

Evaluation Overview

Jaime Arguello

INLS 509: Information Retrieval

jarguell@email.unc.edu

Information Retrieval

- Given a **query** and a **corpus**, find **relevant** documents

Query: user's textual description of their information need

Corpus: a repository of textual documents

Relevance: satisfaction of the user's information need

Information Retrieval

- In some disciplines, you can prove that solution **A** is better than **B** without experimentation.
- Is information retrieval one of these disciplines?
- If not, why not?

Uncertainty in Information Retrieval

Uncertainty in Information Retrieval

- The user may not know what they want
- Even if they do, they may not know how to describe it

Uncertainty in Information Retrieval

- ▶ soft surroundings
- ▶ trains interlocking dog sheets
- ▶ belly dancing music
- ▶ christian dior large bag
- ▶ best western airport sea tac
- ▶ www.bajawedding.com
- ▶ marie selby botanical gardens
- ▶ big chill down coats
- ▶ www.magichat.co.uk
- ▶ broadstone raquet club
- ▶ seadoo utopia
- ▶ seasons white plains condo
- ▶ priority.club.com
- ▶ aircat tools
- ▶ epicurus evil
- ▶ instructions
- ▶ hinds county city of jackson
- ▶ last searches on aol a to z

(AOL query-log)

Uncertainty in Information Retrieval

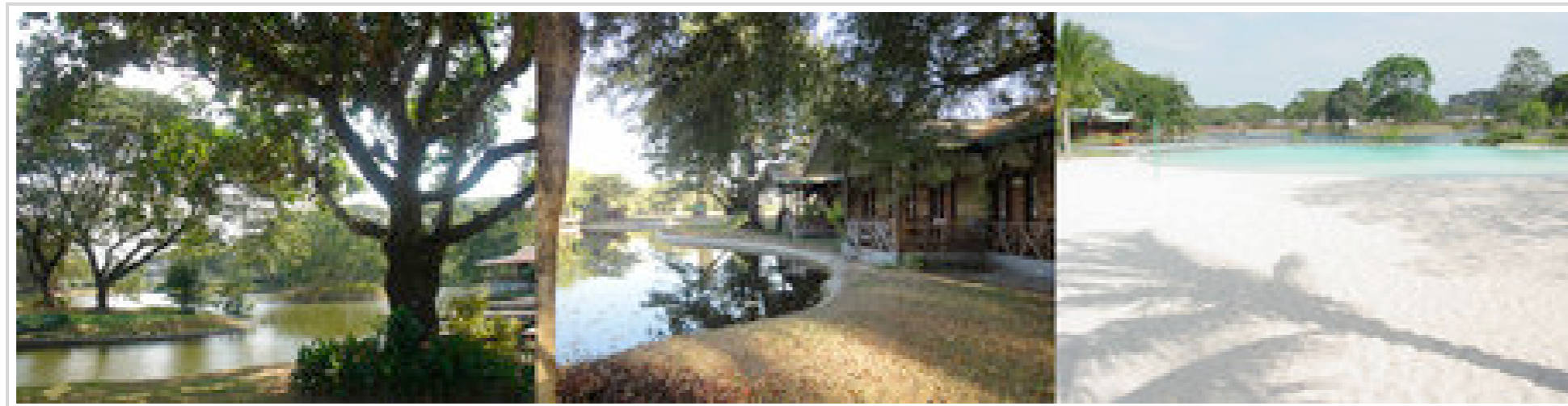
- Even if the user knows how to describe their information need, it will probably still be ambiguous
 - ▶ to the system
 - ▶ to a human trying to see if system **A** is better than **B**

Uncertainty in Information Retrieval

- **Query:** curbing population growth

Dengue in Philippines taken seriously at one hotel in Clark Philippines

Jul 18, 2010 23:58 EDT



Philippines government department of Health has issued warning of Dengue epidemic in July 2010 which is expected to last until the end of the year.

While most hotels and resorts are slow to take steps to address this health and safety issues, one hotel in Clark Philippines is already well on its way to implementing proactive measures to prevent the spread of Dengue by curbing population growth of mosquito.



Uncertainty in Information Retrieval

- Document relevance is subjective
 - ▶ many factors affect whether a document is relevant to a query
 - ▶ relevance is difficult to define in a way that generalizes across users, tasks, and environments

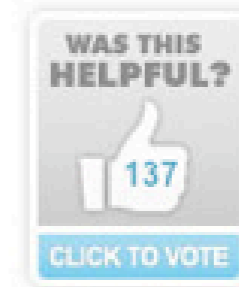
Uncertainty in Information Retrieval

- **Query:** cures for anatidaephobia

Anatidaephobia - The Fear That You are Being Watched by a Duck

December 08, 2008 by [Tammy Duffey](#)

[Single page](#) [Font Size](#) [Read comments \(44\)](#) [Share](#)



Popular searches: [YouTube](#) | [Rihanna](#) | [Tiger Woods](#) | [Search more](#)

What Is Anatidaephobia?

Anatidaephobia is defined as a pervasive, irrational fear that one is being watched by a duck. The anatidaephobic individual fears that no matter where they are or what they are doing, a duck watches.

Anatidaephobia is derived from the Greek word "anatidae", meaning ducks, geese or swans and "phobos" meaning fear.

An advertisement for Aflac. It features a white duck with a yellow beak on a blue background. The text reads: "Aflac can help attract and retain employees, at no direct cost to your company." Below this is the Aflac logo, which consists of the word "Aflac" in white with a yellow duck head integrated into the letter 'f'. Underneath the logo is the tagline "We've got you under our wing.™" and a yellow button with the text "Learn More Now".

