Developing Software Agents for Data Mining: [AimM] Agents for Improving Maintenance Management

Report No. [2001-002-B + report no. 5]

The research described in this report was carried out by:

Project Leader: Professor John S. Gero

Project Manager: Dr. Rabee M. Reffat

Team Members: Mr. Wei Peng, Mr. PakSan Liew and Ms. Julie Jupp

Research Program No:

Program Name:

Project No.: 2001-002-B

Project Name: Life Cycle Modelling and Design Knowledge Development in Virtual Environment

Date: 31 March 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

Table of Contents

List of Figures

List of Tables

CRC Construction Innovation
BUILDING OUR FUTURE

1.   SUMMARY

This report demonstrates the development of:

- Development of software agents for data mining
- Link data mining to building model in virtual environments
- Link knowledge development with building model in virtual environments
- Demonstration of software agents for data mining
- Populate with maintenance data

2. DEVELOPMENT OF SOFTWARE AGENTS FOR DATA MINING "AimM"

"Good things of day begin to drop and drowse, whiles night's black agents to their preys do rouse."

William Shakespeare, *Macbeth*, Act 3, Scene 2

People have been doing business for long time. The only intelligent agent whom people had was another human being. A human being is currently the finest technology in the world. Unfortunately, human agents have 168 hours available to work in a week but can only work 48 hours; and at that figure would burn out quickly. We use human agents everywhere in our present day society. We use them to perform services, locate information, and get prices and other information so we can make some decision or can purchase a product or a service. These agents provide a valuable service that makes life easy for customers (Murch and Johnson, 1999). Computationally, an agent is anything that can be viewed as perceiving its environment through sensors and acting upon this environment through effectors (Russell and Norvig, 1995). If we define the environment as whatever provides input and receives output, and take receiving input to be sensing and producing output to be acting, every program is an agent. Thus, if we want to arrive at a useful contrast between agent and program, a distinctive characteristic of an agent is its interactions with the environment. A software agent is basically a software package that exhibits the characteristics of autonomy and goal-directed behaviour.

