MENTORING FOR “GRADUATE-ON-TIME (GOT)”

Professor Dr. Naomie Salim
Faculty of Computing
Universiti Teknologi Malaysia
Outline of Talk

• The challenges to GOT
• Mentoring for GOT throughout the PhD research life-cycle
  – Preparation phase
  – Problem definition phase
  – Literature review phase
  – Methodology design phase
  – Experiment and data collection phase
  – Analysis of results phase
  – Writing-up phase
PhD - Training & Developing Competent Researchers

Domain A
- Knowledge base (A1)
- Cognitive abilities (A2)
- Creativity (A3)

Domain B
- Personal effectiveness
- Professional and career development (B3)
- Self-management (B2)

Domain C
- Research governance and organisation
- Research management (C2)
- Finance, funding and resources (C3)

Domain D
- Engagement, influence and impact
- Working with others (D1)
- Communication and dissemination (D2)

Engagement, influence and impact
The knowledge and skills to work with others and ensure the wider impact of research.
Problem Statement

- High percentage of non-completion
- Long process towards completion
- Candidates regard PhD period as lonely and stressful episode, at high personal cost, and without ‘mentor’ to discuss problems;
- Meetings have inadequate frequency and depth, often no regular performance and progress interviews;
- Completed PhDs have low scientific and social impact, wasteful research time and money.
Sources of Problems

- Poor research design, no focus, no adequate research question;
- Lack of realistic expectations (‘targeting the sky’);
- Inadequate research background; lack of training in methodological and writing skills (inadequate Bachelors and Masters training);
- Problematic research facilities (time, office, computer, assistance, money, flexibility in rules, under-funding of essential tasks);
- Many competing tasks (teaching, consultancies, family life);
- Bad planning, bad phasing, bad time management;
Sources of Problems (cont.)

• Major problems with writing academic English;
• Negligent or inadequate supervision; often unclear, strained relationship;
• Inadequate networking: not aware of others working in the same field of studies, no contact with peers, parochial local research cultures;
• Breakdown of motivation, psychological stress due to isolation, feelings of uselessness;
• Lack of possibilities to participate in a vibrant research culture.
Different Student Needs for Supervision

• Immature, unconfident student - supervisor as “big daddy.”
  – student may have requisite tools and skills to do research, but ability to work independently not well developed.

• Somewhat mature, somewhat confident student – supervisor as “mentoring” colleague
  – Preparatory work helped student be somewhat confident, but the student still needs moderately detailed direction to get going

• Very mature, confident student - supervisor as a “senior” colleague
  – Student recognizes need for guidance and supervision, but the need is at a fairly general level.
Supervision Styles-
Strong Master/apprentice style

• Supervisor is the master, student works as an apprentice on problems selected by master

• Assuming a competent master, the advantages:
  – **significant reduction in the time** to formulate a problem,
  – strong guidance and direction in doing the work,
  – development of **specific skills for the type of problem** with the master.

• Severe disadvantages:
  – may **not develop an ability to formulate research** and conduct it independently.
  – research methods may be limited.
  – focus on the master’s problem may becloud the apprentice student’s understanding.

• Fits best the condition of
  – a relatively immature, inexperienced student who needs strong direction.
  – well-defined, funded streams of research.
Supervision Styles - Collegial master/apprentice style

- Limited domain advising, not restricted by advisor’s current research activity but fit within the general domain of expertise.
- Supervisor willing to advise on problems that are within the scope of his research and methods within his skill set.
- Puts more responsibility on student than master/apprentice style
- Work well when
  - both supervisor and student were interested in a problem
  - supervisor had sufficient expertise to provide good guidance.
Supervision Styles – Collegial development style

- Extended domain advising, not in domain of supervisor’s current or past research but is extended to areas in which the supervisor has an interest and willing to invest in becoming reasonably proficient.
- There is a joint learning experience
  - Supervisor starts with more experience, but both are learning details of research area.
- Fits when
  - Supervisor willing to expand his or her research competence
  - Student willing to engage in a joint learning experience.
- Fits very well for a dissertation that opens up a new or fairly new area of research.
- Fail when supervisor was not willing to make the investment to be competent.
Supervision Styles – Guidance and Suggestion style

• General advising over a range of problem domains.
• Some supervisors have good skills at problem identification and problem formulation over a range of problems and research methods.
  – They conceptualize well and are good at “sense making.”
• Works best with students who are willing and able to take initiative and take responsibility for learning the research domain and the appropriate research methods.
• Student gets good general guidance and good evaluation of the dissertation but usually does not get detailed feedback and detailed mentoring of methods.
• It is not very good for immature students who need more detailed guidance.
• Good with mature students who took initiative.
Supervision Styles – Passive hands-off style

- Laissez faire style, the supervisor takes the role of a general quality control reader.
- Student must take the initiative to define a problem, decide on a research method, develop a research plan, and so forth.
- Supervisor responds to student plans and initiatives with some suggestions, but the responsibility is almost entirely with student.
- Given a competent supervisor who gives good suggestions in response to student initiatives and plans, the advantages are that the student develops independent skills at formulating problems and planning research.
- The disadvantages are that the student may meander from problem to problem and take too long to do a dissertation. Under these conditions, a student may not develop good skills and may drop out of the program.
- For fairly mature students with an ability to take initiative, this style may work well.
- Significant danger with project for which the student does not have the necessary background for doing a good dissertation or the supervisor is unable to do reasonable quality review.
- For immature students, it is likely to be a disaster.
## Supervision Style

<table>
<thead>
<tr>
<th>Style</th>
<th>Advisor Role and Behavior</th>
<th>Student Role and Behavior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strong master/apprentice style</td>
<td>Advisor is master. Advisor has a well specified domain of expertise and set of problems within it.</td>
<td>Student is an apprentice working for the advisor. Student works on advisor’s problems.</td>
</tr>
<tr>
<td>Collegial master/apprentice style</td>
<td>Advisor is expert who limits advising to problems that are within scope of his or her research skill set but will work on student’s problem.</td>
<td>Student develops a problem within advisor’s domain and skills and works under the advisor to develop the research plan and procedures.</td>
</tr>
<tr>
<td>Collegial development style</td>
<td>Advisor is senior colleague who will respond to student research problem and extend his or her advising domain to include new problems and new skills.</td>
<td>Student takes initiative to introduce new problem that requires new skill set and works as a junior colleague with advisor in joint development of new domain.</td>
</tr>
<tr>
<td>Guidance and suggestion style</td>
<td>Advisor is a senior colleague who gives good general guidance over a wide range of problems and methods but does not have personal skill in all of them.</td>
<td>Student is an independent, junior colleague who takes initiative for presenting problems and research plans for discussion and guidance. Student develops required skills.</td>
</tr>
<tr>
<td>Passive hands-off style</td>
<td>Advisor has quality control role and responds only to requests or documents and performs only general quality control review</td>
<td>Student is an independent researcher who takes initiative for developing problem, developing skills, and presenting research plans for general review and approval.</td>
</tr>
</tbody>
</table>
## Advantages & Disadvantages of Styles

