Information Retrieval

Exercise – Winter term 2024/2025

tim.hagen@uni-kassel.de

Agenda

1. Project Status

2. Architecture of a Search Engine

3. Pooling

4. Measures

5. Assignment

Project Status

Each team, shortly give us a summary of
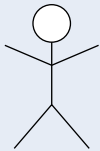how you perceived the annotation process.

Project Status

Each team, shortly give us a summary of
how you perceived the annotation process.

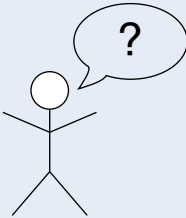What was the most difficult part?
What have you learned?

Architecture of a Search Engine

Karaoke Version
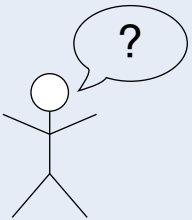
Architecture of a Search Engine

Karaoke Version

Architecture of a Search Engine
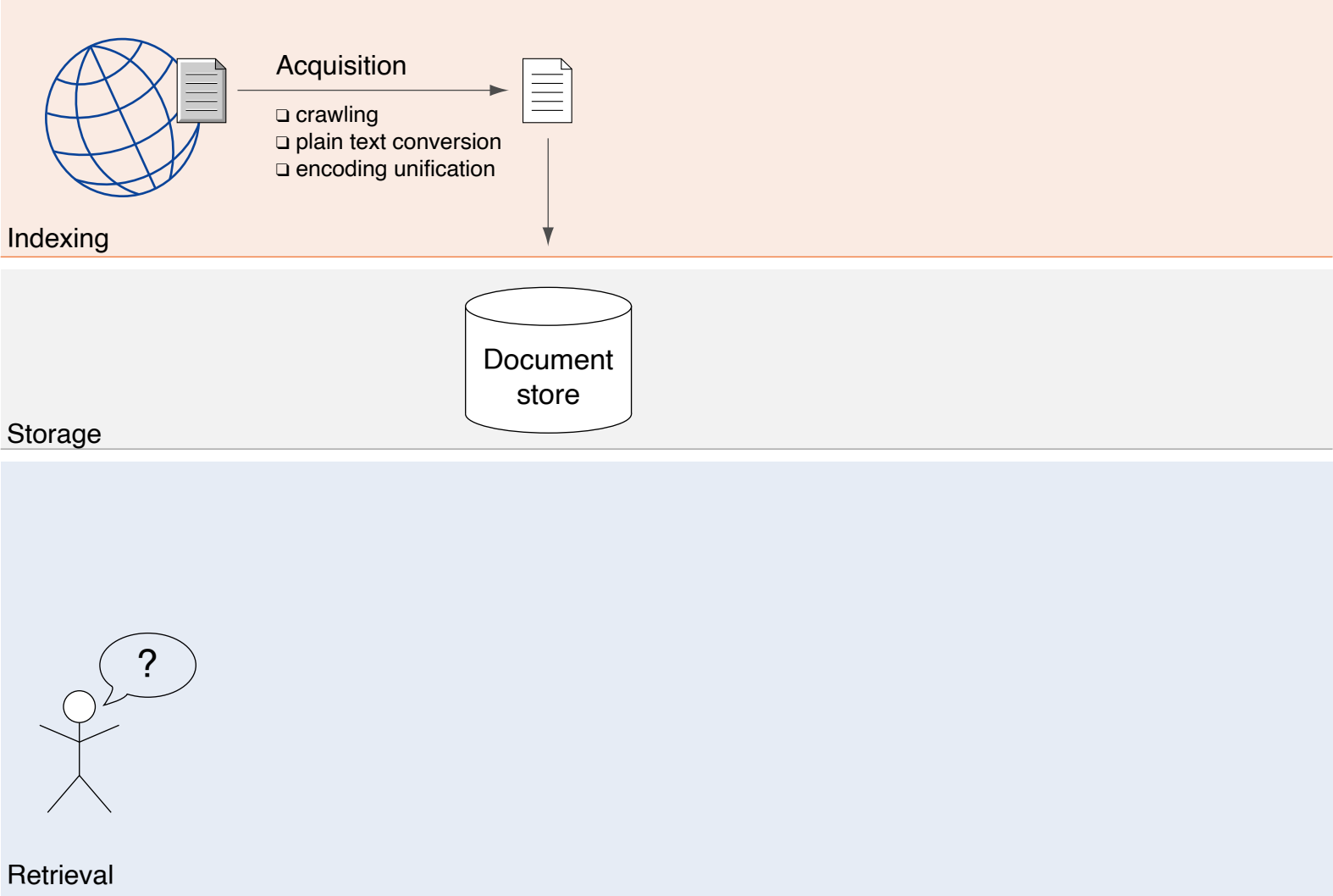
Karaoke Version
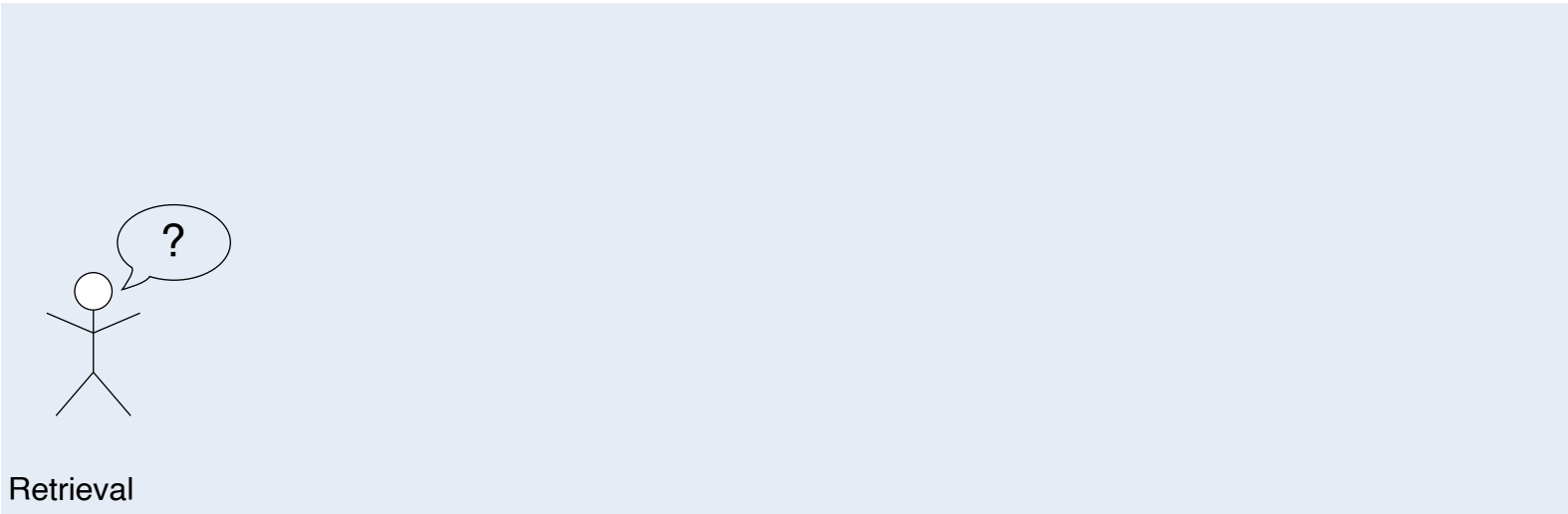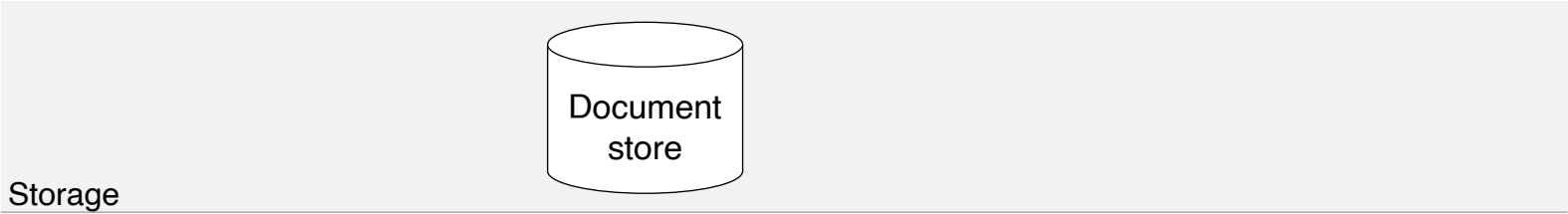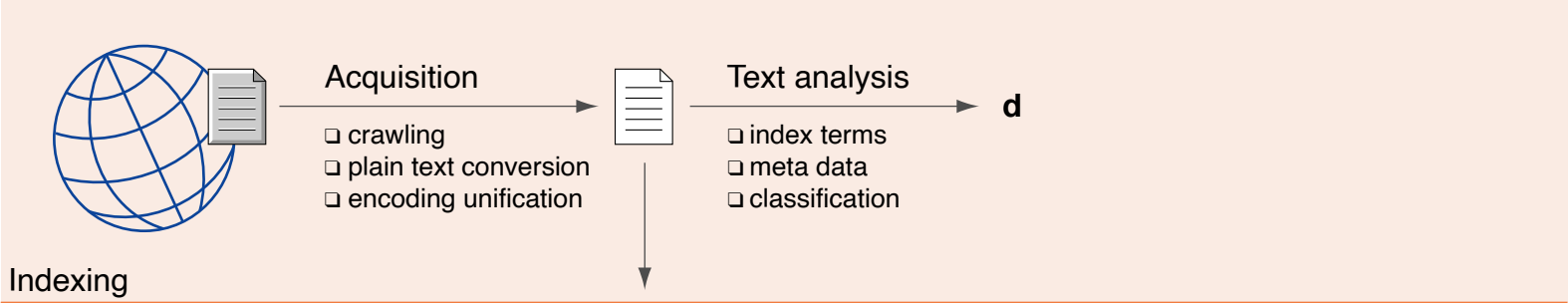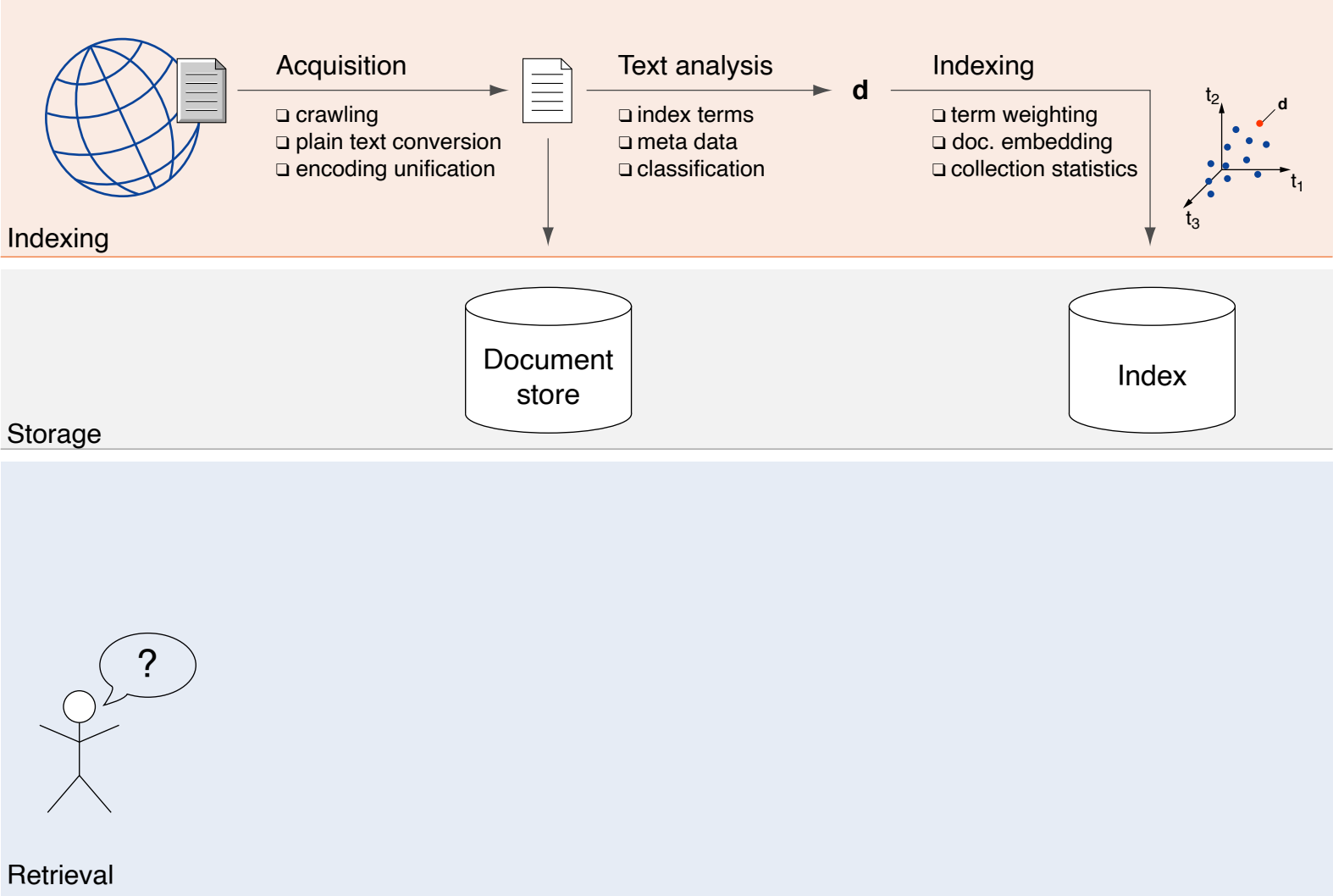

