

Smart Answers, Smarter Context

The Power of RAG + Entity Recognition

Jiri Hana (Geneea)

Geneea

- Developing AI solutions for text processing since 2014
- 20+ team members
- Experience from academia, Google, Reuters, Amazon, Oracle, etc.



Geneea





Radio Télévision
Suisse

- Entities, IPTC Media topics
- Articles, social media posts, radio and TV shows



newton media

- Entities for media monitoring
- 2,000,000 articles per month



RAG: What & Why

RAG: Why

- Limits hallucinations by grounding the AI in facts
- Easy updates for private or new information

RAG by Geneea



World Association
of News Publishers

- 100+ page industry reports
- Presentations, webinars, panel discussions



CZECH NEWS AGENCY

- RAG-based researcher
- 7M archive



Keboola

- Documentation RAG for AI agents

RAG for News Media – Specifics

- Large archives
- Verification
- Internal power-users
- Retrieval more important than prose



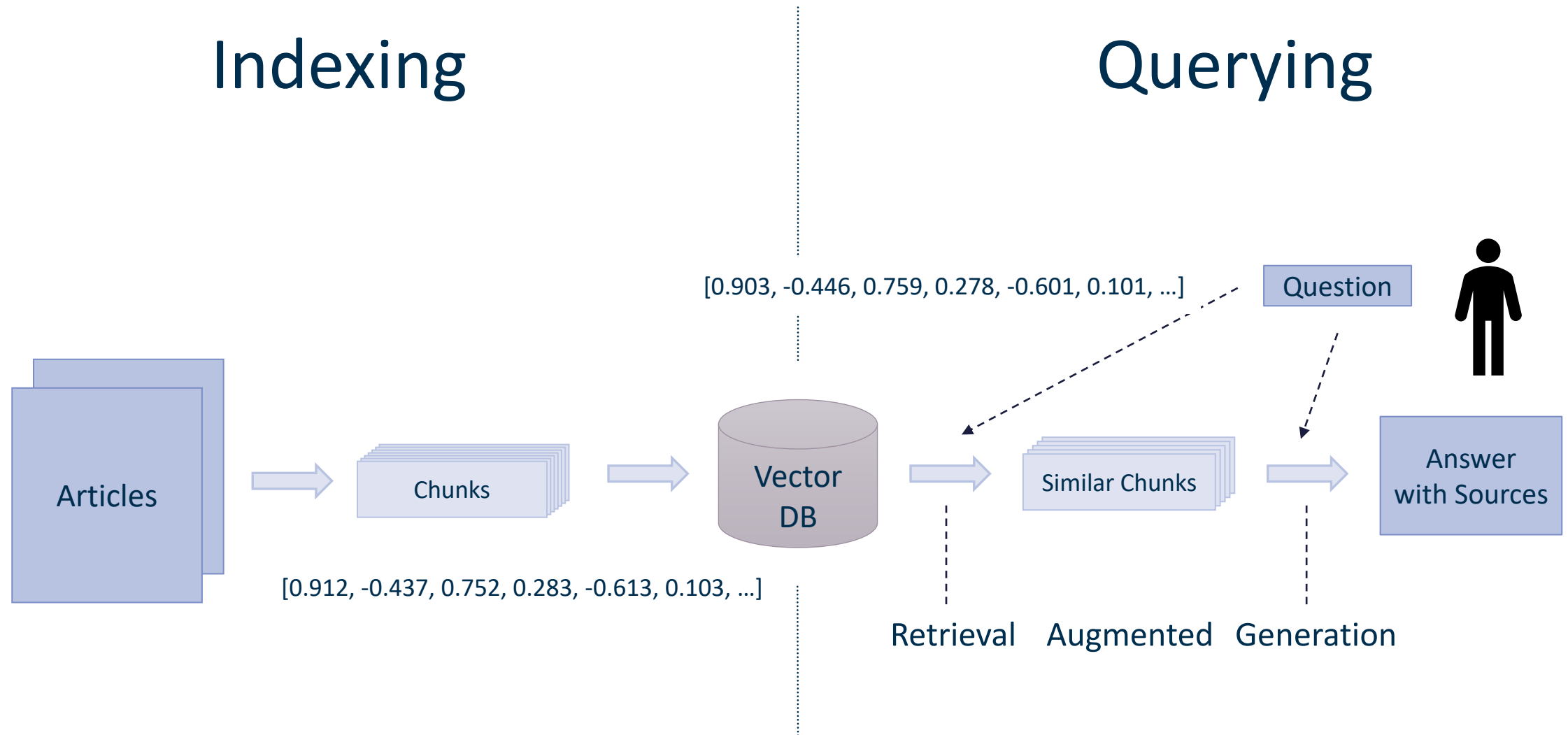
Naïve RAG

What & Why not

Naïve RAG

Indexing

Querying



Problem 1: Context

Small chunks are great for retrieval but not for context

*... Acme Corp announced the acquisition of Globex.
The \$500 million deal marks their largest expansion
into the European market. ...*

