

Developing Research Projects in SE and NLP

Nakul Sharma

Vishwakarma Institute of Information Technology, Kondhwa, Pune, India

*Corresponding author Email: nakul777@gmail.com

Abstract. *Research projects are necessary for conducting research or generating work products. Most of the work however, focuses more on the aspects of research within a domain instead of moving towards interdisciplinary work. In this chapter, the author proposes to develop research projects in perspective of SE and NLP. The future scope is also presented herein.*

1. Introduction

In [1] authors discuss an interdisciplinary technique, IRT. IRT gives a medium of checking artifact if it is Software Engineering (SE) or Natural Language Processing (NLP). There are several authors who have taken a broader study with respect to even subject level integration. For instance, in [2], the authors discuss the problems faced, and the benefits got by interdisciplinary research project. This project under the aegis of European Union (EU) focused on associative autopiotic digital ecosystem. The areas of integration included, the authors conclude that it is not possible to develop a single common framework for interdisciplinary research. The subjects being studied were:-

1. Social Science
2. Computer Science
3. Natural Science

The authors in [7] undertake study of large and diverse, multidisciplinary research projects. The research question answered is quantification of data collection efforts involved in such projects. The authors observe that interdisciplinary research directions are possible based on operational issues involved in such projects.

2. SE Domains

The SE domains are under review of the Software Engineering Body of Knowledge (SEBOK). Figure-1 provides the domains of SE according to SEBOK [5].

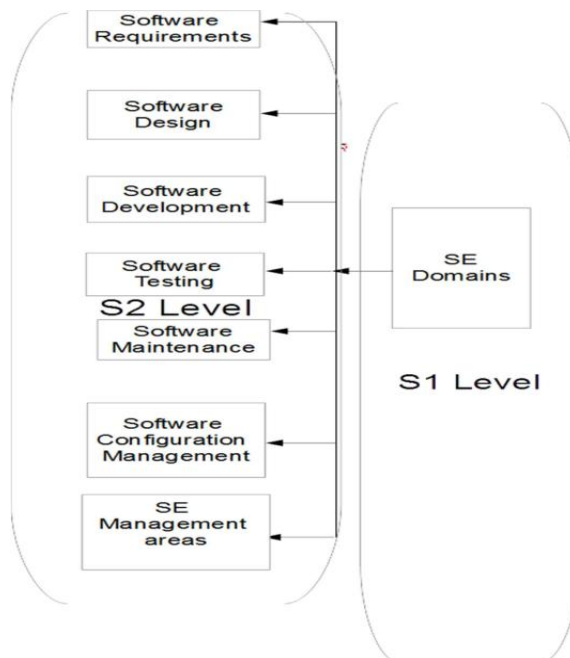


FIGURE 1. SE sub domains at Higher level of Abstraction (Part-1)

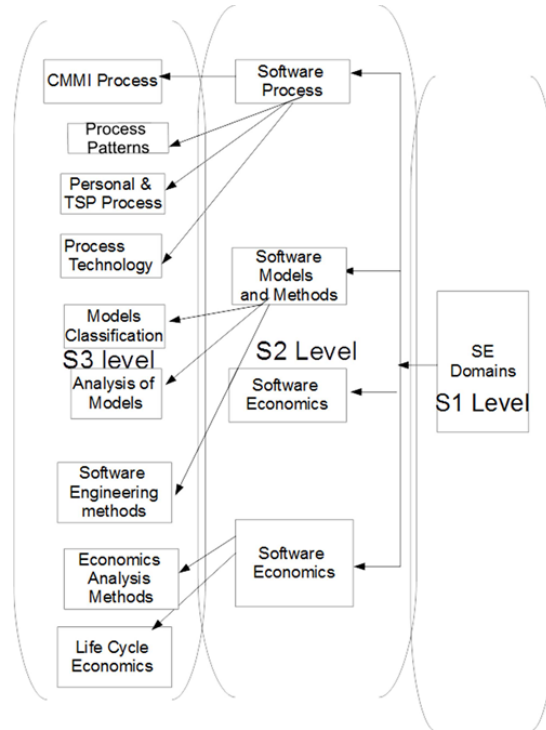


FIGURE 2. SE sub domains at lower level of Abstraction (Part-2)

Artifact or work product is an important component in SE. The authors in the paper [4] had provided a table about the importance of SE and NLP’s interrelation

3. NLP/Computational Linguistics Domain

The tasks actually stem from Machine Learning and various related fields such as text mining, data mining and information retrieval. The basic tasks of NLP/NLU/Computational linguistics include the following [4]:-

1. Analysis at Morphological level
2. POS Tagging
3. Name-Entity Recognition (NER)
4. Shallow and Deep parsing
5. Extraction of semantics behind the sentences/words
6. Discourse Discovery and Pragmatics behind sentences/words.

