Writing academic papers in Computer/ Information Science

Jaap Gordijn, Hans van Vliet
Vrije Universiteit Amsterdam
Computer Science Department
Faculty of Sciences
Contact: j.gordijn@vu.nl
www.cs.vu.nl/~gordijn

Writing academic papers is about

- Interesting and sound content
  - Content is *King*
  - Your *argumentation* should be *sound*
  - You should *very explicit*; not relying on reader's imagination
- Telling *your* content to *others*
  - *A good structure* helps others to understand your content
  - *Knowing your audience* helps to create an understandable story
  - Audience should be *attracted*

Some practicalities regarding *your* paper

- Allowed languages are Dutch and English. Make a choice; do not *mix* the two languages
- See Blackboard

How you should interpret the following slides …

- *Content* (the message that you want to communicate) is the *first class citizen*
- There are many ways to write a paper, therefore the following slides contain *guidelines*, not strict *rules*

Content is King: Choose an interesting topic

- How do you find a topic?
  - Which course(s) did you consider as interesting?
  - Is there some subject area you are knowledgeable about?
- In general:
  - Choose a topic that *you* are familiar with
  - Choose a topic *you* consider worthwhile to spend significant time on

Know your research community

- Understand which *research community* is interested in your topic
- Example communities
  - Software Engineering, Information Systems, Computer Networks, Artificial Intelligence, Management Studies, …
- Explain *explicitly* in your paper why your story is *of interest* for the community
  - "This paper is of interest to … because …"
  - "This paper contributes … to …"
- Explain *assumptions and concepts* not known to the community
  - If you talk about "networked value constellations" in a Software Engineering community, you have to explain what it means, and why it is relevant

Be explicit: Justify your *choices*

- Avoid *interpretation by the reader*, rather *tell* him/her how to interpret your paper
  - " You should understand here 'services' as *commercial* services and not as *web services*".
- *Justify* you choices
  - "We selected this particular group of persons for our test because they all spend a significant amount of time in developing open source software" (if you do research in open source software engineering)

Components of an academic paper

- An academic paper contains the following elements
  - *Abstract*: A short text explaining the added value of your paper (about100 words)
  - *Introduction*: Summary/contribution of the paper, plus rationale, approach and context
  - *Body*: The actual paper with a proper line of reasoning. Usually the body contains multiple *sections*, each discussing a particular *message*
  - *Related work*: Work done by other researcher, which is relevant to your own work
  - *Conclusion*: Statement of the most important finding. The conclusion *does not contain new concepts/arguments/viewpoints*; it just summarizes what has been said
  - *References*: cf. the prescribed style

A paper has several layers of abstraction

- Title: Tells the main message; *Invitation to read the abstract*
  - Abstract: Summary of the paper; *Invitation to read the introduction*
    - Introduction: Tell want you are going to tell; *Invitation to read the body*
      - Section: A particular message
        - Paragraph: Part of the message
          - Sentence: Part of the message
- A paper has a logical flow in sentences, paragraphs and sections
  - Indicators for connections: "Thus …", "However …", "In contrast", "Therefore …", "In conclusion"
- Sentences contribute to the message of a paragraph, a paragraph contributes to the message of a section, etc.
  - If it is not clear *how a sentence contributes to the paper*, you should consider removal of the sentence

Choose a title that *invites* potential people to read the paper

- The title should *invite* to read the abstract, and attract *readers*
  - Not: "The digital stamp"
  - But "The digital stamp has no right to exist"
  - Not "Data mining for marketing"
  - But: "For marketing purposes, the mining of unstructured data is not needed"
  - Not: "Information security"
  - But: "Information security should be driven by business risks"
  - Not "e-Services and their role in on-line shopping"
  - But: "Additional e-services are important for buy-decisions of products via the Internet"
- Avoid "*what is …*" stories; rather *claim something and argue* for it
- *In short: A good title tells in* **one** *sentence the contribution of the paper*