Q: How much did Acme pay for Globex?

Problem 1: Context

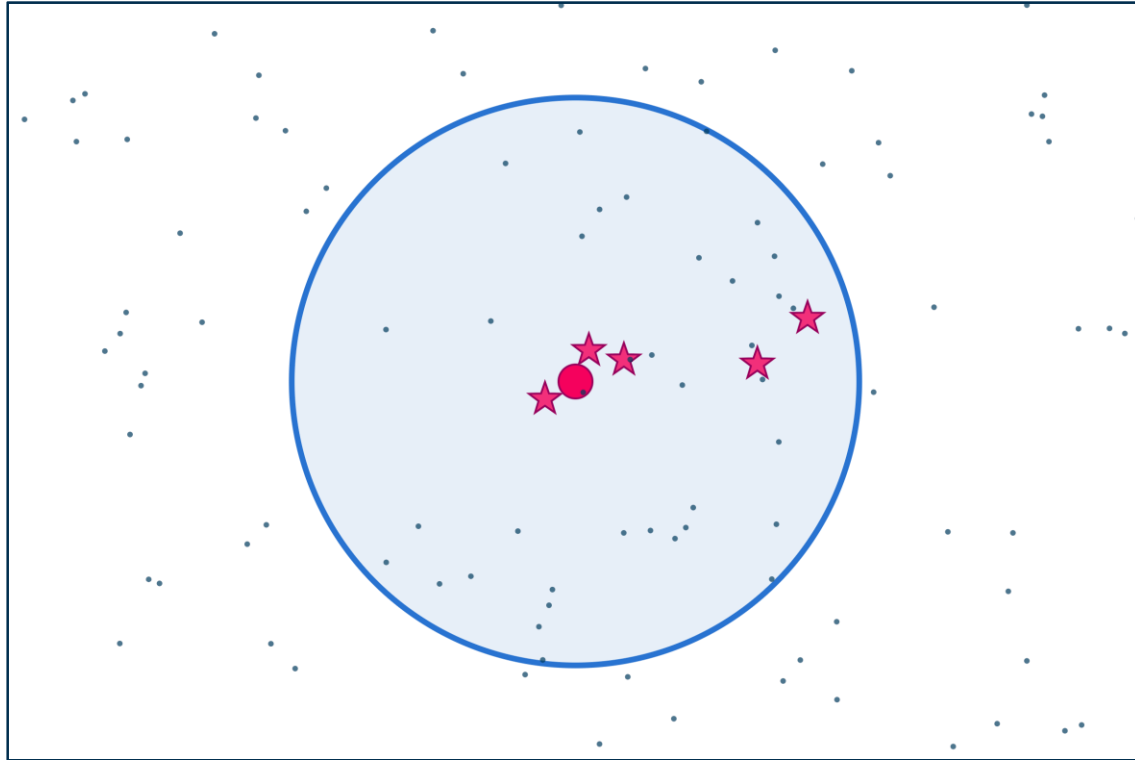
Small chunks are great for retrieval but not for context

... Acme Corp announced the acquisition of Globex.

