

## Developing Research Questions in Natural Language Processing and Software Engineering

Nakul Sharma<sup>#</sup>, Prasanth Yalla<sup>#</sup>

<sup>#</sup> Koneru Laxminah Education Foundation, India  
E-mail: nakul777@gmail.com, prasanthyalla@gmail.com

**Abstract**— This paper endeavors to develop newer medium of developing research questions by keeping in view both fields of SE and NLP in proper perspectives. An overview of the current state of art research in SE and NLP is presented. This is done by referring to the SE Body of Knowledge (SEBOK). Analogues to SEBOK, there are no separate Body of Knowledge available for the NLP/Computational Linguistics (CL). Hence whatever falls within the category of NLP/CL was considered in framing the research categories from the NLP/CL side. The paper concludes with future scope of the research presented.

**Keywords**— Software Engineering, Natural Language Processing, Computational linguistics.

### I. INTRODUCTION

A research question holds the key to successful research. Software Engineering along with Natural Language Processing are emerging areas in Computing environment. They can be applied to various situations and contexts in which software or any textual content are developed and deployed. Software Engineering tasks, methods, processes, formal methods etc. Natural Language Processing also consists of various tasks and methods on which Software Engineering can be applied. The methods can be classical approach to NLP or Empirical approach to NLP.

In the current work, the focus is providing better perspectives in framing the most critical aspect of research, research questions.

In this work, the endeavor is to provide a framework for generating research questions by putting together both the areas of research into perspective. This is given in Equation-1.

$$(NLP)_{sub-area}^{***}(SE)_{sub-area} \quad (1)$$

The concepts in both the areas can be both qualitative as well as quantitative. Equation-1 can be used in different context in developing the research questions. The \* sign can represent a AND sign. The AND sign implies that both research areas must be taken into consideration while doing research. A generic methodology is generation of research questions can be evolved. A cursory reading of Software

Engineering gives an insight to the sub-areas in which research can be conducted.

### II. SOFTWARE ENGINEERING DOMAINS

Table-I summarizes the existing work in regard to the sub-areas in Software Engineering [1] [2].

TABLE I  
TABLE I. SOFTWARE ENGINEERING DOMAINS FROM\*\* [1]

Parameter	Software Engineering Areas
Software Process	<ol style="list-style-type: none"> <li>1. CMMI Process</li> <li>2. Process Patterns</li> <li>3. Personal &amp; TSP Process</li> <li>4. Process Technology</li> </ol>
Process Models	<ol style="list-style-type: none"> <li>1. The most traditional model, waterfall model</li> <li>2. additive process models includes RAD and incremental models.</li> <li>3. There are some process models which increase which are Evolutionary in nature:- <ul style="list-style-type: none"> <li>-Prototyping model</li> <li>-Concurrent model</li> <li>-Spiral model</li> </ul> </li> <li>4. Specialized process models <ul style="list-style-type: none"> <li>-Component Based</li> <li>-Formal methods</li> <li>-Aspect-oriented software development</li> </ul> </li> </ol>
Agile Development	<ol style="list-style-type: none"> <li>1. Basic Methodologies in Agile development</li> <li>2. Extreme Programming</li> <li>3. Adaptive Software Development (ASD)</li> <li>4. Dynamic Systems Development Method</li> <li>5. Scrum</li> <li>6. Feature Driven Development</li> </ol>
Requirement	Tasks

Engineering	1. Inception 2. Elicitation 3. Elaboration 4. Negotiation 5. Specification 6. Validation 7. Requirement Management
Analysis Model	1. Requirement Analysis 2. Data Modeling Concepts 3. O.O. Analysis 4. Modeling using scenarios 5. Modeling using Data Flow 6. Class-Based modeling
Architectural Design	This includes software as a architecture, Designing the Data, its flow within the architecture. The styles and patterns involving architecture. Some alternative ways of designing the architecture. Checking how the data gets flow from the software architecture.
Product Metrics	Software Metrics for each models in SDLC
Quality Management	1. Software Quality Assurance 2. Software Reviews 3. Formal Technical Review

### III. SOFTWARE ENGINEERING DOMAINS ACCORDING TO SEBOK

According the Software Engineering Body of Knowledge, there are following research areas in Software Engineering [2,3]:-

TABLE II  
SOFTWARE ENGINEERING DOMAINS FROM\*\* [2]

