The Analysis of Data Collected by Time and Attendance Systems

Tomasz Jędrzejewski[1], **Bogdan Trawiński**[1], **Aleksander Zgrzywa**[1]

Abstract: Time and attendance software systems are tools for efficient management of labour resources and accurate labour reporting. The system implemented using Microsoft .NET environment is presented in the paper. Analytical capabilities of data collected by the system by means of business intelligence platform of SQL Server 2005 are considered. Some examples charts produced by data mining tools including OLAP, time series and decision trees are given.

Keywords: Time and attendance systems, labour management, OLAP, data mining.

1. Introduction

Time and attendance systems are software tools for efficient management of labour resources and accurate labour reporting. They are powerful time clock systems that collect labour data for both payroll and job costing. They enable managers not only to make cost-effective decisions but also to determine the best way to optimize their workforce. They support different methods of data collection e.g. time clock oriented, manager or timekeeper oriented, and employee oriented. Collecting options include magnetic strip, proximity, barcode, biometric or smart card readers, phone, computerized timesheets, PC or web login/out applications. Such systems automate data collection, eliminate errors and duplication, allow supervisors to easily and quickly authorize and analyze time worked, overtime, holidays, to schedule shifts, and to track employee's compliance to start/stop times. They also improve labour management through comprehensive data analysis and reporting. The latter aspect is the main issue of the present paper.

The system presented in this paper was implemented to cooperate with Microsoft SQL Server 2005 and it was quite natural to explore the capabilities of its business intelligence platform to produce sophisticated and useful analyses that go beyond mere late coming and attendance reports. Microsoft Business Intelligence solutions provide tools for data integration from many sources for data analysis and data reporting. It is possible to analyze data in Microsoft Business Intelligence platform using data warehouses or data mining or both. In SQL Server 2005 there were added more data mining algorithms and the scalability of the platform was extended.

[1] Instytut Informatyki Stosowanej, Politechnika Wrocławska, Wyb. Wyspiańskiego 27, 50-370 Wrocław
e-mail: tomasz@skalmex.com.pl, {bogdan.trawinski, aleksander.zgrzywa}@pwr.wroc.pl

2. Architecture of the Time and Attendance System

The Access Control and Time and Attendance System was implemented using Microsoft .NET environment with SQL Server 2005 as a database management system. It was assumed the system could be exploited using commercial editions of the SQL Server 2005 as well as the free SQL 2005 Express Edition. The limitations of the latter to 4 GB of the database size and the RAM size of 1 GB are not essential for medium and small organizations, so that the system can be deployed in a broad spectrum of companies and institutions. The main functions of the system are as follows:

- tracking of late coming and absence including holidays, sick leave, training, business trips, changed/unscheduled shift work etc.,
- defining individual calendars and schedules,
- calculation of time worked during working days and holidays,
- balancing holidays, work time and overtime,
- producing reports including employees' absence, overtime or punctuality by personnel criteria such as department, position, day, month etc.,
- integration and export of data to payroll, HR and other systems.

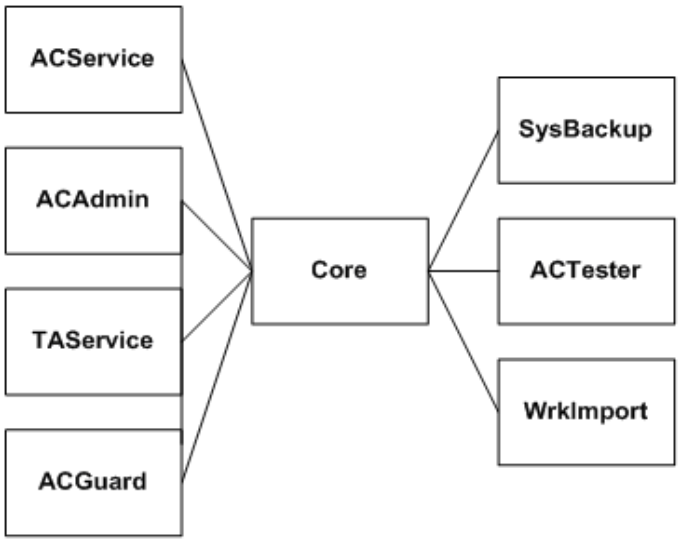

Figure 1. Module architecture of the system

The module structure of the system is depicted in Fig. 1 and its main applications are as follows:

- Core – contains libraries used by other applications, the libraries include among others converter, database access, and communication classes,
- ACAdmin – is the application devoted to access control system administration, and allows for configuration of ports and controllers, granting user rights, introducing employees, browsing event and alert history,
- ACService – performs the communication with online controllers and deals with all controller events,
- TAAdmin – it is the application designed for the time and attendance system administration,
- ACGuard – reports where at a given moment each worker stays and how many persons there are in individual rooms and places,
- WrkImport – enables to import data of employees from other systems and common formats, simplifying therefore initialization of the system.
- ACTester – is used during deployment and enables to test quickly in field conditions all equipment: online controllers, card readers etc.
- SysBackup – enables the user to make database backup by pressing only one button.

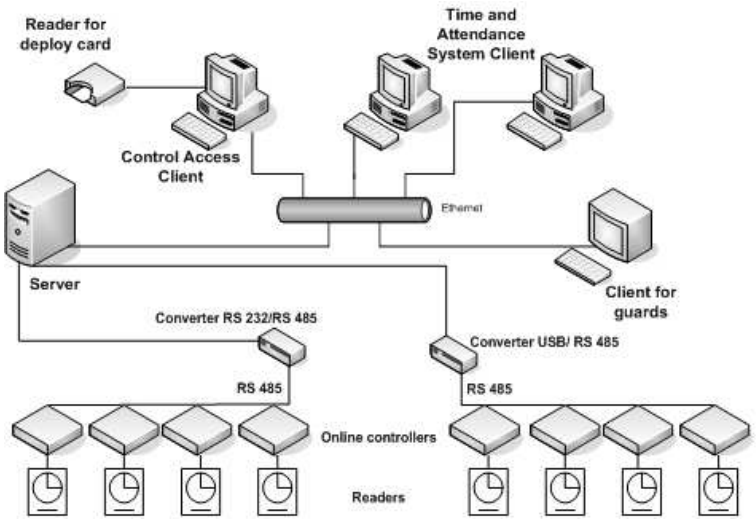

Figure 2. The deployment diagram of the system

The deployment diagram of the system is shown in Fig. 2. The employees' card readers are combined with clocks. They are capable of reading Unique cards as well as programmable Hitag 2 cards. The readers are connected to the controllers

which communicate with the server through the RS 485 bus, and this connection is encoded by means the AES-128 protocol. Due to the computers are not equipped with the RS 485 input, various converters can be used e.g. RS 232/RS 485, USB/RS 485, and Ethernet/RS 485. All the controllers are managed by the server to which they are connected. The service managing the readers runs on the server and all events and alerts are stored in the SQL Server 2005 database. The database is used by client applications such as control and access, time and attendance and guard ones. The reader deploying cards for new employees registered in the system is connected directly to the computer with ACAdmin or TAAdmin application through the USB port. An employee when entering or leaving the controlled area controlled area approaches his card to a reader and his identifier is read and sent to the server. The server application checks whether the card holder is authorized to enter the area. If so the card punch event with its date and time, type and the worker's identifier is recorded in the database. The basis of the reports generated for the management purpose are aggregates calculated for individual employees and for each day. They comprise:

- TT – total time of the attendance at the company,
- WT – worked time,
- NT– nominal work time,
- LO – official leave,
- O1 – overtime of 50
- O2 - overtime of 100
- DN –work during night,
- DS – work during Sunday or holiday,
- DO – work during additional days off,
- DV – work during vacation,
- DD –work during duty,
- LC – late coming,
- AB – absence.

The TAADmin and ACAdmin applications are ready to serve a big office building occupied by several companies, because they are equipped with a especially designed right management system.