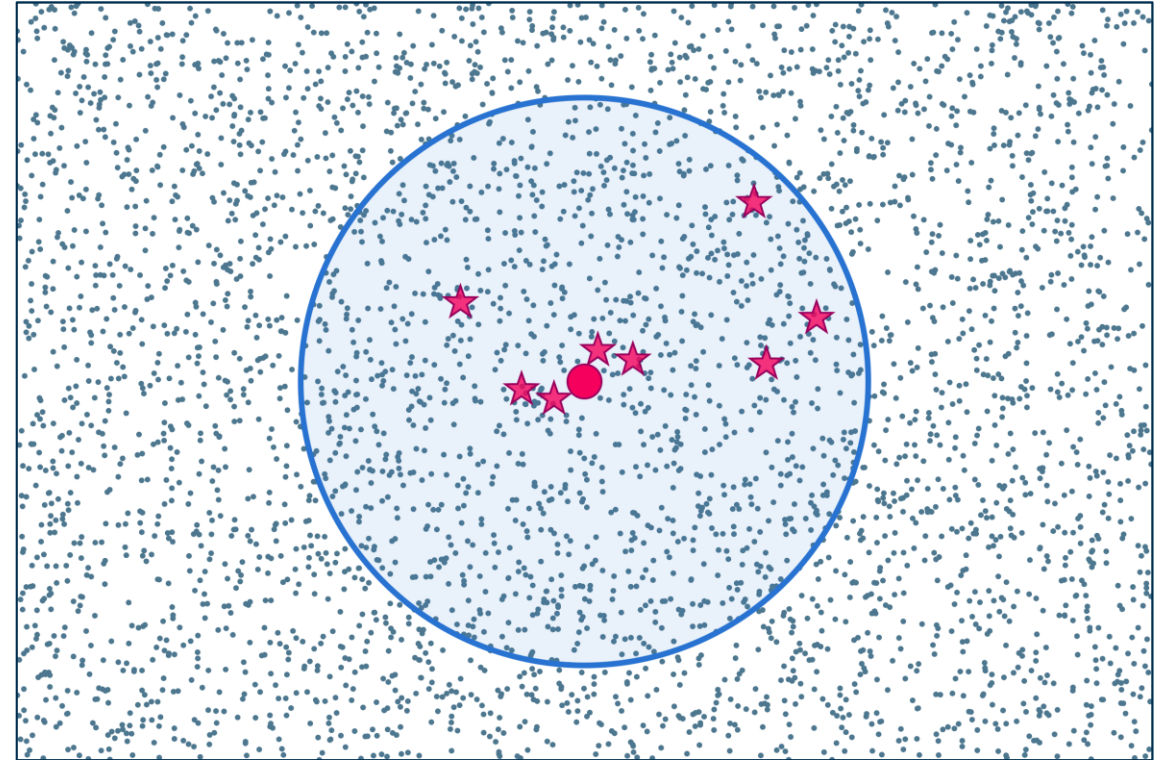
*The \$500 million **deal** marks **its** largest expansion into the European market. ...*

Q: How much did Acme pay for Globex?

Problem 2: Large Archive



Tutorial-size archive



Real-life archive

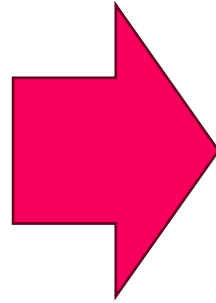


Pragmatic RAG

Fix Context

... Acme Corp announced the acquisition of Globex.

*The \$500 million **deal** marks **its** largest expansion into the European market. ...*



... Acme Corp announced the acquisition of Globex.

***Acme Corp's \$500 million acquisition of Globex** marks the company's largest expansion into the European market.*

Expensive & Slow!

Fix Context

... Acme Corp announced the acquisition of Globex.

*The \$500 million **deal** marks **its** largest expansion into the European market. ...*

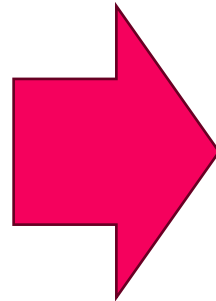
- Traditional NLP
- Late chunking
- Be selective

Efficient, cheap & fast

Improve Retrieval

... Acme Corp announced the acquisition of Globex.

The \$500 million deal marks its largest expansion into the European market. ...



