HOW ENGINEERING DESIGNERS RETRIEVE INFORMATION

K. Wallace

University of Cambridge
Department of Engineering
e-mail: kmw@eng.cam.ac.uk

Keywords: engineering design, information retrieval

Abstract: *Engineering design can be modelled as an information processing activity, so one way to improve the performance of engineering designers is to ensure that the information they require to progress their designs has been captured and stored – and can be retrieved easily. This paper draws together a selection of results from empirical studies into information capture, storage and retrieval undertaken by researchers at Cambridge Engineering Design Centre. These studies, all undertaken in collaboration with two aerospace companies, started in the early 1990s and have continued up to the present day. Two of the most significant conclusions from the research are: first designers retrieve most of their information from colleagues; and second that for many information needs designers are not able to form explicit queries.*

1. INTRODUCTION

The industrial world is changing rapidly and engineering companies are under considerable pressure to adapt in order to remain competitive in a rapidly changing global economy. One of the consequences of this is an increasingly mobile workforce, with experienced engineers frequently moving to different jobs within companies and to new jobs in other companies. Also because of the age profile of the workforce in manufacturing companies in the developed world, many experienced engineers are coming up to the retirement age. In the past, when industrial companies were more stable, much of the experience of these senior engineers would be passed on to younger engineers, who would remain for an extended period with the companies. Because of the current transient situation, this is not happening to the same extent. As the experience of senior engineers is seldom formally recorded, when they retire their knowledge is lost to the companies.

Because the decisions made during the design process influence all downstream costs, the effectiveness of an organisation's design team has a crucial impact on its profitability. The engineering design process can be broken down into phases, e.g. conceptual design, embodiment design, detail design, and then into smaller working steps and activities. These activities can be modelled as a sequence of problems that have to be solved – the direction of the overall solution being determined by the many individual decisions taken by members of the design team. Problem solving is essentially information processing. One starts with either a concrete or vague statement of the problem (input information); processes that information using information recalled from memory and retrieved from external sources; arrives at a possible solution (output information); and then makes a decision as to whether or not that solution is satisfactory. If it is, one proceeds to the next problem solving step; if it is not, the previous step must be repeated as many times as is necessary until a satisfactory solution is found (iteration).

To obtain the required information to solve a problem, the first step is to search one's own memory to see if one has the information already available from having tackled similar or related problems in the past. If sufficient information is not available, or it has been forgotten, one has to retrieve the information, assuming it has been stored, from external sources. These can be the memories of other people or information captured and stored in some form, e.g. documents, drawings, databases, models, etc. Personal experience teaches us that one of the most effective ways of retrieving the required information is to talk to those individuals who have the relevant experience. Even if they do not have the information you require, they can often direct you to a relevant source, whether that is another person or a repository of recorded information, e.g. a particular document. Because of the largely technical nature of the engineering design process and the efforts to systematically record information, it was frequently assumed: (1) that documents and drawings (models) were the main source of information for designers; and (2) that information retrieval was well understood, did not cause particular problems, and did not take up much of the working day. However, there were no quantified data to support these assumptions.

In engineering companies a great deal of time is spent writing and storing reports. This is not a popular task and is usually done under pressure at the same time as work starts on the next project. There are two reasons for recording information: first to provide an audit trail in the event of some future failure and consequent liability; and second to store information for future reuse. As it was commonly assumed that frequent use was made of documented records, large sums of money were spent on storing documents electronically and providing improved access to them through, for example, keyword searches engines.

With almost limitless storage on the horizon, computer support systems would appear to offer the ideal solution. However, although storing information may not be a problem, knowing *what* information to store, and how to structure and retrieve it, are key issues that need to be addressed. The focus of this paper is on information retrieval.

Engineering design therefore involves processing large amounts of information and in this paper the sources of this information will be limited to:

- Persons (memories)
- Documents (alpha-numeric texts)
- Drawings (models)
- Databases.

To retrieve information a designer, either implicitly or explicitly, has to:

- Establish need (for information)
- Form query
- Identify source
- Extract information.

The research at Cambridge started with some generic questions including:

- What are the information needs of engineering designers?
- What is the nature of their information queries?
- What sources do they use to retrieve their information?
- How do they extract information from the sources they select?
- How much time do they typically spend retrieving information?

