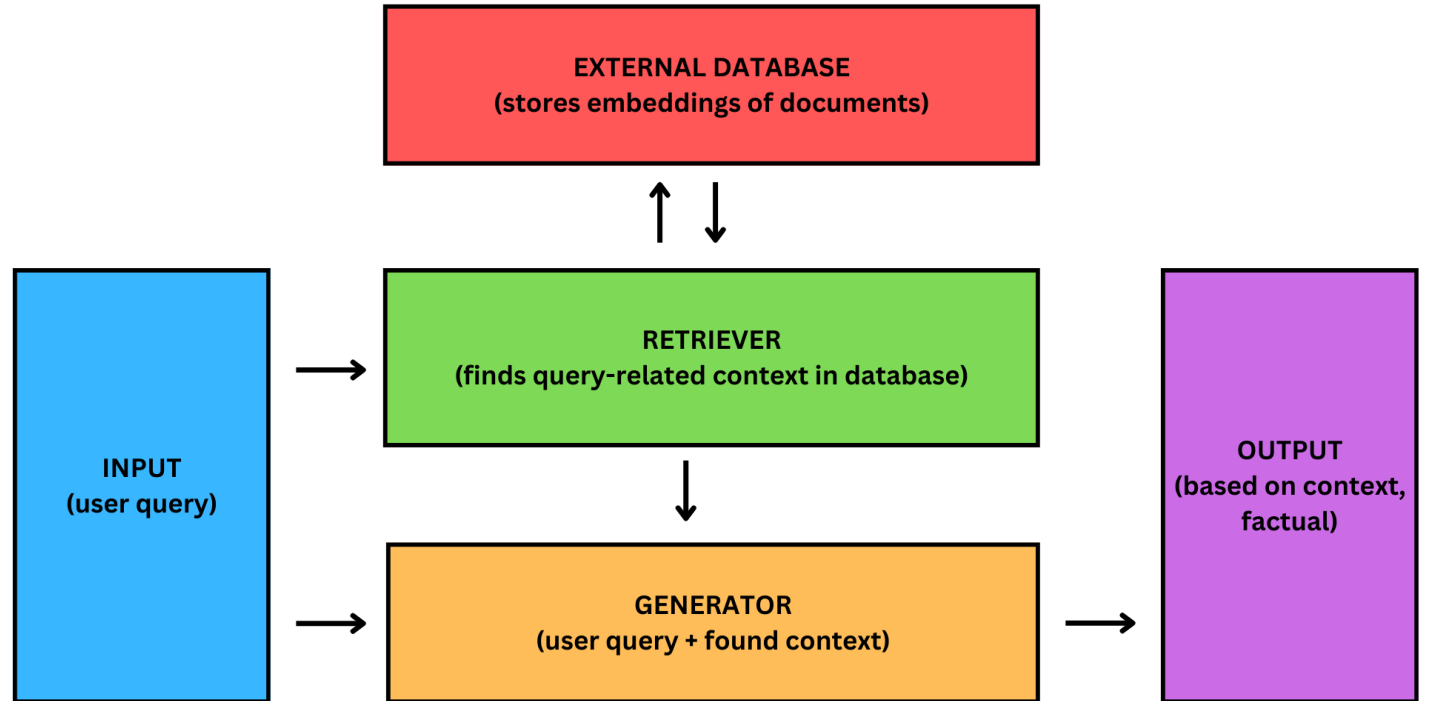


# Retrieval-Augmented Generation implementation for Wikipedia API

Creating specific knowledge assistant

# A brief recap of what is RAG



Knowledge  
base

The diagram features a large dark gray arrow pointing from left to right. Inside the arrow is the text 'Knowledge base'. To the right of the arrow's tip is a central light green box labeled 'Wikipedia Articles'. Surrounding this central box are five other colored boxes, each containing a feature: a red box for 'Similar style & structure' (top-left), a purple box for 'reliability' (top), a cyan box for 'easy access' (top-right), a yellow box for 'wide range of topics' (bottom-right), and a blue box for 'Many tables' (bottom).

