

**800**  
A N N I  
1222 • 2022



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA



# Basic Concepts in Information Retrieval

## Search Engines

Master Degree in Computer Engineering

Master Degree in Data Science

Academic Year 2023/2024

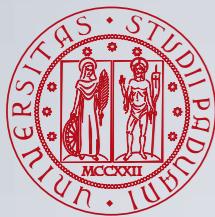


DIPARTIMENTO  
DI INGEGNERIA  
DELL'INFORMAZIONE

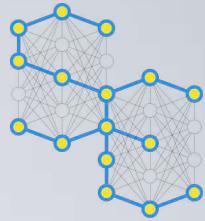
**Nicola Ferro**

Intelligent Interactive Information Access (IIIA) Hub  
Department of Information Engineering  
University of Padua





# Outline



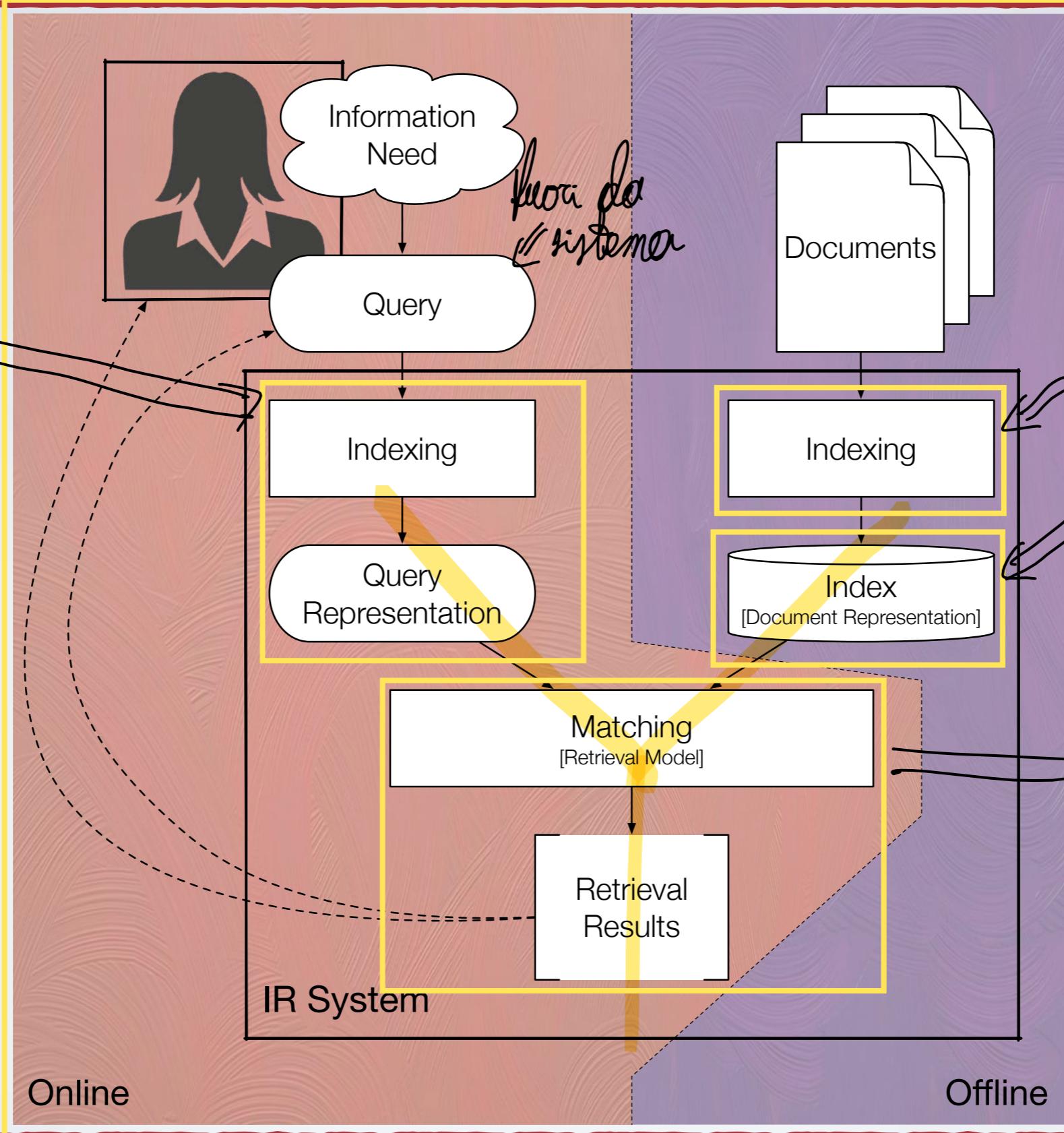
- Basics of Information Retrieval
- Basics of Experimental Evaluation

# Basics of Information Retrieval

# Information Retrieval: The Y

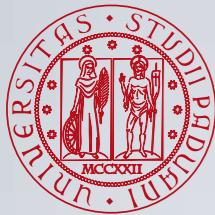
stesso processo  
ma solo per  
documenti  
(non serve  
salvarlo)

VECTOR-SPACE  
model. documenti  
e query come  
vettori in stesso  
spazio

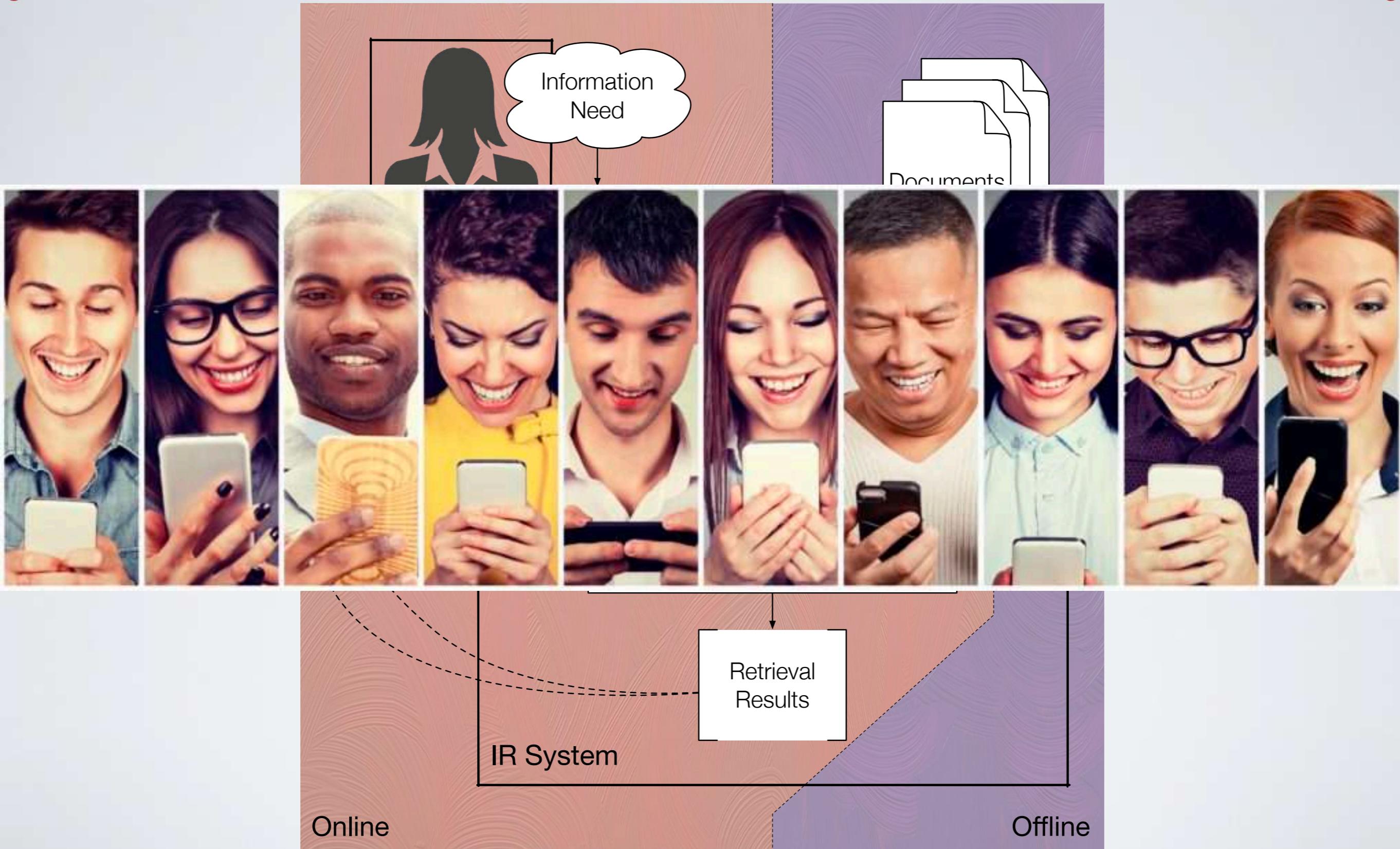
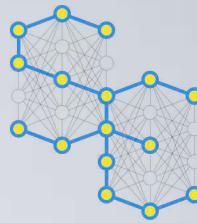


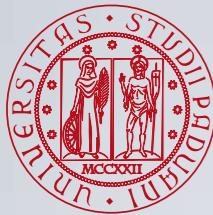
si parte da  
documenti;  
vogliamo  
rappresentarli  
per darli a  
computer  
salvare  
rappresentazione  
su disco  
(offline per user)

si possono usare  
vari tipi di  
apprendimento



# Information Retrieval: The Y





# From Information Need to Query



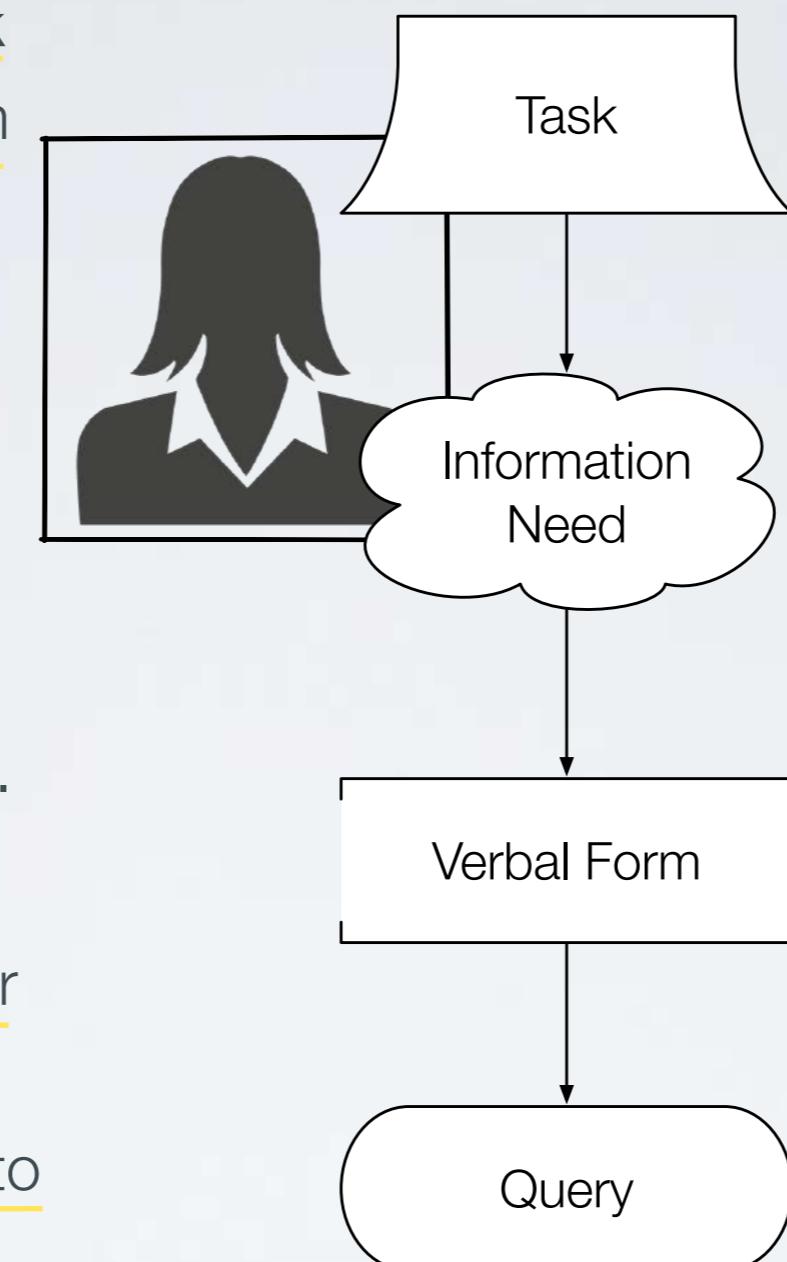
- The **information need** is the lack of knowledge needed to perform a task

- Information required to execute a task for the user's job
- Homepage of a company, government, institution

- The user translates her/his information need into a **query**, i.e. a set of keywords

- A document is **relevant** if the user perceives it as containing information of value with respect to her/his personal information need

- Because the information need is subjective, the notion of relevance is also **subjective**



fluo da zinema ↑

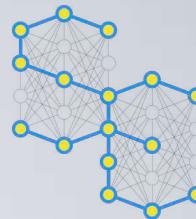
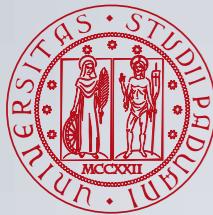
Get rid of mice in a (politically) correct way

Info about removing mice without killing them

How do I trap mice alive?

