MENTORING FOR "GRADUATE-ON-TIME (GOT)"

Professor Dr. Naomie Salim

Faculty of Computing

Universiti Teknologi Malaysia

PROGRAM GOT & U

Outline of Talk

- The challenges to GOT
- Mentoring student to manage for GOT
- Mentoring for GOT throughout the PhD research life-cycle
  - Preparation phase
  - Problem definition phase
  - Literature review phase
  - Methodology design phase
  - Experiment, data collection, analysis and writing-up phase

PhD Training to Develop Competent Researchers

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.
- 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

Style	Advisor Role and Behavior	Student Role and Behavior
Strong master/apprentice style	Advisor is master. Advisor has a well specified domain of expertise and set of problems within it.	Student is an apprentice working for the advisor. Student works on advisor's problems.
Collegial master/apprentice style	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.	Student develops a problem within advisor's domain and skills and works under the advisor to develop the research plan and procedures.
Collegial development style	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.	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.
Guidance and suggestion style	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.	Student is an independent, junior colleague who takes initiative for presenting problems and research plans for discussion and guidance. Student develops required skills.
Passive hands-off style	Advisor has quality control role and responds only to requests or documents and performs only general quality control review	Student is an independent researcher who takes initiative for developing problem, developing skills, and presenting research plans for general review and approval.

Advantages & Disadvantages of Styles

Advising Style	Advantages	Disadvantages
Strong master/apprentice style	Advisor is heavily involved and gives expert direction for research activities. Student learns how to do research within advisor's domain.	Student works on advisor's problems and within advisor expertise and may not develop independence.
Collegial master/apprentice style	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.	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.
Collegial development style	Advisor and student develop together to explore new domain and new research methods. Student develops independence within relationship.	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.
Guidance and suggestion style	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.	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.
Passive hands-off style	Student is able to act independently with little interference from advisor. Student can work on problems of his or her choosing.	Student may make serious mistakes because of lack of advice and suggestions. Student may flounder and not complete on timely basis.

Supervision Style and Fit

Advising Style	Likely Fit with Student Archetypes
Strong master/apprentice style	Good fit with immature, unconfident student. "Do what I do" provides clear direction. Possible good fit with somewhat mature, somewhat confident student, but the style creates some stress between heavy direction of advisor and initiative and independence of somewhat mature student. Likely stress and conflict between this style and the independence of a very mature, confident student
Collegial master/apprentice style	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.
Collegial development style	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.
Guidance and suggestion style	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.
Passive hands-off style	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.

What can take most time ?

What can take most time?

GOT

- Time to formulate problem
- Inadequate background
- Lack of research skills
- Problems in writing
- Problems in data/equipment
- Too many distractions

Mentoring student to manage life

- Learn to super-skim
- Managing and tracking of time
- Note taking
- Log book
- Organization of literature
- Immersion in writing
- Enlisting support, networking and exploiting synergies
- Keeping research narrow
- Prioritization

Helping Students Prepare for a GOT PhD Journey

PROGRAM GOT & U

Successful, Timely Completion of PhD

Help in Analysis and Writing

Coach for Mastery in Research Methods

Planning for GOT

Coach for Deep & Critical Comprehension of Subject Matter

Help Identify Interesting & Suitable Research Topic

Prepare Researcher Mindset

Set-up Supportive Research Environment

SOARING UPWARDS

KEMENTERIAN PENDIDIKAN TINGGI

UNIVERSITI TEKNOLOGI MARA

MYDEGS

PROGRAM GOT & U

Setting-Up of a Supportive Research Environment

Don’t Slow Your Student because of Unconducive Research Environment

SOARING UPWARDS

KEMENTERIAN PENDIDIKAN TINGGI

UNIVERSITI TEKNOLOGI MARA

MYDEGS

Setting Up A Supportive Research Environment

- Have a proposal bank, with supporting literature
- 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

Setting Up A Supportive Research Environment (cont.)

- Schedule regular presentations to group – to be developed into publication
- Monitor progress 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)

Mindset Preparation

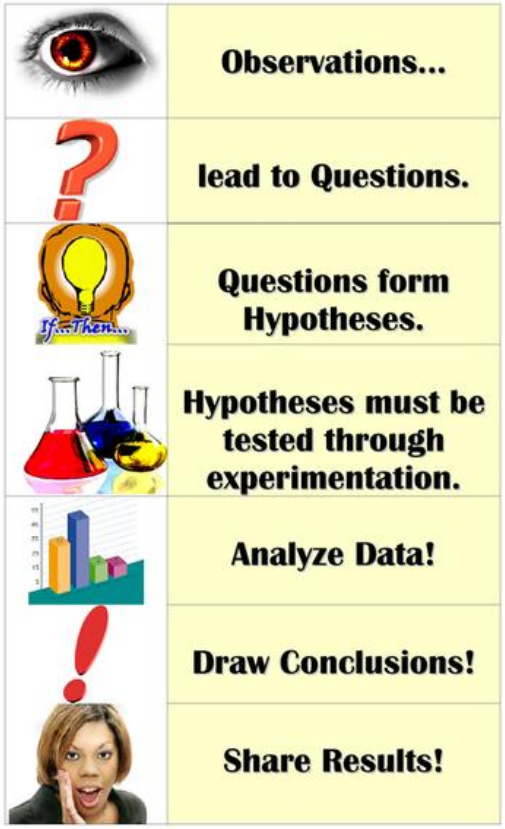
Wrong mindset, and it takes forever to finish PhD

Mindset Preparation (and development)

- 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?

PROGRAM GOT & U

Helping Student Find Good Research Topic

Never underestimate the importance of choosing the right topic

SOARING UPWARDS

UNIVERSITI TEKNOLOGI MARA

MYDEGS

Finding A Good Research Problem

- Novelty of the Idea.

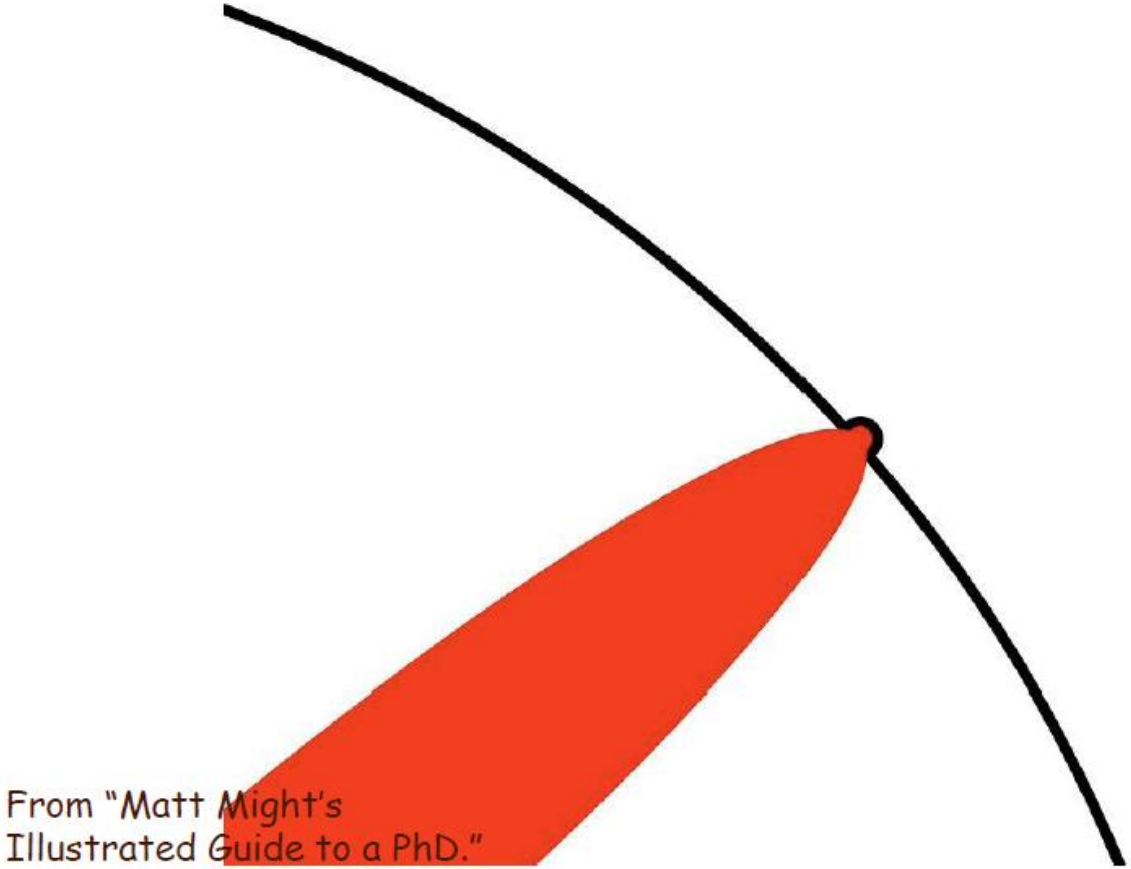
  Research is a study of new ideas in the field
- Significance for the Community.

  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.

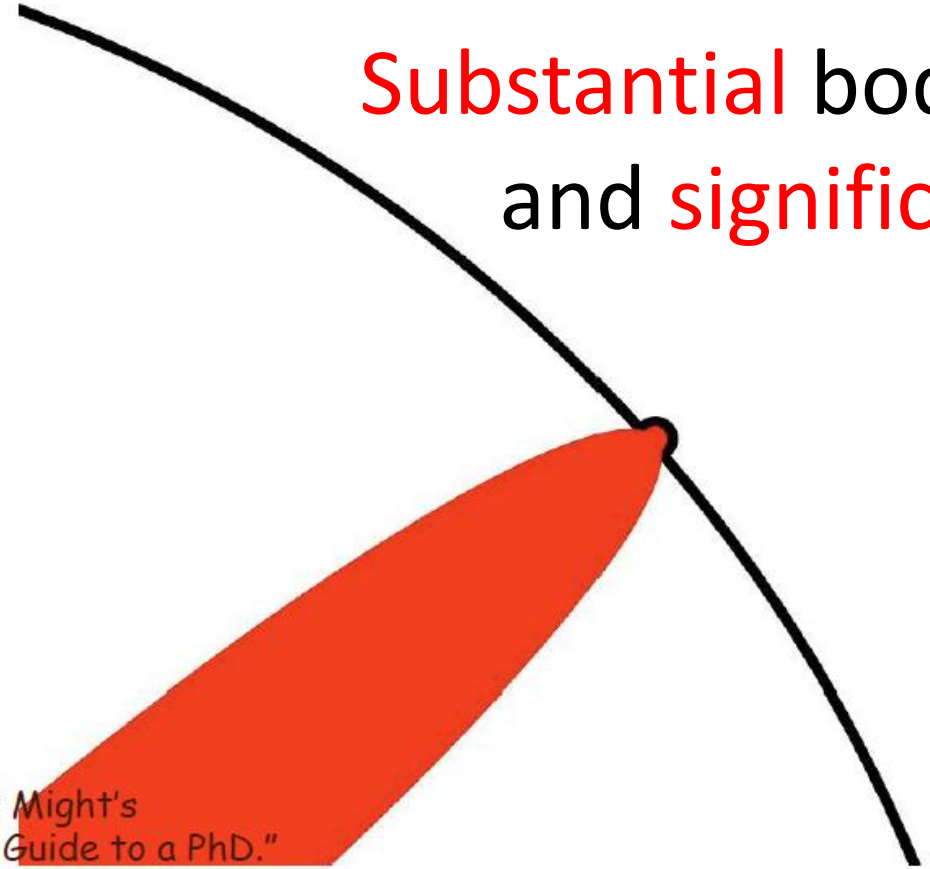
Novelty:

What to Expect of a PhD Research



What to Expect of a PhD Research

Substantial body of original and significant work

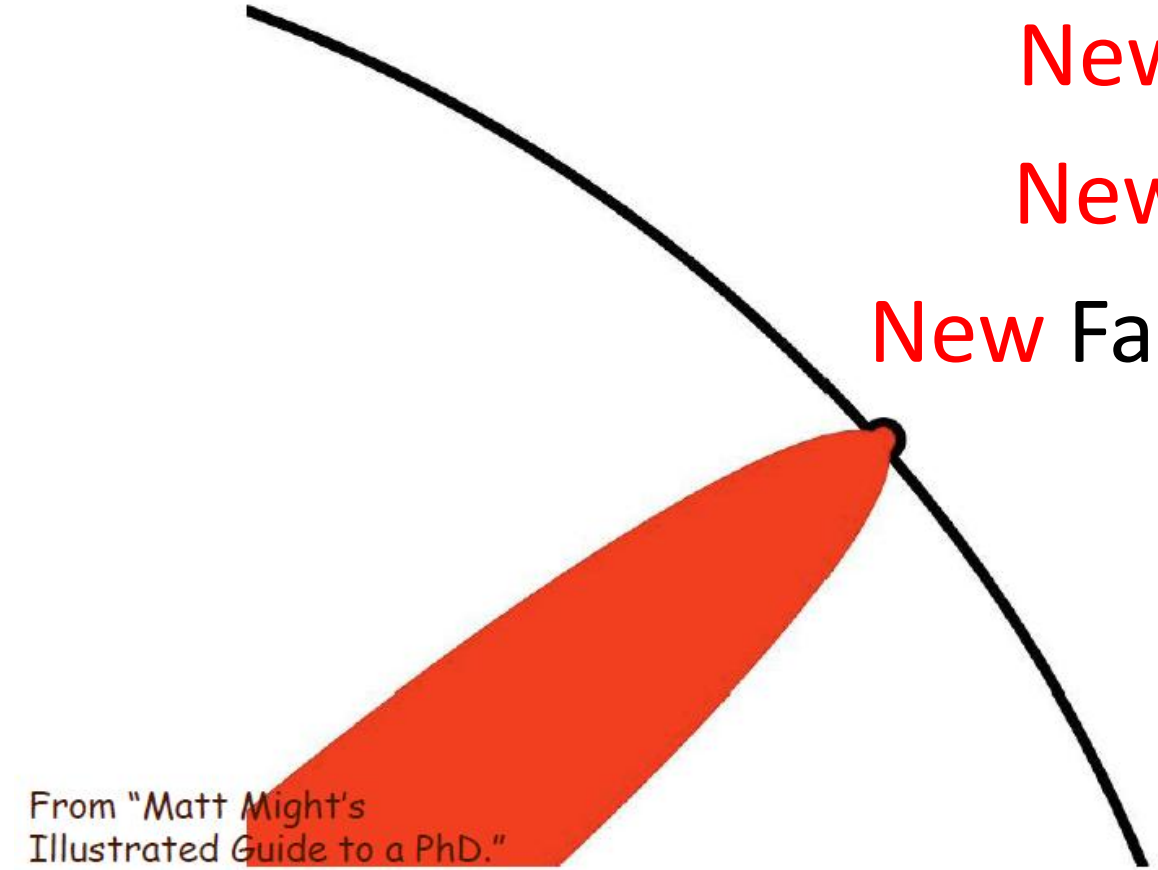


From "Matt Might's Illustrated Guide to a PhD."

