# Suggested practices for conducting scientific research

Student: Sergei Jegorov 204707IASM
Supervisor: Aleksei Tepljakov

School of Information Technologies
Tallinn University of Technology
Tallinn, Estonia

05.05.2021

The advancement of technology would not be possible without proper documentation of discoveries, followed by discussion in scientific conferences. Such discoveries can often shed light on relevant matters, and can also raise new questions and problems which can, in turn, lead to the beginning of new research. With the advancement of information technology, the creation, publishing and exchange of scientific articles has become faster, easier and more accessible than it used to be. However, for the composition of a high-quality scientific paper, there are several things to consider: proper literature review, definition of the state of the art, typography and documentation. Often, one or several of these aspects may get overlooked by young researchers, and this may result in a mediocre paper. In this essay, the author tries to summarize the aspects of scientific research, explain why they are important to consider and provide a short overview of technology that can assist in writing excellent scientific research articles.

To begin with, research is a systematic investigation into, and study of, materials and sources in order to establish facts and reach new conclusions. As this definition suggests, the research process involves study of, and acquisition of knowledge from, previous findings pertaining to the investigated topic, e.g. literature review. In all scientific research, literature review is conducted first and is a highly critical step as it lets researchers gain knowledge on the most relevant discoveries related to the investigated topics, and to find the scientists who best understand and have the most experience with the given topic. While one previously had to search and read through several books and other collections of scientific articles, everything can now be found in online databases consisting of large collections of research papers. Systematic review of such databases allows the researcher to compile the state of the art (SotA) - a critical element indicating the most recent stage in research, without which it would be almost impossible to convince the scientific community of the worth and novelty of the conducted research. Additionally, SotA allows researchers to learn more about their topic of interest, to prove that the research problem has relevance and to help understand the novelty (or lack thereof) of a proposed approach to a problem. There are several methods of systematic review (protocol-driven, personal knowledge, personal contacts), although it is estimated that 51% of sources are collected using a method named snowballing [1].

Snowballing is the practice of using a paper's reference list or citations to locate additional papers. As soon as a relevant research paper is found, researchers can look through that paper's references and, based on the established criteria (such as date of publishing, language, etc.), identify additional, relevant papers. This method is called backward snowballing. Alternatively, forward snowballing can be used, where a paper's citations are followed using a facility known as 'citation

tracking', which is usually provided by large databases such as Google Scholar, IEEE Xplore, ResearchGate, etc. Based on the criteria, citations are reviewed and selected, resulting in additional papers for research. During these processes, the initial review can be carried out simply by reading a paper's abstract, but as soon as all relevant papers are acquired, they are studied in full-text review. Full review allows researchers to indicate potential gaps in research, such as under-researched or overlooked areas, lack of experimental or empirical data, novelty and relevance of methodology used in research, and to formulate problems and questions for their own research.

During literature review, it is highly suggested that the reviewer takes notes of all findings that are of interest, and all questions that appear while reviewing the paper. Since moving to the digital age, this process has become faster and more convenient thanks to the advent of note-taking software. Many of these applications are simple to use and have the support of Markdown - a lightweight markup language for creating formatted text. There are perhaps hundreds of examples of such software, such as *EverNote, Joplin, AppNote, OneNote*, etc. However, the author has focused on using *Joplin*. What makes *Joplin* stand out from other note-taking apps is the fact that it is completely free, open-source (meaning you can contribute to the software's development or program your own features if needed) and cross-platform. Another powerful piece of software used for assisting researchers worldwide is a citation and reference management software named *JabRef. JabRef* utilizes *BibTeX* - a reference formatting system that aids the user by collecting all sources and metadata in one place, and allowing them to be exported to *TeX* documents. It is quite simple to use, especially as many online databases support *BibTeX* citing format.

When the time comes to document the conducted research and resultant findings, *TeX* comes into play. *TeX* is a powerful typesetting system, distributions of which are extensively used in academia for the formatting of documents. However, for creating and writing documents themselves, LaTeX appears to be the most used system. Though it utilizes low-level *TeX* macros, *LaTeX* was created as a high-level software tool for document writing, and is based on the principle that authors should be able to focus on the content of what they are writing without being distracted by its visual representation. The advantages of *LaTeX* and *TeX* are numerous - through the implementation of native syntactical rules, the user can define the structure of a paper and allow *LaTeX* to format the document according to those definitions. This is especially convenient when showing formulas, as in *LaTeX*, instead of manually substituting elements within a template (as in many popular word processors), they can be inputted in a manner more similar to those found in many programming languages. To some extent, *LaTeX* can also be used

for plotting, although it is often better to plot graphs by other means and then export them.

