

Intelligent Information Systems Lab

Winter Semester 2024/2025

Harry Scells, Tim Hagen, and Martin Potthast

About us



Martin Potthast



Tim Hagen



Harry Scells

You can say “you” to us

TEMIR

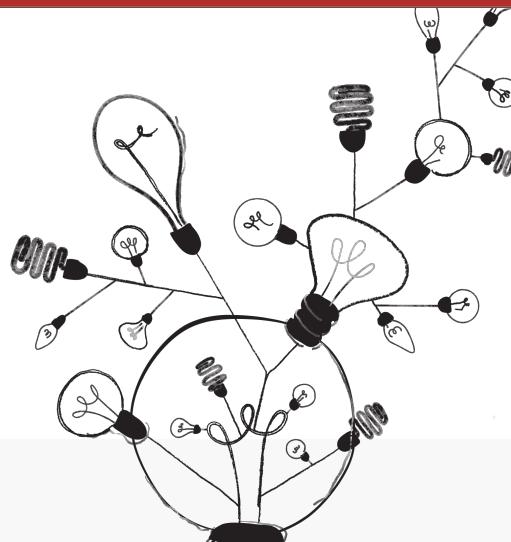
PEOPLE FOR STUDENTS TEACHING RESEARCH PUBLICATIONS DATA EVENTS FACILITIES

**U N I K A S S E L
V E R S I T Ä T**

Deep Semantic Learning

The Temir group carries out basic and applied research with the goal of developing future information systems. Working at the intersection of natural language processing and information retrieval, our focus is on societal challenges concerning quality and trustworthiness of information. Results are integrated into large-scale information systems, making use of crowdsourcing, data mining, machine learning, high-performance computing, and web technologies.

[LEARN MORE](#)



Search Engines Contact

Args Argument search

Deep Semantic Learning Group
University of Kassel

[\[kassel.webis.de\]](http://kassel.webis.de)

Agenda

- ❑ Motivation
- ❑ Information Retrieval
- ❑ Web Search Architecture
- ❑ Retrieval-Augmented Generation
- ❑ What's a good response?
- ❑ Lab Organization
- ❑ Getting started
 - Onboarding
 - Cluster access
 - Cluster use

Search behavior is changing

Provide the answer instead of “10 blue links”

The screenshot shows a search results page from a search engine. The query "Birthplace abraham lincoln" is entered in the search bar. Below the search bar, there are filters: "Alle" (selected), "Bilder", "Videos", "Nachrichten", "Karten", and "Einkaufen". There are also dropdowns for "Immer privat" (selected), "Deutschland" (selected), "Abgesicherte Suche: moderat" (selected), and "Irgendwann". The first result is a link to the Abraham Lincoln Birthplace National Historical Park website: <https://www.nps.gov/abli/index.htm>. The snippet reads: "Abraham Lincoln Birthplace National Historical Park (U.S. National ...)" followed by a summary of the park's history. To the left of the snippet are "Info" and "Maps" sections. To the right are "Alerts" and "Calendar" sections. Below the main result is another link to Wikipedia: "W https://en.wikipedia.org/wiki/Abraham_Lincoln". The snippet reads: "Abraham Lincoln - Wikipedia" followed by a summary of his birthplace. At the bottom, there is a section titled "Bilder für Birthplace abraham lincoln" showing a grid of 10 thumbnail images related to the topic.

VS.

The screenshot shows a search results page with the query "Where was Abraham Lincoln born?". The top result is a snippet from a source that states: "Abraham Lincoln was born in a log cabin on February 12, 1809, in Hodgenville, which was then part of Hardin County, Kentucky (now known as LaRue County). His birthplace is now preserved as the Abraham Lincoln Birthplace National Historical Park." The snippet is preceded by a question mark icon.

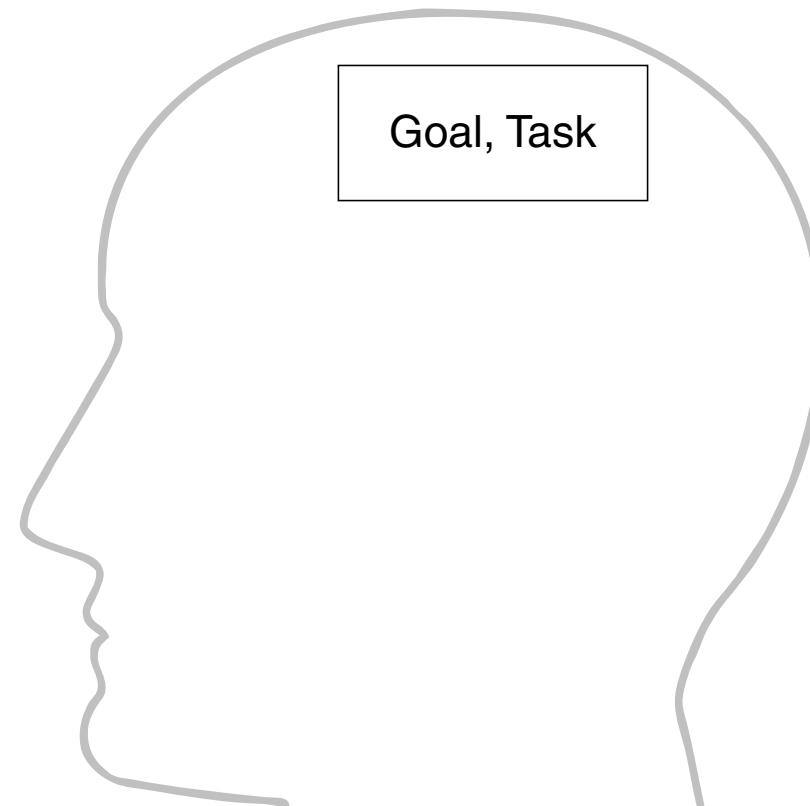
To err is “human”

The screenshot shows a news article from the German news website tagesschau.de. At the top, there is a navigation bar with the tagesschau logo, a "Sendung verpasst?" button, and a menu icon. Below the navigation bar, the breadcrumb trail reads "Startseite > Wirtschaft > Google-Textroboter gibt falsche Antwort". The main content features a large image of a smartphone displaying the Google AI logo and the text "Google AI". Above the phone, the headline "Introducing Bard, an experimental conversational AI service powered by LaMDA" is displayed. Below the phone, a snippet of text says "You can use Bard to — Plan a friend's baby shower Compare two ...". A caption below the phone reads "Panne im Werbeclip für 'Bard'". The main title of the article is "Google-Textroboter gibt falsche Antwort". A timestamp "Stand: 09.02.2023 15:19 Uhr" is shown, followed by a summary text: "Im Rennen um das Geschäft mit Künstlicher Intelligenz hat Google sein Programm 'Bard' vorgestellt. Dabei lieferte der Textroboter jedoch falsche Informationen. Die Aktie des Mutterkonzerns brach ein, 100 Milliarden Dollar Börsenwert verpufften." The entire screenshot is framed by a dark border.

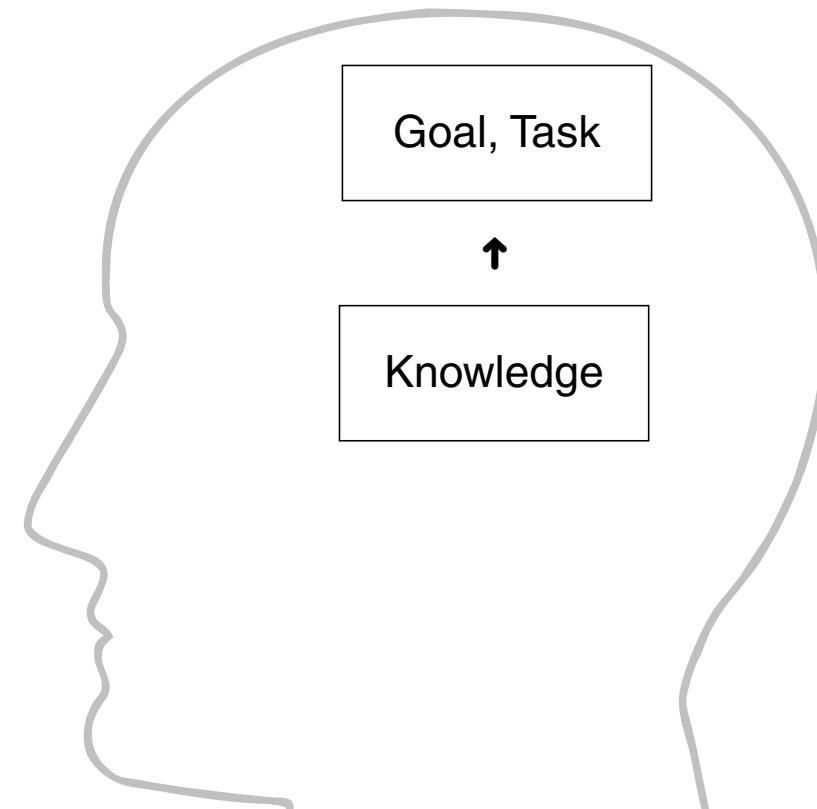
[tagesschau.de]

