"How to write, review and present a Scientific Paper in Computer Science"

Vasco Amaral
FCT/UNL 7/4/2021

online invited talk for ASPI course at NOVA School of Science and Technology

We got it! We did our scientific/technical work...

...and now what?

We need to communicate our findings!

Simplified Publication Process

Produce Content

Produce Report

Quality Control

Present

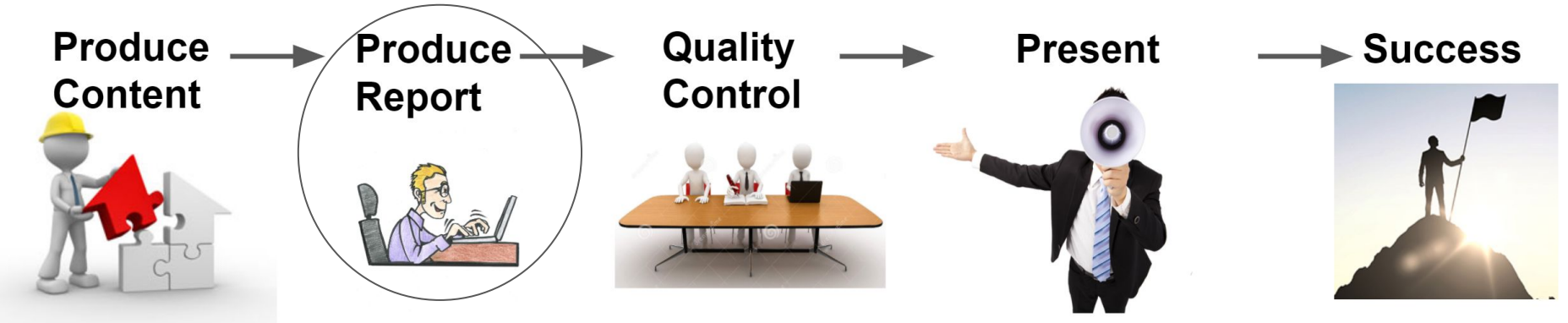
Success

Roadmap

Part I - How to write a CS paper?

Part II - How to criticize?

Part III - How to present?



Part I - How to write a CS paper?

curiosity

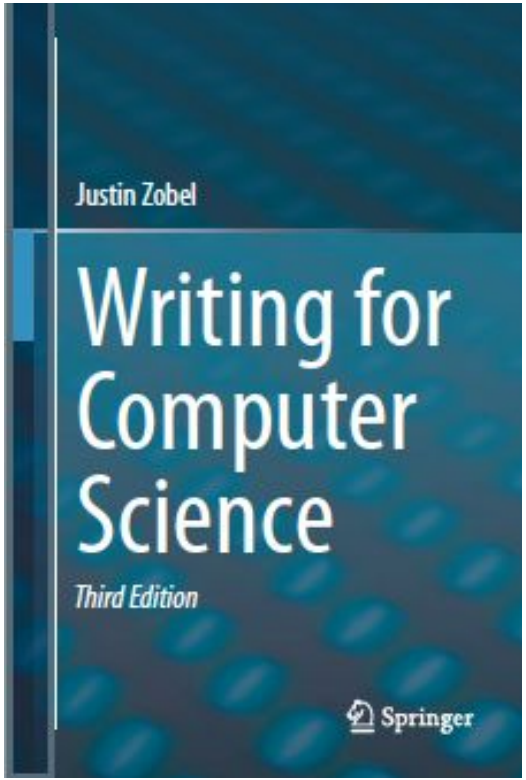
Pick a Text format (template)

https://www.ieee.org/conferences_events/conferences/publishing/templates.html

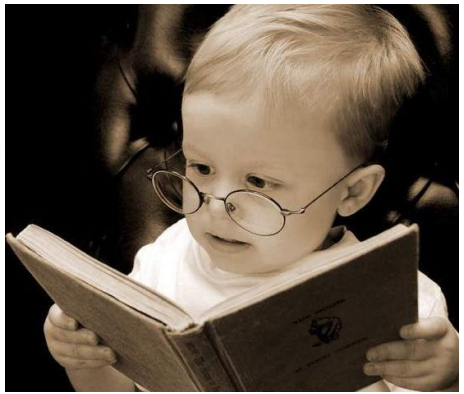
Attention to page limit!

This talk uses material provided by the book:

