

**RESEARCH REPORT:**

The Power of Metadata:

*Structural Topic Modelling for Bibliometric Studies in Technology and Innovation Management (TIM)*

**by**

**Isak Daniel Ungerer**

**29129398**

A project report submitted in partial fulfilment of the requirements for the degree of

**Master in Information Technology**

(by course work)

in the

**FACULTY OF ENGINEERING, BUILT ENVIRONMENT AND INFORMATION TECHNOLOGY, SCHOOL OF INFORMATION TECHNOLOGY**

**UNIVERSITY OF PRETORIA**

**20 October 2021**

**PLAGIARISM FORM**

**DEGREE:** Masters of Information Technology

**DEPARTMENT:** Department of Computer Science

**SUPERVISOR:** Schalk Grobbelaar

**TITLE OF THE MINI-DISSERTATION:** The Power of Metadata: Structural Topic Modelling for Bibliometric Studies in Technology and Innovation Management (TIM)

**DECLARATION BY CANDIDATE:** I declare that this mini-dissertation, submitted by me, is my own work, that I have referenced all the sources that I have used and that no part was previously submitted at any tertiary institution.

Student Name and Signature: **Isak Daniel Ungerer**

Date: **20 October 2021**

**DECLARATION BY SUPERVISOR (AND CO-SUPERVISOR):** I

attach/submit the Turnitin report of the pre-final version and declare that I have analysed the report and is satisfied that it does not reveal severe instances of plagiarism. I declare herewith that I approve the submission of this mini-dissertation.

Supervisor Name and Signature: **Schalk Grobbelaar**

Date: **20 October 2021**

*Dedicated to my wife Liezl and son Lodewyk*

Abstract

Summarise the research report in no more than 500 words.

*Keywords*: Provide up to a maximum of 10 keywords for the study.

Technology and innovation management (TIM) is a rapidly growing research area. Qualitative research reviews in the TIM research area fail to capture the advancements in the field because it is too time-consuming, and reviews considers only a small sample of the research body population. Digital libraries of academic publications have enabled quantitative statistical analysis of large corpora of digital publications. These textual documents' topical bibliometric studies help identify emerging trends and map the importance of various topics in the TIM research area using natural language processing. However, most approaches only consider latent topics from the titles or abstract text of publications. These approaches fail to capture the impact of metadata on the topic classification task – like the author, journal name and the number of citations. This study proposes the use of structural topic modelling for topical bibliometric analysis. Metadata is used to refine the unsupervised topic classification in the TIM research area to identify emerging trends over time.

*Keywords:* technology and innovation management, bibliometrics, structural topic modelling, metadata, unsupervised machine learning

Acknowledgements

Schalk Grobbelaar

Vukosi Marivate

Table of Contents

[1. Background to the Research Study 1](#_Toc465320361)

[1.1. Language and General Requirements 1](#_Toc465320362)

[1.2. Paper Size and Layout 1](#_Toc465320363)

[1.3. Fonts, Point Sizes, Spacing and Equipment 1](#_Toc465320364)

[1.4. Bold, Italics, Underlines, and Justification 1](#_Toc465320365)

[1.5. Corrections 2](#_Toc465320366)

[1.6. Tables and Figures 2](#_Toc465320367)

[1.7. Equation Numbers, Symbols and Abbreviations 3](#_Toc465320368)

[1.8. Headings 3](#_Toc465320369)

[1.9. Word and Text Divisions 3](#_Toc465320370)

[1.10. References 3](#_Toc465320371)

[1.11. Order and Content 3](#_Toc465320372)

[1.12. Pagination 4](#_Toc465320373)

[2. Literature Survey 5](#_Toc465320374)

[2.1. Introduction 5](#_Toc465320375)

[2.2. Conclusion 5](#_Toc465320376)

[3. Conceptual Model 6](#_Toc465320377)

[3.1. Introduction 6](#_Toc465320378)

[3.2. Conclusion 6](#_Toc465320379)

