

# 50 Ways to Tweak Your Paper



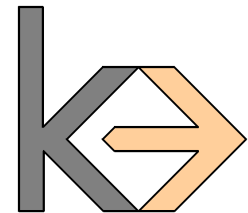
TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

## Some Comments on Paper Writing and Reviewing

**Johannes Fürnkranz**

TU Darmstadt  
Knowledge Engineering Group  
Hochschulstrasse 10  
D-64289 Darmstadt

`juffi@ke.tu-darmstadt.de`



# How did I end up here?



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT



**Bernard Ženko** <bernard.zenko@ijs.si>

May 9 ☆



to juffi, Tom ▾

Dear Juffi,

with Tom we are organizing the PhD Forum at the ECML-PKDD 2017 in Skopje, Macedonia, and we would like to invite you to give a **keynote talk at this forum**. The purpose of the forum is to provide a friendly environment **for junior PhD** students to exchange ideas and experiences with peers in an interactive atmosphere and to get constructive feedback from senior researchers in areas covered by the conference. The format will be similar as in previous years and it will include spotlight talks by students, mini panel and two keynote talks, which should focus on some topics of special interest to researchers at the beginning of their careers. Given that you are well known for your reviewing expertise, we were thinking about **a talk going somewhere along the lines of "How to write a paper to get it accepted** - view from the reviewers perspective." Of course, this is just an idea and you can freely adapt it. We very much hope that you will be able to accept our invitation.

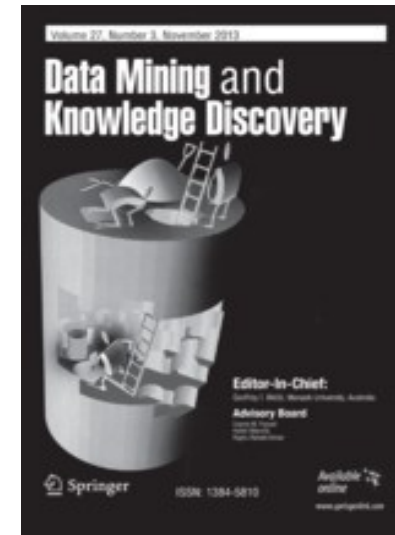
With very best wishes,

Tomislav Šmuc and Bernard Ženko  
PhD Forum at the ECML-PKDD 2017 Program Co-Chairs



# My Credentials

- Editor-in-chief of *Data Mining and Knowledge Discovery* journal
  - since 2014
- 23 years of reviewing experience
  - on both sides



- co-authored 3 submissions with students for ECML/PKDD 2017

all 3 of which were rejected...



# For Starters...

William Somerset Maugham:

There are three rules for the writing of a ~~novel~~ <sup>paper</sup>.

Unfortunately no one knows what they are...



# Caveats

- The following are **very simple observations**
    - chances are good that you already knew all of that
  - These are **subjective** opinions, formed by years of paper-writing and paper-reviewing
    - different people have different opinions
- *Don't blame me if your paper is rejected because you followed my advice*
- but feel free to use me as an excuse...



# 1. What's Your Point?

The number one reason for being rejected.

- Why has the world waited for this great new algorithm that you are proposing?
- Does it solve any real problem or is it just the umpteenth algorithm that can beat your favorite benchmark?

→ *Make it clear from the start, what problem you are trying to solve.*

- And don't forget to show that it actually does (and the others don't)

→ *Start with writing a good abstract!* (Tobias Scheffer)

- Your paper should be interesting without knowing the results!



# Types of Problems

---

## The Good

- We define and study a new type of problem. Nobody has considered such a setting before.
- We noticed that previous solutions to problem X all suffer from a certain problem. We propose an algorithm that can deal with it and show that it actually does.

## The Bad

- We propose a novel algorithm X for solving this well-known and important problem Y. There are many algorithms for solving this problem, but our solution X is novel and outperforms them all.

