PROGRESS REPORT
Reproduce A Paper: A cross-collection mixture model for comparative text mining

1. **Data Collection (Partially Completed)**

   Dataset 1: Laptop reviews (Completed)

   Dataset 2: War news (Not Started)

   This should be done by manually searching and downloading 30 news articles from BBC or CNN for each of the two wars, published in one year span (May 2003 - April 2004 for Iraq war, Nov 2001 to Oct 2002 for Afghanistan war). This would be a very close proximation of the war news dataset that is used in the paper.

2. **Algorithm Implementation (Completed)**

3. **Experimental Results Gathering (Partially Completed)**

   The experimental results should be collected and summarized in a table of theme clusters as the same way in the paper to facilitate comparison.

4. **Results Comparison and Evaluation (Not Started)**

   Baseline model (SimpMix) should be implemented, and the results from the CCMix model should be compared with those from the baseline model to evaluate the performance of CCMix model and compare with the conclusions from the paper.

5. **Preparation of Documentation and Presentation (Not Started)**

   A readme file and a video presentation should be prepared for the final submission.