

# User Models to Compare and Evaluate Web IR Metrics

Georges Dupret  
Yahoo! Labs, Silicon Valley

In order to compare or evaluate the performance of two metrics objectively, we need to define a metric on the metrics. Because this new metric also needs to be evaluated, the problem seems to have no solution. In this work, we propose an alternative route: We argue that all Web IR metrics make assumptions on the user behavior, often implicitly. A Metric can then be judged on how realistic its associated assumptions are. The associated user model can also be evaluated against observations collected in clickthrough logs by search engines. If a model predicts better the user behavior on unseen data, then it is arguably more realistic, and the associated metric is superior. In this work we review some common metrics and propose a user model for each of them. We discuss the different assumptions to highlight their strength and weakness. In particular we illustrate these ideas by a discussion on Discounted Cumulated Gain (DCG) and its user model, we show how the discounting factor can be evaluated and suggest ways to improve it.