Google





# “Imprecision” in IR

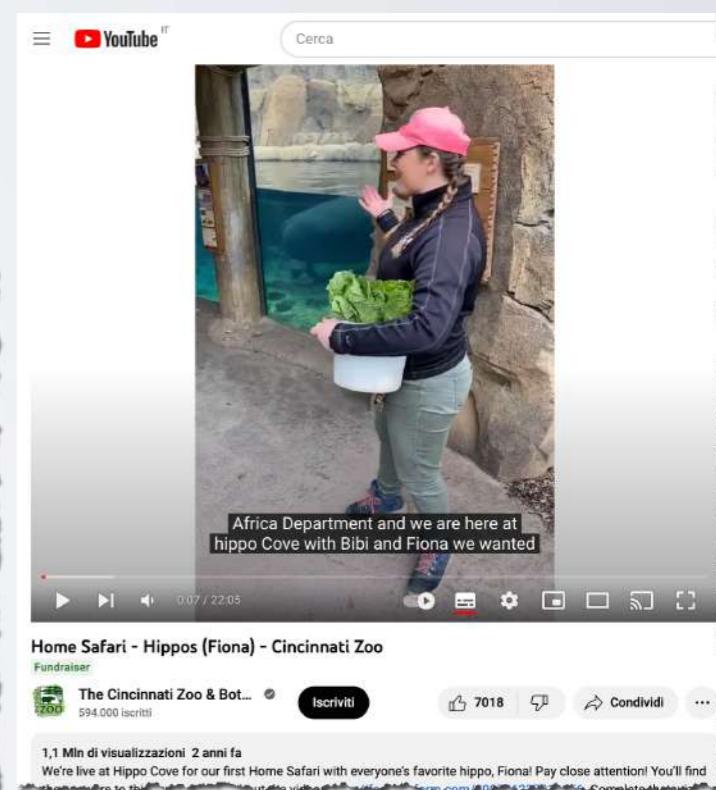
- Most algorithms in computer science have a “correct” answer
  - In contrast, a **heuristic** tries to guess something **close** to the “**correct**” answer
- Consider the following problems:
  - Sort a list of integers
  - Find the highest integer in a list of integers
  - Find beers made by X
- Now consider the following problem:
  - Find the document most relevant to the query “hippos in the zoo”



What Zoos Have Hippos in the US [Top 15 Zoos]  
February 26, 2022 by Team Animalia Facts

Hippos are one of those animals why we take our kids to a zoo. However, not everywhere you can see them. That's why we've made a shortlist about what zoos have hippos.

**Top 15 Zoos Have Hippos in the US**



Africa Department and we are here at hippo Cove with Bibi and Fiona we wanted

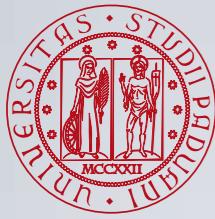
Home Safari - Hippos (Fiona) - Cincinnati Zoo  
Fundraiser

The Cincinnati Zoo & Bot... 594.000 Iscritti

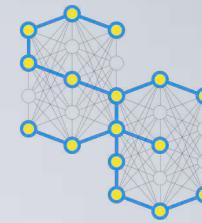
1,1 Min di visualizzazioni: 2 anni fa

We're live at Hippo Cove for our first Home Safari with everyone's favorite hippo, Fiona! Pay close attention! You'll find

- IR techniques are essentially heuristics because we do not know the “**correct**” answer



# What makes a document relevant?



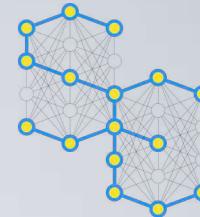
- It contains **any/all** query terms?
- It contains the query terms **many times**?
  - ...but what if it is a **long/short document**?
- It contains the query terms **close together**? *nego proxy per riconoscere entità*
- It contains terms **similar** to the query terms?
  - What does similar mean? E.g., synonym?

- It is **authoritative**
  - What does authoritative mean? E.g., has it many links?
- It doesn't contain too many **ads**?
- It doesn't contain too many different, unrelated words (e.g. **spam**)
- It has been **clicked** by many other people for the same query?
- It is **fresh/recent**?

## How do we rank documents? Which heuristics?



# (Basic) Indexing



- How to represent documents to ease the search for relevant information?

altra rappresentazione elaborabile

## Document Preprocessing

- Lexical Analysis / Tokenization
- Elimination of stopwords
- Stemming

## Index Creation

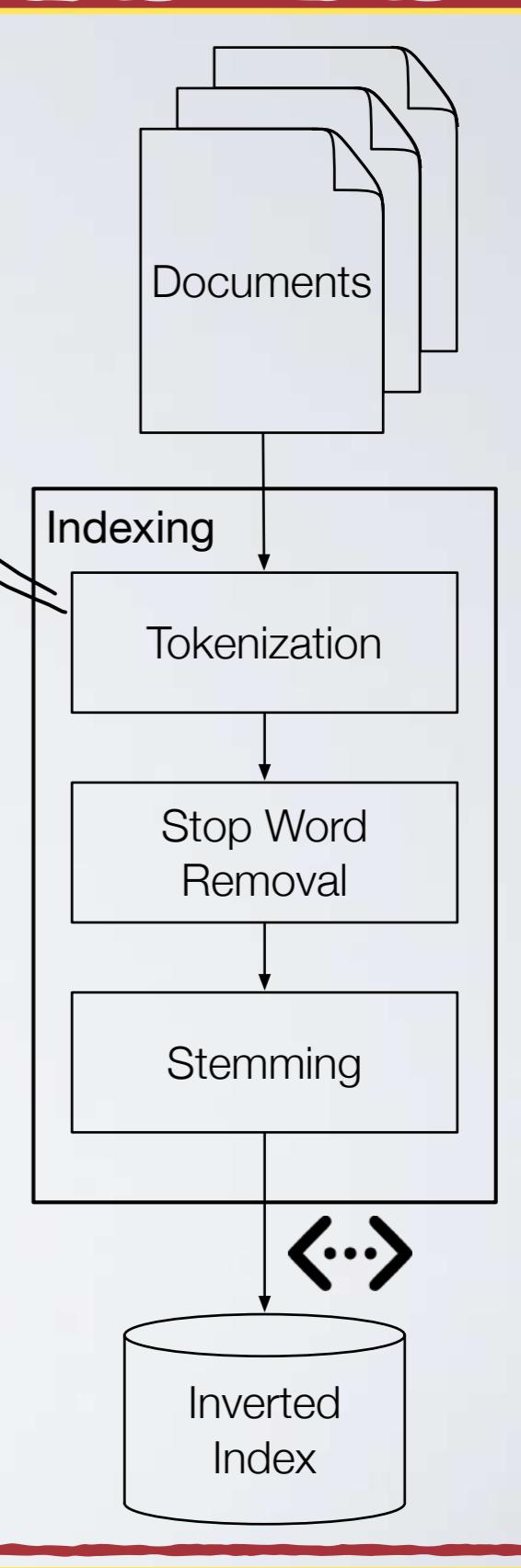
## Inverted Index

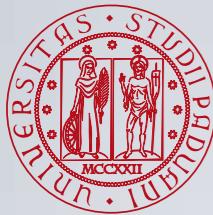
dividere doc in  
TOKENS: parole

importante  
guardare dati  
per sapere come  
implementare

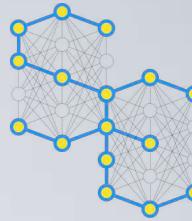
struttura dati per salvare  
indici di documenti

indirizzato per token

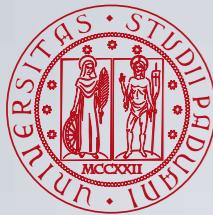




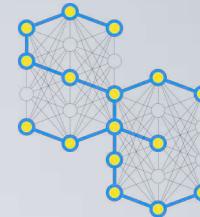
# Lexical Analysis / Tokenization



- **Objective:** treat digits, hyphens, punctuation marks, case of letters
- Convert the text of documents in a stream of words, (candidate words to become index terms)
- Blank spaces and punctuation as word separators and all words converted to lower cases → **tokens**
- Special cases:
  - **Numbers:** their are not good index term without context, too vague, usually disregarded
  - **Hyphens:** example ‘state-of-the-art’ should be treated as a single word
  - **Uppercases:** usually not important, but there are exceptions ‘Bank’ (city in Iran) and bank



# Lexical Analysis: Example



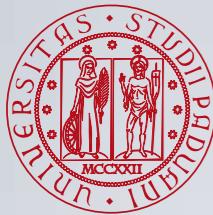
## Original

To be, or not to be: that is the question:  
Whether 'tis nobler in the mind to suffer  
The slings and arrows of outrageous fortune,  
Or to take arms against a sea of troubles,  
And by opposing end them? To die: to sleep;  
No more; and by a sleep to say we end  
The heart-ache and the thousand natural shocks  
That flesh is heir to, 'tis a consummation  
Devoutly to be wish'd. To die, to sleep;  
To sleep: perchance to dream: ay, there's the rub;

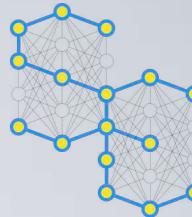
## Tokens

to be or not to be that is the question  
whether tis nobler in the mind to suffer  
the slings and arrows of outrageous fortune  
or to take arms against a sea of troubles  
and by opposing end them to die to sleep  
no more and by a sleep to say we end  
the heart ache and the thousand natural shocks  
that flesh is heir to tis a consummation  
devoutly to be wish d to die to sleep  
to sleep perchance to dream ay there s the rub

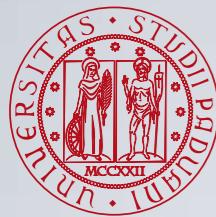
- Words are separated by blank spaces and punctuation
- Punctuation is removed
- Everything is transformed to lowercase



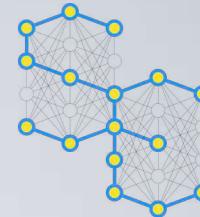
# Elimination of Stopwords



- Objective: filtering words with very low discrimination values for retrieval purposes
- Idea: words which occurs in 80% of the documents in the collection are not informative, these are referred to as **stopwords**
- Examples of stopwords: **articles, prepositions, pronouns, conjunctions, ...**
- Advantage: reduces the size of the index (40% reduction or even more)
- Disadvantage: ‘to be or not to be’ → represented as ‘be’



# Elimination of Stopwords: Example



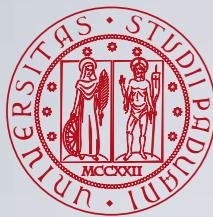
## Tokens

to be or not to be that is the question  
whether tis nobler in the mind to suffer  
the slings and arrows of outrageous fortune  
or to take arms against a sea of troubles  
and by opposing end them to die to sleep  
no more and by a sleep to say we end  
the heart ache and the thousand natural shocks  
that flesh is heir to tis a consummation  
devoutly to be wish d to die to sleep  
to sleep perchance to dream ay there s the rub

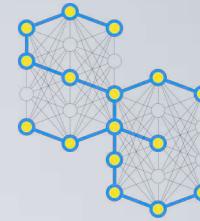
## Stopwords Removal

question  
tis nobler mind suffer  
slings arrows outrageous fortune  
take arms sea troubles  
opposing end die sleep  
sleep say end  
heart ache thousand natural shocks  
flesh heir tis consummation  
devoutly wish die sleep  
sleep perchance dream ay rub

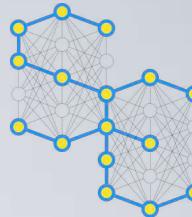
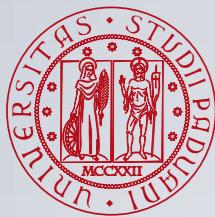
- Some stopwords: to, be, not, that is, whether, in, the, and, of, or, against, a, by, ...



