

# De Autonome Zoekmachine

Maarten de Rijke

# Dankwoord

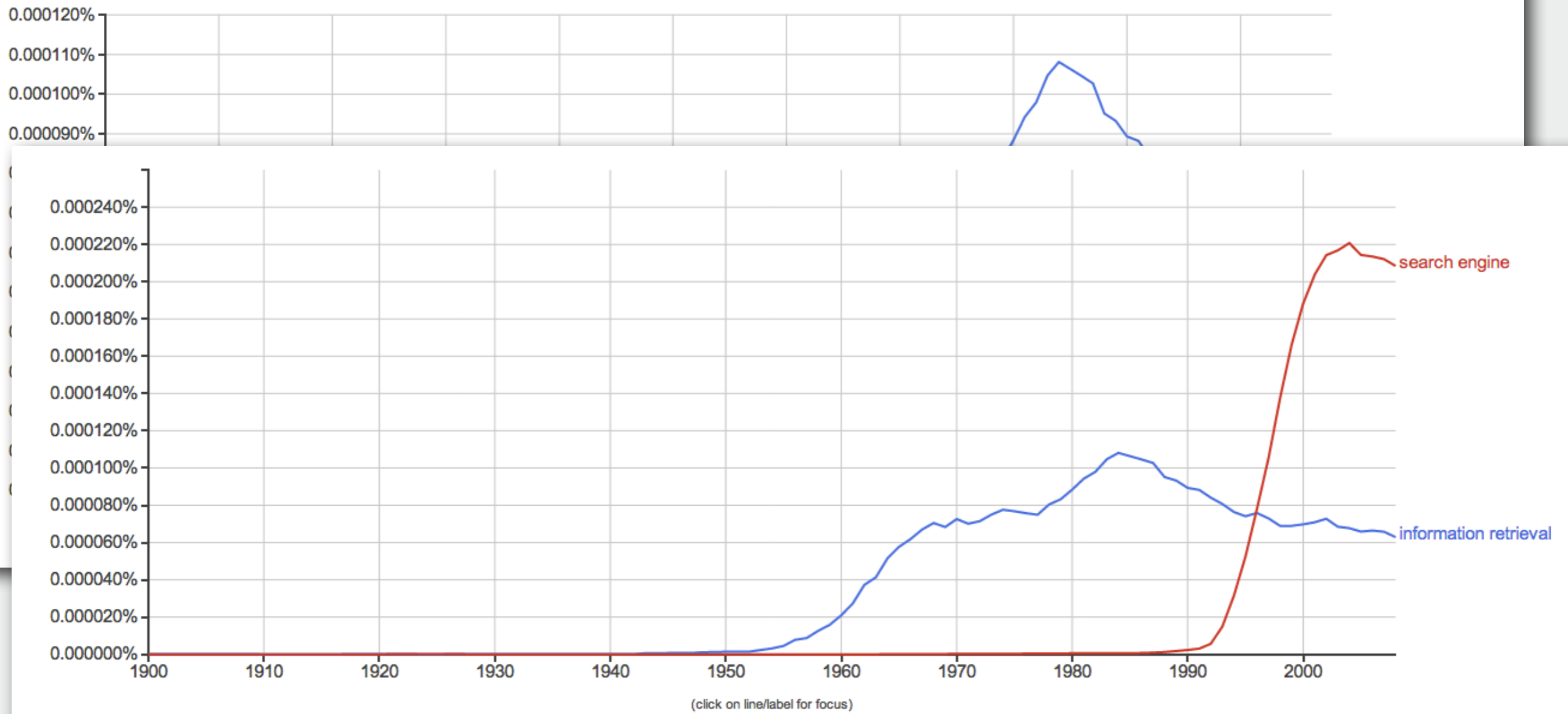
- Gebaseerd op gezamenlijk werk met Aleksandr Chuklin, Katja Hofmann, Damien Lefortier, Remi Munos, Anne Schuth, Floor Sietsma, Pavel Serdyukov, Shimon Whiteson, Masrour Zoghi.
- Gefinancierd door COMMIT/, ESF, EU, KNAW, Microsoft Research, NWO, Yahoo!, Yandex

**Er was eens ...**

# De geboorte van “information retrieval”

- “The problem under discussion here is machine searching and retrieval of information from storage according to a specification by subject ... It should not be necessary to dwell upon the importance of **information retrieval** before a scientific group such as this for all of us have known frustration from the operation of our libraries – all libraries, without exception.”
  - C.N. Mooers. ‘The theory of digital handling of non-numerical information and its implications to machine economics’, in *Association for Computing Machinery Conference*, Rutgers University, 1950.

# En daarna ...



★ My Schedule



# SCHEDULE

## Search Is Dead: The Secret Sauce Behind Discovery



[Tweet](#)

#sxsw #RIPSearch

Next Generation Recommendations: We live in a world where search is dead. Consumers are most responsive to personalized, content-driven and curated experiences and are no longer as willing to take the time to search and discover new brands, products, music, etc. on their own. How is curated content impacting consumption patterns and purchase decisions? How are retailers and vendors responding to this shift and enhancing their offerings to provide consumers the personalized experiences they want? What's the secret sauce behind their method, who will come out on top and why?

This panel of diverse business leaders and industry experts will discuss the importance of delivering a personalized, content-driven experience for customers and why they chose to pursue a recommendation, rather than search, based approach.

### Presenters

★ Monday, March 10  
12:30PM - 1:30PM

[Add to my schedule](#)

#### VENUE INFO

**Driskill Hotel**  
**Driskill Ballroom**  
**604 Brazos St**

Thanks to our sponsors



**esurance**

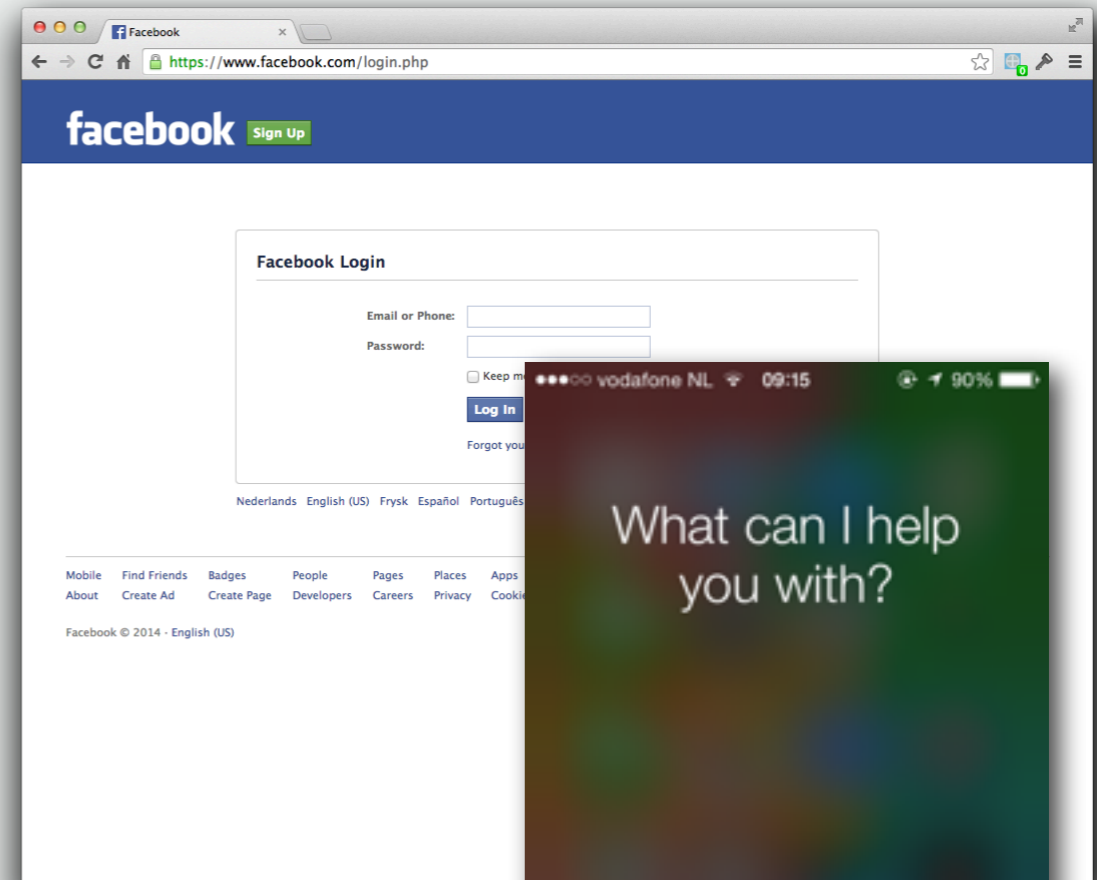


CREDENTIALS WITH ACCESS  
Music Badge, Film Badge,  
Interactive Badge, Gold Badge.



# Nieuwe zoekparadigma's

- Zero query scenario
- “Jij bent de zoekvraag”
- Zoekmachine als assistent bij complexe taken



# Onder de motorkap ...



# De geordende lijst

- Zoekmachines produceren geordende lijsten van items in reactie op query
- Combinaties of honderden of “ranking features”
  - query, document, query +document, gebruiker, wereld, sociale kring, taak, ...

