Relevance Feedback
In An Automatic Document Retrieval System

Today's proliferation of technical and scientific books and articles is too much of a good thing. The so-called "information explosion" has created a need for new methods of document classification and retrieval with the following characteristics:

1) The number of expert indexers and librarians obtainable is completely inadequate for processing such a volume of information, even in a cursory manner. An <u>automatic</u> computerized system is needed.[1]

2) Even with high speed computers, the retrieval operation must be <u>simple</u> in order to minimize time and cost.

3) The user should not be required to understand the detailed operation of the system. For this reason, the system should ideally respond to information requests in <u>natural language</u>.

This study deals with an experimental automatic document retrieval system that meets these requirements. Within this experimental system, several retrieval algorithms that permit user interaction with the search process are evaluated. All of these techniques employ user judgment of the relevance of certain selected documents to his request, and are called "relevance feedback" algorithms.

Section I of this report describes several methods of automatic document retrieval, and details the experimental retrieval system used in this study (the SMART system [3]).

Section II examines several means of user interaction with
an automatic retrieval system, and summarizes the results
of some prior experiments with user interaction in the SMART
system. In Section III the results of earlier experiments
with several relevance feedback algorithms are presented.
Section IV details the features of the experimental environment
of this study. In Section V, means of evaluating the
performance of an information retrieval system are discussed,
and the evaluation measures and statistical tests used in
this study are described. Section VI contains the results
of relevance feedback experiments in five areas. In Section
VI-A relevance feedback results in two document collections
are compared. Section VI-B compares feedback algorithms that
use only information from the documents judged 'relevant'
by the user, and Section VI-C examinesthe effect on these
algorithms of the number of documents used for feedback.
Section VI-D investigates strategies that use information
from relevant and non-relevant documents, and Section VI-E
further studies the comparative usefulness of these strategies
for different types of queries. In Section VII, the last section
of this study, the results of these relevance feedback
experiments are used to support recommendations for interactive
document retrieval systems, and to suggest guidelines for
future experiments with regard to relevance feedback algorithms,
evaluation of feedback performance, partial search strategies,
multiple query strategies, request clustering, and permanent
document vector modification.