→ It is important to give good answers

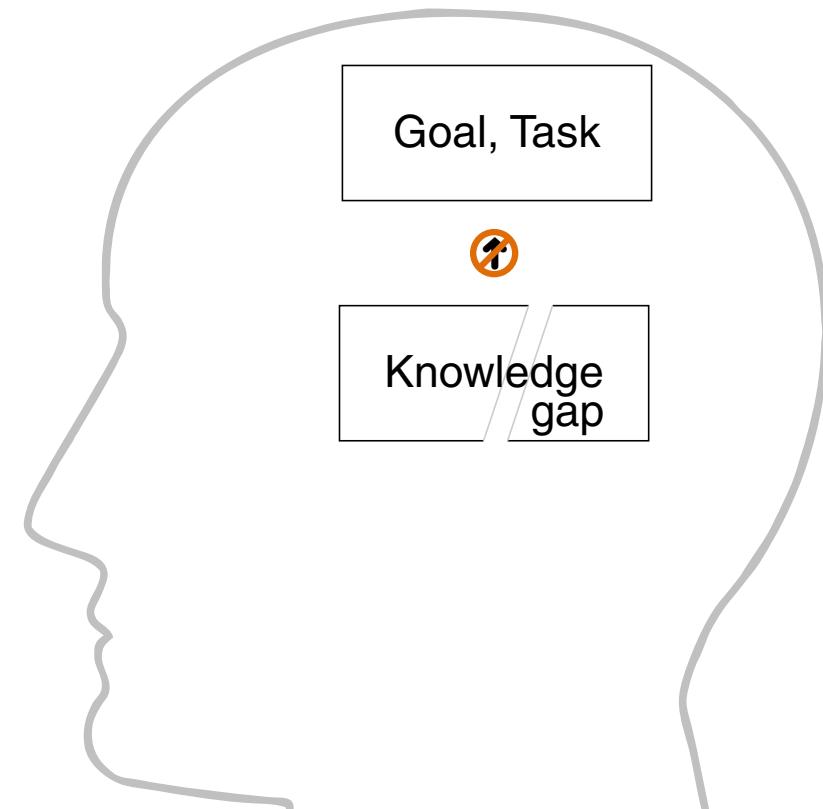
Information Retrieval (IR)



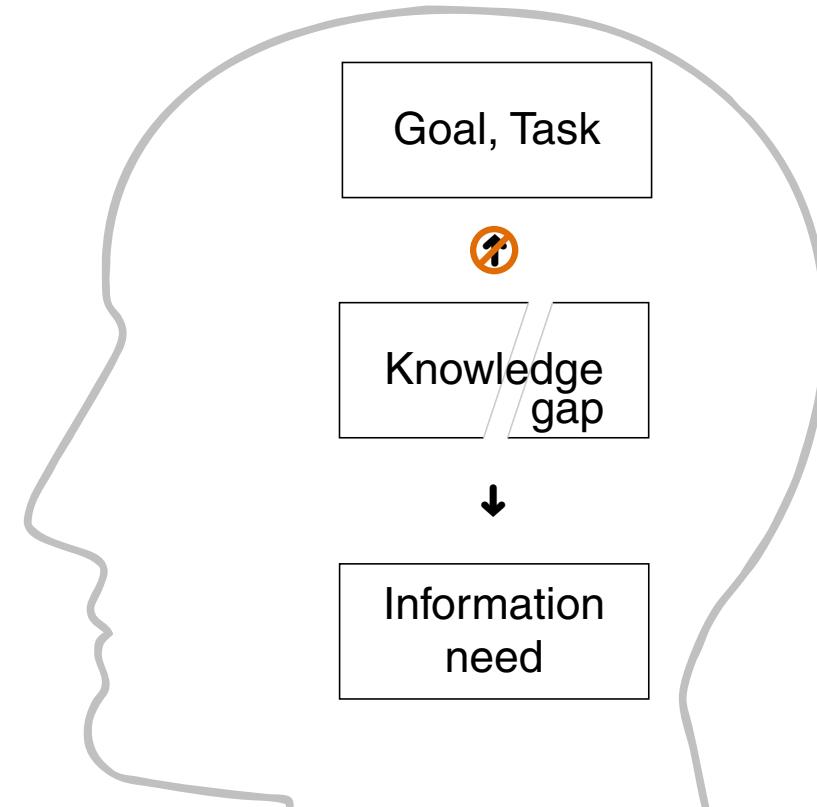
Information Retrieval (IR)



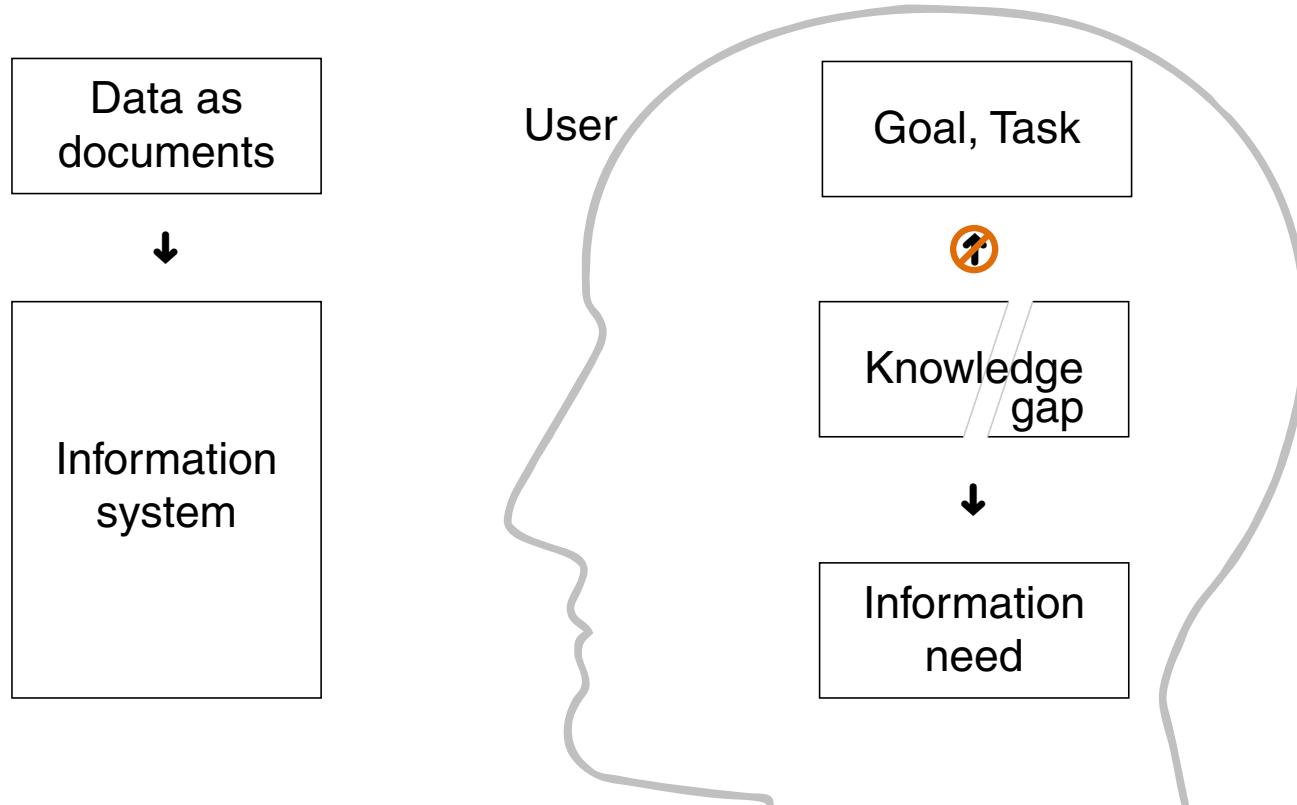
Information Retrieval (IR)