What to Expect of a PhD Research

Winning a Nobel Price through PhD Research is not Neccessary

Expectation of a PhD Research

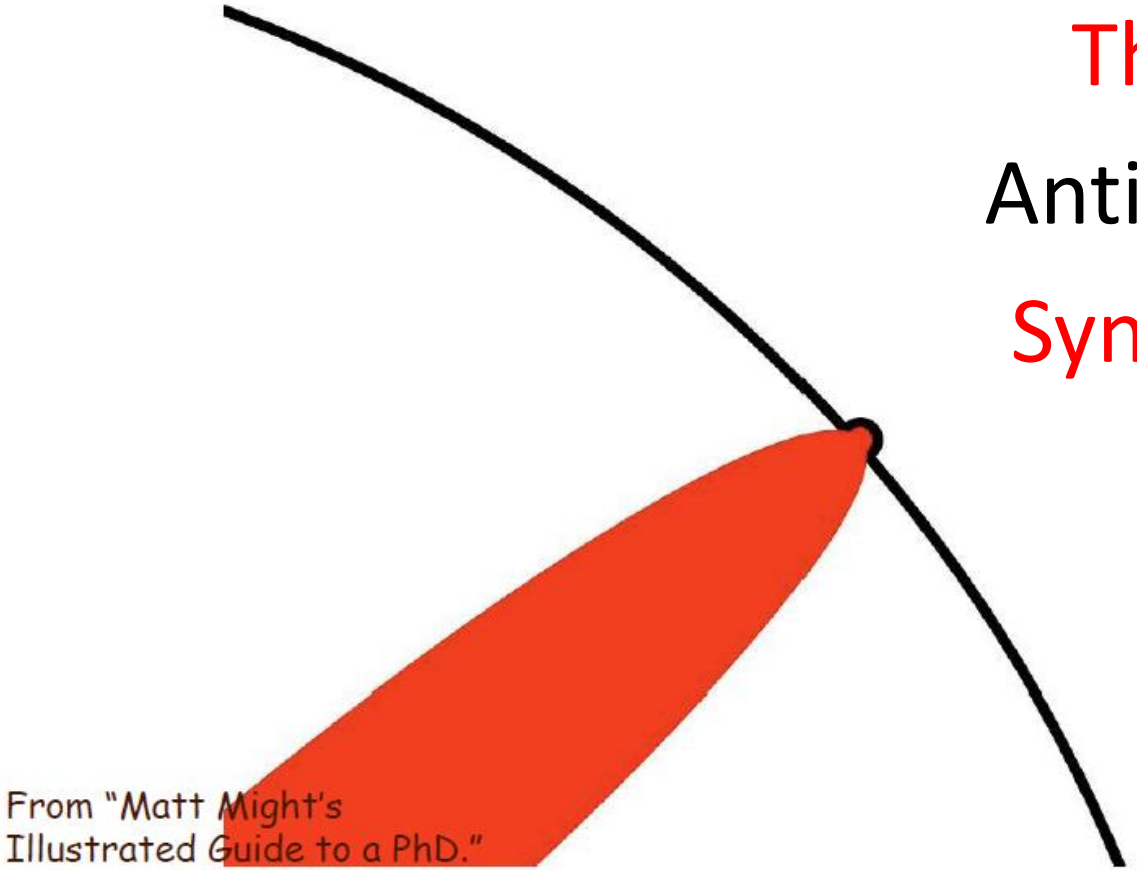


New Facts

New Ideas

New Facts + Ideas

Expectation of a PhD Research



Thesis

Anti-thesis

Synthesis

Example - Identifying Gaps in Information Retrieval Research

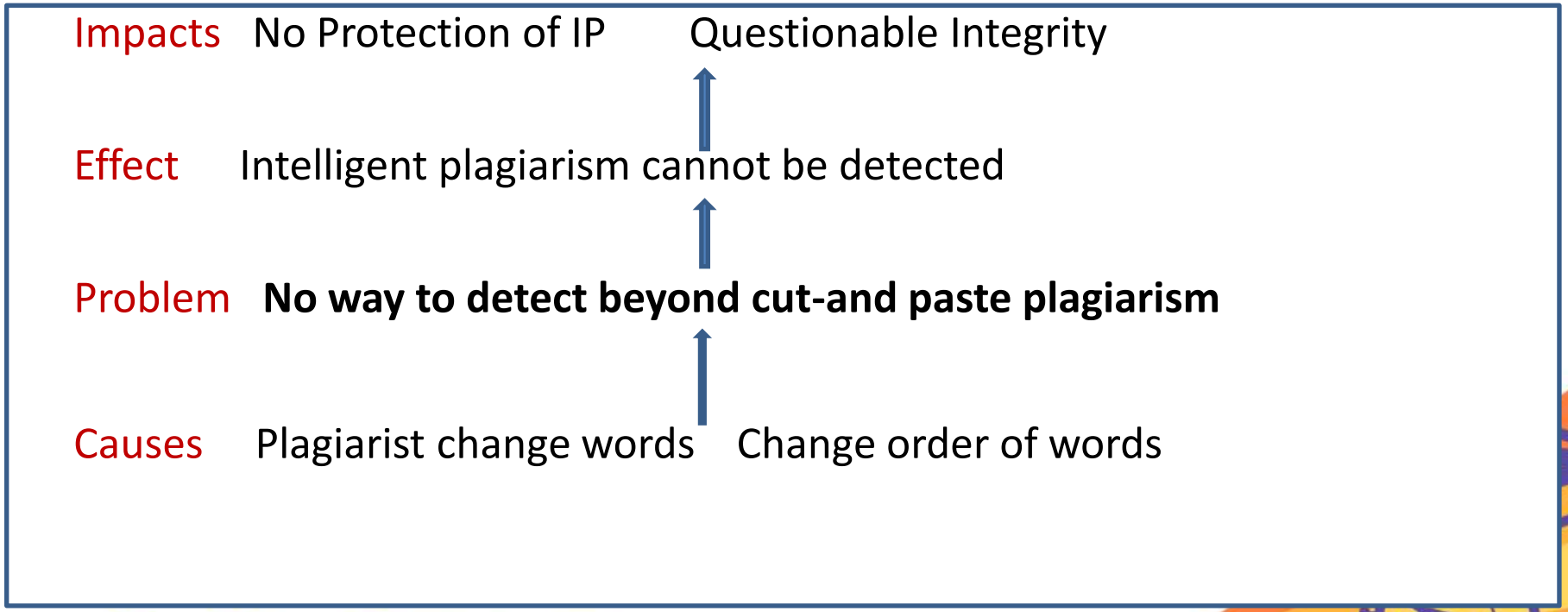
- Other data that can be used to enhance?
  - Eg. Explicit vs implicit, multimedia
- Other ways to represent data?
  - Eg. Graph? Passages?
- Challenge assumptions and rules.
  - Eg.: crisp vs fuzzy? Deterministic vs probabilistics?
- Other external knowledge bases or sources to enhance?
- Combination? Optimization? Weightage?
- Adapting ideas form other fields? Eg. Diversity analysis, Cross-structural theory, Game theory, etc.

Significance to Community

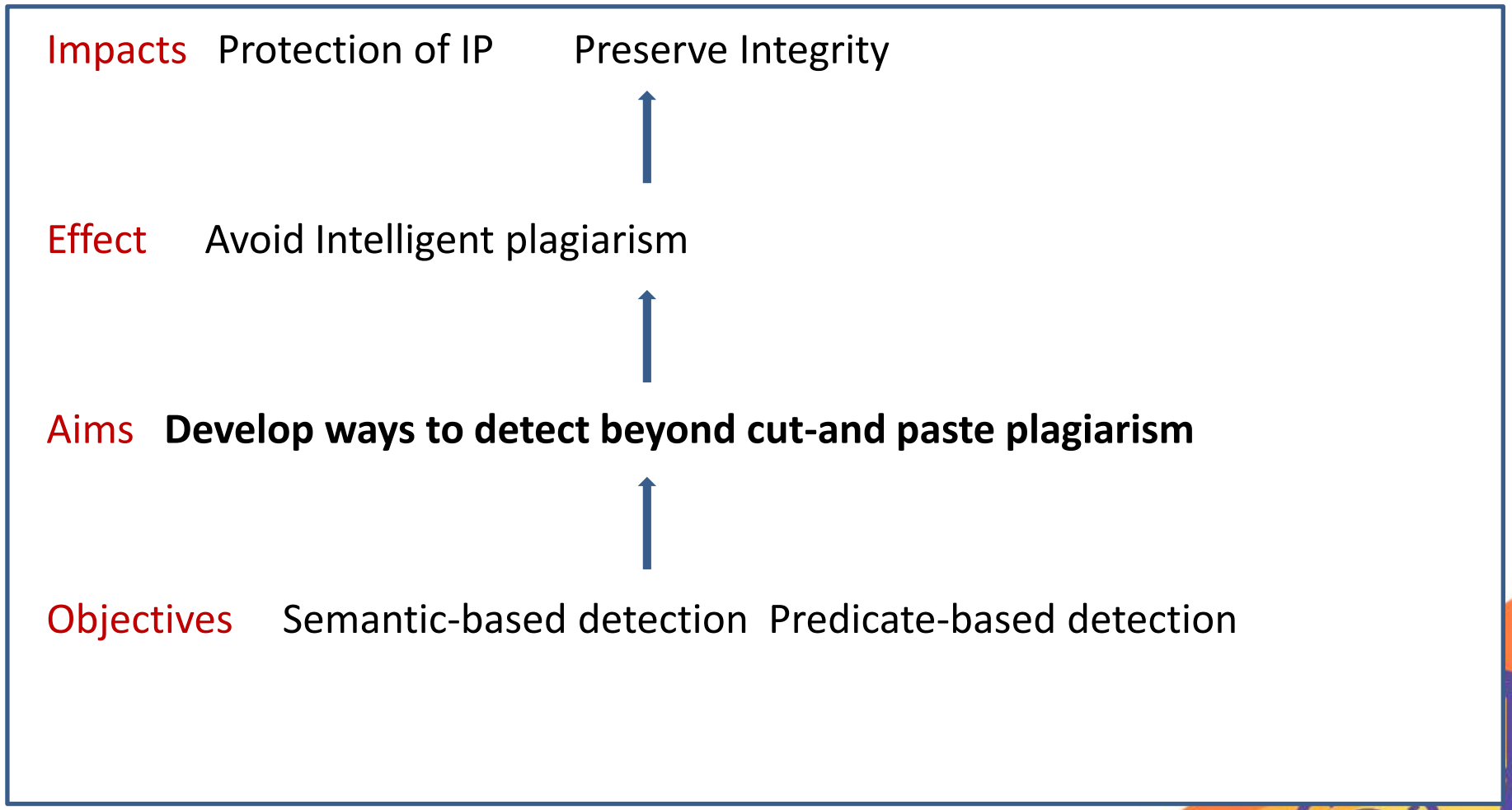
- What is the Need?
- What is your Approach to address the need?
- What is the Benefit of your approach over cost?
- In what ways are your research better than the Competitors?

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

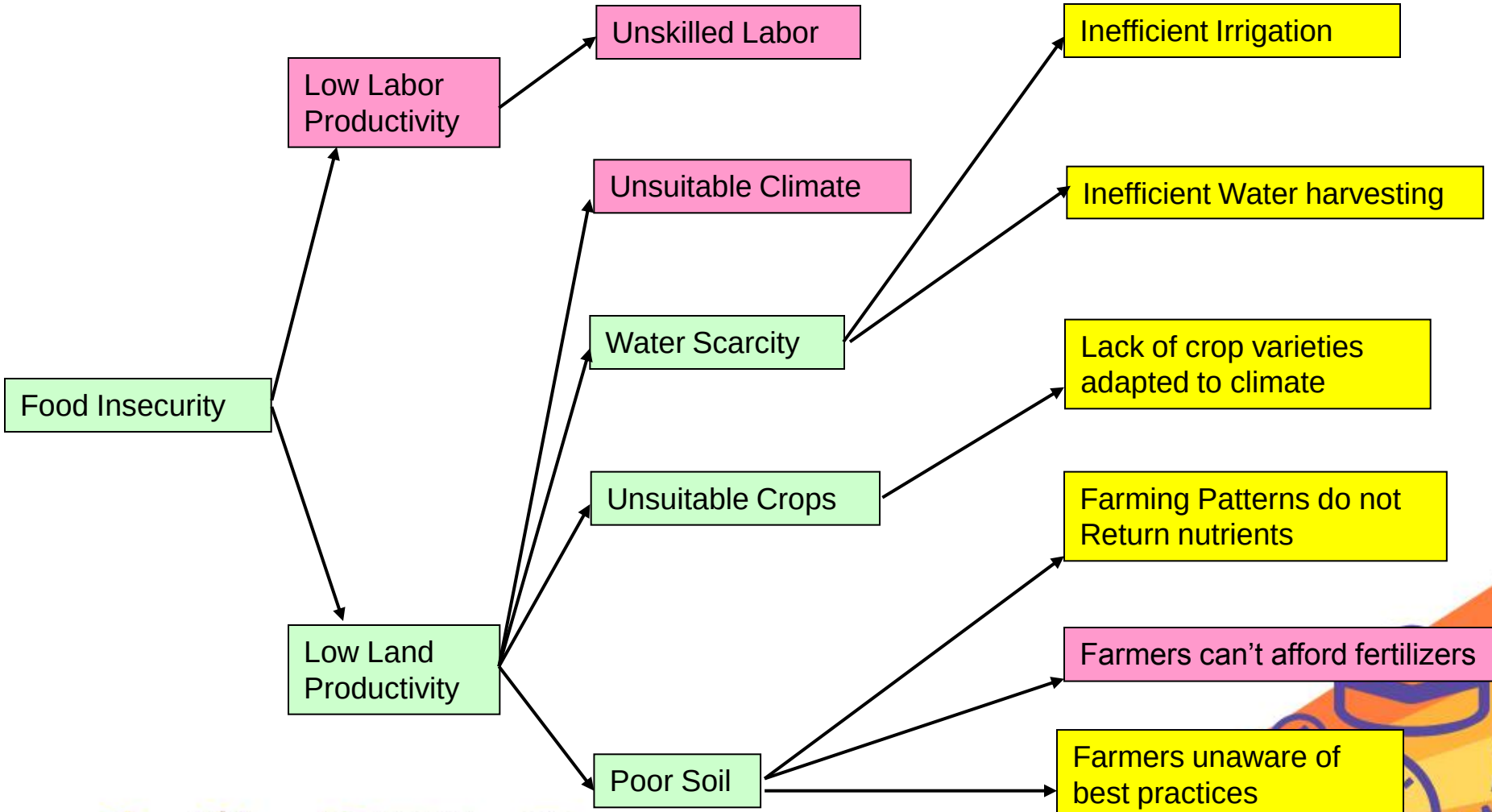


Change Problem Tree to Objective Tree



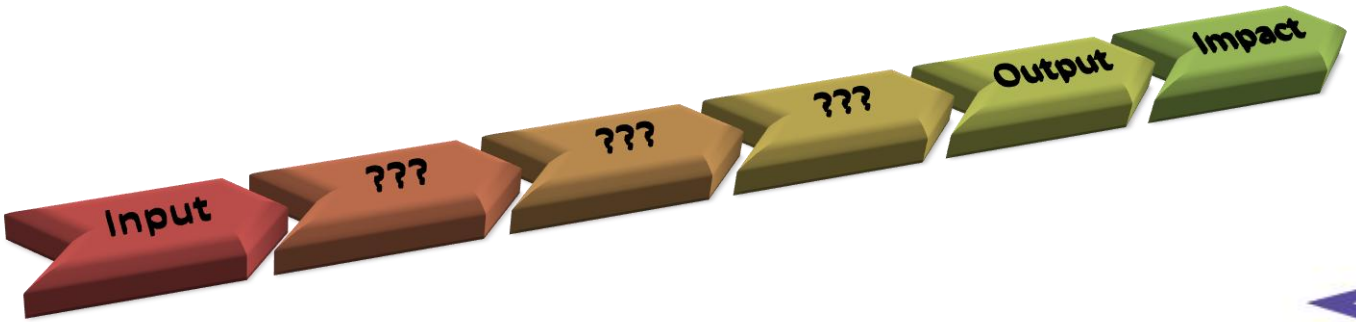
Problem Tree – Keep asking Why?

(SA Prathapar, Research Methodology Slide, 2012)



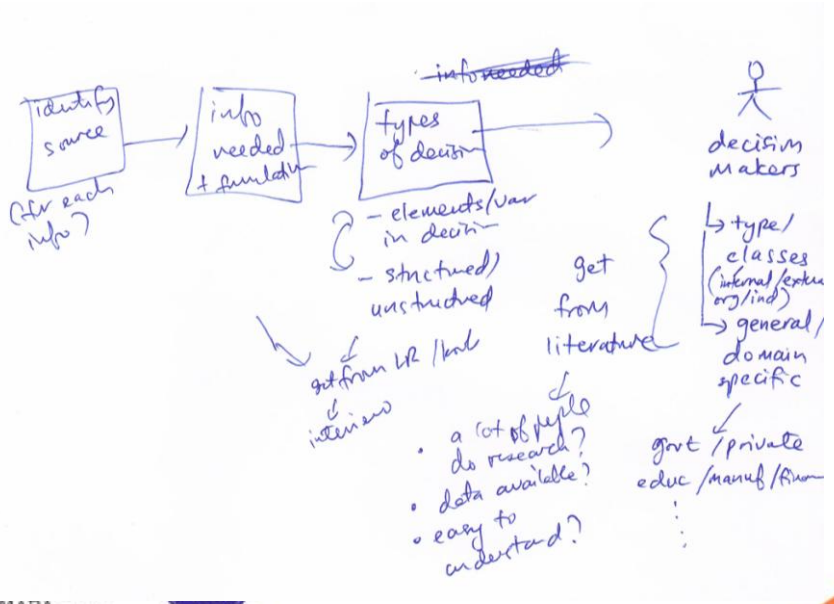
Horizontal Analysis

- What impact?
- What output?
- What input?
- The building blocks in between?



