Highly focused document retrieval in aerospace engineering : user interaction design and evaluation

PETRELLI, Daniela, LANFRANCHI, Vitaveska, CIRAVEGNA, Fabio, BEGDEV, Ravish and CHAPMAN, Sam

Published version

PETRELLI, Daniela, LANFRANCHI, Vitaveska, CIRAVEGNA, Fabio, BEGDEV, Ravish and CHAPMAN, Sam (2011). Highly focused document retrieval in aerospace engineering : user interaction design and evaluation. Aslib Proceedings, 63 (2/3), 148-167.

Highly Focused Document Retrieval in Aerospace Engineering: User Interaction Design and Evaluation

Purpose: This paper describes the preliminary studies (on both users and data), the design and evaluation of the K-Search system for searching legacy documents in aerospace engineering. Real-world reports of jet engine maintenance challenge the current indexing practice, while real users' tasks require retrieving the information in the proper context. K-Search is currently in use in Rolls-Royce plc. and has evolved to include other tools for knowledge capture and management.
Design: Semantic Web techniques have been used to automatically extract information from the reports while maintaining the original context, allowing a more focussed retrieval than with more traditional techniques. We combine semantic search with classical information retrieval to increase search effectiveness. An innovative user interface has been designed to take advantage of this hybrid search technique. The interface is designed to allow a flexible and personal approach to searching legacy data.
Findings: The user evaluation showed the system is effective and well received by users. It also shows different people look at the same data in different ways and make different use of the same system depending on their individual needs, influenced by their job profile and personal attitude.
Research limitations: This study focuses on a specific case of an enterprise working in aerospace engineering. Although the findings are likely to be shared with other engineering domains, e.g. mechanical, electronic, the study did not expand the evaluation to different settings.
Value: The study shows how real context of use can provide new and unexpected challenges to researchers and how effective solutions can be then adopted and used in organizations.

Keywords: organizational archives; semantic web technology; information retrieval; user studies; user interface design; user evaluation.

Classification: Case study

1 Introduction

Organisational memory, the ability of an organisation to record, retain and utilise information from the past to bear upon present activities (Stern, 1995), is a key issue for large organisations. The possibility of observing and reflecting on the past is particularly valuable in highly complex domains as it can inform and sustain decision-making. Civil aerospace engineering is one example: the life cycle of a jet engine can last 40-50 years from initial conception until the last engine is removed from service.

During this long product lifetime a vast amount of information is accumulated. When a new engine is designed, it is important to reflect on previous technical solutions, to identify what has been successful and what instead needs to be revisited. In this context, even if a procedure for capturing newly created information has been introduced recently (e.g. online databases), the potential value of legacy data of the past 15/20 years is high.

The value is, however, only potential as it has to be balanced by the effort to spend to get to the relevant piece of information. A room (or a computer) full of documents in no particular order would discourage even the keen searcher. Easy access is therefore paramount to make organizational memory useful and used.

Another issue is the currency of legacy data. In the case of organizational memory currency does not relate to how recent the information is, but on the validity it has despite the elapsed time. It has much to do with context, more specifically if the context in which the information was created matches (or is important) to the current context of information use then its currency (and its value) is high (Stern, 1995). Contextualized information goes a step further from being a description of data or facts and becomes knowledge (Ackoff, 1989). Retrieving knowledge gives an advantage on retrieving information as the first step for interpretation (information in context) is already outlined. Retaining, organizing and retrieving knowledge poses a number of challenges over traditional information management (i.e., retention, organization and retrieval of information).

This paper describes the first phase of an effort to introduce digital technology in the capturing, organising, searching and sharing operational experience in a complex organisation. The vision is an integrated tool that supports: knowledge acquisition, organisation, retrieval and sharing of corporate memory, knowledge and expertise. The case study reported is that of Rolls-Royce plc. and their aerospace engineering activity. This paper in particular focuses on the retention of legacy information, its automatic organization into knowledge (i.e. information in context), its retrieval and use.

The paper is organized as follows. Section 2 provides a short overview on systems for organizational memory focussing on semantic solutions. Section 3 describes the domain in terms of information practice and type of documents. Section 4 summarises the requirements of highly focussed search while section 5 describes the design of the K-Search system. The user evaluation and general results are discussed in section 6 and 7 respectively. Section 8 reflects on the findings and concludes the paper.

2. Systems to Support Organizational Memory

The need for technology that supports the capturing, classification and finding of information and knowledge in an organization over a period of time has been recognized for quite some time. One of the first systems to have explicitly addressed the issues of organizational memory is Answer Garden (Ackerman and Molone, 1990) that structured into a browsable hierarchy answers provided by experts to problems posed to a mailing list. The classification was manually done; the evaluation showed a clear benefit as users could find answers straightaway and new knowledge was cumulated as new questions were received. This approach has held interest until now (Pipek and Wulf, 2003), but in recent years Web2.0 technologies to support the capturing and classifying of knowledge while at the same time motivating users to contribute their knowledge has stimulated new research. Wikis have been naturally considered as collaborative environments for creating and editing knowledge (White and Lutters, 2007). Various projects have enriched the Wiki model with semantics (Millard et al, 2006, Krötzsch et al, 2006), focusing on collaborative production of semantic knowledge (Alquier et al, 2009). A drawback of these approaches is that they focus on capturing while the retrieval is limited to browsing through organized hierarchies of links (interested readers can find a review of Semantic Media Wikis in (Millard et al, 2008)). Moreover the need for manual labour makes the use of these techniques for archival material too expensive to adopt in any organization.

