

Statistical bases of relevance assessment for the 'ideal' information retrieval test collection

Summary

This Report concerns the statistical basis of relevance assessment for information retrieval experiment, with special reference to the proposed 'ideal' information retrieval test collection. That is, it considers statistical arguments, or methods, for establishing what assessment information is required in given circumstances, from which actual procedures for obtaining this information may be derived. The Report is chiefly devoted to the work done by H. Gilbert on a three month project supported by BLR&DD Grant SI/G/267; but it refers to earlier and related studies, and attempts to provide a self-contained and integrated discussion of the whole question of statistically-adequate assessment for retrieval experiment evaluation, especially where exhaustive relevance assessment is impossible.

The Report is divided into three sections:

- A. This section includes an introduction on the idea and specification of the 'ideal' collection, some illustrative data on the relevance properties of various actual test collections, and a discussion of some of the constraints on any proposed assessment methods and procedures.
- B. This section presents the technical statistical study carried out on the project. Two methods of determining assessment requirements, the 'Pool' method and the 'Squares' method, appear most satisfactory, and these are developed in detail.
- C. This section considers the implications of the methods discussed in B in terms of procedures for obtaining assessments when building the 'ideal' collection, and in relation to the use of the collection for future experiment; it also examines the implications of the study for information retrieval experiment in general.