Real World Example:

NewsCheck

You have a large corpus of scraped

use in NLP experiments.

public news articles stored in S3, for