

Creating A Research Space In Software Engineering: Structure For Writing Introduction

Abdul Rehman Gilal, Adil Khan, Ahmad Waqas, Murk Chohan, Hafiz Ahmed Ali

Abstract— Introduction is an important part of research articles. It is the first opportunity to make good impression and establish the credibility over the reading audience. Well-written introduction of a research article increases the chance of paper acceptance and citations. The objective of this study is to produce a structure for writing introduction section for software engineering (CARSSSE) research articles. Nine highly cited research articles from “IEEE Transaction on Software Engineering” are selected for pattern extractions. Creating a Research Space (CARS) model is kept as a baseline with additional three parameters: sentences, paragraphs and references. Keeping the CARS model in view, from the selected studies, “occupying a niche” move obtains around 48% of appearance in the introduction section whereas “establishing a territory” and “establishing a niche” moves carry 34% and 18% respectively. The proposed structure can further be extended and precise by adding a few more studies.

Index Terms—: Software articles, Writing, Introduction, CARS model, research papers

1 INTRODUCTION

Publication is an important means for communicating research. Two famous models for structuring the scientific papers are usually discussed in the literature [1]: Hourglass Model and King Model. In both models, “introduction” is necessary part. It is also one of the most studied section of research articles [2]. Introduction helps the readers to understand from a general subject of area to a particular field of research [1]. It is the first opportunity to make good impression and establish the credibility over the reading audience. Well-written introduction of a research article increases the chance of paper acceptance and citations. Several studies [2]–[9] have attempted to give a structure for writing effective introduction section. One of the most popular and widely accepted model is “create a research space (CARS)” [10]. The CARS model is comprised of three moves to write effective introduction: (1) establishing territory, (2) establishing a niche, and (3) occupying the niche. Each move in the CARS model is constituted on certain steps. For instance, claiming centrality, making topic generalizations, and reviewing previous research are the steps of establishing territory. In the same way, counter claiming, indicating a gap, question-raising, and continuing a tradition are the steps of establishing a niche move. Lastly, occupying the niche steps are: outlining purposes, announcing present research, announcing principle findings, and indicating the structure of research article.

Anthony conducted a study [6] to verify the accuracies of the standard models for writing introduction in SE. He claimed that a number of important features are not considered while writing introduction of research articles. For example, detailed review of background, evaluation of research, and inclusion of important definitions and examples. In order to discuss on the

CARS model, Anthony [6] also claimed that CARS model is mostly used but it is missing number of important features. Moreover, CARS model [8] also does not provide any solid quantifications for writing introduction in SE. For instance, how many numbers of paragraph should be written to establish a territory? The same question can be aroused for sentences, words or references. Because they make the foundation of any research’s introduction. Therefore, this study is carried out with an objective to define a structure for writing introductions in SE research articles. The CARS model is kept as a baseline with additional three parameters: sentences, paragraphs and references. The purpose of this study is to quantify the structure of introduction so that SE researchers could get some standard guidelines. The following section covers a related work that discuss other models for writing introduction section. Later, the methods of the study are presented with the study selection process. Results and discussion section highlights the significant results and their discussion based on the study objective. At the end of the article, conclusion is drawn from the major outcomes of the study. The very section also presents the future directions for further extensions.

2 RELATED WORK

Introduction is a way of defining purpose and theme of research paper. It leads the reader to understand the view from general to specific area of work. Introduction can be started with the announcement of topic or area on which research has been conducted (i.e., [11]–[18]). Several scholars had used CARS model for writing introductions. The model is proposed by John Swales. Introduction catches the reader’s interest if it is well pronounced and well written with specific tone. Each move of CARS model has its own significance with respect to declaring one’s work. John Swales performed text analysis for explaining genre while using rhetorical-linguistic approach. Similarities and differences has been examined in communicative events and individual users.

