CRC for Construction Innovation, (2001) *Life cycle modelling and design knowledge development in 3D virtual environments : final report.* CRC for Construction Innovation, Brisbane.

CRC Construction Innovation
BUILDING OUR FUTURE

CRC-CI Project 2001-002-B

Life Cycle Modelling and Design Knowledge Development in 3D Virtual Environments

Final Report

Program Number:	B
Program Title:	Sustainable Built Assets
Project Number:	2001-002-B
Project Title:	Life Cycle Modelling and Design Knowledge Development in Virtual Environment
Project Leader:	John Gero, USYD
Project Manager:	Rabee Reffat, USYD
Project Researchers:	Wei Peng, USYD
	Julie Jupp, USYD
	Pak San Liew, USYD
	Lan Ding, CSIRO
Project Contact:	Peter Scuderi
Date:	November, 2004

Distribution List

CRCCI
Authors

Disclaimer

The Client makes use of this Report or any information provided by CRC CI in relation to the Consultancy Services at its own risk. CRC CI will not be responsible for the results of any actions taken by the Client or third parties on the basis of the information in this Report or other information provided by CRC CI nor for any errors or omissions that may be contained in this Report. CRC CI expressly disclaims any liability or responsibility to any person in respect of any thing done or omitted to be done by any person in reliance on this Report or any information provided.

Please direct all enquiries to:

Chief Executive Officer
Cooperative Research Centre for Construction Innovation
9th Floor, L Block, QUT
2 George St
Brisbane Qld 4000
AUSTRALIA
T: 61 7 3864 1393
F: 61 7 3864 9151
E: enquiries@construction-innovation.info

Summary

Experience plays an important role in building management. "How often will this asset need repair?" or "How much time is this repair going to take?" are types of questions that project and facility managers face daily in planning activities. Failure or success in developing good schedules, budgets and other project management tasks depend on the project manager's ability to obtain reliable information to be able to answer these types of questions. Young practitioners tend to rely on information that is based on regional averages and provided by publishing companies. This is in contrast to experienced project managers who tend to rely heavily on personal experience. Another aspect of building management is that many practitioners are seeking to improve available scheduling algorithms, estimating spreadsheets and other project management tools. Such "micro-scale" levels of research are important in providing the required tools for the project manager's tasks. However, even with such tools, low quality input information will produce inaccurate schedules and budgets as output. Thus, it is also important to have a broad approach to research at a more "macro-scale."

Recent trends show that the Architectural, Engineering, Construction (AEC) industry is experiencing explosive growth in its capabilities to generate and collect data. There is a great deal of valuable knowledge that can be obtained from the appropriate use of this data and therefore the need has arisen to analyse this increasing amount of available data. Data Mining can be applied as a powerful tool to extract relevant and useful information from this sea of data.

Knowledge Discovery in Databases (KDD) and Data Mining (DM) are tools that allow identification of valid, useful, and previously unknown patterns so large amounts of project data may be analysed. These technologies combine techniques from machine learning, artificial intelligence, pattern recognition, statistics, databases, and visualization to automatically extract concepts, interrelationships, and patterns of interest from large databases. The project involves the development of a prototype tool to support facility managers, building owners and designers.

This final report presents the AIMM™ prototype system and documents how and what data mining techniques can be applied, the results of their application and the benefits gained from the system. The AIMM™ system is capable of searching for useful patterns of knowledge and correlations within the existing building maintenance data to support decision making about future maintenance operations.

The application of the AIMM™ prototype system on building models and their maintenance data (supplied by industry partners) utilises various data mining algorithms and the maintenance data is analysed using interactive visual tools.

The application of the AIMM™ prototype system to help in improving maintenance management and building life cycle includes: (i) data preparation and cleaning, (ii) integrating meaningful domain attributes, (iii) performing extensive data mining experiments in which visual analysis (using stacked histograms), classification and clustering techniques, associative rule mining algorithm such as "Apriori" and (iv) filtering and refining data mining results, including the potential implications of these results for improving maintenance management. Maintenance data of a variety of asset types were selected for demonstration with the aim of discovering meaningful patterns to assist facility managers in strategic planning and provide a knowledge base to help shape future requirements and design briefing. Utilising the prototype system developed here, positive and interesting results regarding patterns and structures of data have been obtained.

Table Of Contents

1. SCOPE

As the construction industry adapts to new computer technologies, in terms of hardware and software, computerized design, construction, and maintenance data are becoming increasingly available. The growth of many business, government, and scientific databases has begun to far outpace an individual's ability to interpret and digest the data. Such volumes of data clearly overwhelm the traditional methods of data analysis such as spreadsheets and ad-hoc queries. The traditional methods can create informative reports from data, but cannot analyse the contents of those reports. A significant need exists for a new generation of techniques and tools with the ability to automatically assist humans in analysing the mountains of data for useful knowledge.

The increasing use of databases to store information about facilities, their use, and their maintenance provides the background and platform for the use of data mining techniques for future projections. The current technology for facility maintenance uses databases to keep track of information and for notification of maintenance schedules. These databases are so far not well linked with an interactive 3D model of the building and are generally presented in tabular form.

Data Mining (DM) and Knowledge Discovery in Databases (KDD) are tools that allow identification of valid, useful, and previously unknown patterns. These technologies combine techniques from machine learning, artificial intelligence, pattern recognition, statistics, databases, and visualization to automatically extract concepts, interrelationships, and patterns of interest from large databases. The DM and KDD techniques are capable of finding patterns in data that can assist in planning. Patterns and correlations identified from data mining existing records of maintenance and other facilities management activities provide feed back and can improve future maintenance operation decision making, inform strategic planning as well as the design of new facilities.

This research is motivated by several observations of the current situation in the building industry. Since the cost of maintaining a facility over its life span is more than the capital cost, a marginal increase in capital cost can be shown to produce an amplified reduction in facility maintenance with a concomitant reduction in overall cost of ownership. The prototype system developed in this research provides a tool to assist in such decision making during the life cycle of a building

The project develops and assesses the integration of data mining with a number of technologies, including agent-based technology, database management systems, object-based CAD systems and 3D virtual environments. The investigation involves the development of a prototype tool to support automated feed back for building life cycle modelling, planning and decision-making. This final report presents the AIMM™ prototype system and documents how and what data mining techniques can be applied to building maintenance data, its results and the benefits of applying such techniques.

This report also outlines the schema requirements for connecting building asset data with CAD building elements and required mappings for IFC transfer. The report provides assessment of available mapping and translation of building assets to ArchiCAD and IFC (Industry Foundation Class) models. The performance of the prototype system is evaluated and an analysis is provided to assess system frameworks, algorithms and reporting. The report includes discussion and recommendations of the overall performance and capacity of maintenance data requirements, CAD requirements and IFC conversion, the EDM engine, data mining algorithms and virtual environments.

The primary aim of this project is to develop and test a prototype system and report its capabilities as a tool suitable for discovering meaningful patterns, correlations and to report useful information via filtering techniques. The filtering system is based on heuristics derived domain specific knowledge, and provide direct assistance to users.

The interface system has been developed to provide a platform to facilitate set-up, data mining, filtering and reporting capabilities. This is a Java reporting schema and display. The system offers a user-friendly interface to support: data configuration, control of information flows, communication, and effortless interpretation of data mining results.

The scope of this project is outlined as follows:

- Background research on issues of information standardisation, building maintenance, life cycle cost modelling as well as potential future trends of information technology in AEC;
- Description and discussion of pros and cons of existing life cycle models of building data with object-oriented database and virtual environment;
- Review of existing virtual environments and developing tools;
- Survey of existing commercial data management systems and knowledge discovery systems adopted in the construction industry;
- Analysis of data sources from industrial partners and external sources;
- Demonstration of data mining, its algorithms, software and applications on industrial partner data;
- Development of a data mining approach to life cycle modelling of buildings in virtual environments;
- Documentation of systems framework, building life cycle model and integration of design knowledge in a virtual environment;
- Implementation of the prototype system using data mining application (WEKA), Express Data Manager (EDM), IFC, ArchiCAD7.0, and Java programming language;
- Testing AIMM™ prototype system. Experimentation tests three building assets including: air conditioning unit, thermostatic mixing valve, and battery charger components;
- Development of filters and reporting interface system for prototype system using Java
- Demonstration of complete prototype system, its data mining algorithms, filters and reporting interface. This demonstration illustrates the heuristics derived from domain specific knowledge for meaningful feed back.

The Royal Prince Alfred Hospital, Building No.10 was chosen to demonstrate data mining techniques and test the prototype system since the scope of maintenance data specifies a more complete mapping of building assets and CAD data and also maintains a level of complexity practicable for demonstrating the data mining module's capacity to automate knowledge development.

Technologies incorporated within the overall framework of the prototype system include:

- An object-based CAD system: ArchiCAD7.0;
- Industry Foundation Class mapping for component (element) information transfer: IFC2.0 and IFC2.x.;
- EDM database management;
- A 3D virtual environment, Active Worlds;
- WEKA;

- Agent Technology
- Java and Visual Basic C++ programming languages.

The objective of integrating these technologies is to demonstrate the detection of patterns and discovery of knowledge applicable to building maintenance, planning and design strategies from an object-oriented CAD building model in a 3D virtual environment. Such a tool provides an online interactive tool that automates feedback on any combination of assets and components.

This project is carried out at Key Centre of Design Computing and Cognition (KCDCC) of the University of Sydney.

1.1 Report Organisation:

The topics covered in the following sections of this report are:

- Section 2 covers the deliverables of the project.
- Section 3 covers background research on issues of information standardisation, building maintenance, and life cycle cost modelling. Potential future trends of information technology in AEC (Architecture, Engineering and Construction) are addressed and a review of existing life cycle models of building data with object-oriented database and virtual environment is presented. This Section also surveys existing commercial data management systems and knowledge discovery systems.
- Section 4 provides an analysis of two data sources from the industrial partner and an external source.
- Section 5 introduces data mining, its algorithms, software applications, and a data mining approach to life cycle modelling of buildings. This section also summarises the results of a previous demonstration of a variety of algorithms.
- Section 6 introduces a detailed systems framework of life cycle modelling of buildings and design knowledge in virtual environments is presented.
- Section 7 introduces the filtering and interface of AIMM™ prototype system.
- Section 8 discusses the projected estimates, recommendations and future project extensions.
- Section 9 presents the conclusions.
- Section 10 lists all Appendices.
- Section 11 provides a list of References.

2. DELIVERABLES

The deliverables of this project are defined as follows:

1. **Statement**: Approach and how it will improve maintenance planning methodology and knowledge to be used in the maintenance and management of existing facilities.
2. **Demonstration 1**: Data mining and knowledge discovery. The demonstration will use industry partner data (where possible) to show the implications of the availability of such a system.
3. **Development:** Systems framework, specify data mining techniques, agent architecture, links and mappings between CAD, maintenance databases and 3D virtual environments.
4. **Demonstration 2**: Prototype modelling tool that can be attached to asset management systems used by industry and government entities responsible for the management of building assets in a 3D virtual environment.
5. **Outcomes**: Improve connection between maintenance and design knowledge. Higher levels of maintenance knowledge should produce improved building designs and improve collaboration between industry partners.

As a consequence of the deliverables listed above the following milestones were outlined:

- ***Research paper***: present approach to maintenance data mining that includes design data modelling submitted to appropriate conference.
- ***Demonstrate***: a working data mining and knowledge discovery algorithms using industry data to QDPW.
- ***Develop***: object-oriented representation to provide 3D interactive environment using data provided by Woods Bagot.
- ***Implement***: object-oriented representation in virtual environments to provide 3D interaction and a link between knowledge development (as a result of data mining), with the building model.
- ***Establish***: basis of agent technology and develop and implement software agent architecture for data mining.
- ***Implement***: populate maintenance database with selected industry data.
- ***Research paper***: present agent-based approach to maintenance data mining submitted to journal.
- ***Research paper***: present architecture of entire prototype system submitted to appropriate conference or journal.
- ***Workshop***: demonstrating prototype system with data mining of maintenance data.
- ***Flyer***: derived from reports for CRC promotional purposes.
- ***Develop***: proposal for project submission on extension to time-based modelling.
- ***Final research report***: that documents the research basis of the development and its implementation.
- ***Industry report***: focused report for dissemination.

2.1 Deliverable on Statement: How Technology will Improve Maintenance Management and Building Design

A report, CRC2001_002_B_1, documenting potential future trends of information technology in AEC (Architecture, Engineering and Construction) has been completed. This report supported the need for a tool to assist facility managers in decision making and provide feed back for designers. This was established on the basis of relevant issues such as the need for: information standardisation, building maintenance, life cycle cost modelling, existing life cycle models of building data and their pros and cons of these models are discussed. This report also summarised the requirements of the design information and outlined the basic approach to life cycle modelling.

In addition, a refereed research paper has been published titled “Using data mining on building maintenance during building life cycle” at the 38^{th} Annual Conference of Architectural Science Association, ANZAScA 2004.

2.2 Deliverables on Demonstration 1: Data Mining and Knowledge Discovery Algorithms

A report, CRC2001_002_B_2 presenting the application of data mining algorithms and knowledge discovery on industry partner building maintenance data has been completed. The results reported in this document were demonstrated to industry partners in order to illustrate the technique's success in searching for patterns and correlations within the existing maintenance data and provide a method for supporting decision making on future maintenance operations.

In this demonstration various data mining algorithms provided by WEKA were applied and the maintenance data was analysed using interactive visual tools such as stacked histograms. The maintenance data of a variety of asset types were selected for this experiment with the aim of discovering meaningful patterns and provided a knowledge base to help shape future requirements gathering and design briefing. A summary of these results are provided in Section 5.7 and a more detailed analysis is contained within Appendix E.

In addition, a refereed research paper has been published titled “Improving the management of building life cycle: A data mining approach” at the CRC Construction Innovation 2004 Conference on *Clients Driving Innovation*.

2.3 Deliverables on Systems Framework

The overall framework of the prototype system was developed and described across two reports; refer to CRC2001_002_C_3, and CRC2001_002_C_4. The final architecture of the complete system was documented in CRC2001_002_B_5.

Both the overall framework and detailed system architecture have been implemented and tested to demonstrate the performance of a variety of data mining algorithms, filters and reporting schemas. The complete systems architecture and outcomes are presented in Sections 6 and 7 of this report.

2.4 Deliverables on Demonstration 2: Prototype System

The preliminary AIMM prototype system was demonstrated during the March Quarterly Review meeting in Sydney and the results were documented in report: CRC2001_002_C_4. In addition, the July Industry Partner Workshops demonstrated the completed implementation of the AIMM system demonstrating: (i) object-oriented representation in a 3D virtual environment, (ii) software agents (iii) population of the maintenance database with selected industry data, (iv) filtering heuristics, and (v) user-friendly reporting interface. An industry focused flyer has also been designed and published, refer to the Final Technical Report: Appendix A.The July Industry Partner Workshops (refer to Final Technical Report: Appendix B) were held in both Sydney and Brisbane.

The completed AIMM prototype system, results and outcomes are documented in the Final Technical Report and summarised here in Sections 7. An evaluation of the performances and capabilities of the AIMM prototype system is provided in Section 8

2.5 Deliverables on Outcomes: Improved Connection between Maintenance and Design Knowledge

This final Industry Report provides results supporting the capabilities, flexibilities and advantages of automated knowledge development as a result of data mining building models in a virtual environment. Evaluation and testing highlights the following:

- Production of timely data on the effects of different maintenance regimes and provision of proactive information for improving the design, maintenance and management of building facilities.
- Provision of testing methods to validate the usefulness and scope of current databases as a platform for guiding future decisions.
- Linking of a 3D model with maintenance data allows both the facility manager and the designer to gain access to information and knowledge that is currently inaccessible.
- Combination of a 3D model with maintenance and other asset data facilitates the ability of building designers and owners to visually model the impact of design, maintenance, refurbishment and extension decisions on the building's life cycle cost.
- Representation of the facility within the virtual environment provides a basis for linking data mining with emerging technologies (such as connecting to WAP phones and other PDAs both in the office and on site) to address a gap in the construction life cycle.

This Industry Report also provides estimates and recommendations for future works. The recommendations for the extension of this research are described in Section 8. In addition, estimates of the impact of up-scaling the system to handle more complete representations of all building assets, components and systems are provided. A discussion of the prospective areas of development for the CRC is also presented.

3. BACKGROUND

With the boom in of information technology at the end of last century, the increase in information availability has become a dilemma due to inefficiency in processing the information for decision-making. This problem becomes critical in the building industry when we consider the high degree of complexity of work flows involved and the accompanying uncertainty for decision making in the lifetime of a building. Thus, efficiently dealing with information from different stages of a building's life cycle to improve profitability, productivity as well as strategic resource planning are important business forces driving life cycle modelling.

However, the design of new buildings and facilities tends to focus on short-term cost and the immediate needs of the owner for a building that meets various business and functional requirements. Current technologies such as Computer-Aided Design (CAD) have focussed on the needs of designers to develop designs that meet design briefs that do not include life cycle design. Very little attention has been given to modelling of the life cycle costs of buildings at the design and management stages to forecast and achieve the most economical life cycle cost.

There are several life cycle models available for buildings as a whole and for their component systems; although there is no one model that is accepted as a standard, there are some areas of commonality. Life cycle cost models form predictions based on several parameters, some of which include a degree of uncertainty, such as the reliability of a part. These inputs can range from the cost of installation to the cost associated with carrying spare parts in inventory (Siewiorek 1982).

The values of these input variables, along with their probability distributions, can be predicted for each component by the Data Mining system, thus allowing for more accurate estimation of average life cycle cost. By accurately predicting failure rates and repair costs, it is also possible to compute the optimal schedule of preventative maintenance for each asset. What can be predicted and the accuracy of those predictions depends of course on the availability and accuracy of the maintenance data that is available. Furthermore, current life cycle modelling systems fail to provide a seamless integration of hybrid information that provides users with access to previously unreachable knowledge. In this project, our focus is to combine 3D modelling with maintenance database and import data mining agents into the maintenance process to improve knowledge acquisition and accessibility of a Life Cycle Modelling prototype within a highly interactive virtual environment.

A simple statement of the maintenance objective for a building is that building systems should always be available to support building functions. More precisely, the maintenance objective for the building is that the cost of any maintenance activity should be less than the expected marginal value of production enabled by the planned activity. To support this objective, it is essential to tackle the maintenance from multiple facets including interpretation of observed data, diagnosis of problems, planning repair and maintenance, and business evaluation of the value-added from different repair and maintenance options. Equally significant is defining the "value" of maintenance from both engineering and business perspectives.

Figure 1 illustrates a conceptual four-way framework that depicts this research in relation to a variety of existing commercial systems and related state-of-the-art systems' development. The four related areas include: (i) Facilities Management, (ii) Construction Industry, (iii) Architectural and Design Domain, and (iv) University and Research.

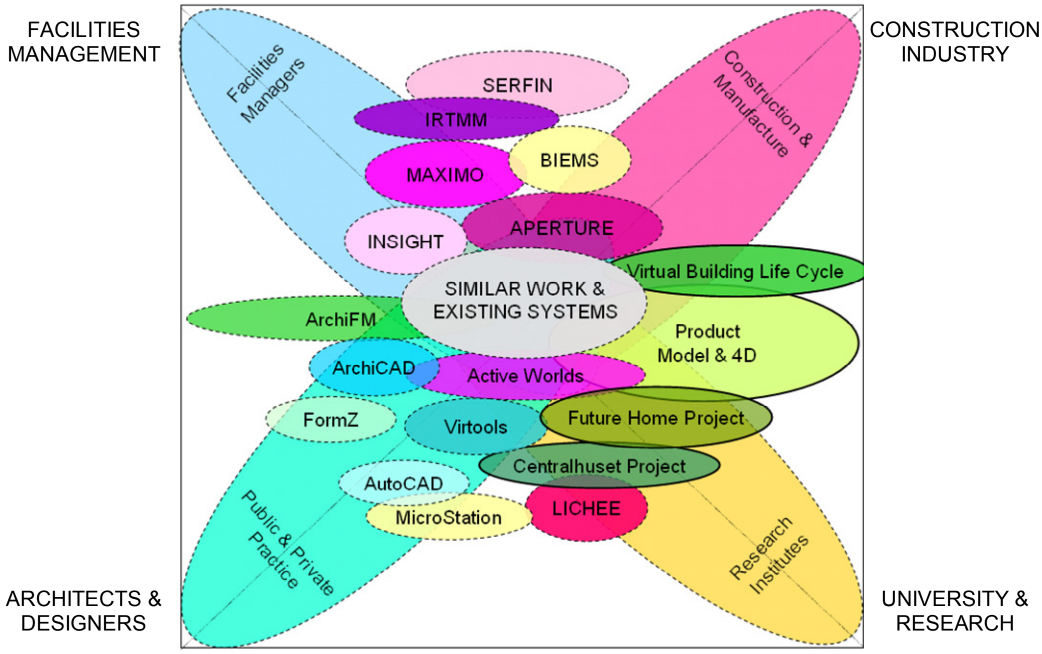


Figure 1. Conceptual framework of related commercial and prototype systems

Based on a general investigation into the four main areas, detailed background studies were carried out which focued primarily on aspects of: (i) information standardisation; (ii) role of 3D models in design and construction; (iii) building maintenance, and (iv) life cycle modelling. Other areas crucial to the development of the project included:

- Trends in Information Technology in AEC Industry;
- Virtual Environments and their significance in modelling;
- Data Management Systems;
- Systems Used for Knowledge Discovery and Data Mining.

The following provides a detailed description of the above areas and relevant case studies surveyed as part of this research project's development.

3.1 Information Standardisation

Standards are defined in the Longman Web Dictionary[1] as a means of measurements that are a fixed official rule for measuring weight, purity, value etc. Another definition more related to the focus of this project is a published document that sets out technical or other specifications necessary to ensure that a method or material will consistently do the job it is intended to do, i.e. what must occur to achieve the desired result[2].

3.1.1 Information Standardisation in CRC Project

To leverage the project data generated in hybrid domains, as well as to improve the efficiency of information sharing, this CRC project has adopted the Industry Foundation Classes (IFC) interoperability standard to exchange project data. The International Alliance of Interoperability (IAI) has been propagating the use of the Industry Foundation Classes (IFC) interoperability standard since 1995 as its major focus. IFC is an interoperable standard that defines a single object-oriented data model to allow different disciplines to accurately share technical information with IFC-compliant tools (Fischer and Kam 2002).

---

[1] http://www.longmanwebdict.com
[2] http://www.iie.qld.gov.au/informationstandards/glossary.asp

IFCs are defined by the AEC/FM industry in which they provide a foundation for shared project models and classes in a common agreed manner. IFC-based objects allow project participants to share a project model and improve efficiency in terms of objects reusability. Interoperability among AEC/FM software applications is achieved via the use of universal AEC/FM objects based on IFC specification. Based on its extensive application in productivity, time and cost control during the design, construction and maintenance life cycle, IFC is becoming a de facto standard for the building industry. Many of the leading CAD software vendors such as AutoDesk and Graphisoft support importing and exporting IFC compliant data that is platform independent. The latest version IFC 2.x is supported by some leading CAD vendors as an add-on feature. Figure 2 illustrates the benefits of IAI's IFC to the AEC/FM (Industrial Foundation Classes, 2002).

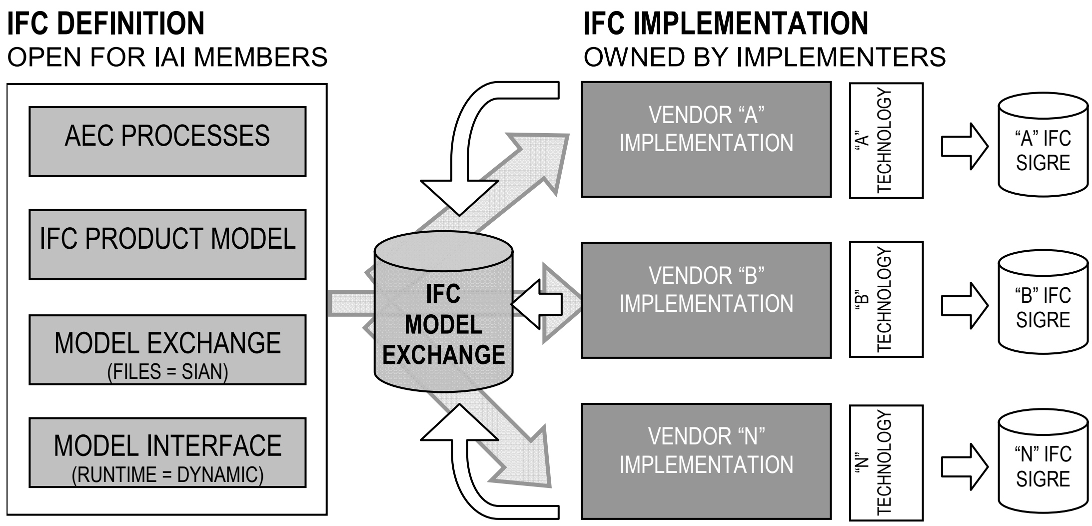


Figure 2. The entire AEC/FM Industry benefit from IAI's IFCs[3]

IFCs allow integration of application modules from different disciplines such as HVAC, Electrical, Architectural, etc. at the domain layer and processes through a multilayer architecture to generate hierarchical resources that are application independent objects. IFC functionality is generalised to include: IFC Model, IFC Views and IFC Object Hierarchy. These functions are illustrated in Figure 3. and Figure 2 illustrates the multilayered architecture of IFCs 2.x.

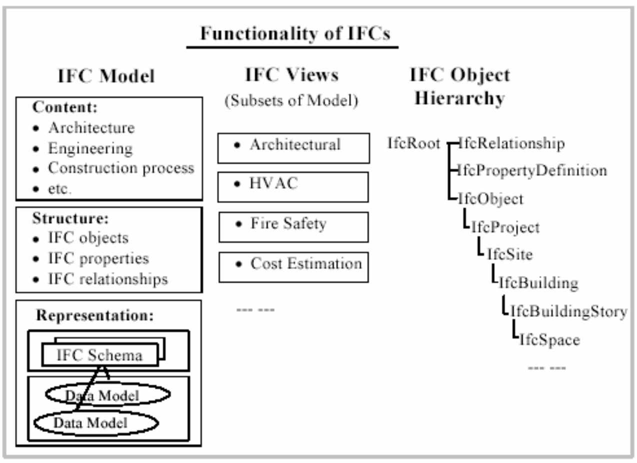


Figure 3. An illustration of IFC functionality (Ding et al., 2003)

3 http://cig.bre.co.uk/iai_uk/iai/page5.htm

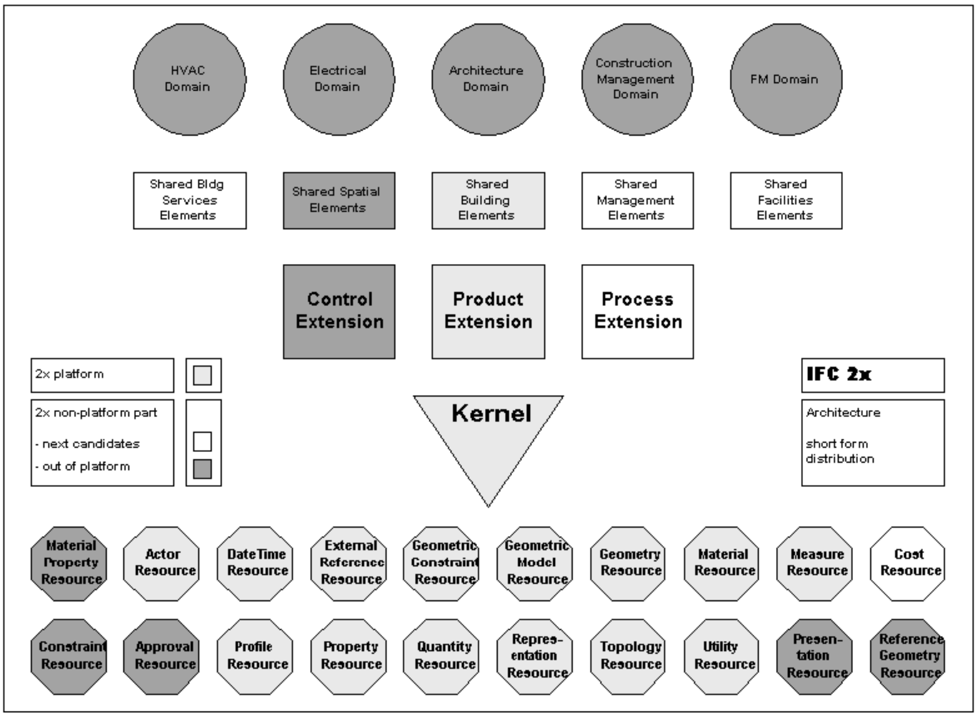


Figure 4. IFC 2x Overall Architecture (IAI, 2000)

3.1.2 Future Trends of Information Technology in AEC Industry

Waugh *et al*. (1996) presented the result of a survey that questioned experts from architecture, engineering, and construction (AEC) to predict the future trends of information technology (IT) in supporting project management into the future over a 20 year timeframe. The survey covered issues of project management, environment, computer systems, application areas, information and integration for project management. Froese et al. (2001) conducted a similar survey with comparable groups from AEC/FM. A general comparison of the key survey results for year 1996 and 2001 are as follows:

– A high percentage responded that new computer technologies will have a positive impact on the market potential/competitive advantage as can be seen in Figures 5 and 6.

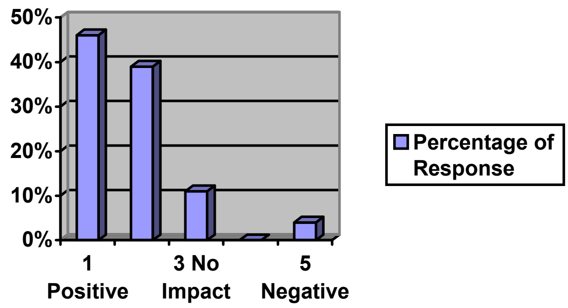


Figure 5. Impact of new computer technologies on market potential / competitive advantage from 1996 survey

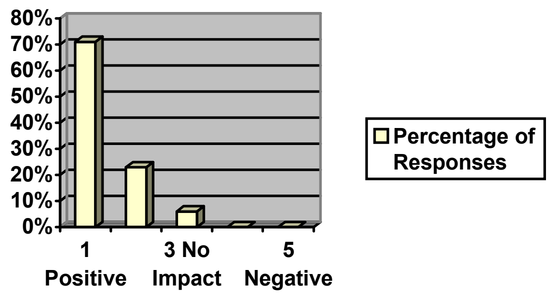


Figure 6. Impact of new computer technologies on market potential / competitive advantage from 2001 survey

– The most expected common types of information standardisation by 2015 and 2020 include:
  - General purpose inter-industry standards for data sharing demand (e.g. HTML, SGML, XML)
  - Construction industry-wide standards (ISO-STEP, IAI-IFC)
  - De facto industry standards (e.g., DXF, WKS)
  - Intelligent software agents that perform translation
  - Project wide standards
  - Company wide standards

– The most expected ways that information technology will change the way project manager's work by 2015 and 2020 are shown in Figure 7 and Figure 8.

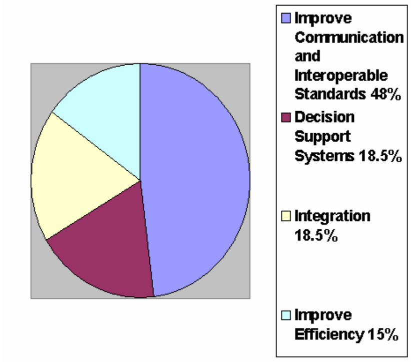


Figure 7. Potential contributions from information technology by 2015 [adapted from Waugh et al. (1996)].

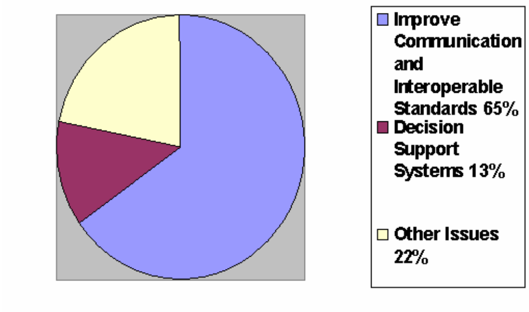


Figure 8. Potential contributions from information technology by the year 2020 [adapted from Froese et al. (2001)].

– The most expected critical areas requiring AEC industrial attention are shown in Figure 9.

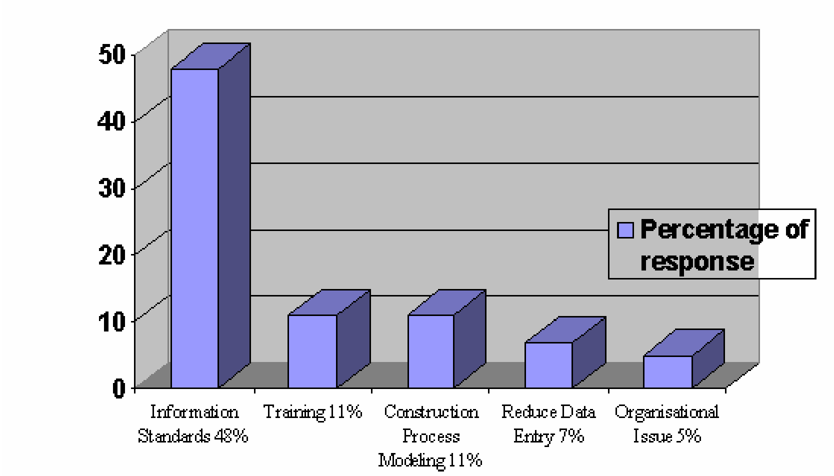


Figure 9. Critical issues that require AEC industrial attention [adapted from Froese et al. (2001)].

Froese et al. (2001) concluded that these surveys provided a fairly strong and consistent support for several important predictions about the future of IT for project management. A number of general attitudes towards information technology are:

- All of the surveys suggested that a company's proficiency in IT would have a positive impact on their competitive advantage, but the belief was noticeably stronger in year 2001 survey than previously.
- Currently, there is a higher expectation on the use of CAD and 3D models, project estimating and scheduling, and automation and robotics.
- In terms of data exchange, the highest expectation is a general-purpose standard such as XML and construction industrial standard IFCs.
- The most likely contribution of information technology to AEC/FM is through improving communication and interoperable standards, particularly in general-purpose standards and IFCs.

3.2 The role of 3D CAD models in design and construction

An understanding of the ways in which 3D CAD modelling techniques can be used to support and reflect design thinking can lead to the development of a greater integration in the building design and construction industry. Since the inception of CAD, computers appear to have played a vital role in the practice of architecture, engineering and their allied professions. This is, however, merely an illusion. It did not happen simply because most designers in practice were not formally trained to use the computer as a productivity tool, and they were therefore unfamiliar with its capabilities. In fact, there are many designers who still develop conceptual sketches for a project, then pass these sketches on to draftsperson who create 2D design development and construction drawings with little integration, if any, with other consultants.

The primary purpose of a 3D CAD model needs to be established at an early stage in a project. 3D CAD modelling can be used in structural, lighting, acoustic, thermal, bio-climatic and spatial analysis. There is still a common misconception that CAD systems are just drafting tools for use in the post-design stages of work rather than having a much richer role to play during designing and construction. 3D CAD models can help to resolve ambiguities, provide linkages to design data and present computerised visualisation. Discovering design conflicts and inconsistencies early is far less costly than repairing design and construction mistakes in buildings as cost increases exponentially at every stage from conceptual design to construction as shown in Figure 10. 3D CAD modelling allows for such inconsistencies to be discovered before construction and therefore for better design and construction decisions to be made in the very early stages of the design and construction phase.

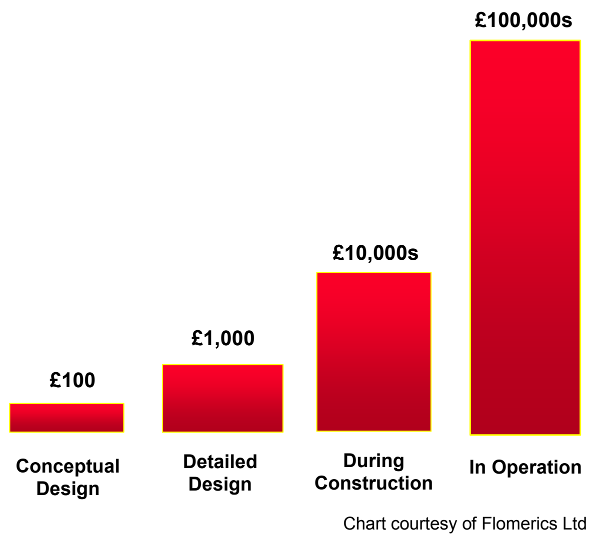


Figure 10. Exponential increase in repairing construction and design mistakes.

3.2.1 Interoperability and Data sharing for 3D CAD models

Interoperability is the capability of devices from different manufacturers to communicate and work together. This meaning has great benefits that cannot be realised without establishing rules of standardisation and compatibility among different sectors in the building and manufacturing industry. The lack of true interoperability is arguably the single largest software problem facing today's manufacturers. It can cause time-to-market delays, bottlenecks, errors, lost data, quality problems, and extensive reworking of parts. This problem also impedes the cost-effective outsourcing of design and production in global manufacturing.

There are currently a variety of software packages available designed to facilitate interoperability between the products of various manufacturers and suppliers. They assist in converting and viewing CAD formats; translating between many different CAD/CAM (computer aided manufacturing) formats; and modifying and sharing objects. Moreover, there are Application Service Providers (ASPs) for the translation and adaptation of 3D solid models, enabling rapid sharing of 3D engineering data across design and manufacturing firms, and their clientele, regardless of their installed CAD systems (Reffat, 2002).

3.2.2 3D CAD modelling as a Means for Knowledge sharing

Building design is a process that has often been considered as an activity carried out only by architects and engineers. They are co-designers in proposing the end product of a building form, materials, supporting structure and environmental control services. However, while these co-designers share the same task, that is designing a specific building, they tend to think largely about concepts with regard to their own particular interest in the evolving building solution. Their co-designing work is in the form of knowledge contribution from each of them based on their particular experience. Their method of working together will therefore be in the form of knowledge sharing in order to evolve a commonly agreed building solution. Knowledge sharing among co-designers occurs when they allow for the exchange of design information that presents their individual knowledge and contribution usually in the form of inputs and outputs (Cornick, 1996). Examples of the knowledge shared among the co-designers include: shapes, proportions, arrangements, and materials of building elements brought together in an overall building form, structural and services elements and enclosures with regard to their structural stability, and size, shape and arrangement of building services.

The development of a 3D CAD model has the essential role of supporting knowledge sharing among co-designers. One set of designers conceives the overall form and how all its parts would fit and work together; other designers then conceive how all the system parts can be engineered as an overall assembly. The degree of realism that can be created in the 3D CAD model and the ease by which 3D views can be generated are essential in knowledge sharing. This is a much richer notion than a simple demonstration of the material, finish and form of building objects (Reffat, 2002).

The Computer User High-Tech dictionary[4], defines an object-oriented database as a database in which data is stored as objects in an object-oriented programming environment, and which is managed by an object-oriented database management system (OODBMS). Object-oriented databases evolved from the increasing demands of integrating data management with object-oriented programming languages such Java, C, and C++. ODBMS is the outcome of this integration. A primary advantage of an OODBMS is that accessing objects in the database is done in a transparent manner such that interaction with persistent objects is no different from interacting with in-memory objects. However, whether an object-oriented database is a good choice depends on the business need, level of performance, and complexity of data. Generally, ODBMS provide the lowest cost for development and best performance combination when using objects because they store objects on disk and have the transparent program integration with object-oriented programming languages (Web service website[5]). Table 1 is an extracted

[4] http://www.computeruser.com/resources/dictionary/
[5] http://www.service-architecture.com/object-oriented-databases/index.html

summary of the advantages and disadvantages of object-oriented databases[6].

Table 1. A Summary of Advantages and Disadvantages of Object-oriented Database

Advantages	Disadvantages
• Composite Objects and Relationships: Objects in an OODBMS can store an arbitrary number of objects, which is a better model of the real world entity than the relational tables. • Class Hierarchy: Data in the real world usually has hierarchical characteristics. • Circumventing the Need for a Query Language by transparently accessing objects • No concern about primary key because of uniqueness of object ID (OIDs). • One data model provides substantial saving in development and maintenance.	• Schema changes in an OODBMS will • Involve a system wide recompile. • Updating all the instance objects within the database can take an extended period of time depending on the size of the database. • An OODBMS is typically tied to a specific language via a specific API, which means language dependency. • Lack of Ad-Hoc Queries

3.3 Building Maintenance

A simple statement of the maintenance objective for a building is that building systems are always available to support building function, and where applicable without ever limiting production. Where production is relevant, the building maintenance objective is that the cost of any maintenance activity should be less than the expected marginal value of production enabled by the planned activity. A managerial challenge is to allow the building and plant operating engineers as well as the owner to share information about current building component status, the business situation and for strategic planning in order for them to meet time-varying objectives.

Supporting this objective is difficult. It is difficult to assess the amount of risk posed by an observed non-critical problem to future production. There are multiple goals (e.g., high long-term availability, minimal short-term cost); goals change (e.g., between availability and cost concerns); goals conflict; indicator data are almost never completely reliable or adequate. The problem has multiple aspects, including interpretation of observed data, diagnosis of problems, repair and maintenance planning, and business evaluation of the value-added of different repair and maintenance options. Finally, significant judgment is needed to interpret both available engineering and business data, and clear business policy is needed to define the "value" of maintenance.

3.3.1 Corrective and Preventative Maintenance

Consider an example of a boiler that has tripped due to a drop off of the primary airflow signal. The technician goes out to the unit and, seeing a differential pressure of zero, replaces the transmitter. Repair work such as this, performed following breakdown, is called *Corrective Maintenance* (CM), also known as *reactive maintenance*. This paradigm can be described as “fix it *when* it breaks”.

The technician wants to prevent this from becoming a problem again and schedules a check on the boiler transmitter every six months. Maintenance performed periodically based on utilisation metrics such as hours of operation or calendar time in order to prevent failures is called *Preventative Maintenance* (PM), also known as *scheduled maintenance*. This paradigm can be described as “fix it at *regular intervals*”.

[6] http://www.25hoursaday.com/WhyArentYouUsingAnOODBMS.html

Most building, building complexes and plants operate with a combination of corrective maintenance, which accounts for over 50% of current practice, and preventative maintenance, which accounts for almost 20% of current practice (Bowers).

While preventative maintenance may avoid the unscheduled down time and costly repairs associated with reactive maintenance, it may be scheduled more often than is necessary. In fact, preventative maintenance is not needed most of the time it is performed, thus introducing costs that are to some degree unwarranted and therefore should be minimised without sacrificing plant performance.

3.3.2 Predictive and Proactive Maintenance

Continuing with our boiler example, once the transmitter has been replaced, the boiler is started up again, but after a short period of time it again trips due to a zero primary airflow signal. Now looking at the situation more closely, the technician finds that the impulse lines from the transmitter to the orifice plate had become plugged and were the true source of the problem. The lines are cleaned out and the unit is brought back up.

The technical team now decides to add on-line instrumentation, like that shown in Figure 11, that measures the boiler's airflow and sends the signal to a processor that monitors this and other signals and uses a model of the system in order to anticipate and diagnose failures before they occur; this is called *Predictive Maintenance* (PdM). By isolating faulty components and calculating best time to repair or replace them, this approach minimizes both maintenance labour and the risk of an unscheduled outage and can be characterised as "fix it *just before* it breaks".

Figure 11. On-line instruments for predictive maintenance

In traditional approaches, the interpretation of a measurement had no correlation to previous tests. As long as an asset's parameters were found to be within specification, it was considered to be fine and no action was taken. In current approaches, when a Computerized Maintenance Management System (CMMS) issues a work order, it identifies the specific test, instrument, and settings to use for the piece of equipment in question. A technician performs the test with the same settings and procedures used the time before. After the test, the technician can download data to a system that aggregates data from previous tests and builds trend reports.

For more than a decade, PdM has enabled facilities technicians to identify and solve problems before they have a chance to damage equipment. PdM now accounts for over 50% of *best* practice (Bowers). For example, Intel uses this approach in its semiconductor manufacturing plants.

In *Proactive Maintenance*, the general time frame that a component will fail is determined *before* the failure occurs or is about to occur. Typically, a model of the system is used to anticipate failure. This approach minimises the risk of failure by eliminating root causes and currently accounts for approximately 10% of best practice (Bowers). The Data Mining component of the system is aimed at allowing proactive maintenance to occur based on the knowledge and patterns extracted from past maintenance records, thus allowing for generalisation from experience of similar assets as well as pinpointed information specifically relevant to a particular asset.

3.4 Life Cycle Modelling

3.4.1 Life Cycle Cost

Life cycle costs (LCC) are summations of cost estimates from inception to disposal for both equipment and projects as determined by an analytical study and estimate of total costs experienced during their life. LCC not only comprises initial acquisition cost, it also contains other cost like 'ownership cost" – operation costs, maintenance costs, logistics costs, etc, which is usually higher than the original acquisition cost. The major objective of LCC analysis is to choose the most cost effective approach from a series of alternatives so the least long term cost of ownership is achieved (Barringer, 1996).

Based on an online report on Life Cycle Cost analysis (Kawauchi and Rausand, 1999), it is believed that a typical range of the ownership costs is 60 percent to 80 percent of the total LCC. For example, according to a standard (SAE ARP-4293), of LCC analysis, the ownership costs of a fighter aircraft are 53 % of the total LCC, and the ownership costs of a basic trainer aircraft occupies 91 % of the total LCC.