Sl. no.	Name of journal	Country	Frequency	Rank
1	Knowledge Organization	Germany	41	1
2	Bell System Technical Journal	USA	37	2
3	IEEE Multimedia	USA	28	3
4	Cataloging & Classification Quarterly	USA	15	4
5	Library Trends	USA	14	5
6	Visual Resources	USA	11	6
7	Communication of the ACM	USA	10	7
8	Library Quarterly			
9	Library Resources & Technical Service			
10	Scientometrics			
11	Extinction's and Corrections to UDC			
12	International Journal of Lexicography			
13	Artificial Intelligence			
14	Human-Computer Interaction			
15	Bibliobase			
16	Byte			
17	Computer and History of Arts			
18	Library Association Record			
19	American			
20	Inform			
21	World			
22	Annuaire			
23	Techn			
24	ACM			
25	Current			
26	Educ			
27	Intern			
28	and Di			
	Librar			

Country/Economy	Rank	Score	Rank*
Singapore	1	6.14	1
Hong Kong SAR	2	5.67	2
Denmark	3	5.41	3
Sweden	4	5.30	4
			6
			12
			10
			5
			8

NAME OF COMPANY AND WEBSITE	POPULOUS TOWN (MILLION)	CEO	NUMBER OF EMPLOYEES	NUMBER OF LOCATIONS (PROCESSED)	NATIONS WHERE OPERATING
European Retail Recycling (ERR) - www.err.nl	11 million	Chris Sheppard	3,000	200	U.K., U.S., Germany, The Netherlands
Stacy Group - www.stacygroup.com	10.7 million	Baruch Uziel (Chief Officer)	7,000	100	Germany, Austria, Poland, Czech Republic, Slovakia, Hungary, Bulgaria, Romania, Turkey, Slovenia, Croatia, Bosnia, Serbia
Stacy Retail Management - www.stacy.com	10.2 million	Daniel David	6,000	200	Europe, North America, Australia, Asia
Stacy Recycling - www.stacy.com	9 million (est.)	Uliak Abulali	1,000	100	
DavidSource Corp./Steel Dynamics Inc. - www.dsource.com	6.2 million	Russell Rice	2,000+	41	
David J. Joseph Co./Honor Corp. - www.honor.com	5 million	Ruth Green	2,200	68	
Take City LLC - www.takecity.com	5 million	Joseph Curtis	3,000	25	
Derks&Berg Environment - www.derksberg.com	4.8 million	Frank Derks/Berg	3,000	148	

Category 1: Stability (weight: 25% of total)

Indicator	Score
Prevalence of petty crime	ES3 rating
Prevalence of violent crime	ES3 rating
Threat of terror	ES3 rating
Threat of military conflict	ES3 rating
Threat of civil unrest/conflict	ES3 rating

福布斯2011中国慈善基金会榜前十名

排名	基金会名称	性质	2010年度 捐赠收入(万元)	2010年度 公益支出(万元)	评分 (总计100分)
1	上海真爱梦想公益基金会	地方性非公募	1719	921	97
2	友成企业家扶贫基金会	全国性非公募			
3	中国红十字基金会	全国性非公募			
4	慈济慈善事业基金会	全国性非公募			
5	深圳壹基金公益基金会	地方性非公募			
6	南都公益基金会	全国性非公募			
7	中国扶贫基金会	全国性非公募			
8	中国扶贫基金会	全国性非公募			
9	中国扶贫基金会	全国性非公募			
10	中国扶贫基金会	全国性非公募			

The screenshot shows a Yandex search engine interface. The search query is "Information retrieval" with 44 million answers. The results page shows several book listings from Amazon.com, including "Introduction to Information Retrieval" by Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze, and "Information Retrieval: Implementing and Evaluating Search Engines" by Stefan Buitner, Charles L. A. Clarke and Gordon V. Cormack. The page also shows a video player and a sidebar with related content.

# “Rankers”

- Traditioneel, handwerk
- Denk goed na over wat het betekent voor document om te **matchen** met query
  - Combinatie van term frequentie, document frequentie, document lengte
  - E.g.,



$$BM25(q, d) = \sum_{q_i: tf(q_i, d) > 0} \frac{idf(q_i) \cdot tf(q_i, d) \cdot (k_1 + 1)}{tf(q_i, d) + k_1 \cdot (1 - b + b \cdot \frac{|d|}{avgdl})} \cdot \frac{(k_3 + 1) \cdot qtf(q_i, q)}{k_3 + qtf(q_i, q)}$$

# Onderzoek naar “rankers”

## ■ Analysis

- Individuele ranking criteria

## ■ Synthese

- Combineren van de uitkomsten van grote aantallen “rankers”

## ■ Evaluatie

- Meten van de kwaliteit van de uitkomsten van (doorgaans gecombineerde) “rankers”

# Analyse van “rankers”

## ■ Inhoud

- Inhoudelijke match
- Semantische analyse
- Subjectieve aspecten (sentiment, framing, reputatie, impact)

## ■ Structuur

- Document structuur
- Link structuur (PageRank)
- Sociale structuur

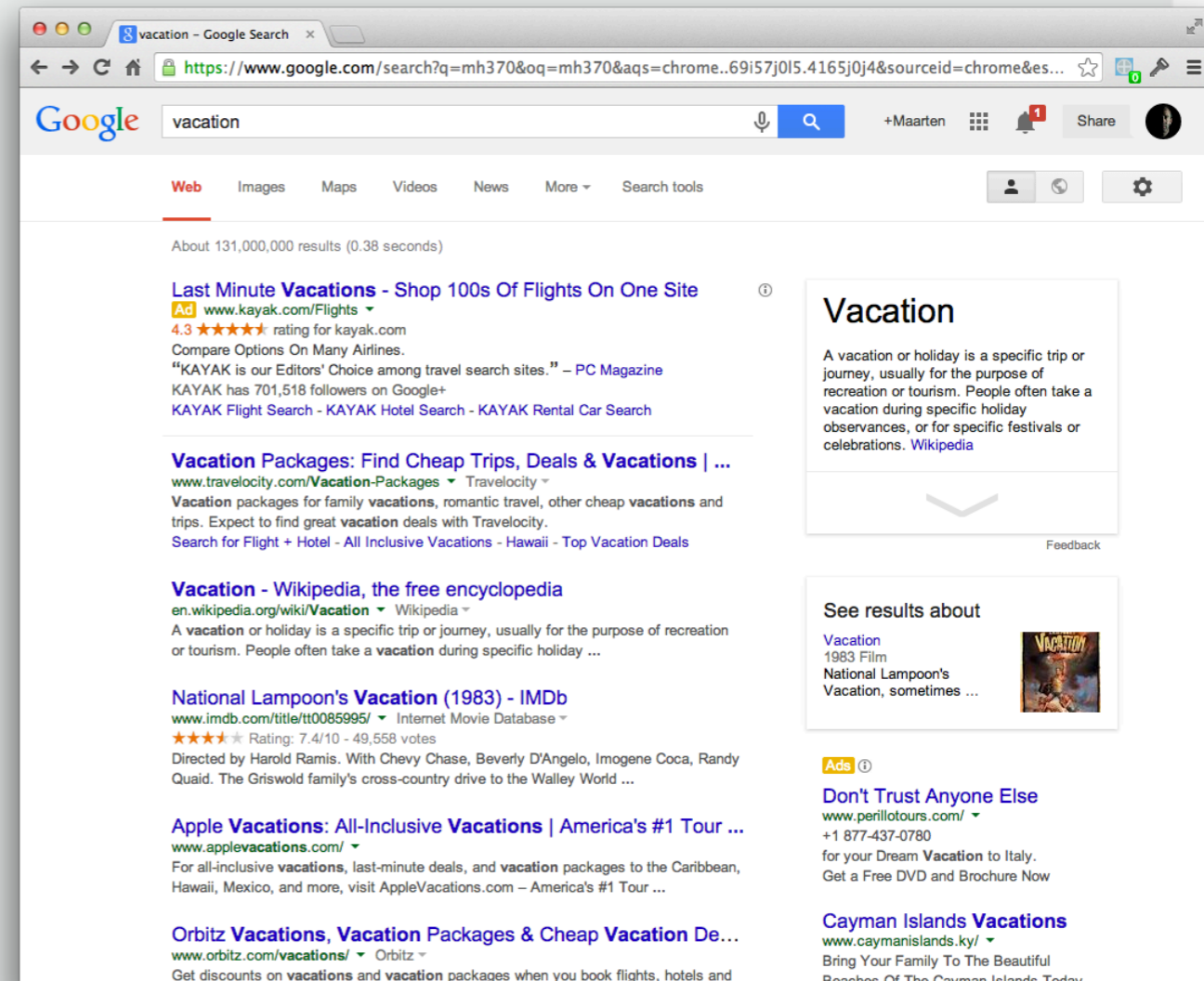
## ■ Gedrag

- Interactie data (“Click Through Rate,” CTR)
- Impliciet vs expliciet



# Synthese van “ranker” resultaten

- Lange lijn van work aan “fusie” of “rank aggregatie”
- Federated search
  - Verticals
- Aggregated search pages
  - Diversiteit
- Learning to rank
  - Point-wise, pair-wise, list-wise



# Evaluatie

## ■ Offline

- Verzamel queries, vraag beoordelaars om te bepalen welke documenten relevant zijn, evalueer “rankers” gebaseerd op de kwaliteit van de ranked lists