These questions, and many more, have been addressed over the past 20 years at the Cambridge Engineering Design Centre. Some of the studies that addressed these questions are now described briefly.

2. EMPIRICAL RESEARCH STUDIES

Marsh started his research project in 1993. Two the objectives of his research were to identify the sources from which designers obtained their information and to determine how much time they spent doing so. An observational study was undertaken. Three groups of four designers were observed undertaking their normal design tasks. This yielded 51 person-days of data for analysis, captured to a resolution of one minute on 1530 data capture sheets [1].

Ahmed started her research project in 1997. One of the objectives of her research was to understand the knowledge needs of novice designers. Part of her empirical research involved analysing 11 discourses between novice designers and experienced designers. These discourses were captured in the same aerospace company that Marsh worked with. In total 633 queries between novice and experienced designers were analysed [2].

del-Rey-Chamorro started his research in 2000. One of the aims of his research was to understand how designers retrieved information from documents, both in hard copy and electronic formats. This empirical study was undertaken in a different aerospace company. Semi-controlled experiments based on a realistic design case study (a military aircraft flight control surface) were undertaken. Sixteen experienced designers took part in the experiments which produced 146 retrieval episodes and 1278 segments for analysis [3].

Aurisicchio also started his research in 2000. One of the aims of his research was to study the nature of information queries and their subsequent information searches. Two main studies were undertaken in the same aerospace company as Marsh and Ahmed worked with. These studies were first a diary study and second observations with shadowing. During the shadowing ten engineering designers were observed for seven hours each. The two studies produced details of 486 original information queries [4].

The following sections present some of the results selected from these studies under the headings: Establish Need; Form Query; Identify Source; and Extract Information.

3. ESTABLISH NEED

From the data gathered from his experiments, del-Rey-Chamorro identified the following three types of information need in descending order of generality:

- Exploring a topic
- Understanding the issues
- Defining product specifications.

Designers need to explore a topic when they have very little previous background knowledge of that topic. The main characteristic of this type of information need is that designers do not know what sort of information is relevant before starting a search, for example:

"We are going to work on this honeycomb construction now. I have never worked with honeycombs before. I do not know what is relevant and what is not. I need to learn a bit more about what is involved in this type of construction."

Once designers have acquired some knowledge on a topic, their information needs change to understanding issues. At this level, designers have a better idea of what they are looking for. They recognise what they are looking for if they see it, but they are still likely to find difficulty in verbalising their needs beforehand, for example:

"I identified that crack propagation is a problem in honeycombs. I would like to know more about what are the risks of the crack propagation in my design case and how these risks could be mitigated."

At the lowest level of generality, designers need to define product specifications (characteristics). In this type of information need, designers know what they are looking for and they are able to verbalise it. These information needs are related to the definition of production specifications of components, for example:

"I am going to select now the type of aluminium alloy that meets the requirements of this honeycomb construction."

By analysing all the data from his experiments, del-Rey-Chamorro showed that the designers in his study spent around 20% of their time exploring a topic; 15% or their time understanding issues; and 65% of their time defining product specifications. An important conclusion from this is that for around one third of their time they are unable to form explicit queries and have to adopt information extraction methods that take account of this.

The type of empirical studies undertaken by Aurisicchio were different, i.e. a diary study and observations, rather than experiments, and took place in a different aerospace company. He identified, from a different perspective, the following three types of need in descending order of generality:

- Reasoning
- Deliberation
- Retrieval-Recognition (Finding).

Reasoning entails making an inference. The response to a reasoning query is logically structured information, for example:

"How can I retain the seal in place?"

Deliberation entails following paths of inference and weighing arguments. The response to a deliberation query is a *network* of queries, responses and arguments, for example:

"Can we increase the outer diameter of Engine XX's shroud tube?"

"How much oil does Engine XX scavenge if the outer diameter of the shroud is increased by 2mm?"

"What is the impact of increasing the outer diameter of the shroud tube on the lower splitter fairing design?"

Retrieval-recognition entails simply identifying and finding the information one needs. The response to a retrieval-recognition query is data, information or logically structured information, for example:

"What material does Engine XX use for this part?"

