Information Retrieval

Introduction

Andrew Yates
a.c.yates@uva.nl

University of Amsterdam

Outline

Outline

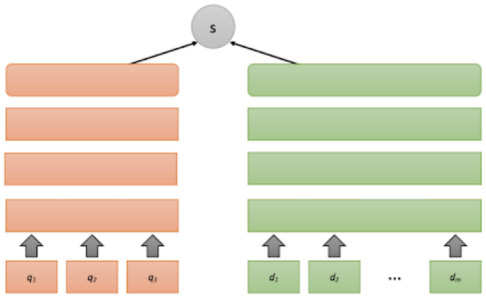
Course coordinator

Andrew Yates

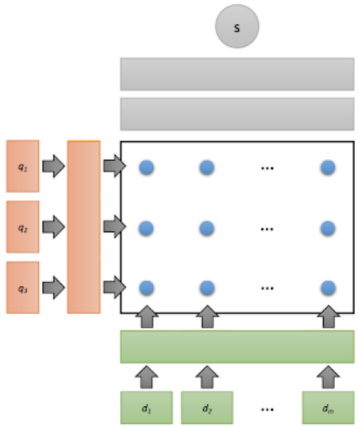
- 2021–now Assistant professor at Information Retrieval Lab (`https://irlab.science.uva.nl`), UvA
- 2018–2021 Senior researcher at Max Planck Institute for Informatics
- 2016–2018 Postdoctoral researcher at Max Planck Institute for Informatics
- 2011–2016 PhD at Georgetown University, DC, USA

Research interests: semantic matching

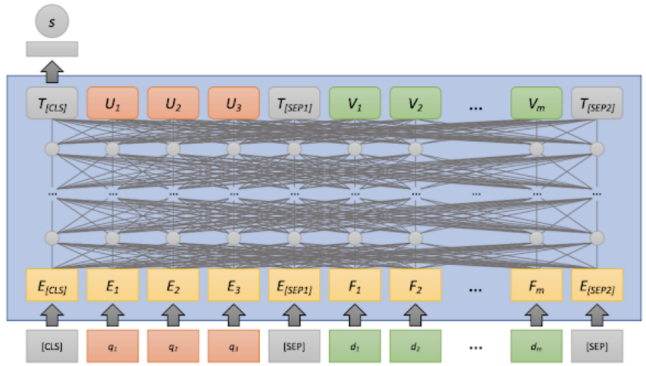
How can we leverage representation learning to improve ranking while considering only textual content?



(a) Representation-Based



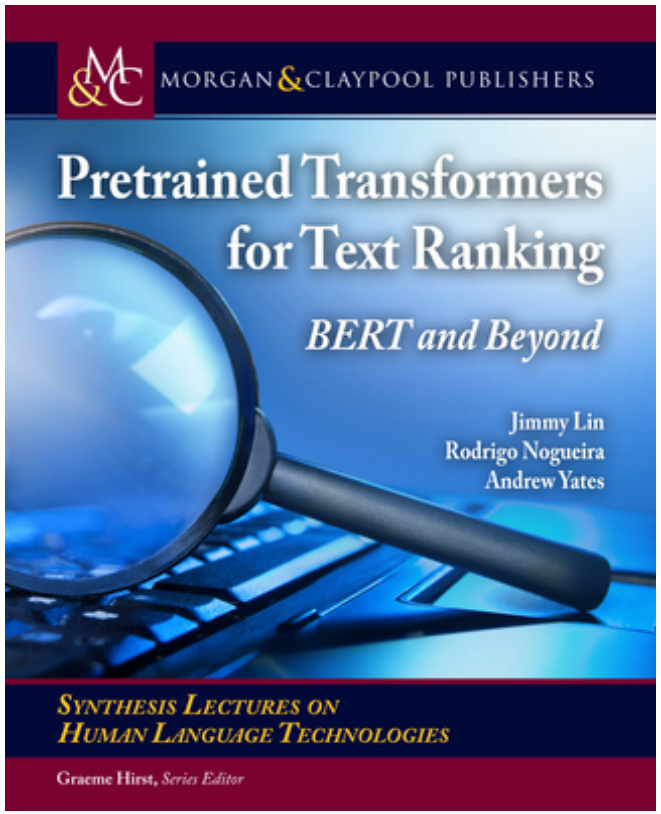
(b) Interaction-Based



(c) monoBERT

Teaching

- MSc courses on IR (previously: core course & advanced lecture)
- Book and tutorials: *Pretrained Transformers for Text Ranking*
  - with Jimmy Lin & Rodrigo Nogueira (U Waterloo)
  - `arxiv.org/abs/2010.06467`



