How to write high quality papers in algorithmic or experimental Computer Science

Veljko Milutinovic
(University of Belgrade, Serbia,
vm@etf.rs)

Jelica Protic
(University of Belgrade, Serbia,
jeca@etf.rs)

Goran Milutinovic
(University of Belgrade, Serbia,
goranmilutinovic2005@yahoo.com)

Goran Rakocevic
(Mathematical Institute, Belgrade, Serbia
goran.rakocevic@turing.mi.sanu.ac.rs)

Nenad Korolija
(University of Siena, Italy,
nenadko@etf.rs)

Oskar Mencer
(Imperial College, London, UK and Stanford University, US,
mencer@maxeler.com)

Abstract: Besides doing innovative and novel research, every scientist faces the question of what, how, and when to publish. Understanding this issue and having a well thought-out publication plan is nearly as important for an academic's career as are the scientific results he or she produces. **This paper is inspired by the success of Maxeler, London and Palo Alto.**

Keywords: Research methodology, generating new ideas in comupter science, writing scientific papers
Categories: A0

1 Introduction

Some universities require PhD students to publish one, two, or three papers in JCR journals (SCI, SSCI, and AHCI), before they can graduate. Others do not pose such direct publication requirements, but the students are, never-the-less, expected to carry out research leading to publishable results. This paper presents a set of guidelines that teach PhD students how to write two different types of JCR journal papers:

(A) Survey papers,
(B) Research papers.

The presented research methodology is justified by the following notions [Aspray 2006], [Leopold-Wildburger 2010]:

(A) One first has to become aware of all existing research concerning the problem as a prerequisite for being able to develop one's own contribution to the field.
(B) Once a good idea (for a new contribution) is generated, it has to be elaborated, using a scientific method.

The presented approach is related to Computer Science and Computer Engineering, but may be used in other domains as well. The presented approach works especially well for those parts of information science that that are algorithmically oriented.

2 Survey Paper

A survey paper can bring lots of citations, if it is the first one in a newly emerging field, is well written, and is published in a good journal. Consequently, selection of the topic for a survey must satisfy the following requirements:

a) The field is newly emerging.
b) Popularity of the field will grow over time.
c) A critical number of papers with new algorithms/approaches does exist.
d) A survey paper does not exist (a discussion is given at the end of the paper).
e) The PhD student worked before in a related scientific field.
f) The PhD student is enthusiastic about the particular field of the tutorial paper.

After the collected papers with original algorithms/approaches have been read and understood, the next step is to think about appropriate classification criteria. With a binary (or n-ary) criteria, one can create either a tree-like classification or a cube-like classification, as indicated in Figures 1 and 2. The presented cube-like scheme is suitable for cases with 3 classification criteria, but may be adapted for less (by moving to a square-like scheme), or more criteria by moving to a hypercube-like scheme. The latter may prove somewhat difficult for visualization, and should only be used if alternatives fail to capture an important point.

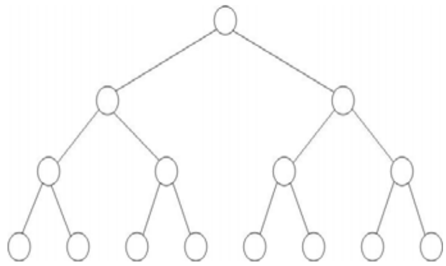

Figure 1: A tree-like classification: Classes are only at the leaves of the tree.

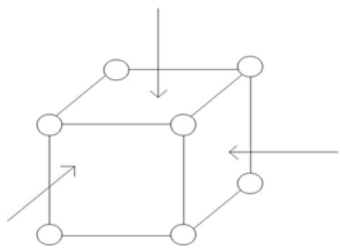

Figure 2: A cube-like classification: Classes can exist also at points inside the cube, as pointed to by the three arrows.

When structuring a survey paper, one can use the template presented next. For short surveys, each template element is a sentence. For long surveys, each template element is a paragraph. For books, each template element can be a page or more (like in [Crouzet 2012]). An example template is:

a) Seven Ws about the survey example (Who, What, When, Where, Why, for Whom, How).
b) Essence (it is extremely difficult to give entire essence in only one sentence).
c) Structure (at this place, one can insert a reference to a figure, like in [Draskovic 2012]).
d) Some relevant details.
e) Example (here one can reference a figure that explains an example using a pseudo-code, like in [Draskovic 2012])
f) Pros and cons.
g) Author’s opinion of this example and its potentials.

3 About Generation of New Ideas

There are a number of ways to inspire creativity: [Doerfler 2008], [Friedman 2010], [Paterson 2001], [Perl 2009]. Here we present a set of methods that can be used by PhD students.

3.1 Mendeleyevization: Catalyst versus Accelerator

If one of the classification classes includes no examples, it first has to be checked why is this is the case. If it is because the technology or the applications are not yet ready for such an approach, one can act in the same way as the famous chemists Mendeleyev: empty positions in any classification are potential avenues leading to new inventions. As indicated in Figure 3, these inventions sometimes need a catalyst, a resource that makes an invention happen. An accelerator is a resource that turns a known, but unrealized idea into a highly valued invention (a number of such examples is given in [Vardi 2011]).

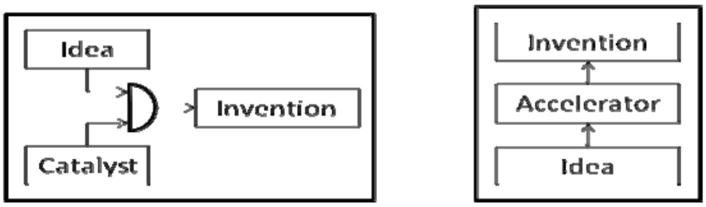

Figure 3: Catalyst versus Accelerator (Mendeleyevization)

3.2 Hybridization: Symbiosis versus Synergy

Sometimes two classification classes can be combined, in order to obtain a hybrid solution (hybridization). Hybrid solutions can be symbiotic (switching from one approach to the other) or synergistic (creating a new approach, which, for each particular solution element takes the better solution element of two different approaches). This is shown in Figure 4. The assumption here is that one solution is better under one set of conditions, and the other solution is better under another set of conditions.

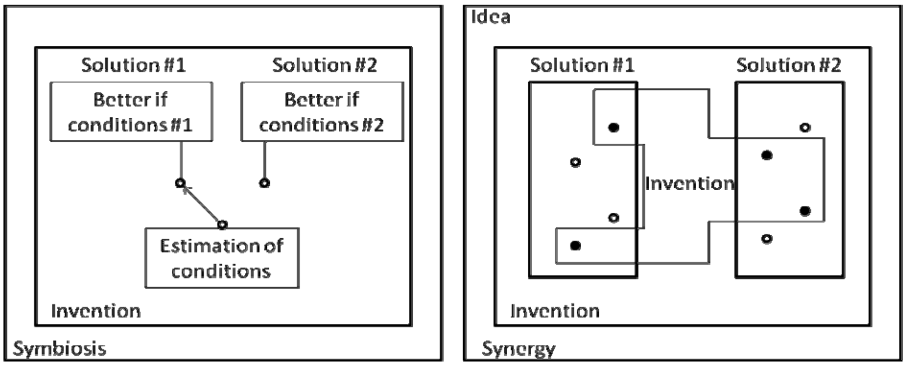

Figure 4: Symbiosis versus Synergy (Hybridization)

3.3 Transdisciplinarization: Modifications versus Mutations

Good ideas are often generated when algorithms, procedures, or ways of thinking, are ported from one field to another (transdisciplinarization). As indicated in Figure 5, for an idea to work better in the new field, either smaller modifications or larger mutations have to be introduced.