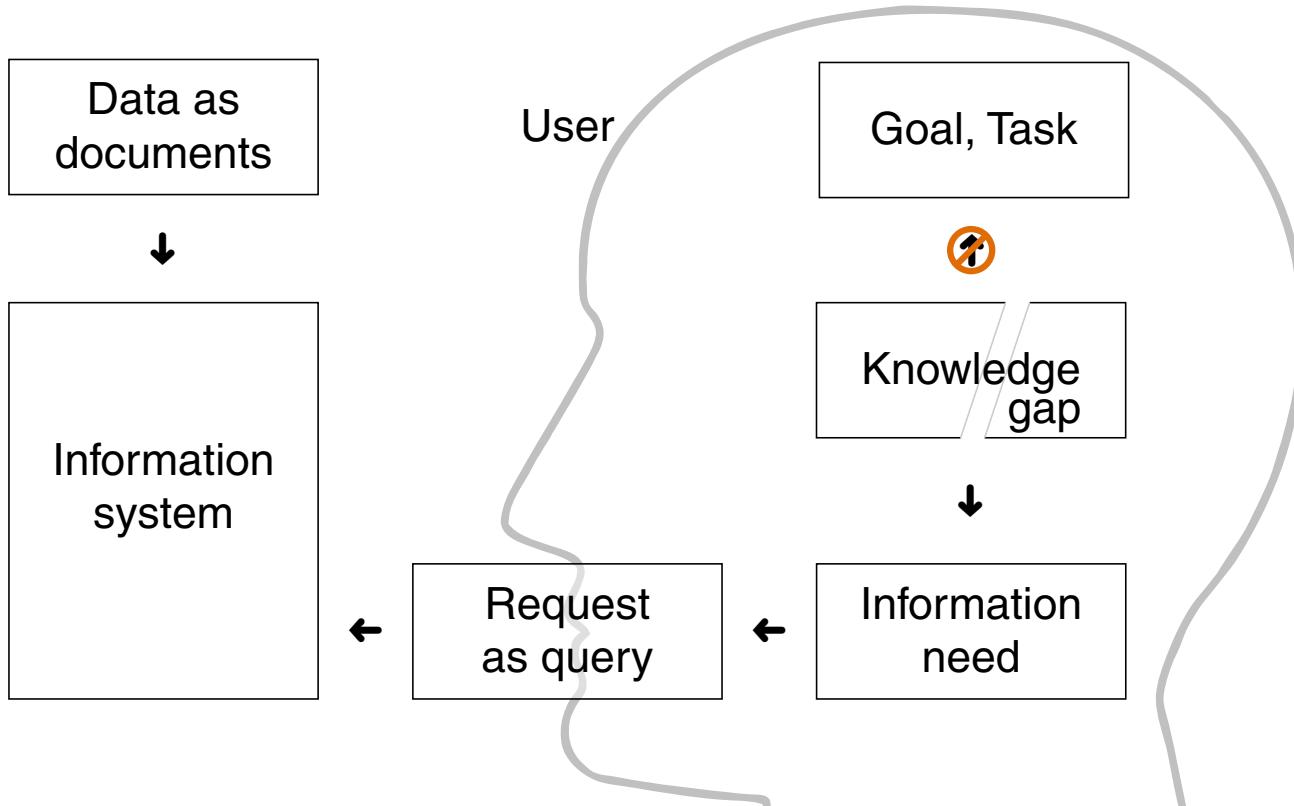
Information Retrieval (IR)



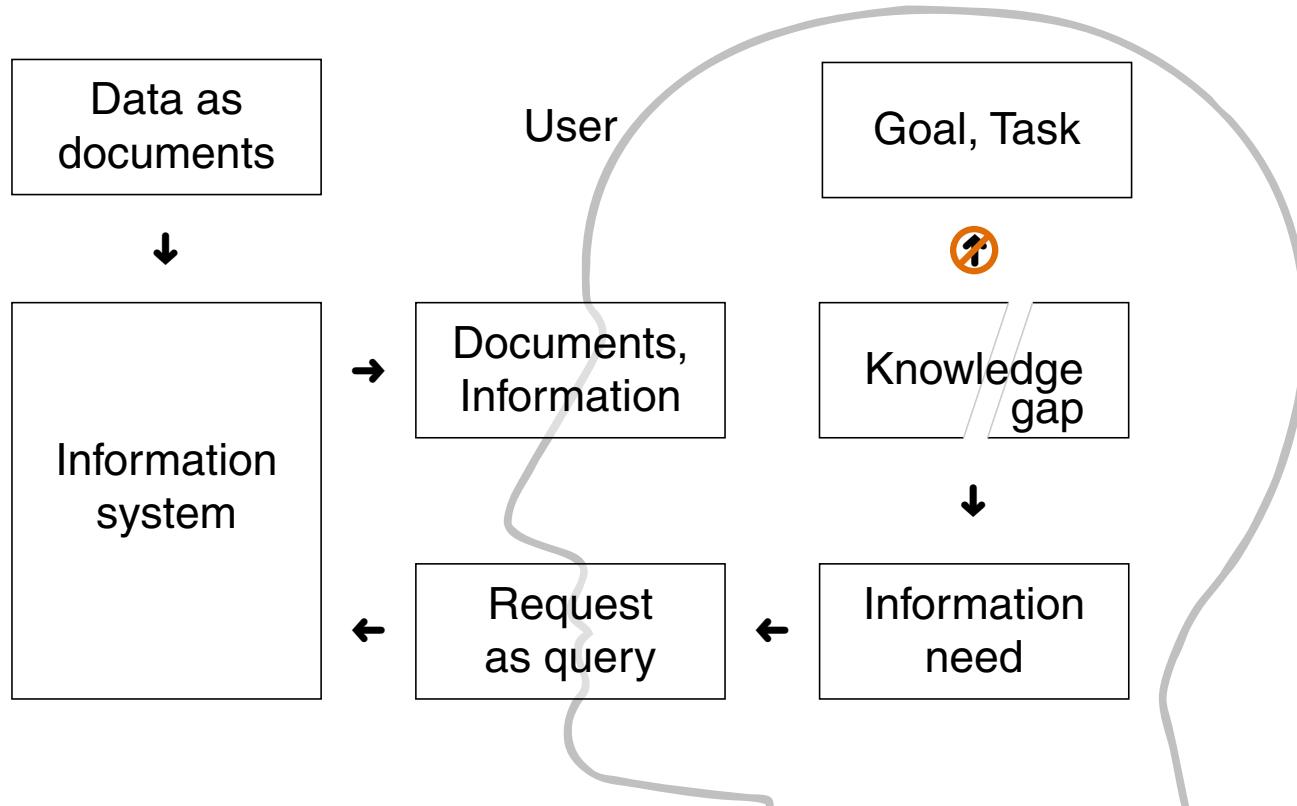
Information Retrieval (IR)



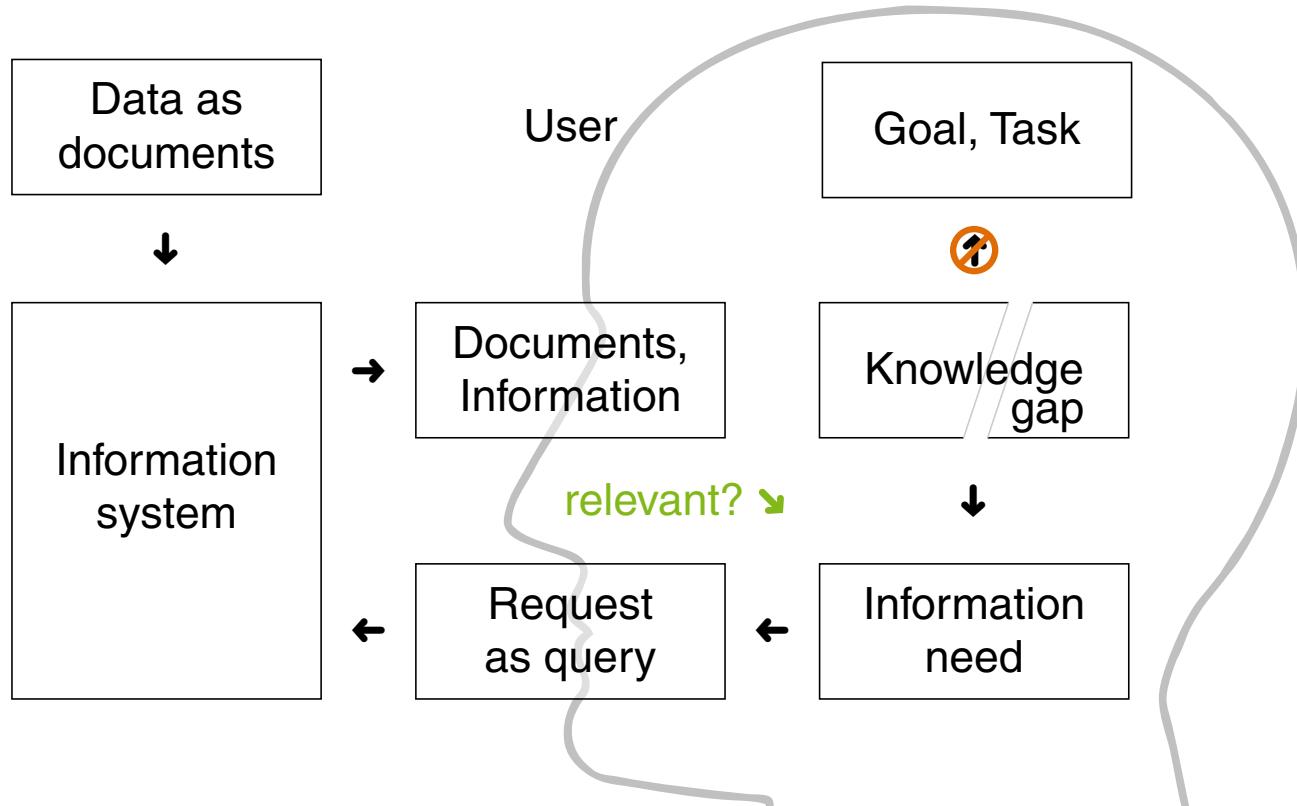
Information Retrieval (IR)



