SALTON, G. 24,370
A new comparison between conventional indexing (MEDLARS) and automatic text processing (SMART).
J. Amer. Soc. Info. Sci. 23, 2 (March-April 1972), 75–84.

The author reports on a comparison of the experimental SMART information retrieval system with the MEDLARS retrieval system. The paper appears to have been written to alleviate possible difficulties in interpreting a previous MEDLARS-SMART comparison [see *CR* **10,** 6 (June 1969), Rev. 16,849].

The experiment involved selection of 30 search requests previously used by F. W. Lancaster in the evaluation of MEDLARS, and the location of one or more documents per query previously determined by Lancaster to be relevant to each of the 30 queries. Using in turn each of the 30 sets of documents, Science Citation Index was consulted to find 15 more documents per query which cited the original set of relevant documents, and which were in the MEDLARS data base. The collection of cited documents (450) was then used as a basis for the comparative evaluation of SMART and MEDLARS with respect to search strategy. One medical student was hired to make judgments on the relevancy of the documents to the queries.

In the first experiment, the MEDLARS controlled vocabulary manual indexing method was compared to a SMART automatic indexing algorithm which used stemming. When the systems were evaluated, SMART had a recall value of 0.1814 compared to MEDLARS' of 0.3117 representing a 42% difference. SMART's precision was 32% less than MEDLARS.

To determine if it was possible to improve the results further, a comparison was made between MEDLARS and the SMART system employing automatic indexing with the aid of a thesaurus. The experiment showed SMART to be 4% better than MEDLARS in recall and showed a zero difference between the two for precision.

The author draws a number of conclusions from the experiments, two of which are presented here: 1) "The Boolean [MEDLARS] search techniques . . . are clearly inferior to vector matching techniques [SMART] producing ranked output in decreasing query-document similarity order." 2) "No technical justification appears to exist for maintaining controlled manual indexing in operational retrieval environments." It would seem necessary to replicate the experiment before such conclusions are offered, especially considering the small data base being used.

One difficulty with the evaluation procedure concerns the way in which relevance judgments are made. Aside from the fact that only one person made the judgments, there is a problem in interpreting the numerical results. The judge was given each of the 30 queries in turn, and each of the associated sets of 15 documents found by Science Citation Index to be potentially relevant. The task of the judge was to decide which of the 15 documents in a particular set were relevant to a query. The judge was *not* asked to determine which of the 450 documents were relevant to the query. However, in the course of the evaluation, both of the retrieval systems searched the entire set of 450 documents to find documents relevant to queries. Thus, in calculating the performance measures, one can only say that the performance values represent a lower bound on the true values since, potentially, there are 435 other documents that might be relevant to a particular query.

Finally, it should be noted that the part of the experiment producing the best results for SMART compared MEDLARS' controlled vocabulary indexing with SMART's automatic indexing algorithm using a manually developed thesaurus. The thesaurus *is* manually developed as C. McAllister [see *CR* **12,** 5 (May 1971), Rev. 21,206] correctly points out. Thus it appears that the title of the paper is somewhat misleading in claiming that the experiments involve comparison of a manual (MEDLARS) and an automated (SMART) system.

M. D. Cooper, Berkeley, Calif.