DV2545 Advanced Topic in Computing- Assignment 1

Sravika Kothawar
9501294582

Neeraj Reddy Avutu
9411053375

Abstract:

A comparative analysis between three articles is being presented on the research topic "Continuous Delivery". The format of the article and the way it is being presented is our major interest. This article is being written to state our perception by analysing the data. The intent of this study is to compare the articles.

Keywords: Continuous delivery; Continuous deployment; Customer feedback; Customer involvement; Software development; User feedback; User involvement, Agile Development, Delivery Pipeline, System Dynamics, Agile methods, Configuration management, Release Management, Software Evolution

Introduction

Continuous Delivery(CD) is a software development discipline where the software is built in such a way that it is released to the production at any time. The main aim is to build, test and release the software faster and frequently. The speed needed to develop and deliver the software in an agile business environment is enabled by continuous delivery, which is the present trending problem in release management. Based on our interest in the field of the agile system, we had considered 'continuous delivery' as a research topic for this assignment. The search string was formulated as "continuous delivery". This search string was used to conduct a search for relevant literature in Scopus. From the 108 results obtained,3 articles were selected. They are found to be apt to the selected title and is according to the abstract. Brief description of selected article is given below:

Article 1: This paper "Customer Involvement in Continuous Deployment: A Systematic Literature Review".

Article 2: "System Dynamics Modelling of Agile Continuous Delivery Process".

Article 3: "Introducing Continuous Delivery of Mobile Apps in a Corporate Environment: A Case Study".

COMPARISON OF ARTICLES:

PROBLEM AREA:

Article 1: The author has described the problem as the need to have a common overview regarding the topic 'Customer involvement in Continuous deployment', although continuous delivery is gaining attention in the software industry as an approach from continuous learning from customers. The motivation for the research is to build successful software product and services within the mentioned problem is Customer misperceptions and gaining customer's trust are being mentioned as major challenges. Research gaps were also mentioned. Related work isn't mentioned specifically but was well written in the results.

Article 2: The authors have stated the major problems in the article as the configuration management problem, lack of testing in a clone production environment and insufficient collaboration between the development time the deployment operations were stated as the major problems. The goals and objectives were stated clearly. Related work regarding problem domain is being mentioned under related work itself. The aim and the focus of the research were mentioned under research focus. Research gap wasn't mentioned.

Article 3: In this paper author mentioned various problems in the article but the main focus was on the applicability and acceptance of continuous delivery depending on its adaptability. They have aimed to find a better release management workflow, i.e. research methods were conducted using rugby's release management workflow, where its applications in industrial projects were verified. Info regarding research gap wasn't mentioned throughout the article.

Comparison of three articles based on problem area: The problem was well described in article1, whereas several problems were mentioned in article 2 and article 3. The objectives and goals regarding problem area were well explained in article 2, this wasn't found in article 3. In article 1 gaps in research were mentioned, but in article 2 and 3, it was not mentioned precisely. Related work was well written in article 2. Related work was specified in the results of article 1, but it was not specified in article 3.

RESEARCH QUESTION:

Article 1: Research questions were clearly stated in the article. They were formulated from the research on customer involvement in CD. The aim for the questions was clearly stated in the SLR. The main motive is to provide structured knowledge of research area and clarify the underlying factors related to customer input during CD.

Article 2: The research questions are being stated clearly based on the research topic. The research questions were formulated for the early satisfaction of customers and the continuous delivery of valuable software. The major objectives and goals with these research questions were stated.

Article 3: In this paper, research questions were formulated based on the goal to measure the effect of workflow and the impact of continuous integration and delivery in mobile projects and based on interests in continuous delivery and continuous integrations.

Comparison of three articles based on research questions: In article 1 aims for the research questions were mentioned, whereas in articles 2 and 3 they were just stated.

RESEARCH METHOD:

Article1: SLR was used in this paper, where guidelines established by Kitchenham and Charters were followed in this SLR The reason behind choosing SLR wasn't stated, but they had decided to use it before starting their research as the title suggests. RQ1 was examined by descriptive analysis. RQ2, RQ3, and RQ4 by thematic analysis. The need for the research enforced from "Finnish research program" that aimed to enhance Finnish ICT companies' ability to deliver in real time.

Article 2: The authors have used Interview, Literature review, Questionnaire and Author's discretionary assumption as their research methods. Reasons for the use of these methods have been stated. The process followed were explained briefly in the article. The simulation was mentioned as a subtopic in research method. Simulation is being used in order to overcome the shortcomings of empirical analysis. Validation of the project is being mentioned which is quite extensive, where structural and behavioural validation were discussed.

Article 3: The research method employed in this paper is a case study. Internally, for evaluating the workflow survey was used. Furthermore, informal interview questions were involved in the survey. Even some personal interviews were conducted using a survey. The authors clearly stated the motivation behind using these research methods.

Comparison of articles based on research methods: The authors of all the articles have clearly stated the research methods. In article1, the reason for choosing SLR wasn't stated, whereas it was briefly explained in article 2 and article 3. The motivation behind choosing the research method was stated in only article1 and article 3, but wasn't present in article 2

RESEARCH RESULTS:

Article 1: The results were stated according to the research questions systematically. The results found are in tables and graphs, which were formulated from existing data. The author derived data from existing literature. Techniques used for each result was well described. List of primary studies through which results were obtained was mentioned.

Article 2: There were no results mentioned in the article i.e. which were obtained from the various research methods. This may be due to space constraints. Conclusion and future works weren't mentioned in the article.

Article 3: The results are grouped into 5 categories according to the research questions, for each category the solutions and the project advancements are analyzed and explained in detail. Graphical representations are used for better understanding. Threats to validity and conclusions are stated based on the results. The future scope is not mentioned precisely but the next step in the process is mentioned.

Comparison of three articles based on research results: Results have been explained in the articles 1 and 3, whereas no results were found in article 2. Data collection was well presented in article 1 than article 3, article 2 mentioned only about the methodologies that can be performed. Data analysis process is described only in article 1. The results were well presented in article 1 than article 3.