Writing for Computer Science

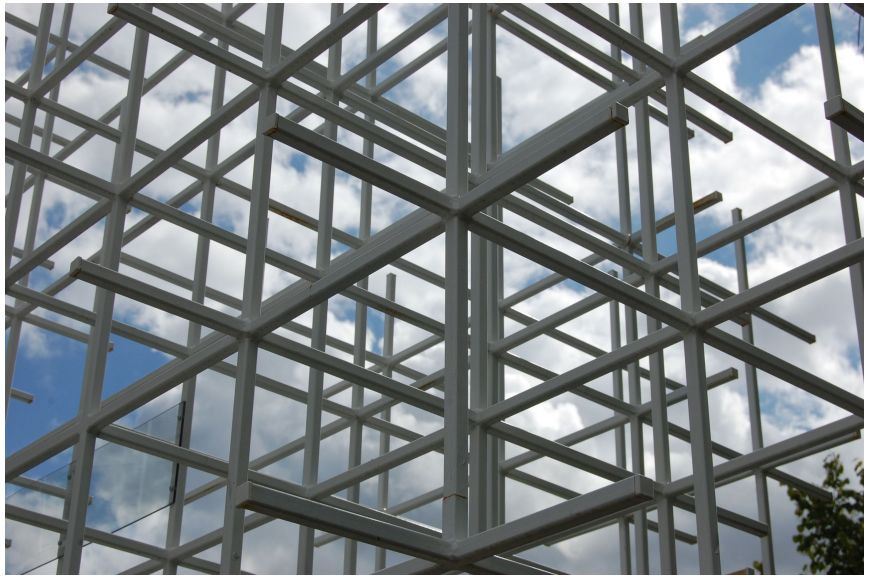


“The only way to produce a well-written paper is to **start early and revise often**.”



General Structure (proposal - depends on the purpose)

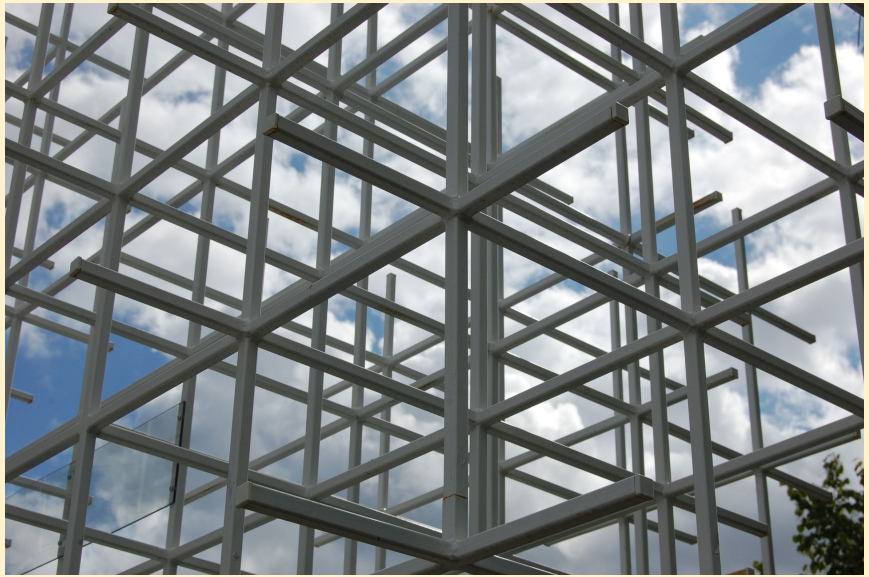
- Title
- Abstract
- Introduction
- State-of-the-art
- Solution/Body
- Results and Discussion
- Conclusions
- Acknowledgements
- Bibliography
- Annexes/Appendices



General Structure (the paper for this course where you survey a topic)

*Disclaimer: this is not a recipe!

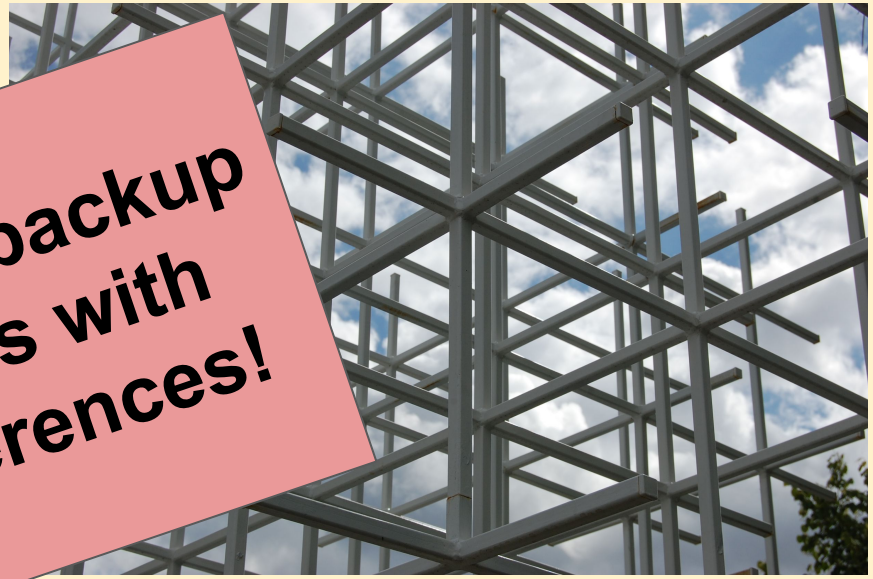
- Title
- Abstract
- Introduction
- Background
- State-of-the-art
  - technologies/techniques
- Discussion backed by:
  - Characterization of the problem
  - Comparative studies
  - Statistics
- Conclusions
  - Your vision of the relevance/future and impact to Society/Professionals/Technology evolution/etc...
- Bibliography



General Structure (the paper for this course)

*Disclaimer: this is not a recipe!

- Title
- Abstract
- Introduction
- Background
- State
  -
- Di
  -
  -
  - S
- Concl
  - You relevance/future and impact to Society/Professionals/Technology evol /etc...
- Bibliography



Don’t Forget to backup all your claims with adequate references!

