

# By Adjusting Keyword Frequency and Separating Keyword into Different Layers to Improve Readability of KeyGraph in Decision Making

Fang-Cheng Hsu<sup>1</sup> Yen-Chih Wang<sup>2</sup>

<sup>1</sup>Department of Information Management, Aletheia University

<sup>2</sup>Graduate School of Management Sciences, Aletheia University

## Abstract

One drawback of KeyGraph is its poor readability. Providing qualified documents are expected helpful for KeyGraph to translate the documents an appropriate and easy read KeyGraph diagram. And further, to discover worthy chances for decision making. We proposed 2 preprocessing strategies for assuring appropriate quality of documents, especially for decision makers when they try to discover chances with KeyGraph. A case study of INSTANT MESSENGER is used to verify the validity and readability of the proposed strategies. The results shown that the proposed strategies are helpful for creating readability KeyGraph diagrams, and further to create clear scenarios for discovering chances.

**Keywords:** KeyGraph, Chance discovery, Value focused think

## 1. Introduction

For introducing computer capabilities into decision making, Simon [7] proposed that decision making can be structured into three major phases, i.e., *intelligence*, *design* and *choice*. The intelligence phase involves searching or scanning the environment for calling for decisions; the design phase involves inventing, developing, and analyzing a set of possible decision alternatives for the problem identified in the intelligence-phase; the choice-phase involves selecting a particular decision alternative from those available.

The KeyGraph [5] has been used to discover decision opportunities hidden in documents for years [6]. It is very important to assure the inputs of KeyGraph are unbiased. To assure the unbiased inputs, suitable preprocessing strategies were expected. The general documents preprocessing consists of two tasks [2]: (1) Document compaction: Use word stems to reduce related words to the same root. For example, words like \create", \creation", and \creating" are

reduced to \create"; (2) Phrase construction: Let longer phrases with higher frequency. Except that, Sakakibara and Ohsawa developed other preprocessing methods to generate more simplified KeyGraph [8].

Providing qualified sentences is helpful for KeyGraph to translate the input sentences into an appropriate and easy read graph, and further, to discover worthy chances for decision making. However, when we adopted focus group [9] collecting narratives and dialogues of participants about what were the factors effecting their decision on how to use INSTANT MESSENGER (IM) and their concepts/values of IM. We finally got a great deal of sentences during focus group sessions and found 3 critical factors might cause the collected words biased. Factor A: some participants used to express their concepts or ideas with different words; Factor B: expressing one thing with different layer's words; Factor C: some of the participants are talkative so that their opinions can fully be expressed; while some are not. Hence, the frequency of keywords in verbatim will be biased, we denominated this is "Term Frequency Bias". The biased sentence might affect the inputs of KeyGraph and results in biased KeyGraph Diagrams (KGD).

For factor A, the document compaction is good enough for dismissing the biases. As to the factor B, based on Keeney's Value-focused thinking (VFT) [3] decision makers who describe their fundamental value with the fundamental objective-layer's words, and describe alternatives with the alternative-layer's words. We hence proposed a preprocessing strategy, separating keyword into 2 different layers by following the VFT procedure, to avoid the biases caused by the factor B. For solving the biased problems caused by factor C, we suggested the other preprocessing strategy, adjusting the frequency of keywords by a survey procedure, to avoid the biases.

Finally, a case study of INSTANT MESSENGER was used to verify the validity and readability of the proposed preprocessing strategies.

## 2. Background

Keeney [2] proposed that values are principles for evaluating the desirability of any possible alternatives. It means that values are fundamentally important in any decision situation; and more fundamental than alternative. He emphasized that value should be the driving force for decision making. Thus, although it is useful to iterate between articulating values and creating alternatives, the principle should be value first. This manner of thinking refers to Value-focused thinking.

One of the most important objectives of this study is to discover chances, which might be hidden in users' mind or their usage behaviors, of IM for decision makers who respond to the developments of next generation IM.

