ĐẠI HỌC BÁCH KHOA HÀ NỘI
VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

Reading and Reviewing

Technical Writing and Presentation

SOICT - 2020

Contents

- Reading Literature
- Finding Research Papers
- Critical Reading
- Developing a Literature Review
- Evaluation of Papers

Motivation

- Human knowledge is an infinite treasure
- Do not reinvent the wheel
- What should we do?
  - "Stand on the shoulders of giants", Google scholar
  - Discover knowledge by building on previous discoveries

  ⇒Reading and reviewing papers is an important activity of the scientific process

source: Google

READING LITERATURE

LITERATURE

- Tranditional definition: a collection of written books, Wikipedia
- Other definition: “literary means not only what is written but what is voiced, what is expressed, what is invented, in whatever form”, Greil Marcus & Werner Sollors

source: Google

Reading Literature - Importance

- Understand key concepts, terminologies, theories, discoveries, and debates
- Identify new lines of questioning or investigation
- Discover your work is indeed novel or innovative
- Become familiar with key researchers in the field

Reading Literature - Situation

- Search of literature can lead to hunderds of potentially relevant papers
- Papers are not textbooks, and should not understand every line
- The number of papers that a researcher working on a particular project has to know well is usually small

In dblp, you now find publications from ...

- **5,800+** conference and workshop series
- **1,600+** journals
- **90,000+** table of contents

source: DBLP

Reading Literature - Strategy

- Becoming an effective reader is important
- Give a paper: decides to give it more or less time that it deserves
  - Skim through it to identify the extent to which it is relevant
  - Only read it thoroughly if there is likely to be value in doing so
  - Make effort to understand the details

source: DBLP

FINDING RESEARCH PAPERS

Finding Research Paper

- Each research work builds on prior work
- The number of existing publications is very large
- A consolation is that recent work already explored the older literature => carefully search for current work

source: DBLP

Finding Research Paper - Path

- Use obvious search terms to explore the web: publications, projects, solutions, etc.
- Use special search tools for academic papers such as google scholars
- Seach the publisher-specific digital library such as Springer, ACM, IEEE
- Vist websites of key research groups and researchers working in the area
- Follow up the references in promising research papers
- Browse the recent issues of journals and conferences in the area
- Consider using the citation indexes
- Discuss your work with as many people as possible

Finding Research Paper - Path

- The process of search and discovery of useful papers is a form of learning
- Finding all *relevant* work is hard; finding all *significant* work is a critical part of doing research
- Searching and reading are separate activities, do not try both at one
- If your idea is not so original after exploring the literature, be honest – review your work to see what aspect may be novel

CRITICAL READING

Critical Reading

- Active attempt to identify the contributions and shortcomings rather than simply reading from one paper to the other
- Good researchers should have ability to analyze the work and claims of others
- A paper is refereed is an indicator that it is of value, but it is not a guarantee, because:
  - A paper is a snapshot of research work at a moment in time – what the researchers knew when they submitted
  - Assumptions may be implausible
  - Dataset used may be so tiny, that the results are meaningless

Critical Reading

- Don’t accept something as true just because it was published
- Don’t evaluate researchers being dismissive of their past work
- We should respect published papers, and learn from them about strengths and weaknesses
- Inexperienced researchers see other work either perfect or poor, with nothing in-between. Usually, neither of these extremes is correct

Critical Reading

- Read papers by asking critical questions, such as
  - Is there a contribution? Is it significant?
  - Is the contribution of interest?
  - Are the results correct?
  - Is the appropriate literature discussed?
  - Does the methodology actually answer the initial question?
  - Are the proposals and results critically analyzed?
  - Are all the technical details correct? Are they sensible?
  - Could the results be verified?
  - Are there any serious ambiguities or inconsistencies?

Example 1

- Read paper "**Detecting Spam Web Pages through Content Analysis**"
- **Excerpt**

We continue our investigations of "web spam": the injection of artificially-created pages into the web in order to influence the results from search engines, to drive traffic to certain pages for fun or profit

- **Sample**

In order to design and evaluate our spam detection algorithms,we used a collection of 105, 484, 446 web pages, collected by the MSN Search [22] crawler, to serve as a proxy for the web at large. These pages were collected during August 2004, and were drawn arbitrarily from the full MSN Search crawl.

Problems

- **Is the methodology valid here?**
  - Critical readers would question whether the sample size was big enough to fulfil the aim of this study.
  - They would also question whether the sample was representative enough of the wider group of webpages, as the criterion for inclusion in the sample perhaps created an unrepresentative group.
  - Could such crawling lead to a collection with too many spam pages compromising the results?

Example 2

- Excerpt

The MSN Search crawler discovers new pages using a roughly breadth-first exploration policy, and uses various importance estimates to schedule recrawling of already-discovered pages. Therefore, pages crawled using such a policy may not follow a uniform random distribution; the MSN Search crawler is biased towards well-connected, important, and “high-quality” pages. In addition,the MSN Search crawler already uses numerous spam detection heuristics, including many described in [8]. Although our data set may not correspond to a “random sample” of the web, we believe that our methods and the numbers that we report in this paper still have merit for the following reasons. First, although our crawler focuses on well-connected and important pages, these pages are typically ranked most-highly by search engines. Therefore, the numbers on spam that we report in this paper approximate what will eventually be perceived by users of search engines. Second, since the crawler already discards or downgrades some of the spam, the numbers and metrics that we report in the following sections are ***a conservative estimate of the impact of web spam.***

Problems

- **Has the author overgeneralized the results here?**
  - The author has used the findings from a big enough sample size, that represents a sufficient range of websites, to support a major line of argument about how to recognize spams among websites
  - The authors are inferring that the results gained can be generalized to all set of websites

Example 4

- Excerpt

We trained our C4.5 classifier using the *DS* training set. A portion of the resulting classification tree is shown in Figure 14. To apply the tree to a page, we check the value of the property named in the root node of the tree, and compare it to the threshold associated with the outgoing edges. Depending on the outcome, we follow the left or right edge, and repeat the procedure until a leaf node is reached, assigning a class to the page. For example, considering the tree in Figure 14, a page whose 5-gram independent likelihood value (see Section 4.9) is less than 13.73 and which contains at most 62 of the 1000 most-popular words is classified as non-spam

Problems

Was this training method appropriate?
Are the received result correct?
Is it possible to decide the content of a spam based on the above criteria?

Example 5

class	recall	precision
spam	82.1%	84.2%
non-spam	97.5%	97.1%

Table 1: Recall and precision of our classifier.

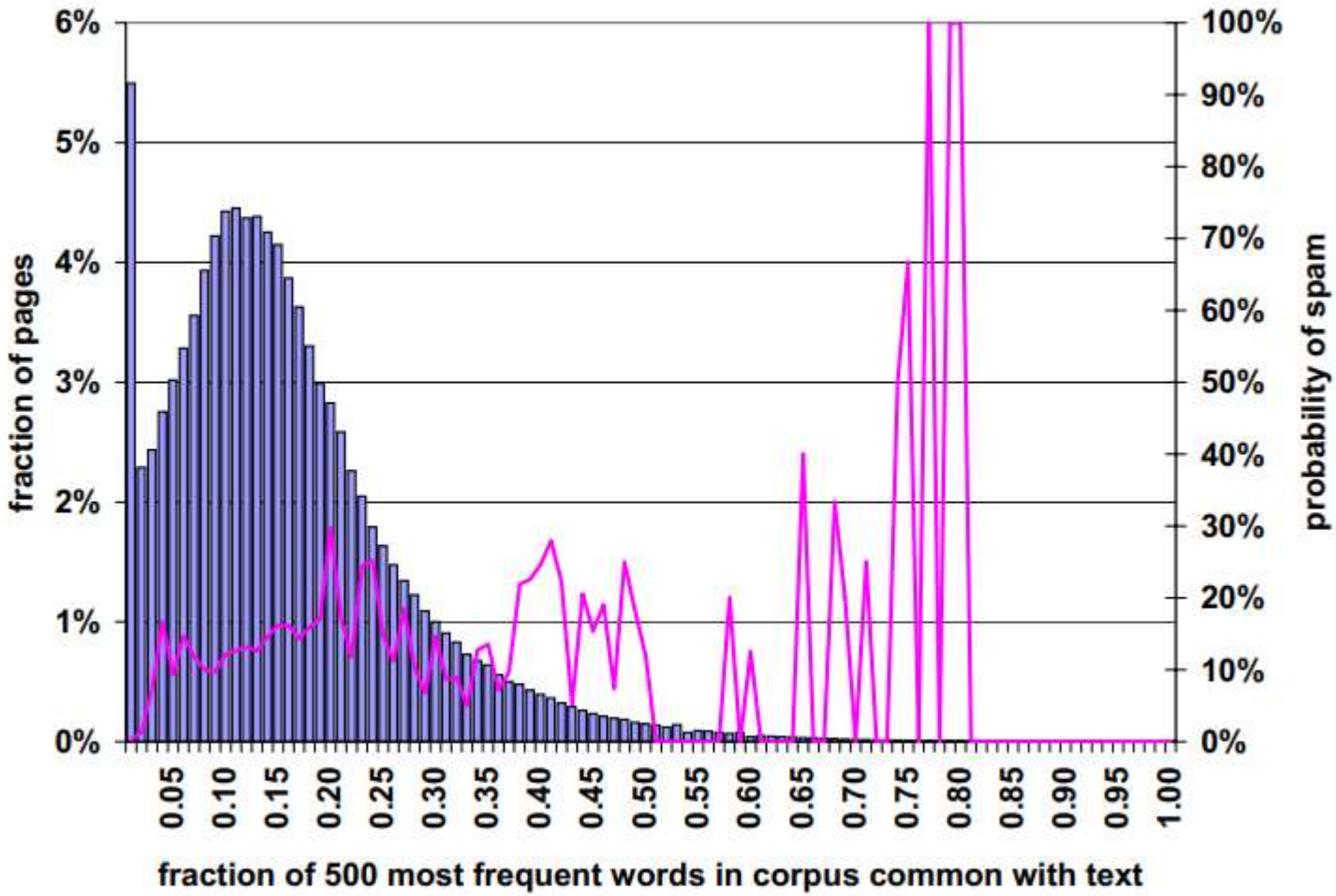
Problems

- Is this measurement suitable for the problem?
- What does this measure reflect?

Example 6

- Excerpt

Figure 11 is based on the fraction of the 500 most popular words that are contained within a page. The bar chart shows a left-truncated Gaussian, with much of the truncated mass appearing as a spike at the left-hand edge. The distribution has a mode of 0%, a median of 13%, and a mean of 14.9%. The prevalence of spam is modest throughout the range, with a dramatic spike for those few pages in which 75% or more of the popular words appear on the page


6%
5%
4%
3%
2%
1%
0%
fraction of pages
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
probability of spam
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
0.45
0.50
0.55
0.60
0.65
0.70
0.75
0.80
0.85
0.90
0.95
1.00
fraction of 500 most frequent words in corpus common with text

Problems

- What evidence does the author provide to support his or her argument?
- Is there evidence provided supporting this?
- Would you accept this as fact? Why?
- Is this the author's opinion or fact?

Example 7

- Excerpt
  - The line graph, depicting the prevalence of spam, rises steadily towards the right of the graph. The graph gets quite noisy beyond a compression ratio of 4.0 due to a small number of sampled pages per range. However, in aggregate, ***70% of all sampled pages with a compression ratio of at least 4.0 were judged to be spam***

Problems

- **Prevalence of spam relative to**
  - Domain name
  - Language
  - Number of words on page
  - Number of words in the page title
  - Average length of words
  - Amount of anchor text
  - Fraction of visible content
  - Compressibility
  - Page drawn from globally popular words
  - Fraction of globally popular words
  - Independent *n*-gram likelihoods
  - What else?
- This point is stated as fact
- What theory is it based on?
- Do you agree with it?

EVALUATION

Contribution

- Contribution is the main criterion for judging a paper
- Typically, a paper is a contribution if it has two main properties: *originality* and *validity*
  - Originality: the degree to which the ideas are significant, new, and interesting.
    - Most papers are to some degree extensions of previously published work
    - Impact of the contribution: how much change would follow from the paper
  - Validity: the degree to which the ideas to be sound
    - Should contain proof or analysis, experiment, simulation to allow verification by other scientists
    - Comparison to existing work is an important part of validity

Evaluation

- Critical questions:
  - Is the contribution timely or only of historical interest?
  - Is the topic relevant to the venue's typical readership?
  - What is missing? What would complete the presentation? Is any of the material unnecessary?
  - How broad is the likely readership?
  - Can the paper be understood? Is it clearly written? Is the presentation at an adequate standard?
  - Does the content justify the length?

DEVELOPING A LITERATURE REVIEW

Literature Review

- A structured analysis of a body of literature, and may cover work from several area of research
- These papers should be grouped by topic, and discussed in a way that allows reader to understand their contributions, limitations, and questions that they leave open

Where is the literature review located in a research text?

- A literature review usually occurs as a section of a paper or report following the Introduction.
- However, parts of the literature review can be integrated into other sections of the paper or report.
- You should check the conventions for the publications you are writing for.

