

Search Components: Development and Evaluation



Open Web Search Webinar, 3 February 2025, Online

Maik Fröbe

[OWS.eu](https://ows.eu)

[@webis_de](https://twitter.com/webis_de)

www.webis.de

ows.eu

Search Components: Development and Evaluation

Motivation

Michael Granitzer

Leiter OpenWebSearch.eu



"I want to
choose my
search engine
like my daily
newspaper"



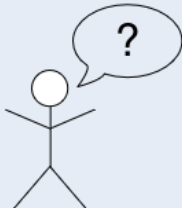
Search Components: Development and Evaluation

What Technology Enables a Diverse and Vibrant Search Ecosystem?



Indexing

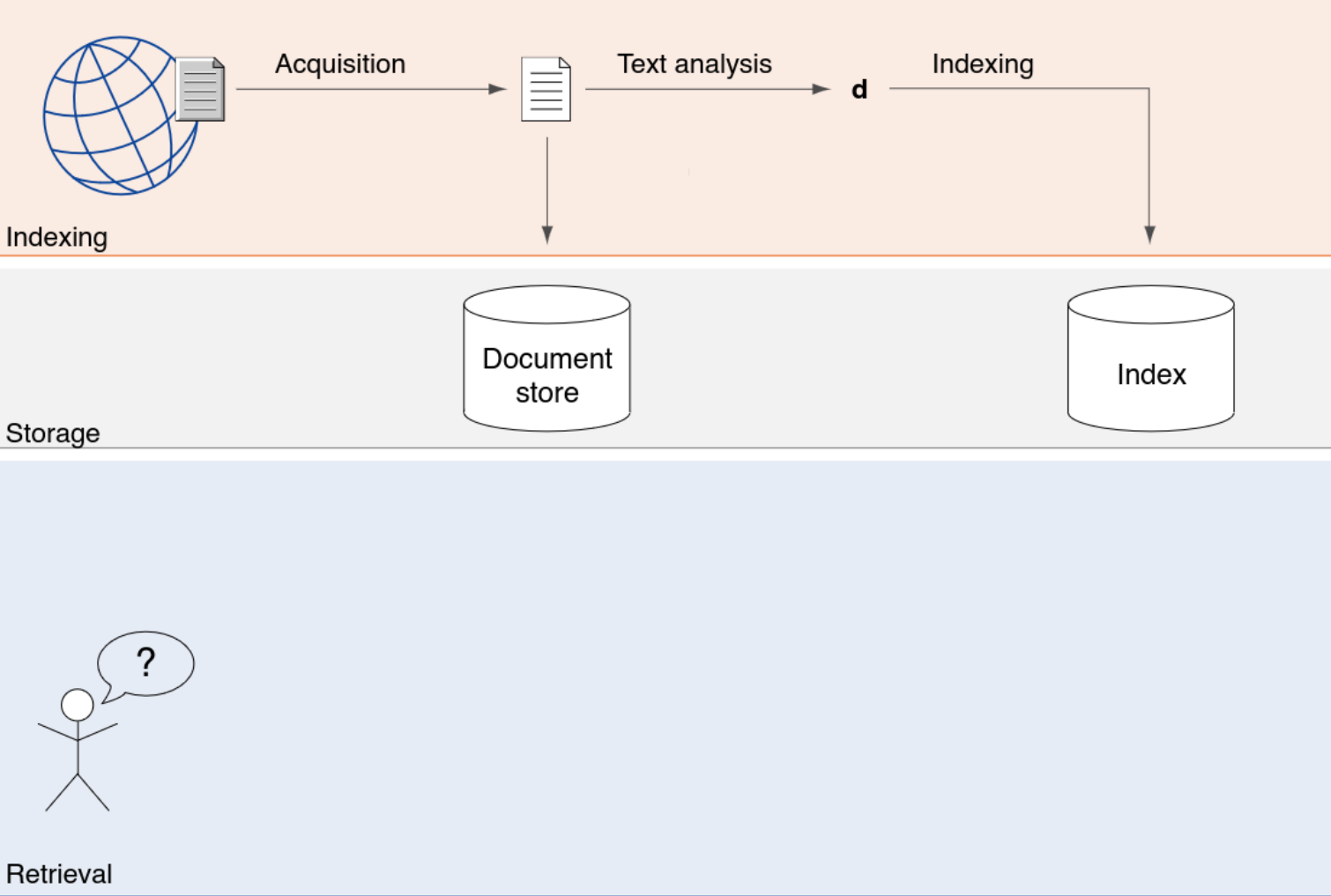
Storage



Retrieval

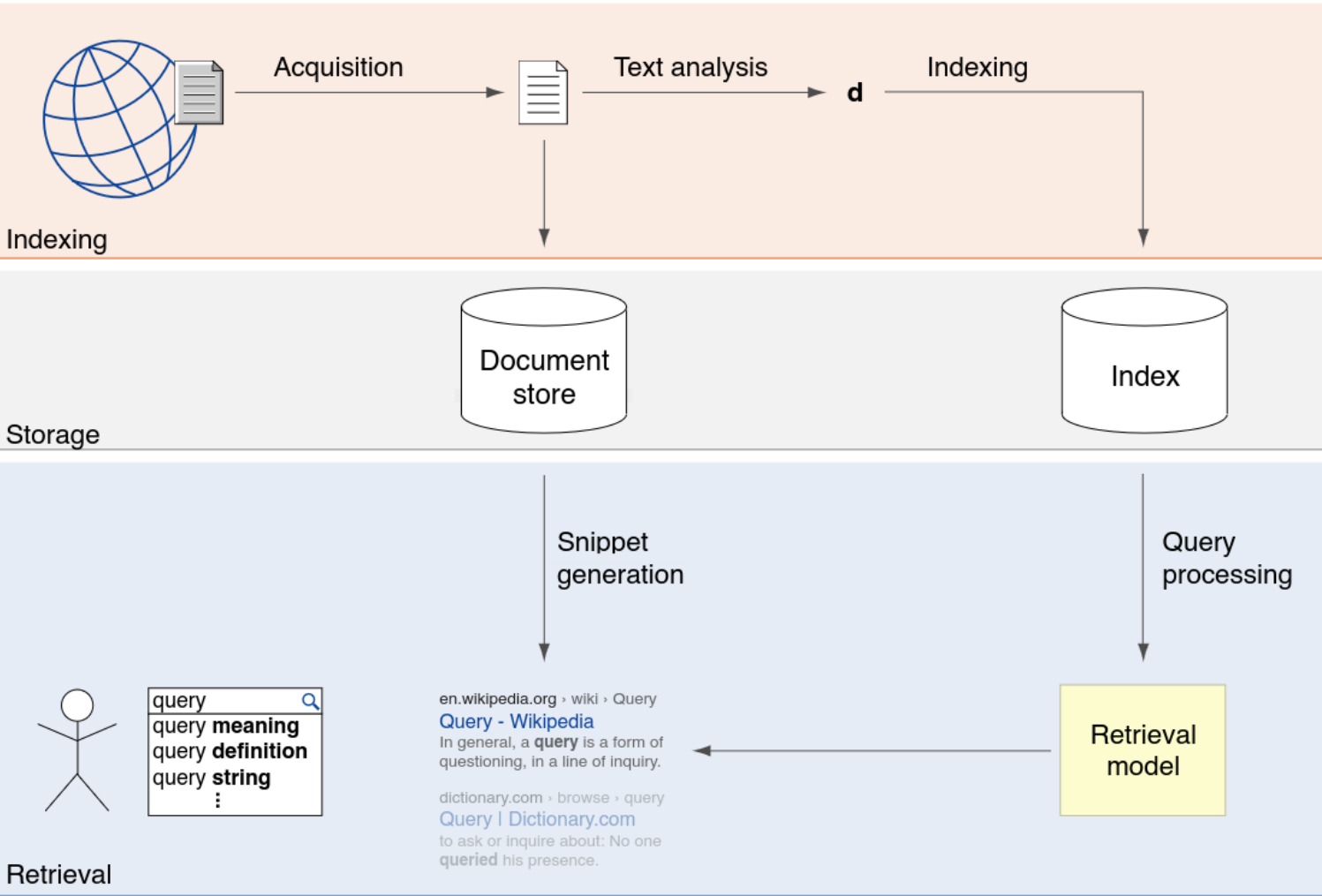
Search Components: Development and Evaluation

What Technology Enables a Diverse and Vibrant Search Ecosystem?



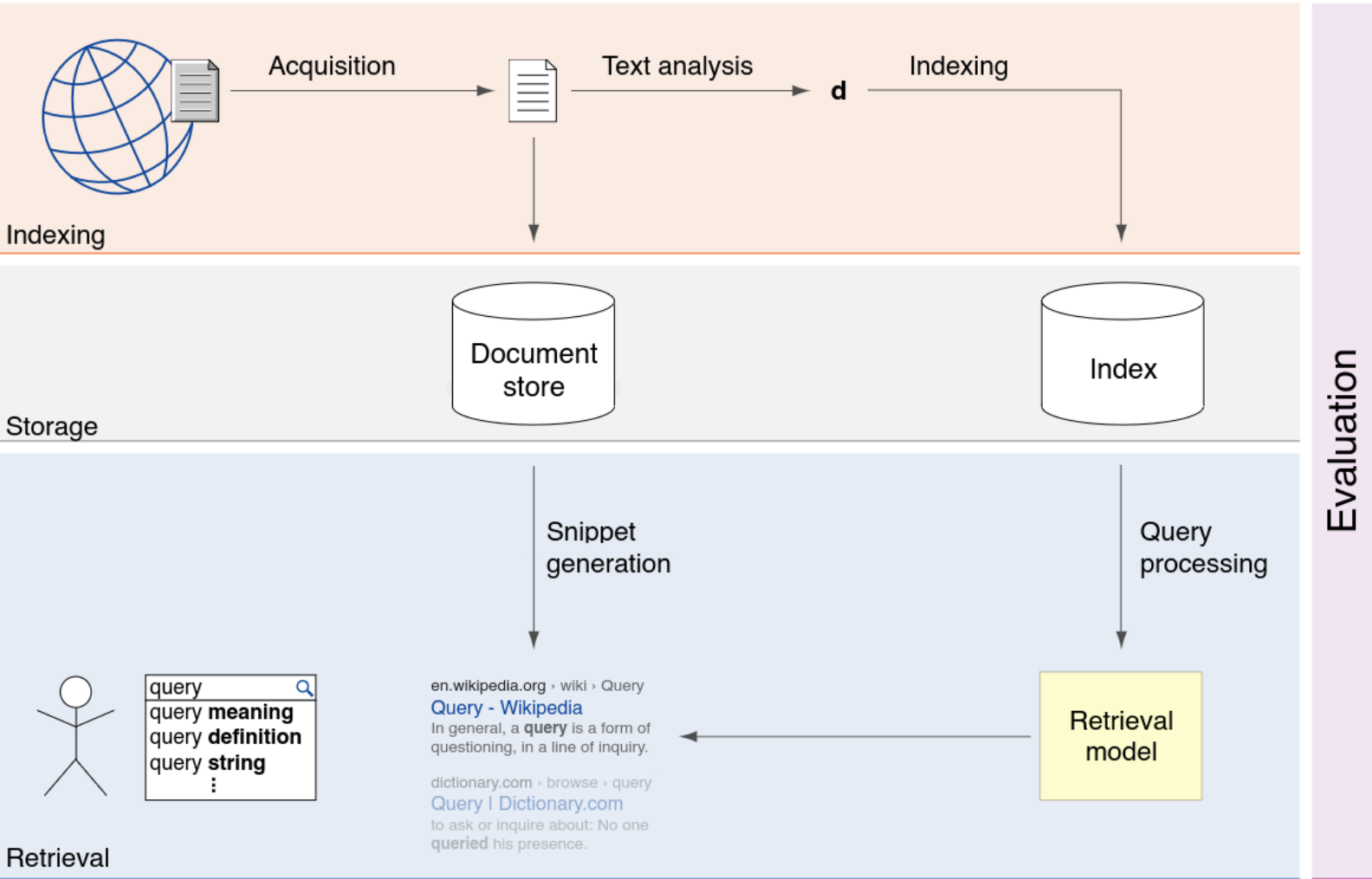
Search Components: Development and Evaluation

What Technology Enables a Diverse and Vibrant Search Ecosystem?



