Construction of a Knowledge-Based Management System for Organizing Energy Management and Indoor Environmental Quality (IEQ) Related Data

Sean Hsieh, University of Nevada, Las Vegas, USA; E-mail: hsiehht@nscee.edu
Kenny Kwan, University of Nevada, Las Vegas, USA; E-mail: kwanl@unlv.edu
Linda D. Stetzenbach, University of Nevada, Las Vegas, USA; E-mail: linda.stetzenbach@unlv.edu
Davor Novosel, National Center for Energy Management and Building Technology, USA; E-mail: dnovosel@ncembt.org

ABSTRACT

The National Center for Energy Management and Building Technology (NCEMBT) initiated a series of research projects related to the indoor environmental quality (IEQ) and energy performance of existing buildings. These projects have generated a vast amount of data on the operation of commercial and institutional buildings. The goal of the NCEMBT is to make these data available online via a knowledge-based resource management system (KBRMS). This paper describes the construction of the KBRMS including the data flow analysis, the data collection and assimilation process and the design, development and prototype demonstration of the KBRMS. The beta version of the KBRMS is operational. The system is divided into public and private sections with the enforcement of role-based access control (RBAC). The external public web portal displays general project information and the research-related publications while the internal web portal accesses project-related data analysis tools, analysis results and intranet communication modules. This paper shows some summary statistics that are automatically calculated through the pre-defined analysis tools using MS SQL Sever stored procedures. More field monitored building data will be put into the system resulting from future projects.

INTRODUCTION

Knowledge discovery in database (KDD) and data mining are emerging fields for extracting useful knowledge from volumes of data. As defined by Fayyad (1996a), "the KDD process is a nontrivial process of identifying valid, novel, potentially useful and ultimately understandable patterns in data". Data sets have little value unless meaningful knowledge can be extracting from the data (Fayyad, 1996b). The multi-step KDD process involves data integration, preparation and transformation, data mining and evaluation, and data visualization. A general framework for the KDD process, proposed by Ingolf (2002), identifies the relationship among management (process view), mining models (model view) and datasets (data view).

The NCEMBT initiated a series of research projects related to the IEQ and energy performance of existing buildings. Researchers from the University of Nevada, Las Vegas are major participants in these multi-year tasks. Each research project generates significant amounts of data. These data may be in form of building IEQ measurements, occupant perception questionnaire responses or underfloor air distribution (UFAD) flow simulation. NCEMBT is obtaining these data sets to transform the built environment by delivering applied research and educational information that results in sustainable building systems which are efficient, productive, secure and healthy.

The true value of the data does not lie within each set but rather in the sum of all the sets. The relationships contained between the individual building data sets are of most value not only to the NCEMBT researchers but also to other research entities, such as the American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc., the U.S. Environmental Protection Agency and private companies. Statistically significant relationships can only be developed and derived from the associations that exist among all the data sets. For example, a statistically valid correlation between ventilation rates and occupants' perception of the indoor environmental quality of their workplace can only be derived from the analysis of complete datasets of several buildings. However, to do so the existing datasets need to be upgraded to a data warehouse, which in turn will allow data mining. The development of the KBRMS is the first step towards the goal of data mining of large building sciences data sets. The concept of organizing and interpreting the discovered rules through the web as proposed by Ma (2000) will be used for this ongoing project.

Organizing such abundant, non-structured data on energy and IEQ performance of buildings using a well-designed, highly flexible internet-based management system, significantly increases the availability the inherent information resulting in greater knowledge generation. As discussed by Imielinski (1996), research on knowledge discovery scenarios can initially be achieved through the construction of performance-driven system and eventually move on to a complete knowledge and data discovery management system. The goal of this project is to create a system with KDD capability using results from multi-year research studies. This paper examines the flow of data collection and assimilation process, the design and development of the resource management system and presents the current stage of the KBRMS.

SYSTEM DESIGN

Based on its industrial experience, the NCEMBT has proposed an organizational framework to accommodate its data flow as shown in Figure 1. Because each functional area consists of various subjects, an analysis of the data flow needed to be completed prior to the implementation of the web-based resource management system. The analysis was followed by the design and development of the portal architecture.