Samraj [2] analysed introduction of research articles from two fields: conservation biology and wildlife behaviour. She selected twelve research articles from conservation biology and wildlife behaviour published in 1995. All the research articles were analysed by using CARS model. After reviewing the introductions, Samraj suggested that how the CARS model can be modified into those two fields. For instance, she modified establishing territory move by adding sub steps into claiming centrality: claiming centrality in either research or in real world. Additionally, topic generalization and reviewing research steps were removed and new step was added:

- Abdul Rehman Gilal is an Assistant Professor in the Department of Computer Science, Sukkur IBA University, Pakistan. E-mail: a-rehman@iba-suk.edu.pk
- Adil Khan is currently pursuing masters degree program in computer science at Sukkur IBA University, Pakistan,
- Ahmad Waqas is an Associate Professor in the Department of Computer Science, Sukkur IBA University, Pakistan. Email: ahmad.waqas@iba-suk.edu.pk
- Murk Chohan is currently pursuing masters degree program in computer science at Sukkur IBA University, Pakistan,
- Hafiz Ahmed Ali is currently pursuing masters degree program in computer science at Sukkur IBA University, Pakistan.

presenting background information. Samraj further mentioned that reviewing previous items of research can be subsumed under the step “indicating a gap” because it can be discussed during creating gap. She also modified “indicating a gap” step by considering either in research or in real world. Author also added new step into establishing a niche move which is “presenting positive justification”.

Introduction, Methodology, Research And Discussion (IMRAD) [9] model is also a well renowned model for writing research articles. IMRAD model mainly focus on all parts and subparts of any research article: from Introduction to discussion focus is majorly distributed. Posteguillo [8] analysed research articles’ structures by taking forty studies from different articles of academic journals in the field of computing. Results of the study [8] indicate that IMRAD pattern are not applied on articles in computer science field systematically. While, analysing introduction based on CARS model, Posteguillo identified “topic generalization” step was mostly used by computer science authors (65.00%) and claiming centrality appeared with 47.50%. Similarly, “counter claiming” step was avoided by CS authors as it obtained 2.5%. “Indicating gap” is highly written step in CS introductions with 57.00%. In the last phase of CARS model in Posteguillo’s study, announcing present research appeared highly frequent by 95% and principle findings and outlining purpose appeared as 70% and 25% respectively. Furthermore, Ozturk [7] also conducted the exploratory research to explore the introductions. He analysed 20 research articles to investigate the difference in the structure of introduction of two sub disciplines of applied linguistics: second language acquisition (SLA) and second language writing (SLW). It was observed that moves of CARS model are not related in two different sub-disciplines: SLA and SLW. This study was only focused on organization of moves rather than steps included in each move. Based on the results, Ozturk claimed that CARS model was inaccurate on the selected studies. However, CARS model was useful for descriptive reference. He also observed that variations can also be identified in related disciplines.

3 METHODOLOGY

In order to obtain the study objective, this study selected 09(these are called as S1 until S9 in this study [19]–[27]) highly cited articles from the journal “IEEE transactions on Software Engineering” between the years 2008 and 2018. The journal is highly reputed in the study domain. The citation based study selection had basically two benefits: (1) citation makes the article’s acceptance which means that article has clear understanding and (2) to reduce the selection bias. Moreover, all of the selected articles were reviewed based on the study measure “create a research space in software engineering (CARSSE)”. The measure was developed based on the CARS model’s attributes. Each attribute of the model was quantified based on the parameters: number of words, number of sentences, tense type, number of paragraphs, and number of references. These simple frequency quantifications had helped us to summarize the general structure of the highly reputed articles which discloses the working weight against each attribute. The number of references also helped us to direct our efforts toward the best practices. In a nutshell, the parameters which were focused are paragraphs, sentences and references.

4 RESULTS AND DISCUSSION

The first thing that is realized during reviewing the articles, most of the cited articles were based on software testing. It is said because seven out of selected nine articles were based on software testing. To summarize the results and discussion part effectively, we have segregated the section into paragraphs, sentences and references sub sections.

4.1 Paragraphs

The paragraph frequency had given us a general

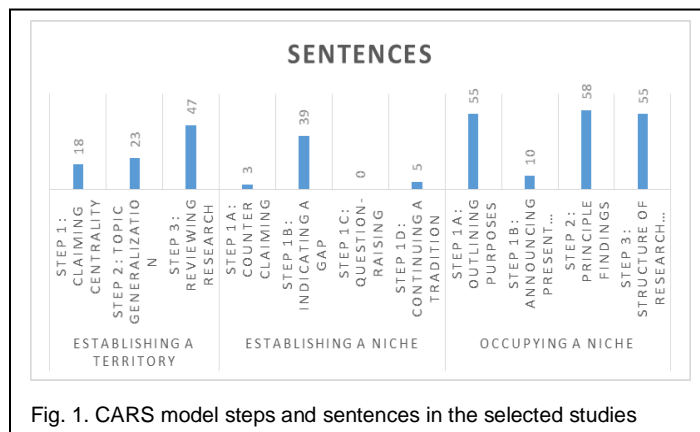


Fig. 1. CARS model steps and sentences in the selected studies

understanding about the overall structure of introduction section. It has also helped to understand the way authors usually discuss different parts of introduction. From the aerial view, the authors have commonly written eight to nine paragraphs in the introduction section. While keeping in view the CARS model, we realized that author usually spares two to three paragraph to establish a territory in the introduction. There was on study S6 which established the territory in 5 short paragraphs. Based on our understanding, authors explained the importance of the topic with the help of literature support during establishing territory as it was a systematic literature review (SLR). Table 1 below presents the number of paragraphs found in each study against each move of CARS model.

TABLE 1
CARS MODEL MOVES AND PARAGRAPHS FREQUENCY IN THE SELECTED STUDIES

Move	S 1	S 2	S 3	S4	S 5	S6	S7	S 8	S 9	Total
Establishing a Territory	2	2	2	4	4	5	3	1	2	25
Establishing a Niche	3	2	1	1	1	1	1	1	2	13
Occupying a Niche	4	4	3	3	2	5	7	4	3	35
Total	9	8	6	8	7	11	11	6	7	

The second move “establishing a niche” appeared to be a short move in the introduction in SE domain. In the selected studies, six articles wrote only one paragraph to discuss the move. However, 1 to 2 paragraphs appeared as an average for the second move of CARS model. Moreover, the last move of CARS model was the most weighted one in the introduction section. We counted total 35 paragraphs from 73 total paragraphs only on this move (i.e., occupying a niche). In other words, occupying a niche move obtains around 48% in

the introduction section whereas establishing a territory and establishing a niche moves carry 34% and 18% respectively.

4.2 Sentences

During paragraph results, we only focused the CARS model's moves. As mentioned earlier that each CARS move has certain steps. This section had also summarized the results by keeping the steps of each move in view. At the first quick sight, we counted total 178, 88 and 47 sentences from the nine selected studies for occupying a niche, establishing a territory and establishing a niche respectively. Total 35 sentences were found as an average in the introduction section of those selected studies. There were three steps in the first move: claiming centrality, topic generalization and reviewing research. Based on the selected studies, we found 10 sentences as an average for establishing territory move. We also found that authors have written 3-5 sentences in each steps. However, little extension was found in the step "reviewing research" in which authors of studies S5, S6 and S7 wrote 13, 14 and 10 sentences respectively. Of course,

reviewing more studies would strengthen the theory grounds. Furthermore, we found a critical fact in the second move "establishing a niche", authors have majorly highlighted one step "indicating a gap" which is a second step in the move. Every single selected study had written 6-8 sentences to discuss about gap. On the other side, counter claiming and continuing a tradition steps were found in only two studies S7 and S9. While, "question raising" step did not appear in the any selected study. In other words, establishing a niche means to indicate gap. We summarize the overall results in the following figure 1. In the third move "occupying a niche" was the most frequent move in the selected studies. This move had three steps: (1a) outlining purpose or (1b) announcing present research, (2) principle findings and (3) structure of the article. Every step was highly frequent. We counted 20 sentences per step in the third move which is the highest average in the overall CARS steps. The following table 2 presents the overall appearance of the sentences in the CARS model.

TABLE 2
OVERALL APPEARANCE OF THE SENTENCES IN THE CARS MODE

Move	Steps	S1	S2	S3	S4	S5	S6	S7	S8	S9	Total sentence	Total words
Establishing a Territory	Step 1: Claiming Centrality	2	1	3	5	2	1	1	2	1	18	394
	Step 2: Topic Generalization	2	3	4	0	7	0	7	0	0	23	599
	Step 3: Reviewing Research	3	0	1	14	13	10	0	4	2	47	1226
	<i>Sub total</i>	7	4	8	19	22	11	8	6	3		
Establishing a Niche	Step 1a: Counter Claiming	0	0	0	0	0	1	0	0	2	3	79
	Step 1b: Indicating a Gap	8	2	6	6	4	2	2	4	5	39	890
	Step 1c: Question-Raising	0	0	0	0	0	0	0	0	0	0	0
	Step 1d: Continuing a Tradition	0	0	0	0	0	0	0	0	5	5	91
	<i>Sub total</i>	8	2	6	6	4	3	2	4	12		
Occupying a Niche	Step 1a: Outlining Purposes	2	17	3	12	0	4	4	7	6	55	1522
	Step 1b: Announcing Present Research	0	0	0	0	10	0	0	0	0	10	217
	Step 2: Principle findings	10	20	4	0	0	9	10	2	3	58	1477
	Step 3: Structure of Research Article	6	0	6	5	9	7	13	6	3	55	1115
	<i>Sub total</i>	18	37	13	17	19	20	27	15	12		
	Total	33	43	27	42	45	34	37	25	27	313	7610

We further realized that authors had generally written the introduction in the present tense. However, some of the sentences were found in the past tense. The past tense was used majorly in three steps: reviewing research, counter claiming. Furthermore, total 7610 words were counted in overall 9 studies. More or less, every sentence was of 20-24 words.

4.3 References

The last parameter of the study was to quantify the references. Total 221 references were counted in the selected studies. In these studies, S5 had reviewed 74 articles which was not observed as a common practice in the other selected studies. Based on the total number of references, the average number

of reference is 25. However, if we consider S5 as an outlier, the average goes blew to 19. Not only every author writes differently but every study demands different structure. Based on our understanding, S5 is a survey based study which might have demanded authors to strengthen their study by adding several other studies. Because S5 study authors had used total 76 references and 74 of them were only in the reviewing research section. Moreover, the major reference's citations were found in the steps: topic generalization, reviewing research, indicating a gap and outlining a purpose. On the other hand, steps like claiming centrality, counter claiming, continuing a tradition, announcing present research and structure of research article were appeared to be weakly cited. Table 3 below displays the number of references used in the different steps of CARS model.

TABLE 3
CARS MODEL AND REFERENCE WEIGHT

Move	Steps	S1	S2	S3	S4	S5	S6	S7	S8	S9	Total
Establishing a Territory	Step 1: Claiming Centrality	0	0	2	2	0	0	0	0	0	4
	Step 2: Topic Generalization	0	13	1	0	0	0	13	0	0	27
	Step 3: Reviewing Research	10	0	7	10	74	7	0	3	14	125
Establishing a Niche	Step 1a: Counter Claiming	0	0	0	0	0	0	0	0	4	4
	Step 1b: Indicating a Gap	6	1	0	2	0	0	0	6	2	17
	Step 1c: Question-Raising	0	0	0	0	0	0	0	0	0	0
	Step 1d: Continuing a Tradition	0	0	0	0	0	0	0	0	0	0
Occupying a Niche	Step 1a: Outlining Purposes	0	11	0	0	0	0	0	1	4	16
	Step 1b: Announcing Present Research	0	0	0	0	2	0	0	0	0	2
	Step 2: Principle findings	5	3	0	0	0	0	4	0	10	22
	Step 3: Structure of Research Article	0	0	0	0	0	0	4	0	0	4
Total		21	28	10	14	76	7	21	10	34	221

At the end of the discussion section, this study summarizes the introduction structure for software engineering research articles based on the CARS model. Table 4 had described the study parameters with minimum and maximum average occurrences. We have deleted the question raising step from

establishing a niche move as it did not appear in the selected studies. Based on the minimum appearance, the structure requires to have 34 sentences in 6 paragraphs with 18 references. Similarly, the maximum appearance includes 70 sentences in 9 paragraphs with 26 references.