3.4.2 The Importance of Life Cycle Modelling

The life cycle cost concept is addressed in the British Standards as 'Terotechnology' which is defined as a combination of management, financial, engineering, building and other practices applied to physical assets in pursuit of economic life-cycle costs.[7] Life cycle cost modelling (LCM) contributes to competitiveness of the company by providing strategic planning on rehabilitation and enhanced information for decision making. LCM helps facility managers in evaluating alternative equipment and process selection based on total costs rather than the initial purchase price. The multidimensional information that LCM presents is merged from hybrid project domains such as management, engineering, as well as finance. LCM may be applied in a wide range of critical functions, including:

- Evaluation and comparison of alternative design;
- Assessment of economic viability of projects and products;
- Identification of cost drivers and cost effective improvements;
- Evaluation and comparison of alternative strategies for product use, operation, test, inspection, maintenance, etc.;
- Evaluation and comparison of different approaches for replacement, rehabilitation/life extension or disposal of aging facilities;
- Optimal allocation of available funds to activities in a process for product development;
- Assessment of product assurance criteria through verification tests and their trade-offs;
- Long-term financial planning.

3.5 Virtual Environments

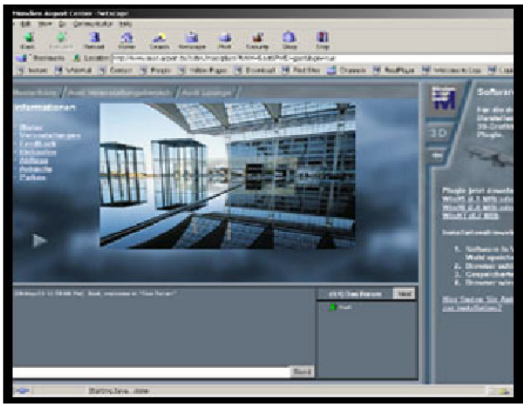
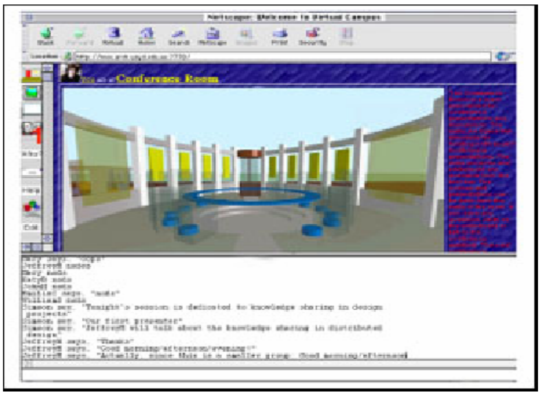
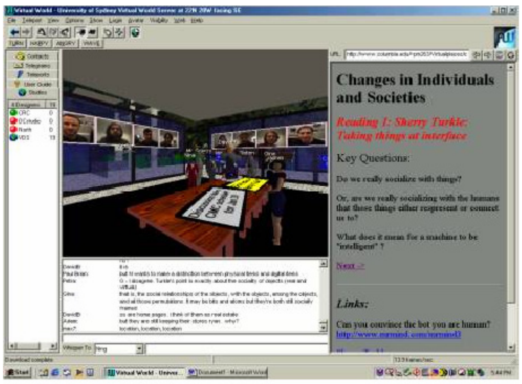
Virtual Environments (VE) are computer generated synthetic environments in which users are provided with multi-modal, highly natural forms of computer interaction. Virtual Environment research is concerned with creating artificial worlds in which users have the impression of being in that world and with the ability to navigate through the world and manipulate objects in the world.[8] An overview of several interactive virtual environments platforms is illustrated in Table 2.

---

[7] http://www.barringer1.com/lcc.htm

[8] http://www.ctit.utwente.nl/programme/areas/vr.html

Table 2. Overview of several interactive Virtual Environments

Product	Functionality	Screenshot
The Blaxxun virtual worlds[9]	• Real time rendering • Avatars navigation • Provides digital communications • Focus on visualisation	 Munich Airport Center Using Blaxxun
LambdaMOO[10]	•Provide visual conference only •Text based approach •Consists of programmable objects •Manually activated function	 A Virtual Seminar Room
Active Worlds[11]	•Multi-user interaction • Avatars navigation •Programmable object with behaviour features •Extensible environment	 A Seminar Room in Active Worlds

[9] http://www.blaxxun.com
[10] http://www.ccs.neu.edu/home/eostrom/muds/lambdamoo.html
[11] http://www.activeworlds.com

The Virtual Worlds ---Microsoft Research[12]	•Built on a persistent and distributed object oriented database •Multi-users interactions •Programmable objects with behaviour features •Avatars navigation	 A Space in Virtual World
Virtools ----Virtools Behavior Company[13]	•An internet gaming environment •3D objects with behaviour features •Virtual programming interface to define behaviour •Avatar navigation and interaction	 A Virtual Room in Virtools Platform

3.5.1 Survey of Existing Life Cycle Models in Virtual Environments

Five existing life cycle modelling prototypes based on buildings with an object-oriented database in virtual environment platforms were surveyed. These projects include: ‘Future Home’, ‘Virtual Building Life Cycle’ (VBLC), LICHEE[14] (Life Cycle House Energy Evaluation), 'Product Model and Fourth Dimension (PM4D)’[15] and 'Centralhuset'.

Future Home Project

The Future Home Project[16] is developed at the Centre for Virtual Environments of University of Salford, UK. FutureHome utilises VR (Virtual Reality) methods and technology to explore its use in the construction of designs from prefabricated components and also the animated simulation of the construction process allowing users to explore constructability of designs.

One important by-product of the use of VR techniques is that building owners and occupiers can

[12] http://www.vworlds.org
[13] http://www.vtools.com
[14] http://www.cmit.csiro.au/innovation/2002-02/lichee.htm
[15] http://www.stanford.edu/group/4D/projects/calvin/PM4D.shtml
[16] http://www.nicve.salford.ac.uk/~norman/futurehome.html

subsequently interrogate an object-oriented database when there is a need to perform some maintenance functions or carry out a repair or replacement. Instead of handing over 'as-built' drawings, the products of FutureHome will be complete with 3D object-oriented models for use and adaptation by both owners and occupiers. This aspect increases the potential to build in some form of 'intelligent' function. Embedded technology could work in harmony with the object-oriented model to support occupiers. Figures 12 and 13 illustrate a building construction in the virtual environment and selecting a building component within the virtual environment respectively.

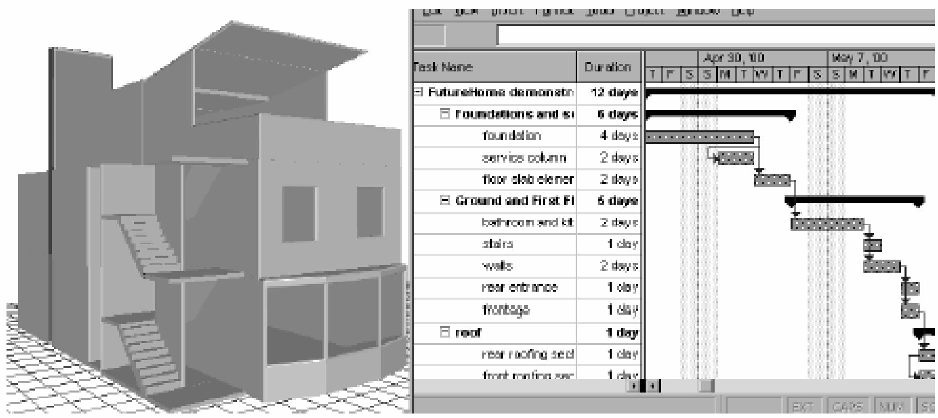
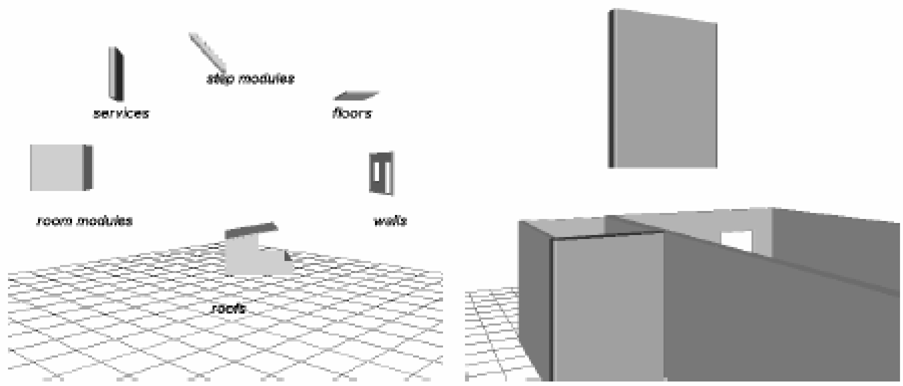


Figure 12. A building constructed in the virtual environment and its construction schedule (Murray, Fernando and Aouad, 2000).



(a) Selection of the components is via a rotating tree menu

(b) Component selected to be added to the environment

Figure 13. A Selecting a building component within the virtual environment (Murray, Fernando and Aouad, 2000)

The 'Virtual Building Life Cycle'

The Virtual Building Life Cycle[17] (VBLC) is a project by Fraunhofer Centre for Research in Computer Graphics (CRCG). This project connects 3D geometrical data to research data such as life expectancy and emissions data to standard database information like prices. The automated visualization of critical points of the structure in the past, present and future is a major advantage helping engineers to improve the duration of the lifecycle and reduce the costs. The project has been implemented as a prototype plug-in to the ARCADE (Advanced Realism CAD Environment) CAD system. The geometry was imported via the standard interface ACIS. The values from the database that belong to certain 3D objects in the scene can be accessed and visualized in various modes. The user is able to access values like "unit price" and "life expectancy" while navigating through the virtual scene and model of a

[17] http://www.crcg.edu/vblc/

building.

According to (Linnert, 1999)[18] the process of accessing data in VBLC through navigation and selecting 3D objects is considered satisfactory. However, the access to the database via an ODBC and export of VRML for World Wide Web visualization was not implemented. The VBLC provides a good research prototype, especially in methodology of building 3D model and linkage of maintenance data with geometrical objects. Figure 14a illustrates a pop-up window showing the information of a selected building element and Figure 14b a visualisation of construction parts respectively.

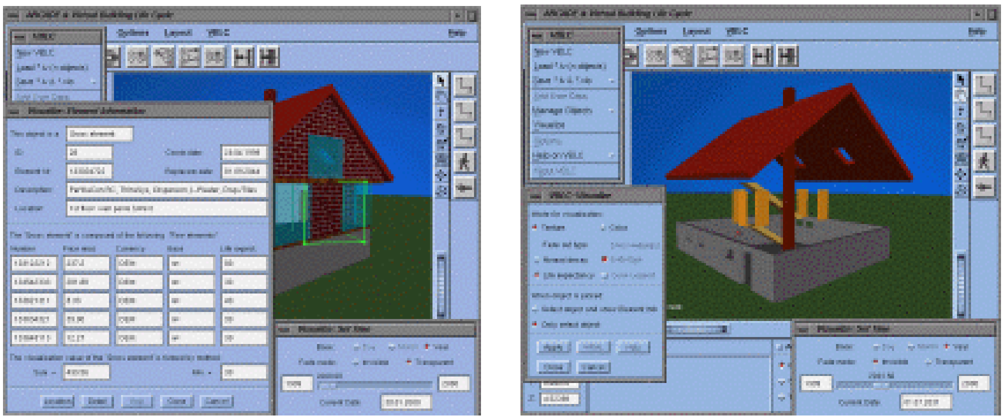


Figure 14. Screenshots form the Virtual Building Life Cycle *(a)* A pop-up window shows the information of picked element.[19] *(b)* Visualisation of the construction parts faded according to life expectancy[20]

LICHEE

LICHEE[21] (Life Cycle House Energy Evaluation) is developed by CSIRO BCE and the Cement and Concrete Association of Australia. The LICHEE system utilizes 3D object-oriented CAD data, together with data about a building's total environmental impact, to create a lifetime profile for a particular design. The object-oriented nature of LICHEE enables the creation of a design with 3D objects such as a room or a roof, and each time an object is altered or modified, existing objects remain unchanged. This makes the entire design easy to modify or to be compared with alternative designs. LICHEE is an IFC-compliant system that improves the interoperability of building industry and transferring of data between CAD systems and databases. Figure 15 shows a screen shot of LICHEE VR interface.

Figure 15. A screen shot of LICHEE VR interface[22]

[18] http://www.crcg.edu/vblc/Download/download.html http://www.crcg.edu/vblc/Download/download.html
[19] http://www.crcg.edu/vblc/Screensh/screensh.html
[20] http://www.crcg.edu/vblc/Screensh/screensh.html
[21] http://www.cmit.csiro.au/innovation/2002-02/lichee.htm
[22] http://www.cmit.csiro.au/innovation/2002-02/pdf/lichee.pdf

Product Model and Fourth Dimension[23]

The project Product Model and Fourth Dimension (PM4D) is developed by CIFE (Centre for Integrated Facility Engineering), Stanford University. CIFE is a concurrent development project that paralleled the construction of HUT-600 in which the project team experimented with 3D stereoscopic virtual reality, the 4th dimension, sustainability, life cycle cost and impact analyses, project databank, and Industry Foundation Class. Both the breadth of product model information-based data and the depth of research experimentation were exceptionally unique in such an active construction project. The PM4D Development Project was one of the pioneering projects that employed IFCs as the primary standard for information exchange and, hence, was a pilot development project. With ArchiCAD from Graphisoft, the architects created a 3D model in the conceptual planning phase and continually maintained and updated the product model. The mechanical engineers employed Progman Oy's MagiCAD to conduct 3D Modelling of cooling and heating systems. Figure 16 shows the scope of testing IFC-based project data exchange of the HUT-600 project.

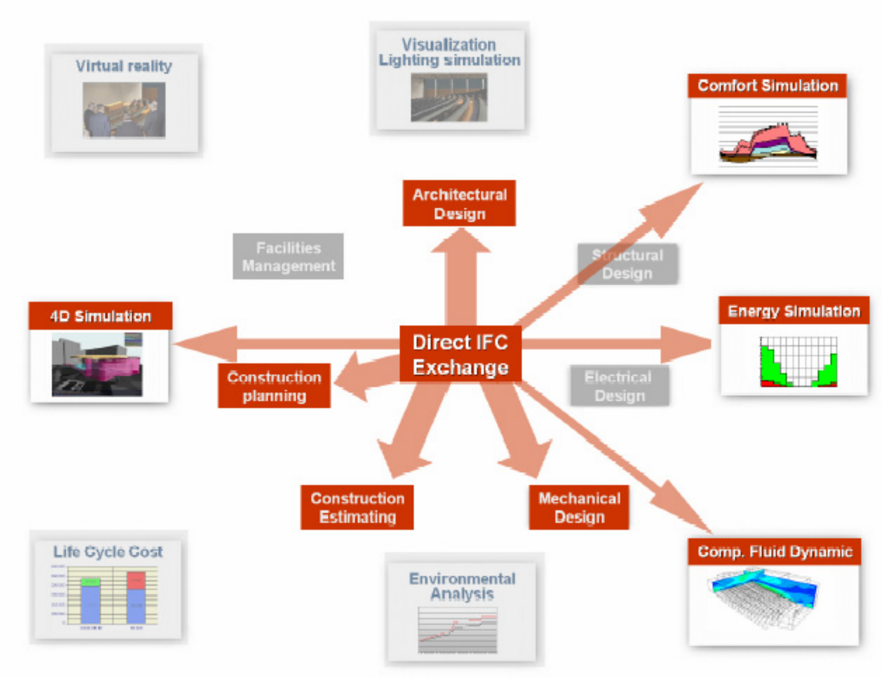


Figure 16. Scope of testing IFC-based project data exchange on the HUT-600 project.[24] (Fischer and Kam, 2000)

Centralhuset Project: Virtual Reality at a Construction Site

A Virtual Reality model of "Centralhuset" -a hotel and office block to be built in Gothenburg, Sweden - was installed at the building site in order to aid the planning and performance of site related activities and related issues. The information was gathered primarily from 2D CAD drawings, but also from some 3D CAD drawings (the steel structure, piles and pile footings). By implementing the VR model at the building site, this Centralhuset model aims to improve communication between all people involved and consequently lower the construction errors and streamline the site-related activities. A 3D VR model was established in the Centralhuset. Figure 17 illustrates the processes of production and distribution of 2D drawings and 3D VR model.

[23] http://www.stanford.edu/group/4D/projects/calvin/PM4D.shtml
[24] http://www.stanford.edu/group/4D/download/PM4D_Final_Report.pdf

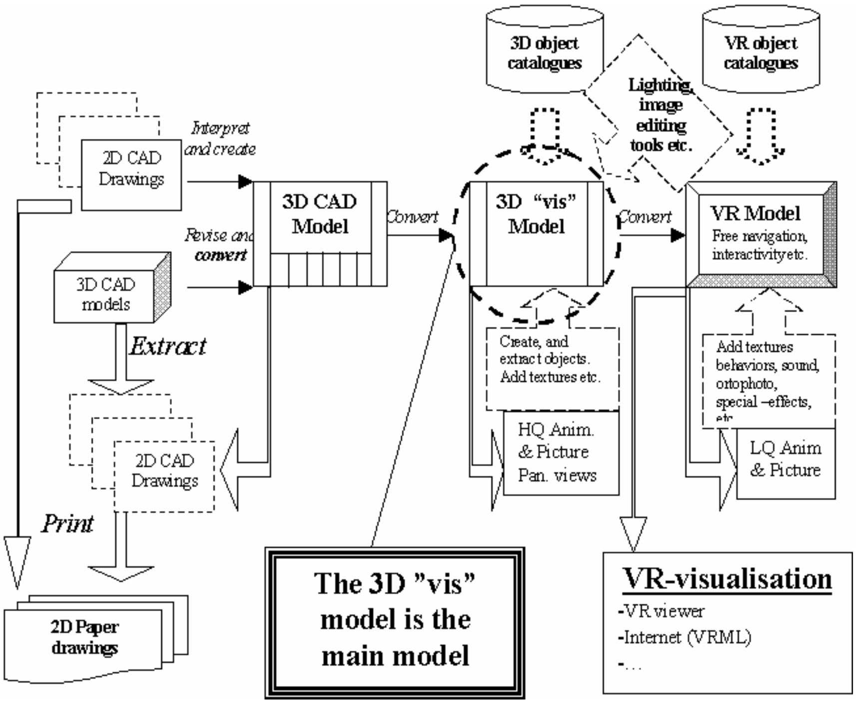


Figure 17. An illustration of the production and distribution process of 2D paper drawings and the 3DVR model[25]

3.5.2 Summary of Existing Life Cycle Model with Object Oriented Database and Virtual Environment

The various benefits derived from the surveyed projects include:

- Use of virtual environments as platforms to provide an interactive interface to improve communications between all team members as well as a simulation tool in enhancing predictable strategic planning within the whole life cycle of selected buildings.
- Application of the IFC-compliant object oriented database in standardizing the data exchange and facilitating the manipulation, reusability of project information.
- Linkage of maintenance data to 3D CAD model provides the potential of future development of intelligent life cycle analysis and control capability.

However these existing prototypes do not provide the capability of data mining of the hybrid data gathered from different stage of a building's life cycle, and do not provide performance-gaining life cycle analysis. This CRC project aims to bridge this gap to accomplish an efficient way in achieving an intelligent knowledge-gain.

3.6 Existing Commercial Data Management Systems

3.6.1 BEIMS

BEIMS[26] is a Facilities Management Software package produced by Mercury Computer Systems. It consists of 5 core modules:

- The Asset Management module provides a centralised mechanism to control, regulate and maintain an organisation's assets, including service histories, financial information, contract details, warranties, and components.

[25] http://vrlcb.sm.chalmers.se/applied.shtml

[26] http://www.beims.com

- The Planned Maintenance module enables the Engineering group to set up a schedule of preventative maintenance for the organisation's assets. This schedule allows many different tasks to be planned against any selected asset, e.g., bi-weekly inspections of a chiller unit through to a yearly overhaul of the chiller. When work is generated from the Planned Maintenance module, it is transferred into the Work Order module as an active job. If unplanned events occur, the maintenance schedule must be adjusted.
- The Work Order Module provides an online computerised method of recording job requests, delegating the tasks and tracking costs involved with engineering work. The work order system interfaces to planned maintenance and handles both pre-scheduled preventative maintenance and ad hoc/breakdown work.
- The Cost Control Module provides users with various costing reports with individual transaction details for work orders or summary information.
- The Information Setup Module has two distinct sections, the general codes and the security and parameter section. The general code option is a centralised area where codes for cost centres, expenses, departments, buildings, etc. are set up and stored. The security and parameter section allows for the definition of user profiles, access rights, financial year starting dates, etc.

3.6.2 INSIGHT

Insight[27] is part of Siemens' APOGEE Building Maintenance System. Its capabilities include system scheduling, report generation, troubleshooting information, system configuration, and system administration. Figure 18 illustrates a screenshot of APOGEE building maintenance system.

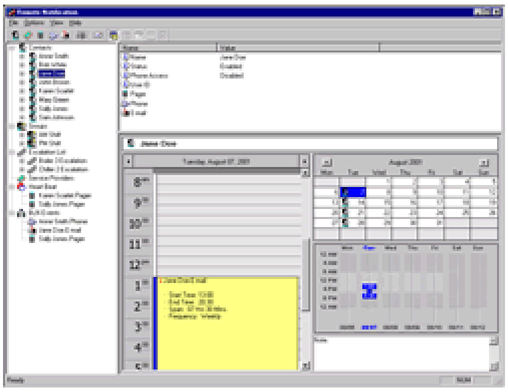


Figure 18. A screenshot of APOGEE Building Maintenance System.

3.6.3 MAXIMO

Maximo's[28] capabilities include:

- Track equipment, costs, histories, and failures
- Schedule work orders and record maintenance work
- View work plans, schedules, costs, labour, materials, etc.
- Generate PM work orders automatically
- Optimise schedules through what-if analysis
- Maintain detailed company, service contract, and tool records to plan and analyse maintenance work
- Maintain personnel files
- Create purchase orders for materials and services
- Track inventory

[27] http://www.sbt.siemens.com/BAU/products/BA_management_insight.asp
[28] http://www.mro.com/corporate/products/maximo.htm

Figure 19 illustrates screenshots of Maximo.

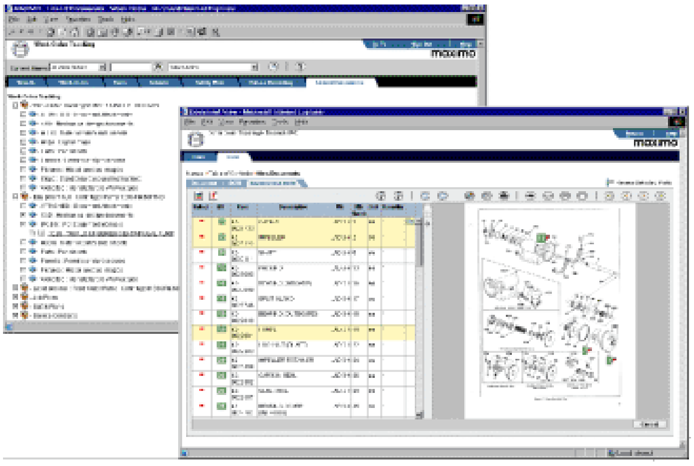


Figure 19. Screenshots of Maximo.

3.6.4 APERTURE

Aperture[29] is a web-based product used to visually track and manage personnel and assets. It provides graphical displays of occupancy and floor plans that are colour-coded by department, vacancy, usage, and personnel type. It also provides forms for requesting personnel moves, space re-allocation, routine facility services, etc. Figure 20 shows screenshots of Aperture.

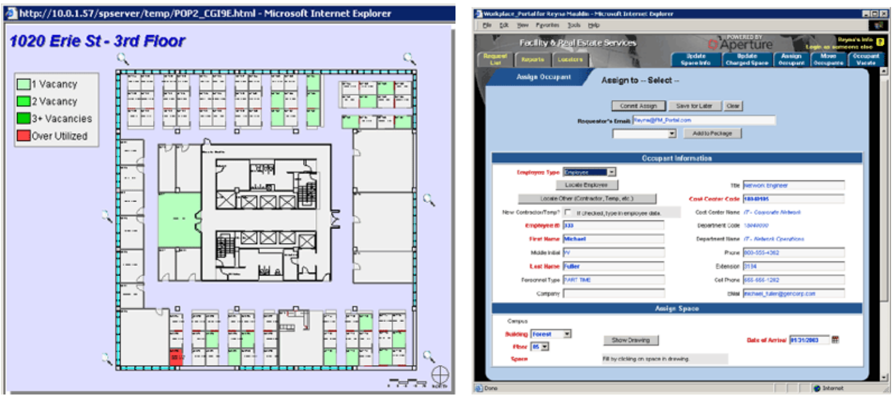


Figure 20. Screenshots of Aperture

3.6.5 ARCHIFM

ArchiFM[30] is a computer-aided facility management tool for:

- Space, Occupancy and Move Management: floor plan visualisation in various modes; occupancy analysis; and move automation.
- Asset Management assigned to person or room and synchronised with floor plan and assets are organised in a hierarchical fashion,
- Drawing Management includes ArchiCAD and customisable Graphical User Interface. Figure 21 shows a screenshot of ArchiFM.

[29] http://www.aperture.com/solutions/workplace_solutions.html
[30] http://www.graphisoft.com/products/archifm/

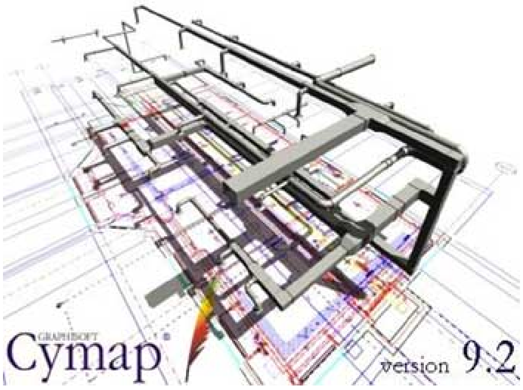


Figure 21. Screenshot of ArchiFM

3.7 Existing Knowledge Discovery Systems

3.7.1 SERFIN

The SERFIN[31] Project is developed at the KBS-Media Lab, Lund University (Christiansson 1998). SERFIN provides facilities management knowledge handling that aims to:

- Identify and capture problems that arise in connection with technical maintenance of buildings
- Make the problem solution process arising during maintenance of buildings more effective
- Make experiences from technical maintenance easily available and accessible in time and space

SERFIN is currently based on web pages on the Internet together with text query. The Demonstrator provides advice on how to fix an item in need of repair. SERFIN is used for searching building maintenance information with the use of hierarchical menus to choose: Building Part, Material, Environment, Problem and Action. The result of search includes similar documents are presented in order of relevance. Figure 22 shows a screenshot of SERFIN.

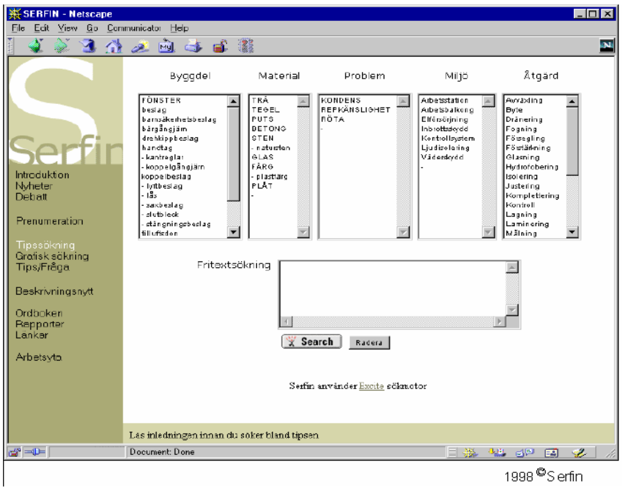


Figure 22. Screenshot of SERFIN.

3.7.2 Intelligent Real-time Maintenance Management System

31 http://it.civil.auc.dk/it/delphi/KBS/projects/serfin.html

The Intelligent Real-Time Maintenance Management (IRTMM) System is a project of the Centre for Integrated Facility Engineering (CIFE) at Stanford University. The objective of the system is to perform *value-based* plant maintenance as needed (Kunz 1995). The IRTMM system provides integrated subsystems for the following aspects of the maintenance and repair planning problem:

Situation Assessment (SA): interprets observed data as being normal or abnormal and diagnoses causes and effects of plant equipment problems. It analyses system performance to identify indications for condition-based maintenance. Given specialised input data from instrumentation (e.g., pressures, flow rates) and expert diagnostic systems (e.g., vibration analysis), the SA focuses on systems diagnosis. It identifies root causes and effects of component problems, where some of those causes and effects are in the component with a problem, and others are in subcomponents or connected systems. The SA uses a combination of methods for assessment:

- model-based diagnosis to identify details of possible problems
- heuristic classification to identify idiosyncratic problems
- case-based reasoning to compare observed data with previous cases

Planning: Given a set of problems to repair, provided by the SA or by a user who is considering maintenance for any reason, the planner builds plans of the activities needed to repair a diagnosed problem. The planner can also merge related plans for the same or different components that can be performed during the same plant outage. Specifically, it identifies required activities to perform particular repairs, suggests alternative start times, identifies plans that can be merged, and identifies dependencies among plans that imply ordering of the sequence of plans.

Value Analysis (VA): identifies the dollar costs and predicted benefits of performing every selected repair plan at different times. The user can consider the repair for any reason. Using predicted power demand and plant operating and maintenance cost data, the VA assesses the net unit operating costs associated with performing plans at different possible future times. The VA uses a decision-analysis procedure: it considers possible choices of repair actions and the chance outcomes that can arise given any choice. The value of a choice depends both on the choices and chances and on the probabilities and costs of each occurrence. Since it is usually impossible to get good probabilities, the system identifies "break-even" probabilities of failure such that an owner is indifferent between two options. The user then judges whether actual probability of failure exceeds the break-even.

The IRTMM system provides interactive analyses to facilitate engineering decision-making, not automate it. Given some data from a data acquisition system, the system identifies candidate causes of problems and predicted effects. The user selects one or more components to analyse in more detail. After reviewing the system-generated plans and value analysis, the user selects one or more components for repair, selecting both the desired repair activity and the planned repair time, after considering the system-generated options.

The system is designed to reside on a computer network. It can receive component status information from an on-line data acquisition system and any available diagnostic expert systems, staff and equipment availability information from a computerized maintenance management system (CMMS), and projected product demand, cost and selling price data from a business database. Recommended work could be logged in the CMMS.

3.8 Available Data Mining Software

3.8.1 Data Mining Software Comparison

There are currently hundreds of mining tool vendors. A comprehensive list of software tools for data mining can be obtained at http://www.andypryke.com/university/software.html and http://www.knowledgestorm.com. A large number of reviews of data mining software are also available at (Goebel and Gruenwald, 1999; Elder and Abbott, 1998). The six leading data mining tools have been reviewed and listed in Table 3. The WEKA package is included into this table, since it is open source software built in Java which can easily be embedded into a pilot decision support system – one of the expected deliverables of this CRC project. Table 4, Table 5, and Table 6 show comparisons between these six tools on classification methods, clustering feature and link analysis, respectively.

Table 3. Data Mining Tools Evaluated

Product	Company	URL
Knowledge Studio	Angoss	www.angoss.com
Clementine	SPSS Inc.	www.spss.com/spssbi/clementine/
Enterprise Miner	SAS Institute	www.sas.com/products/miner
Intelligent Miner	IBM	www.software.ibm.com/data/intellimine/
Oracle 9i Data Mining	Oracle Corporation	www.oracle.com/ip/deploy/database/oracle9i/
S-Plus	Insightful Corporation	www.insightful.com
WEKA	University of Waikato	www.cs.waikato.ac.nz/~ml/weka

Table 4. Classification Methods Comparison

Product	Regression	Decision Tree	Neural Nets	Bayes Class	Meta Learn	K-nearest neighbour	Explain based
Knowledge Studio	X	X	X	X	X		X
Clementine	X		X			X	
Enterprise Miner	X	X	X				X
Intelligent Miner	X	X	X			X	
Oracle 9i Data M,	X	X	X	X			X
S-Plus	X						
WEKA	X	X	X	X	X		

Table 5. Clustering Feature Comparison

Product	Partitioning	Hierarchical	Density	Model	Fuzzy
Knowledge Studio	X				X
Clementine	X				X
Enterprise Miner	X				
Intelligent Miner	X				
Oracle 9i Data Mining	X				
S-Plus	X				
WEKA	X	X	X	X	X

Table 6. Link Analysis

Product	Association Rules	Time Series	Sequential Discovery	Bayes Networks
Knowledge Studio				X
Clementine	X	X	X	
Enterprise Miner	X	X	X	
Intelligent Miner				
Oracle 9i Data Mining				
S-Plus				
WEKA	X	X	X	X

As an outcome of this survey, the focus was narrowed to include only the IBM Intelligent Miner and WEKA since they perform better in scalability and functionality than other packages and are also leaders in commercial and non-commercial data mining software packages. WEKA is freely available, and Intelligent Miner has also been obtained without cost through the IBM Scholars Program.

3.8.2 IBM Intelligent Miner DB2

The IBM DB2 Intelligent Miner Version 8.1 is a set of the following products (IBM Publications)[32]:

- Intelligent Miner Scoring (IM Scoring)
- Intelligent Miner Modelling (IM Modelling)
- Intelligent Miner Visualizing (IM Visualizing)

IM Modelling is a DB2 SQL application programming interface that consists of a set of database objects that enable you to build data mining models from information that is held in IBM DB2 databases. The DB2 SQL interface supports the rapid creation of data mining modelling applications. IM Modelling enables you to build mining models that cover three data mining functions: Association, Classification and Clustering. IM Scoring can be used to deploy PMML (Predictive Model Markup Language) models that were created by one of the Intelligent Miner products or by other applications and tools that support interoperability through the use of PMML models. The PMML is a standard format for data mining models, which is based on XML. PMML provides a standard that enables data mining models to be shared between the applications of different vendors (Data Mining Group (DMG)[33]. IM Visualizing provides the functionality of browsing PMML models that are created by Intelligent Miner products or by other applications and that are PMML-compliant.

Advantages:

- It is a commercial software package, which may be better for the delivered prototype to industrial partners.
- Takes data in SQL format.
- Performs many different data mining methods, including decision trees and neural networks.
- Has text mining capability, which may be useful for free text fields of repair reports (where much of the useful data may be stored).
- Has sophisticated visualisation software.

[32] http://www.elink.ibmlink.ibm.com/public/applications/publications/cgibin/pbi.cgi
[33] http://www.dmg.org

Disadvantages:

- Does not appear to support meta-learning, i.e., combining the results of multiple data mining techniques on the same data.
- Does not appear to be very easily extensible.
- Source code is not provided.

3.8.3 WEKA

WEKA[34] is a collection of machine learning algorithms for solving real-world data mining problems. The algorithms can either be applied directly to a dataset or called from your own Java code. WEKA not only contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization, it is also suitable for developing new machine learning schemes. Figure 23 is a decision tree classifier (weka.classifiers.UserClassifier) that provides users a friendly GUI to allow them to manually construct a decision tree by defining splitting criteria in the instance space. The structure of the tree can be viewed and revised at any point in the construction phase.

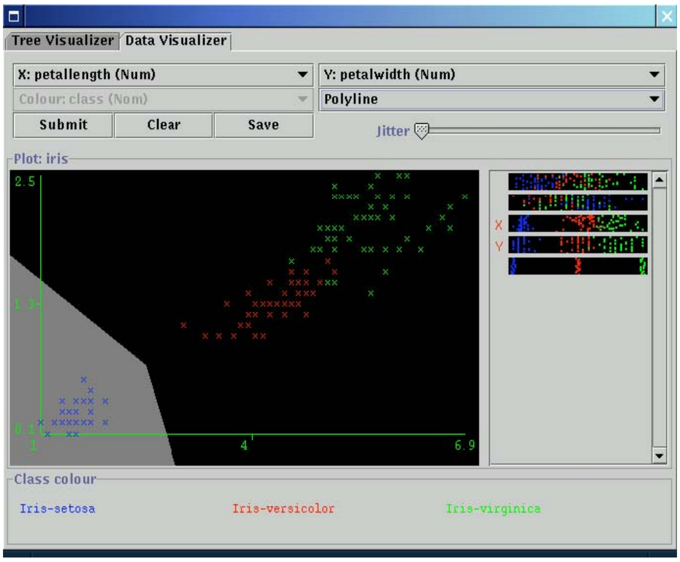


Figure 23. An example of visual analysis using WEKA[35]

Advantages:

- It is an open source software package, making it possible to modify or extend its capabilities.
- It is freeware that can be used in any system as long as its creators are credited.
- Supports meta-learning techniques such as bagging and boosting.
- Has a GUI version that appears to be adequate for our purposes.

Disadvantages:

- Does not appear to include software for neural networks, but other ANN packages can probably be added without too much difficulty.
- Input data must be in their own specialised format (ARFF), but this only requires storing the data in comma-separated values format and adding attribute information and a few tags.

[34] http://www.cs.waikato.ac.nz/~ml/weka/
[35] http://www.cs.waikato.ac.nz/~ml/weka/gui_explorer.html

3.8.4 Conclusions on Data Mining Software

It is noted that commercial data mining vendors in many cases do not disclose the nature of the algorithms. Therefore, our review of these tools from the point of the algorithms each product uses is somewhat limited. Also in the last three years, a significant convergence has occurred between the major tools that not only look very similar from user's point of view but are also similar in terms of their overall performance.

4. SURVEY OF AVAILABLE INDUSTRY DATA

A useful and productive application of data mining for improving the maintenance management of buildings is primarily dependent of having an appropriate, rich and diverse maintenance data set. Acquiring such a building maintenance data set has been an extremely difficult task in this project. This Section introduces a thorough analysis of maintenance data sets obtained from Queensland Department of Public Works and Central Sydney Area Health Service. The project's requirements on the quantity and quality of industry maintenance data are:

- Data must be available down to the element level;
- Data must be accompanied by corresponding 2 or 3D CAD data;
- All assets referred to in maintenance data must also be included in CAD data so that a one-to-one correspondence exists for an item as it appears in both types of data;
- Data must cover a period of at least 2-3 years (the longer the better).

This Section concludes by discussing the minimum requirements listed above in relation to the maintenance and CAD data provided by industry and their appropriateness as maintenance data sets to be used for data mining.

4.1 Queensland Department of Public Works

CAD and maintenance data was received for a building constructed and maintained under the supervision of QBuild, a commercial business unit of the Queensland Department of Public Works (QDPW). The data was provided by David Harrison, Assistant Manager for Product Development at QBuild, with the assistance of Teng Hee Tan, Manager of the Asset Management Research and Development Branch of the QDPW Building Division and member of the CRCCI Life Cycle Modelling and Design Knowledge Development in Virtual Environments project team.

4.1.1 Neville Bonner Building

The data comes from the Neville Bonner Building, shown in Figure 24, a state-of-the-art building at 75 William Street, Brisbane. Named after the first Aborigine elected to the Federal Parliament, it was designed, constructed and completed by Watpac Australia, in December 1998. The six-level building incorporates the latest in building technology, including advanced electrostatic filtration to combat air pollution from the nearby freeway, double glazing of the windows facing the freeway to reduce noise pollution, and precast concrete sunhoods which provide a visual buffer for office workers against the distractions of fast moving traffic. The building houses staff of the Department of Employment, Training, and Industrial Relations, and the Families Youth and Community Care Queensland.

Figure 24. Neville Bonner Building

The Neville Bonner Building (NBB) is under the purview of the Portfolio and Housing Unit of QBuild, who, because they deal with complex high-rise buildings in the CBD such as the NBB, have strict requirements for maintenance data collection and are the most likely to collect data to the component or element level. The data collected for all Portfolio and Housing Unit buildings was deemed as suitable for this project's purposes.

The Portfolio and Housing Unit also recognises the potential benefit of the project. For example, their recent contracts on carpet replacement are based on 15-year life-cycle calculations and corresponding manufacturer and installation guarantees, and any novel information that could be extracted from previous experience would be of great benefit.

4.1.2 CAD Data

The CAD data for the Neville Bonner Building can be grouped into the following seven classes:

1. Air Conditioning:

- Mechanical service layout (150259-02, 03, 05-07, 10, 13, 19, 25, 31, 35-39),
- Duct layout & details (150259-08, 09, 11, 12, 14, 15, 17, 18, 20, 21, 23, 24, 26, 27, 29, 32, 33),
- Toilet Exhaust & Outside (150259-42),
- Chilled and Condenser Water Schematic (150259-43),
- Chilled plant room layout (150259-04),
- Cooling tower plant room (150259-44).

2. Ceilings:

- Ceiling layout details from level 1 to level 6 (L1-CEIL to L6-CEIL)

3. PDT (Powell Dods & Thorpe) Drawings:

- Contains part drawings of stairs, porticos, etc. also includes parts drawing like doors, lights, corresponding e-mailed CAD data which have information of floor plans.

4. Raster images folder:

- Contains images of the selected building and logo of the government as well as other sketches of some glazing works.

5. Weathered Howe (Civil and structure design works by Weathered Howe):

- Civil: 96012c0-c6, civil construction details and William St. kerb and canal layout and detail, Margaret St. ramp and drainage set-out detail.
- All structural design works are put in a subfolder called 'Structural'.

6. Base Drawing:

- Furniture & Equipments (11-16_f-e),
- Room Schedule (I1b-rm to I6-rm),
- Part Drawing (I1_p to I6_p),
- Room drawing (I1_txt to I6_txt).

7. Drawings:

- Landscape Terrace Setout : A-065 to A069 in folder 0000-0999
- Foundation setout plan basement & level1 piles: A103 in folder 0100-0199
- William ST. Plaza Floor Plan: A-250,252,254, Stonework plan A-256, Extend of Tanking A-257,258,260,261, Portico plans A-265,267,

- Margrate St. Ramps floor plan A-270, Precast paving plan A273-274 in folder 02000299
- Elevations: A310-332, Wall elevations A-350-362 in folder 0300-0399
- Column details, tea room and toilets, balcony, joinery details, A551-576 in folder 05000599
- Wall, windows detail A600-630, Atrium Details A650-666 in folder 0600-0699
- Door Jamb detail: A-800 in folder 0800-0899
- Atrium Wall, Louvres, Glazing, Balustrade details and elevations: A-900-931 in folder 0900-0999
- Ceiling Plan of fire related ceiling A-1045, Atrium and lift lobby ceiling details A-1050,1051 in folder 1000-1999
- Stairs details: A-1100-1110, 1128,1129, Metal Balustrade A1120-1127 in folder 1100-1199
- Precast panel plan & elevation A-1200-1206, Roof detail, stonework detail and tanking details A-1214-1259 in folder 1200-1299
- Wall details, lift foyer, lift door, carpet tile setout: A-1550-1579 in folder 1500-1599
- Door signage details, signage layout, setout: A-100-1616 in folder 1600-1699.

4.1.3 Maintenance Data

QBUILD databases use the following three layer tree coding:

- **WIC Number:** a 5 digit number that uniquely identifies a building site. For example, a WIC of 22380 identifies Glenmore State High School.
- **Building Code:** a 3 digit number that identifies an individual building, except for the code 000, which identifies elements pertaining to the overall site. For example, a building code number of 22380000 relates to Glenmore State High School as a whole, and a building code of 22380016 identifies building number 16 at Glenmore State High School.
- **Building Component:** a 4 character alphabetic code that identifies a type of system (fixed plant). For example, 22380016AIRC identifies the air conditioning system in building number 16 at Glenmore State High School. These codes are as displayed in Table 7.
    - **Element Number:** a two digit number that uniquely identifies a specific asset. For example, 22380016AIRC01 identifies air conditioner number 1 in building number 16 at Glenmore State High School.

Table 7. Building component codes

Code	Description
AIRC	AIR-COND SERV
BLDG	BUILDING STRUCTURE
BMSY	BLDG MGT SYS
COMM	COMMUNICATIONS & DATA
EFIN	EXTERNAL FINISHES
EGND	EXTERNAL GROUNDS
ELEC	ELECTRICAL SER
ESTR	EXTERNAL STRUCTURES
FEQT	FIXED EQUIPMT
FINS	FINISHES FI/WI
FIRE	FIRE SERVICES
FITT	FITTINGS
FURN	FURNITURE & FITTINGS
GASS	GAS SERVICES
GIMP	GROUNDS - SITE IMPROVEMENTS
GLAZ	GLAZING
GSTR	GROUNDS - EXTERNAL STRUCTURES

HYDR	HYDRAULIC SERV
IFAB	INTERNAL BUILDING FABRIC
IFIN	INTERNAL FINISHES
LEQT	LOOSE EQUIPMENT
MISC	MISCELLANEOUS
PTEX	PAINTING EXTERNAL
PTIN	PAINTING INT
REFR	REFRIGERATION & ENVIRONMENTAL CONTROL
ROOF	ROOFING
SAFE	SECURITY & SAFETY SYSTEMS
SECR	SECURITY SERV
SUBS	SUBSTRUCTURE
SUPS	SUPERSTRUCTURE
TRAN	TRANSPORTATION
VENT	MECHANICAL VENTILATION

The maintenance database, which consists of work order records, contains a large number of fields, many of which relate to accounting and are not used.

