

67-300 SEARCH ENGINES

---

# IR EVALUATION 2

LECTURER: JOAO PALOTTI ([JPALOTTI@ANDREW.CMU.EDU](mailto:JPALOTTI@ANDREW.CMU.EDU))  
3RD APRIL 2017

---

# RECAP

## PRECISION AND RECALL

**Confusion Matrix**

		Relevant	Non Relevant
		true positives (tp)	false positive (fp)
Retrieved	Not Retrieved	false negatives (fn)	true negative (tn)

- ▶ **Precision:** Fraction of retrieved docs that are relevant

$$(tp) / (tp + fp)$$

- ▶ **Recall:** Fraction of relevant docs that are retrieved

$$(tp) / (tp + fn)$$

# PRECISION AND RECALL

- ▶ Can we go some lessons back to define what we have learnt in terms of precision and recall?
- ▶ Boolean:
  - ▶ OR operator
  - ▶ AND operator

# PRECISION AND RECALL

- ▶ Can we go some lessons back to define what we have learnt in terms of precision and recall?
- ▶ Boolean:
  - ▶ OR operator
  - ▶ AND operator
- ▶ Query expansion with synonyms

# PRECISION AND RECALL

- ▶ Can we go some lessons back to define what we have learnt in terms of precision and recall?
- ▶ Boolean:
  - ▶ OR operator
  - ▶ AND operator
- ▶ Query expansion with synonyms
- ▶ Query expansion with Pseudo Relevance Feedback

# PRECISION AND RECALL

- ▶ Can we go some lessons back to define what we have learnt in terms of precision and recall?
- ▶ Boolean:
  - ▶ OR operator
  - ▶ AND operator
- ▶ Query expansion with synonyms
- ▶ Query expansion with Pseudo Relevance Feedback

**Do Web searchers care for recall?**

# METRICS TO SUMMARIZE PERFORMANCE

- ▶ Binary relevance:
  - ▶ Eleven-point interpolated average precision
  - ▶ Precision@K (P@K)
  - ▶ R-Prec ("P@R")
  - ▶ Mean Average Precision (MAP)
  - ▶ Mean Reciprocal Rank (MRR)
- ▶ Multiple grades of relevance
  - ▶ Normalized Discounted Cumulative Gain (NDCG)

## AT LARGE SEARCH ENGINES

- ▶ Many of the metrics studied in special NDCG
- ▶ But also:
  - ▶ User studies at lab
  - ▶ A/B testing
  - ▶ Result Interleaving

# CRANFIELD EXPERIMENTS GAVE BIRTH TO MODERN IR EVALUATION

- ▶ Cleverdon proposed that everyone would use the same data in an annual meeting in Cranfield University
- ▶ With the same dataset we can fairly compare methods from universities A and B
- ▶ Born the Cranfield Experiments / Cranfield Paradigm

# SONS OF CRANFIELD EXPERIMENTS

- ▶ TREC (Text REtrieval Conference)
- ▶ CLEF (Conference and Labs of the Evaluation Forum)
- ▶ NTCIR (NII Testbeds and Community for Information access Research)
- ▶ FIRE (Forum Information REvaluation)
  
- ▶ TrecVid (TREC Video REvaluation)
- ▶ Mirex (Music Information REvaluation eXchange)
- ▶ MediaEval (Benchmarking Initiative for Multimedia Evaluation)

# TREC (TEXT RETRIEVAL CONFERENCE)

- ▶ Started in 1992, hosted by NIST (National Institute of Standards and Technology)
- ▶ Takes place every year in November in the same city (Gaithersburg, Maryland, USA)
- ▶ Most important benchmarks for IR Studies

# FIRST TREC IN 1992 (TREC 1)

- [http://trec.nist.gov/pubs/trec1/t1\\_proceedings.html](http://trec.nist.gov/pubs/trec1/t1_proceedings.html)
- ▶ In the 30 or 50 years of experimentation there have been two missing elements. First, although some research groups have used the same collections, there has been no concerted effort by groups to work with the same data, use the same evaluation techniques, and generally compare results across systems. The importance of this is not to show any system to be superior, but to allow comparison across a very wide variety of techniques
- ▶ The second missing element, which has become critical in the last 10 years, is the lack of a realistically sized test collection.

# FIRST TREC IN 1992 (TREC 1)

- [http://trec.nist.gov/pubs/trec1/t1\\_proceedings.html](http://trec.nist.gov/pubs/trec1/t1_proceedings.html)
- ▶ In the 30 or 50 years of experimentation there have been two missing elements. First, although some research groups have used the same collections, there has been no concerted effort by groups to work with the same data, use the same evaluation techniques, and generally compare results across systems. The importance of this is not to show any system to be superior, but to allow comparison across a very wide variety of techniques
- ▶ The second missing element, which has become critical in the last 10 years, is the lack of a realistically sized test collection.

THIS FIELD KNOWS WHAT BIG DATA MEANS FOR MORE THAN 25 YEARS....

# DATA COLLECTION

- ▶ About 2 GB:
  - ▶ Wall Street Journal (1986 – 1992)
  - ▶ AP Newswire
  - ▶ Federal register
  - ▶ Abstract from the Department of Energy

# SGML DOCUMENTS IN THE COLLECTION

<DOC>

<DOCNO> WSJ880406-0090 <DOCNO>

<HL> AT&T Unveils Services to Upgrade Phone Networks Under Global Plan <HL>

<AUTHOR> Janet Guyon (WSJ Staff) <AUTHOR>

<DAmLINE> NEW YORK <DAmLINE>

<TEXT>

American Telephone & Telegraph Co. introduced the first of a new generation of phone services with broad implications for computer and communications equipment markets.

AT&T said it is the first national long Distance carrier to announce prices for specific services under a world-wide standardization plan to upgrade phone networks. By announcing commercial services under the plan, which the industry calls the Integrated Services Digital Network, AT&T will influence evolving communications standards to its advantage, consultants said, just as International Business Machines Corp. has created de facto computer standards favoring its products.

<TEXT>

<DOC>

# INFORMATION NEEDS (TOPICS / QUERIES)

<top>

<head> Tipster Topic Description

<num> Number: 066

<dom> Domain: Science and Technology

<title> Topic: Natural Language Processing

<desc> Description:

Document will identify a type of natural language processing technology which is being developed or marketed in the U.S.

<narr> Narrative:

A relevant document will identify a company or institution developing or marketing a natural language processing technology, identify the technology, and identify one or more features of the company's product.

<con> Concept(s):

1. natural language processing
2. translation, language, dictionary, font
3. software applications

<fac> Factor(s):

<nat> Nationality: U.S.

<fac>

<def> Definition(s):

<top>

<top>

<head> Tipster Topic Description

<num> Number: 052

<dom> Domain: International Economics

<title> Topic: South African Sanctions

<desc> Description: Document discusses sanctions against South Africa.

<smry> Summary: Document discusses international sanctions against South Africa.

<narr> Narrative: A relevant document will discuss any aspect of South African sanctions, such as: sanctions declared/proposed by a country against the South African government in response to its apartheid policy, or in response to pressure by an individual, organization or another country; international sanctions against Pretoria imposed by the United Nations; the effects of sanctions against S. Africa; opposition to sanctions; or, compliance with sanctions by a company. The document will identify the sanctions instituted or being considered, e.g., corporate disinvestment, trade ban, academic boycott, arms embargo.

<con> Concept(s):

1. sanctions, international sanctions, economic sanctions
2. corporate exodus, corporate disinvestment, stock divestiture, ban on new investment, trade ban, import ban on South African diamonds, U.N. arms embargo, curtailment of defense contracts, cutoff of nonmilitary goods, academic boycott, reduction of cultural ties
3. apartheid, white domination, racism
4. antiapartheid, black majority rule
5. Pretoria

<fac> Factor(s):

<nat> Nationality: South Africa

</fac>

</top>

# EVALUATION

- ▶ 50 topics were created like these ones
- ▶ Let's say there is 1 million documents in the collection, how can we judge every document in the collection with respect to these 50 topics? 50 million judgements?

# EVALUATION

- ▶ 50 topics were created like these ones
- ▶ Let's say there is 1 million documents in the collection, how can we judge every document in the collection with respect to these 50 topics? 50 million judgements?

THE VERY FIRST TREC HAD AN  
IMPOSSIBLE CHALLENGE

# POOLING METHOD

- ▶ Every group/university can submit up to X results (runs) with up to Y documents per topic



## Participants

```
<DOC>
<DOCNO> WS.I880406-0090 <DOCNO>
<HL> AT&T Unveils Services to Upgrade Phone Networks Under Global
Plan <HL>
<AUTHOR> Janet Guyon (WSJ Staff) <AUTHOR>
<DAI> NEW YORK <DAI>
<TEXT>
American Telephone & Telegraph Co. introduced the first of a new
generation of phone services with broad implications for computer and
communications equipment markets.
AT&T said it is the first national long Distance carrier to announce prices
for specific services under a world-wide standardization plan to upgrade
phone networks. By announcing commercial services under the plan, which the
industry calls the Integrated Services Digital Network, AT&T will influence
evolving
communications standards to its advantage, consultants said, just as
International Business Machines Corp. has created de facto computer
standards by favoring its products.
<TEXT>
<DOC>
```

## TREC topics

# POOLING METHOD

- ▶ Every group/university can submit up to X results (runs) with up to Y documents per topic