TABLE 4
CARSSE STRUCTURE

Move	Steps	Sentences	Paragraph	Reference
Establishing a Territory	Step 1: Claiming Centrality	2-4	2-3	10-14
	Step 2: Topic Generalization	3-5		
	Step 3: Reviewing Research	6-10		
Establishing a Niche	Step 1a: Counter Claiming	1-2	1-2	3-5
	Step 1b: Indicating a Gap	5-8		
	Step 1d: Continuing a Tradition	1-2		
Occupying a Niche	Step 1a: Outlining Purposes	6-12	3-4	5-7
	Step 1b: Announcing Present Research	2-3		
	Step 2: Principle findings	6-12		
	Step 3: Structure of Research Article	2-12		
Minimum		34	6	18
Maximum		70	9	26

5 CONCLUSION

The main objective of the study was to propose a structure for writing introduction sections in software engineering research articles. The CARS model involves the moves and steps to be included within the introduction section. Based on the selected nine studies, "occupying a niche" move had obtained around 48% in the introduction section. On the other hand, "establishing a territory" and "establishing a niche" moves carried 34% and 18% respectively. Additionally, this study had summarized that minimum 34 sentences should be written in 6 paragraphs with 18 references to strengthen the SE research articles. This study had also defined the structure for the studies whose introduction contains huge explanations. For example, one should write 70 sentences in 9 paragraphs with the help of 26 references for long introductions in SE. Moreover, one of the major limitations of the study is the number of selected articles. Hence, the proposed structure

could further be extended and precise by adding a few more studies. One can also add studies from other domains for generalizing it.

REFERENCES

- [1] M. Derntl, "Basics of research paper writing and publishing," *Int. J. Technol. Enhanc. Learn.*, vol. 6, no. 2, pp. 105–123, 2014.
- [2] B. Samraj, "Introductions in research articles: Variations across disciplines," *English Specif. Purp.*, vol. 21, no. 1, pp. 1–17, 2002.
- [3] J. Swales, *Genre analysis: English in academic and research settings*. Cambridge University Press, 1990.
- [4] J. M. Swales, *Aspects of article introductions*, no. 1. University of Michigan Press, 2011.
- [5] J. Swales and H. Najjar, "The writing of research article introductions," *Writ. Commun.*, vol. 4, no. 2, pp. 175–191,

