General comments

- Good that you made the paper in two columns
- Usually scientific papers use "we" instead of "I"
- Your language could be improved, there's a lot of weirdly phrased sentences such as "So I'm going to show and compare the resulting performance for all 3 named methods and discuss about those results" in the abstract
- Short - needs more details

Abstract

- Use `\begin{abstract}` to create an abstract in LaTeX
- An abstract should be about your work and your findings, currently you're describing what the reader can find in the next sections of your paper. To fix this, introduce the methods, a brief explanation of them and lastly your findings, e.g. "all methods performed significantly better than BM25" if thats the case.
- Cool that you included index terms

Introduction

- You introduced ad-hoc retrieval but don't mention the difficulties of the task.
- *Missing/not detailed enough* What is the purpose of using text embedding for this task?
- Very short

Method

BM25

- Using `\left(` and `\right)` will increase size of parantheses in LaTeX
- You mention relevance information r_i but it is not part of your equation?
- I don't believe there is a $_2$ in BM25? Where did you find this BM25 equation?
- Your implementation differs from your reports equation, e.g. you do not have a k_2
- I'm pretty sure the majority of the code is from `https://github.com/fanta-mnix/python-bm25/blob/master/bm25.py`, you should probably credit this to avoid issues.
- Instead of implementing this yourself you could simply use Anserini[1] which was linked to in the assignment text

QE

- Short - Could go more into detail, e.g. what is $qCent$? what is t?
- Missing exponential in your implementation of S_{Cent}
- Gensim does not specify whether similar_by_vector is using cosine similarity or not. I assume it is so i reckon that is fine.
- Your `fusion` is incorrect, seems like you implemented Lq but didn't do the $p(t|q_i)$ part, at least i can't find any division or exponential operators
- Missing interpolation

[1] https://github.com/castorini/anserini/blob/master/docs/experiments-robust04.md

BM+WE

- Multiply (not multiplicate)
- What is BM25 with word embeddings? You briefly explain how you've implemented but you aren't describing what it is

Experiments

- I would like a brief look at some of the files you create here, seeing that i don't have them available, like the first 5-10 lines of each file in an appendix
- Very brief - Can go more into detail, like, which trec_eval paramaters did you pass?

Results

- "mthod"
- You seem to get very poor results, if you look at the Anserini github page you'll find that they get a MAP of 0.2531 on the robust04 dataset so it seems weird that you get a third of that. This leads me to believe that your BM25 implementation is incorrect.
- Your Query Extended results looks quite good in comparison to BM25, but that could be a coincidence.

Discussion

- Compare the results to the theoretics, did the results match your expectations? Your results are lower than the original BM25, was this the intention?

For final handin

- Finish up with BM25 + WE

- Go more into detail in the report based on what i have written above
- Fix your code. You'll probably find that Anserini can make your life easier for this task.