Participants



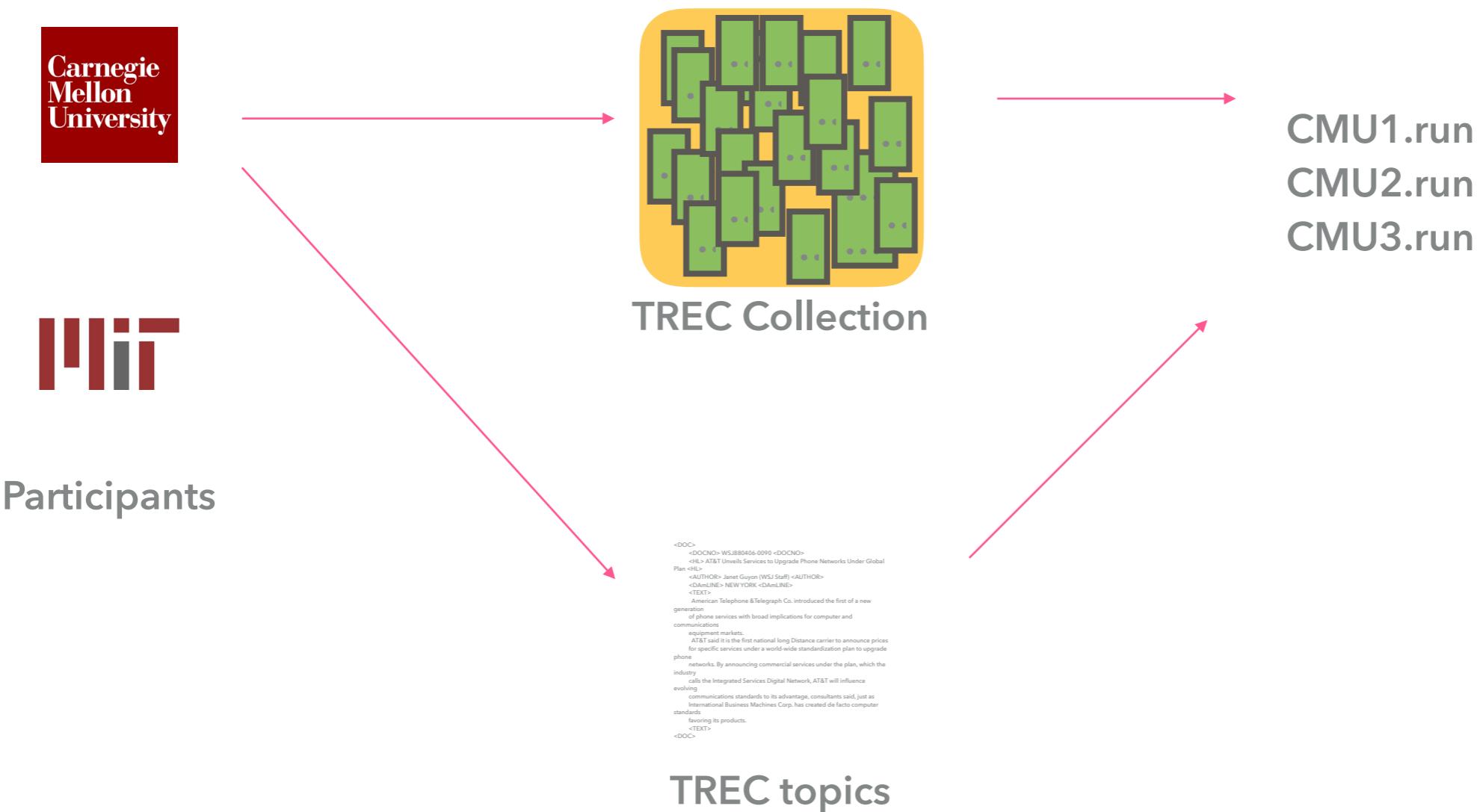
TREC Collection

```
<DOC>
<DOCNO> WS.I880406-0090 <DOCNO>
<HL> AT&T Unveils Services to Upgrade Phone Networks Under Global
Plan <HL>
<AUTHOR> Janet Guyon (WSJ Staff) <AUTHOR>
<DArLINE> NEW YORK <DArLINE>
<TEXT>
American Telephone & Telegraph Co. introduced the first of a new
generation of phone services with broad implications for computer and
communications equipment markets.
AT&T said it is the first national long Distance carrier to announce prices
for specific services under a world-wide standardization plan to upgrade
phone networks. By announcing commercial services under the plan, which the
industry calls the Integrated Services Digital Network, AT&T will influence
evolving
communications standards to its advantage, consultants said, just as
International Business Machines Corp. has created de facto computer
standards by favoring its products.
<TEXT>
<DOC>
```

TREC topics

# POOLING METHOD

- Every group/university can submit up to X results (runs) with up to Y documents per topic



# POOLING METHOD

- Every group/university can submit up to X results (runs) with up to Y documents per topic



Participants



```
<DOC>
<DOCNO> WS.I880406-0090 <DOCNO>
<HL> AT&T Unveils Services to Upgrade Phone Networks Under Global
Plan <HL>
<AUTHOR> Janet Guyon (WSJ Staff) <AUTHOR>
<DArLINE> NEW YORK <DArLINE>
<TEXT>
American Telephone & Telegraph Co. introduced the first of a new
generation of phone services with broad implications for computer and
communications equipment markets.
AT&T said it is the first national long Distance carrier to announce prices
for specific services under a world-wide standardization plan to upgrade
phone networks. By announcing commercial services under the plan, which the
industry calls the Integrated Services Digital Network, AT&T will influence
evolving
communications standards to its advantage, consultants said, just as
International Business Machines Corp. has created de facto computer
standards by favoring its products.
<TEXT>
<DOC>
```

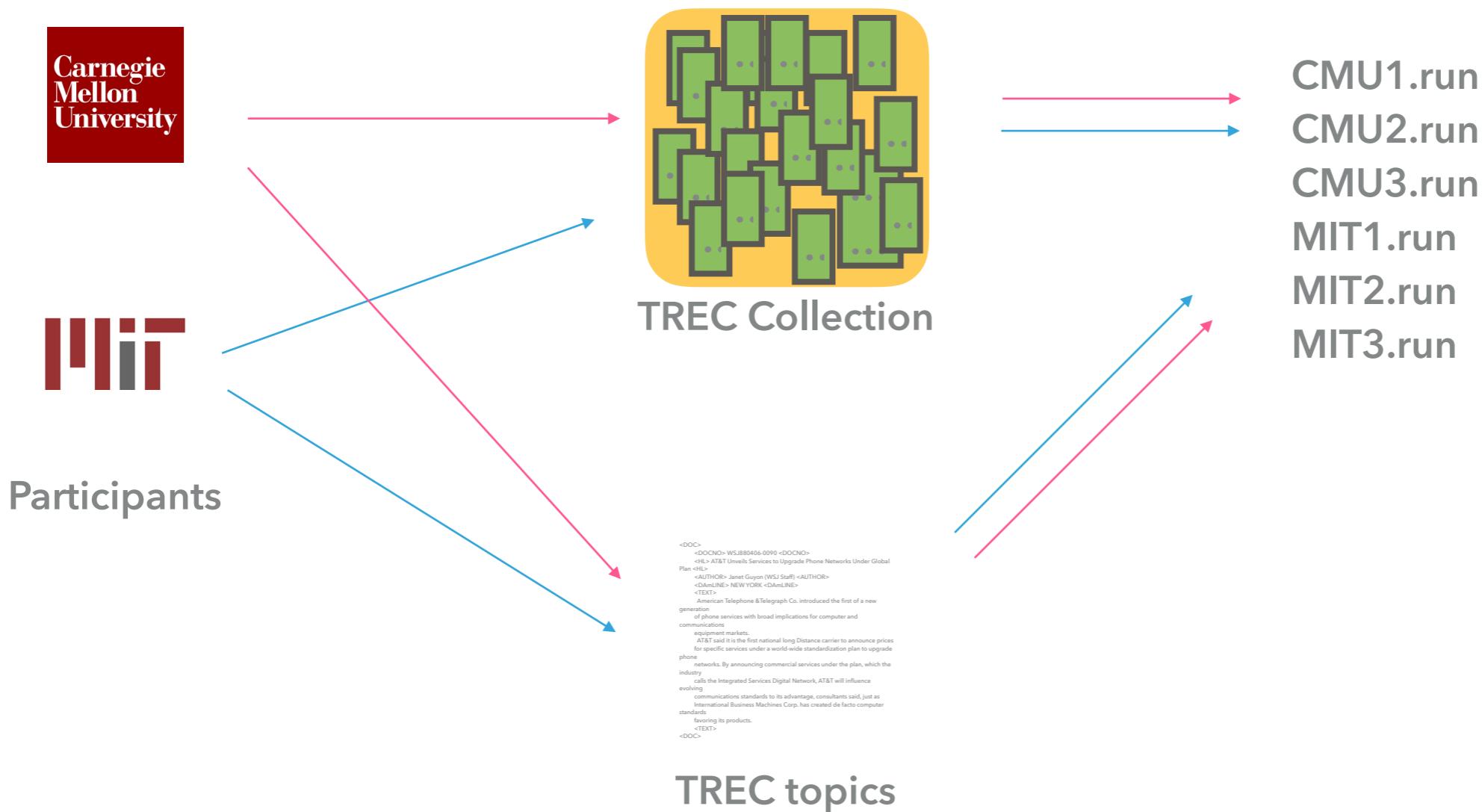
TREC topics

CMU1.run  
CMU2.run  
CMU3.run

**we call 'system'  
a variant of your  
search engine**

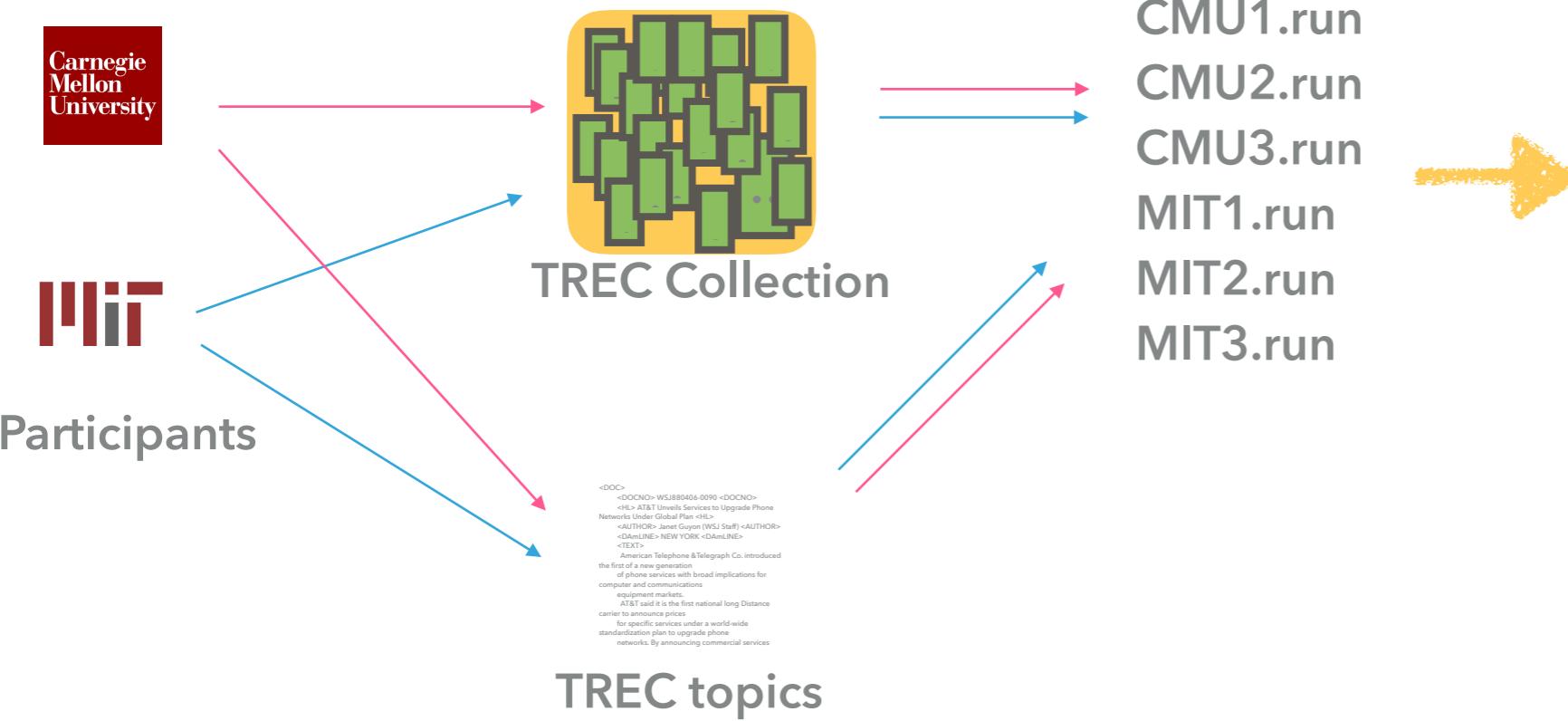
# POOLING METHOD

- Every group/university can submit up to X results (runs) with up to Y documents per topic



