Information Retrieval
Click Models

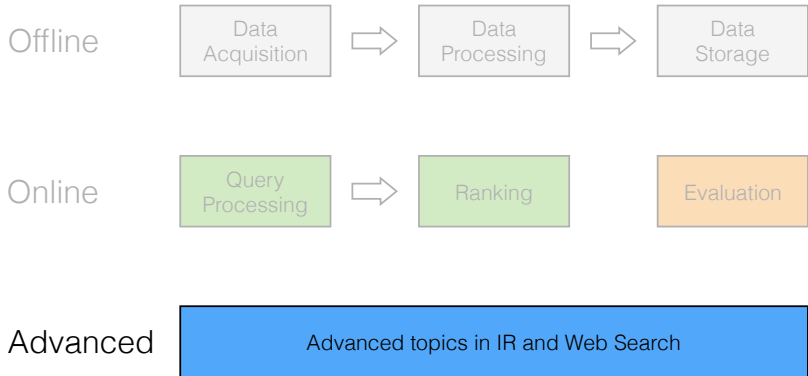**Ilya Markov**
i.markov@uva.nl

University of Amsterdam
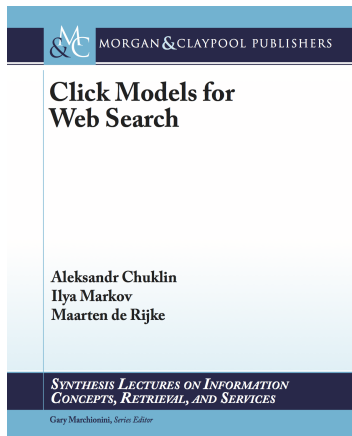
Course overview

Offline — Data Acquisition ⟹ Data Processing ⟹ Data Storage

Online — Query Processing ⟹ Ranking — Evaluation

Advanced — Advanced topics in IR and Web Search

Advanced topics in IR

Offline

| Data Acquisition | $\Rightarrow$ | Data Processing | $\Rightarrow$ | Data Storage |

Online

| Query Processing | $\Rightarrow$ | Ranking | | Evaluation |

Advanced

Advanced topics in IR and Web Search

The book
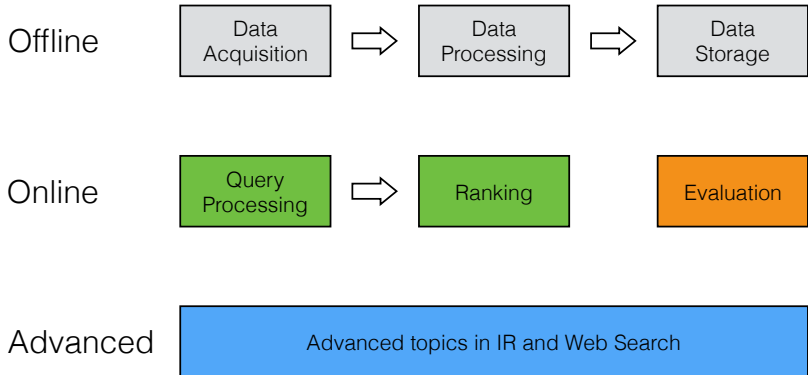
http://clickmodels.weebly.com/the-book.html

Tutorials

- SIGIR 2015, Santiago, Chile
- AINL-ISMW FRUCT 2015, St. Petersburg, Russia
- WSDM 2016, San Francisco, USA
- RuSSIR 2016, Saratov, Russia

  `http://clickmodels.weebly.com/tutorials.html`

Outline

1. Applications

2. Advanced models

3. Summary

Click models summary so far

- Basic click models
  - CTR models
  - Position-based model
  - Cascade model
- Click probabilities
  - Full click probabilities
  - Conditional click probabilities
- Evaluation
  - Perplexity
  - Log-likelihood
- Parameter estimation
  - Maximum likelihood estimation
  - Expectation-maximization

What do click models give us?

General:

- Understanding of user behavior

Specific:

- Conditional click probabilities
- Full click probabilities
- Attractiveness and satisfactoriness for query-document pairs

Applications

Click model's output	Application
Understanding of user behavior	User interaction analysis
Conditional click probabilities	User simulation
Full click probabilities	Model-based metrics
Parameter values	Ranking

Outline

1. Applications

2. Advanced models

3. Summary

User interaction analysis

- Random click model (global CTR): $\rho = 0.122$
- Rank-based CTR:
  $\rho_1 = 0.429, \rho_2 = 0.190, \rho_3 = 0.136, \ldots, \rho_{10} = 0.048$
- Position-based model:
  $\gamma_1 = 0.998, \gamma_2 = 0.939, \gamma_3 = 0.759, \ldots, \gamma_{10} = 0.260$
- Dynamic Bayesian network model: $\gamma = 0.9997$

Click models are trained on the first 10K sessions of the WSCD 2012 dataset.