## ■ Gebruikersstudies

- Voorzie een kleine verzameling gebruikers van verschillende retrieval systemen; vraag ze wat om wat taken uit te voeren; leer meer over rankers door mensen te vragen wat ze doen en waarom ze het doen

## ■ Online

- Kijk hoe normale gebruikers interactie hebben met live retrieval systemen; observeer impliciet gedrag (clicks, skips, saves, bookmarking, ...); probeer verschillende verschillen in gedrag af te leiden uit verschillende smaken van het live systeem

## Take home boodschap

Van supervised

Analyse van ranking features

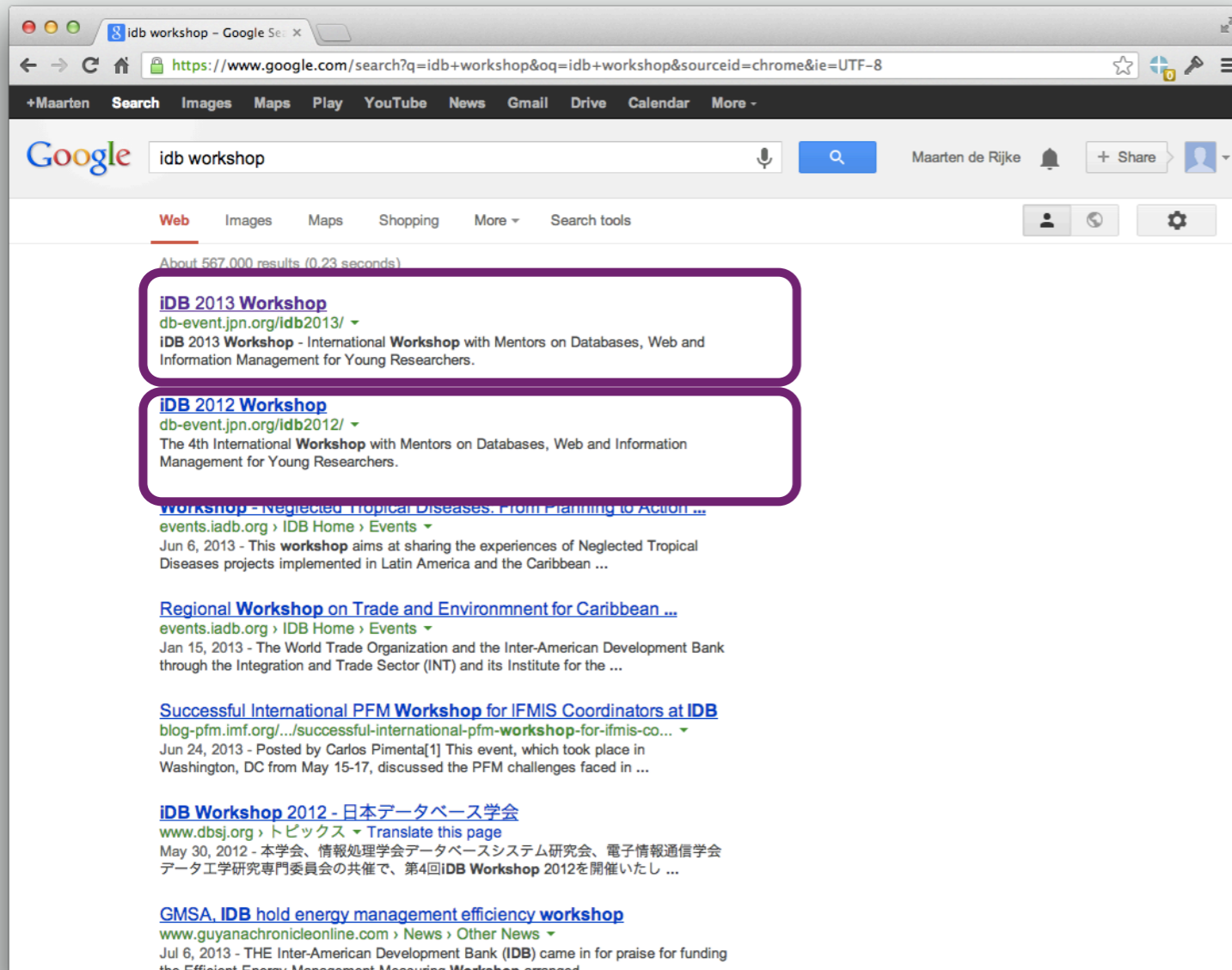
naar

Evaluation van de uitkomsten ranking features

unsupervised



# Evaluation met behulp van clicks

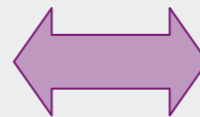


- Clicks zijn goed...
  - Zijn deze twee clicks even “goed”?
- Geen-clicks kunnen veel redenen hebben:
  - Niet relevant
  - Niet bekeken
  - De snippet gaf het antwoord al

# Leren van van natuurlijke interacties met een retrieval systeem

(d.w.z., clicks op zoekresultaten)

- Makkelijk te vergaren terwijl systeem draait
- Weerspiegelt natuurlijk gedrag plus voorkeuren
- Retrieval systeem leert online



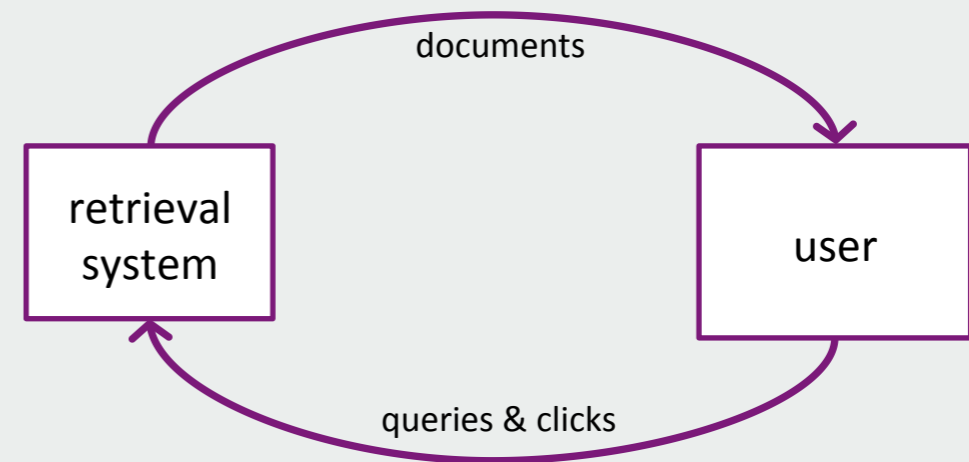
- Ruizig
- Alleen relatieve oordelen en voorkeuren

**Hoe kunnen retrieval systemen betrouwbaar en efficiënt leren van ruizige, relatieve feedback?**

# Benadering

## “Reinforcement Learning” benadering

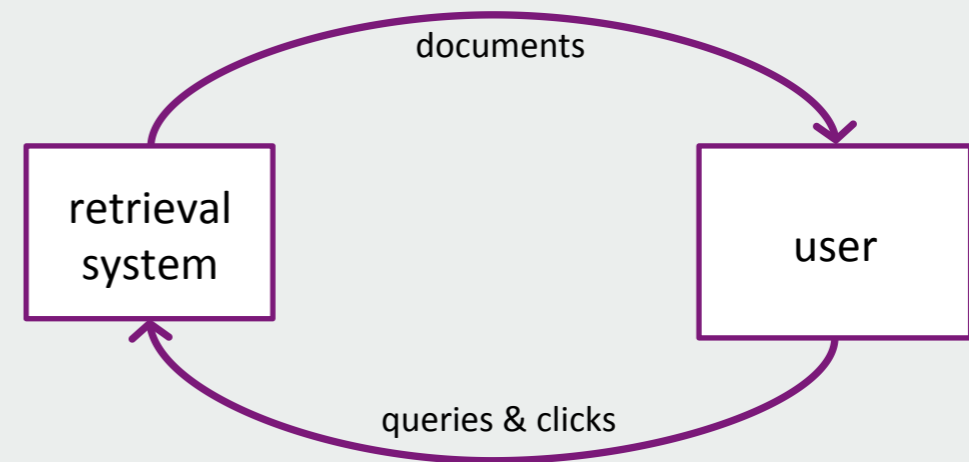
- Leer door verschillende acties (rankings) uit te proberen en feedback te observeren
- Bekijk twee rankers tegelijk



# Benadering

## “Reinforcement Learning” benadering

- Leer door verschillende acties (rankings) uit te proberen en feedback te observeren
- Bekijk twee rankers tegelijk