# POOLING METHOD

- Every group/university can submit up to X results (runs) with up to Y documents per topic



### Topic 1:

WSJ220409-1233  
WSJ251206-0045  
WSJ298506-0094  
WSJ353256-0460  
WSJ450878-5690  
WSJ888479-0090

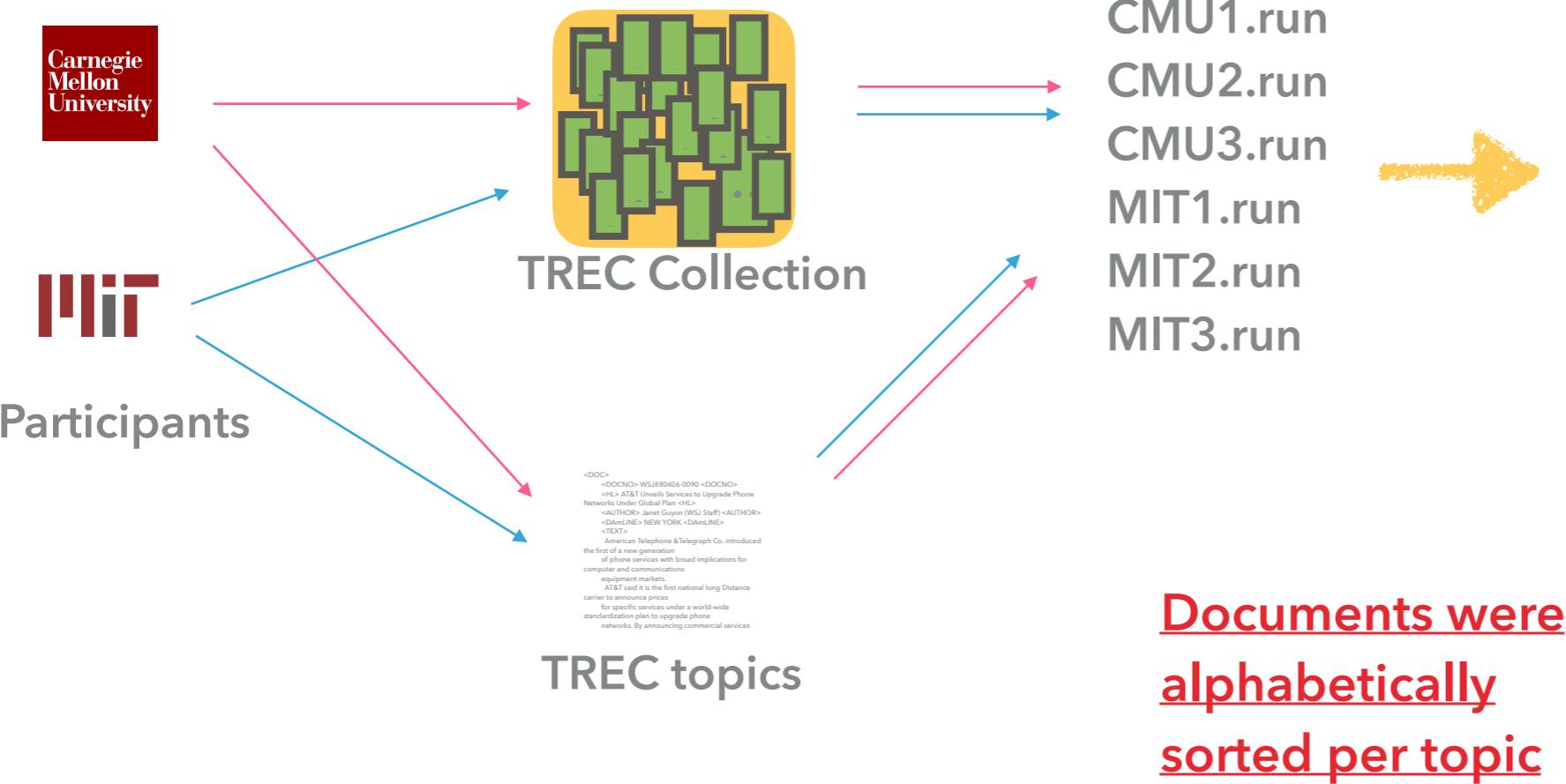
### Topic 2:

WSJ000006-7892  
WSJ050454-0789  
WSJ590656-4658  
WSJ890408-0165

POOL

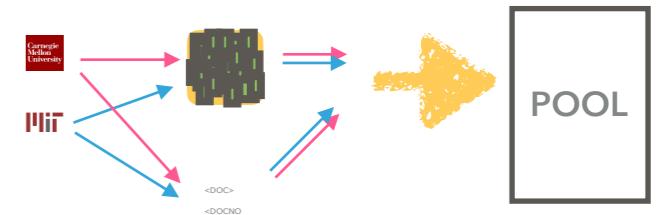
# POOLING METHOD

- Every group/university can submit up to X results (runs) with up to Y documents per topic



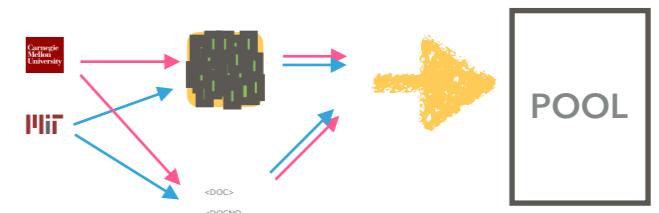
POOL

# GREAT SUCCESS!



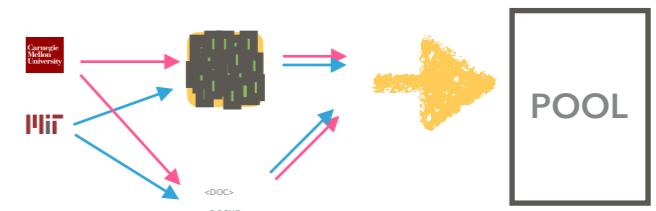
- ▶ 25 universities/industry groups participated
- ▶ Limited number of runs per group (w. 3 runs – total 75 runs)
- ▶ 50 topics
- ▶ Top 200 documents per topic

# GREAT SUCCESS!



- ▶ 25 universities/industry groups participated
- ▶ Limited number of runs per group (w. 3 runs – total 75 runs)
- ▶ 50 topics
- ▶ Top 200 documents per topic
- ▶ Need to judge up to  $25 * 3 * 50 * 200 = 750.000$  (0.75 M)
- ▶ Instead of 50.000.000 (50 M) judgements

# GREAT SUCCESS!



- ▶ 25 universities/industry groups participated
- ▶ Limited number of runs per group (w. 3 runs – total 75 runs)
- ▶ 50 topics
- ▶ Top 200 documents per topic

Why is this the upper bound limit?

- ▶ Need to judge up to  $25 * 3 * 50 * 200 = 750.000$  (0.75 M)
- ▶ Instead of 50.000.000 (50 M) judgements

# THERE ARE OTHER WAYS TO DO POOLING

- ▶ The simplest way (and most used) is to cut at depth-K for each run
- ▶ Other approaches less used or currently being studied:
  - ▶ Stratified Pooling:
    - ▶ Define a policy: 100% of top 50, 20% top 50-200, 1% top 200-500
  - ▶ Active Learning:
    - ▶ Rank documents according to their power to differentiate systems, judge in a pre-defined order according to your budget
  - ▶ Voting schema:
    - ▶ For each topic, judge top X documents according to a voting schema

112 - Documents should discuss ethical issues involved with organ donation, and that any reports that highlight treatment differences in patients who are known organ donors.

[Back to Query](#)[Prev Document](#)

## Document "clueweb12-0005wb-45-41384" (21 / 500)

[Next Document](#)

n  
MATION

[Donation Laws](#)

ee on  
tion

[Reports](#)

[islative History](#)

[al Events](#)

# Advisory Committee on Organ Transplantation

Hilton Rockville Hotel

Rockville, MD

May 15-16, 2007

May 15, 2007

Ms. Gail Agrawal, Chairperson, welcomed the group and read a letter from Larry Hagman, whose last meeting this was to be. Mr. Hagman's letter described his inability to attend due to family obligations and expressed how much he has enjoyed his ACOT membership.

Mr. Remy Aronoff, Executive Secretary, introduced Dr. Russ Wiesner, the newest member, from the Mayo Clinic. Mr. Aronoff thanked ACOT members for the work that everyone has done between meetings and noted that this will be a very full agenda.

### Public Solicitation of Donors – Dr. Douglas Hanto

There are important ethical aims in distributing scarce health care: maximizing the benefits from a limited resource (utility), and distributing the resource fairly or equitably among all possible recipients (justice). These two goals will often conflict, however, and trade-offs will be necessary, but they must be kept in mind. This summary comes from Dr. Dan Brock of Harvard Medical School, who is currently working with the government in Thailand, looking at ways to provide limited dialysis in that country.

Organ donation can be thought of along a spectrum that spans from altruism, solicitation, incentives, coercion and theft. The concept of “sales” can occur along this spectrum in multiple places, depending on your position regarding the ethics of sales. Everyone is familiar with billboards that have been used to advertise for a directed liver donation to that individual.

