



2022 SPE EUROPE ENERGY GEOHACKATHON

9. Azure Cognitive Search

Farzad Sunavala (Microsoft)

6th October 2022

#DatafyingEnergy

Semantic Search: How Improvements in AI are transforming search relevance

Farzad Sunavala

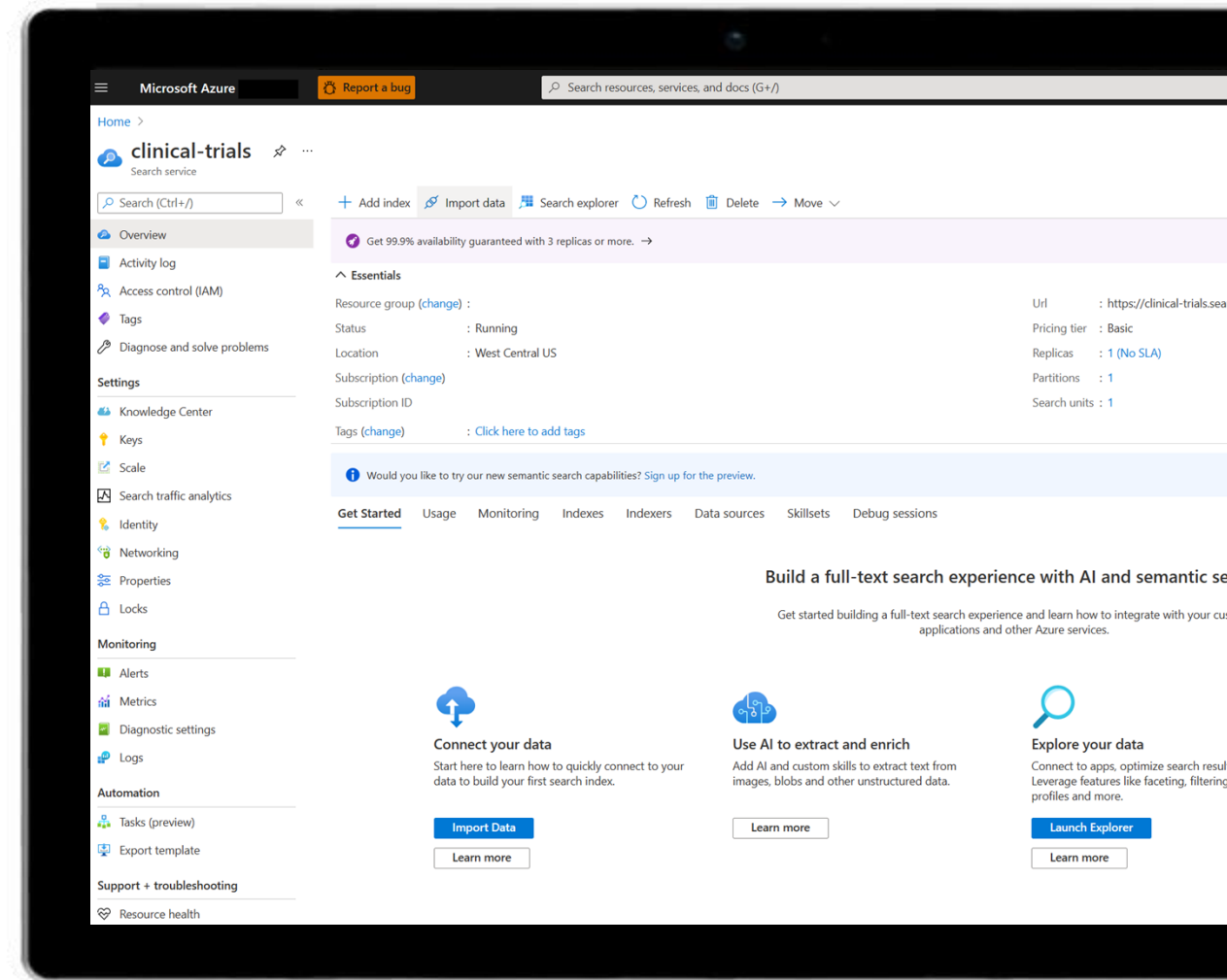
Senior Program Manager

Azure Cognitive Search

Azure Cognitive Search

Platform-as-a-Service

management free
keyword search
faceting
language analyzers
geospatial support
suggestions/auto-complete
customizable scoring
proximity search
synonyms
cognitive skills
etc.





Spelling Mistakes

Suggestions and Auto-complete

Filters and Facets

Geospatial Queries

Paging

Snippets and Highlights

SEARCH

Analyst

STAFF ANALYST

Analyst

Procurement Analyst II

Research Analyst (Part-Time)

PROCUREMENT ANALYST

BUDGET ANALYST

➤ Auditor (17)

➤ Project Manager (17)

➤ College Aide (13)

Agency Attorney (12)

Assistant Corporation Counsel (12)

➤ Construction Project Manager Intern (10)

➤ Administrative Assistant (8)

➤ Procurement Analyst (8)

➤ City Research Scientist (7)

➤ Watershed Maintainer (7)

LOCATION

➤ Internal (1341)

➤ External (1217)

POSTING TYPE

MINIMUM SALARY



Ranking

AVAILABLE JOBS

Relevance



< 1 2 3 4 5 >



5 Nov Business Analyst

150 William Street, New York N

Salary: \$56,937 to \$88,649 Annual

JOB DESCRIPTION: Under supervision of the Business Analyst Manager, with latitude for independent judgment and action, the candidate will serve as an ACS Business Analyst for the development, enhancement and maintenance of ACS systems and applications. The Business Analyst role is critical to the shared resources framework within MIS. Primary responsibilities include: conducting detailed business analysis, creating High-level, Functional and/or Business Requirements documents, producing forms ... [Read More](#)



3 Dec Limited Alteration Application (LAA) Inspector

280 Broadway, 5th Floor, N.Y.

Salary: \$45,162 to \$51,936 Annual

AI improves both Indexing and Querying

Indexing



Document
Cracking

AI Enrichments
(optional)