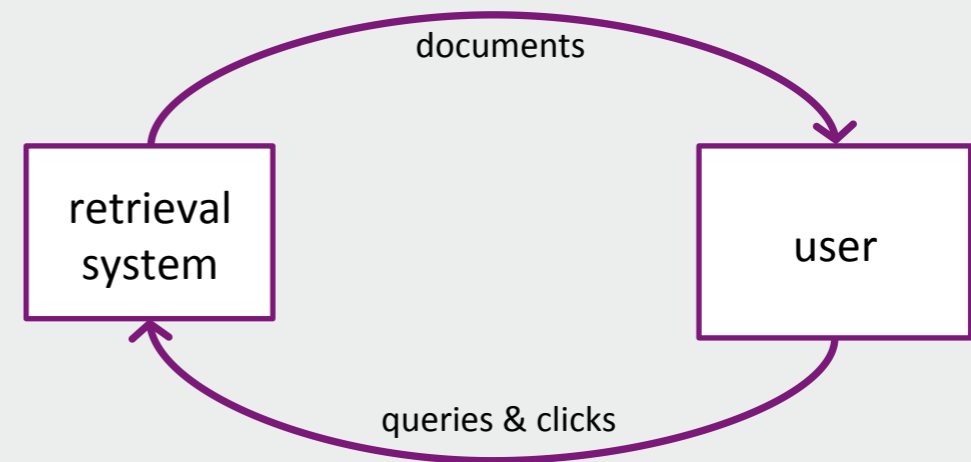


- 1 observeer query
- 2 genereer resultaat lijst
- 3 leid feedback af uit clicks
- 4 update retrieval function

# Benadering

## “Reinforcement Learning” benadering

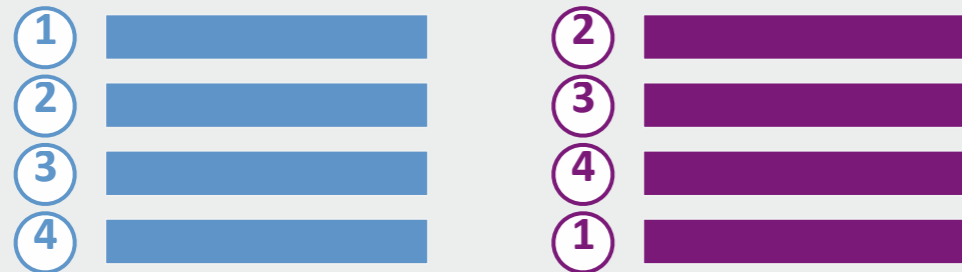
- Leer door verschillende acties (rankings) uit te proberen en feedback te observeren
- Bekijk twee rankers tegelijk



- ① observeer query
  - ② genereer resultaat lijst
  - ③ leid feedback af uit clicks
  - ④ update retrieval function
- } Hoe betrouwbaar feedback afleiden?  
 } Hoe efficient online leren?

# Unsupervised evaluatie

# Interleaved comparisons methode — unsupervised evaluatie



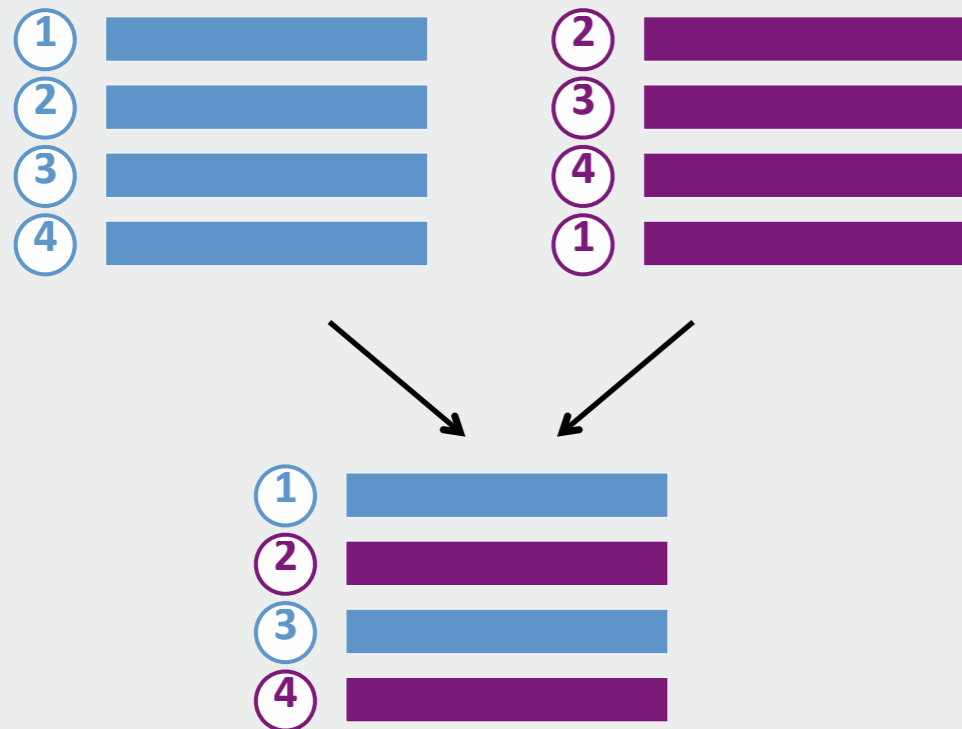
■ **Doel:** Vergelijk twee resultaat lijsten m.b.v. click data

■ **Procedure:**

- 1) Genereer interleaved resultaat lijst
- 2) Observeer gebruiker clicks
- 3) Geef krediet aan oorspronkelijke rankers

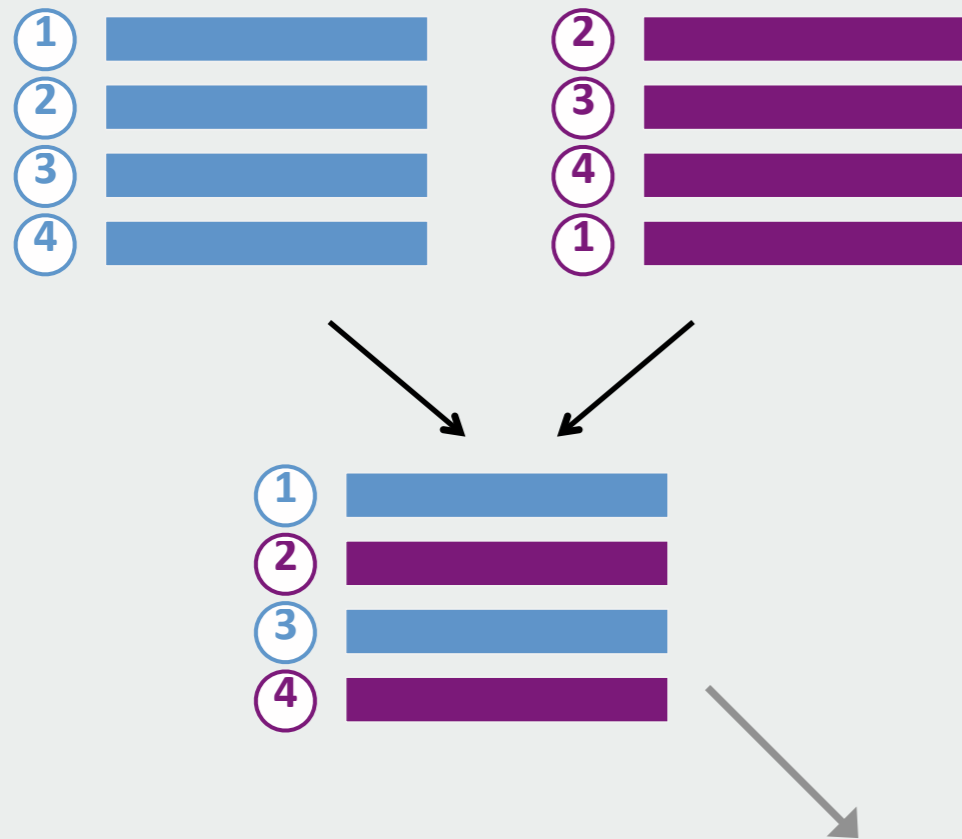


# Interleaved comparisons methode — unsupervised evaluatie



- Doel: Vergelijk twee resultaat lijsten m.b.v. click data
- Procedure:
  - 1) **Genereer interleaved resultaat lijst**
  - 2) Observeer gebruiker clicks
  - 3) Geef krediet aan oorspronkelijke rankers

# Interleaved comparisons methode — unsupervised evaluatie



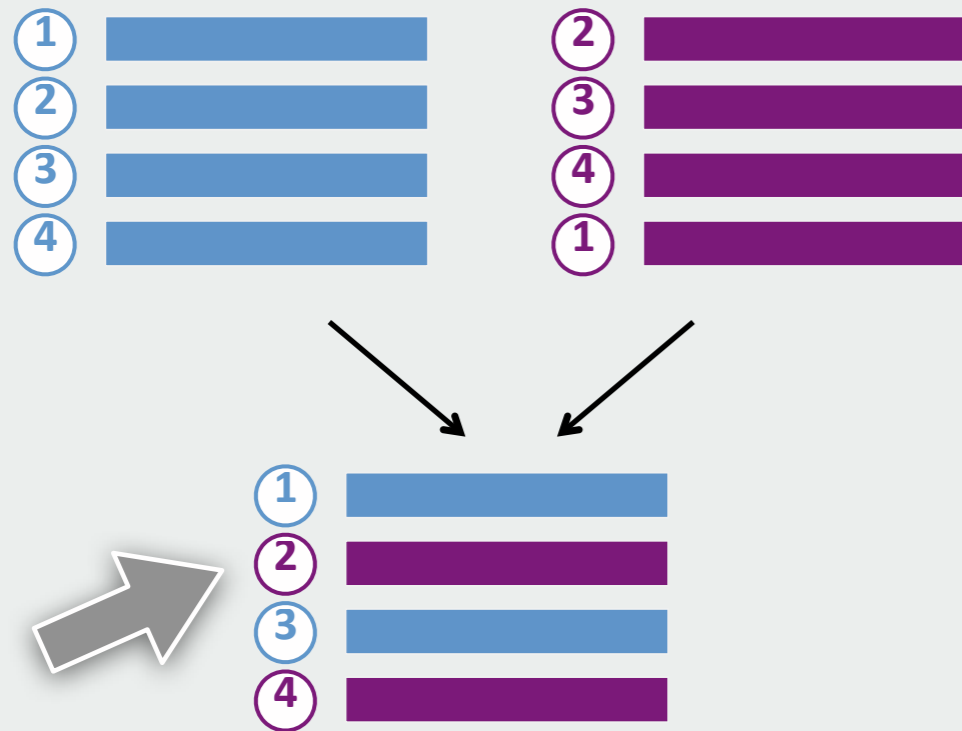
■ Doel: Vergelijk twee resultaat lijsten m.b.v. click data

