Research Methods in Computer Science
(Serge Demeyer — University of Antwerp)

AnSyMo
Antwerp Systems and software Modelling
http://ansymo.ua.ac.be/

Universiteit Antwerpen
Helicopter View

(Ph.D.)
Research

How to perform research? (and get "empirical" results)

How to write research? (and get papers accepted)

How many of you have a systematic way of writing an abstract?
Sometimes I will give advice based on personal experience or representing a particular school of thought. These do not necessarily confirm with what your supervisor says!

Such opinions will be flagged with the Personal Opinion Disclaimer.
2. Reporting & Reviewing Research

Introduction
- The Publication Process
  + Publication Categories
  + Quality indicators
The Review Process
- Identify the Champion
- Implications for Authors
  + The 4-line abstract rule
  + The fish model
  + Natural emphasis of paragraphs
- Things to avoid
  + Method vs. Methodology
The Task of the referee
- Questions to answer ⇒ Review Template

Once Accepted ...
- Tips and Tricks

Conclusion
Publications: Output Measure

“If I have seen a little further it is by standing on the shoulders of Giants.”

(Isaac Newton)

“Are We Polishing a Round Ball?”

(Michael Stonebraker; Panel abstract — Proceedings of the Ninth International Conference on Data Engineering)

Sceptic perspective:
• the quest for the “least publishable unit”
• “publish or perish”

"And since dissertations can be written about everything under the sun, the number of topics is infinite. Sheets of paper covered up with words pile up in archives sadder than cemeteries, because no one ever visits them, not even on All Souls' Day. Culture is perishing in overproduction, in an avalanche of words, in the madness of quantity. That's why one banned book in your former country means infinitely more than the billions of words spewed out by our universities."

(Milan Kundera, The Unbearable Lightness of Being; Part Three: Words Misunderstood — Sabina's Country)
What is (Ph.d.) Research?

http://gizmodo.com/5613794/what-is-exactly-a-doctorate
Publication Categories

Journal Publications
• a1) citation index (ISI web of science)
• a2) international; peer reviewed
• a3) national; peer reviewed
• a4) other

Books
• b1) book
• b2) chapter
• a3) editor (incl. proceedings)

Other
• c1) articles in proceedings
• c2) technical reports; extended abstracts; thesis
• c3) patents

Comparing apples and oranges
International vs. National
• inherently regional research (law, politics, ...)
• vulgarizing research
• scientists taking position in society debates

Publication Culture
• co-authorship (e.g. alphabetical sorting)
• citation behavior
• half-life time of ideas

source: guidelines for project reports
FWO (Research Fund Flanders)
Publication Categories — Computer Science

Journal Publications
- citation index (ISI web of science)
- international; peer reviewed

Conference Publications
- peer reviewed
- (acceptance ratio)

Books
- book
- editor (incl. proceedings)
- chapter

Artifacts
- tools
- patents

Other
- workshops
- technical reports; extended abstracts; thesis

Artifacts ???
The Pipeline Model

new, unpolished idea → good, valuable idea → archival reference

workshops → conferences → journals

Typical for computer science. Not in other scientific disciplines.
Quality Indicators

Proceedings: Acceptance Ratio

  [Acceptance ratio: 37/147 = 25%]

  [Acceptance ratio: 27+4/65 = 42%]