Similar style &  
structure

reliability

easy access

Wikipedia  
Articles

wide range  
of topics

Many  
tables

Feedback we got  
last time

---

# Topic

---

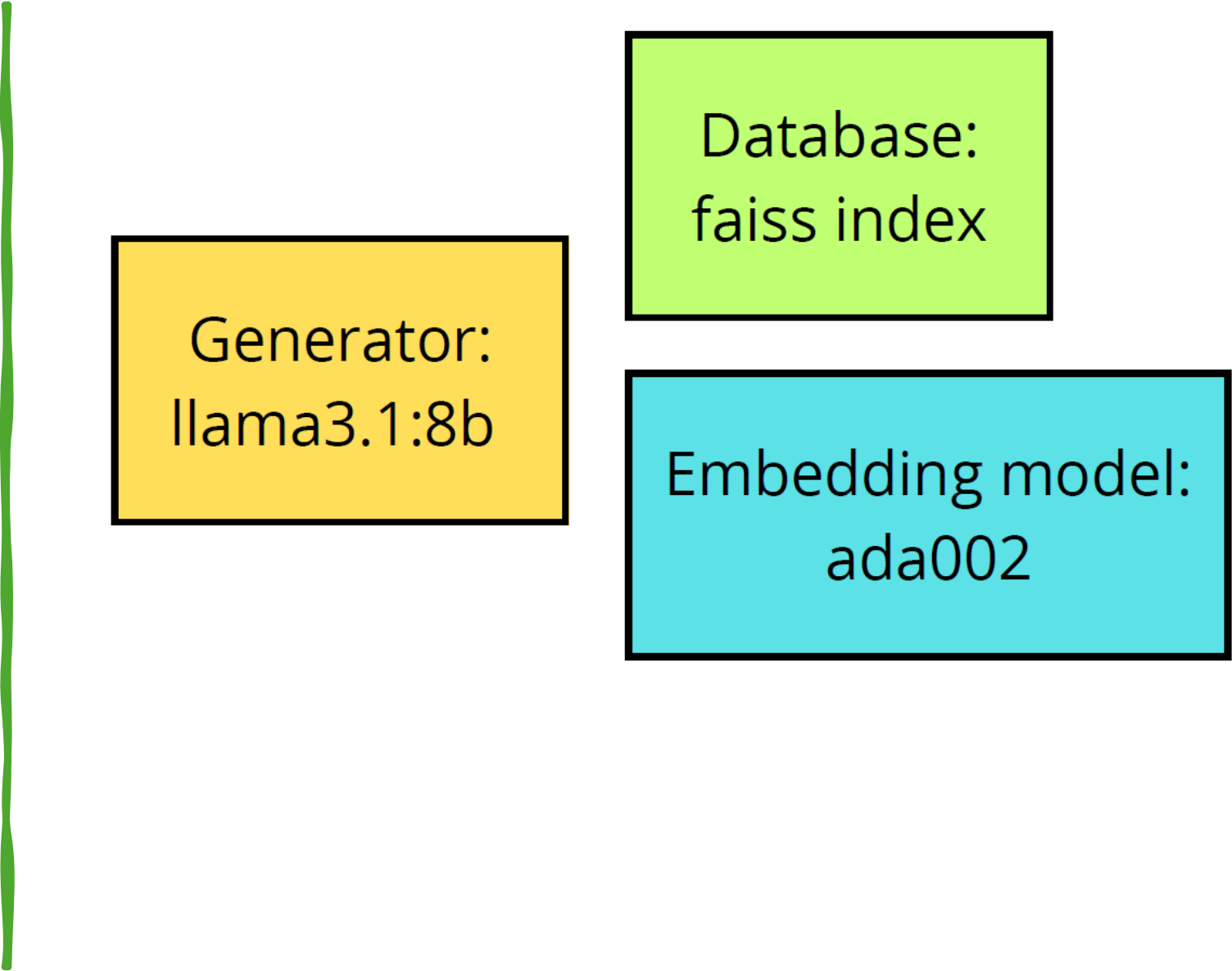
Star Wars  
Universum

```
graph LR; A[Star Wars Universum] --> B[Recent politics in Russia]; A --> C[Paris Olympics 2024];
```