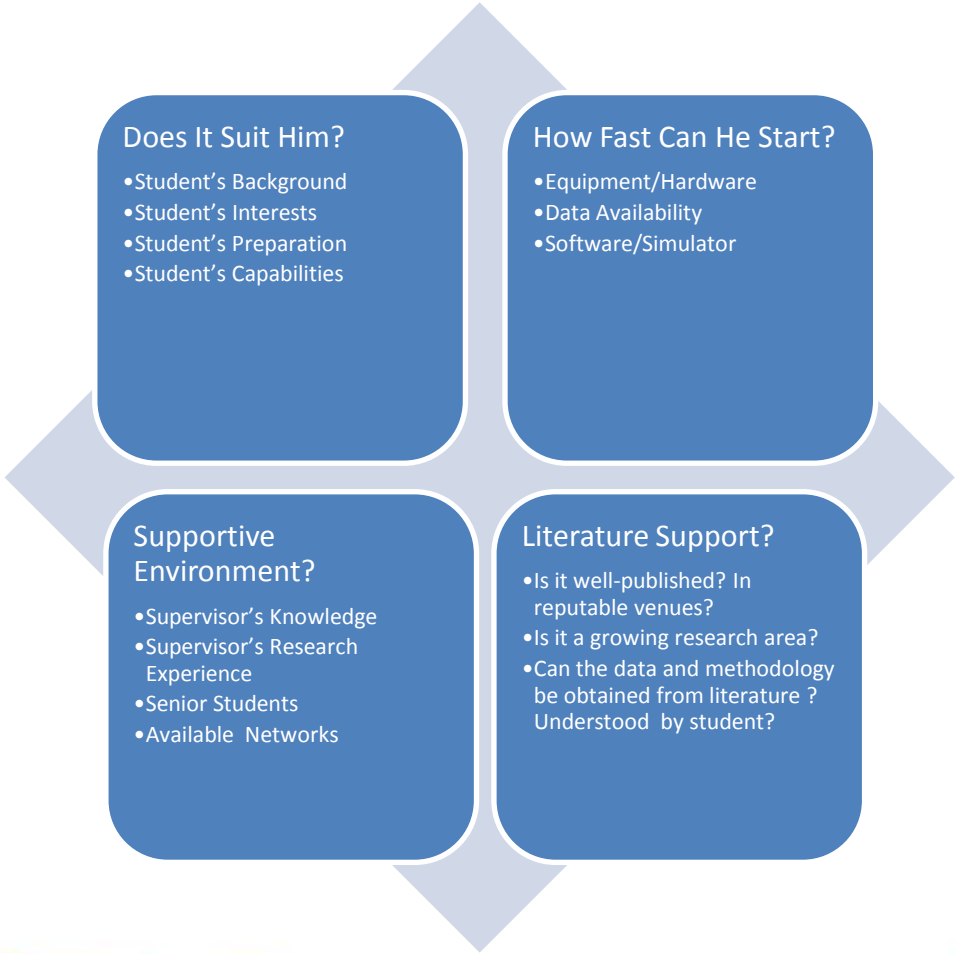
Example of Horizontal Analysis

- Student interested to do text mining to support decision making
- Semantic Question-Answering for Business Intelligence



Topic Identification

– Can he finish it on time?



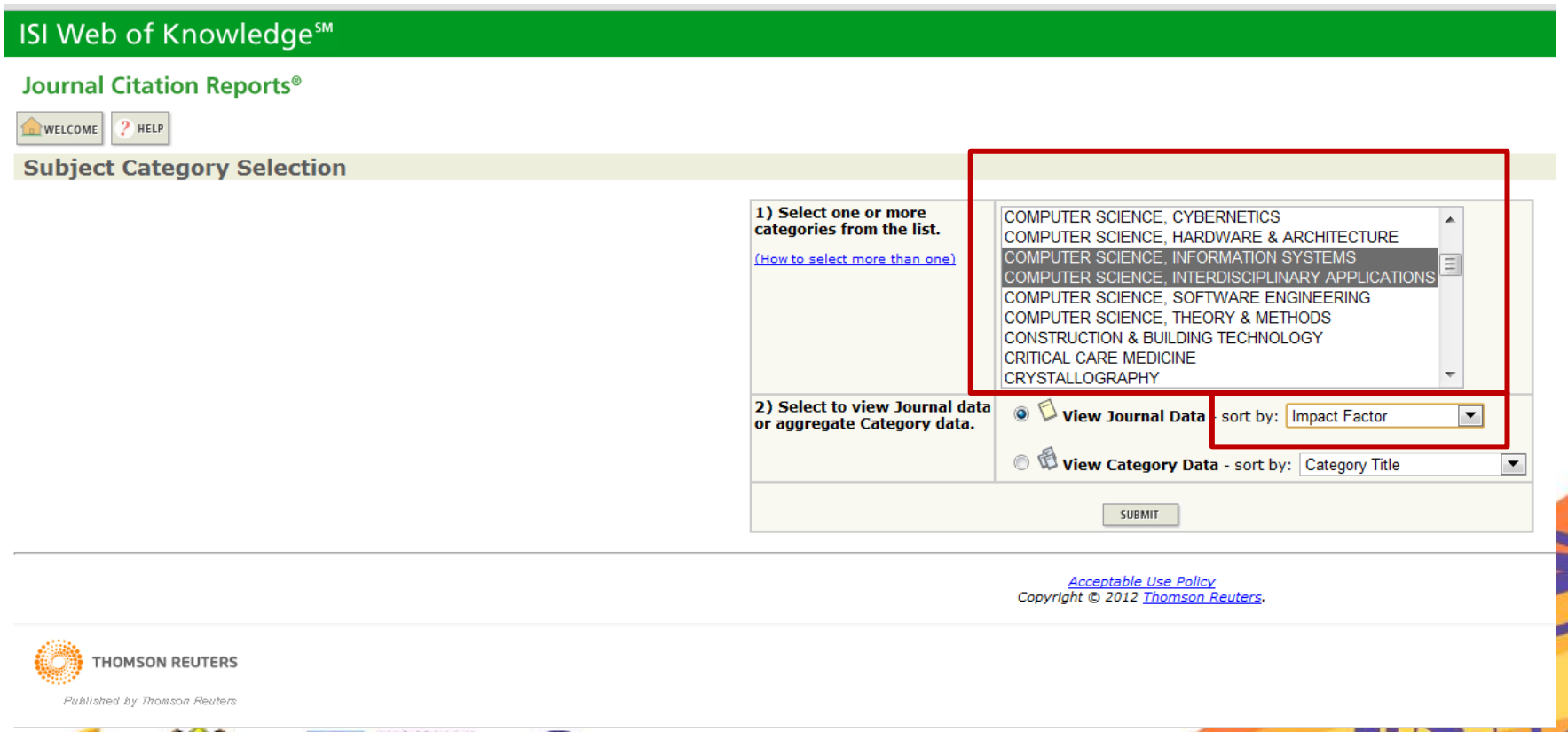
In case he is not prepared at all…

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

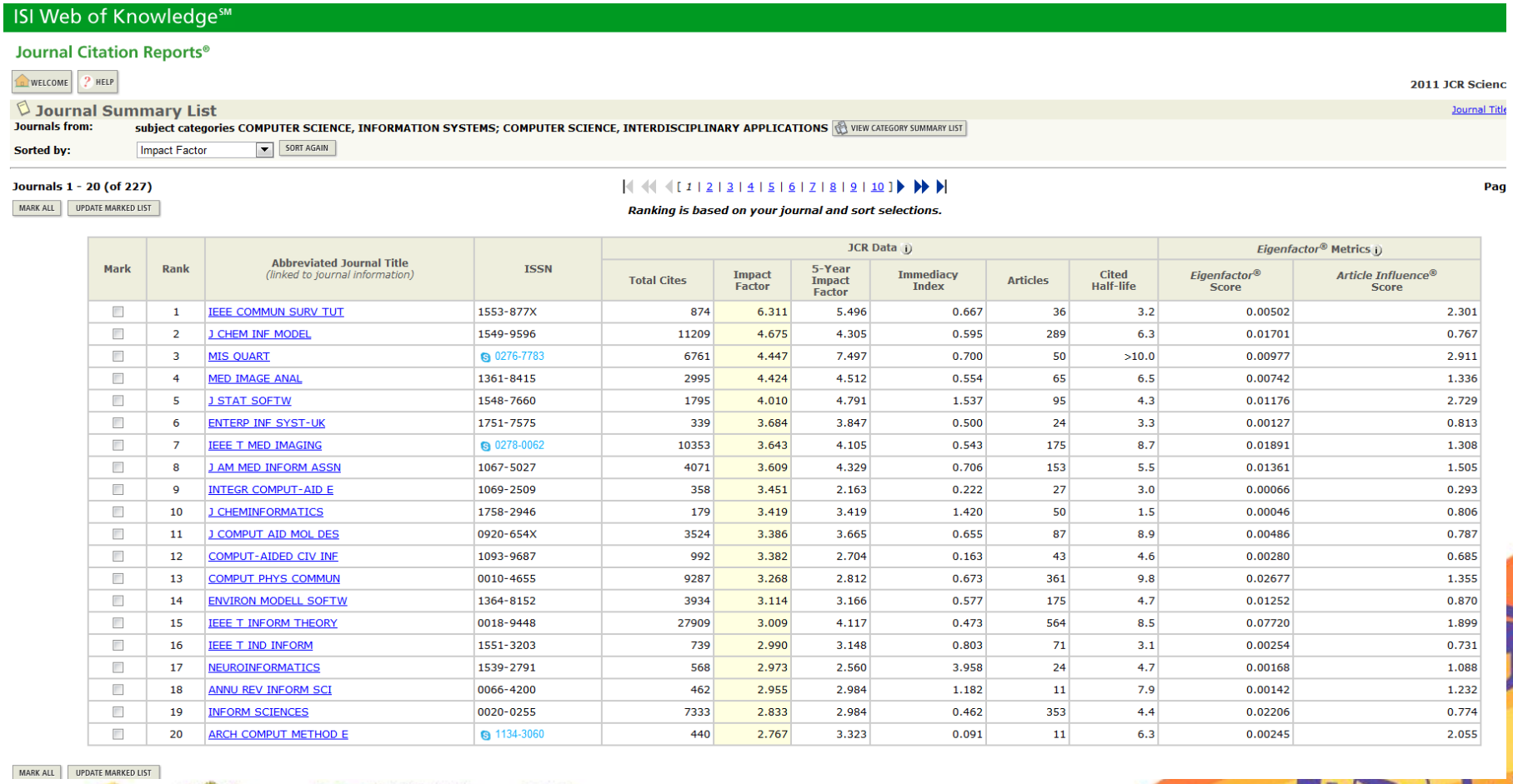
Who are the active researchers in the area?

Guide him to read …

Have a feel of good research and find his interest in good, reputable journals



Finding out about best ranked journals (cont.)



ISI Web of Knowledge℠

Journal Citation Reports®

WELCOME | HELP

2011 JCR Scienc

Journal Summary List

Journal Title

Journals from: subject categories COMPUTER SCIENCE, INFORMATION SYSTEMS; COMPUTER SCIENCE, INTERDISCIPLINARY APPLICATIONS | VIEW CATEGORY SUMMARY LIST

Sorted by: Impact Factor | SORT AGAIN

Journals 1 - 20 (of 227)

[1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10]

Pag

MARK ALL | UPDATE MARKED LIST

Ranking is based on your journal and sort selections.

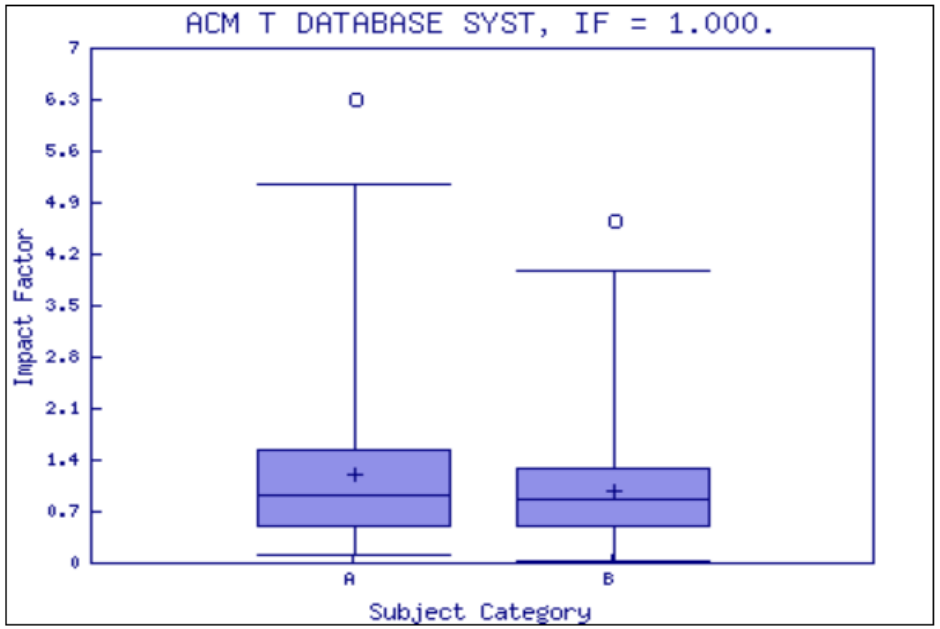
Mark	Rank	Abbreviated Journal Title (linked to journal information)	ISSN	JCR Data						*Eigenfactor*® Metrics	
				Total Cites	Impact Factor	5-Year Impact Factor	Immediacy Index	Articles	Cited Half-life	*Eigenfactor*® Score	*Article Influence*® Score
	1	IEEE COMMUN SURV TUT	1553-877X	874	6.311	5.496	0.667	36	3.2	0.00502	2.301
	2	J CHEM INF MODEL	1549-9596	11209	4.675	4.305	0.595	289	6.3	0.01701	0.767
	3	MIS QUART	0276-7783	6761	4.447	7.497	0.700	50	>10.0	0.00977	2.911
	4	MED IMAGE ANAL	1361-8415	2995	4.424	4.512	0.554	65	6.5	0.00742	1.336
	5	J STAT SOFTW	1548-7660	1795	4.010	4.791	1.537	95	4.3	0.01176	2.729
	6	ENTERP INF SYST-UK	1751-7575	339	3.684	3.847	0.500	24	3.3	0.00127	0.813
	7	IEEE T MED IMAGING	0278-0062	10353	3.643	4.105	0.543	175	8.7	0.01891	1.308
	8	J AM MED INFORM ASSN	1067-5027	4071	3.609	4.329	0.706	153	5.5	0.01361	1.505
	9	INTEGR COMPUT-AID E	1069-2509	358	3.451	2.163	0.222	27	3.0	0.00066	0.293
	10	J CHEMINFORMATICS	1758-2946	179	3.419	3.419	1.420	50	1.5	0.00046	0.806
	11	J COMPUT AID MOL DES	0920-654X	3524	3.386	3.665	0.655	87	8.9	0.00486	0.787
	12	COMPUT-AIDED CIV INF	1093-9687	992	3.382	2.704	0.163	43	4.6	0.00280	0.685
	13	COMPUT PHYS COMMUN	0010-4655	9287	3.268	2.812	0.673	361	9.8	0.02677	1.355
	14	ENVIRON MODELL SOFTW	1364-8152	3934	3.114	3.166	0.577	175	4.7	0.01252	0.870
	15	IEEE T INFORM THEORY	0018-9448	27909	3.009	4.117	0.473	564	8.5	0.07720	1.899
	16	IEEE T IND INFORM	1551-3203	739	2.990	3.148	0.803	71	3.1	0.00254	0.731
	17	NEUROINFORMATICS	1539-2791	568	2.973	2.560	3.958	24	4.7	0.00168	1.088
	18	ANNU REV INFORM SCI	0066-4200	462	2.955	2.984	1.182	11	7.9	0.00142	1.232
	19	INFORM SCIENCES	0020-0255	7333	2.833	2.984	0.462	353	4.4	0.02206	0.774
	20	ARCH COMPUT METHOD E	1134-3060	440	2.767	3.323	0.091	11	6.3	0.00245	2.055

MARK ALL | UPDATE MARKED LIST

Journal Ranking

For **2011**, the journal **ACM TRANSACTIONS ON DATABASE SYSTEMS** has an Impact Factor of **1.000**.

This is a box plot of the subject category or categories to which the journal has been assigned. It provides information about the distribution of journals based on Impact Factor values. median, 25th and 75th percentiles, and the extreme values of the distribution.



Key

A - COMPUTER SCIENCE, INFORMATION SYSTEMS
B - COMPUTER SCIENCE, SOFTWARE ENGINEERING

Scopus - evaluated by SciMago Journal Rank (SJR)*

www.scimagojr.com/journalrank.php?category=1706&area=1700&year=2012&country=&order=sjr&min=0&min_type=cd

SJR SCImago Journal & Country Rank

EST MODUS IN REBUS
Horatio (Satire 1,1,106)

- Home
- Journal Rankings
- Journal Search
- Country Rankings
- Country Search
- Compare
- Map Generator
- Help
- About Us

Journal Rankings

Ranking Parameters

Subject Area: Computer Science
Subject Category: Computer Science Applications
Country: All
Year: 2012
Order By: SJR
Display journals with at least: 0 Citable Docs. (3 years)
Refresh

Subject Area: **Computer Science.**
Subject Category: **Computer Science Applications.**
Year: **2012.**

Download data in MS Excel format (50 Kb)

How to cite this website?

