Using Root Cause Data Analysis for Requirements and Knowledge Elicitation

Zhao Xia Jin[1], John Hajdukiewicz[2], Geoffrey Ho[2], Donny Chan[1], and Yong-Ming Kow[3]

[1] Honeywell Technology Solutions Lab, China
Zhao.Xia.Jin@Honeywell.com
[2] Honeywell Automation and Control Solutions Laboratory, USA
[3] In-Situ Research Pte Ltd, Singapore

Abstract. The purpose of this paper is to present a technique, called Knowledge FMEA, for distilling textual raw data which is useful for requirements collection and knowledge elicitation. The authors first give some insights into the diverse characteristics of textual raw data which can lead to higher complexity in analysis and may result in some gaps in interpreting the interviewees' world view. We then outline a Knowledge FMEA procedure as it applies to qualitative data and its key benefits. Examples from a case study are presented to illustrate how to use the technique. Proposed Knowledge FMEA brings many advantages such as forcing the analysts to become deeply immersed in the raw data, identifying how the information is connected in causation, classifying the data according to why, what, how formulations and quantifying the findings for further quantitative analysis.

Keywords: Root Cause Analysis, Failure Modes and Effects Analysis (FMEA), Thematic Analysis, Qualitative Research.

1 Introduction

Today, user centered design research emphasizes the importance of viewing the world from the stakeholder's point to better understand user needs and to drive design. Mature qualitative techniques, such as interview and focus groups, have been widely used for knowledge and requirements elicitation. Qualitative techniques generate large quantities of raw data from which researchers distill end users' perspectives, knowledge and world view. The large quantities of raw data, which is usually in the format of field notes, transcripts of interviews or focus groups, can be very complicated to analyze due to the nature of the technique utilized and the way stakeholders view the problem or domain. The raw data of qualitative research can be a relatively unorganized knowledge base. In order to uncover the mystery of this unorganized knowledge base, we identify four dimensions of textual raw data that can simplify analysis and aid in interpreting the interviewees' world view: (a) semantic and thematic classification, (b) explicitness, (c) causation and problem solving path and (d) significance and salience.

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 79–88, 2007.

Semantic and thematic classification. Qualitative studies usually rely on inductive reasoning processes to interpret and structure the meanings that can be derived from data [1]. Most analysis methods focus only on classifying the meaning of data according to semantic and thematic frameworks. Typical examples of these methods include thematic analysis and content analysis. Semantical content analysis [2] deals with semantic units (e.g., keywords) of the communication content. Thematic analysis deals with thematic categories (called themes) which are usually semantic entities that have been selected as having a special meaning. Though it is often not explicitly claimed as the method of analysis, Braun and Clarke [3] argued that a great deal of analyses are essentially thematic but is either claimed as something else or not identified as any particular method at all. These methods of analysis are indeed useful but may be limited in extracting more details in tacit knowledge such as interconnections of why and how events occur.

Explicitness. Knowledge is different in the degree of explicitness. Explicit knowledge is transmittable in formal, systematic language. However, tacit knowledge is personal, deeply rooted in a specific context, and hard to formalize and communicate [4, 5, 6, 7]. Commonly claimed notions say, "we know more than we can tell" [4] and "users do not know how to articulate their real needs and requirements". For example, the interviewee may be able to express the information in knowing-that but not in the form of knowing-how. Therefore, the large amounts of raw data consist of unstructured explicit knowledge, tacit knowledge and have complex relationship between data. The difficulty arises because we are trying to translate the tacit knowledge into explicit or codified knowledge. The difficulty is magnified because we attempt to externalize the complex relationship between data.

