

Cognitive Approach Using SFL Theory in Capturing Tacit Knowledge in Business Intelligence

Herison Surbakti^{1,2}
 Information System
 Universitas Respati Yogyakarta¹
 Yogyakarta, Indonesia
<https://orcid.org/0000-0003-4193-446X>

Azman Ta'a²
 School of Computing
 Universiti Utara Malaysia²
 Sintok, Malaysia
azman@uum.edu.my

Abstract— The complexity of Business Intelligence (BI) processes need to be explored in order to ensure BI system properly treats the tacit knowledge as part of data source in BI framework. Therefore, a new approach in handling tacit knowledge in BI system still needs to be developed. The library is an ideal place to gather tacit knowledge. It is a place full of explicit knowledge stored in various bookshelves. Nevertheless, tacit knowledge is very abundant in the head of the librarians. The explicit knowledge they gained from education in the field of libraries and information was not sufficient to deal with a complex and contextual work environment. Complexity comes from many interconnected affairs that connect librarians with the surrounding environment such as supra-organizations, employees, the physical environment, and library users. This knowledge is contextual because there are various types of libraries and there are different types of library users who demand different management. Since tacit knowledge hard to capture, we need to use all possible sources of externalization of tacit knowledge. The effort to capture this knowledge is done through a social process where the transfer of knowledge takes place from an expert to an interviewer. For this reason, it is important for the interview process to be based on SFL theory (Systemic Functional Linguistics).

Keywords—Business Intelligence, Tacit Knowledge, SFL theory

I. INTRODUCTION

Cognitive approach is ideally suited for the capturing knowledge as from among the massive data available these days. The decision maker typically must integrate multiple streams of information from information or other collaboration with the knowledge systems in making decisions [1]. Furthermore, decisions may be based in organizational politics or routines [2], and decision makers may limit themselves to a few choices because of “bounded rationality” [3]. Ducharme and Angelelli [4] invented the use of cognitive as an advanced analytics to capture and extract tacit knowledge by elaborating the predictive analytics, stochastic analytics, and cognitive computing. Moreover, the advanced analytics approach still be implemented in Business Intelligence (BI) environment [17]. Thus, the basic BI framework with involving a tacit knowledge approach can be illustrated as shown in **Figure 1**.

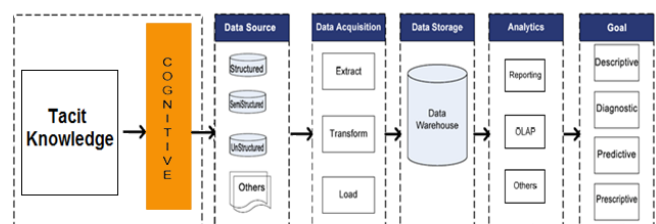


Figure 1 Tacit Knowledge in BI Framework Using Cognitive Approach [17]

The academic library has consumers who are not as heterogeneous as the public library because it serves limited types of consumers, namely students, lecturers, and university staffs. This study limits the context by taking academic libraries as research contexts. Context control is a natural thing because business organizations are also bound to their respective contexts. This can be relevant to the business context where business libraries may only serve the internal needs of an organization with a limited and certain number of organizational structures. Context control also simplifies the problem so that it leaves aspects of the complexity of tacit knowledge in the library.

The simple stages of capturing tacit knowledge can be illustrated in **Figure 2** below. After a theoretical review, the data collection scheme consists of three stages, namely interviews to find out the context for tacit knowledge that is none other than the problems faced by librarians; the survey stage that detects librarians who have the tacit knowledge needed to solve the problem; and the second interview stage, which revealed tacit knowledge from librarians.

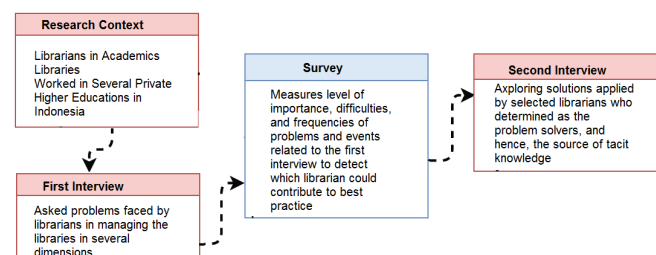


Figure 2 Research Data Collection Scheme

II. DESIGN FOR CAPTURING TACIT KNOWLEDGE

There is small number of earlier research about business intelligence on academic library and library profession.