Course Lecturers

- Andrew Yates (yours truly)
- Ilya Markov
  - Previous course coordinator
  - Learning from user interactions
- Evangelos Kanoulas
  - Full professor at UvA, head of IRLab
  - Evaluation
- Harrie Oosterhuis
  - Assistant professor at Radboud University
  - Offline/online/counterfactual learning to rank
- Mohammad Aliannejadi
  - Assistant professor at UvA
  - Conversational search

Teaching Assistants

- Senior TA: Christos Athanasiadis, c.athanasiadis@uva.nl
- Meeting you (laptopcolleges)
  - Weija Zhang
  - Shuai Wang
  - Ruud van Bakel
  - Arezoo Sarvi
  - Blazej Dolicki
  - Floor van Lieshout
  - Roel Klein
  - Clemencia Siro
- Supporting roles (preparing assignments, etc)
  - Pooya Khandel
  - Ali Vardasbi
  - Gabriel Benedict
  - Xinyi Chen
  - Ming Li

Outline

Resources

- Canvas: `https://canvas.uva.nl/courses/28683`
- Datanose: `https://datanose.nl/#course[99055]`
- Piazza: `https://piazza.com/class/kyiksrdfk0b6te`

Classes and schedule

1. Lectures
   - Pre-recorded, posted on Canvas by each Tuesday
2. Lectures' Q&A and discussion
   - Thursdays, 11:00-13:00: Q&A followed by discussion questions
   - We will assign one third of students to an on-site group each week; the remaining students can only join online
   - `https://uva-live.zoom.us/j/85646263849`
     NOTE: need UvA NetID
   - Not recorded so you can ask questions freely
   - Not mandatory, but this is more than Q&A!
3. Laptopcollege, Assignments Q&A
   - Primarily Wed & Thurs; timeslots vary with group
   - Book slot with your group's TA
   - Not mandatory

How to ask questions

1. Look for an answer on Canvas in slides and videos
2. Look for similar questions in Piazza
3. Ask during a Q&A session
4. Post a new question in Piazza

Assignments

Each assignment is 25% of your final grade.

1. Assignment 1
   - Unsupervised ranking methods
   - Implementation only
   - Released Feb 14, due Feb 28
2. Assignment 2
   - Learning to rank
   - Implementation and report
   - Released March 1, due March 19

Please **form your team** by the end of this week!

Assignments

- One submission per team
- Three students per team (who are in the same LC group)
- Q&A during laptopcolleges (LCs)
- Organize LC timeslots directly with your group TA (at least one full weekday before the timeslot)
- Communicate LC scheduling issues with your TA (e.g., moving online due to COVID)
- Late submissions will not be accepted

Grading

- 50%: assignments (groups of three)
- 50%: final exam (individual; no groups)

Minimum of 5.5 for each

Outline

Course as collaboration

A university course is a collaboration between teachers and students

- Common goal
- Two parties

Response to students' feedback from 2021

1. Continue to improve assignments
   - Clarify and streamline implementation aspect, including autograding
   - Second assignment includes a report
   - Increase TA involvement in assignment checking/polishing
   - Avoid "correct-but-unexpected results"
2. Overwhelmingly positive feedback on flipped classroom
   - Keep dual session types: one day for videos & one for active discussion
   - Stress that Thursday sessions are much more than Q&A
3. Extensive references
   - Videos, slides, and textbook/paper references (as before)
   - Provide pointers by topic (new)

Giving feedback

1. Feel free to send us feedback during the course (to Andrew and Christos)
2. Do not wait: speak up if something does not work for you
3. Be constructive: actionable items instead of aggression
4. Be ready for one of the following responses:
   - Your feedback is implemented this course
   - Your feedback will be implemented in the next edition
   - An explanation of why your feedback cannot be implemented

Program committee (OC)

- If you encounter any problem during program-related events or a course and you want to file a complaint or submit separate feedback, please do not hesitate to contact the program committee at ocai-science@uva.nl
- For more information, please see this page: `https://student.uva.nl/ai/content/az/programme-committee/programme-committee.html`