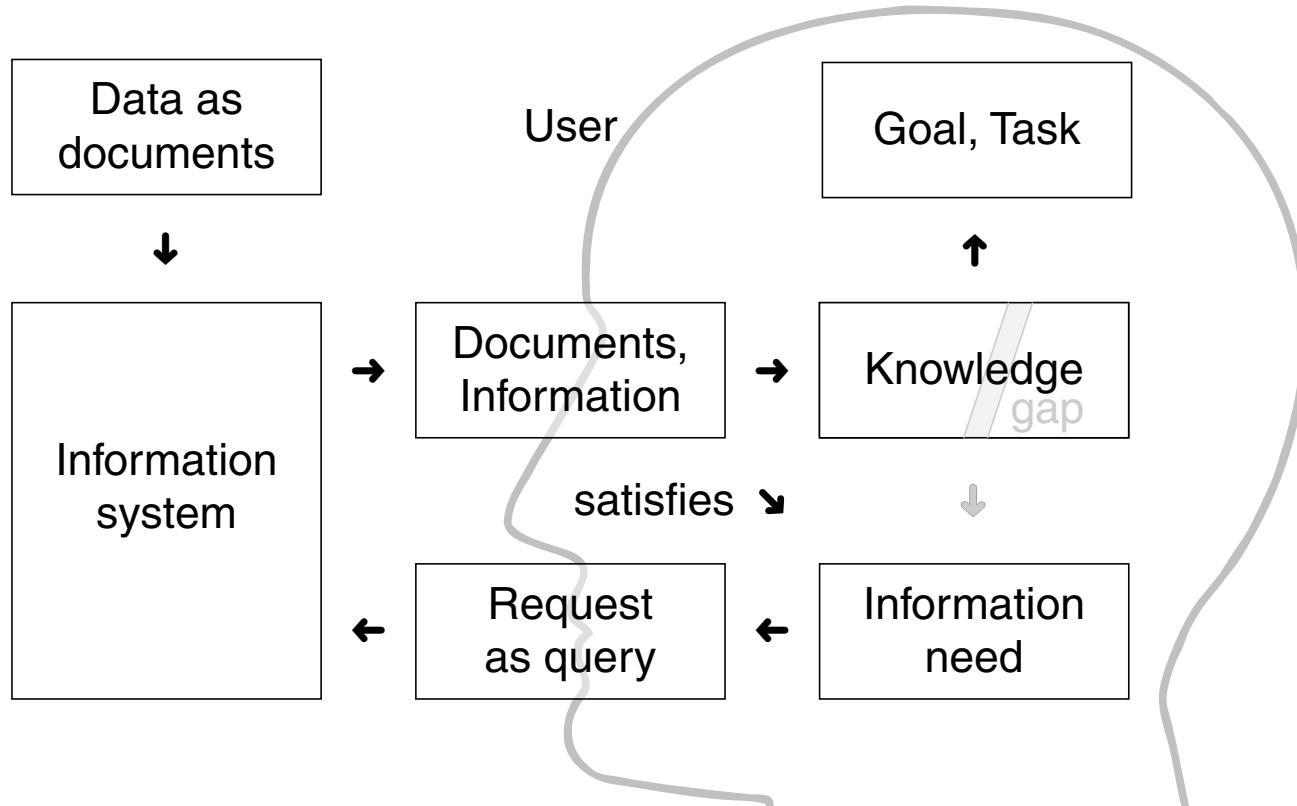
Information Retrieval (IR)



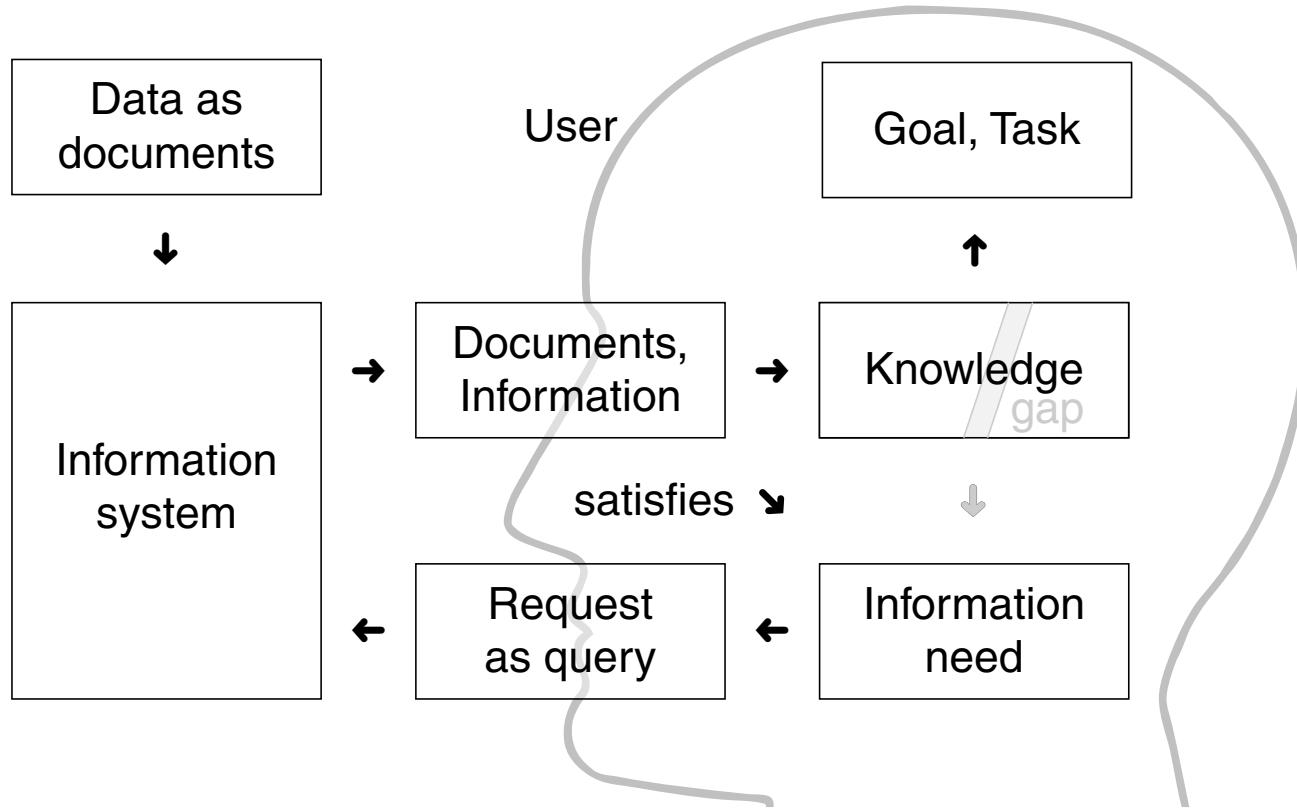
Information Retrieval (IR)



Information Retrieval (IR)



Information Retrieval (IR)



Definition 1 (Information Retrieval, IR)

The activity of obtaining information relevant to an information need from data.

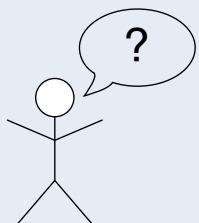
As a research field, information retrieval studies the role of information systems in transferring knowledge via data, as well as the design, implementation, evaluation, and analysis of such systems.