2.1 Revisiting the Interface Scenario of using Software Agents for Data Mining

The scenario shown in Figure 2.1 was proposed for the implementation of using software agents for data mining. This scenario assumed familiarity of the user with Data Mining techniques and algorithms and the structure of maintenance data. It requires certain abilities to deal with data processing as shown in Figure 2.2 and to examine the results of classification as shown in Figure 2.3. However, such approach is useful but its use might be limited and requires an extensive preparation and background knowledge from the user in Data Mining. This may not always be obtainable in the AEC (architecture, engineering and construction) industry. Hence, a better alternative is proposed to maximise the potential and utilisation of using software agents for data mining.
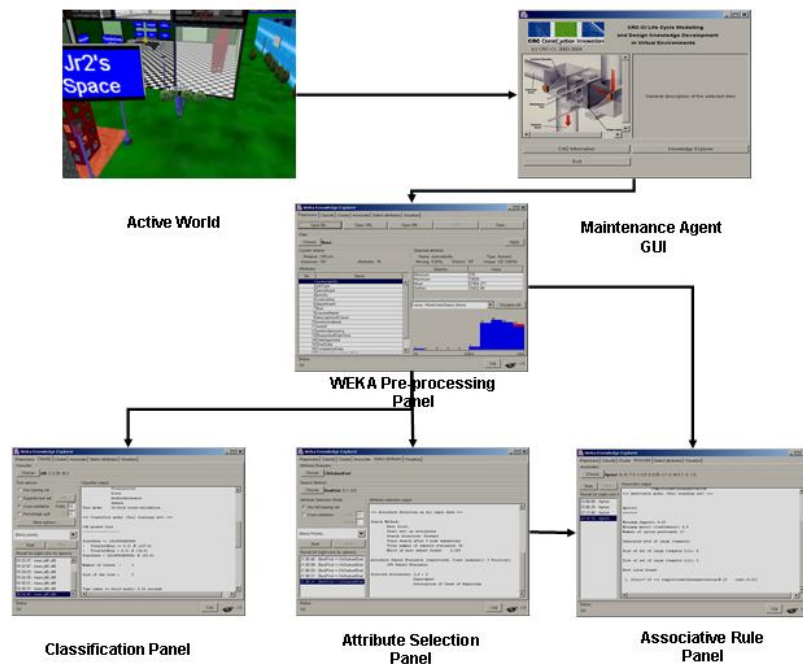
Figure 2.1 Previous Scenario of using software agents for Data Mining.
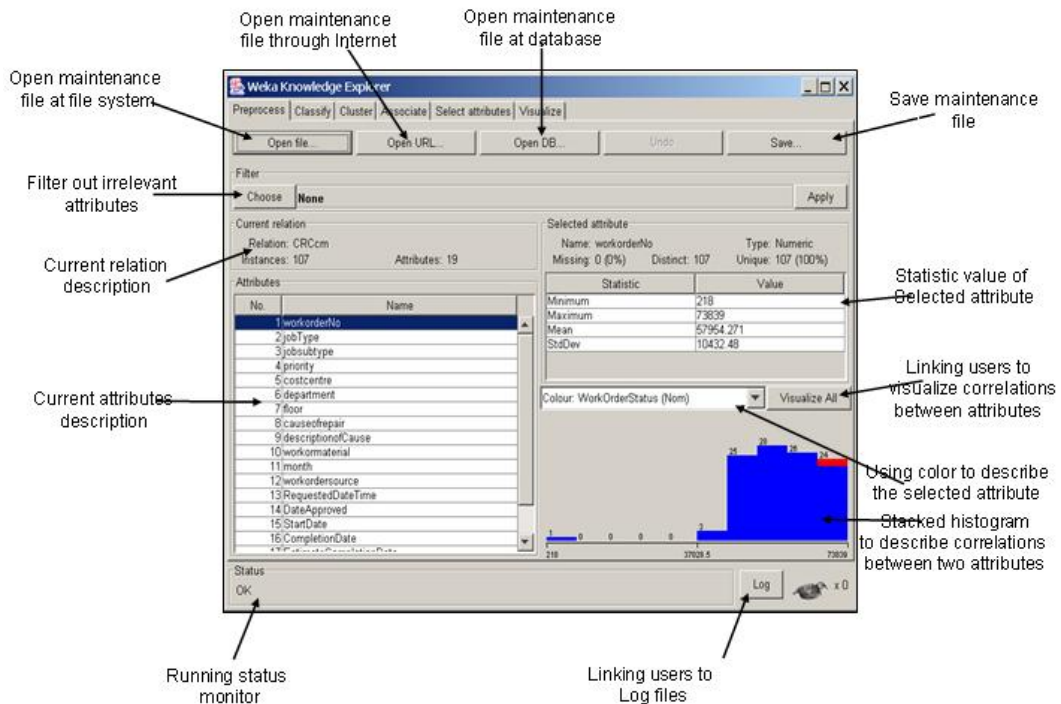
Figure 2.2 Pre-processing data in the previous of scenario using software agents for Data Mining.
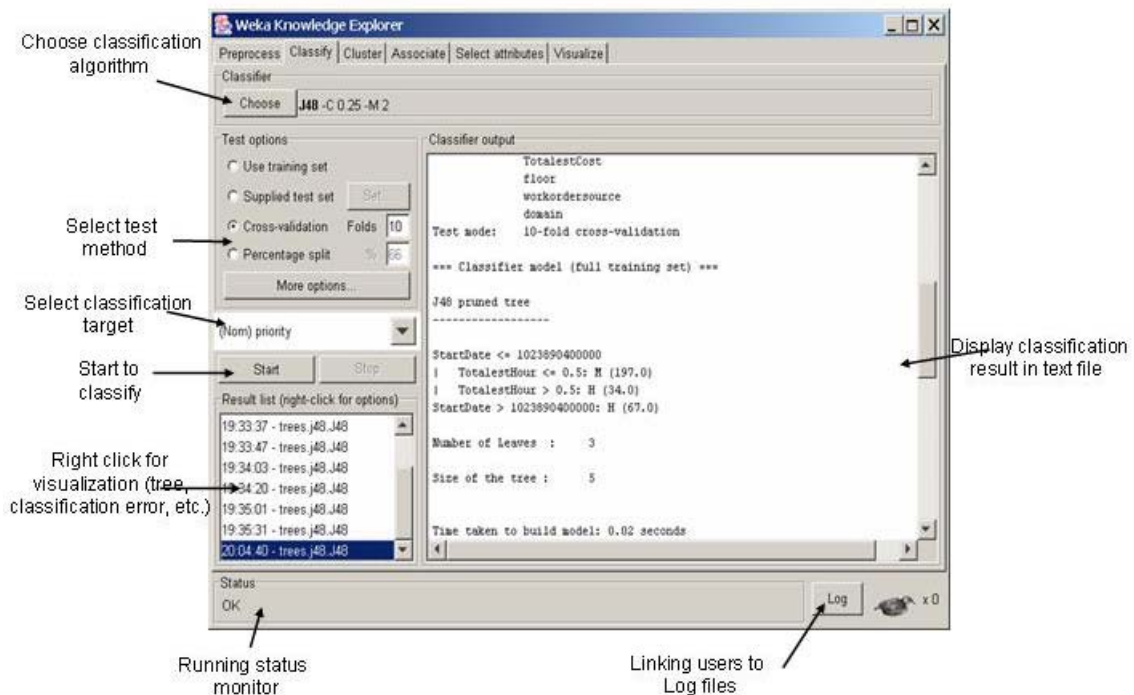
Figure 2.3 Examining classification results in the previous scenario of using software agents for Data Mining.

