Chapter IR:II

II. Architecture of a Search Engine

- Acquisition
- Text Analysis
- Indexing

- User Interface
- Query Analysis and Synthesis
- Retrieval

- Evaluation

Remarks:

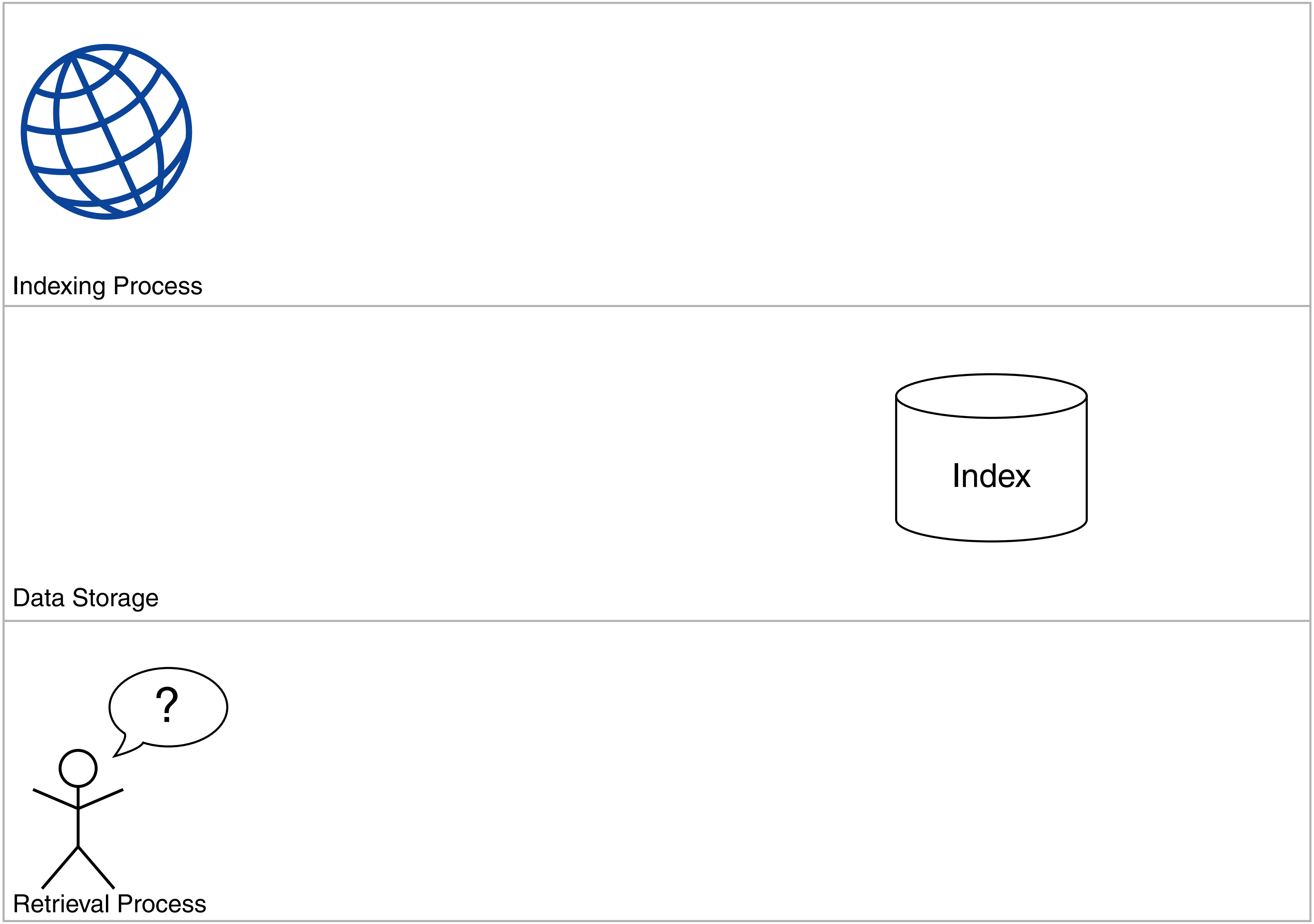
- Software architecture refers to the high level structures of a software system. These structures are needed to reason about the software system. Each comprises software elements, relations among them, and properties of both elements and relations. [Wikipedia]

- Software architecture can be specified at various levels of abstraction, also called views. We adopt a high-level functional view, showing what a search engine does, not how it is implemented.

- The implementation of a search engine must meet two requirements: effectiveness and efficiency. Effectiveness refers to retrieval quality, efficiency to retrieval speed. Other requirements boil down to these two categories. Examples: Scalability demands efficiency; result freshness improves effectiveness and demands efficiency.

- Search engines basically implement two processes, indexing and retrieval, on top of a storage layer. Indexing is a background process to prepare to-be-searched data for efficient search, as well as updating it. Retrieval offers a user interface for query submission, and implements query analysis and synthesis, and retrieval. The storage layer implements a data model for storing documents, index, and logs so that distributed and parallel search are possible.

- Compare with Google's early architecture described in “The Anatomy of a Large-scale Hypertextual Web Search Engine” by Brin and Page 1998.

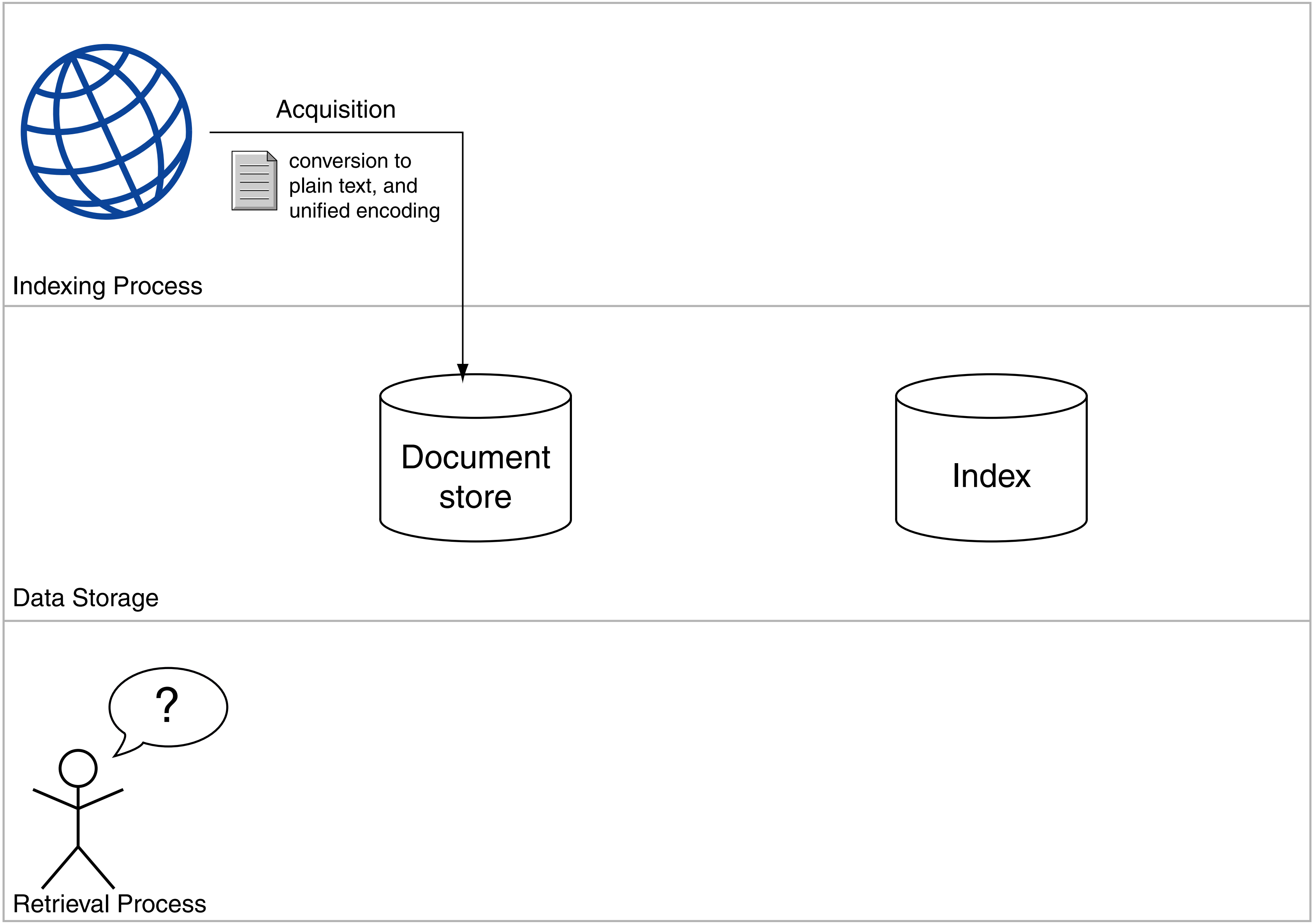


Indexing Process

Index

Data Storage

Retrieval Process



Acquisition

conversion to plain text, and unified encoding

Indexing Process

Document store

Index

Data Storage

?

Retrieval Process

Acquisition

In the acquisition step, documents are collected, prepared, and stored.

Key components:

- Crawler
- Converter
- Document Store

Acquisition

Crawler

A crawler discovers and acquires documents.

Web crawler

- Discovers new web pages via hyperlinks
- What are challenges for web crawling?

Site crawler / focused crawler / topical crawler

- Web crawler for websites / that discards documents (wrt. topics, genres, etc.)
- May exploit structured sitemaps, RSS, or Atom feeds on sites.
- May require a document classifier to identify matching documents.
- Examples: academic search, news search, business search, job search, etc.

Document crawler

- Scans local directories, emails, databases, etc.
- Examples: enterprise search, desktop search

Acquisition

Crawler

A crawler discovers and acquires documents.

Web crawler

- Discovers new web pages via hyperlinks
- Exploration policy (e.g., avoid spider traps), duplicate identification (e.g., URL normalization), parallelization, revisit / update policy, politeness

Site crawler / focused crawler / topical crawler

- Web crawler for websites / that discards documents (wrt. topics, genres, etc.)
- May exploit structured sitemaps, RSS, or Atom feeds on sites.
- May require a document classifier to identify matching documents.
- Examples: academic search, news search, business search, job search, etc.

Document crawler

- Scans local directories, emails, databases, etc.
- Examples: enterprise search, desktop search

Acquisition

Converter

A converter unifies documents as follows:

Reformatting / text extraction

- Documents come in a variety of formats. Examples: HTML, PDF, DOC
- Implementation benefits from unified formats (HTML in web search)
- Plain text extraction from documents is lossy (e.g., formatting / layout is lost)
- Layout analysis is still useful for retrieval (see segmentation)
- Document formats are often invalid, they must still be parsed

Encoding normalization

- Plain text documents come in a variety of encodings (e.g., ASCII, Unicode)
- Subsequent processing steps require unified input encoding (e.g., Unicode)
- Encoding specifications are untrustworthy, encodings must be detected
- Documents' encodings are often invalid, they must be repaired

Errors propagate; when visible in search results, the search engine is blamed.

Acquisition

Document Store

The document store manages all documents acquired:

Original and converted documents

- Why is it sensible to mirror documents locally?

Document metadata

- Provenance data, such as origin, crawl date, etc.

Version history

- Every recrawl of a document is kept.
- Older document versions are useful for later analyses.

Scale often demands for a distributed document store.

Acquisition

Document Store

The document store manages all documents acquired:

Original and converted documents

- Original may not always be available
- Fast reprocessing (e.g., when a processing step is improved)
- Fast snippet generation

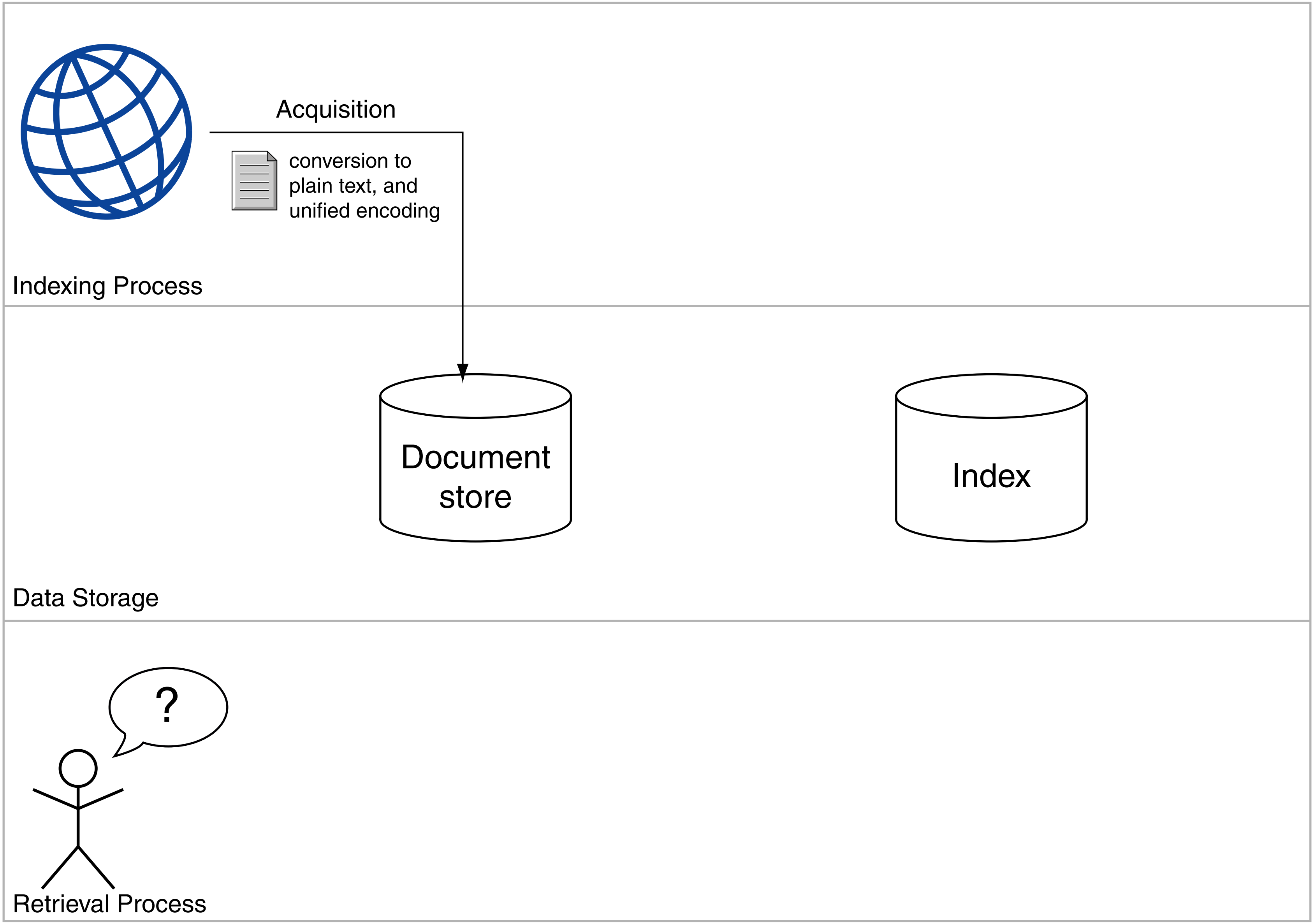
Document metadata

- Provenance data, such as origin, crawl date, etc.

Version history

- Every recrawl of a document is kept.
- Older document versions are useful for later analyses.

Scale often demands for a distributed document store.



