Data Mining to Support Engineering Design Decisions

Pooja Jadhav, Sylvia C Wong, Gary B Wills, Richard M Crowder and Nigel R Shadbolt

School of Electronics and Computer Science,
University of Southampton, UK
{sw2, gbw, rmc, nrs} *at* ecs.soton.ac.uk

Abstract. The design and maintenance of an aero-engine generates a significant amount of documentation. When designing new engines, engineers must obtain knowledge gained from maintenance of existing engines to identify possible areas of concern. Firstly, this paper investigate the use of advanced business intelligence tenchniques to solve the problem of knowledge transfer from maintenance to design of aeroengines. Based on data availability and quality, various models were deployed. An association model was used to uncover hidden trends among parts involved in maintenance events. Classification techniques comprising of various algorithms was employed to determine severity of events. Causes of high severity events that lead to major financial loss was traced with the help of summarization techniques. Secondly this paper compares and evaluates the business intelligence approach to solve the problem of knowledge transfer with solutions available from the Semantic Web. The results obtained provide a compelling need to have data mining support on RDF/OWL-based warehoused data.

1 Introduction

The design and maintenance of large and complex engineering systems requires a significant amount of documentation, particularly if the system being considered is a turbofan engine used on the current generation of aircraft. These engines are amongst the most complex machine ever designed, incorporating a wide range technologies including high temperature materials, complex fluid dynamics and high speed rotating components.

A fundamental shift is currently occurring in the aerospace industry away from selling products to providing services. Companies such as Rolls-Royce aim to make an increasing number of its engine fleet subject to long-term maintenance service agreements [1]. Essential to the success of this market shift is to design new products with lower and more predictable maintenance costs. To minimize maintenance costs throughout the engine's life cycle, engineers must obtain knowledge gained from maintenance histories of similar products during the design phase of new products. This will help engineers identify parts most likely to be problematic throughout the engine's entire life cycle. It should be noted that engine design is typically undertaken by a number of teams who are responsible for individual engine modules, e.g compressor or turbine. Therefore it is impossible for any single member of a design team to access more than a fraction of the available documentation. As is widely recognized, information systems usually develop over time into a set of heterogeneous resources with ill-defined metadata. As a result, it becomes difficult for engineers to follow a trail through the resources [2]. The challenge for organizations is therefore to develop an information system that is both comprehensive and will satisfy the increasing demands from industry for up-to-date and easily accessible information.

In response to these challenges, the Integrated Products and Services (IPAS) project is developing a Semantic Web based document repository to support engineers to design for the aftermarket [3]. Semantic Web technologies are used for this project because of their ability to easily integrate distributed resources. In the aerospace industry, maintenance documents are created by service teams located all over the world, at airports and overhaul facilities. Design documents are created by multiple design teams, which can also be based at multiple sites. In this paper, we compare business intelligence techniques with solutions from the Semantic Web in answering maintenance related questions raised by design engineers.

The paper is organized as follows. Section 2 explains the motivation behind our document repository project, and the objectives we are trying to achieve. Section 3 describes the type of knowledge design engineers specifically sought from our document repository. Section 4 describes the system we use for data mining using the business intelligence approach, and the result we obtained from such system. Section 5 explains the Semantic Web approach, and the kind of knowledge you can harvest from Semantic Web queries. The paper finishes with discussion in Section 6 and conclusions in Section 7.

2 Motivation

As is well recognized in engineering design, the use of past experiences and previously acquired knowledge, either from the designer's own experiences or from resources within their organization forms an important part of the design process. It has been estimated that 90% of industrial design activity is based on variant design [4], while during a redesign activity up to 70% of the information is taken from previous solutions [5]. A cursory consideration of these figures identifies two immediate challenges – how to capture knowledge, and how to retrieve it. The purpose of our document repository is thus to enable the transfer and retrieval of knowledge across the organization to support design activities.

Figure 1 shows the key information flow for the different stages in the life of an engine. Concept design is the first stage of an engine's life cycle. Given a set of broad requirements, such as thrust, range of the target aircraft and fuel burn, engineers determine the approximate dimensions, weight, power and other physical characteristics for the engine design. The engineers also make estimates of the manufacturing costs of the engine. In the design stage, engineers transform the preliminary abstract design into a set of concrete plans that can be used in production. In production, engines are built according to the design plans. Traditionally, after production and sales, responsibility for the engine passes from the manufacturer, to the airlines, who own the engines. The airlines are responsible for maintaining the engines. This maintenance activity is supported by the manufacturer's technical support and operations team. To assist maintenance engineers to identify problems before a breakdown occurs, engines are commonly equipped with sensors for engine monitoring. This monitoring information can be analysed for abnormal operating conditions, such as temperature or pressure. However until fairly recently, the monitoring data was only used to support maintenance activities, even though it is a rich source of information for the designers of future engines.