Indexing

Storage

Retrieval

Architecture of a Search Engine
Karaoke Version

Indexing

Acquisition

❑ crawling
❑ plain text conversion
❑ encoding unification

Storage

Document store

Retrieval

?

Architecture of a Search Engine
Karaoke Version

Indexing

Acquisition
- ❑ crawling
- ❑ plain text conversion
- ❑ encoding unification

Text analysis
- ❑ index terms
- ❑ meta data
- ❑ classification

d

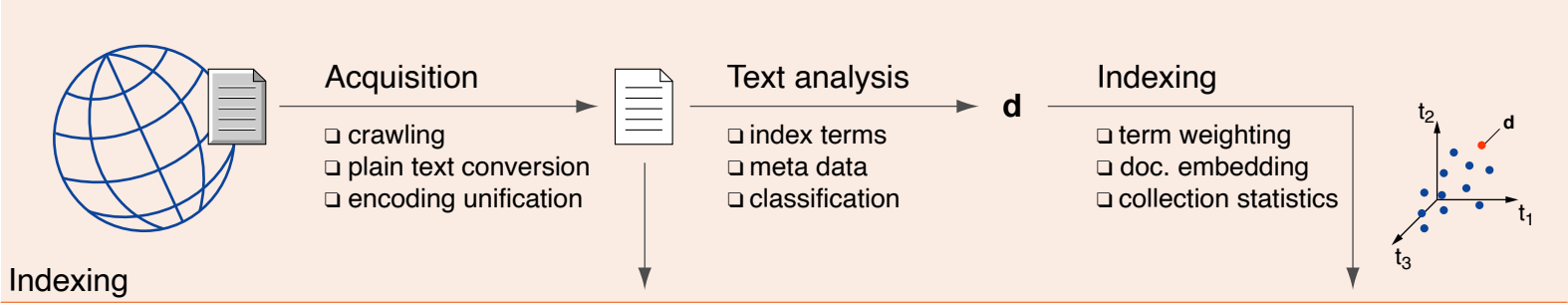**Storage**

Document store

Retrieval

?

