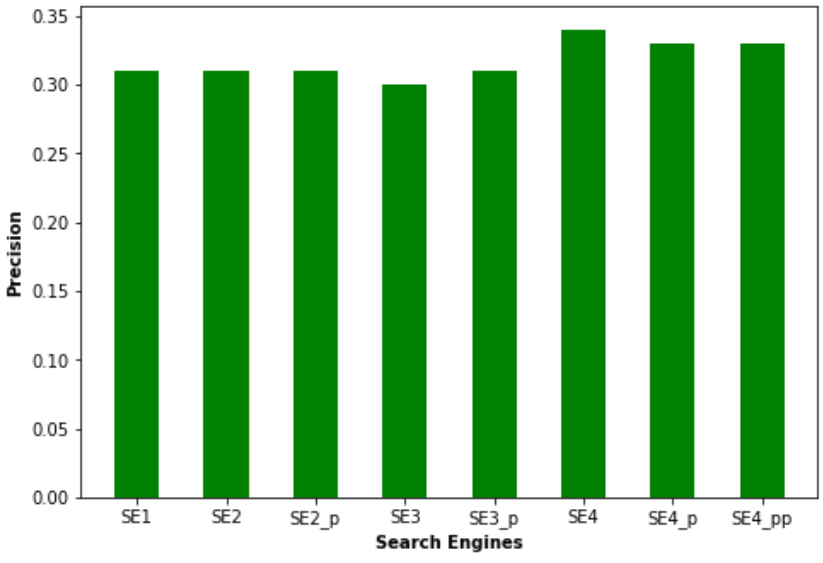
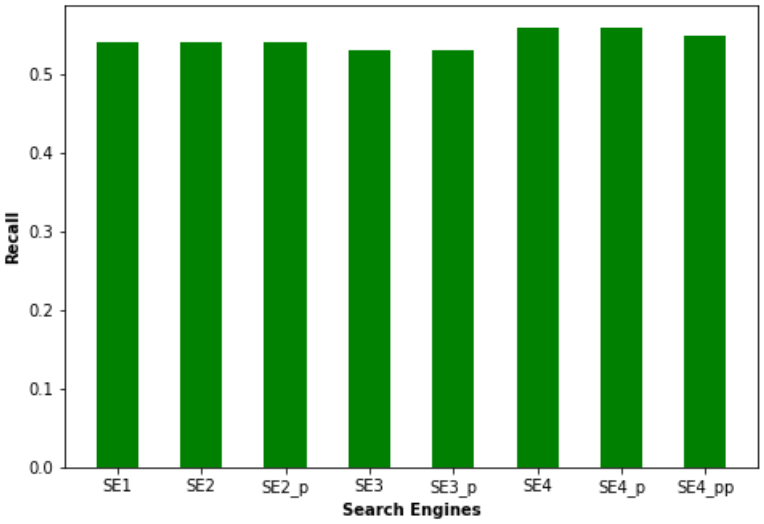
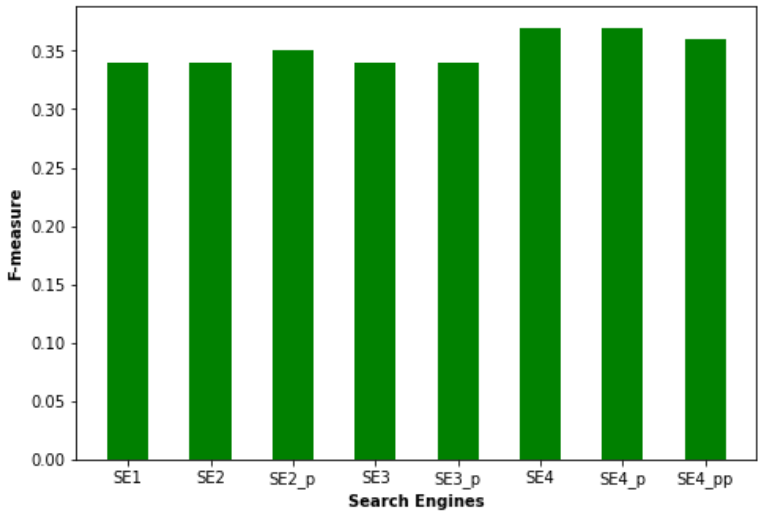
A Report On "Implementation Of Elasticsearch on a Collection of Persian Poems"

