**Start with these resources. We encourage you to add on to the list any items you believe would benefit the group. Feel free to modify/add notes to any of the rows/columns. This is for you. A big part of AI/computer programming/software design/other cs fields is knowing how to make the right search queries.**

|  |  |  |
| --- | --- | --- |
| **Overview (relevant to our project)** | **Useful for/when...** | **Link** |
| Official: Common Repertoire | **[model building repertoire with packages]** | <https://scikit-learn.org/stable/tutorial/text_analytics/working_with_text_data.html> |
| University Resouce: Rationale | **[rationale for our methods]** | <http://www1.se.cuhk.edu.hk/~seem5680/lecture/vector-classify.pdf> |
| Official: sklearn’s cosine\_similarity | **[sklearn’s cosine similarity implementation]** | <https://scikit-learn.org/stable/modules/generated/sklearn.metrics.pairwise.cosine_similarity.html> |
| Blog: Example TFIDF + cosine similarity implementation from scratch | **[TFIDF+ cos sim]**  Getting a grip on the math behind the implementations | <https://janav.wordpress.com/2013/10/27/tf-idf-and-cosine-similarity/> |
| Blog: Example TFIDF + cosine similarity implementation using Python packages | **[TFIDF+ cos sim]**  Blackbox version of cosine\_similarity implementation | <https://stackoverflow.com/questions/12118720/python-tf-idf-cosine-to-find-document-similarity> |
| Stackoverflow: Incompatible TFIDF matrix shape in computing cosine similarities | **[error resolution]** | <https://stackoverflow.com/questions/40366175/handle-incompatible-matrices-shapes-in-tf-idf> |
| Google: finding the source of a Memory error | **[error resolution]**  Re-evaluating exactly what we’re doing? | <https://www.google.com/search?ei=Y9c1Xbj3LeWyggfV1r3IDw&q=cosine+similarity+memory+error&oq=cosine+similarity+memory+error&gs_l=psy-ab.3..0j0i22i30l2.2040.14526..14638...5.0..0.161.3173.27j7......0....1..gws-wiz.....10..35i39j0i273j0i131j0i67j0i10j0i20i263.Z1TaP49A5Ak&ved=0ahUKEwj4tNfu7MjjAhVlmeAKHVVrD_kQ4dUDCAo&uact=5> |
| Paper: Description of various baseline implementations for FNC-1 | **[classifier, rationale]**  Identifying baseline classifier for TF-IDF + cosine similarity & looking for a different approach to problem | <https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1174/reports/2761028.pdf> |
| sklearn.linear\_model.SGDClassifier | **[classifier]**  SVM with SGD optimizations | <https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.SGDClassifier.html> |
| Text Classification Methods | **[classifier]**  Searching for good classifier | <http://www1.se.cuhk.edu.hk/~seem5680/lecture/vector-classify.pdf> |
| Repertoire for making predictions and evaluations | **[making predictions]**  Evaluting the accuracy of our predictions | <https://towardsdatascience.com/machine-learning-nlp-text-classification-using-scikit-learn-python-and-nltk-c52b92a7c73a> |
| Blog: More on Cosine Similarity | Understanding the Math | <https://www.machinelearningplus.com/nlp/cosine-similarity/> |
| Blog: Overview of text (headline vs. body) similarity metrics | You have extra time to explore beyond cosine similarity | <https://towardsdatascience.com/overview-of-text-similarity-metrics-3397c4601f50> |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |