

# Search Engine Personalization Tool Using Linear Vector Algorithm

Wadee Al Halabi  
University of Miami  
w.alhalabi@umiami.edu

Miroslav Kubat  
University of Miami  
mkubat@miami.edu

Moiez Tapia  
University of Miami  
mtapia@miami.edu

**Keywords:** Search Engine, personalization, vector space, machine learning, parallel programming.

**Abstract:** Internet search engine result is a dynamic research field, full of challenging tasks. In this paper, we introduced the linear vector algorithm (LVA) to provide an efficient way to rank the search engine's result. We conducted several experiments to compare the new algorithm with an existing one (Vector Space Algorithm) and found a promising result. This method uses a knowledge-based system. The information supplied to the knowledge-based system is acquired from the user behavior. The system monitors the user behavior and feeds positive and negative examples to the knowledge-based system. The system is able to be trained and produces good result compared with other algorithms. The search engine's result was improved tremendously when this system was applied to the search engine.

## 1. Introduction

In this study, we attempt to determine if a search engine's result could be enhanced by exploring web page's contents. Presuppositions in this work appear to lend weight to our hypothesis that a new algorithm could be developed to help inexperienced users utilize search engines more efficiently. It was our belief that if the web pages text can be obtained, then the rank according to each user's interest would be improved.

Search engines are the only available search tool on the web today. Although they are widely used, search engine results are not efficient to satisfy the user. We suggest two main reasons that clarify this inadequacy:

1. The tremendous amount of information on the web, which is an advantage to the web.
2. Our tool seems to have more promising solutions.

Exploring the text content of a web page with a personalized search engine using new tools such as linear algorithm, concurrent programming, and agent technology, in addition to the use of learning ability, would result in a much greater ranking efficiency.

Our study is centered on the hypothesis that we can design an algorithm which explores the search engine result, and downloads the text contents of each URL appearing in the search result. After that, the algorithm stores the text of each URL in a distinct

memory location. Then, it compares the content of each URL with a reference document provided by the user. Lastly, the algorithm assigns a weight to each document and ranks them accordingly. If the suggested algorithm is intelligent enough to learn, we can apply its knowledge to the ranking task. This hypothesis implies that the new rank is much better than the original search's result. It also implies that the algorithm achieves the new rank using its knowledge.

We designed an algorithm to embark on this quest. To show that this algorithm improved the search engine rank result, we conducted several experiments. Finally, we compared our algorithm "Linear Vector Algorithm", with a widely used algorithm "Vector Space Algorithm" The findings were spectacular.

## 2. Assumptions

There has been a great deal on personalized search engines in earlier literature [1, 2, 6, 9]. Fan et al. have published an interesting paper on personalization of search engine [1]. The authors implemented their framework as a specific approach rather than a generic one. They showed how this method improved the search's efficiency. The authors believe that the personalized information retrieval and delivery is an imminent problem facing search engines. They argue that the term-weighted strategy should be context-specific, which we did not use for simplicity because this strategy would increase the new algorithm complexity.

Sufficient amount of recent activity in machine learning contributed to information retrieval and web search. Boyan et al. in [2] introduced a heuristic method to optimize the search's result, and therefore, improve the overall system performance. In their work, Boyan et al. showed how they successfully implemented a machine learning approach to improve the search's retrieval efficiency. "Learning Architecture for Search Engine Retrieval" is the name of their system. They designed it to enhance search engine's performance. The authors implemented a learning algorithm to assign different weights according to the word location in a text. Their algorithm assigns a weight to a word in headlines, for instance, and assigns a totally different weight to the same word in the text body. Their method is considered as content-based in the way it assigns weights.