



UNIVERSITY OF AMSTERDAM

Challenges in Professional Search

Evangelos Kanoulas
e.kanoulas@uva.nl

Who am I?

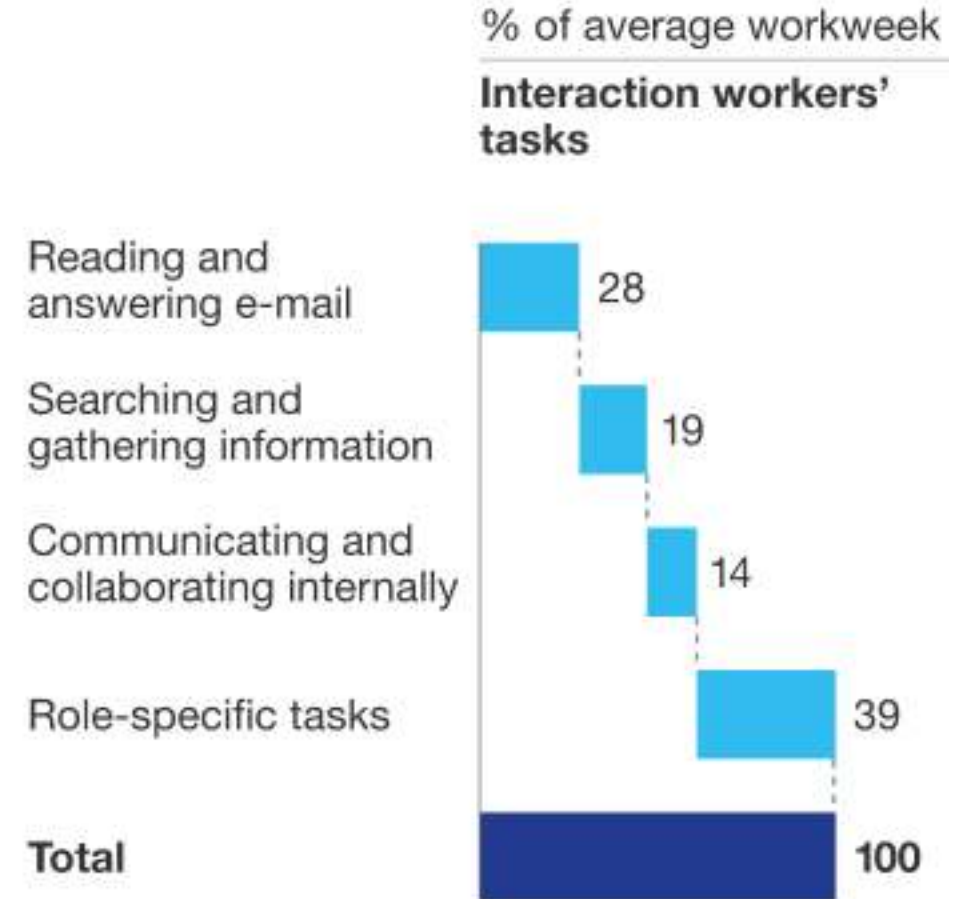
- Assistant Professor at Institute of Informatics (UvA)
 - Director of Data Science MSc program (UvA, VU, ADS)
- Before that:
 - Google Research & University of Sheffield
- My background:
 - Computer Science (PhD and MSc, Northeastern Univ.)
 - Joint degree on Informatics & Economics (BS)
- My expertise:
 - **Information Retrieval**, Text Mining, and Natural Language Understanding



Professional Search

“... employees spend 1.8 hours every day— 9.3 hours per week, on average— searching and gathering information.” – source: McKinsey

“the knowledge worker spends about 2.5 hours per day, or roughly 30% of the workday, searching for information” – source: IDC



Web Search Engines

Great at answering simple user questions

A screenshot of a Google search interface. The search bar contains the text "how old is barack obama". Below the search bar, there are navigation tabs for "Web", "Images", "News", "Videos", "Shopping", "More", and "Search tools". The search results show "About 110,000,000 results (0.78 seconds)". On the left side, there is a knowledge panel for Barack Obama, showing his age as "54 years" and his birth date as "August 4, 1961". Below this, there is a small photo of Michelle Obama with the text "Michelle Obama Spouse 51 years".



why are sunsets red

Web Images Videos Shopping News More S

About 18,800,000 results (0.53 seconds)

When the sun is low in the sky, this long journey through the means the colors with shorter wavelengths, like blue, have a scattered or bounced off in numerous directions. Orange su (yellow and red light waves) appear when the air is clean.

Why are sunsets red?

sci.odu.edu/sci/Scire/05Edition/sunset.html

A screenshot of a Google search interface. The search bar contains the text "tests for pregnant mothers". Below the search bar, there are navigation tabs for "Web", "Images", "Videos", "News", "Shopping", "More", and "Search tools". The search results show "About 48,800,000 results (0.47 seconds)". The first result is "Common Tests During Pregnancy" from www.stanfordchildrens.org. The second result is "Prenatal Tests You Need for Your Baby's Health - WebMD" from www.webmd.com. The third result is "Prenatal Testing: Ultrasounds, Blood Tests, and Your Baby" from www.webmd.com. The fourth result is "[PDF] Routine Tests During Pregnancy - American College of ..." from www.acog.org. The fifth result is "Prenatal care and tests | womenshealth.gov" from www.womenshealth.gov. The sixth result is "Pregnancy & Prenatal Testing - Lab Tests Online" from https://labtestsonline.org. The seventh result is "Screening tests in pregnancy - NHS Choices" from www.nhs.uk. The eighth result is "Antenatal checks and tests - Pregnancy and baby guide - NHS" from www.nhs.uk. The ninth result is "Routine blood tests in pregnancy - BabyCentre" from www.babycentre.co.uk.

Web Search

- Find **one** (or a few) good web-pages
 - High redundancy in information on the web
 - High redundancy in user signals
 - E.g. clicks on documents, query re-writes

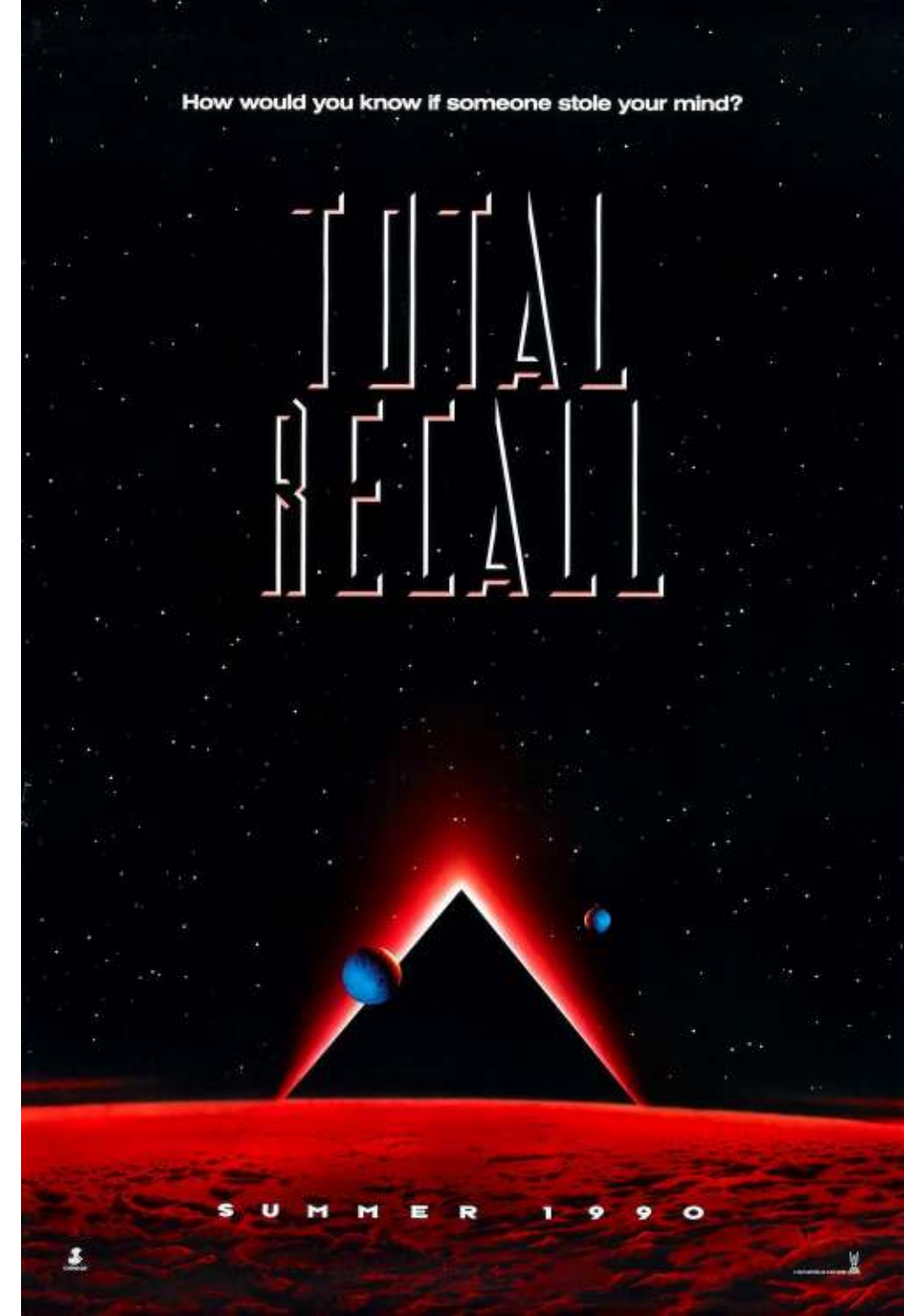
Professional Search

- (Often) **exploratory** search
 - Users do not know exactly what they are looking for or ...
 - ... how to phrase their request (query)