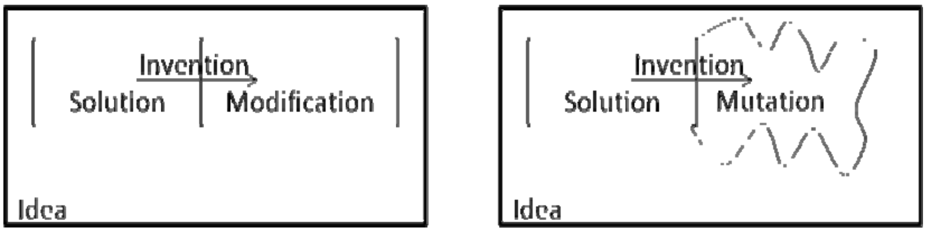

Figure 5: Modification versus Mutation (Transdisciplinarization)

3.4 Retrajectorization: Reparametrization versus Regranularization

Sometimes it is simply the best to take a research trajectory different (even opposite) compared to what others take (retrajectorization). The different research trajectory may make sense either due to technology changes, or due to application changes), as indicated in Figure 6. The two alternatives are referred to as a reparametrization and regranularization.

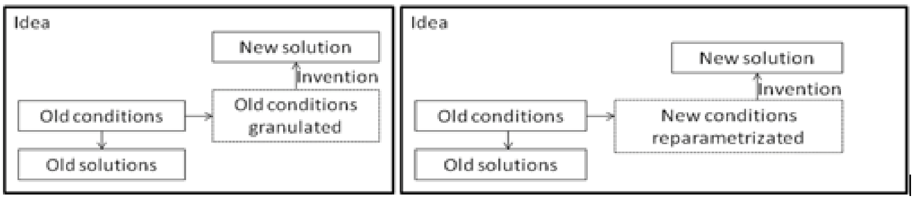

Figure 6: Re-parameterization versus Re-granularization

3.5 Unorthodoxization: ViewFromAbove versus ViewFromInside

This category encompasses the approaches that are difficult to classify (Figure 7): Sometimes one sees something that others did not see for decades or centuries (ViewFromAbove) or one gets struck by an idea of a genius with no ground in existing research (ViewFromInside). Popular examples include the contributions of Nobel Laureates Martin Perl [Perl 2009] and Jerome Friedman [Friedman 2010].

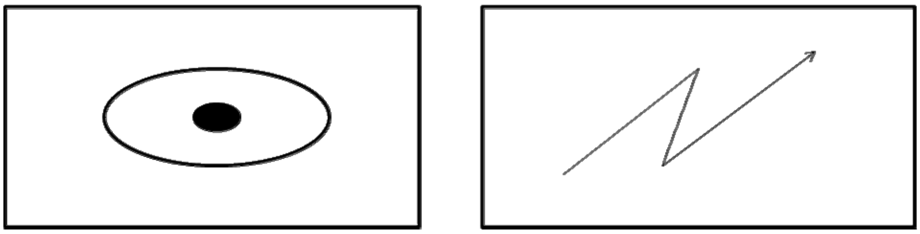

Figure 7: Unorthodoxization (ViewFromAbove versus ViewFromInside)

4 Research Paper

The major purpose of the research paper is to describe an innovation and to demonstrate that, under certain conditions, it has a better performance and/or complexity, compared to the best one from the open literature. The major steps in the process are:

a) To create an invention.
b) To perform a rigorous analysis, to demonstrate that the invented solution is better in performance, and to show under which condition.
c) To perform a analysis showing complexity and performance/complexity.

As far as the presentation of the research results, each research paper should contain the following twelve sections:

The major purpose of the research paper is to describe an innovation and to demonstrate that, under certain conditions, it has a better performance and/or complexity, compared to the best one from the open literature. The major steps in the process are:

a) To create an invention.
b) To perform a rigorous analysis, to demonstrate that the invented solution is better in performance, and to show under which condition.
c) To perform a analysis showing complexity and performance/complexity.

As far as the presentation of the research results, each research paper should contain the following twelve sections:

1) Introduction:

The minimum introductory text should contain the following three paragraphs:

a) About the general field of this research.
b) About the specific field of this research.
c) About the viewpoint of this research, as well as the goals of this research.

2) Problem statement:

The following elements are obligatory:

a) Problem definition.
b) Why is the problem important.
c) Why will the importance of the problem grow over time.

3) Existing solutions and their drawbacks, looking from the viewpoint defined in introduction, and having in mind the goals defined in introduction. Elements of this section are:

a) A brief classification of the best solution from the open literature.
b) Short description of each relevant solution.
c) A detailed criticism of each presented solution, especially in the domains in which the proposed solution is expected to be better.

4) The proposed solution and its essence, and why is it supposed to be better compared to the best solution from the open literature; elements of this section are:

a) Philosophical essence of the proposed solution.
b) Why the proposed solution is without drawbacks of existing solution(s).
c) What is the best methodology to prove the superiority of the proposed solution, and under what conditions that holds.

5) Details:

This section should contain details of the best one among the existing approaches and of the proposed solution. The relevant details should be grouped into categories. For example:

a) Hardware details.
b) System software details
c) Application software details.

6) Axioms, conditions, and assumptions of the analysis to follow:

a) Axioms refer to axiomatic standpoints.
b) Conditions refer to realistic specifiers of the environment.
c) Assumptions refer to simplifications that make the analysis easier, without jeopardizing on the quality of the final result.

7) Mathematical analysis:
   a) Axioms, conditions, and assumptions are described mathematically.
   b) Closed or open form formulae are derived for the major performance measures.
   c) Closed or open form formulae are derived for the major complexity measures.

8) Simulation analysis to show performance:
   a) Simulator, logical structure and user interface are described.
   b) Simulation experiments are described.
   c) Simulation results are discussed.

9) Implementation analysis to show complexity:
   a) Implementation strategy is discussed for the chosen technology.
   b) Implementation details and complexity are presented.
   c) If a prototype is implemented, show some characteristic measurement.

   If a prototype is not implemented, give some implementation guidelines.

10) Conclusion:
   a) Summary of what was done and to what extent are the initial goals achieved.
   b) To whom is that of benefit.
   c) Newly open problem for further research.

11) Acknowledgments:
   a) To all those who patiently listened to your ideas and especially to those who volunteered to share with you some of their own ideas for further benefit of your research. Also, it is obligatory to cite the relevant work of all those who volunteered the improvement ideas.
   b) To all those who helped provide the infrastructure for your research. If this is related to one or more research project, list them.

12) Annotated references:

The references are more useful if listed in groups. Each topic requires different grouping. The grouping that seems most appropriate for this paper includes:

   a) References related to methodology.
   b) References related to examples.
   c) References related to success of past students.

5 Analysis of the past

Finally, we present a short analysis based on a number of papers created following a framework similar to the one described here. The period in which these publications were generated spans over a period of over a decade, and the topic vary.

Figure 8 presents the impact of the existing, prior, surveys upon the citation count of a survey paper. In contrast to what may expected, paper which were not the first survey on a topic generated more citations than the ones who were. The reasons for this may vary, the field may have evolved, leaving room for a better understanding of issues, novel solutions may have been introduced, making the old survey obsolete, etc. Of cores, the presented numbers are by no means statistically significant, nor do we wish to claim that this is a general rule. The discussion is only presented as a demonstration that even with existing surveys, it sometimes well worth to invest the time and effort into a new one.

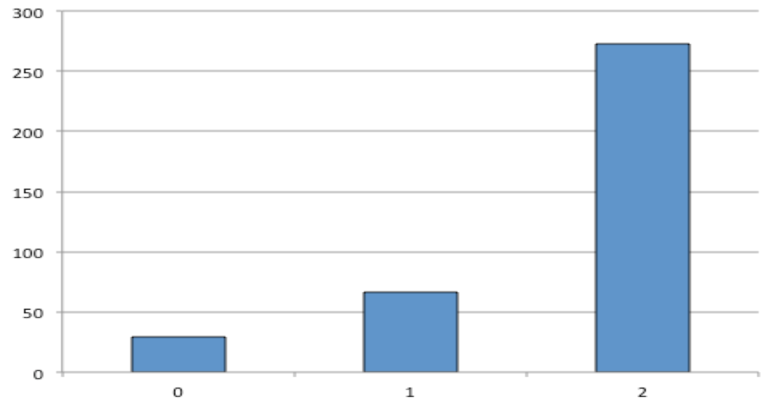

FIGURE 8: Impact of the existence of another survey paper. Explanation: This figure gives a result, which was absolutely unexpected. The expectation was that existence of a survey would decrease citations of our survey, but absolutely the opposite happened.