Journal Publications: Impact factor

  [SCI impact factor 1.967, ranked 7 / 79]
## Acceptance Rates

Source http://people.engr.ncsu.edu/txie/seconferences.htm

<table>
<thead>
<tr>
<th>Top General SE Conferences</th>
<th>ICSE</th>
<th>FSE/ESEC</th>
<th>ASE</th>
<th>OOPSLA</th>
<th>ECOOP</th>
<th>ISSTA</th>
<th>FASE</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009</td>
<td>50/405(12%)</td>
<td>32+7/217(15%)</td>
<td>38+33/222(17%)</td>
<td>25/144(17%)</td>
<td>25/117(21%)</td>
<td>25/93(27%)</td>
<td>30/124(24%)</td>
</tr>
<tr>
<td>2008</td>
<td>56/371(15%)</td>
<td>31/152(20%)</td>
<td>34+36/280(12%)</td>
<td>33/117(28%)</td>
<td>27/138(20%)</td>
<td>26+9/100(26%)</td>
<td>30/141(21%)</td>
</tr>
<tr>
<td>2007</td>
<td>49/334(15%)</td>
<td>43+20/251(17%)</td>
<td>37+40/312(12%)</td>
<td>33/156(21%)</td>
<td>25/160(16%)</td>
<td>22/10?/(21%)</td>
<td>30/141(21%)</td>
</tr>
<tr>
<td>2006</td>
<td>36/395(9%)</td>
<td>25/125(20%)</td>
<td>22+12/121(18%)</td>
<td>26/157(17%)</td>
<td>21/160(13%)</td>
<td>22/84(26%)</td>
<td>27/166(17%)</td>
</tr>
<tr>
<td>2005</td>
<td>44/313(14%)</td>
<td>32/201(16%)</td>
<td>28+35/291(10%)</td>
<td>32/174(18%)</td>
<td>24/172(14%)</td>
<td>--</td>
<td>22/99(22%)</td>
</tr>
<tr>
<td>2004</td>
<td>58/436(13%)</td>
<td>25/169(15%)</td>
<td>25+26/183(14%)</td>
<td>27/173(16%)</td>
<td>25/132(19%)</td>
<td>26+2/93(28%)</td>
<td>22/91(24%)</td>
</tr>
<tr>
<td>2003</td>
<td>42/324(13%)</td>
<td>33+9/168(20%)</td>
<td>22+20/170(13%)</td>
<td>26/147(18%)</td>
<td>18/88(20%)</td>
<td>--</td>
<td>20/89(22%)</td>
</tr>
<tr>
<td>2002</td>
<td>48/303(15%)</td>
<td>17/128(13%)</td>
<td>19+19/94(20%)</td>
<td>25/125(20%)</td>
<td>24/96(25%)</td>
<td>18+8/97(19%)</td>
<td>21/60(35%)</td>
</tr>
<tr>
<td>2001</td>
<td>47/268(18%)</td>
<td>29/137(21%)</td>
<td>32+28/164(20%)</td>
<td>27/145(18%)</td>
<td>18/108(17%)</td>
<td>--</td>
<td>22/74(30%)</td>
</tr>
<tr>
<td>2000</td>
<td>49/335(14%)</td>
<td>17/92(18%)</td>
<td>23+22/100(23%)</td>
<td>26/142(18%)</td>
<td>20/109(20%)</td>
<td>17+4/73(23%)</td>
<td>21/60(35%)</td>
</tr>
<tr>
<td>1999</td>
<td>50/269(19%)</td>
<td>29/141(21%)</td>
<td>25+25/123(20%)</td>
<td>30/152(20%)</td>
<td>20/183(11%)</td>
<td>--</td>
<td>13/?</td>
</tr>
<tr>
<td>1998</td>
<td>41/209(20%)</td>
<td>19%</td>
<td>24+20/150(16%)</td>
<td>?</td>
<td>24/124(19%)</td>
<td>16/47(34%)</td>
<td>18/59(31%)</td>
</tr>
<tr>
<td>1997</td>
<td>50/219(23%)</td>
<td>27/194(14%)</td>
<td>32+15/108(30%)</td>
<td>?</td>
<td>20/103(19%)</td>
<td>--</td>
<td>?</td>
</tr>
<tr>
<td>1996</td>
<td>52/213(24%)</td>
<td>?</td>
<td>?</td>
<td>16%</td>
<td>21/173(12%)</td>
<td>16+8/69(23%)</td>
<td>?</td>
</tr>
<tr>
<td>1995</td>
<td>28/155(18%)</td>
<td>29/150(19%)</td>
<td>?</td>
<td>18/90(20%)</td>
<td>--</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

| Submission Deadline | Aug 29 | March 16 | May 4 | March 19 | Dec 17 | Jan 30 | Oct 2 |

- [100% - 50%]: not selective
- [50% - 30%]: reasonably selective
- [30% - 15%]: selective
- [15% - 0%]: too selective !?
Ranking of Conferences

http://portal.core.edu.au/conf-ranks/

http://valutazione.unibas.it/gii-grin-scie-rating/
### Impact Factor — Citation Index

