Department of Computer Engineering &
Informatics, University of Patras

Information Retrieval

Laboratory Exercise

Winter Semester 2022

Teacher: H. Makris

Pronunciation

In the context of this lab exercise, you are asked to implement a book search engine that will be based on Elasticsearch and will decide the order of presentation of the results using machine learning techniques. No implementation language is defined but the use of Python and the libraries pandas, scikit-learn, tensorflow and keras is recommended.

Question 1

First, you will need to install Elasticsearch on your system and write a small program that will read the records contained in the BX-Books.csv file and import them into Elasticsearch. Next, you should write a program that will accept as input (either as a command-line argument or at runtime) an alphanumeric and an integer, the user's ID. This program will return the list of books that match the alphanumeric input ordered in descending order according to a metric that you create that takes into account Elasticsearch's similarity metric and the user's score on the book (if she is available). The returned list should only show the top 10% of books with the best match. The user ratings for each book can be found in the file BX-Book-Ratings.csv.