Acquisition

conversion to plain text, and unified encoding

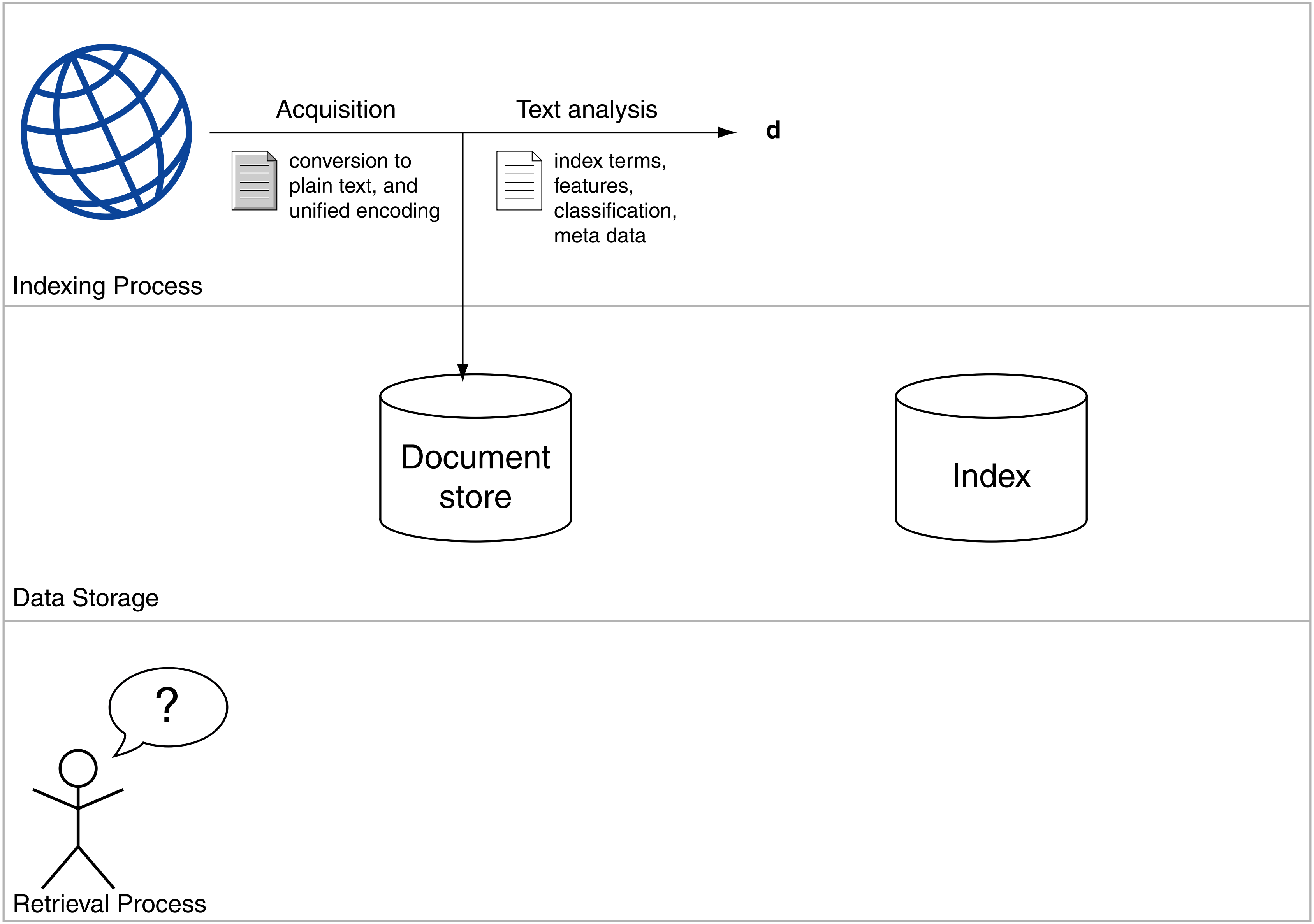
Indexing Process

Document store

Index

Data Storage

Retrieval Process


Acquisition
Text analysis
d
conversion to plain text, and unified encoding
index terms, features, classification, meta data
Indexing Process
Document store
Index
Data Storage
?
Retrieval Process

Text Analysis

The text analysis extracts from a document the “keys” by which it can be looked up in the index. Two kinds of keys are distinguished:

Index terms (terms, for short)

- Words or phrases from a document’s text
- Their purpose is to represent what a document is about
- All index terms of all documents combined form the terminology

Features

- A feature is a measurable property of a document, a feature set represents it
- Different feature sets are suitable for different classification targets
- Example: relevance, spam, language, genre, . . .

Text Analysis

The text analysis extracts from a document the “keys” by which it can be looked up in the index. Two kinds of keys are distinguished:

Index terms (terms, for short)

- Words or phrases from a document’s text
- Their purpose is to represent what a document is about
- All index terms of all documents combined form the terminology

Features

- A feature is a measurable property of a document, a feature set represents it
- Different feature sets are suitable for different classification targets
- Example: relevance, spam, language, genre, . . .

Key components:

- Segmenter
- Stopping
- Stemmer / Lemmatizer
- Link Extraction
- Information Extraction
- Classification

Text Analysis

Segmenter

A segmenter breaks down a document into its constituent parts at two levels:

Page segmentation

- Analysis of the HTML source of a web page with regard to its structure
- Extraction of main content vs. ads, navigation, header, footer, etc.
- HTML pages often do not meet the specification (e.g., semantic elements)
- Extraction of text structure and text formatting (e.g., headings, paragraphs)

Text segmentation

- Analysis of a plain text with regard to linguistic units (e.g., words, sentences)
- Tokenization turns a text into a token sequence (token ≈ word).
  (a) White space tokenization: tokens are separated by white space characters
  (b) Regular expression tokenization: a token is an alphanumeric string
  – Why are these definitions insufficient?

- Lower-casing of words ensure more matches with queries words

Text Analysis

Segmenter

A segmenter breaks down a document into its constituent parts at two levels:

Page segmentation

- ❑ Analysis of the HTML source of a web page with regard to its structure
- ❑ Extraction of main content vs. ads, navigation, header, footer, etc.
- ❑ HTML pages often do not meet the specification (e.g., semantic elements)
- ❑ Extraction of text structure and text formatting (e.g., headings, paragraphs)

Text segmentation

- ❑ Analysis of a plain text with regard to linguistic units (e.g., words, sentences)
- ❑ Tokenization turns a text into a token sequence (token ≈ word).
  - (a) White space tokenization: tokens are separated by white space characters
  - (b) Regular expression tokenization: a token is an alphanumeric string
  - – Words are not only separated by white space, but also punctuation
  - – Contractions and words with special characters are neglected
- ❑ Lower-casing of words ensure more matches with queries words

Text Analysis

Stopping

Stopping (also stop word removal) discards a selection of words from the set of index terms of a document. Candidates for stop words:

- ❑ Function words

  Words that carry little semantics, are ambiguous, serve only grammatical purposes, or specify attitude or mood. Examples: the, of, to, for.

- ❑ Frequent words

  The most frequently appearing words of a language, or within a collection of documents. Example: “Wikipedia” appears on every Wikipedia page.

- ❑ Domain-specific words

  Words that do not discriminate in a given search domain. Example: “learning” may be ignored in the education domain, regardless of its frequency.

- ❑ Upsides: reduced index size, faster query processing speed, reduced noise.
- ❑ Downsides: many special cases missed. Example: “to be or not to be”.
- ❑ Retrieval effectiveness improvement depends on the retrieval model.
- ❑ Stopping on index terms is often more conservative than on queries.

Text Analysis

Stemming / Lemmatization

Stemming reduces inflected index terms to a common stem.

Example: “statistics”, “statistic”, and “statistical” refer to basically the same abstract concept.

Two approaches to stemming can be used:

- (Heuristic) Stemming

  Rule-based affix elimination (i.e., suffixes, prefixes, and infixes). Examples: “work*er*”, “*mega*volt”, “un-*bloody*-likely”. Naive heuristic: truncate word at letter 4.

- Lemmatization

  Mapping of a word to its root form, even if it is spelled differently. Example: “saw” and “see”.

- Upside: finding more relevant documents (better retrieval effectiveness)
- What are problems related to aggressive stemming?

Text Analysis

Stemming / Lemmatization

Stemming reduces inflected index terms to a common stem.

Example: “statistics”, “statistic”, and “statistical” refer to basically the same abstract concept.

Two approaches to stemming can be used:

- (Heuristic) Stemming

  Rule-based affix elimination (i.e., suffixes, prefixes, and infixes). Examples: “work*er*”, “*mega*volt”, “un-*bloody*-likely”. Naive heuristic: truncate word at letter 4.

- Lemmatization

  Mapping of a word to its root form, even if it is spelled differently. Example: “saw” and “see”.

- Upside: finding more relevant documents (better retrieval effectiveness)

- Downsides: conflations of unrelated words, lemmatization is expensive

  Example: “university”, “universe”, and “universal” are mapped to “univers” by common stemmers. Stemmed words may not be real ones. Lemmatization requires dictionaries.

- Retrieval effectiveness improvement varies with language.

  Example: English vs. Arabic. Alternative approach: query expansion.

Text Analysis

Link Extraction

Extraction of links and anchor texts from a document. This serves two purposes:

Link analysis

- Hyperlinks induce a graph among web pages.
- Link analysis traverses this graph to identify authoritative web pages.
- Algorithms for link analysis include PageRank and HITS.

Text augmentation with anchor texts

- The text found on a web page may be insufficient to describe its contents.
- Examples: product pages or pages showing only images.
- Anchor texts as well as the text before and after may justify a link.
- Anchor texts are added to the text extracted from a linked page.

Text Analysis

Information Extraction

Information Extraction identifies more complex index terms by means of natural language processing technology:

- Noun phrases

  Phrases which have a noun as its head word (i.e., a noun and any word that modifies it).
  Examples: “*The yellow house* is for sale.”, “I want *a skateboard*.”

- Named entities

  Words or phrases that designate something in the “real” world.
  Examples: places, people, organizations, etc.

- Coreference resolution

  Coreferences (i.e., anaphora and cataphora) are expressions that refer backward or forward in a text, respectively. Resolving *them* is important for text understanding, yet, still unsolved.

- Relation detection

  Extraction of relations between named entities mentioned in the text.
  Example: “*Bill* lives in the *USA*. *He* works at the *United Nations*.”

- Semi-structured information extraction

  Extraction and analysis of tables, quotes, references, comments, etc.

Text Analysis

Classification

Machine learning is applied to classify or categorize documents based on features. Common classification goals are:

- ❑ Language identification

  Determines the (main) language of a web page.

- ❑ Spam detection / malware detection

  Determines if a website/web page is trying to undermine a search engine's ranking (spam), or harm its users (malware).

- ❑ Topic categorization

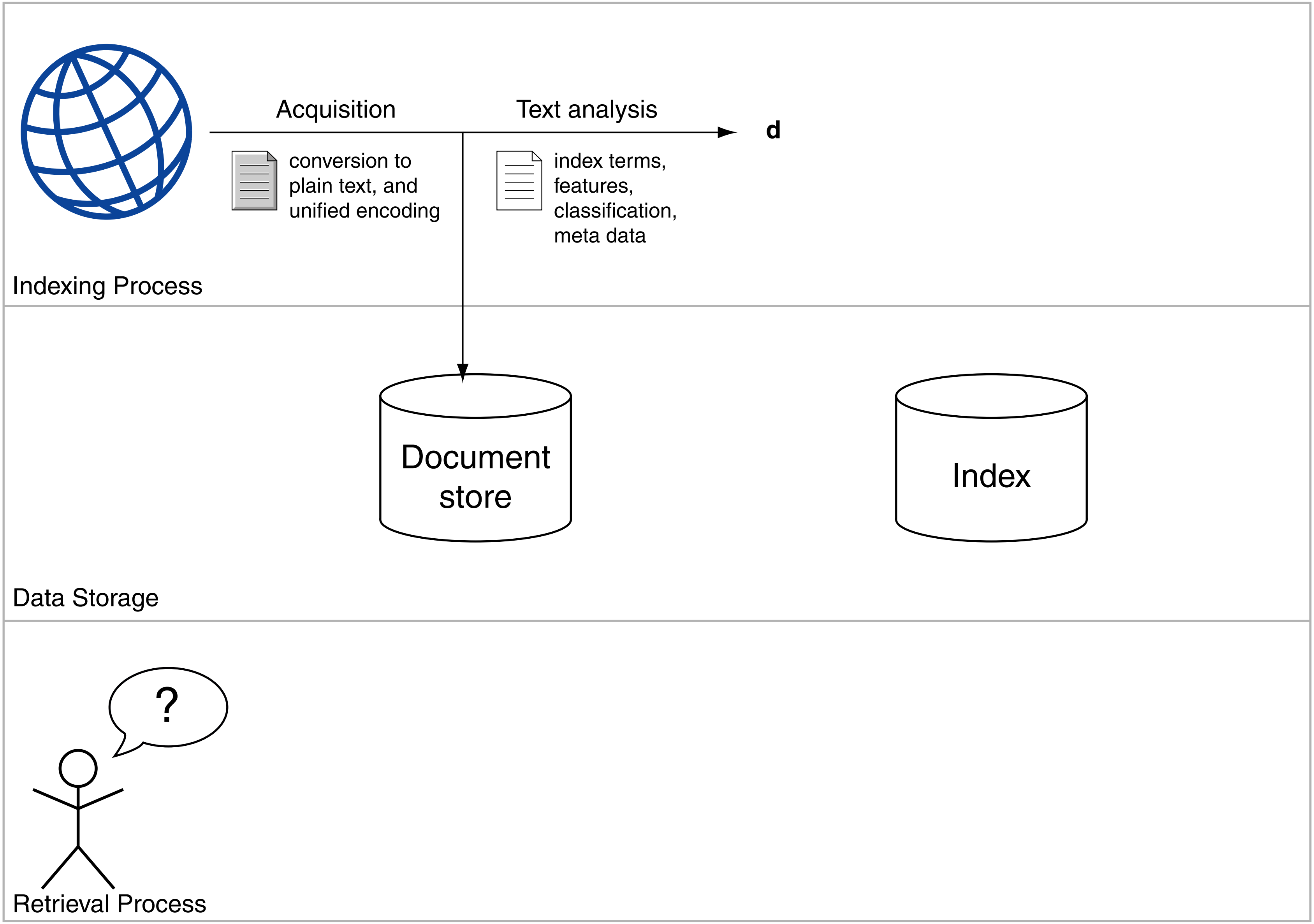
  Determines the topic of a document. Topics overlap, form hierarchies and depend on the search domain. Examples: sports, politics, technology, etc., or user-/domain-specific ones.

- ❑ Cluster analysis

  Determines previously unknown topics; automatic cluster labeling required.

- ❑ Genre categorization

  Determines the genre of a web page. Genres overlap, form hierarchies, and depend on the search domain. Examples: personal home page, message board, blog, shop, etc.



Acquisition

Text analysis

d

conversion to
plain text, and
unified encoding

index terms,
features,
classification,
meta data

Indexing Process

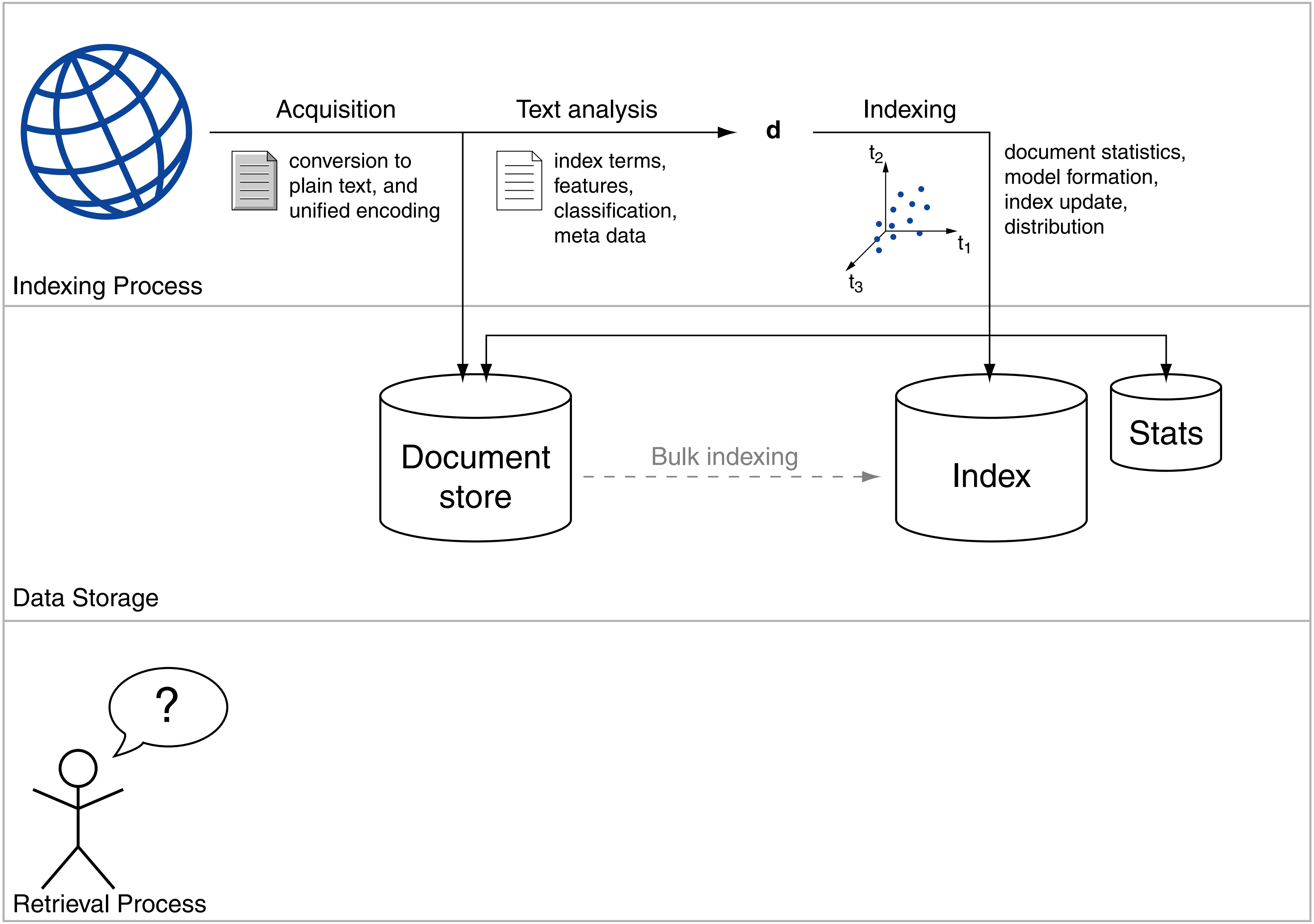
Document store

Index

Data Storage

?