Architecture of a Search Engine

Karaoke Version

Architecture of a Search Engine
Karaoke Version

Indexing

Acquisition
- ❑ crawling
- ❑ plain text conversion
- ❑ encoding unification

Text analysis
- ❑ index terms
- ❑ meta data
- ❑ classification

d

Indexing
- ❑ term weighting
- ❑ doc. embedding
- ❑ collection statistics

t_2 t_1 t_3 **d**

Storage

Document store

Index

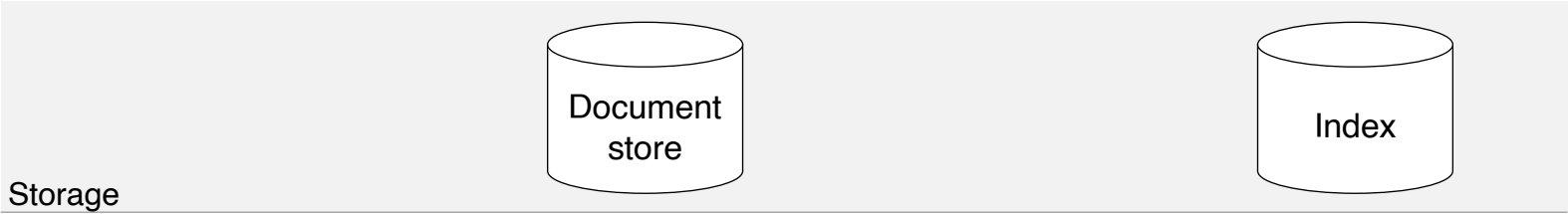Retrieval

query 🔍

Architecture of a Search Engine

Karaoke Version

Architecture of a Search Engine
Karaoke Version

Indexing

Acquisition
- ❏ crawling
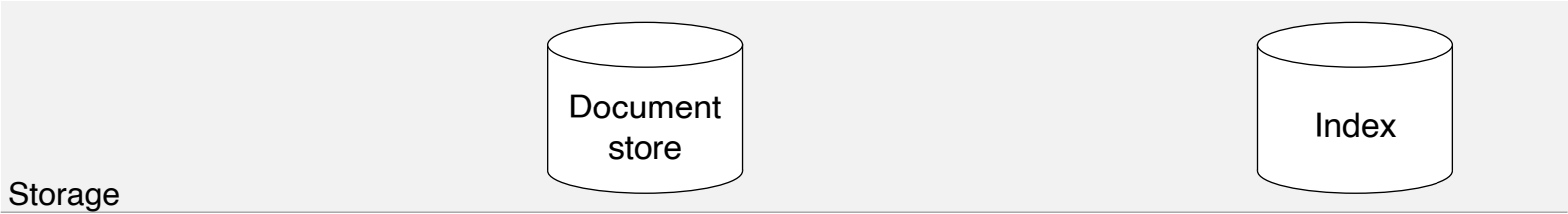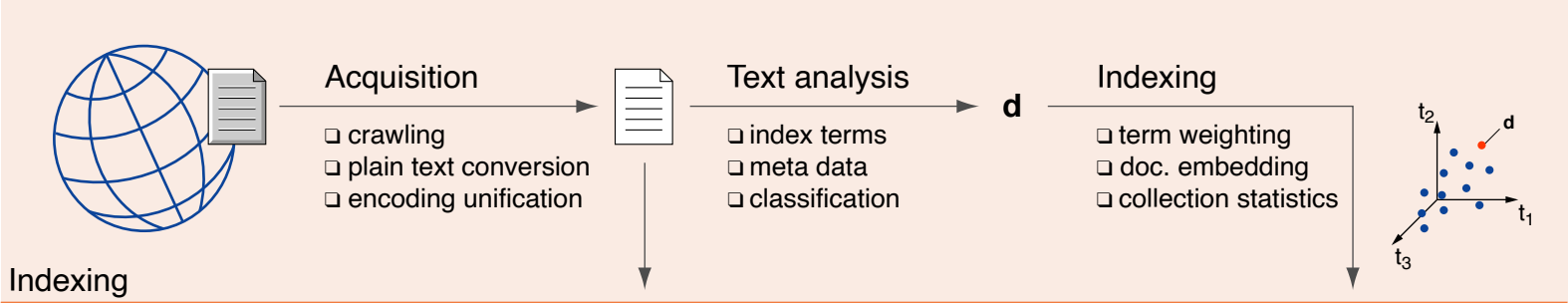- ❏ plain text conversion
- ❏ encoding unification

Text analysis
- ❏ index terms
- ❏ meta data
- ❏ classification

d

Indexing
- ❏ term weighting
- ❏ doc. embedding
- ❏ collection statistics

Storage

Document store

Index

Retrieval

Query analysis

q

query

Snippet generation

en.wikipedia.org › wiki › Query
Query - Wikipedia
In general, a **query** is a form of questioning, in a line of inquiry.

