

**Relevance Feedback
In An Automatic Document Retrieval System**

**A Thesis
Presented to the Faculty of the Graduate School
of Cornell University for the Degree of
Master of Science in Computer Science**

by

**Eleanor Rose Cook Ide
January, 1969**

Biographical Sketch

Eleanor Ide, born in Flint, Michigan on September 7, 1940, earned a Bachelor of Arts Degree with Distinction and High Honors in Sociology from the University of Michigan in 1962. She was a research assistant for one year at the Population Studies Center, University of Michigan, and a computer programmer for three years studying adaptive logic and pattern recognition at IBM Systems Development Division, Endicott, New York. There she co-authored with Doctor Cyril J. Tunis "An Unsupervised Adaptive Algorithm," in IEEE Transactions on Electronic Computers, Volume EC - 16 #12, December, 1967, and co-invented with Dr. Tunis and Mitchell P. Marcus the 'Adaptive Logic System for Unsupervised Learning' described in Patent Application IBM Docket EN966005, now being filed.

Eleanor Ide, a member of Phi Beta Kappa, Sigma Delta Epsilon, and the Association for Computing Machinery, is presently employed as a Senior Systems Analyst by Hydra Computer Corporation in Raleigh, North Carolina.

Dedication

This thesis is gratefully dedicated
to my husband, Robert Ide, whose help
and patience made my graduate work
possible.

Acknowledgments

The author wishes to acknowledge the programming support of Robert Williamson, and the advice and encouragement of Professor Gerard Salton and Michael Keen. The free computer time supplied by Hydra Computer Corporation is also appreciated.

Abstract

Many automatic document retrieval systems represent documents and requests for documents as numeric vectors that indicate the subjects treated by the document or query. This study investigates relevance feedback, a process that allows user interaction with such a retrieval system. The user is presented with a small set of possibly relevant documents, and is asked to judge each as relevant or non-relevant to his request. The numeric vectors representing the judged documents are used to modify the numeric vector representing the query, and the new query vector is used to retrieve a more appropriate set of documents. The relevance feedback process can be iterated as often as desired. Several feedback algorithms are investigated in a collection of 200 documents and 42 queries. Two distinct viewpoints are taken in evaluation; one measures the movement of the query vector toward the optimum query defined by Rocchio, the other measures the retrieval experienced by the user during the feedback process. Several performance measures are reported from each evaluation viewpoint. Both evaluation methods indicate that relevance feedback is an effective process.

All algorithms tested that use only relevant document vectors for feedback provide equally good retrieval. Such algorithms should supply additional documents to any user

who judges every document presented for feedback to be non-relevant. Algorithms using non-relevant document vectors for feedback improve the retrieval obtained by these users without requiring additional relevance judgments.

The relevance feedback algorithms tested are based on the assumption that the vectors representing the documents relevant to a query are clustered in the same area of the document space. The conclusion that no tested relevance feedback algorithm is completely appropriate for the experimental environment is supported by a hypothesis that explains the observed contrasts between the behavior of strategies using only relevant documents for feedback and that of strategies using non-relevant documents. This hypothesis states that for most queries, some relevant document vectors are separated from others by one or more non-relevant document vectors. The implications of this result for future research in relevance feedback, partial search or multi-level strategies, multiple query strategies, request clustering, and document vector modification are discussed, and useful evaluation measures and new algorithms for these areas are suggested.

Table of Contents

	Page
Biographical Sketch	ii
Dedication	iii
Acknowledgments	iv
Abstract	v
Table of Contents	vii
List of Illustrations	ix
Introduction	1
I. Automatic Document Retrieval Systems	3
II. User Interaction With an On-Line Retrieval System	8
III. Prior Investigations of the Relevance Feedback Retrieval Algorithm	14
IV. Environment of the Reported Experiments	19
V. Evaluation of Retrieval Performance	23
A. Measures of Performance	23
B. Statistical Significance Tests	28
C. The Feedback Effect in Evaluation	29
VI. Experimental Results	33
A. Comparison of the Cranfield and ADI Collections	33
B. Strategies Using Relevant Documents Only	36
C. Amount of Feedback Output	42
D. Strategies Using Non-Relevant Documents	54
E. Characteristics of Query Subgroups	75

	Page
VII. Recommendations Based on Present and Prior Experiments	101
A. Relevance Feedback Recommendations for Concept Vector Document Classification Systems	101
B. Evaluation of Relevance Feedback Experiments	106
C. Feedback of Non-Relevant Documents	130b
D. Partial Search and Multiple Query Algorithms	147
Summary	199