Recent politics  
in Russia

Paris Olympics  
2024

In our  
project:



Generator:  
llama3.1:8b

Database:  
faiss index

Embedding model:  
ada002

# Creating the database from Wikipedia articles

Recursive search

Subcategories  
with more  
articles



Articles  
directly in  
category



## Category:2024 Summer Olympics

Category [Talk](#)

From Wikipedia, the free encyclopedia

### Subcategories

This category has the following 9 subcategories, out of 9 total.

#### 0–9

- ▶ [Coaches at the 2024 Summer Olympics](#) (1 C, 186 P)
- ▶ [2024 Summer Olympics templates](#) (5 C, 2 P)

#### B

- ▶ [2024 Summer Olympics bids](#) (11 P)

#### C

- ▶ [Competitors at the 2024 Summer Olympics](#) (37 C)

#### E

- ▶ [Events at the 2024 Summer Olympics](#) (38 C, 36 P)

#### N

- ▶ [Nations at the 2024 Summer Olympics](#) (1 C, 206 P)

### Pages in category "2024 Summer Olympics"

The following 20 pages are in this category, out of 20 total. [This list may not reflect recent changes.](#)

#### 0–9

- [142nd IOC Session](#)
- [2024 Summer Olympics](#)

#### C

- [2024 Summer Olympics and Paralympics cauldron](#)
- [Chronological summary of the 2024 Summer Olympics](#)
- [2024 Summer Olympics closing ceremony](#)
- [2024 Summer Olympics closing ceremony flag bearers](#)
- [Concerns and controversies at the 2024 Summer Olympics](#)

#### L

- [List of 2024 Summer Olympics broadcasters](#)

