Research Methods INFO5993

Writing the literature review for the research proposal & research article

The Learning Centre
Semester 2 2014
Dr. Bronwyn James
Aims for today

• Identify the purpose and structure of a literature review for:
  – a journal article
  – a research proposal

• Develop analytical and critical ways of thinking about and writing your literature review
Writing the Literature Review

The research proposal

The literature review

The research article
A research process

The following process may not be applicable to all research projects but it is applicable to most:

1. **Choose a problem/question**
2. **Review the published literature**
3. Define the research problem/question
4. Develop a hypothesis [or objectives]
5. Choose a method
6. Carry out the research
7. Analyse your data
8. Write up your results AND conclusions
9. Identify significance, limitations & future research

c/- A.Prof. Peter Hyland University of Wollongong
What do you already know?

• What is the purpose of a literature review?
Google Scholar’s Ranking Algorithm: The Impact of Citation Counts (An Empirical Study)

Jivan Baud
Otto-von-Guericke University
Department of Computer Science
Vahrenwalder Landstrasse 10
Maggdeburg, Germany
j.baud@ovgu.de

Bola Gips
Otto-von-Guericke University
Department of Computer Science
Vahrenwalder Landstrasse 10
Maggdeburg, Germany
b.gips@ovgu.de

ABSTRACT
Google Scholar is one of the major academic search engines, and it ranks research articles based on the number of citations. This study investigates the impact of citation counts on Google Scholar’s ranking algorithm. The study found that the higher the citation count of an article, the higher its ranking on Google Scholar. This finding is important for researchers as it suggests that obtaining a higher number of citations can lead to increased visibility and recognition of their work.

2. RELATED WORK

In the field of academic search engines, the ranking of results is crucial. Various factors influence the ranking of papers, including citation count, relevance, and quality. In this study, we focus on the impact of citation counts. Previous research has shown that citation count is a significant factor in determining the ranking of articles. However, the exact mechanism through which citation count affects ranking is not well understood. This study aims to contribute to the understanding of this mechanism.

1. INTRODUCTION

Today’s academic search engines play a crucial role in the dissemination of research. The accuracy of these search engines is essential for researchers to find relevant and high-quality papers. Google Scholar is one of the most popular academic search engines, and its ranking algorithm is based on various factors. In this study, we focus on the impact of citation counts on the ranking algorithm.

For more information, contact the authors via email at baud@ovgu.de and gips@ovgu.de.

Keywords: Ranking algorithms, Google Scholar, Citation counts.

1. INTRODUCTION

The increasing importance of academic search engines has led to the development of various ranking algorithms. Google Scholar is one of the most widely used academic search engines, and its ranking algorithm is based on multiple factors, including citation counts. The impact of citation counts on Google Scholar’s ranking algorithm is a critical aspect that requires further investigation.

2. RELATED WORK

Previous research has shown that citation count is a significant factor in determining the ranking of articles. However, the exact mechanism through which citation count affects ranking is not well understood. This study aims to contribute to the understanding of this mechanism.

1. INTRODUCTION

Today’s academic search engines play a crucial role in the dissemination of research. The accuracy of these search engines is essential for researchers to find relevant and high-quality papers. Google Scholar is one of the most popular academic search engines, and its ranking algorithm is based on various factors. In this study, we focus on the impact of citation counts on the ranking algorithm.

For more information, contact the authors via email at baud@ovgu.de and gips@ovgu.de.

Keywords: Ranking algorithms, Google Scholar, Citation counts.

1. INTRODUCTION

The increasing importance of academic search engines has led to the development of various ranking algorithms. Google Scholar is one of the most widely used academic search engines, and its ranking algorithm is based on multiple factors, including citation counts. The impact of citation counts on Google Scholar’s ranking algorithm is a critical aspect that requires further investigation.

2. RELATED WORK

Previous research has shown that citation count is a significant factor in determining the ranking of articles. However, the exact mechanism through which citation count affects ranking is not well understood. This study aims to contribute to the understanding of this mechanism.

1. INTRODUCTION

Today’s academic search engines play a crucial role in the dissemination of research. The accuracy of these search engines is essential for researchers to find relevant and high-quality papers. Google Scholar is one of the most popular academic search engines, and its ranking algorithm is based on various factors. In this study, we focus on the impact of citation counts on the ranking algorithm.