Example of this research is Cox and Janti [5] on Library Cube project, a business intelligence system that demonstrate the value that can be provided by academic libraries. However, the research is not targeting the tacit knowledge at all since it is only targeting the provided information in academic information system. Heims et al [6] mentions that reporting BI research and creating BI reports are the key area of responsibility of librarians in information era. We addressed the problem by open dialog with librarian, which actually what considered would happen between BI manager and librarian to develop clear communication channels [7]. Noted that for librarian, BI is part of their challenge in information era [8].

Since tacit knowledge is hard to capture, we need to use all possible sources of externalization of tacit knowledge [9]. The effort to capture this knowledge is done through a social process where the transfer of knowledge takes place from an expert to an interviewer. For this reason, it is important for the interview process to be based on SFL theory.

According to SFL theory, only a fraction of “can do” turned into “can mean” and only a fraction of “can mean” turned into “can say” [9]. This is what is meant by Polanyi when he said “we know more than we can tell” [10]. Hence, only a portion of tacit knowledge can be captured by linguistic means. We need other means that came up from “can mean” which anything that could analyse semiotically. It could be non-verbal cues or drawing, written text, etc. We refer to drawing, photograph, videos, written text, and others as documented source and beyond our analysis. Here we just focused on non-verbal cues. However, whenever documented sources considered relevant, we could use it as source of tacit knowledge.

A. Linguistic Source of Tacit

According to SFL theory, language is realized in four strata: semantic, lexicogrammar, phonology, and phonetics [11]. Semantics is the highest level that explains the hidden meaning of language. Lexicogrammar is an aspect of language that explains real meaning, can be seen from the choice of words and grammar used. Phonology is the meaning that exists in sound. Phonetics is speech that arises from language activities. It can be seen that this stratification moves from something abstract (semantic) to something concrete (phonetic).

Someone will choose a word to represent his experience when speaking. What word or wording chosen can distinguish whether the experience or knowledge expressed is inheritance or not. In fact, sometimes, a person will find it difficult to find the right words to describe their knowledge so that they choose new words, ask the right words, or state their difficulties in describing them.

From the LCM (Linguistic Categorization Model) and SFL, it can be concluded that the effort to explore linguistic knowledge linguistically must be directed to the question "how" and the words action verbs. This is referred to as grammar-targeted questioning (GTQ). GTQ are questions that focus on the word "how" in the interview. It is distinguished from Content Targeted Questioning (CTQ) which focuses on "what", "when", and "where" or Semantic Targeted Questioning (STQ) which focuses on "why".

Zappavigna [9] used GTQ as a supplement to CTQ to express one's personal knowledge found that GTQ is able to encourage the concretization of CTQ. In this study, a common

response arises when the resource person is asked "how" is exposition ('in other words'), clarification ('to be precise'), or exemplification ('for example'). Responses like these contain a high load of tacit knowledge because they reach deeper descriptions of one's knowledge than can be achieved by content-focused strategies that might only reveal something very general like 'good', 'well', and 'alright'. If the participant expresses this in the interview, it is the job of the interviewer to elaborate this answer in more depth. The interviews need to be carried out sequentially with the first content-based interview followed by the grammar-based interview in order to minimize the substantial learning effect. Substantial learning effects occur when grammar-based interviews make participants rethink and reflect on themselves so that it impacts on subsequent content questions. This results in the answers given not being comprehensive and depending on the order. The protocol for running GTQ according to Zappavigna [9] is as follows:

1. The interviewer asks a general question of the form "Tell me about your particular area of expertise" (or if the task/domain is sufficiently specified "Tell me about task X").
2. The interviewee responds.
3. The interviewer interrupts the interviewee when he/she has identified a grammatical feature of under-representation about the particular content-area of interest and asks a question aimed at unpacking the grammatical feature.
4. The interviewee responds to the question
5. The interviewer repeats steps 3 and 4 until he/she has constructed a coherent argument for a particular reading of the interviewee's tacit knowledge about some topic. The argument is coherent in the sense that it is supported by multiple patterns of grammatical features.
6. When the interviewer has achieved an understanding of the concept/skill that is the topic of the interview he may present his 'reading' to the interviewee, asking a question of the form "Now I think I have understood what you mean about X. How do you feel about this reading?".
7. The interviewer and interviewee may engage in unstructured discourse relating to what has occurred in the interview.
8. The interviewer will conclude the interview at an interpersonally appropriate juncture.