Professional Search

- **Total-recall** search
 - Users need to find (nearly) **everything** about a **topic X**
 - Exhaustive research
 - X = me, my PhD topic, ebola
 - Investigation
 - X = somebody or something or some activity
 - Systematic review
 - X = studies measuring a particular effect
 - Patent search
 - X = prior art



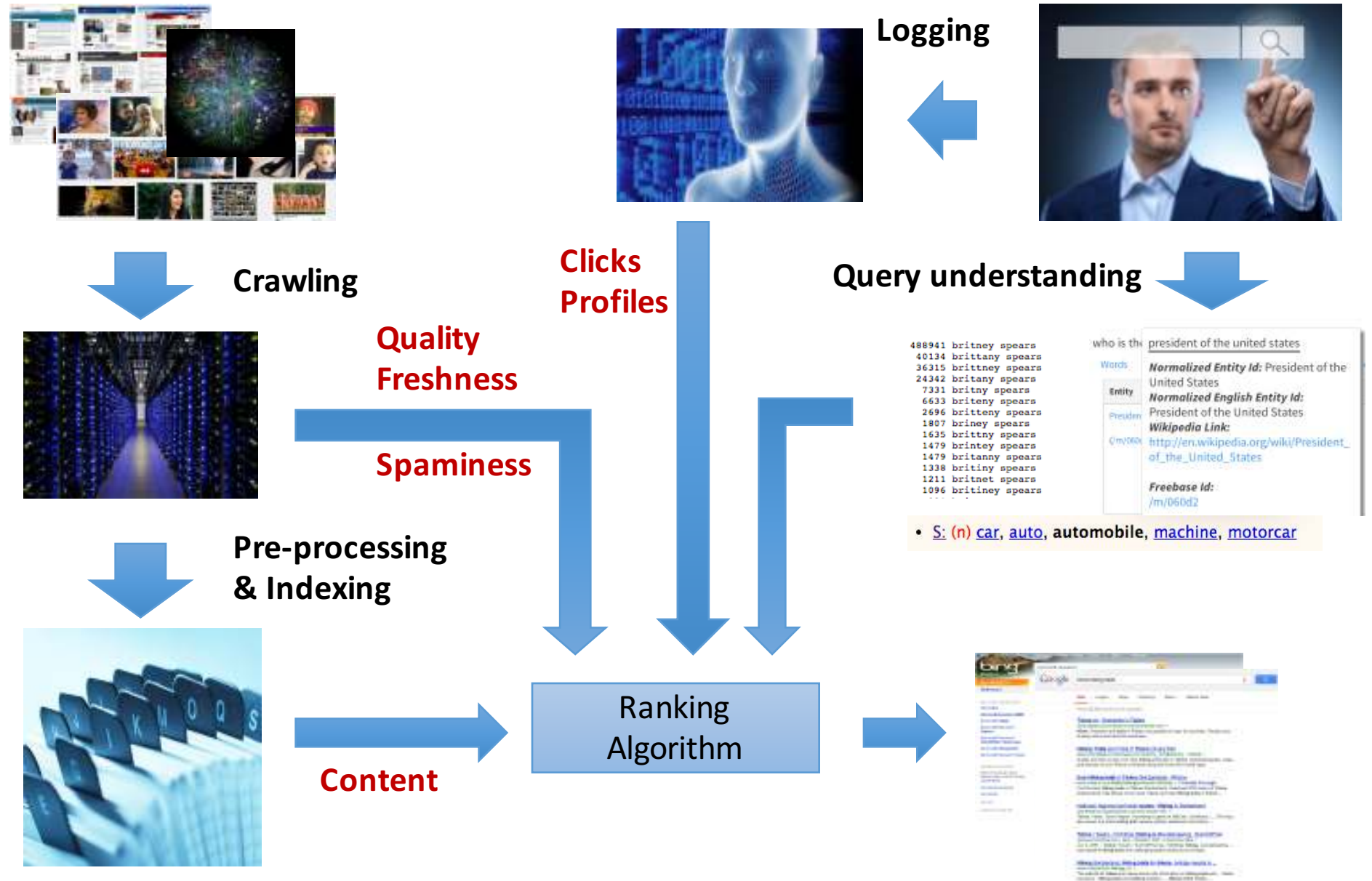
Professional Search

- There is **no (simple) single query**

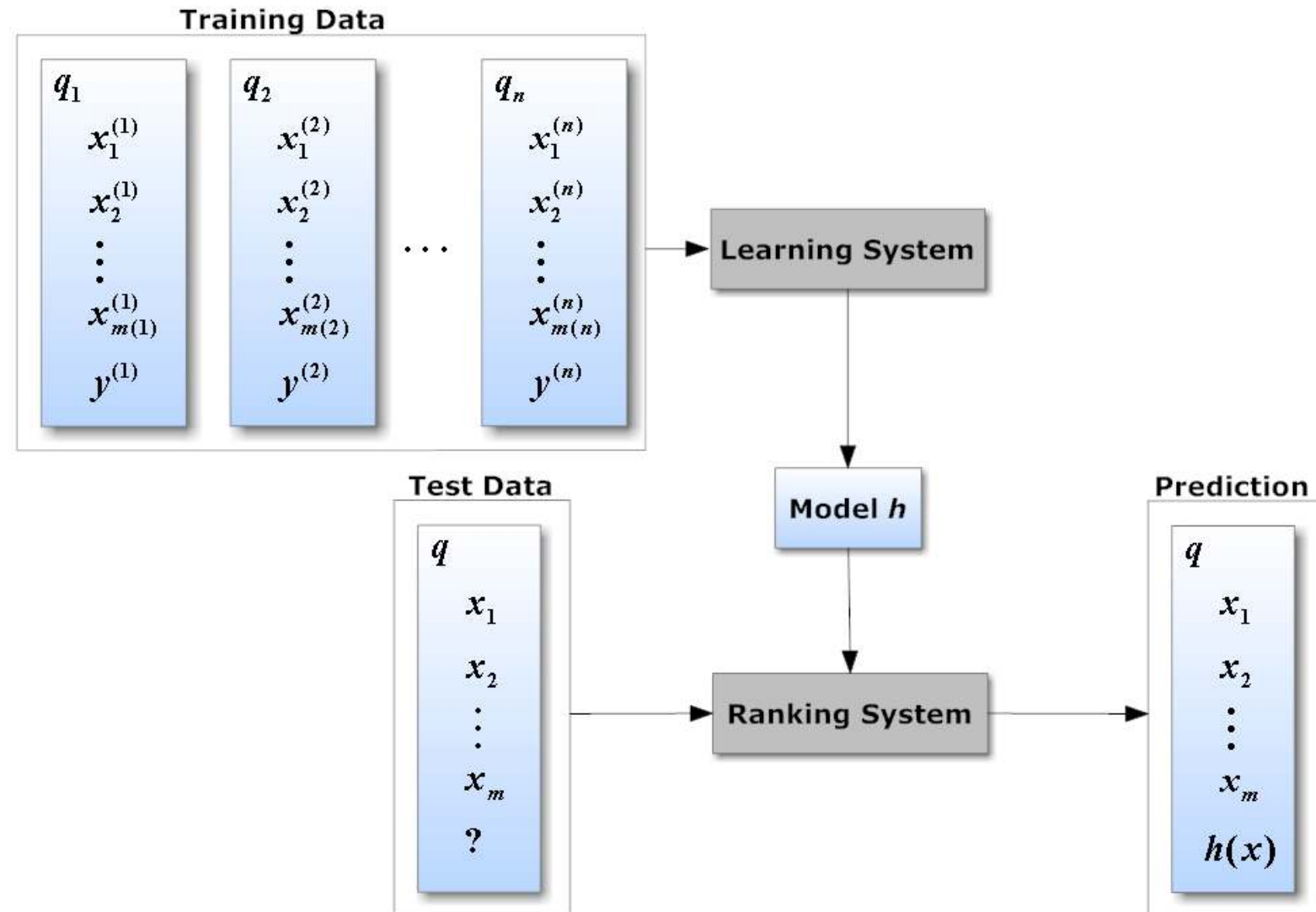
Main Question: Is adjunctive vitamin A effective in children diagnosed with non-measles pneumonia?