Decreased organ donation allocation: organs are allocated in a non-directed fashion to

### Judgement

- Unjudged
- Highly relevant
- Somewhat relevant
- Not relevant

[Save](#)[Save & Next](#)

112 - Documents should discuss ethical issues involved with organ donation, and that any reports that highlight treatment differences in patients who are known organ donors.

[Back to Query](#)[Prev Document](#)[Next Document](#)

## Document "clueweb12-0005wb-45-41384" (21 / 500)

n  
MATION

[Donation Laws](#)

ee on  
tion

[Reports](#)

[islative History](#)

[al Events](#)

# Advisory Committee on Organ Transplantation

Hilton Rockville Hotel

Rockville, MD

May 15-16, 2007

May 15, 2007

Ms. Gail Agrawal, Chairperson, welcomed the group and read a letter from Larry Hagman, whose last meeting this was to be. Mr. Hagman's letter described his inability to attend due to family obligations and expressed how much he has enjoyed his ACOT membership.

Mr. Remy Aronoff, Executive Secretary, introduced Dr. Russ Wiesner, the newest member, from the Mayo Clinic. Mr. Aronoff thanked ACOT members for the work that everyone has done between meetings and noted that this will be a very full agenda.

### Public Solicitation of Donors – Dr. Douglas Hanto

There are important ethical aims in distributing scarce health care: maximizing the benefits from a limited resource (utility), and distributing the resource fairly or equitably among all possible recipients (justice). These two goals will often conflict, however, and trade-offs will be necessary, but they must be kept in mind. This summary comes from Dr. Dan Brock of Harvard Medical School, who is currently working with the government in Thailand, looking at ways to provide limited dialysis in that country.

Organ donation can be thought of along a spectrum that spans from altruism, solicitation, incentives, coercion and theft. The concept of “sales” can occur along this spectrum in multiple places, depending on your position regarding the ethics of sales. Everyone is familiar with billboards that have been used to advertise for a directed liver donation to that individual.

Decreased organ donation allocation: organs are allocated in a non-directed fashion to

### Judgement

- Unjudged
- Highly relevant
- Somewhat relevant
- Not relevant

[Save](#)[Save & Next](#)

112 - Documents should discuss ethical issues involved with organ donation, and that any reports that highlight treatment differences in patients who are known organ donors.

[Back to Query](#)[Prev Document](#)[Next Document](#)

## Document "clueweb12-0005wb-45-41384" (21 / 500)

n  
MATION

[Donation Laws](#)

ee on  
tion

[Reports](#)

[islative History](#)

[al Events](#)

# Advisory Committee on Organ Transplantation

Hilton Rockville Hotel

Rockville, MD

**May 15-16, 2007**

**May 15, 2007**

Ms. Gail Agrawal, Chairperson, welcomed the group and read a letter from Larry Hagman, whose last meeting this was to be. Mr. Hagman's letter described his inability to attend due to family obligations and expressed how much he has enjoyed his ACOT membership.

Mr. Remy Aronoff, Executive Secretary, introduced Dr. Russ Wiesner, the newest member, from the Mayo Clinic. Mr. Aronoff thanked ACOT members for the work that everyone has done between meetings and noted that this will be a very full agenda.

### Public Solicitation of Donors – Dr. Douglas Hanto

There are important ethical aims in distributing scarce health care: maximizing the benefits from a limited resource (utility), and distributing the resource fairly or equitably among all possible recipients (justice). These two goals will often conflict, however, and trade-offs will be necessary, but they must be kept in mind. This summary comes from Dr. Dan Brock of Harvard Medical School, who is currently working with the government in Thailand, looking at ways to provide limited dialysis in that country.

Organ donation can be thought of along a spectrum that spans from altruism, solicitation, incentives, coercion and theft. The concept of “sales” can occur along this spectrum in multiple places, depending on your position regarding the ethics of sales. Everyone is familiar with billboards that have been used to advertise for a directed liver donation to that individual.

Decreased organ donation allocation: organs are allocated in a non-directed fashion to

### Judgement

- Unjudged
- Highly relevant
- Somewhat relevant
- Not relevant

[Save](#)[Save & Next](#)

112 - Documents should discuss ethical issues involved with organ donation, and that any reports that highlight treatment differences in patients who are known organ donors.

[Back to Query](#)[Prev Document](#)

## Document "clueweb12-0005wb-45-41384" (21 / 500)

[Next Document](#)

n  
MATION

[Donation Laws](#)

ee on  
tion

[Reports](#)

[islative History](#)

[al Events](#)

# Advisory Committee on Organ Transplantation

Hilton Rockville Hotel

Rockville, MD

**May 15-16, 2007**

**May 15, 2007**

Ms. Gail Agrawal, Chairperson, welcomed the group and read a letter from Larry Hagman, whose last meeting this was to be. Mr. Hagman's letter described his inability to attend due to family obligations and expressed how much he has enjoyed his ACOT membership.

Mr. Remy Aronoff, Executive Secretary, introduced Dr. Russ Wiesner, the newest member, from the Mayo Clinic. Mr. Aronoff thanked ACOT members for the work that everyone has done between meetings and noted that this will be a very full agenda.

### Public Solicitation of Donors – Dr. Douglas Hanto

There are important ethical aims in distributing scarce health care: maximizing the benefits from a limited resource (utility), and distributing the resource fairly or equitably among all possible recipients (justice). These two goals will often conflict, however, and trade-offs will be necessary, but they must be kept in mind. This summary comes from Dr. Dan Brock of Harvard Medical School, who is currently working with the government in Thailand, looking at ways to provide limited dialysis in that country.

Organ donation can be thought of along a spectrum that spans from altruism, solicitation, incentives, coercion and theft. The concept of “sales” can occur along this spectrum in multiple places, depending on your position regarding the ethics of sales. Everyone is familiar with billboards that have been used to advertise for a directed liver donation to that individual.

Decreased organ donation allocation: organs are allocated in a non-directed fashion to

### Judgement

- Unjudged
- Highly relevant
- Somewhat relevant
- Not relevant

[Save](#)[Save & Next](#)

112 - Documents should discuss ethical issues involved with organ donation, and that any reports that highlight treatment differences in patients who are known organ donors.

[Back to Query](#)[Prev Document](#)[Next Document](#)

## Document "clueweb12-0005wb-45-41384" (21 / 500)

n  
MATION

[Donation Laws](#)

ee on  
tion

[Reports](#)

[islative History](#)

[al Events](#)

# Advisory Committee on Organ Transplantation

Hilton Rockville Hotel

Rockville, MD

**May 15-16, 2007**

**May 15, 2007**

Ms. Gail Agrawal, Chairperson, welcomed the group and read a letter from Larry Hagman, whose last meeting this was to be. Mr. Hagman's letter described his inability to attend due to family obligations and expressed how much he has enjoyed his ACOT membership.

Mr. Remy Aronoff, Executive Secretary, introduced Dr. Russ Wiesner, the newest member, from the Mayo Clinic. Mr. Aronoff thanked ACOT members for the work that everyone has done between meetings and noted that this will be a very full agenda.

### Public Solicitation of Donors – Dr. Douglas Hanto

There are important ethical aims in distributing scarce health care: maximizing the benefits from a limited resource (utility), and distributing the resource fairly or equitably among all possible recipients (justice). These two goals will often conflict, however, and trade-offs will be necessary, but they must be kept in mind. This summary comes from Dr. Dan Brock of Harvard Medical School, who is currently working with the government in Thailand, looking at ways to provide limited dialysis in that country.

Organ donation can be thought of along a spectrum that spans from altruism, solicitation, incentives, coercion and theft. The concept of “sales” can occur along this spectrum in multiple places, depending on your position regarding the ethics of sales. Everyone is familiar with billboards that have been used to advertise for a directed liver donation to that individual.

Decreased organ donation allocation: organs are allocated in a non-directed fashion to

### Judgement

- Unjudged
- Highly relevant
- Somewhat relevant
- Not relevant

[Save](#)[Save & Next](#)

# JUDGEMENTS

- After judgements are done, a file called Qrels (query relevance set) is created:

<QID, 0, DocNo, Rel>

1 0 WSJ220409-1233 1



Document WSJ220409-1233  
was judged as relevant for  
topic/query 1

1 0 WSJ251206-0045 0

1 0 WSJ298506-0094 1

1 0 WSJ353256-0460 1

1 0 WSJ450878-5690 0



Document WSJ450878-5690  
was judged as non relevant for  
topic/query 1

1 0 WSJ888479-0090 0

# TREC\_EVAL

- ▶ trec\_eval is just a program that implements many of the methods we surveyed last lecture:

[http://trec.nist.gov/trec\\_eval/trec\\_eval\\_latest.tar.gz](http://trec.nist.gov/trec_eval/trec_eval_latest.tar.gz)

- ▶ Homework 3 is about running it:
  - ▶ `./trec_eval <qrel> <your_run>`
  - ▶ Writing your findings in a documentation

Metric	Description
num_ret	Total number of documents retrieved over all queries
num_rel	Total number of relevant documents over all queries
num_rel_ret	Total number of relevant documents retrieved over all queries
map	Mean Average Precision (MAP)
gm_ap	Average Precision. Geometric Mean, $q\_score = \log(\text{MAX}(\text{map}, .00001))$
R-prec	R-Precision (Precision after R (= num-rel for topic) documents retrieved)
bpref	Binary Preference, top R judged nonrel
recip_rank	Reciprocal rank of top relevant document
ircl_prn.0.00	
ircl_prn.0.10	
ircl_prn.0.20	
...	Interpolated Recall Precision Averages at X recall
ircl_prn.0.90	
ircl_prn.1.00	
P5	
P10	
P15	
...	Precision after X docs retrieved
P500	
P1000	

# TREC DOCUMENTATION

- ▶ Vector Space Model using a cosine distance measure.  
Furthermore, the vector space does not fix the document vector length component of the cosine distance formula. Rather, the length of any document vector is allowed to vary depending on the terms present in the query; only the terms in the query vector are considered to be 'active' for any given distance calculation, and all other terms in the document are ignored. Since query vectors may include zero-coefficient distractor terms, each query can specify its own list of 'active terminology', thereby allowing query-specific control of the document representation.