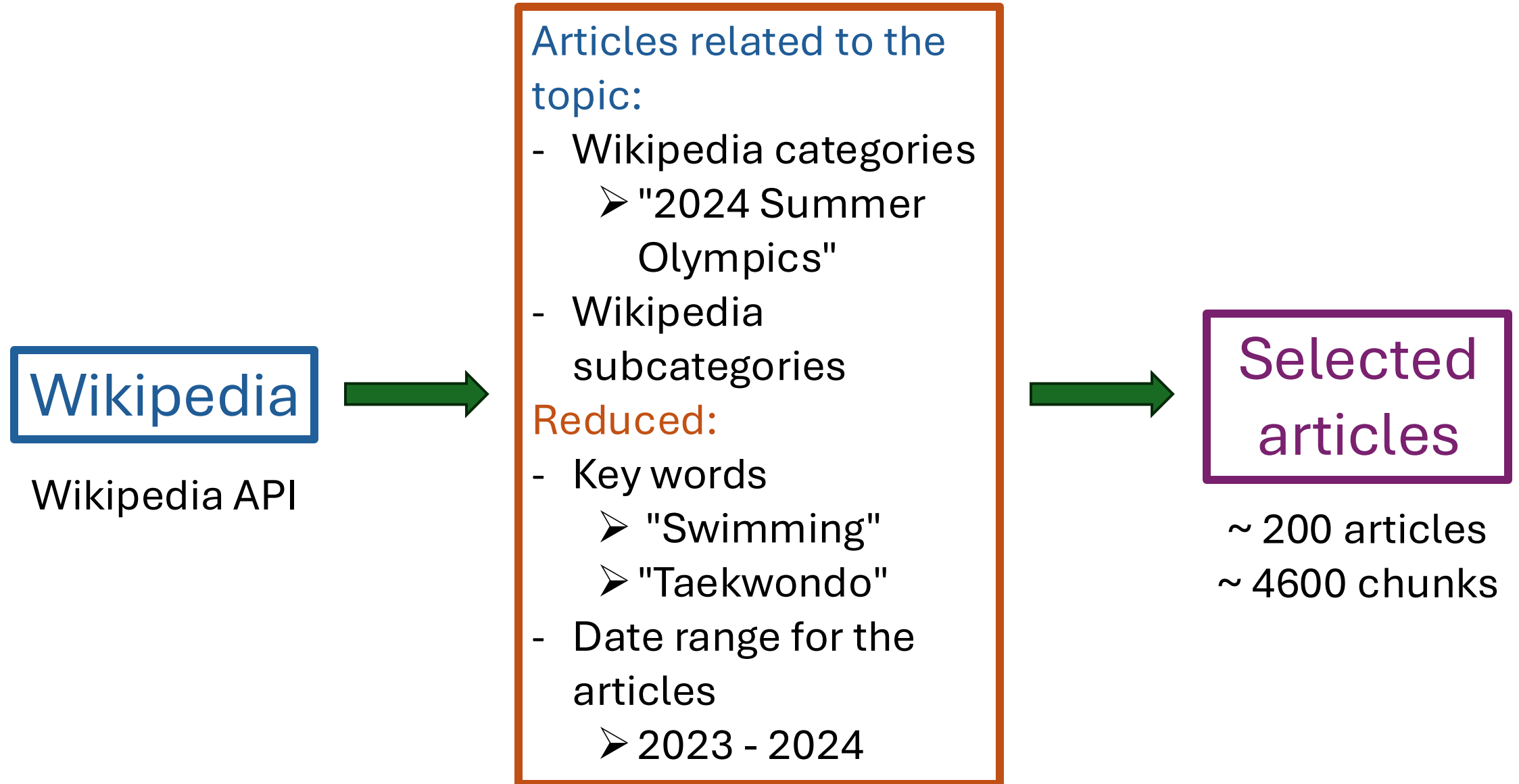
#### M

- [2024 Summer Olympics medal table](#)

#### O

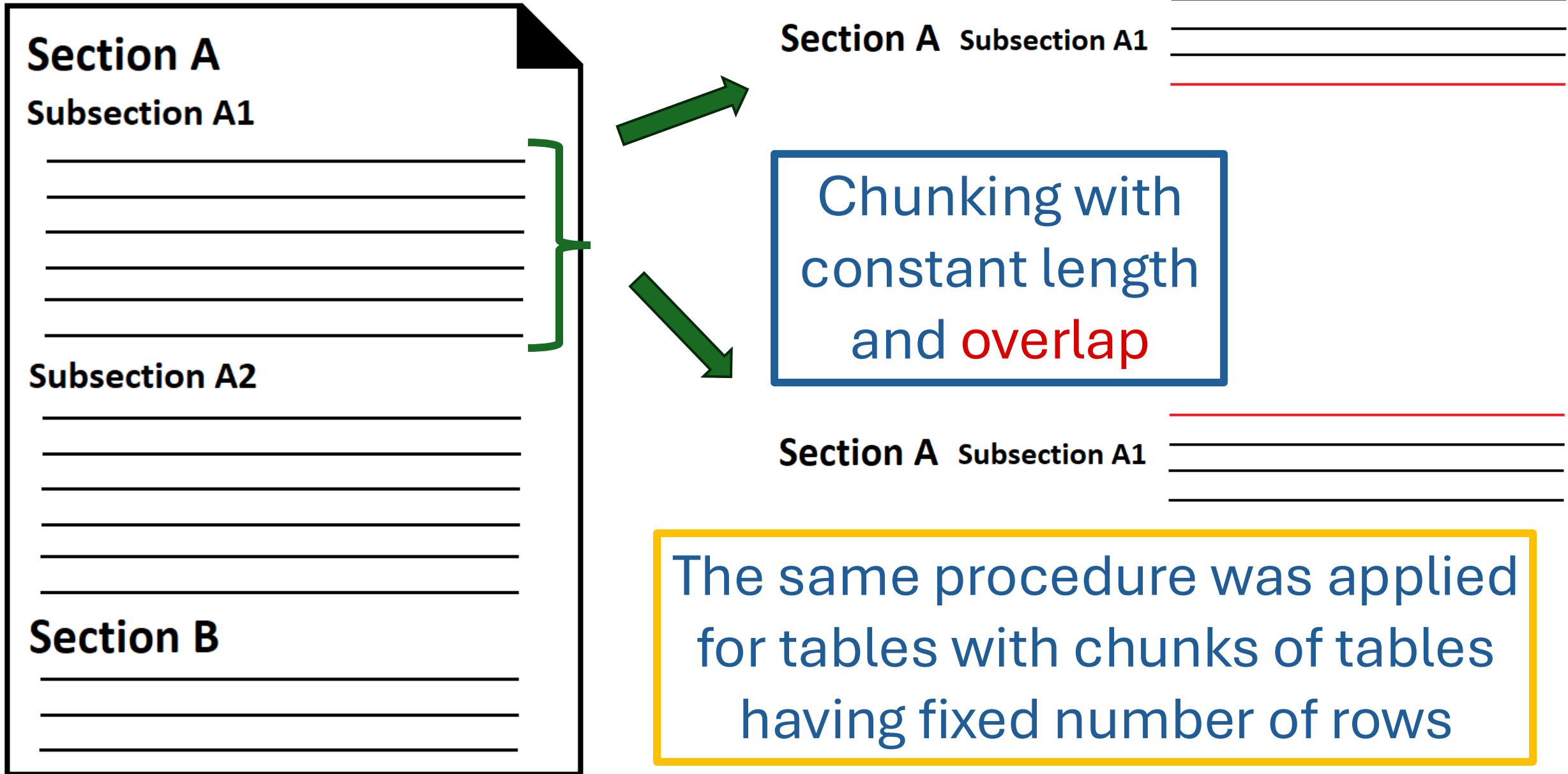
- [Olympic Village \(Paris\)](#)
- [2024 Summer Olympics opening ceremony](#)

