

## Chapter 72

# Context Window Based Co-Occurrence Approach for Improving Feedback Based Query Expansion in Information Retrieval

**Jagendra Singh**

*Jawaharlal Nehru University, India*

**Aditi Sharan**

*Jawaharlal Nehru University, India*

### **ABSTRACT**

*Pseudo-relevance feedback (PRF) is a type of relevance feedback approach of query expansion that considers the top ranked retrieved documents as relevance feedback. In this paper the authors focus is to capture the limitation of co-occurrence and PRF based query expansion approach and the authors proposed a hybrid method to improve the performance of PRF based query expansion by combining query term co-occurrence and query terms contextual information based on corpus of top retrieved feedback documents in first pass. Firstly, the paper suggests top retrieved feedback documents based query term co-occurrence approach to select an optimal combination of query terms from a pool of terms obtained using PRF based query expansion. Second, contextual window based approach is used to select the query context related terms from top feedback documents. Third, comparisons were made among baseline, co-occurrence and contextual window based approaches using different performance evaluating metrics. The experiments were performed on benchmark data and the results show significant improvement over baseline approach.*

DOI: 10.4018/978-1-5225-5191-1.ch072

## 1. INTRODUCTION

The field of information retrieval system (IR) is as old as the computer itself. According to authors (Mooers, 1950; Savino & Sebastiani, 1998), “Information retrieval is the name of the process or method whereby a prospective user of information is able to convert his need for information into an actual list of documents in storage containing information useful to him”. IR are useful in large number of applications such as search engines (Singh et al., 2013), media search, digital libraries, recommender system, information filtering and many other’s applications so there is a constant need to improve such information systems. In this context, information retrieval is an active research field in computer science area.

The most critical problem for retrieval effectiveness is the term mismatch problem (Furnas et al., 1997; Xu, 1997): the indexers and the users do often not use the same words for the same concept or idea. One of the most feasible and successful technique to handle the problem of term mismatch is to expand the original query(Query Expansion) with other words that describes the user intention or a query that is more likely to retrieve only the relevant documents. In order to consider the above problem, there is a need of automatic query expansion techniques that can assist the user in formulating the query. The query expansion may be done in different ways: manual, interactive and automatic. The type of interactive query expansion is better than automatic query expansion because both the user and system are involved in the process. But in most of the time it is not feasible to involve the user in the process of query expansion, therefore a lot of researcher’s are trying to develop efficient techniques for automatic query expansion. Researchers work with co-occurrence information for expanding user query, but it has many drawbacks.

The concept of term co-occurrence has been used since the 90’s for identifying some of the semantic relationships among terms present in text documents. According to Rijsbergen (Rijsbergen, 1997), the idea of using co-occurrence statistics is used to detect some kind of semantic relations between query and document terms and exploiting it to expand the user’s queries. In fact, this idea is based on the following hypothesis: “*If an index term is good at discriminating relevant from non-relevant documents then any closely associated index term is likely to be good at this*”. Following are some well known co-occurrence coefficient measuring methods:

$$\text{Jaccard}(t_i, t_j) = \frac{d_{ij}}{d_i + d_j - d_{ij}} \quad (1)$$

$$\text{dice}(t_i, t_j) = \frac{2d_{ij}}{d_i + d_j} \quad (2)$$

$$\text{cosine}(t_i, t_j) = \frac{d_{ij}}{\sqrt{d_i d_j}} \quad (3)$$

15 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:  
[www.igi-global.com/chapter/context-window-based-co-occurrence-approach-for-improving-feedback-based-query-expansion-in-information-retrieval/198616](http://www.igi-global.com/chapter/context-window-based-co-occurrence-approach-for-improving-feedback-based-query-expansion-in-information-retrieval/198616)

## Related Content

---

### Intent-Based User Segmentation with Query Enhancement

Wei Xiong, Michael Recceand Brook Wu (2013). *International Journal of Information Retrieval Research* (pp. 1-17).

[www.irma-international.org/article/intent-based-user-segmentation-with-query-enhancement/109659](http://www.irma-international.org/article/intent-based-user-segmentation-with-query-enhancement/109659)

### A Hybrid Classification Approach Based on Decision Tree and Naïve Bays Methods

Saed A. Muqasqas, Qasem A. Al Radaidehand Bilal A. Abul-Huda (2014). *International Journal of Information Retrieval Research* (pp. 61-72).

[www.irma-international.org/article/a-hybrid-classification-approach-based-on-decision-tree-and-nave-bays-methods/127364](http://www.irma-international.org/article/a-hybrid-classification-approach-based-on-decision-tree-and-nave-bays-methods/127364)

### To Implement an Open-MAS Architecture for Semantic Web Services Discovery: What Kind of P2P Protocol Do We Need?

Mohamed Gharzouliand Makhlof Derdour (2018). *Information Retrieval and Management: Concepts, Methodologies, Tools, and Applications* (pp. 1184-1199).

[www.irma-international.org/chapter/to-implement-an-open-mas-architecture-for-semantic-web-services-discovery/198594](http://www.irma-international.org/chapter/to-implement-an-open-mas-architecture-for-semantic-web-services-discovery/198594)

### Taxonomy Based Fuzzy Filtering of Search Results

S. Vrettosand A. Stafylopatis (2004). *Intelligent Agents for Data Mining and Information Retrieval* (pp. 226-240).

[www.irma-international.org/chapter/taxonomy-based-fuzzy-filtering-search/24166](http://www.irma-international.org/chapter/taxonomy-based-fuzzy-filtering-search/24166)

### Biometrics: Electronic Resources Management in Jaykar Library, University of Pune, India

Prakash Dongardiveand Neela Deshpande (2013). *Information Retrieval Methods for Multidisciplinary Applications* (pp. 226-249).

[www.irma-international.org/chapter/biometrics-electronic-resources-management-jaykar/75910](http://www.irma-international.org/chapter/biometrics-electronic-resources-management-jaykar/75910)