Retrieval Process



Acquisition

conversion to plain text, and unified encoding

Text analysis

index terms, features, classification, meta data

d

Indexing

t_2 t_1 t_3

document statistics, model formation, index update, distribution

Indexing Process

Document store

Bulk indexing

Index

Stats

Data Storage

?

Retrieval Process

Indexing

The indexing step creates the index data structures required for fast retrieval from the set of acquired documents.

The most commonly used data structure is the inverted index.
Data structures for special purposes include the suffix array and the signature file.

Key components:

- Term Weighting
- Index Construction
- Distribution
- Document Statistics

Indexing

Term Weighting

For each index term of a document, calculate a weight indicating its importance with respect to the document. Basic term weighting schemes:

- Term frequency (*tf*)

  Logarithm of the number of occurrences of a term in a document.

- Inverse document frequency (*idf*)

  - Document frequency (*df*): Number of documents containing a term
  - Logarithm of the total number of documents divided by *df*.

- $tf \cdot idf$

  One of the most well-known term weighting schemes in IR.

- BM25

  Similar to $tf \cdot idf$, but yields better retrieval performance.

Term weights are pre-computed and store in the index to speed up document scoring during retrieval.

Indexing

Index Construction

Index construction creates an inverted index data structure by inverting the document-term data to term-document data.

T	$\rightarrow$	**Postings** (Posting Lists, Postlists)			
t_1	$\rightarrow$	$\boxed{d_1, w_{1,1}}$	$\boxed{d_2, w_{1,2}}$		
t_2	$\rightarrow$	$\boxed{d_1, w_{2,1}}$	$\boxed{d_2, w_{2,2}}$	$\boxed{d_4, w_{2,4}}$	
t_3	$\rightarrow$	$\boxed{d_1, w_{3,1}}$	$\boxed{d_2, w_{3,2}}$	$\boxed{d_4, w_{3,4}}$	$\boxed{d_5, w_{3,5}}$
t_4	$\rightarrow$	$\boxed{d_2, w_{4,2}}$			
t_5	$\rightarrow$	$\boxed{d_1, w_{5,1}}$			
$\vdots$					

- Bulk indexing

  Creates an index offline by processing all acquired documents. Once ready, the currently used index is replaced with the new one.

- Index update

  Updates a currently used index online with new documents as they appear.

Indexing

Distribution

Large indexes are distributed (sharding, partitioned) across machines:

Document distribution

- Split collection; smaller indexes for sub-collections on different machines

Term distribution

- Split the index for the entire collection by terms
- Different machines serve different terms

Replication

- Copies of (parts of) indexes at multiple sites

What are arguments for sharding by documents / terms, and for replication?

Indexing

Distribution

Large indexes are distributed (sharding, partitioned) across machines:

Document distribution

- Split collection; smaller indexes for sub-collections on different machines
- Enables parallelism for indexing and query processing
- Small indexes are faster due to caching; more search results can be retrieved

Term distribution

- Split the index for the entire collection by terms
- Different machines serve different terms
- Not all machines have to process every query

Replication

- Copies of (parts of) indexes at multiple sites
- Reduced delays during query processing
- Fault tolerance

Indexing

Document Statistics

Auxiliary information about documents is gathered for fast online query processing:

- Term frequencies per document, topic, and genre

  E.g., to compute query-dependent relevance scores.

- Term positions per document

  E.g., to speed up snippet generation.

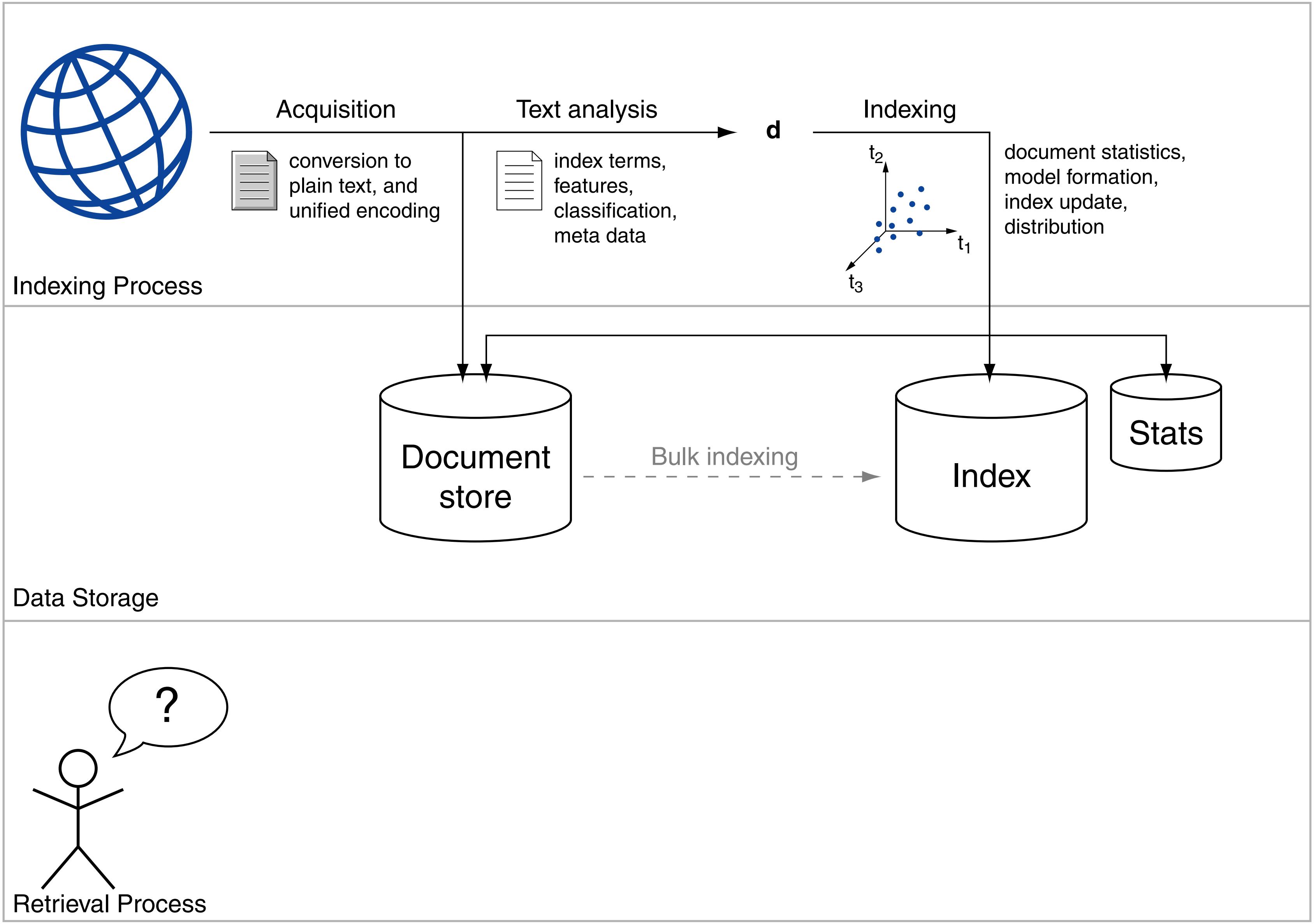
- Document frequencies per term

  The number of documents a term occurs in; required for relevance scoring.

- Document lengths

  Required for relevance scoring.

The data structure used is a basic key-value store (i.e., a hash map).



Acquisition

conversion to plain text, and unified encoding

Text analysis

index terms, features, classification, meta data

d

Indexing

t_2

t_1

t_3

document statistics, model formation, index update, distribution

Indexing Process

Document store

Bulk indexing

Index

Stats

Data Storage

?

Retrieval Process

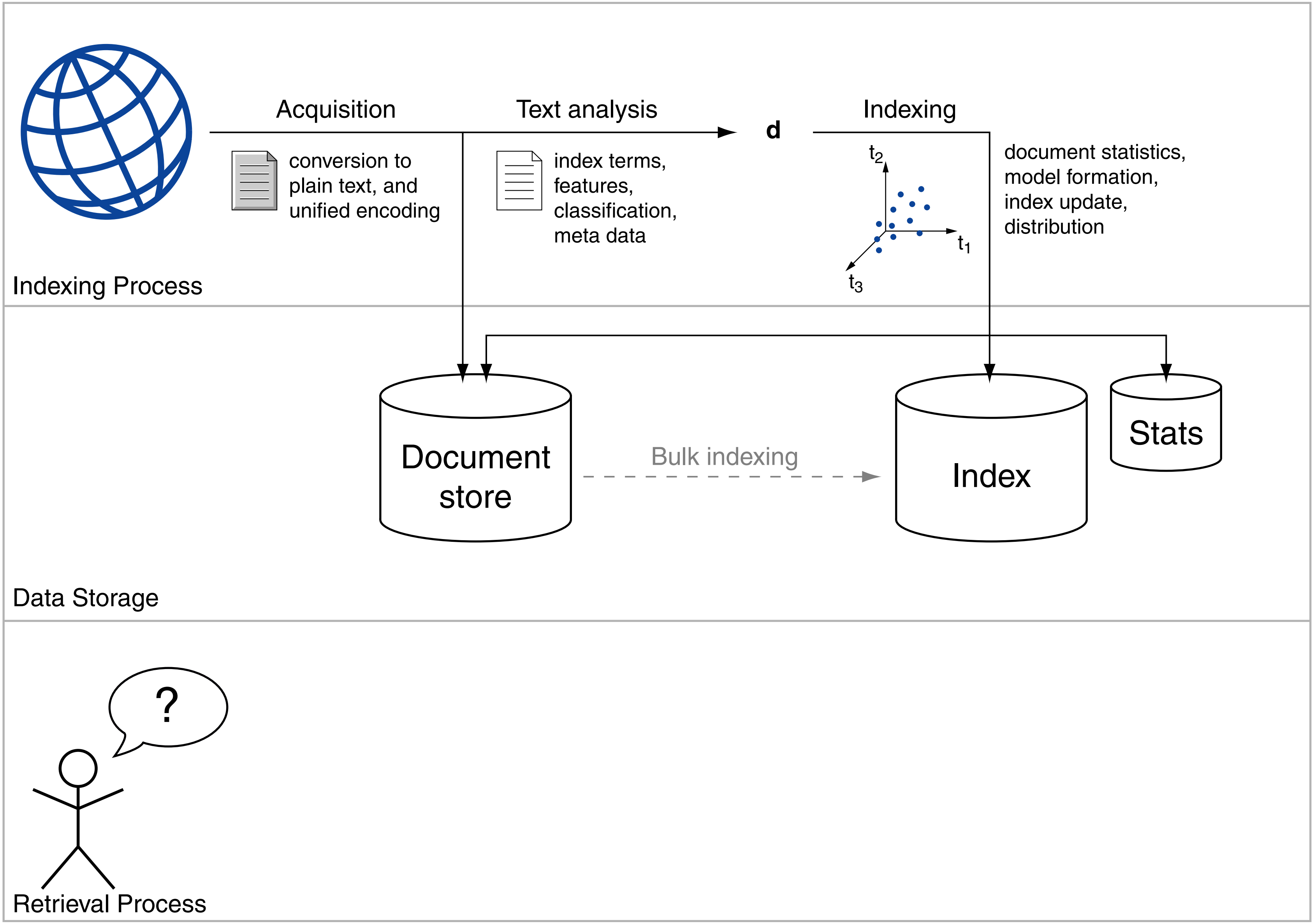
Chapter IR:II

II. Architecture of a Search Engine

- ❑ Overview
- ❑ Acquisition
- ❑ Text Analysis
- ❑ Indexing

- ❑ User Interface
- ❑ Query Analysis and Synthesis
- ❑ Retrieval

- ❑ Evaluation



Acquisition

conversion to plain text, and unified encoding

Text analysis

index terms, features, classification, meta data

d

Indexing

t_2 t_1 t_3

document statistics, model formation, index update, distribution

Indexing Process

Document store

Bulk indexing

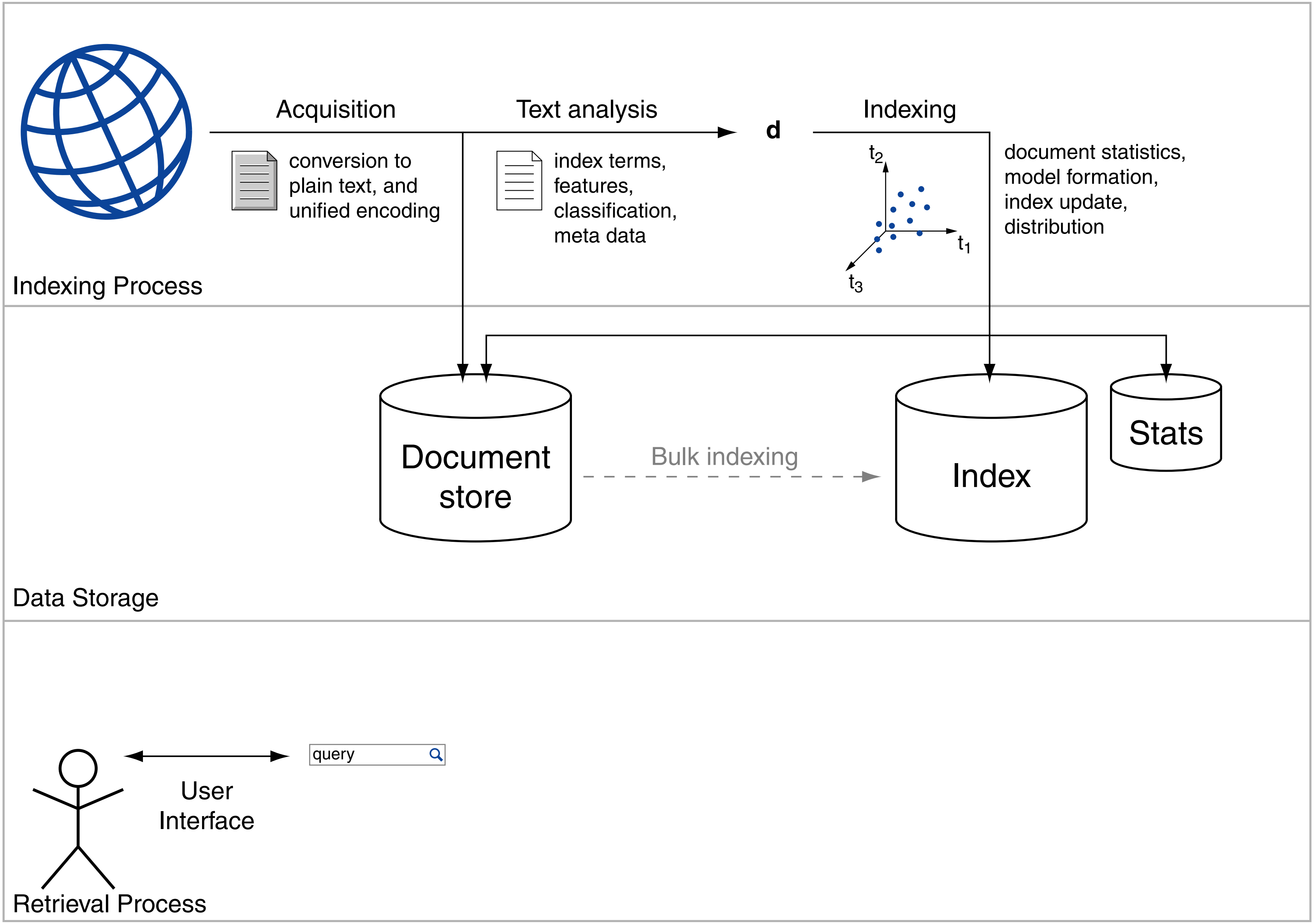
Index

Stats

Data Storage

?