dictionary.com › browse › query
Query | Dictionary.com
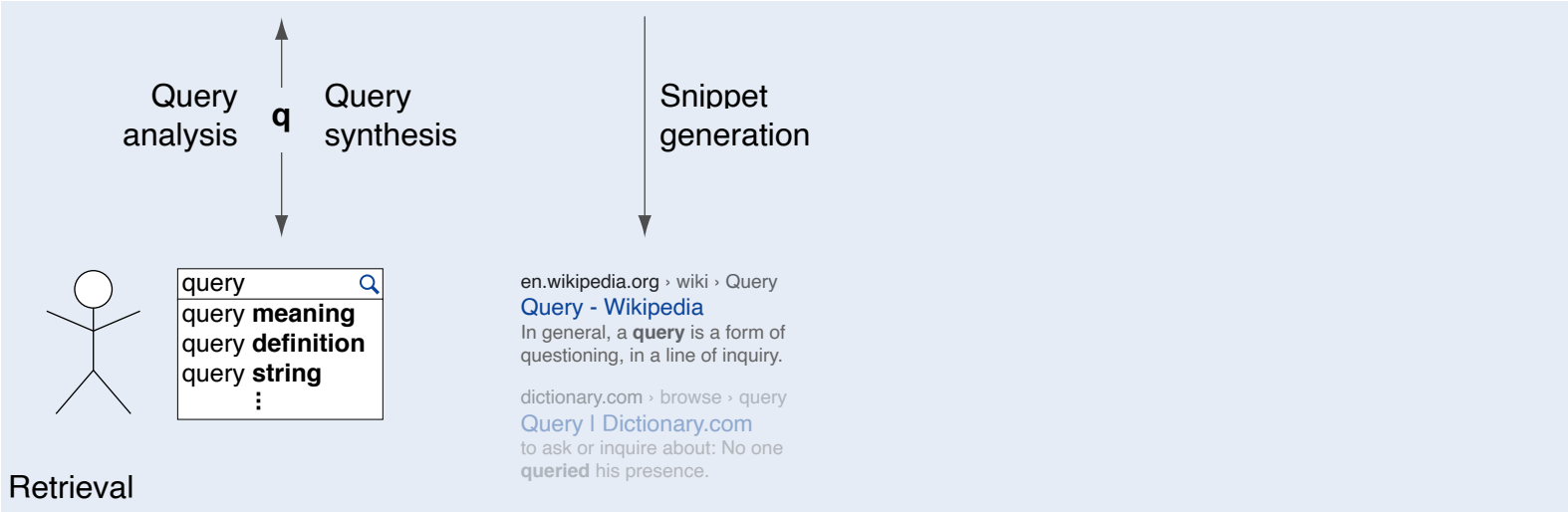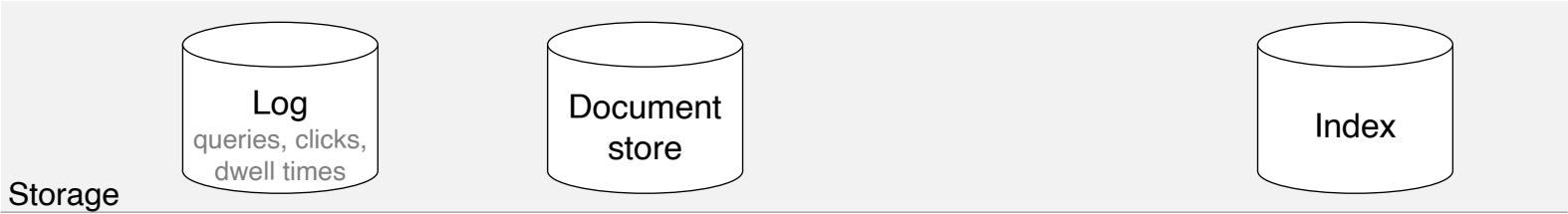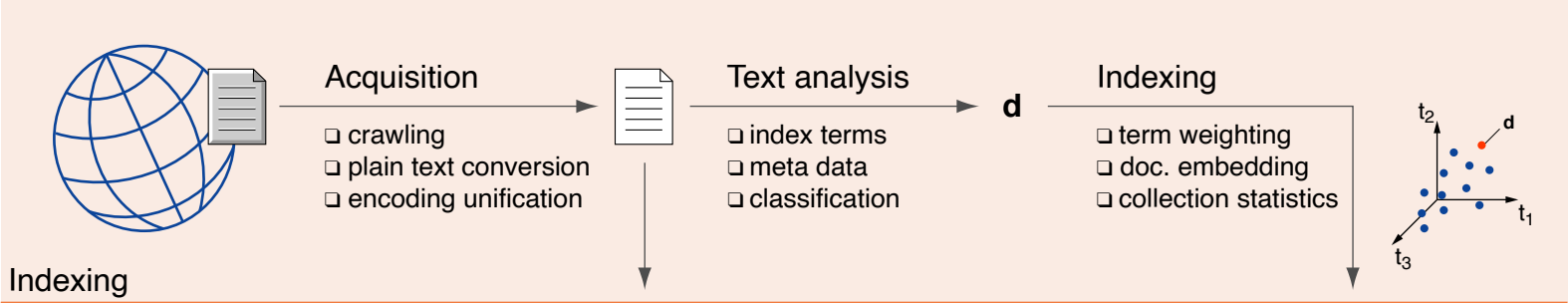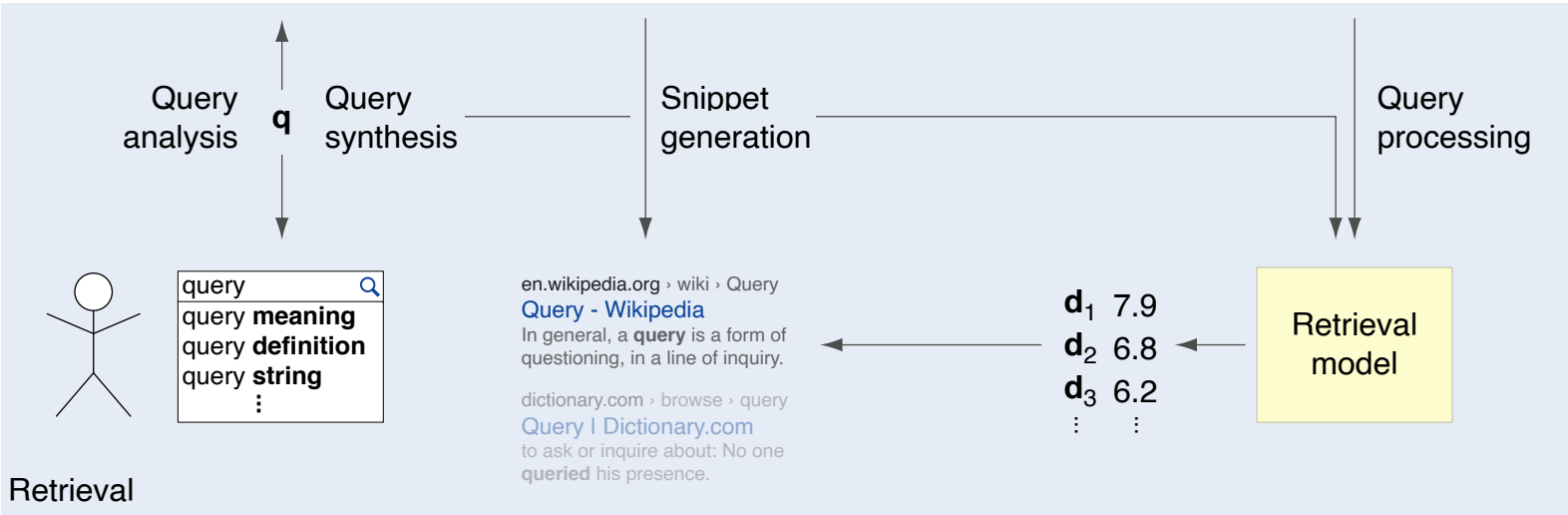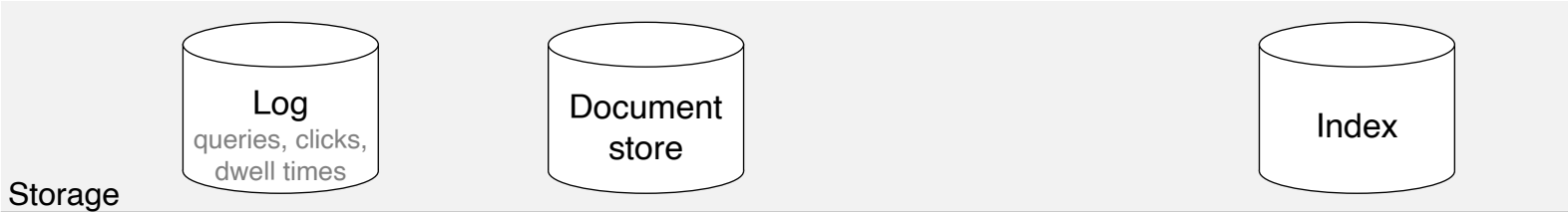to ask or inquire about: No one **queried** his presence.