1 - 50 of 230 << First | < Previous | Next > | Last >>

	Title	SJR	H index	Total Docs. (2012)	Total Docs. (3years)	Total Refs.	Total Cites (3years)	Citable Docs. (3years)	Cites / Doc. (2years)	Ref. / Doc.	Country
1	Bioinformatics	Q1 4,223	204	747	2.251	17.874	14.933	2.160	5,38	23,93	
2	Argument and Computation	Q1 3,173	5	8	17	377	68	17	4,00	47,13	
3	Journal of Field Robotics	Q1 3,161	46	48	148	1.610	492	134	3,39	33,54	
4	Journal of Informetrics	Q1 3,075	26	78	172	2.242	787	158	4,80	28,74	

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

SJR is developed by:

Analyze trend

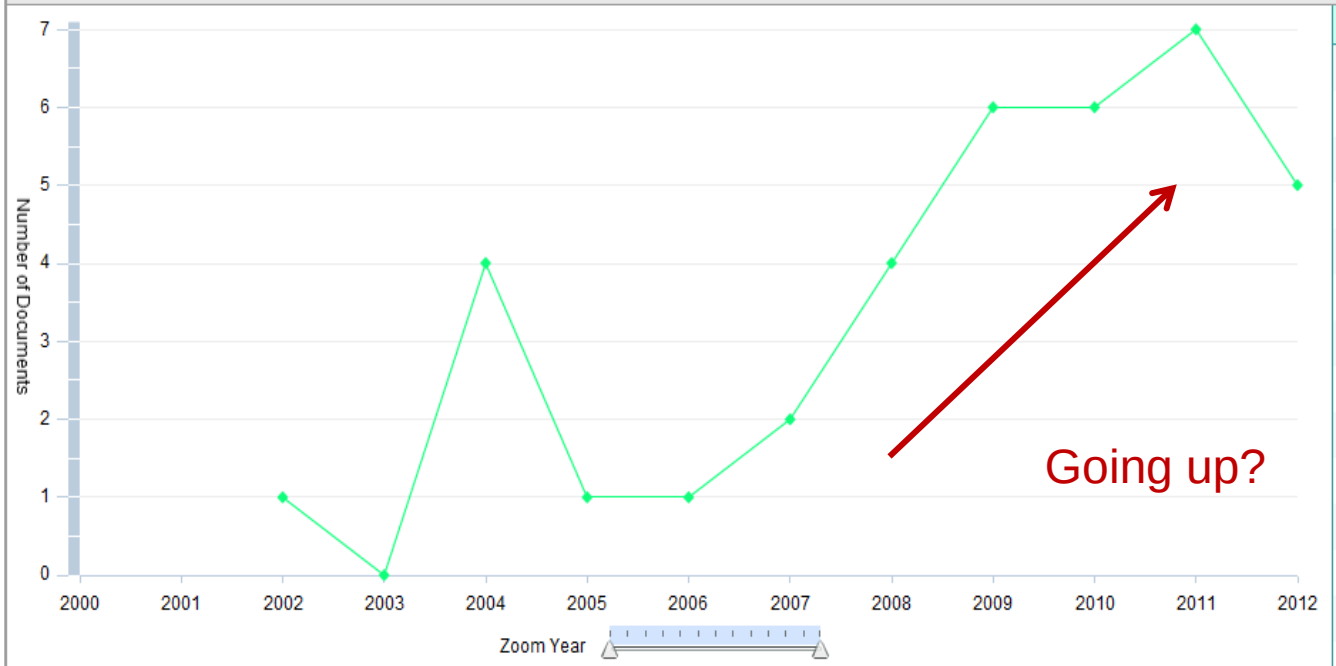
Analyze results | Back to results

Date range 2000 to 2012 Analyze Document results 37

Year | Source title | Author name | Affiliation name | Country | Document type | Subject area

Export | Print | E-mail

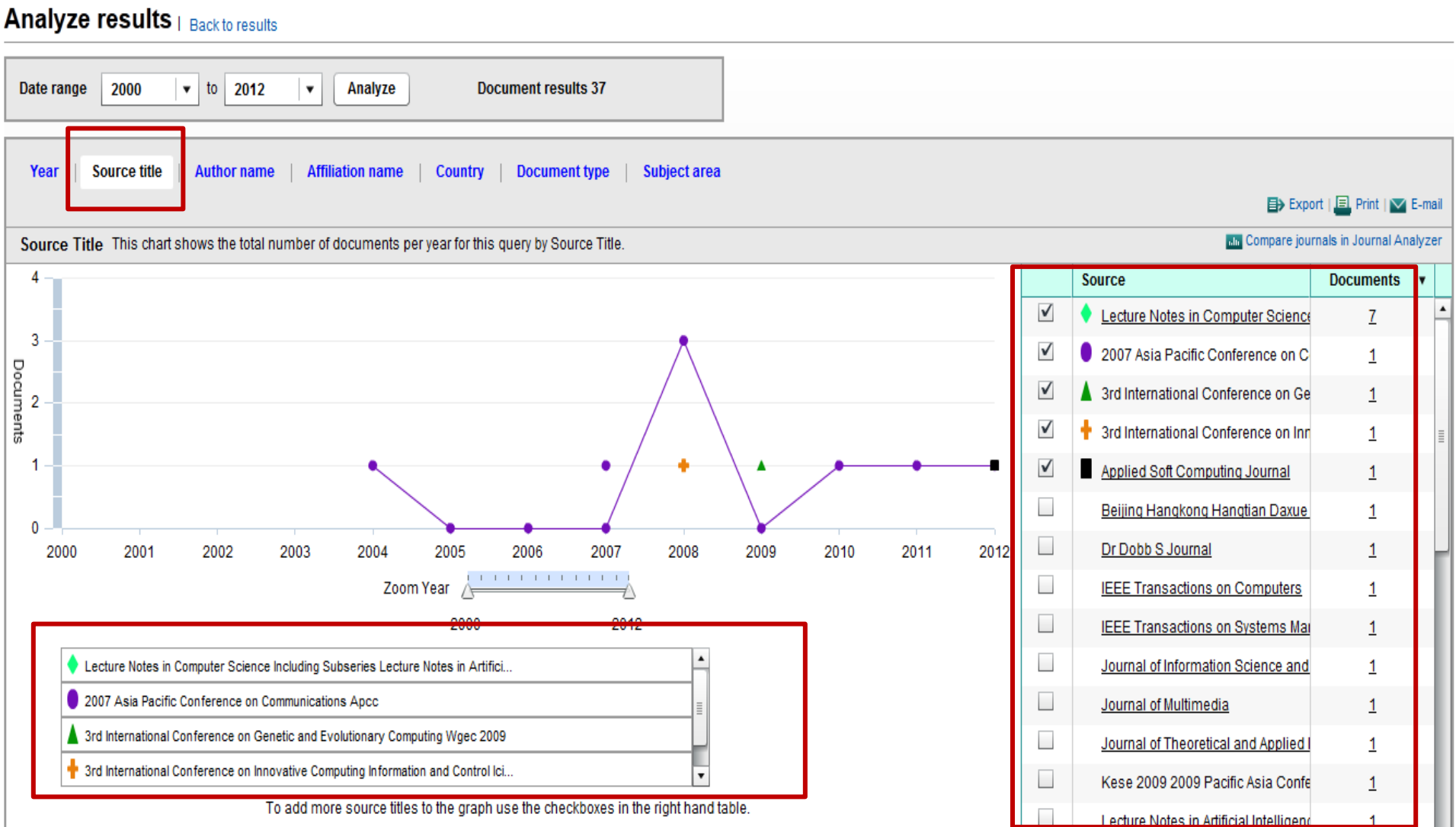
Years This chart shows the total number of documents for this query by Year.



Year	Documents
2012	5
2011	7
2010	6
2009	6
2008	4
2007	2
2006	1
2005	1
2004	4
2002	1

PROGRAM GOT & U

Conference & Journal for Area



People who work in same area

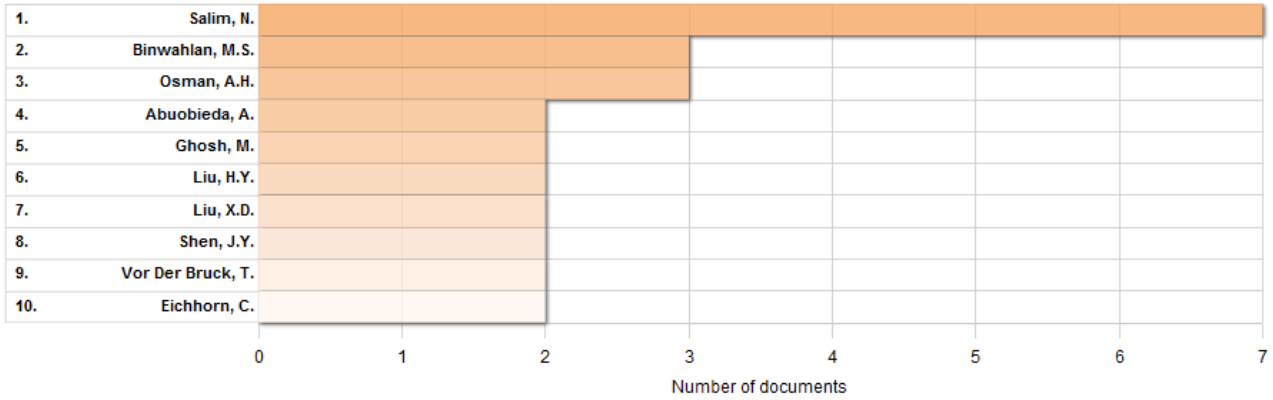
Analyze results | Back to results

Date range 2000 to 2012 Analyze Document results 37

Year | Source title | Author name | Affiliation name | Country | Document type | Subject area

Export | Print | E-m

Author Name This chart shows the total number of documents for this query by Author.



	Author	Documents
☑	Salim, N.	7
☑	Binwahlan, M.S.	3
☑	Osman, A.H.	3
☑	Abuobieda, A.	2
☑	Ghosh, M.	2
☑	Liu, H.Y.	2
☑	Liu, X.D.	2
☑	Shen, J.Y.	2
☑	Vor Der Bruck, T.	2
☑	Eichhorn, C.	2
☐	Bao, J.P.	2
☐	Abraham, A.	2
☐	Zhang, X.D.	2
☐	Hartrumpf, S.	2

To add more authors to the graph use the checkboxes in the list on the right.

Institution Doing Active Research in Area

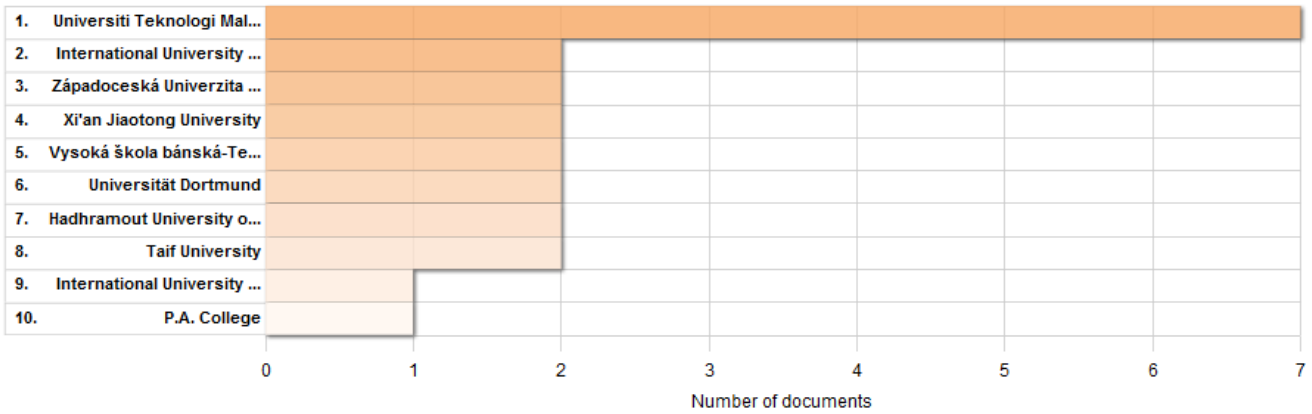
Analyze results | Back to results

Date range 2000 to 2012 Analyze Document results 37

Year | Source title | Author name | Affiliation name | Country | Document type | Subject area

Export | Print

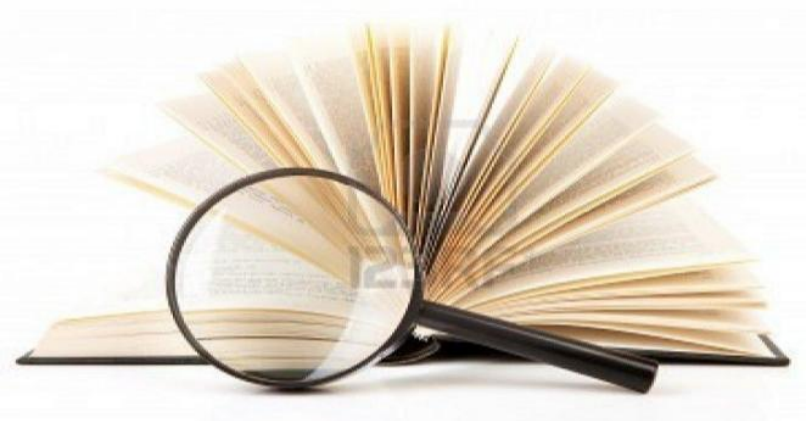
Affiliation Name This chart shows the total number of documents for this query by Affiliation Name.



	Affiliation	Documents
☑	Universiti Teknologi Malaysia	7
☑	International University of Africa	2
☑	Západoceská Univerzita v Plzni	2
☑	Xi'an Jiaotong University	2
☑	Vysoká škola bánská-Technická unive	2
☑	Universität Dortmund	2
☑	Hadhramout University of Science and	2
☑	Taif University	2
☑	International University of Africa	1
☑	P.A. College	1
☐	School of Management	1
☐	Zeidman Consulting	1

PROGRAM GOT & U

Coaching For Deep & Critical Apprehension of Literature



SOARING UPWARDS

UNIVERSITI TEKNOLOGI MARA

MYDEGS

Literature Review

- **Analysis** and **critical synthesis** of primary source materials
  - Not a *summary* - recap of the important information of the source
  - *synthesis* is a re-organization, or a reshuffling, of that information
- The evaluation of the literature that leads logically to the research question.

Guiding Students to Do Critical LR

- Conventional Review
- System Literature Review
- Experimental Review

After initial topic/area has been determined, ask student to find a focus

- A review is usually organized around ideas
  - not the sources themselves as an annotated bibliography would be organized.
  - not just simply list your sources and go into detail about each one of them, one at a time.
- Construct thesis statement
  - 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 docu-

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.

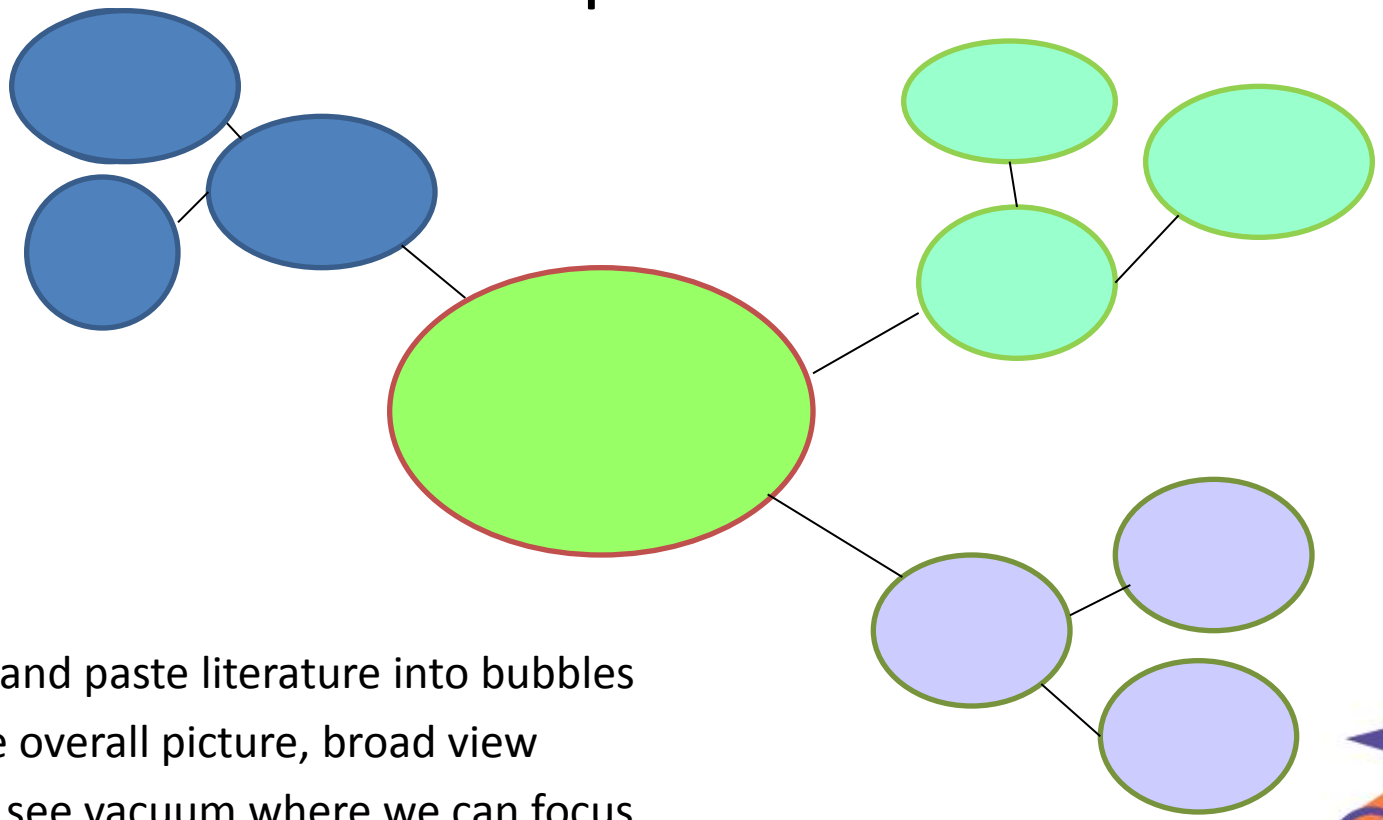
Organizing Your Review: The Analysis

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

Feature map	Classifies and categorises your thought in tabular form
Concept map	Links between concepts and processes, or shows relationship between ideas and practice
Tree construction	Shows how topic branches out into subthemes and related questions or represents stages in the development of a topic.