Inverted
Index

Querying

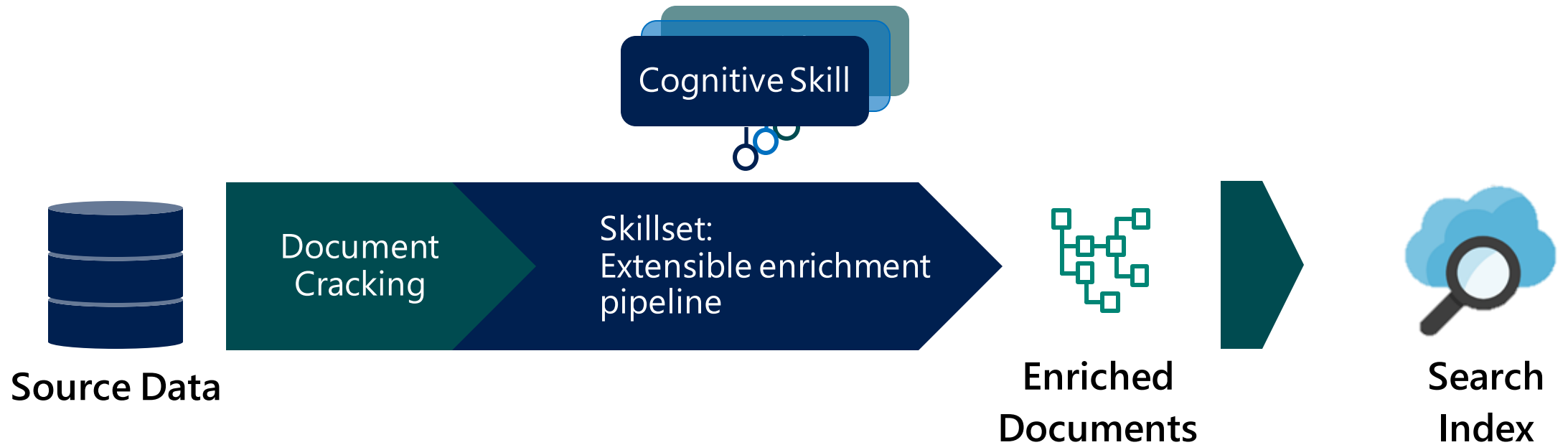
How to build a better mousetrap 🔍

Retrieval

Ranking

Query Results

Knowledge Mining with Azure Cognitive Search



INGEST

ENRICH

EXPLORE

The Metropolitan Museum of Art

Enabling Art Education and Exploration

Solution

Art exploration application allowing the general public to explore The Met's open access collection

Results

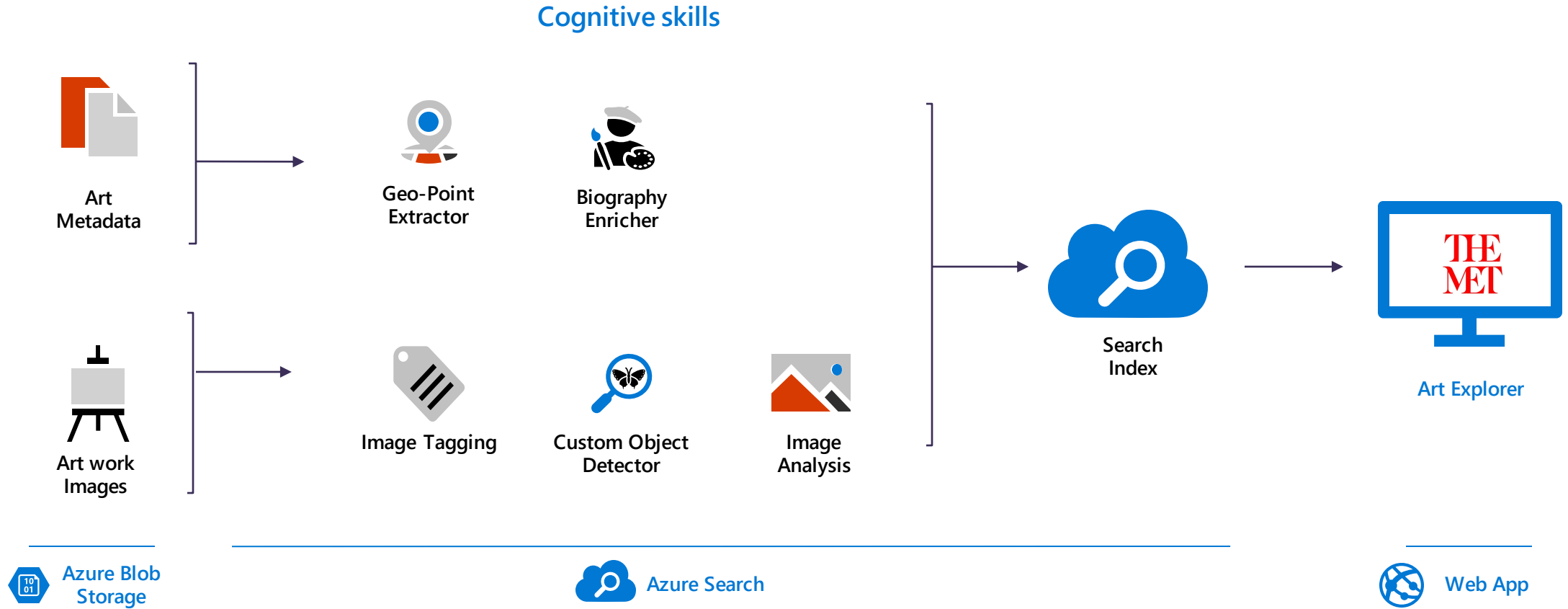
- Automatically tag art pieces and explore similar or related art works
- Enrich information about the artist and their work
- Explore art works based on location and time period of the piece



THE
MET

Art Explorer

Enrich The Met's Open Access Data



AI improves both Indexing and Querying

Indexing



Document
Cracking

AI Enrichments
(optional)