Web Search Architecture



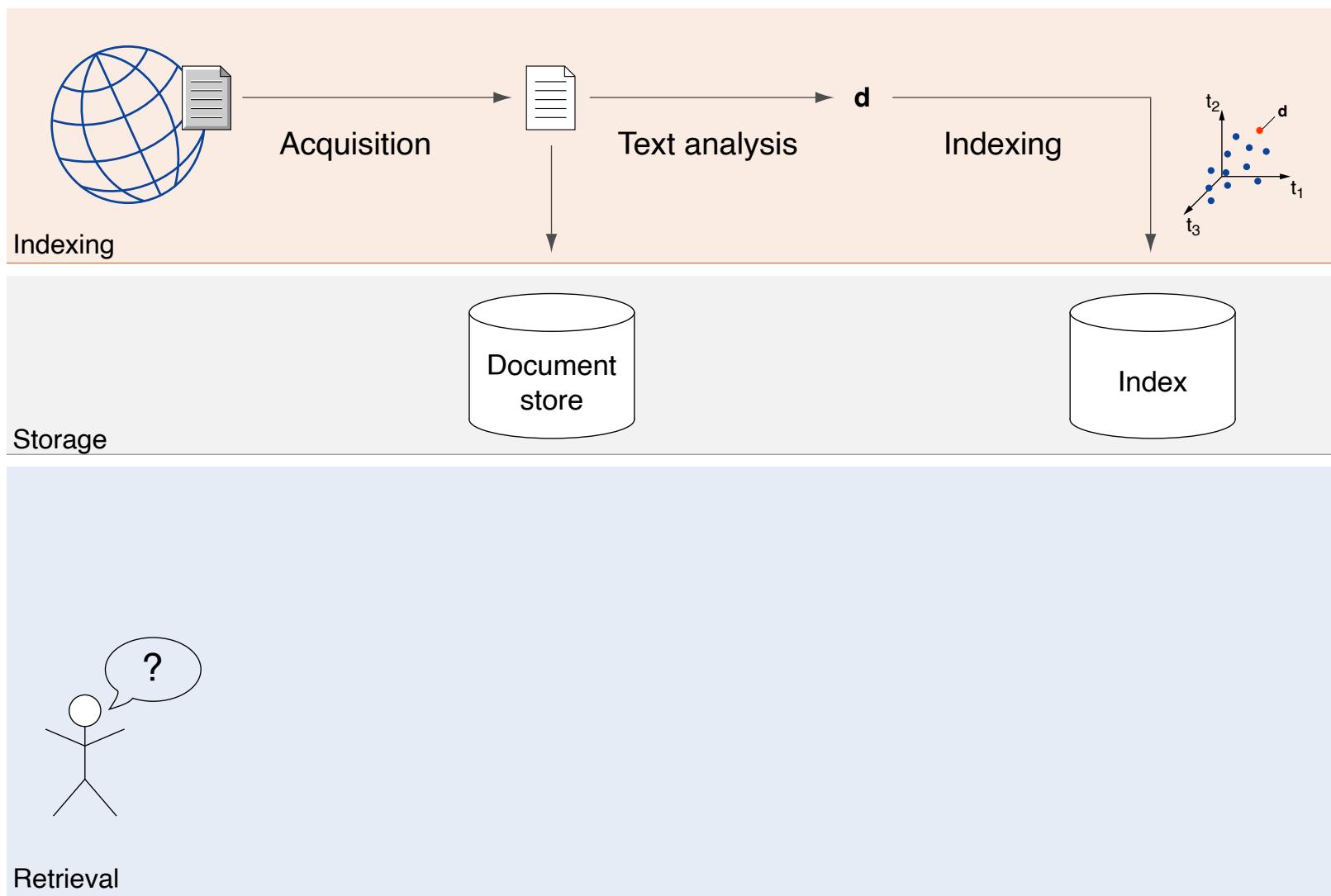
Indexing

Storage

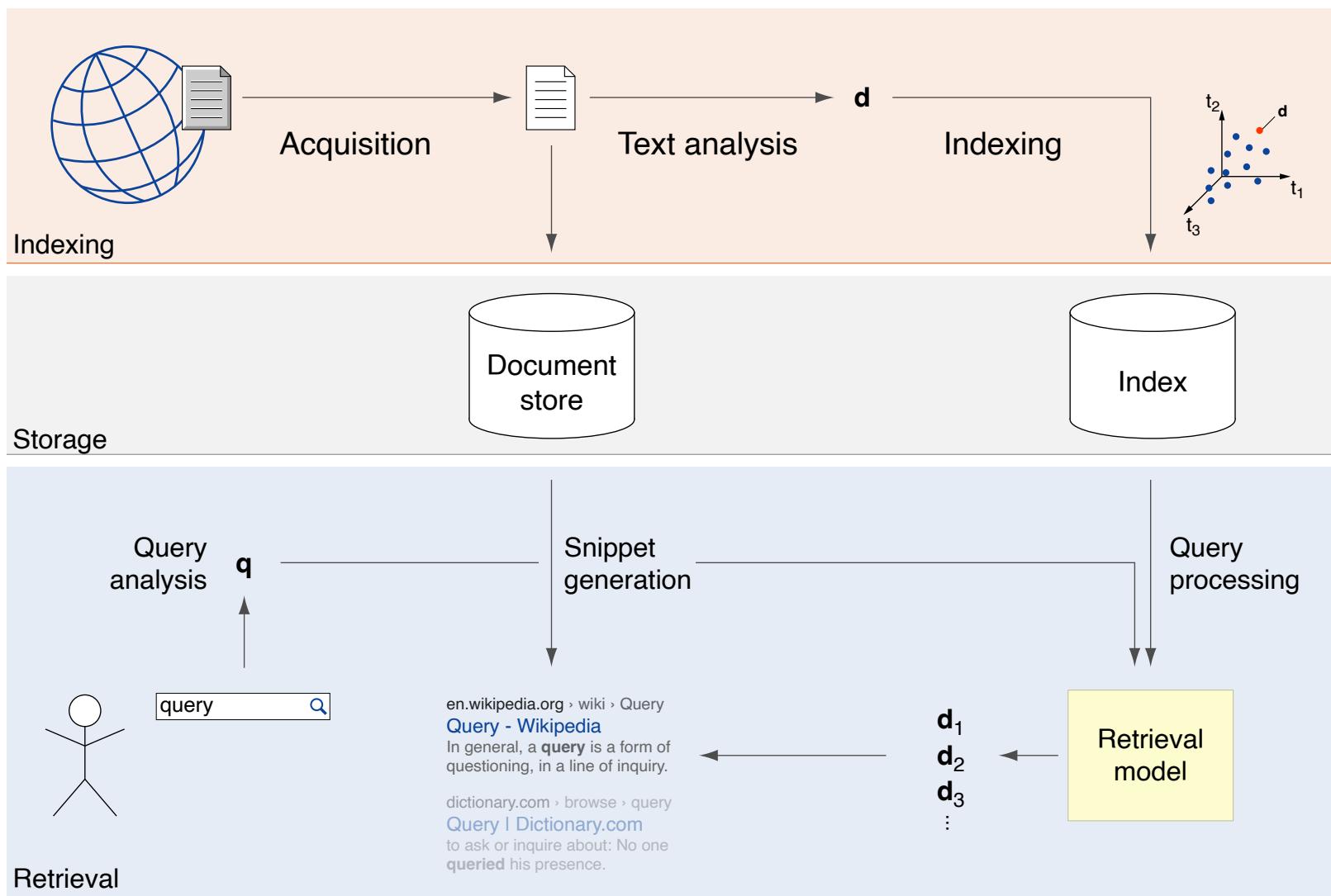


Retrieval

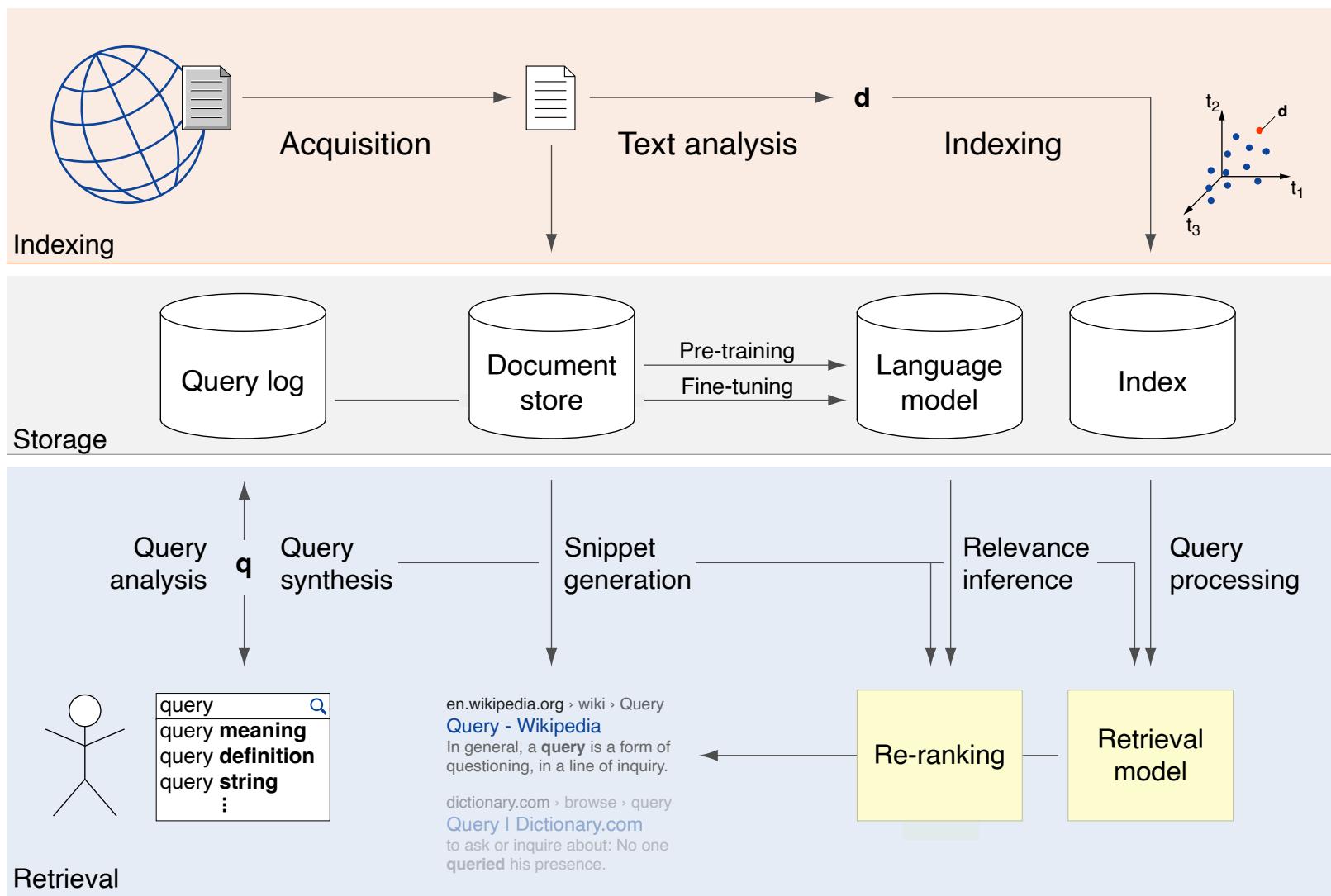
Web Search Architecture



Web Search Architecture



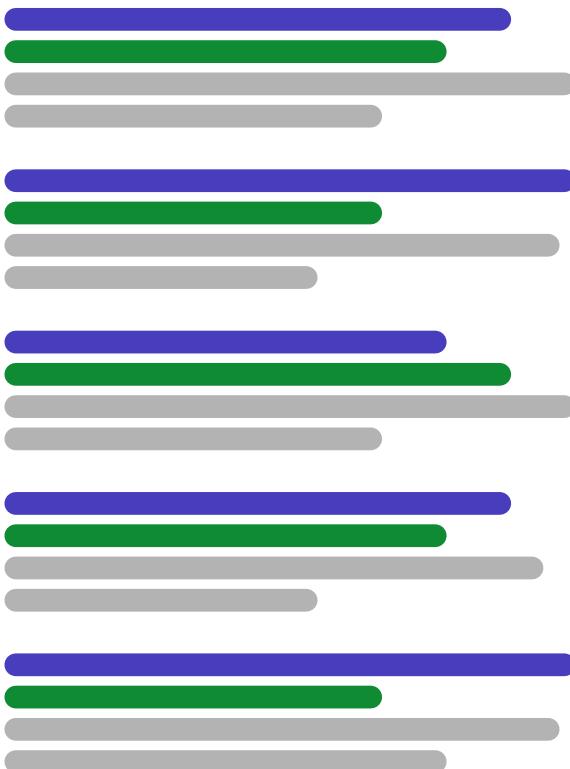
Web Search Architecture



Retrieval-Augmented Generation

How to satisfy an information need:

Query



Query



- ↑ Convenient, but uncertain veracity
- ← Authoritative, but tedious to analyze