“Procrastination is the enemy of good writing. There are always plenty of things
you might do first—whether they are sufficiently important is another question. Start now.”

Clear sign of procrastination...

Anonymous “Graffiti”, Wall at ED.II (in front of the cafeteria), FCT/UNL, 2018.

Some suggestions

- **Aim for simplicity and clarity**
- Start by the **background**, **Solution/Body** and **related work**
- Next write the results and discussion
- Leave for the **end the introduction and conclusions** and for the **very last the abstract**
- English text
  - Be consistent to use either american or british, consult this link)
  - Use spell checkers
  - Only capitalize proper names and abbreviations
  - Avoid contractions (don’t, isn’t, etc.)
  - Avoid colloquial terms
  - Avoid figures of speech
  - Avoid the use of adjectives
    - “our work to get there was overwhelming”
    - “we came up to a great solution”

Some suggestions

- Have **one idea per sentence or paragraph** and one topic per section.
- Have a straightforward, **logical organization**.
- Use **short words**.
- Use **short sentences** with simple structure.
- Keep **paragraphs short**.
- **Avoid buzzwords, clichés, and slang**.
- Avoid **excess, in length or style**.
- **Omit unnecessary material**.
- **Be specific**, not vague or abstract.

About gender

- Forms of expression that unnecessarily specify gender are widely regarded as sexist
- Avoid to use constructs such as “s/he” “he/she” “(s)he”, use “they”/”their” (Oxford dictionary).

About gender examples

From dictionary.com:

2 a form of the possessive case of singular they used as an attributive adjective, before a noun:

a (used to refer to a generic or unspecified person previously mentioned, about to be mentioned, or present in the immediate context):

Someone left their book on the table.

A parent should read to their child.

b (used to refer to a specific or known person previously mentioned, about to be mentioned, or present in the immediate context):

I'm glad my teacher last year had high expectations for their students.

c (used to refer to a nonbinary or gender-nonconforming person previously mentioned, about to be mentioned, or present in the immediate context):

My cousin Sam is bad at math, but their other grades are good.

Title

Condenses the paper’s content in a few words (10 to 12 words)

Captures the reader’s attention (Think about terms people would use to search for your study and include them in your title)

Differentiates the paper from other papers of the same subject area

Abstract

1. Introduction. **What's the topic?**
2. State the problem you tackle. **What's the key research question?**
3. Summarize **why nobody** else has adequately **answered** the research question yet?
4. How did you tackled the research question? **What's your big new idea?**
5. How did you go about doing the research that follows from your big idea? **Did you run experiments? Build a piece of software? Carry out case studies?**
6. **What's the key impact of your research?** Here we're not looking for the outcome of an experiment. We're looking for a summary of the implications. **What does it all mean? Why should other people care**? What can they do with your research.

Remember, the word 'abstract' means a summary of the main ideas with most of the detail left out. So feel free to omit detail!

Based on http://www.easterbrook.ca/steve/2010/01/how-to-write-a-scientific-abstract-in-six-easy-steps/

Abstract - Example

(1) In widgetology, it's long been understood that you have to glomp the widgets before you can squiffle them. (2) But there is still no known general method to determine when they've been sufficiently glomped. (3) The literature describes several specialist techniques that measure how wizzled or how whomped the widgets have become during glomping, but all of these involve slowing down the glomping, and thus risking a fracturing of the widgets.

(4) In this thesis, we introduce a new glomping technique, which we call googa-glomping, that allows direct measurement of whifflization, a superior metric for assessing squiffle-readiness.

(5) We describe a series of experiments on each of the five major types of widget, and show that in each case, googa-glomping runs faster than competing techniques, and produces glomped widgets that are perfect for squiffling. (6) We expect this new approach to dramatically reduce the cost of squiffled widgets without any loss of quality, and hence make mass production viable.

Based on http://www.easterbrook.ca/steve/2010/01/how-to-write-a-scientific-abstract-in-six-easy-steps/

Abstract - Example

(1) **Topic?** In widgetology, it's long been understood that you have to glomp the widgets before you can squiffle them.

(2) **Research question?** But there is still no known general method to determine when they've been sufficiently glomped.

(3) **Why nobody answered?** The literature describes several specialist techniques that measure how wizzled or how whomped the widgets have become during glomping, but all of these involve slowing down the glomping, and thus risking a fracturing of the widgets.

Based on http://www.easterbrook.ca/steve/2010/01/how-to-write-a-scientific-abstract-in-six-easy-steps/

Abstract - Example

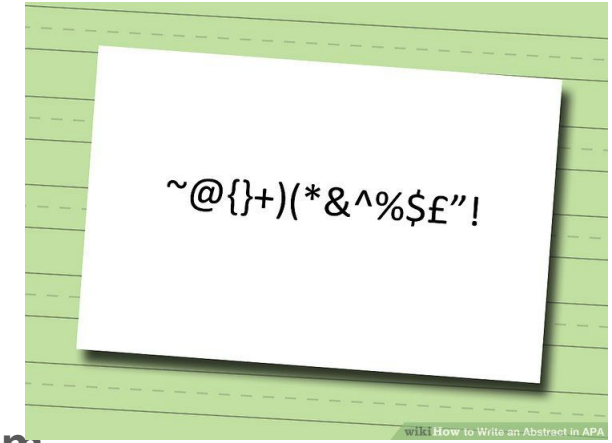
(4) **What is your idea?** In this thesis, we introduce a new glomping technique, which we call googa-glomping, that allows direct measurement of whifflization, a superior metric for assessing squiffle-readiness.

