INTERNATIONAL CONFERENCE ON ENGINEERING DESIGN
ICED 03 STOCKHOLM, AUGUST 19-21, 2003

STRUCTURING ENGINEERING DESIGN INFORMATION USING A MODEL OF HOW ENGINEERS' INTUITIVELY STRUCTURE DESIGN INFORMATION

Russell Japikse, Patrick Langdon, Ken M. Wallace

Abstract

This paper details a research programme that has established a better understanding of how to identify conceptually similar pieces of design information. This research investigates the conceptual framework that engineers intuitively place design information into. Results from a pilot study and preliminary results from the main study have been evaluated. A map of how concepts are related within a specific engineering domain has been created, as well as a basic understanding of how designers intuitively process information.

Keywords: information structuring, design guidelines, cluster analysis, conceptual similarity

1 Introduction

When engineering design information is stored for future retrieval and use, some sort of organisational structure is usually applied. Structuring permits increased retrieval precision and ease in obtaining the desired information [1]. Structuring is used in both linking one section of information to another, e.g. hypertext links, and for creating a framework from which similarity values may be obtained. The ability to generate meaningful similarity values is crucial to electronic information searches [2]. Previous methods of structuring information have varied greatly. A few recent approaches have based the structure upon behaviour characteristics observed in designers [3], [4]. A possible approach is to structure design information in the same manner that engineers intuitively use when conceptualising design information. This should allow for the searching and learning of design information to more closely match the cognitive preferences of designers. To this end, the following hypothesis has been posed: structuring information the same way that designers intuitively structure knowledge will provide for faster and more accurate information access. This hypothesis leads to a several follow up questions:

1. Do engineers order their knowledge in a recognisable pattern?
2. If information is organised in a non-random fashion, is this behaviour consistent amongst engineers?
3. If information relations are non-random, can a model be developed that allows information to be appropriately classified?

2 Exploratory Studies

In order to explore the above hypothesis, two exploratory studies have been carried out in cooperation with Rolls-Royce and BAE SYSTEMS. Both studies have used empirical methods of observation in an attempt to discern how design knowledge is conceptualised. In the first study, designers directly ranked the level of perceived similarity between twenty guidelines. This study used a set of guidelines representing basic engineering design knowledge. This was a broad study that was intended to validate the research method, to answer the first and second research questions, and to contribute basic knowledge to a follow on study. The second study was much more comprehensive. In it, the conceptual similarities of seventy-five pieces of design information were indirectly ranked. This design information is much more detailed and represents key aspects of wing fuel tank design. This design information was supplied by one of our industrial partners.

3 Pilot Study - Overview

In the first study, twelve designers from BAE SYSTEMS and Rolls Royce completed a guideline similarity ranking exercise. This exercise measured the conceptual distance between each guideline. From this set of distances, the cognitive structure may be inferred. For this exercise, the subjective similarity of every unique combination of guidelines was ranked. The only criteria was, how similar is guideline *a* to guideline *b*? This allowed the study to be conducted without any presuppositions about possible relationships between the guidelines. Cluster analysis and multidimensional scaling (MDS) were used for the exploratory data analysis. This data analysis was integral to understanding the structure, and presenting it in a useful form for subjective analysis.

3.1 Pilot Study –Parameters

The guidelines in this study were obtained from a set of 3500 guidelines contained in the Cambridge Engineering Designer (CED). The CED is a database of guidelines meant to augment decisions made during the conceptual phase of engineering design. Since the participants all had varying backgrounds and experience within aerospace design, the use of basic design guidelines reduced the possibility of field specific experience skewing the results.

3.2 Pilot Study – Data Processing for a Common Group Solution

The raw similarity data formed a non-metric set of distances defining the relative positions of the guidelines in a space of n-1 dimensions. The matrices were processed and combined to a single solution using non-metric MDS (Proxscal) in SPSS. The data was then scaled to two, three, four, and five dimensions. This is the valid range for this data set [5]. The stress values were then plotted against the number of dimensions. In MDS, stress is a measure of the amount of numerical distortion that occurs when a set of points is fitted to specific number of dimensions. Figure 3-1 displays the typical progression of decreasing stress as the number of dimensions increase. While the number of dimensions to use is subjective, there is some statistical merit in selecting the point where there is an elbow in the curve [6]. This occurs at three or four dimensions. As four dimensions have a lower level of stress than three dimensions, this value was chosen for further data processing.