The new scenario does not require the user to be familiar with the techniques and algorithms but rather use them as a tool to improve the management of building maintenance. The architecture for AIMM (Agents for Improving Maintenance Management) has been modified and the Maintenance Agent is responsible to apply a set of selected Data Mining algorithms and provide the user in the virtual environment with patterns of knowledge discovered from mining the maintenance data. This knowledge could be utilised to improve the management of future building maintenance and operations. The new architecture of AIMM is illustrated in the following Section.

2.2   Architecture of AIMM

The architecture for AIMM (Agents for Improving Maintenance Management) is illustrated in Figure 2.4. The resources that should be made available for the software agents (AIMM) to function include 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. AIMM commences by converting the 3D CAD object model into IFCs (Industry Foundation Classes) to the EDM (Express Data Manager) database. The EDM interface agent converts the model into a Renderware (RWX) format and presents the 3D object model at the virtual environment (Activeworlds) as shown in Figures 2.5 and 2.6. The User may choose to navigate and walkthrough in the 3D object model at a real time virtual environment and may select a building asset to discover useful patterns of knowledge and utilise them to enhance the management of future maintenance operations. 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. One function of the Maintenance Interface Agent is to connect data contained within the Maintenance Database with data contained within the EXPRESS Data Manager (EDM) Database via the virtual environment, Active Worlds. The Maintenance Agent performs data mining and life cycle analysis on the data in the Maintenance Database through the Maintenance Interface Agent. In subsequent development of the system, the Maintenance Agent communicates directly with a Situated Agent to assist in improving maintenance management by providing life cycle implications as feedback whenever building assets (mechanical and electrical elements) are selected. A Situated Agent is basically a software agent that has the sole purpose of performing design actions. 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.
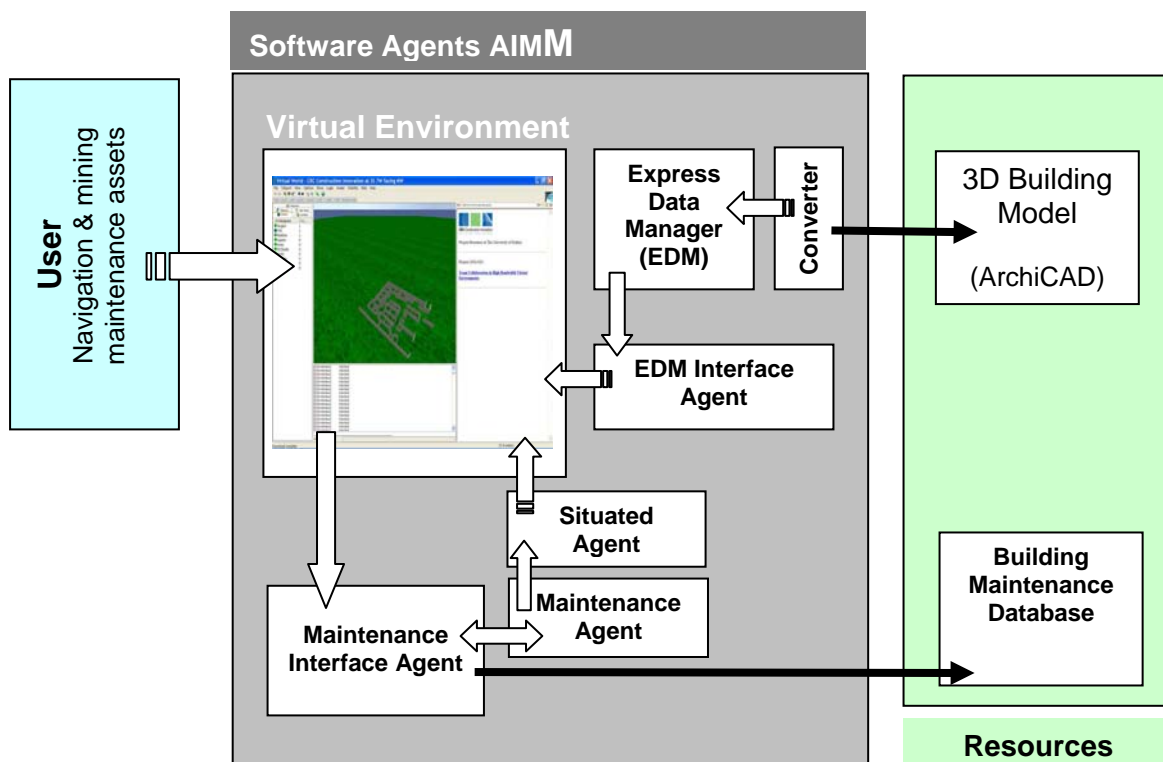
Figure 2.4 The architecture of AIMM (Agents for Improving Maintenance Management) for life cycle analysis of building maintenance.

