

On Laboratory Testing of Text Retrieval Systems

Stephen Robertson
Microsoft Research Cambridge

Abstract

The 45-year history of information retrieval evaluation includes a range of experiments which might be categorised in the light of the tradition division between in vitro and in vivo experiments in biology. In this talk, I will explore some of the characteristics of laboratory or in vitro experiments, and contrast them with operational system or in vivo experiments. The present state of the field shows a clear predominance of the laboratory approach, characterised by TREC, NTCIR and various other similar endeavours. However, there are research questions that can only be answered in an operational environment; the two approaches are complementary. It is in any case not a simple division: it is more like a spectrum, one end represented by complete experimental control and the other by complete realism. Actually the extremes are neither possible nor interesting - all real experiments involve some degree of compromise between the two. In particular, the trend in TREC towards a wider variety of tasks reflects in part various attempts to introduce at least some of the conditions of real-world systems. There is always conflict between the requirements of laboratory control and those of realism; compromise is both difficult and necessary.