At the opposite end of the spectrum lies the retrieval of structured knowledge instead of sparse information. Semantic search has been developed for this purpose and makes use of ontologies and knowledge bases to drive and execute the user's queries, returning the results with ontology-related visualizations (e.g. (Maedche et al, 2003) (Popov et al, 2004) (Castells et al, 2007). What distinguishes semantic search systems is whether they keep (and display) a reference to original documents (by storing triples linked to the documents with the annotations (Guha et al, 2003)) or are limited to storing and displaying ontology triples (Castells et al, 2004; Maedche et al, 2003). Most of these research systems, however, are limited to semantic search only, missing the advantage of full keyword search (Lei, Uren and Motta, 2006), but enterprises have started to understand the advantage of adopting semantic technology to improve the performance of their retrieval systems. However the potential of semantic technology is not exploited and its use is limited to matching query terms to similar concepts (synonyms)[1].

3 Information in Aerospace Engineering

3.1. Information Production and Use

Previous research in the area of information behaviour of engineers has focused on theoretical aspects and aimed at defining models (Ellis, 1997), supporting theories (Sonnenwald, 1999) and assessing social dynamics (Fidel et al., 2004). Instead, we see the study of how information is produced and used (i.e. information flow) as fundamental to inform the design of technology and create systems that support information retrieval in context (Ingwersen and Järvelin, 2005). A number of user studies have been conducted to better understand the users, their tasks, the data and the environment of use. This sub-section summarises part of the results relevant for this paper; a full account of the study is in (Petrelli et al., 2006).

[1] For an extended survey interested readers may refer to http://swuiwiki.webscience.org/index.php/Semantic_Search_Survey (accessed 10.12.2010).

Every Rolls-Royce (RR) jet engine currently in use is monitored via internal and external sensors; the data is sent to the control centre in Derby (UK). Every time a RR jet engine is serviced in any airport around the world, an Event Report (ER) is written by a Service Representative (SR) and submitted to the control centre.

While currently this information is remotely archived in a database by the SR, until a few years ago ERs were sent as email attachments (Word files). ERs are short documents (about one page, Figure. 1) that contain key information on the event (generally in tabular form) such as engine type and number, airline operator, location, event description and actions taken, plus a short natural language text describing the event. The tabular part contains structured information; the text contains diagnoses and actions written in free text. The structured information is captured in the database; the free text is not. In order to fully understand what happened and why, it is necessary to read the document (retrievable through a link stored in the database record).

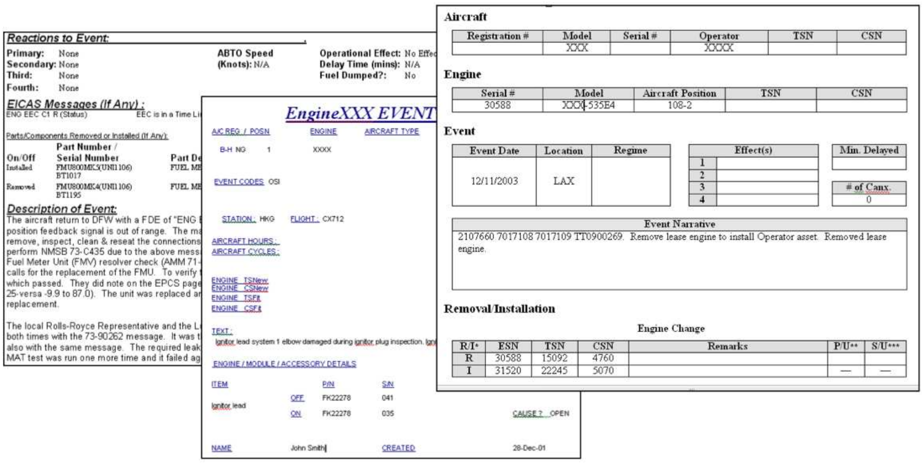

Figure 1 – Three different layouts of Event Reports.

The history of each engine and its components is captured in a series of ERs[2]. Currently, searching for information contained in ERs requites complex queries, several search steps and manual filtering of the results. For example, let us suppose John works in the control centre and is in charge of minimising the impact of unplanned, extra maintenance on scheduled flight time[3]. He has noticed a number of flight delays connected to the installations of a specific component, a Fuel Metering Unit. Many can be the reasons, but if a pattern is discovered, then, by anticipating the installation of the new component in a scheduled service, the flight disruption is

[2] Other documents are produced when, for example, an engine is taken off the wing and serviced in a workshop (Petrelli et al., 2006).

[3] This example, though inspired by observations, is fictitious and has been created for illustrative purposes only; it is not an accurate representation of RR activities.

eliminated with consequent reduction of the costs. John starts investigating the history of a fleet (engines of the same family) by searching which past events have caused:

1) flight delay or cancellation, and
2) required the installation of a new Fuel Metering Unit (FMU); and
3) in which a fuel leak was discovered.

John can retrieve from the database recent events that caused delay or cancellation and required the installation of a new FMU as these are available fields in the database. However, to find out in which cases a fuel leak was discovered he has to manually inspect every retrieved report because the event cause is not made explicit in the record but is written in the free text.

If legacy data is important, John has to manually browse and read old ERs (sent as email attachments and organized in folders) to find the relevant ones, as those are not in the database.

Service engineers are not the only user group interested in ERs. Designers involved in the planning of new engines are interested, for example, in discovering how the component or the part they are responsible for wears with use. An example of designers' information need when redesigning the FMU is "*find all ERs in which the FMU has been replaced on an engine type PQ123*". The type of engine is core as the component has to be designed to fit with the closest parts and different engines have different layouts.

Service engineers and designers both look for FMU on the same data (past ERs), but apply different perspectives and intentions of use.

Both types of user perform recall-oriented search: it is essential that all instances are retrieved. Currently both service engineers and designers spend considerable time searching and reading ERs. However, despite their (often extended) effort, they may end up with just a handful of documents and no confidence they retrieved all the relevant ones.

3.2. Corpus Analysis

Traditional information retrieval is challenged by very short text (e.g. image captions (Smeaton and Quigley, 1996)) or technical documents with very limited vocabularies (e.g. in car manufacturing (Ciravegna, 1995)). To properly direct the design of the retrieval functionalities, a corpus analysis was performed. Although different types of documents are produced and used at RR, we focus here on ERs.

The corpus is composed of 18,097 ERs gathered over 3 years (2000-2003): each and every document describes a single event occurring within a particular engine. Figure 1 shows three very different examples. A large part of the ERs consists of tabulated information; though the information type is largely equivalent in every ER, the presentation can vary considerably. The descriptive text can also vary greatly from few words (Figure 1, right) to some paragraphs (Figure 1, left).

The text uses technical and terse English. The quality of the writing varies greatly, sometimes just a few words among abbreviations and technical jargon, e.g.: "*EICAS msg ENG EEC C1 R with MM 73-70232. Trouble shooting carried out IAW AMM resulting in pin 3 & 4 of FMU excitation coil found to have low resistance to ground. FMU replaced*."

Some terms are very common across the entire corpus, e.g.: *installed, engine, aircraft, removed, hazard, category, nrep, pse, blade, replaced, hkg, esn, csn*, each occur in at least 50% of the ERs. These terms, frequent in users' queries, have limited discrimination power.

Retrieval is made more complex by a variety of terms mapping to the same concept; for example the component '*Fuel Metering Unit'*, is described in the ERs by many synonyms: FMU, Metering Unit, part numbers fmu701mk5 and even unique serial numbers. This variety is due to the personal style of the writer, despite directives from RR to use only terms in the official dictionary.

The context in which each term is used varies too: for example the component part *Fuel Metering Unit* (and its various synonyms) is mentioned in 25% of the documents in different contexts, typically in routine check, e.g. *FMU leak check performed. FMU OK*. Only in a few cases is FMU the subject of the report. To provide an effective retrieval mechanism, it is essential to model the context, i.e. provide a way to distinguish <u>why</u> FMU is mentioned in the ER. For example to effectively find all cases in which the FMU was removed, it is necessary to distinguish them from the documents where the FMU was just tested.

4 Requirements

The analysis of the user and their task, and the analysis of the corpus resulted in a list of requirements that directed the design of both user interaction and retrieval functionalities. The search system should:

1. make legacy data available for contextual search via concepts at limited cost;
2. support the composition of complex, context-rich queries in an easy to understand and simple way;
3. retrieve contextualized information, i.e. only documents where the term occurs in the context specified by the user in the query should be returned;
4. retrieve with high recall, i.e. *all* documents containing knowledge (information in context) relevant to the query;
5. retrieve with high precision, e.g. *only* documents where FMU was removed and not those where FMU was just tested;
6. support users in quickly changing their search strategy or focus;
7. accommodate individual perspectives; different user classes (e.g. service engineers or designers) and their specific terminology should be supported;

8. support the statistical analysis of retrieved results via dynamically generated charts that are:
   – customizable: the user selects dimensions to explore alternative views on the retrieved data;
   – interactive: support drilling into specific aspects captured in the chart;
9. provide multiple views on the same retrieved set, e.g. multiple charts, many documents opened simultaneously, to support data exploration.

The next section describes the design rationale, i.e. how requirements drove the K-Search system design.

5 K-Search Design

5.1 The Design of the Search Functionality

The requirements collected point to a solution for very specific queries, more than traditional IR could provide. Semantic Web Technologies can be used for this purpose, as they enrich document content by overlapping metadata (or tags) onto the free text, thus opening possibilities for automatic processing (Berners-Lee et al., 2001). An ontology that represents the domain is needed to describe the semantics of the tags. The process of annotating content (or tagging) makes the relation between the relevant part of the document and its domain explicit. By associating content and concept it is possible to aggregate synonyms to the same concept (e.g. *Fuel Metering Unit, Metering Unit, FMU*, *fmu701mk5* are all instances of the same concept *component*). Concepts can also be connected through logical statements (e.g. *discoloration* on *blade*). Annotated documents are then stored and retrieved with respect to the semantics they contain, thus supporting highly focused queries.

However, the cost of manual annotation could undermine any project as the real value of legacy data has to be very high to justify the cost of manual annotation. Information extraction, the automatic identification of structured information in unstructured text, can be used to minimise the acquisition costs. In specific domains, ontology-based information extraction can learn from examples of manually-annotated text to automatically annotate additional documents (Petrelli et al. 2005). In this way an information extraction process automatically annotates documents and contextualises pieces of information in the specific domain.

A semantic approach, although powerful, is limited: the search is constrained to the information that matches concepts captured by the ontology, and it requires users to compose queries in Boolean logic, a skill that has to be learned. Conversely, keyword-based information retrieval has the advantage of being flexible - any term can be searched for independent of pre-processing - and straightforward to use: the user just types in terms.

A hybrid approach that fuses keyword-based and ontology-based search is able to combine the advantages of both techniques, providing effective, flexible and focused search that each method alone cannot achieve. In summary, a hybrid search approach would satisfy the user requirements discussed above.

A hybrid search engine, K-Search, was therefore designed and developed: semantic search retrieves contextualized information; keyword-based search captures what is missed by the semantic; and the combination supports highly focussed retrieval. To achieve hybrid search, the data is pre-processed to create two different views: an inverted-index for keyword retrieval as well as an RDF repository for the semantically annotated counterpart. The RDF repository is created automatically by applying information extraction techniques during the pre-processing phase.

At access time, K-Search transforms the user query into two specific formats used to search the two views independently. The retrieved sets of the parallel queries are then combined into a single list of results (the technical detail, and underlying architecture and algorithms can be found in (Bhagdev et al., 2005)).

5.2 The Design of the User Interface and User Interaction

The biggest challenge we faced when designing K-Search's interaction was how to support the formulation of complex Boolean queries in a simple way. Previous research shows that a graphical representation of Boolean logic is not understood by users (Hertzum and Frokjaer, 1996; Shneiderman, 1997), but some highly structured queries must be input if the potential of semantic search is to be exploited.

In discussions with RR engineers and observation of their work we noticed their information behaviour is focused on one case at a time, and they pursue a single thread until all information is collected. This suggests that the OR operator to combine multiple topics in the same query is not used much. The design decision taken was then to reduce the logical expressiveness (to allow only the most frequently used combinations of AND and OR), to favour simplified interaction: different concepts are combined with AND, while different values for the same concept are distinguished by OR.

When the user selects a concept in the ontology (left hand side in Fig. 2), a new line is created on the search panel (right, Fig. 2), and a text field in which to type the value for that concept appears; the (last or currently) selected concept is highlighted in the ontology in italics. Figure 2 shows the concept "*removed component - description*" with the value "*Fuel Metering Unit*". Other values for the same concepts can be added by clicking on the "*[or]*" operator and can be removed by clicking on the "x" icon next to the value. It is therefore possible to simultaneously search for synonyms, e.g. "*Fuel Metering Unit" OR "FMU" OR "Metering Unit".*

Additional concepts are automatically added in AND when other ontology elements are selected. In Figure 2 a second concept "*operational effect*" has been added and two different values have been specified: "*delay"* and "*cancellation*". The semantic query displayed maps to "how many times the removal of a fuel metering unit caused delay or cancellation” and its logical representation is *(part-removed FMU) AND (operational-effect (delay OR cancellation))*.

Figure 2: The part of the K-Search interface used for inputting a query.

Keywords for the free text retrieval are typed in the very top line. This field is always available, i.e. the default input panel resembles the typical free-text search layout.

The query formulation is very flexible: Users can use just the keyword-based search by typing in the first text field, the semantic search described above or a combination of the two as in Figure 2 where the concepts are refined by the keyword ”*fuel leak*”. Changing from one search mode to another is straightforward and users can adopt the strategy they prefer for the task in hand.

The result set contains the ERs where the concepts and the keywords in the query co-occur. It is displayed as a list on the top, right panel of the interface (Fig. 3); each item has a link to the document and the snippet of text that contains the keywords. When selected, individual ERs are shown on the bottom right. Multiple documents can be opened simultaneously, each displayed in a different tab.

Annotations are made evident through colour highlighting (as in (Ciravegna et al., 2002; Dzbor et al., 2003)) and are the means to access advanced features or services (Dzbor et al., 2003; Lanfranchi et al., 2005): for example clicking on “engine ABC” in Figure 3 triggers a query expansion to that specific engine.

The last feature is the automatic generation of graphs from the content of the retrieved document set. The user selects the concept for the grouping and layout (pie or bar chart). The graph in Figure 3 plots the result set by engine type and airport. Each graphic item (each bar in the example) is active: clicking on a bar filters the retrieved set to only those documents with the value matching the bar one.

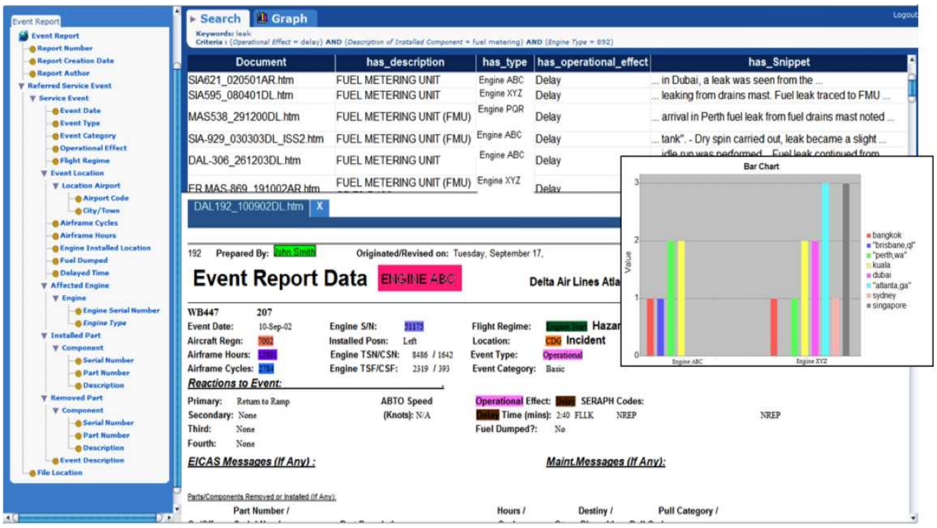

Figure 3 The display of the results and the generated chart.

This final design was reached after a number of user validations of prototypes and some interface features. For example, use of multiple tabs to show results was actually proposed by users in those meetings. Similarly the ontology and the identification of the important parts of the text to be extracted were defined and discussed with users.

6 User Evaluation

6.1 Method

The purpose of the user evaluation was twofold: 1) to assess the usability and acceptability of K-Search; 2) to gain an impression of how the system would be used when installed on the user's desktop. The method adopted was therefore a combination of a controlled setting, i.e. every participant had to carry out the same task, with a more naturalistic approach, i.e. every participant also searched on a topic of their own choice and relevant to their job. This combined approach allows making the most of the two conditions while mitigating their limitations (Petrelli, 2008):

- *assigned tasks* allow assessing the system performance and its usability as all participants perform the same activity and therefore fall under the same (experimental) condition;
- *self selected tasks* support a semi-naturalistic condition that broadens the view on the potential utility of the system in users' everyday jobs.

By adopting this experimental setting we were able to measure and compare performance as well as to obtain an initial understanding of how K-Search would be used in a natural setting, to determine, for example, which parts of ERs designers or engineers look at.

6.2 Procedure

At recruiting time participants were informed about the project objectives and K-Search goals and requested to prepare an individual task.

At arrival, they filled in a personal profile questionnaire and were shown a short video (2 minutes) on the system and its use.

To familiarize themselves with K-Search, participants had to carry out an assisted task: the experimenter sat nearby, and helped and provided explanations where requested. No time limit was given, although participants were encouraged to stay within 10 minutes.

The test task was carried out without any help from the experimenter who, however, was present in the role of an observer.

Both familiarization and test tasks were derived from real tasks discussed by RR employees in previous meetings, they were written as work task simulations (Borlund, 2000).

A short questionnaire aimed at assessing how familiar participants were with the given task was filled in at this point.

Participants were invited to try their self-selected topic next: they were asked to explain what was the task was about and what they aimed at getting. They were observed while translating their goals into interface actions and asked to write down if they thought they were successful in finding an appropriate answer. If problems arouse, e.g. no document could be retrieved, the experimenter explained why and indicated other search strategies, if any were possible.

After this stage participants were requested to fill in an extended user satisfaction questionnaire that covered several aspects of the system, e.g. layout, ease of use, ease of learning, perception of speed and accuracy, range of functionalities, etc. Questions were mainly on a 5 point Likert scale measuring the rate of agreement with a specific statement. They were also invited to write what they would suggest as improvements and conversely, what they liked.

The user evaluation ended with an interview aimed at eliciting explanations about the participants' behaviour, impressions of the system, and at collecting personal opinions.

In advance of the evaluation at the Rolls-Royce premises, a pilot test was carried out at Sheffield University to discover and resolve problems, e.g., failure of the automatic log or unclear questions.

6.3 Data Collection and Analysis

The evaluation took place in two RR premises in April-May 2007. Thirty-two (32) people took part (16 in Derby, 16 in Bristol) and tested the system individually. Each session lasted about 90 minutes and two different classes of users were involved: designers and service engineers.

The data collected was both objective (logs of the interaction, screen activity) and subjective (questionnaires, interviews), and has been analysed quantitatively and qualitatively.

In order to gain a precise understanding of the data collected, the interaction was split in 3 separate phases:

1. input: K-Search input features and functionalities, i.e. semantic search, keyword search or both;
2. search: the success of the retrieval process itself;
3. output: the layout of the results and interaction with the system, the graph visualization and its use.

Following the ISO measures for usability (Van Welie, 1999), effectiveness, efficiency, and user satisfaction were calculated using a combination of objective and subjective data.

6.4 Results

6.4.1 Input: query formulation

Inputting a query required the user to understand the function of the ontology, which value would be meaningful for a concept, how multiple concepts were composed in the query and what would be the contribution of a keyword to the search. Questions examining all these were posed in the final questionnaire.

Overall, participants found it easy to formulate queries mapping to their goal (easy 56%) and only a minority (6%) found it hard, while 38% said it was neither difficult nor easy. Similarly, the experience of composing the query was positive for 62%, who found it intuitive, and only 10% who considered it cryptic, while 28% were neutral.

Two questions addressed the formulation of the semantic search in more detail (Fig. 4). Choosing the concept in the ontology was considered easy/very-easy by 31%, difficult by 28% and average by 41%; specifying a concept value was easy/very-easy for 37%, difficult for 28% and average for 35%. These values seem to show that participants are more confident once a concept has been selected, to in choose its value, as the positive response increased from 31% to 37%. These values are however lower than the general judgement given on the semantic search: this could be explained by some limitations discovered during the test, for example it was impossible to search for a range of values, e.g. from 1998 to 2002, or to know which values would be valid for a certain concept, e.g. the concept 'Operational effect' has valid values like 'delay', 'cancellation', 'none', etc.

Figure 4 Perceived difficulty in selecting a concept from the ontology (left) and assigning it a value (right).

To balance users' opinions, the number of iterations carried out to complete the task was calculated: on average this decreased from 4 in the training phase to 3 in the task with some users going down from a high number of iterations in the training (as much as 8) to just 1 in the task phase. We observed also different strategies, with some users composing the final query in different steps, adding one condition at a time, while others composed them all in a single step[4].

6.4.2 Search: matching the user query

Search effectiveness was measured by successful completion. All participants were successful in the training task, but 3 (9%) failed to retrieve the full set and stopped after finding a few. This seems to be due to "forgetting" part of the task requirements, e.g. requirement to retrieve ER in 2002 and 2003 but searched only for 2002.

The subjective perception of K-Search's effectiveness was measured in the questionnaire by the question "understanding why a report was retrieved was": 68% answered it was easy/very-easy, 3% found it difficult and 29% neither difficult nor easy. K-Search was also considered fast (97% vs. 3% average) and reliable in retrieving relevant ERs (87%, 3% unreliable and 10% average, Fig. 5). When questioned why they felt the system was unreliable or average those participants commented on not being able to retrieve the documents they wanted; sometimes they were looking for documents other than the ERs, i.e. documents that were not in the K-Search dataset; at other times the failure was due to the concept not being modelled by the current ontology and therefore missed by the current system, e.g. cost detail. Both issues have been fixed by adding more document formats and re-applying information extraction on the newly identified concepts.

[4] Precision (P) and Recall (R) have not been used because they would have penalized the cautious behaviour of those participants who did many iterations. In addition P and R had been previously assessed in an extended laboratory test (Bhagdev et al., 2005).

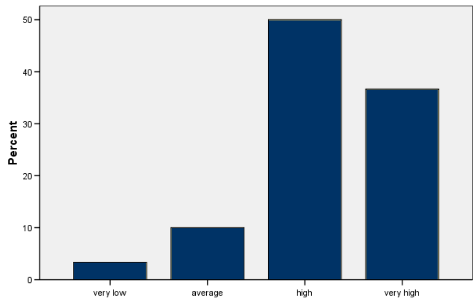

Figure 5 - Perceived system reliability in retrieving the desired results in the allotted time

6.4.3 Output: result exploration

Participants' opinions of the output display and browsing of the results were positive: 77% said it was easy/very-easy to use, 12.9% found it average and 9.7% difficult (see Figure 6).