Inverted
Index

Querying

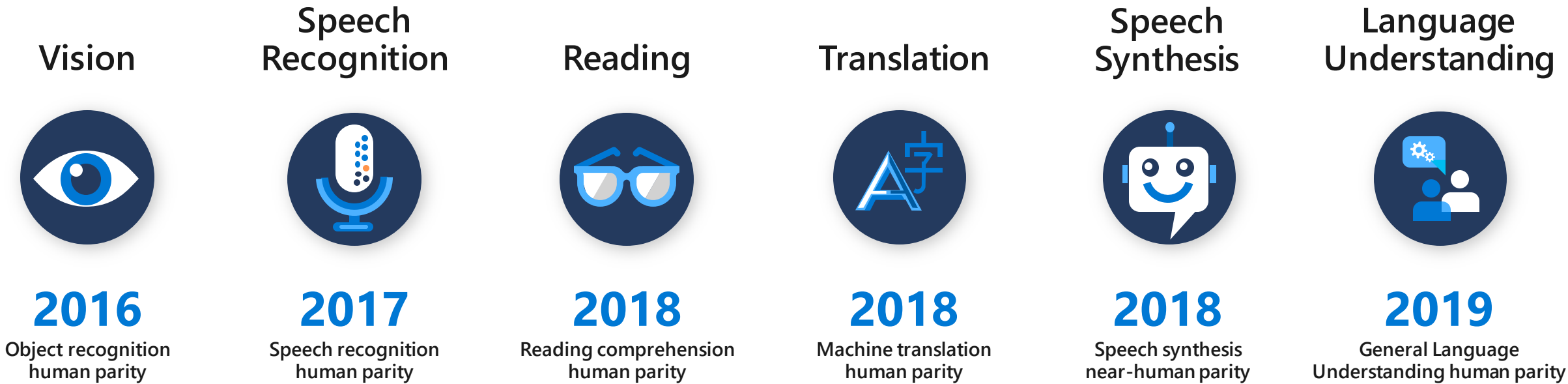
How to build a better mousetrap 🔍

Retrieval

Ranking

Query Results

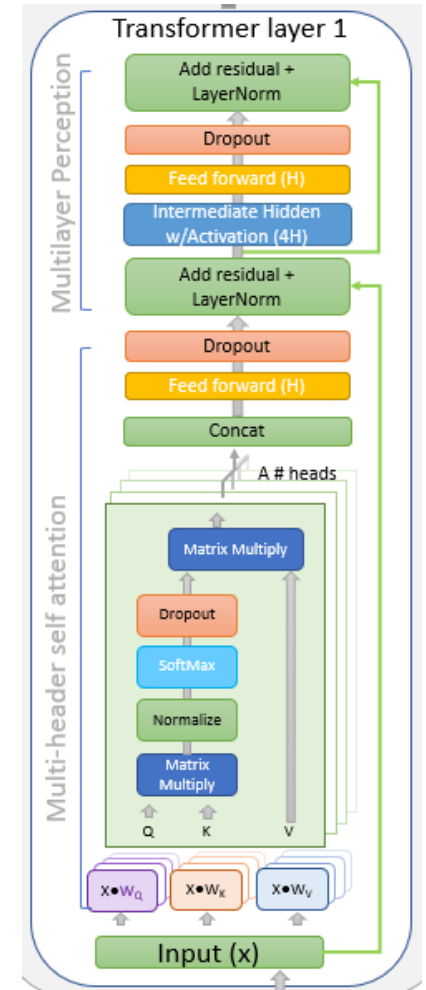
Microsoft AI advancements



Transformer Models + GPUs = SOTA gains in NLP

Since 2018, new deep neural network architecture (the Transformer) + availability of GPUs has led to breakthroughs in model that are:

- **Capable language models** (e.g. Understand structure, grammar, tense, relationships, etc)
 - Work on arbitrary text passages
 - Large vocabulary (expandable into other languages)
 - Reason between and within passages of text
- **Reusable** (across different tasks)
- **Trainable**
 - Without manually labeling
 - Efficient convergence
- **Manageable** (configurable size, decent inference and training perf)



Keyword based ranking

"Build a better mousetrap, and the world will beat a path to your door" is a phrase attributed to Ralph Waldo Emerson in the late nineteenth century. The phrase is actually a misquotation of the statement:

"If a man has good corn or wood, or boards, or pigs, to sell, or can make better chairs or knives, crucibles or church organs, than anybody else, you will find a broad hard-beaten road to his house, though it be in the woods."

— Ralph Waldo Emerson

The phrase has turned into a metaphor about the power of innovation and is frequently taken literally, with more than 4,400 patents issued by the United States Patent and Trademark Office for new mousetraps, with thousands more unsuccessful applicants, making them the "most frequently invented device in U.S. history"

	frequency
better	1
beat	2
build	1
corn	1
...	
waldo	2
wood	1

$$\text{score}(D, Q) = \sum_{i=1}^n \text{IDF}(q_i) \cdot \frac{\boxed{f(q_i, D)} \cdot (k_1 + 1)}{\boxed{f(q_i, D)} + k_1 \cdot \left(1 - b + b \cdot \frac{|D|}{\text{avgdl}}\right)}$$