Extract from CMU's paper at TREC4 (1996)

<http://trec.nist.gov/pubs/trec4/appendices/B/sysdes/done.sysdes.clarit.new>

# TREC EVOLUTION

- ▶ Initial track is named ad-hoc retrieval
- ▶ Many other **tracks** appear in later editions of TREC:
  - ▶ Motivated by real problems or challenges
  - ▶ Retrieval in other languages (Spanish, Chinese)
    - ▶ TREC 3: 208 Mb of Mexican newspaper text in the Spanish language (>150k docs)
    - ▶ TREC 5: 170 Mb of Chinese newspapers (>150k docs)
  - ▶ Specific tasks:
    - ▶ Filtering: text classification (e.g. In 2005 participants should classify text as being either Spam or not Spam)

[http://trec.nist.gov/pubs/trec5/papers/dbmerge\\_track.ps.gz](http://trec.nist.gov/pubs/trec5/papers/dbmerge_track.ps.gz)

# TREC EVOLUTION

## ▶ TREC 5 (1996):

### ▶ Database Merging Track:

- Data is distributed over many different sources, goal is to find whatever is relevant across different dataset
- Almost nobody participated (only 3 teams)

[http://trec.nist.gov/pubs/trec5/papers/dbmerge\\_track.ps.gz](http://trec.nist.gov/pubs/trec5/papers/dbmerge_track.ps.gz)

# TREC EVOLUTION

## ▶ TREC 5 (1996):

### ▶ Database Merging Track:

- Data is distributed over many different sources, goal is to find whatever is relevant across different dataset
- Almost nobody participated (only 3 teams)

**Can you give two actual examples of this task?**

[http://trec.nist.gov/pubs/trec5/papers/dbmerge\\_track.ps.gz](http://trec.nist.gov/pubs/trec5/papers/dbmerge_track.ps.gz)

# TREC EVOLUTION

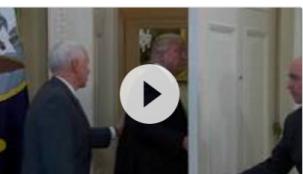
- ▶ TREC 5 (1996):
  - ▶ Database Merging Track:
    - Data is distributed over many different sources, goal is to find whatever is relevant across different dataset
    - Almost nobody participated (only 3 teams)

**Can you give two actual examples of this task?**

- 1 - Metasearch engines
- 2 - Vertical search engines

About 961,000,000 results (0.96 seconds)

## Top stories

[Trump leaves executive order ceremony without signing](#)

CNN · 1 hour ago

[→ More for trump](#)[Donald Trump's train-wreck presidency](#)

Los Angeles Times · 2 hours ago

[Judge rejects Trump's defense against claim he incited violence at rally](#)

The Guardian · 5 hours ago

**Donald Trump**

45th U.S. President

 [donaldjtrump.com](#)Donald John T  
politician, and

Born: June 14,

Height: 6' 2"

Spouse: [Melar  
Trump](#) (m. 197Children: [Ivank  
Barron Trump](#)Education: Wh  
MoreParents: [Fred](#) 1

trump

[Web](#)[Images](#)[News](#)

18,300,000 RESULTS

Date ▾

Language ▾

Region ▾

**Trump Organization - Official Site**[www.trump.com](#) ▾

Trump Luxury Real Estate redefines what is meant by luxury living, built to be the absolute best in the world.

**Donald J. Trump - Official Site**[www.donaldjtrump.com](#) ▾

Donald J. Trump is the very definition of the American success story, continually setting the standards of excellence in business, real estate and entertainment.

**Trump - YouTube**[www.youtube.com/user/TrumpSC](#) ▾

Trump puts his namesake to the test by strategizing, being ruthlessly efficient, making value plays and thus trumping the competition! By optimizing and push...

**Images of trump**[bing.com/images](#)[See more images of trump](#)**Donald J. Trump (@realDonaldTrump) | Twitter**[https://twitter.com/realdonaldtrump](#) ▾

Verified · 27M followers · 35K tweets

34.7K tweets · 1,998 photos/videos · 27M followers. Check out the latest Tweets from Donald J. Trump (@realDonaldTrump)

**Donald J. Trump (@realDonaldTrump) · Twitter**[https://twitter.com/realdonaldtrump](#)

Thank you @JCLayfield – will get even better as my Administration continues to put #AmericaFirst

[pic.twitter.com/AQQzmt1...](#)

14 hours ago · Twitter

...not associated with Russia. Trump team spied on before he was nominated." If this is true, does not get much bigger. Would be sad for U.S.

19 hours ago · Twitter

Wow, @FoxNews just reporting big news. Source: "Official behind unmasking is high up. Known Intel official is responsible. Some unmasked...."

20 hours ago · Twitter

## Profiles



Twitter

## People also



Hillary Clinton

[http://trec.nist.gov/pubs/trec6/papers/hp\\_track.ps.gz](http://trec.nist.gov/pubs/trec6/papers/hp_track.ps.gz)

# TREC EVOLUTION

- ▶ TREC 6 (1997):
  - ▶ High-precision Track:
    - The TREC High Precision (HP) track compares systems on the simple “Real World” task of having users finding a few relevant documents **as quickly as possible.**

[http://trec.nist.gov/pubs/trec6/papers/hp\\_track.ps.gz](http://trec.nist.gov/pubs/trec6/papers/hp_track.ps.gz)

# TREC EVOLUTION

- ▶ TREC 6 (1997):
  - ▶ High-precision Track:
    - The TREC High Precision (HP) track compares systems on the simple “Real World” task of having users finding a few relevant documents **as quickly as possible.**

## Track Description

The High-Precision Track is a new track for TREC. It has a very simple short description: for each query, a user should find the best 10 documents possible within 5 minutes clock time. One realistic scenario corresponding to this task might be that your boss asks you for a quick report on some area and you need to find some information on the area fast.

<http://trec.nist.gov/pubs/trec6/papers/sdr97.ps.gz>

# TREC EVOLUTION

- ▶ Also TREC 6 (1997):
  - ▶ Spoken Document Retrieval Track:
- ▶ Two different source of data:
  - ▶ “Easy” Speech: native speaker of American English, quite conditions, high fidelity recording
  - ▶ “Difficult” Speech: lower quality records, spontaneous speech, non-native speaker, noisy conditions
- ▶ For a topic like: “*Does the Olympic torch ever travel by motorcycle?*”, the goal is finding the point (minute) in the record (type) that a relevant story starts

<http://trec.nist.gov/pubs/trec6/papers/sdr97.ps.gz>

# TREC EVOLUTION

- ▶ Also TREC 6 (1997):
  - ▶ Spoken Document Retrieval Track:
- ▶ Two different source of data:
  - ▶ “Easy” Speech: native speaker of American English, quite conditions, high fidelity recording
  - ▶ “Difficult” Speech: lower quality records, spontaneous speech, non-native speaker, noisy conditions
- ▶ For a topic like: “*Does the Olympic torch ever travel by motorcycle?*”, the goal is finding the point (minute) in the record (type) that a relevant story starts

[http://trec.nist.gov/pubs/trec7/papers/CLIROverview\\_Trec7L.pdf.gz](http://trec.nist.gov/pubs/trec7/papers/CLIROverview_Trec7L.pdf.gz)

# TREC EVOLUTION

- ▶ TREC 7 (1998)
  - ▶ Cross-Language Information Retrieval:
  - ▶ Queries in many languages, documents in many languages (English, German, French, Italian)
  - ▶ Issue your query in German/French/Italian: retrieve documents in any language (or in your language + English)

[http://trec.nist.gov/pubs/trec7/papers/CLIROverview\\_Trec7L.pdf.gz](http://trec.nist.gov/pubs/trec7/papers/CLIROverview_Trec7L.pdf.gz)

# TREC EVOLUTION

- ▶ TREC 7 (1998)
  - ▶ Cross-Language Information Retrieval:
  - ▶ Queries in many languages, documents in many languages (English, German, French, Italian)
  - ▶ Issue your query in German/French/Italian: retrieve documents in any language (or in your language + English)

Do you think this is useful? Where? How?

<http://trec.nist.gov/pubs/trec8/papers/qtrack.pdf>

# TREC EVOLUTION

- ▶ TREC 8 (1999)
- ▶ Question Answering Track (sample questions)

- How many calories are there in a Big Mac?
- What two US biochemists won the Nobel Prize in medicine in 1992?
- Who was the first American in space?
- Who is the voice of Miss Piggy?
- Where is the Taj Mahal?
- What costume designer decided that Michael Jackson should only wear one glove?
- In what year did Joe DiMaggio compile his 56-game hitting streak?
- What language is commonly used in Bombay?
- How many Grand Slam titles did Bjorn Borg win?
- Who was the 16th President of the United States?

<http://trec.nist.gov/pubs/trec8/papers/qtrack.pdf>

## TREC EVOLUTION

- ▶ TREC 8 (1999)
- ▶ Question Answering Track (results)

Run Name	Participant	MRR	# not found
textract9908	Cymfony, Inc.	.660	54
SMUNLP1	Southern Methodist U.	.555	63
attqa50e	AT&T Research	.356	109
IBMDR995	IBM	.319	110
xeroxQA8sC	Xerox Research Centre Europe	.317	111
umdqqa	U. of Maryland	.298	118
MTR99050	MITRE	.281	118
IBMVS995	IBM	.280	120
nttd8qs1	NTT Data Corp.	.273	121
attqa50p	AT&T Research	.261	121
nttd8qs2	NTT Data	.259	120
CRL50	New Mexico State U.	.220	130
INQ634	U. of Massachusetts	.191	140
CRDBASE050	GE/U. of Pennsylvania	.158	148
INQ638	U. of Massachusetts	.126	158
shefinq50	U. of Sheffield	.081	182
shefatt50	U. of Sheffield	.071	184
UIowaQA3	U. of Iowa	.018	188
UIowaQA4	U. of Iowa	.017	193

→ Have you heard of IBM Watson?

- a) Runs with a 50-byte limit on the length of the response.

<http://trec.nist.gov/pubs/trec8/papers/qtrack.pdf>

## TREC EVOLUTION

- ▶ TREC 8 (1999)
- ▶ Question Answering Track (results)

Run Name	Participant	MRR	# not found
textract9908	Cymfony, Inc.	.660	54
SMUNLP1	Southern Methodist U.	.555	63
attqa50e	AT&T Research	.356	109
IBMDR995	IBM	.319	110
xeroxQA8sC	Xerox Research Centre Europe	.317	111
umdqqa	U. of Maryland	.298	118
MTR99050	MITRE	.281	118
IBMVS995	IBM	.280	120
nttd8qs1	NTT Data Corp.	.273	121
attqa50p	AT&T Research	.261	121
nttd8qs2	NTT Data	.259	120
CRL50	New Mexico State U.	.220	130
INQ634	U. of Massachusetts	.191	140
CRDBASE050	GE/U. of Pennsylvania	.158	148
INQ638	U. of Massachusetts	.126	158
shefinq50	U. of Sheffield	.081	182
shefatt50	U. of Sheffield	.071	184
UIowaQA3	U. of Iowa	.018	188
UIowaQA4	U. of Iowa	.017	193

→ Have you heard of IBM Watson?