Simulating users

Algorithm Simulating user clicks

Input: click model M, query session s

Output: vector of simulated clicks $(c_1, \ldots, c_n)$

1: **for** $r \leftarrow 1$ to $|s|$ **do**

2:     $P_r \leftarrow \underbrace{P_M(C_r = 1 \mid C_1 = c_1, \ldots, C_{r-1} = c_{r-1})}_{\text{conditional click probability}}$

3:     Generate c_r from $Bernoulli(P_r)$

4: **end for**

Model-based metrics

- Utility-based metrics

$$uMetric = \sum_{r=1}^{n} P(C_r = 1) \cdot U_r$$

- Effort-based metrics

$$eMetric = \sum_{r=1}^{n} P(S_r = 1) \cdot F_r$$

Expected reciprocal rank

$$ERR = \sum_r \frac{1}{r} \cdot P(S_r = 1)$$

$$= \sum_r \frac{1}{r} \cdot R_{u_r q} \cdot \prod_{i=1}^{r-1} \left(\gamma \cdot (1 - R_{u_i q}) \right)$$

Features for ranking

Yandex   san francisco — 62 million answers   Search

Web

Images

Video

Translate

More

San Francisco Travel
sanfrancisco.travel ▾
San Francisco is home to a little bit of everything. Whether you're a first time visitor or a long-time local. This is the place to find out about all things **San Francisco**.

$\alpha_{u_1 q}$

San Francisco - Wikipedia, the free encyclopedia
en.wikipedia.org › **San Francisco** ▾
San Francisco (/sæn frənˈsɪskoʊ/), officially the City and County of **San Francisco**, is the commercial, cultural, and financial center of Northern California and the only consolidated city-county in California.

$\alpha_{u_2 q}$

San Francisco travel guide - Wikitravel
wikitravel.org › en/**San_Francisco** ▾
San Francisco is a major city in California, the centerpiece of the Bay Area, well-known for its liberal community, hilly terrain, Victorian architecture, scenic beauty, summer fog, and great ethnic and cultural diversity.

$\alpha_{u_3 q}$

San Francisco City Guide | Hotels, Restaurants, Nightlife, Real...
sanfrancisco.com ▾
The job market may seem hard to navigate these days, but employment and career opportunities can be found in **San Francisco**'s Financial District and Silicon Valley's...

$\alpha_{u_4 q}$

City and County of **San Francisco**
sfgov.org ▾
SFGov Visitors Key Services: **SF** Travel Resources. ... Table of links to **San** Franisco districts and supervisors. District. Supervisor.

$\alpha_{u_5 q}$

Outline

1. Applications

2. Advanced models

3. Summary

Aggregated SERPs

Search tasks

Using features

ID	Feature description
1	Term frequency (TF) of body
2	TF of anchor
3	TF of title
4	TF of URL
5	TF of whole document
6	Inverse document frequency (IDF) of body
7	IDF of anchor
8	IDF of title
9	IDF of URL
10	IDF of whole document
11	TF*IDF of body
12	TF*IDF of anchor
13	TF*IDF of title
14	TF*IDF of URL
15	TF*IDF of whole document
16	Document length (DL) of body
17	DL of anchor
18	DL of title
19	DL of URL
20	DL of whole document
21	BM25 of body
22	BM25 of anchor
23	BM25 of title
24	BM25 of URL
25	BM25 of whole document

Beyond clicks
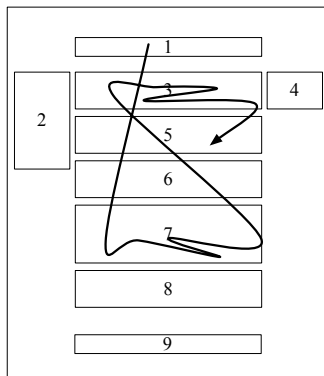
Picture taken from F. Diaz, R.W. White, G. Buscher, and D. Liebling. Robust models of mouse movement on dynamic web search results pages. In *CIKM*, 2013. ACM Press

Outline

Outline

1. Applications

2. Advanced models

3. Summary

Materials

- Aleksandr Chuklin, Ilya Markov, Maarten and de Rijke
  Click Models for Web Search
  Morgan & Claypool, 2015

Advanced topics in IR

Offline

| Data Acquisition | ⟹ | Data Processing | ⟹ | Data Storage |

Online

| Query Processing | ⟹ | Ranking | | Evaluation |

Advanced

| Advanced topics in IR and Web Search |