Case Authoring from Text and Historical Experiences

Marvin Zaluski[1], Nathalie Japkowicz[2], and Stan Matwin[2]

[1]Institute for Information Technology, National Research Council of Canada, Ottawa Ontario, Canada, K1A OR6
marvin.zaluski@nrc.ca
[2] School of Information Technology and Engineering, University of Ottawa, Ottawa, Ontario, Canada, K1N 6N5
{nat, stan}@site.uottawa.ca

Abstract. The problem of repair and maintenance of complex systems, such as aircraft, cars and trucks is certainly a nontrivial task. Maintenance technicians must use a great amount of knowledge and information resources to solve problems that may occur. This paper describes a semi-automated tool that sorts through the mass of information a maintenance technician must consult in order to make a repair, thus helping him decide how to tackle the problem and thereby increasing his efficiency and, possibly, his reliability. Our tool was developed using state-of-the-art Case-Based Reasoning and Information Extraction technologies. More specifically, we developed a semi-automated Case Authoring method that creates a Case-Base in two steps. It begins by extracting knowledge from readily available resources such as technical documents and follows by complementing those cases using individual experiences in the maintenance organization. The case-base developed is a reflection of the knowledge encoded in the technical documentation and an authentication of the cases with real historical instances. Our case authoring approach is applied to the real world in the aerospace domain.

1   Introduction

A variety of Case-Base Reasoning (CBR) applications have been fielded since the idea of CBR was founded. These applications range from helpdesk (Aha 1998) to tutorial applications (Brüininghaus & Ashley 2001). The most important prerequisite in any CBR application is a collection of experiences in the form of a case-base (Bartsch-Sporl et al. 1999). Case authoring is the acquisition of new experiences that are not represented in the case-base. Many CBR applications use manually intensive approaches to represent these experiences in their case-bases. There has been little or no research done to facilitate automatic or semi-automatic approaches to author cases for the case-base (Aha 1998). This paper will describe a semi-automated approach to case authoring that utilizes readily available resources in an organization.

The case-base is a reflection of the experiences that have occurred, but may not be as comprehensive as other knowledge resources such as manufacturers' manuals. Domain experts rely on these other resources to assist them in solving new problems. Our approach to case authoring does not start at the individual experience, but at the

documents that contain domain knowledge. It would therefore be useful to develop a case base in two steps. The first step is to build a generic, comprehensive case-base from technical documentation. In the second, continuing step the case-base grows incrementally with experiences of the organization that uses the CBR system. This approach eliminates the manual processing previously documented experiences and allows the domain expert to focus their time authoring cases from the anomalous ones. Finally, the effectiveness of a case can be determined from the historical statistics compiled from past experiences.

This paper will outline the approach taken for case authoring using technical manuals and historical experience. This case authoring approach will be implemented within the context of a commercial airline's maintenance and repair facility. The paper will establish the viability of using technical manuals to create cases for a case-base to represent the knowledge that is already documented for the aircraft. Also, we will demonstrate that case enhancements such as historical statistics could affect the retrieval process from the case-base. The paper will discuss the results of creating the case-base from technical manuals and the validation of those cases with correlation with historical data. A specific example will be used to demonstrate the steps taken in this case authoring approach.

The paper proceeds as follows. The second section describes the background information for case authoring in CBR applications and the application domain of maintenance and repair. The third section outlines the approach of case authoring from structured documents or manuals. The fourth section describes the results of this approach and discusses issues related to applying this approach. The final section describes conclusions reached from the experimentation and desired future work.

2 Background

2.1 Case Authoring

Case authoring has been implemented with approaches that rely on interaction with the domain expert to handcraf cases. In the case of the aerospace maintenance domain—our domain of interest—several approaches have been sought. The first one used decision tree induction and expert interaction to construct cases that troubleshoot problems on jet engines (Heider 1995). Decision tree induction was used to determine relevant slots in the parametric data and use them in the retrieval of cases in the case-base. The textual information in the repair reports had valuable information for constructing cases, but had to be interpreted by the domain experts. Further work resulted in the evaluation of the effectiveness of the case-base in their work in troubleshooting jet engines (Heider et al. 1997). This helped in the development of a smaller case-base that resulted in more accurate retrieval. The second approach, which results in the Integrated Diagnostic System (IDS), used a custom designed case authoring tool to create cases (Wylie et al. 1997). This manual approach resulted in a small case-base that was not used to its fullest potential within the maintenance organiza-

tion. In order for a case authoring to be successful in a CBR application many constraints must be considered. Constraints such as access to domain experts and the time required to author cases are factors that need to be addressed when fielding a successful CBR application. In many maintenance organizations the access to domain experts is very limited and their time is very valuable. Therefore, a manually intensive case authoring approach is not the optimal solution in the maintenance domain.

Information Extraction (IE) has been successfully demonstrated in the construction of cases from the text in the area of court cases (Bruninghaus & Ashley 2001). Information recorded in court case documents was extracted using Natural Language Processing (NLP) techniques to construct the case-base. Even though full understanding of the text is not achieved, IE is a useful technique in the identification of information in text and the development more complex structures from the text. Therefore, it can be used to process technical manuals to author cases. For instance, preprocessing of a priori knowledge from documentation can benefit the case authoring process. Case authoring approaches for plan creation found that manually eliciting knowledge from a textual doctrine is critical in establishing planning knowledge in case authoring (Aha et al. 2001).

2.2 Aerospace Maintenance and Repair Domain

Aircraft, cars, trucks, computers, and people have documentation written about them that allow a person to diagnose and repair problems that occur. Domain experts use this documentation to make timely decisions on what actions should be taken to resolve a problem. After the solution has been applied, the domain expert may record this experience in textual form for future reference. Domains such as aerospace and open-pit mining maintenance implement computer applications to track maintenance activities. Other domains such as the medical domain may use more traditional methods such as paper to achieve a similar functionality.

Aircraft are very complex systems with a variety of sensors, computers, and communication equipment. This makes the troubleshooting of aircraft difficult even with the onboard diagnostic capabilities and the extensive documentation developed and provided by the aircraft manufacturer. The diagnostic information and documentation is distributed over many different systems and is consulted before a diagnosis is made. It would be beneficial to automatically collect this information for the maintenance technician in order for them to make more timely accurate decisions.

The majority of knowledge about the aircraft is found in the aircraft's manuals written by the aircraft manufacturer (e.g. Trouble Shooting Manual (TSM), Illustrated Parts Catalogue (IPC)). Information related to the repair and maintenance of the aircraft is found in the Aircraft Maintenance and Tracking And Control (AMTAC) recording system. The information from these two information resources are critical in documenting reoccurring experiences, which have good potential for cases in a case-base.
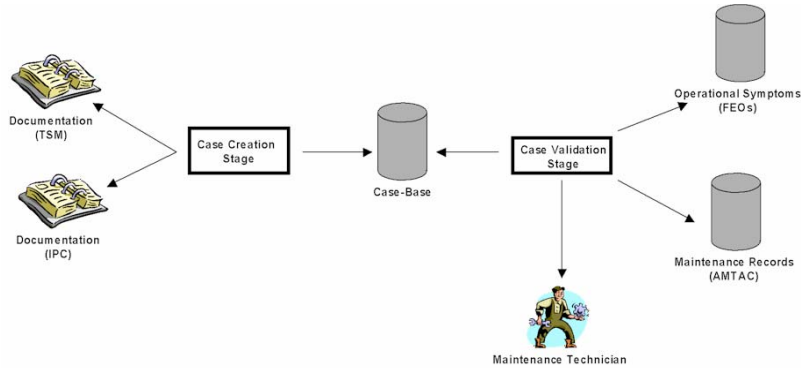
Fig. 1. Case Authoring Process using TSM and AMTAC

3   Case Authoring Approach

Figure 1 describes the overall process of case authoring from the TSM. This case authoring approach is fundamentally different from previous approaches by not considering individual experiences first. The first stage is to automatically create cases for the case-base from the manufacturer's documentation using IE techniques. The individual experiences are then used to update the cases within the case-base. The TSM and IPC are resources for aircraft knowledge developed by the designers and engineers of the manufacturer. This approach to case authoring captures the knowledge encoded in these manuals and uses it as a starting point to gain further knowledge about the aircraft. Our case authoring approach uses readily available resources such as manuals, operational data, and repair data. This proposed approach to case authoring is a two-stage process: case creation and case validation.

3.1 Case Creation

The case structure used in this case authoring approach is the same as the one used in IDS (Wylie et al. 1997). The features used for case retrieval are related to the symptoms that describe a problem handled by the case. These symptoms are the automatically generated messages from the built-in test equipment onboard the aircraft. The case separates these symptoms into different aggregations according to textual similarity, time proximity, TSM reference, and human association grouping. The component and action taken on the component is stored in the case as the recommended solution. Additional information like historical statistics and recorded incidents are also stored in the case. The historical statistics and recorded individual experiences are captured from the organization's historical data in the case validation stage. The case creation stage will focus on automatically extracting TSM reference symptoms

for the case and extracting the actions and components used in the solution for the case using the TSM.

