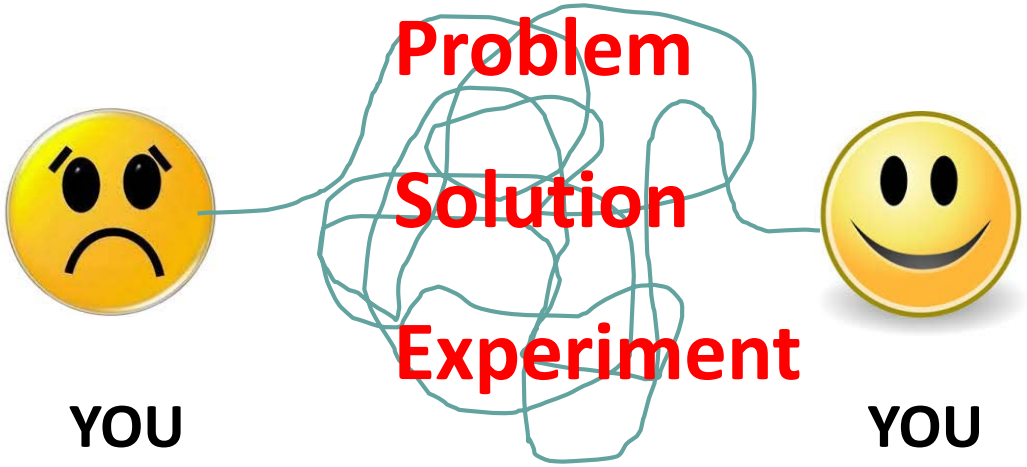
Tips for Writing Technical Papers

DR. YANYAN SHEN (沈艳艳)

SHANGHAI JIAO TONG UNIVERSITY

SHENYY@SJTU.EDU.CN

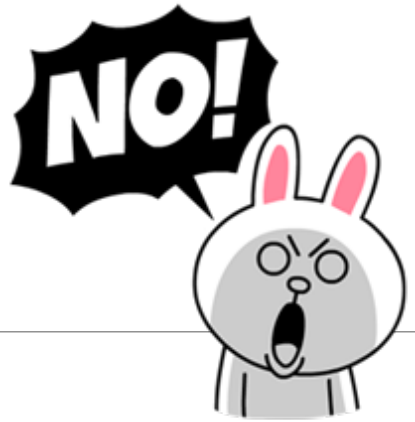
Writing is easy?

