

Toni Linnusmäki

# **INCREASING E-COMMERCE CONVERSION RATE WITH RELEVANT SEARCH RESULTS USING ELASTICSEARCH**

Master's thesis  
Faculty of Information Technology and Communication Sciences  
Examiner: Prof. Martti Juhola  
April 2020

# ABSTRACT

Toni Linnusmäki: Increasing e-commerce conversion rate with relevant search results using Elasticsearch

Master's thesis

Tampere University

Master's Programme in Computer Science

April 2020

---

The abstract is a concise, self-containing one page description of the work: what was the problem, what was done, and what are the results. Do not include charts or tables in the abstract. Put the abstract in the primary language of your thesis first and then the translation when that is needed. International students do not need to include an abstract in Finnish.

Keywords: relevant search, e-commerce, Elasticsearch

The originality of this thesis has been checked using the Turnitin OriginalityCheck service.

# TIIVISTELMÄ

Toni Linnusmäki: Increasing e-commerce conversion rate with relevant search results using Elasticsearch

Maisterin tutkielma

Tampereen yliopisto

Tietojenkäsittelyopin maisteriopinnot

Huhtikuu 2020

---

Tiivistelmä on suppea, 1 sivun mittainen itsenäinen esitys työstä: mikä oli ongelma, mitä tehtiin ja mitä saatiin tulokseksi. Kuvia, kaavioita ja taulukoita ei käytetä tiivistelmässä. Laita työn pääkielellä kirjoitettu tiivistelmä ensin ja käännös sen jälkeen. Suomenkieliselle kandidaatintyölle pitää olla myös englanninkielinen nimi arkistointia varten.

Avainsanat: hakutulosten relevanssi, verkkokauppa, Elasticsearch

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck -ohjelmalla.

## PREFACE

This thesis is mean to serve as a starting point when a organization wants to improve their search application. The thesis introduces and utilizes the Elasticsearch to implement methods which purpose is to increase the conversion rate by offering more relevant search results for the customer.

Acknowledgements to the Power International As for letting me utilize their existing Elasticsearch cluster and allowed testing the implementations with real-world customers.

Tampere, 26th April 2020

Toni Linnusmäki

**CONTENTS**

1 Introduction . . . . . 1

References . . . . . 3

# 1 INTRODUCTION

Search engines have been a significant element of the internet for a long time, and some of the most visited websites in the world are search engine interfaces, like Google and Baidu [1]. By visiting different websites, people might interact with dozens of search engines, for example, while writing text to the address bar of a browser or looking for information about a topic of interest. In addition to interacting with search engines, people might not be aware that they are utilizing one of the most powerful tools of parsing and matching text.

The history of search engines goes decades back, but the era of modern search engines can be thought to have begun when Brin and Page [2] introduced their search engine, Google. They proposed methods to index a vast amount of documents, in their case web pages, and to use a ranking algorithm, PageRank, to rank the documents [2]. The usage of indexing and ranking algorithms has been the foundation for search engines ever since. While using ranking algorithms and the ranking search results by relevancy was not a new idea, it was revolutionary to utilize the PageRank ranking algorithm in search engines to provide more relevant search results to the queries of the users.

Besides handling most of the navigation on the internet, searching is a crucial part of an e-commerce platform. Sorokina and Cantu-Paz [3] introduced the usage of a search application in the largest e-commerce in the world, Amazon, and how the search application powers an extensive amount of sales of Amazon. Turnbull and Berryman [4] introduced text analysis and result ranking methods that can be efficiently utilized in a search engine to provide relevant search results to users.

While different relevancy methods and the relevancy of the search results has been investigated extensively in the past, little research has included the use of the relevancy methods in practise in search applications with real-world customers in e-commerce. The research has tended to focus on applying the methods in theory to a predetermined dataset to get measurable results as opposed to focusing on implementing the methods and measuring the performance in practice.

The primary focus of this research was to investigate how the relevancy of the search results affected the conversion rate of an e-commerce platform by implementing a selection of methods from the literature. The conversion rate measures what proportion of customers who visited a website bought something, and it provided a generic measurement of how the implemented methods performed in providing relevant search results to

customers. However, the relevancy of the search result is subjective since not all customers are searching for the same thing making the relevancy of a search result different for each customer.

The implementation of the methods in this research was done to an existing e-commerce platform, which serves tens of thousands of customers daily across the Nordic countries. Using this platform, a set of tests was performed to a set of customers by offering different search results and measuring how the implementation affected the conversion rate of the e-commerce platform. Furthermore, by making sure the set of customers is large enough and randomized, the relevancy of a different implementation could be compared against a baseline.

The significance of a search application to the conversion rate of the e-commerce platform is approximately 2.5x increase in conversion rate when customers interact with the search functionality on the website compared to customers who do not interact with the search while visiting the website [5]. On the other hand, the increase shows that there is a link between showing the search results shown to customers and the conversion rate of the e-commerce, on the other hand, the increase does not measure the relevancy of the search results. So, depending on the size of the e-commerce, the increase in conversion rate can lead to a significant amount of revenue for the organization when customers use their search. However, in this research, the goal is to increase the conversion rate of the proportion of customers who use the search when they visit the website. Therefore, how the proportion of customers who use the search can be increased is out of the scope of this research.

If the conversion rate changes based on the performed tests, then the relevancy of the search results affects the conversion rate of e-commerce. However, tests performed for a few weeks might not even capture the full effect of this research since more relevant search results might have a lasting effect on the reputation of the organization, and the increase of a reputation is hard to measure with these kinds of tests. Generally, the more satisfied customers are with their experience on a website, the more likely they are to return to revisit the website.

The current search application of the e-commerce platform uses Elasticsearch, which was the platform that the methods used in this research were implemented on. Elasticsearch is a distributed search and analytics engine which provides a real-time search and text analysis tools in addition to document storage and indexing capabilities. Turnbull and Berryman [4] introduced how to build a production scale search engine with Elasticsearch in their book. While building a large-scale search engine is out of the scope of this study, a selection of the text analyzing and search result ranking methods Turnbull and Berryman [4] proposed were utilized in this research. Therefore making it crucial to understand how the search engine performs in large scale search applications.

## REFERENCES

- [1] Alexa, A. *The top 500 sites on the web*. Available: <http://www.ctan.org/pkg/listings>. Checked 20.01.2020.
- [2] Brin, S. and Page, L. The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems* 30.1 (1998). Proceedings of the Seventh International World Wide Web Conference, 107–117. ISSN: 0169-7552. DOI: [https://doi.org/10.1016/S0169-7552\(98\)00110-X](https://doi.org/10.1016/S0169-7552(98)00110-X). URL: <http://www.sciencedirect.com/science/article/pii/S016975529800110X>.
- [3] Sorokina, D. and Cantu-Paz, E. Amazon Search: The Joy of Ranking Products. *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*. SIGIR '16. Pisa, Italy: Association for Computing Machinery, 2016, 459–460.
- [4] Turnbull, D. and Berryman, J. *Relevant Search: With applications for Solr and Elasticsearch*. USA: Manning Publications, 2016.
- [5] Power Internation As. *Internal Analytics of the company*. Checked 20.01.2020.