Retrieval-Augmented Generation

How to satisfy an information need:

Query



Query



The dilemma of the direct answer: [\[Potthast et al., 2020\]](#)

A user's choice between convenience and diligence
when using an information retrieval system.

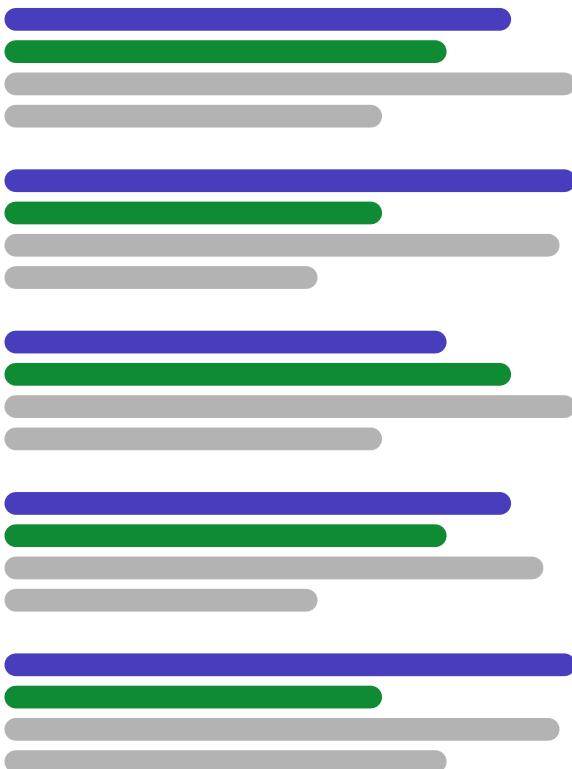


- ↑ Convenient, but uncertain veracity
- ← Authoritative, but tedious to analyze

Retrieval-Augmented Generation

How to satisfy an information need:

Query

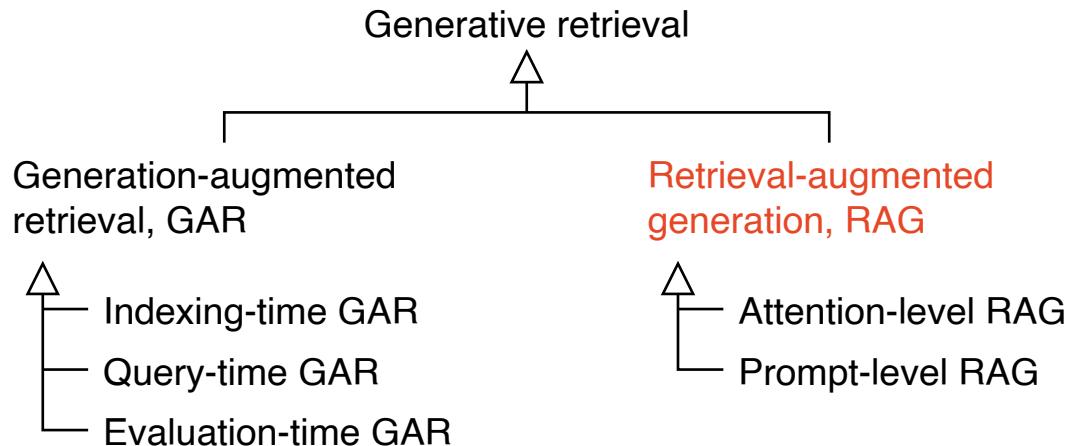


Query

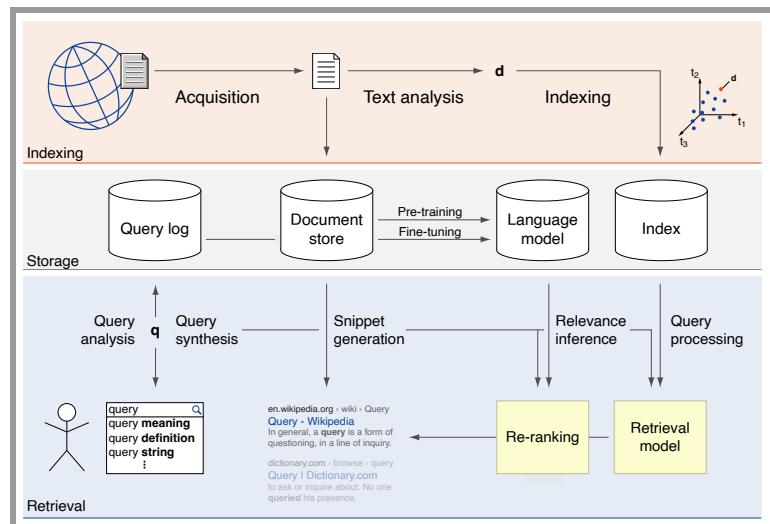
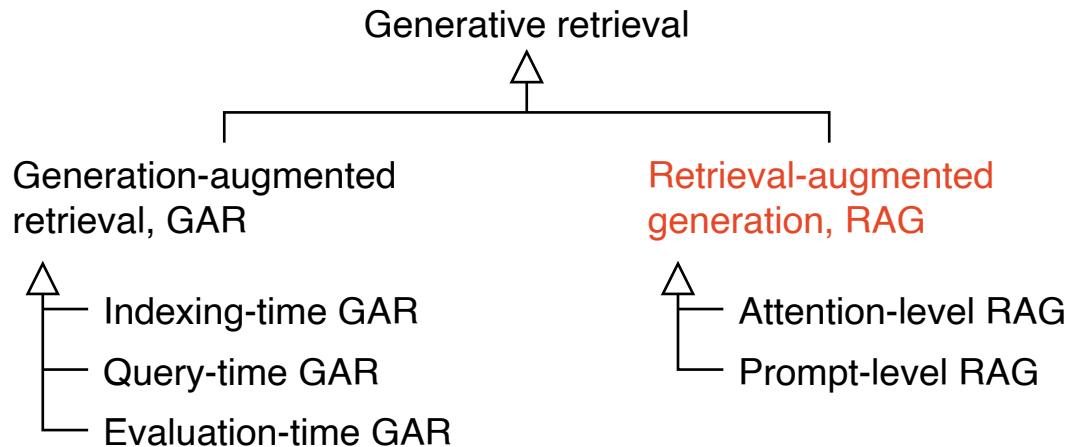


- ↑ Convenient, but uncertain veracity
- ← Authoritative, but tedious to analyze

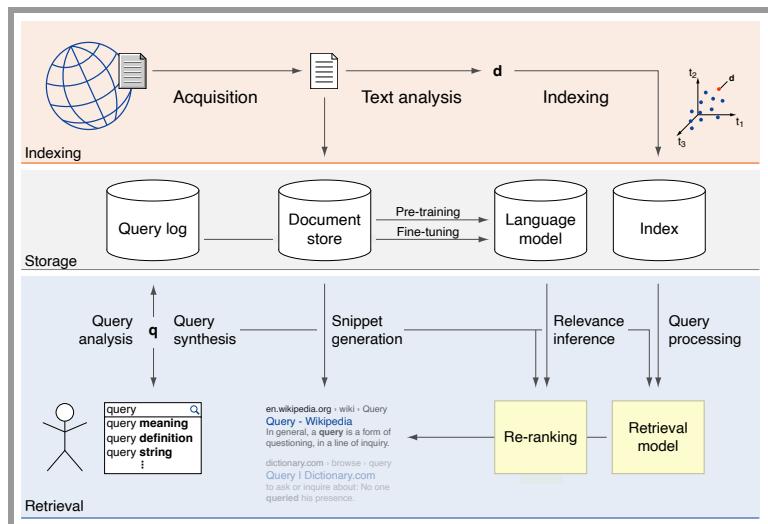
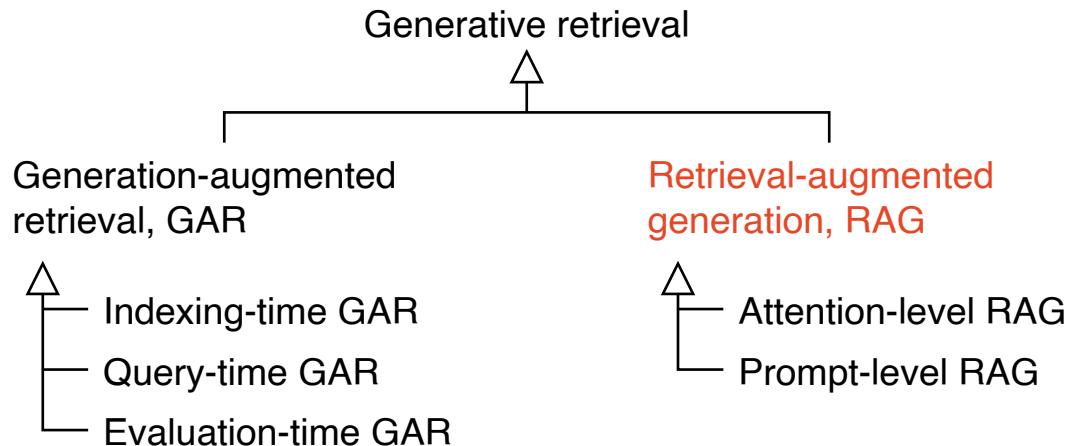
Retrieval-Augmented Generation



Retrieval-Augmented Generation



Retrieval-Augmented Generation



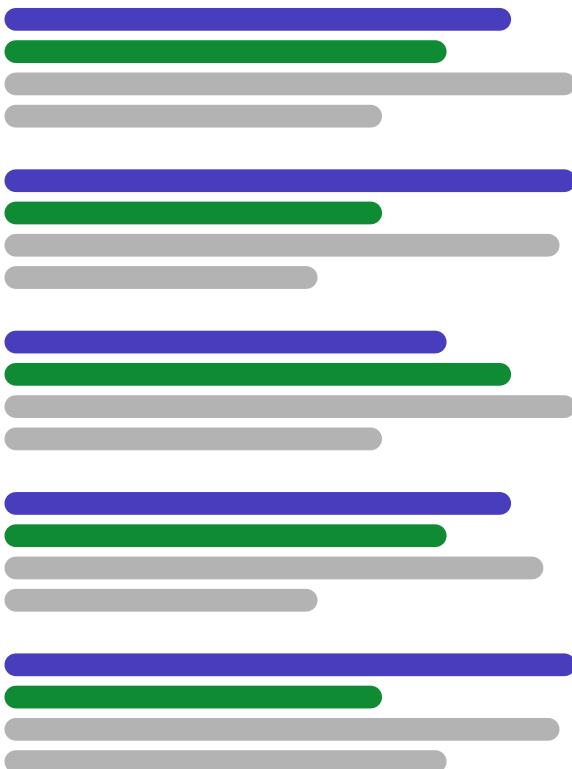
Attention-level RAG: [Lewis et al., 2020]

- Neural retrieval of document embeddings at query time
- Attention to retrieved document embeddings during text generation
- End-to-end tuning of generator and retriever at training time

Retrieval-Augmented Generation

How to satisfy an information need:

Query



Query

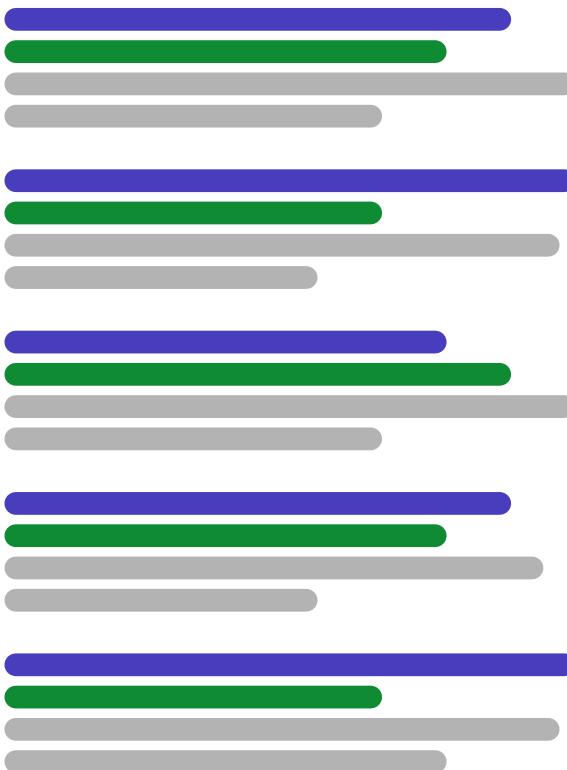


- ↑ Convenient, but uncertain veracity
- ← Authoritative, but tedious to analyze

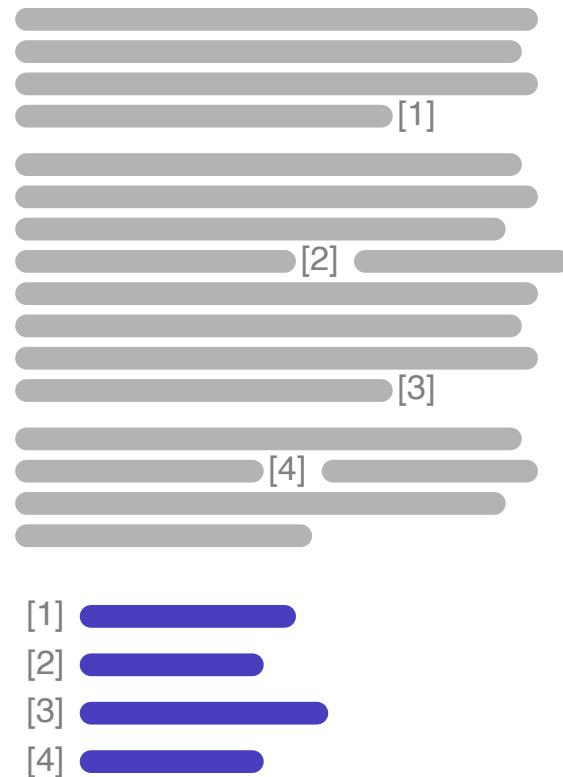
Retrieval-Augmented Generation

How to satisfy an information need:

Query



Query



What makes a good response?

What makes a good response?

How tall is the Empire State Building?

A tabby cat is distinguished by its unique coat patterns, which typically include the classic "M" shape on the forehead, stripes, dots, or swirling patterns on the body, and ring-like markings on the legs and tail. The four primary tabby patterns are classic, mackerel, spotted, and ticked. Tabby cats also have agouti hairs, which are hairs that have multiple colors along their length, contributing to their mottled appearance. Their coat can come in a variety of colors such as brown, grey, and orange.

The Empire State Building stands at 381 meters without the antenna and reaches 443 meters when the antenna is included.

What makes a good response?

How tall is the Empire State Building?

A tabby cat is distinguished by its unique coat patterns, which typically include the classic "M" shape on the forehead, stripes, dots, or swirling patterns on the body, and ring-like markings on the legs and tail. The four primary tabby patterns are classic, mackerel, spotted, and ticked. Tabby cats also have agouti hairs, which are hairs that have multiple colors along their length, contributing to their mottled appearance. Their coat can come in a variety of colors such as brown, grey, and orange.

The Empire State Building stands at 381 meters without the antenna and reaches 443 meters when the antenna is included.

Quality aspects:

1. Relevance: The user's information need must be addressed

What makes a good response?

How tall is the Empire State Building?

The Empire State Building stands at 381 meters without the antenna and reaches 443 meters when the antenna is included.

The Empire State Building stands at 381 meters without the antenna and reaches 443 meters when the antenna is included. [1]

[1] https://de.wikipedia.org/wiki/Empire_State_Building

Quality aspects:

1. Relevance: The user's information need must be addressed

What makes a good response?

How tall is the Empire State Building?

The Empire State Building stands at 381 meters without the antenna and reaches 443 meters when the antenna is included.

The Empire State Building stands at 381 meters without the antenna and reaches 443 meters when the antenna is included. [1]

[1] https://de.wikipedia.org/wiki/Empire_State_Building

Quality aspects:

1. Relevance: The user's information need must be addressed
2. Grounding: The response must be verifiable from sources

What makes a good response?

How tall is the Empire State Building?

Ah, an inquiry regarding the towering vertical majesty that is the Empire State Building! Well, let us first dissect this marvel of architectural splendor by considering its height in two distinct forms. If one is to contemplate the structural essence of the building sans antenna, it reaches an elevation of precisely 381 meters—yes, 381, that rather specific number which seems to land with elegant finality. However, if one dares to include the soaring antenna that crowns this iconic skyscraper, a further extension of the structure's grandeur is achieved, bringing its total height to a dizzying 443 meters! Indeed, a height that challenges one's very perception of the skyward limit. Thus, the Empire State Building, in its duality of heights, stands either at 381 meters in a minimalist contemplation of its mass or at 443 meters when fully embracing its towering ambition.

The Empire State Building stands at 381 meters without the antenna and reaches 443 meters when the antenna is included.

Quality aspects:

1. Relevance: The user's information need must be addressed
2. Grounding: The response must be verifiable from sources

What makes a good response?

How tall is the Empire State Building?

Ah, an inquiry regarding the towering vertical majesty that is the Empire State Building! Well, let us first dissect this marvel of architectural splendor by considering its height in two distinct forms. If one is to contemplate the structural essence of the building sans antenna, it reaches an elevation of precisely 381 meters—yes, 381, that rather specific number which seems to land with elegant finality. However, if one dares to include the soaring antenna that crowns this iconic skyscraper, a further extension of the structure's grandeur is achieved, bringing its total height to a dizzying 443 meters! Indeed, a height that challenges one's very perception of the skyward limit. Thus, the Empire State Building, in its duality of heights, stands either at 381 meters in a minimalist contemplation of its mass or at 443 meters when fully embracing its towering ambition.

The Empire State Building stands at 381 meters without the antenna and reaches 443 meters when the antenna is included.

Quality aspects:

1. Relevance: The user's information need must be addressed
2. Grounding: The response must be verifiable from sources
3. Presentation: The response must be readable

Which response is better?

How do I tell that I have a virus?

You might have a virus if your computer is slow or crashes often. Look for unexpected pop-ups and missing files. Strange behavior, like settings changing on their own, can also indicate a virus. Finally, check for unusual network activity.

To determine if you have a virus, watch for symptoms like fever, cough, fatigue, muscle aches, sore throat, headache, runny or stuffy nose, nausea, or diarrhea. If you experience multiple symptoms, especially if they worsen, consult a healthcare professional for diagnosis and testing.

Quality aspects:

1. Relevance: The user's information need must be addressed
2. Grounding: The response must be verifiable from sources
3. Presentation: The response must be readable

Which response is better?

What society issues did goth rock address?

Gothic rock delves into a variety of thought-provoking topics, focusing on societal issues such as racism, war, and group hatred. The music often presents a unique perspective on life, seeking to find beauty in life, pain, and death [0].

[0] <https://www.example.org/a.html>

Gothic rock delves into a variety of thought-provoking topics, focusing on societal issues such as racism, war, and group hatred. The music often presents a unique perspective on life, seeking to find beauty in life, pain, and death [0].

[0] <https://www.example.org/b.html>

Quality aspects:

1. Relevance: The user's information need must be addressed
2. Grounding: The response must be verifiable from sources
3. Presentation: The response must be readable

Which response is better?

How do I divide two fractions?

To divide fractions, multiply the first fraction (the dividend) by the reciprocal of the second fraction (the divisor).

To divide fractions, flip the second fraction upside down and then multiply it with the first fraction. After multiplying, simplify the answer.

Quality aspects:

1. Relevance: The user's information need must be addressed
2. Grounding: The response must be verifiable from sources
3. Presentation: The response must be readable

Agenda

- ❑ Motivation
- ❑ Information Retrieval
- ❑ Web Search Architecture
- ❑ Retrieval-Augmented Generation
- ❑ What's a good response?
- ❑ Lab Organization
- ❑ Getting started
 - Onboarding
 - Cluster access
 - Cluster use

Your tasks

- ❑ Develop and implement heuristics for automatically evaluating RAG responses regarding the quality aspects above:
 1. Retrieval aspect
 2. Grounding aspect
 3. Presentation aspect
- ❑ Create a simple UI for users to analyze RAG responses
- ❑ Deliverables
 - Presentation (30 min per group)
 - Lab report (4-10 pages)
 - System and code
- ❑ All course material can be found on temir.org

Remarks: (learning objectives)

- Work in a structured and self-supervised manner
- Work on a project of a larger scope
- Deal with open-ended tasks
- Groupwork and communication skills
- Apply and extend current research and tools in the field of generative IR
- Develop and carry out experiments
- Scientific writing
- Evaluate RAG systems
- Train and apply transformer models
- Provide a user interface for a machine learning model
- Apply machine learning to a real life problem

Getting Started

- ❑ Onboarding (cluster access & Discord)
- ❑ We need from you: e-mail address & matriculation number
- ❑ Until next Week:
 - Completed onboarding
 - Work on the introduction notebook
 - Form groups
- ❑ Papers & useful resources will be published to temir.org