Retrieval Process



Acquisition

conversion to plain text, and unified encoding

Text analysis

index terms, features, classification, meta data

Indexing

document statistics, model formation, index update, distribution

Indexing Process

Document store

Bulk indexing

Index

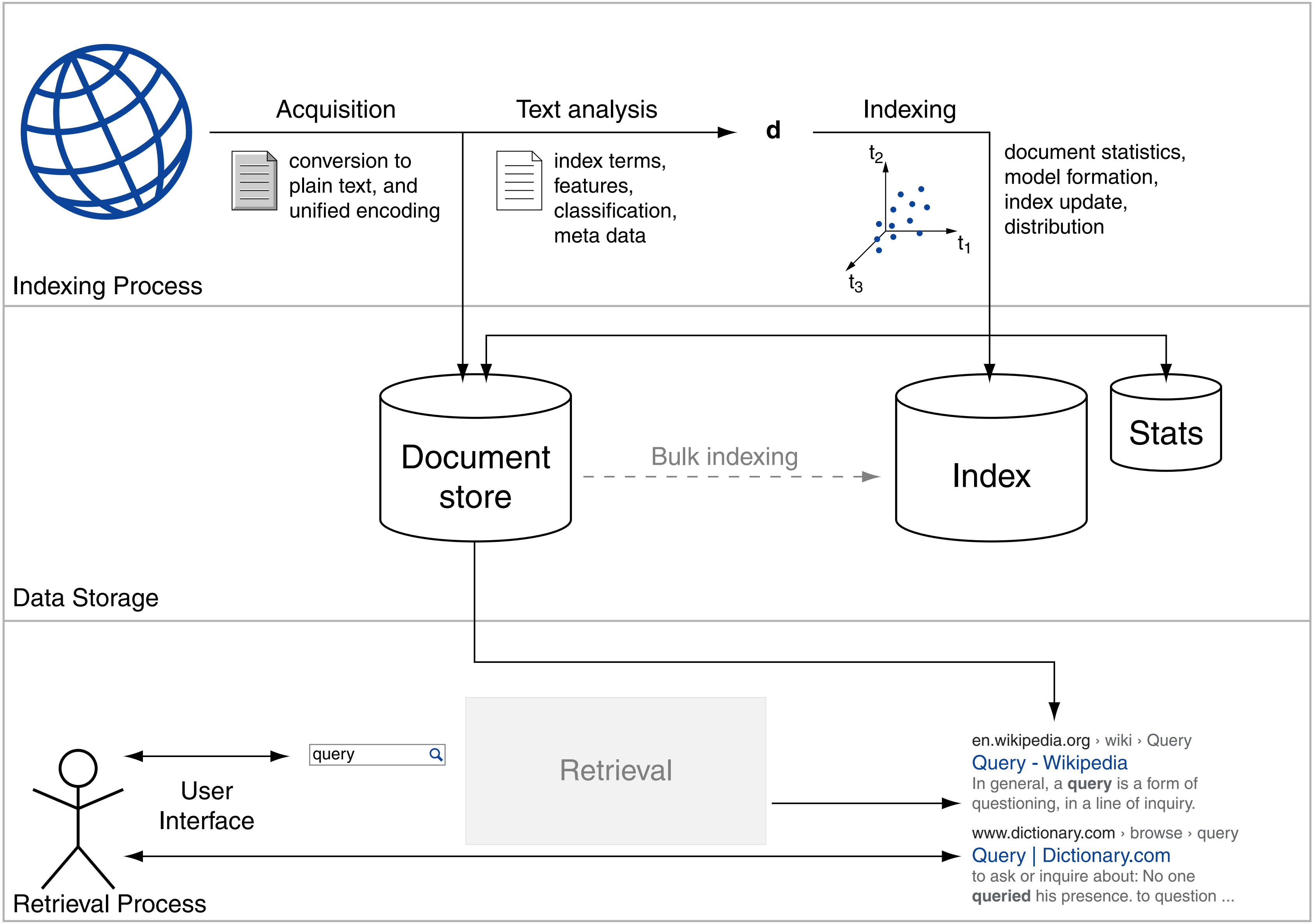
Stats

Data Storage

User Interface

query

Retrieval Process



Acquisition

conversion to plain text, and unified encoding

Text analysis

index terms, features, classification, meta data

Indexing

document statistics, model formation, index update, distribution

Indexing Process

Document store

Bulk indexing

Index

Stats

Data Storage

User Interface

query

Retrieval

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

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

Retrieval Process

User Interface

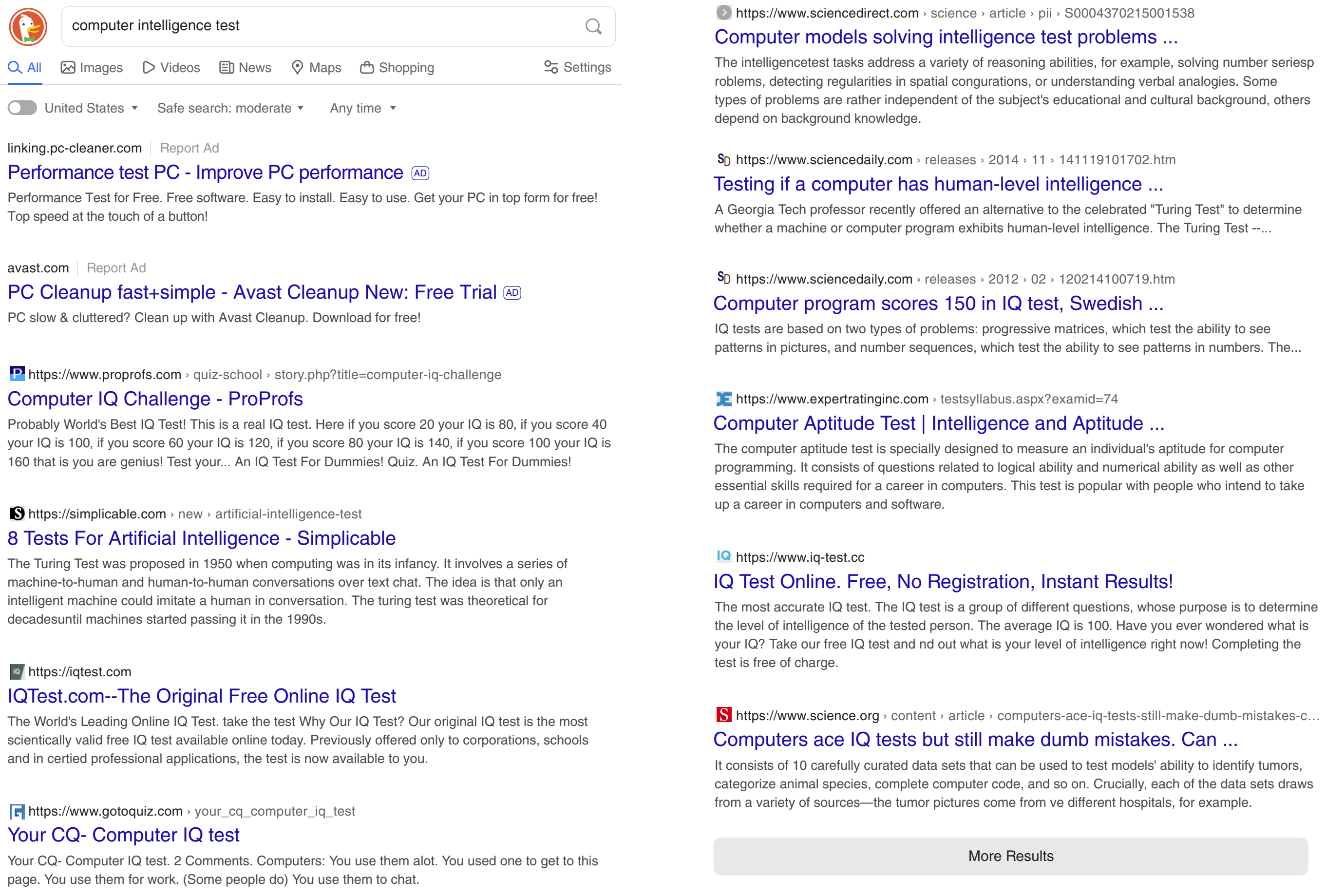
The user interface of a search engine allows the user to interact with the system. Different interfaces are designed for different usage scenarios.

Key components:

- Query Language
- Result Presentation

User Interface

Example



computer intelligence test

linking.pc-cleaner.com | Report Ad

Performance test PC - Improve PC performance AD

Performance Test for Free. Free software. Easy to install. Easy to use. Get your PC in top form for free! Top speed at the touch of a button!

avast.com | Report Ad

PC Cleanup fast+simple - Avast Cleanup New: Free Trial AD

PC slow & cluttered? Clean up with Avast Cleanup. Download for free!

https://www.proprofs.com › quiz-school › story.php?title=computer-iq-challenge

Computer IQ Challenge - ProProfs

Probably World's Best IQ Test! This is a real IQ test. Here if you score 20 your IQ is 80, if you score 40 your IQ is 100, if you score 60 your IQ is 120, if you score 80 your IQ is 140, if you score 100 your IQ is 160 that is you are genius! Test your... An IQ Test For Dummies! Quiz. An IQ Test For Dummies!

https://simplicable.com › new › artificial-intelligence-test

8 Tests For Artificial Intelligence - Simplicable

The Turing Test was proposed in 1950 when computing was in its infancy. It involves a series of machine-to-human and human-to-human conversations over text chat. The idea is that only an intelligent machine could imitate a human in conversation. The turing test was theoretical for decadesuntil machines started passing it in the 1990s.

https://iqtest.com

IQTest.com--The Original Free Online IQ Test

The World's Leading Online IQ Test. take the test Why Our IQ Test? Our original IQ test is the most scientifically valid free IQ test available online today. Previously offered only to corporations, schools and in certified professional applications, the test is now available to you.

https://www.gotoquiz.com › your_cq_computer_iq_test

Your CQ- Computer IQ test

Your CQ- Computer IQ test. 2 Comments. Computers: You use them alot. You used one to get to this page. You use them for work. (Some people do) You use them to chat.

https://www.sciencedirect.com › science › article › pii › S0004370215001538

Computer models solving intelligence test problems ...

The intelligencetest tasks address a variety of reasoning abilities, for example, solving number seriesp roblems, detecting regularities in spatial configurations, or understanding verbal analogies. Some types of problems are rather independent of the subject's educational and cultural background, others depend on background knowledge.

https://www.sciencedaily.com › releases › 2014 › 11 › 141119101702.htm

Testing if a computer has human-level intelligence ...

A Georgia Tech professor recently offered an alternative to the celebrated "Turing Test" to determine whether a machine or computer program exhibits human-level intelligence. The Turing Test --...

https://www.sciencedaily.com › releases › 2012 › 02 › 120214100719.htm

Computer program scores 150 in IQ test, Swedish ...

IQ tests are based on two types of problems: progressive matrices, which test the ability to see patterns in pictures, and number sequences, which test the ability to see patterns in numbers. The...

https://www.expertratinginc.com › testsyllabus.aspx?examid=74

Computer Aptitude Test | Intelligence and Aptitude ...

The computer aptitude test is specially designed to measure an individual's aptitude for computer programming. It consists of questions related to logical ability and numerical ability as well as other essential skills required for a career in computers. This test is popular with people who intend to take up a career in computers and software.

https://www.iq-test.cc

IQ Test Online. Free, No Registration, Instant Results!

The most accurate IQ test. The IQ test is a group of different questions, whose purpose is to determine the level of intelligence of the tested person. The average IQ is 100. Have you ever wondered what is your IQ? Take our free IQ test and find out what is your level of intelligence right now! Completing the test is free of charge.

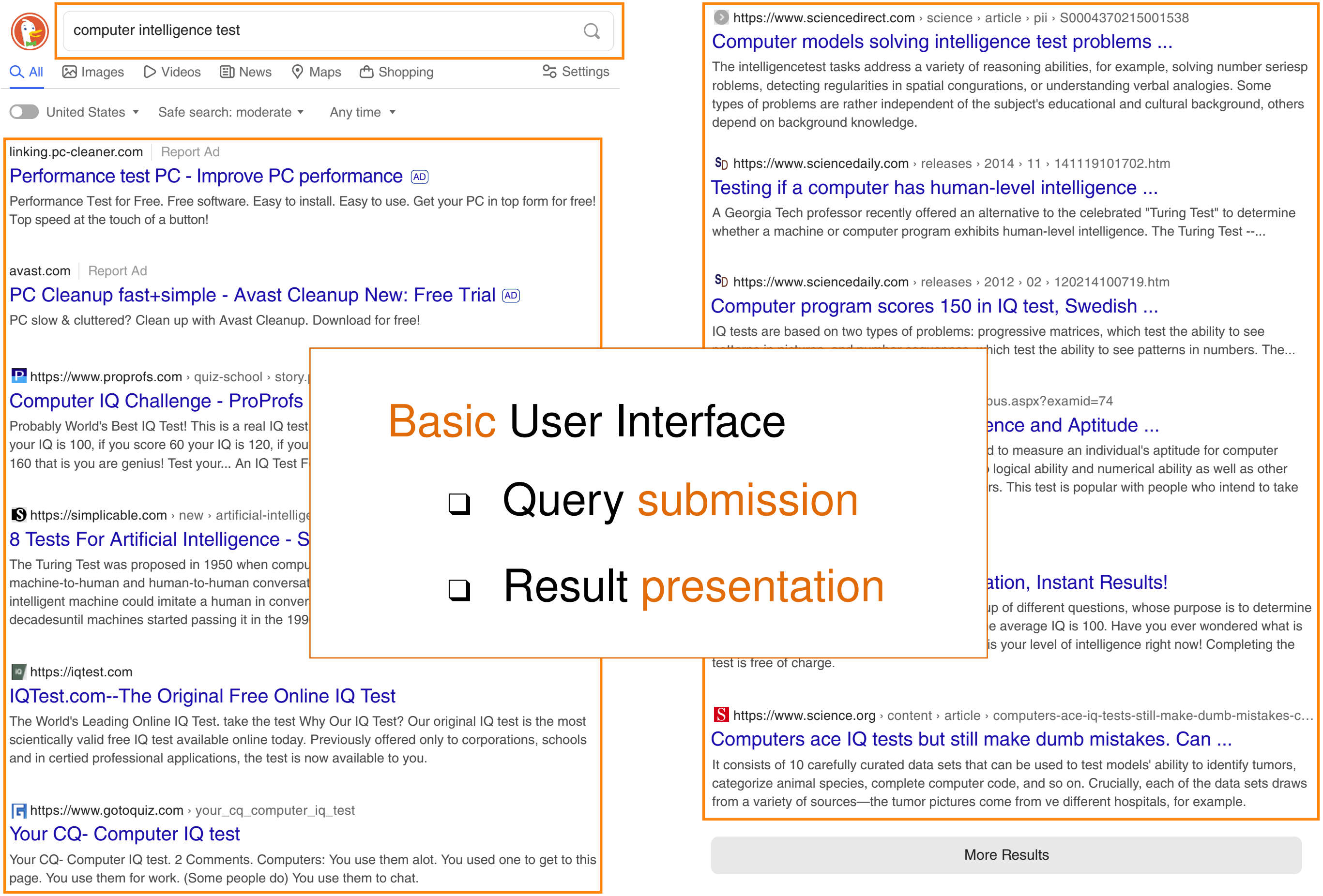
https://www.science.org › content › article › computers-ace-iq-tests-still-make-dumb-mistakes-c...

Computers ace IQ tests but still make dumb mistakes. Can ...

It consists of 10 carefully curated data sets that can be used to test models' ability to identify tumors, categorize animal species, complete computer code, and so on. Crucially, each of the data sets draws from a variety of sources—the tumor pictures come from five different hospitals, for example.

User Interface

Example



computer intelligence test

linking.pc-cleaner.com | Report Ad

Performance test PC - Improve PC performance AD

Performance Test for Free. Free software. Easy to install. Easy to use. Get your PC in top form for free! Top speed at the touch of a button!

avast.com | Report Ad

PC Cleanup fast+simple - Avast Cleanup New: Free Trial AD

PC slow & cluttered? Clean up with Avast Cleanup. Download for free!

https://www.proprofs.com › quiz-school › story.