DATA COLLECTION AND DATA FLOW

Information from NCEMBT projects was modularized and dissected into categories using the classification scheme depicted in Figure 1. Each category was further divided into (sub)levels depending on the underlying data structures (Figure 2). Such a classification approach provides strong data assimilation and reusability capabilities. Accessibility to individual modules was designed to be controlled by the user right definition. The system integrates and organizes all available information under designated repositories based on a customized user definition.

Figure 1. Requirement analysis performed based on the NCEMBT's functional areas

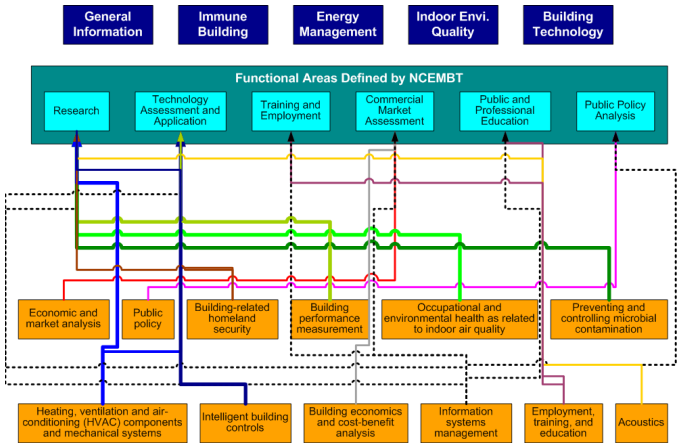
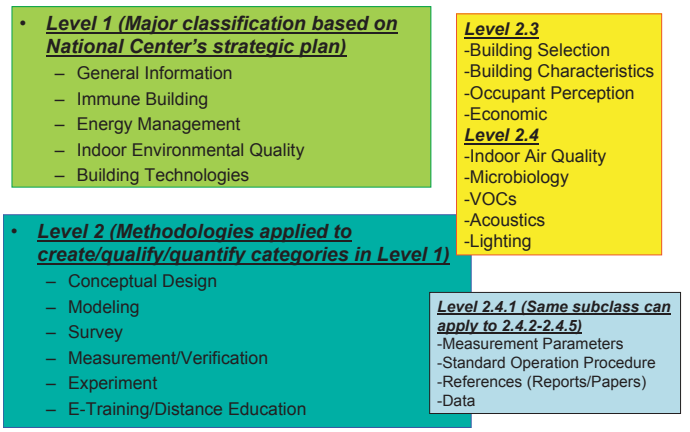


Figure 2. Hierarchy definition used for the portal system



Therefore, the same resources can be accessed by various users at pre-defined data resolutions. One primary development strategy was to minimize the administration effort for content managers and maximize the utilization of analysis and knowledge retrieval tools.

DATA FLOW

Analysis of the data flow as shown in Figure 3 defined the required modules and interfaces needed for the portal system. The data flow follows the path by which public users access the general information without undergoing a registration process. This path is also followed by the content manager to log in. This system is designed to allow

- building managers who participated in the building monitoring projects to access their respective building data and compare the performance of their building to all others contained in the data set
- users to download hourly averaged data sets through the web portal; and
- project team members to access all raw data up to a 10-second resolution.

SYSTEM DESIGN AND DEVELOPMENT

The web portal covers materials from passive information, such as general project introduction and organizational summary, to data-centered information pages, such as results of building occupant surveys or IEQ measurements.

A multi-tier application architecture was employed. The portal contains two data sources. The configuration settings are stored in an xml-formatted file while the content for the application is stored in a SQL Server database. Data access is provided through a Microsoft .NET assembly using the stored procedures. The portal framework was built using of a number of components that handle both the security and portal configuration information. The presentation layer is organized by several web forms and user controls that handle the display and management of the portal data for end users.

All portal contents are stored in MS SQL Server databases. The basic approach uses an xml file which contains all the configuration settings for the portal. This method provides a simple and easy way to manipulate configuration settings. The configuration file is used to store all high level portal, tab and module definitions.