Better vector matching

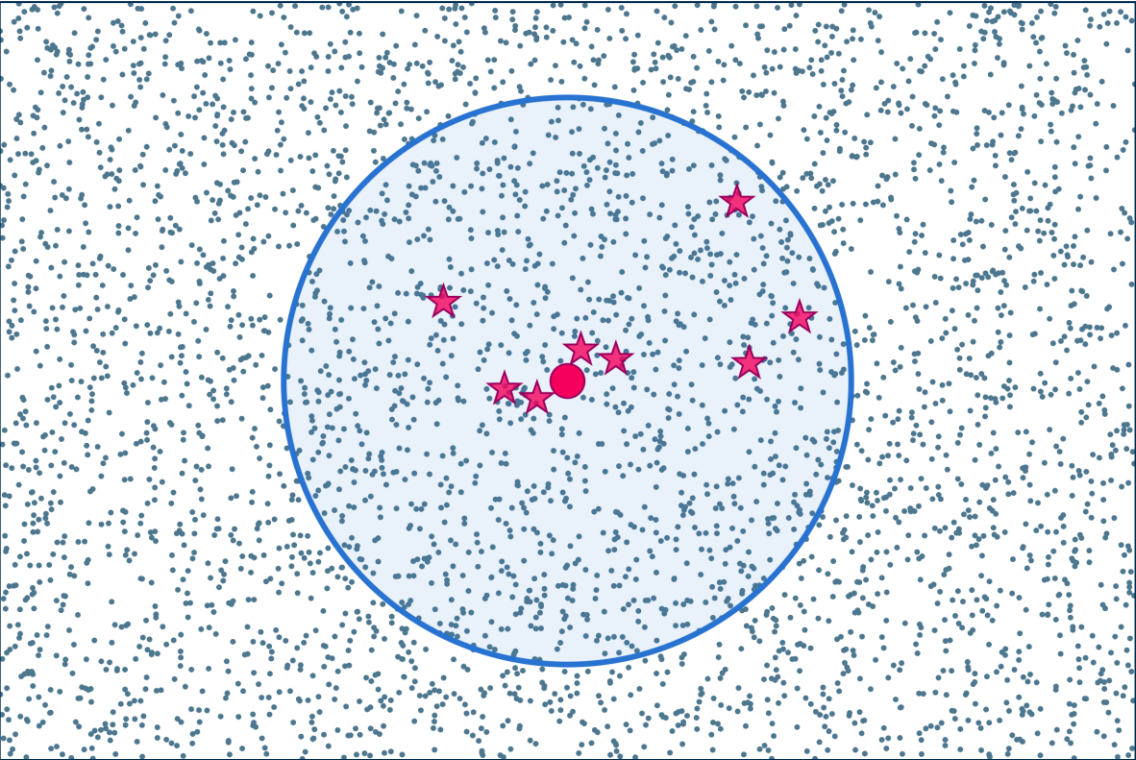
How is Acme Corp

Has Globex been acquired

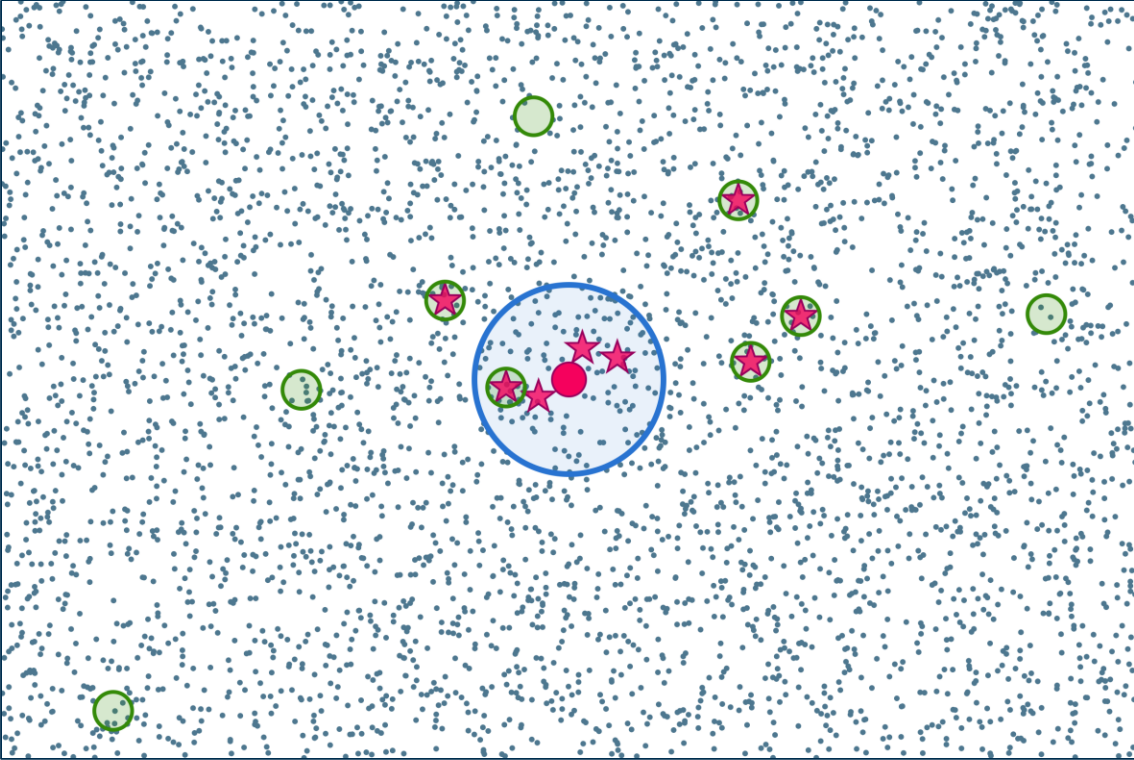
How much did Acme Corp pay to acquire Globex?

