

# Journal of Information Science

<http://jis.sagepub.com/>

---

## Improving pseudo relevance feedback based query expansion using genetic fuzzy approach and semantic similarity notion

Pragati Bhatnagar and Narendra Pareek

*Journal of Information Science* published online 19 May 2014

DOI: 10.1177/0165551514533771

The online version of this article can be found at:

<http://jis.sagepub.com/content/early/2014/05/19/0165551514533771>

A more recent version of this article was published on - Jul 9, 2014

---

Published by:



<http://www.sagepublications.com>

On behalf of:



Chartered Institute of Library and Information Professionals

Additional services and information for *Journal of Information Science* can be found at:

**Email Alerts:** <http://jis.sagepub.com/cgi/alerts>

**Subscriptions:** <http://jis.sagepub.com/subscriptions>

**Reprints:** <http://www.sagepub.com/journalsReprints.nav>

**Permissions:** <http://www.sagepub.com/journalsPermissions.nav>

Version of Record - Jul 9, 2014

>> OnlineFirst Version of Record - May 19, 2014

[What is This?](#)

# Improving pseudo relevance feedback based query expansion using genetic fuzzy approach and semantic similarity notion

Journal of Information Science

1–15

© The Author(s) 2014

Reprints and permissions:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/0165551514533771

jis.sagepub.com

**Pragati Bhatnagar**

Department of Computer Science, M.L. Sukhadia University, India

**Narendra Pareek**

Department of Computer Science, M.L. Sukhadia University, India

## Abstract

Pseudo relevance feedback-based query expansion is a popular automatic query expansion technique. However, a survey of work done in the area shows that it has a mixed chance of success. This paper captures the limitations of pseudo relevance feedback (PRF)-based query expansion and proposes a method of enhancing its performance by hybridizing corpus-based information, with a genetic fuzzy approach and semantic similarity notion. First the paper suggests use of a genetic fuzzy approach to select an optimal combination of query terms from a pool of terms obtained using PRF-based query expansion. The query terms obtained are further ranked on the basis of semantic similarity with original query terms. The experiments were performed on CISI collection, a benchmark dataset for information retrieval. It was found that the results were better in both terms of recall and precision. The main observation is that the hybridization of various techniques of query expansion in an intelligent way allows us to incorporate the good features of all of them. As this is a preliminary attempt in this direction, there is a large scope for enhancing these techniques.

## Keywords

Genetic fuzzy algorithm; information retrieval; pseudo relevance feedback; query expansion; semantic similarity

## 1. Introduction

This section presents an overview of information retrieval system and focuses on need for query expansion. Further, it discusses the appropriateness of evolutionary approaches and the need for incorporating semantics in the field of query expansion.

### 1.1. Information retrieval: an overview

The history of information retrieval (IR) parallels the development of libraries. The first civilizations had already come to the conclusion that efficient techniques should be designed to fully benefit from large document archives. Only recently, IR has radically changed with the advent of computers and digital technologies. Digital technologies provide a unified infrastructure to store, exchange and automatically process large document collections. The search for information consequently evolved from the manual examination of brief document abstracts within predefined categories to algorithms searching through the whole content of each archived document. There are a large number of applications in which information retrieval is useful. Nowadays, automatic retrieval systems are widely used in several application domains, including digital libraries, information filtering, recommender systems, media search and search engines, and

---

### Corresponding author:

Pragati Bhatnagar, M.L. Sukhadia University, Udaipur, Rajasthan, India.

Email: pragatibhat@gmail.com