# A Practical Guide to Computing Evaluation Measures and Comparing Systems: Twelve Small Tips

Tetsuya Sakai
Waseda University

## ABSTRACT

The NTCIRPOOL toolkit creates pool files. The NTCIREVAL toolkit computes evaluation measures based on relevance, diversity, group-fairness, and so on. The Discpower toolkit computes p-values. While demonstrating how I actualy use these toolkits, I will provide some small but practically important tips for IR experimenters. The recommendations are based on a few decades of my experience as an "NTCIR researcher." This tutorial complements my NTCIR-16 tutorial (see http://sakailab.com/publications/ for slides and videos).

## BIOGRAPHY

Tetsuya Sakai is a professor at the Department of Computer Science and Engineering, Waseda University, Japan. He is also a General Research Advisor of Naver Corporation, Korea (2021-), and a visiting professor at the National Institute of Informatics, Japan (2015-). He spent 20 years in industry, including 1.5 years at the University of Cambridge as a visiting researcher from Toshiba, and 4.5 years at Microsoft Research Asia. He has subsequently spent 10 years in academia so far. He is an ACM distinguished member and an IPSJ Fellow. In 2023, he was inducted into the SIGIR Academy.