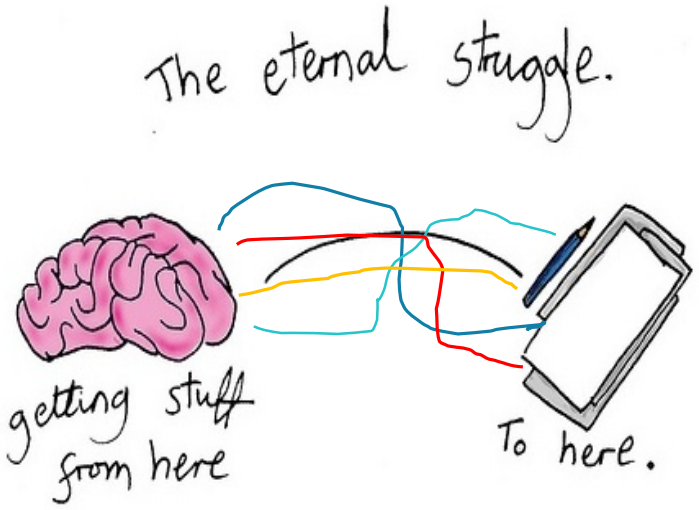


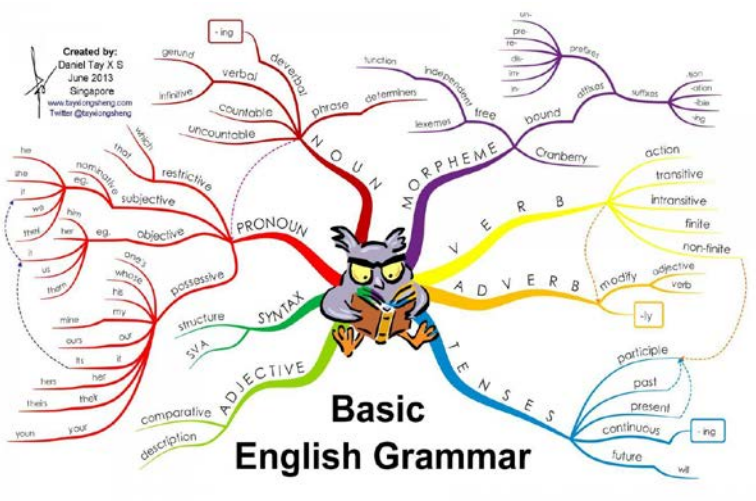
After a looooooooooooong journey…

Imagination

Writing is easy?







Correct?

Understandable?

Good?

Too many ways to say the same thing

Rules of English grammar

Requires: experience, patience, …

Writing is important?

From your supervisors:

- Very badly written!
- The English really sucks!!
- Don’t translate from Chinese!!!
- For goodness sake, learn to write better…

Keep Editing for You

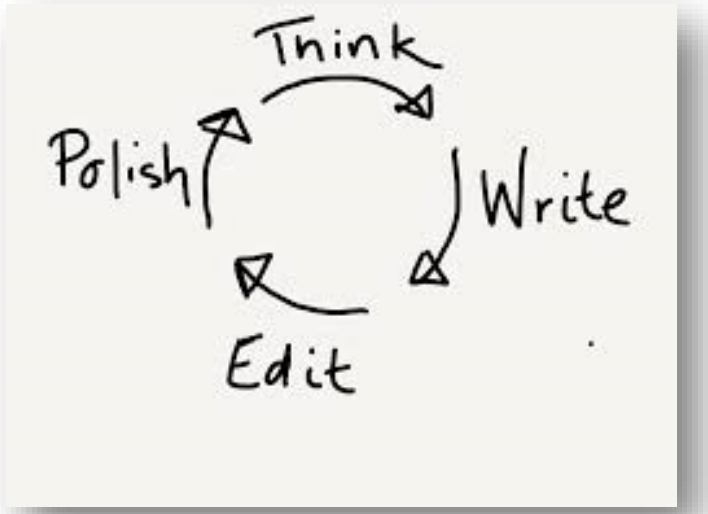
From reviewers:

- This paper is poorly written.
- The paper is hard to follow.
- The poor English makes understanding difficult in many places.

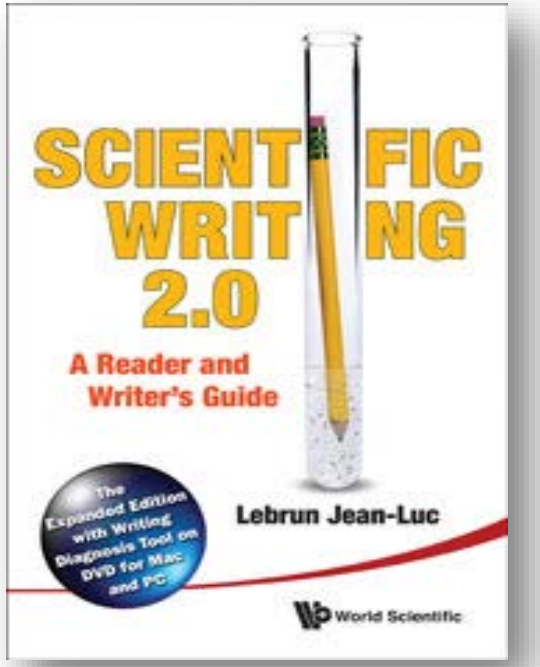
Reject!!!

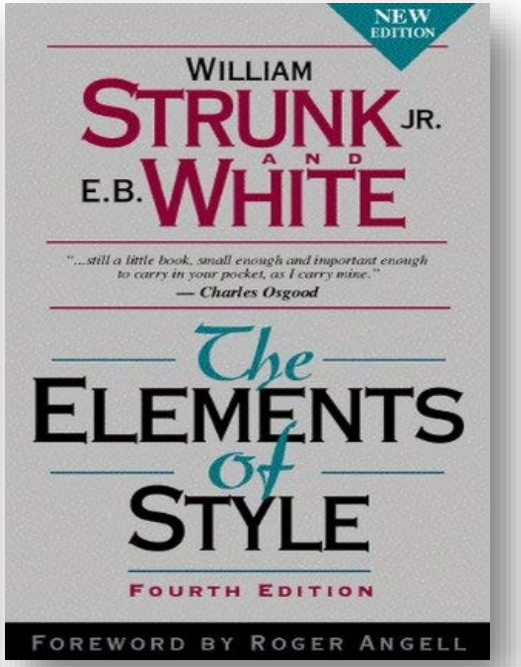
Good writing skills can be learned?

When there is a WILL, there is a WAY



Hard working





References : http://www.scientific-writing.com/

Paper organization

Paper Title

The Abstract

The Introduction

Related work

The Body

Experiments

The Conclusions

The Acknowledgements

Citations

Appendices

The “key” to good writing

Read more?

Write more?

Having something to say ---- 有感而发

（VS 无病生吟）

The Title

Unique

- Pregel: a system for large-scale graph processing
- Mizan: a system for dynamic load balancing in large-scale graph processing
- Graphchi: large-scale graph computation on just a pc
- GraphX: graph processing in a distributed dataflow framework
- Trinity: a distributed graph engine on a memory cloud

What is the difference?

What is the key feature?

The Title

Concise and Clear

- 10-12 words
- Prefer short titles to long ones

Fast

~~Partition-based cost-sensitive~~ failure recovery in distributed graph processing systems

Shared

A~~n adaptive load-balance~~ recommender system for taxi drivers ~~based on reinforcement learning~~

The Title

Catchy

- WTF: the who to follow service at Twitter

Name your approach

- Want to be retweeted? Large-scale analytics on factors impacting…

Use question

- Show, attend and tell: neural image caption generation with visual attention

Rhythm

The Abstract

Problem + Solution + Main Contributions

Keep it short, but not too short (1/3-1/2 column)

I have been informed that some of the papers have been registered with either **no abstract or with only a one line abstract**. The abstracts are used by PC members to determine which papers they are qualified to review. Therefore, submissions without abstracts or with a one/two line abstract **that does not properly explain the content of the paper are inappropriate and a violation of submission rules**.

The Introduction

5-point structure for Introductions from Stanford InfoLab

- **What is the problem?**
- **Why is it interesting and important?**
- **Why is it hard? (e.g., why do naïve approaches fail?)**
- **Why hasn't it been solved before?**
- **What are the key components of my approach and results?**

2-extra points (unless reach space limit):

- Summary of main contributions
- Paper organization

The Introduction

Power of illustration

Use Figure (background/application)

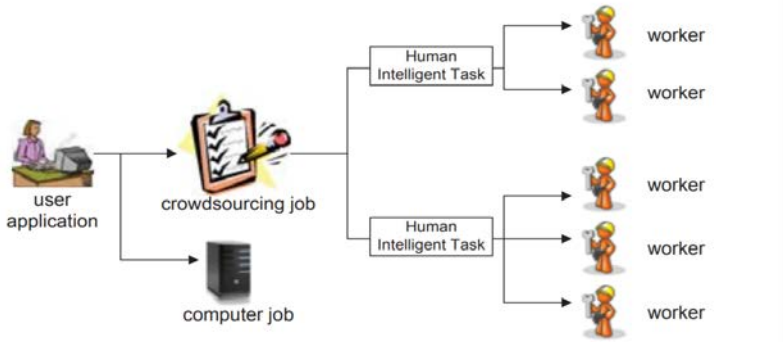


Figure 1: Crowdsourcing Application

Use Example (problem)

EXAMPLE 1. *Consider the simple sales and employee database at a computer retailer (e.g., Best Buy) shown in Figure 1. The database contains 3 fact tables (shown in green): 'Sales' where each row stores which customer bought which device with which applications (apps) installed, 'Owner' that stores which employee owns which device with which app installed and 'ESR' (Employee Service Requests) that stores which employee submitted service request for which app along with a textual description. There are 4 dimension tables: 'Customer', 'Device', 'App' and 'Employee' containing names of customers, devices, apps and employees respectively. Each dimension table has a primary key and the fact tables contain foreign key references to those primary keys (shown using directed edges).*

Use Table (comparison)

TABLE I
COMPARISON OF NVM CHARACTERISTICS

Technology	Read Operations		Write Operations	
	Latency	Bandwidth	Latency	Bandwidth
DRAM	15ns	15GB/s	15ns	10GB/s
NAND Flash [10], [11]	25μs	25-400MB/s	200-500μs	10-25MB/s
PCM [12]	50ns	10GB/s	350-1000ns	2GB/s
Memristor [13], [14]	100ns	10GB/s	100ns	5GB/s
STT-RAM [15]	40ns	15GB/s	40ns	10GB/s

Related Work

Listing

[31] presents the Explicit Factor M[illegible] (EFM) to generate recommendations according to the specific produ[illegible] applies the Tensor Factorization technique to learn the ranking of t[illegible]nces over various aspects of an item. [12] applies a vertex ranking appr[illegible]-partite graph of users-items-aspects to provide better recommendations of items using reviews.

Related Work

Compare

[18] uses the LDA-based approach combined with Matrix Factorization for better prediction of unknown ratings. Th highly interpretable textual labels for latent rating dimensions. More re and [27] went beyond [18] by using more complicated graphical edict unknown ratings based on collaborative filtering and topic mod user reviews. Their models are able to capture interpretable aspects and the sentiments on each aspect of a review.

Related Work

Contrast -- state the differences

In contrast to all the previous works, we not only predict unknown ratings of items based on user reviews as done in th k reviewed above, but also estimate the sentiments that a user would exp aspects in the review and determine the impacts of the aspects predicted rating of the review about an item. Moreover, we use ted impacts to recommend the most valuable aspects to the users to enha eir experiences with the recommended items. Finally, ….

Related Work

Beginning (section 2), if it can be short yet detailed enough, or if it's critical to take a strong defensive stance about previous work right away.

End (before conclusion), if it can be summarized quickly early on (in the Introduction or Preliminaries), or if sufficient comparisons require the technical content of the paper.

The Body

Key components:

- Preliminaries
- Problem statement
- Framework overview OR system architecture
- Methodology (algorithm + running example)
  - Design + implementation
  - Basic solution + optimization
- Theoretical analysis (e.g., complexity, approximate ratio)

The Body

Tips:

- Each section tells a story (don't put a long story in one section).
- Keep readers engaged at every step and looking forward to the next step.
  - E.g., preamble

> 3. XXX
> In this step we apply the model trained in Section 3.5 to determine the most important aspects of user's potential experiences with the item that were discussed at the beginning of Section 3.
> ……
> Note that they can be positive or negative, and we can use them to recommend positive and avoid negative experiences when users consume the recommended items, as explained in the next section.

- Move "interruptions" to appendix.

Experiments

Set experimental goals

In this section, we present an experimental evaluation of the techniques proposed in this paper. The goals of our study are:

- To compare the performance of our FILTER approach with VERIFYALL and SIMPLEPRUNE
- To evaluate the sensitivity of VERIFYALL, SIMPLEPRUNE and FILTER on various example tables
- To compare our FILTER approach with WEAVE algorithm proposed in [18]

In this section, we conduct experiments with the aim of answering the following research questions:

RQ1 Do our proposed NCF methods outperform the state-of-the-art implicit collaborative filtering methods?

RQ2 How does our proposed optimization framework (log loss with negative sampling) work for the recommendation task?

RQ3 Are deeper layers of hidden units helpful for learning from user–item interaction data?

Experiments

Key components:

- Datasets
- Baselines
- Measurements
- Parameter settings and runtime environment
- Results (Fulfill the goals)

The Conclusions

Two purposes:

- Restate the contributions with concrete results.
- Show how the work sets new research directions, or encourage future collaboration.

In this paper, we studied the problem of discovering minimal project join queries based on an example table. The main technical challenge is to efficiently verify which queries, among a candidate set of queries, are valid answers. We formalize the problem as the filter selection and develop a novel solution. Our experiments demonstrate that our filter-based approach is much more efficient than straightforward adaptation of known techniques.

