3. NLP (Natural Language Processing)
4. Elasticsearch
- 5. Q&A**



## 5. Q&A





ありがとうございます

