ISSN No:-2456-2165

Information Technology: Robust Data Gathering and Analysis Techniques

Nyamwamu Roseline Wangui
Department of Information Technology,
Kibabii University, Kenya

Dr Peters Anselemo
Department of Information Technology,
Kibabii University, Kenya

Abstract:- Data collection constitutes a methodological process of collecting, collating, and analyzing information to present solutions that are relevant to research questions. There are a myriad of data collection methodologies (Schick-Makaroff et al., 2016). The data collection aspect of research is mutual to all fields of research including Information Technology (IT). While the data collection methods vary, they are founded on some basic data collection tools and focus on capturing accurate data that remains the same (Kabir, 2016). The objective of all data collection/requirements determinations is to get access to quality evidence that leads to quality data analysis and permits the development of credible and convincing answer to research questions or solutions (Goertzen, 2017). Regardless of the area or preference of study, either quantitative or qualitative, accurate requirements, accurate requirements gathering methods and analysis techniques are critical to ensuring the integrity of research. In this paper, robust data collection methods and analysis techniques in Information Technology are unveiled.

Keywords:- *Requirements, Data, Analysis, Information Technology*.

I. INTRODUCTION

With massive data available than ever before, well chosen, well implemented methods for requirement gathering and analysis are essential for all types of research. Determining the requirements and analysis is very critical in developing systems/applications. The requirement gathering process entails the 'what' rather than the 'how'(Javaria M. 2016). Presentation of the data collection tools varies depending on the type of data being collected: primary or secondary data (Ajayi, 2017). The tools used ought to make the data collected unique until the research is published. The main tools and techniques used in data collection traditionally include questionnaires, focus group interviews, interviews, surveys, observations, diaries, case studies, memo motion studies, activity sampling techniques, link analysis, telephone interviews, census etc(Fontana & Prokos, 2016). In IT, methods such as documents reviews, brainstorming, business process reengineering (BPR) and prototyping are useful in gathering data/requirements from the users. Alternative these techniques are combined with the traditional tools depending on the deliverable that is expected. In addition to these methods, computer assisted methods of data collection are also available for use. (Stansberry, 2019).

II. DATA GATHERING TECHNIQUES IN IT

- *Document review*

Document review is a way that is used by analyst to assist them recognize how the contemporary system is supposed to work. With time documents get outdated calling for their review especially when business processes have changed. This technique includes analyzing of relevant commercial enterprise, gadget and task documentation with the goal of information the commercial enterprise, the challenges so as they can figure out requirements or opportunities for improvement. This technique may be used to accumulate records before scheduling interviews or using other elicitation mechanisms (Mallela, 2020) . document evaluation can complement different elicitation techniques like workshops, interviews and prototyping. a number of the files that can be reviewed consist of the enterprise process documentations, organizational structure files, job descriptions/activity roles files, patron feedback, various reports and so on.

- *Brainstorming*

This is another popular method of obtaining ideas and building them up into a solution in computing. Brainstorming is a facilitated group creativity technique where each member belonging to a group in engaged in finding a solution for a specific problem by gathering a list of ideas spontaneously. The small group discuss a specific problem, opportunity, or issue.

Prior to engaging in brainstorming sessions to address the primary issue, individuals with limited experience may find it advantageous to participate in practice sessions. The insights generated during brainstorming are typically organized and visually represented through a mind map, illustrating shared concepts and divergent viewpoints while fostering the generation of fresh ideas. This method promotes the emergence of novel concepts, encourages team collaboration, and empowers participants to expand upon the notions and suggestions put forth by their colleagues. Widely

utilized techniques that incorporate variations of brainstorming include Joint Application Development (JAD) and SCRUM (Mushtaq, 2016).

➢ *Business Process Reengineering (BPR)*

Business Process Reengineering (BPR) encompasses the pursuit and execution of drastic transformations in business processes to attain significant advancements in products and services. Its objective is to revamp essential business processes with the aim of enhancing product output, quality, or cost reduction. A crucial requirement for BPR is comprehending and recognizing the organization's key business processes, which are the processes that directly impact organizational outcomes and customer satisfaction.

Conducting BPR requires understanding which processes need to change, thus it requires the analysis of the organization's workflows, finding processes that are inefficient and figuring out the ways that can be used to get rid of them or changing them. These processes are what is used as input or data agents that are used to reengineer the processes to improve the organizations productivity. IT plays a pivotal role in providing value to organizations by assisting in the reengineering of business infrastructure and processes. It offers various components that enhance performance, leading to a competitive edge. Moreover, IT facilitates the redesign of business processes to align with overall corporate objectives and create value for customers (Susanto, 2020).