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

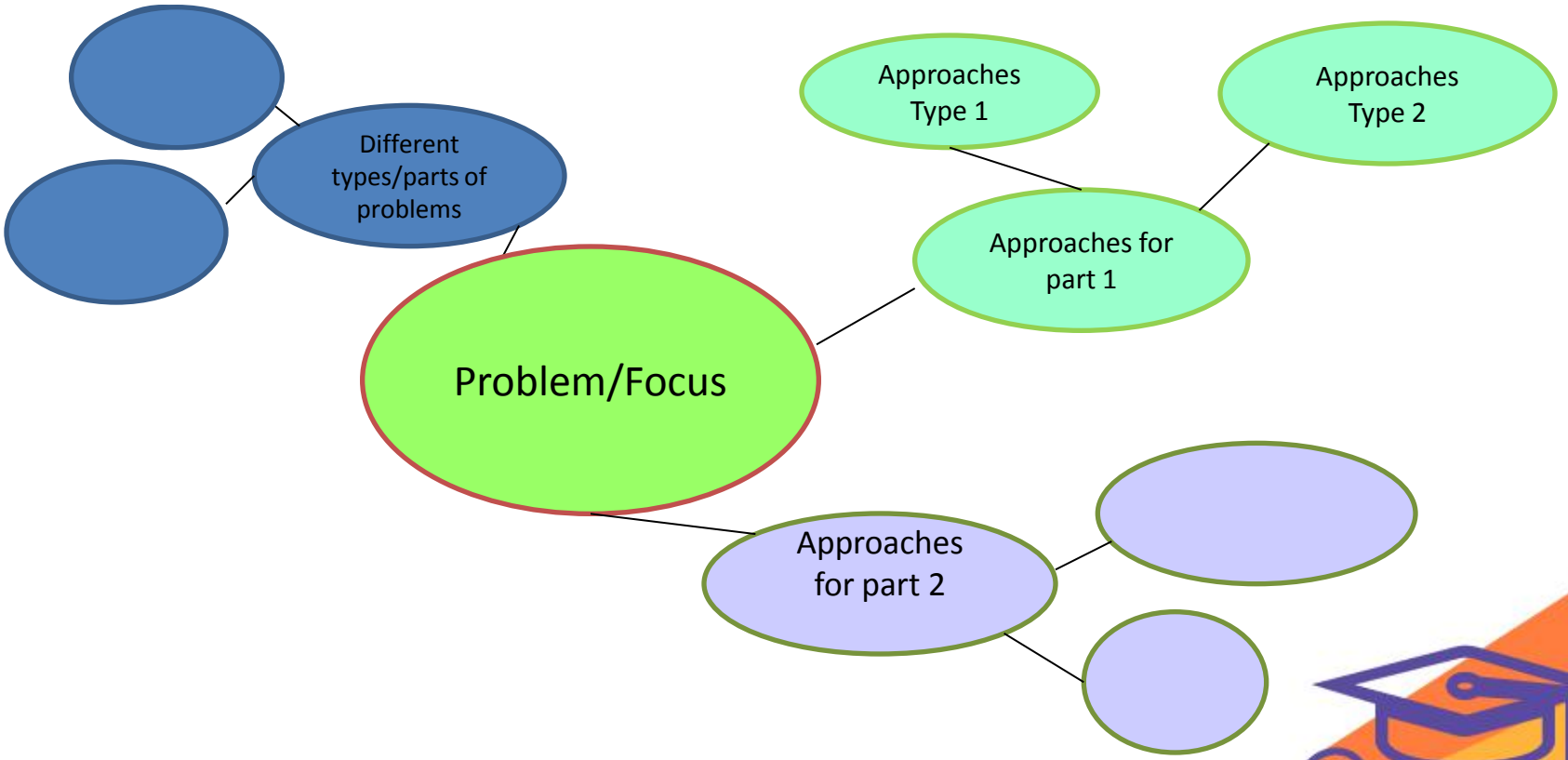
Conceptual Framework

- Building up of concept of work through literature

Theoretical Framework

- What theory support each component of the mind-map

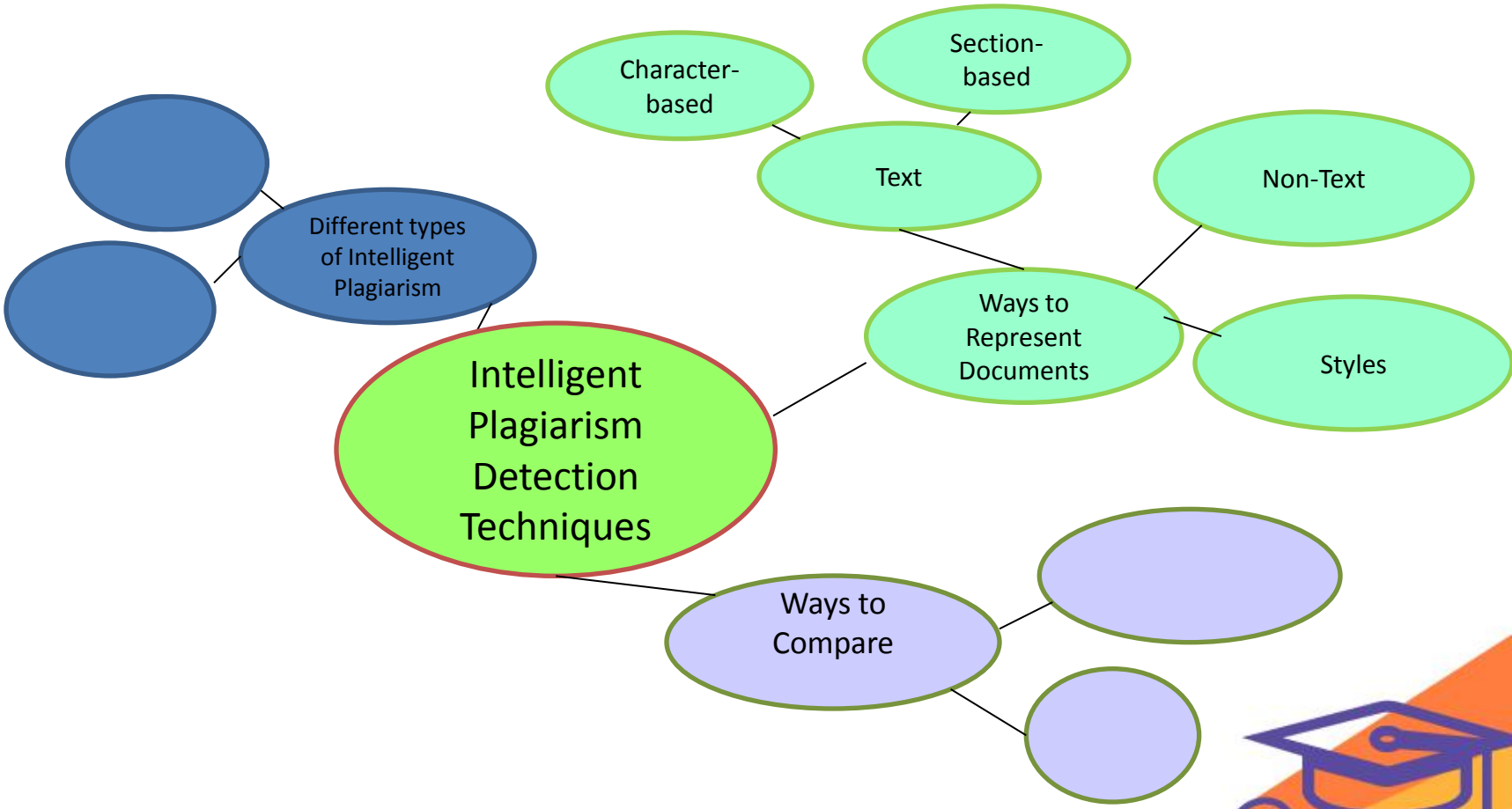
Conceptual framework



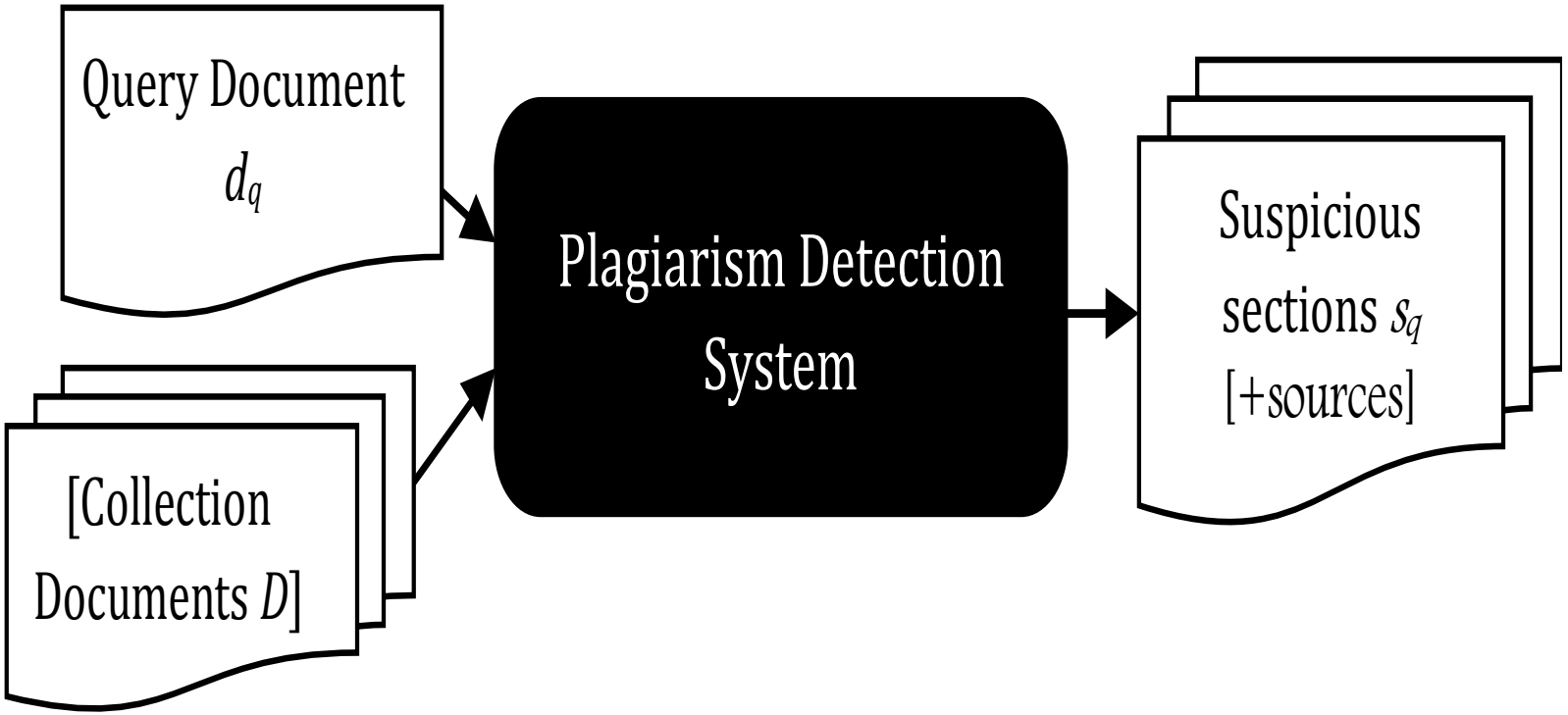
Example of Analysis Framework

- 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

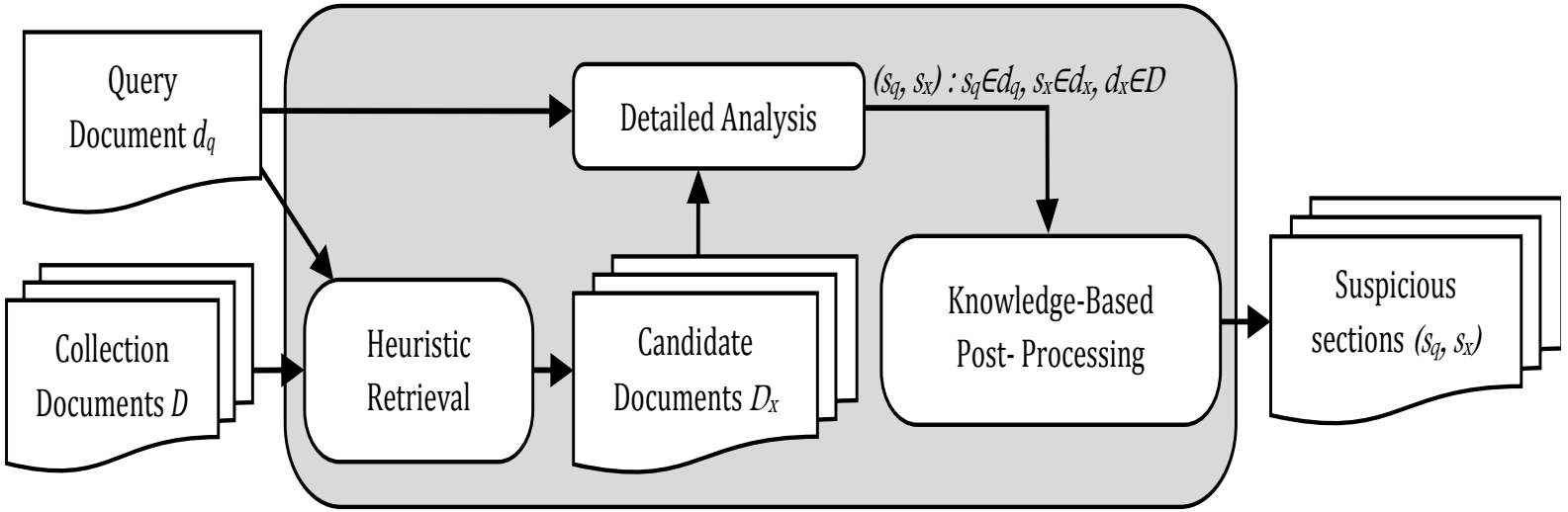
Conceptual framework



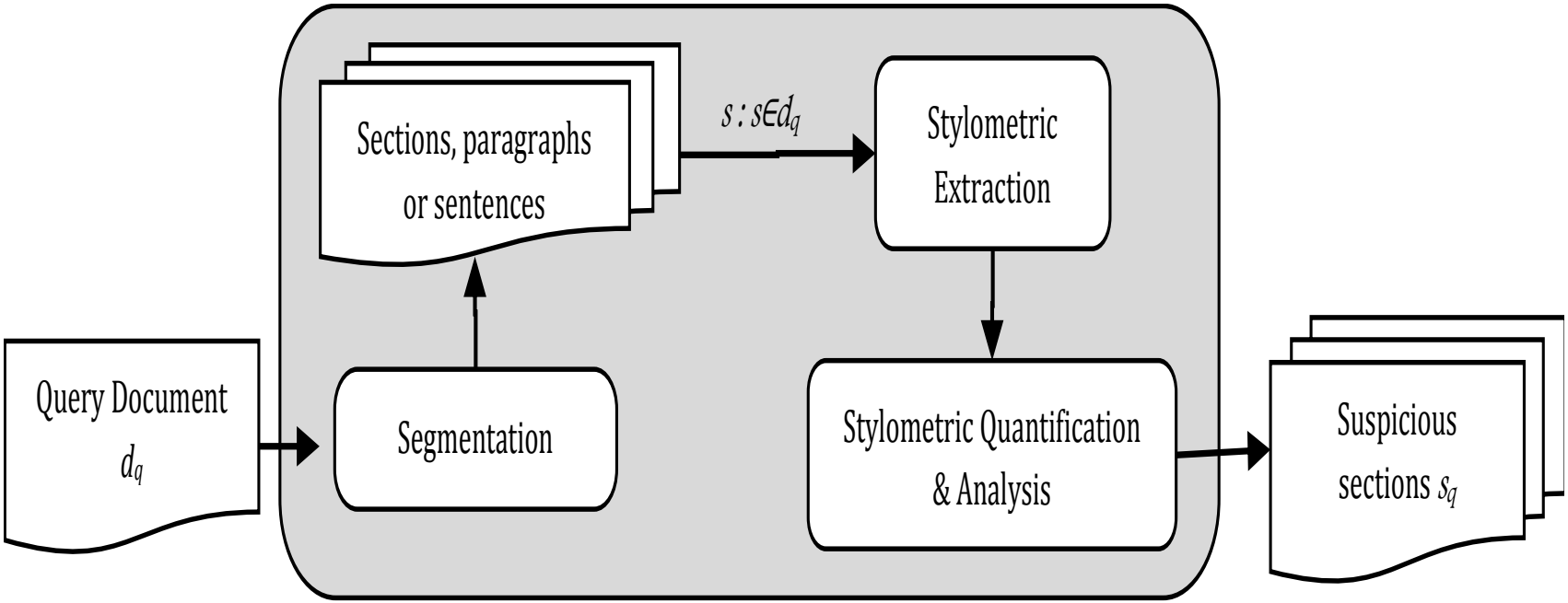
Eg. Framework of Problem Area



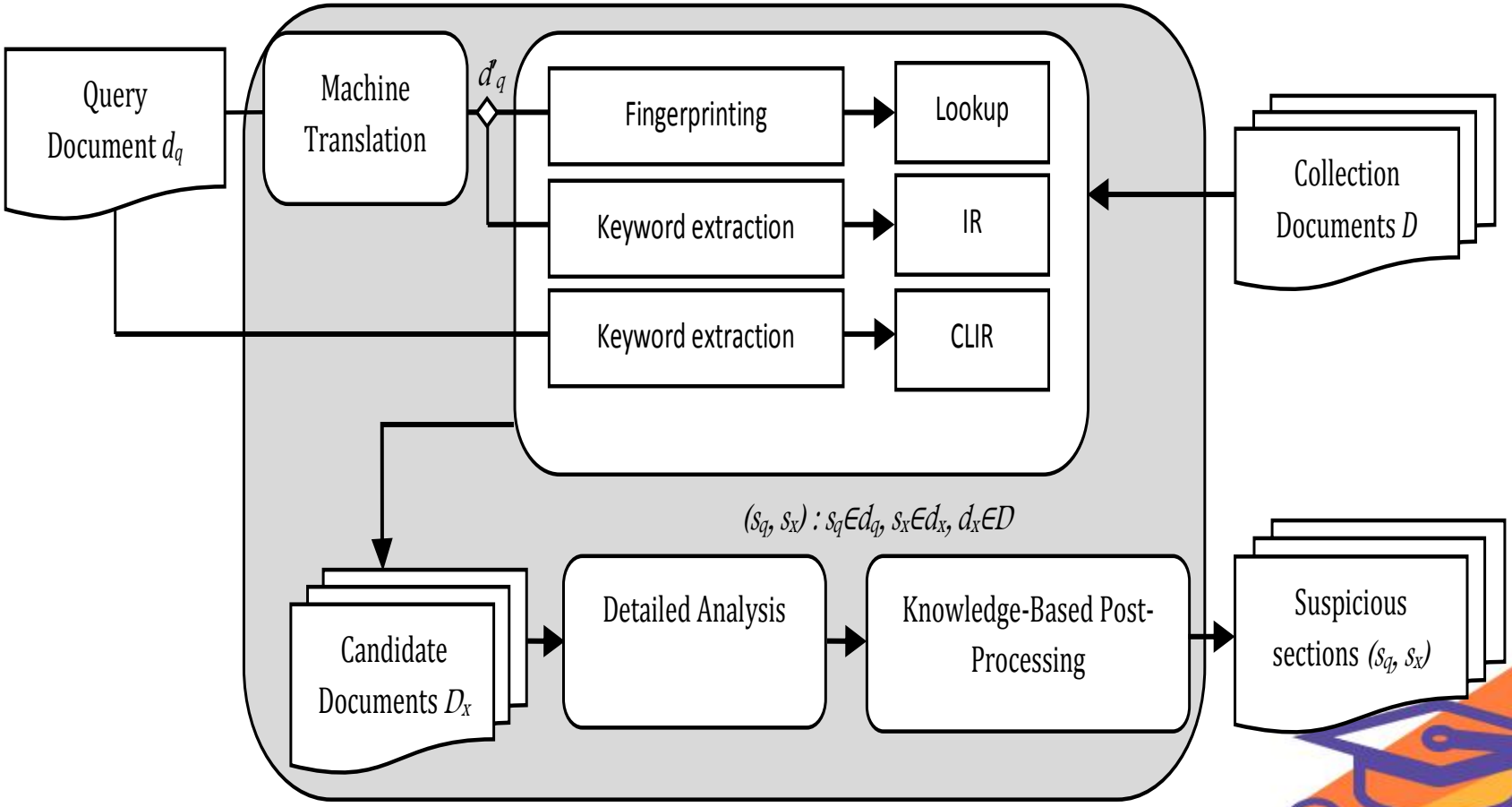
White-box design for extrinsic plagiarism detection system