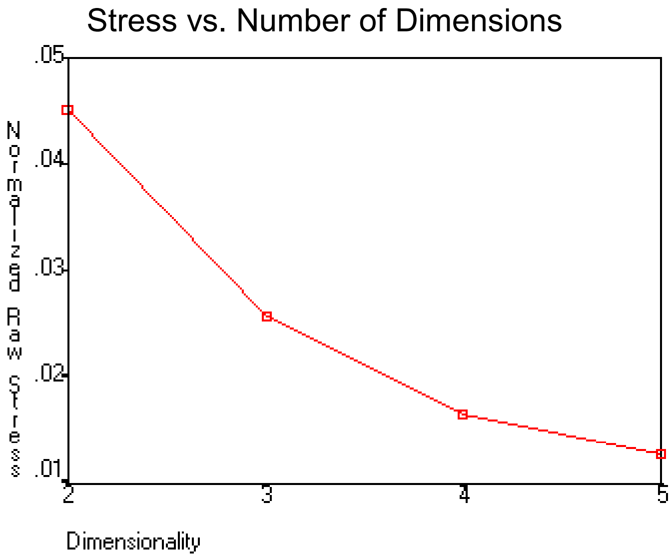


Figure 3-1 Pilot Study, Stress vs. Dimensionality

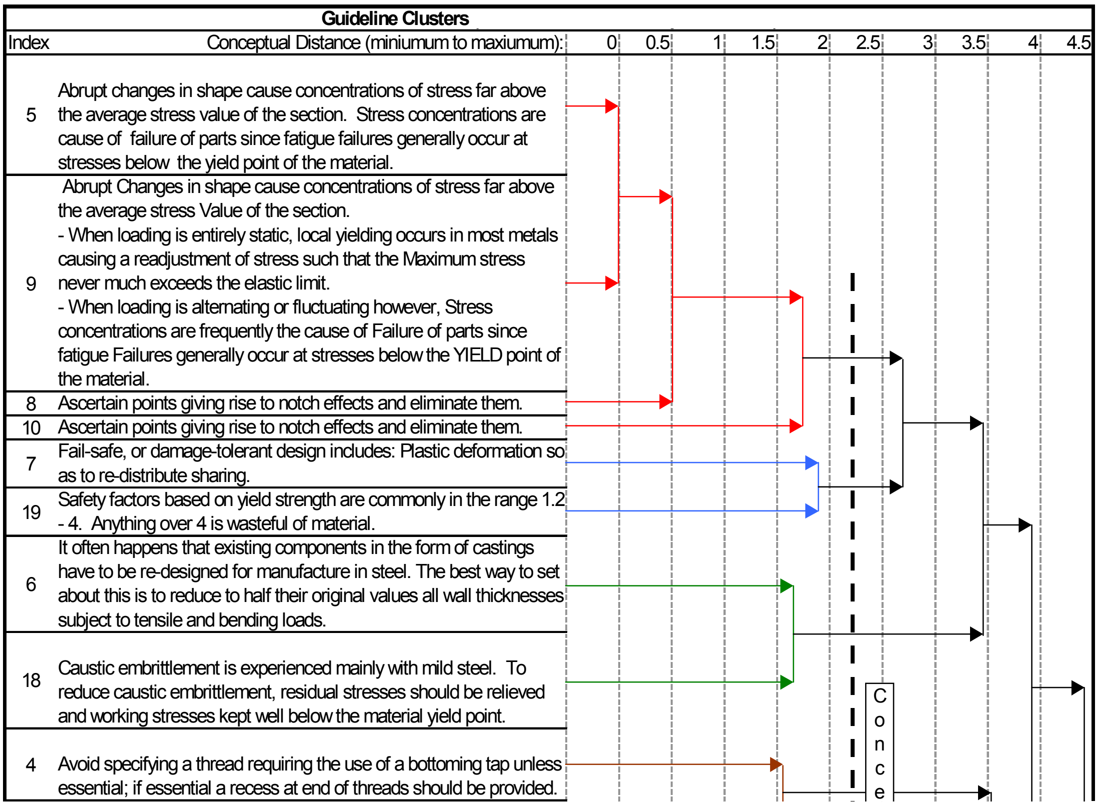
The data was also processed in two and three dimensions. The results were very similar to those from the four-dimension solution, but there was some distortion.

As it is impossible to directly visualise four dimensions, an intermediate method must be used to view the data. Agglomerative hierarchical clustering was chosen for this step. It was chosen because it concisely displays proximity information. The following clustering methods were used: centroid, single, ward, average, and complete. These methods are constructed with different assumptions about how the data is structured. Since this is exploratory research, it is impossible to tell which method is best *a priori*. Some of the methods, such as the centroid method, seem logical and likely to fit the data. However, using this method without closely examining the results from the other methods risks reaching poor conclusions. If a solution exists, it will ultimately be found when all the pieces fit in place. Examining all of the clustering methods also provides a check of how robust the data is. A further check of data integrity is provided by the use of repeated guidelines, eight and ten. Because of the distortion encountered when a non-metric space is adjusted to be metric, the slight separation between these two guidelines is acceptable.

3.3 Pilot Study – Analysis of Group Data

At close conceptual distances, all of the methods produced very similar cluster patterns. At greater conceptual distances, the cluster patterns were very much alike when similar clustering methods were compared. This indicated that the data was robust. The *complete* method fit the data noticeably better than the other methods. A section of the dendrogram is given below.

The general behaviour depicted in the dendrogram below (Figure 3-2) is one of clustering about a common subject. More abstract guidelines about the same subject are added as conceptual distance increases. With the exception of one cluster, the commonality of the guidelines in the clusters is clear. However, this is only true in hindsight for some of the clusters. Knowing that guidelines cluster by subject and then by level of abstraction, it would be reasonable to predict that guidelines 5, 8, 9, and 10 would be in the same cluster.



Index	Conceptual Distance (miniumum to maxiumum):
5	Abrupt changes in shape cause concentrations of stress far above the average stress value of the section. Stress concentrations are cause of failure of parts since fatigue failures generally occur at stresses below the yield point of the material.
9	Abrupt Changes in shape cause concentrations of stress far above the average stress Value of the section. - When loading is entirely static, local yielding occurs in most metals causing a readjustment of stress such that the Maximum stress never much exceeds the elastic limit. - When loading is alternating or fluctuating however, Stress concentrations are frequently the cause of Failure of parts since fatigue Failures generally occur at stresses below the YIELD point of the material.
8	Ascertain points giving rise to notch effects and eliminate them.
10	Ascertain points giving rise to notch effects and eliminate them.
7	Fail-safe, or damage-tolerant design includes: Plastic deformation so as to re-distribute sharing.
19	Safety factors based on yield strength are commonly in the range 1.2 - 4. Anything over 4 is wasteful of material.
6	It often happens that existing components in the form of castings have to be re-designed for manufacture in steel. The best way to set about this is to reduce to half their original values all wall thicknesses subject to tensile and bending loads.
18	Caustic embrittlement is experienced mainly with mild steel. To reduce caustic embrittlement, residual stresses should be relieved and working stresses kept well below the material yield point.
4	Avoid specifying a thread requiring the use of a bottoming tap unless essential; if essential a recess at end of threads should be provided.

Figure 3-2 Partial Dendrogram of Guideline Cluster Patterns

More ambiguous guidelines such as, 'A general rule applicable to allowable dimensional tolerances is that such tolerances should be at least half the minimum shrinkage allowable for the material involved', could easily be expected to be grouped with other guidelines pertaining to dimensioning. But seen grouped with this guideline, 'By welding on the step-by-step principle the tendency for the work to twist round at right angles to the weld is avoided', the intent of this grouping becomes apparent. Both give advice dealing with thermal distortion. Guidelines couched in this manner are likely to be misinterpreted or viewed as out of context. Instead of forming large groups with many distantly related guidelines, it appears that engineers conceptualise engineering knowledge in tight groups formed around subject areas.

3.4 Pilot Study – Processing of Individual Data Sets

Two methods were employed to examine the amount of variation between individual conceptualisation patterns. In the first method, the amount of distortion required to fit each set of conceptual distances to a common space was measured. A Euclidian distance model was used to create the common space and each case was weighted in each dimension to fit it to a single solution. These weights were then examined to determine how much each participant's response varied from the mean. To ease visualisation, the weights from scaling the data to two dimensions were used.

3.5 Pilot Study – Analysis of Individual Data Sets

Referring to Figure 3-3 one can see that there is a core group of designers who share a similar view of the data.

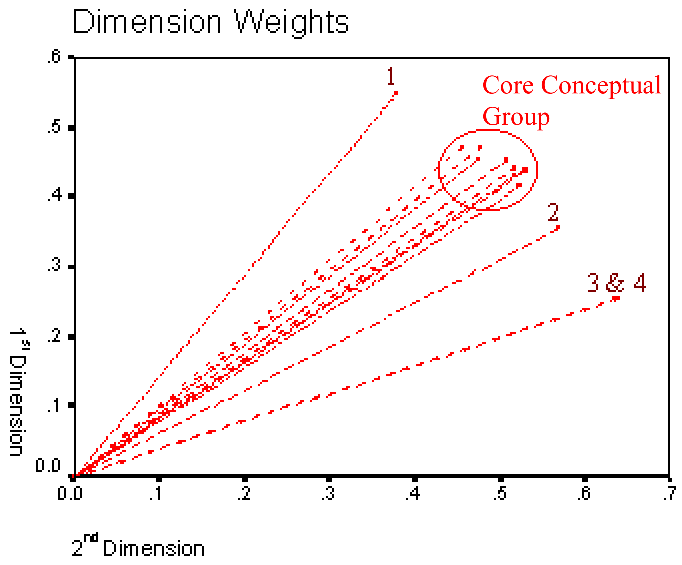


Figure 3-3 Dimensional Weights of Individual Participants

The outlying points, 1-4, represent individuals who view the data with a different conceptual emphasis. Due to the small sample size, it is only possible to state that this indicates the possibility that other groups of engineers are underrepresented in this study.

In addition, the individual cluster plots were compared with each other. The individual sets of data were clustered using the same method as before.

Guidelines that had a clearly defined meaning and area of application consistently grouped together and formed the core of larger clusters. Guidelines that were more vaguely worded were placed into varying clusters. As there was not sufficient context to fully define these guidelines, it is presumed that the participants created sufficient context in their minds and this was used to place the guidelines into pertinent groups. This reasoning is also supported in part by a study conducted by Kuffner and Ullman [7]. As there were only twelve participants, the possibility for distinct sub patterns to emerge was limited. One clear group of four designers was identified. All but one of the remaining responses were consistent in their placing of well-defined guidelines, but varied in the clustering of more broadly applicable pieces of information.

Two very distinct groups of core guidelines emerged. These guidelines formed the nucleus of larger cluster patterns in all but one response. An example of these core guidelines would be guideline number eight. This guideline was always placed with guideline number nine, and almost always with guidelines number ten and eleven. Nearly all of the participants placed the guidelines into three clusters. Some designers left one or two odd guidelines in outlying positions; however, this behaviour was not consistent.

The two well defined groups of guidelines pertained to stress and dimensioning. The third group of guidelines could best be described as containing general manufacturing information. These guidelines tended to be somewhat ambiguous in application and not easily related to the first two groups. In the first two conceptual groups, *stress* and *dimensioning*, additional guidelines were added to the core guidelines to round out the clusters. As previously mentioned, these guidelines could easily be viewed as related to the rest of the group, but they could just as easily be applied to another engineering domain.

3.6 Pilot Study Conclusions

It is clear that unambiguous guidelines are conceptualised in a similar manner by engineers. These pieces of design information clearly address a specific design issue. While information that is applicable across several areas of engineering design may appear to be a more fundamental truth of designing, these guidelines lack sufficient context to reliably describe how engineers will interpret them. Guidelines are also clearly clustered by subject. There is no information indicating that the guidelines in this study were clustered by any other rational. However, the number and scope of guidelines in this study are very limited. If other trends exist for conceptualising design information, it would be more difficult to identify them using this data set.

4 Primary Study

4.1 Primary Study Overview

A second exploratory study is currently underway. It utilises information gathered from various design manuals that are in use with each partner firm. This study is intended to be much larger than the first, with a total of thirty participants and over 150 design guidelines. To give the guidelines context and to relate them to the experience of the participants, the guidelines are selected around a common design problem. A separate set of 75 to 100 guidelines is being selected with each partner company. Because these guidelines represent the current working practices of these firms, they cannot be published. In brief, these guidelines are much more detailed then the guidelines listed in the first study. Typically they are comprised of a short paragraph of text and frequently have an accompanying diagram.

4.2 Primary Study Method

As it would be very time consuming to rank every unique distance in a data set of seventy-five guidelines, an alternate method of ranking conceptual distances is being used. For this, the repertory grid analysis method was chosen. In repertory grids, a set of constructs, or conceptual scales, is elicited from the participant, and then the data set is ranked against these constructs. Experiments with repertory grids, led to the use of eight constructs for similarity ranking. It was found that using eight constructs to rank the data from the first experiment would typically reproduce the same patterns that direct pair-wise ranking did.

To elicit a construct, three items are presented to a participant. In our case, these are three randomly selected guidelines. Based upon their experience, the designer then identifies the 'odd one out'. Thus, the other two guidelines are more similar to each other than any other combination with the third. After determining the most similar pair of guidelines, the designer then has to articulate what they share in common. The resulting statement, e.g. *it* [the two guidelines] *pertains to electrical bonding*, forms the construct. The guidelines in the survey are then ranked against this construct on a scale of zero (no agreement with the construct) to ten (closely matches the construct). Thus, the construct forms a half dimension, an axis that starts at zero. The resulting similarity data is metric and describes the location of each engineering guideline in a space where each dimension is a known concept. Because of time constraints, only forty guidelines are ranked by any one person.

4.3 Primary Study – Initial Data

This study will require a significant amount of time (a total of ninety hours is the current estimate). In order to reduce risk, several test studies have been conducted within the

Department of Engineering at Cambridge University. Data from these studies will be discussed here. Eight engineers, six post graduate students and two faculty, completed the test study. This study was comprised of seventy-five design guidelines, all of which pertained to wing fuel tank design. The first group of four generated their constructs from four identical sets of twenty-four randomly selected guidelines. These constructs were then used to rank four identical sets of forty randomly selected guidelines. For construct generation, the second group of three received three different sets of twenty-four guidelines of randomly selected guidelines, and three different sets of forty randomly selected guidelines for ranking. An additional person ranked all seventy-five guidelines against eight constructs generated from a random set of twenty-four guidelines. Both groups received the same written instructions. They also received an instructional presentation. This presentation was changed from the first to the second group to reflect questions asked and issues raised by the first group.

4.4 Primary Study – Data Processing

Several steps were taken in the data processing to combine the data to create a common set of guideline similarity distances for analysis. The inter-point distances were first calculated. This resulted in a 40 x 40 dissimilarity matrix (the numerical inverse of a similarity matrix). These distance values were then combined by averaging the individual distance values to the appropriate location in a 75 x 75 matrix representing the entire guideline set. The resulting non-metric space was scaled to both two and three dimensions using the Proxscal variant of MDS in SPSS. An *ordinal* proximity transformation model was chosen. The resulting data was then clustered using the *centroid* method. There was no significant difference between the cluster patterns formed from either of these dimensions. In addition, the stress value at two dimensions was low enough (0.055) to be considered adequate [5]. As a result, direct analysis of the conceptual space may be easily done in two dimensions, but the use of three or four dimensions in dendrogram plots would give a slightly more accurate depiction of the data.

4.5 Primary Study – Data Analysis

Using the two dimensional view of the data in Figure 4-1, one can see three outlined groups. Of course, at finer levels of resolution, more groups are apparent. However, these groups represent three clearly defined areas of the conceptual space. Although these groups may not be immediately apparent, they are the result of applying the *centroid* method of clustering. In this, clusters are formed based upon nearest distances between cluster centroids. To gain further insight in how engineers understand this information, it is worthwhile to examine this figure in detail.

On a very broad level, the concepts can be described as varying along two major axes. While these axes aren’t depicted, they would extend from one corner of the plot to the diagonal opposite. Proceeding clockwise from the upper right corner, the concepts that anchor these axes are, sealing, fasteners/bonding/physical connections, piping and access, and layout/construction/general assembly. This layout of basic ideas and the further mapping of detailed concepts to spatial locations, help to define the location of a specific piece of information. After matching the guideline numbers to the corresponding design information, the concepts for each cluster became apparent. The first cluster deals with the structural aspects and the components of fuel tanks, the second cluster generally pertains to sealants, and the third group is focused on electrical bonding. These divisions give a broad indication as to how the design information is segmented.

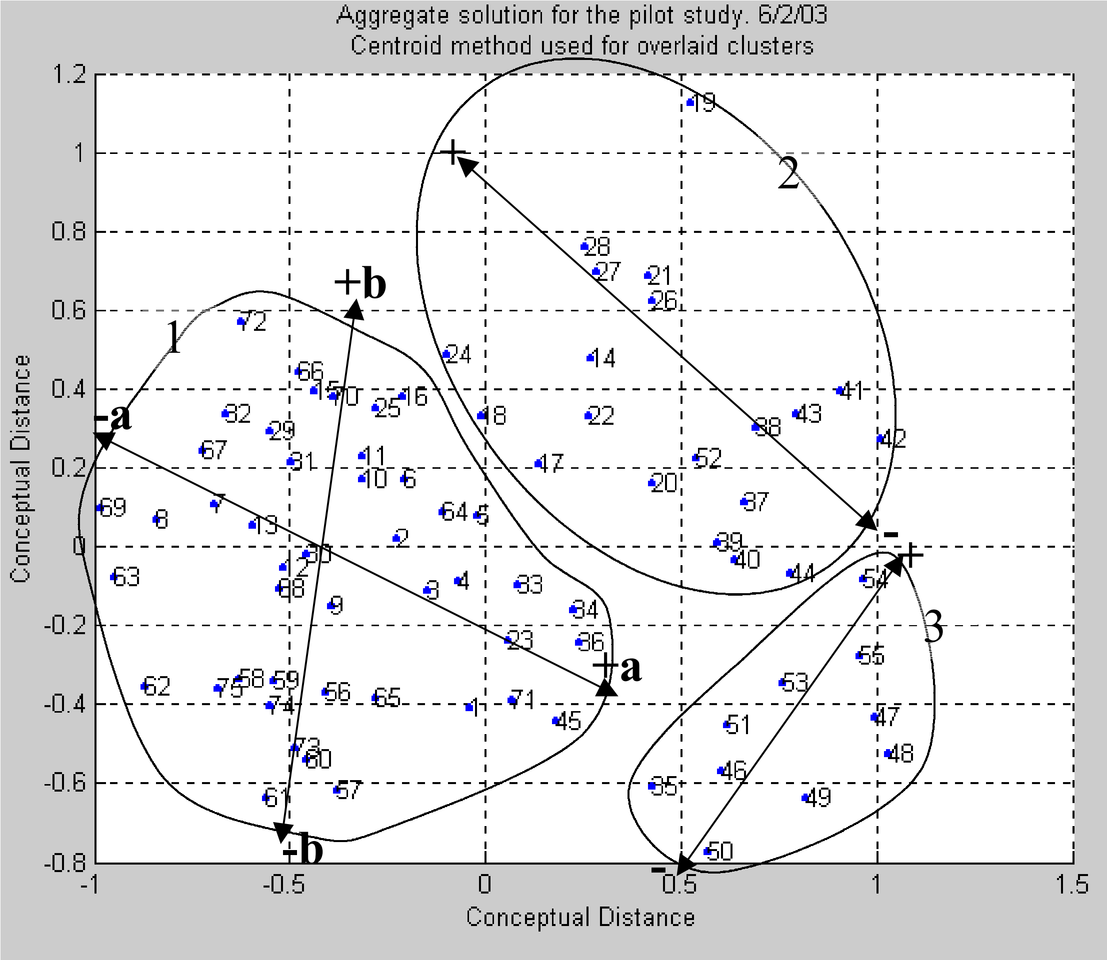


Figure 4-1 The Conceptual Location of Design Information

An application of this would lead to the conclusion that if one were faced with a problem about fuel tank ribs (1st group), then information about fasteners (1st group) would be more pertinent then information about sealants (2nd group). Although this conclusion may seem very basic, it is the first quantifiable data about how engineers conceptualise design knowledge. From high-level divisions in the data, the basic structure for defining conceptual similarity may be formed.

When the first cluster was subjected to further scrutiny (Figure 4-1), it was noted that there were two major axis of variation. While these axes aren't orthogonal, this is acceptable as they adequately describe the conceptual space. The negative end of the *b* axis has guidelines about fuel pipe specifications and properties, such as pressure, material, routing and safety requirements. The positive end of this axis has guidelines that relate to issues about the layout or construction of fuel tank boundaries. The midpoints of this axis (and the *a* axis) relate to fuel tank access. In defining conceptual similarity, this shows a progression of design concepts along the topic of fuel tanks. The *a* axis in this group represents a set of concepts which shares a common point with the *b* axis. In addition, these conceptual axes accurately describe the location of the other guidelines in this cluster. In Figure 4-2, the first and third quadrants of the first cluster are given. Comparable trends exist in the second and forth quadrants.

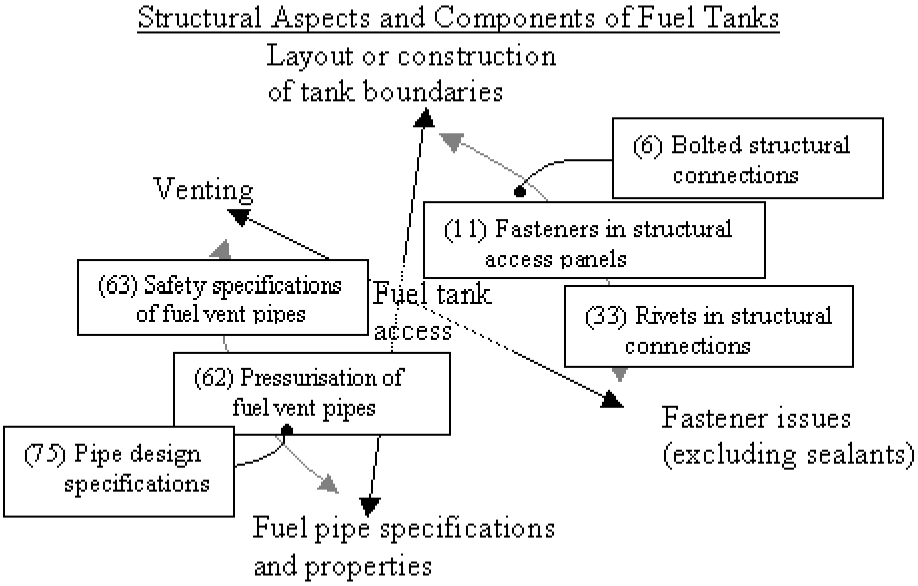


Figure 4-2 Cluster 1, First and Third Quadrants

In the first and third quadrants a clear progression of guideline topics can be seen as one moves from one axis to the other. When travelling along an approximately constant radius arc, the two different concepts represented by the axes *a* and *b* vary in proportion to the conceptual distance. A chunk of design knowledge, such as guideline eleven, which pertains to three aspects of fuel tank design, is placed closer to the origin. The placement of this and other guidelines indicates that this is a remarkably continuous conceptual space. Although not yet verified, it is very likely that these criteria could be used to accurately place additional guidelines into this conceptual space.