ArchiCAD Education version, not for resale. Courtesy of Graphisoft

GRAPHISOFT®
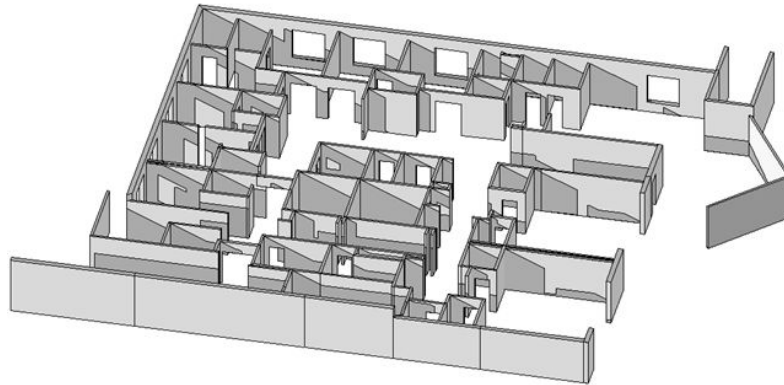
Figure 2.5 Part of the typical floor plan of an object oriented CAD model of building 10 at RPAH, Sydney modelled in the ArchiCAD package.
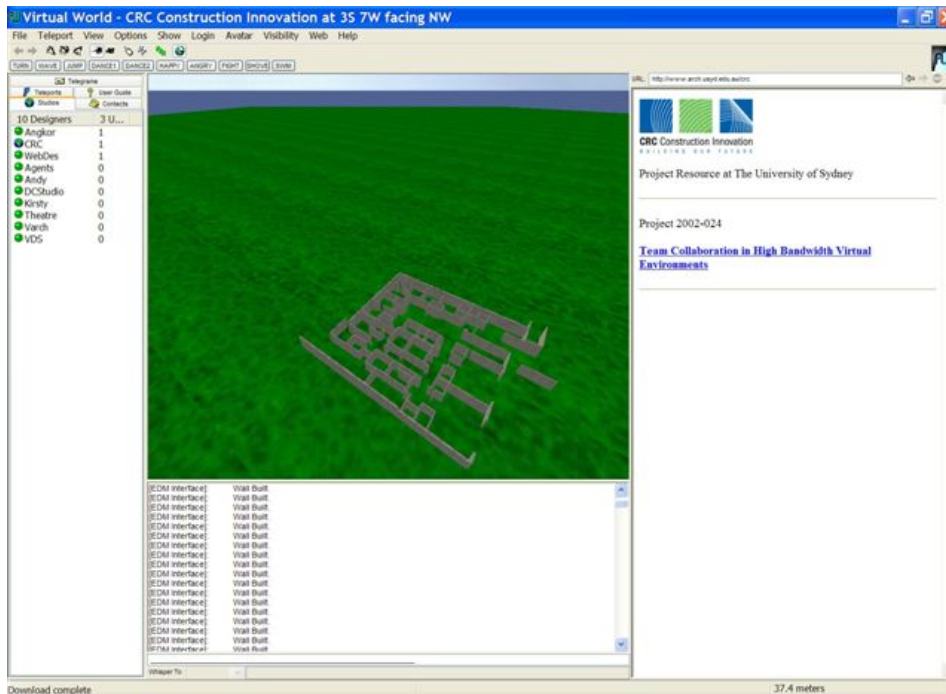
Figure 2.5 A 3D building model presented in the virtual environment (Active Worlds) after being converted by the EDM Interface Agent.

2.3   Link Data Mining to Building Model in Virtual Environments

Linking Data Mining techniques to building model in the virtual environment (Active Worlds) is achieved by the Maintenance Agent that accesses the maintenance database and applies its mining algorithms on it. However, the appropriate mapping between the building assets of the building model in the virtual environment is maintained by the Maintenance Interface 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.

2.4   Link Knowledge Development with Building Model in Virtual Environments

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). However, such feedback of useful knowledge is discovered by the Maintenance Agent that applies four data mining techniques and algorithms for discovering various classification of knowledge. The four mining algorithms are clustering using SimpleKmeans, learning rules by applying Associative Rule Algorithm "Apriori" and Classification using C4.5 and ID3. Mining algorithms and linking knowledge development with building model in virtual environments have been implemented within the system prototype of AIMM.

3.   DEMONSTRATION OF USING AIMM™ ON BUILDING NO. 10 AT RPAH

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

This Section illustrates a demonstration of running the system prototype of AIMM. This demonstration is applying AIMM on Building No. 10 at RPAH, Sydney. AIMM starts off 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, then AIMM 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 3.1. The results are presented in a single scrolling window. However as illustrated here in Figures 3.1,2,3,4 and 5 that show the complete results as the scroll bar is used. The results shown are the aggregation of all results produced by applying each of the four Data Mining Techniques.
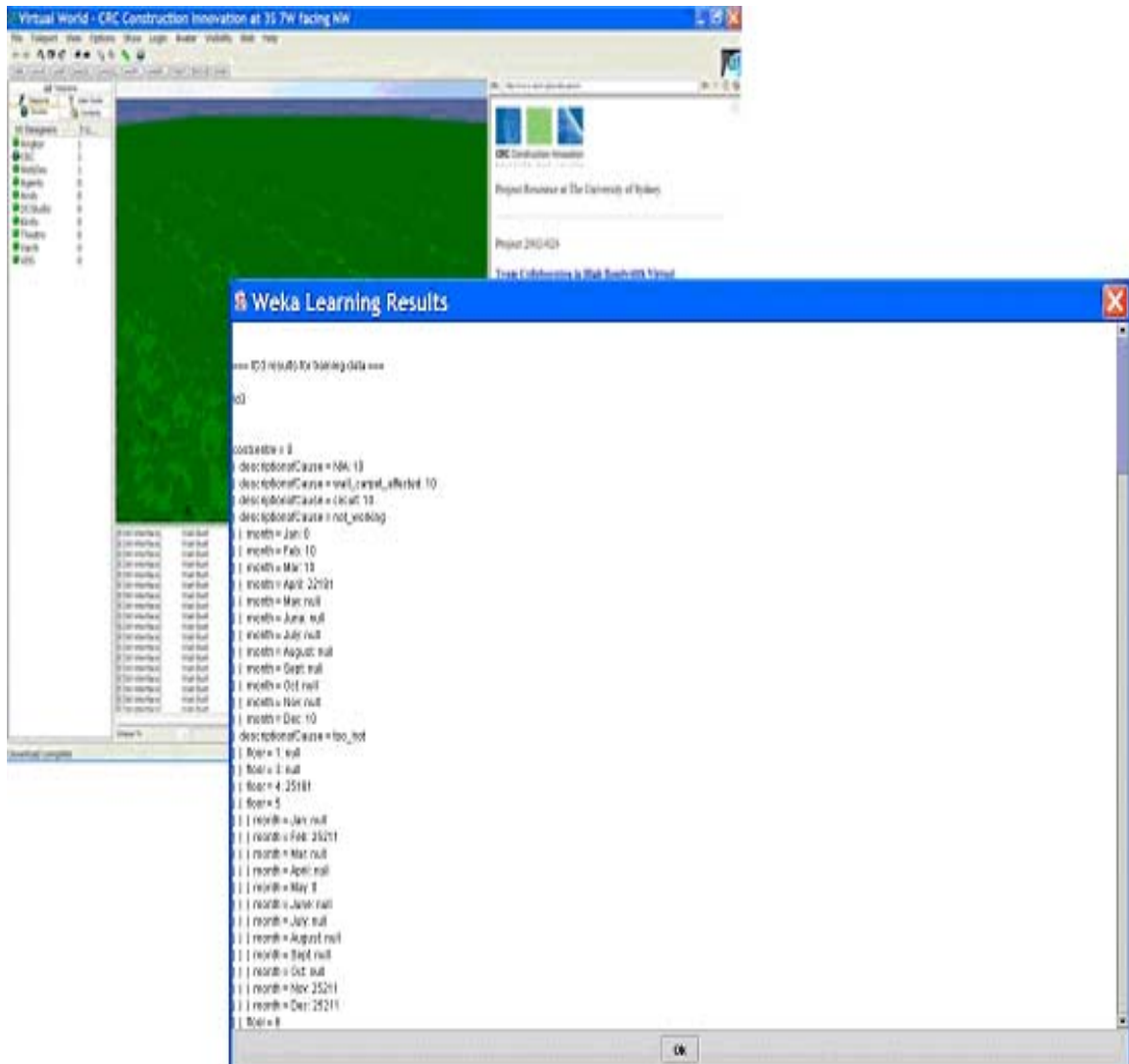

Figure 3.1 Demonstration output (display Part I) of applying Data Mining on the maintenance data of the Air Handling Unit at Building no. 10, RPAH.

```
🔹 Weka Learning Results                                                    X

| floor = 6
| | workormaterial = N/A: 21271
| | workormaterial = Valves_closed: null
| | workormaterial = Circuit_locate: null
| | workormaterial = Adjust: 10
| | workormaterial = Valves: null
| | workormaterial = Actuator_motor: null
| | workormaterial = A/C_Registers: null
| | workormaterial = Check_VSD_Supply: null
| | workormaterial = Lock: null
| | workormaterial = Vents: null
| | workormaterial = Motor: null
| | workormaterial = cut_hole: null
| | workormaterial = V/Belt: null
| | workormaterial = shut_down_air_flow: null
| | workormaterial = Check: null
| | workormaterial = Reset: null
| floor = 7: 0
| floor = 9: null
| floor = 10: null
descriptionofCause = too_cold
| month = Jan: null
| month = Feb: null
| month = Mar: null
| month = April: null
| month = May: 12
| month = June: null
| month = July: 10
| month = August: null
| month = Sept: null
| month = Oct: 0
| month = Nov: 21271
| month = Dec: 10
descriptionofCause = fire_trip: 10
descriptionofCause = faulty_actuator_motor: 10
descriptionofCause = shut_down: null
descriptionofCause = air_velocity_alarm: 10
descriptionofCause = dust_out: 21271
descriptionofCause = VSD_Supply: null
descriptionofCause = lock_on_firedoor: 10
descriptionofCause = blowing_hard: null
descriptionofCause = cut_hole_ceiling: null
descriptionofCause = shut_down_air: 10
descriptionofCause = internal_drains_blocked: null
descriptionofCause = noise: null
descriptionofCause = after_fire_alarm: 10
descriptionofCause = A/C_sensor: 10
descriptionofCause = cold&noise: 25191
descriptionofCause = Motor_burnt_out: 10
descriptionofCause = V/Belt: 10
descriptionofCause = A/C_Registers: null
descriptionofCause = Vents_Taped_Up: null
descriptionofCause = smelly: null
descriptionofCause = cover_off: null
descriptionofCause = water_leaking: 22392
descriptionofCause = faulty_valves: 10
costcentre = 14404

                              Ok
```

Figure 3.2 Demonstration output (display Part II) of applying Data Mining on the maintenance data of the Air Handling Unit at Building no. 10, RPAH.