Plots, graphs, pictures and any other kinds of graphical content can prove extremely valuable nowadays. This becomes especially important when research involves the use of complex set-ups, processes and systems. Just as mathematicians may use a single formula to replace hundreds of words of explanation, researchers and engineers can use drafts, plots and diagrams to better and more quickly explain the design and flow of various processes. Frequently, such elements appear to be too complex and time-consuming, and thorough observation of every single element is needed. Moreover, readers with vision problems may need to zoom in and look more closely at the whole paper. Yet, where is the benefit in zooming in on a jagged figure when it is impossible to understand the scale readings of a plot? One could suggest increasing the size of the picture, but then the file size will also increase. This is the disadvantage of using raster graphics in documents and sometimes it can significantly impact the review of the research work. In most cases, researchers should use vector graphic (SVG) figures, which expand infinitely, stay sharp and occupy less space. Nevertheless, in the case of gradients and spectrograms, raster graphics still remains a go-to tool. This is mainly because, in the case of highly-coloured plot, it may take too long for a figure to load if SVG is used. Once again, multiple tools are available here: *Adobe Illustrator, Inkspace* for vector images drawing and *draw.io, yEd, Dia* for graphs, diagrams, UML and flowcharts. The author has been using *draw.io* extensively throughout his studies and has successfully produced numerous quality diagrams in a short period.

It is perfectly normal to conduct research alone, but commonly researchers work in teams. Simultaneous work on the same project, document or file can get confusing, especially when several versions of the one document may exist. Version control software (VCS) is of help here - it allows researchers to manage changes to a file over time, storing modifications in an online database. Nowadays, it is impossible to imagine a large project being managed without a VCS. In the case of a mistake being made which would otherwise require teammates to return to a previous version of a file, it is available, and there is no need to manually revert anything (a process which would be time- and resource-consuming). Additionally, VCS helps resolve any inconsistencies in modifications to a file. For example, if the same section of text in a document were simultaneously overwritten by two collaborators, the VCS would warn each collaborator that a conflict exists in the merging of documents, thus allowing them to resolve that conflict without needing to create separate copies. There are several VCS available, but ultimately the "king" of VCS is *Git*. *Git* appears to be the most popular tool of the trade for

programmers, but notes and text documents are also supported. It does, however, require quite some skill to get comfortable with, especially with its CLI (command line interface) commands. Therefore, for some people it may be more preferable to use a graphical user interface (GUI) such as *GitKraken*.

An often overlooked part of composing a paper nowadays is quality typography. Matthew Butterick, the author of the book "Practical typography" wrote:

> "Good typography reinforces the meaning of the text".[2]

People tend to think that using standard style fonts is a one-for-all solution, but this is not always true. Different fonts serve different purposes. As an example, serif fonts better suit large documents that contain a lot of text, as they visually create a 'line structure' that helps to grasp a reader's attention and keep it on one line. On the other hand, sans-serif fonts are generally preferred for banners and other elements that a person would typically look at from afar. In the scope of scientific writing, using the wrong font families in documents can repel readers, as they may be difficult to follow, or even distracting. Butterick argues that a reader's experience is not often taken into account and it is assumed that a reader's attention to the text is broad [2]. Such a misconception can lead to poorly designed documents.

All things considered, conducting research and writing scientific articles are by no means easy skills. The discussed aspects of conducting research require constant repetition to perfect. Writing good scientific papers takes effort but is ultimately rewarding. As once mentioned in the course, one learns better by doing. The author of this essay has found room for self-improvement in many areas as he has followed the course, and is now confident that the acquired knowledge will allow him to conduct higher quality research in the future.

# References


[1] Trisha Greenhalgh and Richard Peacock. “Effectiveness and efficiency of search methods in systematic reviews of complex evidence: audit of primary sources”. In: *BMJ* 331.7524 (Oct. 2005). (visited 2021-05-01), pp. 1064–1065. DOI: 10.1136/bmj.38636.593461.68. URL: https://doi.org/10.1136/bmj.38636.593461.68.

[2] Matthew Butterick. *Butterick’s Practical Typography*. 2nd ed. https://practicaltypography.com/(visited 2021-05-02). O’Connor’s, 2021.