KeyGraph has two important parts: (1) the associated frequency between keywords, (2) key value: refers to importance and contributions of a keyword on the structure of documents. KeyGraph has been applied to a variety of topics. Such as: finding areas with the highest risks of near-future earthquakes [6], discovering emerging topics from WWW [4].

In this paper, the procedure of original chance discovery was divided into 3.

**Step1. Focus Group:** Encourage participants talking about the factors affecting their decisions on how to use IM and their concepts/values of IM. A video camera was used to record their talks through entire sessions. The WITI (What Is That Important?) test [1] was used to distinguish words of value from words of alternatives.

**Step2. Generating verbatim:** Took out the sentences recorded with the video camera and transformed them into verbatim scripts.

**Step3. KeyGraph Analysis:** A document, in this study is a verbatim script, is composed of sentences, and a sentence is composed of words. The KeyGraph analysis consists of 2 sub-steps: one is the general document preprocessing strategies: document compaction and phrase construction. The other sub-step is the process of generating an original KGD.

## 3. The Proposed Strategies

### 3.1. KGD-AKF: Adjusting Keyword Frequency

It is very difficult to avoid the "Term Frequency Bias" when using techniques of narrative or dialogue to

collect information. For preventing the bias, an adjusting procedure called KeyGraph Diagram – Adjusting Keyword Frequency (KGD-AKF) was proposed. The adjusting procedure has 4 steps:

**Step1 and step2:** same with the original procedure.

**Step3 Adjusting Frequency:** Requesting the participants of focus group to check the arranged sentences and verbatim scripts. After that, we develop a 110-item questionnaire based on the validated verbatim and deploy the questionnaire on a web page. We mixed convenience sampling and **snowball sampling** [10] techniques to develop a research sample (210 subjects; all the subjects has at least 1 year experience in using IM).

**Step4. KeyGraph Analysis:** The same with step3 of Original Procedure.

### 3.2. KGD-VFT: Separating Keyword into Two Layers

Based on the concept of Value Focused Thinking, we proposed the other procedure, KGD-VFT, for improving the readabilities of KeyGraph Diagrams. The procedure has 5 steps:

**Step1, Step2 and step3** are the same with the KGD-AKF procedure,

**Step4. Separating Keywords into 2 Layers:**

**Step4.1** Coding form of questionnaire items

Each item of the 110-item questionnaire is composed of several keywords, and the keywords can be categorized into Two different sets: (Value-layer, Alternative-layer.) Each item was coded with a form of (KW#1, 2-tuple value; KW#2, 2-tuple value; ...; KW#i, 2-tuple value); KW#i means the item has i Keywords and the 2-tuple value will be 0 (belongs to Value-layer), 1 (belongs to Alternative-layer). We have three coders to discuss and decide which word should be categorized into which set. It means that the words in an item might be categorized into 2 sets, or even only one set. An item is coded as (KW#1, 1; KW#2, 0; ... KW#7, 0) means that the item has 7 keywords, and the first keyword belongs to Alternative-layer; the second keyword belongs to Value-layer; ..., and the 7th keyword belongs to Value-layer.

**Step4.2** Generating Sentences and Documents

Collecting all questionnaire items was picked up by each subject and categorizing every keyword into different set based on its triple-code value. Every picked up item by a subject will generate at most 3 sentences (pure alternative, pure value, and constituted by alternative and value). Finally, all of the sentences generated by all subjects were integrated into a document. For example, If subject x agree with item

#1, #3, #7, and the item #1 = (“a”, 1; “b”, 1); item #3 = (“a”, 1; “c”, 1; “d”, 0; “b”, 1); item #7 = (“b”, 1; “c”, 1). #1 and #3 to belong topic 1, #7 to belong topic 2 (topic is defined in focus group participant.) The item #1 and #3 will generate 3 sentences: s11 = “a, b, a, c, b”, s01 = “d”, s21= “a, c, b, d”. The item #7 will generate 1 sentence: s12 = “b, c”.

#### 4. KeyGraph Diagrams and Scenarios

The objective of this research is compared to the readabilities of KeyGraphs created by three different kinds of chance discovery procedures. Three procedures are: *Original KGD Procedure*, *KGD-AKF Procedure*, and *KGD-VFT Procedure*.