Task evolved as time went by:

– from finding documents to finding passages...

– from finding passages to structuring answers...

- a) Runs with a 50-byte limit on the length of the response.

<http://trec.nist.gov/pubs/trec9/papers/web9.pdf>

# TREC EVOLUTION

- ▶ TREC 9 (2000)
  - ▶ Web Track (substitute of AdHoc and Very Large Collection Tracks)
  - ▶ Encourages the use of Web structure

## 2 Main Web Task

None of the participants in the TREC-8 Small Web Task, using a two gigabyte corpus (WT2g), managed to demonstrate any benefit whatever from using hyperlink methods in that particular retrieval task. Given that most commercial Web search engines exploit hyperlinks to apparently great effect, this result may seem surprising.

Accordingly, a new task was devised for TREC-9 which removed possible impediments to good link-based performance which were perceived to be inherent in last year's task.

<http://trec.nist.gov/pubs/trec9/papers/web9.pdf>

# TREC EVOLUTION

- ▶ TREC 9 (2000)
  - ▶ Web Track (substitute of AdHoc and Very Large Collection Tracks)
  - ▶ Encourages the use of Web structure

## 2 Main Web Task

None of the participants in the TREC-8 Small Web Task, using a two gigabyte corpus (WT2g), managed to demonstrate any benefit whatever from using hyperlink methods in that particular retrieval task. Given that most commercial Web search engines exploit hyperlinks to apparently great effect, this result may seem surprising.

Accordingly, a new task was devised for TREC-9 which removed possible impediments to good link-based performance which were perceived to be inherent in last year's task.

GOOGLE WAS A FEW YEARS OLD AT THIS TIME

# TREC EVOLUTION

- ▶ TREC 2001
  - ▶ Video Track
  - ▶ True multimedia territory (video, audio, transcripts, text in video, music, natural sound...)
  - ▶ 11 hours of video 74 topics

Topic #	Inter-active	Auto-matic	Text description of needed information/shot	Number of examples			Known items
				Video	Image	Audio	
1	Y		<b>number of spikes on Statue of Liberty's crown</b>		1		10
2		Y	<b>liftoff of the Space Shuttle</b>	4			
3		Y	<b>vehicle traveling on the moon</b>	1			2
4		Y	<b>mountains as prominent scenery</b>	1			8
5		Y	<b>water skiing</b>	1			5
6		Y	<b>scenes with a yellow boat</b>	1			4
7		Y	<b>pink flower</b>		1		1
8	Y	Y	<b>the planet Jupiter</b>		2		6
9		Y	<b>people who are water skiing</b>	1			
10		Y	<b>swimming pools</b>	1			

# TREC EVOLUTION

- IBM Groups Almaden and T.J. Watson (US)

SBD: used the IBM CueVideo toolkit;

Search: with, and without, speech recognition, automatic and interactive searching tasks; based on the semi-automatic construction of models for different kinds of scenes, events and objects - extensive experiments;

- Carnegie Mellon University

Search: with, and without, the Sphinx speech recognition system, both automatic and interactive searches; Minor changes to the Informedia system; Used colour histogram matching, texture, video OCR, face detection and speech recognition;

Topic #	Inter-active	Auto-matic	Text description of needed information/shot	Number of examples			Known items
				Video	Image	Audio	
1	Y		<b>number of spikes on Statue of Liberty's crown</b>		1		10
2		Y	<b>liftoff of the Space Shuttle</b>		4		
3		Y	<b>vehicle traveling on the moon</b>		1		2
4		Y	<b>mountains as prominent scenery</b>		1		8
5		Y	<b>water skiing</b>		1		5
6		Y	<b>scenes with a yellow boat</b>		1		4
7		Y	<b>pink flower</b>		1		1
8	Y	Y	<b>the planet Jupiter</b>		2		6
9		Y	<b>people who are water skiing</b>		1		
10		Y	<b>swimming pools</b>		1		

<http://trec.nist.gov/pubs/trec11/papers/NOVELTY.OVER.pdf>

# TREC EVOLUTION

- ▶ TREC 2002
  - ▶ Novelty Track: Goal is finding relevant and **novel** information
  - ▶ Much harder to evaluate than Ad-Hoc search
  - ▶ Later task evolved to find **diversity (in 2009!!!)**

<http://trec.nist.gov/pubs/trec18/papers/WEB09.OVERVIEW.pdf>

# TREC EVOLUTION

```
<topic number="21" type="faceted">
  <query>volvo</query>
  <description>
    I'm looking for information on Volvo cars and trucks.
  </description>
  <subtopic number="1" type="nav">
    I'm looking for Volvo's homepage.
  </subtopic>
  <subtopic number="2" type="inf">
    Find reviews of the Volvo XC90 SUV.
  </subtopic>
  <subtopic number="3" type="inf">
    Where can I find Volvo semi trucks for sale (new or used)?
  </subtopic>
  <subtopic number="4" type="inf">
    Find a Volvo dealer.
  </subtopic>
  <subtopic number="5" type="inf">
    Find an online source for Volvo parts.
  </subtopic>
</topic>
```

### Examples from TREC Web 2009

```
<topic number="19" type="ambiguous">
  <query>the current</query>
  <description>
    I'm looking for the homepage of The Current, a program
    on Minnesota Public Radio.
  </description>
  <subtopic number="1" type="nav">
    Take me to the homepage of The Current, a program on Minnesota
    Public Radio.
  </subtopic>
  <subtopic number="2" type="nav">
    I'm looking for the homepage of The Current newspaper in New Jersey.
  </subtopic>
  <subtopic number="3" type="nav">
    I want to find the homepage of The Current newspaper in Hartford.
  </subtopic>
  <subtopic number="4" type="nav">
    I want to find the homepage of The Current magazine in San Antonio.
  </subtopic>
</topic>
```

# TREC EVOLUTION

- ▶ TREC 2003
  - ▶ Robust Track: All queries that participants struggled before! <http://trec.nist.gov/pubs/trec12/papers/ROBUST.OVERVIEW.pdf>
  - ▶ Genomics Track: find information about genes.

For gene X, find all MEDLINE references that focus on the basic biology of the gene or its protein products from the designated organism. Basic biology includes isolation, structure, genetics and function of genes/proteins in normal and disease states.

<http://trec.nist.gov/pubs/trec12/papers/GENOMICS.OVERVIEW3.pdf>

<http://trec.nist.gov/pubs/trec13/papers/HARD.OVERVIEW.pdf>

# TREC EVOLUTION

- ▶ TREC 2004
- ▶ HARD (High Accuracy Retrieval of Documents) Track:
  - ▶ Propose the use of user context to improve retrieval
  - ▶ Metadata with: familiarity, genre and geography

<http://trec.nist.gov/pubs/trec13/papers/HARD.OVERVIEW.pdf>

# TREC EVOLUTION

<number>

HARD-428

</number>

<title>

International organ traffickers

</title>

<description>

Who creates the demands in the international ring of organ trafficking?

</description>

<topic-narrative>

Many countries are institutionalizing legal measures to prevent the selling and buying of human organs. Who, in the ring of international organ trafficking, are the "buyers" of human organs? Any information that identifies 'where' they are or 'who' they may be will be considered on topic; the specificity of info does not matter. Also, the story must be about international trafficking. Stories that only contain information about the "sellers" of organs or those that focus on national trafficking will be off topic.

</topic-narrative>

<http://trec.nist.gov/pubs/trec13/papers/HARD.OVERVIEW.pdf>

# TREC EVOLUTION

<metadata-narrative>

Subject (CURRENT EVENTS) is chosen as it is expected that such articles will have more information about the identities of the parties involved. Genre (NEWS) is expected to exclude stories that tends to focus on ethical matters.

</metadata-narrative>

<retrieval-element>

passage

</retrieval-element>

<metadata>

  <familiarity>

    little

  </familiarity>

  <genre>

    news-report

  </genre>

  <geography>

    any

  </geography>

</metadata>

</topic>

<http://trec.nist.gov/pubs/trec14/papers/ENTERPRISE.OVERVIEW.pdf>

# TREC EVOLUTION

- ▶ TREC 2005
  - ▶ Enterprise Track:
    - ▶ Enterprise data: intranet pages, email archives, document repositories, people (experts) search

<http://trec.nist.gov/pubs/trec14/papers/ENTERPRISE.OVERVIEW.pdf>

# TREC EVOLUTION

- ▶ TREC 2005
  - ▶ Enterprise Track:
    - ▶ Enterprise data: intranet pages, email archives, document repositories, people (experts) search

In the expert search task, participants could use all 331,037 documents in order to rank a list of 1092 candidate experts. This could involve creating a document for each candidate and applying simple IR techniques, or could involve natural language processing and information extraction technologies targeted at different document types such as email. Results are presented in Table 5.

[Who does that nowadays?](#)

# TREC EVOLUTION

- ▶ TREC 2006
  - ▶ Blog Track:
    - ▶ Searching blogs (weblogs) for an entity (location, person, organization, product name, brand).
    - ▶ Task is to summarize people's opinion
  - ▶ Legal Track:
    - ▶ "e-discovery"
    - ▶ Recall-oriented task

# TREC EVOLUTION

- ▶ TREC 2007
- ▶ Million Queries:
  - ▶ Not one million queries, but a large number (10.000)
  - ▶ Collection with 426GB
  - ▶ Investigating whether it is better to evaluate using many shallow judgments or fewer thorough judgments

<http://trec.nist.gov/pubs/trec19/papers/SESSION.OVERVIEW.2010.pdf>

# TREC EVOLUTION

- ▶ TREC 2010
  - ▶ Session Track: Going beyond single query retrieval
  - ▶ Query Reformulation were classified into 3 categories:
    - ▶ Specifications: the user enters a query, realizes the results are too broad or that they wanted a more detailed level of information, and reformulates a more specific query.
    - ▶ Drifting/Parallel Reformulation: the user entered a query, then reformulated to another query with the same level of specification but moved to a different aspect or facet of their information need.
    - ▶ Generalizations: the user enters a query, realizes that the results are too narrow or that they wanted a wider range of information, and reformulated a more general query.