# Creating the database from Wikipedia articles





# Chunking process for text



# Results

**Question:** What was the key outcome of the Russia-Africa Summit held in July 2023?

## RAG Answer

According to the context provided, **many agreements were signed** with participating African countries on July 28, 2023. The summit concluded with a final **declaration and an official plan for implementing a Partnership Forum for 2023-2026**, along with other documents. Therefore, the key outcome of the Russia-Africa Summit held in July 2023 **was the signing of multiple agreements** between Russia and African countries, as well as the adoption of a final declaration and a **partnership framework for the next few years**.

## Correct Answer

The summit aimed to strengthen ties between Russia and African countries, focusing on economic cooperation, trade, and military support.

# Results

**Question:** How many teams will participate in the men's football tournament at the 2024 Summer Olympics, and how is qualification achieved?

## RAG Answer

A total of **16 teams** will participate in the men's football tournament at the 2024 Summer Olympics.

Qualification is achieved as follows:

- \* The **host nation, France, earned automatic qualification.**
- \* From the six continental confederations:
  - + **AFC** (Asian Football Confederation): **3** automatic slots + **1 additional slot** through a play-off match against CAF
  - + **CAF** (Africa): **3** automatic slots + **1 additional slot** through a play-off match against AFC
  - + **UEFA** (Union of European Football Associations): **3** slots
  - + **CONCACAF** (Confederation of North, Central America and Caribbean Association Football) and CONMEBOL (South American Football Confederation): **2** slots each
  - + **OFC** (Oceania Football Confederation): **1** slot
  - + **CONMEBOL : 2 slot**

## Correct Answer

The men's football tournament at the 2024 Summer Olympics in Paris will feature **16 teams**. **France automatically qualifies** as the host nation.

Each regional confederation organizes qualification tournaments or designates existing competitions to determine its representatives. The allocation is as follows:

- AFC (Asia) **3**
- CAF (Africa) **3**
- CONCACAF (North & Central America, Caribbean) **2**
- CONMEBOL (South America) **2**
- OFC (Oceania) **1**
- UEFA (Europe) **3**
- AFC-CAF play-off **1**

# Results - tables

**Question:** How many medals in total did Italy acquire in swimming competitions in 2024 Olympics? How many of them were gold?