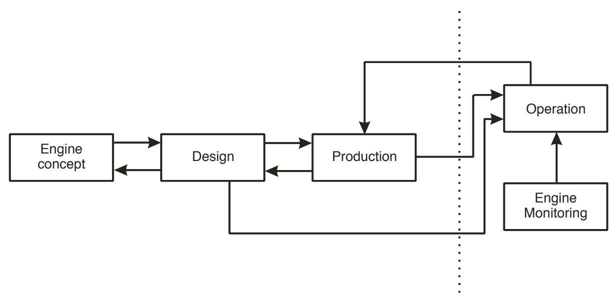

Fig. 1. Information flow between the different stages in the life of an aero-engine. The vertical line between production and operation represents the transfer of the engine from its manufacturer to an airline. Operations is the generic term for maintenance and aftersale support.

As can be seen in Figure 1, there are interactions and information flow between neighbouring stages in the production maintenance process. This is due to the iterative nature of engineering processes. Design knowledge is also passed to operations in the form of 'owner's manuals'. Sometimes, information also passes between unconnected stages, for example, between operations and engine concept. However, the flow is weak and may take the form of informal and personal networking between engineers.

While the process works very well, it does have significant disadvantages. In particular, design engineers are remote from the problems experienced in the field by operations. Due to the importance of increasing operational reliability and minimizing maintenance costs in the new market paradigm of product support, information gained in the operation of a fleet of engines needs to be fed back to the designers of subsequent engines. However, the current information infrastructure makes this difficult as concept design engineers do not have access to maintenance knowledge. Similarly, design engineers should consult existing maintenance documents to help design parts with more predictable maintenance costs.

As a result, we need to strengthen and help formalize the information flow between the company's aftermarket operations and the design teams. Figure 2 shows the information and knowledge flows our research aims to build. This would allow the knowledge gained during the design, production and operation of an engine to advise the design of the next variant.

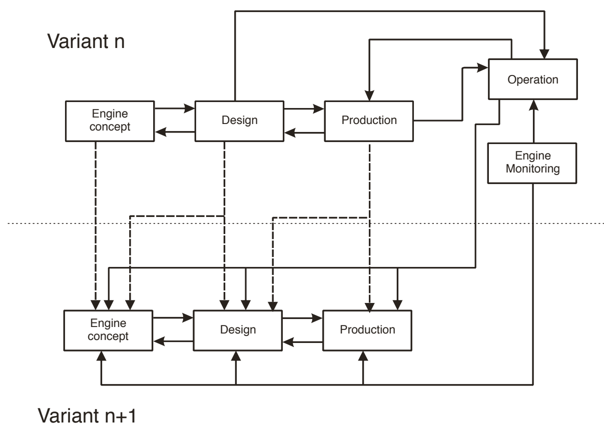

Fig. 2. We aim to facilitate the flow of information gained during the life-cycle of one engine variant to inform the design of the next variant. The dotted arrows indicate the flow of design rationale and similar knowledge. The solid arrows represent all other information flows including real time engine information and design documentation.

The following scenario[1] illustrates the potential use and benefit from such a document repository. The scenario involves three separate and different groups of users that are involved in the life of a jet engine. Front-line maintenance engineers are involved in the day to day servicing of the engines, and thus responsible for populating the document repository with maintenance reports and other similar documents.