<table>
<thead>
<tr>
<th>Advising Style</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strong master/apprentice style</td>
<td>Advisor is heavily involved and gives expert direction for research activities. Student learns how to do research within advisor’s domain.</td>
<td>Student works on advisor’s problems and within advisor expertise and may not develop independence.</td>
</tr>
<tr>
<td>Collegial master/apprentice style</td>
<td>Advisor knows the research domain and research methods and can give expert advice on them. Student can take initiative in formulating problem and working with advisor.</td>
<td>Student is limited to advisor research domain and advisor research skills. Student may do work that is not within his or her long term research plan.</td>
</tr>
<tr>
<td>Collegial development style</td>
<td>Advisor and student develop together to explore new domain and new research methods. Student develops independence within relationship.</td>
<td>Risk of exploring new research area that does not work. Risk that necessary development of both advisor and student does not occur or occurs unevenly.</td>
</tr>
<tr>
<td>Guidance and suggestion style</td>
<td>Student is able to develop independent research and research management skills while receiving guidance and suggestions. The student may research a broad range of topics and employ broad range of methods.</td>
<td>Student may not get expert advice from advisor on many issues, so student must search for expert advice. Student has significant responsibility for research quality and management of process.</td>
</tr>
<tr>
<td>Passive hands-off style</td>
<td>Student is able to act independently with little interference from advisor. Student can work on problems of his or her choosing.</td>
<td>Student may make serious mistakes because of lack of advice and suggestions. Student may flounder and not complete on timely basis.</td>
</tr>
</tbody>
</table>
# Supervision Style and Fit

<table>
<thead>
<tr>
<th>Advising Style</th>
<th>Likely Fit with Student Archetypes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strong master/apprentice style</td>
<td>Good fit with immature, unconfident student. “Do what I do” provides clear direction. Possible good fit with</td>
</tr>
<tr>
<td>Colleagial master/apprentice style</td>
<td>Some stress but a workable fit with immature, unconfident student because some initiative is required. Fairly good fit with somewhat mature, somewhat confident student because it provides boundaries for what is expected and allows some independence. Some stress from this style for mature, confident student because of constraints on what can be done by student.</td>
</tr>
<tr>
<td>Colleagial development style</td>
<td>Stressful for immature, unconfident student because of need for significant student initiative. Good fit with somewhat mature, somewhat confident student because it builds confidence through development interactions. Reasonable fit and reasonably low stress for very mature, confident student if advisor and student are compatible relative to problems and methods.</td>
</tr>
<tr>
<td>Guidance and suggestion style</td>
<td>Very stressful for immature, unconfident student because of vagueness of process and need for initiative that may exceed capacity of novice. Stressful but workable relationship for somewhat mature, somewhat confident student because of high initiative required from student. Good fit and reasonably low stress fit for very mature, confident student who is given much freedom and good feedback.</td>
</tr>
<tr>
<td>Passive hands-off style</td>
<td>Likely disaster for advising relationship with immature, unconfident student because not sufficient guidance. Stressful relationship between advisor with this style and somewhat mature, somewhat confident student because of insufficient feedback. This style may work for a very mature, confident student but introduces risks because of lack of clarity in expectations.</td>
</tr>
</tbody>
</table>
Helping Students Prepare for a GOT PhD Journey

Successful, Timely Completion of PhD

- Help in Analysis and Writing
- Coach for Mastery in Research Methods
- Help Plan for GOT
- Coach for Deep & Critical Comprehension of Subject Matter
- Help Identify Interesting & Suitable Research Topic
- Prepare Researcher Mindset
- Set-up Supportive Research Environment
Setting-Up of a Supportive Research Environment

Don’t Slow Your Student because of Unconducive Research Environment
Setting Up A Supportive Research Environment

• Build systematic datasets/gold standards
• Have organized, well-documented standard programs (pre-processing, benchmarks)
• Have well-administered servers & tools
• Provide non-research support
• Develop mentoring system
• Build a strong, cohesive and family-like research group
How to Help Student to Publish (cont.)

• Schedule regular presentations to group – to be developed into publication
• Monitor progress (eg. every 2 months) through paper
• Set appointment by draft paper
• Give fast feedback
• Suggest and ready to pay for proof-reading of papers
• Have list of journals and reviewers for papers
• Suggest collaborator(s)
Monitoring Attendance in WA

Attendance for group meeting

X absent
!? MIA (missing - did not contact supervisor)

📅 21 September 2017

1. Ng Liang Shen 🎤
2. Sharhan Shawkat 🎤
3. Maged 🎤
4. Akram 🎤
5. Laika 🎤
6. Hafiz 🎤
7. Zuriati X
8. Kasturi X
9. Abbas ⚫
10. Anser ⚫
11. Mutasem 🎤
12. Altaf ⚫
13. Omayma X
14. Shirin X
15. Hamza X
16. Samaa ⚫
17. Phuripoj ⚫
18. Anupon ⚫
Remote PhD
Mindset Preparation
Wrong mindset, and it takes forever to finish PhD
Mindset Preparation

• Matured
• Ability to focus and concentrate
• Disciplined
• Independent
• Hardworking
• Innovative & Creative
• Critical Thinking
• Available (time)
Help Them Understand Characteristics of Research

• Systematic
  – Work in a sequence of steps which were in order and thus systematic
  – Order: Observation, Problem Definition, Hypothesis, Testing, Conclusion

• Follows a scientific method of enquiry
  – Researcher should not just jump at the conclusions,
  – But used a scientific method of enquiry in reaching conclusion
Emphasize What Research is NOT

• Research is not information gathering
  – Gathering information from resources such as books or magazines.
• Research is not the transformation of facts
  – No contribution to new knowledge although this might make knowledge more accessible
• Research is not about having successfully developed something
  – What can others learn from it?
Helping Student Find Good Research Topic

Never underestimate the importance of choosing the right topic
Finding A Good Research Problem

- Novelty of the Idea.
  Research is a study of new ideas in the field
- Significance for the Community.
  What idea is actually needed for the community “today”.
- Contribution from the Researcher.
  An amount of efforts made by a researcher to study the idea.
Novelty:

