Πανεπιστήμιο Κρήτης

Τμήμα Επιστήμης Υπολογιστών

HY463 Συστήματα Ανάκτησης Πληροφοριών

Εξάμηνο: Άνοιξη 2022

**Γραπτή Αναφορά Έργου**

**BioMedic Indexer**

Στοιχεία Φοιτητών

|  |  |
| --- | --- |
| Ονοματωπώνυμο | Manos Chatzakis |
| ΑΜ | 4238 |
| Email | csd4238@csd.uoc.gr |

**Table of Contents**

[1 Introduction 3](#_Toc471646822)

[2 Implementation 3](#_Toc471646823)

[3 Experimental Evaluation 3](#_Toc471646826)

[4 Conclusion 3](#_Toc471646827)

[5 references 3](#_Toc471646828)

# Introduction

This project is the implementation of a BioMedical Search Engine over a biomedical document collection of 5GB.

**How to run:**

# Implementation

In this section, the basic methods used to implemented both parts of BioMedic Indexer (Index Creation and Query Answering are described).

## Index Creation

BioMedic Indexer index a selected directory in three steps: (a) Partial Indexing (b) Partial Merging and (c) Document Norm Calculation.

**Partial Indexing.**

BioMedic Indexer

**Merging.**

**Document Norm Calculation.**

## Query Answering

**Vector model.**

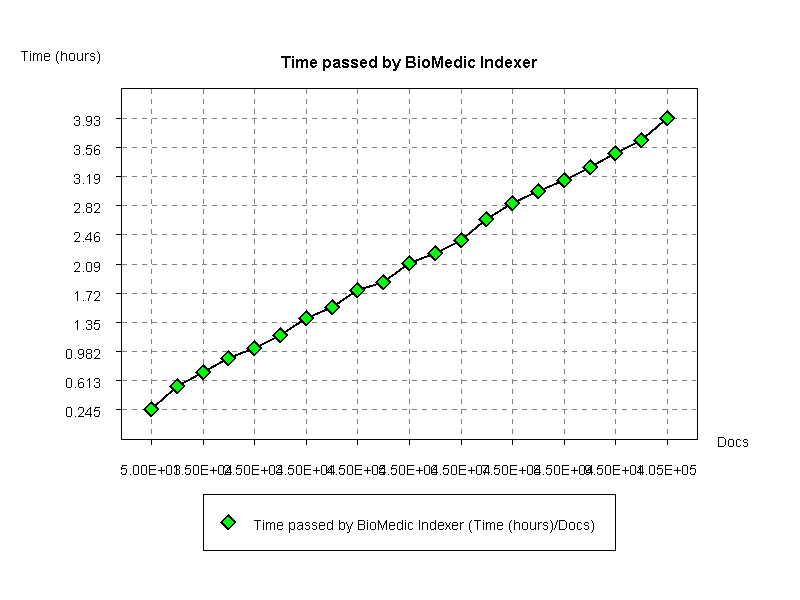
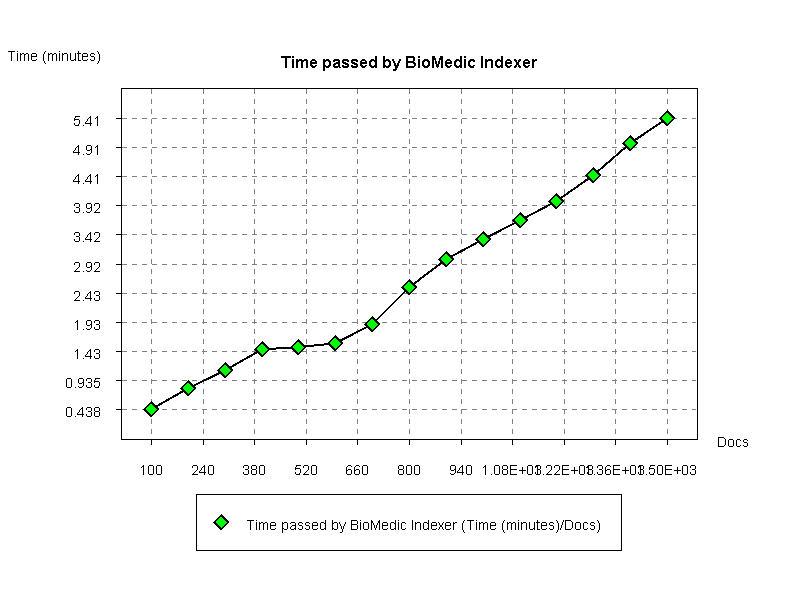
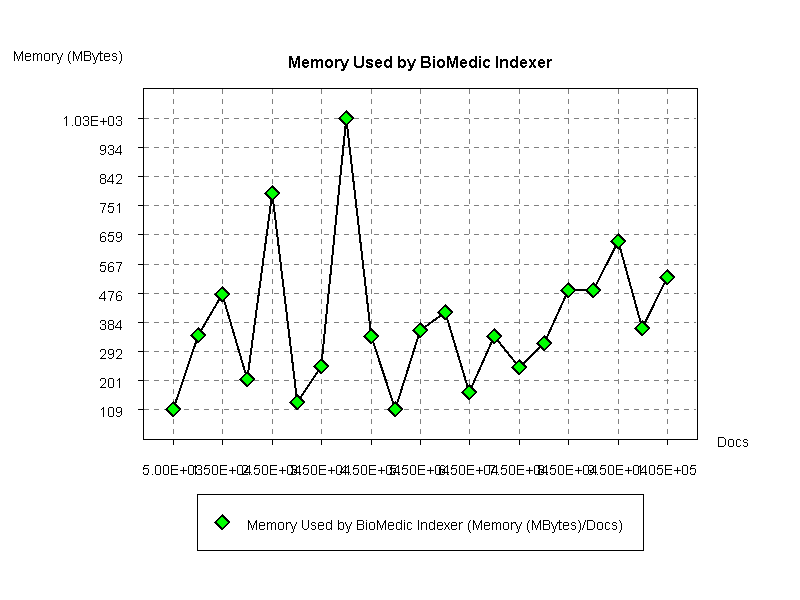
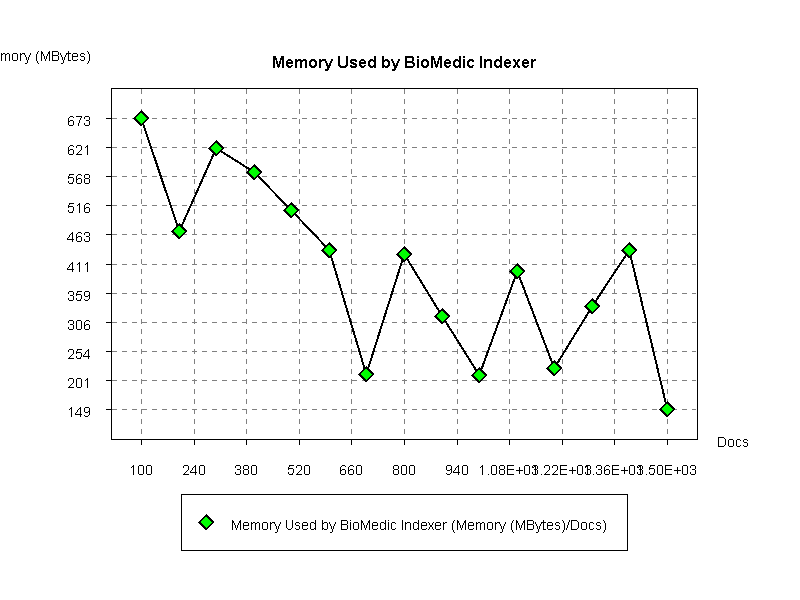
**Vector model with Examination type support.**