Ranking Knobs

language analyzers

custom analyzer

tokenizers

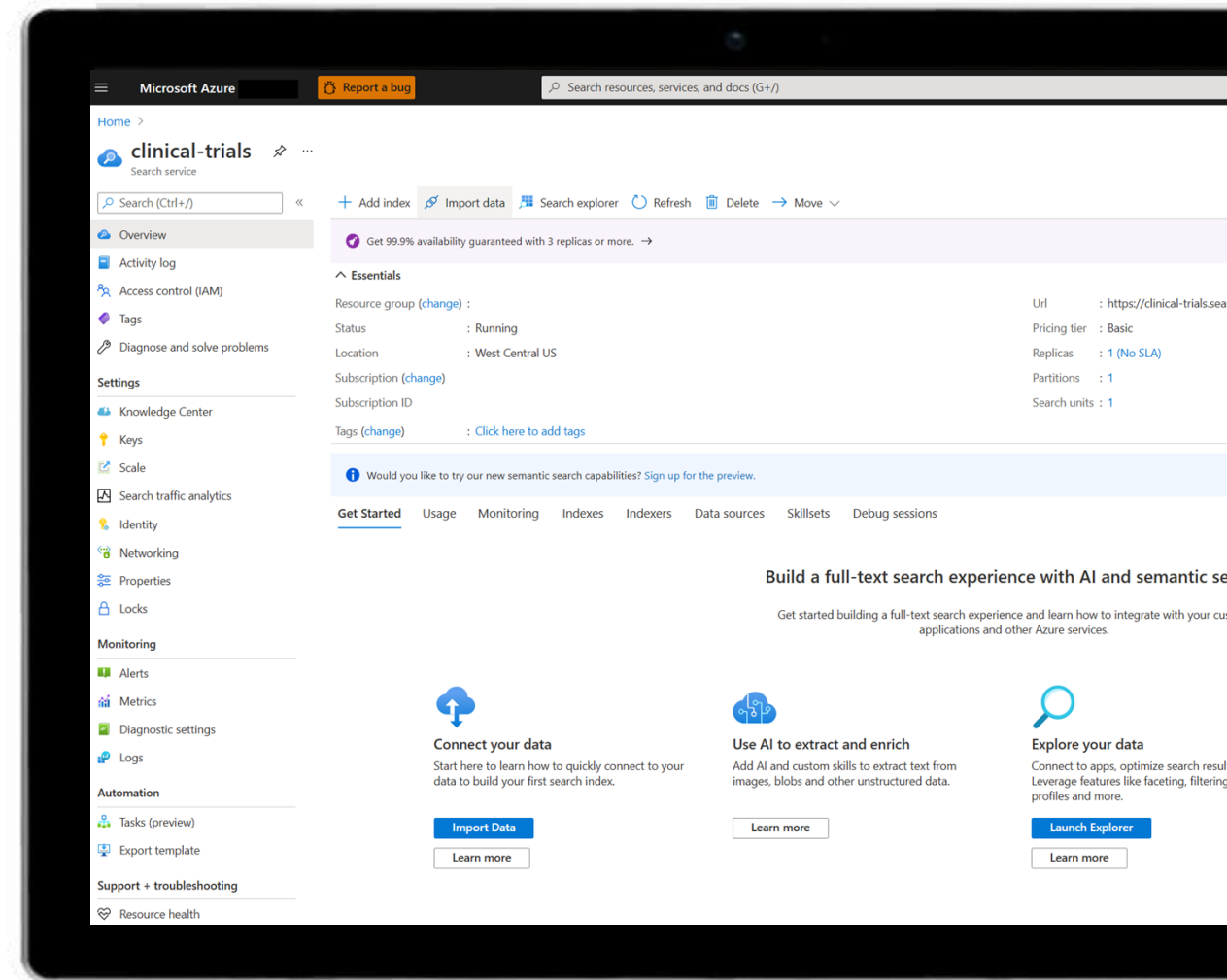
scoring profiles

customizable scoring

proximity search

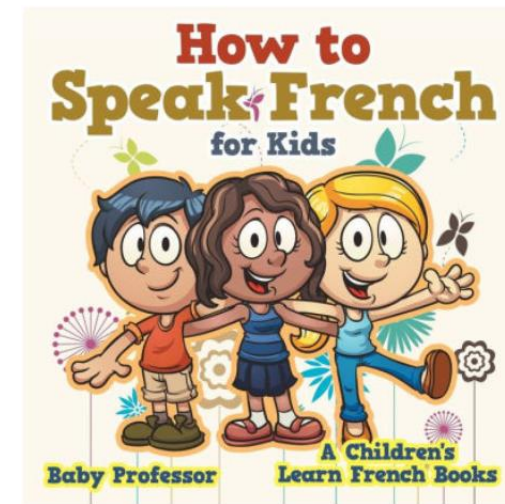
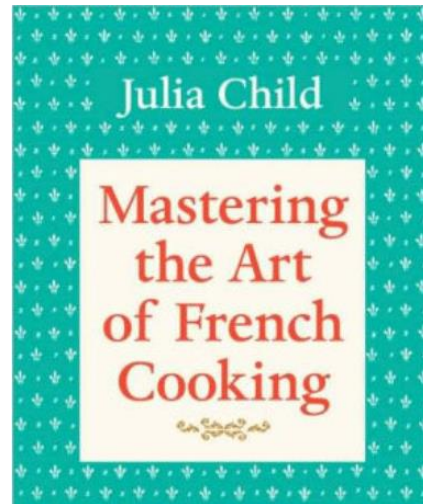
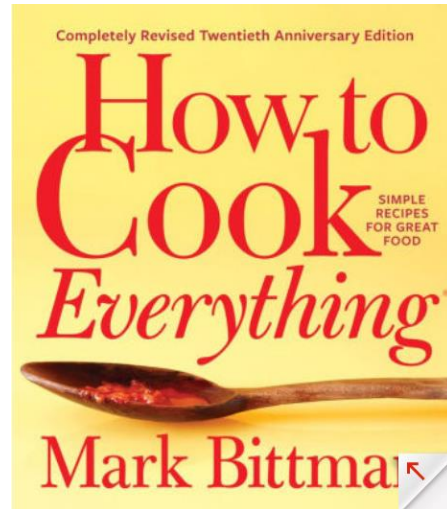
field selection and boosting

synonyms



But there is still a problem...

How to cook French food



What do we really want??

"How to cook French food"

Syntactically similar tokens

how

to

cook

french

Food

cooked

cooking

Semantically similar concepts

introduction

learn

class

master

France

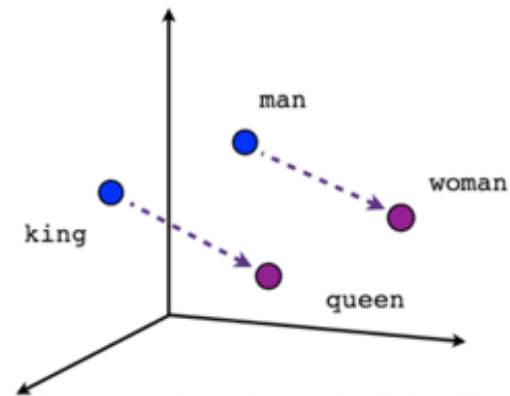
Paris

culinary

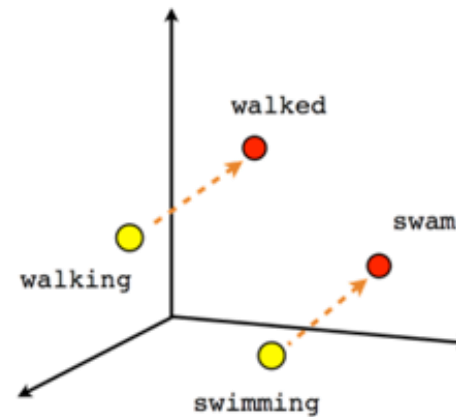
chef

Deep Neural Networks can help find semantically similar terms

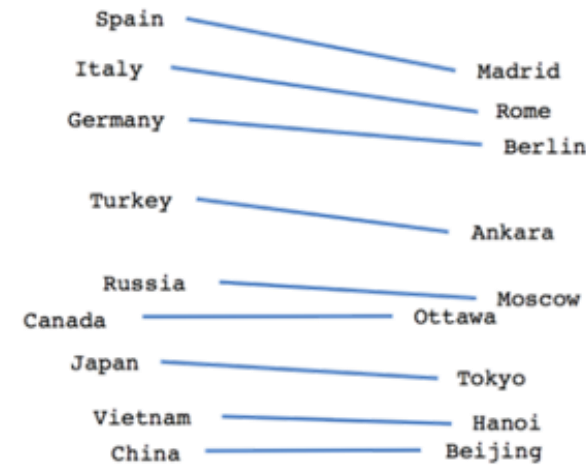
Everyday Bing uses trained deep neural networks that map words and sentences into a multi-dimensional space.
(embeddings)



Male-Female



Verb tense



Country-Capital

Deep Neural Networks can help ranking documents



pasteurization process

ALL WORK IMAGES VIDEOS MAPS NEWS SHOPPING

1,630,000 Results Any time

Pasteurization

Pasteurization or pasteurisation is a process in which packaged and non-packaged foods (such as milk and fruit juice) are treated with mild heat, usually to less than 100 °C (212 °F), to eliminate pathogens and extend shelf life.

[Pasteurization - Wikipedia](https://en.wikipedia.org/wiki/Pasteurization)
<https://en.wikipedia.org/wiki/Pasteurization>

See more about Pasteurization

PEOPLE ALSO ASK

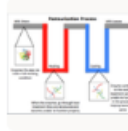
- What are the steps in pasteurization?
- What is the process of pasteurization and its purpose?
- How does pasteurization preserve food?
- What is the purpose of pasteurization?

Feedback

Pasteurization - Wikipedia

<https://en.wikipedia.org/wiki/Pasteurization>

< Overview **Pasteurization process** History Effects on nutritional and >



Pasteurization is a mild heat treatment of liquid foods (both packaged and unpackaged) where products are typically heated to below 100 °C. The heat treatment and cooling process are designed to inhibit a phase change of the product. The acidity of the food determines the parameters (time and temperature) of the heat treatment as well as the duration of shelf life. Parameters also take into account nutritional and sensory qualities that are sensitive to heat.


Wikipedia · Text under CC-BY-SA license

pasteurization | Definition, Process, Inventor, & Facts ...




<https://www.britannica.com/technology/pasteurization>

Pasteurization, heat-treatment process that destroys pathogenic microorganisms in certain foods and beverages. It is named for the French scientist Louis Pasteur, who in the 1860s demonstrated that abnormal fermentation of wine and beer could be prevented by heating the beverages to about 57° C (135° F) for a few minutes.

Deep Neural
Networks can help
predict what
sentence or caption
directly answers
a query




When was Emmental invented



ALLWORKIMAGESVIDEOSMAPSNEWSSHOPPING

838,000 ResultsAny time ▾



1293

Emmental was first written down in records in **1293**, but first called by its name in 1542.


Pasteurized: Not traditionally



Region, town: [Bern](#), [Emmental](#)

Source of milk: [Cow](#)

Texture: Medium-hard

[Emmental cheese - Wikipedia](#)

 en.wikipedia.org/wiki/Emmental_cheese

Was this helpful?  