Search Components: Development and Evaluation

What Technology Enables a Diverse and Vibrant Search Ecosystem?



Search Components: Development and Evaluation

The Open Web Index To the Rescue?

 **Open** WebSearch

OpenWebSearch.eu - Building an Open Web Index for an open web search ecosystem

 UNIVERSITÄT
PASSAU
Fakultät für Informatik und Mathematik

Open Web Search 

 Funded by
the European Union

SUPPORTED BY 

Search Components: Development and Evaluation

The Open Web Index: Overview of Partners

14 partners + 3rd party calls

- Research, infrastructure, industry, and NGOs

Research

Information Retrieval, AI, HCI, Geo-spatial Data Processing



Infrastructure Organisations

Data Storage, HPC, Services and Scientific Computing



ICT Solutions for Brilliant Minds



Associations
for a future Web /Internet



Companies
for a future Web /Internet

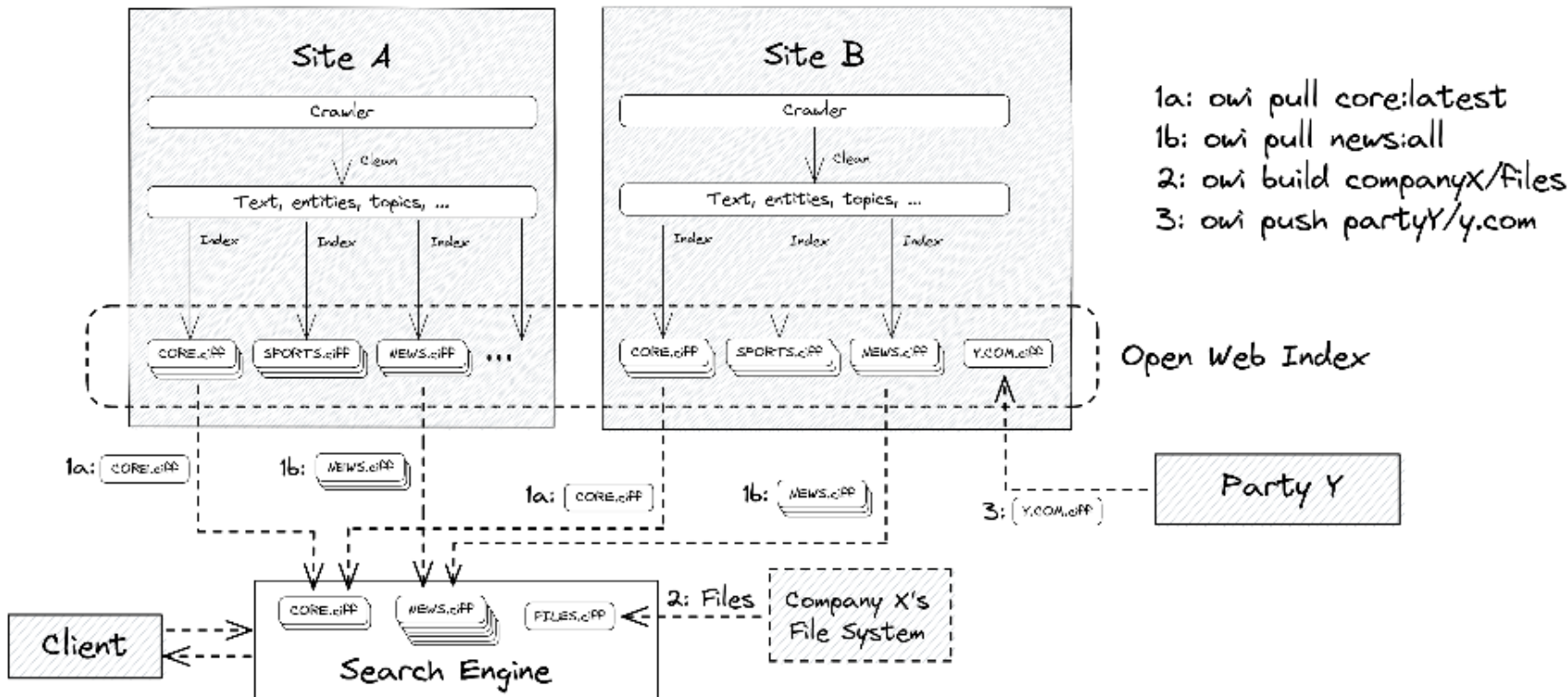
Search Components: Development and Evaluation