What Causes Anatidaephobia?

As with all phobias, the person coping with Anatidaephobia has experienced a real-life trauma. For the anatidaephobic individual, this trauma most likely occurred during childhood.

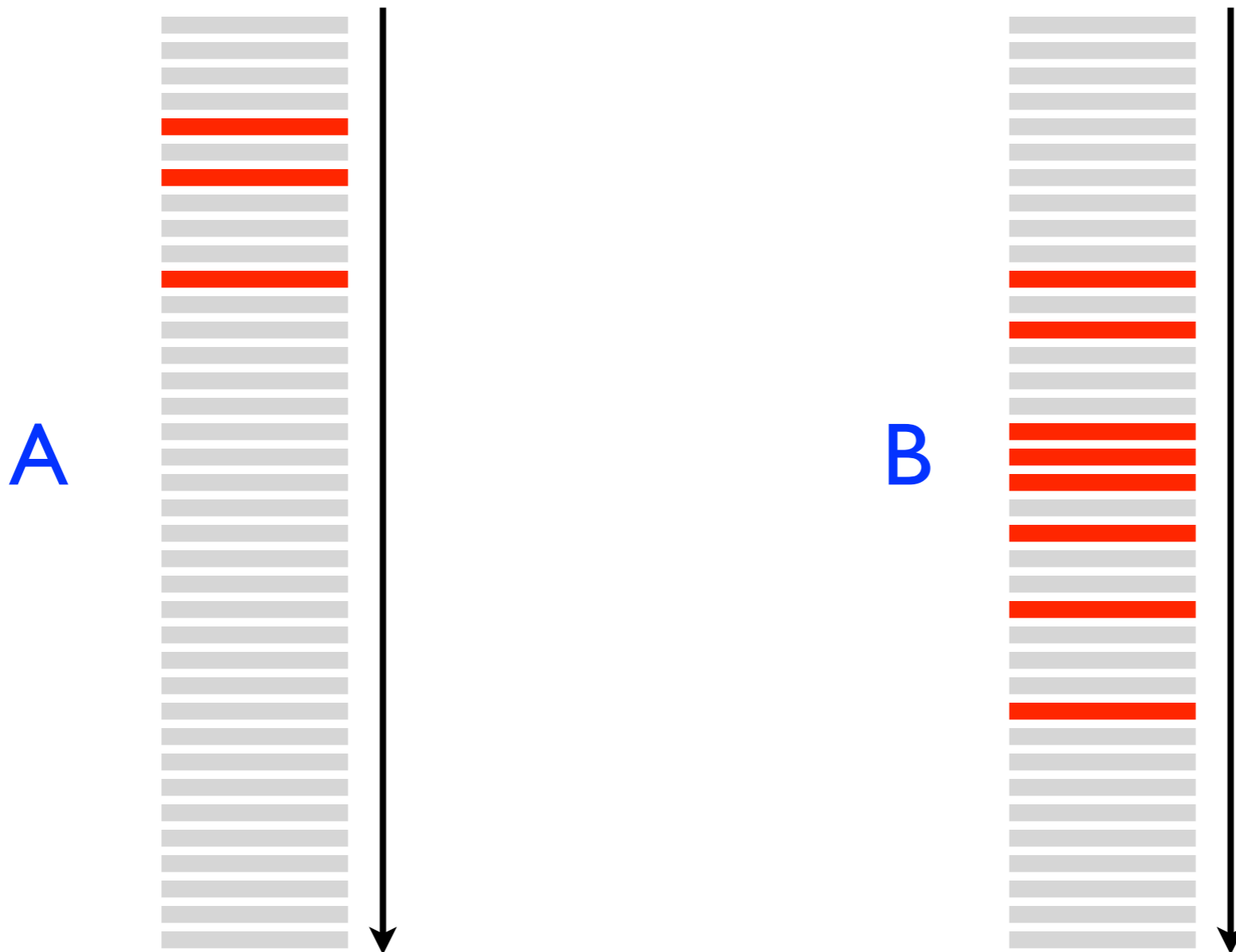
Perhaps the individual was intensely frightened by some species of water fowl. Geese and swans are relatively well known for their aggressive tendencies and perhaps the anatidaephobic person was actually bitten or flapped at. Of course, the Far Side comics did little to minimize the fear of being watched by a duck.

Uncertainty in Information Retrieval

- Even if we know which documents are relevant, the preferred ranking may be difficult to determine without understanding the user population and their objectives

Uncertainty in Information Retrieval

- Which ranking is better?



Information Retrieval

- In some disciplines, you can prove that solution **A** is better than **B** without experimentation.
- Is information retrieval one of these disciplines?
- No, it is not. There is too much uncertainty
- For now, the only way to show that **A** is better than **B**, is through extensive experimentation

Information Retrieval Evaluation

- Evaluation is a fundamental issue of information retrieval
 - ▶ an area of IR research in its own right
- Evaluation methods:
 - ▶ batch evaluation
 - ▶ user-study evaluation
 - ▶ online evaluation
- Each method has advantages and disadvantages

Batch Evaluation

overview

- Collect a set of queries (to test average performance)
- Construct a more complete description of the information being sought for each query

Batch Evaluation

overview: query + description (example)

- **QUERY:** pet therapy
- **DESCRIPTION:** Relevant documents must include details of how pet- and animal-assisted therapy is or has been used. Relevant details include information about pet therapy programs, descriptions of the circumstances in which pet therapy is used, the benefits of this type of therapy, the degree of success of this therapy, and any laws or regulations governing it.