In the data received, 2048 entries are recorded. The fields that describe maintenance and their contents are as follows:

1. Plan/Site WIC Number

`32573`	(for Neville Bonner Building complex)	2048 entries

2. Plant Number/Building
   One of the following:

`32573000`	(Neville Bonner Building complex)	199 entries
`32573001`	(Neville Bonner Building)	1834 entries
`32573001`	+component code (specific asset in NBB)	15 entries

for example, one plant number is: `32573001FIRED001`

3. Complex

`BRISBANE NEVILLE BONNER`	2048 entries

4. Component Code
   One of the following:

`AIRC`	441 entries
`BLDG`	1 entry
`BMSY`	6 entries
`ELEC`	390 entries
`FEQT`	7 entries
`FINS`	1 entry
`FIRE`	61 entries
`FITT`	325 entries
`FURN`	300 entries
`GLAZ`	5 entries
`HYDR`	273 entries
`IFIN`	1 entry
`MISC`	12 entries
`PTEX`	8 entries
`PTIN`	2 entries

REFR	25 entries
SAFE	2 entries
SECR	41 entries
TRAN	19 entries
VENT	1 entry
none	126 entries

5. BOMA Code

A description of the category of work performed is indicated using a set of codes put forward by BOMA, the Building Owners and Managers Association of Australia[36] (BOMA 1998). All but 32 work orders in the QBUILD data contain a BOMA code; some of the codes that appear include:

237	A/C Filter Maintenance	4 entries
238	A/C Water Treatment	6 entries
240	A/C Repairs	156 entries
243	A/C Spare (meaning unclear)	211 entries
291	Fire Sprinkler Systems	7 entries
292	Fire Protection – Thermal Alarms	39 entries
331	Electrical Repairs and Maintenance	179 entries
332	General Repairs and Maintenance	804 entries
333	Lamps and Tubes Repairs and Maintenance	229 entries
334	Locks, Keys, and Card Keys	9 entries
336	Plumbing Repairs and Maintenance	257 entries

6. Work Order Type
One of the following:

Client Funded		523 entries
Planned	(Scheduled Maintenance)	130 entries
Responsive	(Corrective Maintenance)	1395 entries

7. Maintenance Type
One of the following (meanings unknown at this stage):

CB	(all are Planned)	27 entries
CN	(all are Client Funded)	1 entry
ID	(all are Responsive)	1 entry
NC	(all are Client Funded)	521 entries
RB	(all are Responsive)	1394 entries
SM	(all are Planned)	81 entries
SN	(all are Planned, except 1 Client Funded)	23 entries

8. Quote Value (Estimated Invoice Amount)

Quotes range from $0 to $81,233. The frequency with which the quotes falls within certain dollar value ranges is shown as a histogram in Figure 25; a quote value of $0 occurred 607 times.

[36] http://www.boma.org/

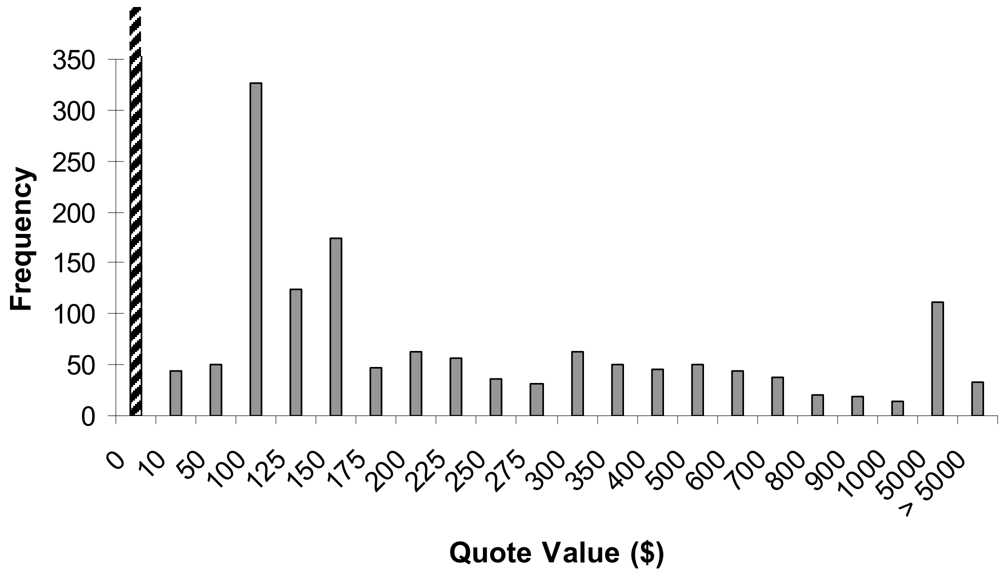


Figure 25. Frequency of various quote value ranges (in dollars)

9. Work Order Status
   One of the following:

A	(meaning unknown)	125 entries
C	Completed	1907 entries
O	Outstanding	16 entries

10. Received Date
    Ranges from 18/05/1999 to 27/02/2003 — 1986 entries

11. Authorised Date
    Ranges from 18/05/1999 to 27/02/2003 — 1979 entries

12. Required By Date
    Ranges from 20/10/1999 to 27/02/2003 — 1920 entries

13. Element — 195 entries
    Examples:
    - 8.4.02 Jim Keim advised Graham Douglas requested this
    - 8.6.01 MATERIALS $320.00 TRFD TO 01244484 BUT PUT IN SUB CON
    - 9.10.00 $8.40 Materials TRANSFERED FROM 01242150 -
    - 9/04/02 Advised by Alex Kovchenko that B767059 should not
    - AC system are worn.
    - Advised Leanne Reid that workorder should be cancelled..jad
    - air con problems on all floors & now urgent.

14. Work Order Extended Description — 2048 entries
    Examples:
    - AIR COND FAULTS ON VARIOUS LVLS
    - LVLS 1,2,3,4,5 LIGHTS OUT PER LIST CONTACT SECURITY ***URGENT*** FOR FRIDAY
    - SEE LIST OF LIGHT PROBLEMS WITH SECURITY

- SUPPLY LOCKABLE STEEL CUPBOARD FROM LOAN FURN DEPOT TO IBIS PROJECT L3B & SAFETY TO BE REMOVED TO BASEMENT STORAGE
- **URGENT** BLOCKED MALE TOILET LVL 6B SEE SECURITY x64195
- 1B SOUTH - LIBRARY AREA A/C TOO COLD meryl 73962
- 2 LEAKING WATER BOILERS BOTH ON L5B ONE NORTHERN END AND ONE CENTRAL - SEE SECURITY SEE TROY CRANG x43518
- 2A SOUTH, AIR CONDITIONING COLD & DRAFTYSee Security, Ext: 73001.

15. Defect Description — 91 entries

    Examples:
    - 11.12.01 TRFD 3HRS FROM W/O 01041712 ON (863), INCORRECTLY CHARGED TO WORK ORDER ...SS
    - 14.05.01 $1025.33 sub con tfrd from 01013707 - j000001509 incorrect work order - requested by b.boshard refers mh.
    - 14.11.01 MICHAEL RANG - PLUMBERS HAVE NOT RETURNED FIXTURE WHEN WILL IT BE BACK. RANG DARREN - HE WILL CHECK INTO IT AND RING MICHAEL C.G.
    - 15.11.01 VAN STOCK J2049 $0.90 FROM VAN STOCK W/O 01034728 ...SS
    - ASM TO THERMO MIXING VALVES
    - ASM WATER TREATMENT
    - ASM-A/C - CENTRIFUGAL CHILLER SETS CONTRACT-BC40/146 FREQ-MONTHLY/6 MONTHLY/YEARLY
    - INSTALL WHITEBOARD LVL 1 BLOCK B
    - REPAIR TABLE L4B

16. Customer Number / Customer Name (2 separate fields)

    Many different customers, including the following:

1021 / ACQUISITION & ACCOUNTING SE	423 entries
100041 / GOV OFFICE ACCOM (DPW)	2 entries
104229 / DISABILITY SERV CENTRAL OFF	76 entries
900012 / DEPT OF PUBLIC WORKS (PORTFOLIO)	1515 entries

17. Work Order Number

    Numbers fall into the following ranges:

2 - 3309	157 entries
10000 – 13000	67 entries
1000000 – 10000000	1802 entries
71001000 – 71004000	22 entries

4.2 Central Sydney Area Health Service

Data was also provided by the Engineering Division of the Central Sydney Area Health Service (CSAHS) for a hospital building.

Queen Elizabeth II Building of the RPAH

Data was received for the Queen Elizabeth II Building of the Royal Prince Alfred Hospital (RPAH), hereafter referred to simply as Building 10. It is a five-storey building, three levels of which are occupied by the Institute of Rheumatology and Orthopaedics.

4.2.1 CAD Data

CSAHS divided CAD data of Level 3-8 of Queen Elizabeth II Building (Building 10) which fall in 7 domains – Architecture and related raw drawings, BMS (Building Management and Control System) which is in charge of controlling mechanical system as swell as coordinating other building services, Electrical Engineering, Hydrant, Lifts, Mechanical Engineering. CAD data is in AutoCAD 2D. Each domain can be generally described as below:

- Architecture and detailed raw drawing: Includes overall floor plan of level 3-8 and corresponding detailed part plans. Some specific domains such as dirty utilities and medication rooms have been covered in this part.
- BMS -- The Building Management and Control System (BMS) is a Siemens Building Technologies SYSTEM 600 Apogee in order to provide direct digital control to Mechanical as well as other building services including Electrical Services, Hydraulic Services, Fire Services, Lift Services, Power Monitoring, Medical Gases.
- Electrical: includes lighting and power supply layout, switchboards layout and arrangements, communications and alarm layout, etc.
- HYD: Hot water and cold water schematic view and circulation pump control, water and gas supply, sanitary service, etc.
- Lifts: Includes landing appointment, car refurbishments, etc.
- MECH: Mechanical services for level 3-8, Steam schematic, duct cleaning, fire sprinklers and hydrants, heated water, etc.

4.2.2 Maintenance Data

CSAHS currently uses the BEIMS system (described above in Section 3.1) to store maintenance data. Data for the last two and a half years is available in SQL format and contains data that is highly detailed and structured. In particular, it contains asset numbers that match those contained in current CAD data. The analysis that follows is based on the approximately 5000 work orders recorded for Building 10 in the period from 1/1/2001 to 9/12/2002. More recent data has also been available to the project.

Data for the period from 1/1/1995 to 31/12/2000 is stored in the Paradox database format. The records are less detailed and contain references to assets that no longer exist because the building was renovated; nonetheless, the enormous dataset provides a rich resource for the data mining system to draw from. There are over 2000 work orders recorded for Building 10 in the period from 1/1/1995 to 31/12/2000, as well as tens of thousands of records for other buildings on the RPAH campus.

The detailed work order records kept since 2001 contain the following fields:

Work Order

1. Work Order Number

 Work orders are prefixed by a code as follows:

 `P` Preventative Maintenance (PM)
 `Z` Corrective Maintenance (CM)
 `R` Corrective Maintenance submitted electronically

 Corrective Maintenance (Z and R) accounts for approximately 55% of all work orders. This prefix is then followed by a 7 digit number, e.g., P0070374 or Z0043925.

2. Requested By

 For CM, the name of the person who lodged the request; for PM it simply states: `Planned Maintenance`.

3. Description

 For CM, this field contains a description of the fault to be repaired.

For PM, it refers to one of a set of pre-defined scheduled upkeep tasks.

4. Priority Code

Each work order is assigned a priority of Low (L), Medium (M), or High (H), or a number in the range of 0-9. As shown in Figure 26, over half of the work orders in the Building 10 data are assigned a priority of M, with another 20-25% assigned priorities of 5 and 6 (presumably roughly equivalent to M), with the rest being assigned a priority of L or H. Less than 1% of the work orders were assigned to numerical values other than 5 or 6.

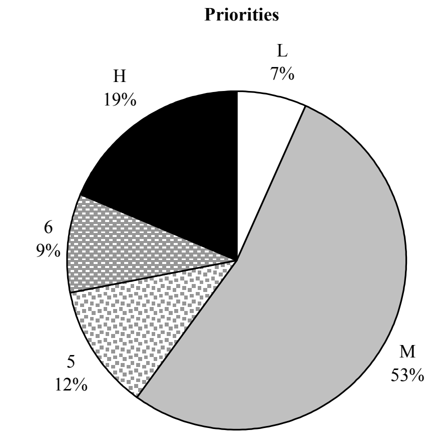


Figure 26. Distribution of Priorities

5. Work Order Status Code

The current disposition of the work order, which is one of:

C	Completed	92.5%
O	Outstanding	5.9%
X	Cancelled	1.6%

Location

6. Department Code

There are 835 distinct department codes, which are numbers of up to 5 digits such as:

12104	Biomedical Engineering
14404	Plant Operations & Maintenance
68112	Institute of Forensic Medicine

A code of 10 is used to indicate an item that pertains to Building 10 as a whole or to a non-specific area of the building.

7. Room Code

This contains codes such as:

MAINSWB	Main switch room
MPR10/08-2	Plantroom - roof west
BJG.01-01 to BJG.01-08	1 bed room - 6E
SHN.07	Disabled toilet
E3.03	PABX Room

This field is only completed in approximately 4% of the records.

8. Building Code

The code designating the building in question; in this data set it is always 10.

9. Floor Code

The floor code has one of the following values:

`GND`	Ground Floor
`L3-L9`	Floor number, starting Level 3, including upper levels for plant rooms
`L6E`, `L6W`	East or West wings of Level 6
`R1`	Roof
`ZZZ`	Miscellaneous

10. Phone

A contact telephone for the person who requested the CM; rarely completed for PM.

11. Cost Centre Code

The cost centre code indicates the department to be charged for accounting purposes; a value for it has been entered for just over 40% of the work orders. Some of the most common entries are:

`14405`	Building Services	6.1%
`25211`	Physiotherapy	4.4%
`25211`	Rachel Forster Breast Clinic	3.6%
`21271`	Q6W – Haematology	3.3%
`26341`	Q6E – Rheumatology	3.2%
`14404`	Plant Operations & Maintenance	3.1%

Associated with many cost centre codes there are also the names of a contact person and a phone number, as well as a division code.

Job Type

12. Job Type Code

One of the following:

`AD`	Administration / Development
`CM`	Corrective Maintenance
`MAC`	Major Capital Works
`MIC`	Minor Capital Works
`MW`	Minor Works
`PM`	Planned Maintenance
`SM`	Statutory Maintenance

13. Job Sub-Type Code

There are 48 different Job Sub-Type Codes, approximately half of which are actively in use, such as:

`ACU`	Air conditioner service
`BATT`	Maintenance to battery systems
`KEYS`	Key cutting
`LIFT`	Lift Service (Unplanned)
`TANK`	Water storage tank clean/test
`TMV`	T/static mixing valve service

`ZZZZ`		Miscellaneous

Unfortunately this field is filled in as ZZZZ (Miscellaneous) in almost 90% of the records.

14. Trades

There are over 100 different Trade Codes, which indicate what type of trades people are needed for completing the work order. There are 22 codes currently in use for Building 10, the most common of which are:

`CR`	Control Room	19%
`HM`	Handyman	15%
`PL`	Plumbers Workshop	15%
`EL`	Electrical Workshop	11%
`CA`	Carpenters Workshop	10%
`FT`	Fitters Workshop	10%
`RF`	Refrigeration Workshop	6%
`PC`	Pest Control	6%

Dates

Dates are entered in the format dd/mm/yyyy and if a time is also provided it is in the format hh:mm:ss { AM | PM }.

15. Requested Date/Time

The date and hour at which the CM is requested or the PM is scheduled.

16. Entry Date/Time

The date and hour at which the work order is entered into the system.

17. Date Approved

Date of work order authorisation.

18. Start Date

The date (and sometimes the hour) on which the maintenance action commences. This is the same day as the request date over 99% of the time, but occasionally can be as much as 3 weeks later.

19. Completion/Cancel Date/Time

This is the date and time at which the work was either completed or cancelled. A work order may be cancelled, for example if a long time has elapsed for a PM action and the next scheduled maintenance is already imminent, so that the old PM is redundant. The date entered for this value is within 24 hours of the request date/time 33% of the time. At the time this data was gathered, no completion date had been entered yet for roughly 6% of the work orders. The percentage of work orders completed (or cancelled) within a certain number of days is shown in Figure 27.

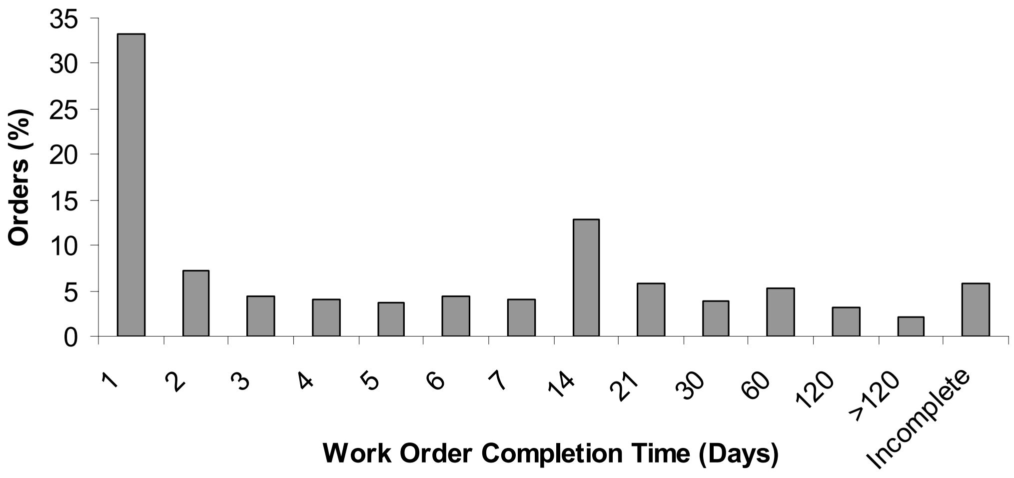


Figure 27. Percentage of work orders completed (or cancelled) by the number of days indicated. The duration is measured from the time the work order was requested.

20. Estimated Completion Date

An estimated completion date is entered for many work orders, presumably as the order is entered into the system. The time difference between the actual and estimated completion dates is shown in Figure 28. No value is calculated if either date was not entered for a work order.

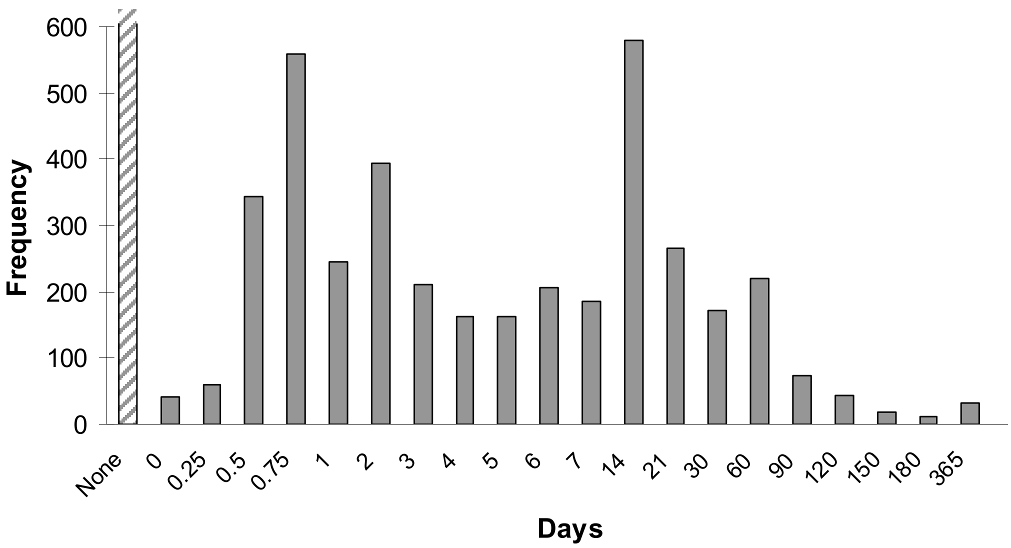


Figure 28. Difference between actual and estimated completion dates.

Asset Description

21. Asset Number

The work orders often indicate an asset number that refers to a specific building element or component. A description is provided for each asset, along with detailed location information in the form of a building code, room code, level code, and department code. Additional information is also offered when available, such as installation date, purchase value, make, model, and serial number. For each asset, there is also an indication of the asset category to which it belongs. For example, the relevant information for some assets is shown in Table 8.

Table 8. Sample Asset Information

Asset Number	COM1000-02	RPA1003-CHL-01	RPA1005-ADR-01	RPA1008-BLR-01	RPA1008-EXF-12
Description	Air Compressor - unit 2	Chiller #1 - water cooled recip.	Automatic Door - L5 front entry	Steam Boiler-1	Exhaust Fan 8-12
Asset Cat.Code	COMP	CHL	ADR	BLR	EXF
Purchase Value	0	0	5000	24500	0
Make	KAESER Dental	Carrier	AUTO ENTRY	Hunt	Fan & Blower Co of Aust.
Model	Duplex 2/350	30LA40091003	TGSC	VS 200	PCD456DD
Serial	195001	61746	TG 381362	1791	2578 95
Warranty Period	0	0	12	12	0
Depart.Code	10	10	14400	14400	14400
Building Code	10	10	10	10	10
Room Code		MPR10/03-1	SHN.10.01	PLQ-08-02	R1
Floor Code	ZZZ	L3	L5	L8	L8
Instal. Date	30/06/1995	30/03/1995	1/04/2000	30/03/2000	16/05/2002
Location	Basement	pump area	front entrance	west side pl/room	near m/up tank

An asset number is provided in approximately 50% of the work orders in our sample data set. Most often the asset is simply indicated as BLD-10, which refers to Building 10. There are over 500 asset category codes, of which 22 are used in this data, including:

BLD	Building	59.64%
TMV	Thermostatic Mixing Valve	12.06%
EDG	Emergency Diesel Generator	4.98%
ZBC	Battery Charger	4.09%
PMPS	Sump Pump	2.83%
CFA	Filter-Air	2.79%
BED	Standard Ward Bed	2.59%
AHU	Air Handling Unit	2.47%
PCT	Pest Control Trap	1.90%
KIT	Kitchen	1.66%
PFC	Power Factor Correction Gear	1.34%
SSC	Soil Stack Collection System	1.21%
PMP	Pump	1.01%

Task Details

For Preventative Maintenance work orders, a description of the planned task is provided from a list of pre-defined descriptions. A short textual description is given, as well as an extended description. Further information is referenced using the task code.

22. Task Code

A number is also provided that refers to the task list, where extensive details about the task is given in a highly structured format, including task frequency, job type code, job sub type code, and priority code. There are 795 such task codes, of which approximately 40 are in use for this building site. The most frequently occurring PM tasks, along with a short description and an indication of their relative frequency, are shown in Table 9.

Table 9. Most commonly occurring Preventative Maintenance tasks by code.

Task Code	Short Description	% of PM Tasks
DPC001	Pool check - daily	20.2%
DLS001	Log sheet - daily, BLD-10	17.7%
PCT00D	Pest control trapping	10.2%
TMV004	Thermostatic mixing valve - monthly temp. check	9.6%
BAT001	Battery service - weekly	7.7%
DBC00D	Daily Boiler Check - boilers	5.4%
PMP004	Pump inspection - 4 Monthly	3.4%
BLR0DD	Boiler check, daily - gas fired or electric [CR]	3.3%
A01	Light globe/tube replacement	3.0%
BED006	6 monthly check - ward beds	2.5%
FR001	Filter replacement - monthly	2.0%

An example of the extended details for some task codes is given in Table 10.

Table 10. Detailed description of Preventative Maintenance tasks.

Task number	Freq. type	Freq. no.	Job type	Job subtype	Priority code	Short description
BAT001	Weekly	1	PM	BATT	5	Battery service - weekly
Extended Description	1 : Clean batteries & check all connections 2 : Check water level , adjust if necessary. 3 : Check battery specific gravity. Log result in ' Work Done ' section of work order. 4 : Measure output of battery charger. Log as above. 5 : Load test batteries. Log any faults found.					
DBC00D	Daily	1	PM	ZZZZ	M	Daily Boiler Check - boilers
Extended Description	1. Check boiler, associated pipes & equipment for gas leaks. Use soapy water to confirm gas leak. 2. Check salt & chemical levels. 3. Blow down gauge glasses. 4. Sight burner flame thru observation glass at front of boiler. Clean as necessary. 5. Blow down boiler to check water level alarm operation. 6. Change boiler chart & sign off daily log sheet.					
PMP004	Monthly	4	PM	PMP	5	Pump inspection - 4 Monthly
Extended Description	Check / inspect and adjust if necessary the following. a).Pump gland or mechanical seals. If gland is of the packed type and leak is excessive adjust or re-pack gland as required. Remove one or two turns of old packing and top up with fresh packing using stagger-packing & mitre cuts. Refer to manufacturers recommendation for correct packing material. If mechanical seal requires replacement obtain pump name and serial no. and report to leading hand or supervisor. b).Check operation of pump for vibration and differential pressure across suction/discharge outlets (if gauges are fitted). If pump is air balance type, check air pressure in chamber and cut-in/cut-out frequency to assess if bladder is ruptured (NB. if bladder is faulty cut-in/cut-out will be within a few seconds) c). Check all holding down bolts and flanges for tension. Report any gasket leaks to supervisor. d).Clean out gland drain. Ensure water will flow freely to waste drain. e).Grease pump and motor as required. DO NOT OVERLOAD BEARINGS. Refer to manufacturers recommendations for correct lubricant. f).Replace all safety guards.					
TMV004	Monthly	1	SM	ZZZZ	H	Thermostatic mixing valve – monthly temp. check
Extended Description	ALL WORK TO BE CARRIED OUT IN ACCORDANCE WITH AUST. STANDARDS 1. Warm water temp. is to be taken at one fixture off each TMV, preferably a handbasin. 2. Run water at fixture for 2 minutes. 3. Take a reading of temp. at fixture, with a calibrated & tested thermometer. 4. If temp. is within acceptable range ADULT - 40.5 to 43.5 deg. C, CHILD - 38.0 to 40.5 deg C Record temp on monthly log sheet & initial. 5. If temp is 2 deg. above maximum ADULT - 43.5 or CHILD - 40.5 --- ISOLATE TMV IMMEDIATELY. Inform NUM or dept. head. 6. Inform leading hand of failed valve. Failed valve is to receive a full service (see annual service text) and a work order docket is to be raised. 7. After failed valve is serviced & reinstated - recheck temp. & record on log. 8. Include temperature readings & W/O numbers on BEIMS docket.					

4.3 Industry Data Conclusions

In order for the data mining process to succeed, it must have data of sufficient quality so that knowledge can be extracted, as well as sufficient quantity so that there can be a reasonable belief that the extracted knowledge reflects true patterns in the data, rather than mere statistical anomalies. The data we have received from QBuild for the Neville Bonner Building does not adequately meet the minimum requirements that must be satisfied, in terms of both quantity and especially quality, in spite of it being the data set that was deemed to be the best for the purposes.

It is critical that the maintenance data be available at the element level. The QBuild databases use a three layer tree coding, consisting of a WIC Number, Building Code and Building Component with Element Numbers, and is considered ideal to use in data mining. However, of the 2048 data entries for the Neville Bonner Building, only 15 entries specified the asset component.

Although it might be possible to make use of the component code or BOMA code, which are specified separately in another field (see Section 4.1.3), since these codes are present in the majority of entries, in order to distinguish between different assets, they would have to be used in conjunction with other information, such as location. However this information is not available. Without the ability to uniquely identify an asset, it is impossible to create a history of maintenance work done on that asset.

In addition, there must be corresponding CAD data available for the building, and the assets referred to in the maintenance data must also be included in the CAD data in such a way that it is possible to create a one-to-one correspondence between the same items as it appears in both types of data. While substantial CAD data for the building does exist, it is presented in a manner and is labelled in such a way that the assets are not included in the drawings nor identified uniquely. It may be possible to obtain better CAD models for later buildings.

The maintenance data acquired from the Central Sydney Health Area Services (CSAHS) meets all of the specified requirements to allow data mining. A large number of the entries in the CSAHS maintenance database contain a reference to an asset number, which refers to a specific building element or component. A detailed description is also provided for each asset, including an indication of the asset category, along with detailed location information in the form of a building code, room code, level code, and department code. Additional information is also offered when available, such as installation date, purchase value, make, model, and serial number. There is also a large amount of additional information provided in detail and in a highly structured manner and for Preventative Maintenance work orders, a description of the planned task is provided from a list of pre-defined descriptions. The maintenance data for Building 10 spans a period of almost 7 years, from January 1995 to September 2002 and contains over 7000 entries, 5000 of which were entered since January 2001, reflecting the higher level of accountability and detail that has been the practice in recent years.

The CAD data for Building 10 is also well organised and clearly labelled. In particular, the assets are identified uniquely in the drawings using the same asset codes as are used in the maintenance database. This allows an asset to be located with respect to the building, providing for further features that can be used in data mining. This also allows an end-user to access an asset's maintenance history and relevant discovered knowledge by selecting that asset in the 3D virtual environment that will be created from the CAD data.

5. APPROACH TO LIFE CYCLE MODELLING

Past experience often plays an important role in building management. “How often will this asset need repair?” or “How much time is this repair going to take?” are the types of questions that project managers face daily in their planning activities. Failure or success in developing good schedules, budgets and other project management tasks depends on the project manager's ability to obtain reliable information in order to be able to answer these types of questions. Students and young practitioners tend to rely on information that is a regional average provided by various publishing companies. This is in contrast to experienced project managers who tend to rely heavily on their personal experience.

Another aspect of building management is that many seek to improve the available scheduling algorithms, estimating spreadsheets and other project management tools. Such a “micro-scale” level of research is important in providing the required tools for the project manager's tasks. However, even with the best of such tools, low quality input information will produce inaccurate schedules and budgets as output. Thus, it is also important to have a broad approach of research in a "macro-scale" level.

The Architectural, Engineering, Construction (AEC) industry is seeing explosive growth in its capabilities to both generate and collect data. Advances in scientific data collection, the introduction of bar codes for almost all commercial products, and computerisation have generated a flood of data. Advances in data storage technology, such as faster, higher capacity, and cheaper storage devices (e.g. magnetic disks, CD-ROMS), better database management systems, and data warehousing technology, have allowed the transformation of this enormous amount of data into computerised database systems. As the AEC industry is adapting to new computer technologies in terms of hardware and software, computerised building data is becoming increasingly available. Yet, in many cases, this data may not be used, or even properly stored. Several reasons exist (Soibelman 2002):

1. Project managers do not have sufficient time to analyse computerised data,
2. Complexity of the data analysis process is beyond the capabilities of the relatively simple building maintenance systems commonly used,
3. No well defined or automated mechanisms to analyse data and interpret results so that site managers can use it.

However, there is a great deal of valuable knowledge that can be obtained from an appropriate use of this data; there is a need to analyse this increasing amount of available data and Data Mining can be applied as a powerful tool to extract relevant and useful information.

5.1 System Composites

The framework of this CRC project is established based on three modules: (i) a 3D Object-Oriented Database module, (ii) an agent-based virtual environment module, and (iii) a data-mining module. The general description of this framework is illustrated in Figure 29.

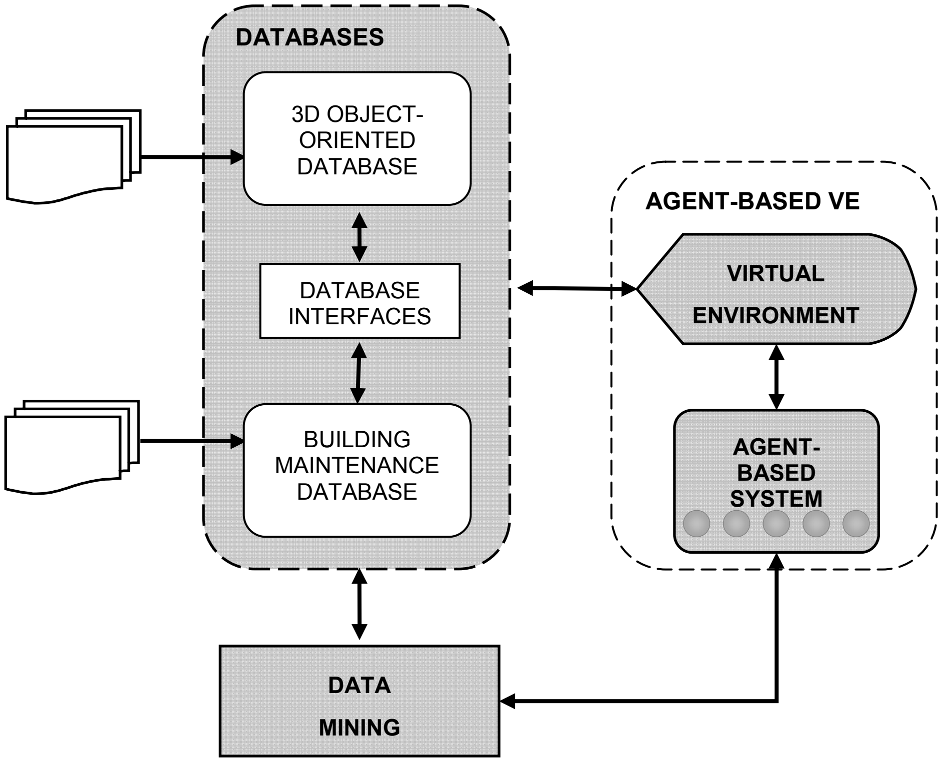


Figure 29. Proposed framework of life cycle modelling in virtual environments

Data Mining (also known as Knowledge Discovery in Databases, or KDD) has been defined as the "nontrivial extraction of implicit, previously unknown, and potentially useful information from data" (Frawley 1992). Data mining uses machine learning, as well as statistical and visualization techniques to discover and present knowledge in a form that is easily comprehensible to humans.[37]

Mining data enables us to understand how systems that were once thought to be completely chaotic actually have predictable patterns (Peitgen 1992). Through KDD, patterns and causal relationships behind apparently random data in AEC projects can be found. By applying KDD to identify novel patterns, project managers will be able to build knowledge models that may be used for recurrent activities of on-going construction projects, as well as for a future project activities, and avoid unanticipated consequences (Soibelman 2002). Data Mining also presents the potential to address the problem of transforming knowledge implicit in data into explicit knowledge for decision making.

In contrast to traditional methods of statistical data analysis, KDD is an automated process that discovers new trends and patterns without the need for human intervention. KDD takes input variables whose relevance may not be obvious to a designer but which becomes evident as a result of this process. In addition, KDD makes no prior assumption about the probability distribution of the input variables (Gaussian, Poisson, etc.), as is required in statistics, and is therefore more robust and general. However, like other methods, the process of transforming the data into a format suitable for knowledge discovery is not automated and has a large impact on the results obtained.

Thus, the approach is based on a comprehensive view of the building management problem. It views the process of building design, maintenance, and replacement as a process generating an enormous

[37] http://www.andypryke.com/university/dm_docs/dm_intro.html

amount of information. While current practice addresses parts of this information generation and management, our approach attempts to account for the life cycle flow of this information.

The costs of designing and building structures are much smaller than the costs of operating a building or other structure over the course of its life span. The knowledge that becomes available through data mining enables a building owner to make important decisions about life cycle costs in advance, thereby significantly affecting and improving design decisions.

The rich set of building data that is created during the design and documentation phase of the building remains relevant even after the building is constructed, and the data only becomes richer as maintenance data is added. Architects, interiors designers and engineers, as well as contractors, marketing and sales personnel, building managers and owners can extract information from the databases for the building's renovation, maintenance, and operation.

Figure 30 outlines our model of the flow of information in building design and maintenance. The bold arrows depict the functionality provided by our proposed approach while the dashed arrows describe the scope of present approaches to building information management.

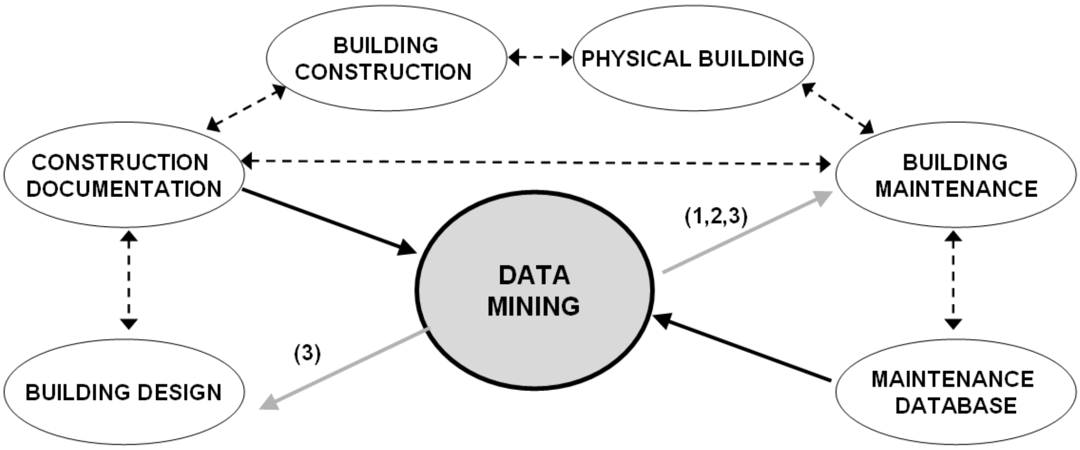


Figure 30. Integrating data mining within the life cycle of building information management.

Data Mining techniques can be used effectively on data stored in a Building Maintenance System (BMS) by creating knowledge that can be used in future management and design decision making. Knowledge that implicitly resides in BMS databases includes information about:

1. Components that frequently need maintenance and therefore need to be inspected carefully,
2. Historical consequences of maintenance decisions that may inform future decisions,
3. Components of buildings that significantly determine maintenance cost and therefore may inform future building designs, as well as refurbishment of the building in question

This information can be extracted using Data Mining techniques and used to improve all phases of the building life cycle, both for current and future buildings, as indicated by the numbers (1, 2, 3) and (3) marked on the arrows in Figure 30, which correspond to those listed above.

5.1.1 3D Object-Oriented Database and Building Representation

3D building models are useful across the entire spectrum of architecture, engineering and construction (AEC) practices. Architects and their clients use 3D building models to observe and evaluate building designs before construction, while there is still a chance to make substantial changes at a reasonable cost. Engineers use 3D building models for energy, lighting, acoustics and fire simulations. The results of these simulations give valuable insight into building useability and safety. Professionals in the construction industry utilise 3D models to estimate costs and to plan cost-effective construction sequences. This process often leads to the early discovery of design conflicts that would otherwise result in expensive construction mistakes. Even for an existing building, it is often desirable to have a 3D model to facilitate analysis of the energy properties of the building or to predict how a potential fire might spread or to study potential changes to the building, or to identify possible uses of existing building spaces (Lewis and Sequin, 1998).

The 3D Object-Oriented CAD module consists of 3D CAD files and an object-oriented database. The object-oriented database called Express Data Manager (EDM)[38] is used for managing CAD objects. EDM is a STEP database supporting data exchange using ISO 10303-21 (STEP Physical File), 10303-22 SDAI (Standard Data Access Interface) and long-term archiving. EDM was adopted based on the considerations of construction industry requirements for interoperability and the complexity of data in AEC domain. EDM provides the following advantages and benefits:

- A complete database management system according to the STEP standard from which the IAI/IFC standard of AEC Industry evolves
- Enables product data to be effectively managed, exchanged and shared across radically different systems, independent of location, type or network design.
- Allows access to this data throughout the life cycle of the product and ensures that the information is in a form that can be accessed and interpreted for decades to come.
- Provides a generic software integration that aimed to improve functionality and performance of data model implementation.

5.1.2 Agent-Based Virtual Environments

The application of Virtual Environments in the building industry can contribute to improving the communications between different project partners around an intuitive, 3D representation of the planned process. Apart from visualization and communicational capabilities, this CRC project is also concerned with life cycle control within virtual environments. Unfortunately, most artificial intelligence explorations on VE were carried out in unrelated AEC areas. The integration of life cycle modelling with VE has the potential to be viable with the inclusion of intelligent agents and data mining. Through the agent-based virtual world each object represents an agent with attributes and behaviours.

5.1.3 Data Mining Techniques

Several research projects have explored the use of Data Mining in design (Reich et al. 1993), as well as civil engineering and other relevant disciplines for:

- Evaluating structures during preliminary design (Arciszewski et al. 1987)
- Learning from failures of structures (Stone and Blockley 1989)
- Learning environmental evaluations (Julien et al 1992)
- Building a cable stayed bridge design assistant (Reich and Fenves 1995)

The data mining module incorporates WEKA because of its advanced scalability and functionality. WEKA[39] is a collection of machine learning algorithms for solving real-world data mining problems. The algorithms can either be applied directly to a dataset or called from your own Java code. WEKA not only contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization, it is also suitable for developing new machine learning schemes.

[38] http://www.epmtech.jotne.com/products/index.html
[39] http://www.cs.waikato.ac.nz/~ml/weka/

5.2 3D Object-Oriented CAD Database & Building Representation Module

EXPRESS Data Manager 4.5 is a powerful object-oriented database management system that provides greater functionality for building application and data modelling. It also serves as a multi-user rule engine to define one's own constraints to improve knowledge management. EXPRESS is the unified modelling language adopted by EDM database in which heterogeneous data mapping is defined by ExpressX through *EDMmodelConverter*. With the interactive working environment of *EDMvisualExpress*, a user can define and document object models that are capable of integrating domain specific rules for knowledge-based explorations. The additional feature of EDM includes the use of EDM query schema to define and query domain specific views of objects.

The EDM can be accessed via the EDMI/SDAI API (Application Programming Interface) that was implemented in C and C++ binding. The Java-binding has been developed consequently in supporting Java program or applet for accessing an EDM database.

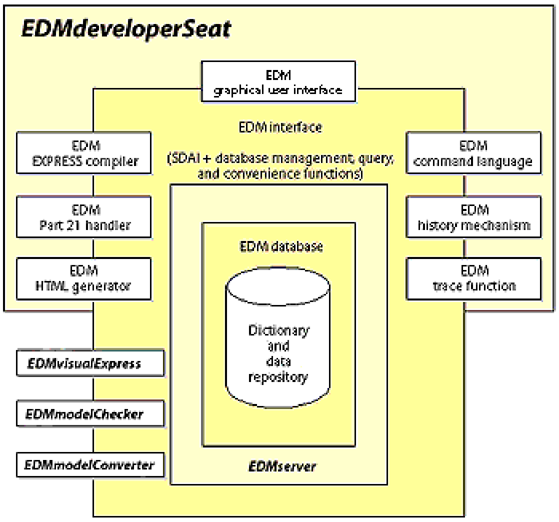


Figure 31. A Comprehensive view of EDM products and functions (EDM Website)

Using IFC-compliant design software such as ArchiCAD or AutoCAD, different kinds of CAD data ranging from three-dimensional geometric data to space identity to building material information can be stored in the EDM database for further processing.

However, like other methods, the process of transforming the data into a format suitable for knowledge discovery is not automated. Pre-processing is shown schematically in Figure 32.

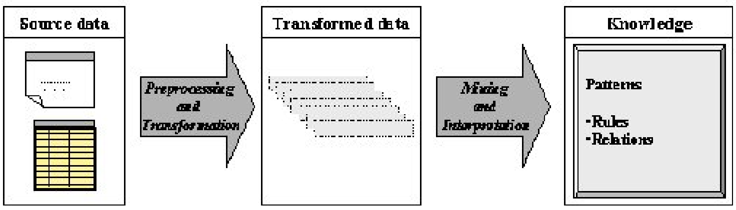


Figure 32. Pre-processing and transformation of data into a suitable format for data mining

5.2.1 Object-Oriented 3D Building Representation

The representation of building objects is not limited to the graphical information that illustrates only the structure of these objects. There are various other types of non-graphical information that are not less important than the structural representation of these objects. Non-graphical information includes functional, behavioural and semantic properties. Each building object is created to perform certain functions. These functions should produce the required set of behaviours. An integrated view of both graphical and non-graphical information can be looked at as functional, behavioural and structural scheme of object representation. Structure is what the object is, function is what it does, and behaviour is how it does it. Furthermore, the purpose of creating this object or building element is “what for”. The structure of an object exhibits behaviour; behaviour affects function; and function enables purpose (Rosenman and Gero, 1998). For instance, the structure of a hydraulic elevator includes piston and oil column, its functions are to contain loads and move them vertically, its behaviours are to push loads up and hold them by compression and its purpose involves transferring people and goods from one storey to the other. This is a simplified description of the non-graphical information related to an object or a building element The non-graphical information is not limited only to the description and properties of the object but can also include the relationships of this object to other objects in the building.

Being able to exchange non-graphical information in building projects using the computer is quite useful. A general approach has been developed by employing computer techniques that were first applied in the field of artificial intelligence (AI). In this approach objects have attributes, one of which is their geometry, which can be viewed by CAD systems, databases that contain non-geometric attribute descriptions of objects and data-structures that follow agreed standards and format. Integrated software systems would assist in describing the geometry of architectural form, the geometry of the structural frame, the structural calculations of the frame, and the work sequence and time duration of construction. Using the AI approach, when the architectural form (graphical information) is changed the other three systems (graphical and non-graphical) automatically change to suit the new form. The importance of this approach lies in the capability of computer systems to recognise the graphically represented objects as real objects with attributes rather than just geometrical forms. Object-oriented databases incorporated within CAD systems allow such a facility to be utilised. Object-oriented technologies also facilitate collaborative design.