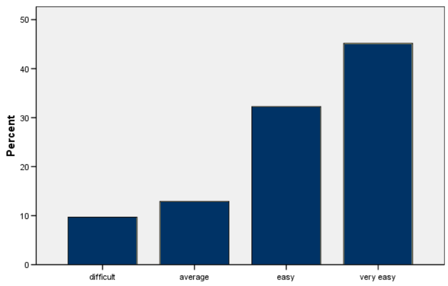

Figure 6 - User satisfaction in browsing the results

A special feature of K-Search's result display is the highlighting of (instances of) ontology concepts in the document that, when clicked, trigger query expansion, and the display of the result set in an interactive chart (see Figure 3). Despite the large number of colours in the documents participants found the highlighting helpful (75%) while only 7% judged it confusing and 16% were neither positive nor negative (Figure 7b). This very high percentage could indicate the value of the highlighting as a means toward a better understanding of the ontology and its mapping onto document content.

The generation of the chart was considered by 75.1% easy to understand, 12.9% average and 9.7% difficult (Figure 7a). Some participants used the chart generation as a way to carry their queries further, for example, instead of adding a new concept to the query they generated a chart with respect to that concept.

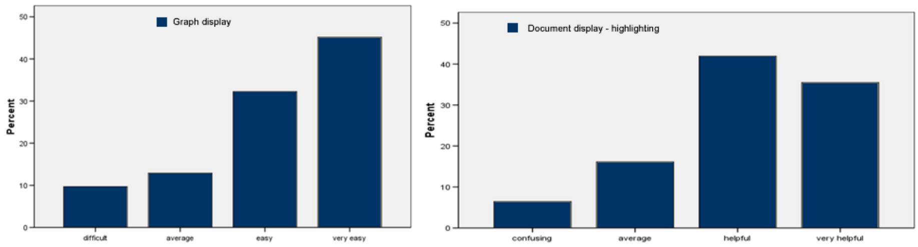

Figure 7 - Perceived layout effectiveness of (a) data result as graph and (b) text highlighting.

7 The Effect of the User's Goal

The flexibility of the interface in accommodating searching styles emerged in several situations. There was no prescription on how to compose a query: some users started with keyword search and then switched, others started with concepts, others combined immediately the two from the start.

The exploration of the results seems based on individual attitude and preference too. Some participants used the chart or the highlighted concepts as query expansion tools; all opened several documents (in tabs) and moved between them to perform comparisons.

The clearest indication of individual use of a common system emerged when analysing the self-selected task. Different users looked for different content in the same documents: at design time ERs are important for the facts they report, but some participants were interested in the cost of the event and not in its occurrence. K-Search failed to satisfy those user class needs, but simple re-training to extract new data allowed those needs to be satisfied in the following version of the system.

Differences emerged between groups. Participants were classified as service engineers (32%), designers (39%) and others (e.g. cost engineers, 29%).

Designers searched mainly for instances of components (37%) and engines (37%); concepts (19%) or event (7%) were present in a minority of queries.

Service engineers looked to the data from the point of view of finding events (35%) and engines (32%), in a much smaller proportion from the point of view of problem (15%) and component (18%).

Participants belonging to other groups had a less distinctive search strategy: 37% started from the component, 37% from the engine, 16% from an event, 10% from concepts and none from a problem (Figure 8).

Figure 8 Search strategies participants used when carrying out the self-selected task, grouped by user class.

Service engineers and designers differ also in search strategy. Service engineers clearly preferred hybrid search, 61%, with 24% preferring semantic and 15% keyword search.

Designers favoured keyword search, 43%, hybrid was used in 30% of cases and semantic only in 27%. The preference for using keyword search could be explained by the potential it offers to explore the data irrespective of classification: designers were able to investigate components from many and diverse points of view.

Other groups used semantic search most, 66%, 24% hybrid, and 15% keyword. A possible explanation is that semantic search is driven by the concepts in the ontology and therefore provides an easier way to search unfamiliar data.

In summary, the user and their task are fundamental to capturing the actual context of use. This in turn is fundamental to directing the design, both interaction and functionalities, that is able to accommodate a wide variety of styles and goals.

8 Discussion

The study was successful in showing the possibilities technology offers for the automatic capturing of archival material into a usable format that can be easily accessed and used by everyone in an organization. This is an indispensable step as the value of legacy documents is uncertain and the cost of manually acquiring knowledge from them is prohibitive (e.g. reading them one by one and inputting the core information in a database). The value of such an effort was made explicit during the user evaluation as more than one participant asked to be able to use the system in their everyday job: one person ended up printing a few of the retrieved documents as relevant for a task in hand but which they had not previously discovered. Having the historical archive available to everyone opens up new uses of this information, particularly by employees who would not normally access it, for example those planning business cases would not look at ERs individually but at digested summaries; being able to break down the summaries into single events and access the facts directly allows in-depth analysis that was not possible before.

The outcome of the user evaluation was used to fix minor issues on the user interface, e.g. search for range of dates, and re-run the information extraction module to capture new knowledge that emerged as relevant. The system was then delivered to Rolls-Royce for actual use. The data set has been progressively enriched with other kinds of documents selected from the large Rolls-Royce archive. By re-creating electronically the rich legacy, the organizational memory is progressively composed into a coherent and organic set that overcomes the fragmentation that occurs when paper copies or different databases are used by different departments in charge of storage.

As with all automatic acquisition techniques, those used by K-Search may also miss some relevant information or misrepresent others, with the consequence that errors are stored in the database used[5]. Users accustomed to the certainty of manually created and validated centralised databases, (nearly) always correct and exhaustive, need to be made aware of the limitations of these techniques. When questioned about this issue participants stated that the advantage of having archives easily accessible outweighs any issues of correctness, accuracy and exhaustiveness.

The quality of the automatically acquired data depends on the initial format of the archive and on the number of steps applied to process archival documents into the repository. The original form of the Event Reports used in this study was Microsoft Word; this was then converted into HTML, on which information extraction was applied. The conversion Word-HTML is a potential source of noise, i.e. errors, that impact on the following information extraction, that in turn can fail to detect features. It goes without saying that the higher the number of the steps the greater the number of the errors: during the acquisition of new archives we came across a repository of files that contained scanned documents on which manual annotations had been made. The original Word documents were no longer available and to include this set into K-Search required converting the scan into HTML. Both the conversion from scan to

[5] Tests showed P=0.85, R=0.83, and F-mes=0.84 (Bhagdev et al. 2008) indicating very good effectiveness of the approach adopted.