Figure 3.3 Demonstration output (display Part III) of applying Data Mining on the maintenance data of the Air Handling Unit at Building no. 10, RPAH.
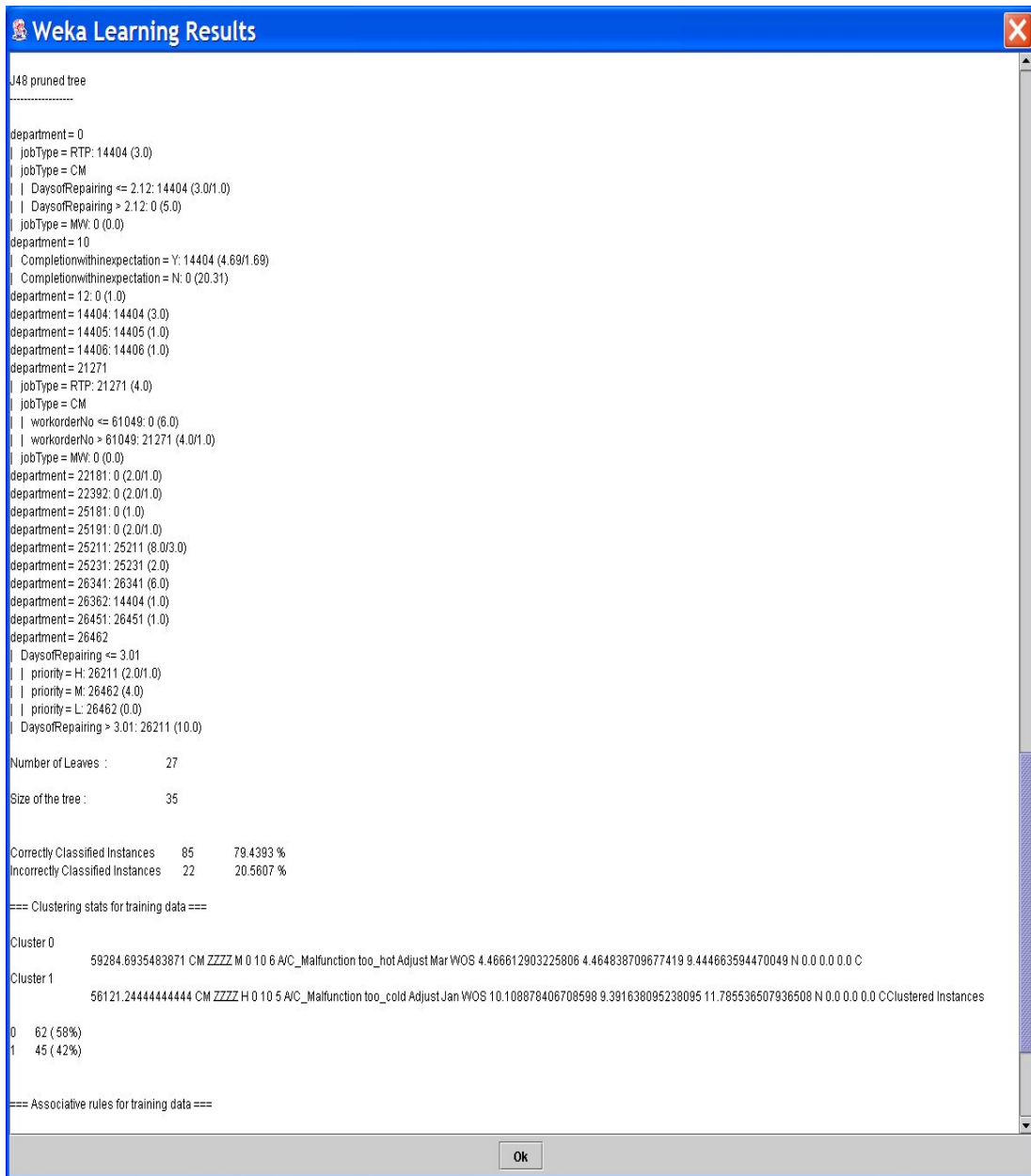
Figure 3.4 Demonstration output (display Part IV) of applying Data Mining on the maintenance data of the Air Handling Unit at Building no. 10, RPAH.
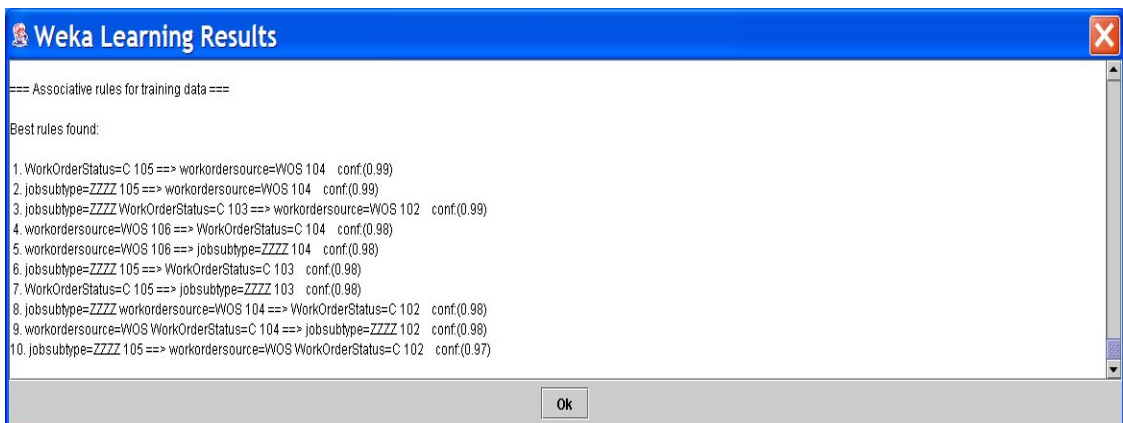

Figure 3.5 Demonstration output (display Part V) of applying Data Mining on the maintenance data of the Air Handling Unit at Building no. 10, RPAH.

3.2   Limitations of the result in this Demonstration

From the result generated through a single run of these algorithms, it can be observed that the rules generated are not very interesting. The rules that were obtained and described in the second report (an example shown in Table 3.1) were pulled out manually in a strategic manner. The pull process involved cleaning and modification of data files; selecting various attributes alternatively. For Instance, when a data file is feed into associative rule algorithm, meaningless and less useful rules are 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 as illustrated 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 3.1. 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.

4. IMPROVING THE RESULTS OBTAINED FROM USING AIMM™

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. A more realistic scenario is to communicate with the data mining system, using additional questions to examine the findings and direct the mining process; some of these questions 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 quickly change variables to be included in the mining run, in combination with filters that can be defined for all variables; e.g. only focus on times with a resultant temperature above 27^{0}C, occupied period, time of high occupancy densities, etc.).