REFERENCES

Article 1: The APA (American Psychological Association) reference style has been used in the article. The author referred 25 articles in total which include conference papers, workshop proceedings, journals. Articles were cited with numbering style.

Article 2: The references in this article are stated using the IEEE format. 22 articles were referred to this article which consists of conference papers, journals, magazine article and workshop article.

Article 3: A total of 31 articles are referred in the paper which includes conference papers, journals, systematic literature reviews. The articles are cited with numbering style. The references are stated in IEEE format.

Comparison of three articles in references: Article 1 follows the APA reference style whereas articles 2 and 3 follows IEEE format for referencing the relevant architecture. All three had correctly mentioned the references.

Conclusion

Out of the evaluated three articles, article 1 is well-reported article according to our perspective. Its problem area was well pointed out in the abstract itself, where its motivation was well stated. Related work was found in results. Research questions were mentioned with its aim in article 1. Research questions were clearly stated and formulated from the topic. The motivation of the article was explicitly stated. Research method opted is SLR. Results were stated precisely and to the point. Limitations and future work were also discussed. Overall, the paper is well structured. It is moderately a good paper, which had few shortcomings, where research methodology could be specified in a better way. Article 3 can also be considered as a good article but it has more shortcomings compared to article1 and article 2 is below average compared to articles 1 and 2, with results being not mentioned.

REFERENCES:

[1] S. G. Yaman, Sauvola Tanja, Riungu-Kalliosaari Leah, H. Laura, K. Pasi, O. Markku, and M. Tomi, "Customer Involvement in Continuous Deployment: A Systematic Literature Review," 2016, pp. 249–265.

[2] O. Akerele, M. Ramachandran, and M. Dixon, "System Dynamics Modeling of Agile Continuous Delivery Process," in *Agile Conference (AGILE), 2013*, 2013, pp. 60–63.

[3] S. Klepper, S. Krusche, S. Peters, B. Bruegge, and L. Alperowitz, "Introducing Continuous Delivery of Mobile Apps in a Corporate Environment: A Case Study," in *2015 IEEE/ACM 2nd International Workshop on Rapid Continuous Software Engineering (RCoSE)*, 2015, pp. 5–11.

APPENDICES

Appendix A: Systematic Literature Review (RM paper 1)
Appendix B: Systematic Literature Review (RM paper 2)
Appendix C: Comparison of Articles (ATC)
NOTE: Two papers are added as the team members of ATC were in different group for RM course.

BIG DATA ANALYTICS AND PRIVACY ISSUSES

Systematic Literature Review

Nidhi Veereddy
940929-7422
nive16@student.bth.se

Sravika Kothawar
950129-4582
srko16@student.bth.se

I. GROUP MEMBERS PARTICIPATION

The group members participated in idea creation and in report writing with the following amount of involvement.

Group members	Idea Creation	Report Writing
Nidhi Veereddy	50%	50%
Sravika Kothawar	50%	50%

***Abstract*—Big data is a term that refers to massive datasets having complex structures. It is difficult to store, analyse and process such large volumes of data. Linking of personal data of individuals poses a severe threat to the privacy of individuals. Through this systematic literature review the main characteristics of big data and the privacy issues related to big data are addressed. Big data is the current topic in the technological world and has privacy problems. Hence research work is being conducted for the easier and faster analyses of big data and to solve the privacy issues**

Keywords: Big data analytics, characteristics, privacy issues, systematic literature review.

II. INTRODUCTION

Context:

Due to the evolution of technologies such as computers, satellites etc., there is a considerable increase in the information age. Technology has evolved in such a rapid pace that information generated is increasing in the current age. Initially with the development of computers the means for the mass storage of such data began. As the volume of data increased rapidly due to the increased usage of internet globally, it has become difficult to store such data. This lead to the concept of big data. Linking of an individual's data leads to the factor of privacy in the big data analytics. It becomes difficult to process big data using on hand database management tools[7]. Thus these properties and challenges have become the basics for conducting the review on big data. This SLR is conducted following Kitchenham and Chartes guidelines.[8]

Objectives:

- The main aim of the research is to give a review of the big data and its evolution**.**
- Understand the importance of big data analytics in today's world.
- Know what are the privacy issues related to an individual in today's big data world
- Motivation in doing this SLR is to improve the future scope of research on the privacy issues for the improvement if privacy in big data.

Methods:

In order to do this SLR, IEEE Explore is chosen as the database to search for the related search papers. The main focus of search was on the privacy issues and the characteristics of big data. The data extraction details are mentioned in the table.

Results:

The main results of the SLR are shown in the table. The papers are selected based on inclusion/exclusion criteria and quality assessment factors and then analysed. The results for both research questions were formulated from these papers. The issues mentioned give a complete picture of the privacy issues and made a way to future research.

The rest of the document is as follows.

Section III: gives the research questions based on which the review is conducted.

Section IV: gives the research methodology applied on the review

Section V: gives the rationale behind the inclusion and exclusion criteria of the articles used in the review

Section VI: states the final results of the research.

Section VII: contains discussions and views on the research.

Section VIII: gives the limitations of the research.

Section XI: gives the conclusions of the research.

Section X: the references used to conduct the review.

III. REVIEW QUESTIONS

The following questions are formulated in order to conduct our Systematic Literature Review.

RQ.1) What are the characteristics of Big data?

RQ1 gives us details about the review context. Evaluation of this question gives us the knowledge of what exactly big data is and on what basis is it analysed.

RQ.2) What are the privacy issues in Big Data Analytics that are addressed in the present literature review?

While conducting a pilot study on RQ1, the research question 2 was framed. The review on big data analysis made the authors come to a frame a search string which enlightened the privacy issues related to big data analytics. The evaluation of RQ2 gives us the details of various privacy issues related to big data.

IV. REVIEW METHODOLOGY

Kitchenham's guidelines are followed while conducting the SLR [8].