# Stemming



- Objective: removing prefixes and suffixes to allow for syntactic variations
- Examples of syntactical variations: plurals, gerund forms, past tense suffixes, ...
- **Stem:** portion of a word which is left after the removal of prefixes and suffixes, root form
- Example: “connect” is the stem for the variants connected, connecting and connections
- Advantages: reduce variants of the same root word to a common concept and reduce the size of the indexing structure
- Disadvantages: **controversy** in the literature and many conflicting conclusions



# Stemming: Example

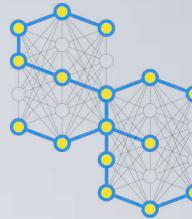
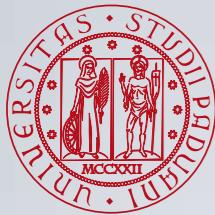
## Stopwords Removal

question  
tis nobler mind suffer  
slings arrows outrageous fortune  
take arms sea troubles  
opposing end die sleep  
sleep say end  
heart ache thousand natural shocks  
flesh heir tis consummation  
devoutly wish die sleep  
sleep perchance dream ay rub

## Stemming

question  
ti nobler mind suffer  
sling arrow outrag fortun  
take arm sea troubl  
oppos end die sleep  
sleep sai end  
heart ach thousand natur shock  
flesh heir ti consumm  
devoutli wish die sleep  
sleep perchanc dream ay rub

- Stemming examples: “arrows” → “arrow” (plural),  
“outrageous” → ‘outrag’ (root), “opposing” → “oppos”  
(gerund form)



# European and Non-European Languages

## ● Tokenization issues common to many European languages that are not shared by English:

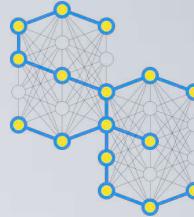
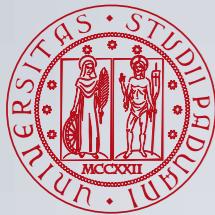
- Accents and other diacritical marks are almost nonexistent in English:  
example Spanish “cuna” is a cradle but “cuña” is a wedge
- Compound nouns written as a single word: example Danish  
“universitetsuddannelse” composed by “universitet” (university) and  
“uddannelse” (degree) and an extra “s” character

## ● Non-European languages may have different characteristics

- different script (but this also happens for some European languages)
- different direction, e.g. Arabic is written right-to-left
- no easy boundaries between “words”, e.g. Chinese does not split “words” at spaces

## ● Mixed languages in a document

# **Basics of Experimental Evaluation**



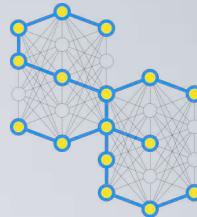
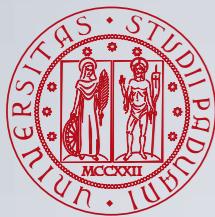
# Why Evaluation?



“To measure is to know”

“If you cannot measure it,  
you cannot improve it”

Lord William Thompson,  
first Baron Kelvin (1824-1907)



# What to Evaluate?

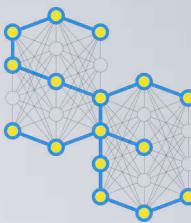
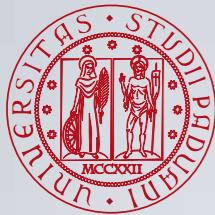
Efficiency



Effectiveness

VS



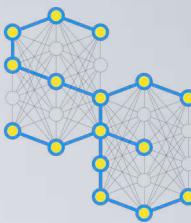
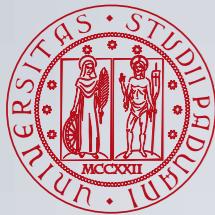


# Evaluation in Action

(c) IBM Corporation. <http://www.youtube.com/watch?v=3G2H3DZ8rNc>



IBM Watson:  
Deep QA Project

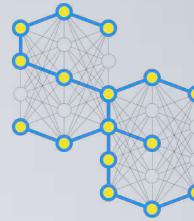
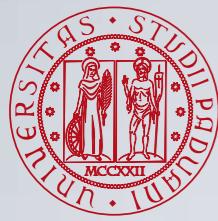


# Evaluation in Action

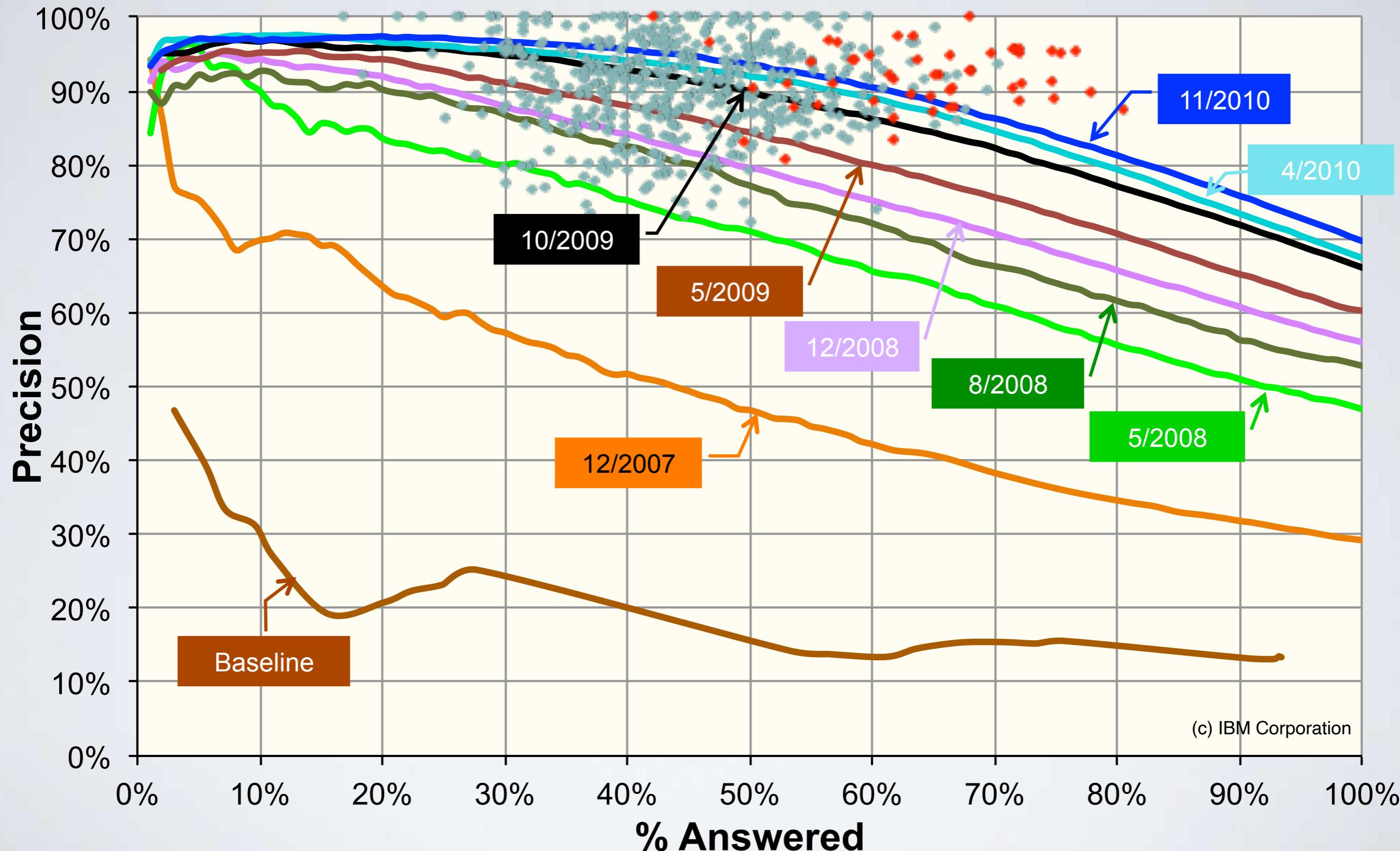
(c) IBM Corporation. <http://www.youtube.com/watch?v=3G2H3DZ8rNc>

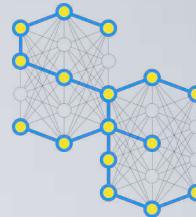
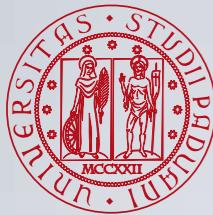


IBM Watson:  
Deep QA Project



# Evaluation in Action





# Critical Issues in Evaluation

- It must be scientifically valid

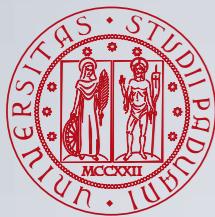
- valid methodology, measures, and statistics
- large-scale enough to be statistically valid
- must be “repeatable” if possible

- It must be realistic for the applications that will be using the information retrieval systems

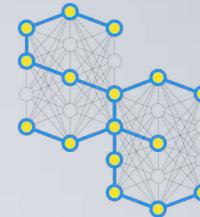
- **task** and use cases

- It must be understandable to your audience/client

Harman, D. K. (2011). *Information Retrieval Evaluation*. Morgan & Claypool Publishers, USA.



# Evaluation Spectrum



TREC-style Studies

“Users” make relevance assessments

Filtering and SDI

Log Analysis

TREC Interactive Studies

Experimental Information Behavior

Information-Seeking Behavior with IR Systems

Information-Seeking Behavior in Context

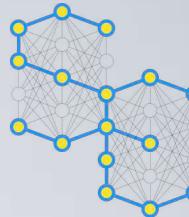
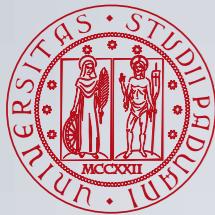
System Focus

Archetypical IIR Study

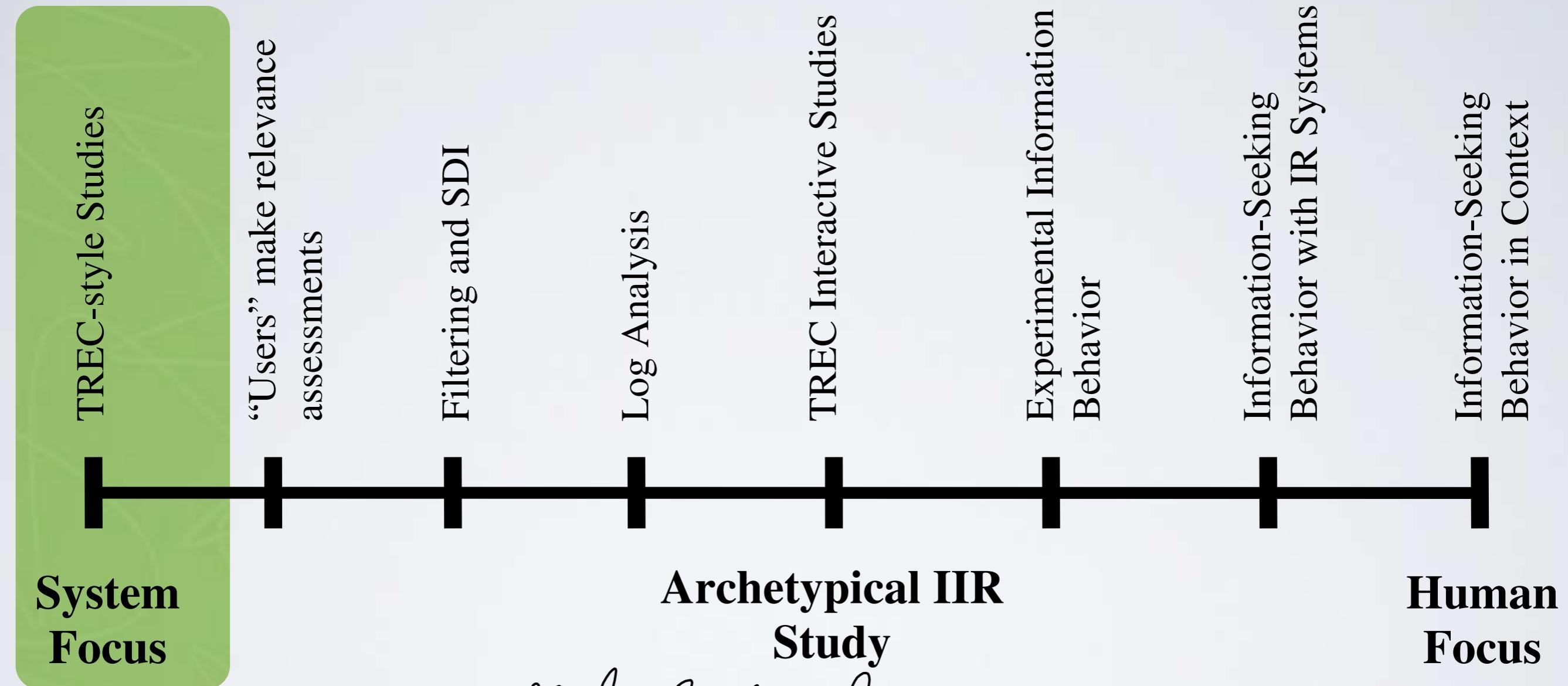
Human Focus

*mettre per evitare di rendere utenti insoddisfatti*

Kelly, D. (2009). Methods for evaluating interactive information retrieval systems with users. Foundations and Trends in Information Retrieval (FnTIR), 3(1-2), 1-224.

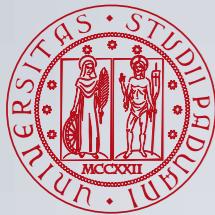


# Evaluation Spectrum

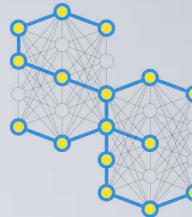


INTERLEAVING: unire risultati di due alg.  
HARD TESTING: attenti hanno uno, retenti hanno l’altro

Kelly, D. (2009). Methods for evaluating interactive information retrieval systems with users. Foundations and Trends in Information Retrieval (FnTIR), 3(1-2), 1-224.