Computer IQ Challenge - ProProfs

Probably World's Best IQ Test! This is a real IQ test
your IQ is 100, if you score 60 your IQ is 120, if you
160 that is you are genius! Test your... An IQ Test F

https://simplicable.com › new › artificial-intellige

8 Tests For Artificial Intelligence - S

The Turing Test was proposed in 1950 when compu
machine-to-human and human-to-human conversat
intelligent machine could imitate a human in conver
decadesuntil machines started passing it in the 199

https://iqtest.com

IQTest.com--The Original Free Online IQ Test

The World's Leading Online IQ Test. take the test Why Our IQ Test? Our original IQ test is the most scientifically valid free IQ test available online today. Previously offered only to corporations, schools and in certified professional applications, the test is now available to you.

https://www.gotoquiz.com › your_cq_computer_iq_test

Your CQ- Computer IQ test

Your CQ- Computer IQ test. 2 Comments. Computers: You use them alot. You used one to get to this page. You use them for work. (Some people do) You use them to chat.

https://www.sciencedirect.com › science › article › pii › S0004370215001538

Computer models solving intelligence test problems ...

The intelligencetest tasks address a variety of reasoning abilities, for example, solving number seriesproblems, detecting regularities in spatial configurations, or understanding verbal analogies. Some types of problems are rather independent of the subject's educational and cultural background, others depend on background knowledge.

https://www.sciencedaily.com › releases › 2014 › 11 › 141119101702.htm

Testing if a computer has human-level intelligence ...

A Georgia Tech professor recently offered an alternative to the celebrated "Turing Test" to determine whether a machine or computer program exhibits human-level intelligence. The Turing Test --...

https://www.sciencedaily.com › releases › 2012 › 02 › 120214100719.htm

Computer program scores 150 in IQ test, Swedish ...

IQ tests are based on two types of problems: progressive matrices, which test the ability to see
hich test the ability to see patterns in numbers. The...

ous.aspx?examid=74

ence and Aptitude ...

d to measure an individual's aptitude for computer
logical ability and numerical ability as well as other
rs. This test is popular with people who intend to take

ation, Instant Results!

up of different questions, whose purpose is to determine
e average IQ is 100. Have you ever wondered what is
is your level of intelligence right now! Completing the
test is free of charge.

https://www.science.org › content › article › computers-ace-iq-tests-still-make-dumb-mistakes-c...

Computers ace IQ tests but still make dumb mistakes. Can ...

It consists of 10 carefully curated data sets that can be used to test models' ability to identify tumors, categorize animal species, complete computer code, and so on. Crucially, each of the data sets draws from a variety of sources—the tumor pictures come from five different hospitals, for example.

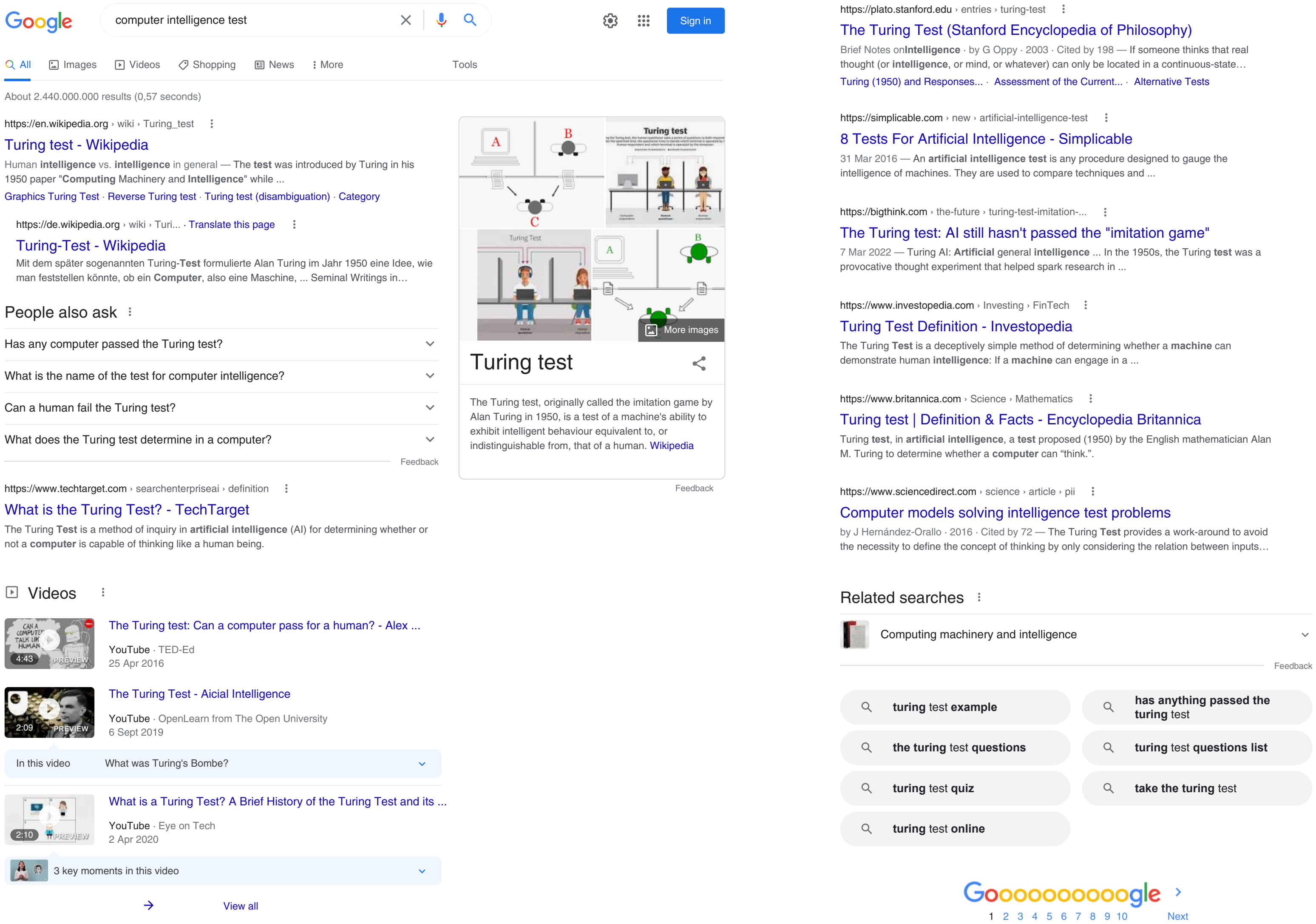
More Results

Basic User Interface

- ❑ Query submission
- ❑ Result presentation

User Interface

Example



computer intelligence test

All · Images · Videos · Shopping · News · More · Tools

About 2.440.000.000 results (0,57 seconds)

https://en.wikipedia.org › wiki › Turing_test
Turing test - Wikipedia
Human **intelligence** vs. **intelligence** in general — The **test** was introduced by Turing in his 1950 paper "**Computing** Machinery and **Intelligence**" while ...
Graphics Turing Test · Reverse Turing test · Turing test (disambiguation) · Category

https://de.wikipedia.org › wiki › Turi... · Translate this page
Turing-Test - Wikipedia
Mit dem später sogenannten Turing-**Test** formulierte Alan Turing im Jahr 1950 eine Idee, wie man feststellen könnte, ob ein **Computer**, also eine Maschine, ... Seminal Writings in…

People also ask
- Has any computer passed the Turing test?
- What is the name of the test for computer intelligence?
- Can a human fail the Turing test?
- What does the Turing test determine in a computer?

Feedback

https://www.techtarget.com › searchenterpriseai › definition
What is the Turing Test? - TechTarget
The Turing **Test** is a method of inquiry in **artificial intelligence** (AI) for determining whether or not a **computer** is capable of thinking like a human being.

Videos

The Turing test: Can a computer pass for a human? - Alex ...
YouTube · TED-Ed
25 Apr 2016
4:43 PREVIEW

The Turing Test - Aicial Intelligence
YouTube · OpenLearn from The Open University
6 Sept 2019
2:09 PREVIEW
In this video — What was Turing's Bombe?

What is a Turing Test? A Brief History of the Turing Test and its ...
YouTube · Eye on Tech
2 Apr 2020
2:10 PREVIEW
3 key moments in this video

View all

Turing test

The Turing test, originally called the imitation game by Alan Turing in 1950, is a test of a machine's ability to exhibit intelligent behaviour equivalent to, or indistinguishable from, that of a human. Wikipedia

Feedback

https://plato.stanford.edu › entries › turing-test
The Turing Test (Stanford Encyclopedia of Philosophy)
Brief Notes on**Intelligence** · by G Oppy · 2003 · Cited by 198 — If someone thinks that real thought (or **intelligence**, or mind, or whatever) can only be located in a continuous-state…
Turing (1950) and Responses... · Assessment of the Current... · Alternative Tests

https://simplicable.com › new › artificial-intelligence-test
8 Tests For Artificial Intelligence - Simplicable
31 Mar 2016 — An **artificial intelligence test** is any procedure designed to gauge the intelligence of machines. They are used to compare techniques and ...

https://bigthink.com › the-future › turing-test-imitation-...
The Turing test: AI still hasn't passed the "imitation game"
7 Mar 2022 — Turing AI: **Artificial** general **intelligence** ... In the 1950s, the Turing **test** was a provocative thought experiment that helped spark research in ...

https://www.investopedia.com › Investing › FinTech
Turing Test Definition - Investopedia
The Turing **Test** is a deceptively simple method of determining whether a **machine** can demonstrate human **intelligence**: If a **machine** can engage in a ...

https://www.britannica.com › Science › Mathematics
Turing test | Definition & Facts - Encyclopedia Britannica
Turing **test**, in **artificial intelligence**, a **test** proposed (1950) by the English mathematician Alan M. Turing to determine whether a **computer** can “think.”.

https://www.sciencedirect.com › science › article › pii
Computer models solving intelligence test problems
by J Hernández-Orallo · 2016 · Cited by 72 — The Turing **Test** provides a work-around to avoid the necessity to define the concept of thinking by only considering the relation between inputs…

Related searches
- Computing machinery and intelligence

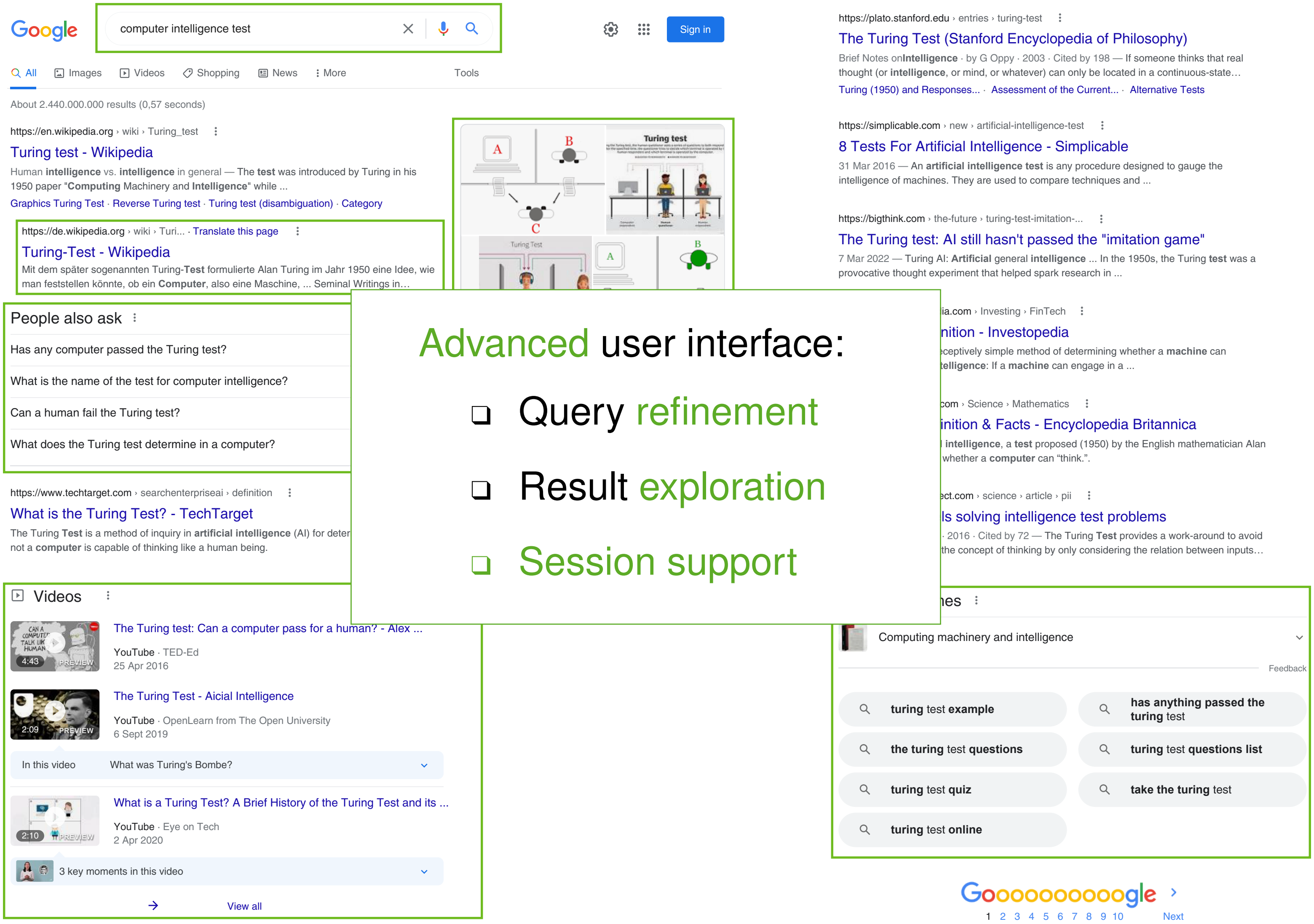
Feedback

- **turing** test **example**
- **has anything passed the turing** test
- **the turing** test **questions**
- **turing** test **questions list**
- **turing** test **quiz**
- **take the turing** test
- **turing** test **online**

Goooooooooogle
1 2 3 4 5 6 7 8 9 10 Next

User Interface

Example



About 2.440.000.000 results (0,57 seconds)

https://en.wikipedia.org › wiki › Turing_test

Turing test - Wikipedia

Human **intelligence** vs. **intelligence** in general — The **test** was introduced by Turing in his 1950 paper "**Computing** Machinery and **Intelligence**" while ...

Graphics Turing Test · Reverse Turing test · Turing test (disambiguation) · Category

https://de.wikipedia.org › wiki › Turi... · Translate this page

Turing-Test - Wikipedia

Mit dem später sogenannten Turing-**Test** formulierte Alan Turing im Jahr 1950 eine Idee, wie man feststellen könnte, ob ein **Computer**, also eine Maschine, ... Seminal Writings in...

People also ask

Has any computer passed the Turing test?

What is the name of the test for computer intelligence?

Can a human fail the Turing test?

What does the Turing test determine in a computer?

https://www.techtarget.com › searchenterpriseai › definition

What is the Turing Test? - TechTarget

The Turing **Test** is a method of inquiry in **artificial intelligence** (AI) for deter not a **computer** is capable of thinking like a human being.

Videos

The Turing test: Can a computer pass for a human? - Alex ...

YouTube · TED-Ed
25 Apr 2016

The Turing Test - Aicial Intelligence

YouTube · OpenLearn from The Open University
6 Sept 2019

In this video What was Turing's Bombe?

What is a Turing Test? A Brief History of the Turing Test and its ...

YouTube · Eye on Tech
2 Apr 2020

3 key moments in this video

View all

https://plato.stanford.edu › entries › turing-test

The Turing Test (Stanford Encyclopedia of Philosophy)

Brief Notes on**Intelligence** · by G Oppy · 2003 · Cited by 198 — If someone thinks that real thought (or **intelligence**, or mind, or whatever) can only be located in a continuous-state...

Turing (1950) and Responses... · Assessment of the Current... · Alternative Tests

https://simplicable.com › new › artificial-intelligence-test

8 Tests For Artificial Intelligence - Simplicable

31 Mar 2016 — An **artificial intelligence test** is any procedure designed to gauge the intelligence of machines. They are used to compare techniques and ...

https://bigthink.com › the-future › turing-test-imitation-...