The Open Web Index: Overview of Partners



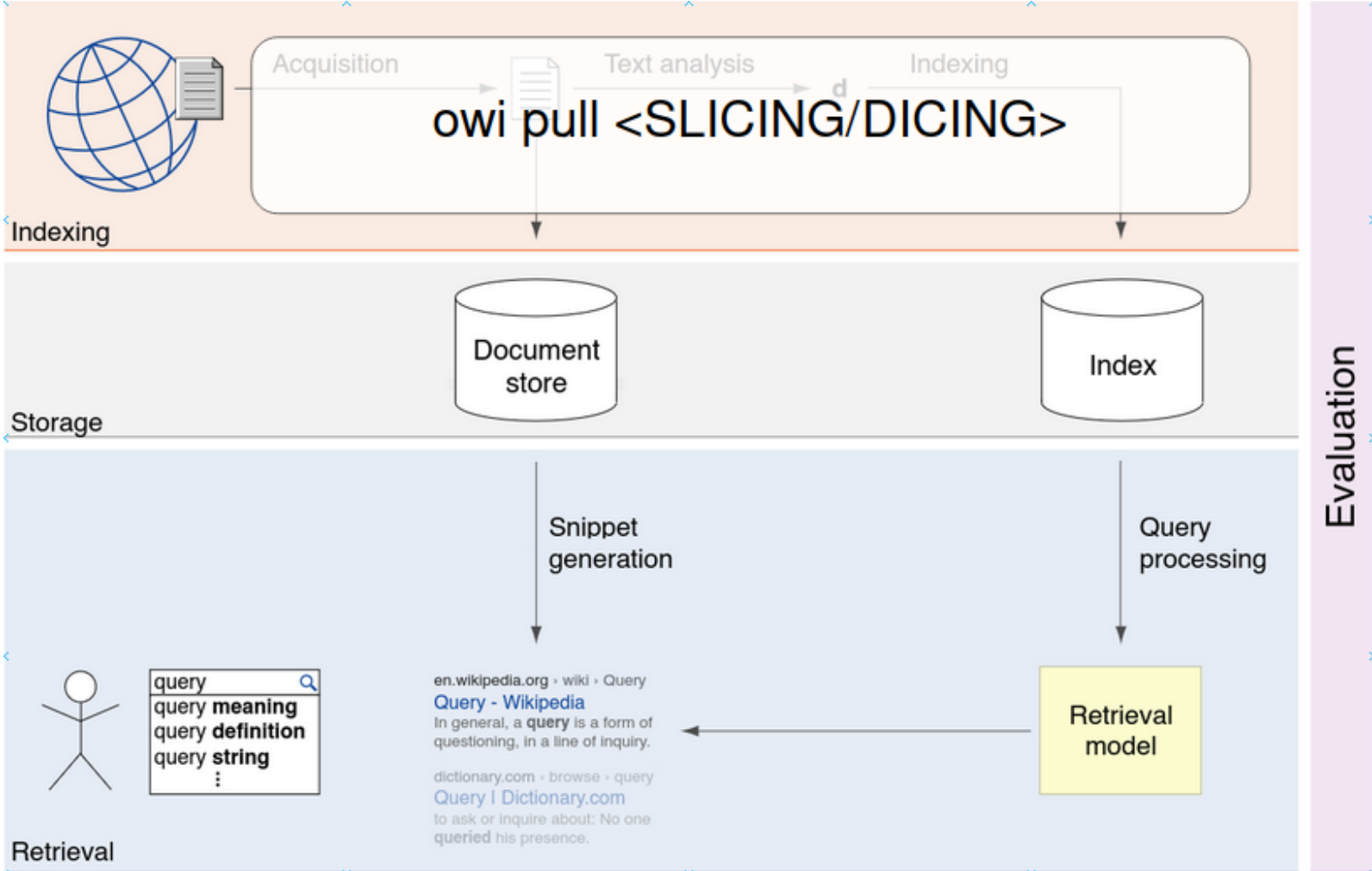
Search Components: Development and Evaluation

Slicing and Dicing the Open Web Index for a Diverse Search Ecosystem



Search Components: Development and Evaluation

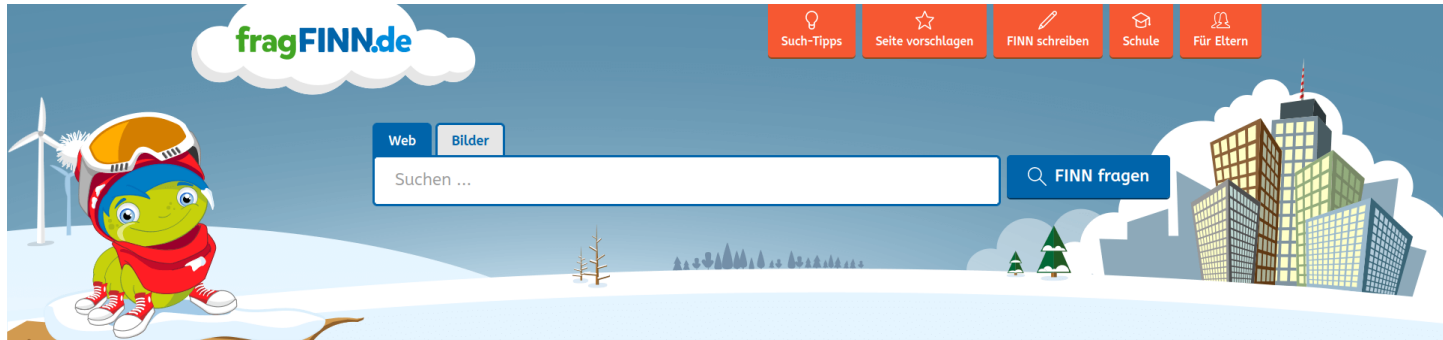
Slicing and Dicing the Open Web Index for a Diverse Search Ecosystem



Search Components: Development and Evaluation

A (biased) Selection of Use-Cases of Slicing and Dicing the Open Web Index

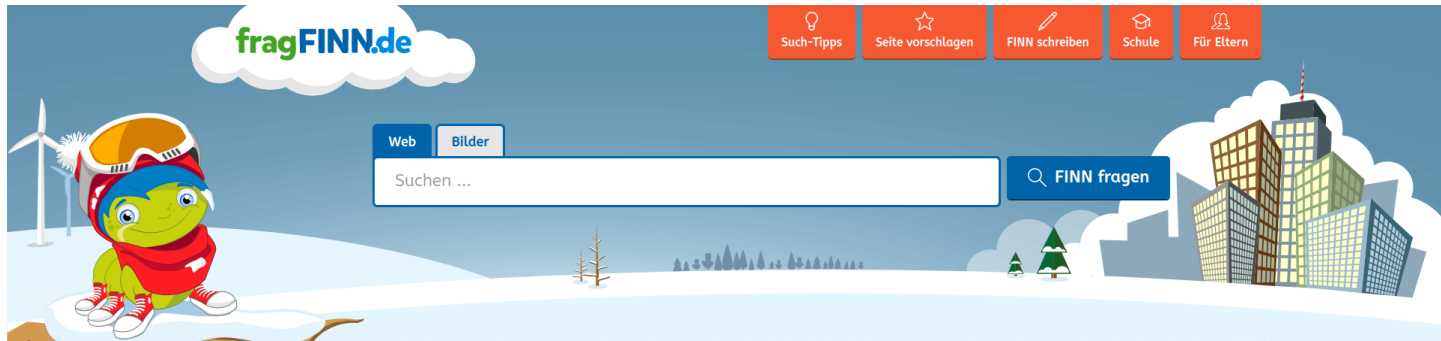
Search engine for kids:



Search Components: Development and Evaluation

A (biased) Selection of Use-Cases of Slicing and Dicing the Open Web Index

Search engine for kids:



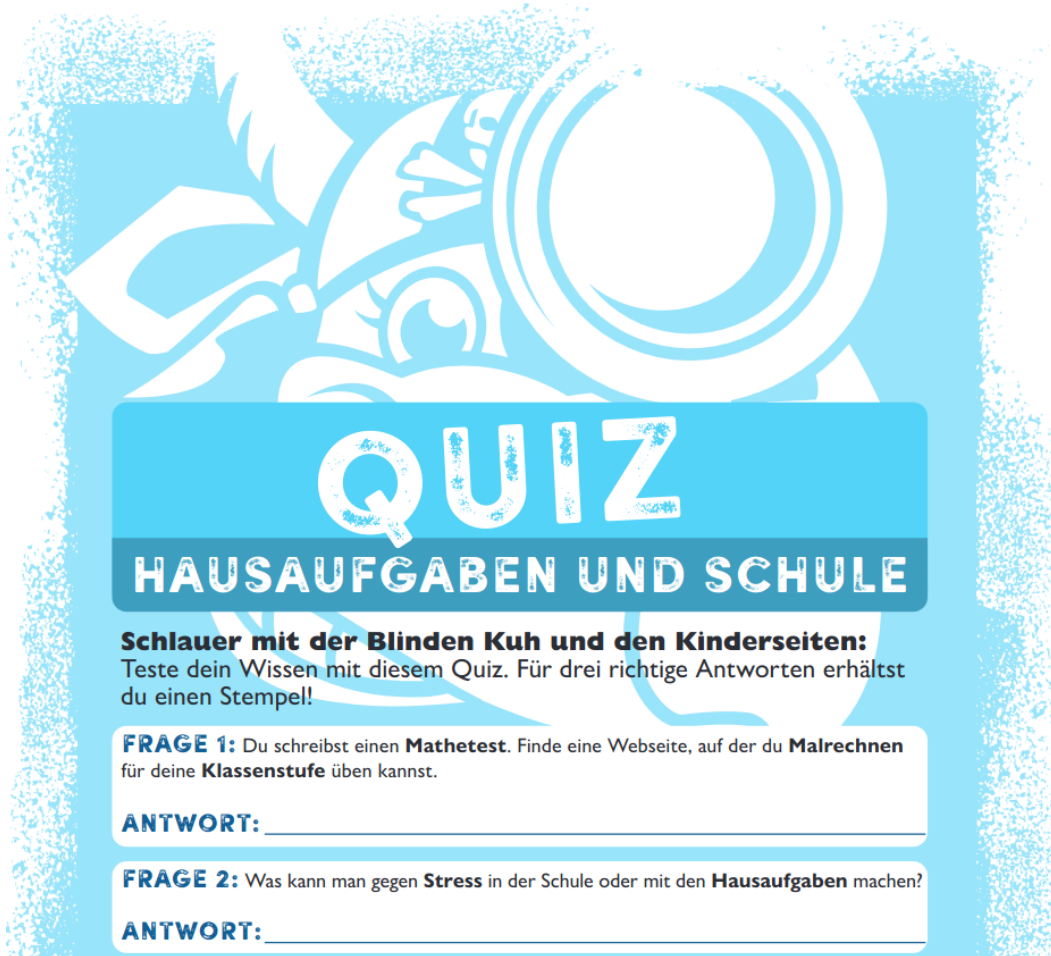
Substantial differences to commercial search:

- ❑ Index size: 1GB
- ❑ Encoded values, guidance, manual curation, ...

Search Components: Development and Evaluation

A (biased) Selection of Use-Cases of Slicing and Dicing the Open Web Index

Search engine for kids:



QUIZ

HAUSAUFGABEN UND SCHULE

Schlauer mit der Blinden Kuh und den Kinderseiten:
Teste dein Wissen mit diesem Quiz. Für drei richtige Antworten erhältst du einen Stempel!

FRAGE 1: Du schreibst einen **Mathetest**. Finde eine Webseite, auf der du **Malrechnen** für deine **Klassenstufe** üben kannst.

ANTWORT: _____

FRAGE 2: Was kann man gegen **Stress** in der Schule oder mit den **Hausaufgaben** machen?

ANTWORT: _____

Search Components: Development and Evaluation

A (biased) Selection of Use-Cases of Slicing and Dicing the Open Web Index

More Ideas than time

- ❑ Build your own Pokemon search engine, or Harry Potter, etc.
- ❑ A climate change search engine (MANILA workshop at SIGIR)
- ❑ A search engine to support financial experts, lawyers, etc.

Search Components: Development and Evaluation

A (biased) Selection of Use-Cases of Slicing and Dicing the Open Web Index

More Ideas than time

- ❑ Build your own Pokemon search engine, or Harry Potter, etc.
- ❑ A climate change search engine (MANILA workshop at SIGIR)
- ❑ A search engine to support financial experts, lawyers, etc.

Use Cases: The Web as Searchable Resource for AI

Web data drives innovation beyond search, particularly in AI



Not only AI:
GDELT: Web-scale Event Analytics



GPT-3: 80% of training data was Web data

DALL-E 2

DALL-E 2 is an AI system that can create realistic images and art from a description in natural language.