Batch Evaluation

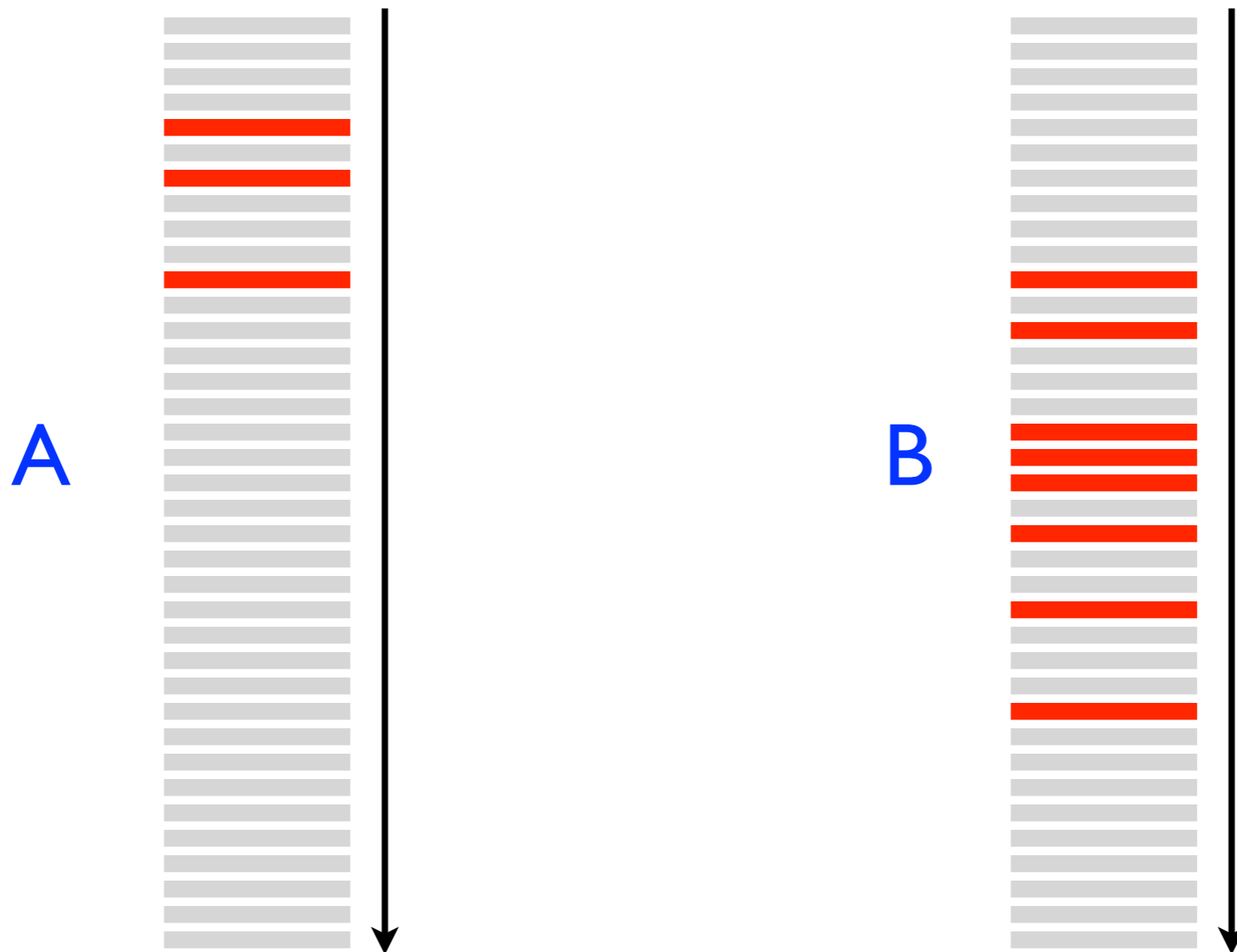
overview

- Using these descriptions, have human judges determine which documents are relevant for each query
- Evaluate systems based on their ability to retrieve the relevant documents for these queries
 - ▶ **evaluation metric**: a measurement that quantifies the quality of a particular ranking of results with known relevant/non-relevant documents

Batch Evaluation

overview: metrics

- Which ranking is better?