B. Contextual Resources

Even after knowledge has been expressed verbally and non-verbally, there is still space where the knowledge of tacit cannot be expressed at all and can only be demonstrated by behavior. Apart from observations requiring precise and specific time, experts generally do not like being observed while working [12]. In addition, observations become more complicated when several experts are involved [12]. This can only be done in a non-intrusive manner such as a surveillance camera, but it can be a problem with privacy issues. Alternatively, observations can be made through third-person testimonies. In this case, the interview was conducted on the third person who had witnessed the behavior of the first person who was the target to reveal the knowledge of his possessions.

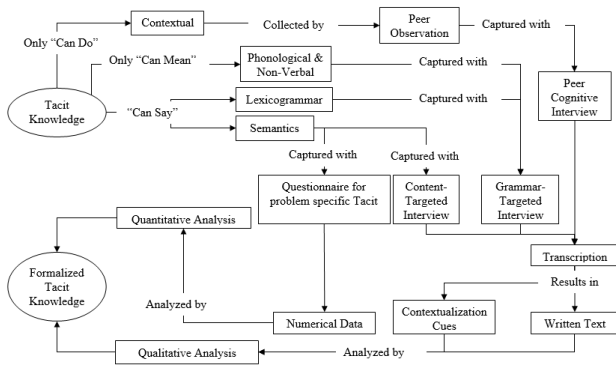


Fig 3 Design for Capturing Tacit Knowledge

The framework above shows the design used to capture comprehensive knowledge of experts. Based on SFL theory, tacit knowledge consists of three levels. The first level is the most basic level where a person can only do but cannot interpret it, let alone say it. This knowledge is contextual tacit knowledge because it can only be raised in a supportive context. It can only be collected through observation. Even so, because the context is very specific, in terms of space and time, only people present in that context can see and understand from their perspective what the tacit knowledge is. In this study, it is assumed that the person is a peer. Researchers collected data on tacit knowledge from peers through cognitive interviews. Furthermore, we can conclude there are two ways to collect tacit knowledge:

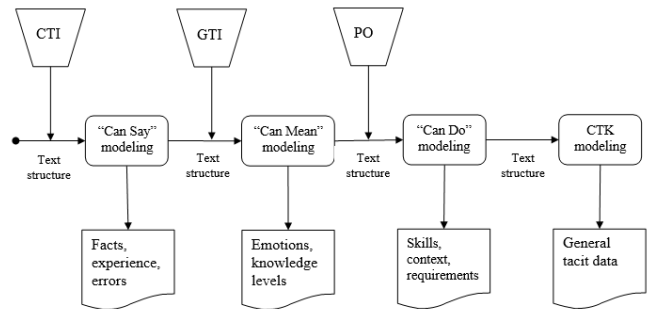
1. **Focused on a stated problem.** Participant presented with a problem which needs tacit knowledge to be solved. The tacit knowledge needed to solve this problem can be collected with interview, based on respondents chosen with questionnaire. Questions in the interview informed by problems urgency, detected by questionnaire. Here, sequences of the steps determine the completeness of tacit knowledge. Figure below show the connection between questionnaire design and decision.



Figure 4 General flow of information to collect problem focused tacit data

2. **Comprehensive tacit data collection.** The technique above only collects “can say” dimension of tacit knowledge. Furthermore, the “can say” in this sense only focused on the problem stated, not all the tacit knowledge possessed by the participant, at least for the problem field. The solution is to collect tacit knowledge data more comprehensive by three means:
 - a. Using content targeted interview (CTI) and grammar targeted interview (GTI) to collect “can say” dimension of tacit knowledge, not bounded by a problem.
 - b. Using grammar targeted interview to collect “only can mean” dimension of tacit knowledge.
 - c. Using peer cognitive interview (PCI) to collect “only can do” dimension of tacit knowledge.

The data collection and analysis process are illustrated in **Figure 5**.



Note: CTI = Content-targeted interview, GTI = Grammar-targeted interview, PO = Participant observation, CTK = Comprehensive tacit knowledge

Figure 5 Comprehensive tacit data collection

As illustrated in **Figure 5**, data collection process includes four steps. Each step supplemented by its tool. Notes that text structures are the input and output from each step. Data results from “can say” modelling specified as facts, experience, errors, and anything which only can be described by words. For example, numerical codes or particular reference used by participants. Data from “Can mean” consisted of words pattern loaded with emotions or particular stress on something that signify confidence to a statement. Skills, context, requirements, or anything which can’t be said by words could result from “can do” modelling.