The abstract *motivates* to read your contribution

- The abstract is the *intro for introduction*
  - Purpose: *Invites* to read *at least* the introduction section, therefore the abstract should be **motivating**.
  - Contains the problem discussed in the paper **plus your own contribution**
  - Bad abstracts
    - Focus on the problem only and not really on the contribution
    - Discuss the structure of the paper; Do this in the last part of the introduction
  - An abstract may contain sentences of a well-written introduction
- An abstract contains no references, references to figures or tables or other side issues; If necessary do this in the introduction section. Only the *core issues* in the abstract. Usage of '(' and ')" in the abstract is usually a bad sign as they signal side issues.
- *Write the abstract first* (then you are forced to tell very briefly what you are going to tell) but *rewrite your abstract* when you finish the paper
- Tips: http://writing.wisc.edu/Handbook/presentations_abstracts.html

Exercise: Improve this abstract

- "*It is indicated that 80% of a company's information is contained in text documents (Tan, 1999). Some of the vendors claim that text mining has higher potential than data mining because analyzing 20% of the data would not give accurate predictions. In the first part of this paper we try to examine the information needed for marketing functions, the data type and the value of analyzing structured and unstructured data. In the second part we will try to give some drawbacks mining textual data. In the third part we give a small framework for data mining for marketing functions.*"

Structure in the abstract helps

[Context] Increasingly, software is developed in an open source setting. **[Problem]** Coordination while sharing files between individuals participating in an open source project is a major problem. **[Objective]** The aim of our research is to facilitate efficient and effective coordination between open source developers while editing source files. **[Contribution]** In this paper, we contribute a source control system that facilitates the editing of a source file by *thousands of users* at the same time. **[Method]** We evaluate our source control system in an open source project with a large amount of users. [**Conclusion**]. It shows that our system increases the efficiency and effectiveness significantly but only for open source communities with a maximum of 10,000 developers.

The introduction tells the complete story (1)

- Tell what you are going to tell
- The structure of an academic introduction section *usually* is:
  - *Context*: discussion of the problem area
    - "Increasingly, software is developed in an *open source* setting …"
  - *Problem* identification: The problem that you (partly) solve in the paper
    - "A key problem in open source software engineering is cooperation between individuals."
  - *Objective* of the research. Want are a goals?
    - "The aim of our research is to facilitate the concurrent editing of source files by many open source developers at the same time"
  - An indication of the *solution(direction)*.*Indicate clearly* your *contribution*, why should we use your work*?*
    - "We facilitate cooperation between individuals with a source control system. The key contribution of our system this allows for thousands of developers to work concurrently on the same source code repository"

The introduction tells the complete story (2)

- *Method* used: We *develop a prototype* of your source control system and *test* this prototype in a real-life large scale open source project with respect to increase in efficiency and effectiveness of the open source software developer
- An explanation of the *structure* of the paper.
  - "Section 2 presents an overview of existing source control systems. In section 3 we discuss our large scale source control system, followed by an evaluation in section 4. Finally, section 5 highlights our conclusions"
- Note: The introduction is an intro, not the story itself– which is in body of the paper (1 page introduction for a 4 page paper is quite lengthy; Then it takes too long before the reader encounters the 'meat' of the paper – the contribution of your paper)

Ensure that the introduction is attractive to the target audience

- Do not state the *obvious*:
  - "Information systems are important"
  - "Information is nowadays core to most business activities"
  - Rather, try to find motivating opening sentences for your paper
    - "The business model of software development is in a substantial transition phase: From closed source software developed by large companies to open source software by individuals"
- *Be consistent with abstract*: The introduction details the abstract, and can not suddenly discuss another topic.
- Write abstract – introduction – abstract – introduction …

Tips to evaluate your own introduction

- Ask *someone else* to read your introduction and ask him/her to summarize the message of the introduction
  - Check if the reader interprets the text as you intended
- *Read the text aloud* to yourself and others
- Check *connections* between sentences and paragraphs. Is there a *clear and logic story line*?
  - Remove sentences which do not fit into the flow

Body: For real; not for a course

- Use *formal language*
  - Not: "What are web-services actually?" (To indicate context)
  - But: "Web-services are…"
  - Not: "aren't"
  - But "are not"
  - Tip: Use a style that activates the reader: "We evaluate our source control system for open source projects by … " rather than "Our approach was evaluated …"
- Write a paper for a *conference or journal* (not for an academic course)
  - Not "This paper about digital stamps is a paper for the course …"
  - But "This paper argues that digital stamps are useful for…"
- No *judgment by yourself*
  - Not "The arguments are sufficient for our claim."
  - But leave the judgment to the reader
- Choose your best language (if it is allowed). Focus on communication of the message of the paper rather than writing in another language.
  - However: it *is* useful to practice writing in English

