

## 4. Introduction to Information Retrieval

### a. Flat Clustering- **Chapter 16**

- i. Clustering can also speed up search.
- ii. For search result clustering, we may want to measure the time it takes users to find an answer with different clustering algorithms. This is the most direct evaluation, but it is expensive, especially if large user studies are necessary.
- iii.
- iv. K Means
- v. Fuzzy K Means
- vi. EM
- vii.

### b. Hierarchical Clustering - Chapter 17 - Youtube was better

### c. Labels - Pick terms with high info gain?

- i.

Having more data for training neural networks is usually desirable, but it has to be of good quality to give good results.

Bunch of ES books -

<https://github.com/BlackThursdays/https-github.com-TechBookHunter-Free-Elasticsearch-Books>