PDF and then from PDF to HTML introduced errors. However, once more, being able to search those highly important documents overcame any drawbacks. What is worth pointing out is that strategies could be put in place to help users discover data mismatches and support updating and corrections that will eventually converge the data set to a correct and complete repository (Petrelli et al., 2005)

Since K-Search was introduced for everyday use in Rolls-Royce, it has evolved into a suite of tools (K-Tools) for capturing, managing and organizing knowledge using semantic technology. The whole knowledge lifecycle is now supported: the users can define their own knowledge representation and use self-defined forms to formalize and capture new knowledge (using K-Forms, (Bhagdev et al. 2008)) that is then stored and available for sharing and reuse via K-Search. By adding a way to capture new knowledge, K-Tools allows the recovery of organisational memory from the past and combine it with newly captured knowledge, creating a sustainable and consistent way of linking past information and current issues.

9 Conclusions

To study a case of interactive information retrieval requires considering other elements beside the system: its user, their task, and the data are all equally important in setting up the context of use. When designing such systems, decisions have to be taken on which features are important and what should the system do; evaluations then have to be set up in order to test if the system is as effective as expected.

We showed how the analysis of users and data affected the design of the system in terms of both search functionalities and interaction. An innovative search paradigm that combines ontology-based and keyword-based search in a hybrid system was devised and tested with users in the aerospace industry.

The purpose of the user evaluation was twofold: to measure usability and to investigate how the different user groups would use the system. The results show the interaction design was successful while the data needed to be processed again to accommodate different users' interests. It also showed how users approach the system in very different ways, sometimes showing characteristics of a group, sometimes as singular individuals.

Since the study reported here K-Search has been updated: different document types have been added and the user interaction has been modified where needed (e.g. suggesting possible values for a selected component). After a monitored trial, K-Search is currently in use at RR in Derby.

Acknowledgments. The research reported in this paper was in part supported by the IPAS (Integrated Products And Services) project, funded by UK DTI Department of Trade and Industry and Rolls-Royce plc. Although realistic, none of the examples used correspond to actual situations.

References

Ackerman, M., Malone P. (1990) “Answer Garden: a tool for growing organizational memory.” In Proc. COCS '90 Proceedings of the ACM SIGOIS and IEEE CS TC-OA conference on Office information systems, 31-39.

Ackoff, R. L. (1989) “From Data to Wisdom.” Journal of Applied Systems Analysis, 16, 3-9.

Alquier, L., McCormick, K. and Jaeger, E. (2009) “knowIT, a semantic informatics knowledge management system”. In Proc. of the 5th International Symposium on Wikis and Open Collaboration (WikiSym '09).

Bhagdev, R., Chapman, S., Ciravegna, F., Lanfranchi, V., Petrelli. D. (2005) “Hybrid Search: Effectively Combining Keywords and Semantic Searches”. 2nd European Semantic Web Conference ESWC 2005.

Bhagdev, R., Chakravarthy, A., Chapman, S., Ciravegna, F., and Lanfranchi, V. (2008) “Creating and Using Organisational Semantic Webs in Large Networked Organisations.” In Proceedings of the 7th International Semantic Web Conference – ISWC08.

Berners-Lee, T. Berners-Lee, Hendler, J., Lassila, O. (2001) “The semantic web”. Scientific American.

Borlund, P. (2000) “Experimental Components for the Evaluation of Interactive Information Retrieval Systems”. In: Journal of Documentation , 56 (1) 71-90.

Castells, P., Fernández, M., and Vallet, D. (2007). “An Adaptation of the Vector-Space Model for Ontology-Based Information Retrieval.” IEEE Transactions on Knowledge and Data Engineering, Special Issue on Knowledge and Data Engineering in the Semantic Web Era, 19(2), 261-272.

Castells, P., Foncillas, B., Lara, R., Rico, M., and Alonso, J. L. (2004) “Semantic Web Technologies for Economic and Financial Information Management.” In Proc. of 1st European Semantic Web Symposium (ESWS 2004). Lecture Notes in Computer Science, Vol. 3053. Springer Verlag, Berlin Heidelberg (2004) 473-487

Ciravegna, F. (1995) “Understanding messages in a diagnostic domain”. Information Processing and Management, 31(5) 687–701.

Dzbor, M., Domingue, J. B., Motta, E. (2003) “Magpie - towards a semantic web browser”. 2nd International Semantic Web Conference (ISWC), Sanibel Island, Florida, USA.

Ellis, D. (1997) “Modelling the infrmation seeking patternss of engineeers and research scientists in an industrial environment”. Journal of Documentation, 53(4) 384–403.

Fidel, R., Pejtersen, A. M., Cleal, B., Bruse, H. (2004) “A multidimensional approach to the study of human-information interaction: A case study of collaborative information retrieval”. JASIST, 55 (11) 939-953.

Guha, R. V., McCool, R., and Miller, E. (2003) “Semantic search.” In Proc. of the 12th International World Wide Web Conference (WWW 2003), 700-709.

Hertzum, M. and Frokjaer. E. (1996) “Browsing and querying in online documentation: a study of user interfaces and the interaction process”. ACM Transaction in Computer-Human Interaction, 3(2) 136–161.

Ingwersen, P and Järvelin, K. (2005) “The Turn – Integration of Information Seeking and Retrieval in Context”. Springer.

Krötzsch, M., Vrandecic, D. and Völkel. M. (2006) "Semantic MediaWiki". In Proceedings of the 5th International Semantic Web Conference (ISWC-06).

Lanfranchi, V., Ciravegna, F. and Petrelli, D. (2005) “Semantic web-based document: Editing and browsing in AktiveDoc.” European Semantic Web Conference ESWC 2005, 623–632.