Our expectations

Preparing a course is hard

We appreciate it if you appreciate this work

- Be a good citizen
- Send us likes

Our expectations

- Honesty, no plagiarism
- Study is prioritized over job
- Follow course guidelines and instructions
- Value our time: check materials before asking questions
- Timely communication
- Constructive feedback

Outline

Ethical issues and stress

Potential sources:

- Team work (assignments)
- Thesis
- TAing

If you experience ethical issues or stress:

- Immediate communciation
- Contact course coordinator
- Contact FNWI study adviser `https://student.uva.nl/is/shared/studentensites/fnwi/esc-gedeelde-content/en/az/study-adviser-esc/study-adviser-esc.html`

Responsible communication

- Always two parties, always two-way communication
- Ask for communication, be open to communication
- Talk about yourself only, do not give judgements
- Be explicit, ask for help

Outline

Outline

What is IR?

Information Retrieval is about technology
to connect people to information

Why study IR?

Nowadays, IR problems are everywhere

- Text processing and analysis
- Various forms of ranking
- Ranked offline/online evaluation
- Learning from user interactions
- Etc.

What is so special about IR?

1. Relevance
   - "No one ever saw me but everyone knows I exist"
   - No precise definition
   - Highly subjective
   - Different in different scenarios
2. Ranking
   - Depends on relevance
   - Dependencies between ranked items

IR and AI

- IR uses AI
- IR learns from users (and, thus, contributes to AI)
- IR + NLP = set of techniques to work with text

Outline

What should we teach in IR?

OPINION PAPER

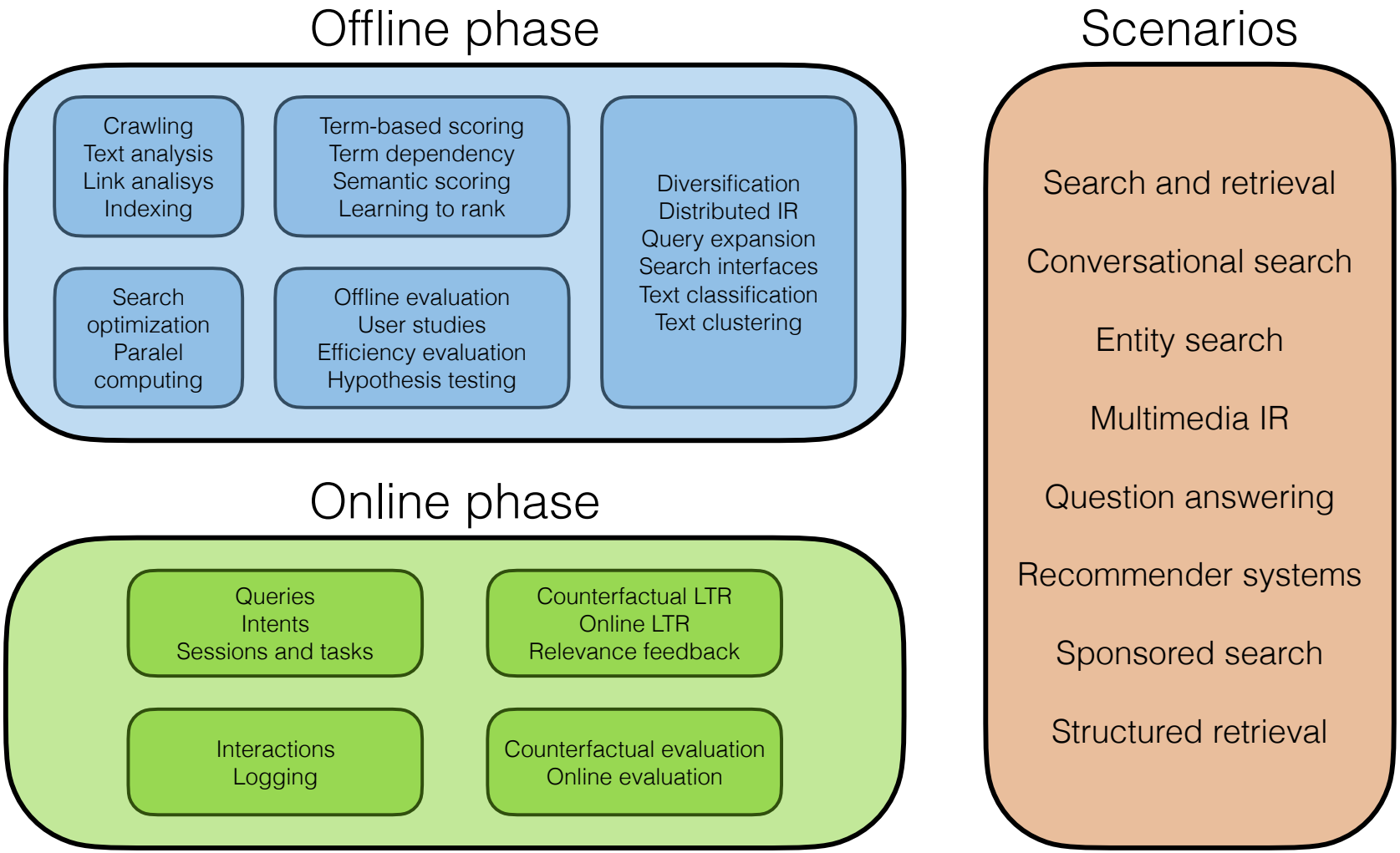
What Should We Teach in Information Retrieval?