The Turing test: AI still hasn't passed the "imitation game"

7 Mar 2022 — Turing AI: **Artificial** general **intelligence** ... In the 1950s, the Turing **test** was a provocative thought experiment that helped spark research in ...

ia.com › Investing › FinTech

nition - Investopedia

ceptively simple method of determining whether a **machine** can **telligence**: If a **machine** can engage in a ...

com › Science › Mathematics

inition & Facts - Encyclopedia Britannica

intelligence, a **test** proposed (1950) by the English mathematician Alan whether a **computer** can "think.".

ect.com › science › article › pii

ls solving intelligence test problems

2016 · Cited by 72 — The Turing **Test** provides a work-around to avoid the concept of thinking by only considering the relation between inputs...

hes

Computing machinery and intelligence

Feedback

turing test **example**

has anything passed the **turing** test

the turing test **questions**

turing test **questions list**

turing test **quiz**

take the turing test

turing test **online**

Advanced user interface:

- ❑ Query refinement
- ❑ Result exploration
- ❑ Session support

Remarks:

- Despite the obvious differences between the two exemplary search engines, neither is DuckDuckGo’s user interface extremely simple nor is Google’s extremely advanced. On the one hand, DuckDuckGo implements features like search suggestions and the display of related searches. On the other hand, Google’s search interface offers little support for exploration nor for search sessions or search missions.
- Many different user interfaces for search engines have been proposed and developed. [Kules et al., 2008]

User Interface

Query Language

The query language defines the syntax and semantics of valid queries. It may include commands to influence the search, so-called query operators.

Common query types:

- Structured query
- Keyword query
- Question query
- Query by example

Common operators:

- Boolean operators (AND, OR, NOT (or –))
- Which Boolean operator is implicitly assumed in a keyword query (e.g., in web search)?
- What other operators do you know?

User Interface

Query Language

The query language defines the syntax and semantics of valid queries. It may include commands to influence the search, so-called query operators.

Common query types:

- Structured query
- Keyword query
- Question query
- Query by example

Common operators:

- Boolean operators (AND, OR, NOT (or –))
- Quotes / phrasal search (“phrase of text”)
- Field search (title, text, url)
- Wildcards (*, ~synonym)
- Site search (site:example.com)

The most basic form of a query language is the keyword search.

Only about 1% of web queries contain operators. [White and Morris 2007]
Web search engines cannot expect users to be experts of the query language.

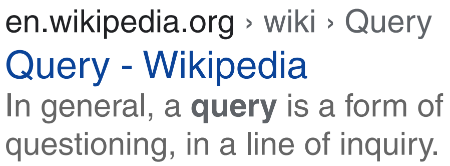
Domain-specific search engines often have specialized query languages, allowing for fine-grained control of retrieval behavior.

User Interface

Result Presentation

Search results are shown in ranking order on the search results page (SERP). Several additional processing steps are required to compile the page:

- ❑ Snippet generation
  Accesses the stored original web page and extracts sentences and phrases that summarize it, dependent on the query. The query's terms are highlighted in the snippet, including alternative inflections or synonyms.



en.wikipedia.org › wiki › Query

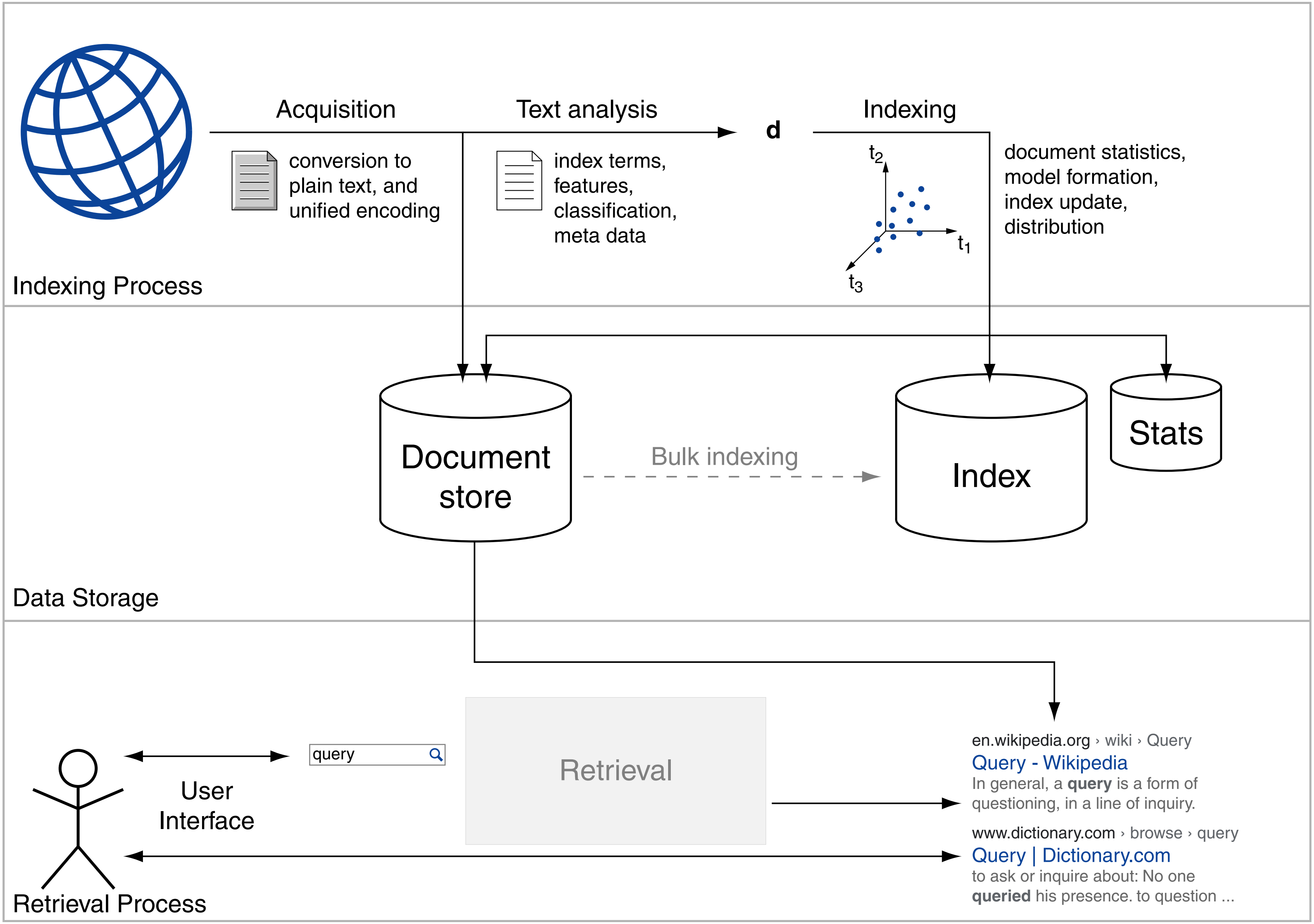
Query - Wikipedia

In general, a **query** is a form of questioning, in a line of inquiry.

- ❑ Universal search (e.g., oneboxes)
  Determines whether other specialized search engines can supply relevant results. Ranks the oneboxes into the SERP as per their importance compared to the organic web results.

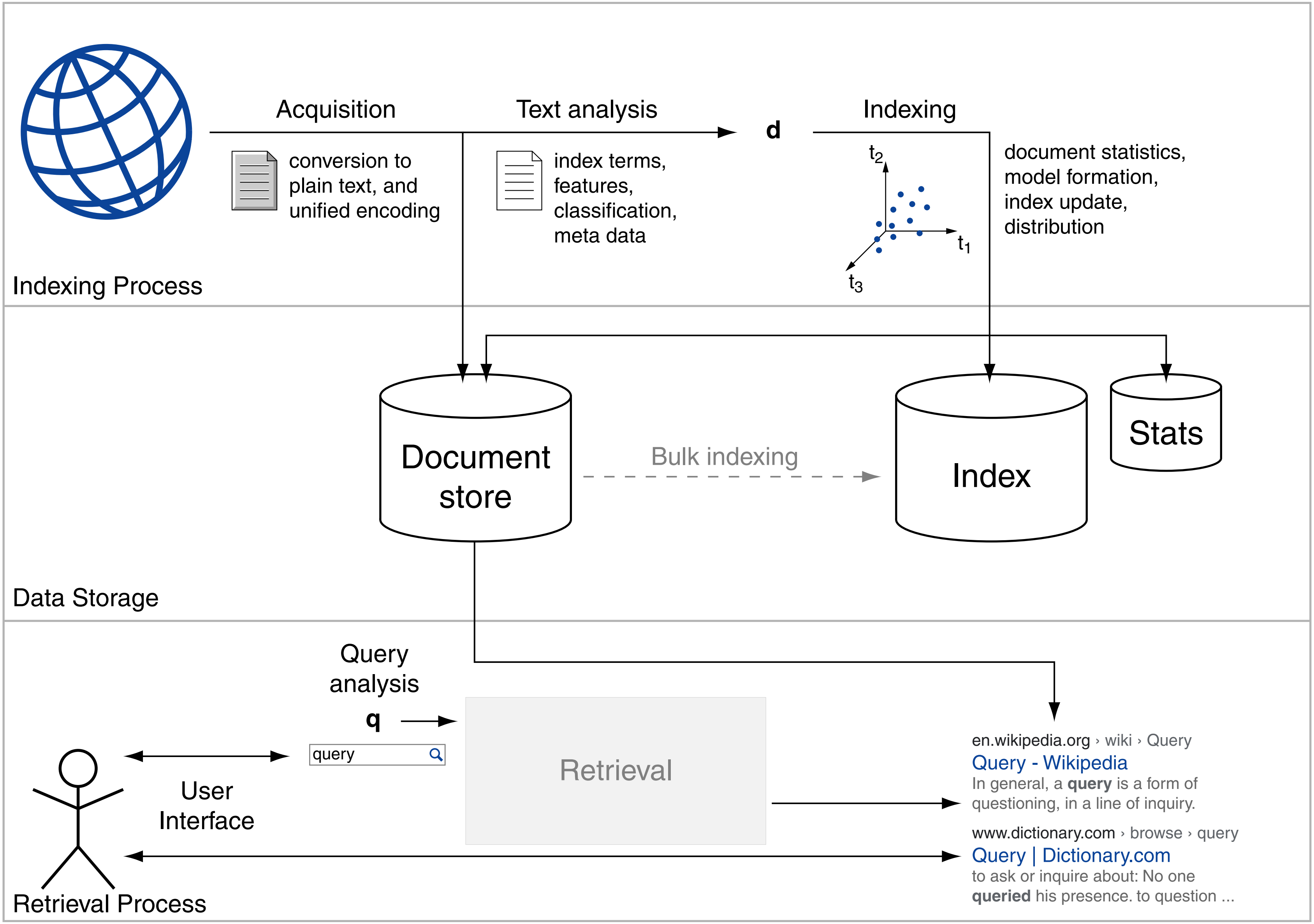
- ❑ Ad retrieval
  Accesses an internal ad search engine tailored to the retrieval of ads relevant to a query from all ads offered by advertisement partners.

- ❑ Facets / Categorization
  If metadata about documents is available, a side bar that allows to set constraints about the metadata can be displayed. Example: Shopping search engines. Alternatively, documents may be categorized using cluster analysis and cluster labeling.


Acquisition
Text analysis
d
Indexing
conversion to plain text, and unified encoding
index terms, features, classification, meta data
t_2
t_1
t_3
document statistics, model formation, index update, distribution
Indexing Process
Document store
Bulk indexing
Index
Stats
Data Storage
query
User Interface
Retrieval
Retrieval Process

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

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



Acquisition

conversion to plain text, and unified encoding

Text analysis

index terms, features, classification, meta data

d

Indexing

t_2

t_1

t_3

document statistics, model formation, index update, distribution

Indexing Process

Document store

Bulk indexing

Index

Stats

Data Storage

Query analysis

q

query

Retrieval

User Interface

en.wikipedia.org › wiki › Query

Query - Wikipedia

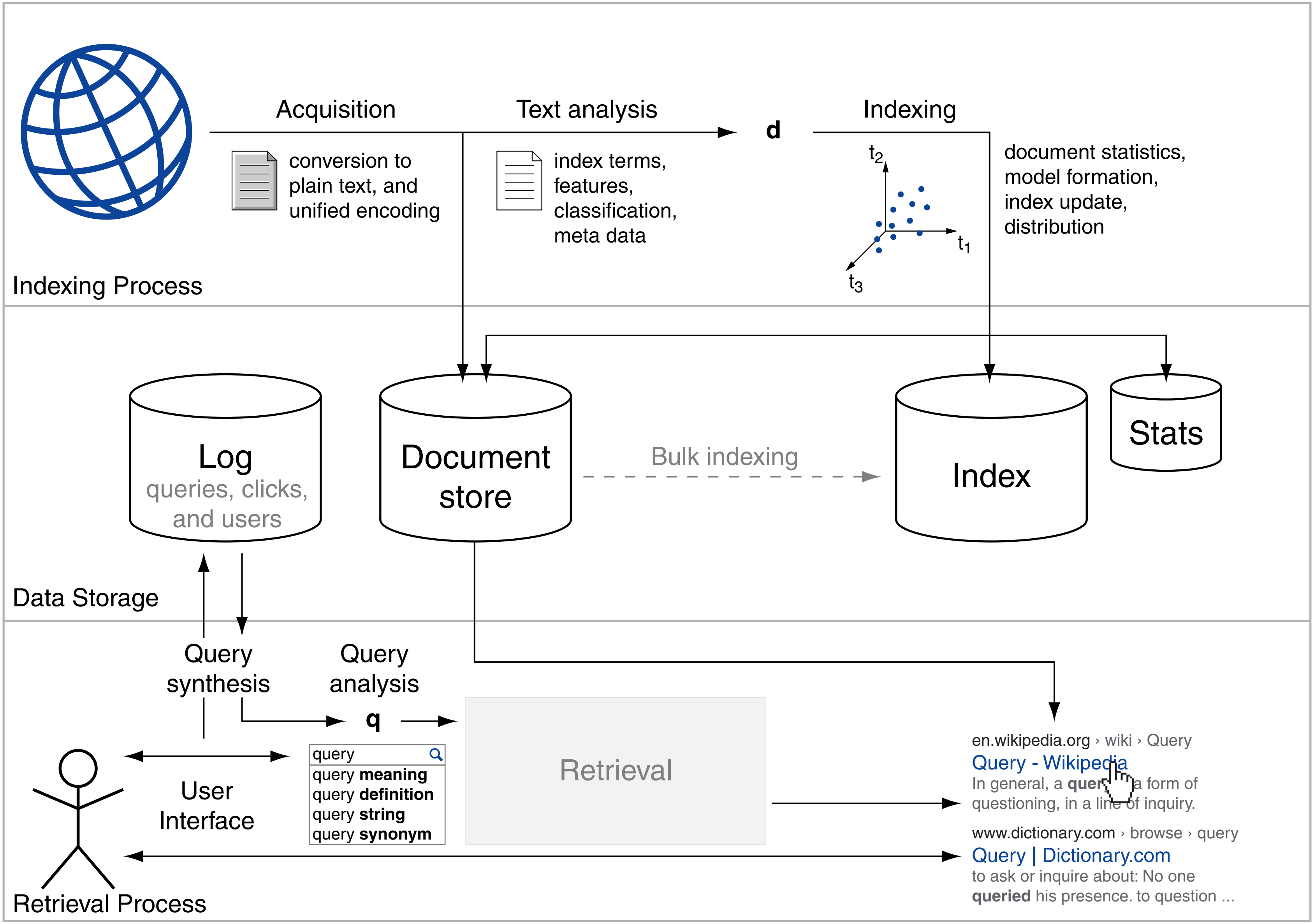
In general, a **query** is a form of questioning, in a line of inquiry.

www.dictionary.com › browse › query

Query | Dictionary.com

to ask or inquire about: No one **queried** his presence. to question ...

Retrieval Process



Indexing Process

Acquisition

conversion to plain text, and unified encoding

Text analysis

index terms, features, classification, meta data

d

Indexing

document statistics, model formation, index update, distribution

Data Storage

Log
queries, clicks, and users

Document store

Bulk indexing

Index

Stats

Retrieval Process

Query synthesis

Query analysis

q

User Interface

query
query **meaning**
query **definition**
query **string**
query **synonym**

Retrieval

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

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

User Interface

Query analysis (query understanding) maps the keywords of a query to the index terms of the terminology to enable retrieval.

Query synthesis generates a new query to represent the user's information need.