The case creation stage is the process of automatically extracting knowledge from the TSM in order to create a case-base for the maintenance organization. Identification of symptoms and recommended solutions is critical in the case creation stage. The first part of case creation is to identify the symptom sets for the cases. In IDS, a set of rules is extracted from the TSM (Wylie et al. 1997). The Left Hand Side (LHS) of these TSM rules describes symptom sets in the form of automatically generated diagnostic message information. These symptom sets described in the LHS of the rules become the TSM reference symptoms in the case. The recommended solution information for the case is found inside the fault isolation procedures described in the Right Hand Side (RHS) of these IDS rules. Using IE techniques, it is possible to extract action and component information from the text in the TSM and correlate it with the symptom set information to create cases.

Our initial approach to IE is very simple and is outlined in Figure 2. We scan the text inside TSM for occurrences of important actions, and extract the surrounding information. Scanning, at this early stage, is performed by regular expressions that encode what we are looking for, and are matched against the text.
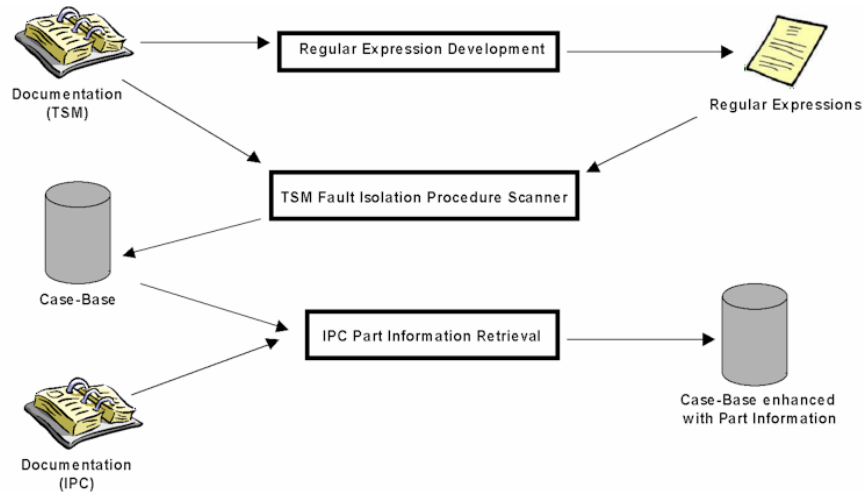

Fig. 2. Case Creation Stage

Table 1. Regular Expressions used in the Case Creation from TSM

ID	Regular Expression	Application Frequency
1	/(replace) the (.*)/i.	10,608
2	/do a check of the (.*) and (replace) it/i.	26
3	/make sure that the (.*) is not clogged. If necessary, (replace) it/I	11

3.1.1 Regular Expression Development

Regular expressions were developed to extract the actions and components found within the TSM fault isolation procedure. After some manual analysis of the text in the fault isolation procedure, the verb 'replace' was identified as the most frequently used action. A set of regular expressions was developed using the most frequently referenced action 'replace'. Table 1 outlines the three expressions used and a text scanner uses these regular expressions to identify the components that are replaced in the TSM. Further regular expression development must be completed to cover other action words used in the solution.

3.1.2 TSM Fault Isolation Procedure Scanner

The TSM Fault Isolation Procedure Scanner was developed to create the cases from the TSM. The TSM Fault Isolation Procedure Scanner uses both the IDS rule set and the TSM fault isolation procedures to create the case-base. Each individual IDS rule is processed by the TSM Fault Isolation Procedure Scanner for symptoms located in the LHS of the rule and the corresponding procedure on the RHS. The automatically generated diagnostic messages are extracted from the LHS and used to create a template case. A template case is created because a symptom set can have more than one recommended solution. The corresponding procedure from the RHS is scanned using the TSM Fault Isolation Procedure Scanner with the regular expressions developed in the previous step. Once an action and component are identified within the TSM fault isolation procedure, a new case is duplicated from the template case. This new case has its component and action fields populated with the action and component information identified from the TSM fault isolation procedure. After the IDS rule set has been processed, a case-base is built from the IDS rule set and TSM documentation.

This case-base can be perceived as a duplication of the IDS rule set, but the case base has the ability to be enhanced and updated with supplementary information. Further enhancements can be in the form of additional information gained from other manuals such as the IPC. Another form of enhancement is the recording of individual experiences that validate the case's usefulness. Once the case-base is enhanced with additional information, these cases contain more knowledge and information than the IDS rule set and can easily be updated. This up to date knowledge affects the way the cases are organized and retrieved and represents the current knowledge of the organization.

3.1.3 IPC Part Information Retrieval

The first enhancement of the TSM case-base helps identify components in the case validation stage. The IPC contains information about the specific part number and manufacturers. A correlation between components in the TSM and IPC are established through a code called the Functional Item Number (FIN). Not all components in the TSM case-base have a FIN number associated with them. If a FIN code is found, it is used to identify IPC part number and manufacturer information and add this component information to the correlated TSM case. Since the identification of components in the AMTAC reports, which will be needed during the case-validation stage (see below), is difficult, any additional part information could be useful in this

identification process. The resulting TSM case-base is ready to be validated with related individual historical experiences.
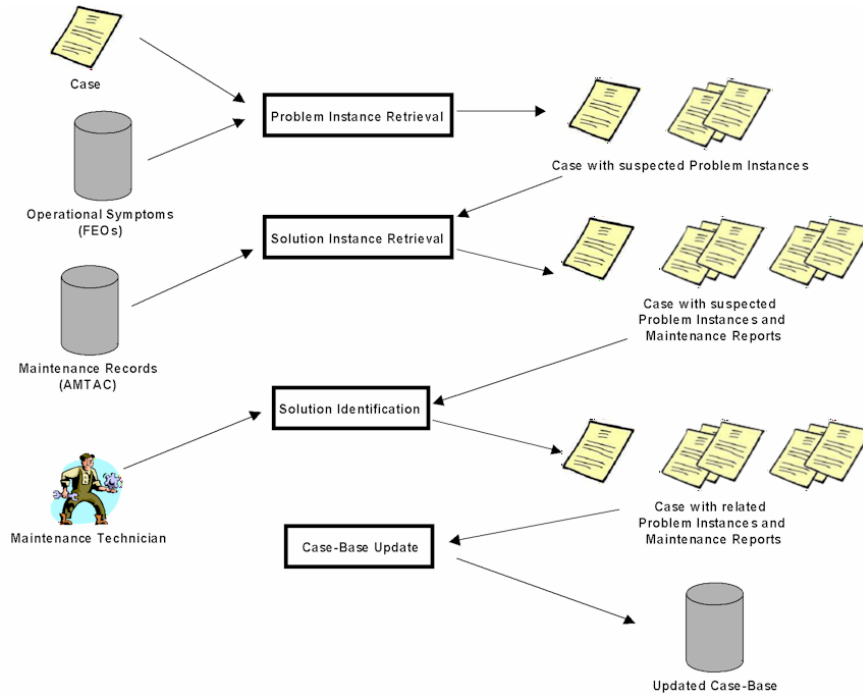

Fig. 3. Case Validation Stage

3.2 Case Validation

The case validation stage is the process that further enhances the case-base by capturing the organization's maintenance history inside the case-base. Even though a large set of cases may have been extracted from the TSM, the aircraft may not have generated all the problem symptoms described by the TSM. For case-base performance, it is desired to minimize the number of cases, but still achieve the same amount of coverage (Lenz et al. 1998). Since the TSM contains comprehensive information about problems on board the aircraft, an analysis of the aircraft's maintenance history can help reorganize the cases used in the case-base created from the TSM. This reorganization can be done in a hierarchy of cache memory where the most referenced cases are retrieved first before ones that have never been referenced. This applicability metric is established by validating the cases with historical experience. The case validation stage of case authoring can be broken down into four parts: problem instance retrieval, solution instance retrieval, case solution identification, and case-base update. The steps involved in the case validation stage are outlined in Figure 3.

3.2.1 Problem Instance Retrieval

Problem instance retrieval uses the symptom set from the case to identify instances of problems that occur in the operation of the aircraft. First, an arbitrary case is retrieved from the case-base to initiate the validation process. The symptom sets described in the case are used to identify instances of problems called *Fault Event Objects* (*FEO*s). FEOs are created within IDS whenever identified symptoms to problems related to the TSM occur on board the aircraft and the FEO information is stored in a database (Wylie et al. 1997). The FEO database is searched to obtain FEO problem instances related to a specific case. If no FEO problem instances were found then the case would not have been retrieved at any time during the aircrafts' recorded history. Cases that have identified FEO problem instances record the FEO information for later processing. This FEO information is used to determine the aircraft identification and the period of time when the symptoms occurred. Once the FEOs have been retrieved from the FEO database, they are clustered with respect to aircraft identity and time. Two FEOs that happen on the same aircraft within a 24-hour period of each other are considered to be the same problem to eliminate problems that are sporadic in nature. The result at the end of the problem instance retrieval process is a correlation between a case and suspected problem instances where this case could have potentially been applied.

3.2.2 Solution Instance Retrieval

The problem instances retrieved in the Problem Instance Retrieval step contain the aircraft's identity and the period of time when the problem occurred. This information is used to construct a query that will retrieve all the repair documentation for the aircraft during the time of the problem. The repair documentation is in the form of AMTAC reports that are stored in the AMTAC database. Solution instance retrieval process uses the AMTAC database to retrieve suspected AMTAC reports that happened during the time of the problem instance. The query results in a collection of AMTAC reports related to a specific aircraft for the problem instance's period of time. We will use five fields in the AMTAC report to establish its relevance to the problem instance. The five fields are problem description, solution description, parts installed/removed, and temporal information. These suspected AMTAC reports are correlated to their respective problem instances. Since not all the retrieved AMTAC reports are related to the problem instance, the next step established the relevancy of each of the AMTAC reports to the problem instance.

3.2.3 Case Solution Identification

Case solution identification reduces the collection of suspected AMTAC reports to only the ones related to the problem instance. The information stored in the case is used to establish the relevancy criteria by using the words in the symptoms and solution for the case and part information retrieved from the IPC. The first step was to compare the field that identified the parts replaced in the AMTAC report to the IPC part information stored in the case. The information in the parts removed/installed fields of the AMTAC report is helpful in establishing the component replacement with confidence, but this field is sometimes not filled in. Therefore, other fields in the

AMTAC report must be processed in order to determine the component and action taken in the solution. The text from the AMTAC report was searched for the words established in the criteria for AMTAC report applicability. The text in these fields was processed using a Bag of Words (BOW) approach to determine similarity to the case's word criteria (Craven et al. 1998). A BOW approach was used to determine the existence of terms used in both AMTAC report and the case's word criteria . In order to improve the accuracy of this second step, a stop list of words was manually created. This stop list contains frequently occurring words such as 'FAULT' and 'MSG'. The presence of a stop word is disregarded unless there is a presence of a nonstop word. After this automatic processing the suspected AMTAC reports is done, only the AMTAC report(s) relevant to the case's suspected problem instances are identified. A final evaluation from the maintenance technician is done to assure that the AMTAC messages are related to the problem instance. This manual evaluation consists of the maintenance technician making a binary decision on whether the solution was applicable or not. After the case solution identification is completed, the case's problem instances contains correlated TSM fault isolation procedures to related AMTAC reports and these problem instances can be used to update the case-base.

3.2.4 Case-Base Update

The case-base update implements a strategy for evaluating the effectiveness of the case solution with respect to the case's problem instances. Each of the case's problem instances contains the solution recommended by the case. The related AMTAC report is used to establish the date and time for the repair. If the repair occurred after the presence of symptoms for the case disappeared then the repair was successful. If not then the repair was unsuccessful. The result of this evaluation is used to update the information inside the case. The case is updated in two ways. The first is the AMTAC report identifier and the result of the repair are added as additional information to the case. The second is the statistical information regarding the result of the application of the case is updated by adding one to the applicable success or failure count. Once the case information is complete, the case is modified in the case-base to reflect these historical experiences.

The case authoring approach results in a case-base that is constructed from the TSM manual with enhancements from other manuals and historical experiences. This semi-automated case authoring approach facilitates constructing a case-base from scratch or modifying an existing case-base to reflect changes in the aircraft's documentation. When an update to the TSM is issued, this approach is reapplied to update the existing case-base. Also, this approach can be used as a complimentary approach to the other case authoring approaches implemented in IDS. A case-base can be created using this case authoring approach and can later be extended with the Automated Case Creation System in IDS (Yang et al. 2002). This approach eliminates problems solved by the technical documentation and allows for more computationally intensive approaches to focus on the undocumented problems that occur on the aircraft.

4 Results and Discussion

The case authoring approach was used to create a case-base for the Airbus A320/A319 aircraft. The latest TSM and IPC manuals were used in the case creation stage. Over six years of historical data from the AMTAC reporting system was used in the case validation stage. This approach resulted in a case-base that could be used by an airline's maintenance technician in IDS. The three results are from the two stages in the approach, an example using a specific TSM fault isolation procedure and some discussion concerning the results.