CAD systems are now becoming increasingly object- based and web-based systems are moving in a similar direction. Recent trends in CAD systems development (Szalapaj, 2001) show attempts to integrate GIS (geographic information systems) with facilities management. This type of development involves making connections between graphical and non-graphical information, and object-oriented environments for supporting such integration are increasingly being used. The integration of graphical and non-graphical information is of paramount importance in achieving the best results in planning and co-ordinating building systems.

Object-oriented computer-aided design is an important new development in the architecture and engineering professions. Traditional CAD systems were developed to mimic the processes of hand drafting and overlay graphics. A traditional CAD drawing has little more intelligence than a hand drafted paper document. Object-oriented computer-aided design represents the next generation of CAD applications predicated on the concepts of object-oriented design that has been used successfully in the software industry to build much larger and more complex applications than were ever possible using older design methods. It is only recently that object-oriented design has been applied as a way to conceptualise and communicate design solutions.

5.2.2 3D Object-Oriented Representation for Interactive Virtual Environments

The research team has coordinated with Woods Bagot (an Industry Partner in this project), to model the selected building (Building no. 10, Royal Prince Alfred Hospital, Sydney) into object-oriented 3D CAD. Woods Bagot has modelled Building 10 using ArchiCAD, an object oriented CAD modelling system. One floor only has been modelled as a prototype for this project. Captions of the 3D model are shown in Figures 33, 34 and 35.

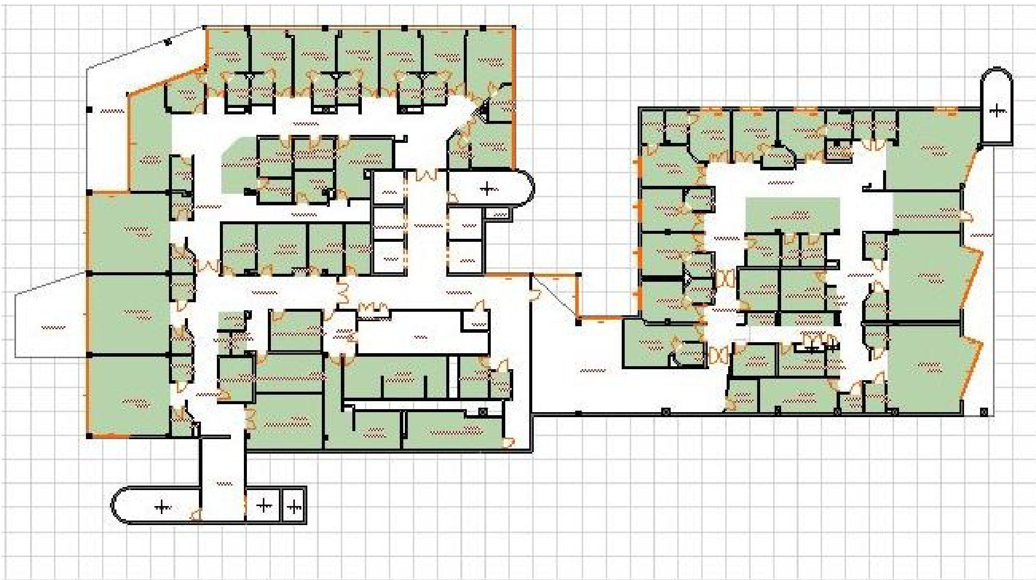


Figure 33. Top view of typical floor of Building no. 10, Royal Prince Alfred Hospital, Sydney.

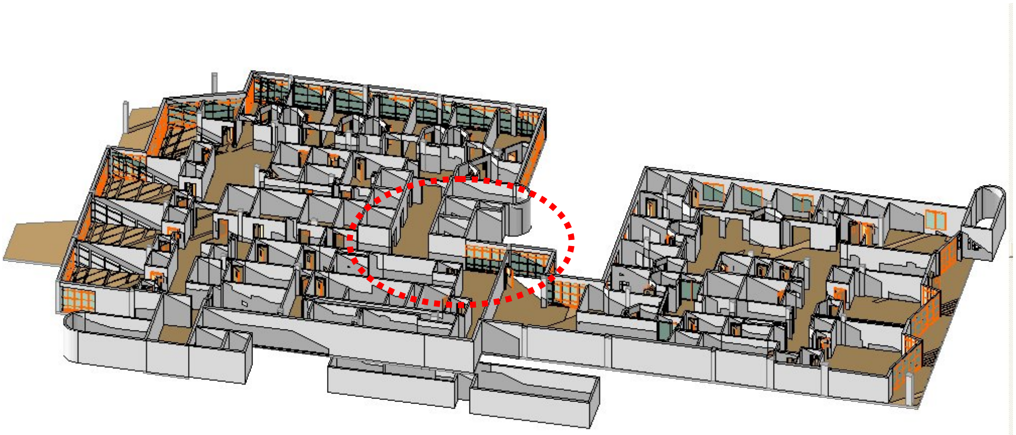


Figure 34. 3D view of typical floor of Building no. 10, Royal Prince Alfred Hospital, Sydney.

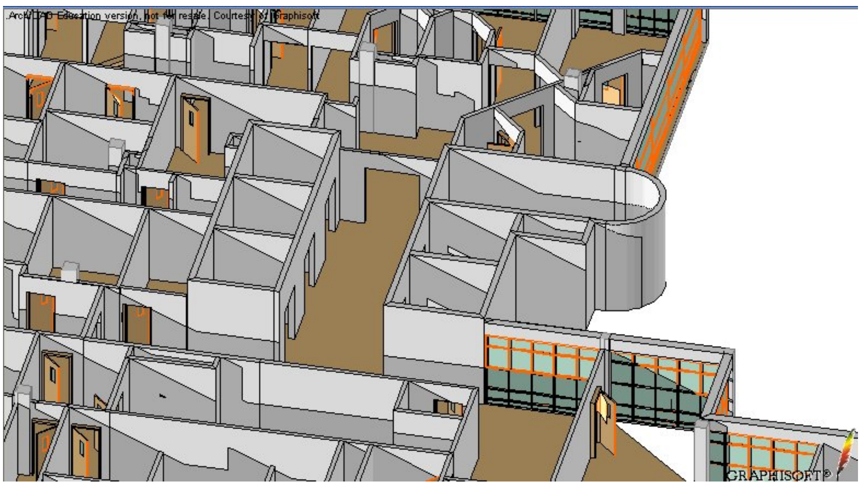


Figure 35. 3D view of vertical transportation area (Lifts and stairs) at Building no. 10, Royal Prince Alfred Hospital, Sydney.

For this 3D CAD model to useful for interactivity in a virtual environment (Active Worlds), it must be converted to RWX format to enable object screening and interaction. All objects in the 3D model must also be exported as Industry Foundation Classes using the IFC2x schema. This exchange enables EDM to handle both graphical and non-graphical attributes of building objects. The elements exported from this 3D CAD model into the IFC2x format are shown in Table 11.

Table 11. Elements exported from the 3D CAD model of Building No. 10, RPA into the IFC2x format.

Elements	Quantity
IFC2DCOMPOSITECURVE	72
IFCAPPLICATION	1
IFCARBITRARYCLOSEDPROFILEDEF	1107
IFCAXIS2PLACEMENT2D	339
IFCAXIS2PLACEMENT3D	3310
IFCBOUNDINGBOX	1165
IFCBUILDING	1
IFCBUILDINGSTOREY	6
IFCCARTESIANPOINT	12761
IFCCIRCLE	21
IFCCOLUMN	104
IFCCOMPLEXPROPERTY	1425
IFCCOMPOSITECURVESEGMENT	106
IFCCONNECTIONCURVEGEOMETRY	866
IFCCONNECTIONSURFACEGEOMETRY	63
IFCCONVERSIONBASEDUNIT	1
IFCCURVEBOUNDEDPLANE	63
IFCDIMENSIONALEXPONENTS	1
IFCDIRECTION	1215
IFCDOOR	198
IFCDOORSTYLE	198
IFCEXTRUDEDAREASOLID	1425
IFCGEOMETRICREPRESENTATIONCONTEXT	2
IFCLINE	28
IFCLOCALPLACEMENT	1822
IFCMATERIAL	16
IFCMATERIALLAYER	2650
IFCMATERIALLAYERSET	932
IFCMATERIALLAYERSETUSAGE	932
IFCMEASUREWITHUNIT	1
IFCOPENINGELEMENT	260
IFCORGANIZATION	3
IFCOWNERHISTORY	650
IFCPERSON	2
IFCPERSONANDORGANIZATION	2
IFCPLANE	63
IFCPOLYLINE	3497
IFCPRODUCTDEFINITIONSHAPE	1425
IFCPROJECT	1
IFCPROPERTYSET	1425
IFCPROPERTYSINGLEVALUE	18680
IFCRECTANGLEPROFILEDEF	318
IFCRELAGGREGATES	4

IFCRELASSOCIATESMATERIAL	1036
IFCRELCONNECTSPATHELEMENTS	866
IFCRELCONTAINEDINSPATIALSTRUCTURE	2
IFCRELDEFINESBYPROPERTIES	1425
IFCRELDEFINESBYTYPE	260
IFCRELFILLSELEMENT	260
IFCRELSPACEBOUNDARY	63
IFCRELVOIDSELEMENT	260
IFCSHAPEREPRESENTATION	3520
IFCSITE	1
IFCSIUNIT	9
IFCSLAB	2
IFCSPACE	129
IFCTRIMMEDCURVE	49
IFCUNITASSIGNMENT	1
IFCVECTOR	28
IFCWALLSTANDARDCASE	930
IFCWINDOW	62
IFCWINDOWSTYLE	62

5.3 Agent-Based Virtual Environments Module

The virtual environment of an agent-based system serves as an interactive interface for linking users to the underlying 3D and data-mining module. The agent model in the virtual environment is represented as objects that have agency and are capable of sensing, reasoning and affecting the environment. In an agent-based virtual world, a behaviour may be triggered by any change about the world. For example, triggers can occur either when a user 'clicks on' a 3D building element or when their avatar enters a room. In doing so, specific actions are performed according to agent's rules. An agent-based virtual environment is thus comprised of objects that have a 3D model and an agent model which encapsulates the following five operations: 'sensation', 'perception', 'conception', 'hypothesizer' and 'action' as shown in Figure 36 (Maher and Gero, 2002). The linking of the agent-based virtual environment with CAD modelling is as shown in Figure 37.

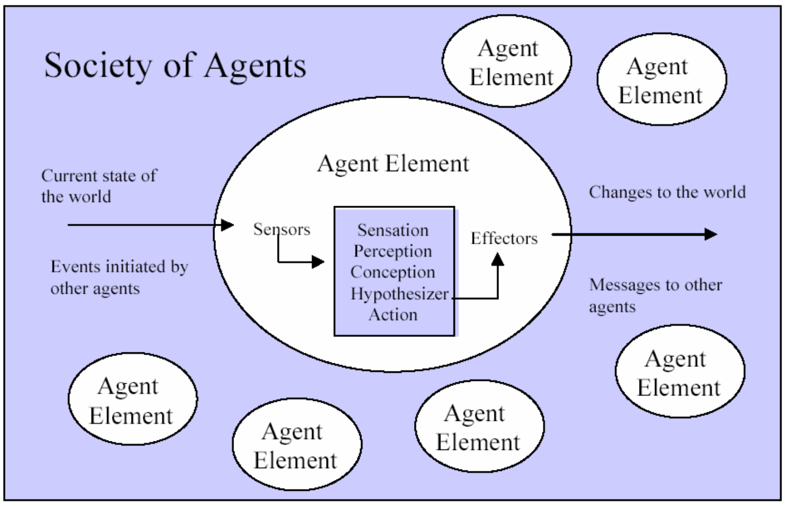


Figure 36. A Virtual Environment as a Society of Agents (Adopted from Maher and Gero 2002)

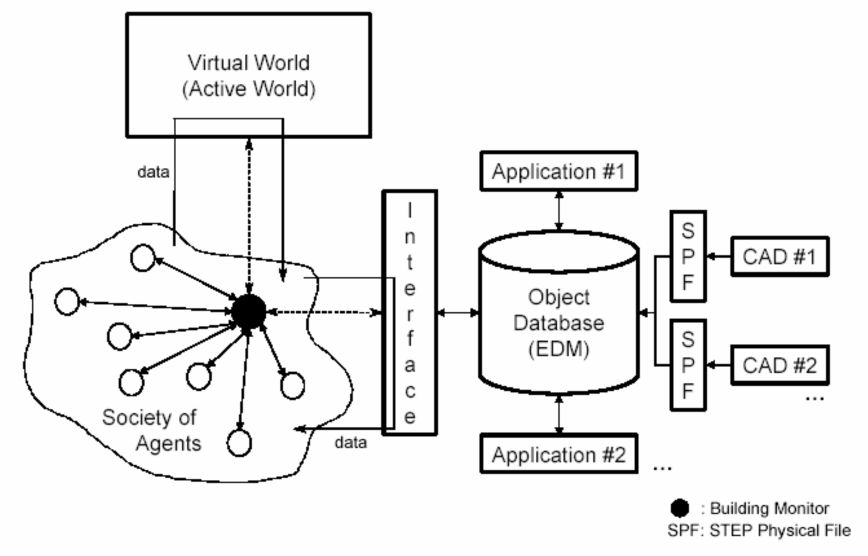


Figure 37. Linking of Agent-based Virtual World with CAD Model (Ding et al 2003)

For each object in the virtual environment, besides the 3D model itself, an agent with a sensor and an effector has been implemented. Sensors are able to sense the agent's environment and can be described as a monitoring state. Sensors 'listen' to actions and effectors perform rule-based functions. Sensors and effectors have been implemented using Java classes and linked to Active Worlds SDK via the Java Native Interface (JNI) which allows Java code that runs within a Java Virtual Machine (VM) to be portable across applications and libraries written in other programming languages. An agent model that provides a basis for setting up an agent-based virtual environment can be later developed by extending the capability of the ReteAgent (Smith and Maher, 2003). A ReteAgent is an implementation of an agent that uses Jess for everything except sensors and effectors. (A Society can contain one or more agents that are instances of ReteAgents).The creator of a ReteAgent configures the agent by specifying a set of sensors. Figure 38 illustrates the ReteAgent architecture.

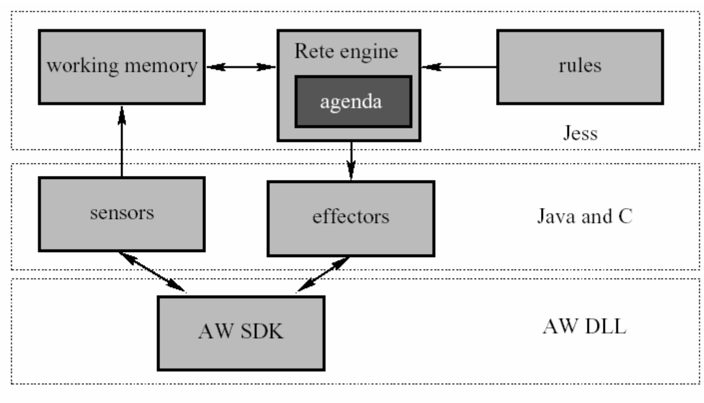


Figure 38. Architecture of ReteAgent (Smith and Maher 2003)

5.4 Data Mining Module

For this project, two types of building descriptions exist that can be exploited for data mining purposes; the complete set of features that can be derived from the object-oriented 3D CAD model of a building, and the aggregated features derivable from the building maintenance system. Based on an evaluation of the building maintenance and CAD data that has been made available to us from QBUILD and

CSAHS (described in detail in Section 4.2), we make certain observations regarding the nature of the data that will be used for data mining:

- Data consists of both numerical and nominal values
- Information will be incomplete (fields may be empty)
- Information will be noisy (there may be errors or contradictory data)
- Some fields will consist of unstructured English text
- Some data will be in the form of a 3D object-oriented model of a building
- There will be large data sets available

Based on the above properties, a variety of approaches to data mining were evaluated including: Decision Trees, Clustering and Classification, Association Rules, Artificial Neural Networks, and Hidden Markov Models. Each of these approaches has certain advantages and disadvantages, and can best be used in a complementary fashion. For example, rule-based representations can be used for further reasoning, explanation of the discoveries, formulating goals and reasoning chains. Some other representation schemata may have computational advantages but can be difficult for humans to understand. A trained neural network, for example, is capable of computing output values from the input values that it is supplied with, but it can not provide any meaningful explanation of how it is computing those values. Each of these algorithms has been described in report CRC2001_001_1 and a summary of their review can be found in Appendix E.

5.5 Stages of the Knowledge Discovery Process

The process of Knowledge Discovery from Databases occurs in several stages, of which Data Mining is one intermediate phase. The stages, illustrated in Figure 39, are as follows:

1. Select the data to be used for mining purposes
2. Pre-process the data to remove or correct errors, select feature subsets, create additional fields from existing data, etc.
3. Transform raw data into an appropriate format
4. Mine the data for emergent patterns
5. Evaluate the data to decide which patterns are relevant and useful

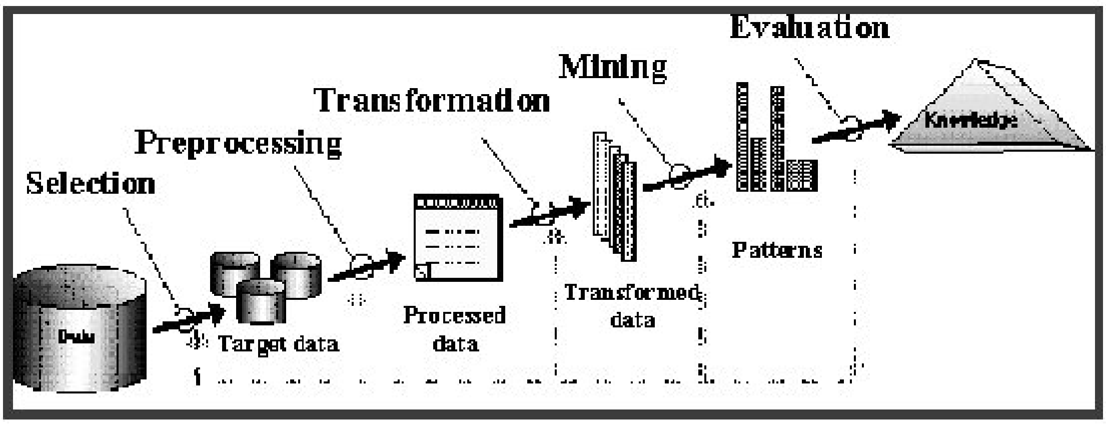


Figure 39. Stages of Knowledge Discovery Process

Data mining requires many significant steps from problem specification to the implementation of tools, and monitoring of the model (Feelders et al. 1999). Successful data mining requires several collaborative expertises such as; subject area expertise, data expertise, and data analysis expertise.

5.5.1 Data Mining Process

Data mining the CSAHS maintenance database is an explorative process since new knowledge is discovered and new hypotheses can be formed. The data mining process for extracting hidden knowledge from large databases can be depicted as shown in Figure 40. The process focuses on finding interesting patterns that can be interpreted as useful knowledge and the five stages outlined above have been elaborated by (Hui and Jha 2000) and consist of the following seven steps:

- Establishing the mining goals. This involves the understanding of building maintenance process and its acquired database.
- Selection of data. This step identifies a subset of variables or data samples, on which mining can be performed. There are many tables in the database in which not all are suitable for mining since they are not sufficiently rich.
- Data pre-processing. This step aims to remove the noisy, erroneous and incomplete data. The presence of too many different categories of data makes visualisation of the displayed information very difficult. Hence, those categories with only a few records are eliminated. Moreover, all the records with missing values are deleted to avoid potential problems in visualisation. Since the proportion of such records is quite small, their deletion will have little effect on the results.
- Data transformation. The data stored in the various tables are required to be in a specified format. Sometimes, it is useful to transform the data into a new format in order to mine additional information.
- Data warehousing. Data warehousing is the process of visioning, planning, building, using, managing, maintaining and enhancing databases. The data suitable for mining are collected from various tables of customer service database and stored in WEKA's data warehouse. WEKA is a collection of machine learning algorithms for solving real-world data mining problems. The algorithms can either be applied directly to a dataset or called from your own Java code. WEKA does not only contain tools for data pre-processing, classification, regression, clustering, association rules, and visualisation, but is also suitable for developing new machine learning schemes.
- Data mining. WEKA is used to perform the data mining functions, including summarisation, association, classification, prediction and clustering.
- Evaluating the mining results. Different data mining functions have been implemented and the information obtained analysed.

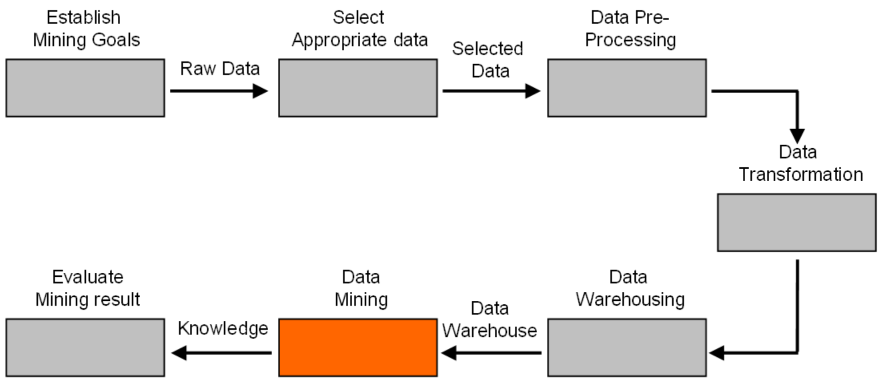


Figure 40. Stages of Data Mining Process (Hui and Jha 2000).

Before Data Mining can take place, it is essential to perform some form of data preparation and pre-processing, both to convert it into a format amenable to Data Mining, and to provide a feature set that is rich enough so that knowledge may be gleaned from it, but not so abundant that the nuggets get lost among a torrent of meaningless information.

To date, most modern KDD tools have focused almost exclusively on building models (Brachman et al. 1997). However, data preparation is a very important process because data itself may have been collected in an ad hoc manner, unfilled fields in records may be found, or mistakes in data entry may have been made (Reich et al. 1994). As a result, the KDD process cannot succeed without a serious effort to prepare the data. Once the quality and details are assessed, serious work is usually needed to get the data in shape for analysis. Typically, 60% of the time in the total KDD process goes into preparing the data for mining, while the actual mining step typically constitutes only about 10% of the overall effort (Cabena et al. 1998). Thus, it can be seen that data preparation is one of the most important parts of the entire process and one of the most time consuming and difficult. In design knowledge discovery tasks this is an extremely important stage, due to the variety of the data formats (usually non-table), and the necessity to deploy a wide variety of data mining techniques to these data. Depending on the quality of the original data, proper data preparation can enable much better models to be produced in much less time (Soibelman 2002).

In addition, representation can be viewed from two perspectives: as data, where the variables are to be considered as individual entities, or as a data set, where the variables are considered together with the interactions and interrelationships between them. Preparation for mining involves looking at the variables individually as well as looking at the data set as a whole. The following section describes the types of actions that need to be taken to prepare for data representation and some of the problems addressed both within the data and data sets.

5.5.2 Preparing Data Representation

Unsupervised pre-processing includes techniques like feature selection, feature construction, discretisation, and clustering using some metric. Supervised pre-processing involves either model based techniques (explicitly using domain knowledge), or providing the classes upon which discretisation and so on work. Further, decisions like whether to sample a database or not, and how to structure the data could be influenced by knowledge of the application domain.

- **Removing Variables**

The number of distinct values and the frequency count of each distinct value are the basic information in KDD applications. From this information, it is determined if a variable is empty. If so, the variable may be discarded. Removing variables becomes more problematic when most of the instance values are empty, but occasionally a value is recorded. The changing value does indeed present some information, but if there are not many actual values, the information density of the variable is low. This circumstance is described as sparsity (Pyle 1999). In general, mining tools deal very poorly with highly sparse data. Also, practical data mining algorithms are known to degrade in performance when faced with many features that are not necessary for predicting the desired output.

A feature subset selection algorithm conducts a search for a good subset using the induction algorithm as part of the evaluation function. The accuracy of the induced classifiers is estimated using accuracy estimation techniques. The wrapper approach (Kohavi 1994) is well known in the machine learning community because of its accurate evaluation and was used in this application. Decision trees select a good subset of features by finding the best single-feature test to conduct at the root node of the tree and continuing recursively, stopping when the features no longer give clear results.

- **Outliers**

An outlier is a single or very low frequency occurrence of the value of a variable that is far away from the bulk of the values of the variable. As a general rule of thumb, if it can be established that it is a mistake, it can be rectified. The problem is what to do if it cannot be pinpointed as an error. First it must be determined that outliers are not due to error. For example, a few of the work orders in the CSAHS data set show the completion time as being 366 days later than the start time, but closer inspection reveals that it was simply a matter of the year being entered incorrectly and the work order was completed in 1 day.

- **Handling Nonnumeric Variables**

Since all tools can handle numerical data but some tools cannot handle text data (e.g., neural networks), it may be necessary to transform text or code labels into appropriate numerical values. What must be avoided at all costs is an arbitrary assignment of numbers to text labels. The initial stage in numerating labels is to replace them with an enumeration that has some rationale, if possible. It may be necessary to identify category labels by creating one variable for each possible label; each variable can then take on the value of 0 or 1, and, for a given item, no more than one such variable can have a value of 1 to indicate that the item belongs to that category. However, this can have the unfortunate consequence of increasing the dimensionality of the problem.

- **Handling Numeric Variables**

It may also be desirable to reduce the dimensionality of a feature space by discretisation of continuous numeric variables. This can be achieved by looking at a histogram of the values and dividing them up into bins. Methods such as decision trees can also perform the discretisation as part of the process of choosing the next decision step, using the minimisation of entropy as a guide.

5.5.3 WEKA Algorithms

The search for knowledge in data is not new and with the booming of information technology at the end of last century, the increase of information availability has become the initiatives for development of a variety of data analysis techniques. Morbitzer et al (2003) indicated that emphasis has been put on visual analysis, regression analysis, uncertainty analysis and most recently, data mining techniques. Visual analysis and data mining techniques applied during analysis in the database module include:

- Data analysis through stacked histograms which are provided by pre-process panel of WEKA knowledge explorer. Correlations of various attributes can be visualized in a popup panel;
- Classification through C4.5;
- Clustering using SimpleKmeans;
- Attribute evaluator for finding groups of correlated attributes using the associate rule algorithm "Apriori".

5.5.4 Mining using WEKA clustering algorithm

Clustering techniques are applied when there is no class to be predicted but rather when the instances are to be divided into natural groups (Witten and Frank, 2000). Based on a number of trials with all available clustering algorithm provided by WEKA, it was found that the classic SimpleKmeans which employs the k-means algorithm generates meaningful clusters.

K-means algorithm forms clusters in numeric domains and partitions instances into disjoint clusters. As an iterative distance-based clustering, k-means is very simple. It is performed by first specifying in advance the *k* clusters that are being sought. The *k* points are chosen at random as cluster centres. According to the ordinary Euclidean distance function of instance to the centres, assign each instance to different clusters. Subsequent steps modify the partition to reduce the sum of the distances for each case from the mean (centroid) of the cluster to which the case belongs. The modification consists of allocating each case to the nearest of the k means of the previous partition. This leads to a new partition for which the sum of distances is strictly smaller than before. The improvement step is repeated until the improvement is very small[40]. This clustering method is simple and effective. But for relative big data sets, k-means becomes time-consuming because of the numbers of iterations involved.

[40] http://www.resample.com/xlminer/help/kMClst/KMClust_intro.htm

5.5.5 Mining using WEKA classification algorithm

Decision Trees are a tree-based knowledge representation methodology used to present classification rules. The leaf nodes present the class labels. In this project, various classification algorithms which are offered by WEKA have been applied, however it was found that several algorithms were not able to deal with the data sets derived from the available maintenance data due to limitation in processing certain data types. For instance, some algorithms were not able to accommodate numeric values while others failed to accommodate nominal variables.

After extensive experimentation with a number of classification algorithms, the C4.5 algorithm was selected. The C4.5 algorithm was proposed by Quinlan (1993) and built on top of the ID3 algorithm. The C4.5 algorithm was chosen because of its ability to deal with numeric and nominal variables, undertake pruning and handle missing values. The latter can be done by replacing a whole sub-tree by a leaf node if the expected error rate in the sub-tree for a rule obtained is greater than in the single leaf.

The C4.5 algorithm generates a classification-decision tree for the given data-set by the recursive partitioning of data. Decision trees are constructed based on a "Divide and Conquer" approach in which a search is undertaken to identify the attribute with best information gain at the root node for splitting the tree into sub-trees or branches that can be further and recursively partitioned following the same rules. The splitting stops when there is no longer an information gain or it reaches the leaf node. This process is sometimes called top-down induction of decision trees. Once the tree is constructed, rules can be generated by traversing each branch of the tree and collecting the conditions at each branch of the decision tree.

Although the C4.5 technique is a powerful classifier and is robust to noise, its performance depends on the data sets on which it runs. For example, in this demonstration, when the C4.5 is applied to air handling units in which some attributes with unchanged values, such as "job_subtype" with 105 "zzzz" out of 107 case data sets, "workorder_status" with 105 "C" out of 107 case data sets, the performance is far from expectations. Only several structures of data are located, such as "department 26462 resides only at 1st and 7th floor" and "department 21271 only reside at 6th floor", etc.

5.5.6 Mining using WEKA associative rule algorithm on nominal attributes

Association rule mining involves finding frequent patterns, associations, correlations, or causal structures among sets of items or objects in transaction databases, relational databases, and other information repositories (Han, 2001). The associative rule algorithm adopted in WEKA is "Apriori" which is developed by IBM's Quest project team. Apriori finds all associations that satisfy criteria for minimum support and minimum confidence. Support (also called coverage) is the number of instances predicted correctly. Confidence (also called accuracy) is the same number expressed as a proportion of the number of instances that the rules applied to (Witten and Frank, 2000). Rules with high support are of interest and some rules are pruned out due to their low coverage.

The basic idea of Apriori is to generate item sets which are combinations of attribute-value pairs with the minimum coverage. During the first stage, each search involves passing through the data set to count items. Surviving item sets are stored in a hash table. The second stage takes each item set and generates rules from it. Checks for accuracy of each rule are then made. A simplified description of Apriori is illustrated in Figure 41.

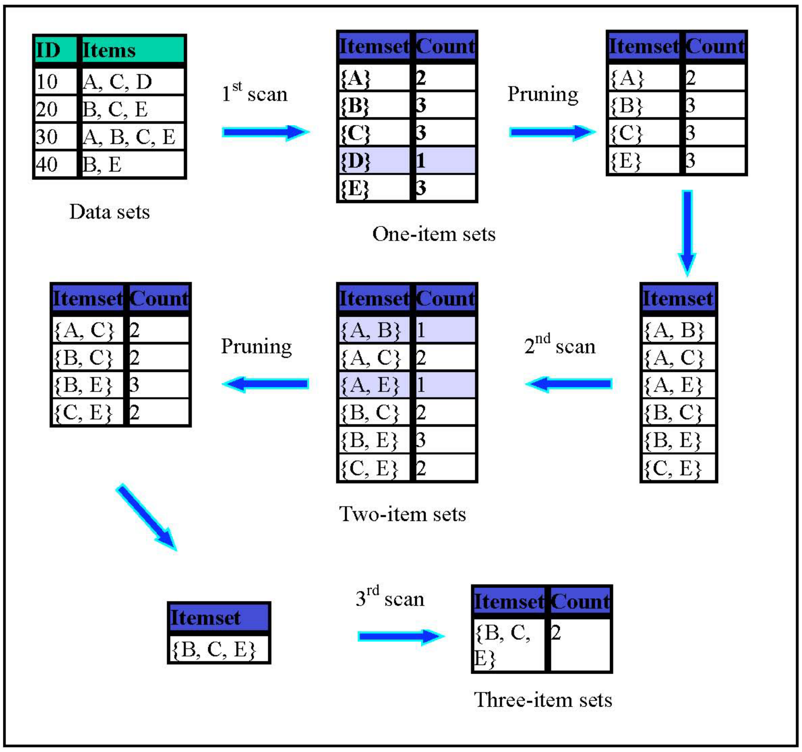


Figure 41. An illustration of Apriori in generating item sets.

Apriori is efficient in searching for associations and correlations between attributes. However shortcomings in this approach arises when large volumes of rules are created and/ or dominant rules not interesting. For example, one of the strong rules obtained during experimentation was:

WorkOrderStatus=C 105 ==> workordersource=WOS 104 conf:(0.99)

The rule does not carry any real meaning since the static attribute “workorderstatus”, and “workordersource” does not change (workorderstatus with 105 “C” out of 107). Hence, there is a need to find the correlated attribute groups and apply rule algorithm on these groups. Also, Apriori deals only with nominal attribute values. Numeric and Date strings are not applicable to most of the associative rule algorithms.

5.5.7 Visual Analysis using stacked histograms

A histogram is defined as a bar graph that shows frequency data. In a histogram, data is collected and sorted into categories. Analysis using histograms is a powerful technique for looking at and processing large amounts of data. Histograms focus on the frequencies and distributions of one particular attribute, for example, the priority description for the whole data set as illustrated in Figure 42 addresses only that Priority attribute. In order to find out correlations between various attributes, there is a need of an interactive visualization rather than a static view of histogram.

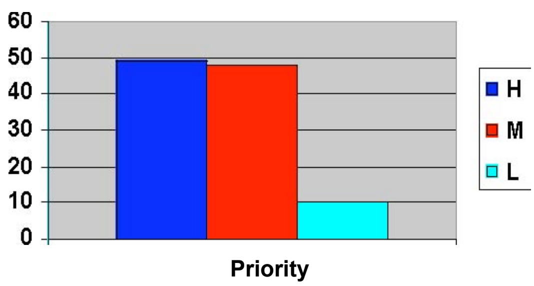


Figure 42. A standard histogram of the “priority” attribute.

WEKA incorporates a stacked histogram which allows three judgments: the trends on the total height of the columns; the proportion of each category within each column; and the trends in the lowest category (Dix and Ellis, 1998). This interactive stacked histogram solves the problem of incapability of cross comparing for standard histograms by allowing different trends to be analysed using the same dynamic graph. Thus, the correlation between attribute "priority" and "cause-of-repair" can be visualized as shown in Figure 43. *A rule can be learned from this interactive stacked histogram, that is about 94% of A/C malfunction belongs to high or medium Priority job.*

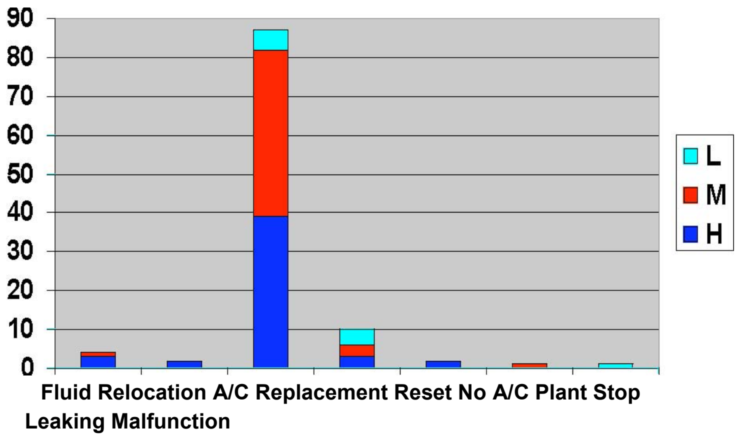


Figure 43. A stacked histogram of correlation between "Priority" and "cause-of-repair"

A number of rules can be generated from the analysis of correlations between various attributes with the assistance of WEKA's stacked histograms function at pre-process panel which is developed in the new version WEKA 3.3.6 knowledge explorer. The stacked histograms are capable of handling both numeric and nominal attribute but they do not perform well in the Date analysis and continuous numeric values.

5.6 Demonstration of Data Mining Techniques

This section discusses the preliminary results of the data mining techniques selected for systems implementation and tested their capacity to search for patterns and correlations of building maintenance data. Based on the industrial building maintenance data, the data mining algorithms provided by WEKA[41], are applied and the results analysed through interactive visual tools such as stacked histograms. The selected industrial building maintenance data covered a variety of domains. The aim of this demonstration was to discover meaningful patterns that can assist facility managers in strategic planning as well as provide a knowledge base to shape future requirements gathering and design briefing.

Through a process of cleaning, integrating meaningful domain attributes, conversion of data formats, a number of experiments were performed in which concentration is given to visual analysis using stacked histograms, classification and clustering techniques, an associative rule mining algorithm such as "Apriori".

Visual data analysis and data mining techniques were applied on three selected data sets. The building maintenance data used in this demonstration are for three asset types including Air Handling Units, Thermostat Mixing Valves and Battery Chargers at Building 10, Royal Prince Alfred Hospital. An example of the raw maintenance data is shown in Appendix D. Some interesting results regarding patterns and structures of data have been obtained and the following provides a summary of results

[41] http://www.cs.waikato.ac.nz/ml/weka/

from applying data mining techniques on Building 10 maintenance data.

5.6.1 Results extracted from the available maintenance data of the asset type 1: Air Handling Units

1. Approximately all A/C malfunction belongs to high and medium priority.
2. A/C malfunction concentrates on the problems of: too_hot 32%, too_cold 28%, not working 7.5%.
3. The lowest levels of A/C Malfunction took place in August followed by June and April while other months share similar high A/C malfunction rate.
4. Approximately all the description of *too_cold* or *too_hot* were associated with high or medium priority
5. All 7th floor jobs were of high and medium priority
6. In all 7th floor jobs the cause of repairing was A/C malfunction
7. Maintenance jobs conducted at Floor 7 did not meet the expectations: with 23 out 25 completion_within_expectation = “N” (92%)
8. In August only department 26464 has corrective maintenance work. (only 1 case).
9. The work at 4, 5, 6, 7th floor constitutes most of the reports of A/C malfunctions, with 86% of A/C malfunction reported from these floors.
10. Floors 4,5,6,7 constitute 80% of completion not meeting expectation.
11. No A/C malfunction was reported at level 9.
12. Higher percentage of users’ unhappiness associated with high and medium priority.
13. Higher level of unhappiness related to completion not meeting expectation with a focus at too_hot and too_cold adjustment activities.
14. The department of Cost centre 0 reports 42% of CM work.
15. Department 26462 only reports A/C malfunction. (all 18 cases)
16. For most costcentre = 0 (45 out of 47), the jobtype = CM
17. For floor 5,6,7, workOrder_Status always was completed

5.6.2 Results extracted from the available maintenance data of the asset type 2: Thermostatic Mixing Valves

1. High priority works constitute of 55/101 monthly work, 22/101 6mthly work, 24/101 12mthly work. (All 6mthly and 12mthly works are of highly priority)
2. 12mthly work happened in the middle of the year – June-Sept, while all 6mthly work is carried out in December.
3. All monthly and 12mthly works were completed. Parts of 6mthly works (50%) were outstanding.
4. All 6mthly and 12mthly works did not meet the expectations of completion date.
5. Monthly work was identified as TMV004, 6mthly = TMV002, 12mthly = TMV003
6. All maintenance of thermostatic mix valves happened at Level 4.
7. All monthly works were supposed to be completed in 0.5 hours and cost $10. All 6mthly and 12mthly works were estimated to be completed in 2 hours and cost $29.
8. All high priority works did not meet the expected completion data.
9. All medium priority works were completed on the expected completion data.
10. There was a trend in emphasizing the maintenance of thermostatic mix valves recently – with increasing of workorder_No, the priority was getting higher.
11. All works between August and December did not meet the expectation of the completion date.
12. All 6mthly works (TMV002) were carried out in December.
13. With higher priority works there was a low level satisfaction of work completion.
14. High priority works take more than 0.5 hour to finish while total hour the duration of <=0.5 must

belong to medium priority works.

15. 50% of 6mthly work did not finish (these unfinished 6mthly works' WorkorderNo> 725085)
16. There is an incremental relationship between the work priority, the estimated time to complete the work and the associated budget.
17. There is a pattern in relation to the work priority, level of meeting the expectation and the frequency of the task TMV004.
18. 100% of medium priority works occurs on monthly maintenance and relates to a specific task TMV004.

5.6.3 Results extracted from the available maintenance data of the asset type 3: Battery Chargers

1. All outstanding works took place at the end of work order list around December 2002.
2. Asset "EPG0101" belongs to cost centre "1000" while the cost centre for the other asset were not available;
3. There is fee charge with asset "EPG0101" while no charge for "EDG1000-01".
4. For all tasks with work order No > 66195, and some tasks with work order No between 48002 and 66195, completions did not meet the expectation of completion date.
5. 57% of workorders were completed within the expected completion date.

5.7 System Components and Results Summary

Although the project focuses on mining the maintenance data in which huge benefits go to the facility management, it is not prevented from attempting to fill the gaps between designing and building maintenance in biasing the future design solutions within the scope of the whole coordinated building life cycle. The major contribution of data mining for this project is to provide a knowledge base which is served as a centre bridge. The impact of this research project can be glimpsed through Figure 44.

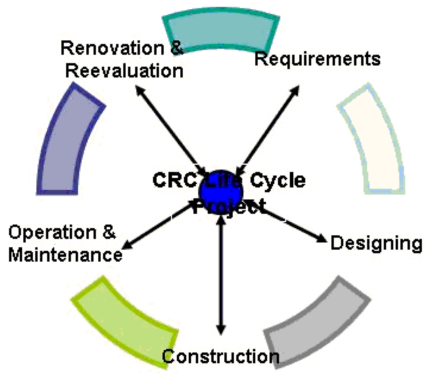


Figure 44. Impact of this project to the whole building life cycle

The detection of previously undiscovered patterns in the maintenance data can be used to determine factors such as the cost effectiveness and expected failure rate of assorted building materials or equipment in varying environments and circumstances. These factors are important throughout the life cycle of a building, and such information could be used in the design, construction, refurbishment, and maintenance of a building site, representing a substantial decrease in cost and increase in reliability.

In order to advance the testing and the feasibility of the proposed approach, a systems prototype was further developed and tested using the building maintenance databases and available CAD drawings. The development of a fully operational prototype system is presented in the following section and the results of further testing are provided in Section 7.

6. AIMM PROTOTYPE SYSTEM

6.1 Architecture of AIMM

As a result of the preliminary data mining demonstration and testing the integration of all system's components and modules were integrated and refined to develop the Agents for Improving Maintenance Management (AIMM) prototype system. The AIMM prototype system includes: an object oriented 3D CAD model of a building modelled in the ArchiCAD package, and a maintenance database in a standard SQL (standard query language) format. The architecture of AIMM was developed to include three agents: Interface, Maintenance and Situated agents, as illustrated in Figure 45. The roles of these agents include:

(a) The appropriate mapping between the building assets of the building model in the virtual environment is maintained by the *Maintenance Agent* that connects data contained within the Maintenance Database with data contained within the EXPRESS Data Manager (EDM) Database via the virtual environment, Active Worlds.

(b) Linking Data Mining techniques to building models in a virtual environment (Active Worlds) is achieved via a *Maintenance Agent* that accesses the maintenance database and applies its mining algorithms on it.

(c) Linking knowledge development with the building model in virtual environments is carried out by the *Situated Agent* that assists in improving maintenance management by providing life cycle implications as feedback whenever building assets (mechanical and electrical elements) are selected in the building model in the virtual environment (Active Worlds).

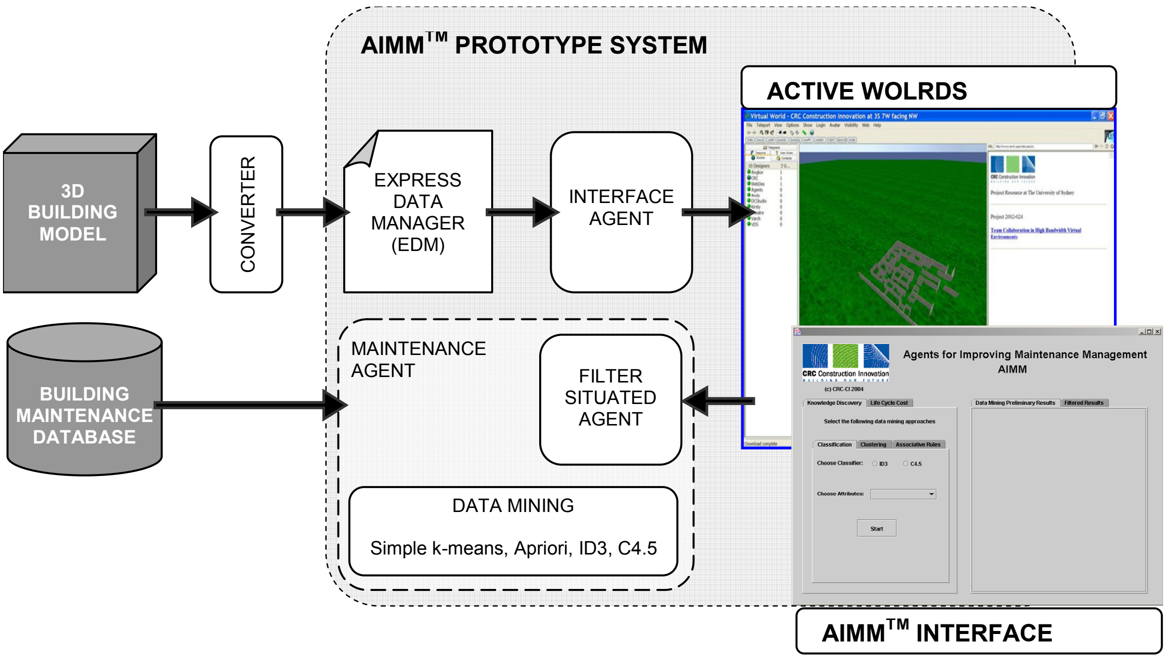


Figure 45. The architecture of AIMM

Feedback of useful knowledge can be discovered by the Maintenance Agent in the application of the four data mining techniques and algorithms (see in Section 5.5) in order to discover various classification of knowledge. The data mining algorithms and the link between its knowledge development and the building model in a 3D virtual environment has been fully implemented.