# How Does Experimental Evaluation Work



## ● Cranfield Paradigm by Cyril W. Cleverdon

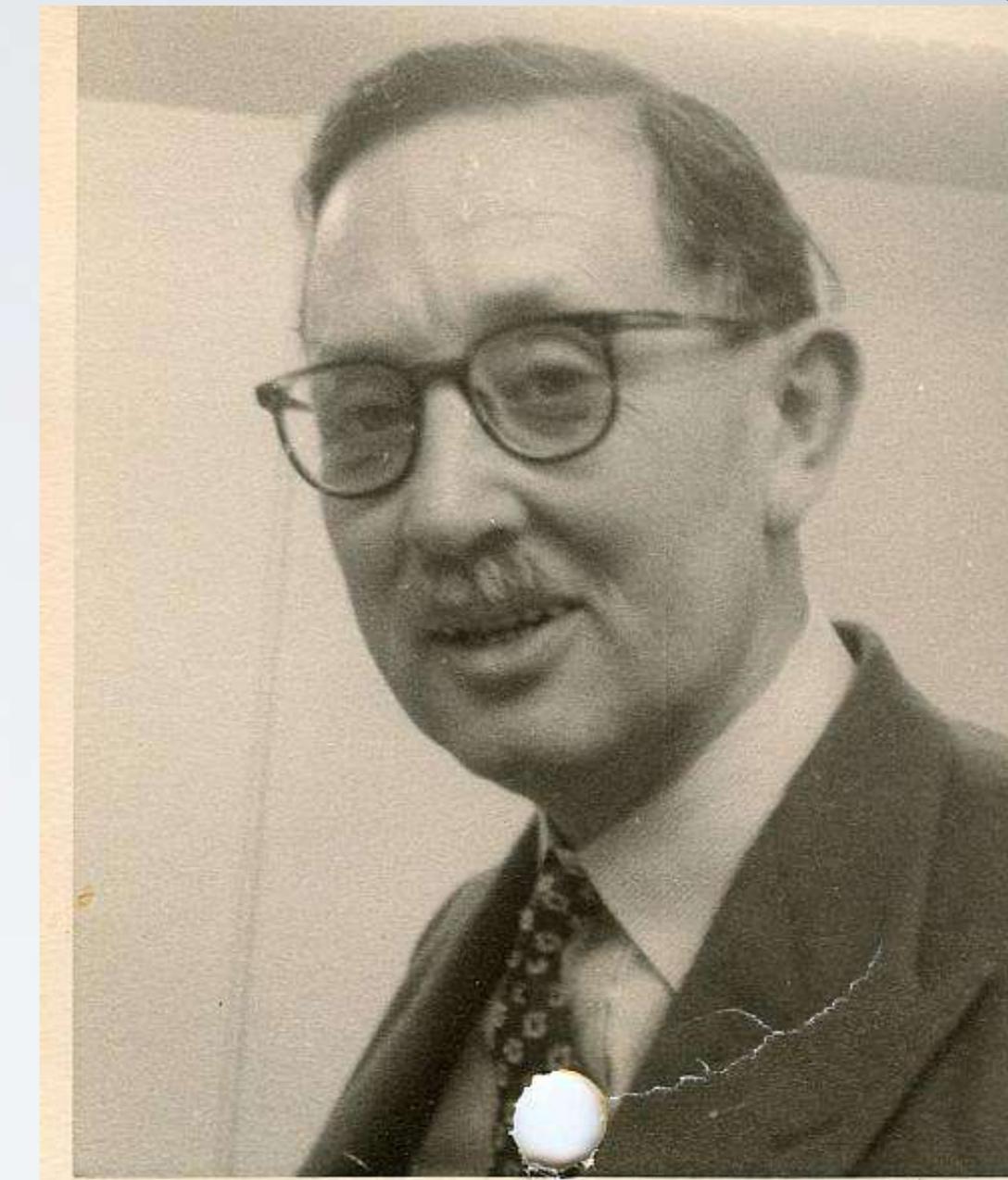
- Dates back to mid 1960s

*ideato da librario*

## ● Makes use of **experimental collections**

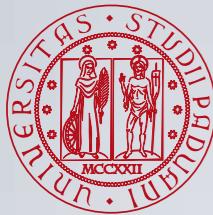
- **documents** (corpora)
- **topics**, which are a surrogate for information needs
- **relevance judgments** (binary or graded)  
also called relevance assessment  
or ground-truth (or qrels)

## ● Ensures comparability and repeatability of the experiments

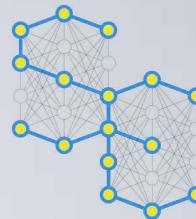


Cyril W. Cleverdon

Cleverdon, C. W. (1962). *Report on the Testing and Analysis of an Investigation into the Comparative Efficiency of Indexing Systems*. Aslib Cranfield Research Project, College of Aeronautics, Cranfield, UK.  
Cleverdon, C. W. (1997). *The Cranfield Tests on Index Languages Devices*. In Spärck Jones, K. and Willett, P., editors, *Readings in Information Retrieval*, pages 47–60. Morgan Kaufmann Publisher, Inc., San Francisco, CA, USA.



# Some Document Collections



## Historical

- **CACM:** 3,024 abstracts from the Communications of the ACM  
[http://ir.dcs.gla.ac.uk/resources/test\\_collections/cacm/](http://ir.dcs.gla.ac.uk/resources/test_collections/cacm/)

## Mid-nineties

- **TIPSTER:** 528,155 documents (news articles, US government reports, ...), Disks 4 and 5 excluding Congressional Record subcollection  
<https://catalog.ldc.upenn.edu/LDC93T3A>

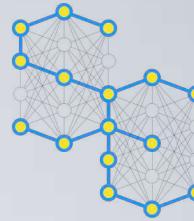
## Early 2000s

- **WT10g:** 1,692,096 Web pages crawled in 2001  
[http://ir.dcs.gla.ac.uk/test\\_collections/wt10g.html](http://ir.dcs.gla.ac.uk/test_collections/wt10g.html)
- **GOV2:** 25,205,179 Web pages crawled fro .gov sites in early 2004  
[http://ir.dcs.gla.ac.uk/test\\_collections/gov2-summary.htm](http://ir.dcs.gla.ac.uk/test_collections/gov2-summary.htm)
- **CLEF Multilingual Corpus:** 4,883,227 multilingual news articles corpus in 13 languages (Bulgarian, Dutch, English, Farsi, Finnish, French, German, Hungarian, Italian, Portuguese, Spanish) gathered in 1994, 1995 and 2002. Topics in 28 different languages (Bengali, Bulgarian, Chinese, Czech, Dutch, English, Farsi, Finnish, French, German, Greek, Hindi, Hungarian, Indonesian, Italian, Japanese, Marathi, Norwegian, Oromo Polish, Portuguese, Russian, Slovenian, Spanish, Swedish, Tamil, Telugu, Thai)

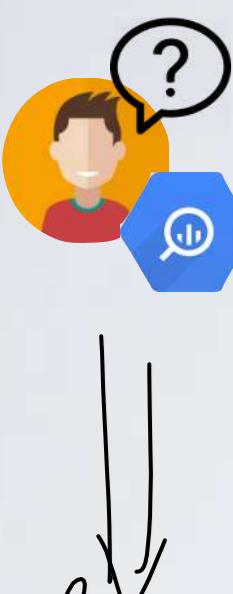
## Today

- **ClueWeb 2009:** 1,040,809,705 Web pages in 10 languages crawled between January and February 2009  
<https://lemurproject.org/clueweb09/>
- **ClueWeb 2012:** 733,019,372 English Web pages crawled between February 10, 2012 and May 10, 2012  
<https://lemurproject.org/clueweb12/>
- **ClueWeb 2022:** 10,000,000,000 Multilingual (40% English, 60% other languages) Web pages, plus additional quality data  
<https://www.lemurproject.org/clueweb22.php/>
- **The New York Times Annotated Corpus:** 1,855,658 news articles from January 1987 through December 2007 from New York Times  
<https://catalog.ldc.upenn.edu/LDC2008T19>
- **TREC Washington Post Corpus:** 595,037 news articles and blog posts from January 2012 through August 2017 from Washington Post  
<https://trec.nist.gov/data/wapost/>
- **MS MARCO:** 3.2 million English documents, 8.8 million passages, 1 million questions  
<https://microsoft.github.io/msmarco/>

# Evaluation with Test Collections in a Nutshell



Topic	Run	Assessed Run	Weighted Assessed Run	Measure Score
-------	-----	--------------	-----------------------	---------------



Relevance Assessment

Highly Relevant
Not Relevant
Partially Relevant
Fairly Relevant
Not Relevant
Not Relevant

3
0
1
2
0
0

$$DCG = 4.6309$$

efficacia, non efficiente

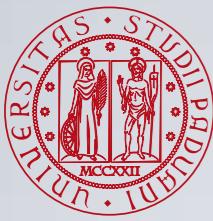
non basta media per confrontare: meglio uscire test statistici



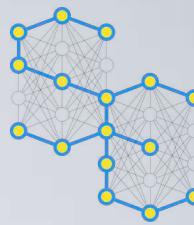
valutare  
ogni doc.  
ritorno

file con  
descrizioni  
di info.  
necessarie

scelgo un topic  
e lo cerco in  
doc. collezione  
tramite sistema  
IR



# Evaluation with Test Collections in a Nutshell



Topic

Run

Assessed Run

Weighted  
Assessed Run

Measure  
Score



Relevance  
Assessment

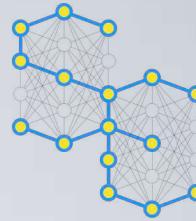
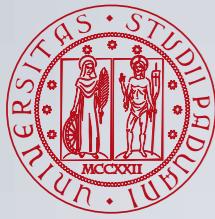
Highly Relevant
Not Relevant
Partially Relevant
Fairly Relevant
Not Relevant
Not Relevant

3
0
1
2
0
0

$$DCG = 4.6309$$



Sanderson, M. (2010). Test Collection Based Evaluation of Information Retrieval Systems. *Foundations and Trends in Information Retrieval (FnTIR)*, 4(4):247–375.



# Example of Topic

```
<?xml version="1.0" encoding="UTF-8"?>
<topic>
  <identifier>41</identifier>

  <title lang="en">Pesticides in Baby Food</title>
  <title lang="fr">Des pesticides dans la nourriture pour bébés</title>
  <title lang="it">Pesticidi negli alimenti per bambini</title>
  <title lang="ru">Пестициды в детском питании</title>
  <title lang="zh">嬰兒食品中含有殺蟲劑</title>
  <title lang="ja">ベビーフード中の病虫害防除剤</title>
  <title lang="th">ยาฆ่าแมลง ใน อาหาร เด็กอ่อน</title>
  <title lang="so">Sunta cayayaanka ee Cuntada Ilmaha</title>
  <title lang="sw">Dawa za kuulia wadudu katika Chakula cha Mtoto</title>

  <description lang="en">Find reports on pesticides in baby food.</description>
  <description lang="fr">
    Rechercher des documents sur les pesticides dans la nourriture pour bébés.
  </description>
  <description lang="it">
    Trova documenti che parlano dei pesticidi negli alimenti per bambini.
  </description>
  <description lang="ru">Найти статьи о пестицидах в детском питании</description>
  <description lang="zh">查詢有關嬰兒食品中含有殺蟲劑的報導。</description>
  <description lang="ja">ベビーフード中の病虫害防除剤に関する記事を探したい。</description>
  <description lang="th">หา รายงาน ที่ เกี่ยวของ กับ ยาฆ่าแมลง ใน อาหาร เด็กอ่อน</description>
  <description lang="so">Hel wargelinada sunta cayayaanka ee cuntada ilmaha.</description>
  <description lang="sw">
    Pata ripoti kuhusu dawa za kuulia wadudu katika chakula cha mtoto.
  </description>

  <narrative lang="en">
    Relevant documents give information on the discovery of pesticides in baby food.
    They report on different brands, supermarkets, and companies selling baby food
    which contains pesticides. They also discuss measures against the contamination
    of baby food by pesticides.
  </narrative>
  <narrative lang="fr">
    Les documents pertinents informent sur la découverte de pesticides dans la
    nourriture pour bébés. Ils contiennent des informations sur les différentes
    marques, les supermarchés et les firmes ayant mis en vente de la nourriture pour
    bébés renfermant des pesticides. Ils relatent également les mesures prises contre
    la contamination de la nourriture pour bébés par les pesticides.
  </narrative>
  <narrative lang="it">
    I documenti rilevanti forniscono informazioni sulla scoperta di pesticidi nei
    cibi per bambini. Riportano i diversi marchi, i supermercati e le ditte che hanno
    venduto alimenti per bambini con i pesticidi. Sono anche rilevanti i documenti
    che discutono le misure contro la contaminazione degli alimenti per bambini con
    i pesticidi.
  </narrative>
</topic>
```

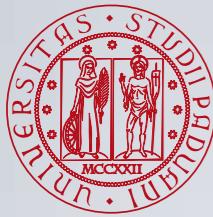
● Topics consists of:

● title: a brief statement expressing the information need.  
It resembles the typical search engine query

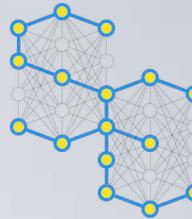
● description: more detailed formulation of the information need

● narrative: instructions for assessors on when to consider a document relevant

● Typical experimental collections make use of 50 topics



# Evaluation with Test Collections in a Nutshell



Topic

Run

Assessed Run

Weighted  
Assessed Run

Measure  
Score



Highly Relevant
Not Relevant
Partially Relevant
Fairly Relevant
Not Relevant
Not Relevant

3
0
1
2
0

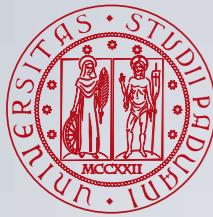
$$DCG = 4.6309$$



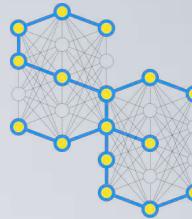
- We can compare two systems A and B run on the same test collection by comparing their average performance or, much better, by using statistical significance tests



Sanderson, M. (2010). Test Collection Based Evaluation of Information Retrieval Systems. *Foundations and Trends in Information Retrieval (FnTIR)*, 4(4):247–375.