The two ways to collect tacit knowledge complement each other. The questionnaire method is quantitative and provides data focused directly on an issue, which is useful for instant and standard situations. The interview method is qualitative and provides more general data. This data must first be translated into useful data in an instant situation but can be useful especially when certain problems cannot be obtained through a questionnaire. General principles from interview data can be used to be translated quickly at an operational level.

Figure 6 below could fit well with proposed BI framework previously modelled in **Figure 1**. In the model, numeric data analysis can be considered as part of cognitive analytics using cognitive mapping process. In itself, problem focused tacit data is the knowledge component of the model. The same goes for general tacit knowledge taken by interview process.

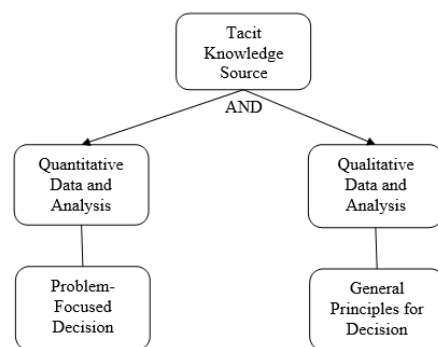


Figure 6 Simple Diagram for Tacit Knowledge Capturing

III. RESEARCH DESIGN

This paper focused on tacit data collection problem. This research is limited to problem focused data collection because the data is sufficiently structured. However, problem focused tacit data can achieve comprehension like general tacit data if the problems are reviewed comprehensively so that they cover all the problems that exist in the work. This is because tacit knowledge is procedural, and these procedures are related to problems. Standard Operating Procedures (SOP) are basically explicit knowledge to deal with problems in the form of how to do the job properly. SOP can be part of general public knowledge if in carrying it out, the resource person is faced with a different situation from the textbook. If they are the same, then the only problems are the components of the knowledge of tacit. The following figure can explain this well.

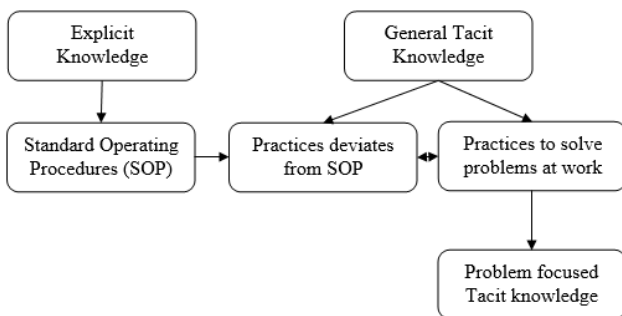


Figure 7 Scope of tacit knowledge

Mathematically, if E is explicit knowledge and T is tacit knowledge, then knowledge K is:

$$K = E + T$$

T consisted of practices deviates from SOP (D) and practices to solve problems not stated in SOP (S), then T is

$$T = D + S$$

While S itself is problem focused tacit data (F):

$$S = F$$

Hence, if D = 0, then T = S = F

This assumes that librarians follow the SOP strictly and all problems they faced at work is not stated in SOP. Guided by this assumption, the author feels save to not conduct observation. Hence, the research steps to collect data illustrated in **Figure 8**.

The steps above described as follows:

A. Identify Interview Panel

According to Marshall et al [13], determining sample size in qualitative Information System (IS) research can be justified by three methods. First by citing recommendation from qualitative methodologists. Second, by citing sample sizes used in studies tackled with similar research problems and design. Third, by internal justification using statistical demonstration of data saturation. This study could consider as case study by focusing in librarian cases. Using first recommendation, Yin [14] recommends at least six sources for case study, while Creswell [15] recommends three to five sources. Meanwhile, looking for case studies in IS research, Marshall et al [13] recommends that the research should contain 15 to 30 interviews. For internal justification, the calculation could infer from Marshall et al [13] graph. The graph relating sample size with number of codes. Generally, larger codes mean smallest sample. As we would see in next section, this research employed 13 codes (questions) for

participants. The value then falls to 7-18 participants in the research.

