This document contains summary results obtained with the hybrid algorithm, using all dump data of English Wikipedia articles.

The doc2vec-dbow algorithm was used as the semantic algorithm. HDP (HierarchicalDirichletProcess) algorithm was used as the Semantic Cluster algorithm. While the number of feature vectors in the Doc2vec-dbow algorithm is taken as 128, the number of topic vectors in the HDP algorithm is kept as 100. Wikipedia number of documents: 5.009.203. Training was carried out by selecting 1 million random documents in the Doc2vec-dbowal algorithm. However, the infer vector number is all documents. All documents were used in the model and infer phase of the HDP algorithm. As a pre-processing process, the removal of stop words and stemming were applied. 17.3 GB text data decreased to 12.1 GB. In order to infer faster in the HDP algorithm, 60 important sentences were extracted using the Sumbasical algorithm. For testing, 2500 sentences were extracted. LexRank, TextRank, Luhn, Lsa algorithms extracted 24 sentences. The success of the ball, which is one in 10 thousand: 500, was taken into consideration. Success was automatically achieved depending on whether the 500 most similar documents in the extracted content contained the original document id or not. Search algorithm C++ and Multithread approach were used. Total Number of Searches: 5,000,000*2,500=over 12.5 billion.

Sonuç-1: (Effect of Min Probe Value)

As the min probe value increases, the number of searches decreases and the search becomes faster. However, success may decrease. When 0.04 is used instead of 0.01 probability, there is a difference of 1 in a thousand. However, it accelerates by around 30%. For 100 topics, the Min probe value can be selected between 0.01 and 0.04. In terms of success and speed balance, 0.04 can be determined.

Min Prob Prm	Acu (%)	Time (second)
0.01	99.84	142.9
0.02	99.88	131.8
0.04	99.76	116.7
0.05	99.68	100.7
0.1	99.32	87.5

Sonuç-2: (Effect of Cluster Number)

As the number of clusters increases, the number of searches increases and the search time also increases. Success is also increasing. The number of clusters can be at least 3. Ideally it should be 5 or above. Min probe value is 0.01. Min Probe: Instead of 0.01 and 5 clusters, min probe 0.04 and 100 cluster number can be preferred. Parameters 0.04 and 100 are both more successful and faster.

Cluster Number	Acu. (%)	Time (Sec)
1	93.04	53.9
2	97.92	81.9
3	99.2	101.2
5	99.68	127.6
100	99.84	142.9

Sonuç-3: (Effect of the Algorithm

The Hybrid Algorithm has almost the same performance as the original algorithm, Doc2vec-dbow. It maintains the success of the original semantic algorithm by searching in the most likely groups without performing a full search. On the lower side, different semantic algorithms such as Doc2vec-dbow can be used. It reduces the search time by approximately 9-10 times for 5 million data and 100 topics. Instead of 0.01 min probe, 0.04 value can be preferred to speed up the probe by around 25-30%.

Algorithm	Acu. (%)	Time (sec)
Doc2vec-Dbow	100.0	1015.3
Hybrid Algorithm (minProb:0.01 – ClusterCnt:100)	99.76	140.4
Hybrid Algorithm (minProb:0.04 – ClusterCnt:100)	99.67	115.3