# Evaluation with Test Collections in a Nutshell



Topic

Run

Assessed Run

Weighted  
Assessed Run

Measure  
Score



- Since we can perform relevance assessment over topics, we can average the scores over them
- We can compare two systems A and B run on the same test collection by comparing their average performance or, much better, by using statistical significance tests

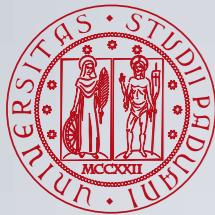


Highly Relevant
Not Relevant
Partially Relevant
Fairly Relevant
Not Relevant
Not Relevant

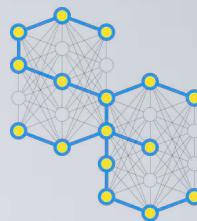
3
0
1
2
0

$$DCG = 4.6309$$

Sanderson, M. (2010). Test Collection Based Evaluation of Information Retrieval Systems. *Foundations and Trends in Information Retrieval (FnTIR)*, 4(4):247–375.



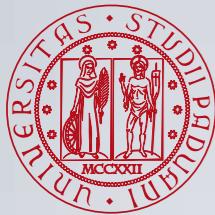
# Example of Relevant Document



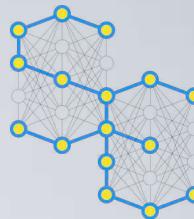
doc. da TIPSTER  
identificatore  
regola: se ci è almeno  
una frase rilevante,  
tutto doc lo è

```
<DOC>
<DOCNO> LA091294-0164 </DOCNO>
<DOCID> 078484 </DOCID>
<SOURCE>
<P>
  Los Angeles Times
</P>
</SOURCE>
<DATE>
<P>
  September 12, 1994, Monday, Home Edition
</P>
</DATE>
<SECTION>
<P>
  Part A; Page 19; Column 1; Metro Desk
</P>
</SECTION>
<LENGTH>
<P>
  1502 words
</P>
</LENGTH>
<HEADLINE>
<P>
  ALAR PANIC SHOWS POWER OF MEDIA TO TRIGGER FEAR; HEALTH: '60 MINUTES' BROADCAST CREATED SCARE AT TIME WHEN THE INDUSTRY WAS ALREADY MOVING AWAY FROM USE OF THE CHEMICAL.
</P>
</HEADLINE>
<BYLINE>
<P>
  By DAVID SHAW, TIMES STAFF WRITER
</P>
</BYLINE>
<TEXT>
<P>
  The warning -- uttered on prime-time network television, against a backdrop of a giant apple marked with a skull and crossbones -- was indeed ominous:
</P>
<P>
  "The most potent cancer-causing agent in our food supply is a substance sprayed on apples to keep them on the trees longer and make them look better."
</P>
<P>
  So said Ed Bradley on "60 Minutes" on Feb. 26, 1989, and so began a nationwide panic, fed by other media, which quickly followed "60 Minutes'"s lead with their own stories on the killer chemical d aminozide (better known by its trade name -- Alar).
</P>
<P>
  Young children, the media reported, were especially vulnerable, because they tended to drink a lot of apple juice and because their digestive and immune systems were not fully developed.
</P>
<P>
  But at the time of the "60 Minutes" broadcast -- which was viewed by an estimated 40 million Americans -- industry was already moving away from Alar, and the nation's three major baby-food makers said they were using non-Alar apples.
</P>
<P>
  The Natural Resources Defense Council, an activist environmental group, wanted the U.S. Environmental Protection Agency to ban Alar and several other pesticides. When the group produced a report ("Intolerable Risk: Pesticides in Our Children's Food") condemning Alar, Newsweek rushed into print with the story before the report was officially released.
</P>
```

<P>  
The report, Newsweek said, would "almost certainly generate frightening headlines."  
</P>  
<P>  
Indeed it did -- three weeks later, right after "60 Minutes" aired its story. News media throughout the country -- effectively manipulated by the Natural Resources Defense Council and aided by public appeals and congressional testimony from that well-known molecular biologist Meryl Streep -- almost made it seem that one bite of an Alar-treated apple or one swig of juice made from Alar-treated apples would mean instant death.  
</P>  
<P>  
Coverage of the Alar scare was "outrageous . . . completely alarmist," says Marla Cone, who writes about the environment for the Los Angeles Times.  
</P>  
<P>  
But Alar was a made-to-order media story. It had apples, kids and cancer. "A lot of (media) people were suckered in," Cone says.  
</P>  
<P>  
The media coverage produced a nationwide hysteria.  
</P>  
<P>  
School boards in Los Angeles, New York, Chicago, Atlanta and many other cities banned apples and apple products from their cafeterias. Some parents raced after their children's school buses to yank apples from their lunch boxes. Supermarkets came under intense pressure to remove apples from their shelves. Uniroyal, the manufacturer of Alar, pulled the product off the market. Sales of apples plummeted, forcing many farmers to dump their crops or give them away -- costing the industry more than \$100 million, according to economists' estimates.  
</P>  
<P>  
Reaction to the Alar scare "set a troubling precedent," the Washington Post editorialized several weeks later. "A complicated scientific issue was allowed to be decided not by officials charged with protecting the public, on the basis of hard evidence, but by a frightened public acting on incomplete and often erroneous press reports."  
</P>  
<P>  
The EPA had expressed concern about the safety of Alar for many years before Newsweek and "60 Minutes" jumped on the story. But the agency had decided that test results were either flawed, contradictory or insufficiently conclusive to warrant an immediate ban and formal action was delayed, pending hearings in late 1990.  
</P>  
<P>  
David Gelber, the producer of the "60 Minutes" Alar story and now the executive producer for ABC's "Peter Jennings Reporting," says EPA and scientific criticism of Alar convinced him the story was worth doing at the time.  
</P>  
<P>  
Dr. John A. Moore, then acting administrator of the EPA, said on "60 Minutes" that Alar "should come off the market" because of what he had earlier described as "an inescapable and direct correlation" between exposure to Alar and "the development of life-threatening tumors."  
</P>  
<P>  
"The public had a right to know that was their view," Gelber says.  
</P>  
<P>  
Al Meyerhoff, senior attorney for the Natural Resources Defense Fund, also defends the Alar story.  
</P>  
<P>  
He says the apple industry launched a "concerted disinformation campaign" in an effort to persuade the media and consumers alike that Alar was not

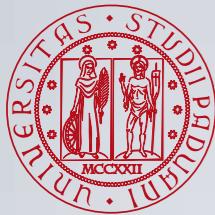


# Example of Relevant Document

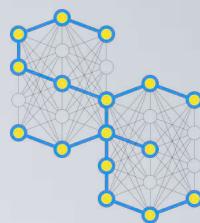


<DOC>  
<DOCNO> LA041694-0248 </DOCNO>  
<DOCID> 032882 </DOCID>  
<SOURCE>  
</P>  
Los Angeles Times  
</P>  
</SOURCE>  
<DATE>  
<P>  
April 16, 1994, Saturday, Home Edition  
</P>  
</DATE>  
<SECTION>  
<P>  
Business; Part D; Page 1; Column 2; Financial Desk  
</P>  
</SECTION>  
<LENGTH>  
<P>  
1135 words  
</P>  
</LENGTH>  
<HEADLINE>  
<P>  
WARNING FLAG OVER GATT; ENVIRONMENTALISTS SAY PACT WILL WEAKEN U.S. SAFEGUARDS  
</P>  
</HEADLINE>  
<BYLINE>  
<P>  
By MELISSA HEALY, TIMES STAFF WRITER  
</P>  
</BYLINE>  
<TEXT>  
<P>  
The environmental community, which recently split deeply over support for the North American Free Trade Agreement, is issuing warnings about the new world trade agreement in a newly unified voice.  
</P>  
<P>  
The environmentalists charge that the comprehensive General Agreement on Tariffs and Trade accord signed Friday in Morocco will erode the United States' -- and California's -- ability to enforce its environmental strictures on everything from recycling to pesticide use to air pollution.  
</P>  
<P>  
"If this . . . is enacted," said Barbara Dudley, executive director of Greenpeace USA, "over two decades of environmental protection could be severely weakened."  
</P>  
<P>  
Nowhere is that more true than in the area of food safety, some environmentalists argue.  
</P>  
<P>  
The United States, with some of the world's most restrictive regulation of pesticides, prohibits the entry of food products with detectable traces of about 40 chemicals -- substances used by many of its trading partners and listed as allowable by the standard-setting organization of the new agreement.  
</P>  
<P>  
But if U.S. Customs Service inspectors begin turning away food imports that bear traces of these chemicals under the new agreement, America's trading partners are almost certain to cry foul, environmentalists warned. If their challenges stand, one activist said, U.S. pesticide protections could topple, one after another.  
</P>  
<P>  
Supporters of the accord acknowledged that the letter of the new agreement may indeed put the United States, with its strong environmental protections, on the defensive. That is because some U.S. laws championed by environmentalists, as well as state laws, do not appear to be based on undisputed scientific evidence demonstrating that a regulation will improve the public's health or mitigate a known environmental hazard.  
</P>

<P>  
Supporters of the agreement also argue that the United States' market power, as well as a growing appreciation for the environmental ethic among America's trading partners, will cause virtually any challenge to U.S. environmental laws to fail.  
</P>  
<P>  
"The practical implementation of trade agreements is often more politically sensitive and realistic than the sheer language of the treaty," said William K. Reilly, administrator of the U.S. Environmental Protection Agency in the George Bush Administration and an avowed free trader.  
</P>  
<P>  
Other free traders cited the power and transparency of the processes by which U.S. laws are passed and regulations are made. In a challenge before the international trade deliberative body, U.S. defenders could produce reams of scientific data, risk assessments and economic analysis to prove that an American environmental standard was established not to keep foreign products out, but to benefit the public's health.  
</P>  
<P>  
"In a fair fight," said Linda Fisher, a Washington-based trade attorney with the Los Angeles law firm Latham & Watkins, "the United States will win."  
</P>  
<P>  
But environmental activists are not comforted by those assurances. They say that the fine print of the new agreement would allow a trading partner to argue before the trade court in Geneva, Switzerland, that Washington's environmental laws -- or those of individual U.S. states -- constitute an unfair barrier to the entry of that country's exports.  
</P>  
<P>  
If the United States regulates more strictly than its trading partners the fuel emissions or stipulates the fuel efficiency of cars sold or operated within its borders (which it does), or prohibits food products that bear traces of certain pesticides it considers hazardous (which it does), a trading partner, in principle, can challenge the federal stricture under the trade agreement.  
</P>  
<P>  
Environmentalists said that under the trade agreement, trading partners could target laws such as the one that grew out of California's Proposition 65, which requires a cautionary label on any product that would expose its user to a carcinogen or a chemical that could be harmful to a developing fetus or pregnant woman.  
</P>  
<P>  
Other California state regulations that could be challenged go beyond federal government requirements by making manufacturers of agricultural chemicals furnish the state with data on the chemical's possible effects on human reproduction, water pollution, exposed workers and endangered species.  
</P>  
<P>  
Another California stricture that could come under attack is one that has required wine manufacturers to sponsor efforts to warn consumers of the possible dangers posed by the lead in the foil that covers wine corks. Finally ... A World Trade Pact  
</P>  
<P>  
Ministers from 124 nations ended seven years of complex negotiations Friday and formally concluded the Uruguay Round of the General Agreement on Tariffs and Trade talks in Marrakesh, Morocco. Following are the main elements of the 26,000-page, 385-pound global world trade pact:  
</P>  
<P>  
MARKET ACCESS - This is the backbone of the act. Countries pledge to cut tariffs on industrial and farm goods by an average of about 37%. The United States and European Union agree to trim tariffs between them by one half.  
</P>  
<P>  
SERVICES - For the first time, rules will govern annual trade in services such as banking, insurance and travel, as well as the movement of labor. The United States reserves the right to deny other countries favorable access to the lucrative U.S. financial services market, but will hold off for at least 18 months. Washington has threatened to challenge EC curbs on audio-visual goods.  
</P>



