Proposed Exercises for "Writing for Computer Science," 3rd ed. (Justin Zobel)

R. Zanibbi (Sept. 16-17, 2014)

The Computing Research Environment

1. (Publication Venues) Define and briefly explain how the following are related in Computer Science. Are these venues used in the same way for engineering disciplines, the natural sciences (e.g. physics) and the humanities?
   - Technical reports
   - Workshops
   - Conferences with peer-reviewed publications
   - Conferences with non-refereed publications
   - Journals
     - Rapid publication journals
     - Open access journals
     - 'Regular'/traditional journals

2. (Publication Types) Briefly define each of the following, and state how they are used by researchers.
   - Survey papers
   - Paper abstracts
   - Extended abstracts
   - Research papers
   - Position papers
   - Magazine articles
   - Letters
   - Grant proposals

3. (Roles in the Research Ecosystem) What role do the following individuals or groups serve in, and what types of work do they perform in their role?
   - The first author on a research paper
   - The last author on a research paper
   - Paper reviewer/referee
   - Program Committee Chair
   - Session Chair at a Conference
- Organizing Chair for a conference
- Area Editor for a journal
- Editor-in-chief for a journal
- Professional organizations (e.g. IEEE, ACM)

4. Discuss and/or write about the criteria for including a person of group in the Acknowledgement of a paper. What constitutes 'enough' help to be acknowledged?

5. Discuss what criteria should be used when deciding to submit to highly competitive (e.g. top-tier journal), somewhat competitive, and non-competitive (e.g. non-refereed) venues. Do all three types of venues serve a meaningful purpose, or should a researcher focus their efforts solely on producing top-tier publications?

6. Read the following two sources, and then discuss the following question: is Computer Science a science?

**Reading Research Papers**

Using literature effectively and efficiently requires skill at both skimming and quickly identifying the organization of a paper, and breaking down complex technical details into their parts. A common mistake is to read research papers similar to a novel, from beginning to end, working to understand confusing parts of the paper one-at-a-time. A better analogy for research papers is the reference manual. People search through manuals to find sections that answer specific questions, after which the manual is put aside. Normally, you are not concerned if you don’t know everything in the manual - only whether the question(s) you had were answered. Particularly if a paper is important for your work, you should expect that you will consult it multiple times as you search for answers to different questions, and that your understanding of the paper will change as you do this.

The following exercises can be done individually or in a class setting. In-class, it is recommended that just one or two papers be used, so that people are able to discuss and compare their analyses of the same document. It is strongly recommended that exercises 1-3 be done using only pen and paper (colored pens and highlighters are quite welcome).

1. Take a research paper that you have not seen before, and start by reading the conclusion of the paper. Then examine the abstract of the paper. How are they related? Are there details in the conclusion that would be appropriate to include in the abstract?

2. Obtain a survey from a Computer Science journal (e.g ACM Computing Surveys) on an unfamiliar topic. In 15 minutes, skim read the article. After this, write down brief answers to the following questions.
   a. How is the paper organized?
   b. What is the main taxonomy, graph or other structure that the authors use to
organize methods presented in the survey?

c. How are figures and tables used for illustration, and where do they appear in the paper?

d. Provide three distinct examples where a statement is supported by a citation, and state the function of the reference(s) for these statements.

e. Do you like the organization and presentation of this survey after this first look? Why or why not?

3. Find a paper from a Computer Science journal that contains a significant amount of mathematical notation.

a. In five minutes, skim the paper, and write down one sentence summarizing the main contribution of the paper.

b. In another five minutes, quickly look through the tables, figures and other graphics in the paper and their captions. Write a short paragraph (e.g. 3-4 sentences) summarizing what these visualizations describe.

c. Now choose what appears to be a challenging section of the paper, with a lot of mathematics. In two minutes, skim read the section, making sure to quickly read through the mathematics, highlighting or making a note where you become confused, but without going back and re-reading. Then answer the following questions.

- What is the main concept presented in the section? If you are uncertain, write down your best guess.
- What additional information would be helpful in understanding this section, and where do you expect that it is located in the paper (without reading these other sections)?
- Of the things that you highlighted or noted as confusing in the section, which would you clarify first if your goal is to strengthen your understanding of the main contribution as quickly as possible?

4. Do the same exercise as in Exercise 3, but instead consider a Human-Computer Interaction paper that involves a substantial human experiment.

5. Discuss in-class or with a colleague on what strategies to use when:

a. Finding literature on a problem that is new to you

b. Finding literature about specific algorithms, models, results or techniques once you have become more familiar with a problem (e.g. have been working on the problem)

6. Discuss how often you should read research papers and for what purpose(s). Richard Hammond commented in his famous talk "You and Your Research" (http://www.cs.virginia.edu/~robins/YouAndYourResearch.html) that it is possible to read "too much." Do you agree with this statement? Why or why not?

7. This exercise is concerned with practicing finding relevant papers. Using online paper databases (e.g. ACM Digital Library, IEEE Xplore, Google Scholar, CiteSeer), create a document containing abstracts from five papers. Above the abstracts, in one or two
sentences identify the research question that you are performing the literature search to answer. Then write a short paragraph that summarizes how you expect the five papers you have selected will address your stated problem.

8. (Annotated bibliographies) Using a bibliographic database tool (e.g. JabRef, BibDesk, Mendeley, Zotero, etc.), create an initial bibliography containing at least five papers on a research question that you are interested in. In the notes or annotation field of the database, provide 1-2 sentences indicating how you expect the paper to be helpful in your search. Within one to two weeks, read the five papers, updating your annotation to briefly record key details about each paper (e.g. results, limitations, insights, opportunities, and questions/confusions about the paper), using point/bullet-form. As you finish reading papers, check your annotations on the other papers to identify how the paper is related to the others, and whether this paper helps clarify questions that you had when reading other papers. **Make sure to update your annotations for papers when you find additional details of interest, clarifications or answers to your questions in other papers.**

9. Prepare a five-minute presentation summarizing the core problem, findings and results of a research paper with a technical result. This can be done very effectively, but requires studying the contents of a paper carefully both before and while preparing the talk.

**Writing and Peer-Review Exercises**

The following are in-class exercises. Peer-review exercises are intended to be completed by pairs of students using printed documents and pens/highlighters. Students should write the documents below with other students in the class as the intended audience, rather than the instructor (i.e. writing for peers, rather than struggling to write for experts).

1. Discuss which is more important - the core content of a paper, or how well a paper is written? How much effort should a researcher put into obtaining significant results versus communicating their work clearly to others?

2. (Class Exercise) Make a (maximum!) one-page summary of a paper assigned by the instructor for your course, writing it in a style intended for other students in the course. When you come to class, exchange a printed copy of your summary with another student. In five minutes, write on the summary to indicate things that are unclear or awkward in the writing, and write down one or two questions whose answer would help you better understand the summary. Exchange your notes and discuss the questions that you have have written down.

3. (Class Exercise) Spend a month preparing an annotated bibliography for a specific research question, and then prepare a two-page draft outline for a paper summarizing what you have learned, using point/bullet form to identify key details in each section of the paper, along with where you will cite references and for what purpose. In-class, exchange your draft with another student, and have each of you make comments on a
written copy of the draft. Return each draft to the author, provide a few minutes for the author to read the comments, and then speak with the other student to clarify questions that you have about their comments.

4. (Class Exercise) Some weeks later, after you have written the literature review outlined in the previous exercise, exchange it with a different student in your class. Take 10 minutes for each of you to skim read and make notes on any questions that you have about the draft, and then write down answers to the questions below (quickly, in about five minutes). Again, take time afterward to share and discuss the comments made on your draft.

   a. What is the research problem summarized in the paper?
   b. What underlying organization used to relate key problems and methods?
   c. Does the organization of the sections of the paper support the presentation of the underlying conceptual organization of the author, or can it be improved?
   d. Are citations used effectively to support claims and facts stated in the paper?
   e. What are the key conclusions about the research problem, and what opportunities for future investigation are identified?
   f. Identify one change that you think would improve the readability/ease of understanding of the paper for students in the class most.