#### Journal Citation Reports®

**Journals from:** COMPUTER SCIENCE, SOFTWARE ENGINEERING

**Sorted by:** Impact Factor

#### Journal Summary List

**Ranking is based on your journal and sort selections.**

<table>
<thead>
<tr>
<th>Mark</th>
<th>Rank</th>
<th>Abbreviated Journal Title (linked to journal information)</th>
<th>ISSN</th>
<th>JCR Data</th>
<th>Eigenfactor™ Metrics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Total Cites</td>
<td>Impact Factor</td>
</tr>
<tr>
<td>✔</td>
<td>1</td>
<td>ACM T SOFTW ENG METH</td>
<td>1049-331X</td>
<td>729</td>
<td>3.956</td>
</tr>
<tr>
<td>✔</td>
<td>2</td>
<td>IEEE T SOFTWARE ENG</td>
<td>0098-5589</td>
<td>5449</td>
<td>3.569</td>
</tr>
<tr>
<td>✔</td>
<td>3</td>
<td>ACM T GRAPHIC</td>
<td>0730-0301</td>
<td>4083</td>
<td>3.383</td>
</tr>
<tr>
<td>✔</td>
<td>4</td>
<td>J WEB SEMANT</td>
<td>1570-8268</td>
<td>438</td>
<td>3.023</td>
</tr>
<tr>
<td>✔</td>
<td>5</td>
<td>COMMUN ACM</td>
<td>0001-0782</td>
<td>12617</td>
<td>2.646</td>
</tr>
<tr>
<td>✔</td>
<td>6</td>
<td>IEEE MICRO</td>
<td>0272-1732</td>
<td>1478</td>
<td>2.565</td>
</tr>
<tr>
<td>✔</td>
<td>7</td>
<td>ACM T MULTIM COMPUT</td>
<td>1551-6857</td>
<td>155</td>
<td>2.465</td>
</tr>
<tr>
<td>✔</td>
<td>8</td>
<td>IEEE T VIS COMPUT GR</td>
<td>1077-2626</td>
<td>2224</td>
<td>2.445</td>
</tr>
<tr>
<td>✔</td>
<td>9</td>
<td>J ACM</td>
<td>0004-5411</td>
<td>5727</td>
<td>2.339</td>
</tr>
<tr>
<td>✔</td>
<td>10</td>
<td>MATH PROGRAM</td>
<td>0025-5610</td>
<td>4658</td>
<td>2.336</td>
</tr>
<tr>
<td>✔</td>
<td>11</td>
<td>IEEE INTERNET COMPUT</td>
<td>1089-7801</td>
<td>1568</td>
<td>2.309</td>
</tr>
<tr>
<td>✔</td>
<td>12</td>
<td>IEEE T MULTIMEDIA</td>
<td>1520-9210</td>
<td>2010</td>
<td>2.288</td>
</tr>
<tr>
<td>✔</td>
<td>13</td>
<td>IEEE MULTIMEDIA</td>
<td>1070-986X</td>
<td>708</td>
<td>2.258</td>
</tr>
<tr>
<td>✔</td>
<td>14</td>
<td>ACM T MATH SOFTWARE</td>
<td>0098-3500</td>
<td>2111</td>
<td>2.197</td>
</tr>
<tr>
<td>✔</td>
<td>15</td>
<td>IEEE SOFTWARE</td>
<td>0740-7459</td>
<td>2371</td>
<td>2.099</td>
</tr>
<tr>
<td>✔</td>
<td>16</td>
<td>COMPUTER</td>
<td>0018-9162</td>
<td>3133</td>
<td>2.093</td>
</tr>
<tr>
<td>✔</td>
<td>17</td>
<td>IEEE T DEPEND SECURE</td>
<td>1545-5971</td>
<td>381</td>
<td>2.093</td>
</tr>
<tr>
<td>✔</td>
<td>18</td>
<td>J DATABASE MANAGE</td>
<td>1063-8016</td>
<td>263</td>
<td>2.000</td>
</tr>
<tr>
<td>✔</td>
<td>19</td>
<td>IBM SYST J</td>
<td>0018-8670</td>
<td>1599</td>
<td>1.883</td>
</tr>
<tr>
<td>✔</td>
<td>20</td>
<td>IEEE COMPUT GRAPH</td>
<td>0272-1716</td>
<td>1930</td>
<td>1.866</td>
</tr>
</tbody>
</table>
The h-index

Represent both
- scientific productivity
- scientific impact
⇒ in a single number (measurement)

A scientist has index h if
- h of [his/her] Np papers have at least h citations each, and
- the other (Np − h) papers have at most h citations each.