The GDELT Project

Watching Our World Unfold

A Global Database of Society

Supported by Google Jigsaw, the GDELT Project monitors the world's broadcast, print, and web news from nearly every corner of every country in over 100

Search Components: Development and Evaluation

Summary

- ❑ The OpenWebSearch.eu project aims to provide an Open Web Index
- ❑ The Open Web Index aims to enable a diverse search ecosystem

Search Components: Development and Evaluation

Summary

- ❑ The OpenWebSearch.eu project aims to provide an Open Web Index
- ❑ The Open Web Index aims to enable a diverse search ecosystem

Evaluation is Key

- ❑ Different organizations with different goals interact on the same fundament

“You don’t want that the client calls you at night.”

Arjen de Vries, OWS.eu

Search Components: Development and Evaluation

Summary

- ❑ The OpenWebSearch.eu project aims to provide an Open Web Index
- ❑ The Open Web Index aims to enable a diverse search ecosystem

Evaluation is Key

- ❑ Different organizations with different goals interact on the same fundament

“You don’t want that the client calls you at night.”

Arjen de Vries, OWS.eu

- ❑ Good evaluation = good sleep :)

Search Components: Development and Evaluation

Summary

- ❑ The OpenWebSearch.eu project aims to provide an Open Web Index
- ❑ The Open Web Index aims to enable a diverse search ecosystem

Evaluation is Key

- ❑ Different organizations with different goals interact on the same fundament

“You don’t want that the client calls you at night.”

Arjen de Vries, OWS.eu

- ❑ Good evaluation = good sleep :)

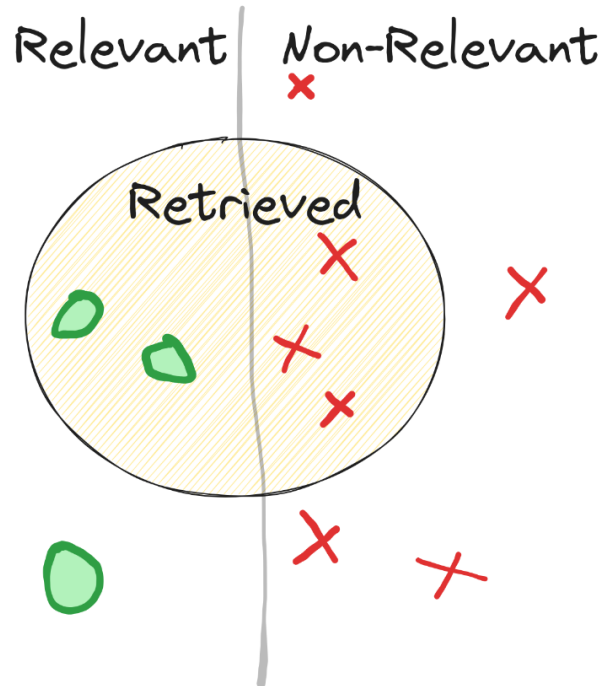
There is no One-Size-Fits-All Evaluation

- ❑ Your use-cases and scenarios determine your evaluation
- ❑ Two competing goals for optimization:
 - Precision: How many retrieved documents are relevant?
 - Recall: How many relevant documents are retrieved?

Basics of the Evaluation of Search Components

Assume we have a corpus with 10 documents:

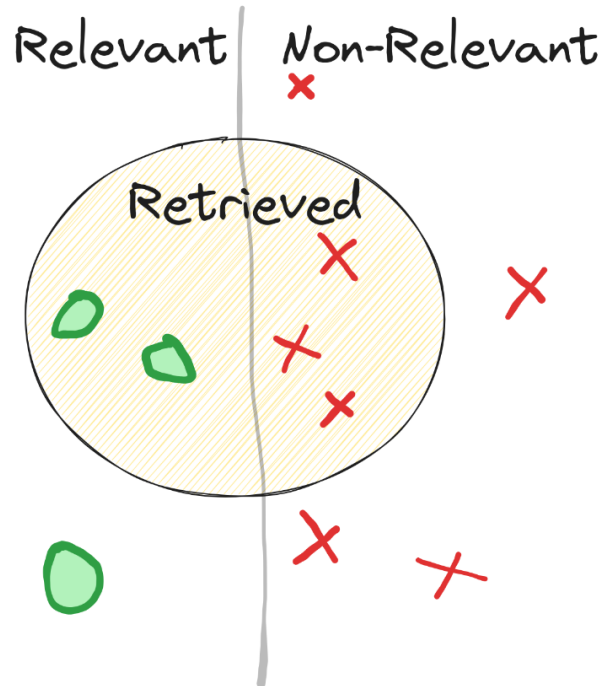
- 3 documents are relevant to a query and 7 are non-relevant.



Basics of the Evaluation of Search Components

Assume we have a corpus with 10 documents:

- 3 documents are relevant to a query and 7 are non-relevant.



$$\text{Precision} = \frac{\text{3 green ovals}}{\text{3 green ovals} + \text{3 red crosses}} = 0.40$$

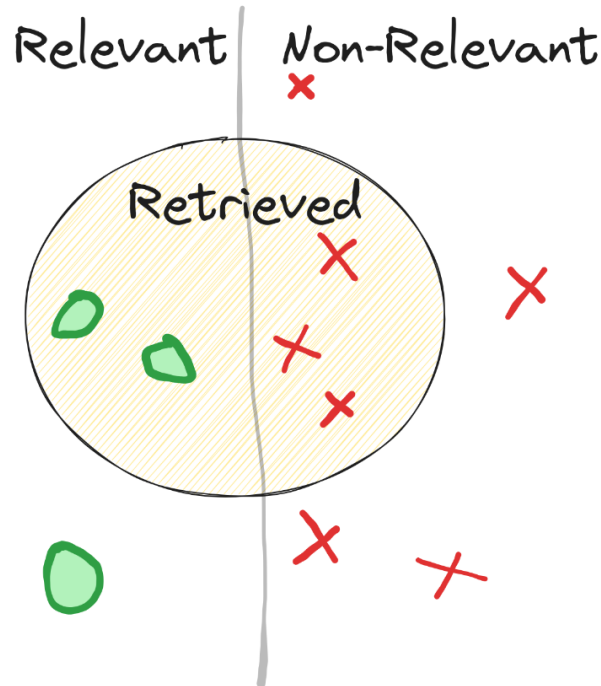
Use-cases:

- Web search, child-safe search, ...

