Jonathan Mandl 211399175

Danielle Hodaya Shrem 208150433

Part 2 – External Word Embeddings

For this section, we loaded pre-trained word-embedding vectors and their corresponding vocabulary, normalized each vector to unit length, and then computed cosine similarity (via dot product) between our five query words and every other word in the vocabulary.

We recorder the top 5 nearest neighbors for each query word:

Top 5 most similar to **dog**:

1. cat (cosine similarity: 0.7709)
2. rabbit (cosine similarity: 0.7320)
3. puppy (cosine similarity: 0.6966)
4. frog (cosine similarity: 0.6485)
5. kitten (cosine similarity: 0.6349)

Top 5 most similar to **england:**

1. ireland (cosine similarity: 0.9009)
2. scotland (cosine similarity: 0.8501)
3. australia (cosine similarity: 0.8063)
4. wales (cosine similarity: 0.7983)
5. europe (cosine similarity: 0.7753)

Top 5 most similar to **<u>john</u>**:

1. george (cosine similarity: 0.9220)
2. robert (cosine similarity: 0.8988)
3. charles (cosine similarity: 0.8947)
4. william (cosine similarity: 0.8894)
5. james (cosine similarity: 0.8783)

Top 5 most similar to **<u>explode</u>**:

1. grenadiers (cosine similarity: 0.5554)
2. slashed (cosine similarity: 0.5508)
3. appendix (cosine similarity: 0.5339)
4. monnet (cosine similarity: 0.5242)
5. instantaneously (cosine similarity: 0.5183)

Top 5 most similar to **<u>office</u>**:

1. board (cosine similarity: 0.6682)
2. court (cosine similarity: 0.6198)
3. offices (cosine similarity: 0.6174)
4. commission (cosine similarity: 0.6167)
5. authority (cosine similarity: 0.6033)