■ Procedure:

- 1) **Genereer interleaved resultaat lijst**
- 2) Observeer gebruiker clicks
- 3) Geef krediet aan oorspronkelijke rankers

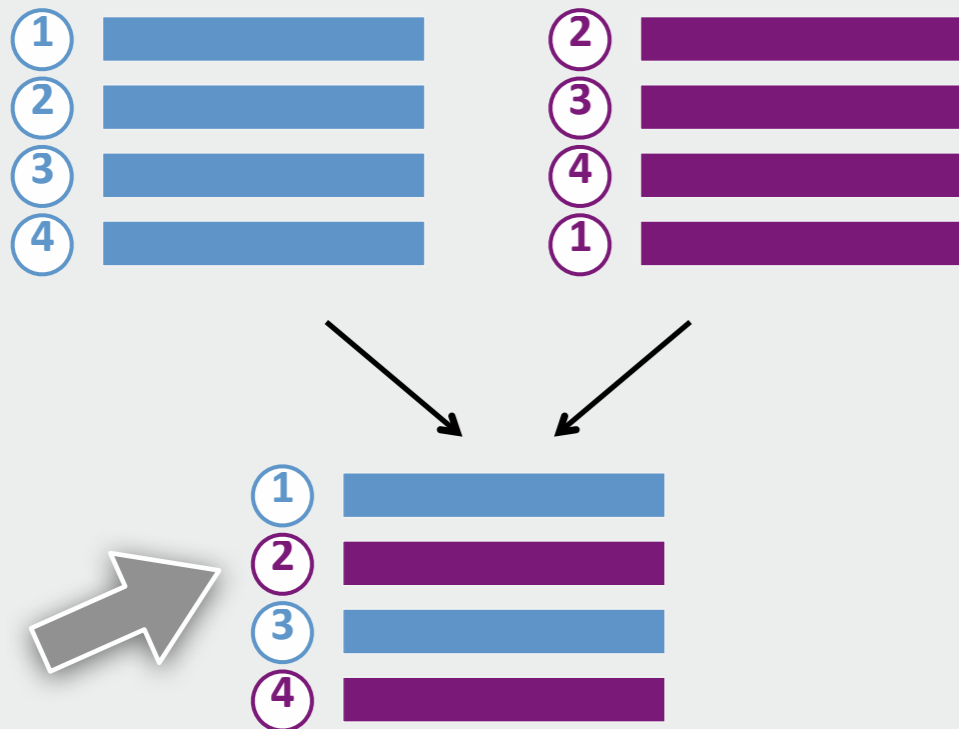
Belangrijk: Houd donaties bij  
(welke lijst heeft welk document bijgedragen)

# Interleaved comparisons methode — unsupervised evaluatie



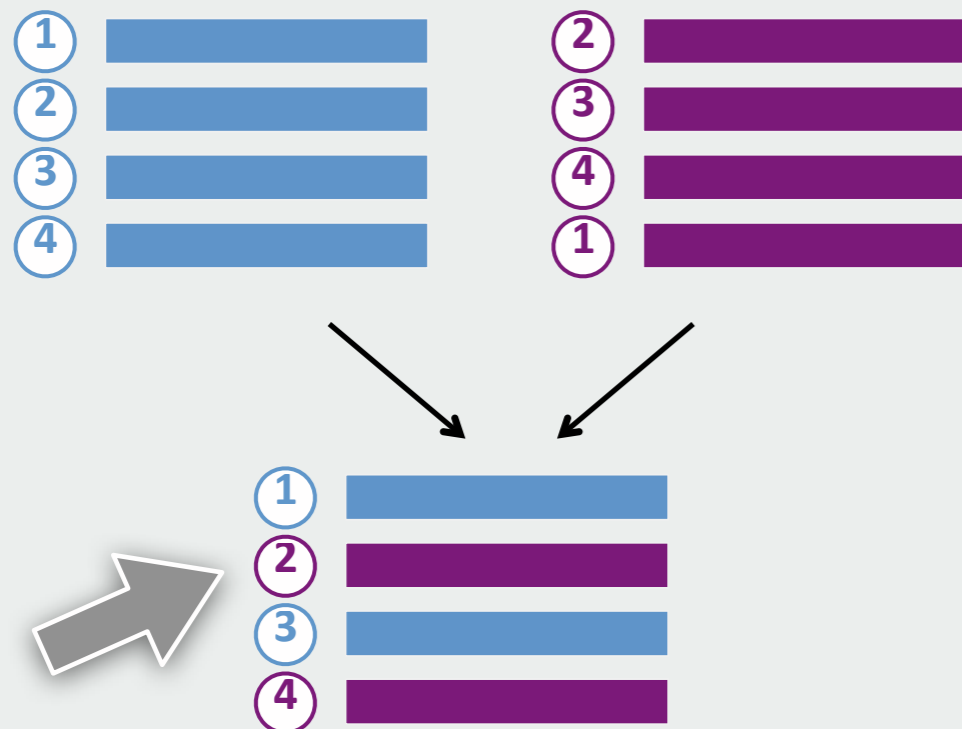
- Doel: Vergelijk twee resultaat lijsten m.b.v. click data
- Procedure:
  - 1) Genereer interleaved resultaat lijst
  - 2) **Observeer gebruiker clicks**
  - 3) Geef krediet aan oorspronkelijke rankers

# Interleaved comparisons methode — unsupervised evaluatie



- Doel: Vergelijk twee resultaat lijsten m.b.v. click data
- Procedure:
  - 1) Genereer interleaved resultaat lijst
  - 2) Observeer gebruiker clicks
  - 3) **Geef krediet aan oorspronkelijke rankers**

# Interleaved comparisons methode — unsupervised evaluatie



Kun je hiermee nauwkeurig verschillen tussen resultaatlijsten afleiden?

- Doel: Vergelijk twee resultaatlijsten m.b.v. click data
- Procedure:
  - 1) Genereer interleaved resultaat lijst
  - 2) Observeer gebruiker clicks
  - 3) **Geef krediet aan oorspronkelijke rankers**

# Recente resultaten

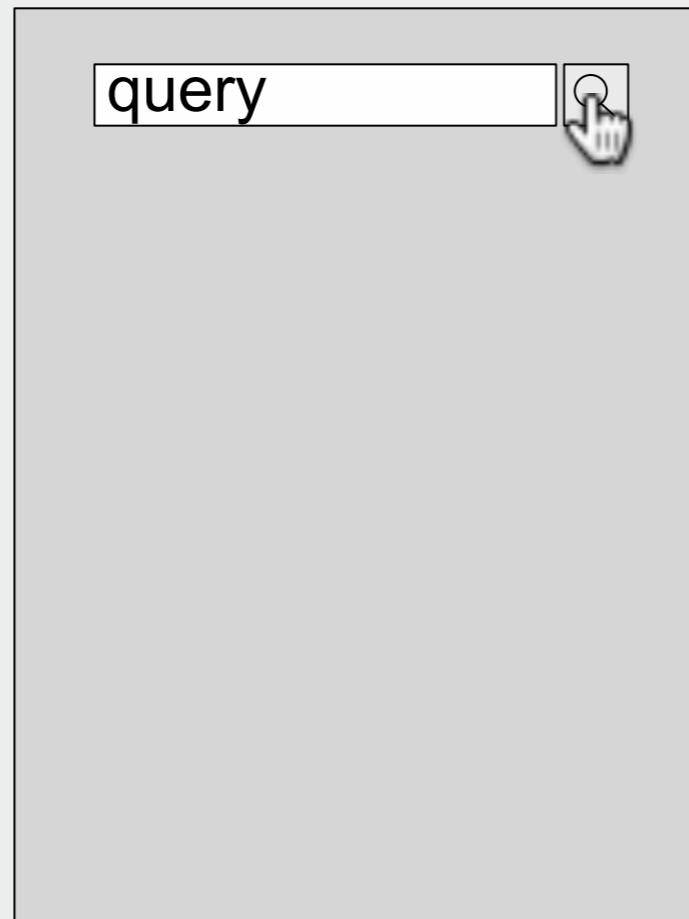
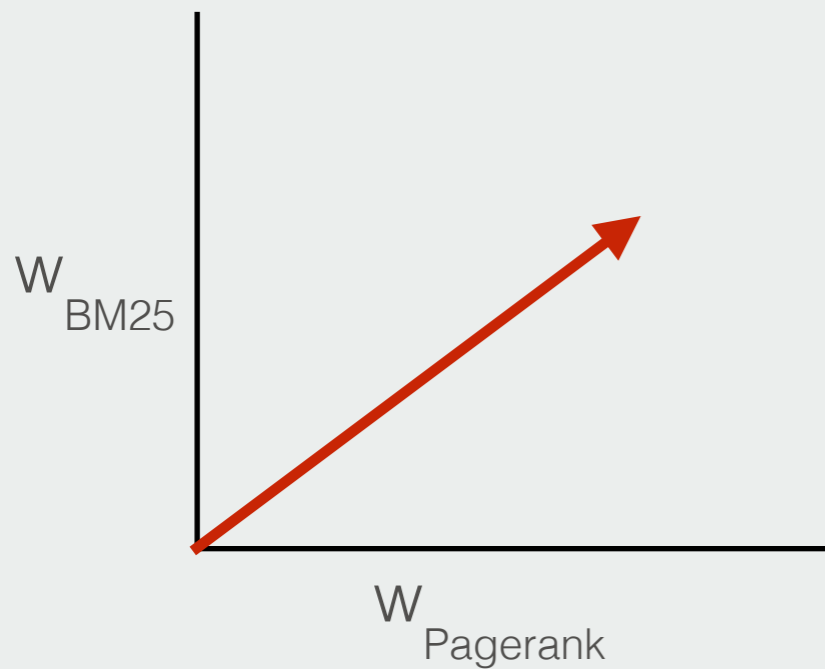
- Betrouwbaarheid (“wordt de juist winnaar aangewezen”) en efficiëntie (“hoeveel interacties zijn nodig”) aanzienlijke verbeterd
  - In productie
- Aanpassingen voor hergebruik van oude gebruikersinteractie
- Uitvoerige vergelijkingen met offline evaluatie en gebruikersstudies

# Unsupervised synthese

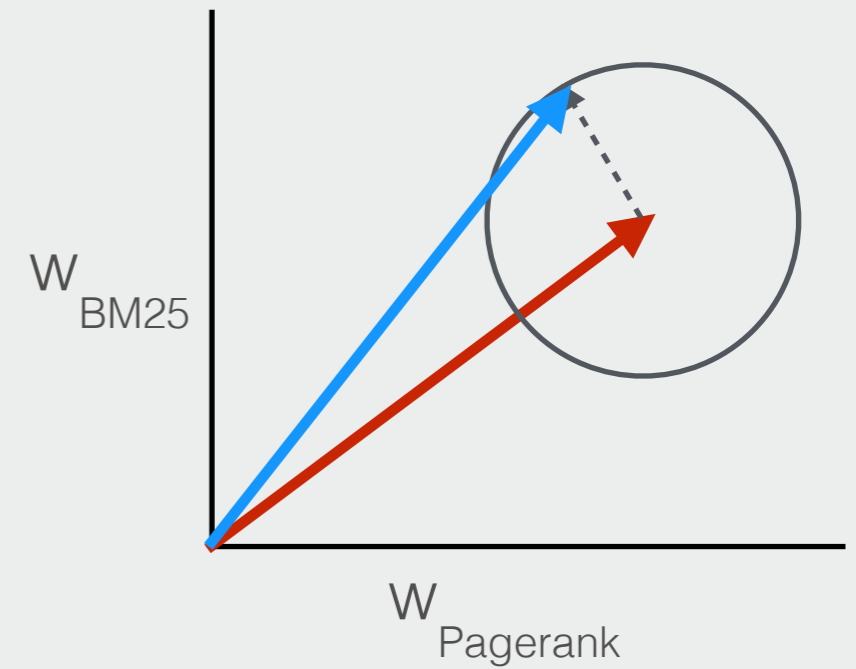


# Unsupervised aanpassen van combinaties van rankers

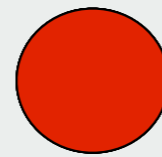
## Exploitative ranker



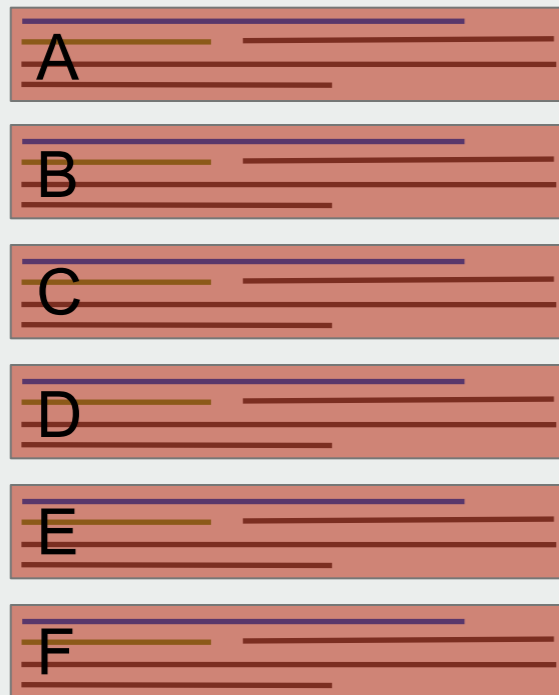
## Explorative ranker



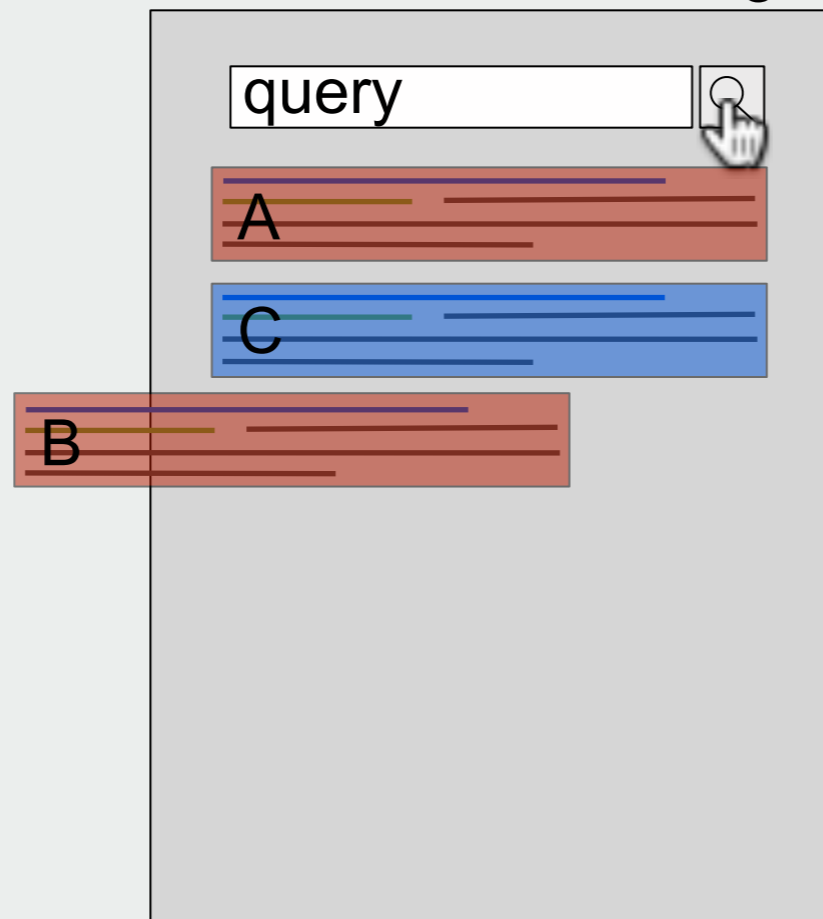
# Unsupervised aanpassen van combinaties van rankers