## RAG Answer

Italy acquired a total of 6 medals in swimming competitions at the 2024 Olympics.

As for how many gold medals Italy acquired, according to the table data provided:

Rank	Nation	Gold	Silver	Bronze	Total
7	Italy	2	1	3	6

So, Italy won 2 gold medals.

## Actual retrieval

Swimming at the 2024 Summer Olympics Medal summary


Start of table data.

Rank	Nation	Gold	Silver	Bronze	Total
...					
6	China	2	3	7	12
7	Italy	2	1	3	6
8	Sweden	2	0	0	2
...					

End of table data.

## Medal summary [\[edit\]](#)

### Medal table [\[edit\]](#)

 Host nation ([France](#))

Rank ↕	Nation ↕	Gold ↕	Silver ↕	Bronze ↕	Total ↕
1	 <a href="#">United States</a>	8	13	7	28
2	 <a href="#">Australia</a>	7	9	3	19
3	 <a href="#">France*</a>	4	1	2	7
4	 <a href="#">Canada</a>	3	2	3	8
5	 <a href="#">Hungary</a>	3	1	1	5
6	 <a href="#">China</a>	2	3	7	12
7	 <a href="#">Italy</a>	2	1	3	6
8	 <a href="#">Sweden</a>	2	0	0	2
9	 <a href="#">Great Britain</a>	1	4	0	5
10	 <a href="#">Germany</a>	1	1	1	3
11	 <a href="#">South Africa</a>	1	1	0	2

# Results - tables

**Question:** Name all candidates who took part in 2023 Primorsky election. Were there any polls? Who had the best chance to win and who won?

## RAG Answer (part 1)

Following candidates took part in the 2023 Primorsky Krai gubernatorial election:

1. Oleg **Kozhemyako** (United Russia) - incumbent Governor of Primorsky Krai
2. Inna **Kosheleva** (Liberal Democratic Party) - Member of Duma of Vladivostok
3. Aleksandr **Andoni** (Communists of Russia) - political scientist, journalist
4. Denis **Kovalyov** (Cossack Party) - Ussuriysk Federation of Hand-to-Hand Combat president
5. Kirill **Rudyuk** (A Just Russia — For Truth) - businessman

Yes, there were **polls** conducted in **August 2023** by **WCIOM**, which showed the following results:

### Polls [\[ edit \]](#)

Fieldwork date	Polling firm	Kozhemyako	Kosheleva	Andoni	Kovalyov	Rudyuk
18–24 August 2023	<a href="#">WCIOM</a> <sup>↗</sup>	76%	10%	7%	3%	2%

### Results [\[ edit \]](#)

Summary of the 8–10 September 2023 Primorsky Krai gubernatorial election results

Candidate	Party	Votes	%
<b>Oleg Kozhemyako (incumbent)</b>	<a href="#">United Russia</a>	<b>476,584</b>	<b>72.78</b>
Inna Kosheleva	<a href="#">Liberal Democratic Party</a>	64,392	9.83
Aleksandr Andoni	<a href="#">Communists of Russia</a>	38,892	5.94
Kirill Rudyuk	<a href="#">A Just Russia — For Truth</a>	31,245	4.77
Denis Kovalyov	<a href="#">Cossack Party</a>	25,503	3.89
Valid votes		636,616	97.22
Blank ballots		18,194	2.78
<b>Total</b>		<b>654,810</b>	<b>100.00</b>
Turnout		654,810	45.55
Registered voters		1,437,494	100.00
Source:		<a href="#">[19]</a>	

# Results - tables

**Question:** Name all candidates who took part in 2023 Primorsky election.  
Were there any polls? Who had the best chance to win and who won?

## RAG Answer (part 2)

Candidate	Support
Oleg Kozhemyako	76%
Inna Kosheleva	10%
Aleksandr Andoni	7%
Denis Kovalyov	3%
Kirill Rudyuk	2%

According to the polls, Oleg Kozhemyako had a significant lead over the other candidates.