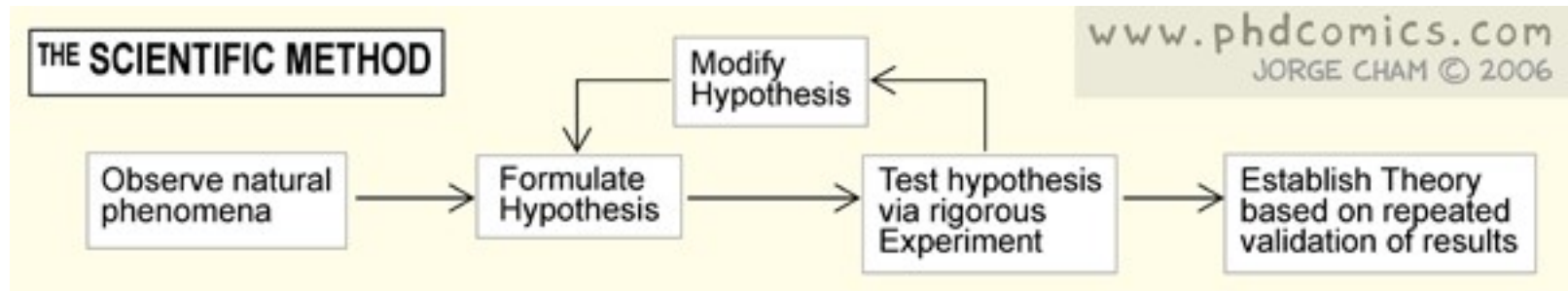
## The Ugly

- We explored a novel combination of genetic-algorithm-based feature selection with fuzzified decision trees trained optimized by ant colonies and showed that it outperforms C4.5 on breast cancer.



# Good Problems don't come for free

- ... sometimes you have to be creative





## 2. Be prepared to be judged by the cover!

- Reviewers will bid for papers by reading their titles
  - sometimes they also read the abstracts
  - they almost never look at the paper itself (at this stage at least)

→ *Try to think of a witty and original title!*



# Pick a Title

- My favorite title template:

*MCA: A new Framework/Method/Approach for X*

- Many paper titles fit this pattern

- Ensemble-Compression: A New Method for Parallel Training of Deep Neural Networks
- ALADIN: A New Approach for Drug--Target Interaction Prediction
- CON-S2V: A Generic Framework for Incorporating Extra-Sentential Context into Sen2Vec
- DeepCluster: A General Clustering Framework based on Deep Learning
- MRNet-Product2Vec: A Multi-task Recurrent Neural Network for Product Embeddings
- GaKCo: a Fast Gapped k-mer string Kernel using Counting
- WHODID: Web-based interface for Human-assisted factory Operations in fault Detection
- Boosted Trees: A scalable TensorFlow based framework for gradient boosting
- TrajViz: A Tool for Visualizing Patterns and Anomalies in Trajectory



# Pick a Title

## ■ My favorite title template:

- MixedTrails: Bayesian Hypothesis Comparison on Heterogeneous Sequential Data
- Vine Copulas for Mixed Data : Multi-view Clustering for Mixed Data Beyond Meta-Gaussian Dependencies
- FCNNs: Fourier Convolutional Neural Network
- PowerCast: Mining and Forecasting Power Grid Sequences
- BeatLex: Summarizing and Forecasting Time Series with Patterns
- Max K-armed bandit: On the ExtremeHunter algorithm and beyond
- PEM: Practical Differentially Private System for Large-Scale Cross-Institutional Data Mining
- Flash points: Discovering exceptional pairwise behaviors in vote or rating data
- TransT: Type-based Multiple Embedding Representations for Knowledge Graph Completion
- zooRank: Ranking Suspicious Activities in Time-Evolving Tensors
- TSP: Learning Task-Specific Pivots for Unsupervised Domain Adaptation
- UAPD: Predicting Urban Anomalies from Spatial-Temporal Data
- DC-Prophet: Predicting Catastrophic Machine Failures in Data Centers
- Delve: A Data set Retrieval and Document Analysis System
- MOB Lit@EVE: Explainable Recommendation based on Wikipedia Concept Vectors
- QuickScorer: Efficient Traversal of Large Ensembles of Decision Trees
- TrajViz: A Tool for Visualizing Patterns and Anomalies in Trajectory



# Find a Cool Title!

→ *You want to be different!*

- The most boring subject has a (small) chance of being accepted if you ent  
cha
  - Try
  - Mo
  - (b
- The Great Time Series Classification Bake-off
  - Tiers for Peers
  - To Tune or Not to Tune
  - ROC 'n' Rule Learning
  - The BOSS is concerned with time series classification in the presence of noise
  - Size matters
  - Research Re: Search and Re-Search
- as a (high)
- acronyms etc.