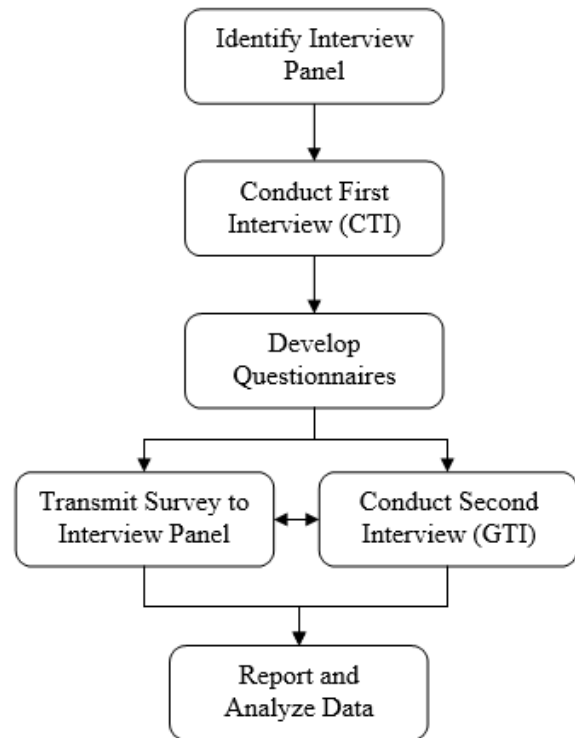


Figure 8 Research Design

The interview panel planned to include 23 librarians, each from different university library in West Java, Indonesia. The number chosen to compensate the possibility that the next rounds there are some librarians which declined to participate. The sample size also fit with average sample size in IS research, which is between 15-30 interviewees [13]. It is also larger than sample size recommended by methodologists and by statistical calculation. If there are reduction of 40% from each round, in the third round there are still eight librarians participated, large enough to analyse qualitatively.

B. Conduct First Interview

The first interview consists of two part. The objective for this interview is to identify points to incorporate into the questionnaire. To do this, CTI conducted.

Themes for interview were adopted from Si and Yujia [16] on collective tacit knowledge of librarians. Si and Yujia [16] mention two forms of librarian's tacit knowledge: personal and collective. Our focused only on collective form since this form ready to share between librarians. Librarian's personal tacit knowledge such as librarians' ability on scientific research, ability to analyze and solve problems in the process of knowledge mining and knowledge reorganization, ability to accept new things and find and solve the question, and other; is more connected to personality and as the name implies, highly personal, hence need deeper reflections and analysis. Furthermore, the collective tacit knowledge list already exhausted. Given the time and energy constraint in the interview process, we left the personal tacit knowledge to further research. The collective tacit knowledge of librarian according to Si and Yujia [16] includes:

1. The library's long-established working methods,
2. The common experience of librarians' dealing with problems at work;
3. The mechanism of knowledge communications among librarians, and between librarians and readers;
4. Library's ability to cope with emergency events and coordination in internal and external environment;
5. The overall level of library service and reputation;
6. Affinity and cohesion within the library and the common work philosophy, moral belief and spiritual outlook embodied in the thoughts and actions of all librarians.

The questions for interview for informant as follow:

1. Introduction session
2. Ask about ...
 - a. The library's long-established working methods
 - b. List of problems librarians faced at work;
 - c. List of events when knowledge communication among librarians needed.
 - d. List of events when knowledge communication between librarians and readers needed.
 - e. List of emergency events in the library.
 - f. List of coordination events between library and external environment.
 - g. List of library current service
 - h. List of library possible new service
 - i. List of library's problems
3. Closing: clarification about next step, ask for second interview, thank you

C. Develop Questionnaires

Based on the interview results, we made a closed answer questionnaire. Each question constructed from the abstraction of each question from CTI. For each question, we assign two to n number of sub questions according to CTI result. For each sub question, we provide answer to choose by the librarian. For example, from CTI, the answer from the question about "list of problems librarians' faced at work" could results in three problems, e.g. noise, broken books, and no returned book. For this question, questionnaire asked two question. First: rank the list from the least to the most frequency. Second: the intensity, asked the respondent to rank the list from the easiest to solve to hardest to solve. The same goes for other questions with different criteria to rank.