The election results were as follows:

1. Oleg Kozhemyako (United Russia) - 476,584 votes (72.78%)
2. Inna Kosheleva (Liberal Democratic Party) - 64,392 votes (9.83%)
3. Aleksandr Andoni (Communists of Russia) - 38,892 votes (5.94%)
4. Denis Kovalyov (Cossack Party) - 25,503 votes (3.89%)
5. Kirill Rudyuk (A Just Russia — For Truth) - 31,245 votes (4.77%)

Oleg Kozhemyako won the election with a significant majority.

## Polls [\[edit\]](#)

Fieldwork date	Polling firm	Kozhemyako	Kosheleva	Andoni	Kovalyov	Rudyuk
18–24 August 2023	<a href="#">WCIOM</a>	76%	10%	7%	3%	2%

## Results [\[edit\]](#)

Summary of the 8–10 September 2023 Primorsky Krai gubernatorial election results

Candidate	Party	Votes	%
<b>Oleg Kozhemyako (incumbent)</b>	<a href="#">United Russia</a>	476,584	72.78
Inna Kosheleva	<a href="#">Liberal Democratic Party</a>	64,392	9.83
Aleksandr Andoni	<a href="#">Communists of Russia</a>	38,892	5.94
Kirill Rudyuk	<a href="#">A Just Russia — For Truth</a>	31,245	4.77
Denis Kovalyov	<a href="#">Cossack Party</a>	25,503	3.89
Valid votes		636,616	97.22
Blank ballots		18,194	2.78
<b>Total</b>		<b>654,810</b>	<b>100.00</b>
Turnout		654,810	45.55
Registered voters		1,437,494	100.00
Source:		<a href="#">[19]</a>	

# Results - tables

**Question:** Who won the gold medal in Taekwondo in Women's 67kg category?

## RAG Answer

Viviana Márton from Hungary won the gold medal in the Taekwondo Women's 67 kg category at the 2024 Summer Olympics. She defeated Aleksandra Perišić of Serbia with a score of 7-4 in the gold medal match.

### Taekwondo at the 2024 Summer Olympics – Women's 67 kg 🌐 17 languages

[Article](#) [Talk](#) [Read](#) [Edit](#) [View history](#) [Tools](#)

#### Results [\[ edit \]](#)

##### Gold medal match [\[ edit \]](#)

Gold medal match

12	<span><span></span></span> <a href="#">Viviana Márton</a> (HUN)	7	4	
3	<span><span></span></span> <a href="#">Aleksandra Perišić</a> (SRB)	1	2	