The portal development was divided into three stages: database development, portal functionality module development and portal skin development. The knowledge management portal has two level of accessibility: external public web access and internal management tools. The internal management tools provide an efficient way for the content manager to update the web information. MS SQL Server was selected as the major data repository. Portal development and cosmetic presentation use MS .NET technology, including ASP.NET, VB.NET and XML.

Role-based access control (RBAC) is a proven technology for enforcing separation of duties, administration and access for web-based system. Ahn (2000) developed models, architectures and mechanisms for examining the application of RBAC in a web-based workflow system and implied suitability to deploy such control in large-scale collaborative environments. The system developed here builds upon Ahn's models.

Figure 3. Data flow definition for the portal

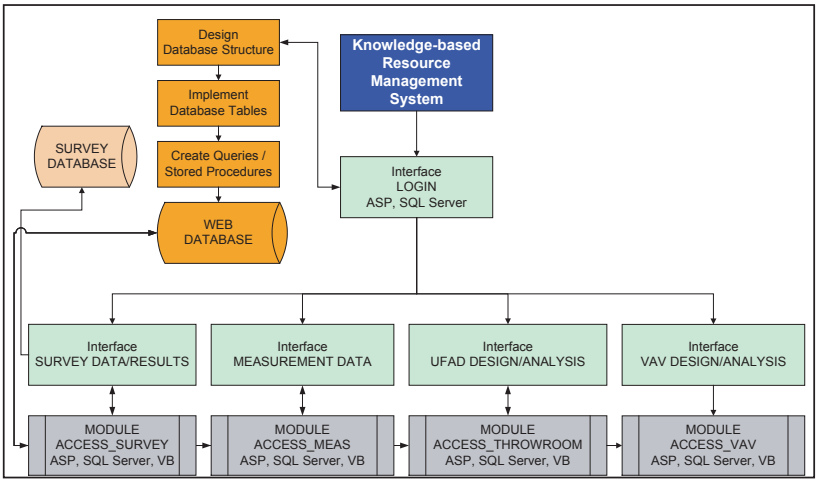
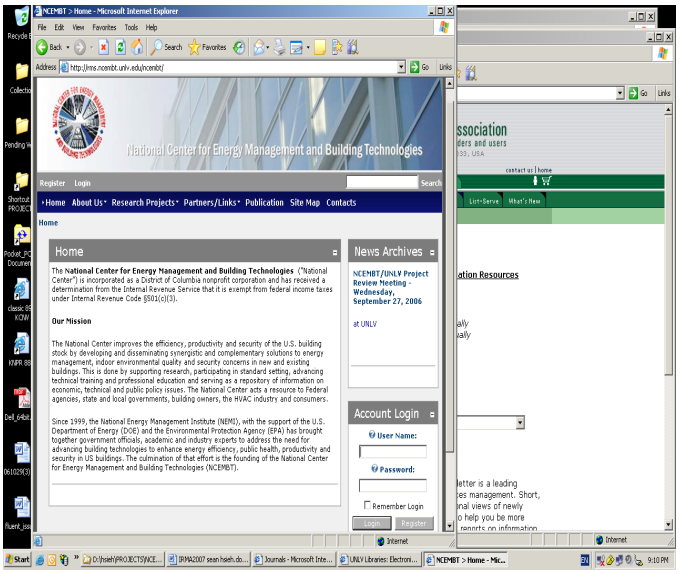


