

University of Waterloo E-Thesis Template for L^AT_EX

by

Zeynep Akkalyoncu Yilmaz

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Mathematics
in
Computer Science

Waterloo, Ontario, Canada, 2018

© Zeynep Akkalyoncu Yilmaz 2019

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner: Bruce Bruce
 Professor, Dept. of Philosophy of Zoology, University of Wallamaloo

Supervisor(s): Doris Johnson
 Professor, Dept. of Zoology, University of Waterloo
 Andrea Anaconda
 Professor Emeritus, Dept. of Zoology, University of Waterloo

Internal Member: Pamela Python
 Professor, Dept. of Zoology, University of Waterloo

Internal-External Member: Deepa Thotta
 Professor, Dept. of Philosophy, University of Waterloo

Other Member(s): Leeping Fang
 Professor, Dept. of Fine Art, University of Waterloo

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

This is the abstract.

Vulputate minim vel consequat praesent at vel iusto et, ex delenit, esse euismod luptatum augue ut sit et eu vel augue autem feugiat, quis ad dolore. Nulla vel, laoreet lobortis te commodo elit qui aliquam enim ex iriure ea ullamcorper nostrud lorem, lorem laoreet eu ex ut vel in zzril wisi quis. Nisl in autem praesent dignissim, sit vel aliquam at te, vero dolor molestie consequat.

Tation iriure sed wisi feugait odio dolore illum duis in accumsan velit illum consequat consequat ipsum molestie duis duis ut ullamcorper. Duis exerci odio blandit vero dolore eros odio amet et nisl in nostrud consequat iusto eum suscipit autem vero. Iusto dolore exerci, ut erat ex, magna in facilisis duis amet feugait augue accumsan zzril delenit aliquip dignissim at. Nisl molestie nibh, vulputate feugait nibh luptatum ea delenit nostrud dolore minim veniam odio volutpat delenit nulla accumsan eum vero ullamcorper eum. Augue velit veniam, dolor, exerci ea feugiat nulla molestie, veniam nonummy nulla dolore tincidunt, consectetur dolore nulla ipsum commodo.

At nostrud lorem, lorem laoreet eu ex ut vel in zzril wisi. Suscipit consequat in autem praesent dignissim, sit vel aliquam at te, vero dolor molestie consequat eros tation facilis diam dolor. Odio luptatum dolor in facilisis et facilisi et adipiscing suscipit eu iusto praesent enim, euismod consectetur feugait duis. Odio veniam et iriure ad qui nonummy aliquip at qui augue quis vel diam, nulla. Autem exerci tation iusto, hendrerit et, tation esse consequat ut velit te dignissim eu esse eros facilisis lobortis, lobortis hendrerit esse dignissim nisl. Nibh nulla minim vel consequat praesent at vel iusto et, ex delenit, esse euismod luptatum.

Ut eum vero ullamcorper eum ad velit veniam, dolor, exerci ea feugiat nulla molestie, veniam nonummy nulla. Elit tincidunt, consectetur dolore nulla ipsum commodo, ut, at qui blandit suscipit accumsan feugiat vel praesent. In dolor, ea elit suscipit nisl blandit hendrerit zzril. Sit enim, et dolore blandit illum enim duis feugiat velit consequat iriure sed wisi feugait odio dolore illum duis. Et accumsan velit illum consequat consequat ipsum molestie duis duis ut ullamcorper nulla exerci odio blandit vero dolore eros odio amet et.

In augue quis vel diam, nulla dolore exerci tation iusto, hendrerit et, tation esse consequat ut velit. Duis dignissim eu esse eros facilisis lobortis, lobortis hendrerit esse dignissim nisl illum nulla minim vel consequat praesent at vel iusto et, ex delenit, esse euismod. Nulla augue ut sit et eu vel augue autem feugiat, quis ad dolore te vel, laoreet lobortis te commodo elit qui aliquam enim ex iriure. Ut ullamcorper nostrud lorem, lorem laoreet eu ex ut vel in zzril wisi quis consequat in autem praesent dignissim, sit vel. Dolor at te, vero

dolor molestie consequat eros tation facilisi diam. Feugait augue luptatum dolor in facilisis et facilisi et adipiscing suscipit eu iusto praesent enim, euismod consectetur feugait duis vulputate veniam et.