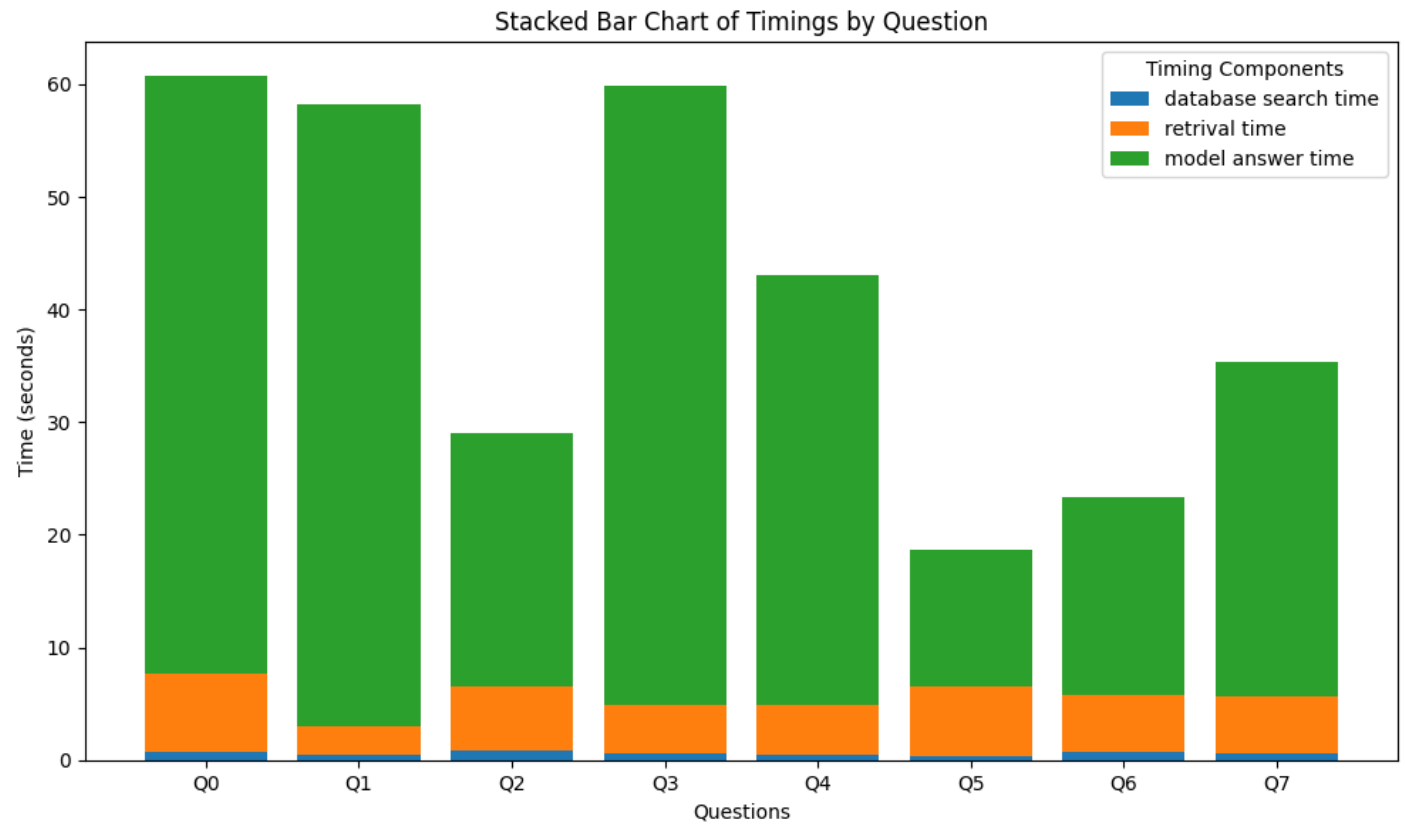
##### Repechage [\[ edit \]](#)

Repechage Bronze Medal

13	<span><span></span></span> <a href="#">Kristina Teachout</a> (USA)	0	8	9	7	<span><span></span></span> <a href="#">Song Jie</a> (CHN)	1	0	
5	<span><span></span></span> <a href="#">Ruth Gbagbi</a> (CIV)	3	5	3	13	<span><span></span></span> <a href="#">Kristina Teachout</a> (USA)	9	3	

Repechage Bronze Medal

Time





# Summary

- Creating the database of Wikipedia articles on a specified topic
- Successfully scraping the text of topic – related articles and tables with additional data
- Successful implementation of RAG and enforcing the ollama model with it
- Successfully extending the model knowledge beyond its training data

# References

- [1] Sebastian Borgeaud et al. “Improving Language Models by Retrieving from Trillions of Tokens”. In: *Proceedings of the 39th International Conference on Machine Learning*. Ed. by Kamalika Chaudhuri et al. Vol. 162. Proceedings of Machine Learning Research. PMLR, 17–23 Jul 2022, pp. 2206–2240. URL: <https://proceedings.mlr.press/v162/borgeaud22a.html>.
- [2] Sebastian Borgeaud et al. “Improving language models by retrieving from trillions of tokens”. In: *CoRR* abs/2112.04426 (2021). arXiv: 2112.04426. URL: <https://arxiv.org/abs/2112.04426>.
- [3] Gordon V. Cormack, Charles L A Clarke, and Stefan Buettcher. “Reciprocal rank fusion outperforms condorcet and individual rank learning methods”. In: *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2009.

# Questions

Thank you

Stanisław Kurzątkowski

Filip Mieszkowski