- 1987.
- [6] L. Anthony, "Writing research article introductions in software engineering: How accurate is a standard model?," *IEEE Trans. Prof. Commun.*, vol. 42, no. 1, pp. 38–46, 1999.
- [7] I. Ozturk, "The textual organisation of research article introductions in applied linguistics: Variability within a single discipline," *English Specif. Purp.*, vol. 26, no. 1, pp. 25–38, 2007.
- [8] S. Posteguillo, "The schematic structure of computer science research articles," *English Specif. Purp.*, vol. 18, no. 2, pp. 139–160, 1999.
- [9] L. B. Sollaci and M. G. Pereira, "The introduction, methods, results, and discussion (IMRAD) structure: a fifty-year survey," *J. Med. Libr. Assoc.*, vol. 92, no. 3, p. 364, 2004.
- [10] J. Swales, "Create a Research Space (CARS) Model of Research Introductions," *ABOUT*, p. 12, 2014.
- [11] U. Lechtenberg, "Research Guides: Organizing Academic Research Papers: 4. The Introduction."
- [12] A. R. Gilal, J. Jaafar, L. F. Capretz, M. Omar, S. Basri, and I. A. Aziz, "Finding an effective classification technique to develop a software team composition model," *J. Softw. Evol. Process*, vol. 30, no. 1, pp. 1–12, 2018.
- [13] A. R. Gilal, J. Jaafar, M. Omar, S. Basri, and A. Waqas, "A Rule-Based Model for Software Development Team Composition: Team Leader Role with Personality Types and Gender Classification," *Inf. Softw. Technol.*, vol. 74, pp. 105–113, 2016.
- [14] A. R. Gilal, J. Jaafar, M. Omar, S. Basri, and I. Din, "Balancing the Personality of Programmer: Software Development Team Composition.," *Malaysian J. Comput. Sci.*, vol. 29, no. 2, pp. 145–155, 2016.
- [15] A. R. Gilal, J. Jaafar, S. Basri, M. Omar, and A. Abro, "Impact of software team composition methodology on the personality preferences of Malaysian students," in *2016 3rd International Conference on Computer and Information Sciences, ICCOINS 2016 - Proceedings*, 2016.
- [16] A. Waqas, A. R. Gilal, M. A. Rehman, Q. Uddin, N. Mahmood, and Z. M. Yusof, "C3F: Cross-Cloud Communication Framework for Resource Sharing amongst Cloud Networks: An Extended Study," *Int. J. Comput. Sci. Netw. Secur.*, vol. 17, no. 8, pp. 216–228, 2017.
- [17] M. Z. Tunio, H. Luo, C. Wang, F. Zhao, A. R. Gilal, and W. Shao, "Task Assignment Model for crowdsourcing software development: TAM," *J. Inf. Process. Syst.*, vol. 14, no. 3, 2018.
- [18] A. R. Gilal, J. Jaafar, M. Omar, and M. Z. Tunio, "Impact of Personality and Gender Diversity on Software Development Teams' Performance," in *International Conference on Computer, Communication, and Control Technology (I4CT 2014)*, 2014, no. 2014 IEEE 2014, pp. 261–265.
- [19] E. T. Barr, M. Harman, P. McMinn, M. Shahbaz, and S. Yoo, "The oracle problem in software testing: A survey," *IEEE Trans. Softw. Eng.*, vol. 41, no. 5, pp. 507–525, 2015.
- [20] C. Le Goues, T. Nguyen, S. Forrest, and W. Weimer, "Genprog: A generic method for automatic software repair," *IEEE Trans. Softw. Eng.*, vol. 38, no. 1, pp. 54–72, 2012.
- [21] Y. Shin, A. Meneely, L. Williams, and J. A. Osborne, "Evaluating complexity, code churn, and developer activity metrics as indicators of software vulnerabilities," *IEEE Trans. Softw. Eng.*, vol. 37, no. 6, pp. 772–787, 2011.
- [22] A. Aleti, B. Buhnova, L. Grunske, A. Koziolok, and I. Meedeniya, "Software architecture optimization methods: A systematic literature review," *IEEE Trans. Softw. Eng.*, vol. 39, no. 5, pp. 658–683, 2013.
- [23] Y. Jia and M. Harman, "An analysis and survey of the development of mutation testing," *IEEE Trans. Softw. Eng.*, vol. 37, no. 5, pp. 649–678, 2011.
- [24] T. Hall, S. Beecham, D. Bowes, D. Gray, and S. Counsell, "A systematic literature review on fault prediction performance in software engineering," *IEEE Trans. Softw. Eng.*, vol. 38, no. 6, pp. 1276–1304, 2012.
- [25] R. Hoda, J. Noble, and S. Marshall, "Self-organizing roles on agile software development teams," *IEEE Trans. Softw. Eng.*, vol. 39, no. 3, pp. 422–444, 2013.
- [26] M. Shepperd, D. Bowes, and T. Hall, "Researcher bias: The use of machine learning in software defect prediction," *IEEE Trans. Softw. Eng.*, vol. 40, no. 6, pp. 603–616, 2014.
- [27] S. Lessmann, B. Baesens, C. Mues, and S. Pietsch, "Benchmarking classification models for software defect prediction: A proposed framework and novel findings," *IEEE Trans. Softw. Eng.*, vol. 34, no. 4, pp. 485–496, 2008.