What to Expect of a PhD Research

From “Matt Might’s Illustrated Guide to a PhD.”
What to Expect of a PhD Research

Substantial body of original and significant work

From "Matt Might's Illustrated Guide to a PhD."
What to Expect of a PhD Research

Winning a Nobel Price through PhD Research is not Necessary

From “Matt Might's Illustrated Guide to a PhD.”
Expectation of a PhD Research

New Facts
New Ideas
New Facts + Ideas

From “Matt Might’s Illustrated Guide to a PhD.”
Expectation of a PhD Research

Discovery
Interpretation
Revision

From "Matt Might's Illustrated Guide to a PhD."
Expectation of a PhD Research

Thesis
Anti-thesis
Synthesis

From “Matt Might's Illustrated Guide to a PhD.”
Problem Formulation

Problem Formulation

A situation, person or things that needs attention and needs to be dealt with or solved

Problem

To develop all the details of a plan for solving problem

Formulation
Problem Identification Process

• Identify Domain Area (Broad Topic)
• Make a conceptual map by identifying related topics (Narrow Topic)
  – Divide the domain area into progressively smaller sub-area until one reaches a subject of interest
  – Tools: mindmap, fishbone
• Identify topic of interests student wish to explore
• Identify “real” problem area in the sub-area
  – What is the most important and possible relative to scope and scale of study?
Problem Solution Process

- Problem analysis/Root-cause analysis
- Duration analysis/Activity-based costing
- Outcome analysis
- Breaking assumptions
- Technology analysis
- Proxy benchmarking
- Activity elimination/process simplification
Example - Identifying Gaps in Information Retrieval Research

• Other data that can be used to enhance?
  – Eg. Explicit vs implicit, multimedia

• Other ways to represent data?
  – Eg. Graph? Passages?

• Challenge assumptions and rules.
  – Eg.: crisp vs fuzzy? Deterministic vs probabilistic?

• Other external knowledge bases or sources to enhance?

• Combination? Optimization? Weightage?

• Adapting ideas from other fields? Eg. Diversity analysis, Cross-structural theory, Game theory, etc.
Significance to Community

• What is the Need?
• What is your Approach to address the need?
• What is the Benefit of your approach over cost?
• In what ways are your research better than the Competitors?
Problem tree analysis

- Analysis of researchable issue to gain insights about its possible cause-effect relationships.
- Helps in identifying the critical areas where an intervention would provide a solution to the problem of concern

<table>
<thead>
<tr>
<th>Impacts</th>
<th>Problem</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Protection of IP</td>
<td>No way to detect beyond cut-and paste plagiarism</td>
<td>Intelligent plagiarism cannot be detected</td>
</tr>
<tr>
<td>Questionable Integrity</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Causes</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Plagiarist change words</td>
<td>Change order of words</td>
</tr>
</tbody>
</table>
Change Problem Tree to Objective Tree

- **Impacts**
  - Protection of IP
  - Preserve Integrity

- **Effect**
  - Avoid Intelligent plagiarism

- **Aims**
  - Develop ways to detect beyond cut-and-paste plagiarism

- **Objectives**
  - Semantic-based detection
  - Predicate-based detection
Root Cause Analysis in Problem Tree
– Keep asking Why?

(SA Prathapar, Research Methodology Slide, 2012)
Horizontal Analysis

Output:
Decisions?  
Impact?
Horizontal Analysis

Input: Data?

Output: Decisions? Impact?
Horizontal Analysis

Input: Data?

Processes?

Output: Decisions? Impact?
Horizontal Analysis

Input: Data?

Processes?

Output: Decisions?
Impact?
Example of Horizontal Analysis

Input:
- Historical Stock Values
- Other Commodities
- Weather
- News

Processes?

Crude Palm Oil Stock Market Prediction:
- Turning Points?
- Values?
How many research objectives?

3 objectives

Objective 1
6 months

Objective 2
6 months

Objective 3
6 months
Make Your RQ Specific Enough

• Very weak (in an empirical sense)
  – *Is the new technique any good?*

• Weak
  – *Is the new technique better than multi-tap?*

• Better
  – *Is the new technique faster than multi-tap?*

• Better still
  – *Is the new technique faster than multi-tap within one hour of use?*

• Even better
  – *If error rates are kept under 2%, is the new technique faster than multi-tap within one hour of use?*
Topic Identification
– Can he finish it on time?

**Does It Suit Him?**
- Student’s Background
- Student’s Interests
- Student’s Preparation
- Student’s Capabilities

**How Fast Can He Start?**
- Equipment/Hardware
- Data Availability
- Software/Simulator

**Supportive Environment?**
- Supervisor’s Knowledge
- Supervisor’s Research Experience
- Senior Students
- Available Networks

**Literature Support?**
- Is it well-published? In reputable venues?
- Is it a growing research area?
- Can the data and methodology be obtained from literature? Understood by student?
In case he is not prepared at all...

Where should I look for ideas? What are the top journals in my area? Hot topics?

Who are the active researchers in the area?
Guide him to read ...

Have a feel of good research and find his interest in good, reputable journals
Select Categories based on Your Area

Select Year 2016 from the dropdown menu.
Choose Quartile, IF Range or Percentile
Click Submit
<table>
<thead>
<tr>
<th>Rank</th>
<th>Journal Title</th>
<th>Total Cites</th>
<th>Impact Factor</th>
<th>Eigenfactor Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>IEEE Communications Surveys and Tutorials</td>
<td>4,172</td>
<td>9.220</td>
<td>0.02037</td>
</tr>
<tr>
<td>2</td>
<td>IEEE TRANSACTIONS ON FUZZY SYSTEMS</td>
<td>9,220</td>
<td>6.701</td>
<td>0.01354</td>
</tr>
<tr>
<td>3</td>
<td>International Journal of Neural Systems</td>
<td>1,235</td>
<td>6.086</td>
<td>0.00247</td>
</tr>
<tr>
<td>4</td>
<td>IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE</td>
<td>31,757</td>
<td>6.077</td>
<td>0.06131</td>
</tr>
<tr>
<td>5</td>
<td>IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION</td>
<td>7,999</td>
<td>5.908</td>
<td>0.00910</td>
</tr>
<tr>
<td>6</td>
<td>MIS QUARTERLY</td>
<td>11,320</td>
<td>5.384</td>
<td>0.01137</td>
</tr>
<tr>
<td>7</td>
<td>COMPUTER-AIDED CIVIL AND INFRASTRUCTURE ENGINEERING</td>
<td>1,935</td>
<td>5.288</td>
<td>0.00367</td>
</tr>
<tr>
<td>8</td>
<td>INTEGRATED COMPUTER-AIDED ENGINEERING</td>
<td>533</td>
<td>4.981</td>
<td>0.00124</td>
</tr>
<tr>
<td>9</td>
<td>IEEE Transactions on Cybernetics</td>
<td>2,248</td>
<td>4.943</td>
<td>0.01017</td>
</tr>
<tr>
<td>10</td>
<td>IEEE Transactions on Neural Networks and Learning Systems</td>
<td>12,919</td>
<td>4.654</td>
<td>0.02246</td>
</tr>
<tr>
<td>11</td>
<td>JOURNAL OF INFORMATION TECHNOLOGY</td>
<td>1,665</td>
<td>4.775</td>
<td>0.00268</td>
</tr>
</tbody>
</table>
Click Rank to Get Quartile of Journal
Find Most Recent, Most Read and Most Cited Articles

<table>
<thead>
<tr>
<th>Title</th>
<th>Authors</th>
<th>Journal</th>
<th>Year</th>
<th>Volume</th>
<th>Issue</th>
<th>Pages</th>
<th>Abstract/Links</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOW-RANK APPROXIMATION BASED NON-NEGATIVE MULTI-WAY ARRAY DECOMPOSITION ON EVENT-RELATED POTENTIALS</td>
<td>Fengqi Cong, Qiuqiu Zhang, Pei Aistikainen, Qiong Zhao, Qiang Wu, Asoke K. Nandi, Jarl K. Hietanen, Tamaz Rustamiani, Andrzei Cichocki</td>
<td>International Journal of Neural Systems Vol. 24, No. 06, 1440005</td>
<td>2014</td>
<td>24</td>
<td>06</td>
<td>14</td>
<td>Abstract [PDF] (441 KB)</td>
</tr>
<tr>
<td>TRANSCRANIAL MAGNETIC STIMULATION (TMS) MODULATES EPILEPTIFORM DISCHARGES IN PATIENTS WITH FRONTAL LOBE EPILEPSY: A PRELIMINARY EEG/TMS STUDY</td>
<td>Vasilios K. Kimsikidis, Dimitris Kougianotis, Sotirios Papagianopoulos, Nikolaos Valvidis</td>
<td>International Journal of Neural Systems Vol. 23, No. 01, 1250035</td>
<td>2013</td>
<td>23</td>
<td>01</td>
<td>1250035</td>
<td>Abstract [PDF] (1416 KB)</td>
</tr>
<tr>
<td>APPLICATION OF EMPIRICAL MODE DECOMPOSITION (EMD) FOR AUTOMATED DETECTION OF EPILEPSY USING EEG SIGNALS</td>
<td>Roshan Joy Martins, U. Rajendra Acharya, Jien Hong Tan, Andrea Petznick, Ratna Yanti, Chia Kuek Chia, E. Y. K. Ng, Louise Tong</td>
<td>International Journal of Neural Systems Vol. 22, No. 06, 1250027</td>
<td>2012</td>
<td>22</td>
<td>06</td>
<td>1250027</td>
<td>Abstract [PDF] (798 KB)</td>
</tr>
<tr>
<td>CORTICAL EXCITABILITY AND REFRACTORY EPILEPSY: A THREE-YEAR LONGITUDINAL TRANSCRANIAL MAGNETIC STIMULATION STUDY</td>
<td>Radiwa A. B. Badawy, Graeme D. Jackson, Samuel F. Berkovic, Richard A. L. McDowall</td>
<td>International Journal of Neural Systems Vol. 22, No. 01, 1250030</td>
<td>2013</td>
<td>22</td>
<td>01</td>
<td>1250030</td>
<td>Abstract [PDF] (1517 KB)</td>
</tr>
</tbody>
</table>
Search by Topic in WOS

Web of Science Core Collection

Basic Search

plagiarism detection system

Timespan

All years

1990 to 2016

More Settings
Result List by Topic in WOS

7. High time for a common plagiarism detection system
   By: Jaric, Ivan
   SCIENTOMETRICS Volume: 106 Issue: 1 Pages: 457-459 Published: JAN 2016

8. DOCODE 3.0 (Document COpy DEctor): A system for plagiarism detection by applying an information fusion process from multiple documental data sources
   By: Velasquez, Juan O.; Covaciavich, Yerko; Molina, Francisco; et al.
   INFORMATION FUSION Volume: 27 Pages: 64-75 Published: JAN 2016

9. Software Plagiarism Detection with Birthmarks Based on Dynamic Key Instruction Sequences
   By: Tian, Zhenzhou; Zheng, Qinghua; Liu, Ting; et al.
   IEEE TRANSACTIONS ON SOFTWARE ENGINEERING Volume: 41 Issue: 12 Pages: 1217-1235 Published: DEC 2015

10. PDLK: Plagiarism detection using linguistic knowledge
    By: Abdi, Asad; Isrini; Norisma; Alguliyev, Rasim M.; et al.
    EXPERT SYSTEMS WITH APPLICATIONS Volume: 42 Issue: 22 Pages: 8936-8946 Published: DEC 1 2015
Click on Journal Title to Find Out Its Quartile and Rank
Create Citation Alert
Scopus - evaluated by SciMago Journal Rank (SJR)*

http://www.scimagojr.com/journalrank.php (developed from the Google PageRank algorithm)

| Subject Area         | Computer Science
|----------------------|------------------|
| Subject Category     | Computer Science Applications
| Country              | All
| Year                 | 2012

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Bioinformatics</td>
<td>4.223</td>
<td>204</td>
<td>747</td>
<td>2.251</td>
<td>17.874</td>
<td>14.933</td>
<td>2.160</td>
<td>5.38</td>
<td>23.93</td>
<td></td>
</tr>
<tr>
<td>2 Argument and Computation</td>
<td>3.173</td>
<td>5</td>
<td>8</td>
<td>17</td>
<td>377</td>
<td>68</td>
<td>1.170</td>
<td>4.00</td>
<td>47.13</td>
<td></td>
</tr>
<tr>
<td>3 Journal of Field Robotics</td>
<td>3.161</td>
<td>46</td>
<td>48</td>
<td>148</td>
<td>1.610</td>
<td>492</td>
<td>134</td>
<td>3.39</td>
<td>33.54</td>
<td></td>
</tr>
<tr>
<td>4 Journal of Informetrics</td>
<td>3.075</td>
<td>26</td>
<td>78</td>
<td>172</td>
<td>2.242</td>
<td>787</td>
<td>158</td>
<td>4.80</td>
<td>28.74</td>
<td></td>
</tr>
</tbody>
</table>
Analyze trend

Going up?
Conference & Journal for Area

Year | Source Title
--- | ---
2000 | Lecture Notes in Computer Science including Subseries Lecture Notes in Artificial Intelligence
2007 Production of Paper 2007 Asia Pacific Conference on Communications Apcc
2009 | 3rd International Conference on Genetic and Evolutionary Computing Wgec 2009
2009 | 3rd International Conference on Innovative Computing Information and Control Ici...
People who work in same area
Institution Doing Active Research in Area

