Manuscript

Why everything in Database Research has already been done (and why this is good for you)... at the SIGMOD 2018 New Researcher Symposium

By Jens Dittrich, Saarland Informatics Campus, http://twitter.com/jensdittrich

I will not use slides for my talk. While preparing this talk, I figured that this kind of speech does not need slides. I will tweet the manuscript on twitter afterwards.

In the following, I will present **25** dos and donts with the hope that they help you somehow in your career.

1. do not work on database performance

database performance is a solved problem

Stonebraker gave an interesting keynote at ICDE 2018 along the same lines: "no major breakthrough in the last 10 years")

not much innovation in query processing anymore, a lot of reassembly or reinventions of ideas plus fine-tuning

2. do not work on unrealistic workloads

hardly used by anyone other than Google, FB, etc. anyways

1 million transactions/sec per server? who cares?

3. accept that many new algorithms are just old wine translated from slower storage layers

=> **storage layer invariance:**

“if an algorithm works in-between say, DRAM and disk, you can often quite easily adapt it to work in-between any two other storage layers anyways.”

just a matter of constant fiddling/calibration,

example: storage with slow writes like flash: we see all kinds of good old logging/diff file techniques there...

joke in my research group: "who reviews the differential files paper this year" (for the past three years we have had at least one differential file paper per year...)

storage layer invariance works from disk to SSD to RAM to NUMA, to distributed computing to entire data centers, and most fancy new hardware

yeah, you can then give new names to make these algorithms look "new",

but:

the core idea of the algorithms will stay the same! And the contribution is kind of incremental.

4. do more experiments papers

we currently got 90% "research ~fancy new stuff" and 10% "experiments" papers

it should be >90% experiments papers, that would be **honest**

5. do decent experiments

I know this is hard.

with every paper, I am still learning more about experiments.

So many mistakes can be done in experiments.

I could talk for hours on this one.

But one thing we should all embrace is:

6. don't play the "let's pick a weak baseline to make my algorithm look good"-game

I know there is this tension: you want to have "good" results! You can easily achieve “good results” by "finding" the right baselines, datasets, parameters...

This really sucks!

=> I bet that >50% of the papers at VLDB/SIGMOD are not repeatable and/or compared against weak baselines, misleading workloads, etc.

in particular, in the area of **indexing**, ...

This diminishes our impact! How do you have impact, if it turns out afterwards that you compared against weak baselines and your results do not hold?

- ⇨ The Repeatability Debate.

BTW: other fields have the same problem. And it is a big issue.

7. Don't interpret "having good results" as "being better than X".

Rather interpret "having good results" as "having **sound** results".

A "sound result" gives guidance, allows readers to understand strengths AND weaknesses of a method, in particular in the light of the state-of-the-art.

The goal of an experiment is **NOT** to show-off that your method is better, the goal is to diligently contrast it with other methods.

This may be painful, but this is what I consider a "research paper".

8. love what you do:

don't do X to have Y

don't do X to become Z

do X because you totally love it

still haven't found X? Keep on looking!

9. trust your gut feeling. Always!

most major decisions I took with my brain turned out to be bullshit afterwards

picking topics, people to work with, things to try out, projects to invest energy in, …

How to decide? At a junction: if you have a choice between two topics,
visualize both topics in front of you, one on the left, one on the right,
feel where the energy goes to.

No energy, no excitement: stay away from it.

There is energy? Go for it!

10. have at least one insane project:

these projects may take time to unfold, it may will take time to convince people,

you will meet a lot of naysayers, but:

ignore them: trust your gut-feeling, don't worry: you will be rewarded, and if it is just for the fun and learning experience

11. view every new wave of Ph.D. students as an opportunity to reinvent yourself

switch topics after every Ph.D. or at least add different complementary topics **rather than subtopics**, try something new, learn!

12. don't force Ph.D. students into particular topics

first try to understand the particular talents of each student

then find a suitable problem for that student

example: students loving incremental work vs risky work

mapping an unsuitable topic to the wrong student is a total waste of time and energy for everyone

And the same applies to yourself as well:

13. understand well where YOU are good at:

YOU have a particular mix of talents: try to find the project where **YOU** can shine with that mix

14. balance incremental with risky work:

what is **incremental work**? e.g. yet another join paper [Yes, we had a SIGMOD 2016 paper on equi-joins…]

yet another "fitting an algorithm to some super-duper hardware"-paper [Yes, we have a DAMON 2018 paper that does exactly that]

this is groundwork, it helps you keep up to date with core database performance bit-fiddling stuff

yet: it has minor impact: it is polishing a very round ball

vs

risky projects: may fail, but if they work, they may make a big splash

what is a risky project?

fine line between stupid and brilliant idea

Albert Einstein once said:
"If an idea does not sound absurd in the beginning, then there is no hope for it."

15. in publishing pick quality over quantity

this helps building up a reputation,

focus on a few very good papers rather than many papers where too many are borderline

16. stay away from people constantly bragging about how cool they are

these people rarely are.

17. stay away from people who take themselves too seriously

seriousness and good research is not necessarily correlated

18. don't try to read everything, focus your reading efforts on important pivot researchers in your area (two or three)

alternatively: make your students read papers for you, it is called “seminar”

19. give credit to other people's work, cite where appropriate

if you know about a work that is relevant, you MUST cite it

this sounds so easy, but this is an issue in a considerable number of papers!

20. once you are a prof, make sure to still get your hands dirty on a regular basis: code and write!

Otherwise you will end up as some research manager playing buzzword bullshit bingo.

21. when in a discussion, if you don't understand anything, ask!

never rush over doubts in understanding something

admit that you do not understand some detail rather than pretending that you do. These "non-understood" things in discussions often lead to surprising research.

22. stay away from research politics (if you are not a politician)

research is a mix of actual research and (a lot of) research politics, too much in fact

personally, I hate research politics, and I am quite bad at it.

however, it is hard to avoid research politics

learn to differentiate:

having a lot of funding does not necessarily make you a better researcher or increases the fun (I once had a 1M Euro grant, it did not make me happy)

to me: a small group with great people is so much more valuable than a larger group where you make compromises on the student's quality

I am happy to have such a great group (Felix Martin Schuhknecht, Ankur Sharma, Immanuel Haffner, and Marcel Maltry), some of them are here at this conf.

And why all of this is good for you:

Currently, there are tons of real world problems out there,

the current data science wave is a **HUGE** opportunity for us

we are one third of what people mean when they say "data science" they just don't realize

help them realize that **data science** has a lot to do with what we have been doing over the past ~40 years

with that in mind:

23. try to solve some real-world problems, and make people happy

see my "**data science not equal machine learning**" **keynote** at the Machine Learning meets Databases Workshop, DEEM on Friday 8:30am for more on this

http://deem-workshop.org/

24. tell yourself every day: what a cool job to be a researcher and teacher!

And I would like to close with a quote by the great Steve Jobs:

25. "stay young, stay foolish"

Thanks!