A. Pilot Study

We began our study by picking some random articles on Big Data Analytics and analysed them thoroughly which helped us in formulating RQ1. The RQ1 gives us the details of characteristics in big data. To answer RQ1 we developed a search string and filtered articles related to characteristics of big data. We went through many articles about privacy issues of big data. Hence we formulated RQ2 and refined our search string to arrive at a solution for RQ2. Evaluating RQ2 will give details about the current privacy issues of big data which help in the literature review and further research.

B. Search Strategy

The crucial step in an SLR is formulating a search strategy. The first step in search strategy is to select a good database. Most of the databases relate to IEEE. Hence, we chose IEEE Explore as our data source. The Figure 1 gives details of the search strategy implemented in the review.

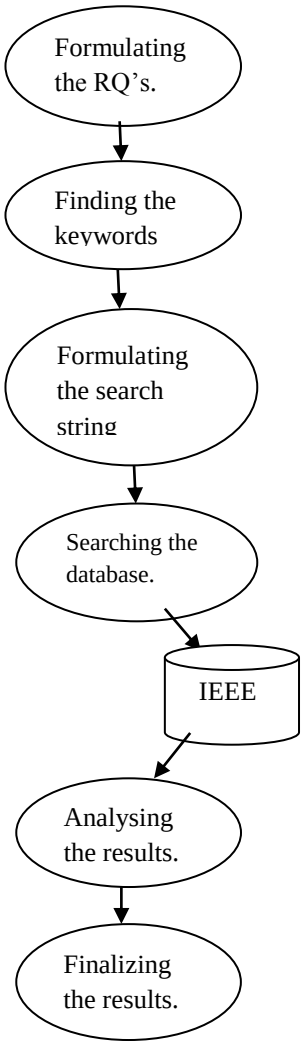


Figure 1: Search Process

C. Refining Research Questions and Search Strings

The initial keywords extracted from the research questions are

"Big data analytics, Privacy Issues"

To collect articles from the database the following search string (search string1).

(("big data analytics") AND ("privacy issues"))

For the above search string a total of 96 results were obtained.

Database	Results
IEEE	96

Table-1: Results for search string-1.

To get results relevant to RQ2 the search string was modified (search string2)

(("big data analytics") AND ("privacy issues") AND ("methods"))

For the above search string 66 results were obtained.

Database	Results
IEEE	66

Table-2: Results for search string-2.

D. Include/Exclude Criteria

The following inclusion and exclusion criteria are used to filter the articles for SLR.

Inclusion Criteria:

- Articles in which the keywords are present in the title are selected.
- Articles that answer both the RQ's are selected.
- Only full text articles are taken.
- Conference and journal articles were prioritized over the others.
- Articles that related to computer science field are only selected amongst the others.
- Articles from the year 2008 are taken.

Exclusion Criteria:

- Articles which take different scenarios into consideration are excluded.
- Articles which emphasize the privacy of data are excluded.
- Duplicates are eliminated.
- Articles that are irrelevant to the research questions are excluded.

E. Quality Assessment Criteria

QC.1) Are the characteristics clearly stated?
QC.2) Are the review questions answered?
QC.3) Are the privacy issues related to the current technology?
QC.4) Does the conclusion of the report support the results?
QC.5) How is the data collection carried out?
QC.6) How is the quality of the document?

After application of inclusion and exclusion criteria to the search results a total of 8 articles were filtered. The quality a criterion of the selected articles is based on the six quality criteria's mentioned and are shown in the Table-3. The quality measure for the first four quality factors is given by Yes/No. the quality measure for the QC5 and QC6 are measured on a scale of G, P and F where G-Good, P-Partial and F-Fail.

Articles	QC 1	QC 2	QC 3	QC 4	QC 5	QC 6
[1]	Yes	Yes	Yes	Yes	G	G
[2]	Yes	Yes	Yes	Yes	G	G
[3]	Yes	Yes	Yes	Yes	G	P
[4]	Yes	Yes	Yes	Yes	G	G
[5]	Yes	Yes	Yes	Yes	G	G
[6]	Yes	Yes	Yes	Yes	G	G
[7]	Yes	Yes	Yes	Yes	G	G
[8]	Yes	Yes	Yes	Yes	G	G

Table-3: Quality Assessment on selected articles

F. Data Extraction Process

A data extraction formed was devised in order to fill the details of the data extracted from the papers. An excel sheet was created as a form inorder to make the analysis of results easier in future.

DE.1) What are the objectives of the article?
DE.2) What are the characteristics the article A stated?
DE.3) Are the privacy issues stated clearly?
DE.4) What kind of study does the paper use?

The Table-4 gives us details of the data extracted from the selected articles using the data extraction objectives mentioned above.

G. Validation of Protocol

Validation of protocol was ensured by asking a few fellow students to make similar search on their systems. Repeating this step several times ensured us that the results were correct and hence we proceeded in the review.

V. INCLUSION AND EXCLUSION STUDIES

Figure 2 depicts the inclusion and exclusion criteria applied at different stages of the research process and the reason for excluding the articles

	What are the objectives of the article?	What are the characteristics the article states?	Are the privacy issues stated?	What kind of study that the paper use?
Ref[1]	To state practices, privacy issues, architecture.	Volume, velocity, variety.	Yes	Experimental study.
Ref[2]	To state characteristics, security and privacy issues.	Volume, velocity, variety.	Yes	Experimental study.
Ref[3]	To state challenges, properties, tools and good practices.	Velocity, volume, variety, variability, complexity, value	Yes	Theoretical study.
Ref[4]	Analysis of data issues and methods.	Variety, velocity and volume	Yes	Experimental study.
Ref[5]	To state the Challenges, privacy issues, techniques	Volume, velocity, variety.	Yes	Theoretical study
Ref[6]	Review of big data	Volume, velocity, variety, value.	Yes	Experimental study.
Ref[7]	To state the characteristics, security issues and privacy issues	Volume, velocity, variety.	Yes	Experimental study.
Ref[8]	To state the challenges and privacy issues of big data.		yes	Experimental study.