Different ways of writing literature reviews

There are different ways of presenting literature reviews in research publications. These include:

- Organization by themes using subheadings
- Discussion of literature integrated throughout the publication
- A separate section following the Introduction

The structure and organization of a literature review wil depend on the kind of publication, its length, the conventions of similar publications.

Literature Review – Questions

- Who wrote the text and what are the author's qualifications?
- When was it written?
- Who is it for?
- Why was the study carried out?
- What is the author's main point, or thesis?
- How has the author collected the data?
- What are the findings?
- What relevant sources does the author use?
- What limits did the author place on the study?
- What aspects are relevant to your research question/area?
- What is your evaluation of the text?

Literature Review - Progress

- When you read a paper that you think will need to be discussed, add it in
- Rough stage: focus on organization and content rather than on presentation
  - group papers by topic and contribution
  - briefly summarize each paper's contributions and evidence used to support the claims
  - add notes to each paper: features that are of interest, shortcomings, how the work might better
  - no need for drafts to be polished, no one but you will early versions
  - Early drafts should be as inclusive as possible

Literature Review - Progress

- Refined stage:
  - Decide whether to include each of the paper you read.
    - Obvious factors: how close some other work to yours, how influential it has been
    - Subtle factors: you find a survey paper or a recent paper with a literature review of older papers; so many older papers do not need to be discussed
  - When you remove a paper, put the discussion in another file (or comment it out) rather than deleting it
  - Steps: rewriting, editing, polishing

Synthesizing

Your literature review will contain considerable analysis of what you have read. In writing up the analysis you bring together other researchers' work making relationships between and among their ideas to produce new ideas.

Language that demonstrates synthesis in a literature review explicitly draws attention to relationships between and among ideas, theories, findings etc by different researchers.

There are many ways to do this, for example:

- **Whereas** X argues... Y suggests a **different** cause which is ...
- X claims ... which **differs** from both Y and Z in that …
- X's research **builds on and expands** Y's initial work by …

Critique: Expressing Judgment and Evaluation

- One **limitation** of X's argument is... X's discussion is **superficial.**..
- X's interviews were **thorough**...
- X carried out an **impressive** number of observations...
- X presents evidence in a **logical and coherent** discussion of …
- Even though X's methodology is **sound**, the conclusions she draws **do not follow**... The experiment was **poorly** constructed...
- X's research demonstrates a **sophisticated** engagement with the literature...

Academic voice

- Your literature review must contain the language of **judgment and evaluation.** The language of judgment and evaluation demonstrates how writers position themselves in the text – any word or phrase that indicates the **writer's attitude** to what is in the research literature.
- Your literature review must be written in your **own academic voice.** This means that your ideas about the literature, how bits of the literature relate to each other, and how useful they are for you must be foregrounded.

Justifying your critique

It is **not** enough to express judgment and evaluation when discussing what you have read. You need to give reasons for your critique. The following examples illustrate this:

- X's discussion is **superficial...in that he does not take account of...**
- X's interviews were **thorough....She ensured that she...**
- X **managed to interview all the..**
- X carried out an **impressive** number of observations...; **he managed to interview all the..**
- X presents evidence in a **logical and coherent** discussion of … She **begins with ... moves to ... and then…**

Sample verbs to introduce others' work

describe identify argue show list

review consider suggest discuss

state distinguish advocat submit

claimi elaborate highlight look at

Evaluation

- Quality of a paper can be reflected in its bibliography
  - How many references are there?
  - Are there recently published references?
  - Are there references to the major journals or conferences in the area?
- Reviewers should make an effort to search for errors that don't affect the quality of the work but should be corrected before going into print
  - Spelling
  - Syntax
  - Errors in the bibliography...

Literature Review Example

Read the literature review example and focusing on the language of critique and synthesis, discuss the language features of each example and answer the following questions for each of the examples:

- Are there any problems with it?
- What else needs to be included to improve this review?

Reading Literature - Sources

- Papers: refereed and published in reputable venues
- Theses: undertaken and examined at reputable instituations
- Books: based on information presented in referred theses or papers
- Other sources: news articles, Wikipedia pages, maganizes, etc. are rarely worth citing

Homework: Literature Review

- Read 3 given papers carefully.
- Write a literature review of the 3 articles ensuring you use the language of analysis, synthesis and critique. The literature review
  - analyses,
  - compares/contrasts and
  - critiques their ideas, theories etc
- Use the language of judgment and evaluation so that your own academic voice is predominant.
- Use linking words and phrases to make your text cohesive.