Number of documents

<table>
<thead>
<tr>
<th>Affiliation</th>
<th>Documents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Universiti Teknologi Malaysia</td>
<td>7</td>
</tr>
<tr>
<td>International University of Africa</td>
<td>2</td>
</tr>
<tr>
<td>Západocešká Univerzita</td>
<td>2</td>
</tr>
<tr>
<td>Xian Jiaotong University</td>
<td>2</td>
</tr>
<tr>
<td>Vysoká škola banská-Technická univerza</td>
<td>2</td>
</tr>
<tr>
<td>Universität Dortmund</td>
<td>2</td>
</tr>
<tr>
<td>Hadhramout University of Science and Technology</td>
<td>2</td>
</tr>
<tr>
<td>Tart University</td>
<td>2</td>
</tr>
<tr>
<td>International University of Africa</td>
<td>1</td>
</tr>
<tr>
<td>P.A. College</td>
<td>1</td>
</tr>
<tr>
<td>School of Management</td>
<td>1</td>
</tr>
<tr>
<td>Zeidman Consulting</td>
<td>1</td>
</tr>
</tbody>
</table>
Coaching For Deep & Critical Apprehension of Literature
Guding Students to Do LR

- Conventional Review
- System Literature Review
- Experimental Review
After initial topic/area has been determined, ask student to find a focus.

ียวA review is usually organized around ideas  mę not the sources themselves as an annotated bibliography would be organized.
רוש not just simply list your sources and go into detail about each one of them, one at a time.

_construct thesis statement  _-mouth Justify the need for research through LR
Example of Thesis Statement for Review

• How effective is semantic-based database integration?
• Can current plagiarism detection techniques identify intelligent plagiarism?
• What are the fuzzy aspects of plagiarism? Can current and fuzzy logic approaches detect these fuzzy aspects in plagiarism?
Parts of a Good Review

• Introduction
  – To the context & importance/significance of the work
• Analysis
  – Main framework used to review a particular topic
• Synthesis
  – Reorganization/re-shuffling of main parts inside the analytical framework
• Evaluation
  – Comparison – similarities, difference
  – Critical discussion of strength, weaknesses
  – Gap analysis
• Suggestion
  – Of further work that can or need to be done
Writing the Introduction Section

1. **What** is the problem? Define.
2. **How** has it been solved? Show the general ways. Broad, to specific (to your focus).
3. **Why** it need solving? Significance, impact.
4. **Unique** viewpoint of your review.
Example of Introduction

I. INTRODUCTION

The problem of plagiarism has recently increased because of the digital era of resources available on the World Wide Web. Plagiarism detection in natural languages by statistical or computerized methods has started since the 1990s, which is pioneered by the studies of copy detection mechanisms in digital documents [42, [43]. Earlier than plagiarism detection in natural languages, code clones and software misuse detection has started since the 1970s by the studies to detect programming code plagiarism in Pascal and C [28, [44]–[47]. Algorithms of plagiarism detection in natural languages and programming languages have noticeable differences. The first one tackles different textual features and diverse methods of detection, while the latter mainly focuses on keeping track of metrics, such as number of lines, variables, statements, subprograms, calls to subprograms, and other parameters. During the last decade, research on automated plagiarism detection in natural languages has actively evolved, which takes the advantage of recent developments in related fields like information retrieval (IR), cross-language information retrieval (CLIR), natural language processing, computational linguistics, artificial intelligence, and soft computing. In this paper, a survey of recent advances in the area of automated plagiarism detection in text documents is presented, which started roughly in 2005, unless it is noteworthy to state a research prior than that. Earlier study was excellently reviewed by [48] and [52]–[55].

This paper brings patterns of plagiarism together with textual features for characterization of each pattern and computerized methods for detection. The contributions of this paper can be summarized as follows: First, different kinds of plagiarism are organized into a taxonomy that is derived from a qualitative study and recent literatures about the plagiarism concept. The taxonomy is supported by various plagiarism patterns (i.e., examples) from available corpora for plagiarism [60]. Second, different textual features are illustrated to represent text documents for the purpose of plagiarism detection. This is achieved by

- **What** is the problem? Define
- **Why** it need solving? Significance, impact.
- **How** has it been solved? Show the general ways. Broad, to specific (to the review focus).
- **Unique** viewpoint of review.
Organizing Your Review: The Analysis

- Set out your thinking on paper through **maps** and **trees**.
  - Build conceptual map/theoretical framework
  - Build taxonomy/trees of area

<table>
<thead>
<tr>
<th>Feature map</th>
<th>Classifies and categorises your thought in tabular form</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concept map</td>
<td>Links between concepts and processes, or shows relationship between ideas and practice</td>
</tr>
<tr>
<td>Tree construction</td>
<td>Shows how topic branches out into subthemes and related questions or represents stages in the development of a topic</td>
</tr>
</tbody>
</table>
Example of Analysis Framework

• What data used
• Different representation of data
• Different ways of comparing data or calculating values of similarity
• Different paradigms used for comparisons
• External knowledge-bases used to enhance. Eg.
  – Wordnet, Social networks
Conceptual framework

Start with a mind-map

- Cut and paste literature into bubbles
- Give overall picture, broad view
- Can see vacuum where we can focus
- Know where to put boundaries, scope, limitations
Conceptual Framework

• Building up of concept of work through literature

Theoretical Framework

• What theory support each component of the mind-map
Conceptual framework

- Different types/parts of problems
- Problem/Focus
  - Approaches for part 1
  - Approaches Type 1
  - Approaches Type 2
  - Approaches for part 2
Conceptual framework

Different types of Intelligent Plagiarism

Intelligent Plagiarism Detection Techniques

Character-based

Section-based

Text

Non-Text

Ways to Represent Documents

Styles

Ways to Compare
Eg. Framework of Problem Area

Query Document $d_q$

[Collection Documents $D$]

Plagiarism Detection System

Suspicious sections $s_q$ [+sources]
White-box design for extrinsic plagiarism detection system

- Query Document \(d_q\)
- Collection Documents \(D\)
- Heuristic Retrieval
- Detailed Analysis
- Candidate Documents \(D_x\)
- Knowledge-Based Post-Processing
- Suspicious sections \((s_q, s_x)\)

\[(s_q, s_x) : s_q \in d_q, s_x \in d_x, d_x \in D\]
White-box design for intrinsic plagiarism detection system
White-box design for cross-lingual plagiarism detection system

- Query Document $d_q$
- Machine Translation
  - Fingerprinting
  - Keyword extraction
  - Lookup
    - IR
    - CLIR