For more information, contact the authors via email at baud@ovgu.de and gips@ovgu.de.
The purpose of a literature review

• summarise and evaluate past research
• show similarities and differences in previous research
• give an overview of controversies in past research
• place one's own research into context
• show a gap in research
• justify one’s own research
• generate new research hypotheses
**Reading strategies for a journal article**

<table>
<thead>
<tr>
<th>Do this</th>
<th>Why?</th>
<th>Possible ways to write about this</th>
</tr>
</thead>
<tbody>
<tr>
<td>Read the title of the article or chapter?</td>
<td>Provides the general focus</td>
<td><em>The focus of this article is...</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>In this article, the authors focus on ...</em></td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>This article is about ...</em></td>
</tr>
<tr>
<td>Read the abstract (if there is one)</td>
<td>Provides an overview of the whole paper (aims, methods, results)</td>
<td></td>
</tr>
<tr>
<td>Read the introduction (usually the first section of the article)</td>
<td>Provides the aims of the paper and more information about the focus</td>
<td><em>The aim of this paper is ...</em></td>
</tr>
<tr>
<td>Read the main headings</td>
<td>Identifies the main issues</td>
<td><em>The main areas covered in this paper are ...</em></td>
</tr>
<tr>
<td>Each topic sentence</td>
<td>Provides details about the main issues</td>
<td></td>
</tr>
<tr>
<td>The conclusion</td>
<td>Identifies the results - recommendations - implications</td>
<td></td>
</tr>
</tbody>
</table>
What is a critical and analytical literature review?

• Focuses on a question or issue or problem

• Organises the literature according to:
  
  themes or
  issues or
  chronology…
  that relate/s to the question, issue, problem

3. Evaluates the literature in relation to question or issue or problem

• Your topic is Social networks but what is your focus: question, issue, problem?

• How is the literature organised in the articles you are reading for your assessment and for your proposal?

• Is the literature evaluated and if so, how?
Defining features

- **Descriptive**: Concrete terms such as names/dates; reference to journal articles.
- **Analytical**: Abstract terms, language to show relationships, clear structure of ideas.
- **Argumentative**: Reference to claims, structure of claims with reason and evidence.
- **Critical**: Reference to the work of others, combined with evaluative language.
## Comparing and contrasting

<table>
<thead>
<tr>
<th>Reading 1</th>
<th>Reading 2</th>
<th>Reading 3</th>
<th>Reading 4</th>
<th>Reading 5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Problem</strong></td>
<td><strong>Problem</strong></td>
<td><strong>Problem</strong></td>
<td><strong>Problem</strong></td>
<td><strong>Problem</strong></td>
</tr>
<tr>
<td>Analysis and tracking of faces and gestures in Real-time Systems</td>
<td>Optimum colour spaces for skin tone detection</td>
<td>A survey of pixel-based skin colour detection techniques</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Method</strong></td>
<td><strong>Method</strong></td>
<td><strong>Method</strong></td>
<td><strong>Method</strong></td>
<td><strong>Method</strong></td>
</tr>
<tr>
<td>Designed an optimum detector using the Neyman-Pearson test on conditional probabilities derived from normalises histogram of a database of images</td>
<td>Identified 3 elementary problems: choice of colour space, modelling of skin colour distribution, developing efficient processing algorithm. Survey first two.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Results</strong></td>
<td><strong>Results</strong></td>
<td><strong>Results</strong></td>
<td><strong>Results</strong></td>
<td><strong>Results</strong></td>
</tr>
<tr>
<td>Bayesian method resulted in ML and MAP depending on assumptions made. ML outperformed MAP technique and the lookup table method</td>
<td>Colour space does not influence performance as long as the optimum skin detector for that space is used</td>
<td>Categorised skin colour modelling techniques Best performance achieved by Bayesian Skin Probability Map method followed by maximum entropy model</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Overall focus</strong></td>
<td><strong>Overall focus</strong></td>
<td><strong>Overall focus</strong></td>
<td><strong>Overall focus</strong></td>
<td><strong>Overall focus</strong></td>
</tr>
<tr>
<td>Skin colour modelling techniques (non parametric)</td>
<td>Colour space</td>
<td>Skin colour and colour space</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Questionnaires and face-to-face interviews are methods of data collection in the social sciences. Questionnaires can be used to sample a large number of people over a wide geographical area. They can normally be conducted by one person at low costs. However, questionnaires do not allow the researcher to ask for extra information.