## LECTURE 8 - IR EVALUATION 2

---

<http://trec.nist.gov/pubs/trec19/papers/SESSION.OVERVIEW.2010.pdf>

# TREC EVOLUTION

```
<topic number="4" type="faceted">
  <query>toilet</query>
  <description> Find information on buying, installing, and repairing
  toilets.
  </description>
  <subtopic number="1" type="inf">
    What different kinds of toilets exist, and how do they differ?
  </subtopic>
  <subtopic number="2" type="inf">
    I'm looking for companies that manufacture residential toilets.
  </subtopic>
  <subtopic number="3" type="inf">
    Where can I buy parts for American Standard toilets?
  </subtopic>
  <subtopic number="4" type="inf">
    How do I fix a toilet that isn't working properly?
  </subtopic>
  <subtopic number="5" type="inf">
    What companies manufacture bidets?
  </subtopic>
  <subtopic number="6" type="inf">
    I'm looking for a Kohler wall-hung toilet. Where can I buy one?
  </subtopic>
</topic>
```

## LECTURE 8 - IR EVALUATION 2

---

<http://trec.nist.gov/pubs/trec19/papers/SESSION.OVERVIEW.2010.pdf>

# TREC EVOLUTION

```
<topic number="4" type="faceted">
  <query>toilet</query>
  <description> Find information on buying, installing, and repairing
  toilets.
  </description>
  <subtopic number="1" type="inf">
    What different kinds of toilets exist, and how do they differ?
  </subtopic>
  <subtopic number="2" type="inf">
    I'm looking for companies that manufacture residential toilets.
  </subtopic>
  <subtopic number="3" type="inf">
    Where can I buy parts for American Standard toilets?
  </subtopic>
  <subtopic number="4" type="inf">
    How do I fix a toilet that isn't working properly?
  </subtopic>
  <subtopic number="5" type="inf">
    What companies manufacture bidets?
  </subtopic>
  <subtopic number="6" type="inf">
    I'm looking for a Kohler wall-hung toilet. Where can I buy one?
  </subtopic>
</topic>
```

Pick a random subtopic



<http://trec.nist.gov/pubs/trec19/papers/SESSION.OVERVIEW.2010.pdf>

# TREC EVOLUTION

```
<topic number="4" type="faceted">
  <query>toilet</query>
  <description> Find information on buying, installing, and repairing
  toilets.
  </description>
  <subtopic number="1" type="inf">
    What different kinds of toilets exist, and how do they differ?
  </subtopic>
  <subtopic number="2" type="inf">
    I'm looking for companies that manufacture residential toilets.
  </subtopic>
  <subtopic number="3" type="inf">
    Where can I buy parts for American Standard toilets?
  </subtopic>
  <subtopic number="4" type="inf">
    How do I fix a toilet that isn't working properly?
  </subtopic>
  <subtopic number="5" type="inf">
    What companies manufacture bidets?
  </subtopic>
  <subtopic number="6" type="inf">
    I'm looking for a Kohler wall-hung toilet. Where can I buy one?
  </subtopic>
</topic>
```

Pick a random subtopic



```
<topic number="1" reformtype="specification" source="webtrack2009">
  <query>toilet</query>
  <reformulation>Kohler</reformulation>
  <description>I'm looking for a Kohler wall-hung toilet.
  Where can I buy one?</description>
</topic>
```

<http://trec.nist.gov/pubs/trec21/papers/CONTEXTUAL12.overview.pdf>

# TREC EVOLUTION

- ▶ TREC 2012
  - ▶ Contextual Track: Somehow evolved from HARD Track
  - ▶ As input to the task participants were given a **set of example suggestions**, a **set of user preference profiles**, and a **set of geotemporal contexts**. The task was to take the profiles and contexts and to produce up to 50 ranked suggestions for each combination of profile and context. Participants gathered suggestions from the open Web.

## LECTURE 8 - IR EVALUATION 2

---

<http://trec.nist.gov/pubs/trec21/papers/CONTEXTUAL12.overview.pdf>

# TREC EVOLUTION

### Example of context (Zero term query)

```
<context number="1">
  <city>New York City</city>
  <state>NY</state>
  <lat>40.71427</lat>
  <long>-74.00597</long>
  <day>weekday</day>
  <time>afternoon</time>
  <season>summer</season>
</context>
```



Given a context, find me 50 suggestions...

## LECTURE 8 - IR EVALUATION 2

---

<http://trec.nist.gov/pubs/trec21/papers/CONTEXTUAL12.overview.pdf>

# TREC EVOLUTION

### Example of context (Zero term query)

```
<context number="1">
  <city>New York City</city>
  <state>NY</state>
  <lat>40.71427</lat>
  <long>-74.00597</long>
  <day>weekday</day>
  <time>afternoon</time>
  <season>summer</season>
</context>
```



Given a context, find me 50 suggestions...

### Example of suggestion (all in Toronto area):

```
<example number="1">
  <title>Fresh on Bloor</title>
  <description>Our vegan menu boasts an array of exotic
  starters , multi-layered salads , filling wraps , high
  protein burgers and our signature Fresh bowls.</description>
  <url>http://www.freshrestaurants.ca</url>
</example>
```

# TREC EVOLUTION

- ▶ TREC 2014
  - ▶ Clinical Decision Support Track

Topic	Type	Summary
1	Diagnosis	58-year-old woman with hypertension and obesity presents with exercise-related episodic chest pain radiating to the back.
11	Test	40-year-old woman with severe right arm pain and hypotension. She has no history of trauma and right arm exam reveals no significant findings.
21	Treatment	21-year-old female with progressive arthralgias, fatigue, and butterfly-shaped facial rash. Labs are significant for positive ANA and anti-double-stranded DNA, as well as proteinuria and RBC casts.

# LECTURE 8 - IR EVALUATION 2

# TREC EVOLUTION

- ▶ TREC 2014
  - ▶ Clinical Decision Support Track

Topic	Type	Summary
1	Diagnosis	58-year-old woman with hypertension and obesity presents with exercise-related episodic chest pain radiating to the back.
11	Test	40-year-old woman with severe right arm pain and hypotension. She has no history of trauma and right arm exam reveals no significant findings.
21	Treatment	21-year-old female with progressive arthralgias, fatigue, and butterfly-shaped facial rash. Labs are significant for positive ANA and anti-double-stranded DNA, as well as proteinuria and RBC casts.

# Have you watched “Idiocracy” ?



<http://trec.nist.gov/pubs/trec25/papers/Overview-T.pdf>

# TREC EVOLUTION

## ▶ TREC 2015

### ▶ Tasks Track:

"cheap hotels in london"



- (a) Price of hotels
- (b) Location
- (c) Hotel review
- (d) Other accommodations in London
- (e) Hotel in a near by location

### ▶ Goals:

- ▶ (1) Task understanding
- ▶ (2) Task completion
- ▶ (3) Adhoc Retrieval

### ▶ Collection: Clueweb12 (1.95 Tb, >50 M. docs)

<http://trec.nist.gov/pubs/trec25/papers/Overview-T.pdf>

# TREC EVOLUTION

- ▶ TREC 2016
  - ▶ Open Search Track: OpenSearch is a new evaluation paradigm for IR. The experimentation platform is an existing search engine. Researchers have the opportunity to **replace components of this search engine** and evaluate these components using **interactions with real, unsuspecting users** of this search engine.
  - ▶ Search Engines accessed thru API:
    - ▶ <http://citeseerx.ist.psu.edu/>
    - ▶ <http://www.ssoar.info/>
    - ▶ <https://academic.microsoft.com/>
  - ▶ Real time search!

<http://trec.nist.gov/tracks.html>

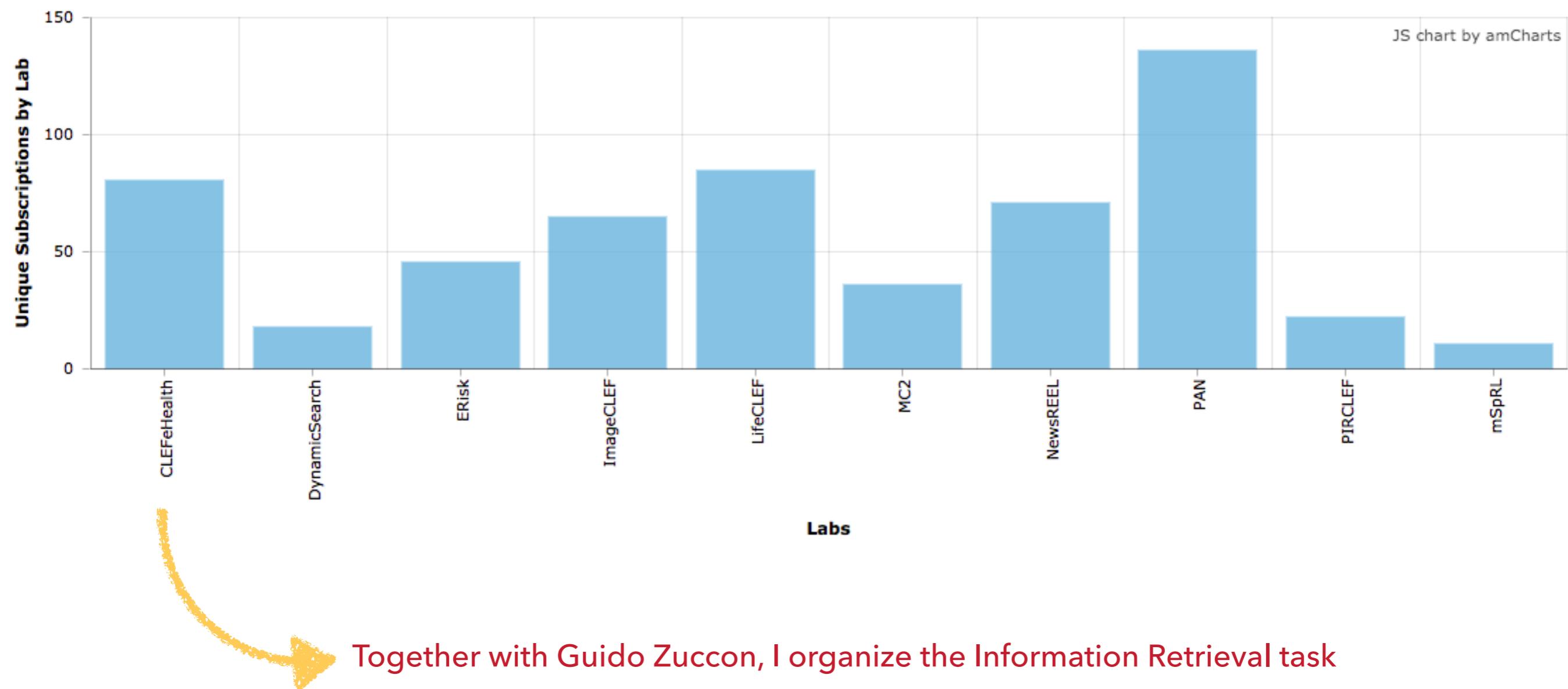
# TREC EVOLUTION

- ▶ TREC 2017:
  - ▶ Common Core Track
  - ▶ Complex Answer Retrieval Track
  - ▶ Dynamic Domain Track
  - ▶ Live QA Track
  - ▶ OpenSearch Track
  - ▶ Precision Medicine Track
  - ▶ Real-Time Summarization Track
  - ▶ Tasks Track

# SONS OF CRANFIELD EXPERIMENTS

- ▶ TREC (Text REtrieval Conference)
- ▶ CLEF (Conference and Labs of the Evaluation Forum)
- ▶ NTCIR (NII Testbeds and Community for Information access Research)
- ▶ FIRE (Forum Information REvaluation)
  
- ▶ TrecVid (TREC Video REvaluation)
- ▶ Mirex (Music Information REvaluation eXchange)
- ▶ MediaEval (Benchmarking Initiative for Multimedia Evaluation)

# CLEF 2017



112 - Documents should discuss ethical issues involved with organ donation, and that any reports that highlight treatment differences in patients who are known organ donors.