Body: Be sufficiently explicit taking into account the audience

- Use only terminology that your audience understands: "cyclomatic complexity", "SCTP", "MGCP"
- Usually, you can only solve *part of the problem*. Say explicitly which part you address
  - If your paper is dealing with reliability of Voice over IP (VOIP), this does not imply that companies *should use* VoIP.
- Be explicit in your reasoning (or: ensure that the reader interprets your as text as intended, no imagination)
- Logical connections between sentences, paragraphs and sentences. Useful check
  - Make a sequential list of your sections (and paragraphs) titles . Your list of section titles should read as a complete story

Body: Be precise and interesting

- Be *precise*:
  - Not" … has several advantages."
  - But: "… has several advantages, such as …"
- Avoid *bold* statements
  - "a biometric solution is *the solution* for …"
- Avoid section and paragraph titles with no *content:*
  - No-content titles: "Problem", "Solution", "Approach", "Method"
  - Check your story line by reading the sections titles aloud. They should tell your story.
  - Exceptions: "Introduction", "Conclusion" (However, a really good author also uses content-driven titles for these sections)
- Watch for *imbalance:*
  - Tell what you want to tell, but nothing more than that. No side-paths. Focus! The main contribution(s) of your work should cover the largest part of your paper

Related work to explain your work

- A "Related Work" section can serve *two purposes*
  - *Summarizing work of others*, which you *use* in your own contribution. Interpret the literature in the context of your own work. In this case, the related work section then usually follows the introduction
  - *Positioning your work in relation to work of others*. The related work section then usually takes the form of a discussion section and presents how you distinguishes yourself (sells the unique features) from others. Usually just before the conclusion section
- For both purposes: There should always be a good *reason* to include related work, it should *contribute* to your story somehow

The conclusion summarizes and suggests further research

- Tell what you have told
- Summarizes your *contribution and findings*
- May contain an outlook on *future work/remaining problems*
  - Motivate other researchers to participate in your area
- A conclusion *does not contain* new messages
- Usually, a conclusion is *short*

References: Be consistent and precise

- Be very *critical* and *careful* with information provided by *suppliers/companies*. In many cases these are not real scientific sources.
  - There are exceptions: Many large companies have their own research departments
- If you use statements of someone else *literally* then use "…". Avoid lengthy quotations. The paper should be *your* contribution
- Refer in text e.g. by [23] to work or opinions of others. Use a consistent style for referencing throughout the paper
- Where do you find academic papers?
  - Good starting points: Digital Libraries of the ACM, IEEE, Springer (LCNS, LNAI, LNBIP), and many more (access via VU network or use proxy)
  - Many conferences and workshops have their own web sites
  - Google scholar

Tip: Write first, revise later

- Writing is an *iterative process*: writing – reading – rewriting – reading …
- Start *writing at day one*; do not wait
- First, focus on *your own understanding* of the subject matter
- Second, focus on *explanation* of your subject matter to *someone else*
- Do *not focus to much on the length* of the paper initially;
  - For a 16 page paper, write first 18/20 pages and then reduce the paper to 16 full pages later on

Lack of space? Tips and tricks

- The '*kaasschaaf*' (cheese slicer) method (about 10/20% reduction)
  - Smaller, condensed figures (many figures contain useless white space)
  - No bullets but " … (1) …. (2) …. (3)" (in text)
  - Author names and affiliations side-by-side rather then vertically
  - Only references which are really needed
- The '*dor hout*' (dead wood) method (about 20/40% reduction)
  - Remove everything that does not directly contribute to the message of the paper
- The '*omhak*' (chop) method (>reduction of 40% or more)
  - Remove parts (paragraphs, sections) of your the paper. This usually results in a *different message* and therefore has a serious impact.
- Reducing a paper by applying the first two methods results in most cases in a better paper. Therefore, each author should apply methods one and two after writing a first version
- *Write first, then reduce.* A well written paper with a sound structure can be reduced easily

In sum …

- First and foremost: *Develop an interesting story.* Content is king
- Understand *your audience*, and know what they consider as interesting
- Be *explicit*; do not rely on the imagination of your reader
- Be *attractive*
- Have a proper consistent and logical story line on *all abstraction levels* (title, abstract, intro, sections, conclusion, paragraphs, sentences)