1. The library's long-established working methods (highly inefficient to highly efficient)
2. List of problems librarians faced at work (least frequent to most frequent)
3. List of problems librarians faced at work (easiest to hardest to solve)
4. List of events when knowledge communication among librarians needed (least frequent to most frequent)
5. List of events when knowledge communication among librarians needed (easiest to hardest to solve)
6. List of events when knowledge communication between librarians and readers needed (least frequent to most frequent)
7. List of events when knowledge communication between librarians and readers needed (easiest to hardest to solve)
8. List of emergency events in the library (least frequent to most frequent)

9. List of emergency events in the library (easiest to hardest to solve)
10. List of coordination events between library and external environment (least frequent to most frequent)
11. List of coordination events between library and external environment (easiest to hardest to solve)
12. List of library possible new service (least possible and most possible)
13. List of library's reputation problems (least frequent to most frequent)

D. Second Meeting

In this meeting, we conduct GTI to delve deeper into interviewee answer to the questions presented in the questionnaire. The interviewer respond the answer by ask deeper into the answer provided, by question such as "why this is the most efficient" → respond → "what do you mean by" → respond → what or how... → respond. If the interviewer understands the answer clearly, give a summary and ask for improvement: "Now I think I have understood what you mean about X. How do you feel about this reading?" After this, move to question two. Same procedure applied. Generally, each question asked by noting the extreme cases answered by participant (most or least) The list below listed about what question to asked after respondent answer a questionnaire question:

1. Introduction session: thank you, warn about the iterative character of the interview; informed consent explanation
2. Based on your questionnaire answer, you stated that:
 - i. The library's long-established working methods (most inefficient to most efficient) → why you think this is the most efficient (most inefficient)? → dig → dig → summary.
 - ii. List of problems librarians faced at work (never to always) → why you think this is the most/least frequent → dig → dig → summary.
 - iii. List of problems librarians faced at work (easiest to hardest to solve) → why you think this is the easiest/hardest problem to solve → dig → dig → summary.
 - iv. List of events when knowledge communication among librarians needed (least frequent to most frequent) → why you think this is the most/least frequent → dig → dig → summary.
 - v. List of events when knowledge communication among librarians needed (easiest to hardest to solve) → why you think this is the easiest/hardest problem to solve → dig → dig → summary.
 - vi. List of events when knowledge communication between librarians and readers needed (least frequent to most frequent) → why you think this is the most/least frequent → dig → dig → summary.
 - vii. List of events when knowledge communication between librarians and readers needed (easiest to hardest to solve) → why you think this is the easiest/hardest problem to solve → dig → dig → summary.
 - viii. List of emergency events in the library (least frequent to most frequent) → why you think this is the most/least frequent → dig → dig → summary.

- ix. List of emergency events in the library (easiest to hardest to solve) → why you think this is the easiest/hardest problem to solve → dig → dig → summary.
 - x. List of coordination events between library and external environment (least frequent to most frequent) → why you think this is the most/least frequent → dig → dig → summary.
 - xi. List of coordination events between library and external environment (easiest to hardest to solve) → why you think this is the easiest/hardest problem to solve → dig → dig → summary.
 - xii. List of library possible new service (least possible and most possible) → why you think this is the most/least possible → dig → dig → summary.
 - xiii. List of library's reputation problems (least frequent to most frequent) → why you think this is the most/least frequent → dig → dig → summary.
3. Ask the possibility that the interviewee have something to say but unable to say because hard to say without demonstration.
 4. If there any, ask permission to interview his/her peer about the action because observation from this peer could tell the difficult thing.
 5. If given permission, ask who he/she would recommend to interviewed.
 6. Ask for the last meeting.

IV. CONCLUSION AND FUTURE RESEARCH

This research presents data collection framework that enables the knowledge of the librarians to be simply captured and efficiently without requiring large resources. This makes this framework suitable for Business Intelligence because data can be integrated and explored quickly. The framework allows the sharp and pace detection of business problems and allows users to decide how to implement solutions in different contexts. The framework is simple and contains only three stages of data analysis. First is an analysis of structured interview data to be translated into questionnaires. Second, analysis of quantitative questionnaire data to be used as questions in capturing tacit knowledge. Third, data analysis uses a cognitive-based approach in the form of a cognitive map to transform tacit knowledge into explicit knowledge.

Our future research is about to develop the framework of data analysis using a cognitive-based approach in the form of a cognitive map after the completion of second interview. Then, the data be analyzed with cognitive mapping technique using Banxia Decision Explorer software. Data collection and analysis will produce basic material for the automation of systems that are relevant for the benefit of BI.