PEOPLE ALSO ASK

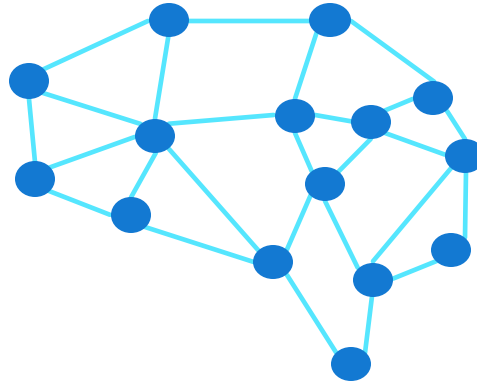
When was Swiss cheese first invented? ▾

What age is Emmental? ▾

What is the difference between Emmental and Emmental? ▾

What is Emmental cheese? ▾

Bing and Azure are partnering to help democratize state of the art search technologies



Microsoft Bing



Microsoft Azure

Semantic Search

Semantic Ranking
Semantic Answers
Semantic Captions
Spell Check



What is the capital of France?



Paris is the capital city of France. It has a population of 2,138,551, and is located on a latitude of 48.85 and longitude of 2.35. **Paris** is also the political center of France, which is considered a Republic, and home to its Executive head of state.

[Where is France?](#)

<https://www.worldatlas.com/eu/fr/where-is-france.html>

[Where is France?](#)

<https://www.worldatlas.com/eu/fr/where-is-france.html>

Paris is the **capital** city of **France**. It has a population of 2,138,551, and is located on a latitude of 48.85 and longitude of 2.35. **Paris** is also the political center of **France**, which is considered a Republic, and home to its Executive head of state.

[Where is France?](#)

<http://www.worldatlas.com/eu/fr/where-is-france.html>

Paris is the **capital** city of **France**. It has a population of 2,138,551, and is located on a latitude of 48.85 and longitude of 2.35. **Paris** is also the political center of **France**, which is considered a Republic, and home to its Executive head of state.

[France](#)

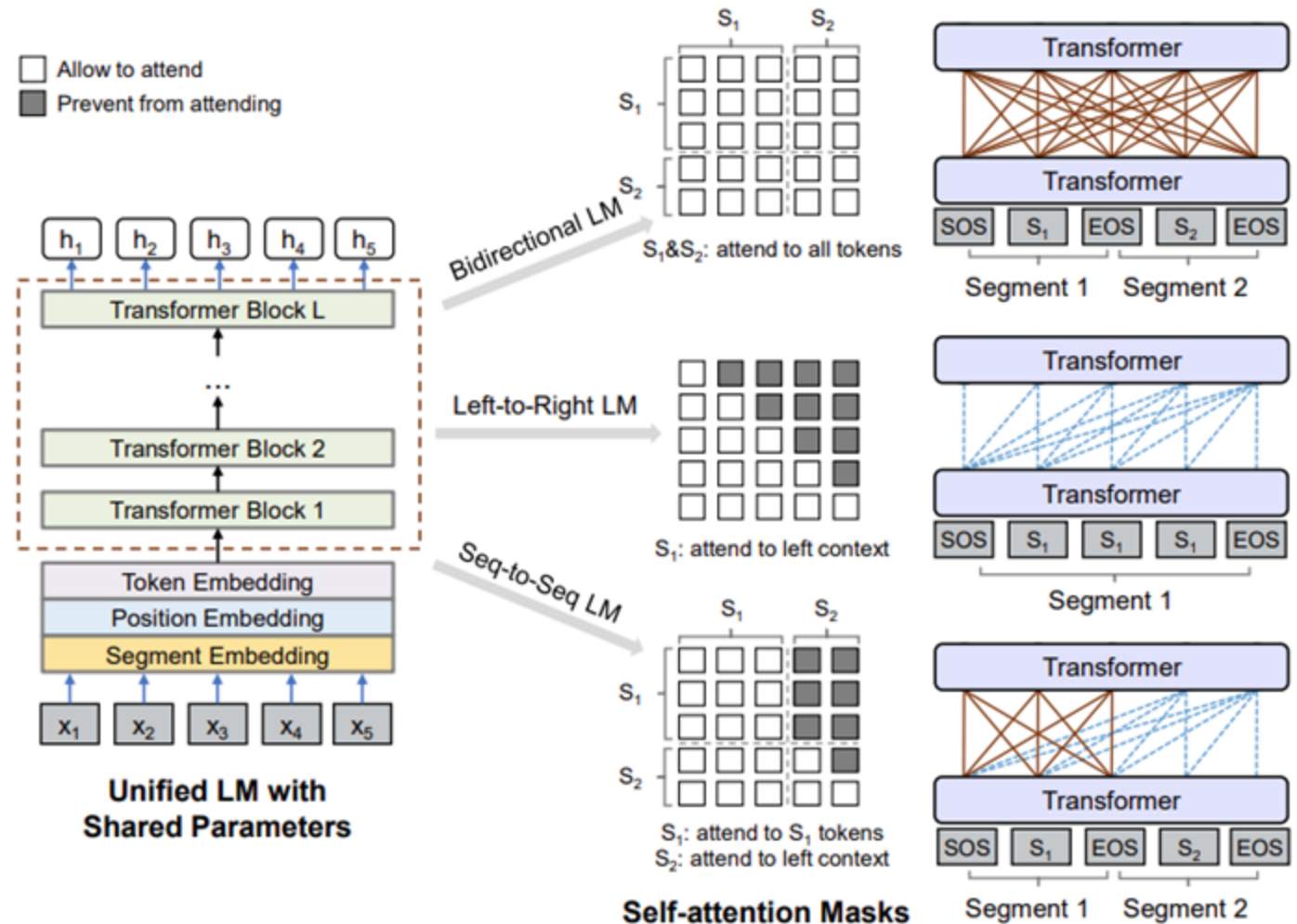
<https://www.iaufrance.org/locations/france>

Paris, the ultra-cosmopolitan **capital** city, is at the heart of **French** culture and an ideal environment in which to learn and perfect **French** language skills. The city has a longstanding global reputation as a mecca of art, culture, fashion, gastronomy, history, and architecture.