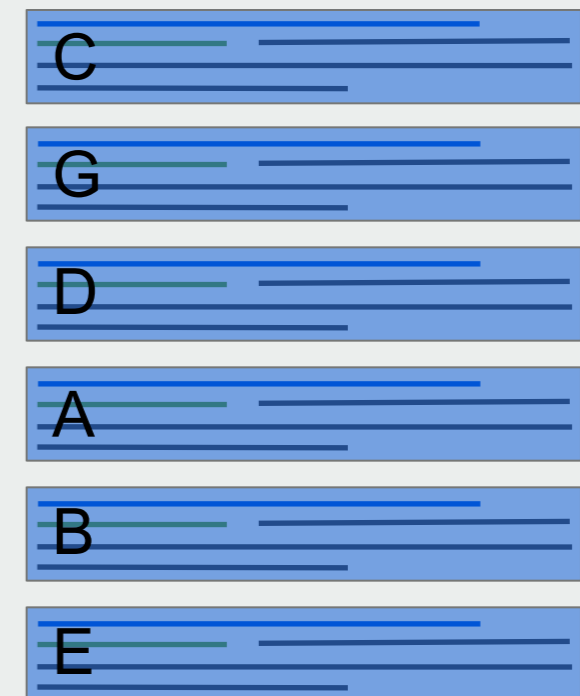
## Exploitative Ranking



## Interleaved Ranking

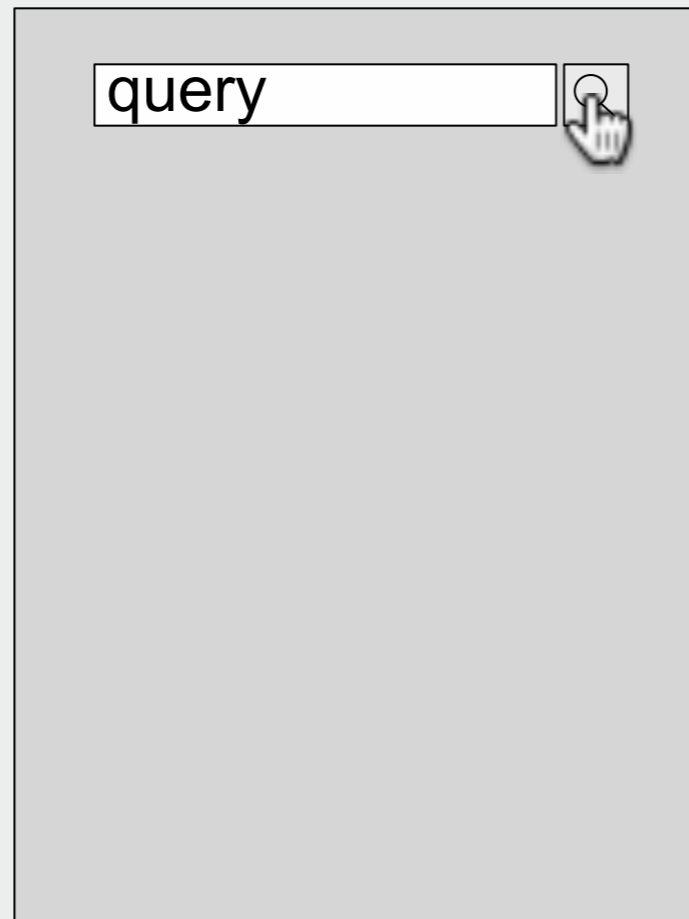
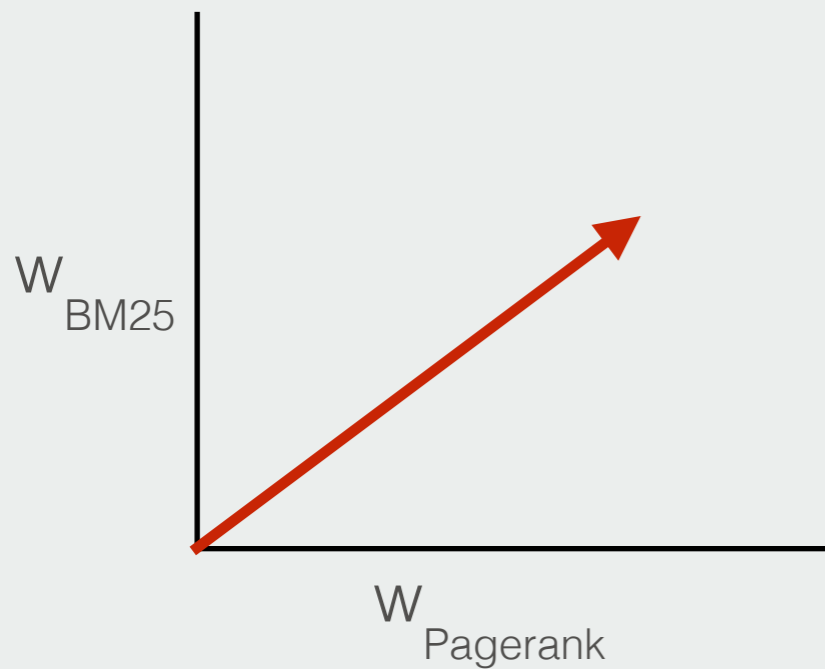


## Explorative Ranking

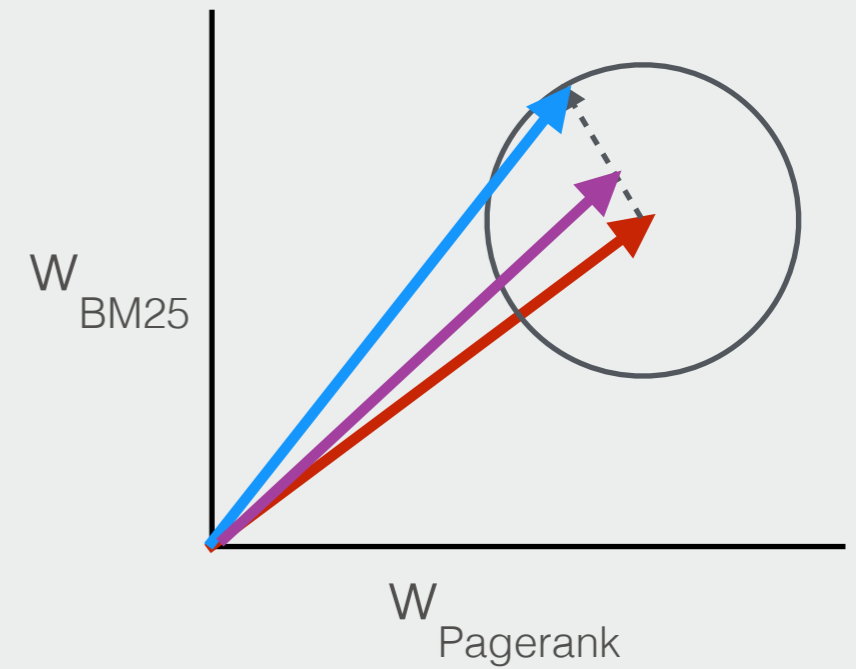


# Unsupervised aanpassen van combinaties van rankers

## Exploitative ranker

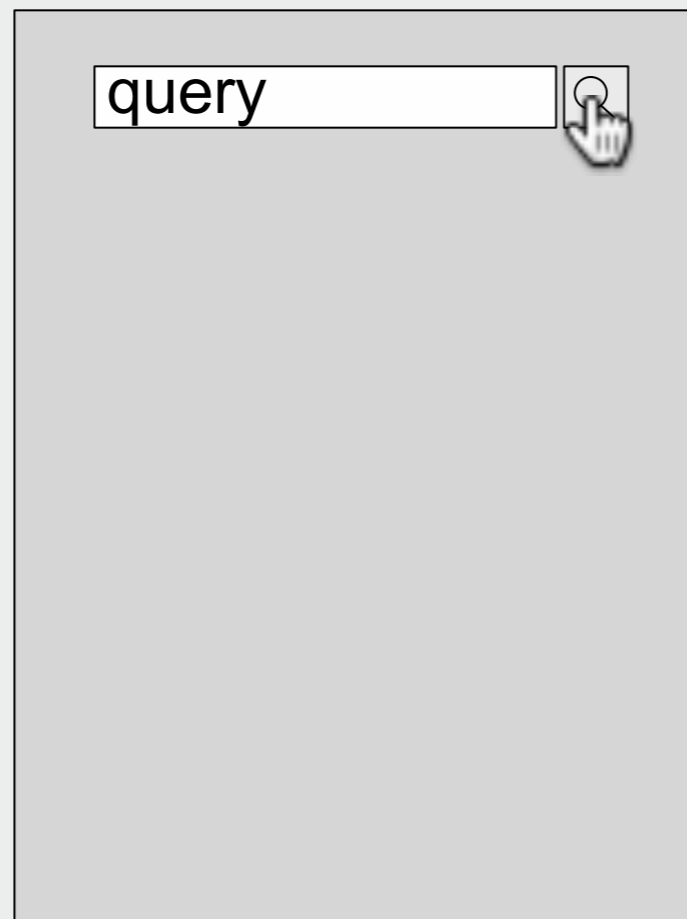
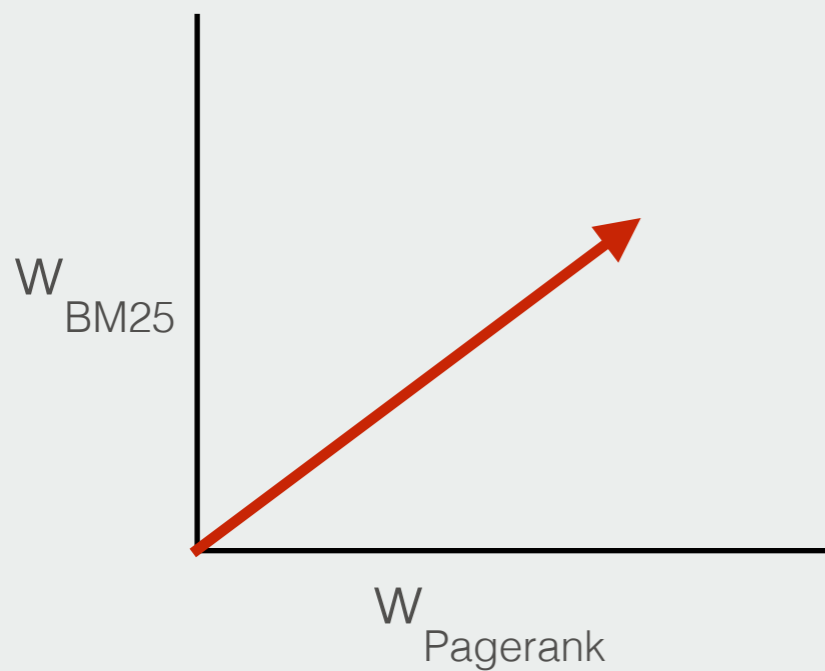