REFERENCES

- [1] C. Imhoff & C. White, "Collaborative BI: Theory Becomes Reality", *Business Intelligence Journal*, 18(2), 40 – 46, 2013.
- [2] K. Kuwashima, "How to Use Models of Organizational Decision Making?" *Annals of Business Administrative Science*, 13(4), 215-230 (2014), DOI:10.7880/abas.13.215, 2014.
- [3] D. Kahneman, *Thinking, Fast and Slow*: Farrar, Straus and Giroux. Macmillan Publishers, 2013.
- [4] F. A. M. Ducharme and L. A. Angelleli, "Cognitive Analytics: A Step Towards Tacit Knowledge?" *Systemics, Cybernetics and Informatics*, 12(4), 32 – 38, 2014.
- [5] Cox, B. L. & Jantti, M. H., "Capturing business intelligence required for targeted marketing, demonstrating value, and driving process improvement", *Library & information science research*, vol. 34, no. 4, pp. 308-316, 2012.
- [6] A. J. Helms, K. D. Bradford, N. J. Warren, D. G. Schwartz, "Bioinformatics opportunities for health sciences librarians and information professionals" *MLA '03, the 103rd Annual Meeting of the Medical Library Association*; San Diego, California; May 5, 2003, *J Med Libr Assoc* 92(4) October 2004.
- [7] E. Cynthia, D Jason, M. Jennifer, K. Sandra, Jackson Sr., J. John, "Collaborate and Innovate: The Impact of Academic Librarians on the Commercialization of University Technology", 2016, 57 (1):36 *Journal of Library Administration*, DOI: 10.1080/01930826.2016.1215674
- [8] J. Philippe, "Business Intelligence: A New Challenge for Librarians?", Conference: « *Spezialbibliotheken heute – Wettbewerb und Kooperation* » 06-09.03.2001 Hannover.
- [9] M. Zappavigna and J. Patrick, "Eliciting tacit knowledge about requirement analysis with a Grammar-targeted Interview Method (GIM)", *European Journal of Information Systems*, 19:1, 49-59, 2010, DOI: 10.1057/ejis.2010.1
- [10] K. U. Koskinen and H. Vanharanta, "The role of tacit knowledge in innovation processes of small technology companies", *Int. J. Production Economics* 80 (2002) 57–64.
- [11] J. Francois, "The Stance of Systemic Functional Linguistics Amongst Functionalist Theories of Language and Its 'Systemic' Purpose", *Perspectives from Systemic Functional Linguistics Journal*, 2018.
- [12] Z. Pourzolfaghar, R. Ibrahim, R. Abdullah, and N. M. Adam, "A Technique to Capture Multi-Disciplinary Tacit Knowledge During the Conceptual Design Phase of a Building Project", *Journal of Information & Knowledge Management*, Vol. 13, No. 2 (2014) 1450013, DOI: 10.1142/S0219649214500130.
- [13] B. Marshall, P. Cardon, A. Poddar, and R. Fontenot, "Does Sample Size Matter in Qualitative Research?: A Review of Qualitative Interviews in IS Research", *Journal of Computer Information Systems*, Fall 2013.
- [14] R. K. Yin, "Case Study Research and Application Design and Methods", *Case Study Research and Applications*, Sixth Edition, SAGE Publications Inc., 2018, LCCN 2017040835 | ISBN 9781506336169 (pbk. : alk. paper) Subjects: LCSH: Case method. | Social sciences— Research—Methodology.
- [15] J. W. Creswell, "Qualitative Inquiry & Research Design Choosing Among Five Approaches", Sage Publications, Inc., 2nd ed. ISBN 978-1-4129-1606-6 (cloth), ISBN 978-1-4129-1607-3 (pbk.), 2007.
- [16] S. Si and Y. Yujia, "Research on Tacit Knowledge Management Strategies in Library Information Consulting Service", *Advances in Economics, Business and Management Research (AEBMR)*, volume 23, 2017 2nd International Conference on Politics, Economics and Law (ICPEL 2017).
- [17] H. Surbakti and A. Ta'a, "Tacit Knowledge for Business Intelligence Framework: A Part of Unstructured Data?," *J. Theor. Appl. Inf. Technol.*, vol. 96, no. 3, pp. 616–625, 2018.