Table-4: Data extraction form

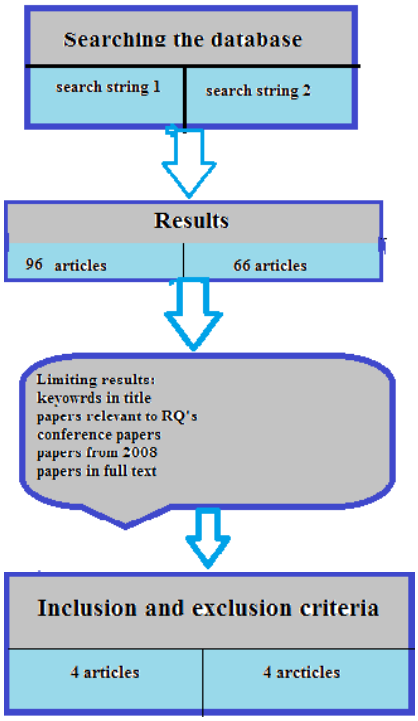


Figure-2: Inclusion and Exclusion results

VI. RESULTS

R.Q.1) What are the characteristics of big data analytics?

Big data refers to large datasets which cannot be managed by normal softwares and hence handling, processing and analysing of big data is the trending factor in today's technological world. The analysis of big data is a step forward to the traditional analysis. Big data is analysed by the concept of 3v's and they are Velocity, Volume and Variety.

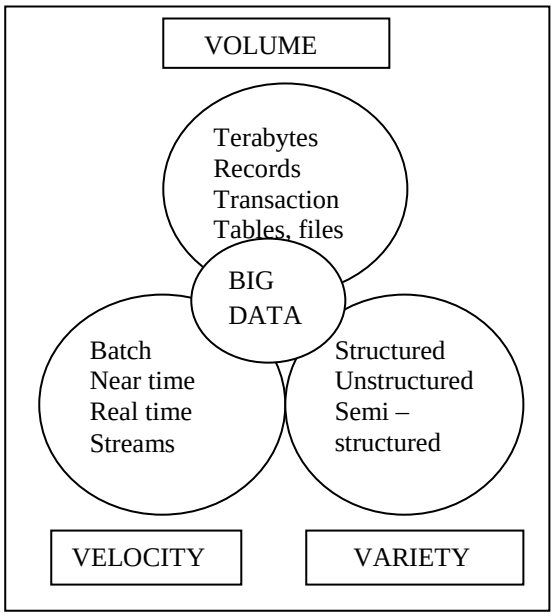


Figure-3: The 3v's of Big Data

Velocity:

Velocity is used to define the speed of any object in the universe. The velocity of big data refers to the speed at which data is coming from different sources. It is not only limited to the incoming data but also refers to the flow of data. [3][4]

Volume:

The term big data refers to the size of the data. Volume of the data is the property which describes the size of the data. The volume of the data in the technological universe will increase from petabytes to zettabyte in the near future.[1][3][4]

Variety:

Variety of anything refers to the types available in it. Variety indicates the heterogeneous data types and sources.[2] These are generally of three types. Structured, Unstructured and Semi-structured data. Structured data is the sorted data and is easily inserted into the available data warehouses. Unstructured data is the raw form of randomly generated. Semi structured data is the data that contains links or tags to the fixed field in the data warehouses but it's not confirmed and hence it is difficult to separate the data. [1][4]

In addition to these characteristics, a new characteristic was developed in the recent times and is called the fourth v of big data and it's the 'value' of the big data.[1]

Value:

Value refers to the data that has a great social value. Value is an important characteristic for a user, as he can run queries in the databases and deduct the required results based on the data value given. The results are filtered based on the queries and the values are given based on the dimensions. This property helps in finding the business trends which may change their strategies.

R.Q.2) What are the privacy issues of big data analytics that are addressed in the present literature?

From our study we found out few privacy issues of big data analytics and reviewed them in the following research question.
The privacy issues are categorized into four parts namely

- Interaction with individuals
    a) Providing transparency.
    b) Getting consent.
    c) Revocation of consent and deletion of personal data.
- Re-identification attacks
    a) Correlation attack
    b) Arbitrary identification attacks.
    c) Targeted identification attacks.
- Provable and probable results.
- Economic factors.
    a) Confusion and Distraction.
    b) Context faults.

Interaction with individuals:

a) Providing Transparency:

Involvement of individuals is one of the most challenging privacy issue in big data analysis. Collecting information from every individual is a necessary process. It must be ensured that the information collected is to be protected and every individual should be aware of its processing. Taking network traffic as an example of big data sources, it is necessary to analyse the huge lists of IP addresses along with the data like url, cookies etc. and information that might be or might not be relevant to the individual. According to European law the collection of personal data of an individual must include the detailed documentation of the process by which this data is processed. Hence it is necessary to provide the details of the all the algorithms and process that are involved in big data analytics. Providing transparency towards individuals with respect to the type of processing used and the data present in it is the most challenging part.

b) Getting Consent:

Consent generally refers to permission. Getting consent refers to getting permission from the individual to process their data. Beyond sheer transparency, any individual may have the right to refuse that their data is being processed by a certain data processor. Thus the privacy laws mention that every individual is to be asked prior to the processing of their data. This process has certain complications. Most of the rules/laws not only require consent but also informed consent which means, the individual must be able to understand what sort of processing is performed.

c) Revocation of Consent and Deletion of Personal Data:

At any stage in the processing of data an individual can decide to revoke its consent. The individual may feel that private data is being leaked or is no longer secure and hence revoke the consent. Once revoked data of individual should be stopped from processing and should be deleted.

Re-identification attacks:

Another major threat to privacy is re-identification attacks which say that from a large dataset, explicit scanning is done for correlation that leads to a unique identity of an individual with the linking of definite data sets together. It has three sub categories.

a) Correlation attacks:

This is mainly about linking of uniform data values in a dataset to other sources to create unique entries. A correlated database has more data, allowing more precise analysis. It consists of

linking datasets up to a point that there is at least one entry in the dataset that is unique.

b) Arbitrary identification attacks:

This type of re-identification mainly focuses on linking an entry in the dataset to an identity of an individual. The attack will be complete once the link between the individual and the correlated dataset is drawn with a sufficient level of priority. As a result lead to the failure of anonymization. The key issue is that all the individuals of the real world are not feasible to link.