Sources to calculate
- ISI web of knowledge
  http://isiknowledge.com/
- UAD - Search
  http://quadsearch.csd.auth.gr/
Quality Indicators — Beware

- impact factor of journal ≠ impact factor of article
  + Joint Committee on Quantitative Assessment of Research (June 12, 2008). "Citation Statistics". International Mathematical Union.

- #citations ≠ impact
  + Carlo Ghezzi; Reflections on 40+ years of software engineering research and beyond an insider's view (ICSE 2009, keynote)

- “The widespread practice of counting publications without reading and judging them is fundamentally flawed.”

- “If used unwisely, as is increasingly the case, they discourage people (young ones in particular) right from the outset from daring to think, from exploring new paths [...]
The Reviewer

- volunteer  
  + don’t waste his/her time

- curious  
  + catch his/her interest

- constructive  
  + supervises other Ph.D.

- influential  
  + wants to support “valuable” papers

- anonymous  
  + avoid tampering

... unfortunately ...

- busy  
  + read’s on train, bus, air-plane, ...
Review Process Steps

Bidding for Abstracts
abstracts + key-words
= “first date” with your reviewer

source: CyberChair (http://www.CyberChair.org)
Providing Keywords

As many as possible? vs. As few as possible?

- Automated reasoning techniques
- Component-based systems
- Computer-supported cooperative work
- Configuration management
- Domain modelling and meta-modelling
- Empirical software engineering
- Human–computer interaction
- Knowledge acquisition and management
- Maintenance and evolution
- Model-based software development
- Model-driven engineering and model transformation
- Modeling language semantics
- Open systems development
- Product line architectures
- Program understanding
- Program synthesis
- Program transformation
- Re-engineering
- Requirements engineering
- Specification languages
- Software architecture and design
- Software visualization
- Testing, verification, and validation
- Tutoring, help, and documentation systems
Writing Abstracts

Descriptive Abstract
• outlines the topics covered in a piece of writing
  + reader can decide whether to read entire document
• \(\approx\) table of contents in paragraph form.

Informative Abstract
• provides detail about the substance of a piece of writing
  + readers remember key findings
  + reviewers find the claims
• \(\approx\) claim and supporting evidence in paragraph form

\(\neq\) executive summary
(abstracts use the same level of technical language)
4-line abstract guideline

- source: Kent Beck “How to Get a Paper Accepted at OOPSLA”

- 1) states the problem
  + WHO is suffering the problem?
  + Connect with your target audience
- 2) why the problem is a problem
  + WHY is it a problem?
  + Cost / Art rather than a science / …
- 3) startling sentence
  + WHAT is the claimed solution?
  + the one thing to say that will catch interest
  … and that you will actually demonstrate in the paper
  ➡ must be falsifiable
- 4) the implication of my startling sentence
  + WHERE can we use this solution?
  + implications for society, community, other researchers, …
Identify The Champion (1/2)

- source: Oscar Nierstrasz, “Identify the Champion,” in Pattern Languages of Program Design 4

- Make Champions Explicit
  - + A: Good paper. I will champion it at the PC meeting.
  - + B: OK paper, but I will not champion it.
  - + C: Weak paper, though I will not fight strongly against it.
  - + D: Serious problems. I will argue to reject this paper.

  ➡ “The most important thing for a reviewer to decide is whether he or she thinks that the paper is worth defending at the PC meeting, not whether it is a great paper or not.”

- Make Experts Explicit
  - + X: I am an expert in the subject area of this paper.
  - + Y: I am knowledgeable in the area, though not an expert.
  - + Z: My evaluation is that of an informed outsider.