List of Illustrations

	Page
Figure 11: An Illustration of the Interpolation Method Used for the Quasi-Cleverdon Recall-Precision Averages.	27
Figure 2: An Illustration of the Interpolation Method Used for the Neo-Cleverdon Recall-Precision Averages.	27
Figure 3: Improvement over Initial Search, ADI and Cranfield Collections	35
Figure 4: Effects of 'Relevant Only' on the Multipliers of Documents Retrieved on Three Successive Iterations.	38
Figure 5: Query Update Parameters for Relevant Only Strategies, Using Only Relevant Documents.	40
Figure 6: Varying the Number of Feedback Documents, Total Performance Recall-Precision Curves	43
Figure 7: Recall and Precision Document Curves, Feedback Increment Strategy.	45
Figure 8: Feedback Effect Document Curves, Recall and Precision Differences, Varying the Number of Documents Retrieved (N=5 normal, N=2 and N=10 compared).	46
Figure 9: Variable Feedback; Feeding Back Only the First or First Two New Relevant Documents Found Within the First Fifteen Documents Retrieved; Total Performance, Q_0 Strategy.	50
Figure 10: Variable Feedback and Combination Strategy Compared to N=5 Norm, Recall and Precision Differences, Feedback Effect Document Curves.	52
Figure 11: Decrementing the Highest Non-Relevant Document, Total Performance Recall-Precision Curves.	56
Figure 12: Normalized Results for Non-Relevant Strategies, Total Performance.	56
Figure 13: Decrementing the Highest Non-Relevant Document on Eleven Bad Queries, Total Performance Recall-Precision Curves.	59

	Page
Figure 14: (Dec H1) - (Q_0) For the 3=Good Queries, a Comparison Between Decrementing the Highest Non-Relevant and Incrementing the Relevant Only, Total Performance Difference.	60
Figure 15: (Dec H1) - (Dec 2 H1) A Comparison of Decrementing the Highest or the Two Highest Non-Relevant Documents on Each Iteration, Total Performance Difference.	62
Figure 16: (Feedback 10) - (Dec H1) A Comparison Between Feeding Back 10 Documents but Increasing Relevant Only, and Feeding Back 5 Documents but Decrementing the Highest Non-Relevant, Total Performance Difference.	63
Figure 17: Statistical Comparison of Feedback Effect Relevant and Non-Relevant Document Strategies, Normalized Recall and Precision.	66
Figure 18: Comparison of First and Second Iterations to Initial Search Differences and Significance Levels - Rocchio Algorithm, Feedback Effect Recall-Precision Curves.	68
Figure 19: Comparison of First and Second Iterations to Initial Search Differences and Significance Levels - Q_0+ Algorithm, Feedback Effect Recall-Precision Curves.	69
Figure 20: Rocchio Strategy on Eleven Bad Queries, Feedback Effect Recall-Precision Curves.	72
Figure 21: Recall and Precision Differences Comparing Rocchio and Dec H1 Strategies to Q_0+ norm, Feedback Effect Document Curves.	73
Figure 22: Some Query Subgroups Investigated.	76
Figure 23: Characteristics of Subgroups Selected By Initial Search Retrieval.	80
Figure 24: Subgroups Selected by Performance, Feedback Effect Recall-Precision Curves, Rocchio Strategy	83
Figure 25: Characteristics of Subgroups Selected By Performance.	84

	Page
Figure 26: Cross-probabilities for Correlation of Modified Query with Original Query.	85
Figure 27: Characteristics of Subgroups Selected By Correlation of Modified Query With Original Query.	87
Figure 28: Subgroups Selected By Strategy, $Q_0 +$ Group, Feedback Effect Recall-Precision Curves, Comparing Positive and Negative Feedback.	89
Figure 29: Subgroups Selected By Strategy, Rocchio Group, Feedback Effect Recall-Precision Curves, Comparing Positive and Negative Feedback.	90
Figure 30: Comparison of Positive and Negative Feedback in Subgroups Selected By Feedback Strategy.	91
Figure 31: Comparison of Negative and Positive Feedback in Subgroups Chosen by Feedback Improvement.	94
Figure 32: Comparing Positive and Negative Feedback In Subgroups Selected By Number of Concepts in Original Query and Number of Relevant Documents.	98
Figure 33: Examples of Performance Equivalence Between Queries as Defined by Different Interpolation Methods.	112
Figure 34: Twenty-Point Interpolation From ExampleQ Query A Using Five Different Interpolation Methods	114
Figure 35: Comparison of Two Evaluation Methods, Total Performance Evaluation with Quasi-Cleverdon Interpolation and Feedback Effect Evaluation with Neo-Cleverdon Interpolation, Comparable Strategies, $N = 5$	121
Figure 36: An Example of an Individual Query With Separate Clusters of Relevant Documents, $Q_0 +$ and Rocchio Strategies	144

	Page
Figure 37: Example of Poor Rocchio Performance Caused by Non-Relevant Document Feedback	145
Figure 38: An Example of Good Rocchio Performance on Separate Clusters of Relevant Documents	145a
Figure 39: A General Two-level Cluster Feedback Algorithm	169
Figure 40: Three Relevant Documents Forming Three Separated Clusters	176
Figure 41: A Multiple Query Algorithm	184e