1. exp vitamin A/
2. vitamin A.mp
3. retinol.mp
4. exp dietary supplements/
5. or/1-4
6. exp pneumonia/
7. pneumonia\$.mp
8. exp pneumonia, bacterial/
9. exp pneumonia, lipid/
10. exp pneumonia, mycoplasma/
- ...
14. exp pneumonia, viral/
15. exp respiratory tract infections/
16. acute adj respiratory.mp
17. respiratory adj infection.mp
18. respiratory adj disease.mp
19. or/6-18
20. 5 and 19

Modern Search Engines



Modern Search Engines



Batch Learning

- Requires **labeling data** (query – document pairs)
 - Time-consuming, and boring



Batch Learning

- Leads to a **static, one-size-fits-all** search engine

User Feedback

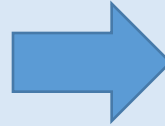
- Leads to a **static, one-size-fits-all** search engine

2

1



adjunctive vitamin A children pneumonia



adjunctive vitamin A children non-measles pneumonia

- [Vitamin A for non-measles pneumonia in children.](#)
 1. Ni J, Wei J, Wu T. Cochrane Database Syst Rev. 2008 Jul 20(3):CD003700. Review. PMID: 18234828 [Similar articles](#)
 2. [Vitamin A for treating measles in children.](#)
 3. Hulming Y, Chaomin W, Meng M. Cochrane Database Syst Rev. 2008 Oct 19(4):CD001476. Review. PMID: 18232020 [Similar articles](#)
 3. [Vitamin A for treating measles in children.](#)
 4. D'Souza RM, D'Souza R. Cochrane Database Syst Rev. 2002 (1):CD001476. Review. Update in: [Cochrane Database Syst Rev. 2002](#); [11\(1\):CD001476](#). PMID: 11895951 [Similar articles](#)
 4. [Vitamin A for treating measles in children.](#)
 5. D'Souza RM, D'Souza R. Cochrane Database Syst Rev. 2001;(2):CD001476. Review. Update in: [Cochrane Database Syst Rev. 2002](#); [11\(1\):CD001476](#). PMID: 11429593 [Similar articles](#)
 5. [Vitamin A supplementation for preventing morbidity and mortality in children from 6 months to 5 years of age.](#)
 6. Indad A, Herzer K, Mayo-Wilson E, Yakoob MF, Bhutta ZA. Cochrane Database Syst Rev. 2013 Dec 8;(12):CD008524. doi: 10.1002/14651099.CD008524.pub2. Review. PMID: 24114289 [Similar articles](#)
 6. [Routine vitamin A supplementation for the prevention of blindness due to measles infection in children.](#)
 7. Bello S, Mamerikou MM, Ejemil-Nwadiaro RI, Oduwole O. Cochrane Database Syst Rev. 2011 Apr 12(4):CD007718. doi: 10.1002/14651099.CD007718.pub2. Review. Update in: [Cochrane Database Syst Rev. 2014\(1\):CD007718](#). PMID: 21481401 [Similar articles](#)
 7. [Interventions for the prevention and management of oropharyngeal candidiasis associated with HIV infection in adults and children.](#)
 8. Penzance FT, Wilson T, Holmes H.

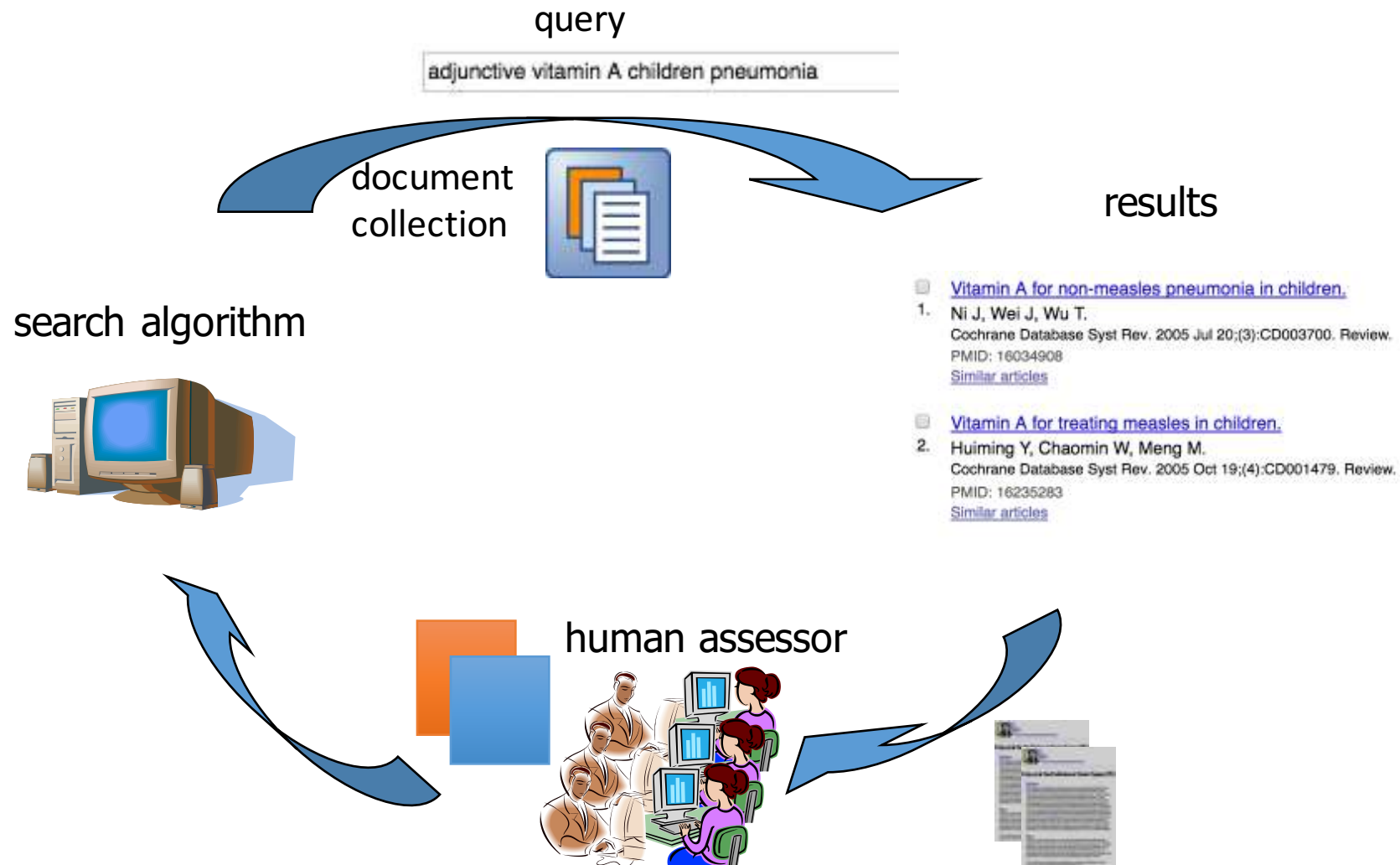


TREC Total Recall

Objective:

1. Find documents containing nearly all relevant information ...
2. ... while uncovering [relatively] few documents

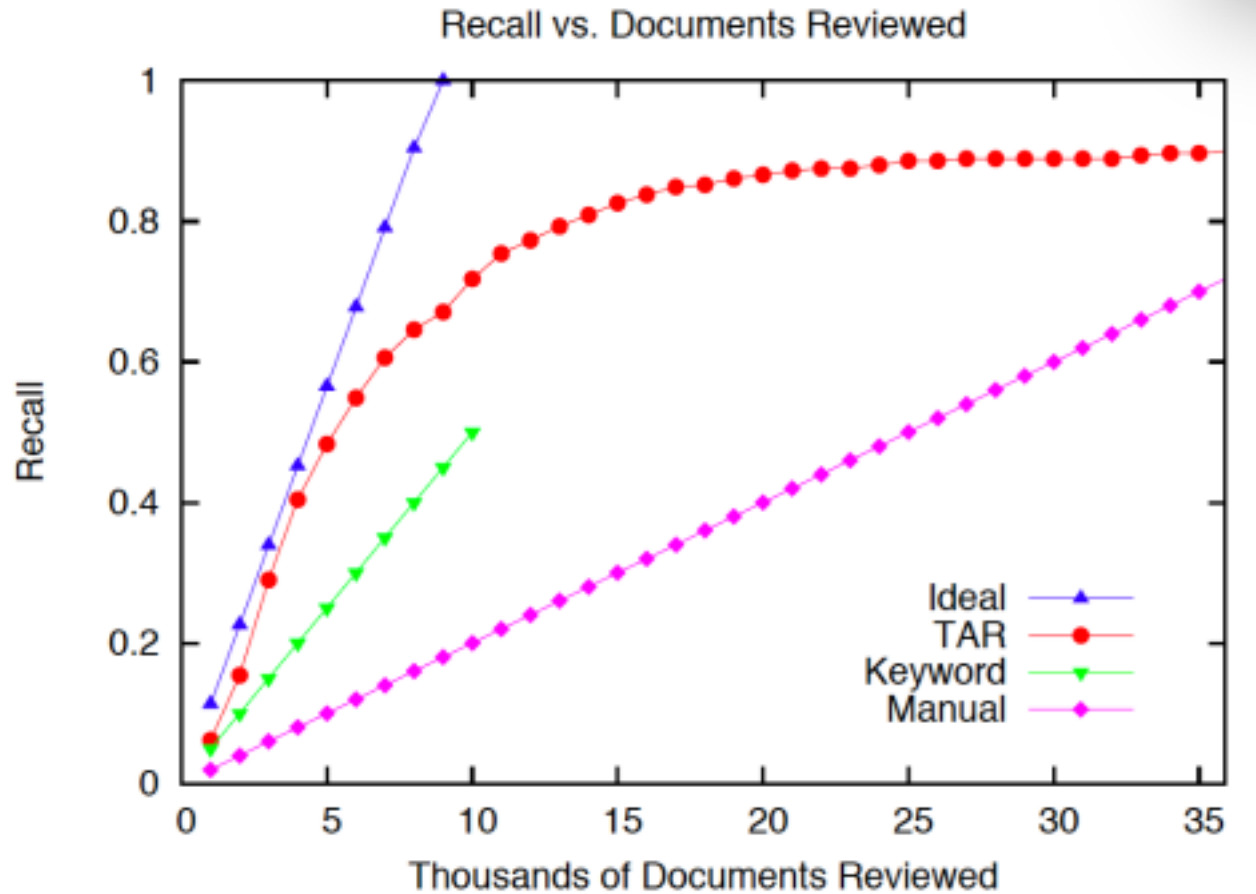
TREC Total Recall: Participation



TREC Total Recall: Participation

- Play-at-home
 - Data collection and queries available via internet
 - Automated assessor accessed via the Internet
- Play-in-sandbox
 - Submit virtual appliance that works isolated from internet
 - Downloads corpus, topic from intranet
 - “Uncover” documents one at a time via intranet