5. (Class Exercise) Do as in the previous exercise, but have the author identify a specific section of the paper that they would like their peer-reviewer to review more carefully in 10 minutes, writing questions and opportunities for clarifying the presentation on the draft. Again, take time afterward to discuss each others' notes.

Experimental Design and Analysis

1. (Visualization) Use a statistical package (e.g. R) to compute descriptive statistics for a large data set (e.g. from the UCI Machine Learning Repository at http://archive.ics.uci.edu/ml/). Compute the mean, standard deviation, median and mode for each of the numerical variables in the data set. Create the following types of plots:
   - An error bar plot showing the mean and standard deviations for different variables;
   - A histogram visualizing the distribution of values for a single variable;
   - A scatter plot visualizing the distribution of values for one variable vs. a second variable;
   - A line plot, showing the values of samples/data parts sorted in increasing order of value.

2. (Visualization and Debugging) Identify how best to visualize results from experiments related to your area of research interest. In particular, consider how these visualizations could be used to help you identify patterns or errors in your implementation’s outputs.

3. (Hypothesis Testing) Acquire a publicly available data set containing results from a paper making use of a statistical hypothesis test (the paper may come from any scientific discipline). Using R or another tool, repeat the statistical test made by the authors,
attempting to produce the same result. Note that this requires properly understanding how data is broken into groups, and which intermediate values need to be obtained to produce the final hypothesis test (e.g. t-test, ANOVA, Wilcoxon rank-sum, Friedman and any post-hoc tests such as the Tukey HSD). Write a brief two-page report summarizing what the data set represents, the null hypothesis to be tested, assumptions of the statistical test used, the procedure for computing the values, and whether you were able to replicate the result.

4. Obtain two research papers using statistical tests of either the same or similar hypotheses. Discuss the extent to which these findings support or contradict one another, and what differences in the experimental design or statistical analysis might have contributed to the observed results.

5. Create a one-page sketch for an experiment in your area of interest. Identify clearly the following: 1) your hypothesis, 2) the independent (i.e. input) and dependent (i.e. output) variables of the experiment, 3) the different conditions (i.e. combinations of independent variable values, e.g. different algorithms or parameter values), 4) visualizations used for outputs and/or distributions of values, and 5) the statistical test(s) that you will use to test your hypothesis, and 6) the possible outcomes of the experiment (in terms of your hypothesis), and what you expect to observe.

Research Programming Exercises

1. Download an open source tool related to your area of research interest. Install and run the program, and try running a couple of the provided examples and/or one of your own creation. Write a two-page document summarizing the tool, state how you think it may be useful in your research, and illustrate a sample input and output for the program. If done as a class exercise, have student present these briefly (e.g in 2-3 minutes).

2. (Scripting and Automation) Write a shell-script program (e.g. in bash shell) that automates the execution of a simple repetitive task for a single file (e.g. sorting and/or filtering columns of a comma-separated-value (CSV) file). Then, modify the script so that it will perform this task for a complete directory of files, writing out results to file related to the name of the input file (e.g. for 'Data.csv,' producing a file 'Data_filtered.csv').

3. Describe in concrete terms the data structures and file encodings that you would use for a set of experiments of interest to you. Your inputs and outputs should provide a human-readable format in addition to any more space-efficient (e.g. binary) encoding.

4. Discuss strategies that you can use to check the accuracy of algorithm implementations that you make, both when replicating the work of others, and when making your own variations or implementing new algorithms. Can these checks be automated to reduce effort, and prevent errors later on?

5. Implement an algorithm described in a paper from your area of research interest, and
design and execute a test script to verify that your implementation is correct. Write a brief report (e.g. two pages) summarizing how you structured your implementation and your replication of the results. If doing this exercise in-class, exchange your implementation with another student, and have them write a two-page report on their experience running your program on an example, running tests, and reading your source code. You should both comment on what is clear and indicate what changes could be made to make using and reading through the implementation easier.