# Example of Not Relevant Document



```
<DOC>
<DOCNO> LA072894-0040 </DOCNO>
<DOCID> 064623 </DOCID>
<SOURCE>
<P>
  Los Angeles Times
</P>
</SOURCE>
<DATE>
<P>
  July 28, 1994, Thursday, Home Edition
</P>
</DATE>
<SECTION>
<P>
  Food; Part H; Page 24; Column 1
</P>
</SECTION>
<LENGTH>
<P>
  3275 words
</P>
</LENGTH>
<HEADLINE>
<P>
  NEW VIETNAMESE CUISINE; MARKETS; THE BEST OF LITTLE SAIGON: MADE IN AMERICA
</P>
</HEADLINE>
<BYLINE>
<P>
  By LINDA BURUM
</P>
</BYLINE>
<TEXT>
<P>
  * Little Saigon Supermarket, 9822 Bolsa Ave., Westminster, (714) 531-7272, 9
  a.m. to 9 p.m. daily.
</P>
<P>
  The powerful whirling helicopter blades kicked up thick clouds of dust as Twan
  Ngo pushed her youngest child into a hovering American chopper. It was April
  27, 1975, three days before the fall of Saigon and the last time Ngo would see
  her suburban Saigon home.
</P>
<P>
  Many of the 165,000 Vietnamese who came to the United States in the first
  post-war wave of immigrants were relocated in Northeastern or Midwestern
  cities. At first the Ngos found themselves in Green Bay, Wis., where, as you
  might expect, they found little in the way of Vietnamese food.
</P>
<P>
  Eventually, along with many fellow transplanted Vietnamese, the Ngos wound up
  in Orange County near the neighborhood centered on Bolsa Avenue -- and running
  from Westminster through Garden Grove to Santa Ana -- that came to be known as
  Little Saigon. Lured by the warmer climate and a growing, cohesive Vietnamese
  community, the Ngos found many ties to their culture here, one of the strongest
  being the ready availability of Vietnamese foods and ingredients.
</P>
<P>
  "You won't find most of these vegetables in Wisconsin," Ngo says, picking over
  the greens in the produce section of the bright, ultra-modern Little Saigon
  Supermarket on Bolsa Ave. By "these vegetables," she means the dozens of
  Vietnamese specialties such as bac ha , the pale green spongy stems that go
  into sour soup, or muop huong , a squash resembling a large zucchini.
</P>
<P>
  *
</P>
<P>
  In recent years, specialty produce farming has turned into a livelihood for a
  number of local Vietnamese farmers, evolving hand in hand with a now-sizeable
  Vietnamese food industry. Many of these California-grown or -made goods are
  distributed nationwide to serve the million or so Vietnamese who have come to
  the United States over the years.
</P>
<P>
  At the center of this commercial activity is Orange County, where Vietnamese
  residents officially increased 271% in the last decade (the Vietnamese-American
  Political Action Committee contends the number is nearly twice what the census
  reports). The sheer size of this local customer base opened up a lucrative
  market for prepared foods as well as shelf-stable items; in much smaller
  Vietnamese communities, it probably wouldn't be economic to market these
  perishables.
</P>
```

<P>

But at Little Saigon Supermarket, Ngo chooses from an increasing range of Vietnamese-style processed meats, all sorts of fresh noodles, herbs, pickled vegetables and soy products (such as fresh tofu and soy milk) that are impossible to import. The selection has made Vietnamese eating here as close to authentic as it gets outside of Vietnam itself.

</P>

<P>

As the selection shows, Little Saigon Supermarket owner David Tran knows the Vietnamese food business inside and out. Tran came to the Little Saigon area when it was still in its awkward growth stages -- a mere stretch of bean and strawberry fields, flower warehouses and bottling plants, and just a few Vietnamese stores and businesses.

</P>

<P>

\*

</P>

<P>

He had left Vietnam in '78 on a crowded boat and ended up temporarily in Escondido, learning English and studying to be an electronic technician. Tran found it impossible to support his family on a technician's wages, but having been a businessman (a bicycle wholesaler, to be exact) in Vietnam, he perceived a growing need for Vietnamese wholesale food distribution. He formed Delta Food Company and began to supply Orange County's escalating number of Vietnamese restaurants and markets with wholesale produce and other ingredients used for Vietnamese and Chinese cooking.

</P>

<P>

As time went on, local Asian farmers began to produce small quantities of specialty herbs and vegetables. "Some would ask him to distribute them," says Denise Tran, who is David's sister-in-law and also vice-president of the market.

</P>

<P>

The 1975 Vietnam trade embargo meant an end to Vietnamese imports. Producers in other Asian countries, particularly Thailand, began putting Vietnamese-language labels on foods they had in common with Vietnam, such as fish paste, dry rice noodles and curry powder, and exporting them to the US for the expatriate market. In the beginning, Tran stocked a lot of these items, but they weren't created specifically for the Vietnamese palate, and many Vietnamese cooks have never ceased to regard them as mere substitutes.

</P>

<P>

Now, however, Vietnamese in this country are no longer dependent on imports. They're producing their own extravagant assortments of Vietnamese-style sauces, spice blends, pickled fish, fresh rice papers, deli foods, sweets, beverages and baked goods, all made in America.

</P>

<P>

\*

</P>

<P>

There's as much competition among the various brands of these foods as there is between American breakfast cereal or coffee companies. You see five or six styles of curry powder and at least that many of the fresh hot chili sauce called tuong ot toi (not to mention many other varieties of hot sauce). Four companies manufacture the skinny fresh rice-noodle cakes called banh hoi and at least half a dozen make the bologna-like Vietnamese sausages chua lua and cha bi. The UPC bar-coded packages of fish dumpling paste in the freezer case also come from several companies.

</P>

<P>

In addition, Little Saigon Supermarket stocks prepared deli foods such as cakes, vegetable-filled rice noodle rolls, taro balls, mung bean desserts and other tidbits, all from bakeries and snack shops in the area. These are displayed on a huge table in the middle of the store and their packaging is still rather basic: foam trays enveloped in shrink wrap or "clam shell" boxes designed for take-out foods. But the manufactured foods' packaging is starting to get more sophisticated.

</P>

<P>

Now there are even convenience products such as spice packets to flavor duck soup, mixes for Vietnamese rice rolls and a Hamburger Helper-like mix to season bo kho , the Vietnamese beef stew. So far though, no Vietnamese TV dinners have become part of this burgeoning industry's repertoire. Shopping List

</P>

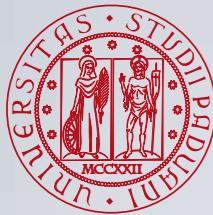
<P>

It's not possible to cover all the locally made Vietnamese products here but this will give you a little taste of what's out there. SOUP SEASONING MIXES

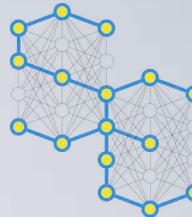
</P>

<P>

Noodle soups are a daily food in Vietnam, as popular as breakfast cereals or hamburgers in this country. Over there, few people bother to make them at home,



# Where Do Topics and Collections Come From?



## Topics

- Collection exploration

- Logs ⇒ e.g. query logs

- Observation of real users

COMPARABLE CORPORA: in + lingue  
Topic devono coprire più lingue  
giornali da paesi diversi possono coprire  
diversi argomenti ⇒ gruppi di persone  
monolingue. Trovano topic, si associano  
anche attenzione a topic s'intendibili  
(e.g. annessi)

## Collections ⇒ devono rappresentare realtà

⇒ dati + profondi,  
no privacy

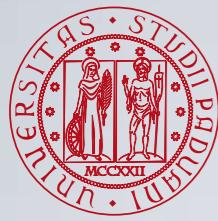
- Opportunistic, e.g. my email or university Web site

- Constructed, e.g. tweets with #hashtag or results of a query to a search engine ⇒ esplorare account / hashtag / handle specifici ⇒ nostri interessi / tropo focus

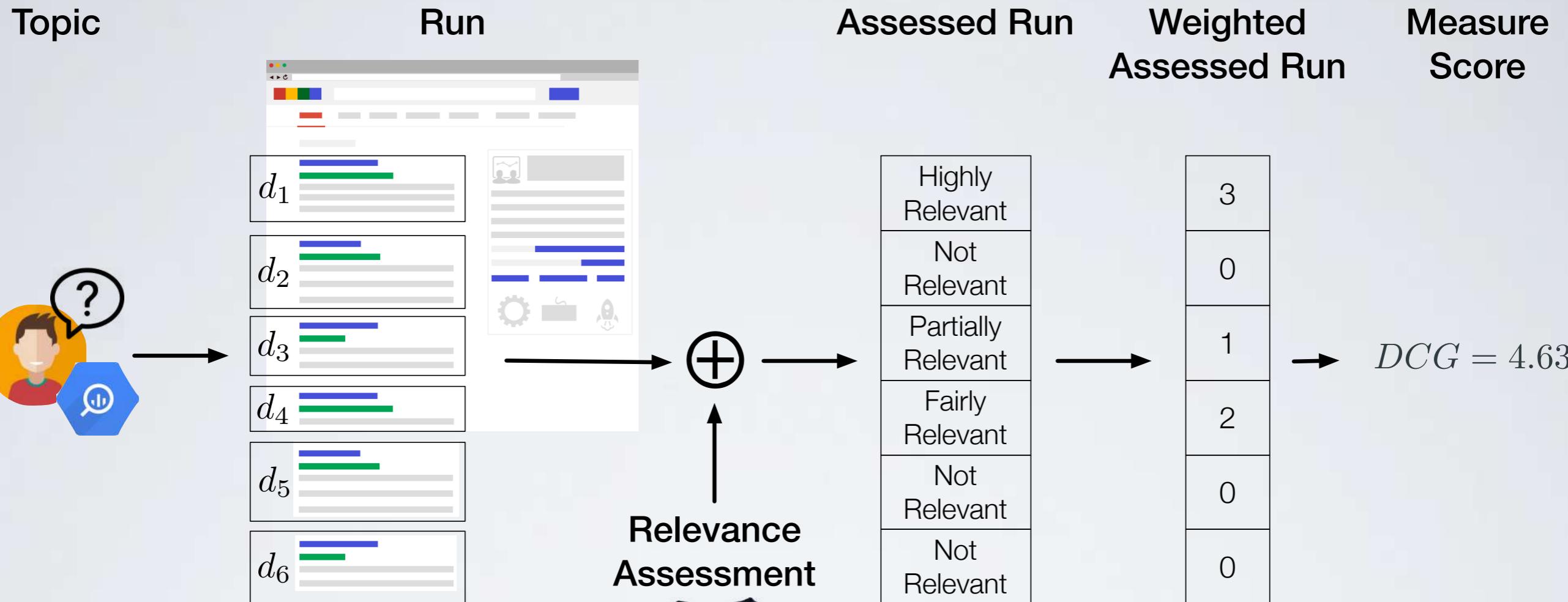
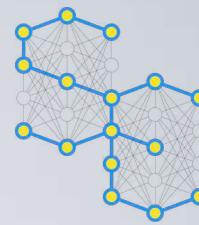
- Naturalistic, e.g. large Web crawl, a year of news, a month of tweets

esplorare web

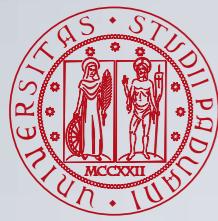
Soboroff, I. (2017). Building Test Collections: An Interactive Guide for Students and Others Without Their Own Evaluation Conference Series. In Kando, N., Sakai, T., Joho, H., Li, H., de Vries, A. P., and White, R. W., editors, Proc. 40th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2017), pages 1407–1410. ACM Press, New York, USA.



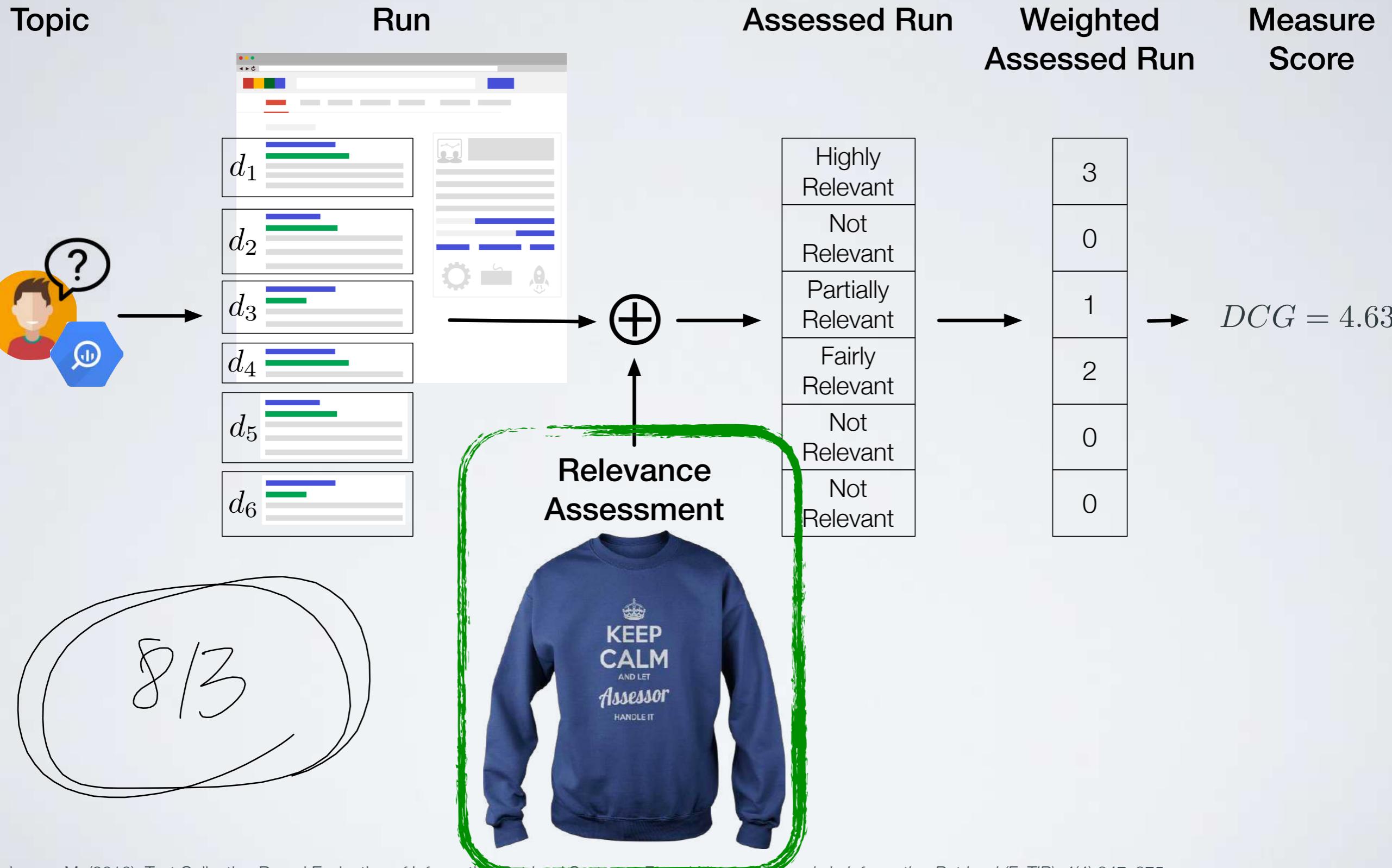
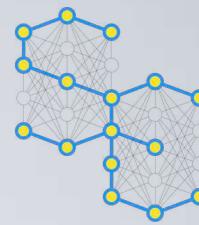
# Evaluation with Test Collections in a Nutshell



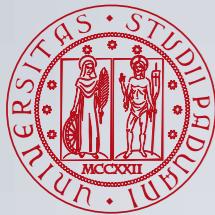
Sanderson, M. (2010). Test Collection Based Evaluation of Information Retrieval Systems. *Foundations and Trends in Information Retrieval (FnTIR)*, 4(4):247–375.