4.1 Case Creation Results

The case creation stage resulted in a case-base from the TSM and IPC with over 10,000 cases. The distribution of the three regular expression frequencies can be seen in Table 1. The most frequent applied regular expression was the regular expression '(replace) the *.' with over 10,000 occurrences. The other two regular expressions are much too specialized, but handle their respective specific situations. These regular expressions capture a large amount of cases represented in the TSM, but this scanning approach does not provide complete coverage for extracting cases represented in the TSM. The word 'replace' has 15, 013 occurrences in the TSM, in which 4,332 are not captured using these regular expressions. These specialized instances are outlier situations and additional regular expressions would have to be developed.

Other regularly used action terms would need to have similar regular expressions developed for them. This regular expression implementation for IE would quickly become unmanageable and too specialized for the problem domain. A better approach using IE would be to implement an NLP parser to provide more robustness and broader coverage when parsing the TSM. An investigation into the development of a semantic grammar was initiated and the discussion around this investigation is in the future work section.

Table 2. Regular Expressions used in the Case Creation from TSM

Case Symptoms Frequency	Number of Cases
0	10085 (81.01%)
> 0 and < 100	1859 (14.97%)
>=100 and < 1000	404 (3.25%)
>=1000	95 (0.77%)
Note: ~19% of Case Symptoms have Frequency > 0	

4.2 Case Validation Results

The results from the case validation stage describe the distribution of the case symptoms with respect to the problems experienced by the maintenance organization. The TSM is a comprehensive document for troubleshooting problems onboard an aircraft.

One issue is to determine what amount of the TSM is used in the everyday trouble-shooting of aircraft. Table 2 displays the distribution of the number of problem symptom set occurrences with respect to the number of cases with that occurrence. Table 2 shows that a majority of the cases represented by the TSM have not occurred, which accounts for 81% of the case-base. The number of cases that have their symptoms present in the operational data of the maintenance organization is less than 19% and further investigation is needed to confirm if these cases were actually applied. This result is not surprising because the TSM should possess more comprehensive knowledge than what is experienced in the maintenance organization. Some of these unused cases may be more applicable in different situations, for instance when the aircraft gets older and these constraints must be taken into consideration in case retrieval.

The second result from the case validation stage is the identification of frequently occurring problems on the aircraft. Out of the 19% of cases that have operational experience, 4% of those have symptoms that occur more than a thousand times in a six year period. One could quickly construct a list of frequently occurring problems that have happened in the past and document these problems for future reference. Also, the maintenance technicians or computationally intense approaches such as data mining can better focus investigative efforts by using this list of frequently occurring problems. Finally, if no extensive domain expertise exists, this list could be used as a benchmark for the maintenance organization.

4.3 Example

This section presents a detailed example of the extraction procedure. The TSM fault isolation procedure 22-83-00-810-849 is used to demonstrate the viability of this case authoring approach. According to the IDS rule set, this fault isolation procedure must be applied whenever the aircraft generates the failure message "AFS BSCU2". Below is sample text from the fault isolation procedure where the actions and components are highlighted by the applicable regular expression patterns.

A. If the test gives the maintenance message AFS: BSCU2 (ISSUED BY: FG1):
- **replace** the **FMGC-1 (1CA1)**
  AMM TASK 22-83-34-000-001 and AMM TASK 22-83-34-400-001 .
  1. If the fault continues:
     - **replace** the **BSCU (10GG)**
       AMM TASK 32-42-34-000-001 and AMM TASK 32-42-34-400-001.
  2. If the fault continues:
     - do a check and repair the wiring of the BSCU OPP VALID COM and BSCU OPP VALID MON discretes (from the FMGC 1 (1CA1) to the BSCU (10GG))
       ASM 22-85/04 .
B. If the test gives the maintenance message AFS: BSCU2 (ISSUED BY: FG2):
- **replace** the **FMGC-2 (1CA2)**

Table 3. Results of Cases after Case Validation Stage

Symptoms	Action	Component	Successful Application	Unsuccessful Application
"AFS BSCU2"	Remove/Install	FMGC-1 (1CA1)	1	0
"AFS BSCU2"	Remove/Install	BSCU (10GG)	11	5
"AFS BSCU2"	Remove/Install	FMGC-2 (1CA2)	2	0