Expensive & Slow!

Improve Retrieval: Hybrid Search



Embedding Search



Hybrid Search

Improve Retrieval: Filtering

Metadata filtering makes a large archive small again

- Cheap & fast
 - But users don't do it
- ⇒ Automatically derive from the query

Improve Retrieval: Filter by Time

Q: How did the market react to the acquisition of Globex in 2020?

→ Ignore the articles published before 2020

Improve Retrieval: Filter by Time

Q: What was the GDP in 2020?

→ in an early 2021 article?

Q: What were the predictions for 2020?

→ in a 1950 article?

Use LLM

Improve Retrieval: Filter by Entities

Q: How much did Acme pay for Globex?

→ Must mention Acme Corp and Globex

Filter by Entities

Improve Retrieval: Filter by Entities

Vincent van Gogh lived and painted in Auvers-sur-Oise.

Q: Did van Gogh live in France?

Use Knowledge Graph

Improve Retrieval: Filter by Entities

Vincent van Gogh lived and painted in Auvers-sur-Oise.

Auvers-sur-Oise *G212406* 

commune in Val-d'Oise, France

CS 

DE 

EN 

FR 

NL 

 data

Derived Entities

Country

France *G142* 1x

Region

Île-de-France *G13917* 1x

District

Val-d'Oise *G12784* 1x

Improve Retrieval: Filter by Entities

Q: When did Einstein move across the Atlantic?

Q: When was the European NASA created?