Semantic Search Demo

Learn about the science behind semantic search

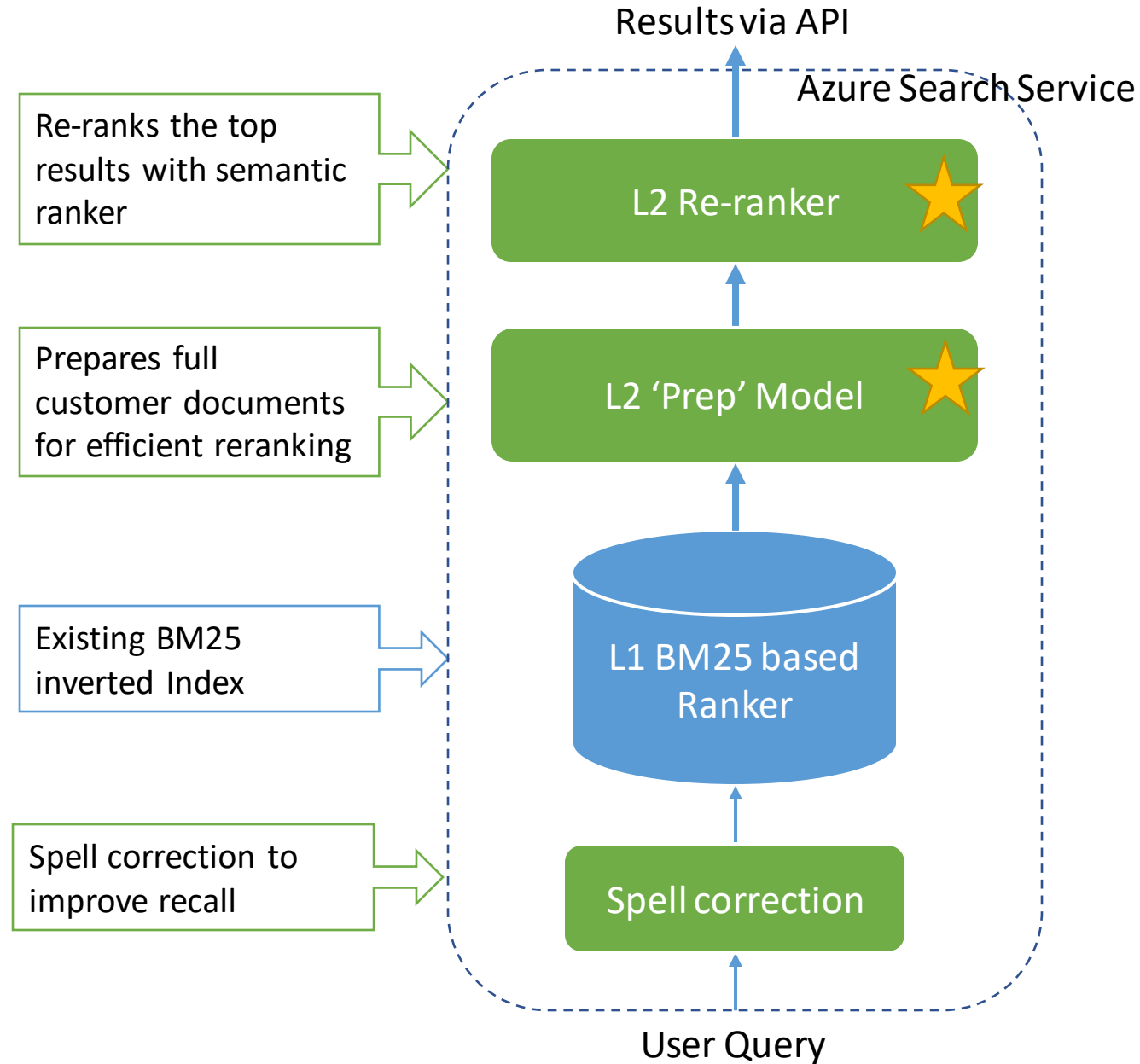
[The science behind semantic search: How AI from Bing is powering Azure Cognitive Search - Microsoft Research](#)



New DNN Model Components

Existing

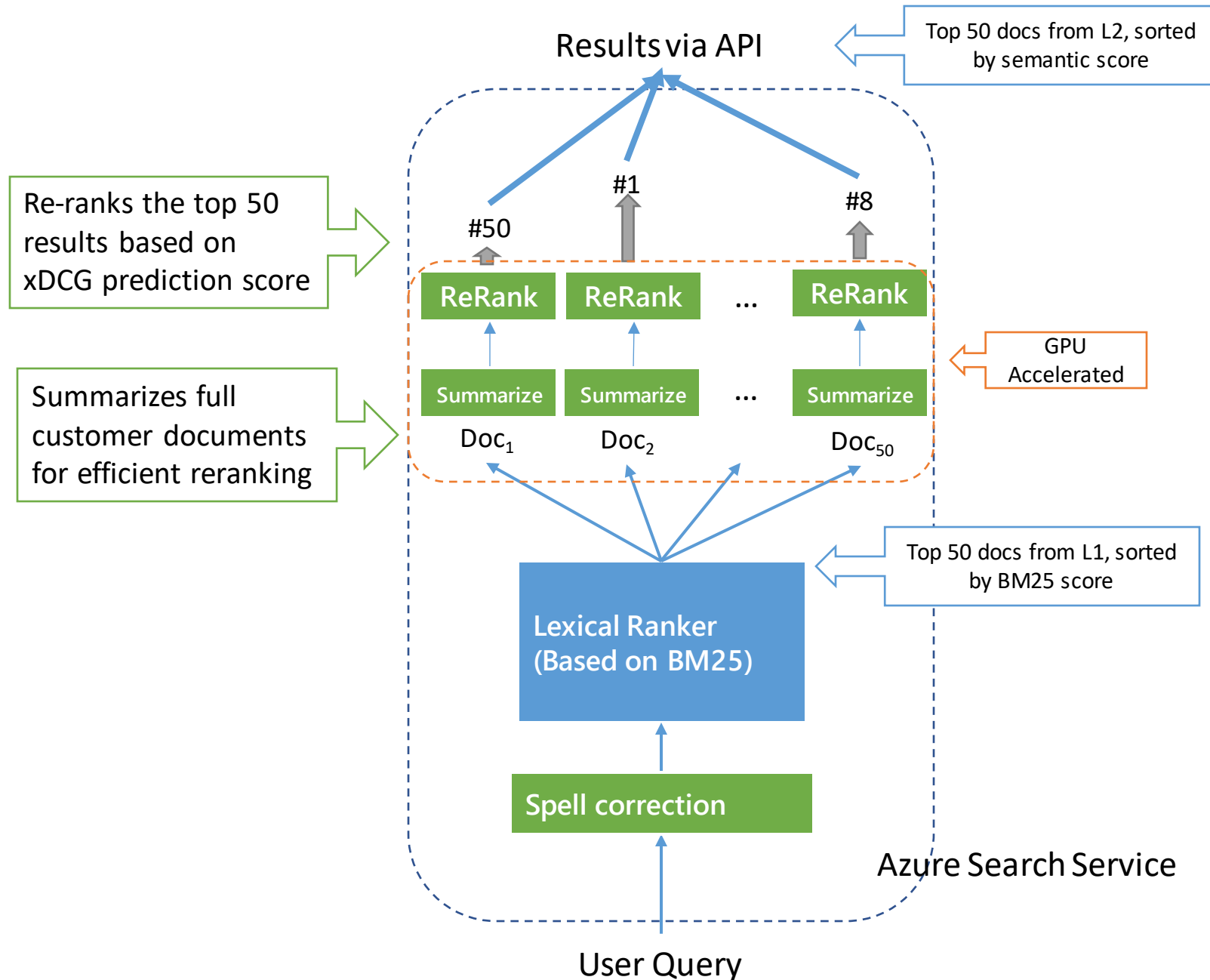
Public Preview (now)