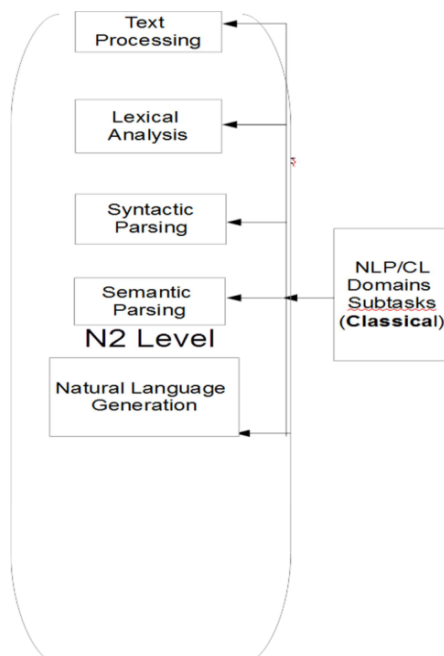


FIGURE 3. NLP/CL Subdomains

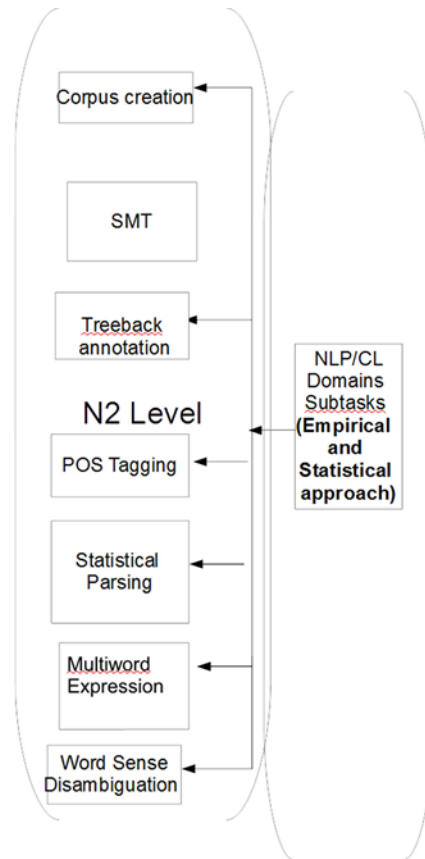


FIGURE 4. NLP/CL Subdomains at Lower level of Abstraction (Part-1)

These are the generic categorization of NLP/SE domains which is given in figure-1,2,3,4. The figures are for SE domain is divided into S1, S2,S3 while figures for NLP domain are divided into N1,N2.

Table-1 and Table-2 give the research projects premise on which future research projects can be build using RQI.

TABLE 1. Research projects premise of SE

Level	Representation
S1	Subject Level
S2	Topic Level
S3	Basic real-life application level

TABLE 2. Research Project Premise For NLP

Level	Representation
N1	Subject Level
N2	Topic Level

4. RQI Technique

As a case study to implement RQI technique, SE artifacts are considered for NLP and vice-versa.

TABLE 3. SDLC Artifact

Sr. No.	Phases in SDLC	Artifact
1	Analysis	Software Requirement Specification, Requirement Document, Use Case Description, Acceptance Test Cases
2	Design	Software Design Specification, UML Diagrams, Test cases
3	Coding	Source Code, executable, user manual, Technical Specification
4	Testing	Test Cases
5	Maintenance	Maintenance log, User Complaints, Customer Executive logs.

Interdisciplinary Technique (RQI):

The technique RQI involves following steps:-

1. Identify if the research is to be done more in SE domain or NLP domain.(Appendix-A)
- 1.1 If the research is done in SE major domain then just take some few forresearch

OR

- 1.2 If the research is done in NLP major domain then include some sub-areas of SE forresearch
2. Repeat
3. Stop after getting necessary researchpapers

Now, considering analysis phase, it has at least 3 different artifacts as Software Requirement Specification, Requirement Document, Use Case Description, and Acceptance Test Cases. All of these of textual specifications which are in natural language hence NLP can be applied. A research project consists of following main highlights:-

1. Project Title

Project title tells about the goals of any project. An example of Project title by applying the RQI technique and Eq-1 from [1] is as follows:-

1. Applying morphological analysis for Use Case description
2. Applying morphological analysis for Acceptance test cases
3. Morphological analysis for Software Requirement Specification.

The above title can also be extended to more specifications like including the sub-domains into the project. This will make project title and hence the project more specific.

The above project title applies NLP in SE domain of analysis phase. The SE artifacts can also be included into a NLP extensive research project. Following will be an example of NLP based project:-

1. Applying Agility in PoS software.
2. Applying Change management in a distributed NLP project.

5. Conclusion

In this chapter, an extension RQI technique is presented. The paper presents various domains of SE and NLP. It also shows the differentiation between SE and NLP domains. Finally how to create research projects using NLP and SE is also presented. The future work also includes making use of “keywords” in each project title and checking the feasibility of such projects.

References

- [1] Nakul Sharma, Prasanth Yalla, (2016), Possibility of Interdisciplinary Research-Software Engineering and Natural Language Processing. Advances in Information Sciences and Service Sciences (AISS), Volume 8, Number 2. Pages 24-31.
- [2] Francesco Nachira, Paolo Dini, Andrea Nicolai "Open Philosophies for associative Autopoietic Digital Ecosystems", [Available Online] <http://www.digital-ecosystems.org/book/pdf/0.3.pdf>
- [3] XXX, YYY, (2018), Developing Research Questions in Natural Language Processing and Software Engineering, International Journal On Informatics Visualization (JOIV), Vol-2, Number 4. Pages 268-270.

- [4] Nakul Sharma, Prasanth Yalla, (2016), Software Engineering and Natural Language Processing-How Can they be Together?, International Journal of Software Engineering and Its Applications, Vol. 10, No. 12 (2016), pp. 389-396, <http://dx.doi.org/10.14257/ijseia.2016.10.12.32>
- [5] Pierre Bourque, Richard E. Fairley, "SWEOK v3.0, Software Engineering of Knowledge", IEEE Computer Society, 2014
- [6] Nitin Indurkha, Fred J Damerou, "Handbook of Natural Language Processing", CRC Press, 2010.
- [7] R.A. Buswell, Lynda Webb, Val Mitchell and Kerstin Leder Mackley, Multidisciplinary research: should effort be the measure of success?. Building Research and Information, 45 (5), pp. 539-555.