[4. Research Design and Methodology 7](#_Toc465320380)

[4.1. Introduction 7](#_Toc465320381)

[4.2. Conclusion 7](#_Toc465320382)

[5. Results 8](#_Toc465320383)

[6. Discussion 9](#_Toc465320384)

[7. Conclusions and Recommendations 10](#_Toc465320385)

[References 11](#_Toc465320386)

[Appendix 12](#_Toc465320387)

List of Figures

[Figure 1.1: Clear, concise captions must go at the bottom of figures 2](#_Toc465320906)

List of Tables

[Table 1.1: Clear, concise captions must go at the top of tables 2](#_Toc465320959)

List of Acronyms/Definitions/Abbreviations

Provide a list as appropriate for the report.

Then make a section break to the first chapter and between subsequent chapters. Notice the changes in the headers and the page number format.

# Background to the Research Study

Section 5 of the Research Proposal template shows the typical sub-headings of the different chapters. Also refer to the Research Proposal template for formatting and document layout issues.

## Language and General Requirements

The report must be in good English. Students are encouraged to seek the assistance of a language editor for a thorough, objective revision of the written text. The final report copy must have a ready-for-publication appearance. It must have standardised features and be attractively reproduced. Introductory material, text, and appendices must all be clearly and consistently prepared and must all meet the specifications stated below.

## Paper Size and Layout

Only good quality A4-size (210mm x 297mm) white paper must be used. Manuscripts must be typed in only one column to the page. Use one side only. All text pages must be laid out in "portrait" orientation. As an exception pages containing only an exhibit (table or figure) may be in "landscape" orientation, if so required.

Top, bottom, and right margins must be 25mm. The left margin must be 30mm, to allow for binding.

## Fonts, Point Sizes, Spacing and Equipment

Text should be printed in Arial with point size 12 (used in this document). An exception to this is the Titles, which are in 14 point. 1.2 spacing should be used (as shown in this paragraph), except where conventional usage calls for single spacing, e.g. footnotes, indented quotations, tables, etc.

A letter quality printer (laser, impact, or ink jet) or an electric typewriter is required for printing the final manuscript.

## Bold, Italics, Underlines, and Justification

Headings must be bold, but not underlined. Text should be left and right justified against the margins (as shown in this document). Text must not be indented under headings.

## Corrections

No corrections with correction fluid or any other method are allowed on the manuscript. Use black ink whenever it is necessary to letter mathematical symbols or other copy by hand.

## Tables and Figures

Tables and figures are exhibits and numbered sequentially according to chapter numbers, i.e. 1.1, 1.2, 1.3, etc. They should be placed as close after their first mention as possible. The table number or figure number as well as heading of each exhibit must appear flush with the left-hand margin directly above (for Tables) or directly beneath (for Figures) the exhibit as in the following example:

###### Figures must go in front of the captions. The formatting must preferably be Arial, bold, 10pt, single line spacing

###### At the very least the text must be readable, and avoid colouring if possible – grey shading is acceptable

###### Figures should not be a repeat of the written text, i.e. they must provide information in a new way that is difficult to describe otherwise

Figure 1.1: Clear, concise captions must go at the bottom of figures

Table 1.1: Clear, concise captions must go at the top of tables

|  |  |
| --- | --- |
| **Headings** | Arial, 11 pt, bold if possible, centre aligned |
| **Body text** | Preferably Arial, 10pt, regular or bold and aligned as required |
| **Line spacing** | Single throughout, with 2pt spacing before and after the text |
| Tables must be easily readable and not repeat any of the written text, but must describe new information that would not fit well in the text | |

Although placement of exhibits directly in the text is preferred, it is also acceptable to place exhibits on separate pages, immediately following the page of first mention. Line drawings and photographs should be reduced to proper size. Ensure that all text, symbols, and information in exhibits are clearly legible in the final manuscript (see the Research Proposal template).