> *During the regular pre-flight checks, a flight crew reported a problem with a leak from an engine's bleed air system. Subsequent inspection which required the removal of the engine revealed that a duct had failed at a joint due to vibration. After repair, the engine was returned to service, and a full maintenance event report submitted to the document repository.*

The document repository can then be used by technical support and operation engineers, who are responsible for improving the performance of existing

[1] The scenario is entirely fictitious and does not derive from any real event.

engines. They can use information collected in the repository to monitor trends that develop over a fleet of engines. Modification can then be designed to mitigate any problems found:

> *Following a review of the maintenance events relating to a specific engine fleet, a trend was noticed in the high than expected number of failure of an air duct joint due vibration. To maintain the reliability of the engine fleet, a modification was developed and implemented.*

The same information in the repository will also be used by design engineers working on a new engine:

> *The design team for the next variant of this engine reviews the performance of the air bleed system across the fleet to learn from previous design rationale and operational history. Finite element analysis showed that a joint failure could occur due to vibration if certain operational conditions were met. It was therefore decided that the future variant of the engine would both eliminate the joint and reroute the duct work. The revised design costs 50% more than the original. However, the saving over the life of the engine will be substantial due to lower likelihood of in-service failure.*

The goal of our work can thus be summarized as follows: To feedback and harvest knowledge gained from the aftermarket operations documents to help (a) operations engineers in designing modifications to existing engines, and (b) design engineers in designing the next variant engine for the aftermarket.

3 User Requirements

To understand the scope of knowledge our users want to gain from an engine's service history, a questionnaire-based study was conducted with design engineers from Rolls-Royce [6]. In the questionnaire, engineers were presented a list of questions relating to maintenance experience with a product. They were asked how often they might ask them when designing a new product. They were also asked what other questions they might want to ask. The result of this questionnaire tells us what are the most important and most common life cycle information design engineers seek from maintenance documents.

The study identified 39 questions commonly asked by design engineers. Only a small number of the 39 questions involve complex mathematical and numerical simulations, which cannot be answered from semantic analysis of maintenance documents. Most of the questions are concerned with textual and semantic information stored within the documents. However, to answer even these textual based questions requires some degree of simple arithmetical manipulation, as demonstrated by the following examples:

1. What are the common deterioration mechanisms associated with this part?

2. How critical is it if and when this part fails?
3. Are there any other mechanisms, which only occur rarely?
4. How many engine removals have been caused by a deterioration of this part?
5. Which parts dominate the reliability and cost drivers in this engine?

The first two questions in this list is picked for the studies carried out in this paper. Criticality of an event is ranked using a scoring system provided by Rolls-Royce.

4 Business Intelligence

Our document repository is based on a collection of web services for processing, and a web portal for user access [3]. In order to predict a suitable pattern among the data and to provide statistical analysis and visualisation, we implemented our data mining models as a web service which can be invoked as an independent service as shown in 3. This helps design engineers analyze maintenance data efficiently. The data mining model used comprises of various techniques like statistical, classification and clustering models. These models are used to answer the questions posed by the design team in Section 3. By providing a data mining facet, there can be a definite measure of support and confidence that can be put forth for each query answered. It can help solve complex queries that involves calculation of statistics over different attributes to uncover hidden patterns in maintenance events. Inputs to the model are complex queries which is processed using the currently available data and the output in the form of XML. The web service further provide the response in the form of SOAP object.

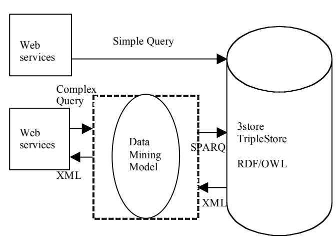

Fig. 3. Architecture for data mining via business intelligence.

The Oracle Data Miner 10.2 is used for our study. This data mining tool requires data to be warehoused in the Oracle 10g database. For our model, we considered two data sets – Dataset 1 is the largest available subset of data that is clean and has well-defined, distinct values for the selected mandatory attributes; Dataset 2 comprises of an outer join of all the important attributes required for data mining. To raise the quality of existing data, cleaner subsets of data was retrieved by discarding missing value records from database tables. Data was then integrated from all the tables by performing join queries. For the association model, dataset 2 is used, as the focus is to have maximum events comprising of number of parts. Since few attributes are involved, the possibility of extracting clean data is high. For the classification model, dataset 1 is used instead. This is because dataset 1 is the largest available subset of data that is clean and has well-defined values for mandatory attributes.

Figure 4 shows the result of the association model implemented. It summarizes the parts that are installed or repaired during a given event. In market basket analysis, 711 rules were observed. After thresholding support and confidence with high values more significant rules were extracted. For instance, Rule 498 (row 2 in Figure 4) suggests that whenever a link (antecedent) is installed or removed, a Support:Assy (precedent) is always installed or removed everytime, that is the conditional probability of Support:Assy getting repaired given that a link has been repaired is 1.0.(P(Support:Assy / Link) = 1) A support of 6.91 suggests the frequency of occurrence of these two events together. The model was implemented for rules having 3 and 4 as the length of attributes (number of antecedents).