Face to face interviews, by contrast, are relatively small scale. They are labour intensive and time-consuming and costly. But interviews have the advantage of allowing the researcher to clarify questions if necessary. Also, during interviews the researcher can ask additional questions. However, questionnaires do not allow the researcher to ask for extra information.
Questionnaires and face-to-face interviews are methods of data collection in the social sciences. Whereas the sample size in face-to-face interviews is normally small, the questionnaire can be sent out to large numbers of people over a wide geographical area. Costs, both in labour and money, are high with face-to-face interviews, while questionnaires can be distributed quickly by just one person at a relatively low cost. However, in terms of quality of information, interviews are often superior to questionnaires as the researcher has the opportunity to ask additional questions and explain what questions mean if this becomes necessary. This is not possible with questionnaires.
Questionnaires and face-to-face interviews are methods of data collection in the social sciences. There are a number of advantages in using questionnaires, including lower costs and sample size. However, the quality of information obtained by interviews is superior to that obtained by questionnaires because the researcher has the opportunity to ask additional questions and explain what questions mean if this becomes necessary. Therefore the importance of quality in the research process can necessitate the use of interviews for data collection.
Two of the main choices available for social science data collection have long been assumed to be questionnaires sent by mail and the face-to-face interview. While both of these methods do have advantages, both have limitations which bring into question their ultimate usefulness as research tools. Although the quality of information obtained by interviews is undoubtedly far superior to that obtained by questionnaires, the cost and time taken to obtain sufficient information on individual interviews is, in the current research climate, no longer tenable. In recent years a number of alternate methods have been used successfully in research. An often neglected method of collection, for example is the focus group, which is far more cost and time efficient compared to individual interviews but enables a much greater depth of information than questionnaires. Therefore such methods need to be given more prominence in social science research methodology.
Draft D

Two of the main choices available for social science data collection have long been assumed to be questionnaires sent by mail and the face-to-face interview. While both of these methods do have advantages, both have limitations which bring into question their ultimate usefulness as research tools. Although the quality of information obtained by interviews is undoubtedly far superior to that obtained by questionnaires, the cost and time taken to obtain sufficient information on individual interviews is, in the current research climate, no longer tenable. In recent years a number of alternate methods have been used successfully in research. An often neglected method of collection, for example is the focus group, which is far more cost and time efficient compared to individual interviews but enables a much greater depth of information than questionnaires. Therefore such methods need to be given more prominence in social science research methodology.
Draft D

Two of the main choices available for social science data collection have long been assumed to be questionnaires sent by mail and the face-to-face interview. While both of these methods do have advantages, both have limitations which bring into question their ultimate usefulness as research tools. Although the quality of information obtained by interviews is undoubtedly far superior to that obtained by questionnaires, the cost and time taken to obtain sufficient information on individual interviews is, in the current research climate, no longer tenable. In recent years a number of alternate methods have been used successfully in research. An often neglected method of collection, for example is the focus group, which is far more cost and time efficient compared to individual interviews but enables a much greater depth of information than questionnaires. Therefore such methods need to be given more prominence in social science research methodology.
Draft D

Two of the main choices available for social science data collection have long been assumed to be questionnaires sent by mail and the face-to-face interview. While both of these methods do have advantages, both have limitations which bring into question their ultimate usefulness as research tools. Although the quality of information obtained by interviews is undoubtedly far superior to that obtained by questionnaires, the cost and time taken to obtain sufficient information on individual interviews is, in the current research climate, no longer tenable. In recent years a number of alternate methods have been used successfully in research. An often neglected method of collection, for example is the focus group, which is far more cost and time efficient compared to individual interviews but enables a much greater depth of information than questionnaires. Therefore such methods need to be given more prominence in social science research methodology.
Two of the main choices available for social science data collection have long been assumed to be questionnaires sent by mail and the face-to-face interview. While both of these methods do have advantages, both have limitations which bring into question their ultimate usefulness as research tools (Smith & Jones 2009). Although the quality of information obtained by interviews is undoubtedly far superior to that obtained by questionnaires, the cost and time taken to obtain sufficient information on individual interviews is, in the current research climate, no longer tenable. In recent years a number of alternate methods have been used successfully in research. Kaufman (2003) used an often neglected method of collection, the focus group, which is far more cost and time efficient compared to individual interviews but enables a much greater depth of information than questionnaires. Therefore such methods need to be given more prominence in social science research methodology.
Writing a literature review within a Research Proposal

• The proposal purpose
• Key proposal components and their structure
• Types of writing in the proposal
• Useful grammatical resources
A research proposal

Definition

• It is a document that presents a case for an idea and the action planned with respect to it (Krathwohl, 1988: 12)

Purpose

• To justify what you plan to do, in order to gain approval for it (Thornquist, 1989:3)
What structure is required for your proposal?