Architecture of a Search Engine
Karaoke Version

Indexing

Acquisition
- ❑ crawling
- ❑ plain text conversion
- ❑ encoding unification

Text analysis
- ❑ index terms
- ❑ meta data
- ❑ classification

d

Indexing
- ❑ term weighting
- ❑ doc. embedding
- ❑ collection statistics

t_2 t_1 t_3 d

Storage

Log
queries, clicks, dwell times

Document store

Index

Retrieval

Query analysis   **q**   Query synthesis

Snippet generation

query
query **meaning**
query **definition**
query **string**
⋮

en.wikipedia.org › wiki › Query
Query - Wikipedia
In general, a **query** is a form of questioning, in a line of inquiry.

dictionary.com › browse › query
Query | Dictionary.com
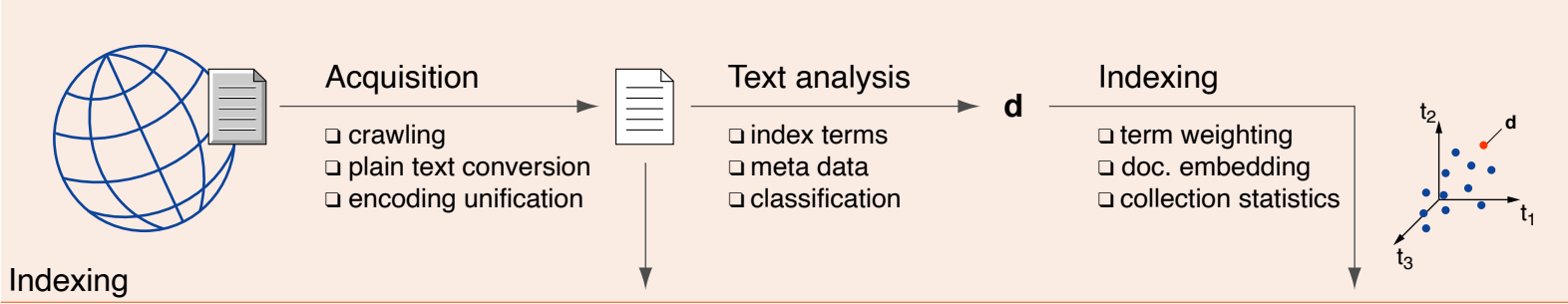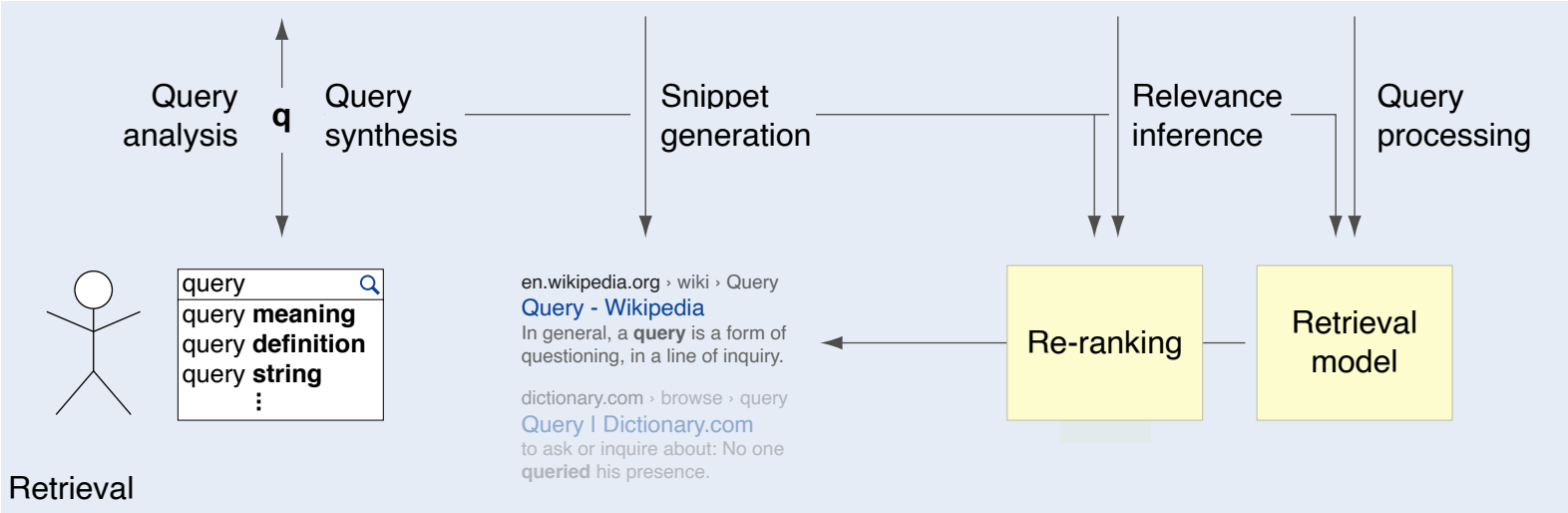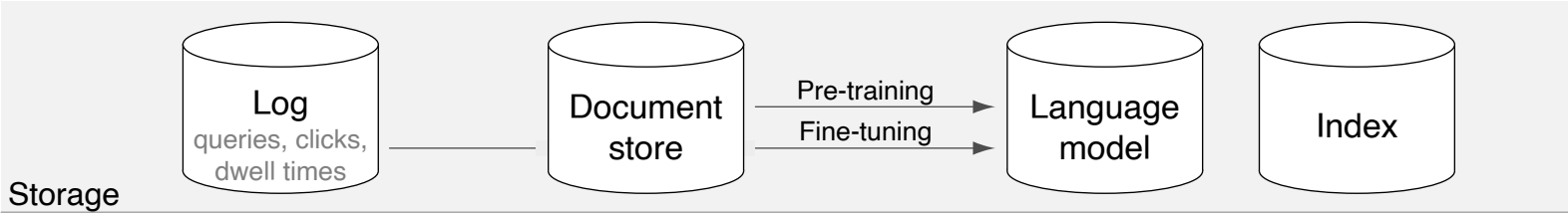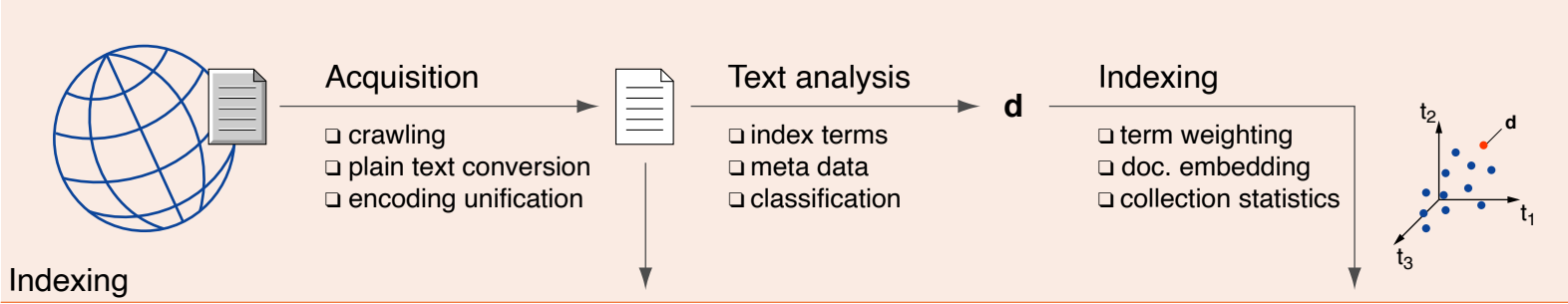to ask or inquire about: No one **queried** his presence.