Causation and problem solving path. Many researchers point out that data analysis should take into account the interrelationship between the themes as they occur within the data [8, 9]. Causation, as an important data relationship, is often overlooked in discussions of qualitative research [10]. Causation is the relationship between cause and effect that holds association between events, properties, causes, influence, determination, contribution, effects, and so on. Cause-effect relationship is the primary aspect that ties into how to prevent the problems from failure again. Baptiste [10] suggested that the analyst should ask questions to reveal the causation but he did not provide an effective method to conduct the analysis. Another exception is pragmatically content analysis, which includes procedures for classifying signs according to their probable causes and effects and emphasizes said the purpose of discourse [2]. The causation and its importance in problem solving drive us to analyze data according to cause-effect relationships and problem solving paths, which involves identifying problems, evidence, root causes, consequences, and providing solutions as corrective and preventive actions.

Significance and salience. Though the usual focus of data analysis of qualitative research is on the meaning of the information collected, quantitative analysis can be followed to uncover the significance and salience of the findings. For example, the quantitative dimension of the content analysis, centers on determining the frequency of keywords or phrases in context [11, 12]. Another example is to simply count the

frequency of occurrences of each of the themes determined [13]. Such analyses are useful for comparing the relative frequency, and arguably the salience, of items of analysis but do not tell us much about the patterns among the items [14]. Also, such analyses involve only simple forms of quantification mostly in the format of frequency but do not consider the severity of the problems which are rooted in the meaning of the textual raw data. Ultimately, "all qualitative data can be coded quantitatively since anything that is qualitative can be assigned meaningful numerical values" as pointed out by Trochim [15].

As summarized by Guest and McLellan [8], during the past few decades, researchers have developed a host of methods for structuring and analyzing textual data (for a review of these, see Dey 1993 [16]; Bernard and Ryan 1998 [17]; LeCompte and Schensul 1999 [18]). Most of the methods only focus on analyzing certain dimensions of knowledge and seldom detail the analysis procedure on how to address the explicitness and root cause paths of the information. This paper applies Root Cause Analysis as a method to analyze the interview data and generate a Knowledge Failure Modes Effects Analysis (FMEA).

2 What Is Knowledge FMEA?

Knowledge FMEA is a FMEA-like data analysis technique designed to decompose and organize the textual data according to cause-effect relationships, quantify the findings by Severity, Occurrence, Detection and Risk Priority Number (RPN) ratings, inform features, functionality and possible solutions ideas. Knowledge FMEA addresses the four dimensions of textual raw data mentioned earlier. Here, it forces the analysts to be very clear about how the information is connected (why, what, and how formulations) and separate the data into root cause and problem solving path.

3 Procedure of Doing Analysis with Knowledge FMEA

The six stages of doing analysis with Knowledge FMEA are summarized in Table 1.

3.1 Get Familiar with Data

Immerse oneself in the data and get a sense of the data as a whole before breaking it into parts. Data collected from an interview might be written information (interview notes, field observational notes) and audio/visual information (audio recordings, video recordings, pictures) that might need transcribing. The analysis begins with getting familiar with the data by repetitious reading to search for the underlining meanings and patterns. Perusing the data allows the analyst to make use of prior knowledge about the data collected and instill initial analytic thought as a whole before breaking it into parts. During this familiarization process, major themes begin to emerge. Making additional notes, color codes and marking initial categorization will be useful for latter stages of coding and will provide a reference point in the later stages. Subsequently, the data can be coded into short statements which will be grouped into the different themes that house them latter.

Table 1. Procedures of Knowledge FMEA Analysis

Analysis Stage	Procedure
State 1: Get familiar with data	(1) Transcribe data (if necessary). (2) Repetitious reading to get familiar with the data and get a whole picture. (3) Code data where the themes start to emerge. (4) Make notes and mark of the initial idea.
State 2: Fundamental Thematic analysis (optional)	(1) Complete coding process with an iterative process. (2) Structure themes and concept map of themes. (3) Classify and group the statements. (4) House the statement according to standardized themes.
State 3: Root cause Analysis	(1) Tailor fields of root-cause path. (2) Conduct root cause analysis iteratively. (3) Standardize the analysis results if necessary.
State 4: Quantitative rating	Rate Severity, Occurrence, Detection and calculate RPN (Risk Priority Number).
State 5: Review the analysis	(1) Review the analysis by self revision or peer review. (2) Follow up or verify with participants if necessary.
State 6 : Quantitative analysis & interpretation	Perform statistical analysis using quantitative RPN data and interpret the analysis.