(5) **What was done?** We describe a series of experiments on each of the five major types of widget, and show that in each case, googa-glomping runs faster than competing techniques, and produces glomped widgets that are perfect for squiffling.

(6) **Impact?** We expect this new approach to dramatically reduce the cost of squiffled widgets without any loss of quality, and hence make mass production viable.

Based on http://www.easterbrook.ca/steve/2010/01/how-to-write-a-scientific-abstract-in-six-easy-steps/

Introduction



Seen as an **expanded abstract**

After reading it the reader must:

- Understand the **scope of the work and the problem**
- The **contributions** of the work
- Be convinced that it **worths reading** the rest of the paper

Novelty: Tells what in the paper is new

Result: What the outcomes are (not focusing on the how)

Introduction - Structure

- Describes the paper topic (**context**)
- The **problem under study well motivated**
  - Why is it interesting?
- **Approach** to the solution
  - Why is it good?
- **Scope and limitations** of the solution and **outcomes**
- (can describe the **structure** of the paper)

Related Work vs. Background

- Background - presents **contextual or prerequisite information** that is important or essential to understand the main body of your work
- Related Work - presents/discussed **similar work** in the presented report done by others

Literature Review

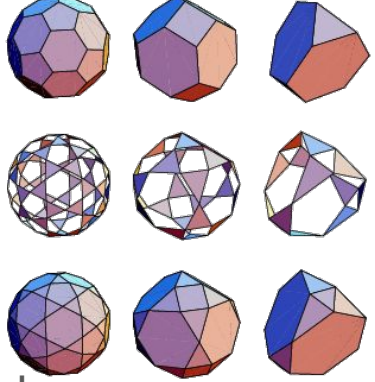
Relationship between new scientific results and proprietary commercial technology

From the perspective of research principles, the existence of a commercial product is irrelevant:

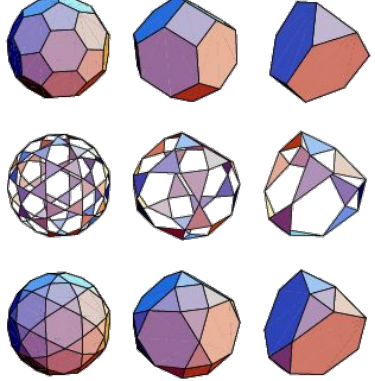
The **ideas are not in the public domain**, it is not known how the problems were solved in the product, and the researcher’s contribution is valid.

However should be mentioned...

Figures



- A figure has a **legend below** it
- The **legend** must be more or less **self-explaining**
- All elements of the figure (axis labeling etc.) must be large enough
- Use appropriate scaling of the axes
- Use clear symbols to distinguish different treatments/plots etc
- **Avoid color figures**
- Make graphs **clear and easy to understand**
- Check graphs from publications in highly ranked scientific journals and **learn** from these **how to create nice graphs**



Figures

Remember that when not yours:

- Proper **copyright permissions**.
- Legend with sentences like “**gently granted by**” (if that is the case).
- Or adapt the figure (if that is possible), by designing your own and put on the legend “**adapted from**”.

Table

Tables have a **header on top and explanatory remarks at the bottom**

Tables can contain lots of information.

To achieve this, **some thinking must be spent how to organize a table**.

TABLE. Estimated numbers* and percentages of HIV/AIDS† and AIDS cases, by year of diagnosis and selected characteristics — United States, 1981–2004

	AIDS						HIV/AIDS	
	1981–1995		1996–2000		2001–2004		2001–2004	
Characteristic	No.	(%)	No.	(%)	No.	(%)	No.	(%)
Sex								
Male	467,286	(84.7)	173,608	(75.9)	120,242	(73.4)	112,237	(71.3)
Female	84,229	(15.3)	55,253	(24.1)	43,576	(26.6)	45,231	(28.7)
Age group (yrs)								
<13	7,668	(1.4)	1,426	(0.6)	341	(0.2)	1,025	(0.7)
13–19	2,748	(0.5)	1,659	(0.7)	1,480	(0.9)	4,336	(2.8)
20–29	98,990	(18.0)	30,161	(13.2)	19,632	(12.0)	31,503	(20.0)
30–44	336,967	(61.1)	137,963	(60.3)	90,581	(55.3)	80,063	(50.8)
45–59	89,530	(16.2)	49,658	(21.7)	44,862	(27.4)	34,882	(22.2)
≥60	15,612	(2.8)	7,996	(3.5)	6,921	(4.2)	5,660	(3.6)
Race/Ethnicity								
White, non-Hispanic	256,460	(46.5)	72,314	(31.6)	46,325	(28.3)	45,497	(28.9)
Black, non-Hispanic	190,561	(34.6)	107,618	(47.0)	81,057	(49.5)	80,310	(51.0)
Hispanic	98,438	(17.9)	45,529	(19.9)	33,185	(20.3)	28,725	(18.2)
Asian/Pacific Islander	3,660	(0.7)	1,868	(0.8)	1,788	(1.1)	1,360	(0.9)
American Indian/Alaska Native	1,490	(0.3)	858	(0.4)	736	(0.5)	768	(0.5)
Transmission category								
Male-to-male sexual contact	282,234	(51.2)	92,301	(40.3)	66,781	(40.8)	68,484	(43.5)
Injection-drug use (IDU)	147,724	(26.8)	63,766	(27.9)	37,308	(22.8)	27,227	(17.3)
Male-to-male sexual contact/IDU	42,966	(7.8)	13,903	(6.1)	7,954	(4.9)	5,725	(3.6)
Heterosexual contact§	55,449	(10.1)	54,384	(23.8)	49,276	(30.1)	53,489	(34.0)
Perinatal	7,028	(1.3)	1,410	(0.6)	333	(0.2)	882	(0.6)
Other¶	16,113	(2.9)	3,098	(1.4)	2,166	(1.3)	1,661	(1.1)
Vital status								
Living	119,606	(21.7)	156,170	(68.2)	141,755	(86.5)	146,431	(93.0)
Deceased	429,582	(77.9)	71,520	(31.3)	21,621	(13.2)	10,957	(7.0)
Total**	**551,515**	**(100)**	**228,863**	**(100)**	**163,818**	**(100)**	**157,468**	**(100)**

