بِسْمِ اللَّهِ الرَّحْمَٰنِ الرَّحِيمِ
How to Read a Research Paper? / Session II
Searching a Paper

Dr. Rao Muhammad Adeel Nawab
زندگی کے سفر میں کچھ کچھ مہنت اور ڈگری پانے پار باقی بھی - قسمت اور دعا بھی جیت جاتی بھی -

روزانہ پیر دعا نال اگئیں:

آہِدِنا الصرَّطِ الْمُسْتَقِيمِ صِرَّطِ الَّذِينَ أَنْعَمْتَ عَلَيهِمْ

بہمنِ سیکھی راه یہا ان لوگوں کی راه جن پر تم لوںے اندام کیا

Dr. Rao Muhammad Adeel Nawab
رسُولُ اللَّهِ صلى الله عليه وسلم نَـے فَرَمَائْا
إنّمَا الأَعْمَالُ بِالنِّيَّةِ
(اعمال کا دارودار نیئت پر ہے۔)

یہ مُہِنَّا اور پُہِنَّا بِامَارِی عبادت چِیتْ-
اللّٰهُ کی رضا اور مَخلوق خدا کی خُرمت کی نیت سے (اللّٰهُ کے دَهِیان کے
سَاتِح) پُہِنَّا اور پُہِنَّا۔-

Dr. Rao Muhammad Adeel Nawab
Dua - Take Help from Allah before starting any Task

اللّهُمَّ خُذِّ لِي وَاخْتَرِ لِي
سُبْحَنَّكَ لَا عَلِمُ لَنَا إِلَّا مَا
عَلِّمَتُنَا إِنَّكَ أَنتَ الْعَلِيمُ الْحَكِيمُ

اللّهُمَّ صَلِّ عَلَيْهِ مُحَمَّدٍ وَآlcِه وَعَتْرَتِه
بَعْدُ ثُلُّثِ كُلِّ مَعْلُومٍ لَكَ

يا فَوِي

Rābʾu ʾAșrāḥ ʾLī ʾIṣlāmī ٦٥ وَيُسَبِّلِ لِي أَمْرَي
وَأَحْلَلُ عَقْدَةَ مِنْ أَسَانِي ٨٧ يَفْقِهُو أَقْوَلُي

Dr. Rao Muhammad Adeel Nawab
Create a profile on Google Scholar

Use Google Scholar to search for articles, theses, books, patents and court reports related to your research area

Use “Alerts” and “Follow” options in Google Scholar to get email alerts on most recent papers published in your research area

Dr. Rao Muhammad Adeel Nawab
For any Task apply - Four Steps Process (Iterative)

1. **PLAN** - in mind
2. **Design** - on paper
3. **Execute** - in real world
4. **Feedback**
   - From domain experts
   - (or audience)

Dr. Rao Muhammad Adeel Nawab
PhD is Game of Patience and Consistency

Dr. Rao Muhammad Adeel Nawab
Balanced Life is Ideal Life

Get **Excellence** in five things

- A Journey from **BEGINNER** to **EXCELLENCE**
- You must have a combination of five things with different variations. However, aggregate will be same.

Dr. Rao Muhammad Adeel Nawab
To Achieve Excellence - Have a Learning and Progressive Attitude

Dr. Rao Muhammad Adeel Nawab
YOU VS MAHOL

Dr. Rao Muhammad Adeel Nawab
Outline

- Answer 4 questions & get a PhD
- Identify sources of research papers
- Searching and saving relevant research papers

Dr. Rao Muhammad Adeel Nawab
Answer 4 Questions & Get a PhD
Answer 4 Questions & get a PhD

1. What is the problem? Research Problem
2. What has been done? Literature Review
3. What you want to do? Proposed Work
4. How it is different? Novelty / Contribution
Research Problem - What is the problem?