## Equation Numbers, Symbols and Abbreviations

When numbering equations, enclose numbers in parentheses and place flush with right-hand margin of the column. Equations themselves should be centred. See sample below. Use only standard symbols and abbreviations in text and exhibits.

 (12)

## Headings

The report must be divided into chapters, and further into paragraphs and subparagraphs (if so required). Do not use more than 3 levels of indenture if at all possible. Each chapter must start on a new page. A blank line should be used to separate headings from text above, and below them in the case of the main chapter heading. Blank lines must be used to separate paragraphs. New paragraphs are not indented, but start flush with the margin. In the event of a subparagraph that is numbered but does not require a heading, the text should start immediately next to the number and not on the next line.

## Word and Text Divisions

Words must be divided correctly at the end of a line and may not be divided from one page to the next. Use a standard dictionary to determine word division. If word-processing is used, it must divide words and text correctly: you may find it advisable to avoid all word division. Avoid short lines that end a paragraph at the top of a page, and any heading or subheading at the bottom of a page that is not followed by text.

## References

List all references by number at the end of the report in alphabetical order, as shown below. Text references should use the author's name, date of publication and page number. For example, "According to Smith (1991:8) ....." or "The use of influence diagrams (Howard and Matheson 1984:22) ......".

Howard, R.A. and Matheson, J.E., 1984. Influence Diagrams. In: Howard, R.A. and Matheson J.E. (eds.), The Principles and Applications of Decision Analysis. Strategic Decisions Group: Menlo Park, California.

Smith, J.P., 1991. The Rise of Engineering Management. Printers Press: Pretoria.

## Order and Content

This document shows the required order and content of the report.

## Pagination

Each page of the manuscript, including all blank pages, and pages with tables, figures, computer program printouts should be assigned a number. Consistent placement of pagination (preferably centre at the bottom of the page), at least 12mm from the paper's edge, should be used throughout the manuscript. The following pagination plan must be used (as is provided in this template):

* For the preliminary pages, use small Roman numerals (i, ii, iii, iv, etc.). The title page does not have a number but counts as page i; the following page is ii.
* For the remainder of the manuscript - continuous pagination for text, illustrations, references and appendices - use Arabic numbers (1, 2, 3, etc.)

# Literature Survey

## Introduction

Each chapter must begin with a short introduction, which tells the reader what is addressed in the chapter, and end with a conclusion of the main outcomes of the chapters and what must now follow. Thereby one establishes a ‘flow’ from one chapter to another.

## Conclusion

# Conceptual Model

## Introduction

## Conclusion

# Research Design and Methodology

## Introduction

## Conclusion

# Results

An introduction and conclusion may not be necessary for this chapter.

# Discussion

An introduction and conclusion may not be necessary for this and the final chapter.

# Conclusions and Recommendations

Clearly stipulate the contribution/s of this study. Also pave the way forward to improve certain aspects and/or for further research.

References

Refer to Appendix 11 of the Research Guide, as provided on ClickUP. You must use the Harvard Referencing method. DO NOT USE ANY OTHER STYLE OR SOFTWARE. A number of examples are given as follows:

De la Tour, A., Glachant, M. & Meniere, Y. 2013. Predicting the costs of photovoltaic solar modules in 2020 using experience curve models. *Energy,* 62**,** pp 341 - 348.

Department of Energy. 2011. Integrated Resource Plan for Electricity 2010-2030, Revision 2, Final Report, Department of Energy (Pretoria).

Department of Energy and Climate Change. 2012. Electricity Generation Costs, (Washington).

Dinkelman, T. 2011. The effects of rural electrification on employment: New evidence from South Africa. *The American Economic Review***,** pp 3078-3108.

Eberhard, A. 2013. Feed-In Tariffs or Auctions? Procuring Energy Supply in South Africa. Viewpoint, The World Bank Group (Washington).

Feldman, D., Barbose, G., Margolis, R., Wiser, R., Darghouth, N. & Goodrich, A. 2012. Photovoltaic (PV) Pricing Trends: Historical, Recent, and Near-Term Projections.

Appendix