* Numbers do not represent reported case counts, but instead are point estimates, which result from adjustments of reported case counts. The reported case counts are adjusted for reporting delays and for redistribution of cases in persons initially reported without an identified risk factor. The estimates do not include adjustment for incomplete reporting. Data are from case reports received by CDC as of June 30, 2005.

† Data include persons with a diagnosis of HIV infection. This includes persons with a diagnosis of HIV infection only, a diagnosis of HIV infection and a later AIDS diagnosis, and concurrent diagnoses of HIV infection and AIDS. Since 2000, the following 35 areas have had laws or regulations requiring confidential name-based HIV infection reporting: Alabama, Alaska, Arizona, Arkansas, Colorado, Florida, Idaho, Indiana, Iowa, Kansas, Louisiana, Michigan, Minnesota, Mississippi, Missouri, Nebraska, Nevada, New Jersey, New Mexico, New York, North Carolina, North Dakota, Ohio, Oklahoma, South Carolina, South Dakota, Tennessee, Texas, Utah, Virginia, West Virginia, Wisconsin, Wyoming, Guam, and the U.S. Virgin Islands. Since July 1997, Florida has had confidential name-based HIV infection reporting only for new diagnoses.

§ Heterosexual contact defined as sexual contact with a person at high risk for or infected with HIV.

¶ Includes hemophilia, blood transfusion, and risk factor not reported or not identified.

** Includes persons with unknown sex, multiple races, unknown race or ethnicity, and unknown vital status. Columns might not sum to the column total because of rounding.

Body

Descriptions of **experiments should permit reproduction and verification**

Careful definitions of the hypothesis and major concepts

Results and Discussion

In this section it is explained **how the results were obtained** and describe the results as well as a **discussion of the outcome**.

The **threats to validity** should be highlighted with an explanation on how the authors took care to minimize them

Conclusions

Concise **statement of the paper's important results and explanation of relevance**

State or restate limitations (shortcomings in the experiment, problems that the theory does not address,...)

Go beyond the current context to problems not addressed, questions not answered or variations that could be explored.

Can speculate and discuss possible consequences.

Acknowledgements

It is really up to you what to write here and whom to thank for support

Bibliography

Set of references cited in the text: thesis, papers, books, reports.

Bibliography

[1] "LaTeX: a document preparation system", Leslie Lamport, 2nd edition (updated for LaTeX 2ε), Addison-Wesley (1994).

[2] "A Guide to LaTeX: document preparation for beginners and advanced users", Helmut Kopka and Patrick W. Daly, Addison-Wesley (1995).

[3] "The LaTeX Companion", Michel Goossens, Frank Mittelbach and Alexander Samarin, Addison-Wesley, (1994).

Appendices vs. Annexes

Both have the purpose to add extra information to the main document to support the flow of thought:

An **appendix** contains data that cannot be placed in the main document and has references in the original copy or file. written by the original author of the document. (ex: ethics approvals, extended tables of data, and transcripts of interviews.

An **annex**, on the other hand, is usually a standalone document that offers additional information than contained in the main document. Assumed to have been written by an external party

SCIgen

Left to right: Dan Aguayo, Max Krohn, and Jeremy Stribling in 2005

Photo: Frank Dabek

SCIgen - An Automatic CS Paper Generator

https://pdos.csail.mit.edu/archive/scigen/

About

SCIgen is a program that generates random Computer Science research papers, including graphs, figures, and citations. It uses a hand-written **context-free grammar** to form all elements of the papers. Our aim here is to maximize amusement, rather than coherence.

One useful purpose for such a program is to auto-generate submissions to conferences that you suspect might have very low submission standards. A prime example, which you may recognize from spam in your inbox, is SCI/IIIS and its dozens of co-located conferences (check out the very broad conference description on the **WMSCI 2005** website). There's also a list of **known bogus conferences**. Using SCIgen to generate submissions for conferences like this gives us pleasure to no end. In fact, one of our papers was accepted to SCI 2005! See **Examples** for more details.

We went to WMSCI 2005. Check out the **talks and video**. You can find more details in our **blog**.