  ➡ detect inexpert champion — expert fence-sitter

These scores are *not* revealed to the authors
Identify The Champion (2/2)

- Identify the Conflicts (classify according to extreme reviews)
  + AA, AB: All reviews are positive, at least one champion.
  + AC: Likely accept; at least one champion, and no strong detractor.
  + AD: This is a serious conflict, and will certainly lead to debate.
  + BC: Borderline papers, no strong advocate nor a detractor.
  + BD: Likely to be rejected.
  + CC, CD, DD: Almost certain rejects.

- Inexpert champion
  + If all champions are Y (or Z)
  + If all reviews are Y or Z
    ➡ solicit extra review

- Expert fence-sitters
  + Experts tend to be more critical
    ➡ B or even C ratings by X may turn out to be champions
      (remember: PC members want to influence the research)
2. Reporting & Reviewing Research

- Clear accept at top
- Clear reject at the bottom (not shown)
- middle area: to discuss
Make it Easy for your Champion

• Select appropriate keywords
  + Why are you in the scope of the conference/journal/...?

• Test the abstract
  + Start early with the abstract
  + Ask for early (external) feedback

• Visible claims
  + Abstract + intro + conclusion have visible claim(s)
  + Ask early feedback to summarize what reviewers think the claim is

• Clear validation
  + Champion is then able to defend it against detractors

• Write to the Program Committee
  + Target a PC member
  + Have a clear picture of your champion
(1) The problem
- who has the problem?
- why is it a problem?
- what is the (sketch of) the solution?

(2) The problem context
- why is it a difficult problem?
  (few bibliographical details)
- which aspect of the problem do you tackle?
- how can you show that you solved the problem?
  (criteria / units of analysis / ...)

(3) The solution
- detailed description
  (sufficient for replication)
- empirical evidence

(4) The problem context revisited
[a.k.a. “Related Work”]
- I only solved one aspect of problem
- others have worked on it
  (many bibliographical details)
- future work (short term)
⇒ together we made progress

(5) The conclusion
- quick summary of solution
  + empirical evidence
- implications of the solution
- future work for the community
  (long term)
2. Reporting & Reviewing Research

Role of “Related Work”

- Problem Statement (beginning of paper)
  - Other researchers define the research agenda
    - ⇒ high entry barrier
      (for experts only)

- Problem Context (end of paper)
  - Other researchers do complimentary work
    - ⇒ crisp problem statement
      (difficult to write)
2. Reporting & Reviewing Research

Target Audience

Experts in sub-domain (in-crowd)

= preaching to the quire

Broader Audience (informed outsider)

= arguing the problem and inviting others to contribute

- Conferences: ICSE, ESEC/FSE
- Journals: TSE, TOSEM
Books on writing

- *The Elements of Style*
  William Strunk Jr., E. B. White

- *Style: Toward Clarity and Grace*
  Joseph M. Williams, Gregory G. Color

- 18 simple guidelines
  + elementary rules of usage
  + elementary rules of composition

- You have to know the rules before you can break them

- guidelines
  + *refactoring* rules

- Give a man a fish and you feed him for a day. Teach a man to fish and you feed him
Skimming texts — Emphasis

“natural” emphasis of paragraphs
• 1rst 1/2 of last sentence (most)
• 2nd 1/2 of first sentence

On section/chapter level
• say what you gonna say
• say it
• say what you have said

Source: Joseph M. Williams, “Style: Toward Clarity and Grace” The University of Chicago Press 1990
How to structure your writing

*The last thing one discovers in writing a book is what to put first* [Blaise Pascal]

all of us ... must understand three things about complex writing:
- it may precisely reflect complex ideas
- it may gratuitously complicate complex ideas
- it may gratuitously complicate simple ideas

© Joseph M. Williams, “Style: Toward Clarity and Grace”

<table>
<thead>
<tr>
<th>FIXED</th>
<th>Issue</th>
<th>Discussion</th>
</tr>
</thead>
<tbody>
<tr>
<td>VARIABLE</td>
<td>Point</td>
<td>—</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>FIXED</th>
<th>Topic</th>
<th>Stress</th>
</tr>
</thead>
<tbody>
<tr>
<td>VARIABLE</td>
<td>Old/Familiar</td>
<td>New/Unfamiliar</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>FIXED</th>
<th>Subject</th>
<th>Verb</th>
<th>Complement</th>
</tr>
</thead>
<tbody>
<tr>
<td>VARIABLE</td>
<td>Characters</td>
<td>Action</td>
<td>—</td>
</tr>
</tbody>
</table>
Things to Avoid