The following four phases reflect the all key composites (outlined in Section 5) of this project:

1. Phase 1: Pre-processing industrial data, selection and application of data mining algorithms on maintenance database;
2. Phase 2: Linking 3D CAD data with Virtual Environment (Active world);
3. Phase 3: Maintenance interface agent and maintenance agent, linking these agents to active world;
4. Phase 4: Filtering agent for post-processing.

6.1.1 Pre-processing and Selection

Phase 1 involves the manual pre-processing of data, which removes noisy, erroneous and incomplete data to derive important attributes from original raw data. For example, the raw text description of time of work orders “1/12/2001” must be converted to a meaningful attribute such as “month”. Moreover, various “testing” algorithms are run through the maintenance data to find out the suitable data mining approaches. From Phase 1, the quality of the data can be improved.

6.1.2 Linking 3D CAD Data with Virtual Environment

Phase 2 adopts the EDM interface agent that was developed in CRC-CI “information flow” project in converting IFCs (Industry Foundation Classes) object model into a Renderware (RWX) format, so as to be presented at the virtual environment (Activeworlds), shown in Figures 46 and 47. The virtual environment provides a collaborative multi-user interface and more importantly, a means for the user to walkthrough 3D object model at a real time. The user is able to navigate and select a building asset type to explore useful knowledge.

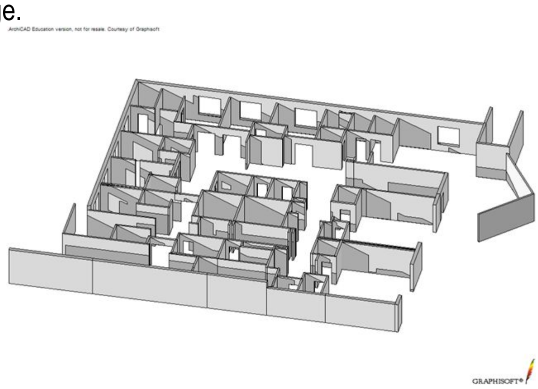


Figure 46. Part of the typical floor plan of an object oriented CAD model of building 10 at RPAH, Sydney modelled in the ArchiCAD package.

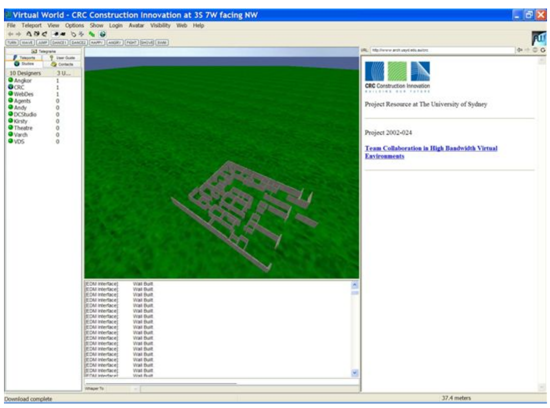


Figure 47. A 3D building model presented in the virtual environment (Active Worlds) after being converted by the EDM Interface Agent.

6.1.3 Maintenance Agents for Data Mining

Phase 3 instantiates the maintenance interface agent and the maintenance agent. Once the user decides to select a certain building asset (right clicking on the object representing this building asset at Active Worlds), the maintenance interface agent is invoked to load related data from database. The Maintenance Agent performs data mining on the selected asset type. The four mining algorithms that have been implemented in the system prototype of AIMM are the same as those demonstrated in Section 5.6, i.e., clustering using "SimpleKmeans", associative rules learning in "Apriori", classification using "C4.5" and "ID3".

6.1.4 Post-processing and Filtering

A situated agent is developed in Phase 4. This is a software agent that performs post-processing of the mined results. The situated agent filters out irrelevant patterns based on the heuristic rules. This architecture has been implemented in the system prototype of AIMM using Active Worlds as the virtual environment platform and developed using Java programming language. Table 12, 13 and 14 provide some examples of the kinds of heuristic rules defined for the filtering agent.

Table 12: Heuristics for Air Handling Units

If Clause	Filtering Display
If the result contains string patterns matching "despcriptionofCause = too_cold ……. | floor = 7 || priority = H:N|| priority = M:N|| priority = L:Y"	Within the complaints related to air conditioning "too_cold", all the "high" and "medium" priority jobs did not complete within expectation. All the "low" priority jobs were finished on time; This may reflect lack of supervision of high and medium air conditioning maintenance works in facility management.
If the result contains string matching "floor=4 …… |month = Dec: O"	For floor 4, air handling unit works that occur in December are not likely to be finished – inspections needed to locate the causes of this failure.
If the result contains string matching "department = 26462 | workorderNo<=50461:1(4.0) | workorderNo > 50461:7(11.0)"	Department 26462 resides only at 1st and 7th floor – this tell you locations of some department
If the result contains string matching "floor=7 25 = = > causeofrepair = A/C_Malfunction 25"	All air handling units maintenance works in floor 7 belong to A/C malfunction – failures concentrate on a particular floor, needs inspections.
If the result contains string matching "deparment = 26462 18 = = > causeofrepair = A/C_Malfunction 18"	All air handling units maintenance works in department 26462 belong to A/C malfunction – failure abnormally concentrate on a particular department, which needs more inspections.
If the result contains string matching "deparment = 21271 16 = = > floor = 6 16"	Department 21271 resides only at 6th floor – this tell you the location of the department
If the result contains string matching "floor=7 25 = = > completionwithinexpectation=N 23"	Floor 7 is most likely not able to meet expectations: with 23 out 25 completionwithinexpectation = "N" (92%), inspection needed to identify the facility management failures.
If the result contains string matching "costcentre=0 47 = = > jobType=CM 45"	96% jobs for costcentre = 0 is corrective maintenance – this is the information about billing.
If the result contains string matching "Cluster0 ….. A/C_Malfunction too_hot …… Cluster 1 …… A/C_Malfunction too_cold"	A illustration of centroid as "too_hot" and "too_cold" and the corresponding percentages "62%" and "45%"

Table 13: Heuristics for Thermostatic Valves

If Clause	Filtered Agent Display
If the result contains string matching "frequency = monthly | priority = H || Completionwithinexpectation = Y:July|| Completionwithinexpectation = N:August"	Appendix D - For all monthly high priority works, all the works in **July** complete within expectation and those in **August** fail – possible failures in relation to seasons or resulting weather and humidity;
If the result contains string matching "frequency = monthly …… | priority = M || Completionwithinexpectation = Y:Jan|| Completionwithinexpectation = N:June"	Appendix E - For all monthly medium priority works, all the works in **January** complete within expectation and those in **June** fail -- possible failures in relation to months or the resulting weather and humidity;
If the result contains string matching "month=Jan:Y month=Feb:N month=Mar:N month=April:N month=May:Y month=June:N ……"	Works in January and May complete within expectation and the rest of the months fail – possibly related to season or temperature or humidity differences. These phenomena may caused by other reasons – different teams or personnel.
If the result contains string matching ":L4"	All thermo static valves maintenance works took place in floor 4
If the result contains string matching "StartDate > 1023926400000: H (61.0)"	Appendix F - Start Date after the date which value equals to "1023926400000" must belong to **high** priority work – A trend of recently emphasizing of thermostatic mix valves maintenance.
If the result contains string matching "CompletionDate > 989741520000: N"	Most thermo static valve maintenance works (211 out of 213) completion after the date "989741520000" did not meet expectations – needs more investigations
If the result contains string matching "workorderNo <= 72085: C (255.0)"	Appendix G - Jobs with WorkorderNo <=725085 are all completed.
If the result contains string matching "Completionwithinexpectation=Y 49 = = > priority=M 47"	96% thermo valve maintenance works which completions meet expectation are medium priority.
If the result contains string matching "frequency=monthly 252 ==> WorkOrderStatus=C 252 conf:(1)"	all monthly maintenance works are finished
If the result contains string matching "Cluster0 ….. monthly SM M …… Jan …… Cluster 1 …… monthly SM H …… Dec"	A illustration of centroid as "Medium Priority Job in January" and "High Priority Job in December" and the corresponding percentages "66%" and "34%"

Table 14: Heuristics Battery Units

If Clause	Filtered Agent Display
If the result contains string matching "workorderNo > 66195: N (30.0)"	Appendix H - All the recently issued jobs are not completed within expectation.
If the result contains string matching "workorderNo <= 70266: C (168.0/1.0) workorderNo > 70266: O (14.0/4.0)"	Appendix I - Most of previously issued works have been completed, while most of (12 out of 14) the recently issued works are not finished. A suggestion for investigation for potential faults.
If the result contains string matching "Cluster0 ….. M May…… EPG0101 …… Cluster 1 …… M April …… EDG1000-01"	A illustration of centroid as "Medium Priority Job in May on EPG0101" and "Medium Priority Job in April on EDG1000-01" and the corresponding percentages "50%" and "50%"

6.2 AIMM Interface

The user is able to navigate a 3D model within a real time virtual environment as shown in the top left of Figure 48. The user is able to instantiate the AIMM prototype system by "right-clicking on the desired building asset or component, thereby presenting the main Maintenance Interface as shown in the bottom right of Figure 48.

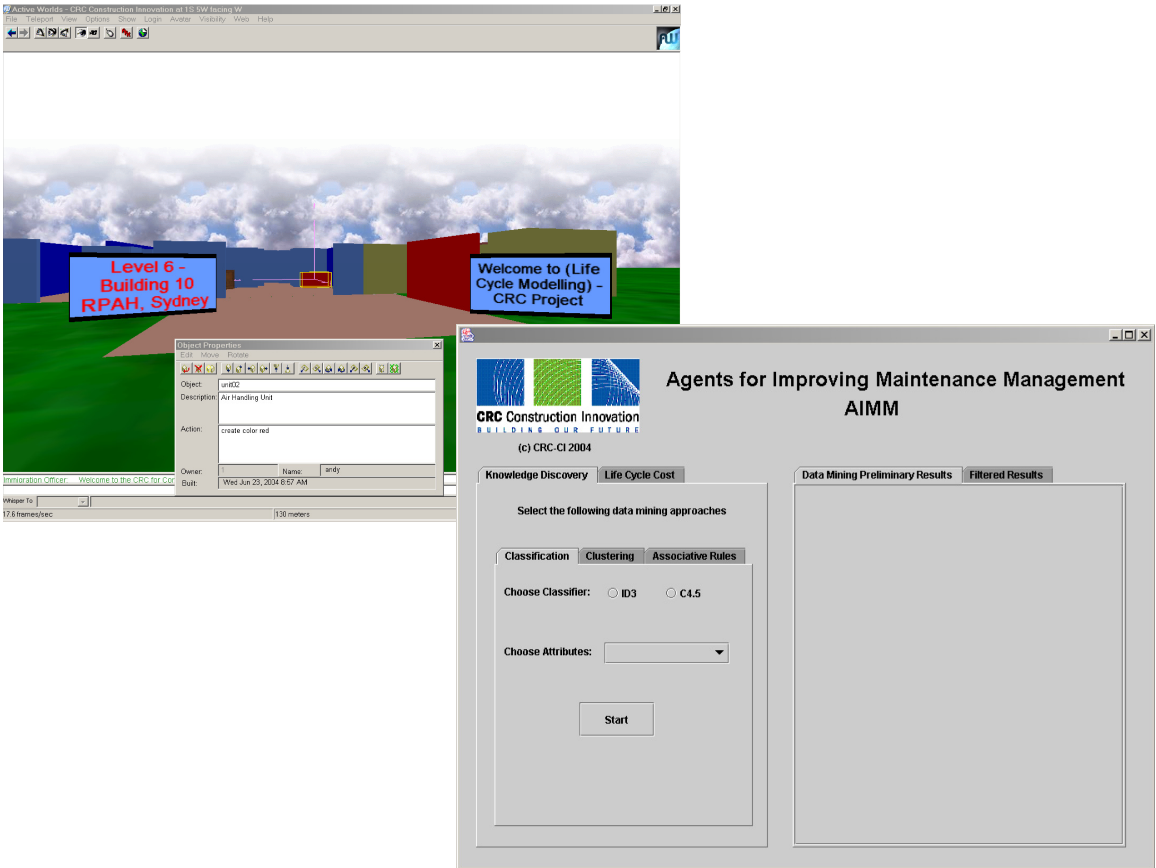


Figure 48. Selecting an asset type in Active World instantiates the maintenance interface agent

The main Maintenance Interface is divided in to equal parts. On the left hand side are two stacked panels: (i) Knowledge Discovery and (ii) Life Cycle Analysis. Here we will only deal with the Knowledge Discovery panel since the Life Cycle panel will be discussed later in Section 8.4. In the Knowledge Discovery panel are located three stacked sub-panels: (I) Classification, (ii) Clustering and (iii) Associative Rules. These panels provide a range of ways for using each different algorithm. This provides the user with greater flexibility and scope since the user may test a variety of data mining approaches for each type of algorithm. On the right hand side are another two stacked panels that are dedicated to reporting results. Results are reported in two ways. The panel named Data Mining Preliminary Results displays the results of the chosen algorithm in their "raw" form. The panel named Filtered Results displays the results in their interpreted form using domain derived heuristics. The overall data mining interface is shown in Figure 49 and illustrates the hierarchy of stacked panels for the different data mining scenarios.

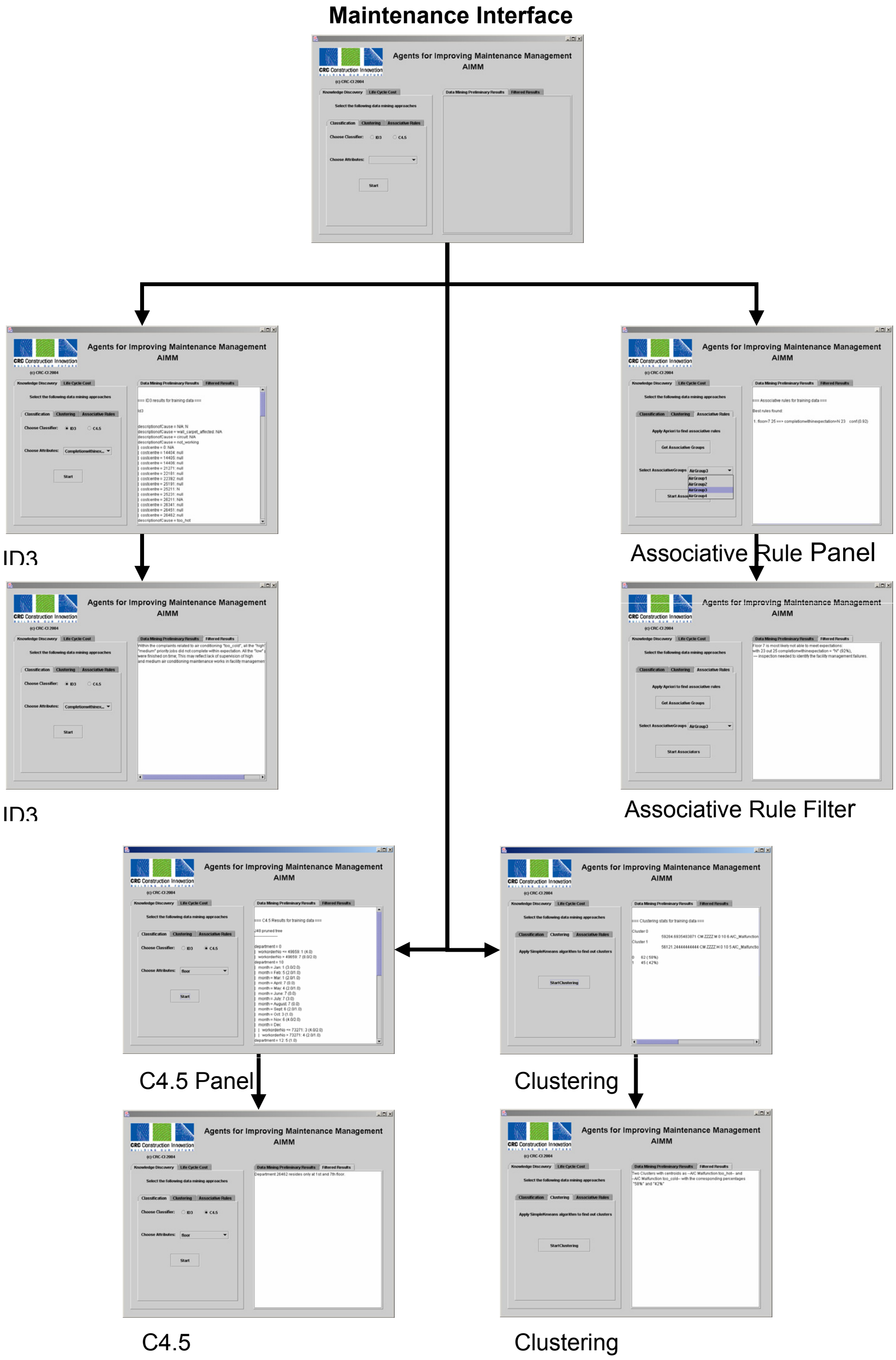


Figure 49. Overall Interface and user workflows

6.2.1 User Scenario

The following scenario is proposed during typical interactions between a user and the proposed system in the previous section. In this scenario, we shall use an Air Handling Unit (AHU) as the chosen building component that a user wishes to apply Data Mining on. The following sequence of actions is then followed:

- The user navigates the building in a real-time and online 3D virtual environment as shown in Figure 50;
- Once the user selects a building asset type such as the Air Handling Unit the object property window pops out describing general information of the selected object as shown in Figure 51;
- AIMM invokes the Interface and the main window pops up to allow selection of algorithms as illustrated in Figure 52;
- The user explores a variety of data mining algorithms and chooses the desired algorithm by clicking on a checkbox or button and running the algorithm as shown in Figure 53;
- AIMM invokes the Maintenance Agent running the algorithm and results are reported first in the Data Mining Preliminary Results panel as illustrated in Figure 54;
- User selects Filtered Results in order to access post-processed knowledge and an example of filtered knowledge is shown in Figure 55.

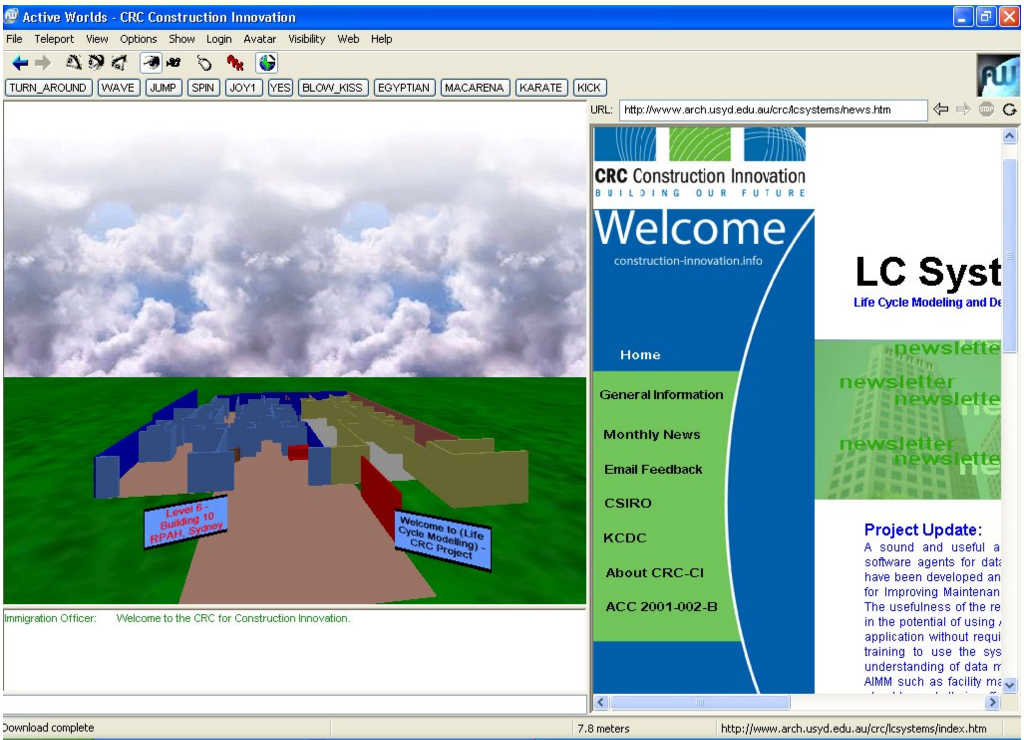


Figure 50 The primary interface of software agents prototype system [AIMM] in an interactive network multi-user environment.

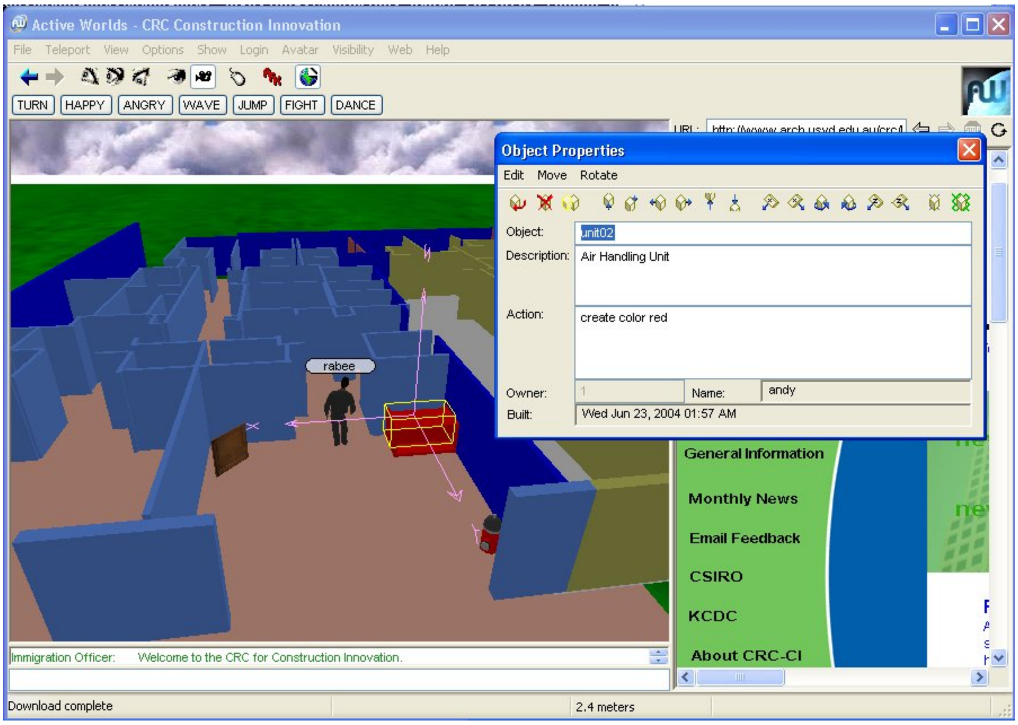


Figure 51. The user selects a building asset type (the Air Handling Unit) and an object property window pops out describing general information of the selected object.

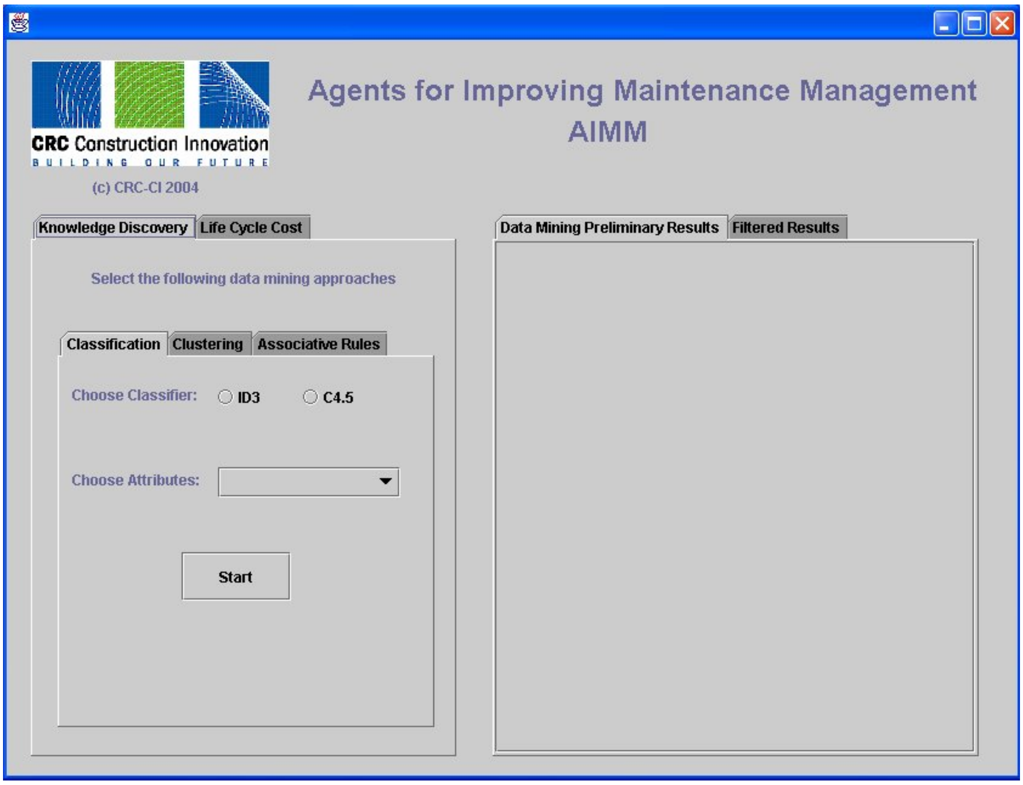


Figure 52. The AIMM prototype system is instantiated once a building asset type has been selected.

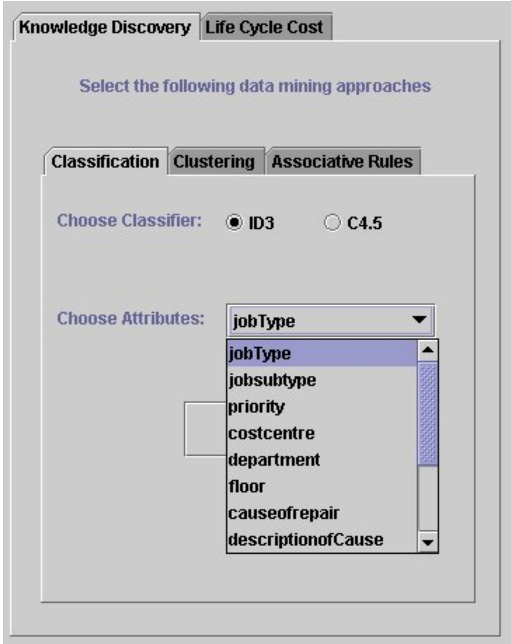


Figure 53. Data mining techniques and different attributes for the user to choose from based on focus and interest.

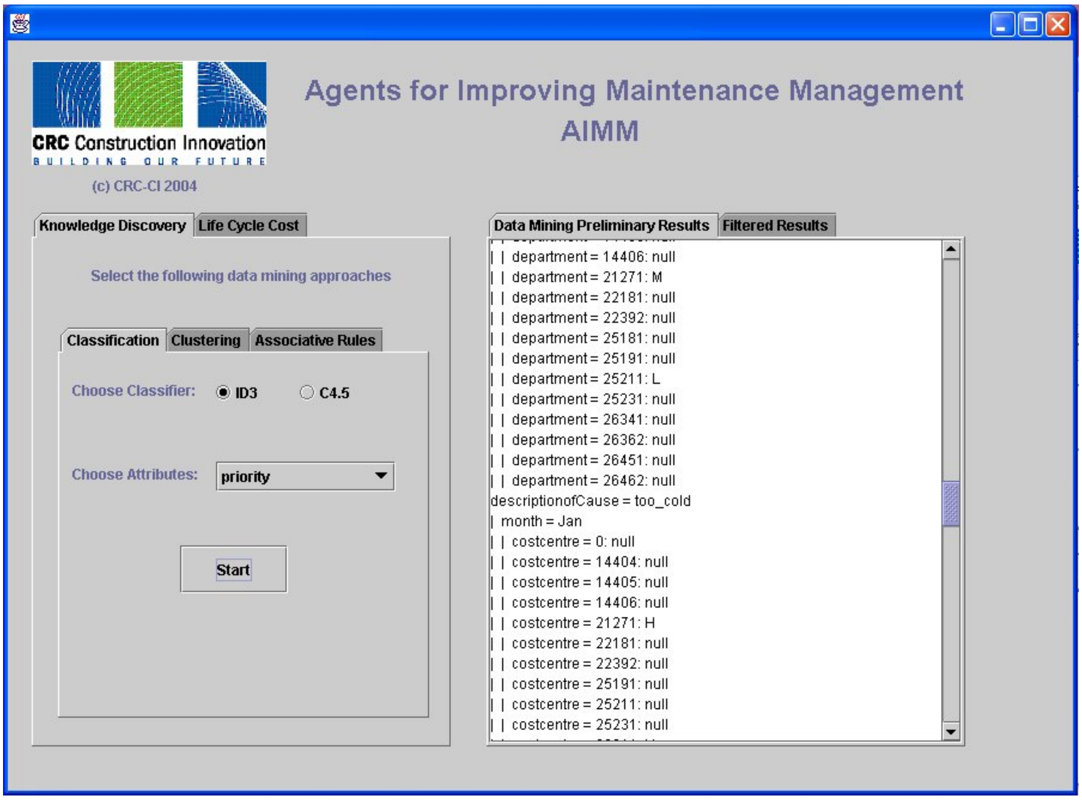


Figure 54. Preliminary results of applying the ID3 with the “Priority” attribute on the maintenance data of Air Handling Unit.

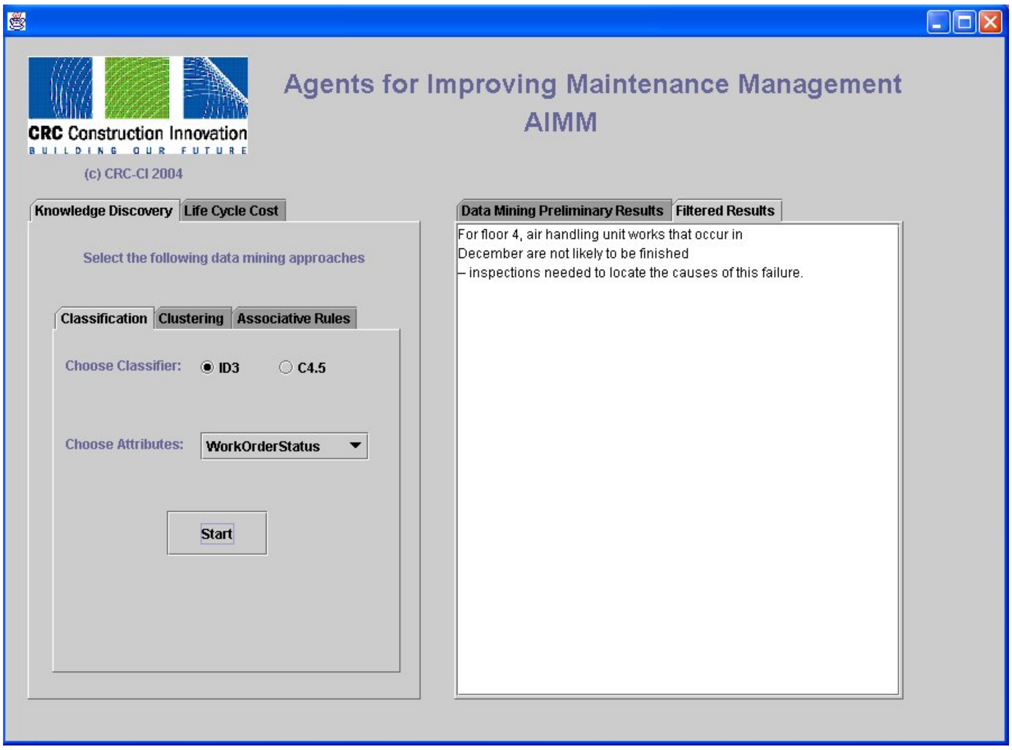


Figure 55. An example of the filtered knowledge presented to the user from the preliminary results of applying the ID3 with the "Work Order Status" attribute on the maintenance data of Air Handling Unit.

7. DEMONSTRATION OF AIMM™ ON BUILDING NO. 10 AT RPAH

7.1 Demonstration of using AIMM on Building no. 10 at RPAH

This Section illustrates a demonstration of a run of the AIMM™ prototype system. This demonstration applies the system on Building No. 10 at RPAH, Sydney. The maintenance data was provided by the Engineering Division of the Central Sydney Area Health Service (CSAHS) and details of the maintenance data can be found in Section 4.2. Maintenance data for the last two and a half years is available in SQL format and contains data that is highly detailed and structured. There are approximately 5,000 work orders recorded for Building 10 in the period from 1 January 2001 to 9 December 2002.

AIMM starts by converting and presenting the 3D Model of Building No.10 in the virtual environment (Active Worlds). The user may navigate the 3D model within a real time virtual environment. We may assume that a user resolves to apply Data Mining on the Air Handling Unit (AHU). Once the user right clicks on the AHU object in the 3D model, the AIMM™ system invokes the Maintenance Agent Interface that activates the Maintenance Agent to apply the four Data Mining techniques and the Situated Agent presents the learned knowledge in a WEKA Learning Results window as shown in Figure 56.

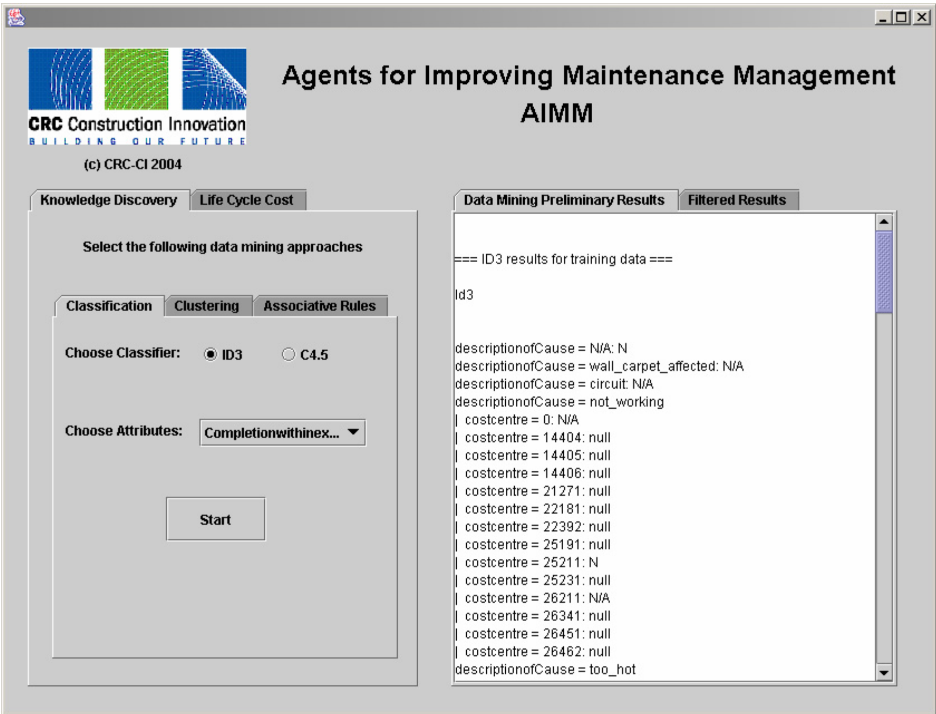


Figure 56. Demonstration output (display Part I) of applying Data Mining on the maintenance data of the Air Handling Unit at Building no. 10, RPAH.

Visual data analysis and data mining techniques were applied on two selected data sets: air handling units and thermostatic mixing valves at Building 10, Royal Prince Alfred Hospital. The evaluation of the results obtained from mining the maintenance data of the above two asset types and their impact on improving the maintenance of existing buildings and the design of future facilities are shown in Tables 15 and 16. For further detail refer to Appendix F.

Table 15. Evaluation of results for air handling units and impact on improving maintenance of existing and design of future building facilities.

Data Mining Technique	Rules Obtained	Potential Impact on Facility Maintenance and Design
Visual Analysis	Approximately all "A/C malfunction"belongs to high and medium priority.	A/C malfunction" is of a major concern in guiding the allocation of maintenance resources.
	"A/C malfunction" is concentrated on the problems of: *too_hot* 32%, *too_cold* 28%, *not_working* 7.5%.	Temperature should be automatically adjusted and a provision of self-reporting faults equipments should be considered.
	The lowest levels of "A/C malfunction" took place in August followed by June and April while other months share similar high rate of "A/C malfunction".	
	The maintenance work on 4, 5, 6 and 7th floors constitute most of the reports of A/C malfunctions, with 86% of A/C malfunction reported from these floors.	
	Approximately all the descriptions of *too_cold* or *too_hot* were associated with high or medium priority.	The appropriate temperature is of high priority from users' perspective.
Visual Analysis + Decision Tree Algorithm (C4.5)	All 7th floor jobs were of high and medium priority and the cause of repairing was "A/C malfunction".	Investigate the possibility of poor design or maintenance of air conditioning function in 7th floor. A special attention in the design should be given to a specific floor due to its high demand of corrective or preventive maintenance or special design of A/C.
Visual Analysis	Higher percentage of user dissatisfaction in terms of work completed not meeting expectations is associated with maintenance work of high and medium priority.	Seeking feedback from users is important in order to improve the efficiency of building maintenance and achieving user satisfaction.
	Higher level of unhappiness related to completion not meeting expectation with a focus at *too_hot* and *too_cold* adjustment activities.	Paying attention to temperature adjustment in maintenance and design phrase may improve users' happiness.
	Cost centre 0 has the highest percentage of maintenance not meeting expectations (44%)	A special attention should be directed to certain places in the building wherein maintenance work is more likely to consume more time and effort than expected compared to normal places.
Decision Tree Algorithm (C4.5)	Department 26462 only reports A/C malfunction. (all 18 cases)	A special attention should be directed to certain places in the building wherein maintenance work is required more often.
	96% jobs for cost_centre = 0 is CM (corrective maintenance).	
Association Rule Algorithm	For floors 5, 6 and 7, the workOrder_Status was always completed.	Benefiting from successful maintenance practices including both equipments and labour is useful to achieve a high level of an overall maintenance performance.

Table 16. Evaluation of results for thermostatic mixing valves and impact on improving maintenance of existing and design of future building facilities.

Data Mining Technique	Rules Obtained	Potential Impact on Facility Maintenance and Design
Visual Analysis	The percentage of high priority work constitutes of 55% of monthly work, 22% of 6mthly work and 24% of 12mthly work.	These percentages should direct the allocation of maintenance resources at the appropriate time of the year to achieve better planning and scheduling of maintenance work.
Visual Analysis + Decision Tree Algorithm (C4.5)	12mthly work occurred during the middle of the year – June-Sept, while all 6mthly occurred in December.	
Visual Analysis	All outstanding works took place in December	
	All monthly and 12mthly works were completed. Parts of 6mthly works (50%) were outstanding.	The 6mthly maintenance work should be thoroughly analysed to identify the actual reasons of incompletion.
	All high priority works did not meet the expected completion date.	Maintenance job required with high priority should be investigated in relation to maintenance labour and the practicality of initial expected date of completion that has not been met.
	All works between August and December did not meet the completion data.	
	All medium priority works were completed on the expected completion data.	Analyse this phenomenon to help identifying the deficiency of not meeting the expected completion with high priority works.
Decision Tree Algorithm (C4.5)	All monthly high priority works are carried out in the later part of the year – July to November.	Distribution of priority of maintenance work is important in planning and scheduling maintenance work and resources.
Association Rule Algorithm	There is an incremental relationship between the work priority, the estimated time to complete the work and associated budget.	A better planning and scheduling will help to advance this pattern of relationship.

7.2 Discussion

From the result generated through a single run of these algorithms, it can be observed that the rules generated are relatively interesting. The rules obtained (an example shown in Table 17) were pulled out manually in a strategic manner. The pull process involved cleaning and modification of data files; and selecting various attributes alternatively. For example, when the associative rule algorithm is used on the data set meaningless and less useful rules were produced. Whereas if a small file (containing just 2 to 3 relevant attributes) is fed into the same algorithm, much more interesting rules are discovered. Currently the four algorithms that have been implemented can only be automated according to the selected attribute for classification. For example, if the attribute "outlook" is selected for splitting in the Weather database a different decision tree is generated from selecting the "humidity" attribute for classification.

Table 17. An example of some of the results pulled out manually when applying data mining techniques on air handling units.

Data Mining Technique	Rules Obtained	Potential Impact on Facility Maintenance and Design
Decision Tree Algorithm C4.5, and Association Rule Algorithm	All 7th floor jobs were of high and medium priority and the cause of repairing was "A/C malfunction".	Investigate the possibility of poor design or maintenance of air conditioning function in 7th floor. A special attention in the design should be given to a specific floor due to its high demand of corrective or preventive maintenance or special design of A/C.
	Department 26462 only reports A/C malfunction. (all 18 cases)	A special attention should be directed to certain places in the building wherein maintenance work is required more often.
	96% jobs for cost_centre = 0 is CM (corrective maintenance).	
	For floors 5, 6 and 7, the workOrder_Status was always completed.	Benefiting from successful maintenance practices including both equipments and labour is useful to achieve a high level of an overall maintenance performance.

7.2.1 Improving the results obtained from using AIMM™

The results indicate that data mining does not automatically extract all available knowledge that is embodied in a data set. Although it may sound at first appealing to have an autonomous data mining system, in practice, such a system would uncover an overwhelmingly large set of patterns, and most of the patterns discovered in the analysis would be irrelevant to the user. Results indicate that a more realistic scenario would be to communicate with the data mining system, using additional questions to examine the findings and direct the mining process; some of these questions might include (Morbitzer, Strachan and Simpson, 2003):

- What is task relevant data?
- What kind of knowledge do I want to mine?
- What background knowledge could be useful?
- How do I want the discovered patterns to be presented?

Consequently, the first analysis will not necessarily provide the required knowledge since the user might have defined a mining exercise that does not reveal important patterns. Hence, the analysis needs to be refined. The creation of different mining exercises is supported by a very flexible definition of a mining task. The user can therefore quickly change variables to be included in the mining run, in combination with filters that can be defined for all variables. For example, only focus on times with a resultant temperature above 27^{0}C, occupied period, time of high occupancy densities, etc.).

7.2.2 Scenarios for improving AIMM

Herein are some of the scenarios utilised for improving the results obtained from using data mining techniques on maintenance data for Air Handling Units. These scenarios have been integrated and implemented within AIMM as shown in Figures 50-55 of the previous Chapter.

Applying ID3

Prior to applying ID3 on air conditioning unit's maintenance data, convert data file into nominal value and replace all the missing values with "N/A" strings.

- The ID3 algorithm takes file CMID3normic.arff and some results are list as below:
  - ID3 classifies on attribute "Completionwithinexpectation"

    For "too_cold" descriptionofCause, all the "high" and "medium" priority job did not complete within expectation. All the "low" priority jobs were finished on time;
  - ID3 classifies on attribute "WorkOrderStatus"

    For floor 4, works that occur in Dec are not likely to be finished (with WorkOrderStatus = "o")

Applying C4.5

Classify against the "floor" and "descriptionofCause" attributes

- The C4.5 algorithm takes file AirCondCMupdate4Nov.csv and some results are list as below:
- **Department 26462 resides only at 1st and 7th floor:**
- **Department 21271 only reside at 6th floor:**
- **Department 26462 only reports A/C malfunction.**

Applying Apriori:

Prior to mining the maintenance data using the Apriori algorithm, apply attribute evaluator "CfsSubsetEval" and search method "BestFirst". The data file is divided into sets of data files.

- Applying Apriori on the data set of (priority, department, floor, causeofrepair, descriptionofCause, workormaterial), the results include:

  All works in floor 7 belong to A/C malfunction
  All works in department 26462 belongs to A/C malfunction
  Department 21271 only resides at floor 6
- Applying Apriori on the data set of (priority, floor, completionwithinexpectation), some of the learned rules include:
  - **Floor 7 is most likely not able to meet expectations: with 23 out 25 completionwithinexpectation = "N" (92%)**
  - **Apply Apriori on relation6.arff (jobType, costcentre, causeofrepair): 96% jobs for costcentre = 0 is CM**

8. RECOMMENDATIONS AND FUTURE WORK

8.1 Data Mining Requirements

The increasing influence and use of data mining in many domains can be largely attributed to its ability to support decision making. As an integrated approach to facilities management, the use of data mining implemented in the AIMM system is a novel approach. Toward this end, the AIMM prototype system requires reliable data and must be able to translate, query, mine and filter this data quickly. For the AIMM system to reach its full potential future research and development must consider data collection, data quality as well as the accuracy and long-term maintenance of the data sets themselves. It is also critical that these aspects also be considered in relation to overall maintenance goals and objectives. Therefore, there is a need to address the (i) necessary specification criteria that enables the identification of user goals, (ii) resulting type and quality of data required and (iii) method of data collection to ensure useful information is maximised and human error is minimised.

