Justification

- I am probably a qualified person to give a talk on this topic... because
  - I'm still STRUGGLING to publish
  - Yes, I still do get rejections 😊
  - I'm still learning from failures

- What’s being presented here is purely my suggestion
- Take it or leave it – up to you !!
The Goal of Research Papers

- Disseminate your ideas to others so that people appreciate/use/cite them
- Graduate… Of course
  - MS: need to write a thesis to graduate…
  - Ph.D: “Publish or Perish”
- Without good publications…
  - No good job, no good career
  - And possibly no good life either
- GPA: nobody cares PhD’s GPA
  - Maintain about 3.0/4.0
Where to Start?

- Given that you have acquired
  - basic theory/knowledge/tools from classes and books...

- Next, first thing to learn:
  - Read others' papers
  - Critique and evaluate them

- Which paper to read?
  - Start from good ones
  - Classical ones
  - Ones from good journals or conferences

Where to Start: eg, Databases

<table>
<thead>
<tr>
<th>DB Conferences/Symposiums/Workshops (81)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADB, ADBIS, ADBT, ADC, ARTDB, Berkeley Workshop, BNCOD, CDB, CIDR, CIKM, CISM, CISMOD, COMAD, COODBSE, CoopIS, DAISO, DANTE, DASFAA, DaWaK, DBPL, DBSEC, DDB, DDW, DEXA, DiWeB, DMDW, DMKD, DNIS, DOLAP, DOOD, DPDS, DS, EDBT, EDS, EFIS/EFDBS, ER, EWDW, FODO, FoIKS, FOAS, Future Databases, GIS, HPTS, IADT, ICDE, ICDM, ICOD, IDA, IDC(W), IDEAL, IDEAS, IDS, IGIS, IWDM, IW-MMDBMS, JCDKB, KDD, KR, KRDB, LID, MDA/MDM, MFDBS, MLDM, MSS, NLDDB, OODBS, OOS, PAKDD, PKDD, PODS, RIDE, RIDS, RTDB, SBBD, SDM-SIAM, Semantics in Databases, SIGMOD, SDD, SDDBM, SWDB, TDB, TSDM, UIDIS, VDB, VLDB, WebDB, WIDM, WISE, XP, XSym</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DB Journals (19)</th>
</tr>
</thead>
</table>

The list excludes Information Retrieval and Digital Library
Where to Start?

- Some good ones:
  - DB: SIGMOD, VLDB, ICDE, EDBT, ...
  - DB Theory: PODS, ICDT, ...
  - Data Mining: KDD, ICDM, SDM, ...
  - Modeling: ER, ...
  - Information Retrieval: SIGIR, CIKM, ...
  - Digital Library: JCDL, ECDL, CIKM, ...
  - Web: WWW, WebDB, ...

Reference Chase

- Don’t trap into the “Exponential Reference Chase” problem
Symptoms

- After chasing relevant works that are increasing super-exponentially fast, you might feel…
  - All relevant problems are ALREADY studied by someone else
    Others have 1000+ history: Mathematics, Art, …
  - Problem is too BROAD for me to tackle
    Divide-n-conquer

TOC

- How to start?
- **How to find research problems?**
- How to write research papers?
- How/Where to submit?
- Ethics
- Misc.
How to Find the DARN Research Problems?

- Easy but non-helpful answer:
  - Read and think and read and think and…
- Subjective but MAYBE-helpful answer
  - MAP approach
  - MATRIX approach
  - DELTA approach
  - DROP approach

What I Call M2D2

1. MAP Approach

- To start a research, initially, you have to read a lot of papers anyway
- While reading those, why don’t you analyze and summarize what you’ve read and put them into your own wording?
  - Good for a survey paper – a MAP for future readers
- To be publishable, your survey must have novel viewpoint, taxonomy, comprehensive analysis, or all of them
2. **MATRIX** Approach

- Now, You have read a lot of papers
- Draw a MATRIX on a specific problem, and map the paper that you read to cells of matrix
- At the end, non-filled cell is the missing work that no one has done
- But wait… first make sure that:
  - The hole is worthwhile to fill in
    - Doable (good as my dissertation topic?)
    - Value (what’s good?)

