



**INTEGRATING GENERAL, NON-TOPICAL TERMS INTO ONLINE  
THESAURI: A NEW APPROACH TOWARDS IMPROVING RETRIEVAL  
RELEVANCE ON THE WEB THROUGH NATURAL LANGUAGE TERMS**

***Rahmatollah Fattahi***

*PhD., Professor, Department of Library and Information Studies, Ferdowsi  
University of Mashhad, Iran ([fattahi@ferdowsi.um.ac.ir](mailto:fattahi@ferdowsi.um.ac.ir))*

***Mehri Parirokh***

*PhD., Associate Professor, Department of Library and Information Studies,  
Ferdowsi University of Mashhad, Iran ([parirokh@ferdowsi.um.ac.ir](mailto:parirokh@ferdowsi.um.ac.ir))*

***Marie Rahimi***

*Master of Library and Information Science, Ferdowsi University of Mashhad  
([marie.rahimi@gmail.com](mailto:marie.rahimi@gmail.com))*

**ABSTRACT**

Thesauri facilitate searching through controlled vocabularies which cover domain specific (topical) terms in a hierarchical or tree structure. One major limitation of thesauri is that they do not cover general terms representing particular aspects of the documents (such as context, approach, readership level, form, etc.) being important to authors as well as to readers. Documents on the same topic and with the same index term(s) differ from one another in some respects. Descriptors assigned to similar documents do not represent the non-topical aspects of documents. General terms can help searchers who are looking for documents with particular approach to content, structure, and/or format. Such terms can be used by the searcher or the system in query expansions through a natural language approach to retrieve more relevant documents. Research was undertaken to identify and study the availability of general, non-topical terms (GNTs) in Web documents as well as in two different thesauri (Hasset and ERIC). Findings show that, while GNTs are frequently present in the titles and URLs of Web documents (before or after topical terms), there are a small number of such terms in the thesauri studied. Searching through a combination of GNTs and descriptors in a natural language string would help users retrieve more relevant documents. Suggestions are provided on how to incorporate GNTs in all levels of the hierarchy (BTs, RTs, NTs) in thesauri based on their appearance in titles and URLs of Web documents. A sample list of GNTs is developed to be used in conjunction with index terms for searching.

**KEYWORDS**

Thesauri, Online thesauri, General, non-topical terms, relevance performance, Web documents

**INTRODUCTION**

Thesauri are extensively used in indexing and retrieval of documents irrespective of their physical formats. They focus on the semantic relationship between and among index terms to provide better access to the content of documents. With the increasing access to online and on-disc databases and with the exponential availability of information on the Web the use of thesauri for searching by end-users has also increased considerably. It is very important for the end-users to find more relevant information when they use online thesauri for searching.