

Overlap of title keywords and assigned descriptors in the OON database

Michael A. Shepherd
Computing Science Division
Dept. of Mathematics, Statistics & Computing Science
Dalhousie University
Halifax, Canada

W. J. Phillips
Department of Applied Mathematics
Technical University of Nova Scotia
Halifax, Canada

T. Lui
Computer Centre
Saint Francis Xavier University
Antigonish, Canada

This study investigated the overlap of document representations in CISTI's OON database. The document representations studied were the keywords from the titles and the assigned Library of Congress Subject Headings (LCSH). Documents were selected randomly from the database to act as queries for retrieval. Retrieval for the keyword representation of a "query" was performed only on the keyword representation of the database documents, and retrieval for the LCSH "query" representation was performed only on the LCSH document representations.

Results indicate that there is little overlap between the clusters retrieved by the two document representations. A test for independence, however, indicated that although the retrieved sets of documents appeared to be almost disjoint, the representations themselves are dependent. The results imply that in order to maximize recall in the OON database, the user must retrieve by both keywords from titles and by the assigned Library of Congress Subject Headings.