Probable vs. Provable results:

Validation of all the gathered results is the key threat to big data analytics and it mainly depends on the query. In a dataset containing correlated email addresses, it is assumed that two entries will have the same email address before correlation will result in a single database entry and will reflect all the information on the same individual. In some cases an individual might spoof the email address and this leads to false correlation and in some cases same email addresses are shown by two people. Complications arise in situations where the linkage is not reliable for e.g., correlation of datasets by IP address. Depending on the data distribution, the distance between the events is established. It is easy to use the same IP address by different individuals at the same time. This is called Probability based linkage and false correlations are high. The risk factor is high in probable links compared to provable links.

Economic factors:

Economic factors also effect the privacy issues of big data. Most of the big data analytics require huge datasets in advance and need to be exchanged between partners. This is based on mutual agreement or by means of economic data market where the data providers sell the data to customers.

a) Confusion and Distraction:

The main idea of this approach is to make a sold dataset less useful to a customer by altering it. For example a company X will sell their data to Y, but will not reveal the number of users using it and will modify it by deleting the common users both have.

b) Context faults:

In this approach the company X sells its data to company Y, but does not give all the details. The threat is that Y interprets the data coming from X. X uses a self-calculated field activity index in percentile(%) and this value makes it difficult for Y to extract information as the index varies from X. this results in false results.[8][5]

VII. Discussions

Through this systematic literature review, the principal findings are that big data is the massive amount of data having complex structures. It is difficult to analyse store and process big data. The main characteristics of big data are the 3v's which are referred from the 7 selected articles. These characteristics mainly give us a clear view on what big data is. This 3v's are velocity, volume and variety. The speed of incoming data is the velocity of big data. The size of the data available is referred to as the volume and variety is the type of data available namely structured, unstructured and semi-structured. When the data of the individuals is linked to the datasets, there is a threat to the privacy of data. The report gives us the details of the privacy issues of big data analytics. The study helped us to define and understand the future scope of research in the field of privacy issues of big data.

VIII. Limitations

The limitations that may arise to our systematic literature review.

- Due to the lack of time we might have missed analysing some articles that are related to our study
- As the topic of big data analytics is very vast, it has become difficult to make our search more precise. Hence the keywords are restructured in order to make search more precise.
- As the big data is trending topic many journals, papers, academic publications magazines reported about it. One research was restricted to academic publications and hence some relevant matter might have missed.
- Research on big data is still continuing and that makes our review limited only to certain privacy issues reported till day.

IX. Conclusions

In order to make a conclusion the SLR conducted brings the facts into consideration that the research in the field of big data analytics is still in the beginning stage. Due to the complexity of data in the real world and with the increasing privacy issues of individual's data due to social networking sites, a solution to the privacy issues has not been proposed as yet. This provides a basis for further research in the area in order to find out the solutions for the privacy issues of big data analytics.

X. REFRENCES

[1] N. W. Grady, M. Underwood, A. Roy and W. L. Chang, "Big Data: Challenges, practices and technologies: NIST Big Data Public Working Group workshop at IEEE Big Data 2014," Big Data (Big Data), 2014 IEEE International Conference on, Washington, DC, 2014, pp. 11-15.
doi: 10.1109/BigData.2014.7004470.

[2] D. S. Terzi, R. Terzi and S. Sagiroglu, "A survey on security and privacy issues in big data," 2015 10th International Conference for Internet Technology and Secured Transactions (ICITST), London, United Kingdom, 2015, pp. 202-207.
doi: 10.1109/ICITST.2015.7412089.

[3] A. Katal, M. Wazid and R. H. Goudar, "Big data: Issues, challenges, tools and Good practices," Contemporary Computing (IC3), 2013 Sixth International Conference on, Noida, 2013, pp. 404-409.
doi: 10.1109/IC3.2013.6612229.

[4] S. Sagiroglu and D. Sinanc, "Big data: A review," Collaboration Technologies and Systems (CTS), 2013 International Conference on, San Diego, CA, 2013, pp. 42-47.
doi: 10.1109/CTS.2013.6567202.

[5] Rongxing Lu, Hui Zhu, Ximeng Liu, J. K. Liu and Jun Shao, "Toward efficient and privacy-preserving computing in big data era," in IEEE Network, vol. 28, no. 4, pp. 46-50, July-August 2014.
doi: 10.1109/MNET.2014.6863131.

[6] B. Matturdi, Zhou Xianwei, Li Shuai and Lin Fuhong, "Big Data security and privacy: A review," in China Communications, vol. 11, no. 14, pp. 135-145, Supplement 2014.
doi: 10.1109/CC.2014.7085614.

[7] A. Katal, M. Wazid and R. H. Goudar, "Big data: Issues, challenges, tools and Good practices," Contemporary Computing (IC3), 2013 Sixth International Conference on, Noida, 2013, pp. 404-409.
doi: 10.1109/IC3.2013.6612229.

[8] B. Kitchenham and S. Charters, Guidelines for performing .Systematic Literature Reviews in Software Engineering. 2007.

Analysis of Automated tools for Requirements Engineering process
Systematic Literature Review

Neeraj Reddy Avutu
9411053375
neav16@student.bth.se

Vinay Kumar Vennu
9408042910
vive16@student.bth.se

I. GROUP MEMBERS PARTICIPATION

The group members participated in idea creation and in report writing with the following amount of involvement.

Group Member	Idea Creation	Report Writing
Neeraj	50%	50%
Vinay Kumar	50%	50%

***Abstract*—Requirements Engineering plays a crucial role in developing a software and satisfying customer requirements. Automation of RE enhances software development and simultaneously reduces effort of the analysts. There are few problems that are addressed in the present literature regarding implementation of automated SRS document. These problems are though challenging, but provide an insight of some of the features that can be implemented in the present automated tools. This paper mainly focuses on these features that can be incorporated in the present tools so that the RE process does not compromise on customer satisfaction.**

Keywords: Automation, Requirements Engineering, Natural Language, Software Requirements Specification.