The second cluster could be broadly described as pertaining to sealants. Starting at the positive end, this axis varies from *sealants and electrically bonded bolts*, to *sealants and general assembly (non bonding)*. In this group and in the third group, only one conceptual axis could be identified. The perpendicular dispersion of points from the axis appears to be random scattering from the axis, not a meaningful dispersion. The relatively small number of points in the second and third groups makes it more difficult to identify additional conceptual dimensions.

The third cluster is clearly focused on electrically bonded connections. At the negative end of the axis, the guidelines deal with metal to carbon fibre composite bonded joints (CFC), while the positive end of the scale terminates with guidelines pertaining to CFC-to-CFC bonded joints. It is intriguing to note that the bonding guidelines in the third group all pertain to composites, while the bonding guidelines in the second group only deal with metal-to-metal connections. This is a clear indication that engineers view the use of composite materials in joints in a very different manner than they view all metal joints. The function of the joints is the same, only the materials varied.

5 Conclusions

The first and second studies have clearly shown that designers view unambiguous engineering design information in a consistent manner. The first study has also demonstrated that conceptualisation of broadly applicable information varies greatly amongst engineers as each constructs a different context in which to interpret the information. Additionally, guidelines that are similar in subject matter have a strong propensity to group together. Building upon the basic conclusions that the first study offered, it became apparent in the second study that engineers have a clearly constructed conceptual space in which they place design information. Knowledge of this conceptual space may be used to increase the ease and precision of information searches. It will also allow for information searches where the engineer has an idea of what type of information they need, but they are not exactly sure of what they are looking for. By understanding what concepts relate design information together, the potential is created for computer searches to return relevant information from a seemingly unrelated domain. Further exploration of this conceptual space, both through additional data analysis and field research, is planned.

6 Acknowledgements

I would like to thank the engineering design staff at Rolls-Royce and BAE SYSTEMS for their time and assistance in this research. In particular, I am indebted to the patience and persistence of my respective points of contact at these firms, Mike Moss and Alastair Stewart.

References

[1] Nowack, M.L., "DESIGN GUIDELINE SUPPORT FOR MANUFACTURABILITY", in "Engineering Department", Cambridge University, Cambridge, 1997, p.212.
[2] Charlton, C.T., "The Retrieval of Mechanical Design Information", in "Engineering Department", Cambridge, Cambridge, 1999, p.165.
[3] Baya, V., "INFORMATION HANDLING BEHAVIOR OF DESIGNERS DURING CONCEPTUAL DESIGN: THREE EXPERIMENTS", in "Mechanical Engineering", Stanford University, 1996, p.167.
[4] Ahmed, S., "Understanding the Use and Reuse of Experience in Engineering Design", in "Department of Engineering", Cambridge, Cambridge, 2000, p.169.
[5] Kruskal, J.B.M., W., "Multidimensional Scaling", Sage Publications, London, 1978.
[6] Schiffman S.S., R.M.L., Young F.W., "Introduction To Multidimensional Scaling", Academic Press, London, 1981.
[7] Kuffner, T.A.U., D. G., "The information requests of mechanical design engineers", DESIGN STUDIES, Vol. 12 No. 1(January 1991f), 1991, pp.42-50.

Corresponding author:
Russell Japikse
Cambridge University
Department of Engineering, Engineering Design Centre
Trumpington Street
Cambridge, CB2 1PZ
United Kingdom
Tel: Int +44 1223 330271
Fax: Int +44 1223 Fax Number
E-mail: rdj26@cam.ac.uk
URL: www-edc.eng.cam.ac.uk/people/rdj26.html