Key components:

- ❑ Text Analysis
- ❑ Logging
- ❑ Query Rewriting
- ❑ Query Expansion

Query Analysis and Synthesis

Text Analysis

Query analysis employs a text analysis pipeline similar to that used for documents:

- ❑ Tokenization

  Tokenization turns a query string into a sequence of tokens.

- ❑ Stopping

  Stopping, also stop word removal, discards a selection of tokens from the set of query terms.

- ❑ Stemming and lemmatization

  Stemming aims at reducing inflected index terms to a common stem. Lemmatization maps a word to its root form independent of its spelling.

- ❑ Named entity recognition

  Recognition of words or phrases that designate something in the “real” world.

Query Analysis and Synthesis

Logging

A search engine keeps logs of user interactions:

- Queries

  Every query submitted to the search engine.

- Result clicks

  Every click on a search result.

- Page interactions

  Data reflecting user behavior on search results pages.

- User tracking

  Association of all of the above data with individual users.

The logs are employed to improve retrieval by synthesizing a better query based on a user's query, and to analyze user experience.

Logs are the most valuable data a search engine collects.

Remarks:

- Further applications of the log include user experience analysis and optimization. The user behavior on the search engine's web pages allows for conclusions about its efficacy in supporting the user.

  Example: Google's optimization of its result link color. A designer, Jamie Divine, had picked out a blue that everyone on his team liked. But a product manager tested a different color with users and found they were more likely to click on the toolbar if it was painted a greener shade. As trivial as color choices might seem, clicks are a key part of Google's revenue stream, and anything that enhances clicks means more money. Mr. Divine's team resisted the greener hue, so Ms. Mayer split the difference by choosing a shade halfway between those of the two camps. Her decision was diplomatic, but it also amounted to relying on her gut rather than research. Since then, she said, she has asked her team to test the 41 gradations between the competing blues to see which ones consumers might prefer. [nytimes.com]

- In 2006, the internet company AOL released a large amount of user search requests to the public [Wikipedia]. Though thought to be sufficiently anonymized, this turned out to not be the case; a lot of personal information could be obtained from the log, and a number of users be personally identified just from their sequence of queries submitted. One particular user's story has been turned into an artistic movie called "I love Alaska" in which the sequence of queries over a long time spells out a tragic personal history.

Query Analysis and Synthesis

Query Rewriting

Query rewriting changes the query to improve it’s chances of retrieving relevant documents. Changes may be suggested to users, or made on-the-fly.

Replacement of terms:

- Query suggestion

  Feedback at various degrees of urgency, ranging from hints to replacement, dependent on confidence. Hints: suggestions while typing, related queries, “Did you mean . . . ” at Google. Replacement: “Showing results for . . . ” at Google.

- Spelling correction

  A language model predicts the probability of a query, and an error model that of an observed misspelling, given an intended query.

Addition of terms:

- Query expansion

  Inclusion of additional terms to a query. Query logs and term co-occurrences in documents are exploited here. Key approach: relevance feedback.

Query Analysis and Synthesis

Query Rewriting (continued)

Query rewriting changes the query to improve it’s chances of retrieving relevant documents. Changes may be suggested to users, or made on-the-fly.

Removal of terms:

- ❑ Query relaxation
  Removal or optionalization of query terms that appear to render a query overspecific, e.g., when too few documents are retrieved. Example: removal of modifiers in a noun phrase.

Specialization:

- ❑ Query segmentation
  Identification of (alternative) phrases and multi-term concepts in a query to enable diversified retrieval. Example query: new york times square dance
- ❑ Query scoping
  Focusing of (parts of) a query on specific fields of a document, e.g., its title, or body.
- ❑ Personalization
  User profiles allow for tailoring search results to the user’s context and interests.

Query Analysis and Synthesis

Query Expansion

The addition of terms to a query so as to retrieve more documents relevant to the user's information need.

Abbreviations:

- ❑ Dictionary-based

  Add long forms of abbreviations in a query found in a dictionary.
  Problem: ambiguity (e.g., “st.” → “saint” or “street”?).

- ❑ Machine learning-based

  Mining of abbreviations in context from document collections as ground truth. Recognition in queries, and addition of their long form dependent on query context.

Synonyms, hypernyms, hyponyms:

- ❑ Dictionary-based

  Add synonyms of query terms found in a dictionary or thesaurus.
  Problems: diversity, exactness (e.g., “computer” → “laptop”, “web” → “internet”).

- ❑ Machine learning-based

  Training of word embeddings based on large document collections, computation of word similarity, and usage of all words similar to a query's terms that exceed a similarity threshold.

Remarks:

- A possible mnemonic to keep in mind that there actually is a difference between the web (content: web pages, links, etc.) and the internet (infrastructure: protocols, cables, servers, routers, etc.) are photos from the W3C20 Anniversary Symposium 2014 on which the inventors Sir Tim Berners-Lee (web) and Vinton G. Cerf (internet) wear t-shirts that explain who did (not) invent what.

[W3C20 Anniversary Symposium]

Query Analysis and Synthesis

Query Expansion: Relevance Feedback

Relevance feedback refines a query in multiple retrieval runs:

1. Retrieval of documents using a given query.
2. Identification of relevant (irrelevant) documents among the top-ranked ones.
3. Extraction of terms related (unrelated) to the query's implied information need from the identified documents.
4. Addition of related terms to the query (weight decrease of unrelated terms). Unless a stopping criterion is met, continue at Step 1.

What are possible sources of relevance feedback?

Query Analysis and Synthesis

Query Expansion: Relevance Feedback

Relevance feedback refines a query in multiple retrieval runs:

1. Retrieval of documents using a given query.
2. Identification of relevant (irrelevant) documents among the top-ranked ones.
3. Extraction of terms related (unrelated) to the query's implied information need from the identified documents.
4. Addition of related terms to the query (weight decrease of unrelated terms). Unless a stopping criterion is met, continue at Step 1.

Sources of relevance feedback:

- ❑ Direct / explicit relevance feedback

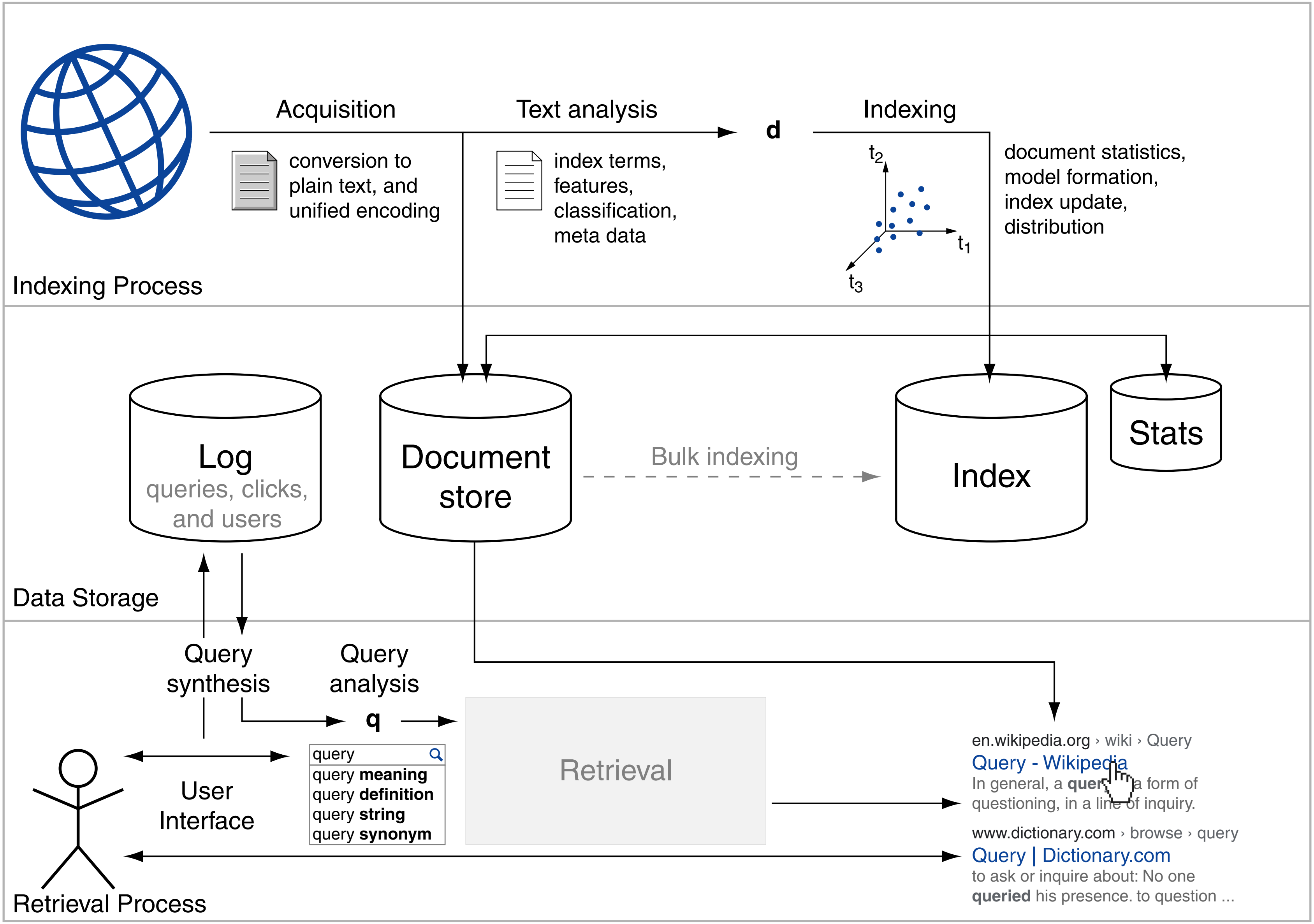
  A user marks retrieved documents as relevant or irrelevant.

- ❑ Indirect / implicit relevance feedback

  Relevant documents are identified by analyzing user behavior on the search results page. Relevance signals include: clicked results, dwell times, search abandonment.

- ❑ Blind / pseudo-relevance feedback

  The k top-ranked documents are considered relevant without checking.



Acquisition

conversion to plain text, and unified encoding

Text analysis

index terms, features, classification, meta data

d

Indexing

t_2 t_1 t_3

document statistics, model formation, index update, distribution

Indexing Process

Log
queries, clicks, and users

Document store

Bulk indexing

Index

Stats

Data Storage

Query synthesis

Query analysis

q

query
query **meaning**
query **definition**
query **string**
query **synonym**

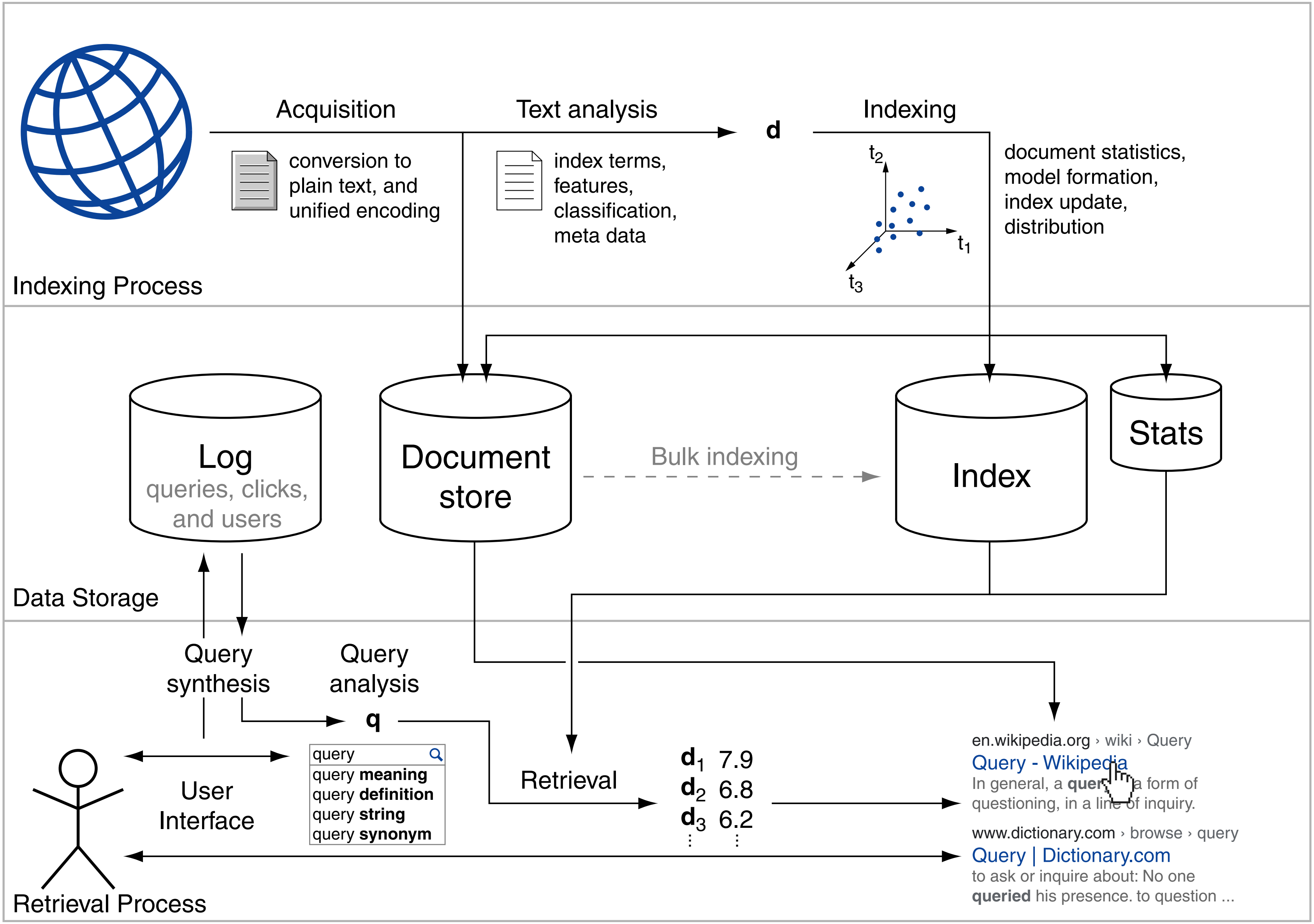
User Interface

Retrieval

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

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

Retrieval Process



Acquisition

conversion to plain text, and unified encoding

Text analysis

index terms, features, classification, meta data

d

Indexing

t_2

t_1

t_3

document statistics, model formation, index update, distribution

Indexing Process

Log

queries, clicks, and users

Document store

Bulk indexing

Index

Stats

Data Storage

Query synthesis

Query analysis

q

query
query **meaning**
query **definition**
query **string**
query **synonym**

User Interface

Retrieval

$\mathbf{d}_1$ 7.9
$\mathbf{d}_2$ 6.8
$\mathbf{d}_3$ 6.2
⋮ ⋮

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

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

Retrieval Process

Retrieval

Given a query representation, the ranking step scores and orders the documents indexed with respect to their relevance to the query.

This step marks the keystone of the implementation of the search engine's underlying retrieval model, a theory of how relevance can be quantified.

Retrieval models consist of

- a function to represent documents
- a function to represent queries
- a function to score a documents relevance to a query

Key components:

- Document Scoring
- Distribution

Retrieval

Document Scoring

Quantification of the relevance of an indexed document d to a query q.

Retrieval

Document Scoring

Quantification of the relevance of an indexed document d to a query q.

Let $t \in T$ denote a term t from the terminology T of index terms, and let $\omega_X : T \times X \rightarrow \mathbf{R}$ denote a term weighting function, where X may be sets of documents D or queries Q. Then the most basic relevance function ρ is:

$$\rho(q, d) = \sum_{t \in T} \omega_Q(t, q) \cdot \omega_D(t, d),$$

where $\omega_Q(t, q)$ and $\omega_D(t, d)$ are term weights indicating the importance of t for the query $q \in Q$ and the document $d \in D$, respectively.

Why are weights for all terms in T considered?
Why not just the weights of terms actually occuring in query q and document d?

Retrieval

Document Scoring

Quantification of the relevance of an indexed document d to a query q.

Let $t \in T$ denote a term t from the terminology T of index terms, and let $\omega_X : T \times X \to \mathbf{R}$ denote a term weighting function, where X may be sets of documents D or queries Q. Then the most basic relevance function ρ is:

$$\rho(q, d) = \sum_{t \in T} \omega_Q(t, q) \cdot \omega_D(t, d),$$

where $\omega_Q(t, q)$ and $\omega_D(t, d)$ are term weights indicating the importance of t for the query $q \in Q$ and the document $d \in D$, respectively.