Table 3 displays the three cases that were extracted from the fault isolation procedure 22-83-00-810-849 after the completion of the case extraction process. The symptoms for these cases were used to retrieve 931 problem instances from the FEO database and 2990 AMTAC reports from the AMTAC database were retrieved using the problem instance information. The case validation stage has identified 19 of these AMTAC messages as solutions to 19 problem instances. The Table 3 outlines the results after each of the 19 solutions is evaluated against the occurrence of the related problem instance. The case authoring approach has identified that the case with the replacement of the BSCU occurs most frequently and has the most frequent application. Currently, the TSM manual recommends the replacement of the FMGC-1 before the replacement of the BSCU, but our case authoring approach uncovers that this historically does not happen. We were expecting to see 16 unsuccessful replacements of the FMGC-1, but the results show no unsuccessful applications of that case. Therefore, there must be additional knowledge that maintenance organization must use in order to make decisions besides the manuals. This additional information is not reflected within the IDS rule set, but this information can be captured in the case-base. The historical statistics on the past applications of a case is additional information that can be used by the maintenance organization to make decisions on current problems. This information influences the retrieval process by ranking the case with replacement of the BSCU higher than the one with FMGC-1, which is recommended by the TSM. Our case authoring approach compiles statistics on the applicability of the case and adjusts the ranking of similar cases accordingly. Ultimately, the maintenance technician has to make the decision, but only after all the information about the problem is displayed.

The coverage of the solutions with respect to the total number problem instances is very low and was less than we anticipated. Despite the small amount of coverage, the solutions give a better idea of how the manuals are used in everyday operation. This allows for the maintenance organization to focus their attention on the problem instances where the solutions are unknown. In the results, problem symptoms were present and then disappeared without any solution from the TSM been recorded. This can be explained in different ways, but two possibilities have been identified. The first possibility is the problem is resolved as a result of an additional test outlined in the TSM. An example of this is the reset of the computer and a test for symptom persistence and the repair action information recorded in the AMTAC report is

"CHECKED OK". The second possibility is that there may exist relationships between components and symptoms that have not been documented. For instance, a set of symptoms may disappear due to a side effect of replacing an unrelated component. Due to the complexity of the equipment and the numerous relationships between components, this possibility is not unrealistic. Since the data has been identified by the case authoring approach, other knowledge discovery techniques may be able to establish these undocumented relationships. It is unclear and not the purpose of this paper to determine what percentage of cases is related to the two different possibilities.

5   Conclusion and Future Work

The extraction of knowledge from text documents for case authoring is a relatively new idea. (Brüninghaus & Ashley 2001) were the first authors to discuss it, but they do not address the automation of the automated process of the creation and validation of the cases from textual documentation (ie. manuals). We have developed this semi-automated process using the knowledge encoded in the manufacturer's documentation. By gaining a better understanding of what is represented in the manufacturer's manual, the domain experts can focus their efforts on the unknown problems. In this paper, we discussed our experience with case authoring using the manufacturer's manuals and historical experience. We also present some insights for continued work on the case authoring process using textual documentation and argue in favour of integration of this approach with other knowledge discovery approaches.

Future work can be divided into two different areas: further development of the case-base authoring process using manufacturer's documentation and integration with other knowledge acquisition processes. Better coverage for extracting cases from the TSM could be achieved by the development of a semantic grammar to parse the fault isolation procedures in the TSM. This tool would be useful to further determine the relevance between a subset of cases, as more information becomes available for the problem instance. This initial investigation into a semantic grammar for the TSM fault isolation procedure started with a randomly chosen fault isolation procedure from the TSM. A lexicon and semantic grammar rules were developed to parse this fault isolation procedure. This TSM Fault Isolation Procedure parser successfully generated parse trees for 49 out of 50 sentences in the procedure. This parse tree was converted into cases. In the attempt to parse the complete TSM, the brittleness of this semantic parser became quickly apparent. The first apparent problem was the incompleteness of the lexicon of words. A more comprehensive lexicon would need to be developed in order for the semantic grammar parser to recognize the semantic and grammatical definitions for the words in the fault isolation procedures. The second problem encountered is the incompleteness of the semantic grammar rules. Domain knowledge and more analysis of the text are needed to further develop this semantic grammar. The resulting semantic grammar is not only useful in parsing TSM fault isolation procedures, but could potentially be used in parsing the AMTAC reports in the case validation stage. Even though the text in the AMTAC reports is cryptic, the

dictionary developed for the semantic grammar could be useful in determining the word usage in the AMTAC reports.

