On Text Ranking for Information Retrieval based on Degree of Preference

Bo-Yeong Kang[1] and Dae-Won Kim[2,*]

[1]Center of Healthcare Ontology R&D, Seoul National University
Yeongeon-dong, Jongro-gu, Seoul, Korea
[2]School of Computer Science and Engineering, Chung-Ang University
221 Heukseok-dong, Dongjak-gu, Seoul, Korea
([*]Corresponding author: dwkim@cau.ac.kr)

Abstract. A great deal of research has been made to model the vagueness and uncertainty in information retrieval. One such research is fuzzy ranking models, which have been showing their superior performance in handling the uncertainty involved in the retrieval process. However, these conventional fuzzy ranking models are limited to incorporate the user preference when calculating the rank of documents. To address this issue, we develop a new fuzzy ranking model based on the user preference.

1 Introduction

In recent years a great deal of research in information retrieval has aimed at modelling the vagueness and uncertainty which invariably characterize the management of information. The application of fuzzy set theory to IR have concerned the representation of documents and the query [1], and many fuzzy ranking models such as MMM, PAICE, and P-NORM have been showing their superior performance in handling the uncertainty in the retrieval process [2–4]. The ranking is achieved by calculating a similarity between two fuzzy sets, a document D and a query Q. However, in spite that the user has an ability to reflect their preference for the information need in searching, these conventional models are limited to incorporate the user preference when calculating the rank of documents. Let us suppose that we are given a vector of query Q with a fuzzy set of the term and its membership degree:

$Q = \{fuzzy(0.8), IR(0.7), korea(0.3), author(0.2)\}$

A document collection consists of four documents (D_1, D_2, D_3, D_4) in which each document is represented as a fuzzy set of the index term and its weight.

$D_1 = \{fuzzy(0.8), IR(0.7)\}$
$D_2 = \{fuzzy(0.2), IR(0.2), korea(0.3), author(0.2)\}$
$D_3 = \{korea(0.7), IR(0.8)\}$
$D_4 = \{fuzzy(0.8), IR(0.7), korea(0.3), author(0.2)\}$

Given a query Q, we are wondering what is the best result of ranking? Intuitively, we know that D_4 is the most relevant document and D_3 is the least