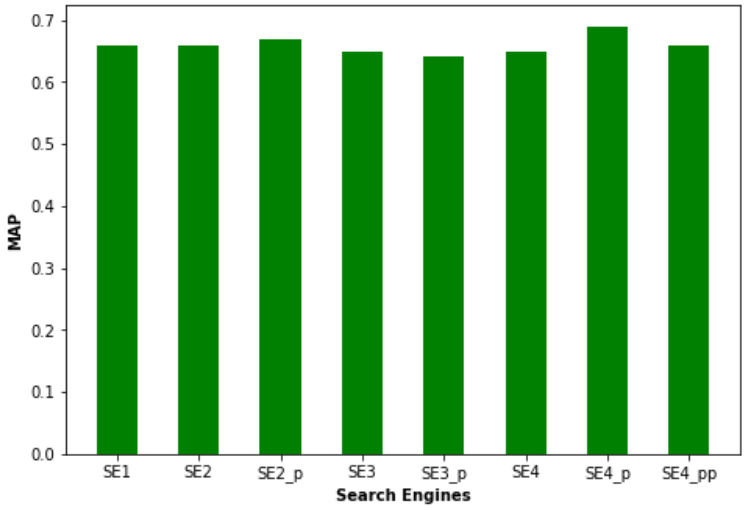
In this project, using *Elasticsearch* we have implemented a few search engines on a collection of persian poems, each one having a different preprocessing stage. The text files' preprocessing is performed according to the mentioned method, using *Hazm* and *Parsivar* libraries. After indexing all files into *Elasticsearch*, we have evaluated the performance of our applications on the given queries with calculating the four common evaluative criteria of an information retrieval system performance which are Precision, Recall, F-measure and MAP. The tables and bar graphs related to the application test results are as follows:

	SE1	**SE4**	**SE4'**	**SE4''**
Precision	0.31	0.34	0.33	0.33
Recall	0.54	0.56	0.56	0.55
F-measure	0.34	0.37	0.37	0.36
MAP	0.66	0.65	0.69	0.66




Recall
0.0
0.1
0.2
0.3
0.4
0.5
SE1
SE2
SE2_p
SE3
SE3_p
SE4
SE4_p
SE4_pp
Search Engines


F-measure
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
SE1
SE2
SE2_p
SE3
SE3_p
SE4
SE4_p
SE4_pp
Search Engines



These applications generally return a very large number of ranked results for queries, while according to the given relevance assessment data, the number of documents associated with each query is much smaller, so the precision of all eight SEs is on average low. Data preprocessing has caused a slight increase in this measure and the highest precision is obtained using the lemmatization function of *Hazm* library which helps to better limit the number of query results. As mentioned, the number of query results is very large and includes most of the desired relative documents, so the recall is on average better than the precision measure and both libraries have similar performance in this regard. The only difference is that the stemming function of *Hazm* has reduced this criterion due to incorrect word stemming in most cases. The range of f-measure is a function of precision and recall. The last criterion to be mentioned is MAP. It shows that all SEs have performed relatively well in ranking a large number of query results. We have reached the highest level of MAP in SE4', which shows that the lemmatization function of *Parsivar* significantly improves the ranking of documents. Overall, compared to the first search engine built with raw data, these two libraries made a slight improvement in system performance. In general, by comparing the two criteria of f-measure and MAP, it can be understood that *Parsivar* has more improved the performance of search engines on average despite producing longer tokens that contain several words, and also the lemmatization function of *Parsivar* works better.