3.2 Fundamental Thematic Analysis

Optional step to classify the data for further analysis. Thematic analysis begins when you start coding the raw data, breaking them into parts and grouping them by proximity. A quick way to describe a phenomenon, identify and verify the patterns from unorganized data is to classify and group the data. A good method to do this is to use thematic analysis as a fundamental step to classify and organize data into themes.

How to use thematic analysis is widely reported in literature [3, 15]. Thematic analysis can be done in a deductive or inductive manner. In the deductive (top-down) method, data analysis is determined by the research objectives and the themes pre-exist in theory or literature. While in the inductive (bottom-up) method, the themes are developed from the empirical data or emerging from reading and re-reading the raw data. Themes should be refined, decomposed and structured so that a concept map (e.g., hierarchy chart) of themes can be created.

In general, thematic analysis helps to select, focus, abstract and transform the data collected into manageable information segments to show patterns. This step lets analysts to familiarize the data further, organize the statement and better prepare analysts to perform further root cause analysis. House the statement according to standardized themes.

3.3 Root Cause Analysis

Next, root cause analysis is conducted to further analyze and understand cause-effect relationship with the statements categorized earlier. A convenient way is adopting existing mature tools like FMEA. While FMEA can be very complex, not all the fields of root cause path (e.g., failure modes, causes, consequences and corrective actions) need to be adopted. These fields shall be tailored to fit the research interest and the level of detail a researcher trying to identify.

Generally, in order to elicit more knowledge, researchers actively ask open questions (e.g. What, When, Where, Why, How) and use probing technique on participant statements. As a result, researcher notes are a collection of answers that

are scattered and do not have coherent relationship. To piece this part and parcel together to tell a story based on cause-effect relationships, researcher continue with further iteration of active reading and asking what, why and how on the collected statement. This process is consistent with widely reported root cause analysis method. By doing so, assumptions might be made on the statements to uncover the relationship. When statements collected are ambiguous, they can be left unclassified with additional notes or markings. Both assumed and uncategorized results need further clarification and verification from participant or peer review to solve the problematic analysis and at the same time eliminate discrepancies.

3.4 Quantitative Rating

This step involves rating all statements after they are categorized and cause-effect relationships are identified. In this step, the analyst needs to rate three items adopted from FMEA analysis tool including 1) Severity (S) – how serious the problem is, 2) Occurrence (O) – how frequent does it happen and 3) Detection (D) – how easily it can be detected. By multiplying these three ratings, a Risk Priority Number (RPN) will be generated that reflects the impact of the happening. These FMEA tool's rating provide a better representation of the statement in a quantitative manner from different angles. For example, RPN (9) = S (9) ×O (1) ×D (1), explains a finding that is serious but might seldom occur or might be easily detected. Another example is, RPN (9) = S (5) × O (9) ×D (5), explains a moderately serious problem that always reoccurs and is oftentimes detectable. The combination of these three ratings creates a high RPN that means it brings high risk and big impact that immediate attention is needed.

Effective and well defined standard rating scales should be used; commonly we use rating scales of 0, 1, 3, 5, 7, 9. Different rating scales might be used separately for Severity, Occurrence and Detection. However, these usually have been standardized in a mature FMEA tool. These S, O, D and RPN provide additional insight that helps to interpret data. The rating process can be performed by a few researchers and can take place at a different time. Results can later be easily shared and combined.

3.5 Review the Analysis

Self review or peer review of the thematic analysis, root cause analysis, and rating help to identify and eliminate discrepancies that arise. Researcher might want to reconsider some categorization that seems to be ambiguous, especially on those results that had been marked with notes or labels. Peer review share more insight of the respondents' world view according to the experience during the study. Any discrepancies found during reviewing process, need be resolve and achieve common agreement among team member. Researcher might need additional follow up with participants and verify the problematic results. Having a reviewing process helps to generate a more reliable (e.g. check and balance by peers) and a more valid data (e.g. standard rating and eliminate discrepancies).