---

Eg, XML-Relational Conversion Problem

<table>
<thead>
<tr>
<th></th>
<th>Schema</th>
<th>Constraint</th>
<th>Query</th>
<th>View</th>
<th>Triggers</th>
<th>Security</th>
<th>Top-K</th>
<th>Temporal</th>
<th>Spatial</th>
</tr>
</thead>
<tbody>
<tr>
<td>XML → Relational</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(40+)</td>
<td>(5+)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relational → XML</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3. DELTA Approach

- Arguably easiest...
  - Pick one paper of your interest
  - Read a lot – more than 10 times
  - Find limitations and Extend it by DELTA
  - Prove or demonstrate that
    The limitation that you pointed out is valid
    Your suggestion improved the problem by DELA
- The more well-known work you choose, the harder to improve, but the better for your reputation...
  - Eg, “E.F. Codd’s relational model is insufficient to handle semi-structured model because…”
- The bigger the DELTA is, the better your paper gets

Eg, The optimal wedding problem

- When a person has a chance to date \( K \) persons, the optimal wedding algorithm is:
  - Date upto \( K/3 \) persons
    Let the best person among \( K/3 \) as \( B \) using a criteria \( C \)
  - Start dating again from \( K/3+1 \) person, \( p \)
  - If \( p \) is better than \( B \) using \( C \)
    Stop and Marry \( p \)
  - Otherwise, keep dating till \( K-th \) person

- How many ways can we improve this algo?
Possible DELTAs

- Parameters fitting:
  - How to determine \( K \)? Estimate?
  - How to determine \( C \)? Comparison?
- Scalability? \( K=10 \) vs. \( K=100,000 \)? Sub-optimal?
- Question the assumptions:
  - Monogamy vs. Polygamy vs. N-gamy? (How to find \( n^{th} \) best spouse fast?)
  - Data distribution? Uniform/Poisson/Scale-free
- Application to another domain?
- System building?
- ...

Which DELTA to Choose

- Pick the DELTA that is the most significant
- Some criteria are:
  - Have practical values
    - Has motivational scenario as of NOW, or
    - Predicted to be useful in \( N \) years
  - Non-trivial
  - Hot topics:
    - Streaming, XML, Sensor, …
4. DROP Approach
(adopted from J. Widom's slides)

- Pick a simple but fundamental assumption of existing theory/model/systems/methods
  - DROP it
- Reconsider to see how the drop affects all aspects of the existing theory/model/systems/methods
  - Many Ph.D. theses

From http://www-db.stanford.edu/~widom/stream.ppt

---

Eg, Two Stanford DB Projects

- The LORE Project
  - Dropped assumption:
    "Data has a fixed schema declared in advance"
  - Semi-structured data (→ XML)
- The STREAM Project
  - Dropped assumption:
    "First load data, then index it, then run queries"
  - Continuous data streams (+ continuous queries)

From http://www-db.stanford.edu/~widom/stream.ppt


**TOC**

- How to start?
- How to find research problems?
- **How to write research papers?**
- How/Where to submit?
- Ethics
- Misc.

---

**Facts on Paper Reviews**

(adopted from J. Cho’s slides)

- 3-4 reviewers per paper
- 10-20% acceptance rate for top-tier venues
  - Very competitive
- Criteria
  - Accept/Weak Accept
  - Neutral
  - Weak Reject/Reject
- One reject kills a paper
  - At least Accept, Weak Accept and Neutral
### About Reviewers

- 15-20 papers per reviewer (for conferences)
- Reviewer cannot spend 5-10 hours per paper
  - $20 \times 10 = 200$ hours $= (40$ hours $\times 5) = 5$ weeks!
  - No reviewers can afford this
- Give a good impression in 1-2 hours!
  - Impression matters the most
  - Content comes next!

**WARNING:** Of course, to start with, your main idea must be good to get into top-tier…

### Good Impression in 1-2 hours?

1. **Good introduction**
   - Everyone reads it
   - If not interesting, people stop reading
2. **Easy to read**
   1. People should understand what you say
   2. Easy to confuse, difficult to understand
3. **Build an excitement and a strong case**
   1. What is good?
4. **Broad reference**
   1. Sometimes kills a paper
   2. Program committee members
### Good Introduction

1. What’s the problem?
2. Why is it important?
   - Mention some application, existing problems
3. Why is it difficult?
   1. Ask some not-very-obvious questions or explain naive approach
4. What others did?
5. What’s my contribution?
   1. Contribution bullet list (paper organization)
6. Build some excitement/surprise
   1. Keep reading! You will find something interesting later
7. Every word should be carefully picked

### How to Write an Introduction

1. Start with 5 bullets
   - What’s the problem?
   - Why is it interesting?
   - ...
2. 1-2 sentence answer to each question
3. Add more content
4. Spend enough time on introduction
   1. Bullet points enough
Easy-to-Read Paper

- You can always make it complicate later
  1. Lots of examples
  2. Figures & Tables – Figure speaks !!
     - Summary of notations
  3. Define assumptions/models/architecture precisely
     - Explicitly write down assumptions
     - Input, output, property, goal function
  4. Make a connection
     - Why this experiment?

Paper Organization (10 pages)

1. Introduction (2 pages)
2. Related Work (half page)
3. Framework (2 pages)
4. Main Ideas (3 pages)
5. Experiments (2 pages)
6. Conclusion (half page)
7. References (half page)

- Actual idea – only 3 pages!!!
  - Even tiny idea can turn into a good paper if you DEVELOP it well
Start Writing Early On...