Use LLM



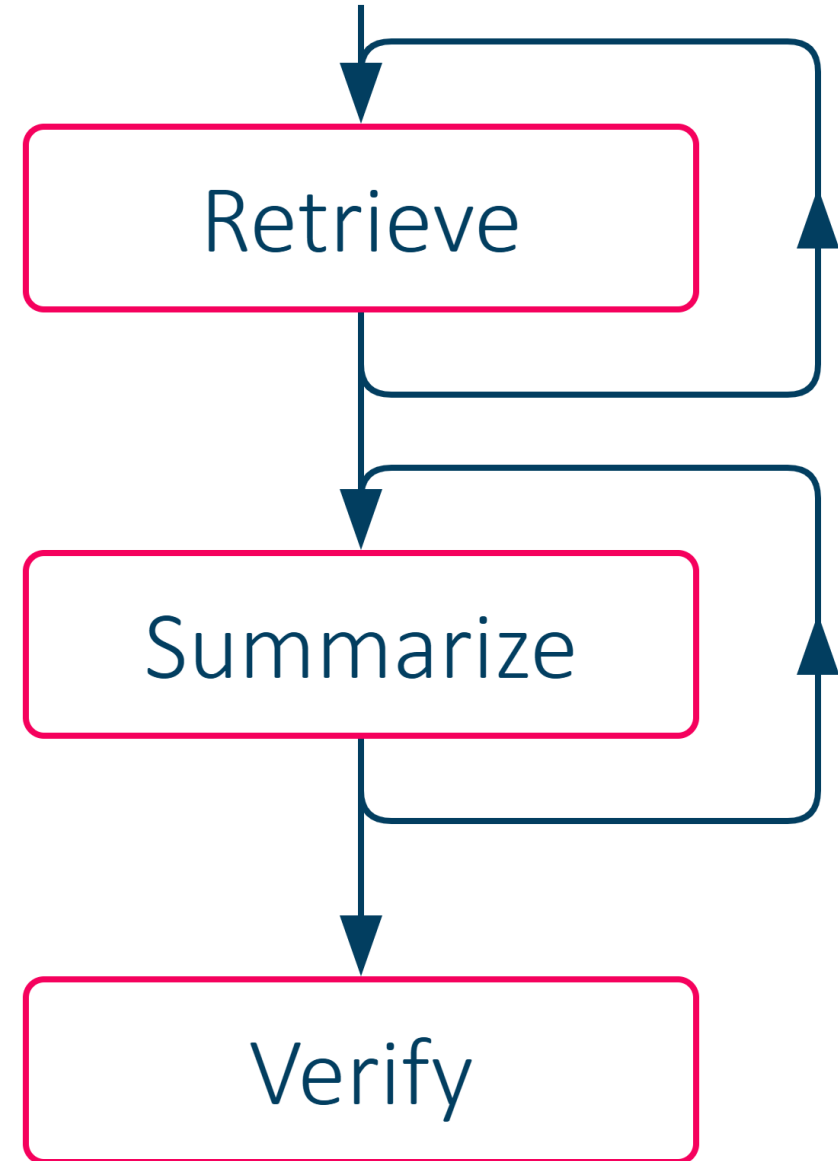
Advanced RAG

Reflective/Agentic RAG

- Reflective Retrieval:
 - AI decides to search again if it lacks context
 - Better results
 - Slower
- Agentic RAG:
 - Highly autonomous
 - Versatile
 - Costs and latency are completely unpredictable