Flow of **argument** in the proposal and links to following sections

Title

Background

Literature review

Aims + research questions

Significance of the study

Proposed methodology

Other sections? (timeline etc.)

Reference list
Requirements of the proposal - also relevant for the research article

Thesis Title
- Indicate topic and provide essential keywords for retrieval

Introduction:
- Nature of the problem
- Why the problem is important
- How your study will contribute to the solution of the problem

Research questions / hypotheses
- This section should also include any discussion of terms (definitions), and any secondary hypotheses, aims or questions
Why draft your title?

• Identifies key words or phrases which position the work

• Assists you to group papers in the literature review

• Helps create links cohesive links with the next section

• Changes and evolves with thesis
Types of titles

Noun string
Quantitative **evaluation** of passage retrieval algorithms for question answering

Question
What have Innsbruck and Leipzig in common? Extracting semantics from wiki content.

Colon
Overcoming the brittleness bottleneck using Wikipedia: Enhancing text categorization with encyclopedic knowledge.
Write your title

Noun string

Question

Colon
# The functions of the Introduction

<table>
<thead>
<tr>
<th>Question</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1 What is the study about?</td>
<td>Orient the reader with general background of the field</td>
</tr>
<tr>
<td>Q2 Why the research is significant?</td>
<td>Justifies your research area- anticipates the worth of outcomes</td>
</tr>
<tr>
<td>Q3 What is some of the most relevant work done in this field?</td>
<td>Summarizes progress or issues in the field</td>
</tr>
<tr>
<td>Q4 What further research is needed in this area?</td>
<td>Provides the reader with weaknesses in current research or omissions in our knowledge</td>
</tr>
<tr>
<td>Q5 What is the purpose of your study?</td>
<td>Addresses the issue, problem, or question</td>
</tr>
</tbody>
</table>
5 stages of an Introduction (Weissberg & Buker 1990)

1. Orientation to the topic
2. Introduce key authors & issue
3. Gap
4. Aim / RQ
5. Scope/methods
Orientations to the topic

- The need to computationally identify certain classes of entities referred to in texts has become a core component of many Natural Language Processing tasks.

- There is a great need for fast, reliable Question Answering (QA) systems to allow users to make full use of the vast amounts of textual data available to them on the internet.

1. X is an important component in the Y, and plays a key role in Z
2. A common observation of X in daily life is Y
3. X is one of the most widely used groups of algorithms for Y
Introductions to key authors

• A more thorough analysis of possible future directions of QA research can be found in the roadmap paper by Burger et al. [2001].

• The task gained specification in the final two DARPA-funded Message Understanding Conferences (MUC-6, MUC-7; see Chinchor (1998)),

1. Field observations (multiple citations) have provided evidence for the significance of X

2. Early investigations of P in Q have been framed in R with S (multiple citations)

3. A recent development in X have heightened the need for Y (single citation)
3 Gap statements

1. Most studies in X have only been carried out in a limited number of areas.

2. Several studies have produced estimates of X (Smith, 2002; Jones, 2003), but unfortunately there is still insufficient data to Y

3. However this analysis does not take account X, nor has Y been examined in sufficient detail.
Aim / Focus /objective

• We therefore see the automatic generation of a high-accuracy corpus of training data for NER as a feasible and necessary task, given the right resource: Wikipedia.

1. This research will focus on the relationship between X and Y
2. The objectives of this research are to determine whether...
3. More specifically this thesis aims to advance understanding in the following areas 1... 2... 3... 4...
4. The purpose of this research is to develop further understanding of X
Scope

• This project will adopt three linked approaches of scaling analysis, numerical simulation and stability analysis.

• Achieving these objectives will require investigations of the following areas: A, B, C.