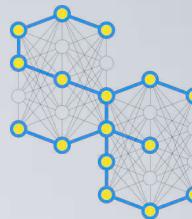
# Evaluation with Test Collections in a Nutshell



Sanderson, M. (2010). Test Collection Based Evaluation of Information Retrieval Systems. In *Information Retrieval Trends in Information Retrieval (FnTIR)*, 4(4):247–375.



# Example of Ground-truth (trec\_eval format)

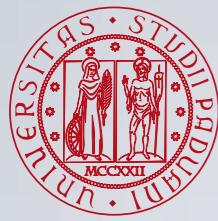


↳ deo si tempe 0

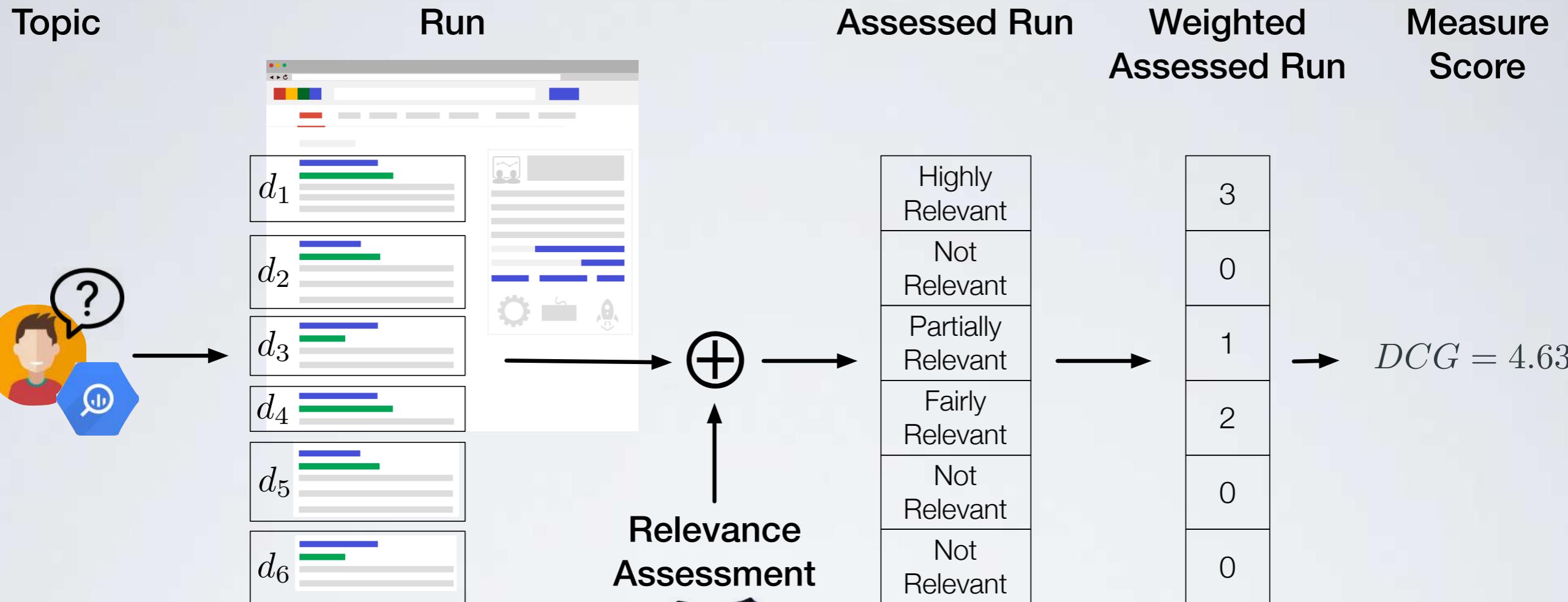
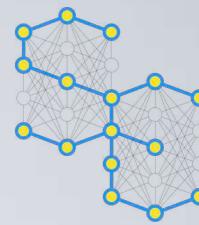
e.g. in cette  
campagne,  
-1 per spam

Topic ID	Fixed	Document ID	Judgement
41	0	LA050394-0237	0
41	0	LA112394-0177	0
41	0	LA091294-0164	1
41	0	LA040594-0187	0
41	0	LA041694-0248	1
...			
42	0	LA031694-0234	0
42	0	LA040494-0111	0
42	0	LA081794-0171	1

- Relevance judgements (qrels) are textual files whose field are separated by tab or space
- Typically, for each topic there are 300-700 judgement documents and the number of judged document vary from topic to topic



# Evaluation with Test Collections in a Nutshell



Sanderson, M. (2010). Test Collection Based Evaluation of Information Retrieval Systems. *Foundations and Trends in Information Retrieval (FnTIR)*, 4(4):247–375.

# Evaluation with Test Collections in a Nutshell



Topic

Run

Assessed Run

Weighted  
Assessed Run

Measure  
Score



Sanderson, M. (2010). Test Collection Based Evaluation of Information Retrieval Systems. *Foundations and Trends in Information Retrieval (FnTIR)*, 4(4):247–375.

# Example of Run (trec\_eval format)

Topic ID	Fixed	Document ID	Rank	Score	Run ID
41	Q0	LA050394-0237	1	0.6342	updrun
41	Q0	LA091294-0164	2	0.4278	updrun
41	Q0	LA040594-0187	3	0.4278	updrun
41	Q0	LA041694-0248	4	0.3197	updrun
41	Q0	LA102394-0113	5	0.3005	updrun
		$\hookrightarrow$ doc trovati per certo topic prioritā, per risolvere parergo: stima di sistema IR			
42	Q0	LA081794-0171	1	0.7687	updrun
42	Q0	LA031694-0235	2	0.7011	updrun
42	Q0	LA031694-0234	3	0.6950	updrun

- Runs are textual files whose field are separated by tab or space
- Typically, there are 50 topics and 1,000 documents are retrieved for each topic (i.e. 50,000 lines)

Eng. ordine in cui doc vengono visualizzati da sistema



# Large-scale Evaluation Initiatives: TREC



- TREC (Text REtrieval Conference), USA, since 1992
- <https://trec.nist.gov/>

**Text REtrieval Conference (TREC)**  
*...to encourage research in information retrieval  
from large text collections.*

[Overview](#)

[Publications](#)

[Information for Active Participants](#)



[Tracks](#)

[Past TREC Results](#)

[Contact Information](#)

[Data](#)

[Frequently Asked Questions](#)

[Call for Participation in TREC 2023](#)

[Celebration of the 25th TREC: November 15, 2016](#)

[TREC Economic Impact Study](#)

[TREC Statement on Product Testing and Advertising](#)

---

The TREC Conference series is co-sponsored by the NIST [Information Technology Laboratory's \(ITL\) Retrieval Group](#) of the [Information Access Division \(IAD\)](#)  
Contact us at: trec (at) nist.gov



Ian Soboroff

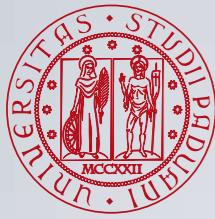


Donna Harman

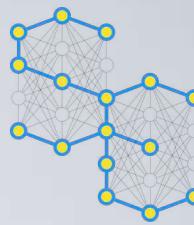


Ellen M. Voorhees

Harman, D. K. and Voorhees, E. M., editors (2005). *TREC. Experiment and Evaluation in Information Retrieval*. MIT Press, USA.



# Large-scale Evaluation Initiatives: NTCIR



- NTCIR (NII Testbeds and Community for Information access Research), Japan, since 1999
- <http://research.nii.ac.jp/ntcir/index-en.html>

NTCIR (NII Testbeds and Community for Information access Research) Project

NTCIR Japanese About NTCIR FAQ Search

Publications/Online Proceedings Data/Tools NTCIR Conference Related URL's Contact us

NTCIR Home

NTCIR 18  
NTCIR 17  
NTCIR 16  
NTCIR 15  
NTCIR 14  
NTCIR 13  
NTCIR 12  
NTCIR 11  
NTCIR 10  
NTCIR 9  
NTCIR 8  
NTCIR 7

What's New

2024.01.29 **NEW** NTCIR-17 Call for Task Proposals released. (due: Feb 9, 2024)

2024.01.29 **NEW** The NTCIR-17 Conference ended successfully. Thank you very much for your participation, cooperation, support, and interest.

2023.10.18 NTCIR-17 Conference is approaching soon! For the detailed information, please refer [NTCIR-17 Conference Webpage](#). We are looking forward to welcoming all of you in Tokyo.

2023.07.12 Task participation was closed for all tasks. Thank you for your registration.

2023.01.06 NTCIR-17 Task Registration Started (Deadline: [\[please see here\]](#))

2022.11.25 The videos and slides for the Tutorial at the NTCIR-16 Conference by Prof. Tetsuya Sakai are now available online.

2022.11.09 The NTCIR-17 Program Committee has accepted the following pilot task, which has been conditionally accepted.  
UFO

2022.09.15 Call for participation to the NTCIR-17 Kick-Off Event

Upcoming Events

December 11-15, 2023 [NTCIR-17 Conference](#)

Around September 2022 [NTCIR-17 Kick-Off Event](#)

Past Events

June 14-17, 2022 [NTCIR-16 Conference](#)

March 29, 2021 [NTCIR-16 Kick-Off Event](#)

December 08-11, 2020 [NTCIR-15 Conference](#)

September 30, 2019 [NTCIR-15 Kick-Off Event](#)

June 10-13, 2019 [NTCIR-14 Conference & EVIA 2019](#)

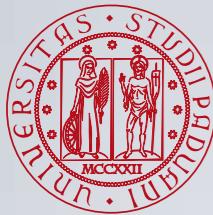
March 20, 2018 [NTCIR-14 Kick-Off Event](#)

February 20, 2018 [NTCIR-14 QALab-PollInfo](#)



Noriko Kando

Sakai, T., Oard, D. W., and Kando, N., editors (2021). *Evaluating Information Retrieval and Access Tasks – NTCIR’s Legacy of Research Impact*, volume 43 of *The Information Retrieval Series*. Springer International Publishing, Germany.



# Large-scale Evaluation Initiatives: CLEF



- **CLEF** (Conference and Labs of the Evaluation Forum), Europe, since 2000

- <https://www.clef-initiative.eu/>

The CLEF Initiative promotes research, innovation, and development of *information access systems* with an emphasis on *multilingual and multimodal information with various levels of structure*. CLEF promotes research and development by providing an infrastructure for:

- multilingual and multimodal system testing, tuning and evaluation;
- investigation of the use of unstructured, semi-structured, highly-structured, and semantically enriched data in information access;
- creation of reusable test collections for benchmarking;
- exploration of new evaluation methodologies and innovative ways of using experimental data;
- discussion of results, comparison of approaches, exchange of ideas, and transfer of knowledge.

**UNLOCKING INFORMATION ACCESS**

The CLEF Initiative (Conference and Labs of the Evaluation Forum) is structured in two main parts:

1. a series of *Evaluation Labs*, i.e. laboratories to conduct evaluation of information access systems and workshops to discuss and pilot innovative evaluation activities;
2. a peer-reviewed *Conference* on a broad range of issues, including
  - investigation continuing the activities of the Evaluation Labs;
  - experiments using multilingual and multimodal data; in particular, but not only, data resulting from CLEF activities;

**Tweets from @clef\_initiative**

t Clef Retweeted ESSIR @essir... · Dec 9, 2022  
Call for bids to host #essir2024 Deadline 5 March 2023. More info at essir.eu/assets/charter...  
@clef\_initiative @ACMSIGIR @ecir2023 @ecir2022 @SIGIRConf

**Bid to Host ESSIR 2024 Deadline: 5 March 2023**

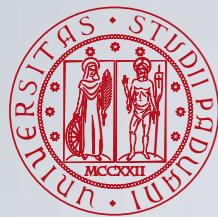


Carol Ann Peters



Nicola Ferro

Ferro, N. and Peters, C., editors (2019). *Information Retrieval Evaluation in a Changing World – Lessons Learned from 20 Years of CLEF*, volume 41 of *The Information Retrieval Series*. Springer International Publishing, Germany.



