Extracting Experience through Protocol Analysis

P C Matthews[*] S Ahmed M Aurisicchio

November 29, 2001

Abstract

Protocol Analysis provides a means for extracting a person's thoughts while they are performing a task. This paper used this method for generating an index of keywords based on the transcript generated through protocol analysis of an engineer working on a design project. An algorithm for processing such a body of text is described. The means for analyzing the index to form a network of keywords is introduced through the use of an illustration. The illustration is taken from a designer working on a new compressor for a gas turbine engine. The use of such an index is compared to similar indices generated using formal documentation, and the impact this would have on knowledge management systems is discussed.

1 Introduction

In the field of mechanical engineering, the act of designing represents the transformation of a requirement into the physical object. The design process gives the framework necessary for the transformation of an engineering need to a product that fulfills that original need. Pahl and Beitz (1996) consider this as a conversion of information, starting with a qualitative need (the requirement or task) and ending with a quantitative result (the detailed design). This transformation is achieved through the application of scientific knowledge and experience to the

[*]Engineering Design Center, Cambridge University Engineering Department, Cambridge, UK. email pm131@eng.cam.ac.uk

problem while working within the constraints of the task requirements and available technology. There is a trend for increased complexity of products, technological improvements and more demanding requirements while, at the same time, product life-cycles are reducing and there is a need for faster responses (Blessing, 1994).

Prescriptive methods in engineering design have largely been developed based on the experiences of their authors, and not on observed activities of designers and some validated theory. In general, design methods were not developed with reference to cognitive models of how designers carry out design work. Hence, design methods provide an overall approach for the design process, but do not supply designers with strategies to design at a problem solving level. For example, methods prescribe that designers carryout task clarification or detailed design work, but they do not suggest how the designers should clarify tasks or generated detailed design work. In addition, very little work has been carried out to evaluate the validity of these methods.

Recent studies suggest that designers rely heavily on their memories and experiences and that research is required to understand how this knowledge and experience are used (Marsh, 1997). In addition, industry has recognized the need to capture 'experience'. In order to capture experience, the nature and use of design experience within industry need to be understood more clearly. Until recently, research into the differences between novices and experts has focussed on problems where constraints and contexts are well defined, and a limited number of rules apply, for example, the game of chess and domain problems such as those of mechanics. Design is frequently described as an ill-defined problem. Usually, many possible solutions exist and there are no defined rules to obtain these solutions. Cross and Cross (1997) describe the limitations in drawing too many conclusions from studies of expertise in disciplines such as psychology as the problems traditionally dealt with are well-defined. The relationship between well-defined problem areas and the ill-defined area of design is poorly understood, and hence the applicability of findings from studies in will-defined areas to design is not clear.

This position paper will explore the use of data mining on transcripts taken of various designers have been working while verbally expressing their actions (protocol analysis). The aim is to identify the relationships between key words and issues from each transcript. Primarily, this would be used for identifying useful indices for formal documentation and for the prompting to designers for generating this documentation. Ultimately, this is to be used as a means of computationally identifying the differences between novice and expert designers. Explicit awareness of this difference will help in accelerating the learning process for novice designers.

The remainder of this paper is structured as follows: an overview to protocol analysis is given in Section 2. This is followed by the description of the transcript processing method in Section 3. This is illustrated with a case study in Section 4. The relationship this method has to knowledge management is discussed in Section 5, and finally Section 6 concludes the paper.

2 Protocol Analysis

Protocol analysis are now becoming increasingly common in design research and have been used as a method of research in the social sciences for many years. They are undertaken in real time and, if carried out in their natural environment, can capture the context of the events (Yin, 1994). Observing in the environment of the designers has the advantage of being a real-life situation, however it is less controlled than a laboratory experiment. As protocols capture data in real time, participant bias, common with retrospective methods such as interviews and questionnaires, is avoided, but the observer's bias may have some effect (Stauffer et al., 1991; Frankfort-Nachmias and Nachmias, 1996; Orne, 1969; Rosenthal, 1966). Protocols are usually carried out in a subject's own environment and usually teams are observed rather than individuals. The researcher may make notes whilst observing the subject, and video or audio-record. The participants are requested to think aloud whilst being observed. Observations, together with thinking aloud, can only provide a short, but detailed, insight into the activities of designers. They do not allow a whole picture of the design process to be formed, as they only captured episodes of design activity. Protocol analysis can be carried out for longer, however they are time consuming. For each hour of protocol, an additional 25 hours may be required 12 hours for planning the observations and transcription and analysis (Ahmed, 2001).

The effects of thinking aloud on the behavior of the designers are considered to be negligible if participants are only prompted to speak during periods of silence. Ericsson and Simon (1993) found no reliable evidence that structural changes occurred to the cognitive processing if subjects were simply asked to verbalize their thoughts, however, they found changes when subjects were asked to explain their cognitive processes.

Orne (1969), and others, state that subjects (participants) can never be neutral to an experiment. The effect of being observed may alter the behavior of the designers. Participants may try and be good subjects by trying to produce the data that they think is desired from them (Rosenthal, 1966). To reduce the effect of subject bias, it can be made clear that the design task and the designer will not be evaluated. Also, if all the participants are in the same situation, any measure is relative. The presence of an experimenter (in this case the researcher) may also effect the behavior of the designers (Rosenthal, 1963). Experimenter bias may influence the outcome of an experiment if the experimenter unintentionally communicates his or her expectations of the results to participants (Rosenthal, 1963). The experimenter may bias the experiment by verbal communication or gestures, e.g. nodding when an expected result is observed. As there is no expectation of the results of the protocols, there is no experimenter bias at the start of the protocols. As the earlier protocols are analyzed, care can be taken with further protocols to ensure that the expected results are not communicated to the participants.

The use of an audio-recorder may also affect the experiment, as the participants may alter their behavior if they are aware that they are being recorded. Permission will be asked from each participant to use the audio-recorder. The audio-recorder will be placed on the table in an attempt to be as discrete as possible.

3 Transcript Processing

The transcript processing is based on a predetermined set of keywords. The aim is to generate a relationship network based on these keywords, that will provide a means for indexing such design work in general. This index will be based on the observed actions a designer takes, as opposed to one generated based on the formal (electronically recorded) documentation.

A metric was constructed to measure similarity between pairs of keywords based on their co-occurrence within the transcript. This metric was computed by analyzing each sentence in the transcript. Each time a pair of keywords co-occurred within a sentence, their similarity score was increased by 5 points and if the two keywords co-occurred in adjacent sentences, their similarity score was increased by 2 point. After all the sentences had been processed, the resulting similarity matrix was linearly normalized.

Based on this similarity matrix, the keywords were clustered. A non-hierarchical clustering method was used for this (Jardine and Sibson, 1968). This non-hierarchical method was selected as keywords will not necessarily fall into a single class (Spärck Jones, 1965; Needham, 1965). The classes that are identified with this algorithm will share certain keywords. This is used to generate the network links between the identified groups of words.

The word network is constructed by analyzing the clustering results. From here, the manner in which keywords are related reveals dependencies within the design domain. These dependencies represent related issues that are addressed by the designer being studied. Further, these dependencies represent the designer's movements across formal documentation and reveal informal networks the designer uses while working on the project. Such knowledge can then be used to augment the formal (recorded) knowledge a company holds.

4 Case Study

This approach is illustrated using a transcript excerpt taken from a real design project. The project involves designing a new compressor for a gas turbine aero-engine. The transcript is broken down into individual sentences (see Appendix A). A set of keywords taken from the transcript (listed in Appendix B) are then used to generate the similarity matrix. Based on this, the following clusters were identified using the non-hierarchical clustering method:

G1 `aerofoil`

G2 `axial, blade, blades, blisc, chord, design, disc, numbers, root, speed, thickness, weight`

G3 `compressor, front, material, nickel, temperature, titanium`

G4 `data`

G5 `design, disc, load, materials, numbers, rim, root, speed, thickness, weight`

G6 `discs, titanium`

G7 `estimate`

G8 `alice, notes`

G9 `numbers, root, schemes, speed, thickness, weight`

G10 `alice, weight`

G11 `rate`

G12 `scheme`

For illustration purposes, the network rooted at `alice` will be considered. `alice` is directly involved in two groups: G8 and G10. G8 contains {`alice, notes`} and G10 contains {`alice, weight`}. This provides a link between `weight` (an important design issue) and `notes` (an informal source of knowledge) via `alice`, a designer. From the transcript (lines 14–18), we can infer that Alice has generated the weight notes, and potentially also be a useful person to talk to in the event that further knowledge is necessary.

Figure 1 illustrates how this network can be further built up. Here, the tree has been rooted on `weight`. The root node contains the intersection of groups 2, 5, 9, and 10. The children of this node contain the elements in the difference between this intersection and the group represented by each child node. This tree can be used in the following manner: a designer is considering the `weight` issue for a project. The root node of the tree provides the most commonly related elements, while each child node provides more specific relationships. For example, `schemes` (G9) represents previous designs that could be referred to, and these are the formal documentation of previous projects. The leftmost child represents further related issues that must be considered in the project, and these are further subdivided into two different major groups (G2 and G5). Finally, the

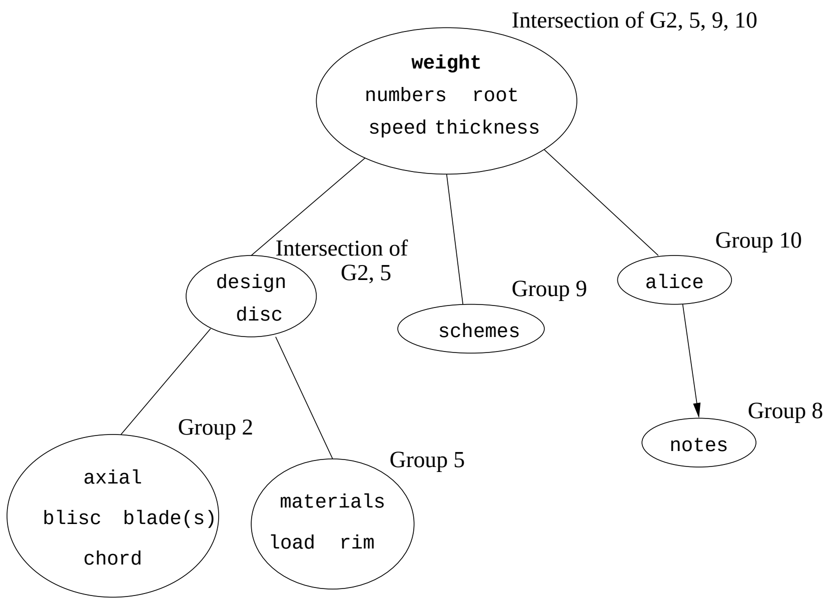

Figure 1: Word network rooted on `weight`

rightmost child node represents informal knowledge sources, namely, Alice and her notes.

5 Impact to Knowledge Management

This form of analysis impacts the corporate knowledge management by augmenting it with an index generated through the analysis of informal knowledge sources, namely, human designers at work. The difference between this and an index based strictly on formal documentation is that here links are made between concepts as the designer works through the project. This includes references to informal knowledge sources (e.g. Alice in lines 15 and 17) and actions performed (e.g. lines 2, 5, 11, and 31). Further, a designer using this index could not only retrieve the relevant formal documentation, but also be given the relevant lines from the transcript. These transcript lines will indicate previous chains of action that were performed in similar situations, thereby assisting the designer with the current project.

6 Conclusions

This analysis method provides an interesting means of associating key concepts based on how humans work through a project as opposed to what is formally recorded. There are two further experiments that can be conducted from this method. Firstly, it would be interesting to compare the index generated from this method to an index generated based strictly on searchable formal documents. The differences between these two indices will highlight the knowledge that is not currently being captured formally, and potentially lead to a methodology for capturing this knowledge. Secondly, indices generated from transcripts from two different designers are not necessarily going to be the same. In particular, comparing the indices generated from an experienced designer and an inexperienced designer will provide an indication to how the inexperienced designer can gain experience more rapidly.

The main difficulty with generating such an index is the cost involved. The process of observing a designer in action (who must unnaturally speak all thoughts aloud) through to the generation of the text transcript takes a large amount of time. Therefore, this is not a process that can take place with every project, as opposed to the generation of formal documentation. However, this should not be a reason not to pursue such lines of analysis, as speech recognition software improves, such experiments could be performed more frequently and potentially be incorporated as part of the documentation process in areas where several similar projects are undertaken. In such projects, a fraction of the designs would undergo such analysis, and this would serve as an 'informal' information source.

Acknowledgements

This work is fully funded by the *University Technology Partnership*, a collaborative research project between the universities of Cambridge, Sheffield and Southampton; and with industrial partners BAE SYSTEMS and Rolls-Royce.

References

Ahmed, S. (2001). *Understanding the Use and Reuse of Experience in Engineering Design*, PhD thesis, Cambridge University Engineering Department.

Blessing, L. T. M. (1994). *A Process-Based Approach to Computer-Supported Engineering Design*, PhD thesis, University of Twente.

Cross, N. and Cross, A. C. (1997). *Designers: The Key to Successful Product Development*, Springer-Verlag, Darmstadt, chapter Expert Designers, pp. 1–16.

Ericsson, K. A. and Simon, H. A. (1993). *Protocol Analysis: Verbal Reports as Data*, MIT Press, Cambridge, MA.

Frankfort-Nachmias, C. and Nachmias, D. (1996). *Research Methods in the Social Sciences*, Arnold, London.

Jardine, N. and Sibson, R. (1968). The construction of hierarchic and non-hierarchic classifications, *The Computer Journal* **11**(2): 177–184.

Marsh, J. R. (1997). *The Capture and Utilisation of Experience in Engineering Design*, PhD thesis, Cambridge University Engineering Department.

Needham, R. M. (1965). Application of the theory of clumps, *Mechanical Translation* **8**(3–4): 113–127.

Orne, M. T. (1969). *Artifact In Behavioral Research*, Academic Press, New York, NY, chapter Demand Characteristics and the Concept of Quasi-Controls, pp. 143–179.

Pahl, G. and Beitz, W. (1996). *Engineering Design: A Systematic Approach*, second edn, Springer-Verlag London.

Rosenthal, R. (1963). The effects of early data returns on data subsequently obtained by outcome-biased experimenter, *Sociometry* **26**(4): 487–493.

Rosenthal, R. (1966). *Experimenter Effects in Behavioral Research*, Appleton Century Crofts, New York, NY.

Spärck Jones, K. (1965). Experiments in semantic classification, *Mechanical Translation* **8**(3–4): 97–112.

Stauffer, L., Diteman, M. and Hyde, R. (1991). Eliciting and analysing subjective data about engineering design, *Engineering Design* **2**(4): 351–366.

Yin, R. K. (1994). *Case Study Research*, SAGE Publications, London.

A Transcript text

1. What I'm trying to do is the disc that hold the blades, this is a stage one blades, typically that would have a blade root and the blade root fixing in the disc, and the disc platform beneath that.

2. Because it's so high speed, last week come to this conclusion that we definitely need a blisc arrangement here.

3. A blisc is a bladed disc.

4. So I's trying to do here, watching this design of the disc set some blisc.

5. First thing is what material I am looking at, as it is temperature related, and I know that the front end of the compressor allows using titanium.

6. Half way to the compressor we have to swap to probably to the nickel, just for the temperature reason.

7. But I know that the front end is happy with the titanium.

8. The reason I know is that we have already got titanium discs at service at the moment.

9. So I'm going to get some materials.

10. What I need to do is we have design program that will design disc for me.

11. I need to give a rim load, which is the weight of all the blades, at that particular rpm.

12. So that we will work out all the rate of the blade is.

13. So what actually need is the aerofoil thickness.

14. I can get the weight of an existing blade from the weight department.

15. I know that Alice has got one.
16. (Get up and gets notes).
17. Alice has been working on the same thing.
18. So what I want to get out of here, is the weight of the aerofoil.
19. So what's the quickest way to get the weight of an aerofoil? Now, look at this.
20. He might have even worked out that aerofoil.
21. (Look through notes).
22. I'm just looking for data, not very exciting stuff.
23. So can't think anything to say for the moment.
24. This is the scheme numbers.
25. I could go and get those schemes out, but I'm not going to.
26. I have to get these numbers which appears there first but I think I'll be good enough for what I want.
27. Here I got a section of a blade, it has a chord and an axial chord on it, 42 blades, and it got a blade thickness.
28. It has a maximum thickness and a blade height.
29. I can relate that thing.
30. I'll measure this blade that I'm looking at, it's 62-mm height and as the same width, and 36 mm wide compare to 42.
31. So I'm going to do is estimate how much bigger this blade is than mine.

B Keywords

aerofoil, alice, axial, blade, blades, blisc, chord, compressor, data, design, disc, discs, estimate, front, load, material, materials, nickel, notes, numbers, rate, rim, root, scheme, schemes, speed, temperature, thickness, titanium, weight.