II. INTRODUCTION

Besides being a challenging task, building a software involves lot of creativity. Keeping in mind the myths of software developers
regarding understandability of requirements, there is a need for filling the gap between construction and design of a software [1].
Requirements engineering, a set of tasks that increases understandability of requirements, fills this gap [1]. Software Requirements
Specification(SRS) is one such process of RE that optimizes the problems of scope, volatility and understandability of requirements [1].
An SRS document contains overall product description, functional and non-functional requirements and system features [1] and it acts as an "agreement between client and analyst"[2]. Even though it is possible to generate SRS document through proper templates, an analyst conducts
questionnaires and holds meetings for elicitation of the requirements and ends up building a document with lot of prolixity and uncertainty [2]. Automation
of every software aspect has been a trending topic these days. The CASE tools do not automate the specification and analysis of requirements. Though there are tools to generate code from models like Rational Rose [3] and Magic Draw [4], they do not support automation of requirements specification and the analysts end up
using a simple text editor like Word, consuming time [2]. NALASS (Natural Language Syntax and Semantics) [5] is one such tool that automates the RE process [2].
Considering a definite number of components in the "discovery, analysis and specification of requirements" is determined [2]. NALASS uses the language semantics for requirements analysis and syntax for requirements specification and later build questions that are answered during discovery of requirements [5]. Based on the answers to these questions, SRS document is generated automatically by this tool. The methodology that NALASS uses is NLSSRE (Natural Language syntax and semantics requirements engineering) [5][2]. This paper besides analyzing the present automated tools for RE, also highlights certain problems faced in implementing them.

The remaining of the paper is organized as follows: section 3 addresses the review questions while section 4 gives details about the review methodology.
The inclusion and exclusion studies were discussed in section 5. The results were illustrated in section 6. Section 7 gives the discussions part, whereas
section 8 provides the limitations of the research. Finally, we conclude with conclusion.

III. REVIEW QUESTIONS

The following review questions were formulated in order to conduct our SLR:

RQ.1) What are the challenges faced in automating the requirements engineering process?
This question was framed in such a way that it keeps us motivating to conduct our literature review. Besides providing context for our review, this question also
encourages us to find the challenges in present environment. These review results will state various issues related to our research area. It provides us with sources
that are enough to gain basic knowledge in the research field.

RQ.2) What are the features that are lacking in the present automated RE tools?

After conducting a pilot study based on RQ1, the search was refined. In this process RQ2 was framed. On examining RQ2, which means refining our search more thoroughly which will finally lead us to the features or ideas that are present in the literature which can improve the implementation of the automated tools for
requirements engineering process.

IV. REVIEW METHODOLOGY

Based on the guidelines provided by Kitchenham we followed a protocol in order to conduct our SLR [6]. It defines the pilot study, data sources and search strategy and the approach to study selection, study quality assessment, data extractions, and data synthesis [6].

PILOT STUDY

Due to immense interest in the topic of software and requirements engineering, we picked some random articles from the database
and analyzed them. We found topics related to Automation more interesting and trending. Later we had formulated RQ1 to address the
challenges faced in automating the requirements engineering process in the present literature. While answering to RQ1, we had gone through
many articles in which very few of them were SLRs. Most of the papers addressed the present automated tools for RE but very few of them elaborated the problems faced in implementing them. Hence we framed RQ2 in this article, in order to present the features that can be included in the present automated tools so that this article can be
used future studies or future researches.

REFINING RESEARCH QUESTION AND SEARCH STRINGS

The initial keywords that were extracted from the research questions are:
"Automated Requirements Engineering, Software Requirements Specification"
The following search string (search string 1) was used in our study to get basic knowledge about the topic and to get SLR's in the literature.
(((Automated requirements engineering) OR RE) AND ((systematic literature review) OR (SLR) OR (literature review)))
From the above search string a total of 4 articles were retrieved. To get relevant studies the following search string was used (search string 2).
(((Automated requirements engineering) OR RE) AND (SRS AND (challenges* OR constraint* OR problem*))).
From the above search string a total of 7 articles were retrieved.

SEARCH STRATEGY

The most important step in our search strategy was selection of a good source. We had selected Scopus as our key database because it is known for its large number of abstracts and citations. The following figure depicts our search strategy.