Other future work is centered on the integration with other knowledge acquisition systems. This strategy is important in demonstrating the complimentary nature of this case authoring approach to other knowledge acquisition approaches. One possible area of integration for the case-base created from the manufacturer's manuals is with other Knowledge Discovery in Database (KDD) approaches. Once a case-base has been created from the manuals and enhanced with operational experience, this knowledge could become valuable information for other knowledge discovery processes like KDD. KDD has been applied to the aerospace domain and one of the issues is the appropriate use of background knowledge during the different phases of data preprocessing and data analysis (Létourneau et al 1997). A process for automatically labeling instances is required to successfully field a KDD application. A potential source for automatically labeled instances can be from the case-base created using our case authoring approach. These instances may be useful in determining the models that are created in the data analysis phase. Therefore, further investigation into the TSM case-base being used as training data for building component failure models in aircraft is outlined as a future direction.

Organizations are developing processes to foster knowledge-based activities. The role of a knowledge worker is critical for today's successful organizations (Drucker 2001). Until recently, knowledge has not been considered as a capital resource, but this intellectual capital has become a very valuable asset within an organization lately (Stewart 2001). There is no process for directly measuring intellectual capital in an organization, but AI technologies can be used to represent and share knowledge within organizations. This knowledge acquisition and distribution process is demonstrated with the case authoring process using knowledge from textual documents and historical experiences.

Acknowledgments

References

1. Aha, D. (1998). The Omnipresence of Case-Based Reasoning in Science and Application. Knowledge-Based Systems, 11(5-6), 261-273.

2. Aha, D.W., Breslow, L.A., & Muñoz-Avila, H. (2001). Conversational case-based reasoning. *Applied Intelligence*, *14*, 9-32.

3. Bartsch-Spörl, B., Lenz, M., Hübner, A. (1999). Case-Based Reasoning – Survey and Future Directions. In the Proceedings of the Fifth Biannual German Conference on Knowledge-Based Systems (XPS), Würtzburg, Germany, March 3-5, 1999.

4. Brüninghaus, S. & Ashley, K. (2001). The Role of Information Extraction for Textual CBR. Proceedings of 4th International Conference on Case-Based Reasoning (ICCBR-01). Vancouver, Canada, July 30 – August 2 2001.

5. Craven, M., DiPasquo, D., Freitag, D., McCallum, A., Mitchell, T., Nigam, K., Slattery, S. (1998). Learning to Extract Symbolic Knowledge from the World Wide Web. In the Proceedings of the 15th National Conference on Artificial Intelligence (AAAI-98), Madison, WI, July 26 – 30, 1998.

6. Drucker, P. (2001). The Essential Drucker: In One Volume the Best of Sixty Years of Peter Druckers's Essential Writings on Management. New York, Harper Collins.

7. Heider, R. (1995), Troubleshooting CFM-56-3 Engines for the Boeing 737 - Using CBR and Data Mining. In the Proceedings of the Third European Workshop, EWCBR-96, Advances in Case-Based Reasoning, Lausanne, Switzerland, November 14-16, 1996, pp512-518.

8. Heider, R. Auriol, E. Tartarin, E. Manago, M. (1997). Improving the Quality of Case Bases for Building Better Decision Support Systems. Proceedings of the 5th German Workshop on Case-Based Reasoning (GWCBR'97), Bad Honnef, March 4-5 1997, p.85

9. Lenz, M., Bartsch-Spörl, B., Burkhard, H., Wess, S. (1998). Case-Based Reasoning Technology: From Foundations to Applications. Berlin, Springer.

10. Létourneau, S., Famili, F, and Matwin, S. (1997). Discovering Useful Knowledge from Aircraft Operation/Maintenance Data. In the Proceedings of the workshop on Machine Learning Applications in the Real World, 14th International Conference on Machine Learning, Nashville, TN, July 8-12, 1997.

11. Stewart, T. (2001). The Wealth of Knowledge: Intellectual Capital and the Twenty-first Century Organization. New York, Doubleday.

12. Wylie, R., Orchard,R., Halasz, M., Dubé, F. (1997). IDS: Improving Aircraft Fleet Maintenance. Proceedings of the 14th National Conference on Aritificial Intelligence and Innovative Applications of Artificial Intelligence (IAAI-97), Providence, Rhode Island.: pp. 1078-1085. July 27-31, 1997.

13. Yang, C., Orchard, R., Farley, B., Zaluski, M. (2002). Automated Case Base Creation and Management. To appear in the 16th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems (IEA/AIE 2003), Loughborough, UK, June 23-26, 2003.