Rule Id	If (condition)	Then (association)	Confidence (%)	Support (%)
538	MODULE:COMPRESSOR:INTER= 1	NOT LISTED= 1	100	6.916579
498	LINK= 1	SUPPORT:ASSY= 1	100	6.3357973
584	MODULE:IP & LP TURBINE= 1	NOT LISTED= 1	99.26471	7.127772
526	LINK:ASSY= 1	SUPPORT:ASSY= 1	99.18699	6.441394
149	BRACKET:ASSY= 1	SUPPORT:ASSY= 1	99.18033	6.3885956
123	BRACKET:ASSY= 1	LINK:ASSY= 1	99.18033	6.3885956
207	DED GEN ROTOR= 1	HP BLEED VALVE= 1	99.166664	6.282999
241	DED GEN ROTOR= 1	VALVE= 1	99.166664	6.282999
213	DED GEN ROTOR= 1	LINK:ASSY= 1	99.166664	6.282999
209	DED GEN ROTOR= 1	IGNITION UNIT= 1	99.166664	6.282999
239	DED GEN ROTOR= 1	SUPPORT:ASSY= 1	99.166664	6.282999
108	DED GEN ROTOR= 1	BRACKET:ASSY= 1	99.166664	6.282999
500	LINK= 1	VALVE= 1	99.166664	6.282999
199	DED GEN ROTOR= 1	DED GEN STATOR= 1	99.166664	6.282999
185	COOLER= 1	OIL LOW PRES SW= 1	98.4	6.494192
183	COOLER= 1	OIL LOW PRES DIFF SW= 1	98.4	6.494192
157	COOLER= 1	DETECTOR= 1	98.4	6.494192
6	COOLER= 1	BLADE:ASSY= 1	98.4	6.494192
411	LINK:ASSY= 1	HP BLEED VALVE= 1	98.373985	6.3885956
528	LINK:ASSY= 1	VALVE= 1	98.373985	6.3885956
124	LINK:ASSY= 1	BRACKET:ASSY= 1	98.373985	6.3885956

Fig. 4. Results of the association model.

Some of the aeroengine parts are installed or replaced along with other parts as a part of planned maintenance. But there can be parts that would be a result of unplanned maintenance. This can highlight hidden sequences of parts that are repaired simultaneously and thus, can be attributed to a design fault in the mechanics of these parts. It can help the design engineers to build a strategy that would help prevent such issues in the future events. Therefore, the association model supports question one in Section 3 by finding the association of parts involved in a failure event using data mining.

A classification model comprised of Adaptive Nave Bayes Network was trained using 60% of data and then was made to predict criticality values (in the form of high, low and medium) on the remaining 40% of data. The models performance, accuracy and cost was evaluated and is shown in Figure 5. Out of 136 High target cases, 90.44% of the values was predicted correctly. The cost of predicting the remaining 10% of False Positive cases is 58.25. The overall predictive confidence of the model is 92.53%.

Target	Total Actual	Correctly Predicted %	Cost	Cost %
High	136	90.44117647	58.25926	66.12573
Low	428	94.62616822	29.8445	33.87427
Medium	4	100	0	0

Fig. 5. Model Performance for Adaptive Nave Bayes.

Events with high criticality and low cycles and hours range are considered to be probable design faults and were filtered from dataset 2. Histogram tools were run on each of the attributes to find the conditions, reactions, symptoms, part names, cycles and hours associated with a high severity design fault event as depicted. Figure 6.

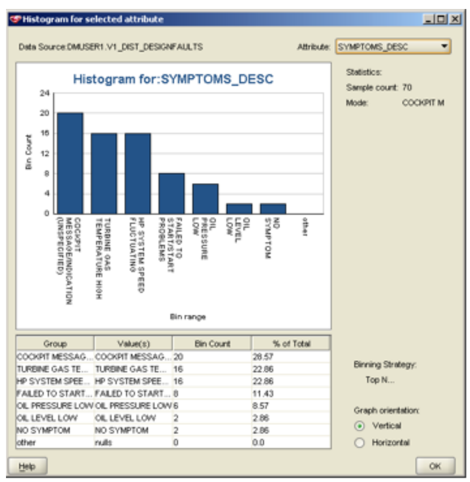

Fig. 6. Common Symptoms for high criticality events.