Software Configuration Management	1. Identification of different configurations for softwares being developed. 2. Controlling the software configuration. 3. Software Configuration Status Accounting 4. How software is delivered and released to the clients. 5. Software Configuration Auditing 6. Software Configuration Management Tools
Management within Software Engineering	1. Initiation and Scope Definition 2. Software Project Planning 3. Software Project Enactment 4. Review and Evaluation 5. Closure 6. Software Engineering Measurement 7. Software Engineering Management Tools
Software Engineering Process	1. Software Life Cycles 2. Software Process Assessment and Improvement 3. Software Measurement 4. Software Engineering Process Tools
Software Engineering Models and Methods	1. Types of Models 2. Analysis of Models 3. Software Engineering Methods
Software Quality	1. Software Quality Management Processes 2. Software Quality Tools
Software Engineering Economics	1. Life Cycle Economics 2. Risk and Uncertainty 3. Economic Analysis Methods

Parameter	Software Engineering Areas
Software Requirements	1. Requirements Basics 2. Requirement Process 3. Requirement induction 4. Requirement Analysis 5. Development of SRS 6. Validation of requirements through test case generation.
Software Design	1. Design of Software Basics 2. Architecture and structure of software 3. UI Design 4. The Design of Software Quality Analysis and its Evaluation 5. Notations for Software Design 6. Software Design Strategies and Methods 7. Tools used in designing the software
Development of Software	1. Managing Construction 2. Construction Techniques 3. Tools employed in development of software.
Software Testing	1. Software Testing Fundamentals 2. The different levels within testing 3. Different Techniques in undertaking testing 4. Test-Related Measures 5. The Process of Testing 6. Tools for testing the software.
Software Maintenance	1. Software Maintenance Basics 2. Main points in Software Maintenance 3. Maintenance Process 4. Techniques for Maintenance

### IV. SOFTWARE ENGINEERING DOMAINS ACCORDING TO SEBOK

There are two types of NLP approaches are classified as [3]:-

1. Classical Approaches to NLP
2. Empirical and statistical approach to NLP

#### *Classical Approaches to NLP*

The classical approach to NLP consists of following major sub-areas:-

1. Text Processing
2. Lexical Analysis
3. Syntactic Parsing
4. Semantic Analysis
5. Natural Language Generation

#### *Empirical and Statistical approach to NLP*

1. Corpus creation
2. Treebank annotation
3. SMT
4. POS Tagging
5. Statistical Parsing
6. Multiword Expressions
7. Word Sense Disambiguation

### V. METHODOLOGY FOR GENERATING RESEARCH QUESTION

In section-II,III,IV different sub-areas within Software Engineering domain as well as Natural Language Processing are shown. Using the equation-1, different sub domains of

SE and NLP can be researched for. As this involves more than one research area, interdisciplinary research methodology is presented named Research Question Interdisciplinary (RQI).

The technique RQI involves following steps:-

1. Identify if the research is to be done more in SE domain or NLP domain.
  - 1.1 If the research is done in SE domain then just take some few for research
- OR
- 1.2 If the research is done in NLP domain then include some sub-areas of SE for research
2. Repeat
3. Stop after getting necessary research papers.

This can hence help in creating a taxonomy of keywords. The keywords can be put in Equation-1 and can be used to get relevant research papers.

There can be 1:m relationship being applied or m:1 relation being applied as given in equation-1. So one or more parameter of SE can be applied to NLP. Reverse is also true. One or more parameter of NLP can be applied to SE.

#### VI. DISCUSSION ON THE PROPOSED WORK

In this paper, a framework for searching NLP and SE research papers is presented. Literature review forms the basis of any research. It can be used in following different ways of research:-

1. Generating Research Questions
2. Generating Research directions
3. Generating Research Projects

As SE and NLP itself contain lot of sub-areas, it is hence possible widen to scope of research and resulting in better products being developed.

#### VII. CONCLUSION

This paper provides a research direction for developing research questions by combining SE and NLP areas of research. The summarizes the essential portions of SE and NLP. The authors provide the methodology for generating research questions. Essentially, the research methodology presented can be applied to generate numerous research questions and hence projects which will increase innovation in the field of interdisciplinary research

#### REFERENCES

- [1] Roger S Pressman, "Software Engineering, A practitioners Approach", 6th Edition, Macgraw Hill Publication, 2012.
- [2] Pierre Bourque, Richard E. Fairley, "SWEOK v3.0, Software Engineering of Knowledge", IEEE Computer Society, 2014.
- [3] Project Management Body of Knowledge, IEEE publishing society, 2010.
- [4] Nitin Indurkha, Fred J Damerau, "Handbook of Natural Language Processing", CRC Press, 2010.  
\*\*-contents changed due to issues of philagiarism inspite of citations.