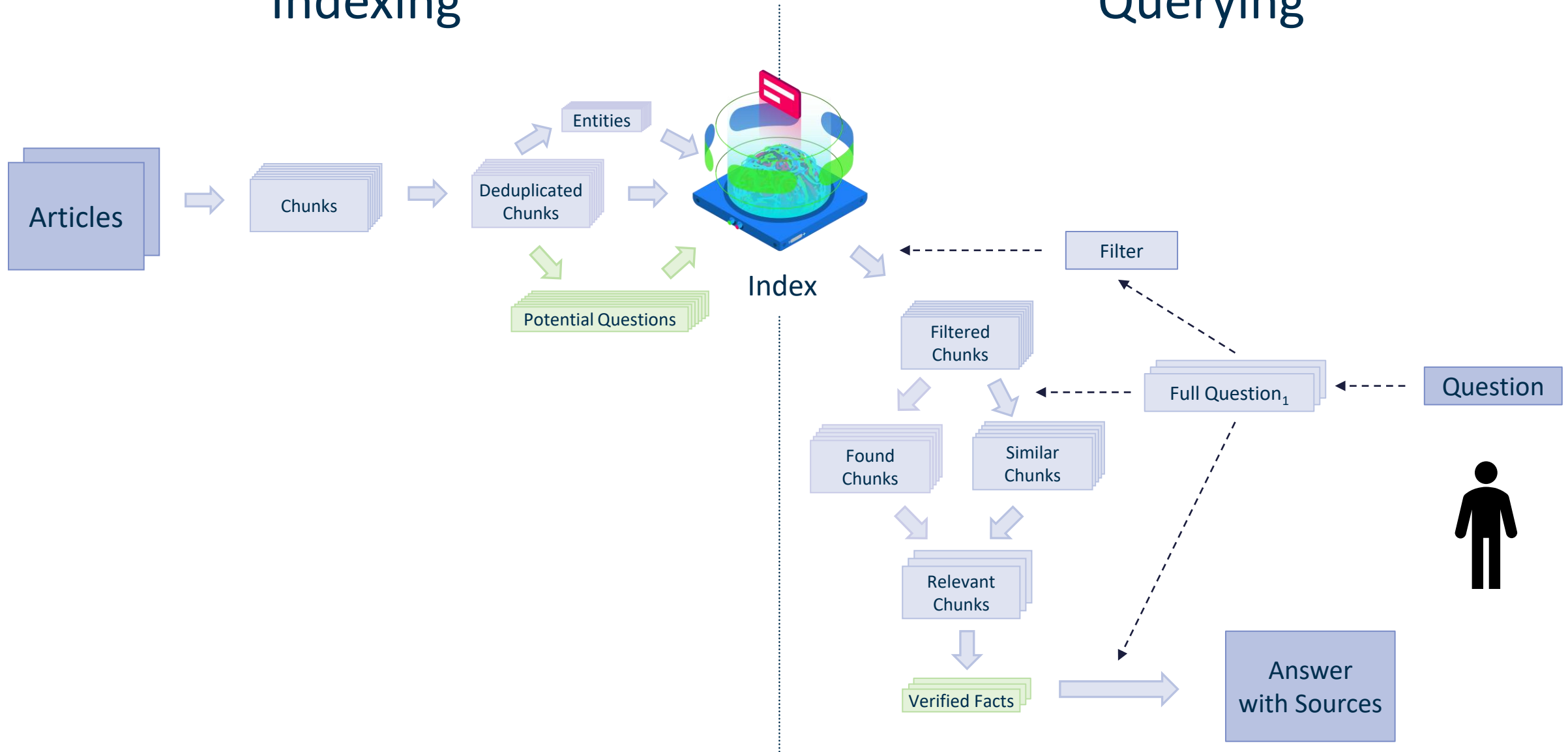
Iterative RAG

- User-driven iterations
- Fast and responsive



Indexing

Querying

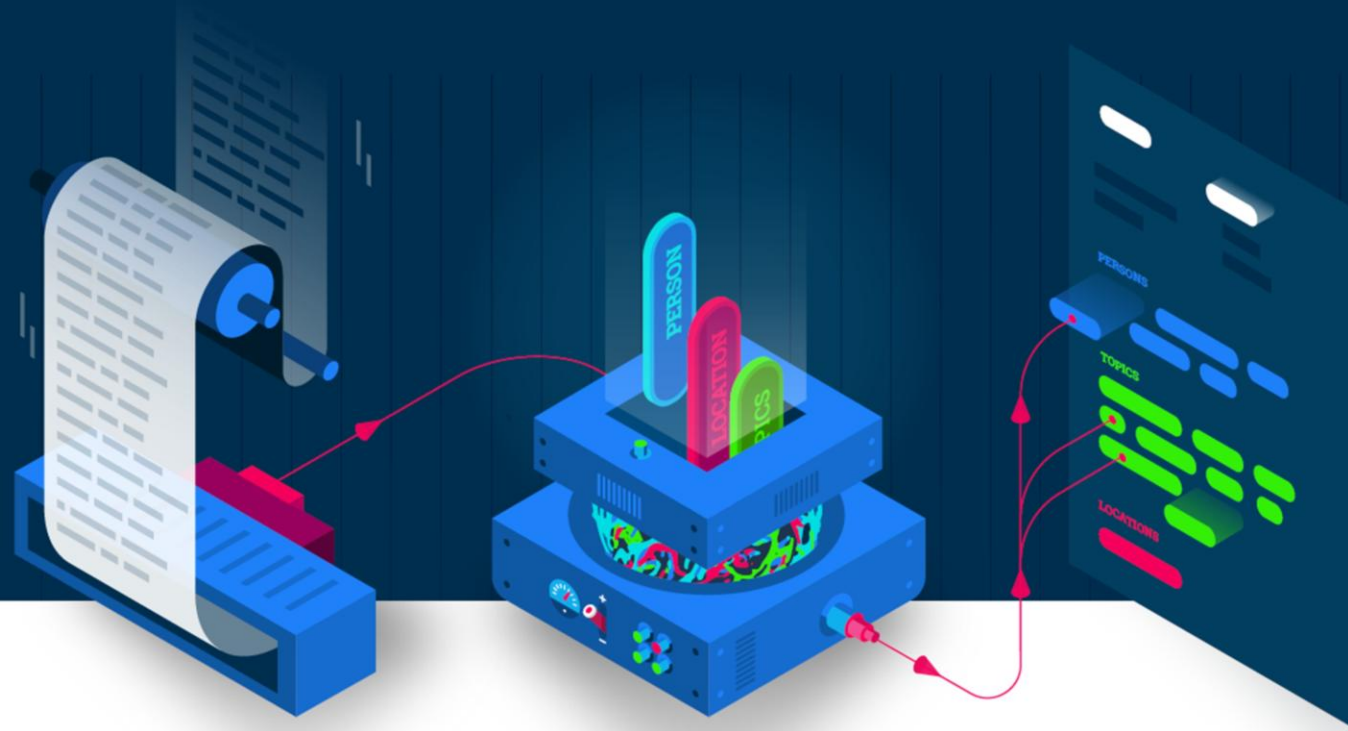


Other Challenges

- Attribution: Who actually said it?
The newspaper, the journalist, or the politician?
- Conflicting views
- Development over time

Conclusion

- RAG over large archives is not trivial
- Semantic search is not enough
- Automatic metadata filtering helps
- Scale requires structure
- Retrieval > Prose



Thank You!