3.6 Quantitative Analysis and Interpretation

With the S, O, D and RPN rating score, further quantitative analysis can be conducted such as descriptive statistic analysis. For example, a matrix with two theme categories

can be created to identify the areas with higher/lower scores. Also, a matrix can be created to analyze the correlation between two themes. Researcher can also compare topics/areas within a sub-theme separately. Finally, these analyses can be interpreted together with qualitative profile and patterns. Refer to examples of Case Study for the more details.

4 Advantage of Using Knowledge FMEA Analysis

i) Researchers are forced to be explicit about the relations in the data. The root-cause analysis forces the analysts to deeply immerse themselves in the raw data and to be very clear about how the information is connected in causation. The method identifies the information that is explicitly or implicitly expressed by the interviewees and thus reflects the interviewees' world view more accurately and with greater detail.

ii) Facilitates problem solving. The analysts are forced to classify the data according to why, what, how formulations and problem solving paths, and thus facilitates problem solving and informs features, functionality, and solutions.

iii) Enables zooming results to areas of interest. The researchers will be aware of the assumptions that they made during analysis and what needs to be followed-up or verified.

iv) S, O, D and RPN ratings quantify the findings for further analysis. These numbers can be further analyzed to achieve greater insight into the meaning of the data; for instance, ratings can provide further prioritization of information or help to examine specific hypotheses.

v) Knowledge FMEA analysis is flexible. The fields that need to be analyzed are customizable and tailor-able depending on the level of detail that the research needs. Also, knowledge FMEA can work in conjunction with other qualitative analytic methods, such as thematic analysis. Finally, even if you just analyze part of the statements stemming from the raw data, the analysis method still provides a deep level of insight.

5 Examples of Case Study

Below is a case study where we used the Knowledge FMEA process.

5.1 Project Background

This field study was a multiple stakeholder study (15 participants from five different stakeholder categories were interviewed) and involved multiple sites (3 different sites). The purpose of the project was to identify the key factors that affect the productivity of work of a complex collaborative project environment in various areas, such as activities, environment, tools, communication and coordination process. Semi-structured contextual inquiry and observation were used to elicit knowledge from participants. For each interview, two researchers conducted the interview as a team.

5.2 Organizing Raw Data and Fundamental Thematic Analysis

As all site visits happened in different stages, a preliminary deductive thematic analysis took place immediately after the interview and produced a report for each interview. The process decomposition of the life cycle of the work was used as a theme category to classify the statements. Also, signs were used to mark the comments (positive, negative, neutral comment) in the reports. This initial familiarization and categorization helped to further improve the site visit design and provided preliminary organized data.

Subsequently, statements categorized in the preliminary report were combined and transferred to a spreadsheet by inserting individual statements into rows. Another new theme category (in here is 'Productivity Area' of from theory [19] for the case study) was added to extend the level of granularity. For the root cause analysis, interested fields from an FMEA (e.g. failure mode, root causes, correcting actions) were tailored into the analysis template. We customized the field names according to the research interest. For example, the term X-Factors replaced the original term, Failure Mode; X-Factors are the known and potential factors (include failures) that could increase, prevent or degrade the productivity, quality and efficiency. Causes of Failures were renamed to Diagnostic Causes referring to any known and potential causes. Correcting Actions was changed to Possible Solutions as we more concerned with the design changes which can correct the failure modes and improve the efficiency of the tools. Also, rating areas such as S, O, D and RPN, were selected. All these fields were combined and form an analysis template as in Table 2. All the statements were coded and analyzed based on this template.

Table 2. Analysis template with two deductive theme criteria and root cause path

Thematic categorization		**Root cause analysis**							**Original Statement**
Process/Sub-function	Productivity Area	X-Factors	Severity	Diagnostic Causes	Possible Solutions	Probability of Occurrence	Detection	RPN	Original statements

5.3 Root Cause Analysis

As mentioned, root cause analysis was conducted to identify or elicit certain root causes behind a problem and ideas to prevent the problems. In Table 3, an inductive example (example 1) and a deductive example (example 2) of conducting root cause analysis are shown by applying the process of asking what, why and how

Statement 1: "There should be no XXX step anymore in HW engineering...This should be automatic...The biggest challenge is if you make a mistake in XXX step, it's about 3 to 5 times the cost to change the mistake..."