Ilya Markov
University of Amsterdam
i.markov@uva.nl

Maarten de Rijke
University of Amsterdam
derijke@uva.nl

`http://sigir.org/wp-content/uploads/2019/01/p019.pdf`

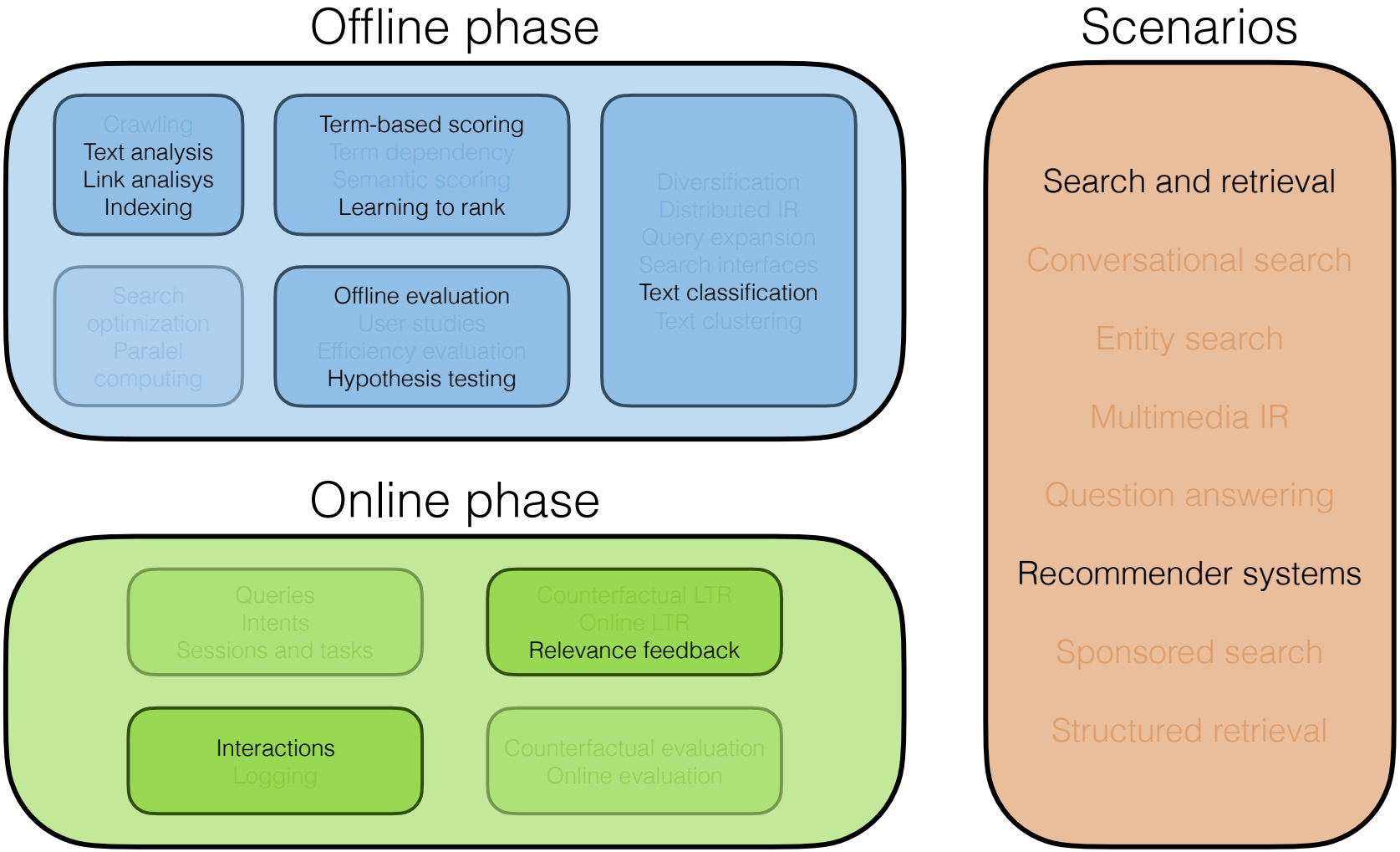
IR topics


Offline phase
Crawling
Text analysis
Link analisys
Indexing
Term-based scoring
Term dependency
Semantic scoring
Learning to rank
Diversification
Distributed IR
Query expansion
Search interfaces
Text classification
Text clustering
Search optimization
Paralel computing
Offline evaluation
User studies
Efficiency evaluation
Hypothesis testing
Online phase
Queries
Intents
Sessions and tasks
Counterfactual LTR
Online LTR
Relevance feedback
Interactions
Logging
Counterfactual evaluation
Online evaluation
Offline phase
Crawling
Term-based scoring
Scenarios
Search and retrieval
Conversational search
Entity search
Multimedia IR
Question answering
Recommender systems
Sponsored search
Structured retrieval

IR courses at UvA

$$\text{IR0} \longrightarrow \text{IR1} \longrightarrow \text{IR2}$$

IR0 ("Zoekmachines"), BSc



IR1, MSc 1st year

Offline phase

Crawling
Text analysis
Link analisys
Indexing

Term-based scoring
Term dependency
Semantic scoring
Learning to rank

Diversification
Distributed IR
Query expansion
Search interfaces
Text classification
Text clustering

Search optimization
Paralel computing

Offline evaluation
User studies
Efficiency evaluation
Hypothesis testing

Online phase

Queries
Intents
Sessions and tasks

Counterfactual LTR
Online LTR
Relevance feedback

Interactions
Logging

Counterfactual evaluation
Online evaluation

Scenarios

Search and retrieval

Conversational search

Entity search

Multimedia IR

Question answering

Recommender systems

Sponsored search

Structured retrieval

IR2, MSc 2nd year

Offline phase