In order to design usable user interface a survey on 10 experienced users was carried out. The results showed that in most systems the flexibility of work calendars

were not sufficient, and the systems missed appropriate overtime calculation. After further study, the designers took care of navigation, worker searching, date input, because these operations can impact on the work with the application. The system interface was designed to resemble other Windows applications, so that the navigation is similar to the internet browser's one. In order to make system applications uniform, icons from www.MouseRunner.com with Creative Commons Attribution-Share Alike 2.5 licence were used. In the Time and Attendance System substantial amount of operations are performed on an employee, therefore there is a search mechanism which enables to filter the list of workers the in each window of the system. The other aspect carefully designed was the way of date input. In this case the DateTimePicker control was used.

3. Analysis of data collected by the System

The above described access control and time and attendance system is quite new and data collected in its database up to now covered the period no longer than about two or three month. Therefore in order to be able to present the data analysis potential, data from the system, which was the ancestor of the new one and had been exploited for several years, were imported. The data were taken from the industrial company manufacturing furniture and employing about 3000 workers. Altogether, data from the period of 2.5 years were analyzed. As an analytical tool SQL Business Intelligence Development Studio was used. After preparing the data a cube with measures: late coming, work time, time of staying within a company area, night work, and with dimensions: sex, position, date of birth, first hierarchical time dimension comprising: year, month, day and second hierarchical time dimension containing year and week was created.

3.1. Late coming analysis using OLAP

The chart in Fig. 3 shows time of coming late for two positions, namely master and storekeeper, in successive weeks of a year. The user can select positions in order to compare how employees of different positions comply with the work time rules.

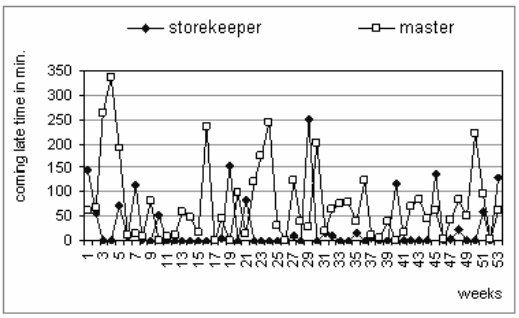

Figure 3. Late coming comparison between two job positions

3.2. Worked time analysis using OLAP

In Fig. 4 worked time of a given position in successive months of a year is depicted. The chart enables the user to analyze how the total worked time changes over time. In the presented case you can observe the trend to lengthen worked time at the end of a year.

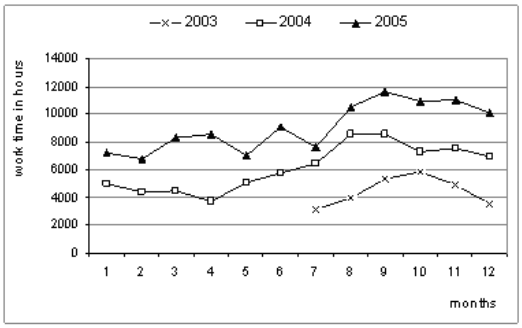

Figure 4. Worked time in individual months of a selected position

3.3. Predicting the number of employees

The Microsoft Time Series algorithm was used to predict mean number of employees in subsequent weeks. The algorithm applied the linear regression decision tree approach to analyze time-related data, such as monthly sales data or other continuous variables and is based only on the trends that the algorithm derives from the original dataset while it is creating the model. Forecast was performed on the basis of the average number of employees in each week within period of two years 2004 and 2005. The number of employees was predicted for 2006 year. Actual data for the first 8 weeks of 2006 enabled to set following parameters of the algorithm to obtain the most reliable predictions:

- PERIODICITY_HINT was set to 23, because basing on historical data, it was assumed that the mean number of workers changes every half year.
- AUTO_DETECT_PERIODICITY - the default value equal 0.6 was changed to 0.8 because setting this value closer to 1 favors the discovery of many near-periodic patterns and the automatic generation of periodicity hints,
- COMPLEXITY_PENALTY was raised from 0.1 to 0.15, because it controls the growth of the decision tree and increasing this value decreases the likelihood of a split
- MINIMUM_SUPPORT was set to 1, this parameter specifies the minimum number of time slices that are required to generate a split in each time series tree and its default value equals 10.