TREC Total Recall

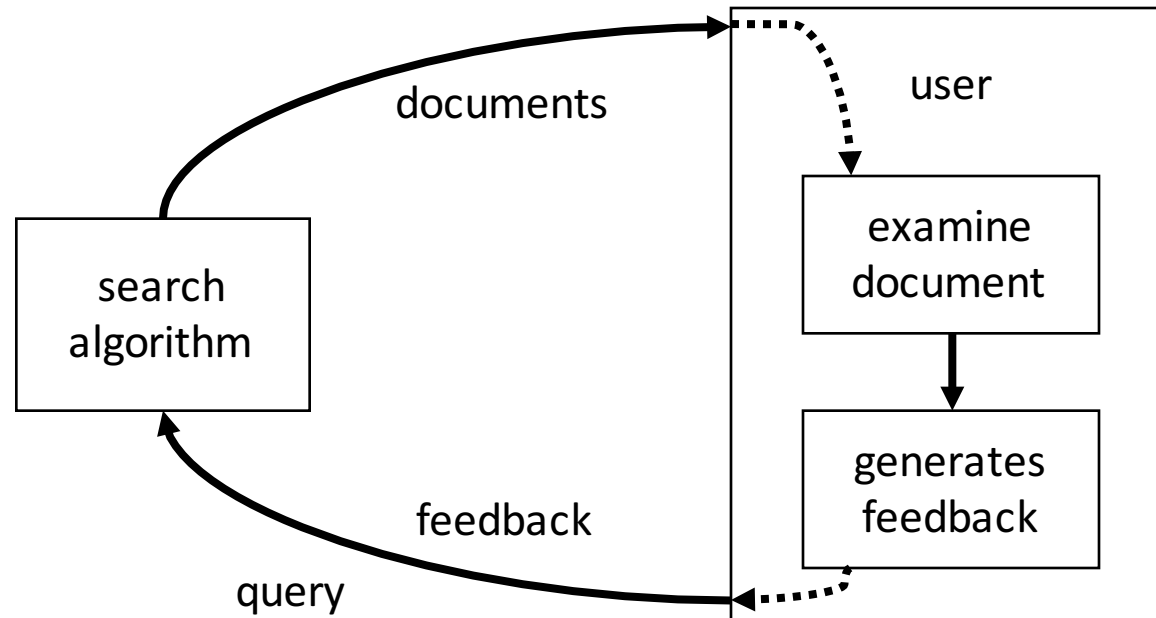


User-in-the-loop strategies

- Extreme relevance feedback
- Batch learning
 - uncover training set; rank
- Online learning [UvA/HvA submission]

Online Learning


- Learn-as-you-go
 - Requires user feedback (implicit or explicit)
 - Serves the user and builds a training collection at the same time



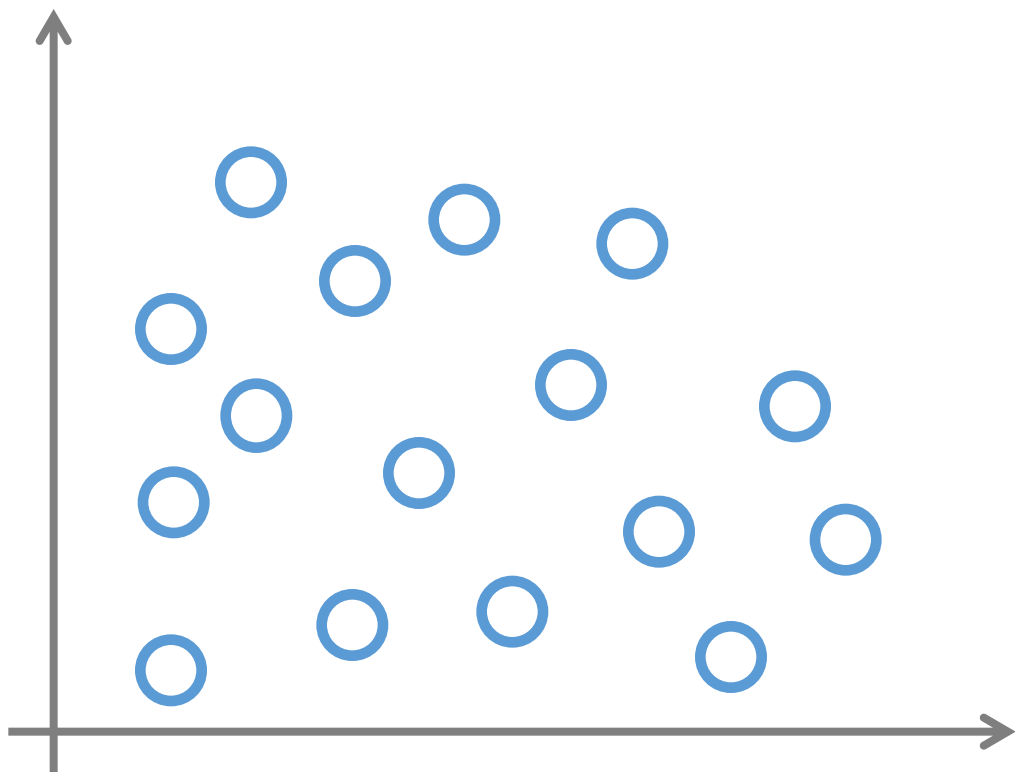
Online Learning

- Learn-as-you-go
 - Requires user feedback (implicit or explicit)
 - Serves the user and builds a training collection at the same time
- The collection contains feedback (e.g. labels) only on items you show to the user

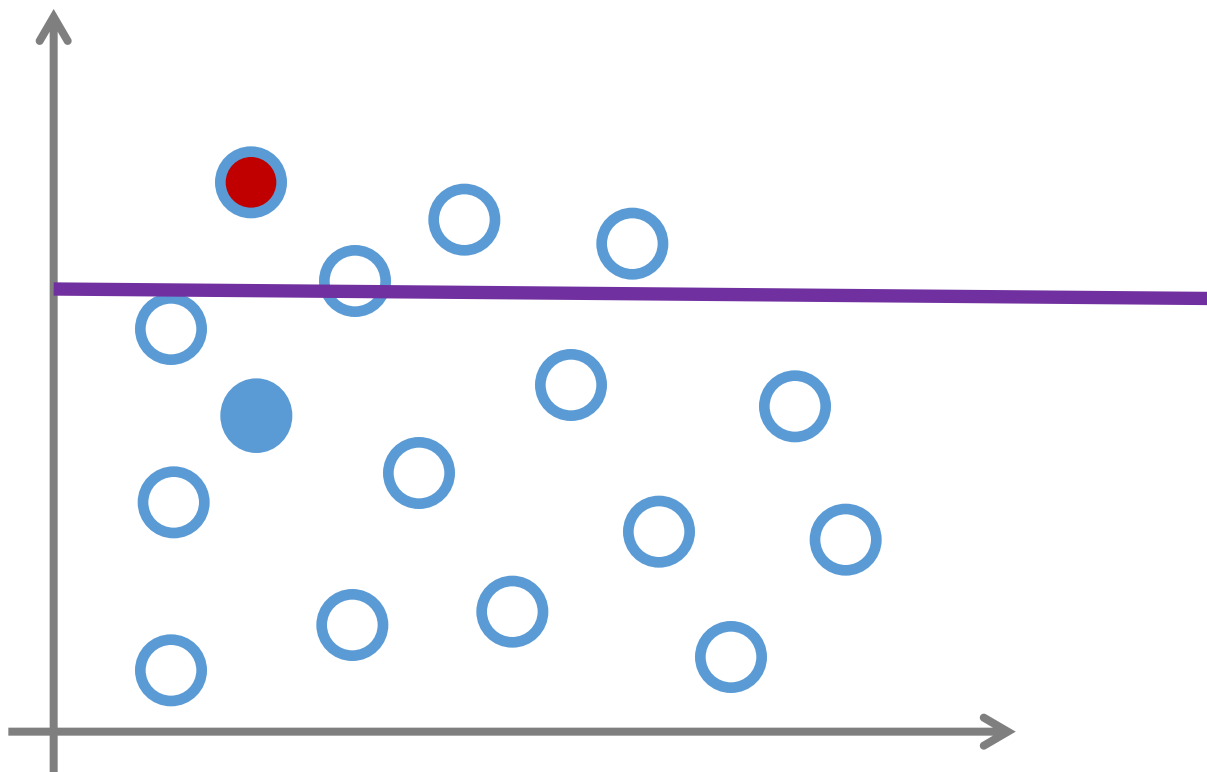
Exploitation: Baseline Model

1. Run ad hoc search to construct a synthetic training dataset
 - Unsupervised method – no training data needed
 2. Train a classifier
 3. Predict relevance for the remaining collection
 4. Select a few **highest-scoring documents** for review.
 5. Review the documents, coding each as “relevant” or “not relevant.”
 6. Add the documents to the training set.
- 