As shown in Row 1, the participant's statement explicitly states that he was referring to a step in the 'HW Engineering' process and that the step was 'poorly supported by Tool Features'. This problem resulted in 'about 3 to 5 times the cost' due

to 'no automation support'. Apparently, this problem can be improved by 'automating X steps in the tool'.

Statement 2: "[We] often go to training and still don't know enough..."

As shown in Row 2, a participant provided feedback of the training they had attended. By just looking at the transcript, various interpretations and assumptions can be summarized about the training. Assumptions were made and translated as 1) product is too complex and difficult to understand, 2) training hours are not sufficient therefore need longer hours, or 3) poor training quality. Based on the interview context, assumptions were made that the participants felt that the training was 'poorly organized', was 'without sufficient information about the product' and had 'no clear training objectives'. Through brainstorming, two solutions, 'setup well organized training' and 'provide sufficient information supported by tool' were provided. Researchers then verified their assumed statements from the participant and confirmed that these were the actual causes of their discontent.

Table 3. Examples of analyzing path of root cause analysis

Process/ Sub-functions	Productivity Area	X-Factors	Root cause	Potential Solutions
HW Engineering	Tools	Poor support of XXX step	[1] No automation support [2] Cost about 3 to 5 times the cost to change the XXX mistake.	Automate XXX
General	Knowledge & skills	Poor product training	[1] Insufficient information about the product. [2] Training poorly set up [3] No clear training objective	Setup up well organized trainings Provide sufficient information supported by tool

Finally, an affinity tree was created by simply sorting the spreadsheet file according to different themes. This helped to bring out the coherent relationship of the results and tell a story by linking the what, why, how formulation. As well, it demonstrated the consequences. We also standardized the X-factors to see patterns created by X-factors.

5.4 Qualitative Analysis and Data Interpretation

After all the statements were categorized according to themes and the root cause analysis was completed, two researchers rated Severity, Priority, and Detection according to a standardized rating scale to generate RPN ratings. Accumulated experience in the field and further understanding during the process of familiarization and categorization helped the researchers rate these statements more accurately. In peer review session, researchers shared their insight and opinions about the statements. At times there were debates to resolve discrepancies in their opinions, particularly when the RPN scores had large inter-rated deviations.

Table 4. Example of matrix table with two theme categories

Process/Sub-Functions Productivity Areas	Productivity Area 1	Productivity Area 2	Productivity Area 3	Productivity Area 4	Productivity Area 5	Productivity Area 6	Productivity Area 7	Productivity Area 8	Productivity Area 9
Process/Sub-Functions 1	34	20	65	6	31	33	31	101	28
Process/Sub-Functions 2	13	26	1	/	/	/	/	23	9
…	…	…	…	…	…	…	…	…	…

Table 4 illustrates a cross-combination of RPN scores to compare main two dimensions. One dimension reflected the work process and their sub-functions and the second dimension reflected key productivity themes or areas). Certain cells were empty because no data was collected for that theme. These empty fields also created a check and balance process where the data collected has to correspond to the role and job responsibilities of the participants.

The S, O, D, and RPN could be easily sorted in the spreadsheet to show the most serious problems, the problems that most frequently, the problems that were not easily detected, and the impact of the overall problem.

6 Conclusion and Discussion

The present paper identified four dimensions of textual raw data that contributes to the complexity of qualitative analysis. These dimensions oftentimes will result in gaps in interpreting the interviewees' world view. The authors focused on the cause-effect relationship and introduced a root cause analysis technique that can be applied to qualitative data analysis. As we have shown throughout this paper, Knowledge FMEA analysis is not a complex method and has many advantages. Used properly, Knowledge FMEA can be both a useful and a powerful technique, but its use must be associated with in-depth knowledge of the raw data. A significant drawback is that it is time-consuming if you try to analyze large amounts of data with this technique. However, you'll get results are easily interpretable and still rich with details. Ultimately, one's approach will depend on the type of data, the granularity of coding, and the research objectives, as Greg Guest [8] pointed out.