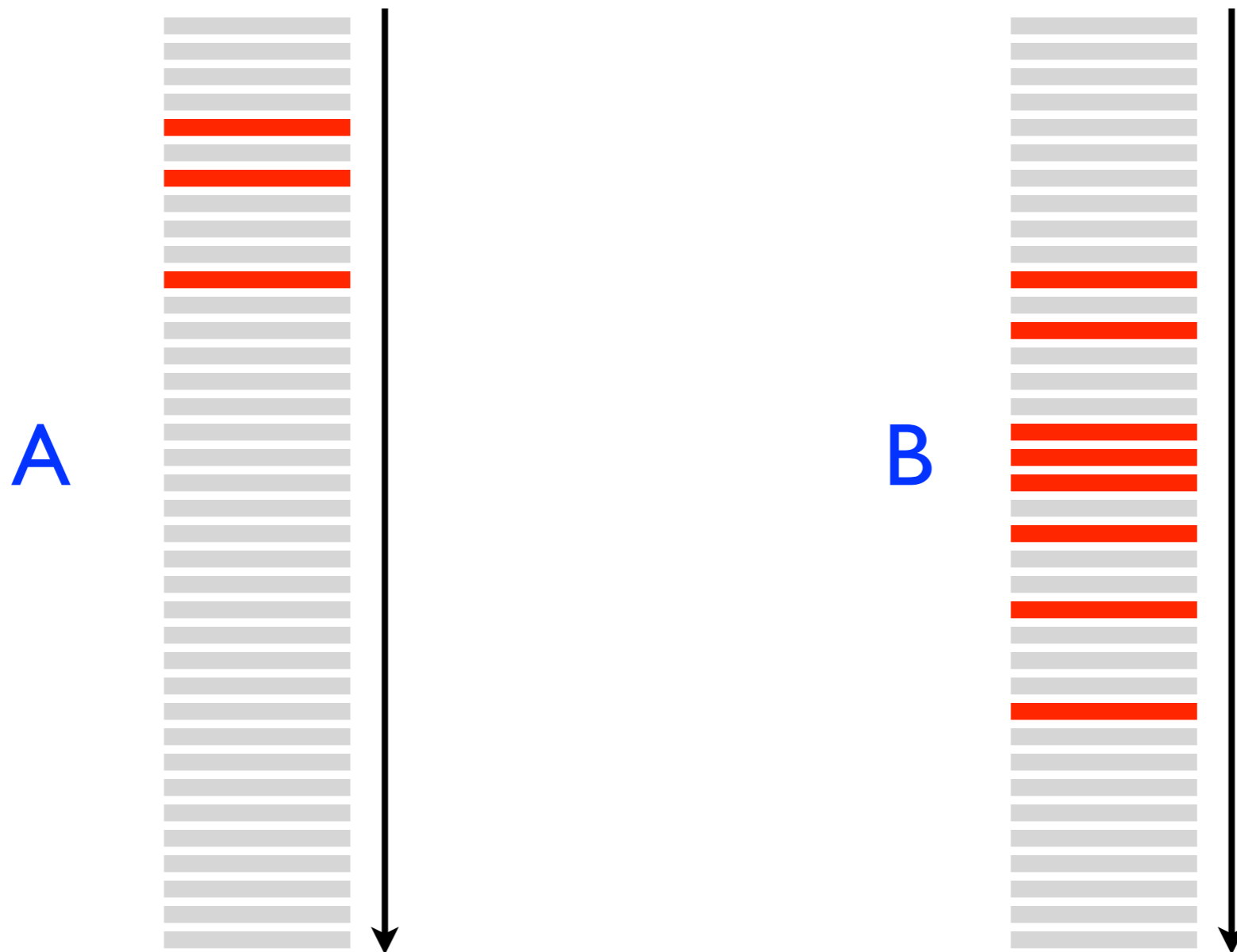


- rank of the first relevant document (lower value is better)

Batch Evaluation

overview: metrics

- Which ranking is better?