• For the purpose of this thesis time constraints will not allow the study of long-term effects such as creep on the serviceability or long term deformations of the slab.
On-line academic phrase bank

http://www.phrasebank.manchester.ac.uk/

Introducing Work
Referring to the Literature
Being Critical
Describing Methods
Reporting Results
Discussing Findings
Writing Conclusions
Your turn

• The topic is computer crime

• Take these notes and decide on 2-3 groupings of the information (your analysis)

• Give each group an abstract label
Sources of information

Peer reviewed
High index & low index journals
Conference Proceedings
Established authors & new author
Web based information
Personal communication
Use an analysis grid to organize the LR

<table>
<thead>
<tr>
<th>Draft</th>
</tr>
</thead>
<tbody>
<tr>
<td>Papers Core</td>
</tr>
<tr>
<td>Directly related</td>
</tr>
<tr>
<td>Indirectly related</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Topic</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Draft Title
Building a Wikipedia-derived training corpus for named entity recognition (NER)

| Topic 1 | Topic 2 | Topic ...
|---------|---------|---------|
| **Topic:** Wikipedia and named entities | **Topic:** Data sources for NER | **Topic:** ??

<table>
<thead>
<tr>
<th>Papers</th>
<th>Core</th>
<th>Directly related</th>
<th>Indirectly related</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extraction and classification</td>
<td>The usefulness of gazetteers for NER</td>
<td>Entity disambiguation</td>
<td></td>
</tr>
<tr>
<td>Toral and Munoz (2006) use approaches that avoid machine learning.</td>
<td>It has often been assumed that a lack of large entity gazetteers across many languages severely impedes recall (Cucchiarelli et al., 1998).</td>
<td>For the purpose of limiting disambiguation, Bunescu and Pasca (2006) used three heuristics based on capitalisation</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Unsupervised approaches ??</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Other Wikipedia entity approaches</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>A different approach to producing more structured data from Wikipedia is given by Suchanek et al. (2007).</td>
<td>??</td>
</tr>
</tbody>
</table>
Create critical comments with an analysis grid

<table>
<thead>
<tr>
<th>Source / Year</th>
<th>Aim of the topic</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Source / Year</td>
<td>Aim of study topic</td>
<td>Comment</td>
</tr>
<tr>
<td>------------------------</td>
<td>-----------------------------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Toral and Munoz (2006)</td>
<td>Don't use machine learning</td>
<td>Words in an article's opening sentence often from an entity class-related synset so they can identify many articles</td>
</tr>
<tr>
<td>Suchanek et al. (2007)</td>
<td>They use Wikipedia's category system,</td>
<td>They use its head noun as a key to associate a WordNet synset with the article</td>
</tr>
<tr>
<td>Bhole et al. (2007)</td>
<td>After several attempts used a standard bag-of-words classification method</td>
<td>Was moderately successful</td>
</tr>
<tr>
<td>Watanabe et al. (2007)</td>
<td>Use more Wikipedia's structural features</td>
<td>Explore entities together in a list are commonly of the same category</td>
</tr>
</tbody>
</table>
A few approaches avoid machine learning: both Toral and Muñoz (2006) and Suchanek et al. (2007) do so by utilising WordNet. Toral and Muñoz (2006) suggest that words in an article’s opening sentence will often come from an entity class-related synset, and so are able to identify a large portion of articles about locations and people, but organisations to a much lesser extent. Suchanek et al. (2007) instead rely on Wikipedia’s category system, which they divide into conceptual, thematic, relational and administrative categories. After heuristically identifying the conceptual category for an article, they use its head noun as a key to associate a WordNet synset with the article. Some initial attempts by Bhole et al. (2007) also showed limited success in using a heuristic approach alone. They first attempted the low-recall method of extracting entities from Wikipedia’s list pages (e.g. “List of biologists”), and eventually used more rule-based heuristics testing the presence of certain infobox templates, of dates of birth, or of geographical coordinates. This again gave very poor results for the ORG category, and only 48% for LOC, although at almost-perfect precision for all classes.
Information transfer

Information Prominent

Research topic concept is the subject

Author Prominent

Author is subject

Reporting verb + information structure
Reporting verbs and argument

Neutral
• present, list, report, state, describe, summarise,

Interpretive
• show, note, proposed, determine, assert, posit, challenge, doubt, question, argue, recommend, illustrated
REFERENCES