Also, check out our 10th anniversary celebration project: **SCIpher**!

Interesting Sources

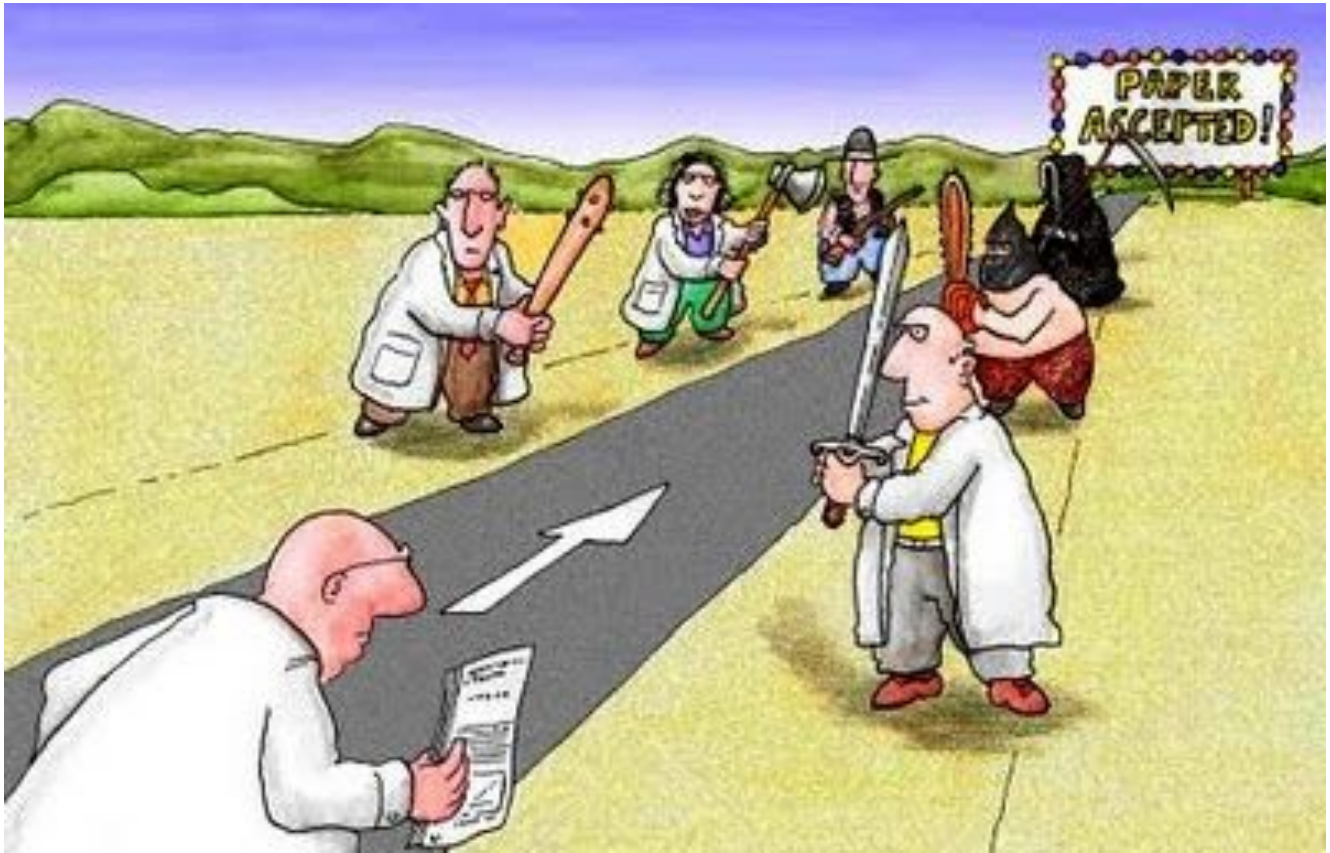
https://www.pordata.pt/

http://ec.europa.eu/eurostat/data/database?p_p_id=NavTreeportletprod_WAR_NavTreeportletprod_INSTANCE_nPqeVbPXRmWQ&p_p_lifecycle=0&p_p_state=normal&p_p_mode=view&p_p_col_id=column-2&p_p_col_pos=1&p_p_col_count=2

http://stats.oecd.org/

Part II - How to criticise?

Reviewing

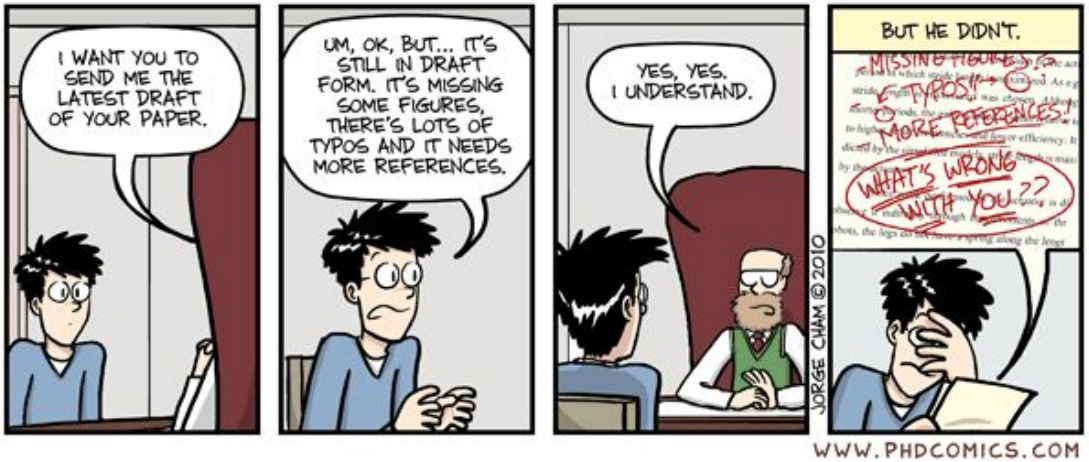


Why reviewing?