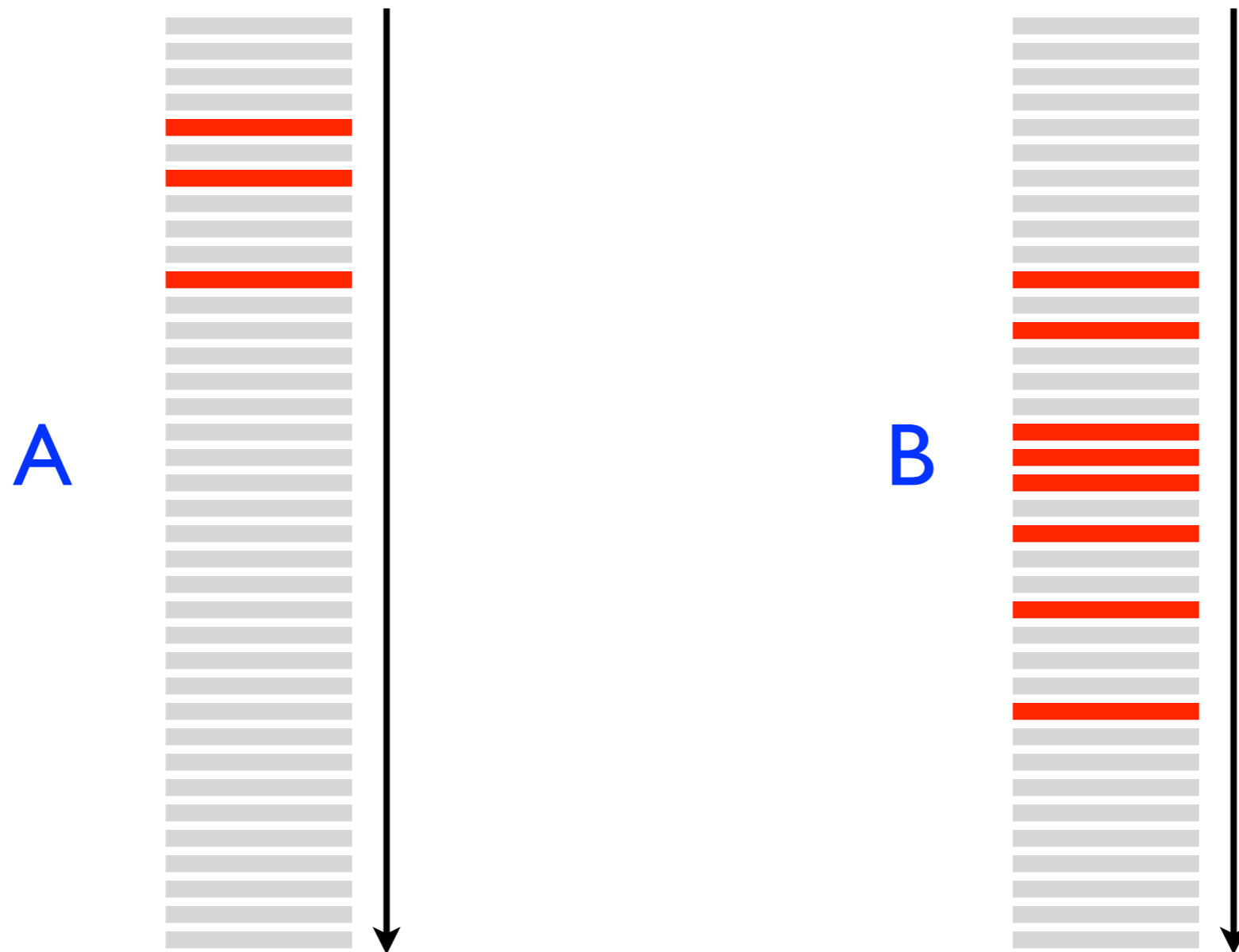


- precision at rank 10 (higher value is better)

Batch Evaluation

overview: metrics

- Which ranking is better?

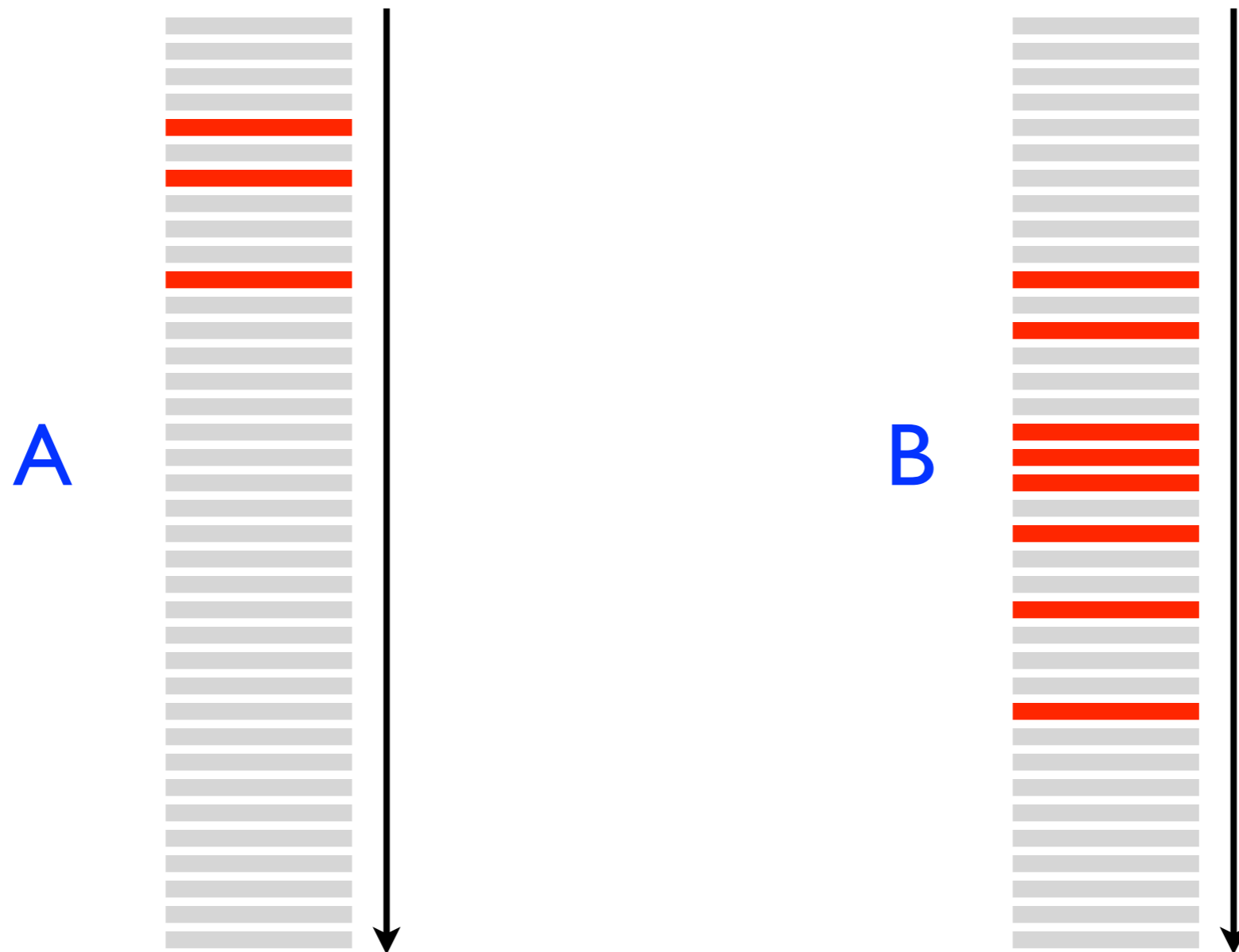


- precision at rank 1 (higher value is better)

Batch Evaluation

overview: metrics

- Which ranking is better?