5 Semantic Web

To answer the two questions selected from Section 3, two SPARQL queries were run against the triplestore. Figure 7 shows a segment of the output formated as

HTML table. The outputs obtained as a result of the two queries directly answers the questions posed by the design engineers. Since the data is structured in the form of knowledge in a knowledge-based repository, it is well-organised.

System_Event_Number	has_symptom	has_affected_engine	has_finding_condition	has_finding_item	has_finding_comment
URI http://www.ipas.ecs.soton.ac.uk/events_trent500.owl#SDMServiceEvent_3179686	engine rundown uncommanded (datatype http://www.w3.org/2001/XMLSchema#string)	URI http://www.ipas.ecs.soton.ac.uk/events_trent500.owl#Trent500_71157	failed (datatype http://www.w3.org/2001/XMLSchema#string)	URI http://www.ipas.ecs.soton.ac.uk/events_trent500.owl#null_HMU500MK3_B15224	Confirmed - LRU - Servo pressure regulator SPR plug released due to oversize bore and poor brazr joint. SPR plug jammed SPR in closed position cutting off servo pressure to MV T/M and MV. (datatype http://www.w3.org/2001/XMLSchema#string)
URI http://www.ipas.ecs.soton.ac.uk/events_trent500.owl#SDMServiceEvent_1188954	power auto shutdown (datatype http://www.w3.org/2001/XMLSchema#string)	URI http://www.ipas.ecs.soton.ac.uk/events_trent500.owl#Trent500_71095	failed (datatype http://www.w3.org/2001/XMLSchema#string)	URI http://www.ipas.ecs.soton.ac.uk/events_trent500.owl#null_HMU500MK3_BT5152	Component Fault: MMV bal seal (datatype http://www.w3.org/2001/XMLSchema#string)
URI http://www.ipas.ecs.soton.ac.uk/events_trent500.owl#SDMServiceEvent_4274732	fuel filter blockage indication (datatype http://www.w3.org/2001/XMLSchema#string)	URI http://www.ipas.ecs.soton.ac.uk/events_trent500.owl#Trent500_71162	confirmed not component (datatype http://www.w3.org/2001/XMLSchema#string)	URI http://www.ipas.ecs.soton.ac.uk/events_trent500.owl#FF_Diff_Pres_Sw_21SN04-338_D009443	Snag came back on 16-Aug-05 (datatype http://www.w3.org/2001/XMLSchema#string)
URI http://www.ipas.ecs.soton.ac.uk/events_trent500.owl#SDMServiceEvent_4274732	cockpit message/indication (unspecified) (datatype http://www.w3.org/2001/XMLSchema#string)	URI http://www.ipas.ecs.soton.ac.uk/events_trent500.owl#Trent500_71162	confirmed not component (datatype http://www.w3.org/2001/XMLSchema#string)	URI http://www.ipas.ecs.soton.ac.uk/events_trent500.owl#FF_Diff_Pres_Sw_21SN04-338_D009443	Snag came back on 16-Aug-05 (datatype http://www.w3.org/2001/XMLSchema#string)
URI http://www.ipas.ecs.soton.ac.uk/events_trent500.owl#SDMServiceEvent_4235110	oil pressure low (datatype http://www.w3.org/2001/XMLSchema#string)	URI http://www.ipas.ecs.soton.ac.uk/events_trent500.owl#Trent500_71307	broken (datatype http://www.w3.org/2001/XMLSchema#string)	URI http://www.ipas.ecs.soton.ac.uk/events_trent500.owl#Oil_Qty_Trans_8TJ146RBV1_MF002329	Found board to have a broken trace that would cause intermittant readings. See attachements for Report. (datatype http://www.w3.org/2001/XMLSchema#string)
URI http://www.ipas.ecs.soton.ac.uk/events_trent500.owl#SDMServiceEvent_4235110	oil level low (datatype http://www.w3.org/2001/XMLSchema#string)	URI http://www.ipas.ecs.soton.ac.uk/events_trent500.owl#Trent500_71307	broken (datatype http://www.w3.org/2001/XMLSchema#string)	URI http://www.ipas.ecs.soton.ac.uk/events_trent500.owl#Oil_Qty_Trans_8TJ146RBV1_MF002329	Found board to have a broken trace that would cause intermittant readings. See attachements for Report. (datatype http://www.w3.org/2001/XMLSchema#string)
URI http://www.ipas.ecs.soton.ac.uk/events_trent500.owl#SDMServiceEvent_4328871	failed to start/start problems (datatype http://www.w3.org/2001/XMLSchema#string)	URI http://www.ipas.ecs.soton.ac.uk/events_trent500.owl#Trent500_71008	no fault found (datatype http://www.w3.org/2001/XMLSchema#string)	URI http://www.ipas.ecs.soton.ac.uk/events_trent500.owl#null_EECS500-07_BX280657	No faults found with unit. RR have agreed to this being classified as a Confirmed - NOT LRU return. M/G Strip Report 144494 / Repair Card RQ6552 covers this return (datatype http://www.w3.org/2001/XMLSchema#string)