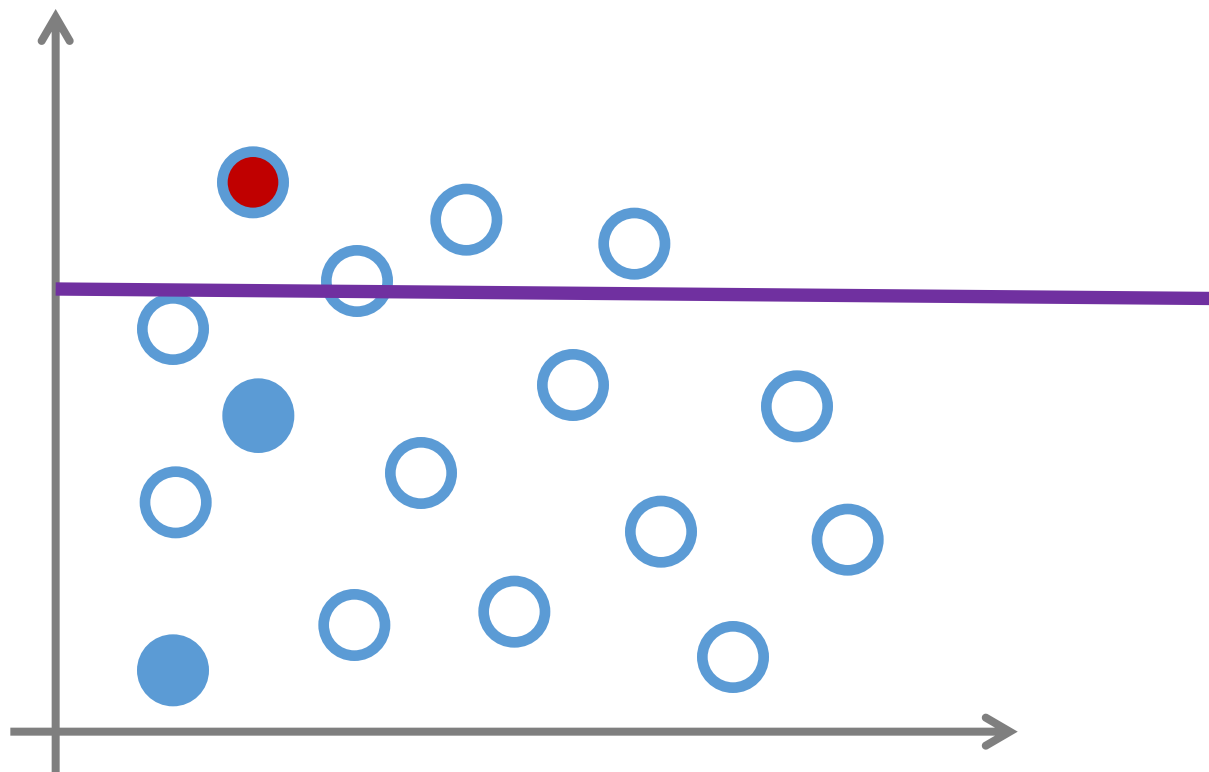
Exploitation: Baseline Model



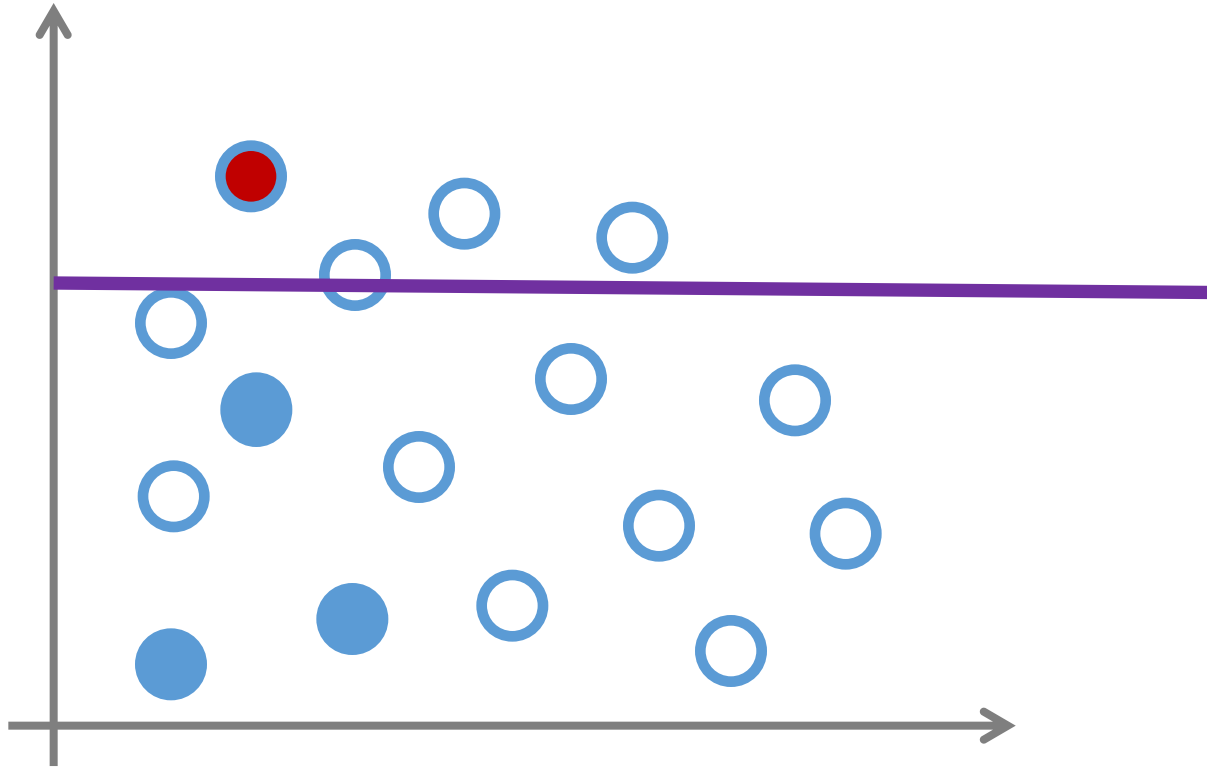
Exploitation: Baseline Model



Exploitation: Baseline Model

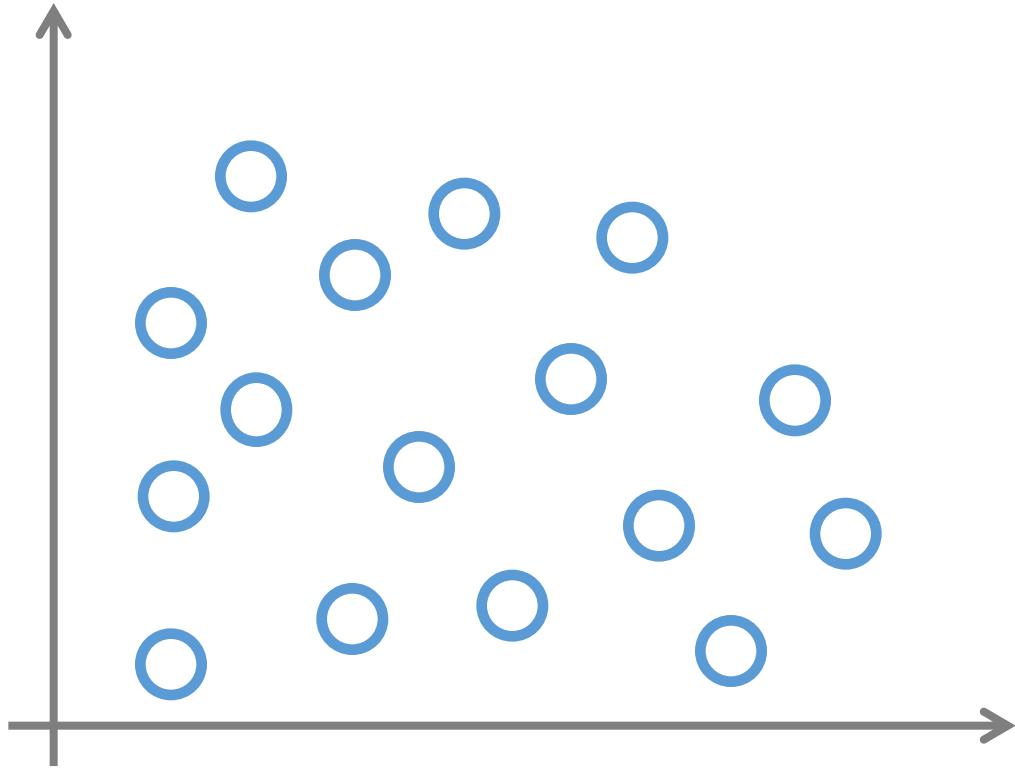


Exploitation: Baseline Model

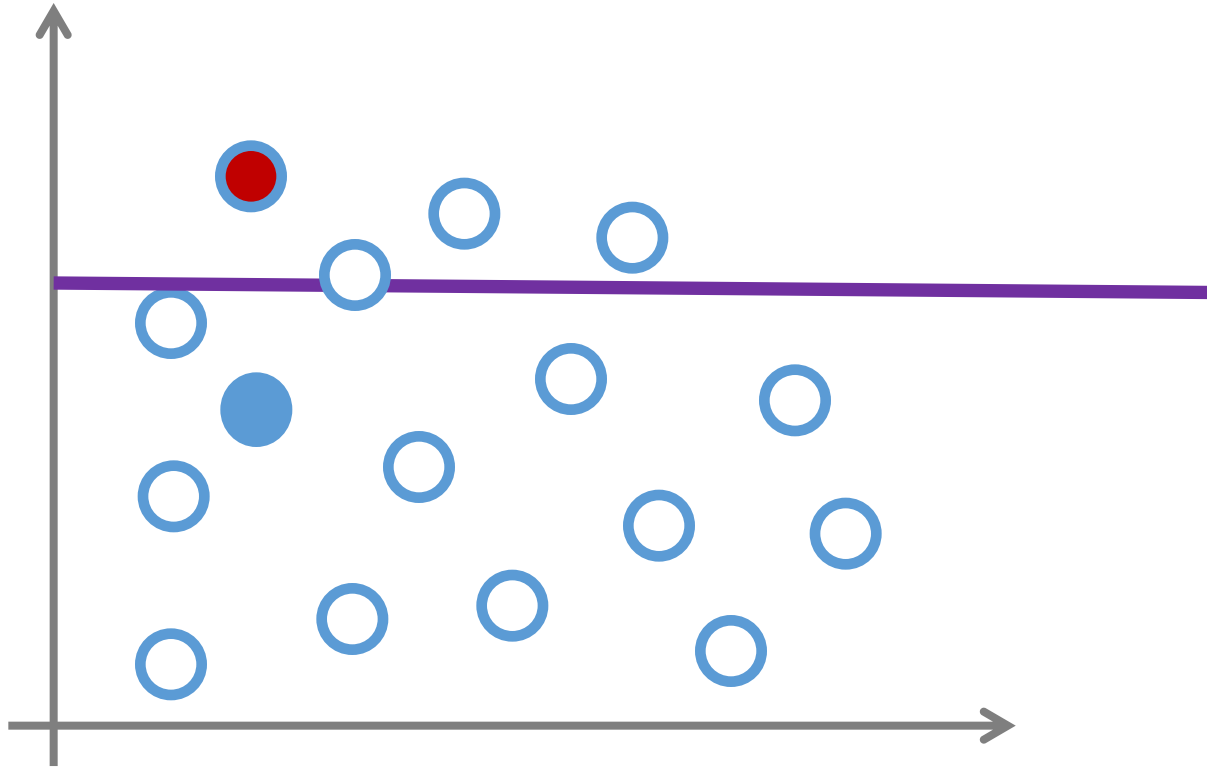