# Large-scale Evaluation Initiatives: FIRE



- FIRE (Forum for Information Retrieval Evaluation), India, since 2008

- <http://fire.irsi.res.in/>

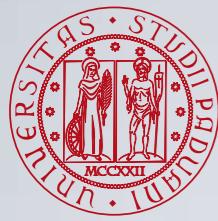
The screenshot shows the homepage of the FIRE 2023 website. At the top left is the logo of the INFORMATION RETRIEVAL SOCIETY OF INDIA (IRSI). The main title "FIRE 2023" is prominently displayed, followed by "Forum for Information Retrieval Evaluation" and the location "Goa University, Panjim". The date "15th-18th December" is also mentioned. Below the title is a black and white illustration of a city skyline featuring several prominent buildings, including a church with a tall tower. To the right of the skyline are social media icons for Facebook, Twitter, and Google+. On the left side of the page is a vertical sidebar with links to "Home", "Call for Papers", "Call for Tracks", "Call for Doctoral Consortium", "Call for Industry Track", "Call for Tutorial", "Organization", "The Journey!", "Contact Us", and "ACM India PhD Clinic". At the bottom of this sidebar is a link to "Archives". The main content area features a "Welcome" section with text about the conference's history and goals, and a cartoon illustration of a person working on a large document labeled "FIRE" while surrounded by books and a magnifying glass. To the right of the welcome section is a "SPONSORS" section listing "acm In-Cooperation", "SIGIR Special Interest Group on Information Retrieval", and "ACM SIGIR". Below this is a "PUBLICATIONS" section with the text "To be announced soon."



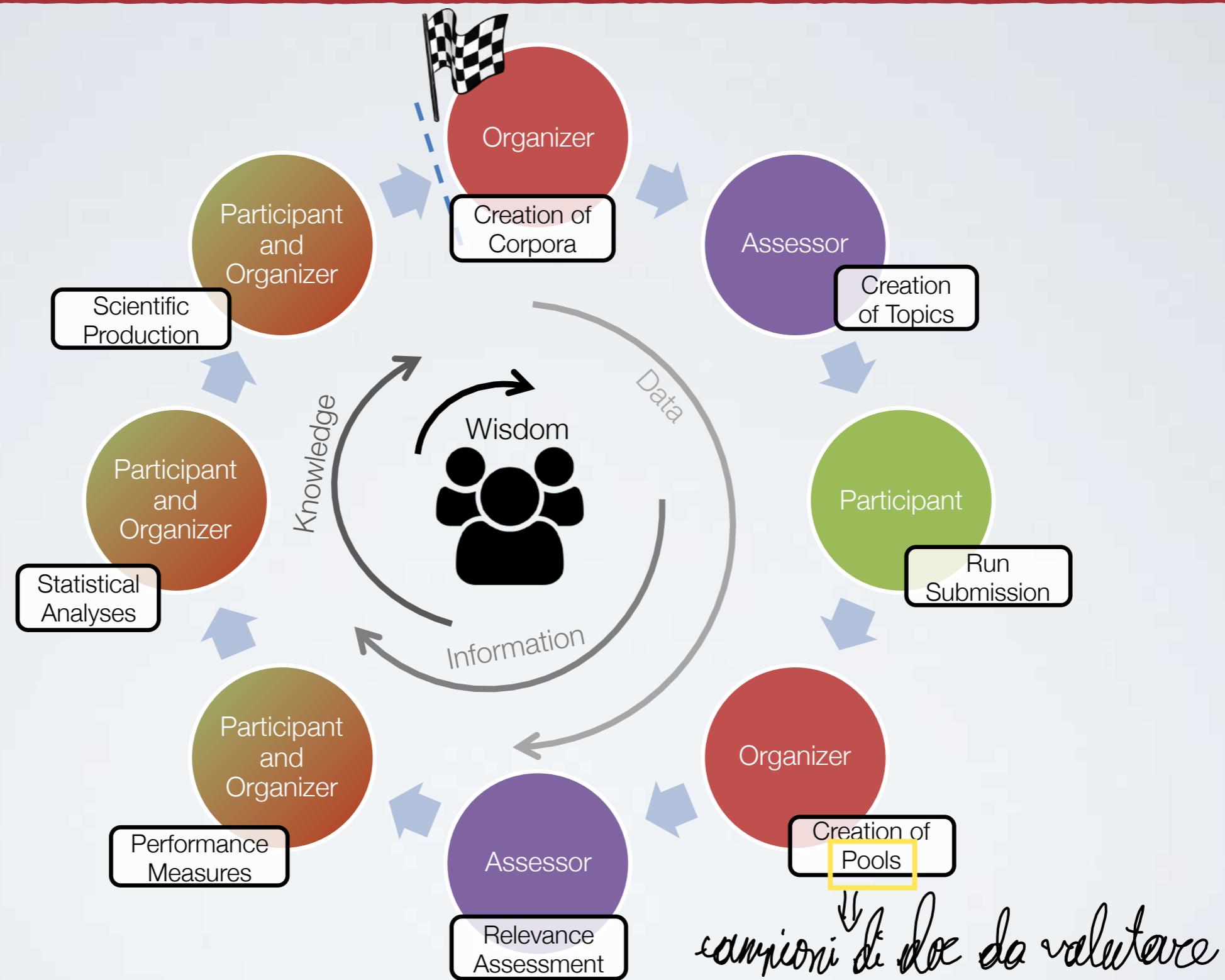
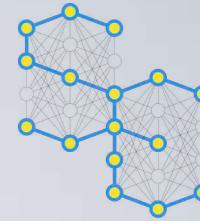
Mandar Mitra



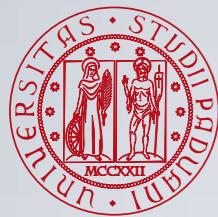
Prasenjit Majumder



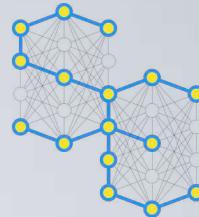
# Evaluation Initiatives: Typical Cycle



Dussin, M. and Ferro, N. (2009). Managing the Knowledge Creation Process of Large-Scale Evaluation Campaigns. In Agosti, M., Borbinha, J., Kapidakis, S., Papatheodorou, C., and Tsakonas, G., editors, Proc. 13th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2009), pages 63–74. LNCS 5714, Springer, Germany.



# The “Ideal Test Collection” Today



**Corpora** → (not historical) corpora are typically OK

- < 500 documents, no real value
- 1-2,000 documents, minimally acceptable
- > 10,000 documents, actually needed  $\Rightarrow$  *molti per l'epoca*



**Topics** → typical size is still 50 topics

- < 75 topics, no real value
- 250 topics, minimally acceptable
- > 1,000 topics, actually needed



**Relevance Judgements** → binary is still most common

option, diversity only recently

- multi-graded (highly and fairly relevant)
- types (novel, stimulating, ...)
- need for **pooling** (still open research issue)

*man mano si  
accumulano*



Karen Spärck Jones

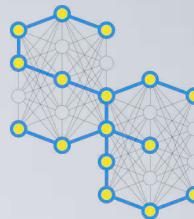


C. J. “Keith” van Rijsbergen

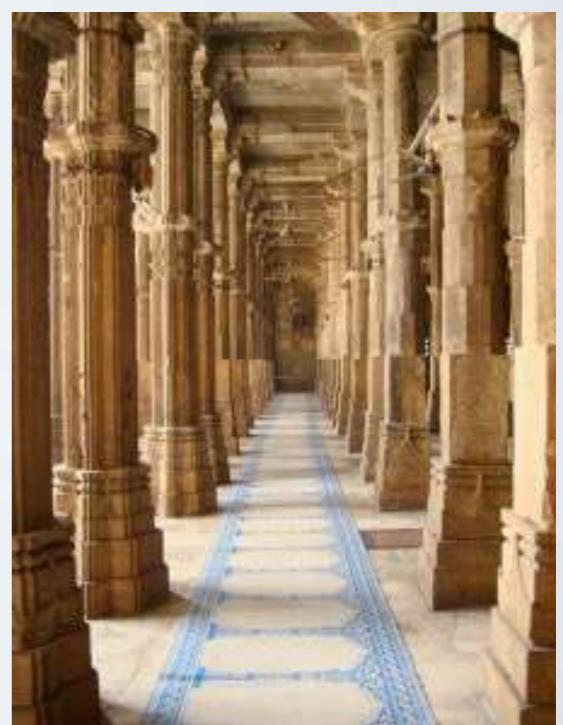
Spärck Jones, K. and van Rijsbergen, C. J. (1975). *Report on the need for and provision of an ‘ideal’ information retrieval test collection*. British Library Research and Development Report 5266, University Computer Laboratory, Cambridge.



# How Valuable is Evaluation?



- The TREC 2010 Economic Impact study estimated in about **30 M\$** the overall **investment in TREC** by NIST
- probably much much more if we had a means to estimate also the investment by participants in TREC
- They are the **pillars** for all the subsequent **scientific research** and **technology development**
- TREC estimated the **return on investment** in the range of **3\$-5\$** for each invested dollar



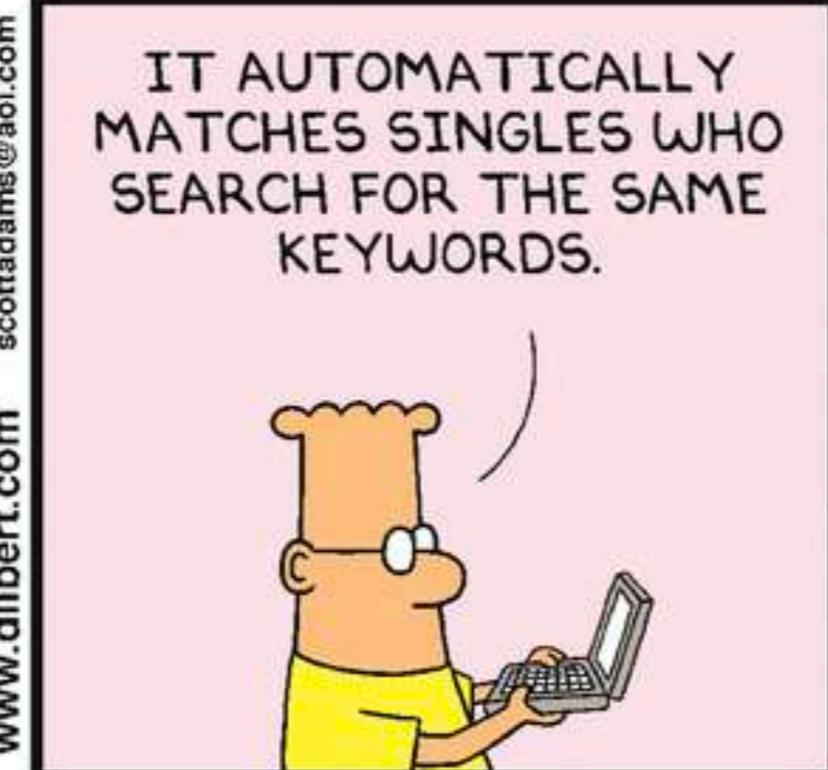
Rowe, B. R., Wood, D. W., Link, A. L., and Simoni, D. A. (2010). *Economic Impact Assessment of NIST's Text REtrieval Conference (TREC) Program*. RTI Project Number 0211875, RTI International, USA. <http://trec.nist.gov/pubs/2010.economic.impact.pdf>.

# questions?

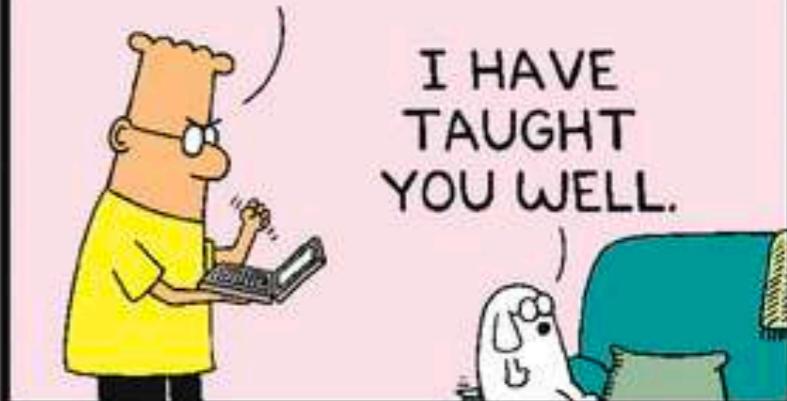
I INVENTED A SEARCH  
ENGINE THAT'S ALSO A  
SINGLES MATCHING  
SERVICE.



IT AUTOMATICALLY  
MATCHES SINGLES WHO  
SEARCH FOR THE SAME  
KEYWORDS.



AND THEN IT THREAT-  
ENS TO E-MAIL THOSE  
KEYWORDS TO THEIR  
MOTHERS IF THEY DON'T  
AGREE TO DATE.



scottadams@aol.com

[www.dilbert.com](http://www.dilbert.com)

5-13-06 © 2005 Scott Adams, Inc./Dist. by UFS, Inc.

I HAVE  
TAUGHT  
YOU WELL.