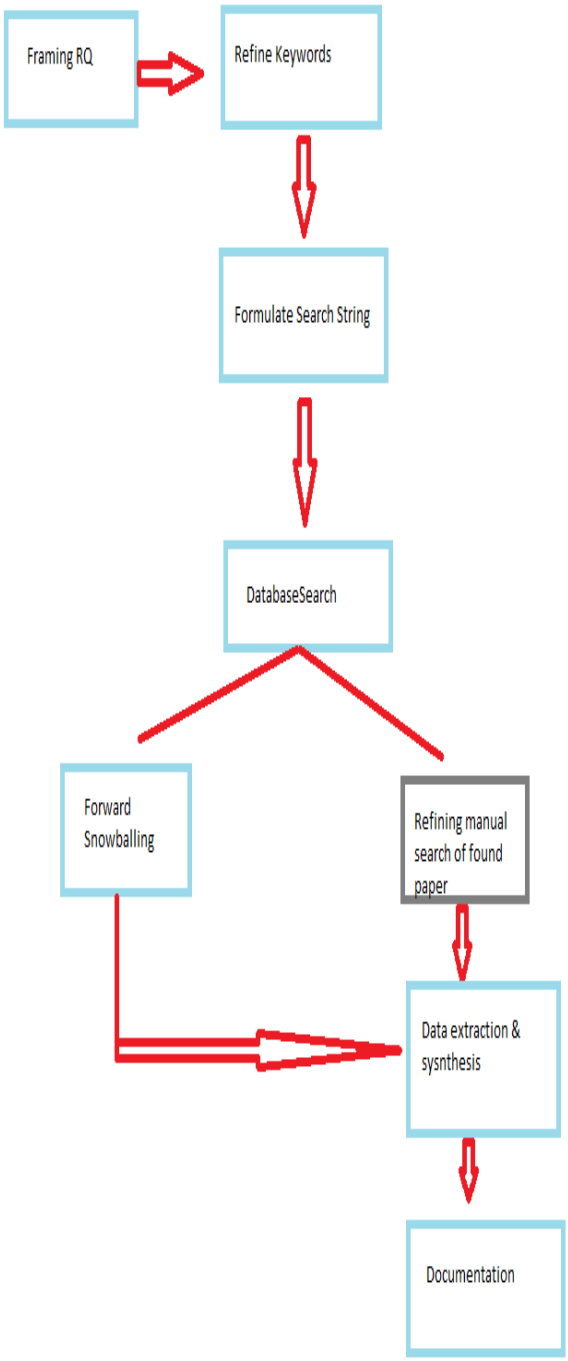


Figure1

INCLUDE/ EXCLUDE CRITERIA

The following inclusion and exclusion criteria was used in order to select the relevant studies for the SLR.

Inclusion Criteria:

1. Articles in which the selected keywords were present in only title and abstract were

included (i.e., the search in database was done only on the title and the abstract).
2. Articles which answers both the RQs.
3. Articles which are available in full text.
4. Articles written in English language.
5. Only Conference and Journal articles were selected.
6. Articles which were written after 2010.
7. Articles which belongs to only Computer Science and Engineering.

Exclusion Criteria:

1. Articles that do not focus on automation.
2. Articles with main emphasis only on the requirements engineering.
3. Duplicate studies were excluded.
4. Studies which are irrelevant to the RQs were excluded.
5. Studies which are relevant but stressed more on automated test cases, and were also excluded.

QUALITY ASSESSMENT CRITERIA

The quality of our chosen articles were assessed by the following Quality Criteria (QC), which was based on the guidelines provided by Kitchenham [6].
It also suggests that quality assessment criteria can also be used for inclusion/exclusion studies.
Quality Criteria for Quantitative Studies:
QC.1) Are the objectives clearly stated?
QC.2) Does the research process allow the questions to be answered?
QC.3) Is the purpose of the analysis clear?

Quality Criteria for Qualitative Studies:
QC.4) How well was data collection carried out?
QC.5) How clear are the links between data, interpretation and conclusions?
QC.6) How adequately has the research process been documented?
The following table measures the quality of the selected articles.

Articles	QC1	QC2	QC3	QC4	QC5	QC6
[1]	Yes	No	Yes	P	G	P
[2]	Yes	Yes	Yes	G	P	G
[3]	Yes	Yes	Yes	G	G	P
[4]	Yes	No	No	P	G	G
[5]	Yes	Yes	Yes	G	G	P
[6]	Yes	Yes	Yes	G	G	G
[7]	Yes	Yes	Yes	G	G	G
[8]	No	No	Yes	P	P	G

Table1: Quality Assessment on selected articles

Table1: Quality assessment on selected articles

Two sample data extraction forms were illustrated in Table 2 and Table 3.

G. Validation of Protocol

The protocol was validated by asking a couple of fellow students to retest the search results obtained by following all the steps that were mentioned above. After retesting a couple of times, all the search results obtained by the two students were.

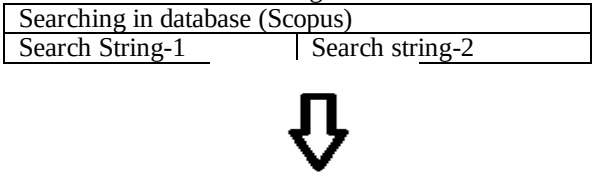
F.DATA EXTRACTION PROCESS
A data extraction form is designed which contains all the necessary information required to address the research questions. During this process one actor acts as
data extractor and the other as data checker. The questions in the data extraction form are answered by the data extractors in Microsoft Excel sheets. Later the checker examines the answers and the two sheets are compared and all the disagreements are solved between the two authors.
DE.1) What are the main objectives of the article?
DE.2) What Research Methodology does the article use?
DE.3) Does the article address both technical and non-technical issues?
DE.4) What are the major automation problems that the article talks about?
DE.5) What were the additional features that are discussed in the article?
DE.6) What extent these features are validated?

Questions	[2]	[5]	[7]
DE.1	Generating automated SRS	NLSAS automated tool	Analyzing challenges
DE.2	Systematic Literature Review	Theoretical analysis	Literature Survey
DE.3	Yes	No	Yes
DE.4	Time constraint	Lack of automation supporting case tools	-
DE.5	Embedding of DFDs	Use Case Description	Scenarios, Diagram and Documentation
DE.6	As this is an SLR, it just discusses automated RE	Yes	No

Table2: Data Extraction form

V. INCLUSION AND EXCLUSION STUDIES

Figure 2 depicts the inclusion and exclusion criteria applied at different stages of the research process and the reason for excluding the articles



Search Results	
3,338	14

Limiting to 1) English 2) 2010-2016 3) Computer Science and Engineering 4) Document type: Conference Paper & articles. 5) Source Type: Conference Proceedings & journals	
1,156	9

Inclusion of articles after studying the abstracts and after applying all the remaining inclusion/exclusion criteria & also quality criteria.	
7	2

Figure 2 Article selection based on inclusion/exclusion criteria

VI. RESULTS

At first the results obtained from the metadata analysis and quality assessment are summarized,
later we discuss the principle findings and the answers to those predefined research questions
After the iterations compiled under the search strings, we finally selected 8 articles in which [2] was defining our review protocol and [7],
[8] are the review articles which were related to our research area.
The remaining 5 articles are journals and conference proceedings.
In table 4 we had illustrated the number of articles that we had taken in a particular year (excluding Kitchenham).