White-box design for intrinsic plagiarism detection system

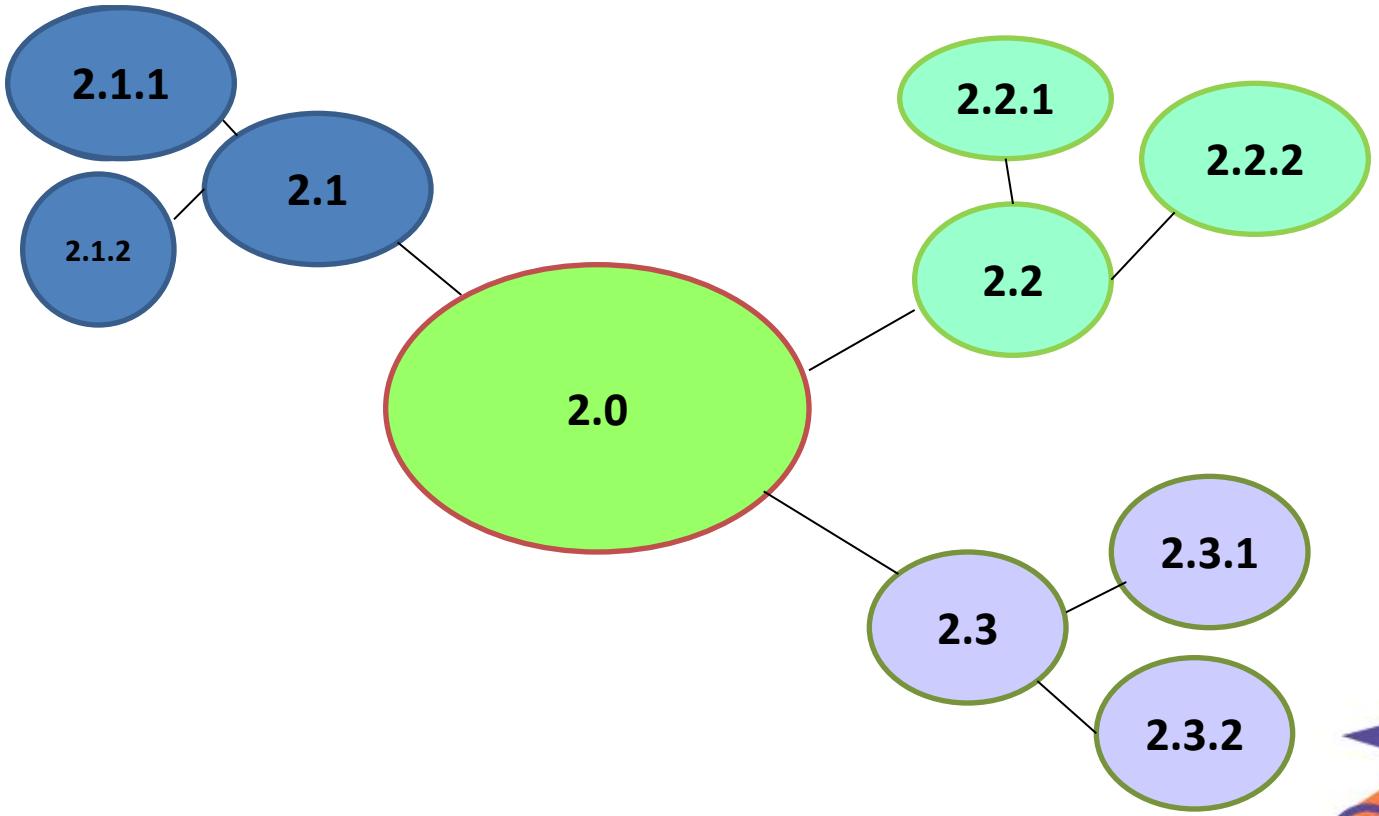


White-box design for cross-lingual plagiarism detection system

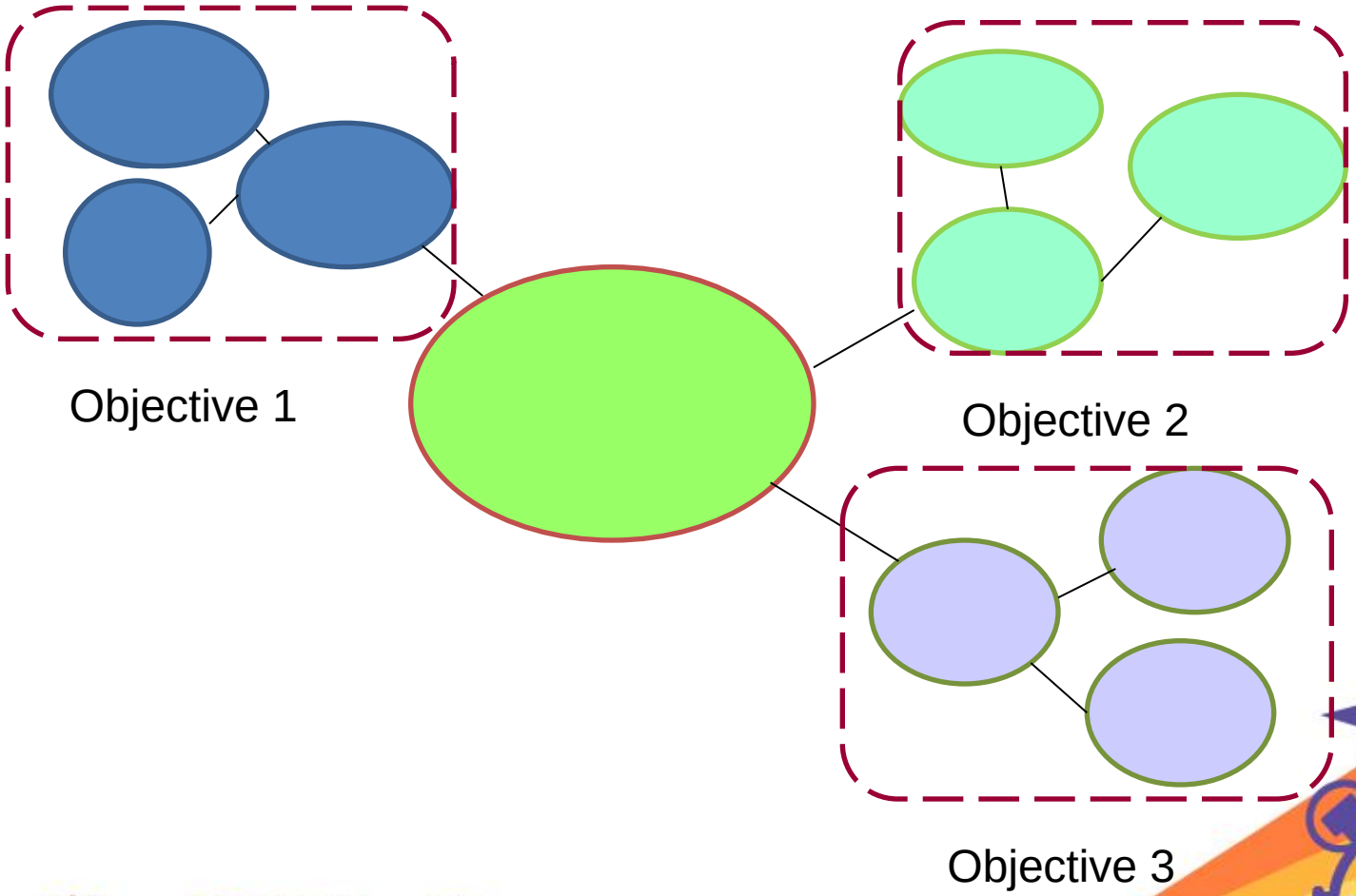


Conceptual framework

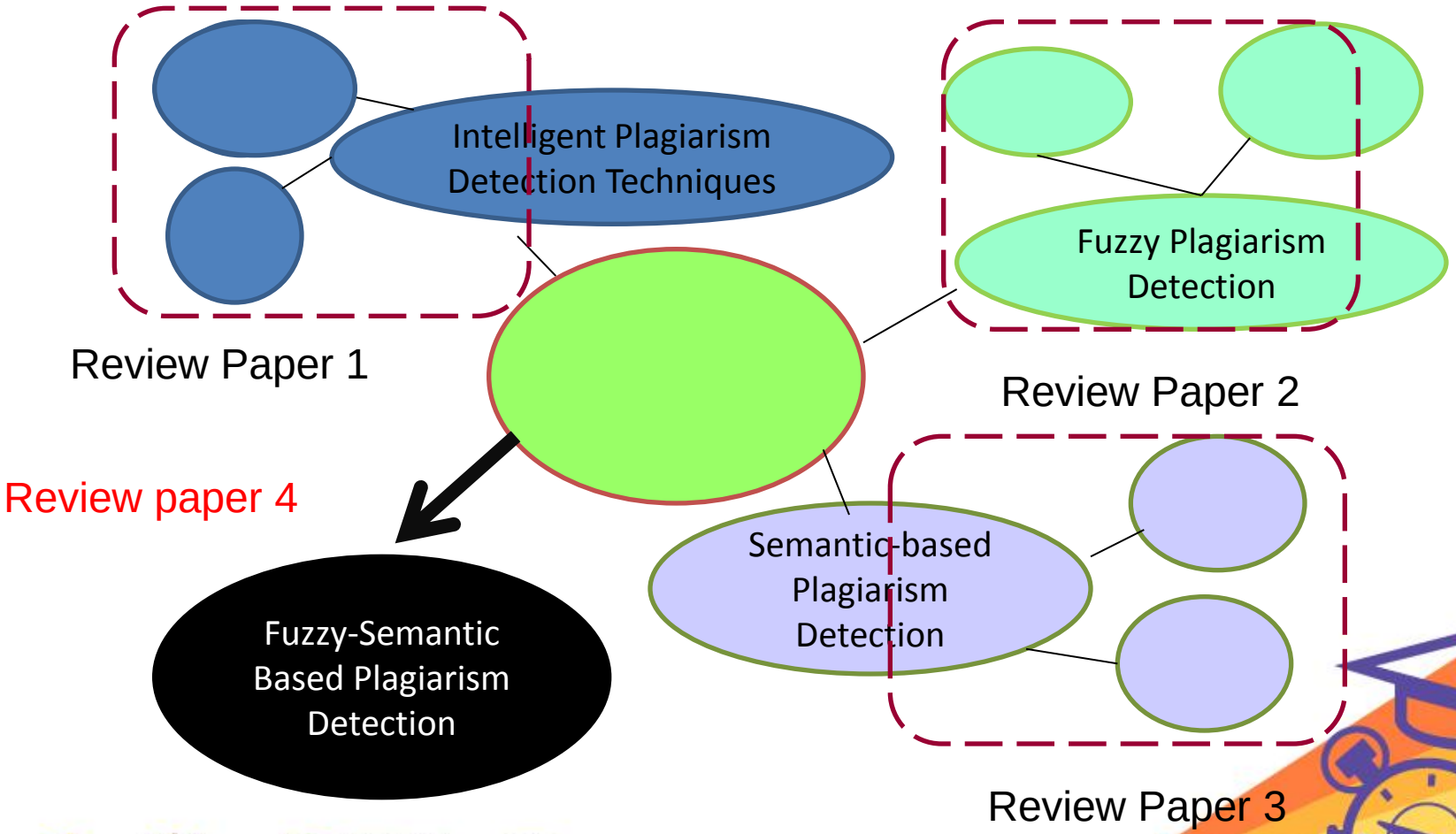
Organize LR chapter from mind-map



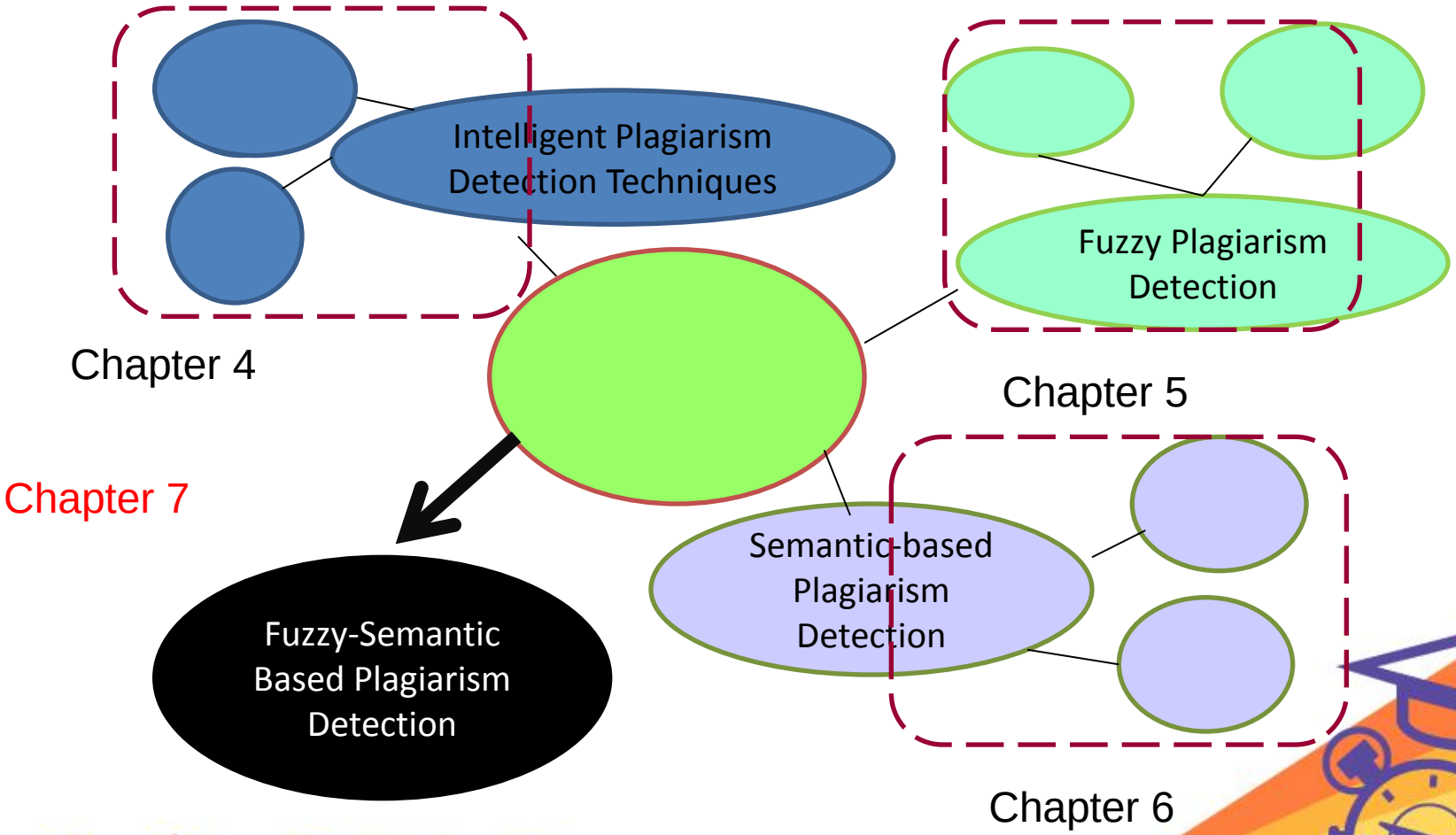
Conceptual framework – can result in research objectives