While the outcomes of both the industry data (Section 4.2) and data mining techniques (Section 5 and 7) surveyed and tested in this demonstration, were separated for purposes of discussion, they are closely linked in practice. Indeed, the power of the data mining scenarios implemented within AIMM come largely from the ability to simultaneously access and process both large quantities of data and high quality 3D building components and their corresponding maintenance data. In addition, whilst the filters developed here for performing secondary analyses are derived from domain knowledge, this phase does not constitute a complete or comprehensive listing of heuristics and further development is required through consultation with a variety of facilities managers, maintenance and design experts.

Data mining is a life cycle process in which the knowledge obtained will affect interpretation of data gathering, in terms of the availability and priority of certain attributes. This process is established via the: setting up of mining goals → pre-processing data → data transformation → data mining → evaluation → refinement of data requirements (according to the knowledge obtained). Therefore to secure the more effective use of the AIMM system, future research should begin with addressing a variety of different management scenarios for a range of maintenance goals in order to identify the resulting data requirements.

8.1.1 Data Collection and Updating

By investigating a variety of different facilities maintenance scenarios it may be possible to build a better understanding of the requirements of maintenance data collection methods during which crucial building component and attribute data is obtained. Since in practice, initial maintenance goals drive the way in which data is collected, there is a real need for further consultation with Industry partners to identify critical relationships between goals, data requirements and their collection. As a result, data mining the building component's attribute data will yield more effective and helpful results.

Collection of maintenance data to be stored in the database is typically a large expenditure in implementing any kind of maintenance management system. For this reason, in conjunction with general data mining requirements, maintenance data must be as reliable and accurate as possible. Complete and accurate documentation of all maintenance operations is important to assure the integrity of the AIMM system, the reliability of subsequent analyses as well as the ability to maintain the system over time. Remembering that the utility of data mining is a direct function of the data contained within the system, a commitment must be made to maintain the database itself. Only dedicated care of the database can ensure that the data mining scenarios and subsequent analysis will produce results of the highest order.

8.1.2 Potential Maintenance Goals and Objectives that Drive Data Collection

The quality of data mining results is a direct consequence of the overall goals of maintenance managers and designers. The following themes are commonly pursued in facilities maintenance management and are considered useful in developing further data mining scenarios, identifying all essential component attributes and structuring the appropriate strategies for data collection:

- Budget and Cost-based analysis – Existing costs, ongoing costs and cost prediction.
- Location-based analysis -- Maintenance anomalies occurring at particular locations within the building or across a complex of buildings.
- Error and Failure-based analysis -- What kind of errors and failures are mostly reported and corrected.
- Temporal and Seasonal analysis – Frequency of failures, occurrence and changes over time, i.e., specific months/ weeks/ days.
- Customer and User satisfaction-based analysis – User satisfaction of facilities maintenance performance and conditions.

8.1.3 Requirements and Problems of Maintenance Data Collection

During the data acquisition stage, special attention must be paid to data accuracy and quality. Most data collected for a building comes from site managers, maintenance officers and repairmen. In the Industry data provided neither the type, nor scale of a maintenance operation relates to the relative accuracy or precision of the collected data. Generally the data collection survey was composed of the following attributes:

- Work Order No,
- General Job Type,
- Job Priority,
- Floor,
- Room Number,
- Asset Number,
- General Repair Description
- Additional Text

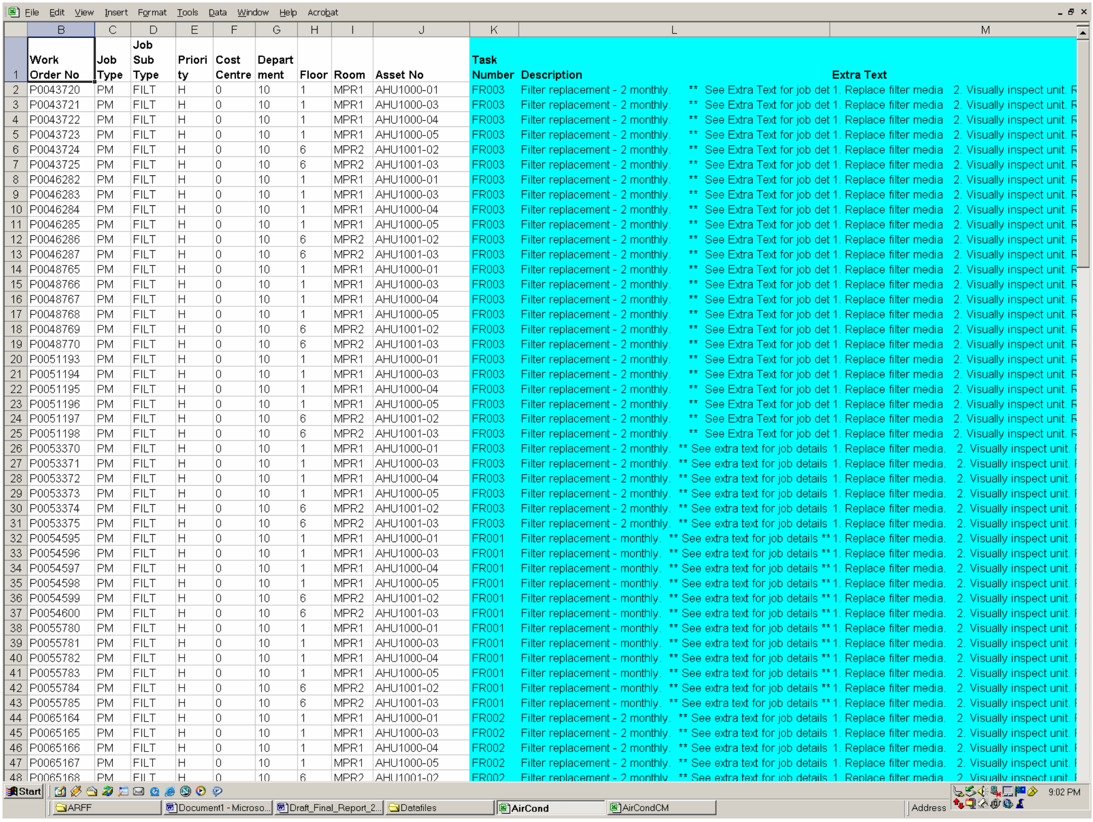
In many cases the “Additional Text” field is a preferred method of entry and often describes valuable information in written form. This kind of entry requires extensive pre-processing before it can be translated into meaningful data that can be used in data mining. The minimum data collection requirements that must be satisfied, in terms of both quantity and quality, and for any type (or scale) of corrective, scheduled or preventative maintenance operation include:

1. Element level maintenance data, involving a three layer tree coding consisting of:
    - WIC Number: a 5 digit number that uniquely identifies a building site
    - Building Code: a 3 digit number that identifies an individual building, except for the code 000, which identifies elements pertaining to the overall site
    - Building Component: a 4 character alphabetic code that identifies the type of system (air conditioning, electrical, etc.)
        - Element Number: a two digit number that uniquely identifies a specific asset
2. The cost of repair or replacement associated to each maintained building asset.
3. The cause of repair or replacement for each maintained building asset
4. Time required to fix or replace each maintained building asset

In particular, if maintenance data is collected about any kind of building component, without all element level data, its utility in data mining is limited. As discussed in the previous section, the degree of accuracy needed will also depend on the goals and objectives identified by facilities and building management.

In addition, data mining does not automatically extract all available knowledge that is embodied in a data set. For example, the industrial maintenance data supplied for Building 10 had to be pre-processed according to the following steps:

1. Removal of “noisy data”: This is maintenance data which contains false or constant values. Attributes such as “Task Number”, “Descriptions”, “Extra Text” which cannot be processed efficiently (highlighted in blue in Figure 57).



	Work Order No	Job Type	Job Sub Type	Priority	Cost Centre	Department	Floor	Room	Asset No	Task Number	Description	Extra Text
2	P0043720	PM	FILT	H	0	10	1	MPR1	AHU1000-01	FR003	Filter replacement - 2 monthly. ** See Extra Text for job det	1. Replace filter media 2. Visually inspect unit.
3	P0043721	PM	FILT	H	0	10	1	MPR1	AHU1000-03	FR003	Filter replacement - 2 monthly. ** See Extra Text for job det	1. Replace filter media 2. Visually inspect unit.
4	P0043722	PM	FILT	H	0	10	1	MPR1	AHU1000-04	FR003	Filter replacement - 2 monthly. ** See Extra Text for job det	1. Replace filter media 2. Visually inspect unit.
5	P0043723	PM	FILT	H	0	10	1	MPR1	AHU1000-05	FR003	Filter replacement - 2 monthly. ** See Extra Text for job det	1. Replace filter media 2. Visually inspect unit.
6	P0043724	PM	FILT	H	0	10	6	MPR2	AHU1001-02	FR003	Filter replacement - 2 monthly. ** See Extra Text for job det	1. Replace filter media 2. Visually inspect unit.
7	P0043725	PM	FILT	H	0	10	6	MPR2	AHU1001-03	FR003	Filter replacement - 2 monthly. ** See Extra Text for job det	1. Replace filter media 2. Visually inspect unit.
8	P0046282	PM	FILT	H	0	10	1	MPR1	AHU1000-01	FR003	Filter replacement - 2 monthly. ** See Extra Text for job det	1. Replace filter media 2. Visually inspect unit.
9	P0046283	PM	FILT	H	0	10	1	MPR1	AHU1000-03	FR003	Filter replacement - 2 monthly. ** See Extra Text for job det	1. Replace filter media 2. Visually inspect unit.
10	P0046284	PM	FILT	H	0	10	1	MPR1	AHU1000-04	FR003	Filter replacement - 2 monthly. ** See Extra Text for job det	1. Replace filter media 2. Visually inspect unit.
11	P0046285	PM	FILT	H	0	10	1	MPR1	AHU1000-05	FR003	Filter replacement - 2 monthly. ** See Extra Text for job det	1. Replace filter media 2. Visually inspect unit.
12	P0046286	PM	FILT	H	0	10	6	MPR2	AHU1001-02	FR003	Filter replacement - 2 monthly. ** See Extra Text for job det	1. Replace filter media 2. Visually inspect unit.
13	P0046287	PM	FILT	H	0	10	6	MPR2	AHU1001-03	FR003	Filter replacement - 2 monthly. ** See Extra Text for job det	1. Replace filter media 2. Visually inspect unit.
14	P0048765	PM	FILT	H	0	10	1	MPR1	AHU1000-01	FR003	Filter replacement - 2 monthly. ** See Extra Text for job det	1. Replace filter media 2. Visually inspect unit.
15	P0048766	PM	FILT	H	0	10	1	MPR1	AHU1000-03	FR003	Filter replacement - 2 monthly. ** See Extra Text for job det	1. Replace filter media 2. Visually inspect unit.
16	P0048767	PM	FILT	H	0	10	1	MPR1	AHU1000-04	FR003	Filter replacement - 2 monthly. ** See Extra Text for job det	1. Replace filter media 2. Visually inspect unit.
17	P0048768	PM	FILT	H	0	10	1	MPR1	AHU1000-05	FR003	Filter replacement - 2 monthly. ** See Extra Text for job det	1. Replace filter media 2. Visually inspect unit.
18	P0048769	PM	FILT	H	0	10	6	MPR2	AHU1001-02	FR003	Filter replacement - 2 monthly. ** See Extra Text for job det	1. Replace filter media 2. Visually inspect unit.
19	P0048770	PM	FILT	H	0	10	6	MPR2	AHU1001-03	FR003	Filter replacement - 2 monthly. ** See Extra Text for job det	1. Replace filter media 2. Visually inspect unit.
20	P0051193	PM	FILT	H	0	10	1	MPR1	AHU1000-01	FR003	Filter replacement - 2 monthly. ** See Extra Text for job det	1. Replace filter media 2. Visually inspect unit.
21	P0051194	PM	FILT	H	0	10	1	MPR1	AHU1000-03	FR003	Filter replacement - 2 monthly. ** See Extra Text for job det	1. Replace filter media 2. Visually inspect unit.
22	P0051195	PM	FILT	H	0	10	1	MPR1	AHU1000-04	FR003	Filter replacement - 2 monthly. ** See Extra Text for job det	1. Replace filter media 2. Visually inspect unit.
23	P0051196	PM	FILT	H	0	10	1	MPR1	AHU1000-05	FR003	Filter replacement - 2 monthly. ** See Extra Text for job det	1. Replace filter media 2. Visually inspect unit.
24	P0051197	PM	FILT	H	0	10	6	MPR2	AHU1001-02	FR003	Filter replacement - 2 monthly. ** See Extra Text for job det	1. Replace filter media 2. Visually inspect unit.
25	P0051198	PM	FILT	H	0	10	6	MPR2	AHU1001-03	FR003	Filter replacement - 2 monthly. ** See Extra Text for job det	1. Replace filter media 2. Visually inspect unit.
26	P0053370	PM	FILT	H	0	10	1	MPR1	AHU1000-01	FR003	Filter replacement - 2 monthly. ** See extra text for job details	1. Replace filter media. 2. Visually inspect unit.
27	P0053371	PM	FILT	H	0	10	1	MPR1	AHU1000-03	FR003	Filter replacement - 2 monthly. ** See extra text for job details	1. Replace filter media. 2. Visually inspect unit.
28	P0053372	PM	FILT	H	0	10	1	MPR1	AHU1000-04	FR003	Filter replacement - 2 monthly. ** See extra text for job details	1. Replace filter media. 2. Visually inspect unit.
29	P0053373	PM	FILT	H	0	10	1	MPR1	AHU1000-05	FR003	Filter replacement - 2 monthly. ** See extra text for job details	1. Replace filter media. 2. Visually inspect unit.
30	P0053374	PM	FILT	H	0	10	6	MPR2	AHU1001-02	FR003	Filter replacement - 2 monthly. ** See extra text for job details	1. Replace filter media. 2. Visually inspect unit.
31	P0053375	PM	FILT	H	0	10	6	MPR2	AHU1001-03	FR003	Filter replacement - 2 monthly. ** See extra text for job details	1. Replace filter media. 2. Visually inspect unit.
32	P0054595	PM	FILT	H	0	10	1	MPR1	AHU1000-01	FR001	Filter replacement - monthly. ** See extra text for job details **	1. Replace filter media. 2. Visually inspect unit.
33	P0054596	PM	FILT	H	0	10	1	MPR1	AHU1000-03	FR001	Filter replacement - monthly. ** See extra text for job details **	1. Replace filter media. 2. Visually inspect unit.
34	P0054597	PM	FILT	H	0	10	1	MPR1	AHU1000-04	FR001	Filter replacement - monthly. ** See extra text for job details **	1. Replace filter media. 2. Visually inspect unit.
35	P0054598	PM	FILT	H	0	10	1	MPR1	AHU1000-05	FR001	Filter replacement - monthly. ** See extra text for job details **	1. Replace filter media. 2. Visually inspect unit.
36	P0054599	PM	FILT	H	0	10	6	MPR2	AHU1001-02	FR001	Filter replacement - monthly. ** See extra text for job details **	1. Replace filter media. 2. Visually inspect unit.
37	P0054600	PM	FILT	H	0	10	6	MPR2	AHU1001-03	FR001	Filter replacement - monthly. ** See extra text for job details **	1. Replace filter media. 2. Visually inspect unit.
38	P0055780	PM	FILT	H	0	10	1	MPR1	AHU1000-01	FR001	Filter replacement - monthly. ** See extra text for job details **	1. Replace filter media. 2. Visually inspect unit.
39	P0055781	PM	FILT	H	0	10	1	MPR1	AHU1000-03	FR001	Filter replacement - monthly. ** See extra text for job details **	1. Replace filter media. 2. Visually inspect unit.
40	P0055782	PM	FILT	H	0	10	1	MPR1	AHU1000-04	FR001	Filter replacement - monthly. ** See extra text for job details **	1. Replace filter media. 2. Visually inspect unit.
41	P0055783	PM	FILT	H	0	10	1	MPR1	AHU1000-05	FR001	Filter replacement - monthly. ** See extra text for job details **	1. Replace filter media. 2. Visually inspect unit.
42	P0055784	PM	FILT	H	0	10	6	MPR2	AHU1001-02	FR001	Filter replacement - monthly. ** See extra text for job details **	1. Replace filter media. 2. Visually inspect unit.
43	P0055785	PM	FILT	H	0	10	6	MPR2	AHU1001-03	FR001	Filter replacement - monthly. ** See extra text for job details **	1. Replace filter media. 2. Visually inspect unit.
44	P0065164	PM	FILT	H	0	10	1	MPR1	AHU1000-01	FR002	Filter replacement - 2 monthly. ** See extra text for job details	1. Replace filter media. 2. Visually inspect unit.
45	P0065165	PM	FILT	H	0	10	1	MPR1	AHU1000-03	FR002	Filter replacement - 2 monthly. ** See extra text for job details	1. Replace filter media. 2. Visually inspect unit.
46	P0065166	PM	FILT	H	0	10	1	MPR1	AHU1000-04	FR002	Filter replacement - 2 monthly. ** See extra text for job details	1. Replace filter media. 2. Visually inspect unit.
47	P0065167	PM	FILT	H	0	10	1	MPR1	AHU1000-05	FR002	Filter replacement - 2 monthly. ** See extra text for job details	1. Replace filter media. 2. Visually inspect unit.
48	P0065168	PM	FILT	H	0	10	6	MPR2	AHU1001-02	FR002	Filter replacement - 2 monthly. ** See extra text for job details	1. Replace filter media. 2. Visually inspect unit.

Figure 57. Unprocessed Industry Data.

2. Re-interpretation of “noisy data”: useful information is extracted from the deleted attributes “Description”, “Extra Text” can be re-interpreted to form new attributes “causeofrepair”, “descriptionofCause” and “workordermaterial” (highlighted in red in Figure 58);

3. Derivation of new information: from an existing attribute new information can be obtained via interpretation or via the combination number of existing attributes new information can be extracted. For example, attribute “month” is derived from “Start Date” and “Completion Date”, “Completionwithinexpectations” is created from cross-comparing “Completion Date” and “Estimated Completion Date” (highlighted in blue in Figure 58);

4. Re-formatting of data: The maintenance data received was in Excel file format. The Excel file must be converted into ARFF file format to be processed by the machine learning algorithms used in the AIMM system.

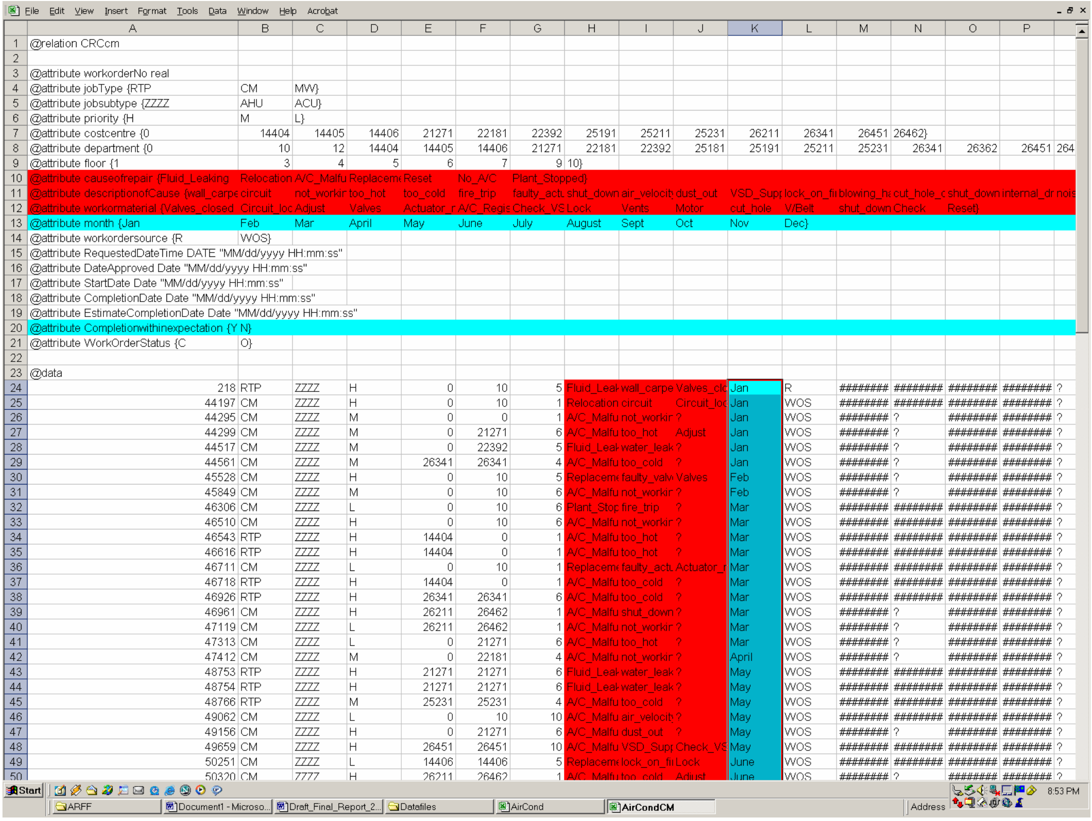


Figure 58. Processed Industry Data (Re-interpreted and Re-formatted).

Crucially, the following data was missing from the existing maintenance data provided by Industry:

- Cost related data – no information was provided of how many resources were involved for each maintenance job;
- Human resources – personnel information is required in order to evaluate work performance;
- Failure information – detailed failure information and related analysis is required for each corrective maintenance job;
- User satisfaction – information on user satisfaction can be derived from some pre-processing techniques, however more detailed information is required to obtain more meaningful analysis.

8.1.4 Importing and Expanding Maintenance Data

Future development of the AIMM system must also allow users to: (i) easily incorporate (import) data from outside sources, (ii) easily update and alter data, and (iii) ask data-related questions of (or query) the database. Since most database management systems such as the EDM system incorporated within AIMM, provide these capabilities, they should also be available through the AIMM systems interface.

Since data must be imported and reformatted from outside sources, this process should be made available through the AIMM system. Currently, data imported into the AIMM system must come in ARFF file format which is similar to CSV format. Only an ARFF file format can be processed by the machine learning algorithms used in the AIMM system. Since maintenance databases often come in the form of SQL, Excel and other similar file formats the AIMM system should provide a process for importing and re-formatting a number of these formats into ARFF files automatically.

A common task in facilities management is updating or editing the database. Since no user can foresee all future data needs and applications, the AIMM system should incorporate ways to easily modify, refine, or correct the database. The AIMM system should also allow for error checking as new maintenance data records are created or existing ones are updated. Not all errors can be eliminated in this way, however, so care must be taken when collecting, automating, and changing the database. Attribute data are seldom static and therefore, maintain the value of the data depends on updating capabilities.

8.1.5 Requirements of Data Mining Algorithms

The data mining algorithms used in the AIMM system have specific data type and format requirements. All algorithms receive data in ARFF file format and the following outlines the general requirements of each algorithm implemented in AIMM:

- **Classification Algorithm: ID3** – Prior to applying the ID3 algorithm on maintenance data, all necessary data must be converted into nominal values and all the missing values must be replaced with “N/A” strings.
- **Classification Algorithm: C4.5** – Although the C4.5 algorithm allows missing values, prior to applying C4.5 algorithm on maintenance data, missing values should be replaced with "?" symbols. The C4.5 algorithm can be applied to both nominal and numeric values.
- **Clustering Algorithm: SimpleKmeans** – Like the C4.5 algorithm, SimpleKmeans allows missing values and prior to applying SimpleKmeans algorithm on maintenance data, missing values should be replaced with "?" symbols. The SimpleKmeans algorithm can also be applied to both nominal and numeric values.
- **Associative Rule Algorithm: Apriori** – Prior to mining the maintenance data using the Apriori algorithm, the data file must be divided into natural associated attribute groups that can be further divided into sets of new data files to allow associations to be found. Like C4.5 and SimpleKmeans, the Apriori algorithm takes nominal value and allows missing attribute value (to replaced with "?" symbols). In addition, an attribute evaluator “CfsSubsetEval” must be applied and the “BestFirst” search method must also be used.

8.2 Requirements for 3D CAD Modelling

The 3D CAD model of a building and its asset should be modelled as object oriented for each component and be transferable to IFCs. The principle benefit of IFC’s is their object description. The IFC protocol preserves the full 3D geometric description in 3D and distinguishes its location and relationships, as well as all the properties and parameter values of each element. This provides access to accurate geometry of building systems, components, structural elements and properties relevant to facilities maintenance requirements.

Since the process of data mining in a 3D virtual environment requires an adequate building model to begin with it is therefore necessary for architects and designers to understand how to better structure building models for data mining a variety of scenarios since information provided within CAD and IFC element descriptions are inadequate for many facilities maintenance descriptions and requirements. It is necessary to extend design information in object oriented CAD systems to support descriptions defined in facilities maintenance operations.

8.3 Potential Industry Partners Survey

The potential benefits gained from the AIMM prototype system have prompted a preliminary survey of large commercial businesses and industries in order to gauge interest levels of such a tool. A flyer was produced in order to present the project to Industry partners and outline participation areas, (refer to Appendix C). A variety of potential industry partners were contacted including:

- Sydney University, Facilities Management Office (FMO);
- Woolworths Pty Ltd, Facilities Management;
- AMP Capital: Commercial & Industrial Management Pty Ltd.;
- Resolve FM Pty Ltd, (Contact: John Smith).

In the initial round of discussions, feedback from the above businesses and industries was very positive. The meeting outline included a presentation of the project, research completed to date, current research maintenance data requirements, and an illustration of the benefits to industry partners. In all preliminary meetings the opportunity for active project involvement was presented. A summary of the responses taken from the minutes recorded at all four meetings are shown below.

Sydney University, Facilities Management Office

The Facilities Management Office (FMO) is responsible for the planning, development, maintenance and operations of the University's Facilities. FMO carries out this responsibility through its three operational groups: Facilities Strategic Planning, Project Services and Facilities Services. The Maintenance Services Group, which is part of the Facilities Management Office (FMO) is responsible for the maintenance of all University buildings and property. FMO manage over approximately 100 buildings and facilities on the main Camperdown and Darlington campuses.

Services are conducted using in-house maintenance personnel, as well as authorised contractors and are carried out across a variety of disciplines including: building, refrigeration, electrical, plumbing, roofing and essential services (fire safety). Requests for maintenance are made through the FMO Service Desk or lodged electronically by utilising the e~Service Desk facility.

The Facilities manager of FMO, Steve Bedford approved the exchange of facilities maintenance information and expressed interest in further project involvement. FMO's maintenance database, the information available, its exact history and form were presented. Available facilities maintenance data for FMO Sydney University can be divided into two general types: *breakdown maintenance data* and *preventative maintenance data*.

Preventative maintenance data:

- Mostly electrical and mechanical building data,
- FMO has a comprehensive database of preventative maintenance,
- History of database: approx. four years,
- Most buildings with maintenance data have 2D CAAD models.

Breakdown Maintenance:

- General building data such as painting re-surfacing, replacement of fittings etc.
- Incomplete database of breakdown maintenance,
- History of database: unknown,
- Some buildings with maintenance data have 2D CAAD models.

In addition, Gerard Gabriel, the Facilities Information Manager, also participated in discussions and provided all relevant maintenance information for a variety of University buildings as well as 2 and 3D CAD models.

Woolworths, Facilities Management

Woolworths operate more than 1400 stores in Australia, plus an additional 33 Dick Smith Electronics stores in New Zealand. Major store names include: Woolworths, Safeway, Food For Less, Woolworths Metro, Dick Smith Electronics and Power House, Tandy, Woolworths Liquor, BWS, First Estate, Dan Murphy's, Plus Petrol, BIG W, Australian Independent Wholesalers (AIW), Woolworths Ezy Banking, Woolworths Home Shop and Green Grocer. Supermarkets are the largest division with stores in all states. Woolworths has over 1000 supermarket sites and is comprised of approximately:

- 700 Supermarkets
- 300 Petrol Stations
- 100 (Unattached) Liquor Outlets.

Woolworths does lease the majority of properties to companies such as Westfield, however they do own and manage a large number of shopping centre properties. Management of property space is generally preventative and/or proactive maintenance and asset management, where major maintenance and refurbishment is undertaken approximately every three years. Corrective maintenance information is available where properties are owned by Woolworths.

Data collection has improved in recent years and potentially presents a high standard of maintenance databases for both refurbishment and general building-type maintenance. Woolworths has over the past 2 years implemented the MAXIMO® 5 Asset Management as their Facilities Management System. MAXIMO provides a set of solutions to meet their strategic asset management needs (for more information, refer to review in Section 3.6.3.). MAXIMO is utilised by Woolworths for asset management, work management, materials management and purchasing capabilities. Comprehensive key performance indicators (KPIs), reporting and analytic capabilities help the Woolworths facility managers track maintenance activities and identify areas that require improvement. Woolworths has approximately 3 500 maintenance jobs raised weekly and 200 000 jobs annually. Aspects significant to maintenance data type, collection and value are:

- System management is not split by asset type.
- Some remaining issues in maintenance reporting which are as yet to be resolved within the system. Thus, there is not complete confidence with the system. It was suggested that by June 2004 all operating issues will be resolved and accurate maintenance data will be thereafter available.

Generally, Woolworths is conservative about involvement with other parties. Previously Woolworths has declined financial involvement in external projects. Their involvement would therefore depend on the size and type of commitment required.

AMP Capital

AMP Capital's Commercial & Industrial Management division has approximately $10 billion in property assets under management, and is one of the largest wholesale property managers in the Asia Pacific region. There are currently 34 commercial buildings directly under AMP Capital's Commercial & Industrial Management, 20 of which are located within the Central Sydney Region. There is $87 million in controllable expenditure, including energy attached to the Central Sydney Region commercial and industrial management.

AMP Capital property portfolio represents over 2 million square metres, and includes the management of a number of Sydney's premier commercial and industrial buildings such as Angel Place in the Sydney CBD, and Rydalmere Metro Centre in Sydney's inner-west. Other commercial and industrial buildings include: National Australia Bank House, Sydney; Coronation Drive Office Park, Brisbane; Jessie Street Centre, Parramatta; Gold Fields House, Sydney; Philips House, Sydney; NRMA House, Sydney; Sydney Cove, Sydney; and AMP Centre, Sydney.

AMP Capital controls property in a multi-asset class environment, adopting a core and satellite approach to property management. Commercial and Industrial Management Pty Ltd provides property management expertise for the majority of AMP Capital Investors' Australian property portfolio on behalf of the investment and superannuation funds and private clients which own the properties.

AMP Capital's Commercial and Industrial Management provides active asset management by: providing independent leasing advice; actively monitoring and reducing vacancies; and reducing operating costs and increasing all building revenues.

Generally, AMP Capital's Commercial and Industrial Management oversees planned maintenance and maintain a general philosophy that resides in "Tenancy is King" and therefore policies tend more toward "Do and Change" rather than any life cycle costing for asset management, since the leasing structure affects maintenance contracts. AMP Capital's maintenance system comprises a maintenance plan, asset registers and a plan regime. Works under $500 are not registered by the system. However AMP Capital's Commercial and Industrial Management does not directly control its properties facilities management. Instead facilities management operations are contracted to Resolve FM.

Since AMP Capital's Commercial and Industrial Management outsource their facilities management to Resolve FM they do not "own" the maintenance data required for the project. AMP Capital is enthusiastic about the future development of the AIMM system and continued involvement in the project has been proposed. AMP expressed a willingness to contribute time in kind as well as resources, however a financial contribution has been declined this financial year. An understanding was reached to approach Resolve FM in a tri-partite agreement with CRC in order to secure both parties' involvement in the project.

Resolve FM

Resolve FM are a management company which provides professional facilities management to large property and infrastructure assets across many market sectors including commercial, industrial, government, education, residential, defence, retail, health and tourism.

Resolve FM use the MAXIMO® 5 Asset Management as their Facilities Management System (refer to Section 3.6.3) and offer the following services to AMP Capital: Building System Infrastructure, Support Services, General Maintenance, High Tech Operations & Maintenance, Facility/Property Management and Strategic Facility Solutions. Of particular relevance to the development of this research are the following services and information derived from:

High Tech Operations & Maintenance

- Critical Facilities
- Data Centres
- Predictive Maintenance
- Computerised Maintenance Management Systems
- HVAC Systems
- Electrical Systems
- Mechanical Systems

Facility/ Property Management

- Supply Chain Management
- Financial Budget Management
- Project Management
- Lease Management
- Utility Management
- Furniture and Office Equipment Management
- Benchmarking
- Capital Planning

Building System Infrastructure

- Building Management Systems
- Intelligent Fire Alarm Systems
- Systems Integration
- System Installation, Retrofit, Upgrade & Commissioning

Support Services

- Cleaning
- Landscaping
- Hygiene
- Security

General Maintenance

- Plumbing
- Carpentry
- Handyman Services
- Low Tech Mechanical & Electrical Services
- Fabric Maintenance

Facility Strategy Solutions

- Program Management
- Occupancy Planning
- Workplace Design
- Critical Facilities
- Energy Management Consulting
- Facility Management Consulting
- Guaranteed Infrastructure Services

Resolve FM agreed in principle to potential involvement in future project development. Participation and contributions were outlined to include: resources such as maintenance data and computing facilities; and time in kind, with Resolve FM having highly skilled staff. However a financial contribution has been declined this financial year. An understanding was reached that the tri-partite agreement with AMP Capital and CRC was a means favourable of securing their future involvement.

8.4 Integrating Life Cycle Cost Analysis in AIMM

Life cycle costs (LCC) are summations of cost estimates from inception to disposal for both equipment and projects as determined by an analytical study and estimate of total costs experienced during their life. LCC comprises initial acquisition cost and also may contain other costs such as operation costs, maintenance costs, logistics costs. Such costs are usually higher than the original acquisition cost. The major objective of LCC analysis is to choose the most cost effective approach from a series of alternatives so the least long term cost of ownership is achieved (Barringer, 1996). Life cycle cost analysis provides strategic planning on refurbishment and enhances information for decision making. LCC helps facility managers in evaluating alternative equipment and process selection based on total costs rather than the initial purchase price. The multidimensional information that LCC presents is merged from hybrid project domains such as management, engineering, as well as finance.

There are various existing life cycle models available for buildings as a whole and for their component systems that demonstrate the multifarious approaches to LCC. Although there is no one model that have been accepted as a standard, there are some areas of commonality. Life cycle cost models form predictions based on several parameters, some of which include a degree of uncertainty, such as the reliability of a part (Siewiorek and Swarz, 1982). Among the inputs whose values could potentially be predicted for each component by the Data Mining system in building LCC (Dhillon, 1989) are:

- mean time between failure,
- mean time to repair,
- average materials cost per repair,
- labour cost per corrective maintenance action,
- average materials cost per preventative maintenance action,
- labour cost per preventative maintenance action,
- spares requirements.

Among the inputs that can potentially be entered as part of the domain knowledge are:

- system listed price,
- warranty period,
- cost of installation,
- cost of component downtime,
- cost of carrying spares in inventory.

The values of input variables, along with their probability distributions, can be predicted for each component, thus allowing for more accurate estimation of average life cycle cost. By predicting failure rates and repair costs, it is possible to compute the optimal schedule of preventative maintenance for each asset. Existing life cycle modelling systems fail to provide a seamless integration of hybrid information that provides users access to previously unreachable knowledge. The proposed building life cycle cost formula to be utilised is expressed by:

$$\boldsymbol{BLCC = \sum c\ LCCc} \quad (1)$$

where:

LCCc	is the life cycle cost of component,
c	is the component

The LCCc, is expressed by:

$$\boldsymbol{LCC = AC_c + DC_c + \sum_{i=1}^{8} (OC_{c\,i} + SMC_{c\,i} + CMC_{c\,i})} \quad (2)$$

where:

ACc	is the acquisition cost of component c
DCc	is the disposal cost of component c
OCci	is the operating cost of component c for year i

SMC_{Ci}	is the scheduled maintenance cost of component c for year i
CMC_{Ci}	is the corrective maintenance cost of component c for year i

The SMC_{Ci} is expressed by:

$$SMC = \sum SMF\ [\ MSMT\ (SMPC + SMLC)] \qquad (3)$$

where:

SMF	is the scheduled maintenance frequency (per year)
SMPC	is the scheduled maintenance parts cost
SMLC	is the scheduled maintenance labour cost (per hour)
MSMT	is the mean scheduled maintenance time (in hours)

The CMC_{Ci} is expressed by:

$$CMC = \sum CAF \left[\frac{SLT}{MTBF} \right] \qquad (4)$$

where:

SLT	is the system's life time (expected)
MTBF	is the mean time between failures
CAF	is the cost of average failure

The CAF is expressed by:

$$CAF = \sum CMPC + [\ (CMLC + LSC) * (MRT + MSRT)\ ] \qquad (5)$$

where:

CMPC	is the corrective maintenance parts cost
CMLC	is the corrective maintenance labour cost (per hour)
LSC	is the loss of service cost (per hour)
MRT	is the mean repair time (in hours)
MSRT	is the mean system response time (in hours)

The proposed life LCC formula will be integrated in the system prototype of AIMM^{TM} via a new Life Cycle Cost (or LCC) Agent. The architecture of the AIMM^{TM} system has been developed so as to include the LCC Agent within the existing maintenance analysis component as shown in Figure 59.

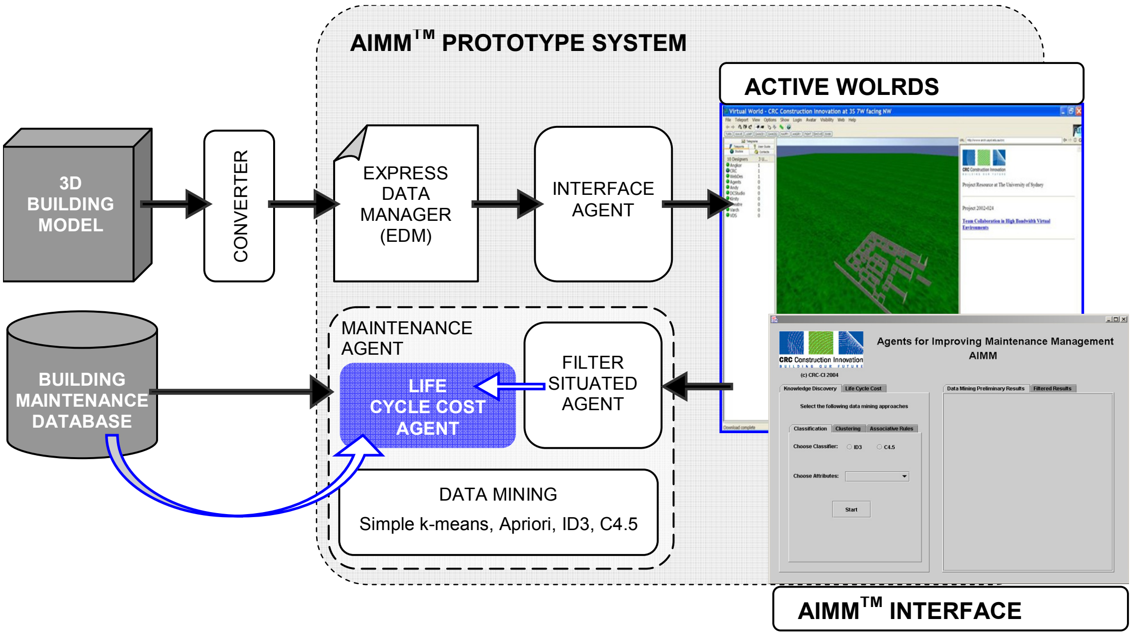


Figure 59. Integrating a new LCC Agent in the architecture of AIMM.

The LCC Agent utilises the results available produced by the Situated Agent and accesses the maintenance database to extract the input variables required in the LCC formula. Figure 60 illustrates the proposed LCC Panel and how it can be integrated with the AIMM prototype system's existing interface.

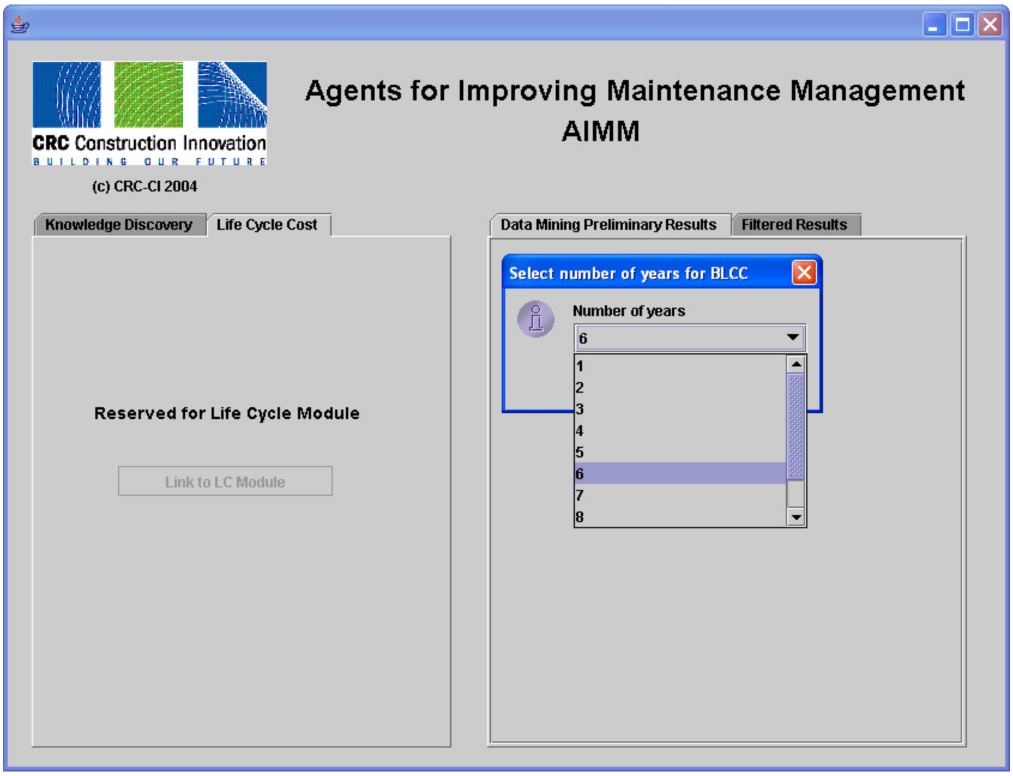


Figure 60. Integrating a new LCC Panel into the AIMM interface.

Some of the possible input variables with the source from where a value may be acquired are shown in Table 18

Table 18. Examples input variables with the source from where a value may be acquired.

Input Variable	Source
Acquisition Cost	Asset Register
Operating Cost (per year) Includes energy consumption	Manufacturer
Scheduled Maintenance Frequency (per year)	Maintenance Database
Scheduled Maintenance Parts Cost	Structural, Mechanical, Electrical Engineers
Scheduled Maintenance Labour Cost (per hour)	Structural, Mechanical, Electrical Engineers
Mean Scheduled Maintenance Time (in hours)	Maintenance Database
System's Life Time (expected)	Manufacturer
Mean Time Between Failures	Maintenance Database
Corrective Maintenance Parts Cost	Structural, Mechanical, Electrical Engineers
Corrective Maintenance Labour Cost (per hour)	Structural, Mechanical, Electrical Engineers
Loss of Service Cost (per hour)	Structural, Mechanical, Electrical Engineers
Mean Repair Time (in hours)	Maintenance Database
Mean System Response Time (in hours)	Maintenance Database
Real Discount Rate	Government or financial institution

Thus, Life cycle cost (LCC) analysis may be introduced to provide strategic planning on refurbishment and enhances information for decision making. The life cycle cost formulas proposed here for building assets and components are appropriate for adoption within the AIMM™ systems architecture and an LCC Agent be implemented may in future research.

8.5 Recommendations for the CRC

The construction industry has adapted the information technology in its processes in terms of computer aided design and drafting, construction documentation and maintenance. Hence, the data generated within the construction industry has become increasingly overwhelming. The growth of many business, government, and scientific databases has begun to far outpace human's ability to interpret and digest this data. This issue becomes critical when the high degree of complexity of work flow is taken into account in the decision making process during the lifetime of a building. Furthermore, past experience often plays an important role in building management. Therefore, applying data analytic techniques to efficiently deal with information at different stages of a building life cycle has great potential.

A large number of buildings maintained in Australia rely on efficient facilities management. Maintaining building facilities is a major task for many in the AEC industry, ranging from the facilities manager to the occupant and client to the designer. Table 19 illustrates this range of professions and stakeholders that can potentially benefit from the system demonstrated by this research project and the life cycle phase that would be influenced in terms of improving standards and best practices.

Table 19. Potential users and beneficiaries and related life cycle phases.

	Stakeholders and Users	Life Cycle Phase
Users: Immediate and potential	Facilities and Asset Managers	Management in-use
	Asset Planners	Strategic Planning
	Facility and Building Operators	Occupancy in-use
	Architects and Designers	Design Life Cycle
Beneficiaries: Direct and indirect	Occupants, Tenants, Owners	Post Occupancy
	Investors	Asset Investment
	Building Developers	Building Development
	Environment: Water, Material, Energy, etc.	Ongoing
Flow-on effects to AEC industry	Consultants: Engineers, Researchers, etc.	In-use Operations
	Building and Quantity Surveyors	Procurement
	Builders	Project Delivery
	Manufacturers, Environmental Control Bodies	Procurement
	Project Managers	Construction

Facilities Managers will benefit from the development of this project in terms of the feed back generated from identifying patterns and correlations in maintenance records that are presented as meaningful asset knowledge. Since integrating informed design decisions for improving building life cycle is a complex task at the early stages of designs, designers will benefit from this project in reducing the risk of design mistakes. Both Facilities Managers and Designers will benefit from the automated feed back in reducing time and cost. Further, the project can be applied in other potential areas - including checking a building model against the information required for project management, and cost estimates as detailed in Section 8.4.