Basics of the Evaluation of Search Components

Assume we have a corpus with 10 documents:

- 3 documents are relevant to a query and 7 are non-relevant.



$$\text{Precision} = \frac{\text{3 green ovals}}{\text{3 green ovals} + \text{3 red crosses}} = 0.40$$

$$\text{Recall} = \frac{\text{3 green ovals}}{\text{3 green ovals} + \text{1 green oval}} = 0.66$$

Use-cases:

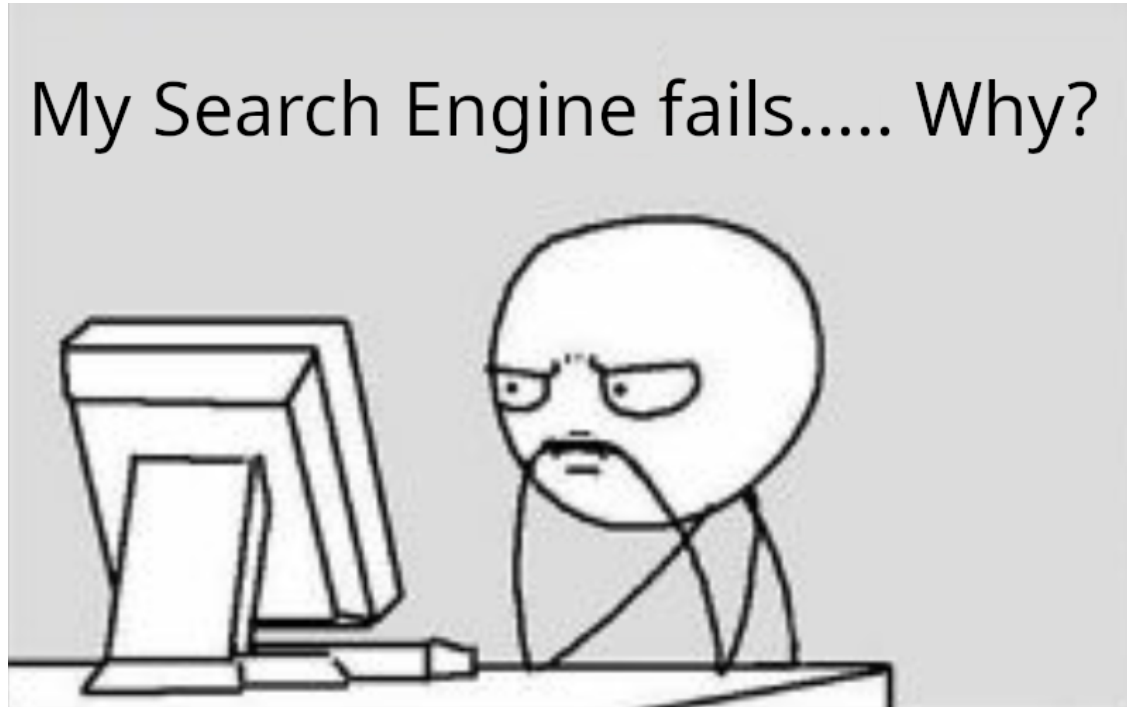
- Web search, child-safe search, ...

Use-cases:

- Scholarly/patent search, ...

Search Components: Development and Evaluation

What can Evaluation do for you?



Search Components: Development and Evaluation

What can Evaluation do for you?



Evaluation can guide the development of your search engine

- ❑ An failure in your search engine is either a **precision** or a **recall** problem
- ❑ Every search component either aims to improve **precision** or **recall**

Do you need to improve precision or recall? \Rightarrow Select a corresponding tool

Search Components of OpenWebSearch.eu

Workshop on Open Web Search 2024

- ❑ Goal: Collect diverse and research-oriented search components
- ❑ Implementations from people with diverse backgrounds:
 - Researchers (PhD-level to Assistant Professor)
 - Bachelor/Master students
- ❑ Hackathon: 1 week to preparation + 1 week hackathon + 0.5 weeks writing



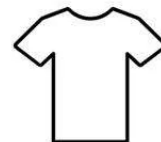
IR Hackathon @ APB

- Team-Projekt (2 – 5 Leute)
- Eine kleine IR Komponente entwickeln
 - Dauer: 1 – 4 Tage je nach Lust und Zeit ☺



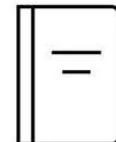
Pizza

+



T-Shirt

+



Anrechenbar
im Studium

Workshop on Open Web Search 2024

Workshop on Open Web Search 2024: Impressios



Workshop on Open Web Search 2024

Workshop on Open Web Search 2024: Impressios



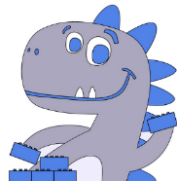
Workshop on Open Web Search 2024

Workshop on Open Web Search 2024: Impressios



Workshop on Open Web Search 2024

Overview of implemented components



Components in TIREx

This is a click dummy/early prototype to search/slice/dice the components/tutorials/resources available in TIREx together with related resources. We currently have an open call for components as part of an EDIR workshop and will update the overview below with all submitted components. Please do not hesitate to contribute to this overview by modifying the underlying yaml file, we would be happy about all contributions!