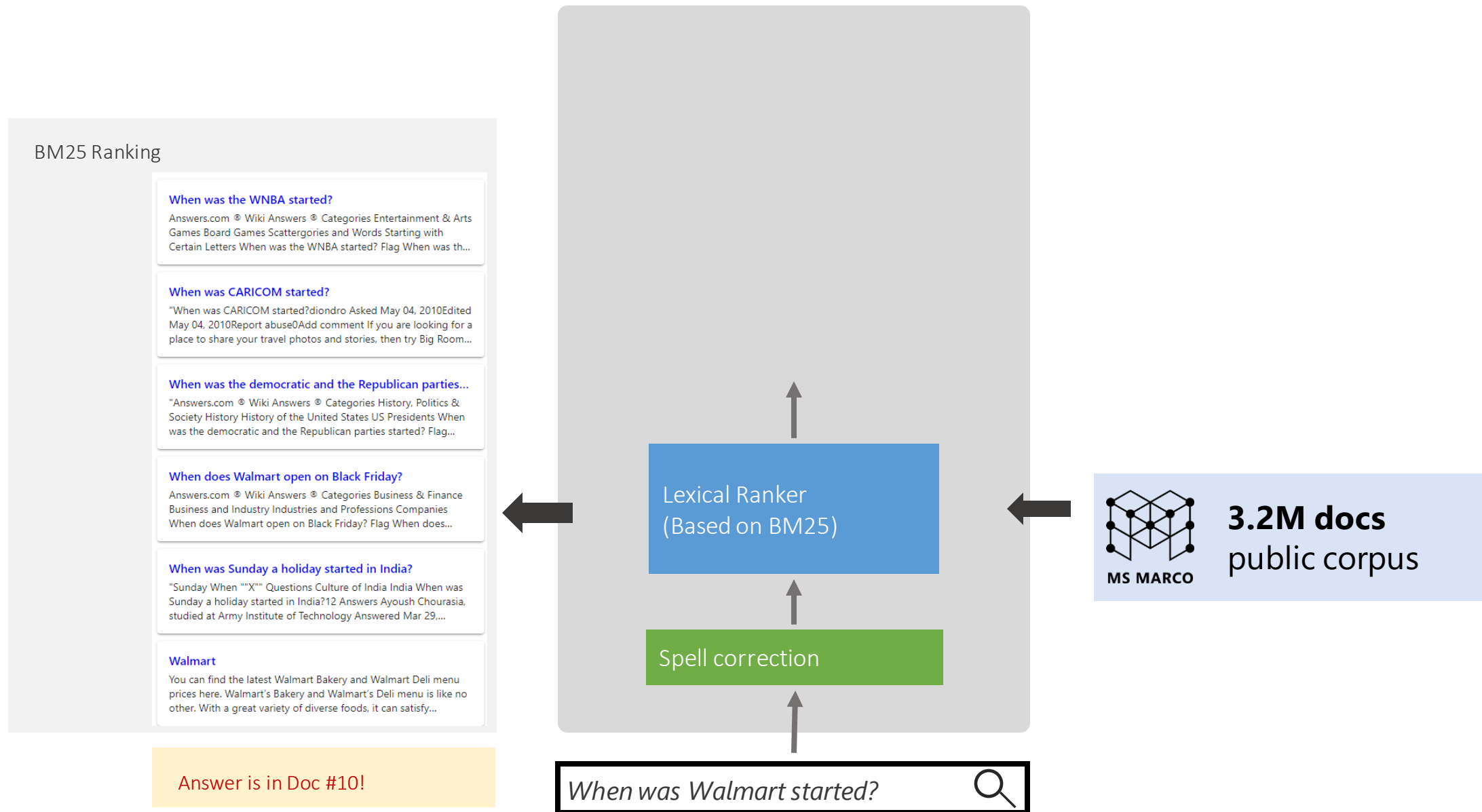
Putting together DNN Model Components

Existing

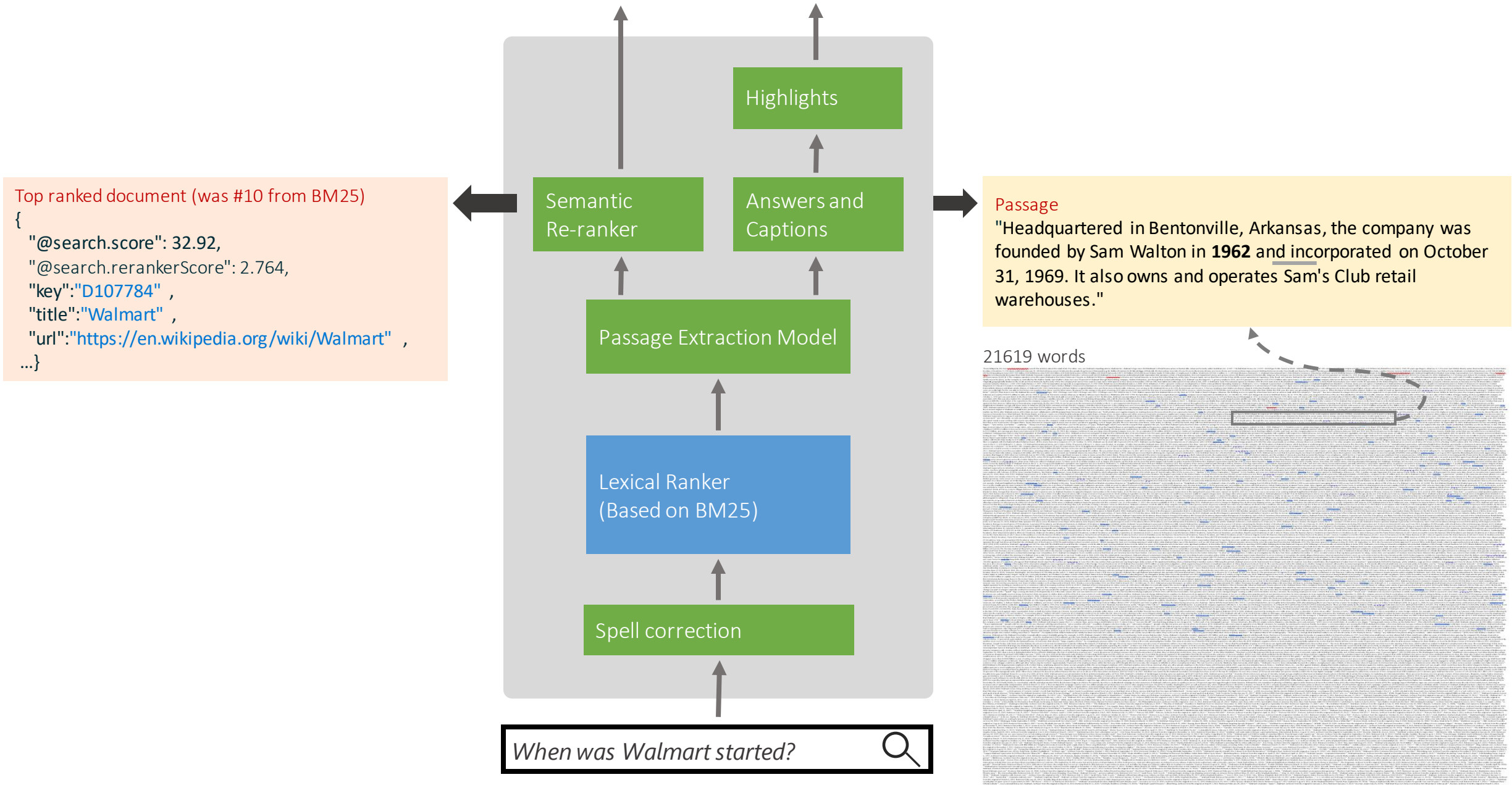
Public Preview (now)



Semantic Search



Semantic Search



Getting started with semantic search

Great! But how do I know how good my current search experience is? And how do I know if semantic search is better?