# Instructions for Finding a Title



## YOUR THESIS TITLE

CONDENSING OVER HALF A DECADE OF  
YOUR LIFE IN ONE SENTENCE.

www.phdcomics.com  
JORGE CHAM © 2006

**the colon**  
Can't decide what to title  
your thesis? Use a colon!

**a preposition**  
A good preposition tells your  
readers "hey, this is not just a  
futile exercise"

**"Witty catch-  
phrase"**

**:**

**Length-enhanced superlative  
verbiage with prolixity**

**in/of/  
for**

**Obscure topic few  
people care about.**

**witty catchphrase**  
Makes people think you're  
hip and culturally relevant.  
Only marginally related to the  
actual thesis? No problem.

**the boring stuff**  
Nothing says "academic rigor" like a  
long string of dry scientific-sounding  
terminology and fancy buzzwords.

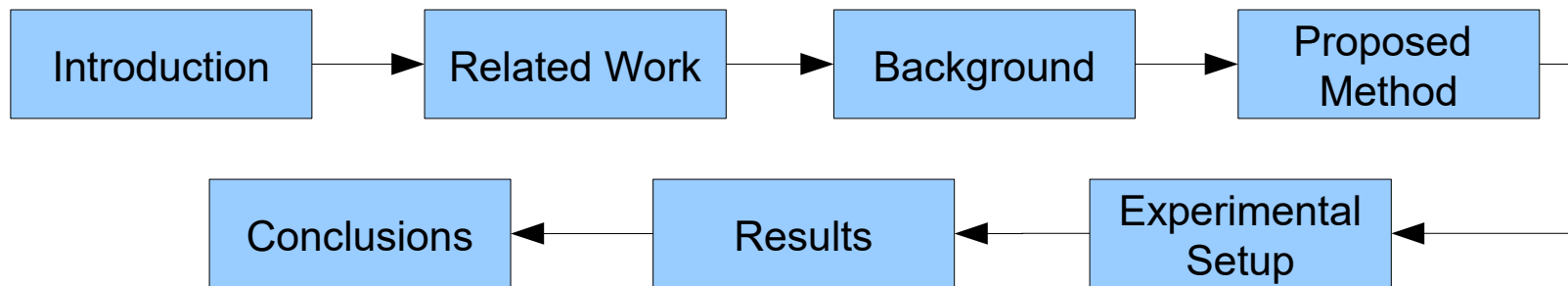
**obscure topic  
few people care  
about**  
Sad, but true.



### 3. Tell me a Story!

Your message needs to be framed into an interesting story

- Usually we have something like



- All this information needs to be there, but the structure may be adapted
  - Where to place “Related Work”?
  - “Proposed Method” is a really bad section name
  - Multiple Stages of the method may be evaluated incrementally, ...

→ *Fit the structure to your story, and not the story to some structure*

- But whatever you do, make sure to explain your structure!



# 4. Loosen up!

- Keep in mind that your reviewer does not want to spend much time in reading your paper
- A page full of text is intimidating...

→ *Add elements that partition the text!*

- figures
- headings and subheadings
- formulae
- ...