Anyone who was response to create scenarios should be asked to understand main features about IM and what users’ purposes are. The main features of IM including: “chat online”, “send instant messages”, “photo sharing & file sharing”, “voice or even video conversation”, “conversations when you can’t be there in person”, etc., The main reasons of using IM including: “to make contact with friends and families”, “to discuss business”, “to deliver the file”, and so on.

##### 4.1. The Scenario of Original KGD

Fig 1 is the KGD based on the original KGD procedure. The scenarios were created such as “If someone has my e-mail account he can send instant message to me,” “Exchange text messages with multiple people in one IM through conference,” “The great quality of Voice is important when I use Voice function to talk with others and sing a song”

##### 4.2. The Scenario of KGD-AKF

Fig 2 is the KGD generated by KGD-AFT. The results show that the subjects created some richer scenarios with KGD-AFT than with original KGD. The scenarios such as “By using conference function, users could exchange text messages with multiple people in one IM,” “The high quality of voice is important when I used Voice function to talk with others and sing a song,” and “Getting someone’s attention by pictures which need interesting, exquisite, and animated.” Besides, some decision opportunities (or chance, Ohsawa, 1998) were easily found on this KGD, include: Skype, curious, and novel. We can create scenarios accordingly. The cause of users starting to use IM is they feel IM is novel and they are curious about what is chat.

Especially they believe IM is helpful for increasing interpersonal relationships.



Fig. 1 KGD based on the original procedure

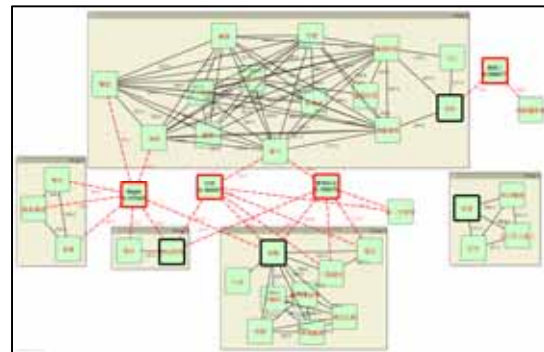


Fig. 2 KGD based on the KGD-AKF procedure

##### 4.3. The Scenario of KGD-VFT

Following the KGD-VFT procedure, three KGDs, Alternative-layer KGD, Value-layer KGD, and Link KGD, will be generated.

The principles of creating scenarios on different layers are: 1. Scenarios on Alternative-layer will focus on what functions are user’s preferences and what features are use’s cares. 2. Scenarios on Value-layer will focus on why users use IM and what is the user’s value. 3. Scenarios on Link-layer should focus on the relationships of values and alternatives.

The three KGDs can be analyzed separately or simultaneously. The followings scenarios were created by following the above principles:

The functions and features which IM users preferences were “send instant message” , “Voice function “, “self-express picture”, “Share file”, etc., (Fig. 3).

Why users like to use IM? Something interesting was found on the Value-layer KGD (Fig.4). The convenience and interest were the reasons caused users used IM. It’s nothing special. But on the same KGD, we found a chance node “perfect life” connects four nodes, interests, convenience, achievement, and

satisfaction life, which belong to four different clusters separately. It means that we can draw a picture: some day people will have a perfect life with IM. What we should do is to share the good experience among the four parties of users. Of course, the chance also provides a good idea to users or IM developers to catch the un-revealed values of IM.

Fig.5 shows what functions were associated with values. In this KGD we can see the value “interest” which was associated with “animated self-express picture” and “chat”. In addition, we can see convenience is associated with “send instant message”, “share file”, “communication with multiple people in the same time”, and so on.