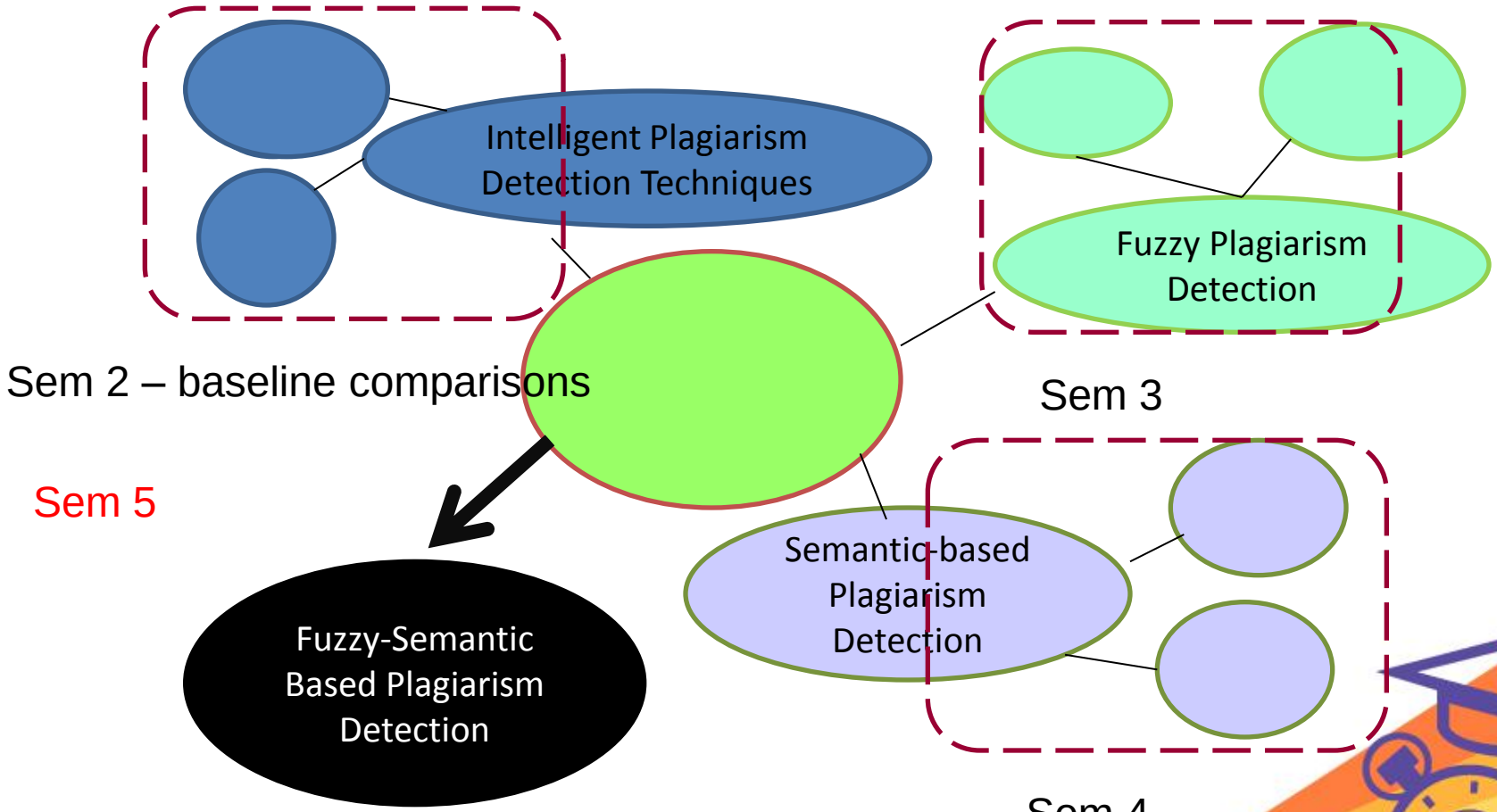
Conceptual framework into Review papers



Conceptual framework into Chapters



Sem 1 – Problem Formulation + LR



Tree Based Organization

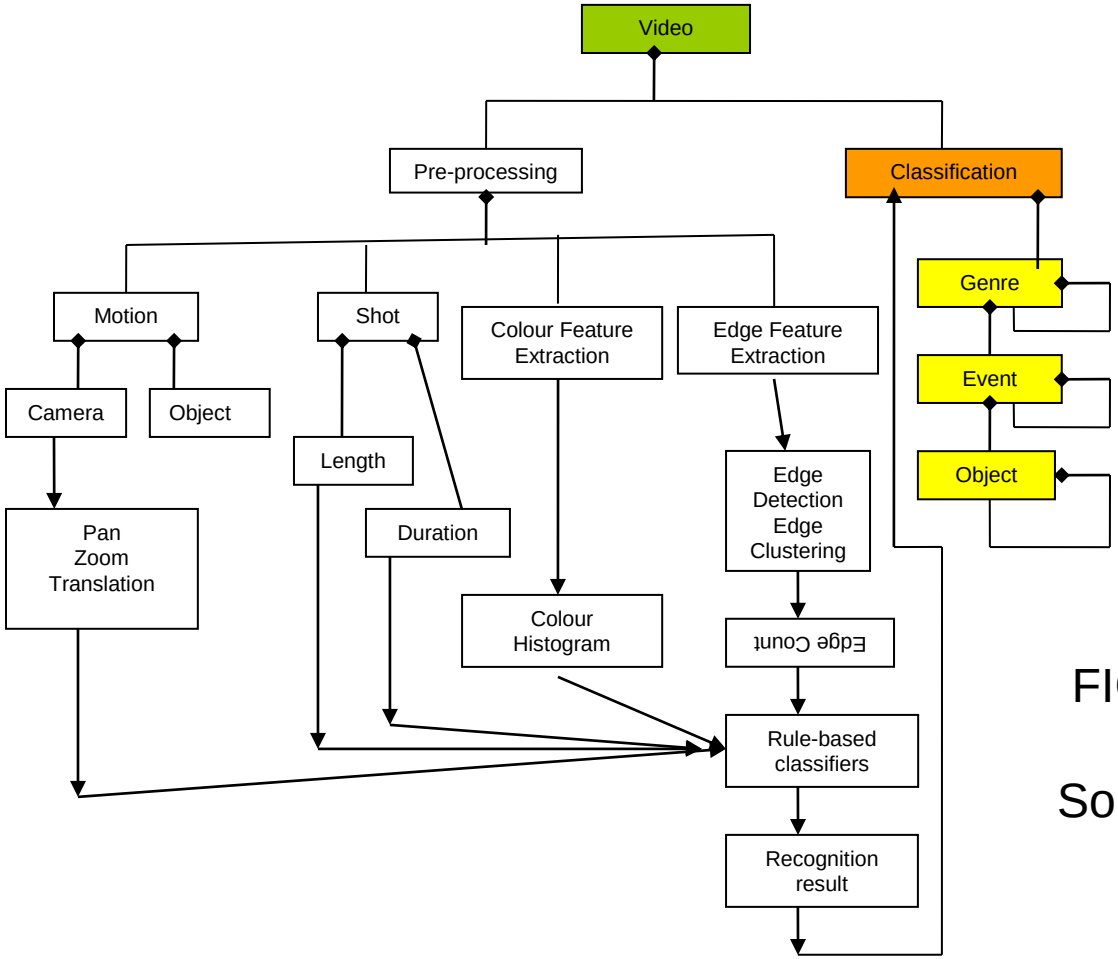
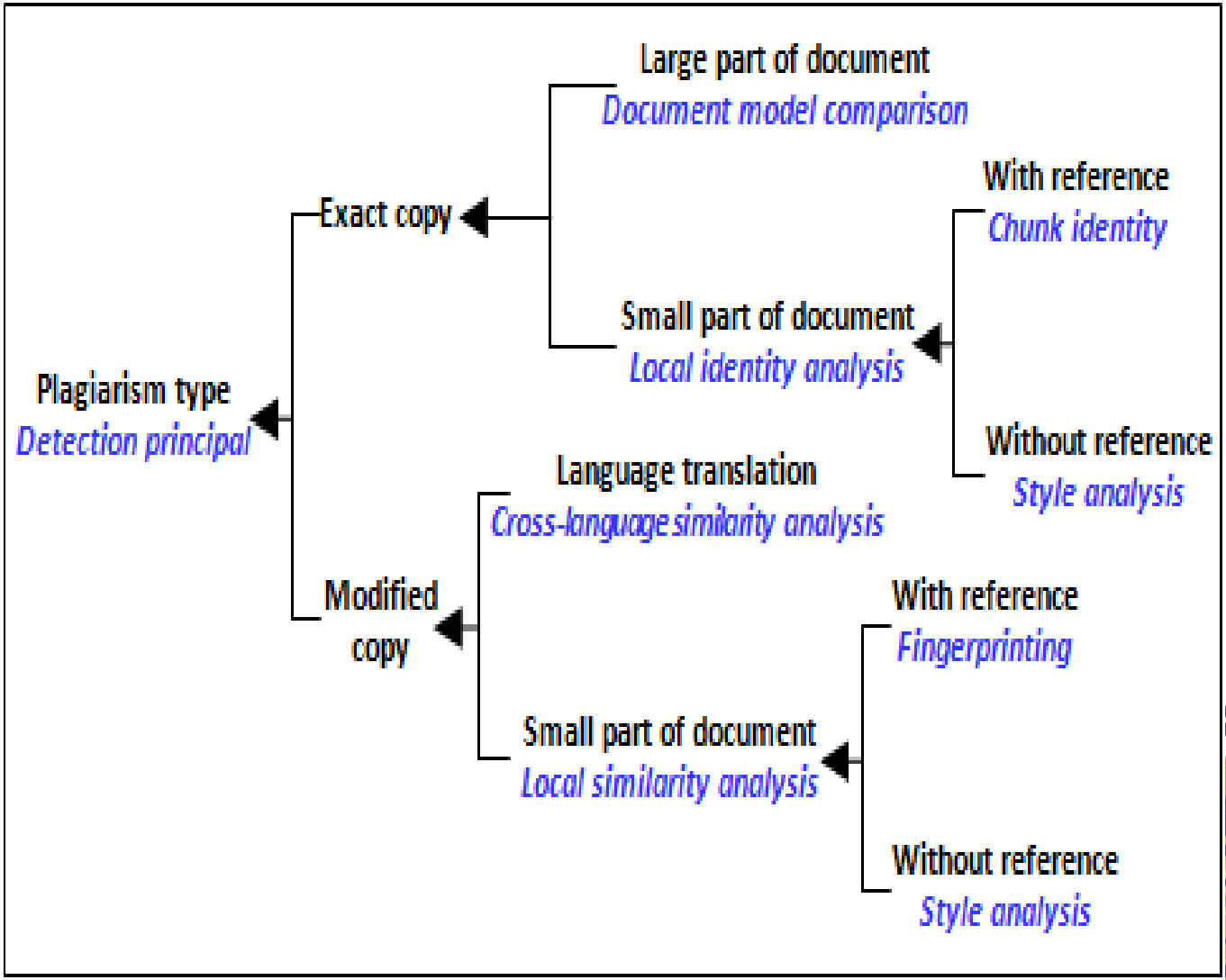
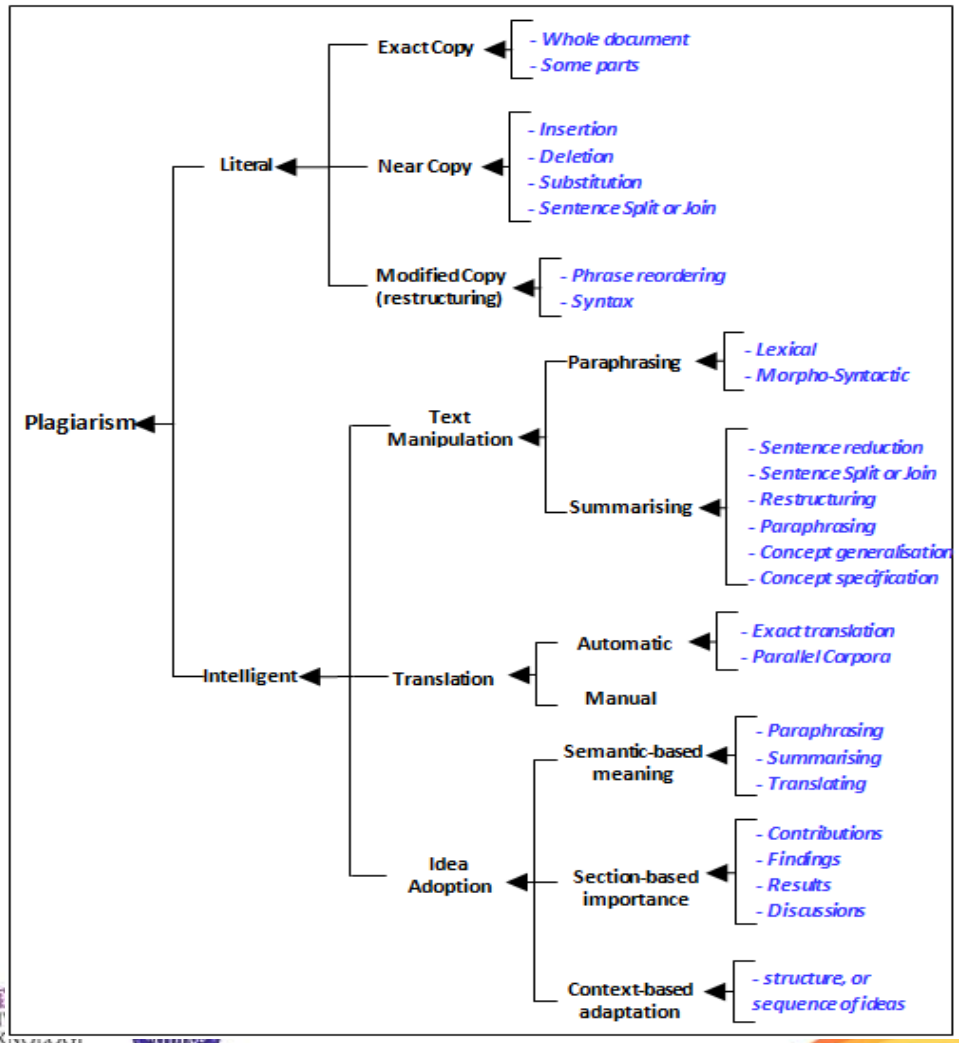


FIGURE 2.3 Schematic diagram for video classification
Source: L.N. Abdullah et al. 2005.