Ad eros odio amet et nisl in nostrud consequat iusto eum suscipit autem vero enim dolore exerci, ut. Esse ex, magna in facilisis duis amet feugait augue accumsan zzril. Lobortis aliquip dignissim at, in molestie nibh, vulputate feugait nibh luptatum ea delenit nostrud dolore minim veniam odio. Euismod delenit nulla accumsan eum vero ullamcorper eum ad velit veniam. Quis, exerci ea feugiat nulla molestie, veniam nonummy nulla. Elit tincidunt, consectetur dolore nulla ipsum commodo, ut, at qui blandit suscipit accumsan feugiat vel praesent.

Dolor zzril wisi quis consequat in autem praesent dignissim, sit vel aliquam at te, vero. Duis molestie consequat eros tation facilisi diam dolor augue. Dolore dolor in facilisis et facilisi et adipiscing suscipit eu iusto praesent enim, euismod consectetur feugait duis vulputate.

Acknowledgements

I would like to thank all the little people who made this thesis possible.

Dedication

This is dedicated to the one I love.

Table of Contents

List of Tables	x
List of Figures	xi
1 Introduction	1
1.1 Contributions	1
1.2 Thesis Organization	2
2 Related Work	3
3 Cross-Domain Relevance Transfer with BERT	4
4 Experimental Results	5
5 Conclusion	6
References	7
APPENDICES	7
A PDF Plots From Matlab	8
A.1 Using the Graphical User Interface	8
A.2 From the Command Line	8

List of Tables

List of Figures

Chapter 1

Introduction

Document retrieval traditionally relies on term-matching techniques such as BM25 to judge the relevance of documents in creating a ranking in response to a query. Unfortunately, the set of techniques based on this approach neglects to exploit rich semantic information embedded in these documents. For this reason, with the advent of neural networks, researchers have been actively pursuing ways to extract and use such semantic information in information retrieval. In this setup, a list of candidate documents are retrieved using a standard term-matching technique, which is then re-ranked with a custom neural model.

However, despite active effort in this field of research, some researchers have recently voiced concern as to whether neural networks have truly contributed to progress in the field of information retrieval. [Yang et al. \(2019\) and other examples...](#)

1.1 Contributions

The main contributions of this thesis can be summarized as follows:

- We present two innovations to successfully apply BERT to *ad hoc* document retrieval with large improvements: integrating sentence-level evidence to address the fact that BERT cannot process long spans posed by newswire documents, and exploiting cross-domain models of relevance for collections without sentence- or passage-level annotations.

- We explore through various error analysis experiments on the effects of cross-domain relevance transfer with BERT as well as the contributions of BM25 and sentence scores to the final document ranking.
- With the proposed model, we establish state-of-the-art effectiveness on three standard TREC newswire collections at the time of writing. **neural or otherwise**
- something about demo, TREC DL...

1.2 Thesis Organization

The remainder of this thesis is organized in the following order: **add link to actual chapters** Chapter 2 reviews related work in neural document retrieval, particularly applications of BERT to document retrieval. Chapter 3 motivates the approach with some background information on the task, and introduces the datasets used for both training and evaluation as well as metrics. Chapter 4 proposes an end-to-end pipeline for document retrieval with BERT by elaborating on the design decisions and challenges. **What about TREC DL? MS MARCO?** Chapter 5 describes the experimental setup, and presents the results on three newswire collections – Robust04, Core17 and Core18. Chapter 6 concludes the thesis by summarizing the contributions and discussing future work.

Chapter 2

Related Work

Chapter 3

Cross-Domain Relevance Transfer with BERT

Chapter 4

Experimental Results

Chapter 5

Conclusion

APPENDICES