uncover the
most **relevant**
document to
present

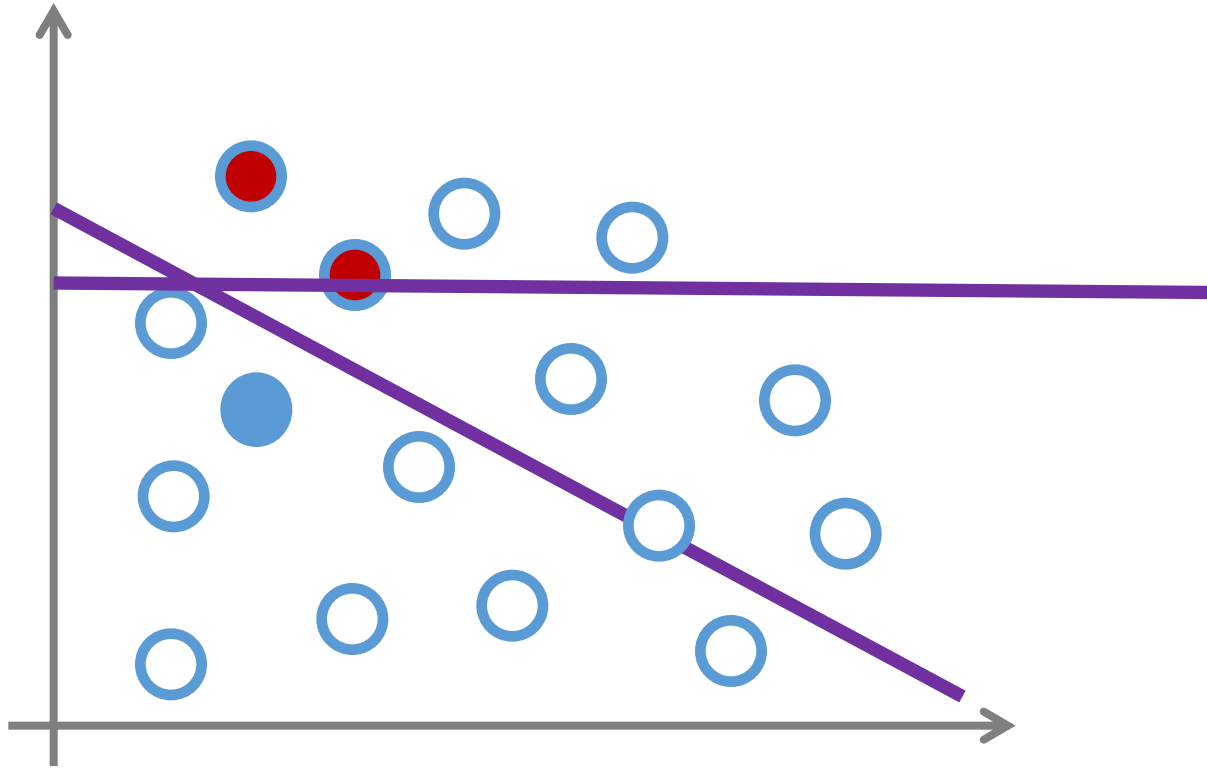
Exploitation: Active Learning



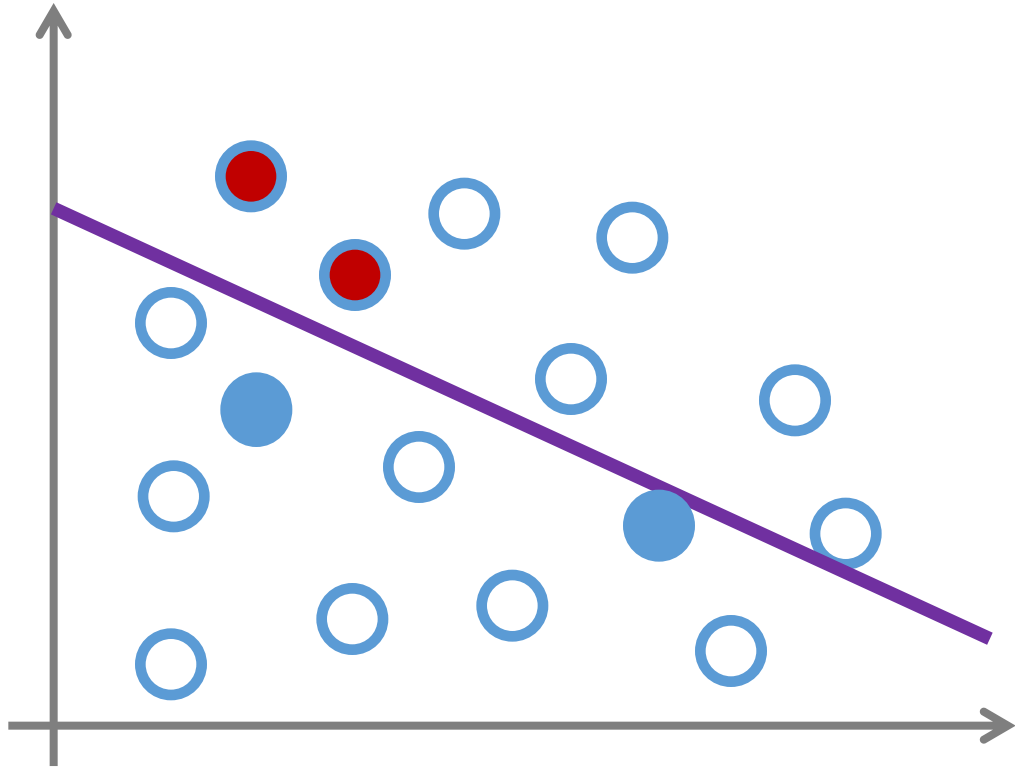
Exploitation : Active Learning



Exploitation : Active Learning

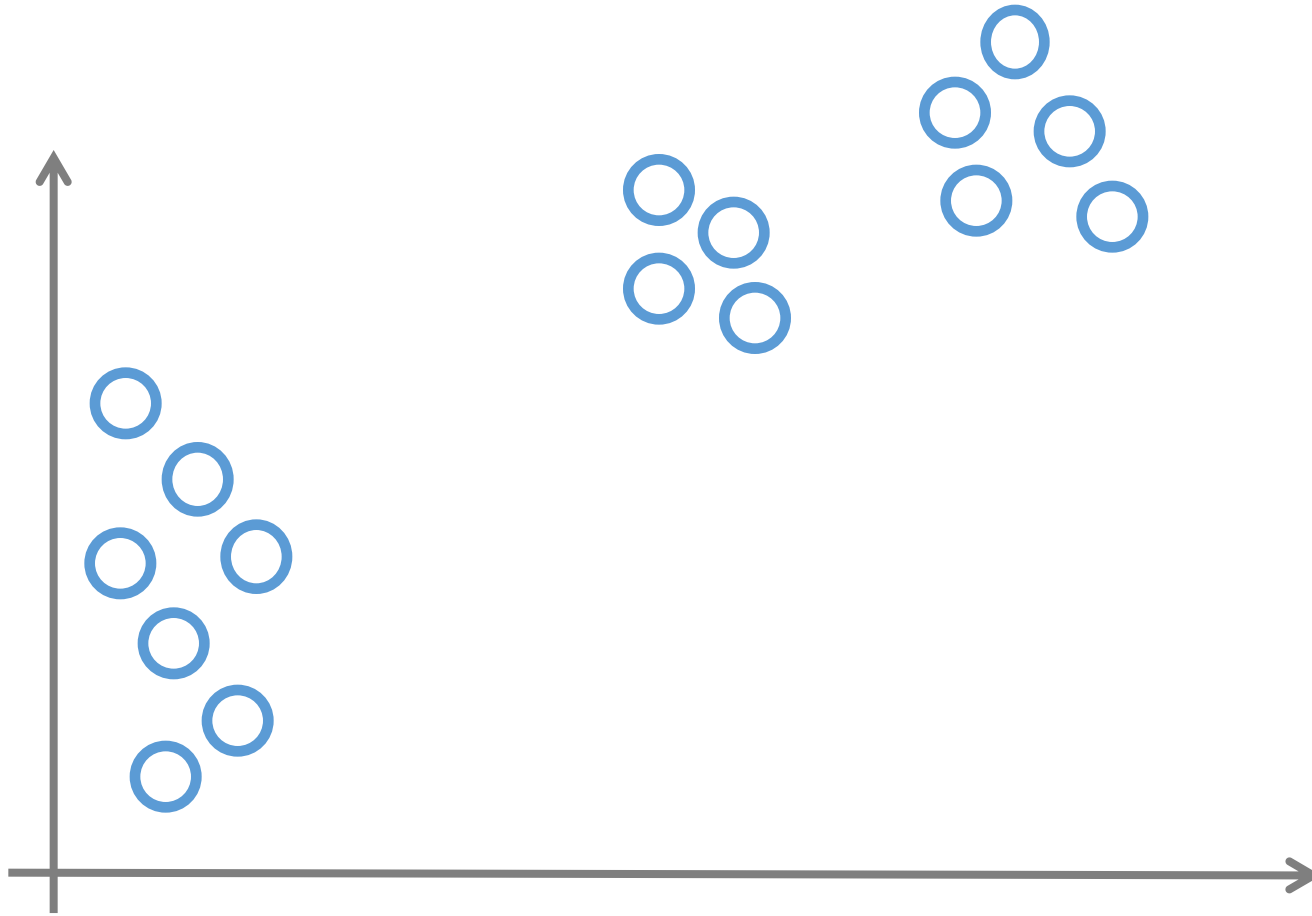


Exploitation : Active Learning