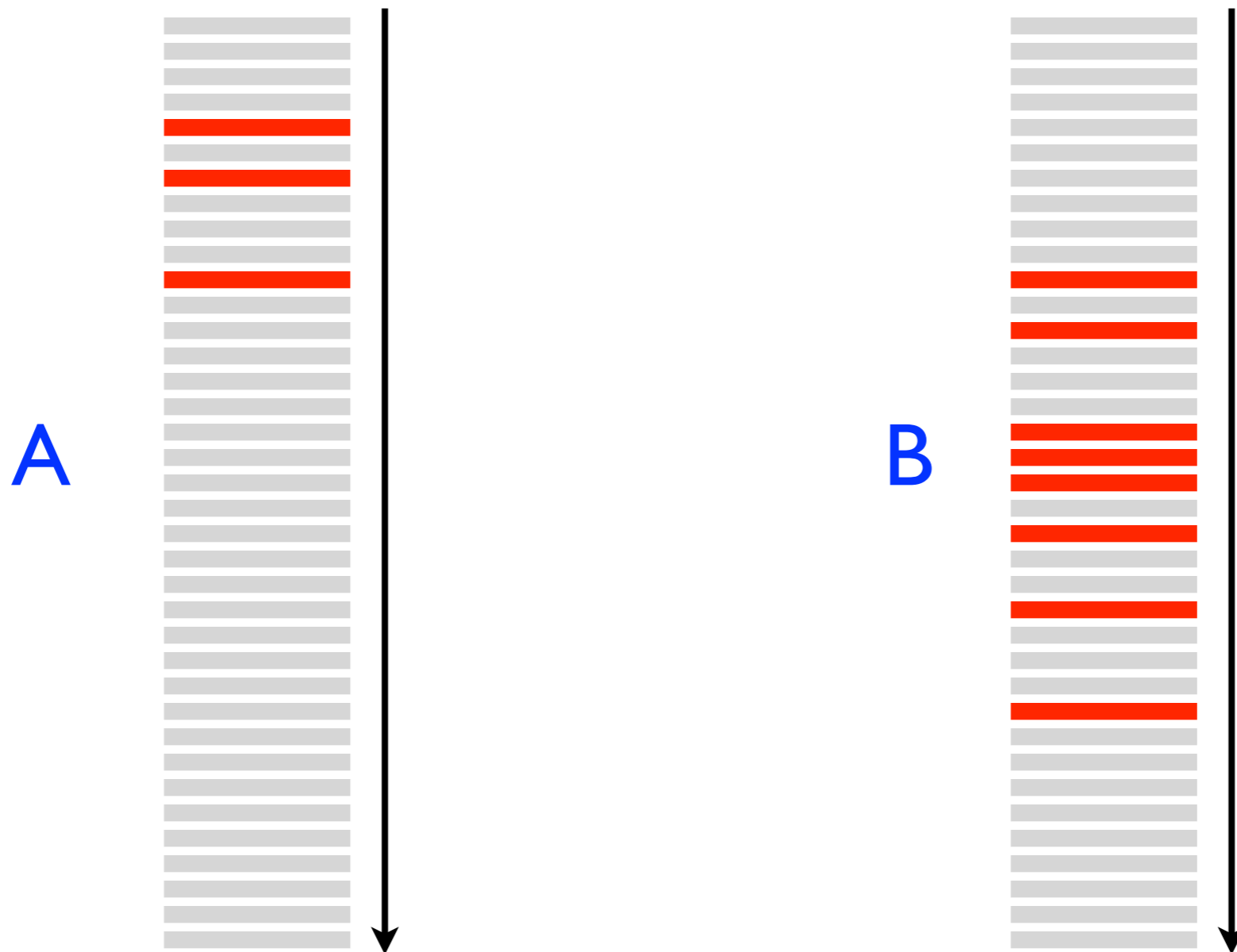


- recall at rank 10 (higher value is better)

Batch Evaluation

overview: metrics

- Which ranking is better?



- recall at rank 30 (higher value is better)

Batch Evaluation

overview: trade-offs

- Advantages:
 - ▶ inexpensive (once the test collection is constructed)
 - ▶ the experimental condition is fixed; same queries, and same relevance judgements
 - ▶ evaluations are reproducible; keeps us “honest”
 - ▶ by experimenting on the same set of queries and judgements, we can better understand how system **A** is better than **B**

Batch Evaluation

overview: trade-offs

- Disadvantages:
 - ▶ high initial cost. human assessors (the ones who judge documents relevant/non-relevant) are expensive
 - ▶ human assessors are not the users; judgements are made “out of context”
 - ▶ assumes that relevance is the same, independent of the user and the user’s context

Batch Evaluation

overview: trade-offs

- Many factors affect whether a document satisfies a particular user's information need
- Topicality, novelty, freshness, authority, formatting, reading level, assumed level of expertise
- **Topical relevance:** the document is on the same topic as the query
- **User relevance:** everything else
- Which kind of relevance does batch-evaluation address?