Architecture of a Search Engine
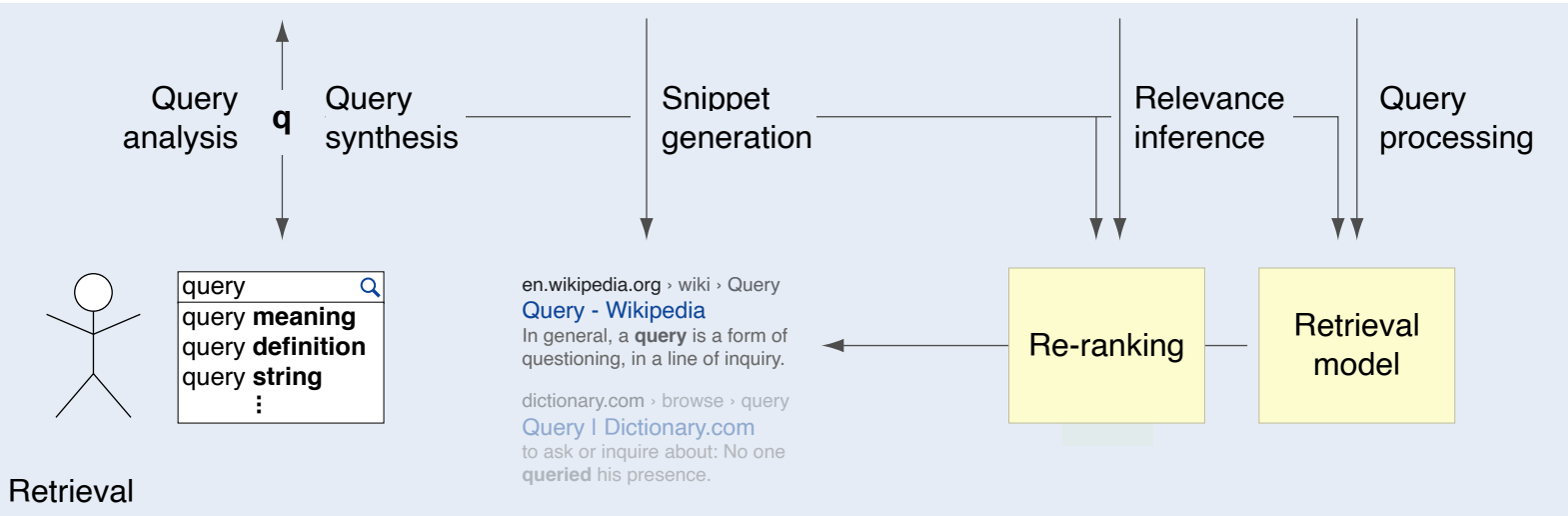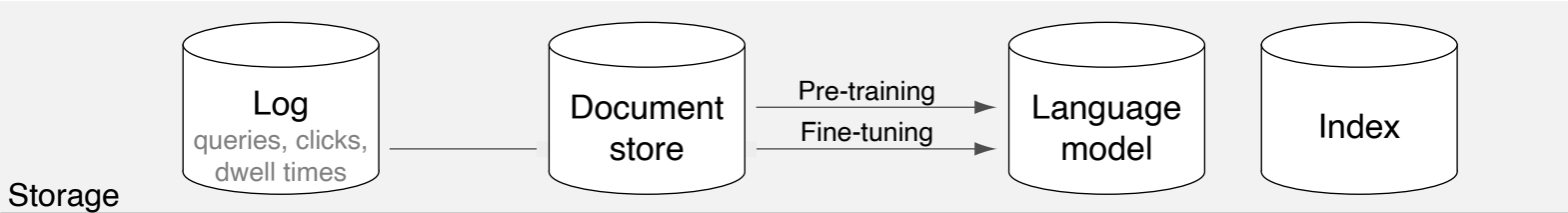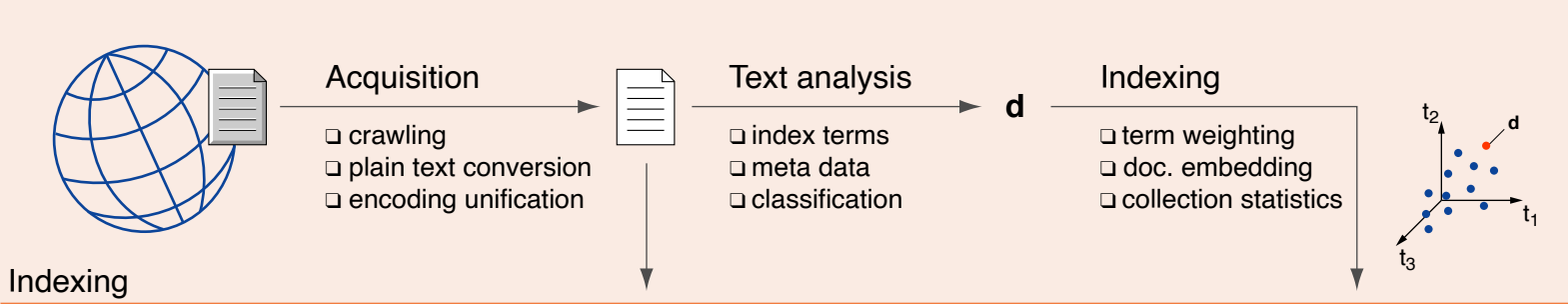## Karaoke Version