By analysing all the data from his diary study and observations with shadowing, Aurisicchio showed that engineering designers spent around 30% of their time on reasoning queries; 20% of their time on deliberation; and 50% of their time on retrieval-recognition (finding). Based on this broader study, the engineering designers observed spent around half their time on higher level information retrieval activities, i.e. those queries where they are not simply looking for a straightforward answer to a well defined query.

Bearing in mind the different types of study, undertaken in two different aerospace companies, and the relatively small data sets, the results from del-Rey-Chamorro and Aurisicchio are in broad agreement. An important conclusion is that engineering designers spend a considerable amount of information retrieval time (between one third and a half) on higher level information retrieval activities. For these activities they are unable to explicitly form a query, i.e. a keyword search would be of little use – even if the information had been stored in the first place.

4. FORM QUERY

From the discussion above, it is clearly important to support information retrieval where it is not possible to explicitly form a query. However, it is reasonable to ask if this is a more widespread problem. The research undertaken by Ahmed focused on the different ways that experienced engineering designers and less experienced ones (novices) tackled design problems – including how they retrieved their information. Clearly experienced designers have a greater range of experiences stored in their memories on which they can draw. However, these were not the only issues. By analysing the queries that novice designers addressed to experienced designers, Ahmed showed that only 35% of their queries were clearly stated, i.e. the novices knew what information they required and what precise questions to ask.

When analysing the discourses between novice and experienced designers, Ahmed identified five query patterns:

Explicit question and answer: the trainees asked a question and the interviewee simply answered it (35%), for example:

Query: *“What is delta P?”*

Response: “*The change in the pressure*.”

Rephrased or irrelevant queries: the interviewee rephrased a question or statement, or described it as irrelevant (10%), for example:

Query: *“What is the relevance of ambient pressure to the upper pumping level pressure?”*

Response: *“We are only interested in the ambient pressure under normal operation of the system - we should be considering the typical pressure.”*

Additional information provided: the experienced designers offered additional information on top of that necessary to simply answer a question or respond to a statement (15%). As a consequence, the number of queries was far less than the number of responses in topics where a lot of additional information was provided, for example:

Query: *“Is it a case of saying we want a pump delivery pressure rise?”*

Response: *“That's right for simple checks.”* The designer then provided additional information by explaining how to read graphs that showed how pump delivery pressure varies: *“Graph shows how the pressure varies, you can make an assumption from the graph if the data is not available from suppliers. A twenty-five percent increase in pressure decreases the flow by fifty percent. This is what is plotted on the graph.”*

Statements: No explicit questions were asked as the trainees simply expressed statements (30%), suggesting that the trainees required further information but were unsure of the questions to ask, for example:

Statement: *“The actual high-pressure pump requirements would be quite high.”*

Confirming queries: The novices were observed to make queries to confirm the answers of the experienced designers (10%), for example:

Experienced designer: *“This normally coincides with the maximum temperature and is towards the descent of the flight.”*

Novice designer: *“So, this is at the descent of the flight and at maximum temperature?”*

A question-based method to help novice designers become more aware of what they need to know, along with suggesting questions to ask, has been proposed and tested by Ahmed based on the eight strategies she observed that experienced designers adopted when tackling design tasks. The eight strategies are Consider Issues; Aware of Reason; Refer to Past Designs; Question is it Worth Pursuing; Question Data; Keep Options Open; Aware of Trade-Offs; and Aware of Limitations [5].

These findings suggest that supporting novice designers by simply making information available in a repository, e.g. a database, may not be enough – they also require support in identifying *what they need to know* and *what questions to ask.* These conclusions are also true for more experienced designers when they are working on higher level information retrieval needs, e.g. exploring a topic and understanding the issues.

5. IDENTIFY SOURCE

In the early 1990s it was assumed that designers extracted most of the information they needed from documents and drawings that stored information about past and current projects. To determine the validity of this assumption, Marsh, in collaboration with the University of Bath, undertook a survey to establish document usage. The company maintained 31 *series* of documents. The results were normalised to estimate how frequently an *average* designer referred to these documents and this demonstrated conclusively that documents are referred to far less than was previously assumed. The two most frequently referred to series were: (1) procedures to be followed to meet specific requirements; and (2) materials selection – and on average these were only referred to once every two weeks. The most surprising finding was that documents in over half the document series were effectively not referred to at all.

Possible reasons why documents are not referred to include: (1) the document indexing is poor and the information cannot easily be retrieved; and (2) the documents do not contain the required information or it is out-of-date. These are important issues but the main question was: How do engineering designers obtain their information if they do not use documents?

The data gathered by Marsh from his observational study were carefully analysed and one of the key findings was that for nearly 90% of their information queries, designers used another person as their source – and documents, drawings and other repositories were only used for 10% of their queries. Two people are therefore involved in each information retrieval episode: the person asking the question and the person answering it. In nearly 80% of the cases observed, the information was provided from the person’s memory. This finding is supported by Court who also found that, in general, designers made extensive use of their memories to retrieve information rather than searching in documents [6].

Marsh also observed that in the dialogues that took place the persons approached for information quickly established the context of the query – and on around 50% of the occasions revised the original question being asked. This finding reinforces the findings of del-Rey-Chamorro and Ahmed that frequently even experienced designers do not know exactly what information they are seeking and cannot form precise queries.

Aurisicchio started his research in the same aerospace company as March seven years later. In the intervening period, many IT improvements had been introduced by the company including the widespread use of email and the setting up of a company intranet providing on-line access to documents, drawings and databases. It was interesting to determine whether the sources used by designers had changed as a result. His data showed that on 70% of the occasions when a designer needed information another person was approached, c.f. Marsh 90%. Drawings, reports and databases were each used as sources on around 10% of the occasions. The main source had not changed as much as might have been expected – though the way a person was contacted had. On 65% of the occasions, the person was contacted face-to-face, on 25% by telephone, and on 10% by email.

It is clear that the strongly preferred source of information is another person. It is therefore important to establish why this is so, and to determine whether other information sources can be improved to reduce the valuable time taken up by those involved in these discussions to retrieve information.

6. EXTRACT INFORMATION

Once a promising source has been found, the required information has to be extracted from it. This should not be difficult if one knows exactly what one is looking for, i.e. one can form a precise query, and the information in the source is well indexed or can be rapidly searched electronically. However, from the discussion so far, it is clear that designers are not always fully aware of the information they require and in many instances cannot form precise queries, e.g. select keywords. This is clearly one of the reasons why designers approach colleagues so frequently since, during a dialogue, the context of the query can be established and the queries can evolve. But Marsh showed that these dialogues take up around a quarter of the working day. If this is to be reduced, then more information must be extracted from non-human sources. It is therefore important to determine just how designers extract information from documents, whether in paper or electronic formats.

del-Rey-Chamorro found that designers adopted three strategies for extracting information from documents:

- Discovering
- Recognising
- Finding.

Discovering is the retrieving strategy used by designers when they do not have tacit search criteria for the required information. The designers simply read through documents to find explicit statements related to their information needs. For example, one designer was searching for issues related to the design of honeycombs. Stress corrosion is a relevant issue, but he only discovered this when he came across a statement in a document explicitly indicating its importance. This was the strategy most frequently used when designers were *exploring a topic*.

Recognising is the strategy used by designers when they have tacit search criteria, but are not able to verbalise them. This retrieving strategy consists of scanning through documents and focusing on those promising pieces of information that match the tacit criteria. In these cases, if designers are presented with relevant chunks of information, they are able to recognise them. For example, one designer was looking for the principles of cracking in metallic structures under stress and corrosion. This designer was not able to verbalise what he understood by *principles*, but he had a feeling for what he was looking for. He went into a document and concentrated on recognising the information that could match these tacit criteria. The information finally retrieved was related to threshold stress, residual and assembly stress, and protective treatments. This was the strategy most frequently used when designers were *understanding the issues*.

Finding is the strategy used by designers when they can verbalise their search criteria. This retrieving strategy is based on selecting a precise set of keywords and synonyms. For example, one designer was looking for blind bolts for the attachment of the spigot to the flying control surface. The context in which these bolts were going to be used was known, so the designer was able to verbalise the requirements of the blind bolts and establish precise keywords. This was the strategy most frequently used when designers were *defining product specifications*.