Lei, Y., Uren, V. and Motta, E. (2006) “Semsearch: A search engine for the semantic web.” In Proc. 5th International Conference on Knowledge Engineering and Knowledge Management Managing Knowledge in a World of Networks, Lect. Notes in Comp. Sci., Springer, 238-245.

Maedche, A., Staab, S., Stojanovic, N., Studer, R., and Sure, Y. (2003). “SEmantic portAL: The SEAL Approach.” In Fensel, D., Hendler, J., Lieberman, H. and Wahlster, W. (eds.)

Spinning the Semantic Web. MIT Press. 317-359.
Millard, D., Bailey, C., Boulain, P., Chennupati, S., Davis, H., Howard, Y., and Wills, G. (2008) “Semantics on demand: Can a Semantic Wiki replace a knowledge base?” New Rev. Hypermedia Multimedia, 14 (1), 95-120.
Millard, I., Jaffri, A., Glaser, H. and Rodriguez-Castro, B. (2006) “Using a Semantic MediaWiki to Interact with a Knowledge Based Infrastructure.” In Proc 15th International Conference on Knowledge Engineering and Knowledge Management.
Petrelli, D., Lanfranchi, V., and Ciravegna, F. (2005) “Working Out a Common Task: Design and Evaluation of User-Intelligent System Collaboration” 10th IFIP TC13 International Conference on Human-computer Interaction, INTERACT’05.
Petrelli, D., Lanfranchi, V., Moore, P., Ciravegna, F. and Cadas, C. (2006) “Oh my, where is the end of the context?: dealing with information in a highly complex environment” IIiX 1st international conference on Information interaction in context, Copenhagen, Denmark, October 2006.
Petrelli, D. (2008) “On the role of User-Centred Evaluation in the Advancement of Interactive Information Retrieval.” Information Processing and Management, 44 (1), 22-38.
Pipek, V. and Wulf, V. (2003) “Pruning the Answer Garden: knowledge sharing in maintenance engineering” In Proc. ECSCW'03 Proceedings of the eighth conference on European Conference on Computer Supported Cooperative Work, 1-20.
Popov, B., Kiryakov, A., Ognyanoff, D., Manov, D., and Kirilov, A. (2004). “KIM – A Semantic Platform for Information Extraction and Retrieval.” Journal of Natural Language Engineering, 10, 375-392.
Shneiderman, B. (1997) “Designing the User Interface.” (3rd ed.) Addison-Wesley, 1997.
Smeaton, A. F., Quigley, I. (1996) “Experiments on using semantic distances between words in image caption retrieval.” In SIGIR ’96: Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval, 174–180.
Sonnenwald, D. (1999) “Evolving perspectives of human information behaviour: context, situations, social networks and information horizons.” In Exploring the Contexts of Information Behaviour – proc. 2nd International Conference on Information Needs. 176-190.
Stern, E. W. (1995) “Organizational memory: Review of concepts and recommendations for management.” In International Journal of Information Management. 17–32.
Van Welie, M., van der Veer, G. C., and Eliens, A. (1999). “Breaking down usability.” In Proc. INTERACT99, 613–620.
White, K. and Lutters, W. (2007) “Midweight Collaborative Remembering: Wikis in the Workplace,” CHIMIT ’07 – Computer Human Interaction for Management of IT.

Daniela Petrelli
Department of Information Studies
University of Sheffield
Regent Court, 211 Portobello Street
S1 4DP – Sheffield, UK
d.petrelli@shef.ac.uk

Bio: An expert in Human-Computer Interaction and User Centred Design, Dr. Petrelli is a member of the Information Retrieval group. Her research focuses on methods for and practice of designing interactive information systems to ease their use. She has explored different domains, e.g. multimedia and multilingual, in different contexts, e.g. information professionals, digital archives for scholars and the public, organizational memory and its use in large corporations. Other research interests include the design of innovative devices for accessing personal multimedia memories.

Vitaveska Lanfranchi
Department of Computer Science
University of Sheffield
Regent Court, 211 Portobello Street
S1 4DP – Sheffield
v.lanfranchi@shef.ac.uk

Bio: Dr. Vitaveska Lanfranchi is a research fellow in the Organizations, Information and Knowledge - OAK group at the department of Computer Science. Her research interests are: human computer interaction for web intelligence and the semantic web. Dr. Lanfranchi work focuses on understanding how to make Semantic Web technology usable so that users can exploit all its potential. She has coordinated the design and development of knowledge management systems in both industrial and government settings.

Fabio Ciravegna
Department of Computer Science
University of Sheffield
Regent Court, 211 Portobello Street
fabio@dcs.shef.ac.uk

Bio: Fabio Ciravegna is Full Professor of Language and Knowledge Technologies and coordinates the Organisations, Information and Knowledge (OAK) Group at the Department of Computer Science. His research field concerns methodologies and technologies for the Intelligent and Semantic Web, with focus on Knowledge Management applications in different domains. He is also director of research of K-Now, and director for Research and Innovation for the Digital World.

Ravish Begdev

Knowledge-Now Ltd,
The Innovation Centre,
217 Portobello Street,
Sheffield - S1 4DP, UK
ravish@k-now.co.uk

Bio: Ravish Bhagdev is a PhD student in Computer Science at the University of Sheffield and Senior Analyst at Knowledge-Now Ltd, a spin-off company of the University of Sheffield focusing on supporting dynamic distributed communities in large organizations. He specializes in use of Semantic Web Technology for search and large-scale Knowledge Management.

Sam Chapman
Knowledge-Now Ltd,
The Innovation Centre,
217 Portobello Street,
Sheffield - S1 4DP, UK
sam@k-now.co.uk

Bio: Dr Sam Chapman is the Executive Director and co-founder of Knowledge Now Limited, a growing state of the art knowledge management company which focuses upon applications of advanced semantic technologies, indexing and knowledge acquisition within large organisations. Dr. Chapman has also been involved in many high impact projects both academically and industrially at national and international level. Sam's work and interests focus upon state of the art approaches applied to knowledge management for organisational optimisation.