➢ *Prototyping*

Prototyping has been used as both a requirement gathering technique as well as a validation process. Building a prototype allows the user to view a model of his verbal description and then engages the users. A prototype allows the designer to learn how stakeholders and users feel about the product, identify areas for improvement and pave the way for a quality product devoid of errors. Thus, the feedback from users is used as input to make revisions of the product as continuous testing is carried out until an acceptable product is designed. Feedback from user is a useful resource that is used to build a product that meets their requirements (Mushtaq, 2016).

III. ANALYZING DATA (RAW DATA)

Raw data has been used to refer to information gathered for a study before it has been analyzed or transformed. The term applies to data as soon as it has been collected or after it has undergone cleaning, but it cannot be used to described data that has been analyzed or transformed. In most cases, the issue is not lack of data but proper ways analyzing the data to present meaningful information that can be used to make decisions. In short, there is a great need for proper ways of analyzing data.

➢ *Verification and Validation of data*

V & V are done on systems to ensure that the software systems/applications meet the users' needs. Verification entails ensuring that the software conforms to its specification i.e., Are you building the product right? Validation entails ensuring that the software does what the user really requires ie., Are you building the right product?

Before analysis, data collected from various sources using various techniques/methods is verified to ensure its credibility and reliability. Occasionally in IT, test data is used or created to satisfy the execution preconditions as inputs to execute one or more test cases. In order to test a software application effectively, one needs a representative data set. The test data can be inform of system test data, SQL test data, performance test data or in XML test data. Various tests can be carried out using the test data to validate the test results i.e. to establish that the software is able to meet the needs of the user. Peer reviews, scenarios, and walkthroughs are some of the techniques that are used to validate and verify requirements results in a more accurate, specification and higher customer satisfaction (Javaria M. 2016).

➢ *Types of test data*

- *Normal use data*

This is the data that is expected to be entered into the application. An input form for a username can be used as an example of normal use data. The user is expected to type in some letters and numbers without any errors being expected.

- *Borderline / Extreme data*

Borderline test data is utilized to assess the limits of acceptable data within a system. Similar to normal data, borderline data is processed in a similar manner. For instance, if a user is required to input a username consisting of 1 to 10 characters, the borderline test data would include a username with only 1 character and another with exactly 10 characters. This type of data is particularly useful for testing the software's robustness against defined boundaries and ensures that the application functions correctly when handling such data.

- *Invalid data*

Invalid test data refers to data that the program rejects due to various reasons, such as being of the wrong data type, containing characters that are not allowed, or falling outside the accepted parameters of the program. For instance, if a user is required to enter a username within the range of 1 to 10 characters, invalid test data would include a blank username, a username with 13 characters, or one that includes non-standard letters and foreign characters. It is essential for the application to handle such invalid data accurately without crashing. Typically, the system provides feedback to the user indicating the rejection of the data or that the provided data is invalid (QATestLab, 2021).

ISSN No:-2456-2165

IV. DATA ANALYSIS TECHNIQUES IN IT

Data analysis is defined as a process that involves cleaning, transforming, and modeling of data to discover useful information for business decision-making. The objective of this process is to extract useful information from data and making decision(s) based on the analysis. After preparing the requirements, the following analysis can be carried out:

- *Descriptive Analysis*

Descriptive analysis, also known as descriptive statistics, represents the initial stage of data analysis. It involves examining and summarizing data in a meaningful way to describe or present key insights and patterns. Descriptive analysis can be applied to various types of data, including customer feedback, sales figures, website traffic, or any other recorded information of past events that allows for business analysis up to the present moment. Its purpose is to provide a comprehensive overview and understanding of the data through meaningful summaries and visualizations. Descriptive analysis does not, however, permit the making of conclusions beyond the data that has been analyzed in order to conclude hypotheses that have been made in the research (Vigliarolo, 2019).

Characteristically, there are two main forms of descriptive statistics which are measures of central tendency and measures of dispersion or spread. Measures of central tendency are statistical techniques used to describe the central or typical position of a dataset. They provide insights into where the data is concentrated. Common measures of central tendency include the mean (average), mode (most frequent value), median (middle value), frequency (count of occurrences), percentage (portion of total), and range (difference between the highest and lowest values).

On the other hand, measures of dispersion summarize how data points are spread out or dispersed within a dataset. They provide information about the variability or spread of the data. Standard deviation and variance are commonly used measures of dispersion. Standard deviation quantifies the average amount by which data points deviate from the mean, while variance measures the average squared deviation from the mean.

Together, measures of central tendency and measures of dispersion provide a comprehensive understanding of the characteristics and distribution of data. They are essential tools for data analysis and interpretation.