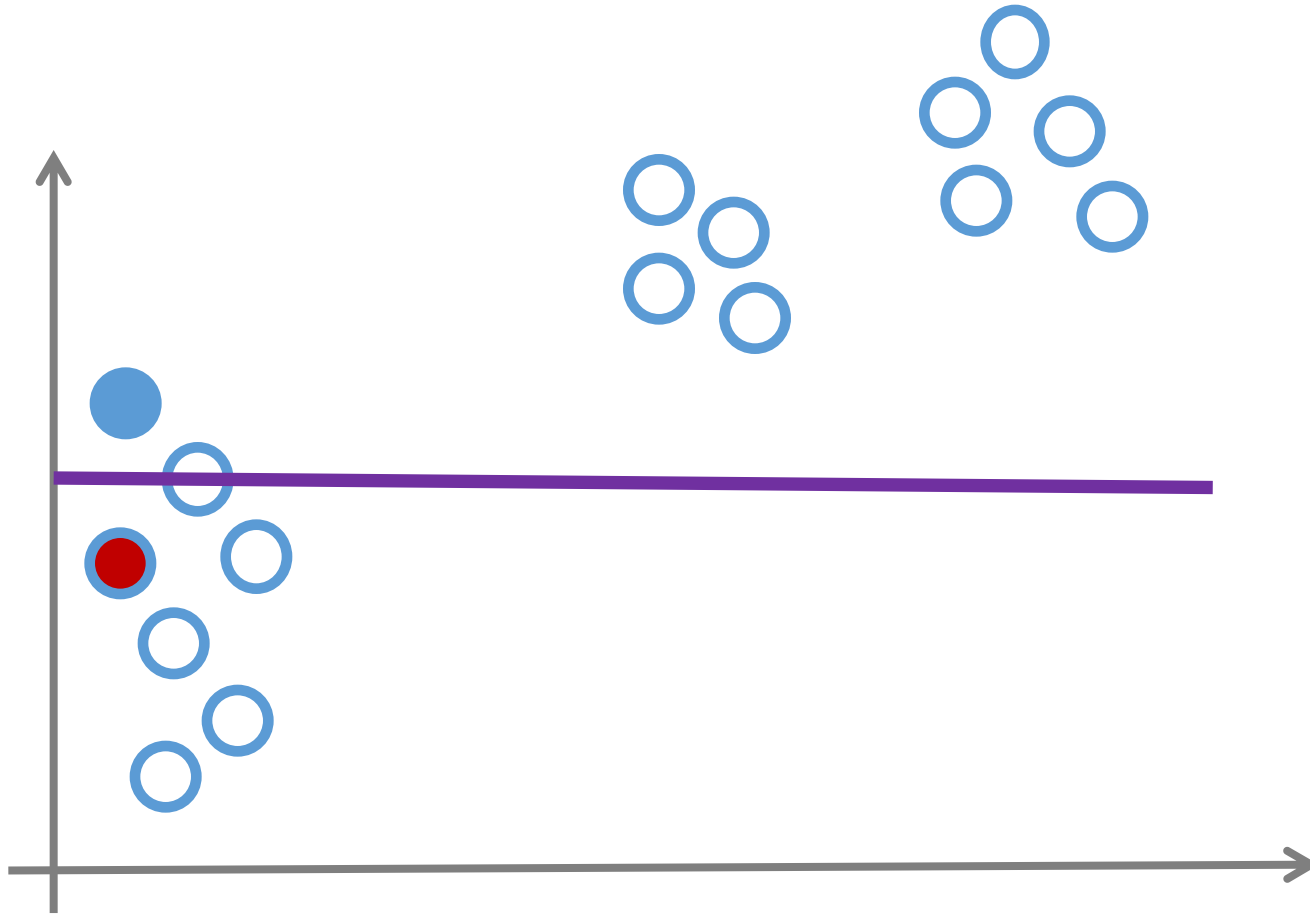


uncover the
most **informative**
document to
present

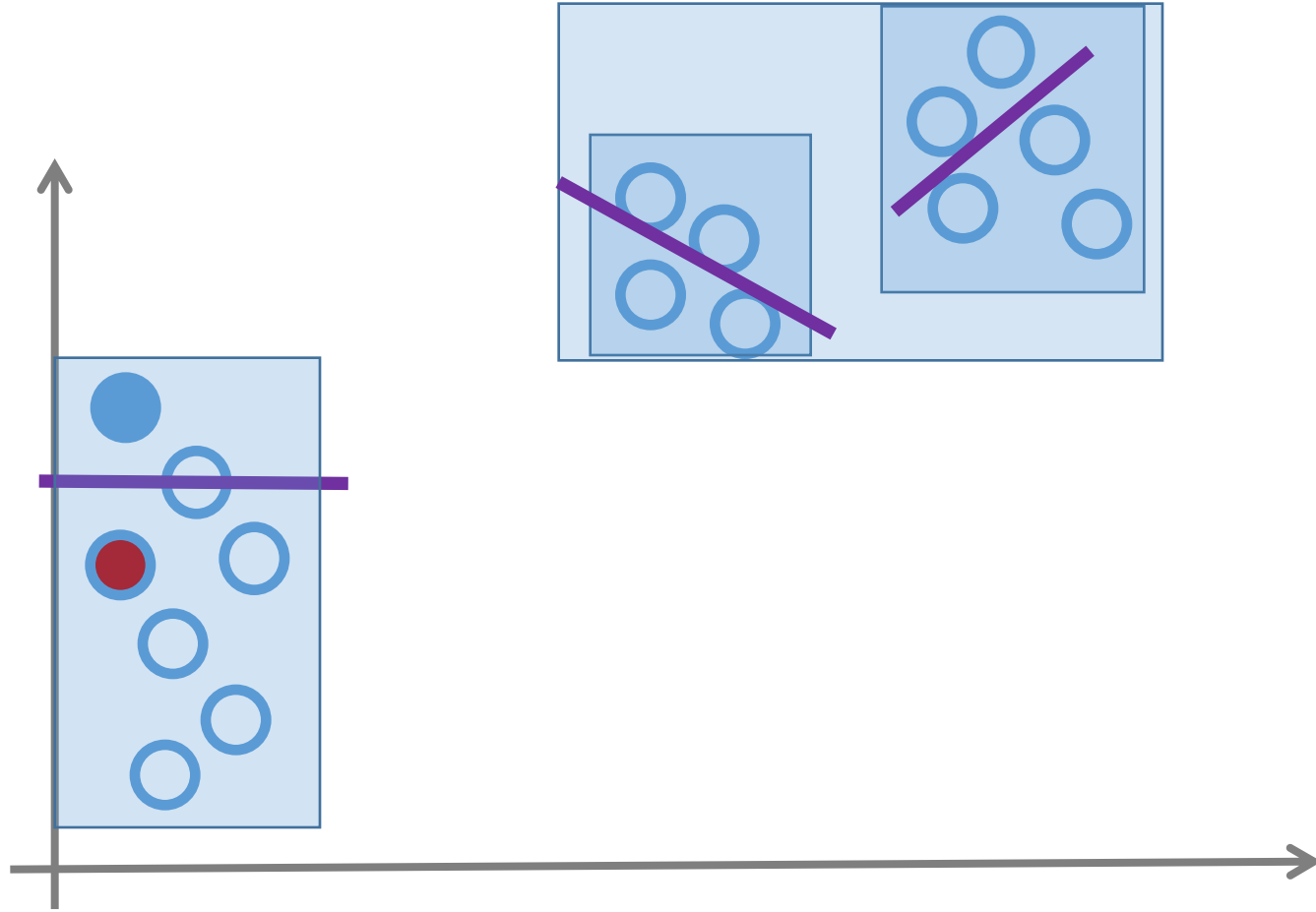
Exploration: Hierarchical Clustering



Exploration: Hierarchical Clustering



Exploration: Hierarchical Clustering



Reinforcement Learning

Balances:

1. **Exploitation** – uncover the most relevant document
2. **Exploitation** – uncover the most informative document
3. **Exploration** – uncover documents from different regions

User Feedback

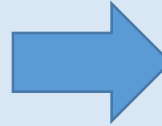
- Leads to a **static, one-size-fits-all** search engine

2

1



adjunctive vitamin A children pneumonia



adjunctive vitamin A children non-measles pneumonia

- 1. [Vitamin A for non-measles pneumonia in children.](#)
Ni J, Wei J, Wu T.
Cochrane Database Syst Rev. 2008 Jul 20(3):CD003700. Review.
PMID: 18334938
[Similar articles](#)
- 2. [Vitamin A for treating measles in children.](#)
Hulming Y, Chaomin W, Meng M.
Cochrane Database Syst Rev. 2008 Oct 19(3):CD001476. Review.
PMID: 18333203
[Similar articles](#)
- 3. [Vitamin A for treating measles in children.](#)
D'Souza RM, D'Souza R.
Cochrane Database Syst Rev. 2002;(1):CD001476. Review. Update in: [Cochrane Database Syst Rev. 2002](#);
[1\(1\):CD001476](#).
PMID: 11895951
[Similar articles](#)
- 4. [Vitamin A for treating measles in children.](#)
D'Souza RM, D'Souza R.
Cochrane Database Syst Rev. 2001;(2):CD001476. Review. Update in: [Cochrane Database Syst Rev. 2002](#);
[1\(1\):CD001476](#).
PMID: 11429593
[Similar articles](#)
- 5. [Vitamin A supplementation for preventing morbidity and mortality in children from 6 months to 5 years of age.](#)
Indad A, Herzer K, Mayo-Wilson E, Yakoob MF, Bhutta ZA.
Cochrane Database Syst Rev. 2013 Dec 8;(12):CD008524. doi: 10.1002/14651099.CD008524.pub2. Review.
PMID: 24134289
[Similar articles](#)
- 6. [Routine vitamin A supplementation for the prevention of blindness due to measles infection in children.](#)
Bello S, Mameritoku MM, Ejemol-Nwadiaro RI, Oduwole O.
Cochrane Database Syst Rev. 2011 Apr 12;(4):CD007718. doi: 10.1002/14651099.CD007718.pub2. Review. Update in:
[Cochrane Database Syst Rev. 2014;1\(1\):CD007718](#).
PMID: 21481401
[Similar articles](#)
- 7. [Interventions for the prevention and management of oropharyngeal candidiasis associated with HIV infection in adults and children.](#)
Ponnan FT, Wilson T, Holmes H.

Session Personalization

paris hotels

3660 Hotels in Paris - Best Price Guarantee - booking.com

Paris Hotels - TripAdvisor.com

Hotels in Paris - Save on Hotels with Expedia

Hotels near paris

Paris France Hotel	€120
Four Seasons Hotel George V Paris	€220
Hotel Lutetia Paris Tour Eiffel	€210

The 10 Best Paris Hotels - TripAdvisor

Top 10 Best Hotels in Paris | Hotels.com

Paris Hotels - Cheap Hotels in Paris - lastminute.com

Paris Hotels

Compare 1485+ Hotels in Paris - France Hotel Guide

paris Hilton

Paris Hilton - Wikipedia, the free encyclopedia

Paris Hilton (@ParisHilton) | Twitter

Paris Hilton (Spanish) | Instagram photos and videos

In the news

Paris Hilton - Facebook

Paris Hilton - IMDb

Paris Hilton - Biography - IMDb

Paris Hilton - Reality Television Star - Biography.com

Paris Hilton | People.com



TREC Session

Objective:

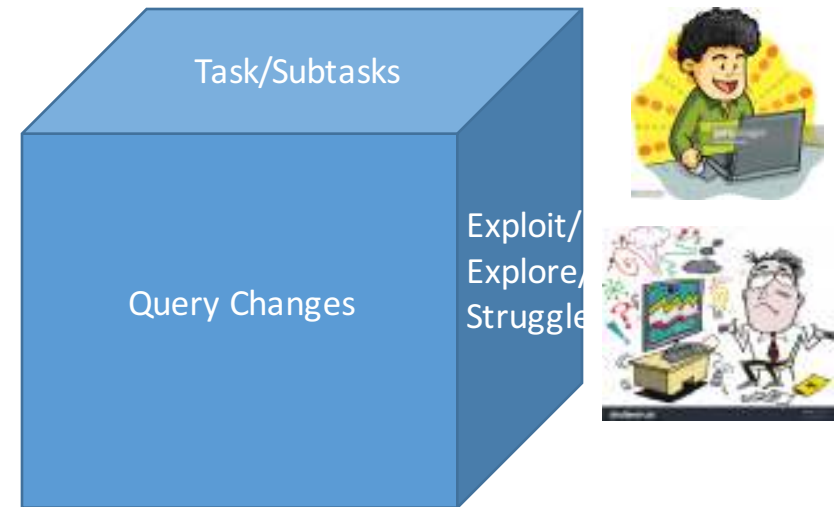
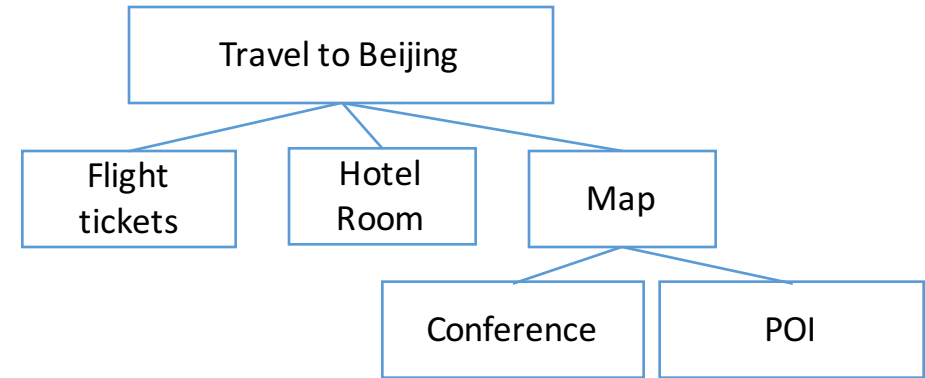
- Improve retrieval performance for a given query by using the session prior to this query

TREC Session: Test Collection

```
<session num="1" starttime="08:59:47.258675">
  <topic>
    <title>peacecorp</title>
    <desc>Find information about the peace corp</desc>
    <narr>When was it started and by whom? What services does it provide and where does it provide these
    services? What is the criteria for applying? What is the salary or stipend? What positions are available?</narr>
  </topic>
  <interaction num="1" starttime="09:00:04.155323">
    <query>peace corp</query>
    <results>
      <result rank="1">
        <url>http://www.peacecorps.gov/</url>
        <clueweb09id>clueweb09-en0011-60-08003</clueweb09id>
        <title>Peace Corps</title>
        <snippet>Fighting hunger, disease, poverty, and lack of opportunity.</snippet>
      </result>
      ...
      <result rank="10">
        <url>http://www.peacecorpsonline.org/</url>
        <clueweb09id>clueweb09-en0133-67-12909</clueweb09id>
        <title>Peace Corps Online</title>
        <snippet>Online community of returned Peace Corps volunteers.</snippet>
      </result>
    </results>
    <clicked>
      <click num="1" starttime="09:00:09.943356" endtime="09:01:13.434255">
        <rank>1</rank>
      </click>
      <click num="2" starttime="09:01:18.582078" endtime="09:02:42.552354">
        <rank>2</rank>
      </click>
    </clicked>
  </interaction>
  <interaction num="2" starttime="09:02:55.569644">
    <query>peace corp apply</query>
    <results>
      ...
    </results>
    <clicked>
      <click num="1" starttime="09:03:02.615239" endtime="09:03:33.507677">
        <rank>2</rank>
      </click>
    </clicked>
  </interaction>
  <currentquery starttime="09:04:03.469341">
    <query>peace corp application</query>
  </currentquery>
</session>
```

Query change

- Changes in the query
 - Adding a term
 - Removing a term
 - Keeping a term
- Correlations between $\Delta(\text{query})$ and feedback
- Task stage
 - Sub-tasks
- User stage
 - Struggling
 - Exploring
 - Exploiting



Dialog Systems

- Fully conversational system
 - Search algorithm asking questions to the user



Conclusions

- Professional Search
 - Exploratory
 - Complex
 - Recall-oriented
- Fully conversational systems
 - Receive feedback
 - Documents
 - Query rewrites
 - Explicitly ask for feedback