Our work can be extended in multiple directions. In this paper, we require the valid query to contain all the tuples in its output; this might sometimes lead to empty answers. How to relax this requirement is an item of future work. How to rank the valid queries is also an open challenge.

The Acknowledgements

May add acknowledgement section during camera-ready

Acknowledge anyone who contributed in any way:

- E.g., shepherd, reviewer, data provider

Citations

Make all citations **complete and consistent**

- Refer to “bibtex” in ACM digital library or DBLP
- Check the final bibliography carefully
- May shorten the references for saving space, but **be consistent**

Example1:
P. Alvaro, N. Conway, J. M. Hellerstein, and D. Maier. Blazes: Coordination analysis for distributed programs. In *ICDE*, pages 52–63, 2014. (author, title, conference, pages, year)

Example2:
S. Agrawal, S. Chaudhuri, and G. Das. Dbxplorer: A system for keyword-based search over relational databases. In *ICDE*, 2002. (author, title, conference, year)

Appendices

Detailed proofs + Algorithms + Extra experimental results

- Should include materials that most readers are not interested in.
- Should NOT contain any contents that are necessary for understanding the paper.
- May or may not include Appendix section in a paper

Are we done?

Not quite…

Presentation Issues

Informal words

- “so”, “a lot of”, …

Imprecise words

- For “various reasons” …

  how many exactly?

Unclear references

- “this”, “that”, “these”, “it”…

  What do they refer to?

Long sentences

Consistent phrasing

- “the black-box attack”
- “the distortion reversal”
- “the human solver attack”

Active voice

- We develop a system…
- We process the data…
- We introduce a method…

Mechanics

Run a spelling checker all the time.

Tables, figures, algorithms are typically placed on the top of a page or column.

Tables, figures, algorithms should appear on the same page as its first reference (at least not too far away).

Ask your seniors/lab mates to do the proofreading.

Print out your paper and read.

About LaTeX

Inter-word space (add space ‘ ’ between sentences):

- “…evaluate the performance.The results show…”

Writing citations in LaTeX, do them in this form:

- text text text~\cite{ABC:2015:DEF}

Use of quotes “ ”

- Two back ticks (``) and two apostrophes (’’)

Summary

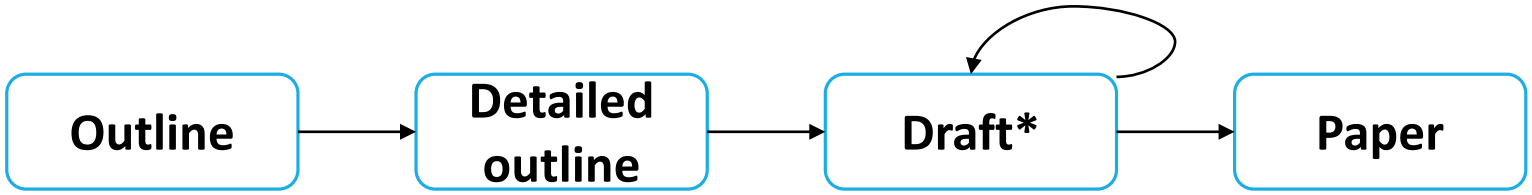
Writing technical papers is not easy.

Writing is important for getting paper accepted.

Good writing skills can be learned.

Good writing is hard work.

Leave enough time for writing, revising, rephrasing, polishing.



Useful links

- Tips for Writing Technical Papers. Jennifer Widom.
- Power Papers I, Power Papers II. Terrance Sim.
- How to Write & Publish a Scientific Paper, Robert A. Day
- How to write a scientific paper.

http://www.scidev.net/ms/howdoi/index.cfm?pageid=60

- How to submit a paper to a scientific journal.

http://www.scidev.net/ms/howdoi/index.cfm?pageid=61

- Jeffrey McQuain. “Power Language: Getting the Most out of Your Words.” Houghton Mifflin, 1996.
- Jean-Luc Lebrun. Scientific Writing: A Reader and Writer’s Guide. World Scientific, 2011.

Q&A

Thank you!

Email: shenyy@sjtu.edu.cn