- Candidate Documents $D_x$
- Detailed Analysis
- Knowledge-Based Post-Processing

- Collection Documents $D$
- Suspicious sections $(s_i, s_x)$

$(s_i, s_x) : s_i \in d_q, s_x \in d_x, d_x \in D$
Eg. Current Taxonomy

Plagiarism type
- **Detection principal**

- Exact copy
- Modified copy

- Large part of document
  - **Document model comparison**
  - With reference
    - Chunk identity
  - Without reference
    - Style analysis

- Small part of document
  - **Local identity analysis**
  - **Cross-language similarity analysis**
  - **Local similarity analysis**
  - **Fingerprinting**
  - **Style analysis**
  - **Without reference**
Modified Taxonomy of Concepts
Conceptual framework

Organize LR chapter from mind-map
Conceptual framework – can result in research objectives
Conceptual framework into Review papers

Review Paper 1

Intelligent Plagiarism Detection Techniques

Review Paper 2

Fuzzy Plagiarism Detection

Fuzzy-Semantic Based Plagiarism Detection

Review Paper 3

Semantic-based Plagiarism Detection

Review Paper 4
Conceptual framework into Chapters

Chapter 4
Intelligent Plagiarism Detection Techniques

Chapter 5
Fuzzy Plagiarism Detection
Semantic-based Plagiarism Detection

Chapter 6
Fuzzy-Semantic Based Plagiarism Detection

Chapter 7
Tree Based Organization

FIGURE 2.3 Schematic diagram for video classification
Synthesizing Sources in Each Section of the Review Paper

• Based on groupings in tables
• Inverted pyramid
• Thematic
• By trend
• Questions for Further Research
• Chronological
# Representation

<table>
<thead>
<tr>
<th>Lexical features</th>
<th>Examples</th>
<th>Required Tools and Resources</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Character n-grams (fixed-length)</td>
<td>-</td>
<td></td>
<td>[1]</td>
</tr>
<tr>
<td>Character n-grams (variable-length)</td>
<td>Feature selector (e.g. n-gram weights)</td>
<td></td>
<td>[16]</td>
</tr>
<tr>
<td>Word n-grams</td>
<td>Tokenizer, [Stemmer, Lemmatizer]</td>
<td></td>
<td>[2, 3, 17, 26] [30]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Syntactic features</th>
<th>Chunks</th>
<th>Tokenizer, POS tagger, Text chunker (Windowing)</th>
<th>[4]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Part-of-speech and phrase structure</td>
<td>Tokenizer, Sentence splitter, POS tagger</td>
<td>[6, 12, 48]</td>
</tr>
<tr>
<td></td>
<td>Word position/order</td>
<td>Tokenizer, Sentence splitter, Compressor (e.g. Lempel-Zif)</td>
<td>[13, 14]</td>
</tr>
<tr>
<td></td>
<td>Sentence</td>
<td>Tokenizer, Sentence splitter, POS tagger, Text chunker, Partial parser</td>
<td>[16, 58]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Semantic features</th>
<th>Synonyms, hyponyms, hypernyms, etc.</th>
<th>Tokenizer, [POS tagger], Thesaurus</th>
<th>[14, 16, 18, 58] [30]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Semantic dependencies</td>
<td>Semantic dependencies</td>
<td>Tokenizer, Sentence splitter, POS tagger, Text chunker, Partial parser, Semantic parser</td>
<td>[14, 61]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Structural features</th>
<th>Block-specific</th>
<th>HTML parser, Specialised parsers</th>
<th>[21, 29]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content-specific</td>
<td>Tokenizer, [Stemmer, Lemmatizer], Specialised dictionaries</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>
## Similarity Evaluation

<table>
<thead>
<tr>
<th>Vector Similarity Metric</th>
<th>Description &amp; Equation</th>
<th>Equation</th>
<th>Range</th>
<th>Example</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matching coefficient</td>
<td>-similar to Hamming distance but between vectors of equal length.</td>
<td>$M(x, y) =</td>
<td>x</td>
<td>-</td>
<td>x \cap y</td>
</tr>
<tr>
<td>Jaccard (or Tanimoto) coefficient</td>
<td>-defines number of shared terms against total number of terms.</td>
<td>$J(x, y) = \frac{</td>
<td>x \cap y</td>
<td>}{</td>
<td>x \cup y</td>
</tr>
<tr>
<td>Dice’s coefficient</td>
<td>-similar to Jaccard but reduces the effect of shared terms between vectors.</td>
<td>$D(x, y) = \frac{2</td>
<td>x \cap y</td>
<td>}{</td>
<td>x</td>
</tr>
<tr>
<td>Overlap (or containment)</td>
<td>coefficient -if $v_1$ is subset of $v_2$ or the converse, then the similarity coefficient is a full match.</td>
<td>$O(x, y) = \frac{</td>
<td>x \cap y</td>
<td>}{\min(</td>
<td>x</td>
</tr>
<tr>
<td>Cosine coefficient</td>
<td>-finds the cosine angle between two vectors.</td>
<td>$\cos(x, y) = \frac{\sum (x_i \cdot y_i)}{\sqrt{\sum x_i^2} \cdot \sqrt{\sum y_i^2}}$</td>
<td>0 to 1</td>
<td>$\cos(x, y)=0.340/0.3421 =0.9939 =1$</td>
<td>[9, 21, 26, 28]</td>
</tr>
<tr>
<td>Euclidean distance</td>
<td>-measures the geometric distance between two vectors.</td>
<td>$Ec(x, y) = \sqrt{\sum</td>
<td>x_i - y_i</td>
<td>^2}$</td>
<td>0 to $\infty$</td>
</tr>
<tr>
<td>Squared Euclidean Distance</td>
<td>-places progressively greater weight on vectors that are further apart</td>
<td>$SEc(x, y) = \sum (x_i - y_i)^2$</td>
<td>0 to $\infty$</td>
<td>$SEc(x, y)=0.01$</td>
<td>-</td>
</tr>
<tr>
<td>Manhattan Distance</td>
<td>-measures the average difference across dimensions and yields results similar to the simple Euclidean distance</td>
<td>$Manh(x, y) = \sum</td>
<td>x_i - y_i</td>
<td>^2$</td>
<td>0 to $\infty$</td>
</tr>
</tbody>
</table>
# Methods and Their Efficiency in Detecting Different Plagiarism Types