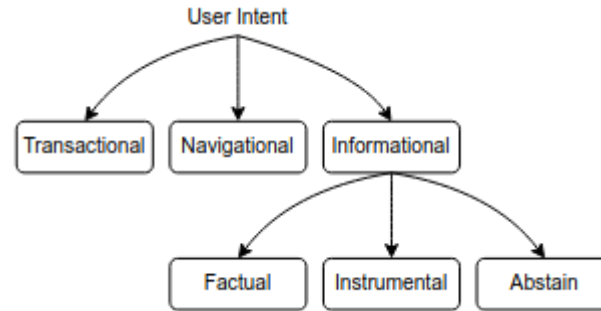
Dataset	Document Processing	Query Processing	Retrieval	Re-Ranking	Evaluation
Args.me (+2)	Keyphrase Extraction (+4)	Query Expansion	Bi-Encoder (+2)	Cross-Encoder (+15)	Bpref
Antique	Health Classification	Query Segmentation (+6)	Late Interaction (+1)	Bi-Encoder (+16)	C/W/L
Web Search (+18)	Stemming (+3)	Query Performance Prediction (QPP) (+12)	Lexical Retrieval (+21)	Late Interaction (+1)	MAP
Medical Search (+3)	Stopword Removal	Health Classification		Lexical Re-Ranking (+20)	MRR
MS MARCO (+2)				Rank Fusion	nDCG
News Search (+5)					Precision@k
Tip-of-the-Tongue					Recall@k
Vaswani					Reproducibility

WOWS 2024: Overview Retrieval Components

Type	Team	Retrieval Component	
		Description	#
Re-Ranking	h2oloo	Citadell	1
	naverlabseurope	Splade	1
	tu-dresden-02	Snippet Extraction	2
Query	QPPTK	Query Performance Prediction	12
	qspell	Spelling Correction	4
	salamander	Comparative Query Classification	1
	tu-dresden-03	LLM Query Expansion	9
	OWS	Query Segmentation	6
	marcel-gohsen	Eintity Linking / Query Interpretation	2
	dossier	Query Intent Prediction	2
Document	fschlatt	Health Classification	2
	seanmacavaney	Corpus Graph	1
	seanmacavaney	DocT5Query	1
	tu-dresden-01	Genre Classification	3
	tu-dresden-04	Readability/Quality/Coherence Features	2
Σ	15		49

Some Example Components: Query-Intent Prediction

[Alexander et al.; SIGIR'22]



Examples:

- ❑ `buy playstation 4` ⇒ **Transactional**
- ❑ `login mattermost` ⇒ **Navigational**
- ❑ `how to build a fence` ⇒ **Informational**

Some Example Components: Document Genre Classification

[Stein et al.; ECAI'06, Erben et al.; WOWS'24]

Classify the aim of a document

Is the goal of the document to ...

- ❑ Educate?
- ❑ Sell?
- ❑ Report?
- ❑ ...



Advertisement

Report

Usage as Search Component

- ❑ Filter based on the predicted query intent
- ❑ Focus: Precision

Some Example Components: DocT5Query

Use a transformer or LLM to predict which queries an document can answer.

Some Example Components: DocT5Query

Use a transformer or LLM to predict which queries an document can answer.

- ❑ Example document:

Barley, a member of the grass family, is a major cereal grain.
It was one of the first cultivated grains eaten widely.

Some Example Components: DocT5Query

Use a transformer or LLM to predict which queries an document can answer.

- ❑ Example document:

Barley, a member of the grass family, is a major cereal grain.
It was one of the first cultivated grains eaten widely.

- ❑ Predicted queries:
 - What is barley used for?
 - What is barley?
 - What cereal grain for cooking?

Some Example Components: DocT5Query

Use a transformer or LLM to predict which queries an document can answer.

- ❑ Example document:

Barley, a member of the grass family, is a major cereal grain.
It was one of the first cultivated grains eaten widely.

- ❑ Predicted queries:
 - What is barley used for?
 - What is barley?
 - What cereal grain for cooking?

Usage as Search Component

- ❑ Add predicted queries to the retrieval index
- ❑ Retrieval now works on original text and predicted queries
- ❑ Focus: Recall

Search Components: Development and Evaluation

Now we get our fingers dirty...



Search Components: Development and Evaluation

Now we get our fingers dirty...



Search Components: Development and Evaluation

Our Goal: Develop an automated Relevance Assessor

Search Components: Development and Evaluation

Our Goal: Develop an automated Relevance Assessor

Input:

- ❑ A query
- ❑ A document

Output:

- ❑ Probability, that the document is relevant to the query

We want to collect such assessors for the WOWS 2025 workshop

- ❑ Goal: enrich the Open Web Index with relevance judgments
- ❑ Will allow first evaluations

During the hands-on part of the webinar, we will collaboratively:

- ❑ We implement a prompted prototype with a (small) language model
- ❑ We embed this relevance assessor as a component in a retrieval pipeline

<https://github.com/OpenWebSearch/wows-code/blob/main/ecir25/baselines/pointwise-autoqrels.ipynb>

Search Components: Development and Evaluation

Conclusions

- ❑ Evaluation can guide development
- ❑ Should your use-case focus on precision or recall?
- ❑ Search components either focus on one of them.
- ❑ We collected 49 re-usable retrieval components
- ❑ We aim to use the potential for creative exploration:
 - How to combine components?
 - What is missing, what is popular?
 - Connected to evaluation campaigns

Workshop on Open Web Search

Conclusions

- ❑ Evaluation can guide development
- ❑ Should your use-case focus on precision or recall?
- ❑ Search components either focus on one of them.
- ❑ We collected 49 re-usable retrieval components
- ❑ We aim to use the potential for creative exploration:
 - How to combine components?
 - What is missing, what is popular?
 - Connected to evaluation campaigns

Future Work

- ❑ WOWS 2025 workshop aims to collect relevance judgments
- ❑ With relevance judgments on the Open Web Index, we can decide and recommend components

Conclusions

- ❑ Evaluation can guide development
- ❑ Should your use-case focus on precision or recall?
- ❑ Search components either focus on one of them.
- ❑ We collected 49 re-usable retrieval components
- ❑ We aim to use the potential for creative exploration:
 - How to combine components?
 - What is missing, what is popular?
 - Connected to evaluation campaigns

Future Work

- ❑ WOWS 2025 workshop aims to collect relevance judgments
- ❑ With relevance judgments on the Open Web Index, we can decide and recommend components

Thank you!