Eg. Current Taxanomy



Modified Taxanomy of Concepts



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

-	Examples	Required Tools and Resources	Ref.
Lexical features	Character n-grams (fixed-length)	-	[1]
	Character n-grams (variable-length)	Feature selector (e.g. n-gram weights)	[16]
	Word n-grams	Tokenizer, [Stemmer, Lemmatizer]	[2, 3, 17, 26] [30]
Syntactic features	Chunks	Tokenizer, POS tagger, Text chunker (Windowing)	[4]
	Part-of-speech and phrase structure	Tokenizer, Sentence splitter, POS tagger	[6, 12, 48]
	Word position/order	Tokenizer, Sentence splitter, Compressor (e.g. Lempel-Zif)	[13, 14]
	Sentence	Tokenizer, Sentence splitter, POS tagger, Text chunker, Partial parser	[16, 58]
Semantic features	Synonyms, hyponyms, hypernyms, etc.	Tokenizer, [POS tagger], Thesaurus	[14, 16, 18, 58] [30]
	Semantic dependencies	Tokenizer, Sentence splitter, POS tagger, Text chunker, Partial parser, Semantic parser	[14, 61]
Structural features	Block-specific	HTML parser, Specialised parsers	[21, 29]
	Content-specific	Tokenizer, [Stemmer, Lemmatizer], Specialised dictionaries	-

Similarity Evaluation

Vector Similarity Metric	Description & Equation	Equation	Range	Example	Ref.
Matching coefficient	-similar to Hamming distance but between vectors of equal length.	$M(x,y) = \|x\| - \|x \cap y\|$	0 to |x| Where |x|=|y|	*x=[0.1, 0.2, 0.3, 0.4] y=[0.1, 0.2, 0.3, 0.5] M(x,y)= 1*	[11]
Jaccard (or Tanimoto) coefficient	-defines number of shared terms against total number of terms. This measure is computed to one if vectors are identical.	$J(x,y) = \frac{\|x \cap y\|}{\|x \cup y\|}$	0 to 1	*J(x,y)=3/5=0.6*	[3, 7, 8, 21]
Dice's coefficient	-similar to Jaccard but reduces the effect of shared terms between vectors. This measure is computed to two if vectors are identical.	$D(x,y) = \frac{2\|x \cap y\|}{\|x \cup y\|}$	0 to 2	*D(x,y)=6/5=1.2*	-
Overlap (or containment) coefficient	-if v_1 is subset of v_2 or the converse, then the similarity coefficient is a full match.	$O(x,y) = \frac{\|x \cap y\|}{\min(\|x\|,\|y\|)}$	0 to 1	*O(x,y)=3/4=0.75 (or 75%)*	[10]
Cosine coefficient	-finds the cosine angle between two vectors.	$Cos(x,y) = \frac{\sum_i (x_i, y_i)}{\sqrt{\sum_i (x_i)^2}\sqrt{\sum_i (y_i)^2}}$	0 to 1	*Cos(x,y)=0.34/0.3421 =0.9939 ≈1*	[9, 21, 26, 28]
Euclidean distance	-measures the geometric distance between two vectors.	$Ec(x,y) = \sqrt{\sum_i \|x_i - y_i\|^2}$	0 to ∞	*Ec(x,y)=0.1*	-
Squared Euclidean Distance	-places progressively greater weight on vectors that are further apart	$SEc(x,y) = \sum_i (x_i - y_i)^2$	0 to ∞	*SEc(x,y)=0.01*	-
Manhattan Distance	-measures the average difference across dimensions and yields results similar to the simple Euclidean distance	$Manh(x,y) = \sum_i \|x_i - y_i\|^2$	0 to ∞	*Manh(x,y)=0.1*	-

METHODS AND THEIR EFFICIENCY IN DETECTING DIFFERENT PLAGIARISM TYPES

Technique	Tasks		IR		Language(s)	Plagiarism Type(s)								Reference
						Literal			Intelligent					
	extrinsic	intrinsic	mono-lingual	cross-lingual		copy	near copy	restructuring	paraphrasing	summarising	translating	idea (section)	idea (context)	
Char-Based (CNG)	☑		☑		any	☑	☑							[1-6]
Vector-Based (VEC)	☑		☑		any	☑	☑	☑						[7-11]
Syntax-Based (SYN)	☑		☑		specific	☑	☑	☑						[6, 12, 13]
Semantic-Based (SEM)	☑		☑		specific	☑	☑	☑	☑	☐				[14, 15]
Fuzzy-Based (FUZZY)	☑		☑		specific	☑	☑	☑	☑	☐				[16-19]
Structural-Based (STRUC)	☑		☑		specific	☑	☑	☑	☐	☐		☐	☐	[21, 29]
Stylometric-Based (STYLE)		☑	☑		specific	☑	☑	☑						[22, 23, 32-35]
Cross-Lingual (CROSS)	☑			☑	cross						☑			[31, 36-38]

Suggested words to discuss 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 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.

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

Example of LR Approach

Example of LR Approach (cont.)

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

Eg. Of Discussion

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

Selection method	Time-complexity	Space-complexity	Applicability
Hierarchic agglomerative (stored matrix algorithm)	O(N^3)	O(N^2)	Small files only
Reciprocal nearest neighbour (Ward's)	O(N^2)	O(N)	Up to a quarter of million molecules
Reciprocal nearest neighbour (Jarvis-Patrick)	O(N^2)	O(N)	Up to more than a million molecules, due to its lower constant proportionality in the time-complexity
Maximum-dissimilarity	O(N^3)	O(N^2)	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].

Systematic Literature Review - SLR

- Formulate the review's research question
- Develop the review's protocol
- Identify the relevant literature by conducting a comprehensive and exhaustive search
- Selection of primary studies based on the inclusion/exclusion criteria
- Extraction of data
- Assessment of studies' quality
- Synthesis of evidence
- Write up the SLR report

Planning

Conducting

Reporting

PROGRAM GOT & U

Review + Experimental Paper

SOARING UPWARDS

UNIVERSITI TEKNOLOGI MARA

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

MAKE YOUR 'VOICE' CLEAR

- Do not just presenting others views or arguments
  - literature review should be more than a catalogue of the literature. It should contain a critical, analytic approach, with an understanding of sources of error and differences of opinion
- It is YOUR perspective, position or standpoint (not only in the LR, but also in the theses as a whole)
- Your theoretical position is clearly and strongly stated
- Your language should indicates YOUR assessment of literature

Planning for GOT

Plan for GOT to GOT

PhD Research Schedule

No.	ACTIVITIES	Year 1				Year 2				Year 3			
		1	2	3	4	1	2	3	4	1	2	3	4
1	Literature Review												
2	Problem Formulation												
3	Initial Results												
4	Proposal Writing (Chapters 1, 2, 3, 4)												
5	Objective 1												
6	Objective 2												
7	Objective 3												
8	Thesis Writing												

No.	MILESTONE	Year 1				Year 2				Year 3			
		1	2	3	4	1	2	3	4	1	2	3	4
1	Review Paper												
2	Problem Formulation												
3	Completion of Proposal Writing & First Assessment												
4	Paper with some results												
5	Completion of Objective 1												
6	Completion of Objective 2												
7	Completion of Objective 3												
8	Thesis Writing Completion/Submission												

Read and Write Throughout PhD

- Ask student to write at least one chapter every semester
  - Can use as progress report
- 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/one semester
- Appointment by paper

Sample Publication Agenda

Time	What	Where	Sample Inclusion
1st semester	Critical Analysis of Literature	• Conference • Journal	• Framework of analysis • New Taxonomy • Specific review • Hypothesis • dy
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 • omparison • ation of • ments

Al Zahrani S., and Salim, N. (2011). "Understanding Plagiarism Linguistic Patterns, Textual Features and Detection Methods". IEEE Transactions on Systems, Man and Cybernetics. Vol 41(1): pp. 1-17. (WOS & Scopus indexed, Impact Factor: 2.06)

Salim,N, Whittle, M.W., Holliday, J.D. and Willett, P., (2003) " Analysis and Display of the Size Dependence of Chemical Similarity Coefficients" , Journal of Chemical Information and Computer Sciences, vol. 43(3): pp. 819-828. (WOS & Scopus indexed, Impact Factor: 3.643)

Salim,N, Holliday, J.D. and Willett, P. (2003), Combination of Fingerprint-based Similarity Coefficients using Data Fusion, Journal of Chemical Information and Computer Sciences, vol. 43(2): pp. 435-442. (WOS & Scopus indexed, Impact Factor: 3.643)

Sample Publication Agenda (cont.)

Time	What	Where	Sample Inclusion
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

A. Abdo, Salim,N (2009), " Similarity-Based Virtual Screening with Bayesian Inference Network", ChemMedChem 4(2): pp. 210-218 (WOS & Scopus indexed, Impact Factor: 3.51)

A. Abdo and Salim,N. (2009) "Bayesian Inference Network Significantly Improves the Effectiveness of Similarity Searching Using Multiple 2D Fingerprints and Multiple Reference Structures", QSAR &Combinatorial Science, 28(11-12):1537-1545, (WOS & Scopus indexed, Impact Factor: 2.594)

Sample Publication Agenda (cont.)

Time	What	Where	Sample Inclusion
5th semester	• 3rd Objective/ Experimental Paper	• Conference • Journal (extended dataset)	• Introduction • Experimental Design • Results • Discussion • ion
	Paper	journal	• ction • Experimental Design • Results • Discussion • Conclusion

A. Abdo, Salim,N, C. Mueller, and P. Willett. "Similarity-Based Virtual Screening Using Belief Network", Journal of Chemical Information and Modeling, 50(6): pp. 1012-1020 (WOS & Scopus indexed, Impact Factor: 3.631)

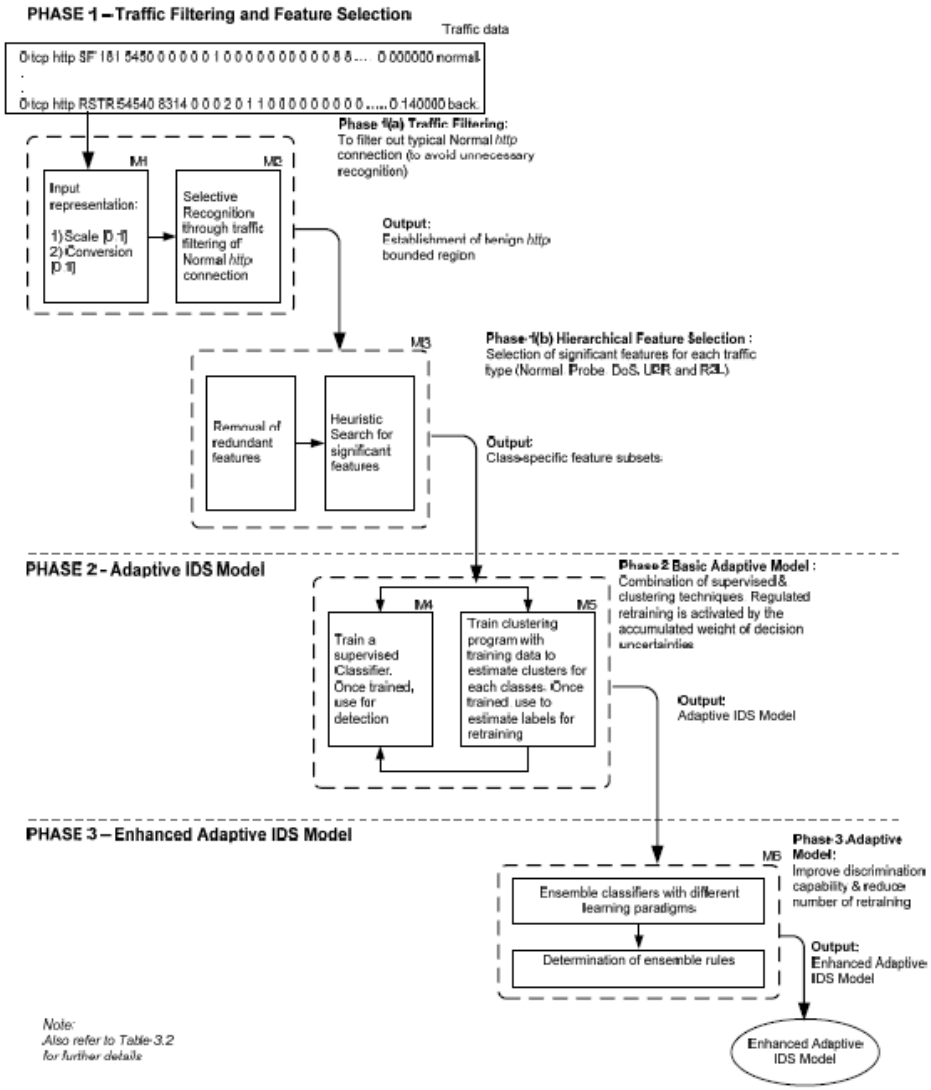
Binwahlan, M. S., Salim, N. and Suanmali, L. (2010). "Fuzzy Swarm Diversity Hybrid Model for Text Summarization". Information Processing & Management. Vol 46(5). Pp. 571-588 (WOS & Scopus indexed, Impact Factor: 2.106)

Mentoring in Experimental Design & Data Collection Phase

Discuss and Enlist Help

Draw Up a Research Framework

- Diagram
- Table
- Description
- Gantt Chart



Phase	Activities	Resources Needed	Benchmark Data	Baselines for Comparison	Perfor-mance Evaluation	Objectives Addressed

Do standard things

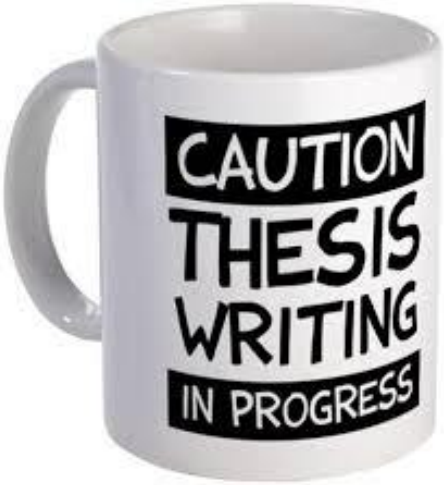
- Get idea from literature or graduated student. Why?
  - Someone has thought method out carefully
  - Saves time
- Learn what those standard things are (add only to test new ideas)
  - Datasets
  - Methods
  - Evaluation
- Statement must be supported by a reference to the scientific literature or by original work.

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

PROGRAM GOT & U

Coaching in Analysis of Results and Writing Up



SOARING UPWARDS

UNIVERSITI TEKNOLOGI MARA

MYDEGS

Ask student to make writing a habit

- Put on the paper even small Ideas, points, thoughts
- Putting an ideas on a paper allows to polish it and invent a new or extend the Idea
- Ask student to put citation alert to keep up with updates in field

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 **cause**s (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!