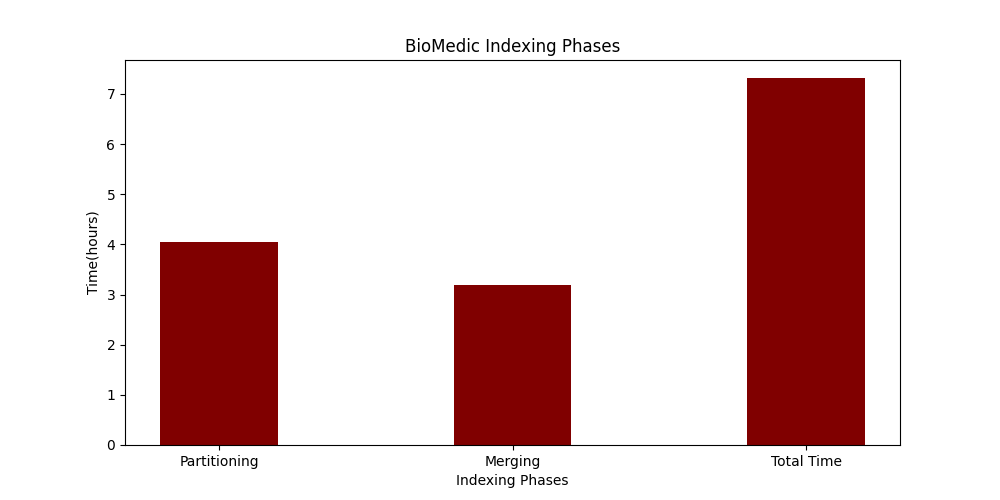
# Experimental Evaluation

The experiments contacted on a machine of 8GB Memory, 256GB Disc, and a 4-Core i5 8th-gen CPU, running Windows 11.

## Index Creation Evaluation

The evalution of the indexing phase is presented in two graphs: A graph showing the memory usage correlated with the document count, and a graph showing the total time passed correlated with the document count. For both of them we show two versions: One for the whole indexing process and one for the indexing of a subset of the collection.





## Query Answering Evaluation

For query answering, we show the total response time needed for the queries created from the files of “topics.txt”.

# Conclusion