<table>
<thead>
<tr>
<th>Technique</th>
<th>Tasks</th>
<th>IR</th>
<th>Language(s)</th>
<th>Plagiarism Type(s)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>extrinsic</td>
<td>intrinsic</td>
<td>mono-lingual</td>
<td>cross-lingual</td>
<td></td>
</tr>
<tr>
<td>Char-Based (CNG)</td>
<td>☑</td>
<td>☑</td>
<td>any</td>
<td>☑ ☑</td>
<td>[1-6]</td>
</tr>
<tr>
<td>Vector-Based (VEC)</td>
<td>☑</td>
<td>☑</td>
<td>any</td>
<td>☑ ☑</td>
<td>[7-11]</td>
</tr>
<tr>
<td>Syntax-Based (SYN)</td>
<td>☑</td>
<td>☑</td>
<td>specific</td>
<td>☑ ☑</td>
<td>[6, 12, 13]</td>
</tr>
<tr>
<td>Semantic-Based (SEM)</td>
<td>☑</td>
<td>☑</td>
<td>specific</td>
<td>☑ ☑ ☑</td>
<td>[14, 15]</td>
</tr>
<tr>
<td>Fuzzy-Based (FUZZY)</td>
<td>☑</td>
<td>☑</td>
<td>specific</td>
<td>☑ ☑ ☑ ☑</td>
<td>[16-19]</td>
</tr>
<tr>
<td>Structural-Based (STRUC)</td>
<td>☑</td>
<td>☑</td>
<td>specific</td>
<td>☑ ☑ ☑ ☑ ☑</td>
<td>[21, 29]</td>
</tr>
<tr>
<td>Stylometric-Based (STYLE)</td>
<td>☑</td>
<td>☑</td>
<td>specific</td>
<td>☑ ☑ ☑</td>
<td>[22, 23, 32-35]</td>
</tr>
<tr>
<td>Cross-Lingual (CROSS)</td>
<td>☑</td>
<td>☑</td>
<td>cross</td>
<td></td>
<td>[31, 36-38]</td>
</tr>
</tbody>
</table>
There are a number of main approaches in....

One of the most popular/used/oldest technique is ...., which has been used by ....(give refs)

Another technique is .... (give refs)

The two techniques are similar in terms of ...... However, the first technique ...(highlight difference)
For each (group) of technique, describe

• What is it?
• How is it done/performed?
• Why is it introduced/proposed? What advantages it offered over other techniques?
Discussion can be ...

• Embedded in each (group of) technique described
• At the end of each section after the techniques are introduced
• In a separate “Discussion” section
Discussion and Evaluation

• Review should be evaluative and not merely descriptive.
  – For example possible reasons for similarities or differences between studies are considered rather than a mere identification of them.
Examples ...

• Similarities
  – As can be observed, all the techniques discussed above used ...

• Weaknesses
  – The problems that could arise with the use of such techniques .... Similar problems have been observed in ....

• Link to research questions
  – A possible ways to improve is ....
Critical Framework

• Regardless of the method of one’s research—subjective, textual, historical, empirical, etc.—an analytical lens must be used to interpret literature and data.

• For quantitative research
  – this framework is the logical or mathematical method by which the data is analyzed

• When analyzing or interpreting qualitative or textual research
  – choose an individual or, more likely, interdependent approaches or lenses through which that data or material is interpreted
Discussion

When choosing between clustering methods, a few factors need to be taken into account. These factors are discussed in the following sections.

Computational efficiency

Table 2.4 summarises the computational complexity of some of the clustering method discussed. Basically, non-hierarchical methods are usually more computationally efficient than hierarchical methods. The Jarvis-Patrick method is very computationally efficient because ……..

Ability to recover natural clusters in dataset

A study by Blashfield [1976] revealed that single linkage has the lowest agreement between cluster solutions and actual structure, whilst Ward’s method has the highest. The superiority of Ward's method in producing meaningful clusters is confirmed by Adamson and Bawden [1981] ……..

Effectiveness for intended application

Empirical results of tests that use evaluation criteria specific to the problem being studied can be used to get an idea of the most suitable clustering method. …….
## Example of Comparisons

<table>
<thead>
<tr>
<th>Selection method</th>
<th>Time-complexity</th>
<th>Space-complexity</th>
<th>Applicability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hierarchic agglomerative (stored matrix algorithm)</td>
<td>$O(N^3)$</td>
<td>$O(N^2)$</td>
<td>Small files only</td>
</tr>
<tr>
<td>Reciprocal nearest neighbour (Ward’s)</td>
<td>$O(N^2)$</td>
<td>$O(N)$</td>
<td>Up to a quarter of million molecules</td>
</tr>
<tr>
<td>Reciprocal nearest neighbour (Jarvis-Patrick)</td>
<td>$O(N^2)$</td>
<td>$O(N)$</td>
<td>Up to more than a million molecules, due to its lower constant proportionality in the time-complexity</td>
</tr>
<tr>
<td>Maximum-dissimilarity</td>
<td>$O(N^3)$</td>
<td>$O(N^2)$</td>
<td>General algorithm implies that it is applicable only to small files. However, $O(N^2)$ time complexity has been described for the MaxMin and MaxSum versions [Holliday et al., 1995; Higgs et al., 1997]. These versions can thus be applied to a million molecules [Higgs et al., 1997].</td>
</tr>
</tbody>
</table>
Review + Experimental Paper
Can Use Empirical Comparison After Dry Review

• Evaluation criteria
• Gold Standard, Benchmark datasets or Development of unique datasets based on criteria
• First, can do Baselines identification
  – Dry comparison based on criteria
  – Selection
• Evaluate alternatives empirically
  – Discuss based on performance criteria (efficiency, effectiveness, ease of use, etc.)
  – Justification, reasoning – look at specific formulation or nature of algorithms, mathematical proving, relate to current findings in the area or other areas
  – Identify weaknesses, gaps that lead to novel technique or fusion or hybrid proposed
Planning for GOT
# PhD Research Schedule

<table>
<thead>
<tr>
<th>No.</th>
<th>ACTIVITIES</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Literature Review</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Problem Formulation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Initial Results</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Proposal Writing (Chapters 1, 2, 3, 4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Objective 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Objective 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Objective 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Thesis Writing</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>No.</th>
<th>MILESTONE</th>
<th>Year 1</th>
<th>Year 2</th>
<th>Year 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Review Paper</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Problem Formulation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Completion of Proposal Writing &amp; First Assessment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Paper with some results</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Completion of Objective 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Completion of Objective 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Completion of Objective 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Thesis Writing Completion/Submission</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Read and Write Throughout PhD

- Ask student to write at least one chapter every semester
- Set-up publication agenda
  - A detailed, realistic, time-bound, publication plan for the research degree, including significant milestones and maintain progress towards its achievement
Publication Agenda

- Concept paper (s) -> Merged concept paper
- LR -> Experimental paper (s) -> Merged experimental paper
- Presentation to group -> paper
- 1 paper every three month
- Appointment by paper
## Sample Publication Agenda