Figure 9 sheds gives an example of the numbers of citations generated by survey and research papers, as well as papers that fall in between (papers with an extensive section surveying the field and presenting novel ideas). Again, we do not claim that these numbers present a general rule, but do show that a survey publication may be well worth the effort, especially given the fact that one must obtain a clear view of a field before conducting research. Thus, when entering into research in a domain new to a person (which is, more or less always the case with a starting PhD student), much of the work will be need to be done anyway.

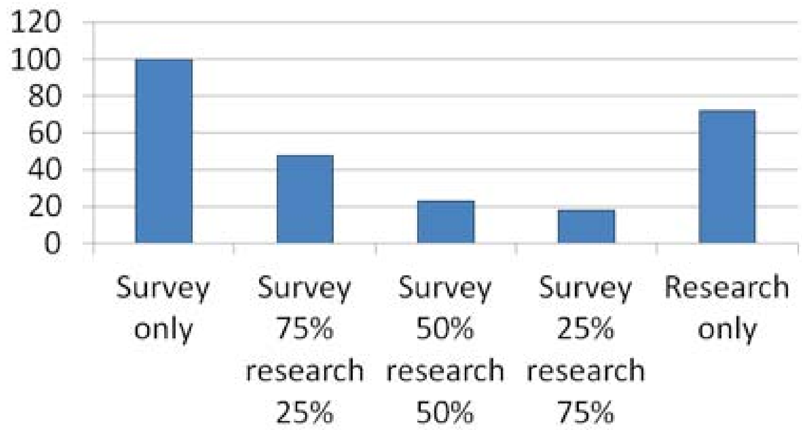

FIGURE 9: Citation counts Survey vs. Research Explanation: Surveys generate more, unless an extraordinary research paper is generated in a popular field.

Acknowledgements

The concepts presented here were generated during a past DARPA project related to a GaAs RISC microprocessor [Milutinovic 1986]. The authors are especially thankful to professor Michael J. Flynn of the Stanford University.

References

[Aspray 2006] Aspray, W., Mayadas, F., Vardi, M.Y., Zweben, S.: Panel Educational Response to Offshoring. *Proc. 37th ACM Symp. on Computer Science Education*, 2006, pp. 330–331.

[Crouzet 2012] Crouzet Y., and K., Kanoun K., System Dependability - Characterization and Benchmarking, *Advances in Computers*, Volume 84, Edition 1 (2012).

[Doerfler 2010] Dorfler, V. Baracskai, Z., Velencei, J., Understanding Creativity *Transactions on Advanced Research,* Volume 6, Number 2, (2010), pp.18-27

[Draskovic 2012] “A Classification of Mutational Approaches to Genetic Search,” *IEEE ICIT Conference*, Athens, Greece, March 19-21, 2012.

[Friedman 2010] Friedman, J., Why We Need Basic Research, *Transactions on Advanced Research*, Volume 4, Number 1, (2010), pp.2-6

[Leopold-Wildburger 2010] Leopold-Wildburger, U., Schutze, J., Verfassen und Vortragen, Wissenschaftliche Arbeiten und Vortraege leicht gemacht, Springer Verlag, Wien, 2010.

[Milutinovic 1986] Milutinovic, V., Fura, D., Helbig, W., *An Introduction to GaAs Microprocessor Architecture for VLSI*, IEEE Computer, Volume 19 Issue 3, 1986.

[Paterson 2001] Patterson, D, “How to have a bad career in research and academia, ” U.C. Berkeley, November 2001.

[Perl 2009] Perl, M., Creativity in Science, *Transactions on Advanced Research* Volume 5, Number 1, (2009), pp. 2-5.

[Vardi 2010] Vardi, M.,Y., Globalization and Offshoring of Software revisited. *Communications of the ACM* 53(5), p. 5, 2010.

[Vardi 2011] Vardi, M.Y., Fumbling the future. *Communications of the ACM* 54(3), p. 5, 2011.