- Both supervisor and student must be interested to work on the research topic.
- Identify your research problem by starting with an inverted triangle.
Example - **Inverted Triangle**

- All research areas
- Natural Language Processing
- Cross Lingual Text Reuse (English-Urdu)
- Text Reuse
- Develop a Cross-lingual Text Reuse Corpus for English Urdu Language Pair

Dr. Rao Muhammad Adeel Nawab
Research Focus

Cross-lingual Text Reuse Detection for English-Urdu Language Pair
Two Main **Research Focus**

1. **New technique / algorithm**
2. **New dataset / corpus**
Template for Technique

- In which category this technique falls?
- How it works?
- What are the strengths of this technique?
- What are the Weaknesses of this technique?
- In which previous studies it has been used?
Template for **Corpus / Dataset**

- In which category this corpus falls?
- Purpose of Corpus
- Corpus Generation Process
- Corpus Characteristics
- In which previous studies it has been used?
Evaluation Measures

- **Definition**
- **Equation (if any)**
- **In which previous studies it has been used?**

Dr. Rao Muhammad Adeel Nawab
### Literature Review - What has been done?

**Literature review can be**

<table>
<thead>
<tr>
<th>History based</th>
<th>Approaches based</th>
</tr>
</thead>
<tbody>
<tr>
<td>1950-1960</td>
<td>Supervised approaches</td>
</tr>
<tr>
<td>1961-1980</td>
<td>Un-supervised approaches</td>
</tr>
<tr>
<td>1980-Date</td>
<td>Semi-supervised approaches</td>
</tr>
</tbody>
</table>

**TIP:** See the outline of Literature Review Chapter from a PhD thesis (preferably latest) related to your research topic

Dr. Rao Muhammad Adeel Nawab
Study State-of-the-art Existing Literature

List Down

- Weaknesses / Limitations of existing approach(es)
- Strengths of existing approach(es)
- Possible future research work directions
Proposed Work - What you Want to do?

Proposed approach:

1. Summarize your literature review
2. List down the limitations of the existing work
3. How would you overcome the limitation(s) of existing approach or (baseline approach)?
4. How will you improve the limitations of existing work (proposed solution)?
Novelty/Contribution - How it is different?

How your proposed solutions is different from existing work?

Your proposed work must be justified/proved theoretically.
Identify sources of research papers
Make a list of top

- Conferences
- Shared Tasks
- Scientific and Professional Societies
- Journals
- Publishers
Calendar of Upcoming Events

Useful to keep a calendar of upcoming conferences and shared tasks in your research area

Dr. Rao Muhammad Adeel Nawab
Two Main Options

1. Search on Google Scholar

2. Discuss with a senior/supervisor
A Possible Google Query Template

[Conferences / journals / shared tasks / scientific societies / publishers] [Research area] [Year (optional)]

Examples:

- Conferences Natural Language Processing
- Journals Natural Language Processing
- Shared tasks Natural language processing 2019
- Scientific societies Natural Language Processing
- Publishers Natural Language Processing

Dr. Rao Muhammad Adeel Nawab
Conference

An academic **conference** is a gathering of scientists or academicians, where research findings are presented or a **workshop** is conducted.
H-Index

H5-index:

- H5-index is the h-index for articles published in the last 5 complete years.
- It is the largest number h such that h articles published in 2013-2018 have at least h citations each.
- It makes the reasoning system more efficient.

H5-median

- For a publication is the median number of citations for the articles that make up its h5-index.
# Conferences (NLP)

<table>
<thead>
<tr>
<th>Conferences (NLP)</th>
<th>H5-Index</th>
<th>H5-Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meeting of the Association for Computational Linguistics (ACL))</td>
<td>87</td>
<td>130</td>
</tr>
<tr>
<td>Conference on Empirical Methods in Natural Language Processing (EMNLP))</td>
<td>76</td>
<td>132</td>
</tr>
<tr>
<td>The North American Chapter of the Association for Computational Linguistics: Human Language Technologies (HLT-NAACL)</td>
<td>56</td>
<td>99</td>
</tr>
</tbody>
</table>

Dr. Rao Muhammad Adeel Nawab
A journal is a scholarly publication containing articles written by researchers, professors and other experts.

- Journals focus on a specific discipline or field of study

**International Standard Serial Number:**

- It is an eight-digit serial number used to uniquely identify a serial publication
- The ISSN is especially helpful in distinguishing between serials with the same title.
The impact factor (IF) is a measure of the frequency with which the average article in a journal has been cited in a particular year.

\[ A = \text{the number of times articles published in 2008 and 2009 were cited by indexed journals during 2010.} \]

\[ B = \text{the total number of "citable items" published in 2008 and 2009.} \]

\[ \frac{A}{B} = 2010 \text{ impact factor} \]

It is used to measure the importance or rank of a journal by calculating the times its articles are cited.