## Explorative ranker

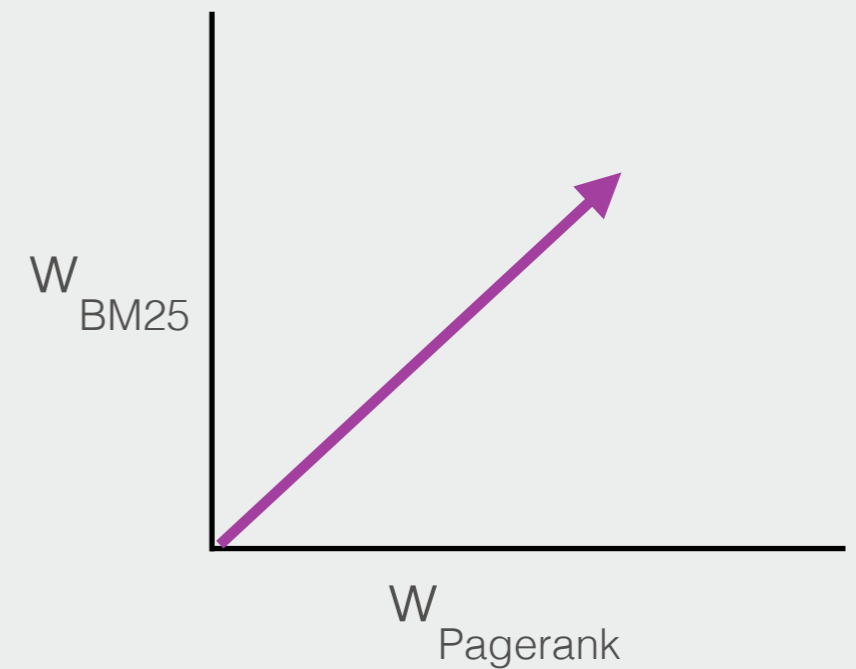


# Unsupervised aanpassen van combinaties van rankers

## Exploitative ranker

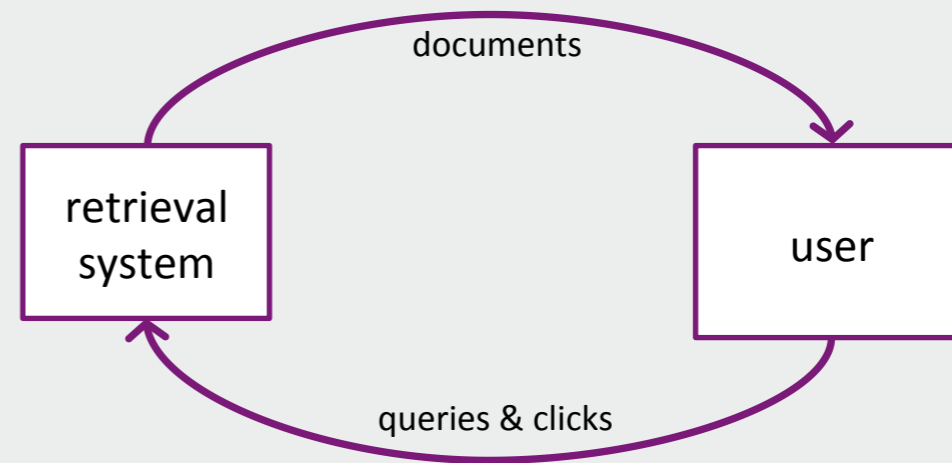


## Explorative ranker

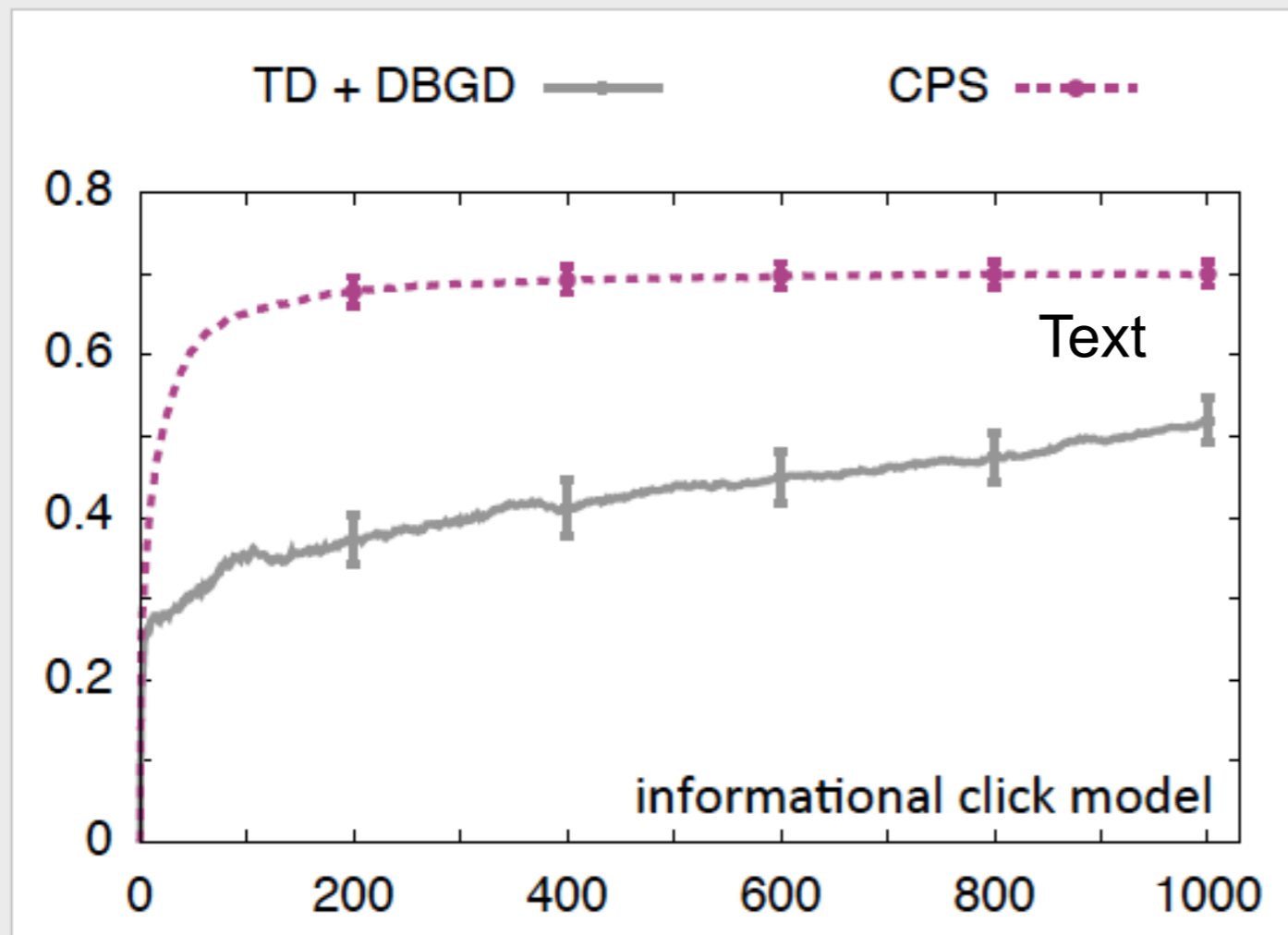


# Experimentele uitdaging

- Onderzoek heeft live systemen nodig
  - **Breng vandaag uw systeem! @mdr**
- Alle gebruikers zijn “guinea pigs”
- Simulaties



# Recente resultaten



## ■ Kandidaat selectie

- Kies kandidaten goed zodat je weinig mogelijk “risicovolle” explorative rankers
- Bijv. draai “achter de schermen” “toernooitjes” tussen alternatieve explorative rankers op historische data

## ■ Risico garanties

# Unsupervised analyse van ranking features





# Ontdek nieuwe rankers

- Traditionele supervised machine leer model
  - Selecteer een aantal ranking features en leer dan hoe je die optimaal combineert om een zo krachtige mogelijke ranker te produceren gegeven de ranking features die je hebt
- **Doel:** genereer een werkelijk autonome zoekmachine die deze ranking features automatisch leert
  - Hamvraag: Kunnen we gebruikersgedrag inzetten om die nieuwe kenmerken te ontdekken?



# Stapje terug

- Bestaande ranking features zijn ontworpen om het redelijk goed te doen “over alle queries”
- Alternatieve benadering
  - Specialistische ranking features, die als doel hebben om uiterst nauwkeurig te werken op een deelverzameling van de queries
  - Zelfs wanneer dat betekent dat ze het over het geheel genomen niet zo goed doen

# Benadering

- Nodig: uiterst gespecialiseerde rankers en een classifier die elke query aan de juiste “specialist” toekent
  - **Query clustering:** aggregeer queries aan de hand van hun “intent,” zo dat soortgelijke rankers effectief zijn voor hen
  - **Ranker specialisatie:** simultaan selectie van queries en features die het goed doen voor die queries
  - **Intent classificatie:** ken elke nieuwe query toe aan de juiste specialisten

# Zoekmachines zonder supervisie

- Leren van interacties met gebruikers
  - ✓ Evaluatie
  - ✓ Synthese van rankers
  - Creëren van individuele rankers ... in uitvoering

- De Autonome Zoekmachine
- Maarten de Rijke
- [derijke@uva.nl](mailto:derijke@uva.nl)
- @mdr