4.1 Some Scenarios for Improving the results obtained from using AIMM for the Air Handling Unit

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.

  o 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

  o 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.

  o 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

o 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**

4.2 Including Filtering Scenarios into AIMM architecture

Filtering scenarios should be investigated for each building asset type that is required to be mined. These scenarios will be integrated within the structure of the Situated Agent as shown in Figure 4.1 in order to improve the results of using AIMM to subsequently improve building life cycle, maintenance and operation.
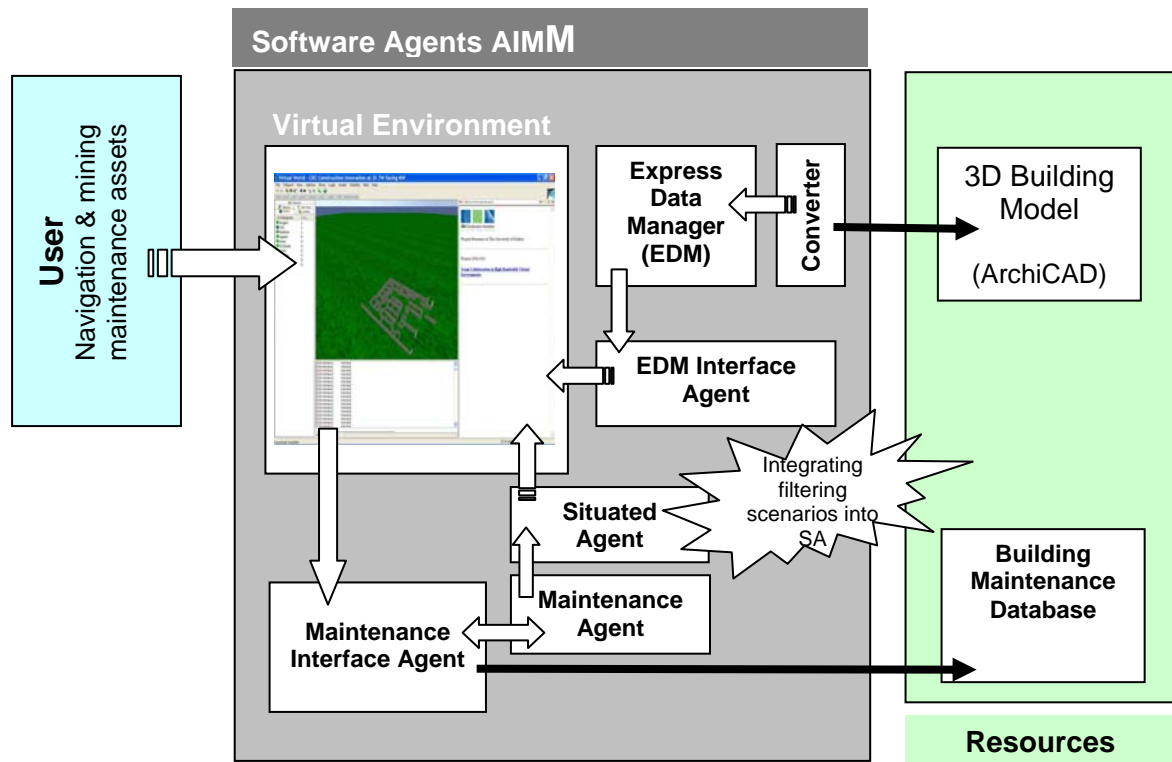

Figure 4.1 Integrating filtering scenarios in the structure of the Situated Agent in AIMM.

5.   CONCLUSION

A sound and useful architecture and scenario of using software agents for data mining on building maintenance have been developed and implemented in an AIMM system prototype. The usefulness of the revised 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 developers should exert their efforts reflecting on the discovered knowledge from applying data mining on the records of maintenance data and utilise these results in improving the management of maintenance and operation of buildings and enhancing building life cycle.

In order to improve the results obtained from applying AIMM on maintenance data, appropriate data mining scenarios and filters are needed to provide the required knowledge. Appropriate data mining scenarios and filters to the task have been proposed and will be further developed and implemented in AIMM system prototype in the coming two quarters of the Project.

Life cycle cost (LCC) analysis has been introduced to provide strategic planning on refurbishment and enhances information for decision making. Appropriate life cycle cost formulas have been proposed and selected for building assets and components. The architecture of AIMM has been adopted to include an LCC Agent and will be implemented in the AIMM system prototype in the coming two quarters of the Project.

The application of AIMM system prototype on the available maintenance data has been utilised on three asset types (air handling units, thermostat mixing valve and battery charger). The AIMM system prototype will be applied to populate with maintenance data of building no.10 at Royal Prince Alfred Hospital, Sydney.

The deliverables in the coming quarter (due on 30 June 2004) include: (a) demonstration of software agents prototype system, (b) draft of research paper on agents-based approach to maintenance data mining, and (c) develop a proposal for project submission on extension to time-based modelling.

6. REFERENCES

Morbitzer, C., Strachan, P., and Simpson, C.: 2003, Application of data mining techniques for building simulation performance prediction analysis, *Proceedings of Building Simulation for Better Building Design*, Eindhoven, Netherlands.

Murch, R. and Johnson, T.: 1999, *Intelligent Software Agents,* Prentice Hall, Upper Saddle River, New Jersey.

Russell, R. and Norvig, P.: 1995, *Artificial intelligence: A Modern Approach,* Prentice Hall, Englewood Cliffs, N.J.

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