Batch Evaluation

overview: trade-offs

- Many factors affect whether a document satisfies a particular user's information need
- Topicality, novelty, freshness, authority, formatting, reading level, assumed level of expertise
- **Topical relevance:** the document is on the same topic as the query
- **User relevance:** everything else
- Which kind of relevance does batch-evaluation address?
- Mostly topical relevance

User-Study Evaluation

overview

- Provide a small set of users with several retrieval systems
- Ask them to complete several (potentially different) search tasks
- Learn about system performance by:
 - ▶ observing what they do
 - ▶ asking about their actions and thought processes
 - ▶ measuring success (task completion, time, etc.)
 - ▶ measuring perceived success (questionnaire data)

User-Study Evaluation

overview: trade-offs

- Advantages:
 - ▶ very detailed data about users' reaction to systems
 - ▶ in reality, a search is done to accomplish a higher-level task
 - ▶ in user studies, this task can be manipulated and studied
 - ▶ in other words, the experimental 'starting-point' need not be the query

User-Study Evaluation

overview: trade-offs


- Disadvantages:
 - ▶ user studies are expensive (pay users/subjects, scientist's time, data coding)
 - ▶ difficult to generalize from small studies to broad populations
 - ▶ the laboratory setting is not the user's normal environment
 - ▶ need to re-run experiment every time a new system is considered

On-line Evaluation

overview

- Given a search service with an existing user population (e.g., Google, Yahoo!, Bing) ...
- Have $x\%$ of query traffic use system **A** and $y\%$ of query-traffic use system **B**
- Compare system effects on logged user interactions (**implicit feedback**)
 - **clicks**: surrogates for perceived relevance (good)
 - **skips**: surrogates for perceived non-relevance (bad)

Implicit Feedback

implicit feedback in information retrieval  Advanced search

[PDF] [Implicit Feedback for Interactive Information Retrieval](#)

research.microsoft.com/en-us/um/people/ryenw/papers/thesis.pdf

File Format: PDF/Adobe Acrobat

by RW White - 2004 - Cited by 30 - Related articles

Implicit Feedback for Interactive. **Information Retrieval**. Ryen William White.

Department of Computing Science. Faculty of Computing Science, Mathematics and ...

click!

[PDF] [Context-Sensitive Information Retrieval Using Implicit Feedback](#)

citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.61.987...

File Format: PDF/Adobe Acrobat - Quick View

by X Shen - 2005 - Cited by 217 - Related articles

exploit **implicit feedback** information, including previous queries and clickthrough information, to improve retrieval accuracy in an in-teractive **information retrieval** ...

can we say that the first result is more relevant than the second?

[Context-Sensitive Information Retrieval Using Implicit Feedback](#)

citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.61.987

by X Shen - 2005 - Cited by 217 - Related articles

CiteSeerX - Document Details (Isaac Councill, Lee Giles): A major limitation ...

 Show more results from psu.edu

[Relevance feedback - Wikipedia, the free encyclopedia](#)

en.wikipedia.org/wiki/Relevance_feedback

Relevance feedback is a feature of some **information retrieval** systems. The idea ... 1


Explicit feedback; 2 **Implicit feedback**; 3 Blind feedback; 4 Using relevance ...

[A Search Engine that Learn from Implicit Feedback](#)

striver.joachims.org/

OSMOT - Learning **Retrieval** Functions from **Implicit Feedback**. ... Such observable behavior gives weak and noisy feedback **information** about which links the ...

Implicit Feedback

implicit feedback in information retrieval  Advanced search

[Implicit Feedback - Under the Reading Lamp](#)

bcao.wikidot.com/implicit-feedback

Xuehua Shen, Bin Tan, and ChengXiang Zhai, "Context-sensitive **information retrieval** using **implicit feedback**," in Proceedings of the 28th annual ...

skip!

[PDF] [Implicit Feedback for Interactive Information Retrieval](#)

research.microsoft.com/en-us/um/people/ryenw/papers/thesis.pdf

File Format: PDF/Adobe Acrobat

by RW White - 2004 - Cited by 30 - Related articles

Implicit Feedback for Interactive. **Information Retrieval**. Ryen William White.

Department of Computing Science. Faculty of Computing Science, Mathematics and ...

click!

[PDF] [Context-Sensitive Information Retrieval Using Implicit Feedback](#)

citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.61.987...

File Format: PDF/Adobe Acrobat - Quick View

by X Shen - 2005 - Cited by 217 - Related articles

exploit **implicit feedback** information, including previous queries and clickthrough information, to improve retrieval accuracy in an in- teractive **information retrieval** ...

can we say that the second result is more relevant than the first?

[Context-Sensitive Information Retrieval Using Implicit Feedback](#)

citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.61.987

by X Shen - 2005 - Cited by 217 - Related articles

CiteSeerX - Document Details (Isaac Council, Lee Giles): A major limitation ...

[+](#) Show more results from psu.edu

[Relevance feedback - Wikipedia, the free encyclopedia](#)

en.wikipedia.org/wiki/Relevance_feedback

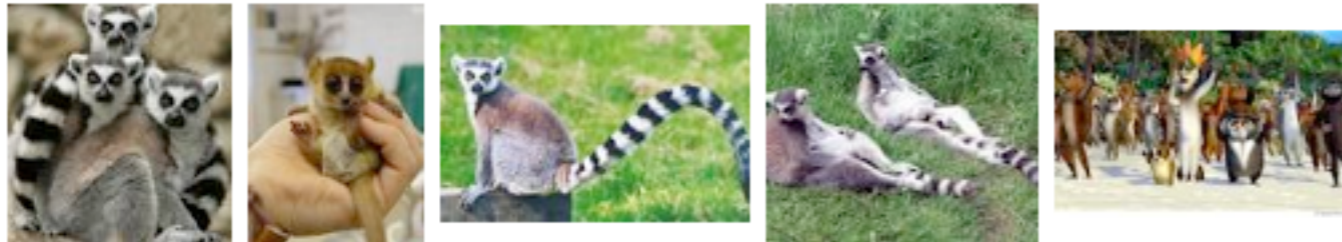
Relevance feedback is a feature of some **information retrieval** systems. The idea ... 1

Explicit feedback; 2 **Implicit feedback**; 3 Blind feedback; 4 Using relevance ...