Fig. 7. Result of SPARQL Query for question one.

However, the following issues are found. (1) There is no mechanism to obtain statistics and measure of support and confidence involved with the output. (2) Aggregate functions, count, minimum and maximum cannot be found using SPARQL queries and hence special characteristics of events cannot be studied.

6 Discussion

Using only Semantic Web tools, it is not possible to obtain statistical measure associated with the occurrence of the aeroengine parts. There is no mechanism to visualize the distribution of aeroengine parts in the data graphically. As a result, understanding the frequency of each part becomes difficult. Further, this impairs the ability to concentrate on single part or a particular type of incident for instance, high, low, medium criticality events. In comparison, the business intelligence approach has definite statistical measures associated with each aeroengine part and each attribute of the data. Thus one can easily concentrate on a particular attribute of interest. The data presented is extremely user friendly and presented using various visualisation and summarisation methods. However, the data in Semantic Web is intelligent and has an inherent association property between its attributes in the form of child and parent class. This can be used to advantage if we can apply data mining techniques using this property. For instance, features of the aeroengine part can be studied as an individual part or as a part of the parent part and its distribution can be well understood. It can help to find the problems associated with an aeroengine part by traversing back to the root or parent class or traversing in the front to the child class given a part.

Currently, in data mining each aeroengine part and attribute is considered to be separate. For understanding the features of the aeroengine part or an event, complex nested queries need to be run.

7 Conclusions

In this paper, we studied the problem of knowledge transfer from maintenance to design in an aeroengine manufacturer. A document repository is created to provide an access point to a wide array of documents. The problem of knowledge discovery is studied using the tradition data mining approach in business intelligence and the Semantic Web. It is found that while the tools from the business intelligence community is more mature and can provide statistical confidence, the Semantic Web can be used to assist data mining techniques with ontological rules about the underlying dataset.

8 Acknowledgements

This research was undertaken as part of the IPAS project (DTI Grant `TP/2/IC/6/I/10292`). The authors would also like to thank the project partners for providing us with data and ontologies. Specifically, we would like to thank Alymer Johnson and Santosh Jagtap from the Engineering Design Centre, Cambridge University for the user requirement analysis.

References

1. Harrison, A.: Design for service – harmonising product design with a services strategy. In: Proceedings of GT2006, ASME Turbo Expo 2006: Power for Land, Sea and Air, Barcelona, Spain (2006)
2. Wills, G., Fowler, D., Sleeman, D., Crowder, R., Kampa, S., Carr, L., Knott, D.: Issues in moving to a semantic web for a large corporation. In: Proceedings of 5th International Conference on Practical Aspects of Knowledge Management (PAKM). Volume 3336 of Lecture Notes in Artificial Intelligence., Springer (2004) 378–388
3. Wong, S.C., Crowder, R.M., Wills, G.B., Shadbolt, N.R.: Knowledge engineering - from front-line support to preliminary design. In Brailsford, D.F., ed.: Proceedings of ACM Symposium on Document Engineering (DocEng), Amsterdam, The Netherlands (2006) 44–52
4. Gao, Y., Zeid, I., Bardasz, T.: Characteristics of an effective design plan system to support reuse in case-based mechanical design. Knowledge-Based Systems Knowledge-Based Systems Knowledge-Based Systems **10**(6) (1998) 337–350
5. Khadilkar, D.V., Stauffer, L.A.: An experimental evaluation of design information reuse during conceptual design. Journal of Engineering Design **7**(4) (1996) 331–339
6. Jagtap, S., Johnson, A., Aurisicchio, M., Wallace, K.: Pilot empirical study: Interviews with product designers and service engineers. Technical Report 140 CUED/C-EDC/TR140- March 2006, Engineering Design Centre, University of Cambridge (2006)