An important insight to emerge from del-Rey-Chamorro’s research was the importance of scanning as an extraction method. When one doesn’t know exactly what one is trying to extract and cannot establish precise keywords, then documents tend to be scanned until relevant information is discovered or recognised. It was clear that designers preferred to scan paper documents rather than electronic ones.

7. DISCUSSION

It is clear that some information retrieval needs are broad and some are very specific. The general, higher level needs are most important during the conceptual design phase and are the most difficult to satisfy. Between 30 and 50% of a designer’s information needs can fall into this group.

Particularly when exploring a topic, and to a lesser extent when understanding issues, designers cannot form explicit queries; so appropriate information extraction strategies have to be used. Most current electronic retrieval systems are based on the assumption that queries can be precisely stated, e.g. in the form of keywords.

Possibly one of the most surprising findings was the reliance of designers on colleagues for retrieving information (Marsh 90% and Aurisicchio 70%). Clearly when one needs to satisfy high-level information needs and one cannot form explicit queries, it is not surprising that one of the best strategies is to enter into a dialogue with a respected colleague.

Reasons why another person is consulted include:

- Speed
- Context
- Query evolution
- Availability (memory)
- Trust
- Confidence.

If the colleague is close to hand and is likely to have the information, this is often the quickest source of information – even for the simple retrieval of factual information, e.g. material data, that could easily be extracted from documents or databases. However, for higher level needs, it is important to establish the context of the need and this is most easily done through a dialogue with another person. During this dialogue the original query evolves – and the information finally supplied can be quite different from that which was initially sought. More information than people realise is not documented and is only stored in the heads of individuals. Marsh showed that in nearly 80% of cases the information provided came from the memory of the person consulted. A further issue is trust. Even if it is known that the required information is available in a document somewhere, there is a tendency not to trust documents, believing them to be out-of-date. If one asks a trusted expert, one's belief in the information being correct and up-to-date is much higher. Also, one's confidence in how one is tackling a design task is greatly enhanced through a fact-to-face discussion with a respected colleague. The support gained from sharing a problem must not be underestimated.

Similar results have been reported from other studies. For example, Frankenberger, who undertook a protocol study in industry, noted: "Very often, the consultation of colleagues in the design process compensates for lack of experience". He also reported that the single greatest factor (30% of cases) attributed to wrong decisions and false analysis was the lack of information rather than the lack of experience [7].

Consulting colleagues for information clearly involves the time of two people and means less time for the main activity of designing. With the aim of making the design process more effective and efficient, anything that can be done to reduce the number of dialogues or make them more efficient should be done, e.g. through specific training. One way is to encourage designers to obtain more information from documents and drawings. del-Rey-Chamorro showed that when doing this either one needed to be able to establish clear keywords or to scan documents. Current electronic retrieval systems do not readily support scanning.

From Marsh's data, it was possible to work out how much time on average a designer spent each day acquiring and giving information and it turned out to be around 25%, with the remaining 75% being divided between design time (65%) and meetings (10%). This division of time would appear reasonable if each of these activities occurred as a continuous period each day. However, information requests do not take place during one continuous period – they occur randomly throughout the day and cause significant fragmentation. It was calculated that there was only a 50% chance of working for eight minutes without needing to obtain or give information. This conflicts with the requirement that creative design work needs intense, uninterrupted periods of concentration. The time spent searching for information is supported by other studies, e.g. the one by Rodgers who found that designers in a telecommunications company spent 20-30% of their time searching for information [8].

Another important insight to emerge is that the *journey* towards finding the information is as important as actually finding the required information, as the following extract from one of del-Rey-Chamorro's transcripts shows:

"I am scanning through this chapter to become aware of the issues involved in the construction of this honeycomb assembly. There is a figure here that has caught my eye. It shows water drainage for flying control surfaces. Obviously water getting into a closed area such as my honeycomb assembly is something we have to take into account. So basically this figure reminds me that allowing water out is another issue for the design of honeycomb construction."