<table>
<thead>
<tr>
<th>Time</th>
<th>What</th>
<th>Where</th>
<th>Sample Inclusion</th>
</tr>
</thead>
</table>
| 1st semester| Critical Analysis of Literature | • Conference  
• Journal | • Framework of analysis  
• New Taxonomy  
• Specific review  
• Hypothesis  
• Pilot study |
|             |                               |                     |                                                                                  |
| 2nd semester| • Concept Paper  
• Empirical Comparison of Techniques  
• Assumption testing  
• Corpus design | • Journal | • Choose a number of performance/selection criteria  
• Select a number of best techniques  
• Critical comparison  
• Identification of possible improvements |

---


### Sample Publication Agenda (cont.)

<table>
<thead>
<tr>
<th>Time</th>
<th>What</th>
<th>Where</th>
<th>Sample Inclusion</th>
</tr>
</thead>
</table>
| 3rd semester | • LR + Suggested framework  
• 1st Objective/Experimental Paper | • Conference (framework)  
• Journal                      | • Introduction  
• Experimental Design  
• Results  
• Discussion  
• Conclusion |
| 4th semester | • 2nd Objective/Experimental Paper     | • Conference  
• Journal (extended dataset) | • Introduction  
• Experimental Design  
• Results  
• Discussion  
• Conclusion |


Sample Publication Agenda (cont.)

<table>
<thead>
<tr>
<th>Time</th>
<th>What</th>
<th>Where</th>
<th>Sample Inclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>5th semester</td>
<td>3rd Objective/Experimental Paper</td>
<td>Conference</td>
<td>Introduction</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Journal</td>
<td>Experimental Design</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(extended dataset)</td>
<td>Results</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Discussion</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Conclusion</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>Paper</th>
<th>journal</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Mentoring in Experimental Design & Data Collection Phase
Do standard things

• Get idea from literature or graduated student. Why?
  – Someone has thought method out carefully
  – Saves time

• Learn what those standard things are (add only to test new ideas)
  – Datasets
  – Methods
  – Evaluation

• Statement must be supported by a reference to the scientific literature or by original work.
Draw Up a Research Framework

<table>
<thead>
<tr>
<th>Phase</th>
<th>Objective</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>To develop an on-line recognition scheme that can perform timely and accurate recognition of CCPs even as they are developing</td>
</tr>
<tr>
<td>2</td>
<td>To develop improved recognisers that can perform accurate classification of partially developed CCPs. In particular, this research focuses on improving input representation and design of the ANN-based recognisers.</td>
</tr>
</tbody>
</table>

Note: Also refer to Table 3.1 for further details.
## Draw Up a Research Framework

<table>
<thead>
<tr>
<th>Phase</th>
<th>Activities</th>
<th>Resources Needed</th>
<th>Benchmark Data</th>
<th>Baselines for Comparison</th>
<th>Performance Evaluation</th>
<th>Objectives Addressed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Diagram
- Table
- Description
- Gantt Chart
Performance Measurement

• **Analytical analysis**
  – will not give the final answers but help understand the concept
  – Eg: proof of validity of the major idea, rough estimation of the performance, rough estimation of the complexity, calculation of initial values for simulation analysis to follow,

• **Simulational analysis**
  – Use simulation

• **Implementational analysis**
  – Actual implementation
Performance Measures

Figure 3.1: Performance measures used in the investigations
Coaching in Analysis of Results and Writing Up
Make an outline

• Think about the plan of chapters and decide what is best to report work.

• Then make a list, in point form, of what will go in each chapter.

• Try to make this rather detailed, ending up with a list of points that corresponds to subsections or even to the paragraphs of thesis.

• If planning each chapter and section before sitting down to write, the result will probably be clearer and easier to read. It will also be easier to write.
Order of Writing

• Draft outline and write literature review section
• Write up a preliminary version of the background section
• Draw up research questions and objectives
• Identify scope of study and (expected contribution)
• Write up the methods section.
  – include a description of the research equipment and relevant calibration plots.
• When you have some data, start making plots and tables of the data.
  – help you to visualize the data and to see gaps in your data collection
  – If time permits, you should go back and fill in the gaps.
  – You are finished when you have a set of plots that show a definite trend (or lack of a trend).
  – Be sure to make adequate statistical tests of your results.
Order of Writing (cont.)

• Once you have a complete set of plots and statistical tests, arrange the plots and tables in a logical order.
  – Write figure captions for the plots and tables.

• Once your plots and tables are complete, write the results section.
  – describe your results, but you must NOT interpret them. Be factual and orderly.

• Once you have written the results section, you can move on to the discussion section
  – talk about your ideas about the data
Order of Writing (cont.)

• In writing the discussion session:
  – explain your results, find reasons, derive conclusions
  – discuss the work of other authors who collected data on the same or related scientific questions.
  – discuss how their work is relevant to your work. If there were flaws in their methodology, this is the place to discuss it.

• After you have discussed the data, you can write the conclusions section.
  – take the ideas that were mentioned in the discussion section and try to come to some closure.
  – If some hypothesis can be ruled out as a result of your work, say so.
  – If more work is needed for a definitive answer, say that.
Order of Writing (cont.)

• Finally, write your recommendation
  – Make recommendations for further research or policy actions
  – If you can make predictions about what will be found if X is true, then do so.

• After you have finished the recommendation section, look back at your original introduction.
  – Your introduction should set the stage for the conclusions by laying out the ideas that you will test

• Write your abstract last.
Results Chapter

• Ask student to draft figures/tables first
• Make captions for every figure and table
• Explain figures and tables
• Discuss and interpret results
• Compare results with previous works
Results vs. Discussion Sections

• Quarantine observations from interpretations.
  – physically separate statements about observations from statements about the meaning or significance of those observations.
Discussion

Start with a few sentences that summarize the most important results. The discussion section should be a brief essay in itself, answering the following questions and caveats:

• What are the major **patterns** in the observations? (Refer to spatial and temporal variations.)
• What are the **relationships, trends and generalizations** among the results?
• What are the **exceptions** to these patterns or generalizations?
• What are the likely **causes** (mechanisms) underlying these patterns resulting predictions?
• Is there **agreement or disagreement with previous work**?
• Interpret results in terms of background laid out in the introduction - what is the **relationship of the present results to the original question**?
• What is the **implication of the present results for other unanswered questions**?
Summary

• Prepare before start of PhD
• Plan for GOT early in the course of study
• Ask student to commit for GOT
• Set up conducive environment for GOT
• Choose a topic that suits student best
• Write early, from the beginning and polish for coherent, smooth flow at the end
• Maintain effective, regular supervision
Thank you