Reviewing is a central part of the scientific process

Criticism and analysis of papers written by other scientists is the main mechanism for identifying good research and eliminating bad

Is arguably **as important an activity as research itself**



Give Fair and Constructive Remarks

Be sensitive to the feelings of the authors (they are your peers)

What to take into account

Is the paper following the given format/template? Are figures readable? Are references formatted properly?

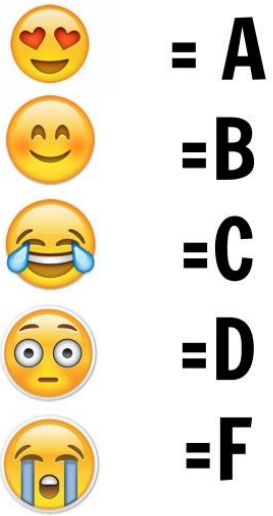
Is it well structured? Does the abstract describe what you read? Do intro and conclusion tell the story on their own?

Is the study rationale clearly articulated?

Is there a logical sequence?

Are the results clear? Does it answer to the question?

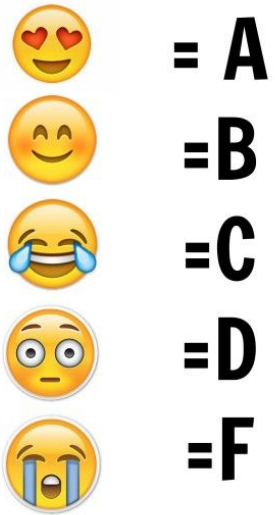
Are there unsubstantiated claims?



What to take into account

Frequent grammar errors, repetitions, incomplete sentences, typos, are **indicators that the paper was written on a rush**.

Therefore, no careful proofreading was done before submission.



Possible review format

1. Summary
2. A brief description using your own words and interpretation of what the work is presenting.
3. Summary of positive and negative remarks
4. Detailed Revision

    4.1. Majors

    4.2. Minors

Grading

From 1 (Bad) to 5 (Excellent)

- Presentation
- Writing
  - Content and relevance
  - Quality of the search
  - Quality of the information
- Global Perspective

Some don’ts...

Never use value judgements

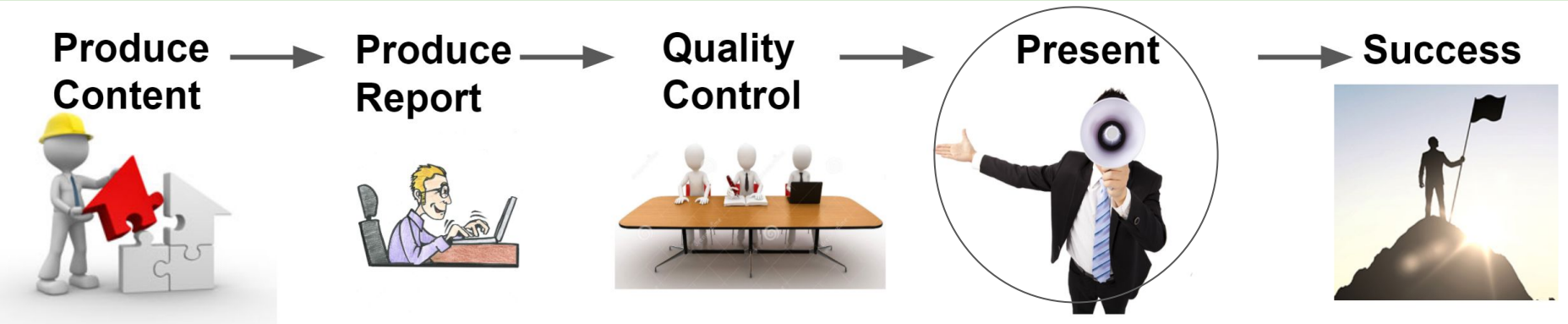
Backup your claims/suggestions (usually the major ones) with citations and strong arguments

Try to be **constructive and give suggestions** to improve the paper.

Etiquette - Use a calm and friendly but also neutral and objective tone

Distinguish statements of **fact** from your own **opinions**.