→ *Yes, this means that you have to sacrifice some text*

- but that's good for the reviewer (less work) and good for you (more focus)



## 5. Be Precise!

- Avoid commonplaces and phrases void of meaning

**DECIPHERING ACADEMESE** YES, ACADEMIC LANGUAGE CAN BE OBTUSE, ABSTRUSE AND DOWNRIGHT DAEDAL. FOR YOUR CONVENIENCE, WE PRESENT A SHORT THESAURUS OF COMMON ACADEMIC PHRASES

"To the best of the author's knowledge..."	=	"WE WERE TOO LAZY TO DO A REAL LITERATURE SEARCH."	"It should be noted that..."	=	"OK, SO MY EXPERIMENTS WERENT PERFECT. ARE YOU HAPPY NOW??"
"Results were found through direct experimentation."	=	"WE PLAYED AROUND WITH IT UNTIL IT WORKED."	"These results suggest that..."	=	"IF WE TAKE A HUGE LEAP IN REASONING, WE CAN GET MORE MILEAGE OUT OF OUR DATA..."
"The data agreed quite well with the predicted model."	=	"IF YOU TURN THE PAGE UPSIDE DOWN AND SQUINT, IT DOESN'T LOOK TOO DIFFERENT."	"Future work will focus on..."	=	"YES, WE KNOW THERE IS A BIG FLAW, BUT WE PROMISE WE'LL GET TO IT SOMEDAY."
			"...remains an open question."	=	"WE HAVE NO CLUE EITHER."

JORGE CHAM © 2004  
www.phdcomics.com

→ *Clearly say what you mean and mean what you say*

- How much "Future Work" promised in the past has been done by now?





## 6. Keep It Simple, Stupid!

Formal rigor and solid math and proofs are important, but

- your paper is not better if you make it harder to read
- It really helps if a reader that does not have a Ph.D. in Math can follow the main argument of your paper and has some guidance in where those details are that he can safely skip over.
- It also helps a reviewer in time pressure....

→ *Your paper should contain as much formal notation and math as necessary, but not more.*

- In particular, don't try to impress people by introducing a complex formalization of things that can be said in a sentence



## 7. Come again?

Some reviewers (not us, of course) are sloppy readers.

- They might miss your main argument if you only mention it once in passing.
- They might have forgotten on p. 12 what you already said or defined on p. 2.
- Or they might have chosen to skip over this part.

→ *Be redundant!*

- It can't hurt to remind the reader here and there that you have previously said
- But don't over-do it either (tricky to strike a balance here)



## 8. You are the best!

Nobody knows more about the subject you write than you

- Seriously.
- You have spent weeks and months on getting your results, you understand the problem very well
- There aren't many people that know as much about your Ph.D. topic than you do, and you have to be awfully lucky to get one of them as a reviewer
  - (and sometimes this means bad luck too...).



→ *Don't assume that the reader knows as much about the problem as you do*

- Try to write it for you grandma (or at least for this bachelor's student you are supervising).



## 9. You are **not** the best!

Nobody knows more about the subject you write than you do

- So your solution must be better than all previous attempts?  
...in all possible dimensions?  
...by a wide margin?

Really?

→ *Don't brag!*

- It doesn't help when you repeatedly emphasize how novel, performant, efficient, useful, robust, scalable your method is

→ *Let the evidence convince the reader*

- Make sure to think about potential disadvantages of your approach, and be candid about them