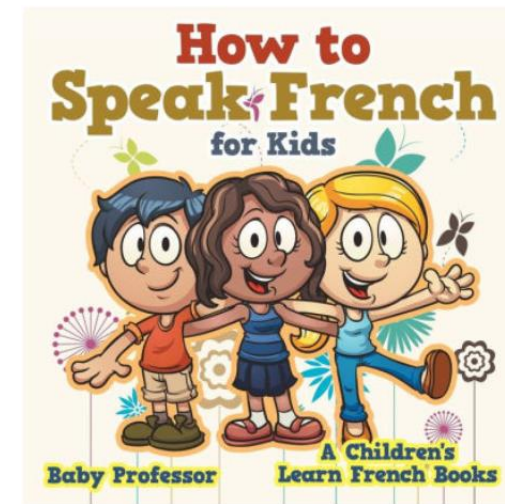
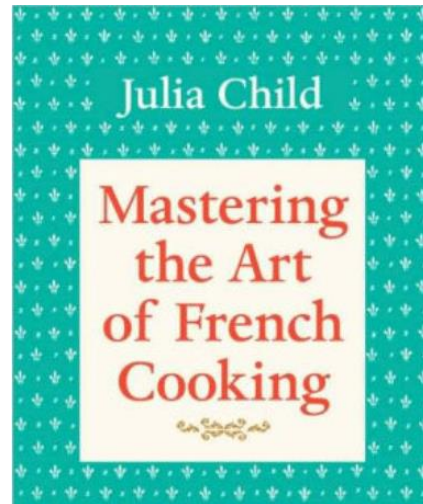
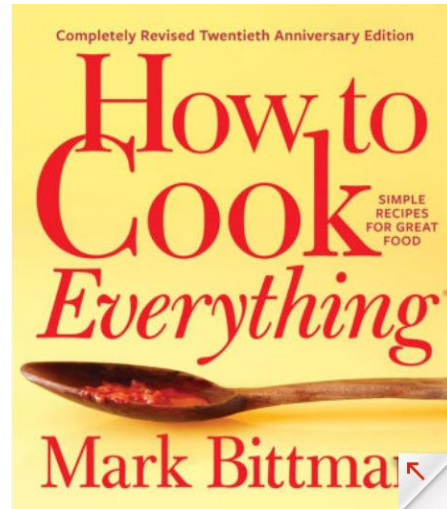
Measuring Search Relevance

Normalized Discounted Cumulative Gain

nDCG is a measure of ranking quality and is used for measuring the effectiveness of search engines

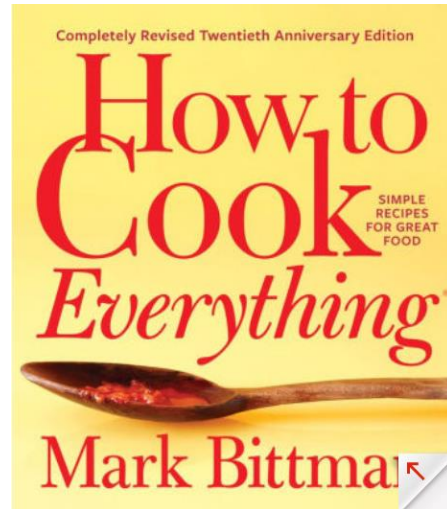
But there is still a problem...

How to cook French food



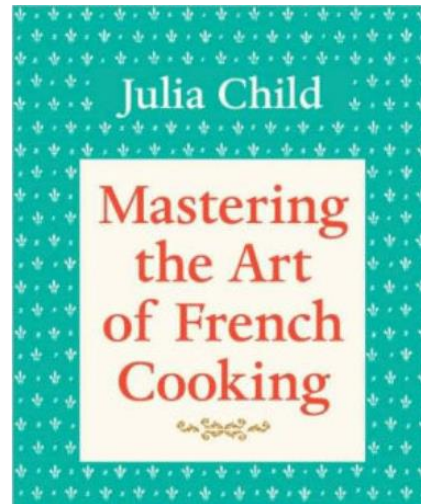
Judging the relevance of each document to the query

How to cook French food



2

Somewhat
relevant



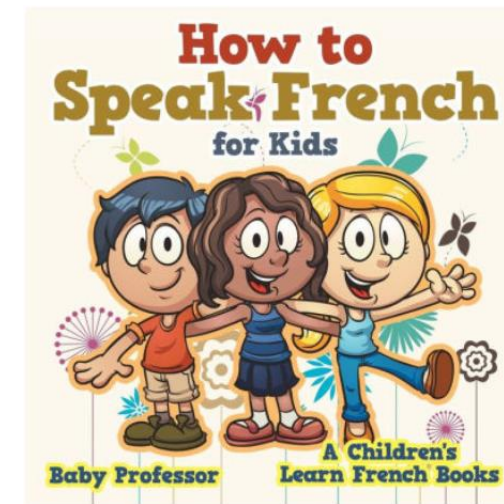
3

Completely
relevant



1

Somewhat
relevant



0

Not
relevant

Normalized Discounted Cumulative Gain

nDCG is a measure of ranking quality and is used for measuring the effectiveness of search engines

$$DCG_p = \sum_{i=1}^p \frac{2^{rel_i} - 1}{\log_2(i + 1)}$$

Ideal order = [3, 2, 1, 0]

Actual order = [2, 3, 0, 1]

$$DCG = \frac{2}{\log_2(1+1)} + \frac{3}{\log_2(1+2)} + \frac{0}{\log_2(1+3)} + \dots$$

$$nDCG_6 = \frac{DCG_6}{IDCG_6} = \frac{6.861}{8.740} = 0.785$$

Computing nDCG with scikit-learn

```
from sklearn.metrics import ndcg_score, dcg_score
import numpy as np

# Relevance scores in the ideal order
true_relevance = np.asarray([[3, 2, 1, 0]])

# Relevance scores in order returned from search engine
relevance_score = np.asarray([[1, 2, 0, 3]])

# Use scikit-learn to calculate the ndcg score
print("nDCG score: ", ndcg_score(true_relevance, relevance_score))
```

Resources

Resources

- <https://aka.ms/SemanticGetStarted>
- [Azure Search - Semantic Demo \(semantic-search-demo-web.azurewebsites.net\)](https://web.azurewebsites.net/SemanticSearchDemo)
- [Azure-Samples/azure-search-knowledge-mining: Azure Search Knowledge Mining Accelerator \(github.com\)](https://github.com/Azure-Samples/azure-search-knowledge-mining)
- [dereklegenzoff/azure-search-retail \(github.com\)](https://github.com/dereklegenzoff/azure-search-retail)
- [farzad528/semantic-search-demo \(github.com\)](https://github.com/farzad528/semantic-search-demo)

Q&A