- Crawling
- Text analysis
- Link analisys
- Indexing
- Term-based scoring
- Term dependency
- Semantic scoring
- Learning to rank
- Search optimization
- Paralel computing
- Offline evaluation
- User studies
- Efficiency evaluation
- Hypothesis testing
- Diversification
- Distributed IR
- Query expansion
- Search interfaces
- Text classification
- Text clustering

Online phase

- Queries
- Intents
- Sessions and tasks
- Counterfactual LTR
- Online LTR
- Relevance feedback
- Interactions
- Logging
- Counterfactual evaluation
- Online evaluation

Scenarios

- Search and retrieval
- Conversational search
- Entity search
- Multimedia IR
- Question answering
- Recommender systems
- Sponsored search
- Structured retrieval

Outline

Objectives

After completing this course you will be able to:

1. Identify the type of problem (ranking, evaluation, interaction, etc.)
2. Identify the range of solutions (from basic to state-of-the-art)
3. Adapt and apply the solutions to the problem
4. Assess the result and adjust the solution

Topics

We will cover one of the following topics each week:

1. Course organization, course content, and IR0 recap
2. Evaluation
3. Document representation and matching
4. Learning to rank (LTR) & interactions
5. Online and counterfactual LTR
6. Conversational search & recommender systems
7. Recap & preparation for final exam

Let's get started

Enjoy the course!

Basic references

- B. Croft, D. Metzler, T. Strohman
  Search Engines, Information Retrieval in Practice
  https://ciir.cs.umass.edu/irbook
- C. Manning, P. Raghavan, H. Schütze
  Introduction to Information Retrieval
  http://nlp.stanford.edu/IR-book
- I. Markov, M. de Rijke
  What should we teach in Information Retrieval?
  http://sigir.org/wp-content/uploads/2019/01/p019.pdf