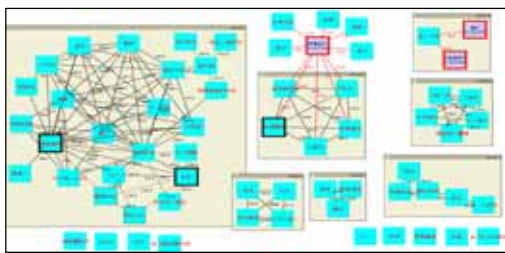


Fig. 3 KGD on Alternative-layers (KGD-VFT)

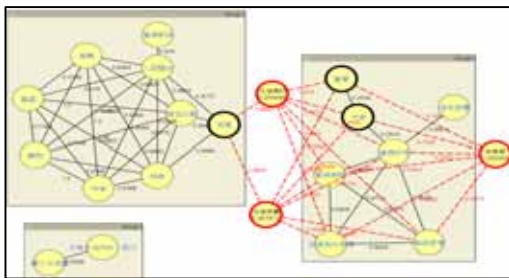


Fig. 4 KGD on Value-layers (KGD-VFT)

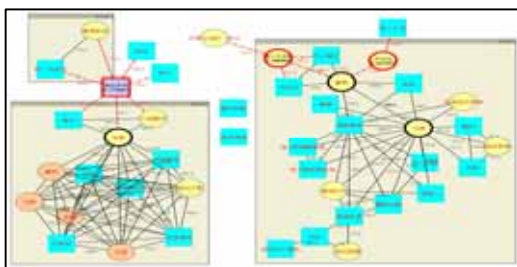


Fig. 5 KGD on Link-layers (KGD-VFT)

## 5. Results and Conclusions

In the paper, two preprocessing strategies were proposed and an instrument was developed to verify the model validity of KGD which created by different procedures. The instrument assessed validities and readabilities by a 3-item questionnaire. The items were on a five-point Likert type scale from 1(total disagree)

to 5 (total agree). Two experts of KeyGraph were invited for evaluating the content validity of the questionnaire.

To avoid the subjects who were not familiar with IM and KeyGraph, we chose the subjects who had at least two years experience on using the IM. The Kruskal-Wallis one-way analysis of variance was used to test the significances.

The experimental results shown that KGD-VFT procedure have high readability, so that it can be used to create appropriate scenarios for discovering chances. In addition, we found the context is full of creativity and is expected to be able to provide decision maker a more creatively decision situation to help them make an appropriate and creatively decision.

## 6. References

- [1] R. T. Clemen, and Reilly T, Making Hard Decisions with Decision Tools, Duxbury Press, 2001.
- [2] D. E. Goldberg, Sastry, K. and Ohsawa, Y. “Discovering Deep Building Blocks for Competent Genetic Algorithms Using Chance Discovery via KeyGraphs,” IlliGAL Report No. 2002026, December, 2002.
- [3] R. L. Keeney, “Creativity in Decision Making with Value-Focused Think,” Sloan Management Review, pp.33-41, summer 1994.
- [4] M. Naohiru, Matsuo, Y., Ohsawa, Y. and Ishizuka, M. “Discovering Emerging Topics from WWW,” Journal of Contingencies and Crisis Management, Vol. 10, No. 2, pp. 73-81, 2002.
- [5] Y. Ohsawa, “Chance Discoveries for Making Decisions in Complex Real World,” New Generation Computing, Vol.20, No.2, Springer Verlag and Ohmsha, 2002.
- [6] Y. Ohsawa, and Yachida, M. “Discovery Risky Active Faults by Indexing an Earthquake Sequence,” Proceedings of Discovery Science, Springer Verlag, Heidelberg, pp.208-219, 1999.
- [7] A. Simon, New Science of Management Decision, NY: Harper and Row, New York, 1960.
- [8] T. Sakakibara and Ohsawa, Y. “Generation of Multiple KeyGraphs with Different Complexity from One Data Set,” in Proceedings of the 4th International Workshop on Chance Discovery: From Data Interaction to Scenario Creation (CDWS4), Germany, August, 2005.
- [9] [http://en.wikipedia.org/wiki/Focus\\_group](http://en.wikipedia.org/wiki/Focus_group), 2006.
- [10] D. D. Heckathorn, “Respondent-Driven Sampling: A New Approach to the Study of Hidden Populations,” Social Problems, Vol. 44, pp.174-193, 1997.