Implicit Feedback

implicit feedback in information retrieval

[Images for lemurs](#) - Report images



click!

[PDF] [Implicit Feedback for Interactive Information Retrieval](#)

research.microsoft.com/en-us/um/people/ryenw/papers/thesis.pdf

File Format: PDF/Adobe Acrobat

by RW White - 2004 - Cited by 30 - Related articles

Implicit Feedback for Interactive. **Information Retrieval**. Ryen William White.

Department of Computing Science. Faculty of Computing Science, Mathematics and ...

[PDF] [Context-Sensitive Information Retrieval Using Implicit Feedback](#)

citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.61.987...

File Format: PDF/Adobe Acrobat - Quick View

by X Shen - 2005 - Cited by 217 - Related articles

exploit **implicit feedback** information, including previous queries and clickthrough information, to improve retrieval accuracy in an interactive **information retrieval** ...

[Context-Sensitive Information Retrieval Using Implicit Feedback](#)

citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.61.987

by X Shen - 2005 - Cited by 217 - Related articles

CiteSeerX - Document Details (Isaac Council, Lee Giles): A major limitation ...

[+ Show more results from psu.edu](#)

[Relevance feedback - Wikipedia, the free encyclopedia](#)

en.wikipedia.org/wiki/Relevance_feedback

Relevance feedback is a feature of some **information retrieval** systems. The idea ... 1

Explicit feedback; 2 **Implicit feedback**; 3 Blind feedback; 4 Using relevance ...

a click is a noisy
surrogate for
relevance!

Implicit Feedback

implicit feedback in information retrieval

[Images for lemurs](#) - Report images



[PDF] [Implicit Feedback for Interactive Information Retrieval](#)

research.microsoft.com/en-us/um/people/ryenw/papers/thesis.pdf

File Format: PDF/Adobe Acrobat

by RW White - 2004 - Cited by 30 - Related articles

Implicit Feedback for Interactive. **Information Retrieval**. Ryen William White.

Department of Computing Science. Faculty of Computing Science, Mathematics and ...

[PDF] [Context-Sensitive Information Retrieval Using Implicit Feedback](#)

citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.61.987...

File Format: PDF/Adobe Acrobat - Quick View

by X Shen - 2005 - Cited by 217 - Related articles

exploit **implicit feedback** information, including previous queries and clickthrough information, to improve retrieval accuracy in an in-teractive **information retrieval** ...

[Context-Sensitive Information Retrieval Using Implicit Feedback](#)

citeseer.ist.psu.edu/viewdoc/summary?doi=10.1.1.61.987

by X Shen - 2005 - Cited by 217 - Related articles

CiteSeerX - Document Details (Isaac Council, Lee Giles): A major limitation ...

[+ Show more results from psu.edu](#)

[Relevance feedback - Wikipedia, the free encyclopedia](#)

en.wikipedia.org/wiki/Relevance_feedback

Relevance feedback is a feature of some **information retrieval** systems. The idea ... 1

Explicit feedback; 2 **Implicit feedback**; 3 Blind feedback; 4 Using relevance ...

user sees the results and closes the browser

Implicit Feedback

capital of honduras

► [Tegucigalpa Honduras](#) [maps.google.com](#)



[Tegucigalpa - Wikipedia, the free encyclopedia](#)

en.wikipedia.org/wiki/Tegucigalpa

As **capital of Honduras**, as department head and as a municipality, the Central District seats ... For all practical purposes the **capital of Honduras** is Tegucigalpa. ...

[Etymology](#) - [History](#) - [Geography](#) - [Cityscape](#)

[Honduras - Wikipedia, the free encyclopedia](#)

en.wikipedia.org/wiki/Honduras

Comayagua was the **capital of Honduras** until 1880, when it was transferred to ...

[Geography of Honduras](#) - [List of cities in Honduras](#) - [Economy of Honduras](#) - [Colón](#)

[+](#) [Show more results from wikipedia.org](#)

[Honduras Facts and Figures, Honduras History, Political, Banking ...](#)

www.ca-bc.com/zip_internacional/about_honduras.html

Tegucigalpa, the **capital of Honduras**, got its tongue twisting name from the ancient Nahuatl language, and translated means "silver mountain" In effect, ...

[What is the capital city of Honduras](#)

[wiki.answers.com](#) > ... > [Countries States and Cities](#) > [Honduras](#)

What is the name of the **capital** city of **Honduras**? Tegucigalpa. What is the **capital** city ... **Honduras** capital city is Tegucigalpa. What city is the Italian **capital** city? ...

the absence of a
click is a noisy
surrogate for non-
relevance

On-Line Evaluation

overview: trade-offs

- Advantages:
 - ▶ system usage is naturalistic; users are situated in their natural context and often don't know that a test is being conducted
 - ▶ evaluation can include lots of users

On-Line Evaluation

overview: trade-offs

- Disadvantages:
 - ▶ requires a service with lots of users (enough of them to potential hurt performance for some)
 - ▶ this is often referred to as the “cold-start problem”
 - ▶ requires a good understanding on how different implicit feedback signals correlate with positive and negative user experiences
 - ▶ experiments are difficult to repeat

Information Retrieval Evaluation

- Evaluation is a fundamental issue of information retrieval
 - ▶ an area of IR research in its own right
- Evaluation methods:
 - ▶ batch evaluation
 - ▶ user-study evaluation
 - ▶ online evaluation
- Each method has advantages and disadvantages