References

1. Thorne, S.: Data Analysis in Qualitative Research. In: Evidence Based Nursing, vol. 3, pp. 68–70. University of British Columbia Vancouver, Canada (2000)
2. Janis, I.L.: The Problem of Validating Content Analysis. In: Lasswell, H.D., Leites, N.C. (eds.) Language of Politics; Studies in Quantitative Semantics, pp. 55–82. MIT Press, Cambridge, MA (1965)
3. Braun, V., Clarke, V.: Using Thematic Analysis in Psychology. In: Qualitative Research in Psychology, vol. 3, pp. 77–101. Edward Arnold, United Kingdom (2006)
4. Polonyi, M.: The Tacit Dimension. Doubleday & Company, Garden City (1996)
5. Crowley, B.: Tacit Knowledge, Tacit Ignorance, and the Future of Academic Librarianship. College and Research Libraries 62 6, 565–584 (2001)

6. Gourlay, S.: Tacit Knowledge, Tacit Knowing or Behaving? In: Third European Conference on Organizational Knowledge, Learning, and Capabilities, Greece (2002) at http://www.alba.edu.gr/OKLC2002/Proceedings/pdf_files/ID269.pdf
7. McInerney, C.: Knowledge Management and the Dynamic Nature of Knowledge. Journal of the American Society for Information Science and Technology 53(12), 1009–1018 (2002)
8. Guest, G., McLellan, E.: Distinguishing the Trees from the Forest: Applying Cluster Analysis to Thematic Qualitative Data. In: Field Methods, vol. 15, pp. 186–201. SAGE, Thousand Oaks, CA (2003)
9. Henderson, C., Yerushalmi, E., Heller, K., Heller, P., Kuo, V.H.: Multi-Layered Concept Maps for the Analysis of Complex Interview Data. In: Roundtable Discussion presented at the Physics Education Research Conference, Wisconsin (2003)
10. Baptiste, I.: Qualitative Data Analysis: Common Phases, Strategic Differences, Forum: Qualitative Social Research, 2(3) (September 2001), at http://www.qualitative-research.net/fqs-texte/3-01/3-01baptiste-e.pdf
11. Berelson, B.: Content Analysis in Communication Research. In: Glencoe, I.L.: The Free Press Publishers (1952)
12. Kaplan, A.: Content Analysis and the Theory of Signs. In. Philosophy of Science 10, 230–247 (1943)
13. Leites, N.C., de S. Pool, I.: On Content Analysis. Library of Congress, Experimental Division for Study War-time Communications. Documents 26, 1–2 (1942)
14. Richards, T., Richards, L.: Using Hierarchical Categories in Qualitative Data Analysis. In: Kelle: Computer-Aided Qualitative Data Analysis: Theory, Methods, and Practice, pp. 80–95. Sage, Thousand Oaks, CA (1995)
15. Trochim, W.M.: The Research Methods Knowledge Base, 2nd edn. (2006), at http://trochim.human.cornell.edu/kb/index.htm
16. Dey, I.: Qualitative data analysis: A user-friendly guide for social scientists. In: Bernard, H.R., Ryan, G. (eds.) 1998 Routledge, New York (1993)
17. Bernard, H.R., Ryan, G.: Text analysis: Qualitative and quantitative methods. In: Bernard, H.R. (ed.) Handbook of methods in cultural anthropology, pp. 595–645. Altamira, Walnut Creek, CA (1998)
18. LeCompte, M.D., Schensul, J.J.: Analyzing and interpreting ethnographic data. Altamira, Walnut Creek, CA (1999)
19. Accel-Team: Productivity Improvement (2006), at http://www.accel-team.com/productivity/productivity_01_what.html