Dr. Rao Muhammad Adeel Nawab
## Journals (NLP)

<table>
<thead>
<tr>
<th>Journal (NLP)</th>
<th>Impact Factor</th>
<th>ISSN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computational Linguistics</td>
<td>2.333</td>
<td>ISSN - 2180-1266</td>
</tr>
<tr>
<td>International Journal of Computational Linguistics and Applications</td>
<td>0.961</td>
<td>ISSN 0976-0962</td>
</tr>
<tr>
<td>Journal of the Association for Information Science and Technology (JASIST)</td>
<td>2.85</td>
<td>ISSN:2330-1643</td>
</tr>
</tbody>
</table>
Shared Tasks

Shared tasks are organized to tackle specific problems which are challenging to tackle by spare research groups for several reasons.

- It's the availability of data.
- One key element of a successful challenge is the goal, it has to be challenging, specific and possibly lacking data.
- To the point where plenty of advances are made because of their organization.
- Why Natural Language Processing (NLP) is one of the fields that has a very long tradition of challenges.
## Shared Tasks

<table>
<thead>
<tr>
<th>Domain</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PAN</strong></td>
<td>PAN is a series of scientific events and shared tasks on digital text forensics and stylometry</td>
</tr>
<tr>
<td><a href="https://pan.webis.de">https://pan.webis.de</a></td>
<td></td>
</tr>
<tr>
<td><strong>SemiEval</strong></td>
<td>International Workshop on Semantic Evaluation</td>
</tr>
<tr>
<td><a href="http://alt.qcri.org/semeval2019/">http://alt.qcri.org/semeval2019/</a></td>
<td></td>
</tr>
<tr>
<td><strong>CoNLL</strong></td>
<td>Multilingual Parsing from Raw Text to Universal Dependencies</td>
</tr>
<tr>
<td><a href="http://www.conll.org">http://www.conll.org</a></td>
<td></td>
</tr>
</tbody>
</table>
Searching and Saving Relevant Research Papers
Searching Related Research Papers

Identify Keywords

- See your research focus statement for keywords

Example

Cross-lingual text reuse detection for English-Urdu language pair

Keywords = Cross-lingual text reuse
Searching Related Research Papers (Cont.)

Search on Google Scholar using Keywords

Google Search

- See “Related Articles”
- See “Cited by”
- See “Recommended Articles” (if you have Google Scholar Profile)
Searching Related Research Papers (Cont.)

Use “Create Alert” option in Google Scholar

- You'll get emails whenever Google Scholar finds matching search results

Alert query: "cross lingual" text reuse detection
Email: adeelnawab@cuilahore.edu.pk
Number of results: Show up to 10 results:
Update results: CREATE ALERT

Sample results since 2019:

Supervised heterogeneous feature transfer via random forests
S Sukhija, NC Krishnan - Artificial Intelligence, 2019
... We also discuss the results of the experiments on two real world cross-lingual text analysis tasks ... Furthermore, we also investigate whether re-training in source/target...
Go to the Google Scholar Profile(s) of author(s) who are working on research fields closely related to your research topic.

Click on Follow to get email alerts about their latest publications.

Dr. Rao Muhammad Adeel Nawab
See “References” of paper(s) you have searched
Organize Collected Data - A Possible Organization

- Literature Review - Folder
  - Papers - Sub-folder
    - Conference - Sub-sub-folder
    - Journal - Sub-sub-folder
  - Thesis - Sub-folder
    - MPhil - Sub-sub-folder
    - PhD - Sub-sub-folder
  - Tutorials - Sub-folder
  - Other - Sub-folder

Dr. Rao Muhammad Adeel Nawab
Possible **Naming Convention**

[Year] - [ conference/journal/workshop/thesis ] - [paper title)]

**Examples:**

- 2012 - ECIR - Retrieving Candidate Plagiarized Documents Using Query Expansion
Take Backup on Daily Basis - Very Important

- An external hard disk
- At least use two sources to take backups
- An online storage source
  - Dropbox / Google Drive / One Drive
Jazak Allah Khair 😊

Dr. Rao Muhammad Adeel Nawab
Email: adeelnawab@cuilahore.edu.pk