## 10. How Come?

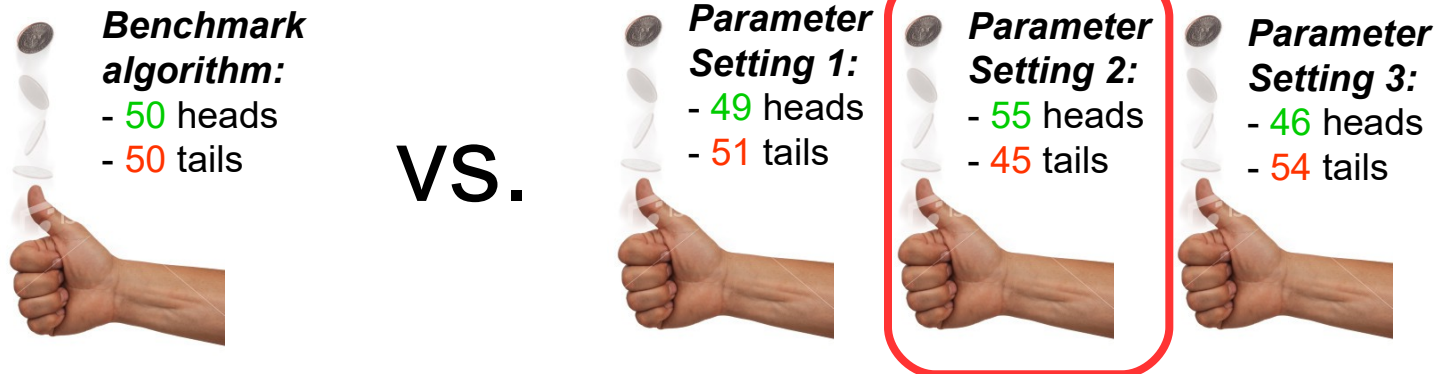
- Many papers have an experimental section that compares the new algorithm with several old algorithms
  - Why?
- *If you do experiments, state what you want to show with them.*
- Comparing to all the algorithms in Weka on all UCI datasets is nice, but you should have a reason for doing so, and you better state it in your article.
  - Better: Clearly state in what way your experiments show that you have sufficiently well solved the problem.
- *What is your criterion for success?*



# 11. Be fair!

Be careful if your algorithm has parameters.

- don't compare your algorithm with x parameter settings to your competitor in default configuration, and happily announce that there is a parameter setting where yours is better.
- That's like giving yourself x trials to toss more heads than you, while your competitor has only one.

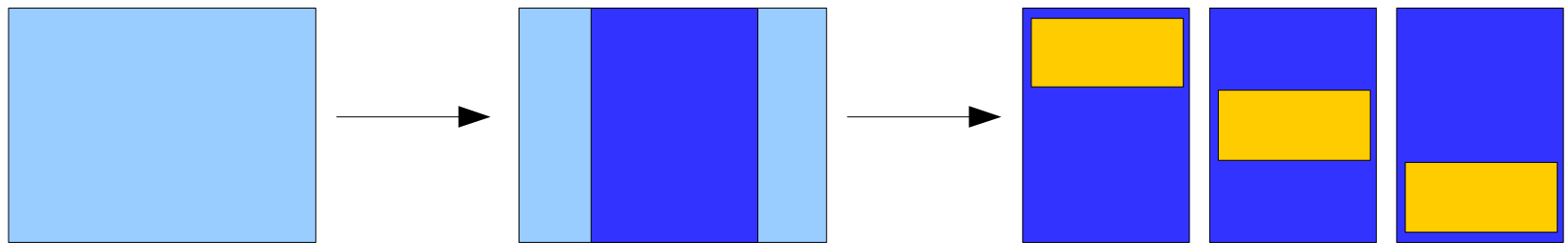


- And it is even worse if you simply don't say how many trials it took you to find that magical parameter setting.



## 12. Be Sound!

- When you do an evaluation, always keep in mind the real-world setting in which the algorithm will be used
- Think carefully about your experimental evaluation, and describe it in enough detail that your reviewer can think about it.
- Be extra-careful with cross-validation!
  - You must never look at the test set.
  - This may sound obvious, but there are countless papers rejected because of little mistakes like:



pre-processing of the dataset  
(discretization, feature subset selection...)

cross-validation  
**but you have already used these data!**



## 13. Spell it out!

- The results of our experiments can be found in Tables 5-10 and in Figures 11-15. We conclude that...
- How?

→ *Take the reader by the hand and explain every important detail.*

- explain what the reader should look at in these graphs and tables that you are showing.
- Don't let the results speak for themselves! They have a tendency to change with the eye of the beholder...





## 14. Make it shine!

- You want to sell something, so you better make it look good
- *Reserve a lot of time for final polishing!*
- Shift figures and tables to nice locations
    - also make sure their fonts are large enough, and that they are also readable in B/W (they will be printed that way)
  - Make nice page and line breaks
    - avoid single lines on top of pages etc.
    - etc.
  - But don't abuse the formatting guidelines!
    - Springer, e.g., will not print the papers camera-ready, but will re-process them from your tex-files
    - Your paper may turn out to be much longer than intended, and all your formatting work has gone.