Research at Cambridge has also been directed at the capture and storage of information in a simple graphical form that is easy to scan. This has resulted in a software tool known as the Design Rationale editor (DRed) that captures a design rationale as it is being developed by a designer [9]. This tool is now being widely adopted throughout Rolls-Royce plc.

8. CONCLUSIONS

In order to remain competitive, there is considerable pressure on industrial companies to improve their design processes. A key area to focus on is the effectiveness and efficiency with which their engineering designers retrieve the information they require to progress their design tasks.

Information retrieval involves four stages and some of the main conclusions for each stage can be summarised as follows:

Stage 1 – Establish Need: Between 35 and 50% of a designers information needs are higher level ones that involve exploring a topic and understanding the issues.

Stage 2 – Form Query: Only when retrieving specific information can designers form precise queries and used indexes or keyword searches. For vaguer queries other extraction strategies have to be used.

Stage 3 – Identify Source: The most important source of information, by a long way, is another person – even for factual information that could easily be retrieved from documents or databases. The reasons for this appear to be speed, context, query evolution, information only available from the memories of others, trust, and confidence. Up to a quarter of the working day is spent acquiring and giving information

Stage 4 – Extract Information: When not consulting a person, the ability to scan documents and drawings appears crucial, and current electronic information support systems do not support scanning very effectively.

These conclusions are worrying when seen in the context of the transient nature of modern industrial companies. Experienced staff now move far more frequently within a company or to another company than they did in the past, and because of the age profile of designers in many industries, many senior ones are retiring – taking the knowledge stored in their heads with them. For these reasons, the opportunities to consult colleagues will therefore diminish. If there are fewer experts available to consult in the future, designers will have to rely more on computer-based information systems. It is therefore important to find new ways of storing information and experience in order to access effectively previous solutions and the rationale behind them.

Currently computers are very effective when one can form precise queries and the information has been clearly indexed or can be rapidly searched, but far less so when one is addressing higher-level information needs. For these needs, when one does not know exactly what one is looking for, the ability to scan documents and drawings – and to trigger memories – are important. There is no doubt that more information could be obtained from electronic sources, thus reducing the amount of time taken up with person-to-person dialogues – but one must not forget the social benefits of such dialogues, e.g. the increased confidence gained from sharing ideas.

It is therefore urgent that information is captured and stored in electronic information support systems in a manner that will make it easy to retrieve and, once retrieved, trusted.

Acknowledgements

This research was funded by the Engineering and Physical Sciences Research Council (EPRSC) and the BAE SYSTEMS/Rolls-Royce University Technology Partnership (UTP) for Design. The UTP is a research collaboration between the two companies and the Universities of Cambridge, Sheffield and Southampton. In particular, the author acknowledges the support of Dr Michael Moss from Rolls-Royce plc.

References

[1] Marsh, J.R. *The capture and structure of design experience*, PhD Thesis, Cambridge University Engineering Department, 1997

[2] Ahmed, S. *Understanding the use and reuse of experience in engineering design*, PhD Thesis, Cambridge University Engineering Department, 2001

[3] del-Rey-Chamorro, F.M. *Understanding the process of retrieving information in aerospace design*, PhD Thesis, Cambridge University Engineering Department, 2004

[4] Aurisicchio, M. *Characterising information acquisition in engineering design*, PhD Thesis, Cambridge University Engineering Department, 2005

[5] Ahmed, S., Wallace, K.M. *Developing a support system for novice designers in industry*, Proceedings: International Conference on Engineering Design (ICED 01), IMechE, 2001, pp 75-82

[6] Court, A., Culley, S., McMahon, C. *Information access diagrams: a technique for analysing the usage of design information,* Journal of Engineering Design, Vol 7 (1), 1996, pp 55-75

[7] Frankenberger, E., Badke-Schaub, P. *Influences on engineering design in industry,* Report, TH Darmstadt, 1996

[8] Rodgers, P.A. *Capture and retrieval of design information: an investigation into the information needs of British Telecom designers*, Cambridge University Engineering Department Report, CUED/C-EDC/TR5, 1997

[9] Bracewell, R.H., Ahmed, S., Wallace, K.M. *DRed and design folders: a way of capturing, storing and passing on - knowledge generated during design projects*, Proceedings: Design Automation Conference, ASME, 2004