- Even if you feel you are NOT ready yet
  - Your advisor will throw away your initial draft anyway
  - Your initial submission will be rejected anyway
- But you get:
  - (good or bad) Experiences and learn from that
  - Writing sharpens your ideas and gives more ideas
  - Writing can be improved only via writing

TOC

- How to start?
- How to find research problems?
- How to write research papers?
- **How/Where to submit?**
- Ethics
- Misc.
Where to Submit?

- **Top-down**
  - Aim at the best venue in the field
  - If rejected, go to next-tier venue
  - If rejected, go to next…

- **Bottom-up**
  - Aim at workshop
  - If accepted, work more and aim at better one (symposium or 2\(^{\text{nd}}\)-tier conference)
  - After making sure that the ideas mature enough, aim at the best conference or journal

Avoid Some Notorious Venues

- “Randomly generated paper got accepted to a conference… MIT Prank” (slashdot, 2005)
  - http://pdos.csail.mit.edu/scigen/
  - Eg, The World Multi-Conference on Systemics, Cybernetics and Informatics (SCI)

- Along your career, you will get emails from unknown venues to submit a paper, to serve as PC, etc
  - Be careful if the venue is not well-known
  - Many of them are NON-REVIEWED, and Profit-Oriented event – no academic values what so ever !!
Some Known Fake Venues

- From http://www.inesc-id.pt/~aml/trash.html:
  - IMCSE: International Multiconference in Computer Science and Computer Engineering
  - WMSCI or SCI: World Multiconference on Systemics, Cybernetics and Informatics
  - ICCCT: International Conference on Computing, Communications and Control Technologies
  - PISTA: Conference on Politics and Information Systems: Technologies and Applications
  - SSCCI: Symposium of Santa Caterina on Challenges in the Internet and Interdisciplinary Research
  - CITSA: International Conference on Cybernetics and Information Technologies, Systems and Applications
  - ISAS: International Conference on Information Systems Analysis and Synthesis
  - CISCI: Conferencia Iberoamericana en Sistemas, Cibernética e Informática
  - SIECI: Simposium Iberoamericano de Educación, Cibernética e Informática
  - WCAC: World Congress in Applied Computing
  - Any IPSI International Conference or journal
  - KCPR: International Conference on Knowledge Communication and Peer Reviewing
  - International e-Conference on Computer Science

TOC

- How to start?
- How to find research problems?
- How to write research papers?
- How/Where to submit?
- Ethics
- Misc.
Fabrication

- From http://en.wikipedia.org/wiki/Fabrication_%28science%29
- “Fabrication, in the context of scientific inquiry and academic research, refers to the act of intentionally falsifying research results, such as reported in a journal article. Fabrication is considered a form of scientific misconduct, and is regarded as highly unethical. In some jurisdictions, fabrication may be illegal…”

Plagiarism

- From http://en.wikipedia.org/wiki/Plagiarism
- “… According to Diana Hacker, "Three acts are plagiarism: (1) failing to cite quotations and borrowed ideas, (2) failing to enclose borrowed language in quotation marks and (3) failing to put summaries and paraphrases in your own words…”
Eg, Fabrication and Plagiarism


- “Constantinos V. Papadopoulos got caught plagiarism at EUROPAR (1995)... 7 papers published and 8 under submission… all plagiarized from Technical Reports…”
  - [http://www.sics.se/europar95/plagiarism.html](http://www.sics.se/europar95/plagiarism.html)

- NEVER, EVER, do these – professional suicide !!

TOC

- How to start?
- How to find research problems?
- How to write research papers?
- How/Where to submit?
- Ethics
- Misc.
Personal Research Log

- Maintain personal research log
  - Sketch your research ideas into a writing
  - Update your ideas as time passes
  - Occasionally go back to old writings
- Prepare a short review for each paper that you read
  - Summary
  - Pros and cons
  - Limitations or problems
  - If needed, contact authors and ask questions
    Usually authors are willing to discuss with their readers

Professional Society

- Be a member of your professional society early on
- Ask your advisor to support membership
- Use the mentor program of societies
References (available at)
http://pike.psu.edu/resources/advice/

- [2002] How to write a paper?, Junghoo Cho, UCLA
- [1996] David Dill's Advice on Choosing an Advisor (or) How to Survive as a Grad Student, David Dill
- [1996] How to Survive as a Graduate Student, Brian Noble, David Dill, Benli Pierce, Jay Sipelstein, Jonathan Shewchuck
- [1997] How to Choose a Thesis Advisor, Michael C. Loui
- [????] How to have your abstract rejected, Mary-Claire van Leunen and Richard Lipton
- [1994] Dissertation Advice, Olin Shivers
- [1999] Advice for Finishing that Damn Ph.D., Daniel M. Berry
- [1999] So long, and thanks for the Ph.D., Ronald T. Azuma
- [2001] How to Have a Bad Research Career, David A. Patterson