Architecture of a Search Engine
Karaoke Version

Indexing

Acquisition
- ❏ crawling
- ❏ plain text conversion
- ❏ encoding unification

Text analysis
- ❏ index terms
- ❏ meta data
- ❏ classification

d

Indexing
- ❏ term weighting
- ❏ doc. embedding
- ❏ collection statistics

t_2 ， t_1 ， t_3 ， **d**

Storage

Log
queries, clicks, dwell times

Document store

Pre-training
Fine-tuning

Language model

Index

Retrieval

Query analysis — **q** — Query synthesis — Snippet generation — Relevance inference — Query processing

query 🔍
query **meaning**
query **definition**
query **string**
⋮

en.wikipedia.org › wiki › Query
Query - Wikipedia
In general, a **query** is a form of questioning, in a line of inquiry.

dictionary.com › browse › query
Query | Dictionary.com
to ask or inquire about: No one **queried** his presence.

Re-ranking

Retrieval model

Architecture of a Search Engine
Karaoke Version

Acquisition
- ❏ crawling
- ❏ plain text conversion
- ❏ encoding unification

Text analysis
- ❏ index terms
- ❏ meta data
- ❏ classification

d

Indexing
- ❏ term weighting
- ❏ doc. embedding
- ❏ collection statistics

t_2 d

t_1

t_3

Storage

Log
queries, clicks, dwell times

Document store

Pre-training
Fine-tuning

Language model

Index

Retrieval

Query analysis **q** Query synthesis

Snippet generation

Relevance inference

Query processing

query
query **meaning**
query **definition**
query **string**
⋮

en.wikipedia.org › wiki › Query
Query - Wikipedia
In general, a **query** is a form of questioning, in a line of inquiry.

dictionary.com › browse › query
Query | Dictionary.com
to ask or inquire about: No one **queried** his presence.

Re-ranking

Retrieval model

Evaluation

Pooling
Exercise

Given the following three rankings:

System	Document at rank									
	1	2	3	4	5	6	7	8	9	10
A	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9	d_{10}
B	d_{12}	d_4	d_6	d_8	d_{10}	d_{11}	d_3	d_5	d_7	d_9
C	d_{10}	d_8	d_9	d_{17}	d_{16}	d_{14}	d_{15}	d_{13}	d_{12}	d_{11}

Pooling
Exercise

Given the following three rankings:

System	Document at rank									
	1	2	3	4	5	6	7	8	9	10
A	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9	d_{10}
B	d_{12}	d_4	d_6	d_8	d_{10}	d_{11}	d_3	d_5	d_7	d_9
C	d_{10}	d_8	d_9	d_{17}	d_{16}	d_{14}	d_{15}	d_{13}	d_{12}	d_{11}

Which documents will end up in a top-5 pooling?

Pooling
Exercise

Given the following three rankings:

System	Document at rank									
	1	2	3	4	5	6	7	8	9	10
A	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9	d_{10}
B	d_{12}	d_4	d_6	d_8	d_{10}	d_{11}	d_3	d_5	d_7	d_9
C	d_{10}	d_8	d_9	d_{17}	d_{16}	d_{14}	d_{15}	d_{13}	d_{12}	d_{11}

Which documents will end up in a top-5 pooling?

Which documents will end up in a top-7 pooling?

Precision–Recall Curve

Exercise

Given the following two rankings:

System	Topic	Relevance at rank									
		1	2	3	4	5	6	7	8	9	10
D	t_1	👍	👎	👍	👍	👎	👍	👎	👎	👎	👍
D	t_2	👎	👎	👍	👍	👎	👎	👍	👍	👎	👍
E	t_1	👎	👍	👍	👎	👍	👍	👍	👎	👎	👎
E	t_2	👍	👎	👍	👍	👎	👎	👎	👍	👍	👎

Precision–Recall Curve

Given the following two rankings:

System	Topic	Relevance at rank									
		1	2	3	4	5	6	7	8	9	10
D	t_1	👍	👎	👍	👍	👎	👍	👎	👎	👎	👍
D	t_2	👎	👎	👍	👍	👎	👎	👍	👍	👎	👍
E	t_1	👎	👍	👍	👎	👍	👍	👍	👎	👎	👎
E	t_2	👍	👎	👍	👍	👎	👎	👎	👍	👍	👎

Draw the precision–recall curves?

Next Steps
Assignment

❑ Exercise sheet on Moodle

❑ Create a GitHub account for the next stage (building an IR system)

❑ Check TIRA login (handed out this week, report if it does not work)

Q&A

Any questions?