Program SIKS course "Advances in Information Retrieval"
(*Each lecture will include 2 short breaks)*

Thursday June 18

09.30 - 09.45 Registration, coffee/tea
09.45 - 10.00 Welcome, introduction

10.00 - 13.00 **Prof.dr.Theo van der Weide: Foundations of IR.**
In this lecture we discuss the basic approaches to Information Retrieval. We discuss the most important models: Boolean Model, Vector Model and Probabilistic Model. We also pay attention to computational aspects of these models. We shortly indicate the implications of big data, and the risks of propagation of computational errors that result from the real arithmetic on digital computers.

13.00 – 14.00 Lunch

14.00 – 17.00 **Dr.Evangelos Kanoulas: Search Engine Evaluation: how can you know you are better than Google?**
Assume that you have come up with an amazing algorithm that can search the Web better than Google. How can you know whether this is indeed a great idea? In this lecture I will discuss two predominant paradigms of search engine evaluation that can allow the data to tell you whether your idea can be the next Google: (a) collection-based offline evaluation, and (b) in-situ online evaluation. In the first part of the lecture I will present how one can build and use a benchmark collection to test the quality of a search algorithm, while in the second half I will talk about A/B testing and interleaving, two powerful techniques that allow users to indicate whether your search algorithm is better than any other (including Google's). By the end of the lecture attendees should be able to understand the pros and cons of the different evaluation methods and apply the discussed techniques to evaluate new search technologies they build in their own laboratory.

Friday June 19

09.30 – 12.30 **Dr.Suzan Verberne: User modelling for Information Retrieval evaluation**.
Most methods for the evaluation of IR systems implicitly assume a 'perfect' user. A perfect user (from the system's viewpoint) would for example examine all results retrieved by the system, and only click on truly relevant documents. If we want to take into account realistic user behaviour, we have three options: (1) bringing in real users, ask what they think and observe their behaviour; (2) using pre-collected log data (queries, clicks); (3) simulate user interactions with user models. Each of these options have their own advantages and disadvantages. In my lecture, I will focus on the exploitation of user simulations for the evaluation of interactive search. I will present several types of user models and explain how we can use them for the simulation of user interaction, and what the use of simulation means for the evaluation of IR methods.

12.30 – 13.30 Lunch

13.30 – 16.30 **Dr.ir.Djoerd Hiemstra: How to build Google in an Afternoon.**
How many machines do we need to search and manage an index of billions of documents? In this lecture, I will discuss basic techniques for indexing very large document collections. I will discuss inverted files, index compression, and top-k query optimization techniques, showing that a single desktop PC suffices for searching billions of documents. An important part of the lecture will be spend on estimating index sizes and processing times. At the end of the afternoon, students will have a better understanding of the scale of the web and its consequences for building large-scale web search engines, and students will be able to implement a cheap but powerful new 'Google'.