• report order ≠ investigate order
  + arguments should appear in order that bests support the claim

• unsubstantiated claims, hopes, assumptions
  + XXX will make it easy/fast/better/integrate with other tools ...
    - do you actually demonstrate these claims in your paper ?
  + We believe ..., We hope ...
    - My favorite reviewing sentence:
      “We are doing science, not religion ...”
  + XXX is valuable ..., XXX can help ..., XXX is an excellent ...
    - My favorite reviewing sentence:
      "Are these opinions? Hypotheses?
       Proven facts? Please add references."

• tackling a non-problem, a problem which you cannot solve
  + A software engineering example
    - papers citing “Software Crisis”
Things to Avoid: Methodology

• “In this paper we propose a methodology for XXX”
  + My favorite reviewing sentence:
    - Do not use the word "Methodology" for something simple like a technique, algorithm or even method; this is inflation of words

• the postfix -OLOGY
  - biology = the study of the living organisms
  - psychology = is the study of the human mind
  - cosmology = is the study of the cosmos
  ➔ methodology = the study of the methods.

• method = a series of steps or acts taken to achieve a goal
  + substeps of method remain a method
  + cfr. Composite design pattern
The Task of a referee (1/2)


Decide
- Makes sufficient contribution?
  + depends on the standards of the journal/conference/workshop/…

Questions to answer
- What is the purpose of this paper?
- Is the paper appropriate? (for computer science / software engineering / reengineering / …)
- Is the goal significant?
- Is the method of approach valid?
- Is the actual execution of research correct?
- Are the correct conclusions drawn from the results?
- Is the presentation satisfactory?
- What did you learn?
The Task of a referee (2/2)

Categories
- (1) Major results; very significant (fewer than 1% of all papers).
- (2) Good, solid, interesting work; a definite contribution (≤ 10 %)
- (3) Minor, but positive, contribution to knowledge (perhaps 10-30 %).
- (4) Elegant and technically correct but useless. This category includes sophisticated analyses of flying pigs.
- (5) Neither elegant nor useful, but not actually wrong.
- (6) Wrong and misleading.
- (7) So badly written that technical evaluation is impossible.
Reviewing Template

Review
- Strong accept / weak accept / weak reject / strong reject
  ➔ Including a solid motivation for your recommendation
- Template
  + summary (neutral)
  + strong points (bullet points)
  + points to improve (bullet points)
  + details
  + PC-only comments

Important for champion/detractor
Time estimation

1 paper = ± 4 hours

- 1,5 hour reading + annotating
  + read on paper
  ➝ submission for review incl. page numbers & white-space
- 1 hour writing review
- 1 hour discussion + adapting reviews
  + over mailing lists etc.
- 0,5 hour overhead
  + print papers (write numbers on them !!!)
  ➝ “first contact” with the papers
  + managing conference reviewing system
  + distribute among co-reviewers
  + ...

Ph.d. students as Co-reviewer

- 2nd opinion (reduces time spent for “reading” and “writing review”)
- Ph.d. students experience “the other side of the fence”
- Mentioned in the proceedings (CV)
Once Accepted ...

... at the Conference

- prepare an elevator-pitch
  + based around “startling sentence” from your abstract

- approach gurus
  + they like it, it's good for their ego

- "explain your Ph.d. topic to at least 3 persons each day"
  + = advice from ICSM 2009 Ph.d. symposium

- submit to Ph.d. symposium
  + receive valuable feedback
  + network with future peers

- participate in workshops
  + test how the community reacts to research questions
  + the gurus struggle too!
Conclusion

Introduction
• The Publication Process
  + Publication Categories
  + Quality indicators
The Review Process
• Identify the Champion
• Implications for Authors
  + The 4-line abstract rule
  + The fish model
  + Natural emphasis of paragraphs
• Things to avoid
  + Method vs. Methodology
The Task of the referee
• Questions to answer ⇒ Review Template

Once Accepted ...
• Tips and Tricks

Conclusion