## 15. Use a Spell-Checker!

- Some reviewers may think that tiny little things like formatting, proof-reading and spell-checking reflect the amount of work and thought you have given to the paper (shame on them!)

→ *Be nice to your reviewer, don't abuse her as a proof-reader!*

### Keogh's Maxim

**I firmly believe in the following:**

*If you can save the reviewer one minute of their time, by spending one extra hour of your time, then you have an obligation to do so.*

(Eamonn Keogh)



## 16. Conclusions, anybody?

Have you ever wondered why it is called “Conclusions” and not “Summary”?

- **Conclusions:**

- What have we learned from the work reported in this paper?
- What do we know why that we did not know before?

- **Summary:**

- What have you done in this paper?  
(You already wrote about that in the abstract, and the introduction)

→ *A good paper should clearly state what conclusions can be drawn from the reported work in the end.*

- You can also summarize on how you arrived at these conclusions, what experiments, proofs, or evidence supports them, but the new insights are the important part.



## 17. – 49. There is much more...

- ... unfortunately we have to omit it due to space and time constraints, so it is left for a future talk

→ *Use this sparingly (if at all)*

- You can't seriously expect to pack 50 things into a single talk
- You will often (always?) have more material to show than you can fit into the 5/8/15 pages that are allowed



→ *Don't try to press in everything, but focus on one (or a few) important aspects that you can present in detail*

**And don't run over time!!!**



# 50. Stay the distance!

(Arno Knobbe)



TECHNISCHE  
UNIVERSITÄT  
DARMSTADT

- You can't expect to get it right the first time

→ *Give your paper the chance to develop*

- **Workshop:**

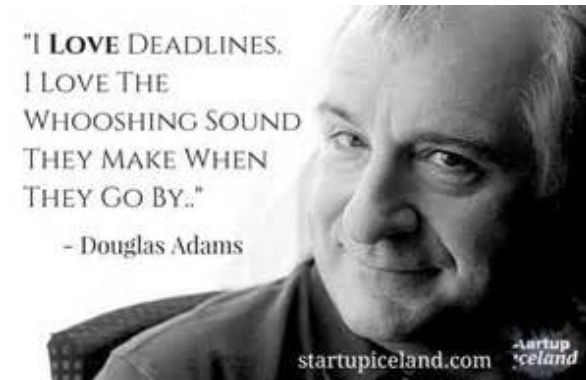
- first rough draft of the idea with preliminary results

- **Conference:**

- a solid version with some reliable results

- **Journal:**

- the definite version that should be the (your) last word on the subject



→ *Take the comments of your reviewers seriously!*

- If they haven't understood your paper, you probably didn't write it well enough!



# Conclusions

- Tell an interesting and coherent story
  - What problem do you solve?
  - How do you solve it?
  - Why is your solution a good solution?
- Tell it well
  - Focus on the message that you want to get across
  - Keep your writing simple and understandable
  - Spend time not only on the content but also on the appearance (layout, typesetting, formulas, figures and tables, references, etc.)
- ... and don't give up!



# Bibliography

- Eamonn Keogh, "How to Do Good Research, Get it Published in SIGKDD, and get it cited", 2009 SIGKDD tutorial (173 slides!)  
[http://www.cs.ucr.edu/~eamonn/Keogh\\_SIGKDD09\\_tutorial.pdf](http://www.cs.ucr.edu/~eamonn/Keogh_SIGKDD09_tutorial.pdf)
  - Excellent introduction into various aspects of doing data mining research, including many real-world examples
- Nikolaj Tatti: A data scientist's guide for writing papers. Tutorial @ECML PKDD 2016. <https://users.ics.aalto.fi/ntatti/howtowrite2016/>
  - I looked at it after the talk and it is indeed very good (thx Albrecht!)
- Bodil Holst: Scientific Paper Writing – A Survival Guide. CreateSpace Independent Publishing Platform, 2015
  - Illustrations by Jorge Cham
- Jorge Cham, Ph. D. Comics, <http://phdcomics.com>
  - A very insightful source of wisdom