RQ.1) What are the challenges faced in automating the requirements engineering process?
The articles [2] [6] addressed most of the answers to the research question 1.
1.The focus of existing tools lies on diagrams, charts and pictures [2].
2. Less importance is given on the "textual specification of requirements"[2].
3. The present CASE tools do not support automation beside developing software.
4. Time constraints regarding specification of requirements.
5. Lack of trustworthy developers and analysts.
6. Not able to meet increased levels of customer satisfaction.
All these challenges motivate us to automate the generation of SRS document and enhance the requirements engineering process.

RQ.2) What are the features that are lacking in the present automated RE tools?
The answers to these questions resulted due to in depth study and analyzing the sources supporting RQ1.
Certain Features like "generation of use case descriptions, scenarios and diagrams" can be implemented in the present tools [2].
"Embedding of DFDs and Class
Diagrams to the right section of the SRS document and assuring full consistency of the tool and the
Documentation component to the methodology" are certain important aspects that can be considered [2].

VII DISCUSSIONS

Through this systematic literature review it can be found out that in the present scenario SRS document mainly faces as a main challenging issue with the automation of requirements engineering which is drawing many organizations a step back. Usually the engineers of an organization take the requirements as the initiative for automating the SRS document issues.
Most of the researchers in the available papers or selected papers were interested in the challenges faced in automation and various features involved in it such as generation of test cases [2].

VIII LIMITATIONS

The following are the limitations of our systematic review.
1.Though certain articles were related to our research area, we have not gone through them due to time constraints.
2.As the field of requirements engineering is very big, and a lot of research was already done in this field. So we narrowed the topic
and restricted our study to only a limited number of keywords.
3.Since we choose Scopus as our key database, we might have not covered some topics and articles present in other databases.
4.The search was performed only on title, abstract and keywords but not on the full text of the articles.
5.Though we addressed the challenges in automation of RE, we found limited papers regarding the lacking features in present tools. So we
could not elaborate all the features.

IX CONCLUSION

The need for automation in every aspect of software is the increasing day by day. Due to the problems faced in generating Natural language for requirements specification, automation of Requirements engineering has been the main topic discussed in this literature. Though the present tools have supported automation of SRS document utilizing verbs, nouns, adjectives, etc. but have not been consistent. The challenges discussed in the literature were mainly related to automation of SRS document.

X REFERENCES

[1] R. S. Pressman and B. R. Maxim, *Software engineering : a practitioner's approach*. New York: McGraw-Hill, 2015.

[2] M. G. Georgiades and A. S. Andreou, "Automatic generation of a Software Requirements Specification (SRS) document," Intelligent Systems Design and Applications (ISDA), 2010 10th International Conference on, Cairo, 2010, pp. 1095-1100.

[3] IBM Rational Rose. Available online as of July 12, 2010 at http://www-306.ibm.com/software/rational/

[4] MagicDraw. Available online as of July 12, 2010 at http://www.magicdraw.com/

[5] M. G. Georgiades, A. S. Andreou and C. S. Pattichis, "A requirements engineering methodology based on natural language syntax and semantics," *Requirements Engineering, 2005. Proceedings. 13th IEEE International Conference on*, 2005, pp. 473-474.

[6] B. A. Kitchenham, "Guidelines for performing Systematic Literature Reviews in Software Engineering," Version 2.3, EBSE Technical Report, EBSE-2007-01, Software Engineering Group, School of Computer Science and Mathematics, Keele Univesity, Keele, Staffs,ST5 5BG, UK and Department of Computer Science, University of Durham, UK, 2007.

[7] M. Bano, D. Zowghi and N. Ikram, "Systematic reviews in requirements engineering: A tertiary study," *Empirical Requirements Engineering (EmpiRE), 2014 IEEE Fourth International Workshop on*, Karlskrona, 2014, pp. 9-16.

Comparison of Articles

Topics and Articles

- Topic: Continuous Delivery.
- Database: Scopus.
- Articles:

1. Customer Involvement in Continuous Deployment: A Systematic Literature Review.
2. System Dynamics Modelling of Agile Continuous Delivery Process
3. Introducing Continuous Delivery of Mobile Apps in a Corporate Environment: A Case Study.

Sravika Kothawar-9501294582
Avutu Neeraj Reddy-9411053375

Comparison of Problem Domain

Criteria	Article 1	Article 2	Article 3
Problem Area	Common over view on Customer involvement in Continuous deployment	configuration management problem, insufficient collaboration between development time the deployment operations	Main focus was on applicability and acceptance of continuous delivery depending on its adaptability
Motivation	Is well stated in the background	Not clearly mentioned	Not clearly mentioned
Related Work	Were not specifically mentioned, was found in results	Is mentioned under related work section	Implicitly stated in conclusion section

Comparison of Research Questions

Criteria	Article 1	Article 2	Article 3
How are they formulated?	Formulated from the research topic	Based on the topic	Based on the goal and interest on the topic
Are the research questions stated clearly?	Are clearly stated	Are clearly stated	Are clearly stated

Comparison of Research Method

Criteria	Article 1	Article 2	Article 3
Research method opted	Systematic Literature Review	Different methods are proposed	A case study using survey
Motivation	Stated clearly	Not stated	Stated clearly

Comparison of Research Results

Criteria	Article 1	Article 2	Article 3
Are all the research questions answered?	Results were precisely stated in the results and discussion section	Results are not mentioned in this article	Results are categorized and clearly discussed
Data collection	Well presented in data collection section	Not mentioned	Discussed in the findings section
How are the results discussed?	Using graphical representation and Tabular form	Not discussed	Using Graphical representations

Comparison of References

Criteria	Article 1	Article 2	Article 3
Reference Style	APA(American Psychological Association)	IEEE format	IEEE format
Citation style	Numbering	Numbering	Numbering

Conclusion

- Article 1 is well reported out of three.
- The problem area, motivation for research, research questions, methods used, results are all precisely stated
- Related work, limitations, future work were all discussed.
- The overall paper was well structured. It is moderately a good paper, which had few shortcomings. The research methodology could have been specified in a better way.
- Article 3 is also considered as a good article with more shortcomings compared to article 1.
- Article 2 is below average compared to article 1, the results were not mentioned in this article.