Figure 4. Public accessible web portal

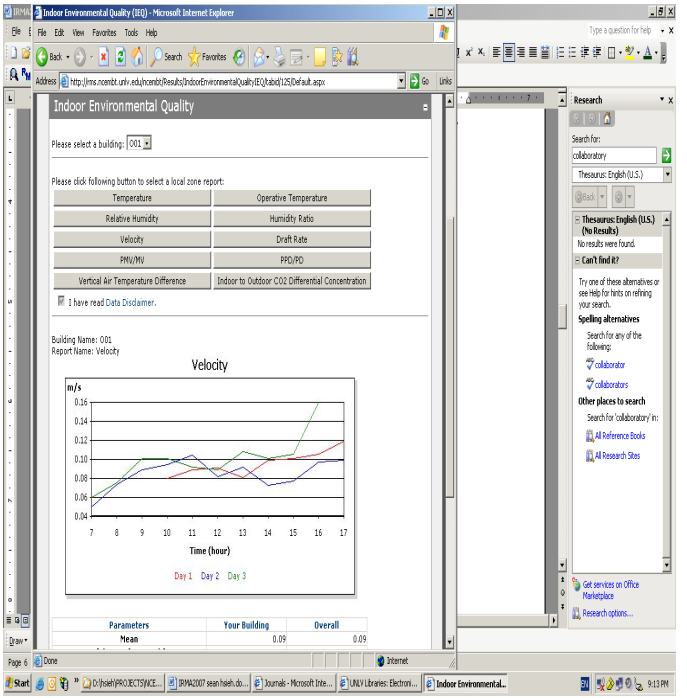


SYSTEM DESIGN

Portal Components

Nine basic portal components were implemented as shown in Figure 4. The web site can be accessed from public (seven modules) or private (nine modules) sections. General information can be viewed by all audiences without system login while private defined components/materials can only be viewed through a password-enabled mechanism.

Figure 5. Summary IEQ data reporting page



Reporting Module Example

Due to the complexity and large quantity of data collected from the first 20 buildings, a reporting module was developed to provide options for data aggregation, access to instant site-specific summary reports and download of the raw data files. Aggregated results and summary statistics of monitored building performance parameters can be accessed via pull-down menus. Figure 5 shows the screenshot of the IEQ parameters that can be displayed for each monitored building. A summary report was developed for each building and incorporates building demographics, recorded IEQ parameters including lighting, microbiological, and sound data and summary results of occupant perceptions of their respective indoor environments.

SUMMARY

An online knowledge-based resource management system to manage, organize and display large data sets of the performance of commercial and institutional buildings has been developed. Handling a large amount of data collected via various measurement techniques and approaches posed significant data storage and management challenges. It also offers a real-world experience on data cleaning, consolidation and manipulation. The knowledge-based resource management system provides pertinent stakeholders (i.e., researchers, engineers, building operators, and facilities managers) with a wealth of building performance data. The system allows users to access various aggregated monitored building parameter sets and provides results of statistical analyses of occupants' perceptions of their indoor environment questionnaire which have been compared to measured building performance parameters. The resulting knowledge is generated to further improve the built environment.

This is a continuous effort that will eventually evolve into the construction of data warehouse and data mining model of building sciences research. More energy management and building technology related knowledge from future projects will be made available along with the growth of the resource management system.

ACKNOWLEDGMENT

This research was supported by the National Center for Energy Management and Building Technology (NCEMBT) under a cooperative agreement with the U.S. Department of Energy.

REFERENCES

Ahn, G., Sandhu, R., Kang, M. & Park, J. (2000). Injecting RBAC to secure a Web-based workflow system. Proceedings of the fifth ACM workshop on Role-based access control, Berlin, Germany, 1-10.

Fayyad, U., Piatetsky-Shapiro, G. & Smyth, P. (1996a). From data mining to knowledge discovery: An overview. In Advances in Knowledge Discovery and Data Mining, U. Fayyad,, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, Eds. AAAI/MIT Press, Cambridge, Mass.,

Fayyad, U., Piatetsky-Shapiro, G. & Smyth, P. (1996b). The KDD process for extracting useful knowledge from volumes of data. Communications of the ACM, 39(11), 27-34.

Imielinski , T. & Mannila, H. (1996). A database perspective on knowledge discovery. Communications of the ACM, 39(11), 58-64.

Geist, I. (2002). A framework for data mining and KDD. Proceedings of the 2002 ACM symposium on applied computing. Madrid, Spain, 508-513.

Ma, Y., Liu, B. & Wong, C. K. (2000). Web for data mining: organizing and interpreting the discovered rules using the Web. ACM SIGKDD Explorations Newsletter. 2 (1), 16-23.