[Back to Query](#)[Prev Document](#)

## Document "clueweb12-0005wb-45-41384" (21 / 500)

[Next Document](#)

n  
MATION

[Donation Laws](#)

ee on  
tion

[Reports](#)

[islative History](#)

[al Events](#)

# Advisory Committee on Organ Transplantation

Hilton Rockville Hotel

Rockville, MD

May 15-16, 2007

May 15, 2007

Ms. Gail Agrawal, Chairperson, welcomed the group and read a letter from Larry Hagman, whose last meeting this was to be. Mr. Hagman's letter described his inability to attend due to family obligations and expressed how much he has enjoyed his ACOT membership.

Mr. Remy Aronoff, Executive Secretary, introduced Dr. Russ Wiesner, the newest member, from the Mayo Clinic. Mr. Aronoff thanked ACOT members for the work that everyone has done between meetings and noted that this will be a very full agenda.

### Public Solicitation of Donors – Dr. Douglas Hanto

There are important ethical aims in distributing scarce health care: maximizing the benefits from a limited resource (utility), and distributing the resource fairly or equitably among all possible recipients (justice). These two goals will often conflict, however, and trade-offs will be necessary, but they must be kept in mind. This summary comes from Dr. Dan Brock of Harvard Medical School, who is currently working with the government in Thailand, looking at ways to provide limited dialysis in that country.

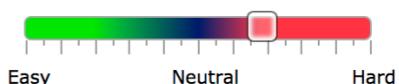
Organ donation can be thought of along a spectrum that spans from altruism, solicitation, incentives, coercion and theft. The concept of “sales” can occur along this spectrum in multiple places, depending on your position regarding the ethics of sales. Everyone is familiar with billboards that have been used to advertise for a directed liver donation to that individual.

Decreased organ donation allocation: organs are allocated in a non-directed fashion to

### Judgement

- Unjudged
- Highly relevant
- Somewhat relevant
- Not relevant

### Understandability



Easy      Neutral      Hard

### Trustworthiness



Low      Medium      High

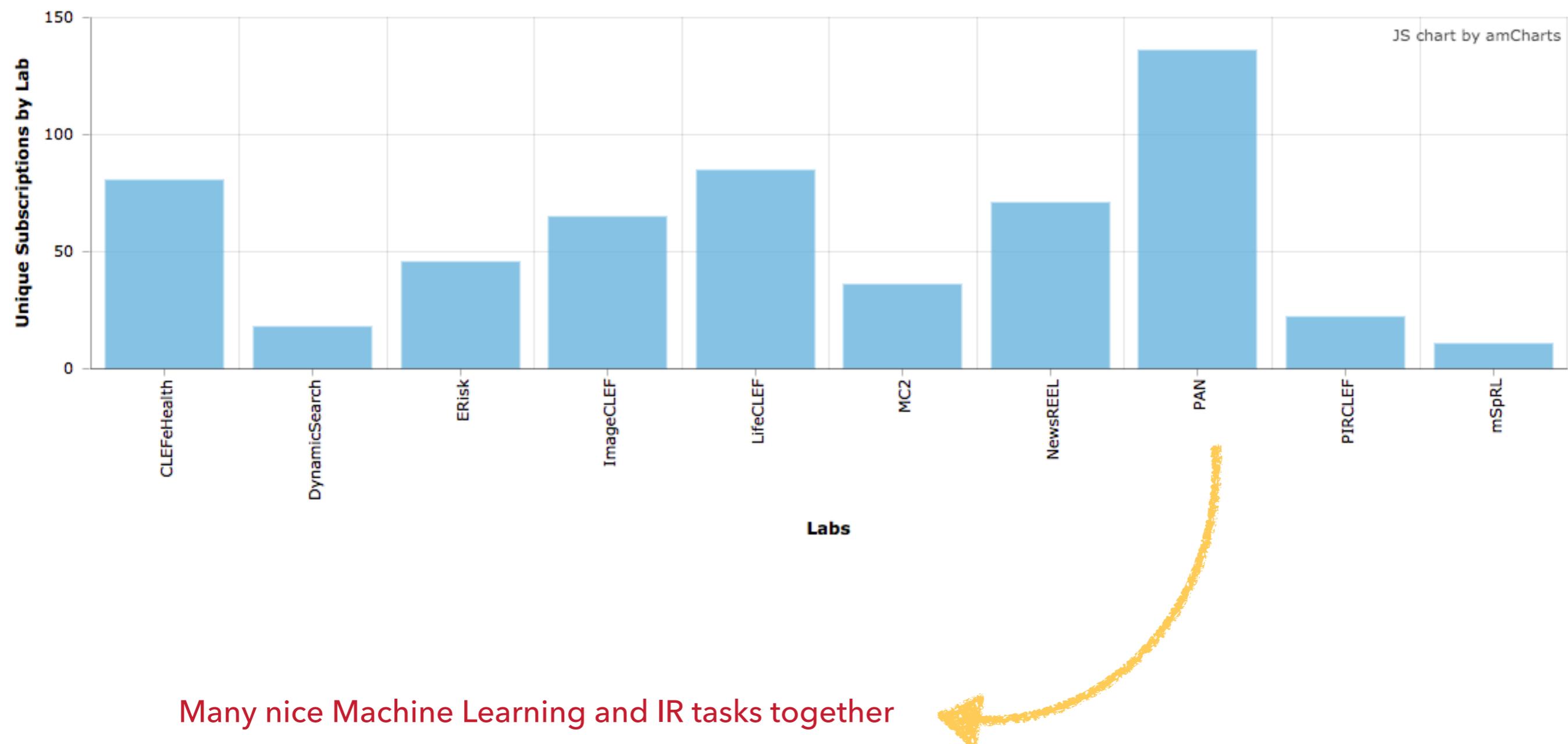
### Comment

(optional)

This talks about treatment differences in LIVING organ donors, probably different from the intent around antemortem treatment of postmortem donors?

[Insert Selected Text](#)[Save](#)[Save & Next](#)

# CLEF 2017



<http://research.nii.ac.jp/ntcir/ntcir-13/>

**NTCIR**

- ▶ How about trying to solve many of the proposed tasks in Japanese?
  - ▶ Ad-hoc retrieval in Japanese/Chinese
  - ▶ User context search in Japan with Japanese data
  - ▶ Q&A in Japanese
  - ▶ + others like: Math retrieval:  $E = mc^2$

<http://fire.irsi.res.in/fire/2016/home>

**FIRE**

- ▶ Focus on languages with very low resource (E.g. >20 languages spoken in India for example)
- ▶ Example: Mixed Script Information Retrieval  
<https://msir2016.github.io/>
  - ▶ Subtask 1: Code-Mixed Cross-Script Question Classification (Bengali-English)
  - ▶ Subtask 2: Information Retrieval on Code-Mixed Hindi-English Tweets

<http://www-nlpir.nist.gov/projects/tv2017/>

# TRECVID 2017

- ▶ Directly evolved from TREC...
- ▶ Ad-Hoc Video Search:
  - ▶ + 400h of video
  - ▶ clock time per query search, reported in seconds (to one decimal place)
- ▶ Surveillance Event Detection (SED):
  - ▶ Task: an observable action or change of state in a video stream that would be important for airport security management
  - ▶ systems should output detection results for any three events in the following list: PersonRuns, CellToEar, ObjectPut, PeopleMeet, PeopleSplitUp, Embrace, and Pointing.

<http://www.multimediaeval.org/mediaeval2017/index.html>

# MULTIMEDIA EVAL 2017

- ▶ Mostly Machine Learning tasks:
  - ▶ Retrieving Diverse Social Images Task: A search for “notre dame cathedral” will not only retrieve photos of the main entrance/facade
  - ▶ Emotional Impact of Movies Task: detect an emotion for a given movie scene
  - ▶ Medico – Medical Multimedia Task: classify diseases according to few images or short clips

[http://www.music-ir.org/mirex/wiki/MIREX\\_HOME](http://www.music-ir.org/mirex/wiki/MIREX_HOME)

- ▶ Some of 2016 tasks:
  - ▶ Audio Music Similarity and Retrieval
  - ▶ Audio Cover Song Identification
  - ▶ Real-time Audio to Score Alignment (a.k.a Score Following)
  - ▶ Discovery of Repeated Themes & Sections
  - ▶ Query by Singing/Humming
  - ▶ Audio classification: Genre, Mood, composer...

# HOMEWORK 2

- ▶ Homework 1 is graded!!!
- ▶ Very very light penalty for errors/mistakes
- ▶ Do not expect me to be so kind in the next HW...
  
- ▶ Grades for methods were updated – see GitHub

[https://github.com/joaopalotti/cmu\\_67300/blob/master/project\\_assignments/assignment2.pdf](https://github.com/joaopalotti/cmu_67300/blob/master/project_assignments/assignment2.pdf)

## WHAT HAVE WE SEEN TODAY?

- ▶ Review/Recap:
  - ▶ Precision Vs Recall in the context of what we have seen so far
- ▶ Motivation for you to be a researcher:
  - ▶ Plenty of Evaluation Campaigns
- ▶ Homework 2