- *Exploratory Analysis*

Exploratory Data Analysis (EDA) is an approach used to analyze and explore datasets in order to uncover hidden relationships and discover new insights. It serves as an initial step in data analysis, performed before applying formal statistical techniques. EDA complements inferential statistics, which typically follows predefined rules and formulas (Glen, 2019). In EDA, analysts aim to gain a "feel" for the dataset, relying on their judgment to identify the most crucial elements. This process involves summarizing the main characteristics of the data and often utilizes data visualization methods.

EDA assists in determining the most effective methods to manipulate data sources in order to obtain the desired insights. It aids data scientists in discovering patterns, detecting anomalies, testing hypotheses, and validating assumptions. Popular tools for performing EDA include Python and R, which provide a range of functions and libraries for data exploration and visualization.

- *Inferential Analysis*

Inferential analysis involves drawing conclusions and making inferences about a larger population based on a sample of data. Once descriptive statistics have been computed and analyzed, further analysis is conducted to extrapolate findings from the sample to the broader population from which it was drawn. The objective of inferential statistics is to derive meaningful insights and make generalizations about the population based on the information obtained from the sample. By employing various statistical techniques, researchers can estimate population parameters, test hypotheses, determine the significance of relationships, and make predictions. Inferential analysis allows researchers to go beyond the specific sample data and draw broader conclusions about the population (Albers, 2017).

- *Predictive Analysis*

Predictive analysis is employed to forecast future events or outcomes based on current and historical data. This type of data analysis utilizes various methods to make predictions about future results by examining patterns and trends in existing or past data. The accuracy of predictive analysis depends on the level of detailed information available and the depth of analysis conducted. Businesses can leverage the vast amount of data they accumulate to describe current trends, anticipate future developments, and most importantly, recommend the optimal course of action to ensure success through prescriptive analytics. Techniques commonly used in predictive analysis include data mining, data modeling, artificial intelligence, machine learning, and other approaches that enable critical predictions (Vigliarolo, 2019).

- *Causal Analysis*

Causal analysis aims to investigate the relationship between variables and understand how changes in one variable affect another. It seeks to identify the underlying reasons or causes behind observed phenomena or outcomes.

Causal analysis is particularly useful in understanding the "why" behind certain events or situations. By examining the relationships between variables, it helps uncover the factors that contribute to specific outcomes or behaviors. This

type of analysis is commonly employed in randomized studies and controlled experiments, where researchers manipulate variables to determine their impact on the desired outcome.

In the context of software quality, causal analysis can be used to investigate why software failures occur. By analyzing the factors that contribute to bugs or data breaches, organizations can identify the root causes and take preventive measures to avoid major setbacks. Causal analysis plays a crucial role in understanding the factors driving certain outcomes and informing decision-making processes to improve performance and mitigate risks.

- *Mechanistic Analysis*

Mechanistic analysis helps researchers understand how changes in certain variables lead to corresponding changes in other variables. It involves using deterministic sets of equations or models to compare outcomes and simulate scenarios.

Mechanistic analysis is based on the idea that understanding the underlying mechanisms and processes driving a system can shed light on the relationships between variables. By dissecting the components and interactions within a system, researchers can analyze how changes in one variable propagate and affect other variables.

Simulation is often employed in mechanistic analysis to explore different scenarios and predict outcomes based on the defined equations or models. This allows researchers to gain insights into the behavior of the system under various conditions and identify the key factors that drive changes in variables.

The primary objectives of mechanistic analysis include understanding the causal relationships between variables, elucidating the mechanisms by which changes in variables impact other variables, and providing clear explanations for past events based on data. It is particularly useful when studying specific activities or subjects where a detailed understanding of the underlying processes is crucial for gaining insights and making predictions.

V. CONCLUSION

This paper has attempted to present meaningful insights by featuring the different types of requirements gathering and analysis techniques adopted in information technology projects. Each technique has a predefine set of specific and unique characteristics and their context of application.

Irrespective of the field of study or the type of data being collected (quantitative or qualitative), accurate data collection or requirement gathering is crucial for maintaining the integrity of research. Properly examining and understanding the requirements before conducting analysis opens up numerous opportunities for businesses and organizations.

The selection of appropriate data collection methods, whether it involves surveys, interviews, observations, or other techniques, is essential to ensure that the data collected is relevant and reliable. Clear and well-defined instructions for collecting and utilizing the data further reduce the likelihood of errors and inconsistencies in the analysis process.

Accurate data collection and requirement gathering contribute to the overall quality of research by providing a solid foundation for analysis. It ensures that the data used in the analysis aligns with the research objectives and enables researchers to draw valid and meaningful conclusions. By maintaining the integrity of the research process through accurate data collection, researchers can enhance the credibility and reliability of their findings. Using proper analysis can help organizations and businesses make predictions, gain insights and conclusions based on the results.