Observations:

- A term t may have importance, and hence non-zero weights, for a query q or document d despite not occurring in them. Example: synonyms.
- The majority of terms from T will have insignificant importance to both.
- The term weights $\omega_D(t, d)$ can be pre-computed and indexed.
- The term weights $\omega_Q(t, q)$ must be computed on the fly.

Retrieval

Document Scoring

Document scoring requires index access. Access strategies determine index organization and what can be accomplished. Two strategies are widespread:

Document-at-a-time scoring

- ❑ Precondition: a total order of documents in the index's postlists is enforced
  Ordering criterion: document ID or document quality
- ❑ Parallel traversal of query term postlists, document ID by document ID.
- ❑ Each document's score is instantly complete, but the ranking only at the end.
- ❑ Concurrent disk IO overhead increases with query length.

Term-at-a-time scoring

- ❑ Iterative traversal of query term postlists (e.g., in order of term frequency).
- ❑ Temporary query postlist contains candidate documents.
- ❑ As document scores accumulate, an approximate ranking becomes available.
- ❑ More main memory required for maintaining temporary postlist.

Safe and unsafe optimizations exist (e.g., to stop the search early).

Retrieval

Distribution

The distribution of query processing depends on that of the index.

Query broker / load balancer

- ❑ Decides which shard and which replicated copies to access.
- ❑ Receives and merges results.

Cache

- ❑ Reduces latency by keeping frequently used data close at hand.
  Example: Lookup table in main memory
- ❑ Caches may include (smaller) indexes containing important documents only, precomputed search results, temporary postlists, or parts of postlists.
- ❑ Caching strategies optimize the usage of the caching hierarchy from operating system to hardware caches.

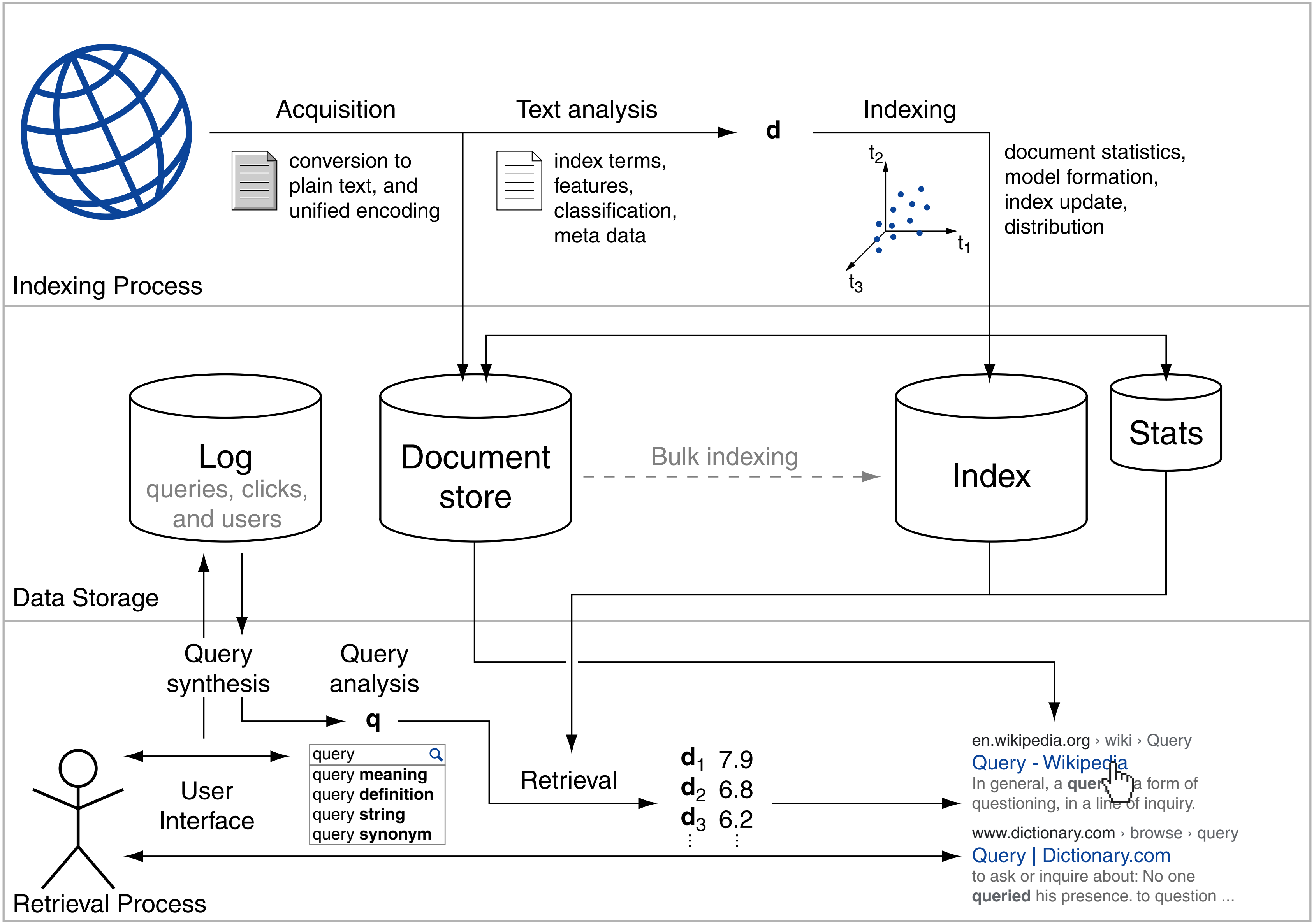
Chapter IR:II

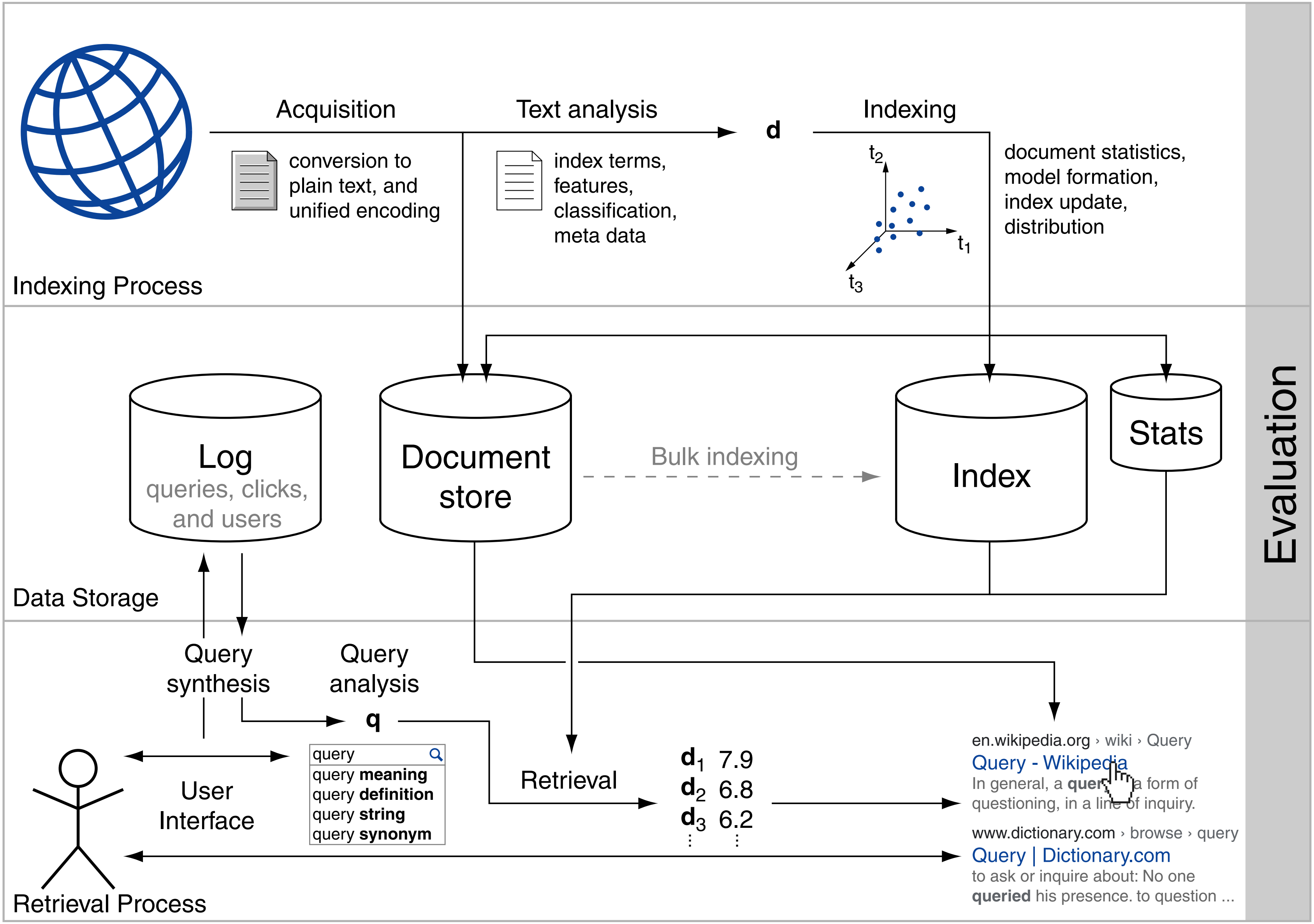
II. Architecture of a Search Engine

- ❑ Overview
- ❑ Acquisition
- ❑ Text Analysis
- ❑ Indexing

- ❑ User Interface
- ❑ Query Analysis and Synthesis
- ❑ Retrieval

- ❑ Evaluation


Acquisition
Text analysis
d
Indexing
conversion to plain text, and unified encoding
index terms, features, classification, meta data
t_2
t_1
t_3
document statistics, model formation, index update, distribution
Indexing Process
Log
queries, clicks, and users
Document store
Bulk indexing
Index
Stats
Data Storage
Query synthesis
Query analysis
q
query
query meaning
query definition
query string
query synonym
User Interface
Retrieval
d_1 7.9
d_2 6.8
d_3 6.2
⋮ ⋮
en.wikipedia.org › wiki › Query
Query - Wikipedia
In general, a query is a form of questioning, in a line of inquiry.
www.dictionary.com › browse › query
Query | Dictionary.com
to ask or inquire about: No one queried his presence. to question ...
Retrieval Process



Acquisition

conversion to plain text, and unified encoding

Text analysis

index terms, features, classification, meta data

d

Indexing

document statistics, model formation, index update, distribution

Indexing Process

Log
queries, clicks, and users

Document store

Bulk indexing

Index

Stats

Data Storage

Evaluation

Query synthesis

Query analysis

q

User Interface

query
query **meaning**
query **definition**
query **string**
query **synonym**

Retrieval

$\mathbf{d}_1$ 7.9
$\mathbf{d}_2$ 6.8
$\mathbf{d}_3$ 6.2
⋮ ⋮

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

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

Retrieval Process

Evaluation

Overview

Evaluation addresses the analysis of search effectiveness and efficiency.

Retrieval analysis

- ❑ Goals: relevant documents first, diversity, novelty, ...
- ❑ (Corwdsourced) acquisition of relevance judgments for query-document pairs
- ❑ Measurement theory (e.g., emphasis on top results is common in web search)
- ❑ Log analysis of search behavior
- ❑ User studies and A/B tests

User experience analysis

- ❑ Goals: usability, user satisfaction, ...
- ❑ Log analysis of user behavior
- ❑ User studies and A/B testing

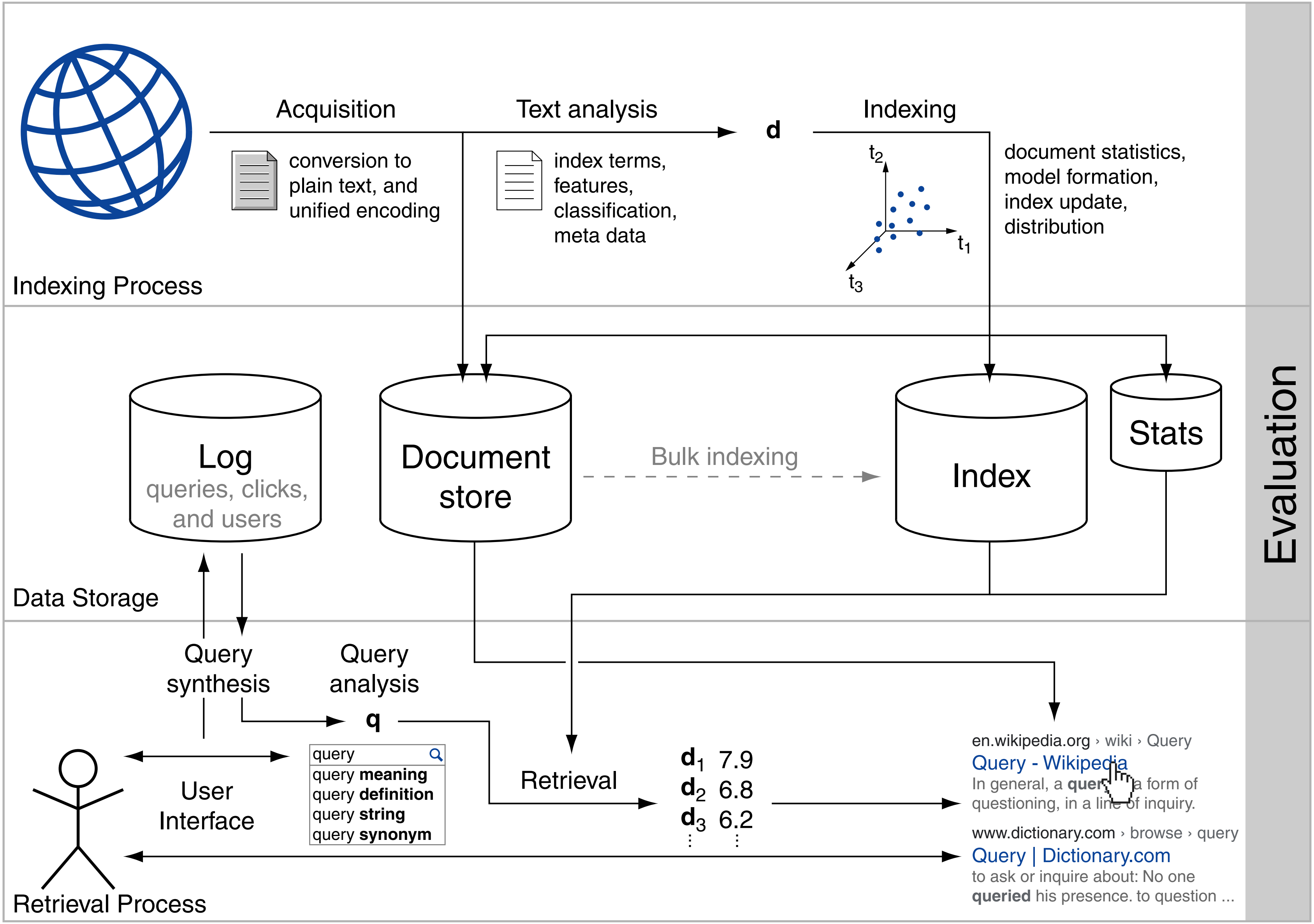
Runtime analysis

- ❑ Goals: throughput, latency, ...
- ❑ Log analysis of system behavior
- ❑ Lab experiments and simulation

Remarks:

- Evaluation is a systematic determination of a subject’s merit, worth, and significance, using criteria governed by a set of standards. It can assist to ascertain the degree of achievement or value in regard to the aim and objectives sought after. The primary purpose of evaluation, in addition to gaining insight into prior or existing initiatives, is to enable reflection and assist in the identification of future change. [Wikipedia]

Architecture of a Search Engine



Indexing Process

Acquisition

conversion to plain text, and unified encoding

Text analysis

index terms, features, classification, meta data

d

Indexing

t_2 t_1 t_3

document statistics, model formation, index update, distribution

Data Storage

Log

queries, clicks, and users

Document store

Bulk indexing

Index

Stats

Evaluation

Retrieval Process

Query synthesis

Query analysis

q

User Interface

query

query **meaning**

query **definition**

query **string**

query **synonym**

Retrieval

$\mathbf{d}_1$ 7.9

$\mathbf{d}_2$ 6.8

$\mathbf{d}_3$ 6.2

⋮ ⋮

en.wikipedia.org › wiki › Query

Query - Wikipedia

In general, a **query** is a form of questioning, in a line of inquiry.

www.dictionary.com › browse › query

Query | Dictionary.com

to ask or inquire about: No one **queried** his presence. to question ...