The resulting model is shown in Fig. 5, where the left part of diagram represents the actual number of workers and the right part shows the prediction.

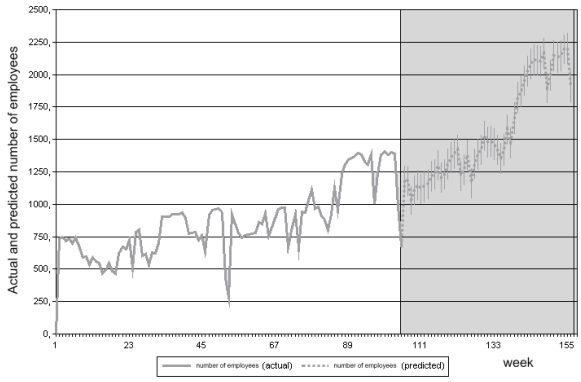

Figure 5. Actual and predicted number of employees

3.4. Factors influencing average work time

Decision trees model was applied to explore which factors influence mean work time of an employee. Data referring to 1000 employees were used to train the model to predict mean work time for the following attributes: position, birth date, sex, mean rate of coming late, total value of coming late, nominal number of working days, actual number of working days. Bayesian Dirichlet Equivalent Prior method was used to calculate the split score, binary and complete method was applied to split the nodes and the minimum support parameter was set to 25. In Fig. 6 a dependency network obtained as the putput of decision trees algorithm is depicted. The parameters are numbered according to their decreasing impact. The strongest effects revealed the actual number of days worked and the position.

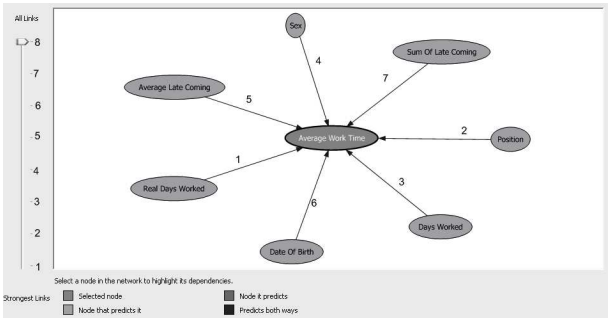

Figure 6. Dependency network of the factors influencing average work time

4. Conclusions and Future Work

Raw data collected by a time and attendance system are valuable for labour management. Data mining tools such as Microsoft SQL Business Intelligence Development Studio make it possible develop applications enabling to meet business information needs in a flexible, comprehensive, intuitive manner. Using reports produced by these applications supervisors could for example monitor their department's performance in terms of absence, overtime, lateness, etc., against the company's average and determine whether necessary steps need to be taken to reduce these incidences. By analyzing reasons for overtime, management can assess whether more employees are required to work in specific positions. Trend analysis enables managers to predict the number of employees available in a given period in the future. Above examples do not exhaust truly broad spectrum of possibilities ranging from simple lists to sophisticated analytical reports.

References

Clifford, C. (2003) *Implementing Time and Attendance Systems: The City of Scottsdale* Arizona Experience. Government Finance Review, Vol. 17, Issue 1.

Praca zbiorowa pod red. dr inż. Andrzeja Wójcika (2001) *Mechaniczne i elektroniczne systemy zabezpieczeń* Verlag Dashöfer sp. z o.o, Warszawa 2001.

Time and Attendance White Paper. Choosing a Time and Attendance Product (2003) WorkForce Software 2003, `www.workforcesoftware.com`.

Payne A (2006) *Business Intelligence and Data Warehousing in SQL Server 2005* Microsoft 2006, `http://www.microsoft.com/technet/`.

ZhaoHui Tang, MacLennan J. (2005) *Data Mining with SQL Server* Wiley 2005.