REFERENCES

[1]. Ajayi, V. (2017, September 7). Primary Sources of Data and Secondary Sources of Data. https://doi.org/10.13140/RG.2.2.24292.68481
[2]. Albers, M. J. (2017). Quantitative Data Analysis—In the Graduate Curriculum. Journal of Technical Writing and Communication, 47(2), 215–233. https://doi.org/10.1177/0047281617692067
[3]. Almoosa, A. S. (2018). A Qualitative Case Study in Augmented Reality Applications in Education: Dimensions of Strategic Implementation. University of Northern Colorado.
[4]. Bacchus, J. (2012, November 13). The Case for Clean Subsidies. Harvard Business Review. https://hbr.org/2012/11/the-case-for-clean-subsidies
[5]. Bazeley, P., & Jackson, K. (2013). Qualitative Data Analysis with NVivo. Sage Publications Ltd.
[6]. Block, E. S., & Erskine, L. (2012). Interviewing by Telephone: Specific Considerations, Opportunities, and Challenges: International Journal of Qualitative Methods. https://doi.org/10.1177/160940691201100409
[7]. Camanho, G. L. (2017). Objectivity in scientific publication. Revista Brasileira de Ortopedia, 52(4), 371–372. https://doi.org/10.1016/j.rboe.2017.07.002
[8]. Creswell, J. W. (2014). Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. SAGE.
[9]. De leeuw, E. (2008). The effect of computer-assisted interviewing on data quality: A review of the evidence.
[10]. Fine, G. A. (2016). Towards a Peopled Ethnography: Developing Theory from Group Life. Ethnography. https://doi.org/10.1177/1466138103004001003
[11]. Fontana, A., & Prokos, A. H. (2016). The Interview: From Formal to Postmodern. Routledge.
[12]. Glen, S. (2019). "Data Analysis & Exploratory Data Analysis (EDA)" From StatisticsHowTo.com: Elementary Statistics for the rest of us!

[13]. Goertzen, M. J. (2017). Chapter 3. Introduction to Quantitative Research and Data. Library Technology Reports, 53(4), 12–18.
[14]. Heru Susanto, F.-Y. L. (2020). Business Process Reengineering, An ICT Approach. Apple Academic Press.
[15]. Hsu, B. (2018). Is literature review a qualitative research method? ResearchGate. https://www.researchgate.net/post/Is-literature-review-a-qualitative-research-method
[16]. Jamshed, S. (2014). Qualitative research method-interviewing and observation. Journal of Basic and Clinical Pharmacy, 5(4), 87–88. https://doi.org/10.4103/0976-0105.141942
[17]. Kabir, S. M. (2016). METHODS OF DATA COLLECTION (pp. 201–275).
[18]. Lavrakas, P. (2008). Encyclopedia of Survey Research Methods. Sage Publications, Inc. https://doi.org/10.4135/9781412963947
[19]. Mallela, S. (2020). Requirement Elicitation Techniques for Business Analysis. DLT Labs.
[20]. McGrath, C., & Laliberte Rudman, D. (2019). Using Participant Observation to Enable Critical Understandings of Disability in Later Life: An Illustration Conducted With Older Adults With Low Vision. International Journal of Qualitative Methods, 18, 1609406919891292. https://doi.org/10.1177/1609406919891292
[21]. Mushtaq, J. (2016). Different Requirements Gathering Techniques and Issues. INTERNATIONAL JOURNAL OF SCIENTIFIC & ENGINEERING RESEARCH VOLUME 7, ISSUE 9.
[22]. Oltmann, S. (2016). Qualitative Interviews: A Methodological Discussion of the Interviewer and Respondent Contexts. Forum Qualitative Sozialforschung / Forum: Qualitative Social Research, 17(2), Article 2. https://doi.org/10.17169/fqs-17.2.2551
[23]. QATestLab. (2021). Types of Test Data Used During Software Testing.
[24]. Sapsford, R., & Jupp, V. (2006). Data Collection and Analysis. SAGE Publications Ltd. https://doi.org/10.4135/9781849208802
[25]. Schick-Makaroff, K., MacDonald, M., Plummer, M., Burgess, J., & Neander, W. (2016). What Synthesis Methodology Should I Use? A Review and Analysis of Approaches to Research Synthesis. AIMS Public Health, 3, 172–215. https://doi.org/10.3934/publichealth.2016.1.172
[26]. Sutton, J., & Austin, Z. (2015). Qualitative Research: Data Collection, Analysis, and Management. The Canadian Journal of Hospital Pharmacy, 68(3), 226–231.
[27]. Vigliarolo, B. (2019). Prescriptive analytics: A cheat sheet. Data, Analytics and AI Newsletter.