This project has developed a prototype to support automated feed back. Therefore future research should look towards applications of the prototype for Industry partners. For continued AIMM^{TM} systems development, a needs-based assessment is perhaps the most important step. Discussions about the systems requirements, an understanding of who are the users and what they demand of the system and an evaluation of the educational and training needs of users are all part of such a needs-based assessment. This process will ensure that proper data sets are collected, that all users understand the technology and its role within the organization, and that the specific goals of maintenance management are identified to generate the proper analyses and output. After a discussion of industry scenarios, and establishment of goals and requirements, work towards education, training, and data compilation (i.e., acquiring and digitising proper data sets) can begin. Based on this approach, valued industry and research benefits can be achieved from further project development including:

- Contribution to long-term scientific and technological research and innovation to Australia's sustainable economic and social development.
- Collaboration between researchers, industry and government, and to improve efficiency in the use of intellectual and research resources.
- Creation of a fully commercialised tool to deliver innovative and sustainable constructed assets to further the financial, environmental and social benefit to the construction industry and the community.

9. CONCLUSION

The development of data mining agents of facilities and building maintenance data in a 3D virtual environment provides useful information for improving the design, maintenance and management of building facilities and guiding future decisions. Virtual environments of building models offer the opportunity for the user to navigate through the model, to manipulate and to interact with its objects. The integration of facilities databases with interactive 3D virtual environments containing building models and data mining techniques provides a visual modelling tool for the simulation and projection of the financial and physical impact of maintenance, refurbishment and major replacement and extension of a building and its components over its life cycle.

A sound and useful architecture and scenario of using software agents for data mining on building maintenance has been developed and implemented in an AIMM™ system prototype. The usefulness of the architecture and scenario lies in the potential of using AIMM™ to maximise the benefits of its application without requiring its users to have an extensive training to use the system or to have a comprehensive understanding of data mining and its techniques. Users of AIMM™ such as facility managers, designers and building developers should be reflecting on the discovered knowledge acquired from the application of data mining on maintenance records and utilise AIMM results to improve overall maintenance management as well as the operation of buildings and their life cycle.

By demonstrating how and what data mining techniques can be applied on maintenance data of buildings this research has discovered and addressed patterns, correlations and useful rules within existing building maintenance data. The results from the initial data mining studies were improved in the final experiments of the prototype system by applying the appropriate data mining scenarios and filters needed to provide the required knowledge. Appropriate filters chosen according to the task were proposed and with future development can be extended further to include a variety of scenarios. The techniques within the tool demonstrate how a building or facility manager can identify meaningful and useful patterns, correlations and trends of knowledge from large amount of building data.

Further, this final report provides results supporting the capabilities, flexibilities and advantages of automated knowledge development that is a result of data mining building models in a virtual environment. Evaluation and testing has highlighted how the connection can be improved between maintenance and design knowledge development in the following ways:

- The combination of a 3D model with maintenance and other asset data facilitates the ability of building designers and owners to visually model the impact of design, maintenance, refurbishment and extension decisions on the building's life cycle cost.
- The linking of a 3D model with maintenance data allows both the facility manager and the designer to gain access to information and knowledge that is currently inaccessible.
- The development of agents for data mining of facilities maintenance and other data provides a method of testing and validating the usefulness and scope of current databases as a platform for guiding future decisions.
- The representation of the facility within the virtual environment provides a basis for linking data mining with emerging technologies (such as connecting to WAP phones and other PDAs both in the office and on site) to address a gap in the construction life cycle.
- The integration of data mining agents into the maintenance process produces timely data for the facility manager on the effects of different maintenance regimes.

- The development of agents (agents are software systems capable of taking autonomous decisions) for data mining of facilities maintenance data in a 3D environment provides proactive information for improving the design, maintenance and management of building facilities.

Applying the AIMM prototype system to the records of facilities provides a tool that is capable of improving the management and maintenance of existing facilities as well as provide valuable knowledge for the design of new facilities. This leads to more efficient and effective facilities maintenance and management through better planning based on models developed from available maintenance data, and therefore results in a more economical life cycle of buildings.

Although the project focuses on mining the maintenance data in which huge benefits go to the facility management, it is not prevented from attempting to fill the gaps between designing and building maintenance in biasing the future design solutions within the scope of the whole coordinated building life cycle. The major contribution of data mining for this project is to provide a knowledge base which is served as a centre bridge. Furthermore, designers and maintenance managers will be better equipped to achieve higher performance by utilising the techniques incorporated in AIMM in their workplace.

For the AIMM™ prototype system to be useful in the AEC industry, there is a set of minimum requirements of building maintenance data that must be satisfied, in terms of both quantity and especially quality (Refer to Section 8.1)

The fundamental directions of expected commercialization potentials of AIMM prototype system involve: (i) enhancing and expediting the traditional maintenance management activities by live and progressive feedback using data mining techniques to improve management of building maintenance; (ii) opening new possibilities in building design and management by exploring 3D multi-user, online, real-time environments; and (iii) providing interface to generally available maintenance databases and 3D CAD software (seamless integration of hybrid information).

Future work of the AIMM prototype system includes comprehensive consultation with Industry experts and the integration of life cycle modelling at the early design stages. The potential benefits of this approach include: (i) recommending quality systems (building assets) which meet users expectations within cost estimates; (ii) recommending systems that are cost-effective to enhance and maintain; (iii) recommending the exploration of alternative concepts and methods to satisfy the need; and (iv) evaluating costs and benefits of alternative approaches to satisfy the basic functional requirements.

10. REFERENCES

Agrawal, R., Srikant., R.: 1995, Mining sequential patterns. *Eleventh International Conference on Data Engineering*, Taipei, Taiwan, 3-14.

Arciszewski, T. and Ziarko, W.: 1987, A methodology of design knowledge acquisition for use in learning expert systems. *International Journal of Man Machine Studies* 27(1):23-32

Arditi & Gunaydin.: 1998, Factors that affect process quality in the life cycle of building projects, *Journal of Construction Engineering and Management*, 124 (3).

Barringer, H. Paul and Weber, David P.: 1996, Life Cycle Cost Tutorial, *Fifth International Conference on Process Plant Reliability*, Gulf Publishing Company, Houston, TX

BOMA.: 1998, *Recommended Chart of Accounts for Commercial Industrial and Retail Properties.* Building Owners and Managers Association of Australia, Sydney.

Bowers, S. V., Piety, K.R. Proactive motor monitoring through temperature, shaft current and magnetic flux measurements. *White Paper*, Computational Systems, Inc., Knoxville, TN.

Brachman.: 1997, Visual data mining: Recognizing telephone calling fraud. *Data Mining and Knowledge Discovery,* 1, 33–46.

Burati,: 1992, Causes of quality deviations in design and construction, J*ournal of Construction Engineering and Management,* 118(1).

Cabena et. al.: 1998, *Discovering Data Mining from Concept to Implementation*, International Technical Support Organization (IBM), Prentice Hall PTR.

Christiansson, P.: 1998, Using knowledge nodes for knowledge discovery and data mining, *Artificial Intelligence in Structural Engineering*, Information Technology for Design, Collaboration, Maintenence, and Monitoring, I. Smith, ed., Springer-Verlag, Berlin, 48-59.

Cornick, T.: 1996, *Computer-Integrated Building Design,* E & FN Spon, London.

Davis et al.: 1989, Measuring design and construction quality costs, *Journal of Construction Engineering and Management,* 115(3).

Dangel, R.: 1969, Integrated logistic support (ILS) implementation in the Naval Ship System Command, *ASE 6th annual technical symposium*, pp. 1-25.

Dhillon, B. S.: 1989, *Life Cycle Costing: Techniques, Models, and Applications*, Gordon & Breach Science Publishers, New York.

Ding L., Liew P., Maher L., Gero J., Drogemuller, R.: 2003, *Integrating CAD and 3D Virtual Worlds Using Agents and EDM*, CMIT, Commonwealth Scientific and Industrial Research Organisation, Key Centre of Design Computing and Cognition, University of Sydney.

Dix A. and Ellis, G.: 1998, Starting Simple - adding value to static visualisation through simple interaction, *Proceedings of Advanced Visual Interfaces AVI98*, Eds. T. Catarci, M. F. Costabile, G. Santucci and L. Tarantino. L'Aquila, Italy, ACM Press. pp. 124-134.

Elder, J., Abbott, D.: 1998, *A Comparison of Leading Data Mining Tools,* tutorial presented at KDD-98, New York, August 28, 1998.

Fischer, M., Kam, C.: 2002, *PM4D Final Report*, CIFE (Center for Integrated Facility Engineering) Technical Report Number 143, Stanford University, Stanford.

Frasconi, P., Soda, G., Vullo, A.: 2002, Hidden Markov models for text categorization in multi-page documents, *Journal of Intelligent Information Systems*, **18**:2/3, 195-217.

Frawley, W., Piatetsky-Shapiro, G., Matheus, C.: 1992, Knowledge discovery in databases: An overview. *AI Magazine,* 57-70.

Froese T. M., Lloyd Waugh, L. M., and Pouria, A., 2001, Project management in the year 2020, Proceedings of 2001 Conference of the Canadian Society for Civil Engineers, Victoria, BC., May 30-Jun 2, 2001. Paper C15.

Goebel, M. and Gruenwald, L.: 1999, A survey of data mining software tools, *SIGKDD Explorations*, 1(1).

Gero, JS and Fujii, H.: 2000, A computational framework for concept formation in a situated design agent, *in Knowledge-Based Systems* **13**(6): 361-368.

Han, J., and Kamber, M.: 2001, *Data Mining: Concepts and Techniques*, Morgan Kaufmann Publishers, San Francisco.

Hui, S.C. and Jha, G.: 2000, Data Mining for Customer Service Support, *Information & Management*, 38, pp.1-13.

Ingirige, B., Aouad, G. and Ming Sun, M.: 2001, Awareness of information standardisation in the UK construction industry: A Survey by the SIENE Network, *Proceedings of International Conference on IT in Construction in Africa,* 30 May – 1 June 2001, Mpumalanga, South Africa.

Julien, B., Fenves, S., Small, M.: 1992, Knowledge acquisition methods for environmental evaluation. *AI Applications* 6(1):1-20

Kawauchi Y., Rausand, M. (1999) *Life Cycle Cost (LCC) Analysis in Oil and Chemical Process Industries,* adopted from http://www.ipk.ntnu.no/rams/notater/lcc.doc on 28th, Mar., 2003.

Kohavi, R., and John, G., 1997: Wrappers for feature subset selection. *Artificial Intelligence*, 97, 273–324.

Kumar, V. and Joshi, M. (1999) *Tutorial on High Performance Data Mining,* Department of computer science, University of Minnesota, Minneapolis, USA.

Kunz, J., Jin, Y., Levitt, R. Lin, S.D., Teicholz, P.: 1995, The intelligent real-time maintenance management (IRTMM) system: Support for integrated value-based maintenance planning."*CIFE Technical Report 100*, Stanford, CA.

Lewis, R. and Sequin, C.: 1998, Generation of 3D building models from 2D architectural plans. *Computer-Aided Design*, **30** (10), 765-79.

Liew, PS and Gero, JS: 2002a, An implementation model of constructive memory for a situated design agent, *in* JS Gero and F Brazier (eds.), *Agents in Design 2002,* Key Centre of Design Computing and Cognition, University of Sydney, Australia, pp. 257-276.

Liew, PS and Gero, JS: 2002b, A memory system for a situated design agent based on constructive memory, *in* A Eshaq, et al. (eds.), *CAADRIA2002,* Prentice Hall, New York, pp. 199-206.

Linnert, C. (1999, Diploma thesis, *Virtual Building Lifecycle -- Visualization of Lifecycle Data on a Virtual Model of a Building in a 3D-CAD Environment,* Institute for Industrial building production, University of Karlsruhe, Germany.

McLaughlin, S. and Gero, J. S.: 1987, Acquiring expert knowledge from characterised designs, *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, **1**(2): 73-87.

Maher, M.L. and Gero, J.S.: 2002, Agent models of virtual worlds, *Proceedings of ACADIA 2002*, Pamona, CA.

Maher, ML, Liew, PS and Gero, JS: 2003, An agent approach to data sharing in virtual worlds and CAD, *in* R Amor (ed.), *Construction IT Bridging the Distance, CIB Report 284*, pp. 204-213.

Menezes, M.: 2000, Improving plant safety & availability through advanced measurement diagnostics. *Instrumentation Systems and Automation Society Conference*, New Orleans, LA.

Morbitzer, C., Strachan, P., Simpson, C.: 2003, Application of data mining techniques for building simulation performance prediction analysis*, in the Proceedings of Building Simulation 2003*.

Murray, N., Fernando, T., Aouad, G.: 2000, A virtual environment for building construction ,*17th ISARC*, 1137-1142, 18-20th September.

Peitgen, H. O., Jurgens, H., and Saupe, D.: 1992, *Chaos and Fractals : New Frontiers of Science*, Springer-Verlag, New York.

Pyle, D.: 1999, *Data Preparation for Data Mining*, Morgan Kaufmann, San Francisco.

Quinlan, R.:1993, *C 4.5 Programs for Machine Learning*, Morgan Kaufmann, San Francisco.

Reich, Y., Konda, S., Levy, S. N., Monarch, I., and Subrahmanian, E.: 1993, New roles for machine learning in design, *Artificial Intelligence in Engineering*, 8(3):165-181.

Reich, Y., Fenves, S. J., and Subrahmanian, E.: 1994, Flexible extraction of practical knowledge from bridge databases, *Proceedings of the First Congress on Computing in Civil Engineering*, Washington, DC, pp. 1014-1021.

Reich, Y., Fenves, S. J.: 1995, A system that learns to design cable stayed bridges. *Journal of Structural Engineering*, ASCE.

Reffat, R.: 2002, Three-dimensional CAD Models: Integrating Design and Construction, *in* R. Best and G. de Valence (eds), *Innovation in Design and Construction: Building in Value*, Butterworth Heinemann, Oxford, pp. 291-305.

Rosenman, M.A. and Gero, J.S.: 1998, Purpose and function in design. *Design Studies*, **19** (2), 161-186.

Rumelhart, D. E.: 1986, Learning internal representations by error propagation. *Parallel Distributed Processing*, J. McClelland, Rumelhart, D.E., ed., MIT Press, Cambridge, MA.

Rumelhart, D. E., Widrow, B., Lehr, M.A.: 1994, The basic ideas in neural networks. *Communications of the ACM*, 37(3) 87-92.

Siewiorek, D. P., Swarz, Robert S.: 1982, *The Theory and Practice of Reliable System Design*, Digital Press, Bedford, MA.

Smith, G. and Maher, M.L.: 2003 *User Guide for KCDCC AW Agent Package*, Key Centre of Design Computing and Cognition, University of Sydney.

Soibelman, L., Kim, H.: 2002, Data preparation process for construction knowledge generation through knowledge discovery in databases, *Journal of Computing in Civil Engineering*, 16(1), 39-48.

Stone, J., Blockley, D.: 1989, Towards machine learning from case histories. *Civil Engineering Systems* 5(3):129-135

Sullivan, D., Keane, M., Kelliher, D.: 2002, Specification of a STEP compliant integrated environment to support performance based assessment and control of building energy systems, *IRUSE,* Department of Civil Engineering, University of College Cork, Ireland.

Szalapaj, P.: 2001, *CAD Principles for Architectural Design*, Architectural Press, Oxford.

Waugh L. M., Froese T. M., St. Pierre N. J.: 1996, Project management, 2015 AD, CSCE Annual Conference (Volume I); *Proceedings of the 1996 Annual Conference of the Canadian Society for Civil Engineering*, Edmonton, Alberta CSCE, Montreal, Canada, 1996. Vol. 1, pp. 458-467.

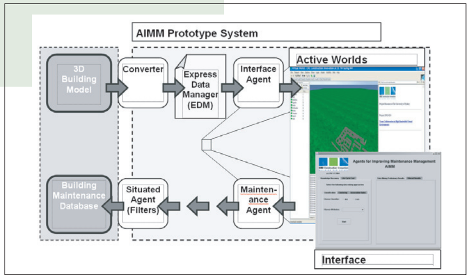
Witten, I. and Frank, E.: 2000, *Data Mining: practical machine learning tools and techniques with java implementations* , M organ Kaufman, San Diego

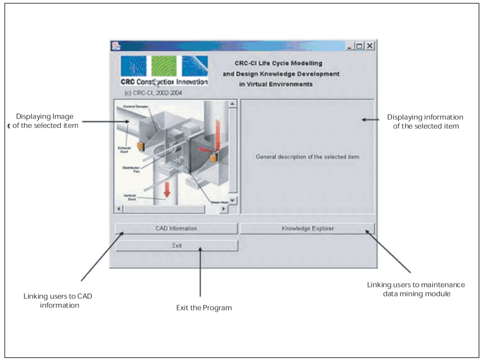
Wooldridge, W.: 2002, *An Introduction to Multiagent Systems*. John Wiley, London.

APPENDIXES

10.1 Appendix A: Industry Flyer A – Research and Development

4





Cooperative Research Centre for Construction Innovation

The Australian Cooperative Research Centre for **Construction Innovation** is a national collaboration involving 19 industry, government and research partners and has been made possible by a $14 Million Federal Government grant throughthe Cooperative Research Centre program.

Construction Innovation is developing the AIMM prototype system to facilitate a paradigm shift to design, construction and facilities managment within architecture, engineering and construction (AEC)sector both in Australia and internationally.

For more information on the CRC CI contact:

Cooperative Research Centre for Contruction Innovation

9th Floor, L Block,
QUT Gardens Point
2 George Street Brisbasne,
QLD 4001 Australia

T 61 7 3864 1393
F 61 7 38649151
E enquiries@construction-innovation.info
www.construction-innovation.info

University of Sydney • QLD Department of Public Works • Woods Bagot • CSIRO

AIMM

Agents for Improved Maintenance Management

Life Cycle Modelling and Design Knowledge Development in Virtual Environment

For information contact:

Project Leader

John Gero
Key Centre of Design Computing and Cognition
University of Sydney
T 61 2 9351 2328
F 61 2 9351 3031
E john@arch.usyd.edu.au
www.construction-innovation.aimm

CRC Construction Innovation
BUILDING OUR FUTURE

AIMM stands for Agents for Improved Maintenance Management. AIMM is a prototype system that enables more efficient and effective facilities maintenance and management through better planning based on models developed from available building maintenance data. This results in more economical life cycle costing of buildings.

Improved skills and competencies for building design professionals and maintenance managers will be possible based on this tool, and the feedback from previous facilities maintenance will result in higher quality designs.

AIMM is specifically designed to:

- Combine a 3D model with maintenance and other asset data in order to enable building designers and owners to visually model the impact of design, maintenance, refurbishment and extension decisions on the building's life cycle cost.
- Provide a basis for linking data mining with emerging communication technologies to address a gap in the construction life cycle.

AIMM improves building life cycle modelling and provides benefits through the:

- Linking of a 3D model with maintenance data to allow both the facility manager and the designer to gain access to building maintenance information and knowledge that is currently inaccessible,
- Integration of data mining agents into the maintenance process to produce timely data for the facility manager on the effects of different maintenance regimes.

2

Building Maintenance & Facilities Management

Building maintenance and management includes a range of building and construction areas, such as:

- Facilities Strategic Planning
- Project services, and
- Facilities serives.

The current technology for facility maintenance uses databases to keep track of information and for notification of maintenance schedules.

The increasing use of databases to store information about facilities, their use, and their maintenance provides the background and platform for the use of data mining techniques for future projections.

These databases are so far not well linked with an interactive 3D model of the building.

AIMM Prototype System

AIMM is a tool for improving the life cycle modelling of building systems by:

- Providing a modelling tool in a 3D environment that can be attached to asset management systems
- Developing agent architectures for data mining within a 3D environment linked to maintenance databases
- Improving the connection between maintenance and design knowledge.

Design Knowledge Development

AIMM provides a more efficient and effective facilities maintenance and management through better planning based on models developed from available maintenance data. This will result in a more economical life cycle costing of buildings.

Improved skills and competencies for designers and maintenance managers will be possible based on these tools. The feedback from previous facilities maintenance will result in higher quality designs.

Knowledge Discovery & Data Mining

Data mining techniques are tools that allow building and facility managers to identify valid and useful patterns of knowledge from large amount of building data.

Integrating facilities databases with interactive 3D virtual environments of building models and data mining techniques provides a visual modelling tool for the simulation and projection of the financial and physical impact of a building and its components over its life cycle.

Agents for Improving Maintenance Management

Software agents for data mining on building maintenance are implemented within the AIMM system.

AIMM's usefulness lies in the maximisation of benefits without requiring users to have extensive training or a comprehensive understanding of data mining techniques.

Users such as facility managers and developers can reflect on discovered knowledge from records of maintenance data and utilise these results in improving management of maintenance, the operation of buildings and enhance building life cycle.

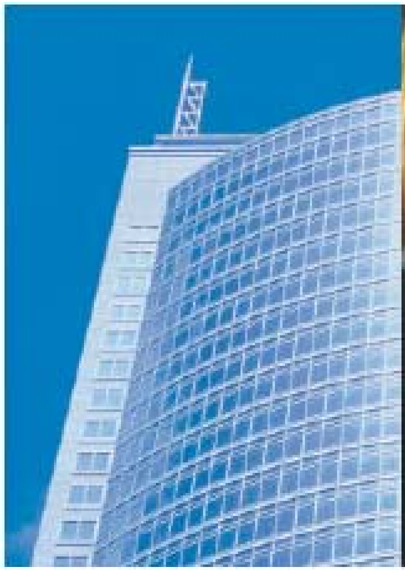
3

Interactive 3D Building Models

Exchanging non-graphical information in building projects using virtual environments is quite useful and the development of an interactive 3D CAD model has the essential role of supporting knowledge sharing among co-designers.

For 3D CAD models to become useful for interactivity in Virtual Environments, it needs to be converted in order to view and interact with objects. At the same time, objects should also be exported in Industry Foundation Class (IFC) format in order to be accessible for data management.

AIMM



10.2 Appendix B: Industry Flyer B – Information for Potential Industry Partners

CRC Construction Innovation
BUILDING OUR FUTURE

Industry Partners

AIMM Industry Partner Information & Benefits

For information contact:

Project Leader

John Gero
Key Centre of Design Computing and Cognition
University of Sydney
T 61 2 9351 2328
F 61 2 9351 3031
E john@arch.usyd.edu.au
www.construction-innovation.aimm

AIMM and the Cooperative Research Centre for Construction Innovation

The Australian Cooperative Research Centre for **Construction Innovation** is a national collaboration between industry, government and research partners and has been made possible by a $14 Million Federal Government grant through the Cooperative Research Centre program.

Construction Innovation is developing AIMM to facilitate a paradigm shift to apply advanced infromation technology to design, construction and facilities management within architecture, engineering and construction (AEC) sector both in Australia and internationally.

How can your company participate?

There are a number of ways your company can contribute to the AIMM Project. Contributions take the form of the following:

Finance

Financial contributions secure direct involvement in project development and demonstrations and early access to techniques and results.

Time

Your company allocates staff for approx. 1-2 hours per week to R&D of the project.

Resources

Supplying existing maintenance data in order to develop, test, analyse and demonstrate AIMM system.

For more information about the CRC CI contact:

Cooperative Research Centre for Contruction

9th Floor, L Block,
QUT Gardens Point
2 George Street Brisbasne,
QLD 4001 Australia

T 61 7 3864 1393
F 61 7 38649151
E enquiries@construction-innovation.info
www.construction-innovation.info

University of Sydney • Woods Bagot • CSIRO • QLD Department of Public Works

AIMM

University of Sydney

CRC Construction Innovation
BUILDING OUR FUTURE

Prospective Partner Information

What is AIMM?

AIMM stands for '**A**gents for **I**mproved **M**aintenance **M**anagement.' The AIMM system is a prototype tool that has developed the state of the art life cycle modelling of buildings through the linking of a 3D model with maintenance data to allow both the facility manager and the designer to gain access to building maintenance information and knowledge that is currently inaccessible. AIMM integrates data mining agents into the maintenance process to produce timely data for the facility manager on the effects of different maintenance regimes.

AIMM is designed:

- To combine 3D modelling with maintenance and other asset data to enable building designers and owners to visually model the impact of design, maintenance, refurbishment and extension decisions on the building's life cycle cost.
- To provide a basis for linking data mining with emerging communication technologies in order to address a gap in the construction life cycle.

AIMM Pro ject i dentifies:

- Better planning based on models developed from available maintenance data resulting in more economical life cycle costing of buildings.
- Increased customer and organisation needs in Life Cycle Modelling and Costing.
- Potential market opportunity from the introduction of new systems to be exploited as a lever to introduce new technology.

Partner Benefits

How will your compan y benefit?

There are a number of ways your business can directly benefit from the AIMM Project. These benefits can be broadly divided into two main areas:

1) Research & Development;

Direct involvement in developing and accessing information on state-of-the- art Life Cycle Modelling Systems with early access to techniques and results. The early adoption of latest technology provides your organisation with competitive advantages.

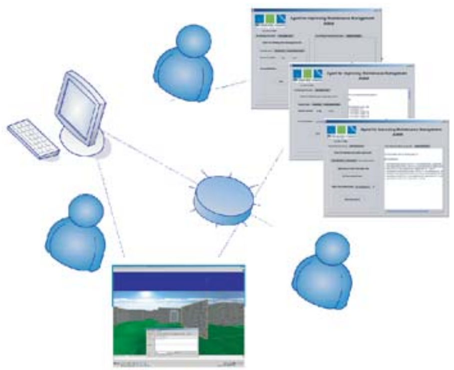
2) CRC Collaboration;

Access to a world-wide network of Industry Partners and Research Institutes. Construction Innovation is collaborating with World Leaders in construction research and development with the collective aim of creating World Best Practice in international construction and property management. The five foundation members of the International Construction Research Alliance (ICALL) include:

Cooperative Research Centre for Construction Innovation, Australia
Centre for Integrated Facilitated Engineering (CIFE), Stanford University, USA
Centre Scientifique et Technique du Bâtiment (CSTB), France
VTT Building and Transport, Finland
Centre for Facilities Management (CFM), The University of Salford, UK

10.3 Appendix C: Industry Flyer C – Workshop Flyer

The AIMM prototype system progress seminar will demonstrate a 3D model with maintenance and other asset data that enables visual modelling of the impacts of design, maintenance, refurbishment and extension decisions on the building's life cycle cost.



The AIMM system will be deomonstrated as a technique for improving the life cycle modelling of building systems by: (1) providing a modelling tool in a 3D environment that can be attached to asset management systems; (2) providing agent architectures for data mining within a 3D environment linked to maintenance databases; and (3) connecting maintenance and design knowledge.

AIMM
Progress Seminar

Dates and Venues

30th June, Sydney:
The 'Sentient'
University of Sydney,
School of Architecture,
Wilkinson Building- G04,

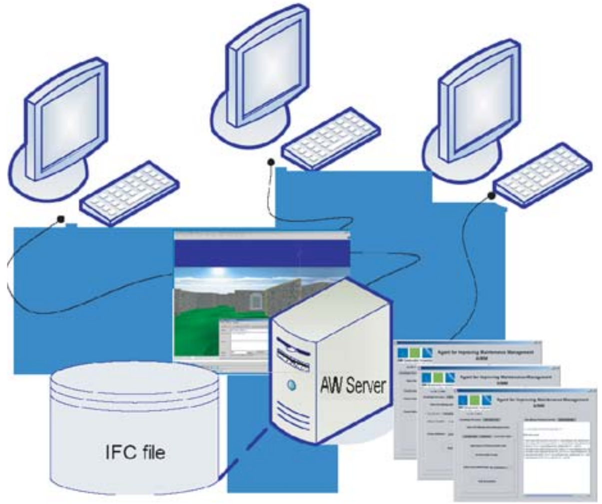
30th July, Brisbane:
Qld Dept of Public Works
80 George Street
Brisbane QLD 4000

Agenda

2:00 Opening
2:15 Project Overview
2:30 Systems Architecture
2.45 Systems demonstration

3:00 Discussion

3:30 Event closure



Team Leaders:

Overall Project Leader	John Gero (USYD)
Technical Project Manager	Dr Rabee Refatt (USYD)
Project Coordinator	Julie Jupp (USYD)

Team Members:

Dr. Mary Lou Maher	(USYD)
Dr. Mike Rosenman	(USYD)
Dr. David Gunaratnam	(USYD)
Mr. Wei Peng	(USYD)
Ms. Julie Jupp	(USYD)
Dr. Lan Ding	(CSIRO)
Mr. David Marchant	(Woods Bagot)
Mr. Dale Gilbert	(Qld Dept of Public Works)
Mr. Teng Hee Tan	(Qld Dept of Public Works)

Industry Partners:

University of Sydney, Prof. John Gero
Woods Bagot, Mr David Marchant
Qld Dept of Public Works, Mr. Dale Gilbert
CSIRO, Lan Ding

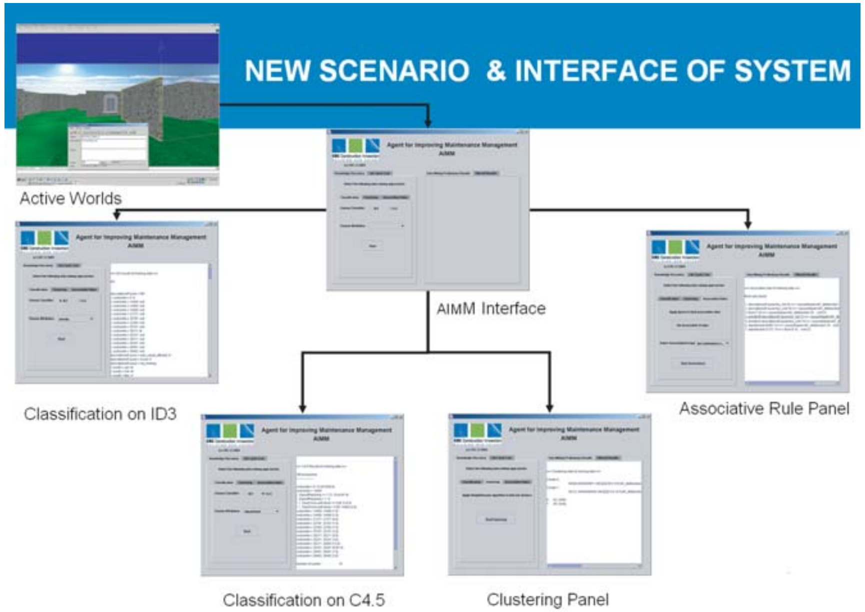
CRC for Construction Innovation Projects:
Life Cycle Modelling and Design Knowledge Development in Virtual Environments [2001-002-B]
Stage 5 – AIMM Systems Demonstration [2001-002-B]

CRC Construction Innovation
BUILDING OUR FUTURE

life cycle modelling & design knowledge development

AIMM

NEW SCENARIO & INTERFACE OF SYSTEM



The AIMM tool:

+ provides a more efficient and effective facilities maintenance and management through better planning based on models developed from available maintenance data,

+ allows building and facility managers to identify valid and useful patterns of knowledge from large amount of building data,

+ provides a visual modelling tool for the simulation and projection of the financial and physical impact of a building and its components over its live cycle,

+ can be used to improve skills and competencies for designers and maintenance managers, providing feedback from previous facilities maintenance will result in higher quality designs.

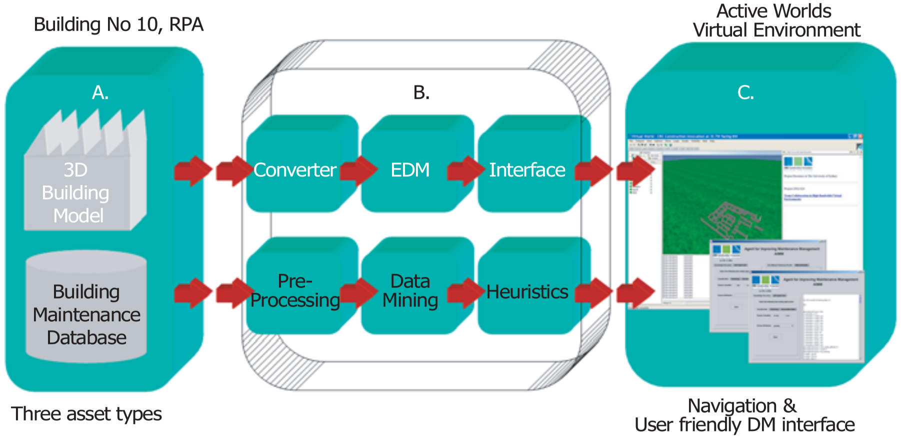
Who should attend?

Our CRC for Construction Innovation Industry Partners involved in:

- Design
- Planning
- Estimating
- Quantity Surveying
- Project Management
- Information Technology

To cofirm your attendance contact:

Julie Jupp
jupp_j@arch.usyd.edu.au



AIMM Progress Seminar:
Life Cycle Modelling and Design Knowledge Development in Virtual Environments

[2001-002-B]

For information contact:
Project Leader
John Gero
Key Centre of Design Computing and Cognition
University of Sydney
T 61 2 9351 2328
F 61 2 9351 3031
E john@arch.usyd.edu.au
www.construction-innovation.aimm

The University of Sydney

Queensland Government
Department of Public Works

AIMM

CSIRO

WOODS BAGOT

CRC Construction Innovation
BUILDING OUR FUTURE

10.4 Appendix D Sample of Available Maintenance Data

An example of the available maintenance data for the Air Handling Units at Building 10, Royal Prince Alfred Hospital, Central Sydney Area Health Service.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Line No	Work Orde	Job Typ	Job Sub T	Priority	Cost C	Depar	Floor	Room	Asset No	Task Num	Description	Extra Text	Comments	Work Orde
2	51	P0043720	PM	FILT	H	0	10	1	MPR1	AHU1000-0	FR003	FILTER REF	1. REPLACE FILTER MEDI		PMS
3	52	P0043721	PM	FILT	H	0	10	1	MPR1	AHU1000-0	FR003	FILTER REF	1. REPLACE FILTER MEDI		PMS
4	53	P0043722	PM	FILT	H	0	10	1	MPR1	AHU1000-0	FR003	FILTER REF	1. REPLACE FILTER MEDI		PMS
5	54	P0043723	PM	FILT	H	0	10	1	MPR1	AHU1000-0	FR003	FILTER REF	1. REPLACE FILTER MEDI		PMS
6	55	P0043724	PM	FILT	H	0	10	6	MPR2	AHU1001-0	FR003	FILTER REF	1. REPLACE FILTER MEDI		PMS
7	56	P0043725	PM	FILT	H	0	10	6	MPR2	AHU1001-0	FR003	FILTER REF	1. REPLACE FILTER MEDI		PMS
8	232	P0046282	PM	FILT	H	0	10	1	MPR1	AHU1000-0	FR003	FILTER REF	1. REPLACE FILTER MEDI		PMS
9	233	P0046283	PM	FILT	H	0	10	1	MPR1	AHU1000-0	FR003	FILTER REF	1. REPLACE FILTER MEDI		PMS
10	234	P0046284	PM	FILT	H	0	10	1	MPR1	AHU1000-0	FR003	FILTER REF	1. REPLACE FILTER MEDI		PMS
11	235	P0046285	PM	FILT	H	0	10	1	MPR1	AHU1000-0	FR003	FILTER REF	1. REPLACE FILTER MEDI		PMS
12	236	P0046286	PM	FILT	H	0	10	6	MPR2	AHU1001-0	FR003	FILTER REF	1. REPLACE FILTER MEDI		PMS
13	237	P0046287	PM	FILT	H	0	10	6	MPR2	AHU1001-0	FR003	FILTER REF	1. REPLACE FILTER MEDI		PMS
14	429	P0048765	PM	FILT	H	0	10	1	MPR1	AHU1000-0	FR003	FILTER REF	1. REPLACE FILTER MEDI		PMS
15	430	P0048766	PM	FILT	H	0	10	1	MPR1	AHU1000-0	FR003	FILTER REF	1. REPLACE FILTER MEDI		PMS
16	431	P0048767	PM	FILT	H	0	10	1	MPR1	AHU1000-0	FR003	FILTER REF	1. REPLACE FILTER MEDI		PMS
17	432	P0048768	PM	FILT	H	0	10	1	MPR1	AHU1000-0	FR003	FILTER REF	1. REPLACE FILTER MEDI		PMS
18	433	P0048769	PM	FILT	H	0	10	6	MPR2	AHU1001-0	FR003	FILTER REF	1. REPLACE FILTER MEDI		PMS
19	434	P0048770	PM	FILT	H	0	10	6	MPR2	AHU1001-0	FR003	FILTER REF	1. REPLACE FILTER MEDI		PMS
20	632	P0051193	PM	FILT	H	0	10	1	MPR1	AHU1000-0	FR003	FILTER REF	1. REPLACE FILTER MEDI		PMS
21	633	P0051194	PM	FILT	H	0	10	1	MPR1	AHU1000-0	FR003	FILTER REF	1. REPLACE FILTER MEDI		PMS
22	634	P0051195	PM	FILT	H	0	10	1	MPR1	AHU1000-0	FR003	FILTER REF	1. REPLACE FILTER MEDI		PMS
23	635	P0051196	PM	FILT	H	0	10	1	MPR1	AHU1000-0	FR003	FILTER REF	1. REPLACE FILTER MEDI		PMS
24	636	P0051197	PM	FILT	H	0	10	6	MPR2	AHU1001-0	FR003	FILTER REF	1. REPLACE FILTER MEDI		PMS
25	637	P0051198	PM	FILT	H	0	10	6	MPR2	AHU1001-0	FR003	FILTER REF	1. REPLACE FILTER MEDI		PMS
26	799	P0053370	PM	FILT	H	0	10	1	MPR1	AHU1000-0	FR003	FILTER REF	1. REPLACE FILTER MEDI		PMS
27	800	P0053371	PM	FILT	H	0	10	1	MPR1	AHU1000-0	FR003	FILTER REF	1. REPLACE FILTER MEDI		PMS
28	801	P0053372	PM	FILT	H	0	10	1	MPR1	AHU1000-0	FR003	FILTER REF	1. REPLACE FILTER MEDI		PMS
29	802	P0053373	PM	FILT	H	0	10	1	MPR1	AHU1000-0	FR003	FILTER REF	1. REPLACE FILTER MEDI		PMS
30	803	P0053374	PM	FILT	H	0	10	6	MPR2	AHU1001-0	FR003	FILTER REF	1. REPLACE FILTER MEDI		PMS
31	804	P0053375	PM	FILT	H	0	10	6	MPR2	AHU1001-0	FR003	FILTER REF	1. REPLACE FILTER MEDI		PMS
32	923	P0054595	PM	FILT	H	0	10	1	MPR1	AHU1000-0	FR001	FILTER REF	1. REPLACE FILTER MEDI		PMS
33	924	P0054596	PM	FILT	H	0	10	1	MPR1	AHU1000-0	FR001	FILTER REF	1. REPLACE FILTER MEDI		PMS
34	925	P0054597	PM	FILT	H	0	10	1	MPR1	AHU1000-0	FR001	FILTER REF	1. REPLACE FILTER MEDI		PMS
35	926	P0054598	PM	FILT	H	0	10	1	MPR1	AHU1000-0	FR001	FILTER REF	1. REPLACE FILTER MEDI		PMS
36	927	P0054599	PM	FILT	H	0	10	6	MPR2	AHU1001-0	FR001	FILTER REF	1. REPLACE FILTER MEDI		PMS
37	928	P0054600	PM	FILT	H	0	10	6	MPR2	AHU1001-0	FR001	FILTER REF	1. REPLACE FILTER MEDI		PMS
38	1010	P0055780	PM	FILT	H	0	10	1	MPR1	AHU1000-0	FR001	FILTER REF	1. REPLACE FILTER MEDI		PMS
39	1011	P0055781	PM	FILT	H	0	10	1	MPR1	AHU1000-0	FR001	FILTER REF	1. REPLACE FILTER MEDI		PMS
40	1012	P0055782	PM	FILT	H	0	10	1	MPR1	AHU1000-0	FR001	FILTER REF	1. REPLACE FILTER MEDI		PMS
41	1013	P0055783	PM	FILT	H	0	10	1	MPR1	AHU1000-0	FR001	FILTER REF	1. REPLACE FILTER MEDI		PMS
42	1014	P0055784	PM	FILT	H	0	10	6	MPR2	AHU1001-0	FR001	FILTER REF	1. REPLACE FILTER MEDI		PMS
43	1015	P0055785	PM	FILT	H	0	10	6	MPR2	AHU1001-0	FR001	FILTER REF	1. REPLACE FILTER MEDI		PMS
44	1890	P0065164	PM	FILT	H	0	10	1	MPR1	AHU1000-0	FR002	FILTER REF	1. REPLACE FILTER MEDI		PMS

Figure 1. Maintenance Data Sample

10.5 Appendix E Data Mining Algorithms

Decision Trees

A Decision Tree is a tree-based knowledge representation methodology used to represent classification rules. For example, consider the problem of predicting whether or not an outdoor game will be played based on the weather outlook, the humidity, and the wind. A decision tree for such a prediction, based on past assessments, might look like the one shown in Figure 2.

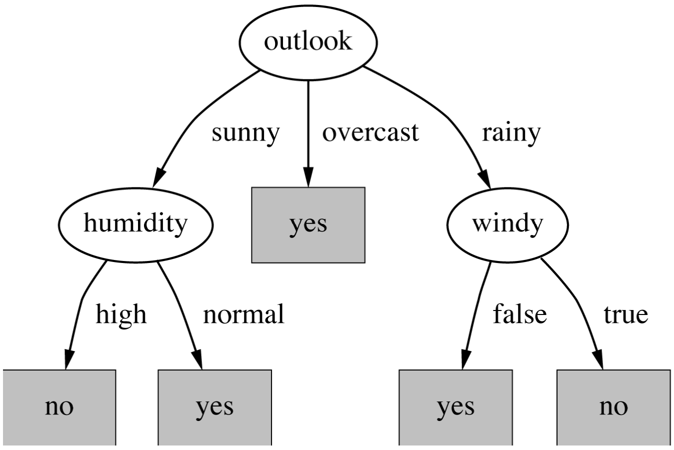


Figure 2. Sample Decision Tree for weather data

The tree starts at the root at the top, which represents all the data to be classified. The branches of the tree represent each possible value of the decision node from which they originate. The leaf nodes represent the class labels while non-leaf nodes represent the attributes associated with the objects being classified. Decision Trees are useful, particularly for solving problems that can be cast in terms of producing a single answer in the form of a class name. Based on answers to the questions at the decision nodes, one can find the appropriate leaf and the answer it contains. This class of data mining algorithm has previously been applied in the construction industry in the area of connecting building environmental performance to design parameters (McLaughlin and Gero, 1987).

The first stage of C4.5 (Quinlan 1993) generates a decision tree. Each level of the decision tree represents a split of the data set. This split is chosen by examining each possible split of the data on each attribute, and choosing the one which best splits the data (according to an information theoretic measure of the distribution of classes in each subset). This continues for each level of the decision tree until there is no benefit from further segmenting the data. Once this has been done, rules are generated by traversing each branch of the tree and collecting the conditions at each branch of the decision tree. Each generated rule has a confidence percentage associated with the class it predicts. The uncertainty is caused by the generalization process, as some leaves on a tree may no longer contain single labels. The rules derived from the tree in the above figure would be:

```
if outlook = sunny          and humidity = high      then play = no
if outlook = sunny          and humidity = normal    then play = yes
if outlook = overcast                                then play = yes
if outlook = rainy          and windy = false        then play = yes
if outlook = rainy          and windy = true         then play = no
```

A decision tree can be used, for example, to classify how long it might take to repair a certain type of item. For example, given the input from a BMS database, a decision tree might create a rule such as:

```
if Job_Type_Code = AD  and Priority = M
   then Completion_Time = MEDIUM
```

where `MEDIUM` might be defined to be greater than 2 weeks and less than 4 weeks, or

```
if Make = Acme and Asset_Category_Code = COMP
   then Job_Type_Code = MAC and Completion_Time = LONG
```

Advantages:

- Rules output by decision trees are easy for humans to interpret
- Tree generation process and resulting rules easily convert to SQL queries
- Robust to noise because they deal with probabilities
- Able to generalise from a small amount of data fairly quickly, and become more accurate as more data becomes available

Disadvantages:

- Poor handling of feature inter-dependence, fuzziness, or non-linearity
- Order of the decision tests relies on measures of the full data set, learning is non-incremental

Association Rules

Association rule mining involves finding frequent patterns, associations, correlations, or causal structures among sets of items or objects in transaction databases, relational databases, and other information repositories (Han 2001). Given a set of records, each of which contains a number of items from a given collection, dependency rules are produced that will predict the occurrence of an item based on the presence of another item. For example, if the contents of shoppers' trolleys are analysed, then a rule might be:

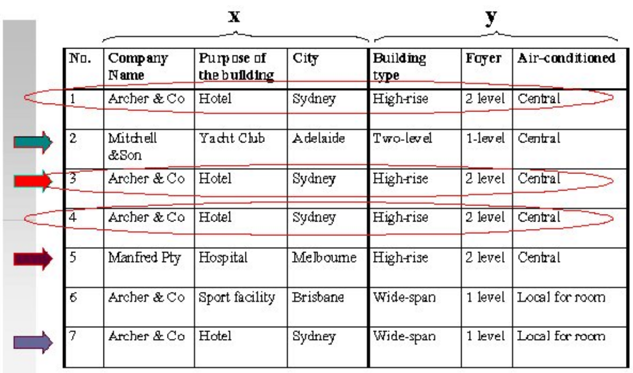
```
if the customer is buying beer,
     then the customer will also purchase pretzels
```

The best known association rule mining algorithm is APriori, which finds all associations that satisfy criteria for minimum support and minimum confidence. Support is the fraction of all instances that satisfy both sides of a rule. Confidence is the fraction of the instances satisfying the antecedent of a rule that also satisfy the consequent (Agrawal 1995).

An example of association rule mining from a database of buildings is illustrated in Figure 3. As indicated by the entries circled in red, there is support for a rule that

```
if the Company Name = "Archer & Co" and Purpose = "Hotel" and
     City = "Sydney", then Building = "High Rise" and
     Foyer = "2 level" and Air-conditioned = "Central"
```

The support for this rule is 0.57, but the confidence is only 0.18, indicating that it is not a rule of good quality.



No.	Company Name	Purpose of the building	City	Building type	Foyer	Air-conditioned
1	Archer & Co	Hotel	Sydney	High-rise	2 level	Central
2	Mitchell &Son	Yacht Club	Adelaide	Two-level	1-level	Central
3	Archer & Co	Hotel	Sydney	High-rise	2 level	Central
4	Archer & Co	Hotel	Sydney	High-rise	2 level	Central
5	Manfred Pty	Hospital	Melbourne	High-rise	2 level	Central
6	Archer & Co	Sport facility	Brisbane	Wide-span	1 level	Local for room
7	Archer & Co	Hotel	Sydney	Wide-span	1 level	Local for room

Figure 3. Association Rule Mining from a Table of Building Information

Most association rule algorithms generate huge numbers of rules. Further, many statistically strong rules will not be viewed by a domain expert as being interesting. To combat this requires either pre-processing using domain knowledge to limit the generation of possible rules, or post-processing by applying a filter to shrink the generated rules.

Advantages:

- Suitable for large data sets
- Model consists of a set of rules that can be easily understand
- Existing tools with exploratory GUIs available

Disadvantages:

- Most association rule algorithm generate huge volumes of rules
- Strong rules are not necessarily interesting

Neural Networks

The foundation of the artificial neural networks paradigm (ANN) was laid in the 1950s and ANN has gained significant attentions in the past decade due to the development of more powerful hardware and neural algorithms (Rumelhart 1994). Neural networks have been adopted in various engineering, business, military, and biomedical domains.

Among the numerous artificial neural networks that have been proposed, Backpropagation Networks have been extremely popular for their unique learning capability. Backpropagation Networks (Rumelhart 1986) are layered, feed-forward models. Activations flow from the input layer through the hidden layer, then to the output layer, as shown in Figure 4. A Backpropagation Network typically starts out with a random set of weights. The network adjusts its weights each time it sees an input-output pair. Each pair is processed at two stages, a forward pass and a backward pass. The forward pass involves presenting a sample input to the network and letting activations flow until they reach the output layer. During the backward pass, the network's actual output is compared with the target output and error estimates are computed for the output units. The weights connected to the output units are adjusted in order to reduce the errors (a gradient descent method). The error estimates of the output units are then used to derive error estimates for the units in the hidden layer. Finally, errors are propagated back to the connections stemming from the input units. The Backpropagation Network updates its weights incrementally until the network stabilizes.

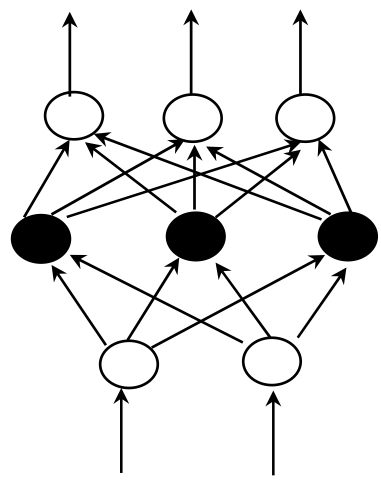


*Figure 4.*Artificial Neural Networks consist of a layer of input units, followed by one or more layers of hidden units, which ultimately lead to output units that indicate the learned response to the given inputs

Advantages:

- Handles large feature spaces

- Handles feature inter-dependence, fuzziness, and non-linearity well
- Robust to noise

Disadvantages:

- Hard to interpret results
- Long training time

Hidden Markov Models

Text mining is about looking for patterns in natural language text, and may be defined as the process of analysing text to extract information from it for particular purposes. Web mining is the slightly more general case of looking for patterns in hypertext and often applies graph theoretical approaches to detect and utilize the structure of web sites.[42] The commonly used text mining tool is Hidden Markov Models (HMMs) which have the capability to take an input of some sequences with some sort of pattern in common and describe what the pattern is in probabilistic language.

HMMs have been introduced several years ago as a tool for probabilistic sequence modelling. "Research on text categorization has been mainly focused on non-structured documents. In the typical approach, inherited from information retrieval, each document is represented by a sequence of words, and the sequence itself is normally flattened down to a simplified representation called bag-of-words. Second, we introduce a hidden Markov model (HMM) that can deal with sequences of bag-of-words. Finally, we solve the categorization problem by running the Viterbi algorithm on the trained HMM." (FRASCONI, P., SODA, G., VULLO, A., 2002).

Advantages

- Extends data mining to text and natural language domain
- Enhanced information retrieval

Disadvantages:

- Data is not well organised
- Natural language has ambiguity

[42] http://www.cs.waikato.ac.nz/~nzdl/textmining/

10.6 Appendix F

Applying stacked histograms on building maintenance data using three asset types

Asset Type 1: Air Handling Units

Stacked Histograms on the basis of the "priority" attribute:

The "priority" attribute is identified as the base attribute and the correlations are conducted between "priority" and all the other attributes as shown in Figure 5.

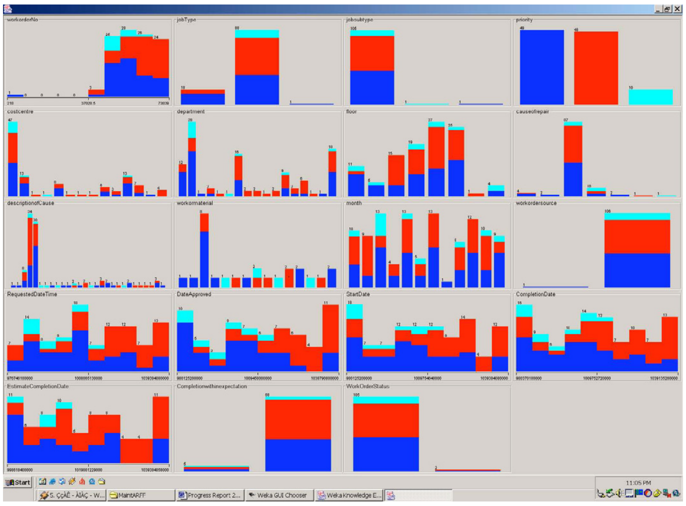


Figure 5. Stacked Histograms on the basis of the "priority" attribute.

There are several rules that can be learned from Figure 5:

- Approximately all A/C malfunction belongs to high and medium priority.
- Approximately all *too_cold* or *too_hot* for attribute *descriptionofcause* are of high or medium priority
- All adjust work for attribute *workormaterial* are high or medium priority Work for attribute *workormaterial* with high priority includes Valves, circuit, V/belt, VSD-supply, reset. These can be viewed as critical works for A/C. Also too_Hot, too_cold consists of large portion of high and medium level priority.

Stacked Histograms on the basis of the "floor" attribute:

There are several rules that can be learned from Figure 6:

- All 7th floor jobs were of high and medium priority
- All 7th floor jobs the cause of repairing was A/C malfunction

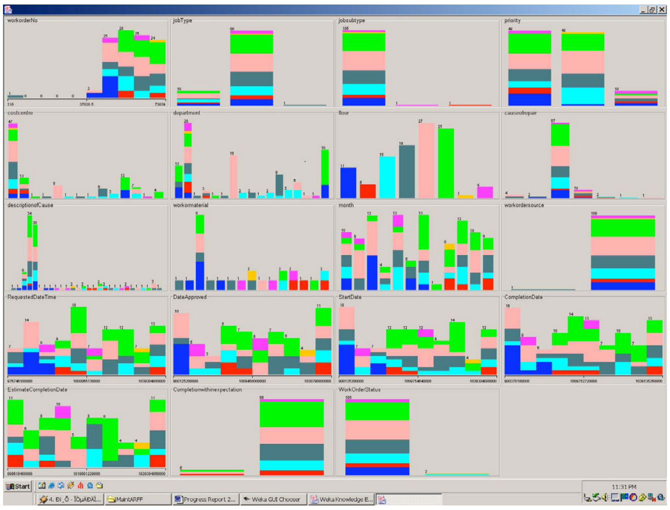


Figure 6. Stacked Histograms on the basis of the "floor" attribute

Stacked Histograms on the basis of the "department" attribute:

There are several rules that can be learned from Figure 7:

- Department 26462 which resides only at floor 7 only involves with A/C malfunction
- Department 21271 only resides at 6th floor.
- In August only department 26464 has corrective maintenance work. (only 1 case).

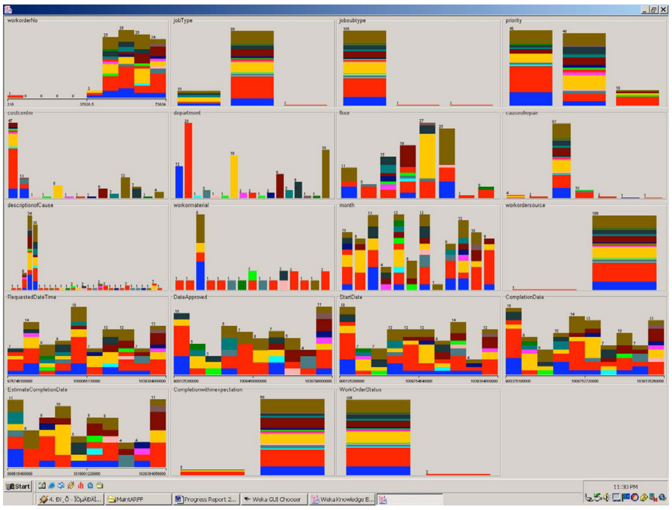


Figure 7. Stacked Histograms on the basis of the "department" attribute.

Stacked Histograms on the basis of the "cause_of_repair" attribute:

There are several rules that can be learned from Figure 8:

- The work at 4, 5, 6, 7th floor constitutes most of the reports of A/C malfunctions, with 86% of A/C malfunction reported from these floors.
- No A/C malfunction was reported at level 9.
- The lowest levels of A/C Malfunction took place in August followed by June and April while other months share similar high A/C malfunction rate.
- A/C malfunction concentrates on the problems of: too_hot 32%, too_cold 28%, not working 7.5%.

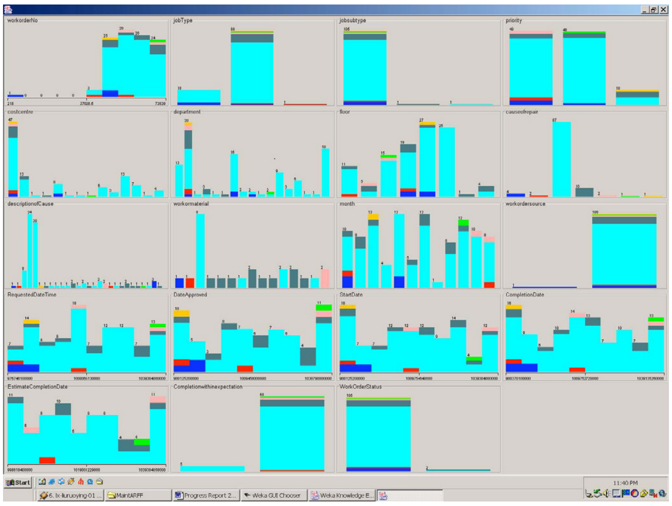


Figure 8. Stacked Histograms on the basis of the “cause_of_repair” attribute.

Stacked Histograms on the basis of the “completion_within_expectation” attribute:

There are several rules that can be learned from Figure 9:

- Floors 4,5,6,7 constitute 80% of completion not meeting expectation.
- Higher percentage of users’ unhappiness associated with high and medium priority.
- Higher level of unhappiness related to completion not meeting expectation with a focus at too_hot and too_cold adjustment activities.
- Cost centre 0 is the highest percentage of not meeting expectation – 44%

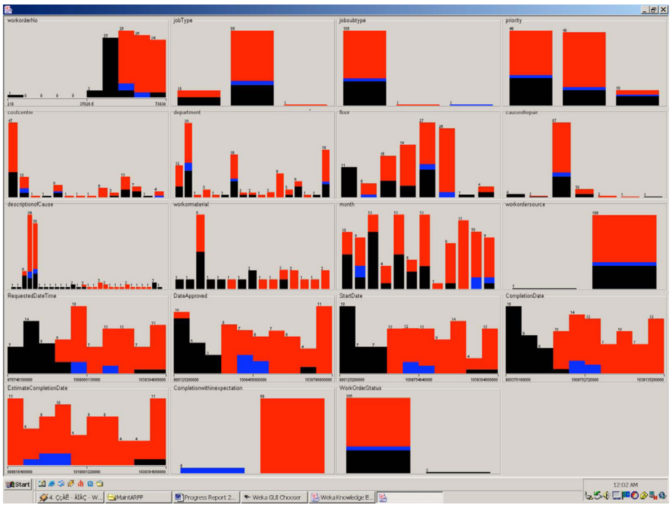


Figure 9. Stacked Histograms on the basis of the “completion_within_expectation” attribute.

Stacked Histograms on the basis of the “job_type” attribute:

There are several rules that can be learned from Figure 10:

- There was no corrective maintenance took place in August;
- All RTP job are with high and medium priority;
- Cost centre 0 reports 42% of CM work.

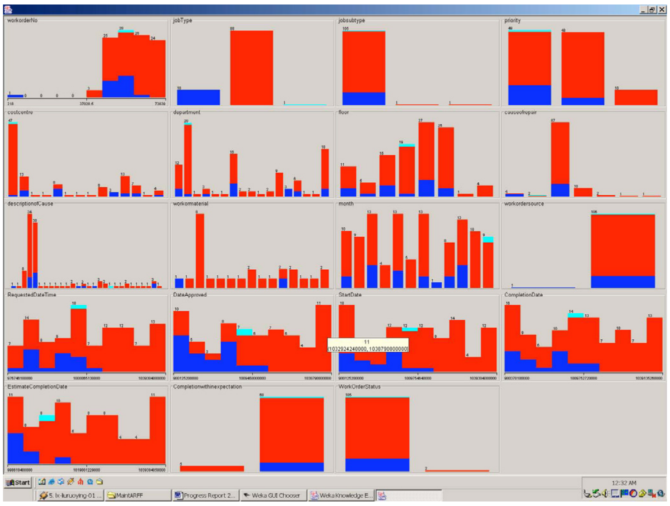


Figure 10. Stacked Histograms on the basis of the “job_type” attribute.

Asset type 2: Thermostatic Mixing Valve Units

Stacked Histogram on the basis of the “Frequency” attribute :

Monthly: 252, 6mthly: 22, 12mthly:24.
There are several rules that can be learned from Figure 11:

- All medium priority works belong to Monthly. High priority works constitute of 55/101 monthly work, 22/101 6mthly work, 24/101 12mthly work. (All 6mthly and 12mthly works are of highly priority)
- 12mthly work happened in the middle of the year – June-Sept, while all 6mthly work is carried out at Dec.
- All 6mthly and 12mthly works were not able to meet completion date expectations.
- All monthly and 12mthly works were completed. Parts of 6mthly works (50%) were outstanding.
- Monthly work was identified as TMV004, 6mthly = TMV002, 12mthly = TMV003
- All maintenance of thermostatic mix valves happened at Level 4.
- All monthly works were supposed to be completed in 0.5 hours and cost $10. All 6mthly and 12mthly works were estimated to be completed in 2 hours and cost $29.

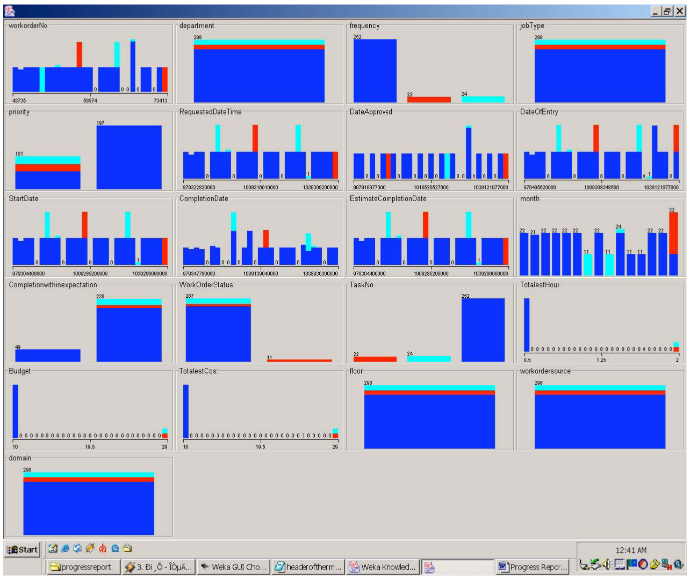


Figure 11. Stacked Histograms on the basis of "maintenance_frequency" attribute.

Stacked Histograms on the basis of the "priority" attribute:

There are several rules that can be learned from Figure 12:

- All high priority works did not meet the expected completion data.
- All medium priority works were completed on the expected completion data.
- There was a trend in emphasizing the maintenance of thermostatic mix valves recently – with increasing of workorder_No, the priority was getting higher.

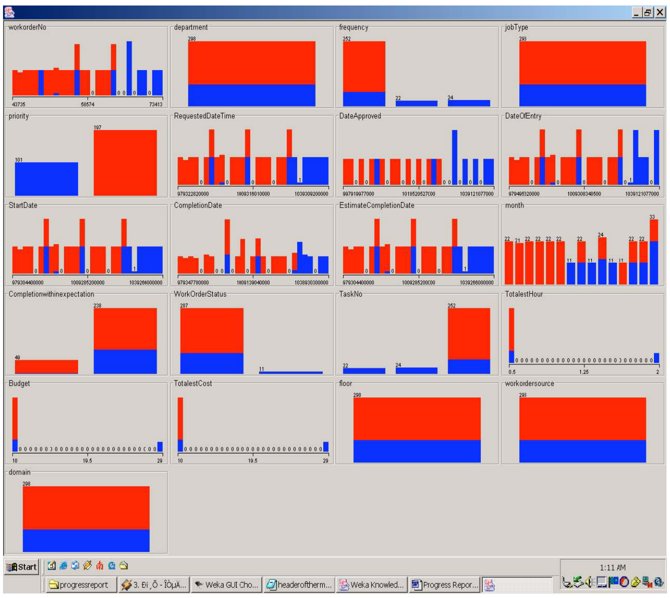


Figure 12. Stacked Histograms on the basis of the "priority" attribute.

Stacked Histograms on the basis of the "month" attribute:

There are several rules that can be learned from Figure 13:

- All works between August and December did not meet the expectation of the completion date.
- All 6mthly works (TMV002) were carried out in December.

- All outstanding works took place in December.

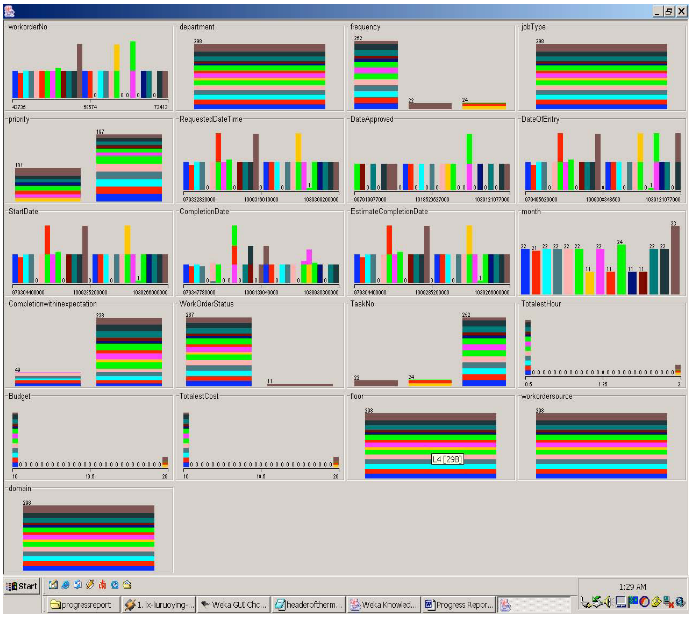


Figure 13. Stacked Histograms on the basis of the “month” attribute.

Stacked Histograms on the basis of the “completion_within_expectation” attribute:

There are several rules that can be learned from Figure 14:

- Nearly all high priority works did not meet the expectations of completion date.
- All 6mthly and 12mthly maintenance work were not satisfactory in meeting deadlines.
- With higher priority works there was a low level satisfaction of work completion.

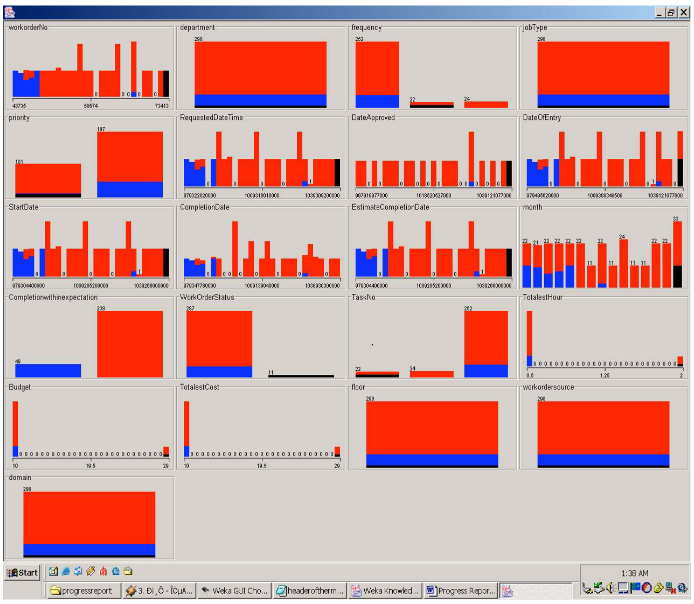


Figure 14. Stacked Histograms on the basis of the “completion_within_expectation” attribute.

Stacked Histograms on the basis of the “workorder_status” attribute:

There are several rules that can be learned from Figure 15:

- The outstanding works were the 6mthly maintenance jobs that took place in December 2002.

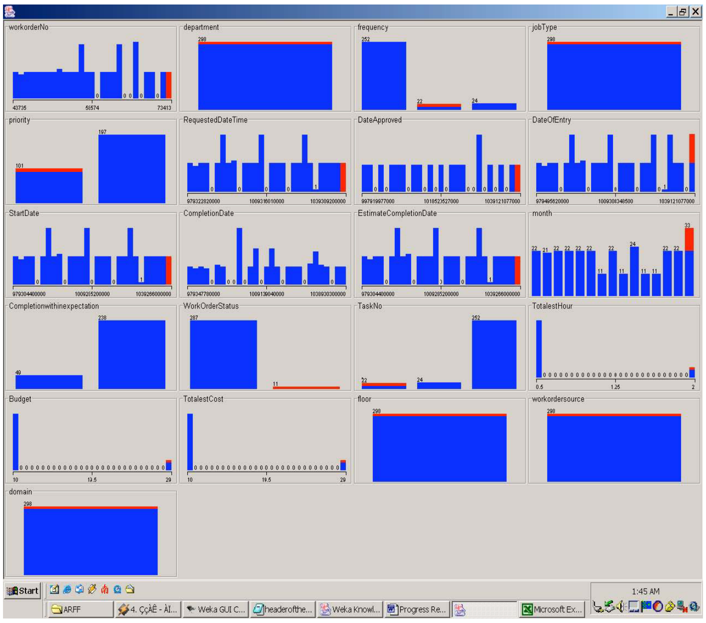


Figure 15. Stacked Histograms on the basis of the "workorder_status" attribute.

Asset type 3: Battery Chargers

Stacked Histograms on the basis of the "workorder_status" attribute:

The rules that can be learned from Figure 16 include:

- All outstanding works took place at the end of work order list around December 2002.

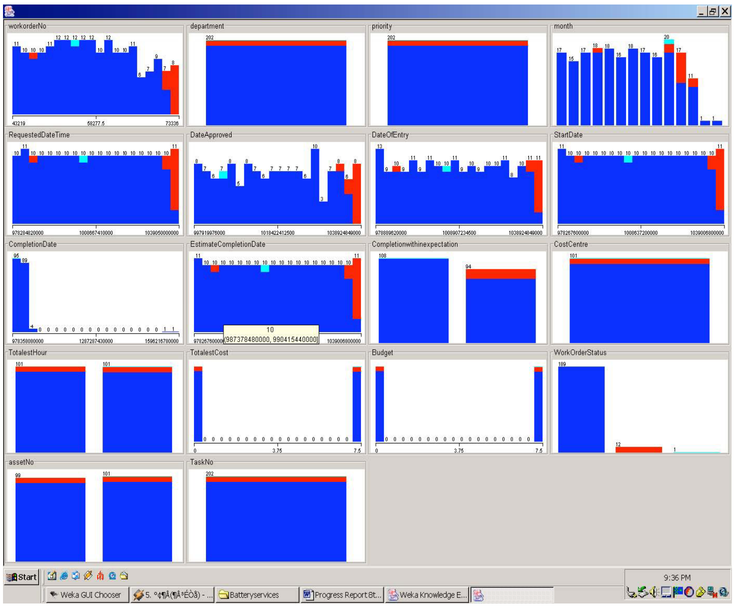


Figure 16. Stacked Histograms on the basis of the "workorder_status" attribute.

Stacked Histograms on the basis of the "asset_No" attribute:

There rules that can be learned from Figure 17 include:

- Asset “EPG0101” belongs to cost centre “1000” while the cost centre for the other asset were not available;
- There is fee charge with asset “EPG0101” while no charge for “EDG100001”.

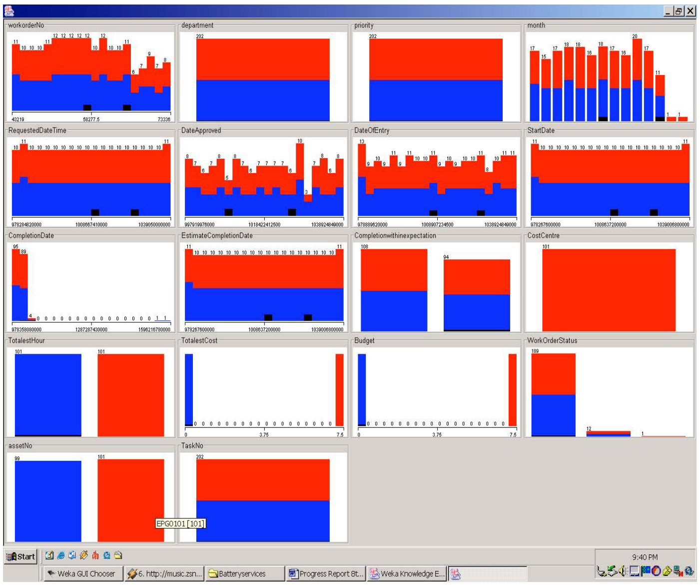


Figure 17. Stacked Histograms on the basis of the “asset_No” attribute.

Learned Rules from applying Clustering using SimpleK-means on maintenance data of the three asset types

Asset type 1: Air Handling Units

```
Cluster centroids:
Cluster 0
60325.098039215685 CM ZZZZ M 0 21271 6 A/C_Malfunction too_hot Adjust Mar
WOS 1.0127091858823529E12 1.0141509239683931E12 1.0127128611764706E12
1.0130870229019608E12 1.0210212768711112E12 N C
Cluster 1
55795.125 CM ZZZZ H 0 10 7 A/C_Malfunction too_cold Adjust Jan WOS
1.0052290457142858E12 1.0067068180170569E12 1.0053050014285714E12
1.0056454119285714E12 1.015600460645E12 N C
Clustered Instances
0      51 ( 48%)
1      56 ( 52%)
```

Applying the data mining technique of clustering using SimpleK-means a cluster was found that divides two major types of A/C malfunction: too_hot, and too_cold with an approximately even distribution around 50%.

Asset type 2: Thermostatic Mixing Valve

The clustering algorithm was applied on two data sets of the Thermostatic Mixing Valve and two different clustering were found. The first data set is the complete data set of the Thermostatic Mixing Valve while the second only contains nominal attributes. The clusters do not seem to be helpful since they only show that demarcations were centred at monthly maintenance in January and December.

Applying SimpleK-means to the complete data set of Thermostatic Mixing Valve:

```
Cluster centroids:
Cluster 0
53927.68020304568 10 monthly SM M 1.001709071573604E12 1.0137665240175492E12
1.001736772675127E12 1.0016730456852792E12 1.0025591802030457E12
1.0016730456852792E12 Jan N C TMV004 0.5 10.0 10.0 L4 PMS RPA
Cluster 1
65177.0 10 monthly SM H 1.0244114352475248E12 1.027678321276379E12
1.0245714955841584E12 1.0243691287128713E12 1.0231385369448372E12
1.0243691287128713E12 Dec N C TMV004 1.183168316831683 18.653465346534652
18.653465346534652 L4 PMS RPA
Clustered Instances
0      197 ( 66%)
1      101 ( 34%)
```

Applying SimpleK-means to the partial data set with nominal attributes of Thermostatic Mixing Valve:

```
Cluster centroids:
Cluster 0
10 monthly SM M Dec N C TMV004 0.5 10 10 L4 PMS RPA
Cluster 1
10 monthly SM H Jan Y C TMV002 0.5 10 10 L4 PMS RPA
Clustered Instances
0      296 ( 99%)
1      2 ( 1%)
```

Asset type 3: Battery Chargers

The clustering algorithm was applied on two data sets of the Battery Chargers and two different clusters were found. The first data set is the complete data set of the Thermostatic Mixing Valve while the second only contains nominal attributes. The clusters do not seem to be helpful since the only show that demarcations were centred at monthly maintenance in April and May.

Applying SimpleK-means to the complete data set of Battery Chargers:

```
Cluster centroids:
Cluster 0
57414.56435643564 10 M May 1.0088072530693069E12 1.0184626676395468E12
1.0088231315247524E12 1.0087689742574258E12 1.0096200989015115E12
1.0087689742574258E12 Y 1000 0.5 7.5 7.5 C EPG0101 BAT001
Cluster 1
57250.574257425746 10 M April 1.0085473283168317E12 1.0182709640850913E12
1.0085536774752476E12 1.0085097742574258E12 1.0202540907826996E12
1.0085097742574258E12 Y 1000 0 0.0 0.0 C EDG1000-01 BAT001
Clustered Instances
0      101 ( 50%)
1      101 ( 50%)
```

Applying SimpleK-means to the partial data set with nominal attributes of Battery Chargers:

```
Cluster centroids:
Cluster 0
10 M May Y 1000 0.5 7.5 7.5 C EPG0101 BAT001
Cluster 1
10 M April Y 1000 0 0 0 C EDG1000-01 BAT001
Clustered Instances
0      101 ( 50%)
1      101 ( 50%)
```

Learned Rules from applying Classification using C4.5 on maintenance data of the three asset types

Asset type 1: Air Handling Units

Although the classification technique C4.5 is a powerful classifier which is robust to noise, it had a poor performance on the available maintenance data because some attributes have almost constant values, such as "job_subtype" with 105 "zzzz" out of 107 cases within the data set, "workorder_status" with 105 "C" out of 107 cases. Furthermore, the lack of meaningful domain data is the major issue for Air Handling Unit maintenance data which deteriorate the performance of C4.5. The findings are primarily representing the distribution of geographical locations wherein the maintenance work was conducted and its type.

1. Department 26462 only reports A/C malfunction. (all 18 cases)
2. For most cost centres = 0 (45 out of 47), the jobtype = CM
3. Department 26462 resides only at 1st and 7th floor
4. Department 21271 only reside at 6th floor

Asset type 2: Thermostatic Mixing Valves

Applying C4.5 on the "frequency" attribute:
The constructed decision tree as shown in Figure 18 describes the structure of maintenance data:

1. Task TMV002 is monthly work;
2. Task TMV003 is 12mthly (yearly) work;
3. Task TMV003 is 6mthly (half-yearly) work.

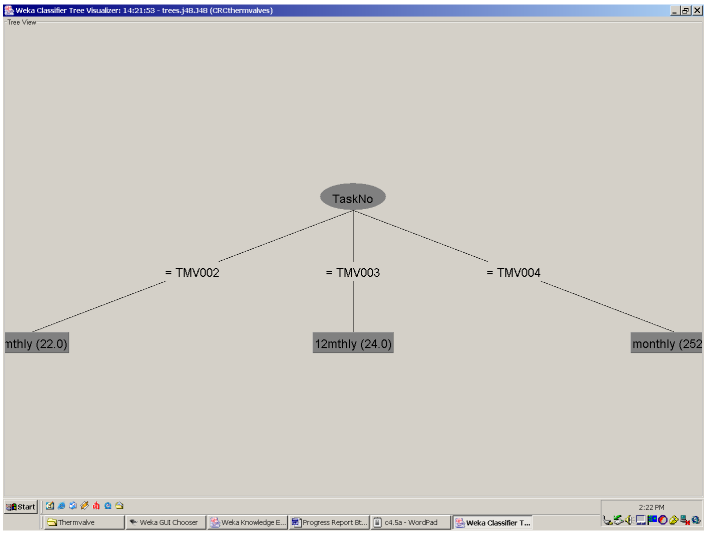


Figure 18. A Decision Tree of the application of C4.5 algorithm on “frequency” attribute.

Applying C4.5 on the “priority” attribute:

Rules associated with the decision tree from the application of the C4.5 algorithm on the “priority” attribute as shown in Figure 19 include:

1. High priority works take more than 0.5 hour to finish while those works with total duration of less than or equal to 0.5 hour belong to medium priority works.

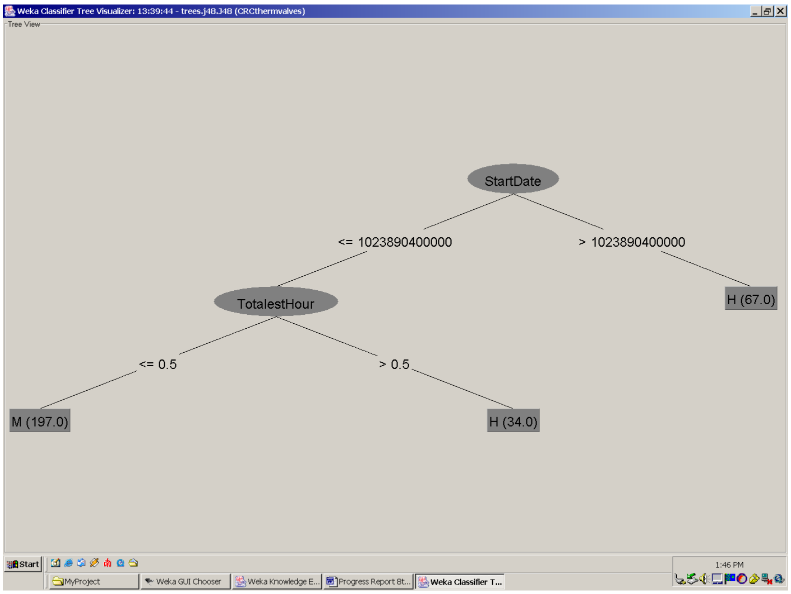


Figure 19. A Decision Tree of the application of C4.5 algorithm on “priority” attribute.

Applying C4.5 on the “month” attribute:

Rules associated with the decision tree from the application of the C4.5 algorithm on the “month” attribute as shown in Figure 20 and 21 includes:

1. All 6mthly works happen in December, and all 12mthly works are in June – September.
2. All high priority works are carried out in the later part of the year July to November.

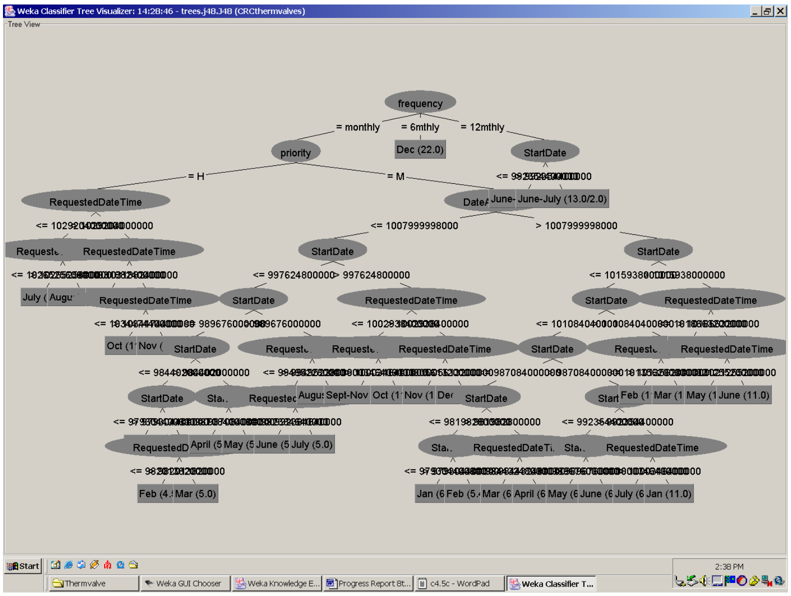


Figure 20. A Decision Tree of the application of C4.5 algorithm on “month” attribute.

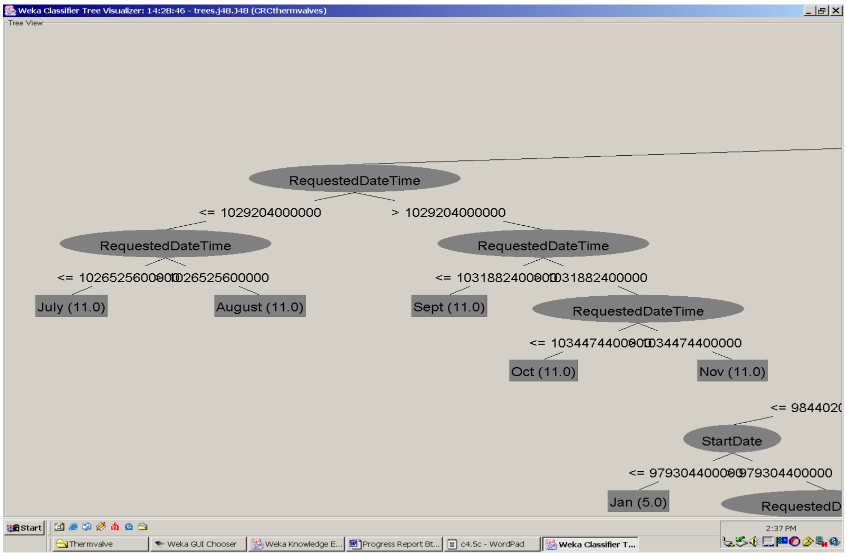


Figure 21. A portion of treeC to show all high priority works

Applying C4.5 on the “workorder_status”:

The Rules associated with decision tree from applying C4.5 on the “workorder_status” attribute as shown in Figure 22 include:

1. Maintenance jobs with WorkorderNo <=725085 are all completed.
2. 50% of 6mthly work did not finish (these unfinished 6mthly works’ WorkorderNo> 725085)

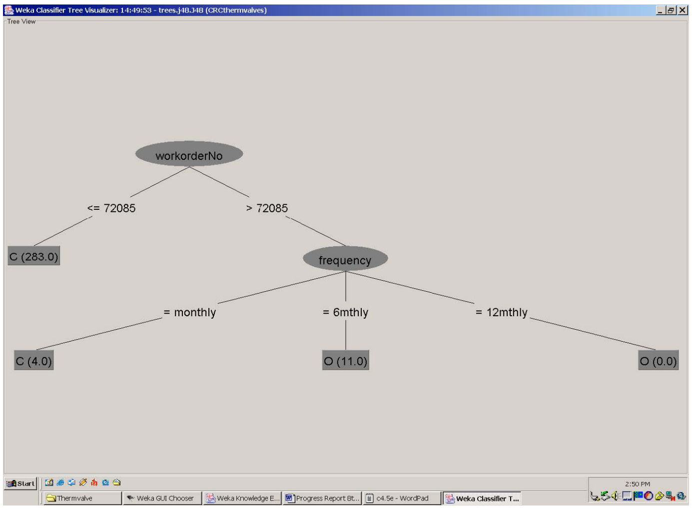


Figure 22. A Decision Tree of the application of C4.5 algorithm on "workorder_status" attribute.

Asset type 3: Battery Chargers

Applying C4.5 on the "completion_within_expectation" attribute:

The Rules associated with decision tree from applying C4.5 on the "completion_within_expectation" attribute as shown in Figure 23 include:

1. For all tasks with work order No > 66195, and some tasks with work order No between 48002 and 66195, completions did not meet the expectation of completion date;
2. There is a trend that the higher a work order number is, the lower the ability will be to meet the work order's deadline.

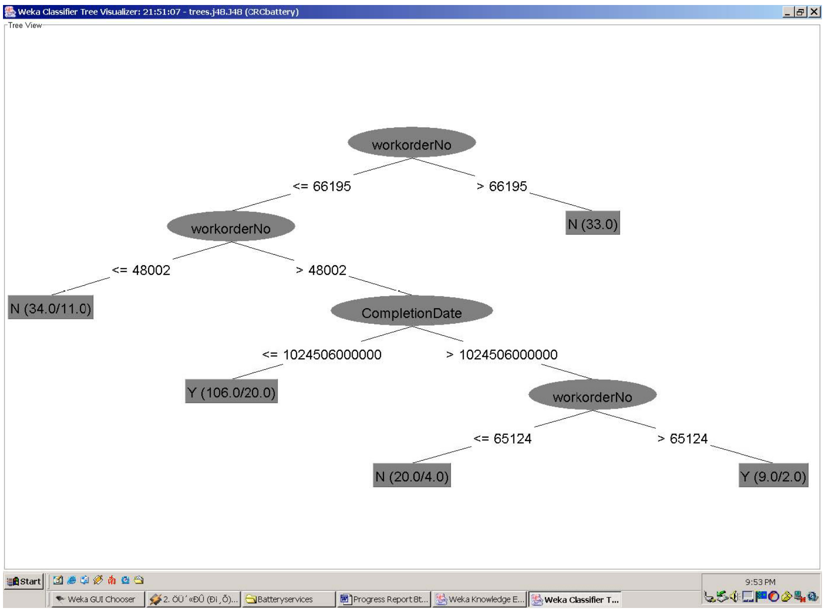


Figure 23. A Decision Tree of the application of C4.5 algorithm on "workorder_status" attribute.

Learned Rules from applying associative rule algorithm "Apriori" on maintenance data of the three asset types

Asset type 1: Air Handling Units

After applying the attribute evaluator and using the "BestFirst" search on the available maintenance data of the Air Handling units, there were 4 meaningful relational groups in which the association rule algorithm "Apriori" was applied to them. The learned rules include:

Group1: with the attributes floor, description_of_Cause and workOrder_Status

For floor 5,6,7, workOrder_Status always was completed
For all too_hot and too_cold descriptions, workOrder_Status was completed

Group 2: with the attributes priority, department, floor and cause_of_repair, description_of_Cause, and workormaterial.

All works in floor 7 belong to A/C malfunction
All works in department 26462 belongs to A/C malfunction
Department 21271 only resides at floor 6

Group 3: with the attributes priority, floor, completion_within_expectation

Maintenance jobs conducted at Floor 7 did not meet the expectations: with 23 out 25 completionwithinexpectation = "N" (92%)

Group 4: with the attributes job_Type, costcentre, cause_of_repair

96% jobs for costcentre = 0 is CM (corrective maintenance)

Asset type 2: Thermostatic Mixing Valves

After applying the attribute evaluator and using the "BestFirst" search on the available maintenance data of the Thermostatic Mixing Valves, there were 3 meaningful relational groups in which the association rule algorithm "Apriori" was applied to them. The learned rules include:

Group 1: with the attributes priority, month, total_est_Hour and budget

The rules associated with Group2 reflect the relationships between the work priority, the estimated time to complete the work and the associated budget.

priority=M 197 ==> Total_est_Hour=0.5 Budget=10 197 conf:(1)

Total_est_Hour=2 46 ==> priority=H Budget=29 46 conf:(1)

Group 2: with the attributes month, frequency, priority and completion_within_expecation.

The rule associated with associated with Group2 indicates the pattern of in relation to the work priority, level of meeting the expectation and the frequency of a specific task.

priority=M Completion_within_expectation=N 150 ==> frequency=monthly TaskNo=TMV004 150 conf:(1).

This rule indicates that 100% of medium priority works occurs on monthly maintenance and relates to a specific task TMV004.

Group 3: with the attributes completion_within_expectation, priority and month.

The rule associated with associated with Group 3 indicates that 96% of maintenance works which completions meeting expectation are of medium priority.

Completion_within_expectation=Y 49 ==> priority=M 47 conf:(0.96)

Asset type 2: Battery Chargers

After applying the attribute evaluator and using the “BestFirst” search on the available maintenance data of the Battery Chargers, there was one meaningful relational group in which the association rule algorithm “Apriori” was applied to it. The learned rules include:

Group 1: with the attributes completion_within_expectation, workorderstatus and month.

The rule associated with associated with Group 3 indicates 57% (107 out of 189) of workorderstaus = “c” was completed within the expected completion date.

Completionwithinexpectation=Y 108 ==> WorkOrderStatus=C 107 conf:(0.99)