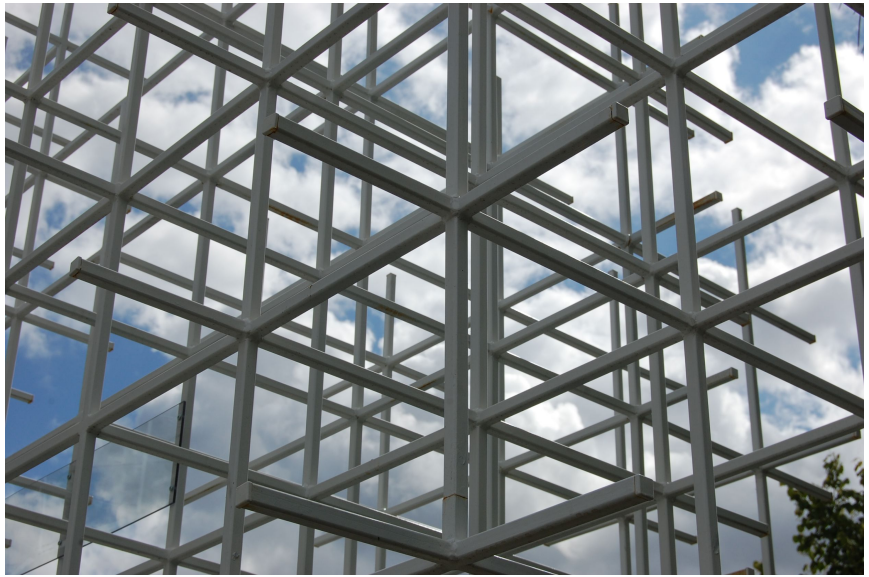
Remember that we are **critiquing the work**, not the authors.



Part III - How to present?

Slides: General Structure

- Title slide
- Index (optional)
- Motivation
- Research Question
- Background and State of the Art
- Content
- Analysis
- Final Conclusions



Some guidelines

- **Number your slides** and make sure page numbers are visible on each slide.
- **Avoid jargon** and define abbreviations that might not be obvious to your audience.
- **Be consistent** throughout your presentation in the use symbols and nomenclature, both in figures and text
- **Make sure figures are large enough and good quality**: legends, axis labels and axis titles should be readable

Some guidelines

- The specific layout of your slides is something personal, but please **use common sense** when preparing your presentation.
- Make sure all text is **easily readable on a beamer**: use large enough fonts, avoid dark backgrounds, avoid yellow or green font colours.
- **Slides should not be too dense**: as a rule of thumb, more than 7 lines of text on a slide is too much.
- Finally, **run a spell-checker** over your slides.

Some guidelines

- Roughly estimate **1 slide per available minute** (15 minutes means 15 slides)
- Make use of comparison tables, diagrams and schematics (1 picture worths a thousand words)
- Avoid excessive animation (distracts the audience)

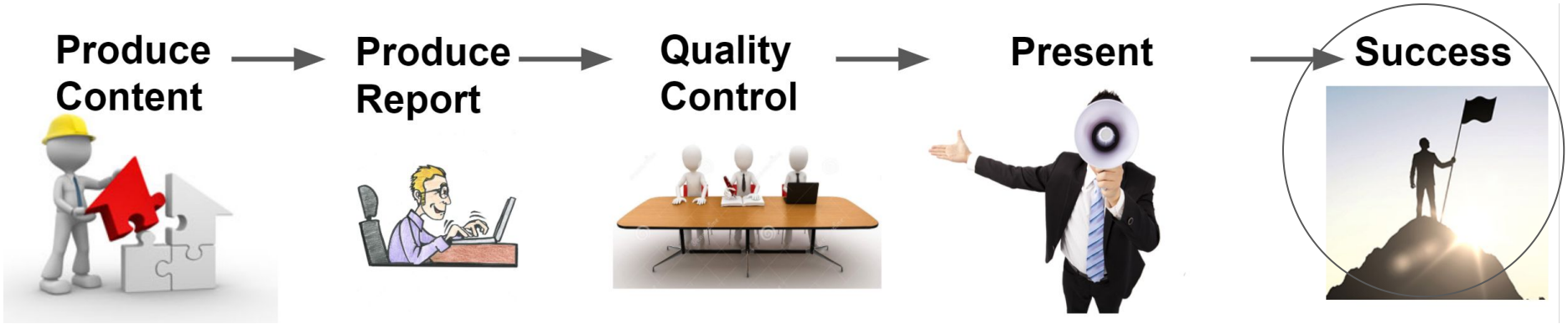
Oral presentation guidelines

- You should show up looking like you feel **it is an important event**
- You don’t need ties and jackets and sophisticated dresses, but you should **show that you put some thought** into the event
- **Speak with confidence**, with good speaking pace, good volume and sound that you know what you are talking about.
- **Practice** your talk out loud (to avoid being nervous) if possible **with someone** else to tell you how you are doing
- Look at your audience while you are speaking (COVID19: **keep your camera on**)

Oral presentation guidelines

- If you **practise** time enough **you will see that the beginning will come almost automatically**, which is nice to boost your confidence when you are nervous and your mind is “frozen”. (when you realize that you have not died yet your brain will kick back in and do the rest)
- Also, practicing **helps you relax and reduce nervous gestures** (like walking back and forward systematically) **or use of repetitive words**.
- **Don’t read your slides!** (the audience reads faster than you can talk…)
- Don’t be afraid to use some **humour** on your slides (but do not exaggerate as you are not doing stand-up comedy…)

Conclusion



Reporting is an essential part of research and development.
To master it is a warrant that you will pass your message with success.

However...

There is no point if there is no content! So, start working!

Thank you!

Thesis Life-cycle