Automation in Construction 70 (2016) 51–61

Contents lists available at ScienceDirect

Automation in Construction

journal homepage: www.elsevier.com/locate/autcon

Building performance optimisation: A hybrid architecture for the integration of contextual information and time-series data

Shushan Hu [a,b], Edward Corry [a,b], Edward Curry [c], William J.N. Turner [b], James O'Donnell [a,*]

[a] School of Mechanical and Materials Engineering, University College Dublin, Belfield, Dublin 4, Ireland
[b] UCD Energy Institute, University College Dublin, Belfield, Dublin 4, Ireland
[c] School of Computer Science and INSIGHT Research Centre, University College Dublin, Belfield, Dublin 4, Ireland

ARTICLE INFO

Article history:
Received 5 June 2015
Received in revised form 6 April 2016
Accepted 22 May 2016
Available online 18 July 2016

Keywords:
Building performance optimisation
Data interoperability
Linked data
BMS
Time-series data
Relational database

ABSTRACT

Buildings tend to not operate as intended, and a pronounced gap often exists between measured and predicted environmental and energy performance. Although the causes of this 'performance gap' are multi-faceted, issues surrounding data integration are key contributory factors. The distributed nature of the Architecture, Engineering and Construction (AEC) industry presents many challenges to the effective capture, integration and assessment of building performance data. Not all building data can be described semantically, nor is it feasible to create adapters between many different software tools. Similarly, not all building contextual data can easily be captured in a single product-centric model.

This paper presents a new solution to the problem based upon a hybrid architecture that links data which is retained in its original format. The architecture links existing and efficient relational databases storing time-series data and semantically-described building contextual data. The main contribution of this work is an original RDF syntax structure and ontology to represent existing database schema information, and a new mechanism that automatically prepares data streams for processing by rule-based performance definitions. Two test cases evaluate the concept by 1) applying the hybrid architecture to building performance data from an actual building, and 2) evaluating the efficiency of the architecture against a purely RDF-based solution that also stores all of the time-series data in RDF for a virtual building. The hybrid architecture also avoids the duplication of time-series data and overcomes some of the differences found in database schemas and database platforms.

1. Introduction

Buildings are responsible for 40% of EU energy consumption and 36% of the EU's CO_2 emissions. Improved environmental and energy performance of buildings is therefore a key EU objective [1]. A recognised and multi-faceted performance gap exists between measured and predicted building performance. De Wilde attributed components of the gap to three broad stages of the building lifecycle: design, construction (including handover) and operation [2].

One key contributing factor to the performance gap is the accuracy of physical measurement of building performance information. One cannot manage what one does not measure, and existing measurement practises can be inadequate or difficult to implement effectively [3,4]. The poor transfer of performance information [2] throughout the building lifecycle [5] and poor interoperability and data integration between toolsets in the domain can lead to inadequate levels of performance assessment [6].

* Corresponding author.
E-mail address: james.odonnell@ucd.ie (J. O'Donnell).

http://dx.doi.org/10.1016/j.autcon.2016.05.018
0926-5805/© 2016 Published by Elsevier B.V.

Schein [7] and Granderson [8] have provided robust and effective methods of building performance assessment, while approaches to data integration such as Building Information Modelling (BIM) and the Internet of Things (IOT) broadly promise a resolution to the data integration problem. The limiting factor restricting optimal performance assessment remains the use of integrated performance data.

The AEC industry presents some unique challenges to data integration due to the distributed nature of the industry and the presence of many interacting but poorly linked domains. A direct consequence of this is that building managers do not have access to the data and information they need in order to optimally manage buildings [9].

1.1. Computer science approaches to data integration

When considered more closely, there are three main computer science approaches which might be applied to the data integration issue; the common data model, the adapter approach, and semantically-described data. Each approach has significant drawbacks when applied in the AEC industry.

Proprietary software suites use an underlying data model to describe data exchange in specific domains for instance [10,11]. The common data model approach uses a central underlying data model as a data hub through agreement between tool vendors. The data hub is then interpreted by domain-specific applications. In response, the buildingSMART alliance provides the open Industry Foundation Class (IFC) data model to describe data exchange in specified domains in an open manner [12], though significant challenges remain relating to the description of diverse building data domains, and the broader acceptance and adoption of the standard in the wider BIM community.

Rather than attempting to convert all data to a particular data model, the adapter approach uses adapters to integrate different data models, representing different AEC domains [13]. Previous efforts such as the HESMOS and Cooperate projects [14–16] used such an approach, enabling data transfer between different native data formats. However, significant data loss is inherent in such an approach [17]. Also, in an industry as diverse as AEC, it is not feasible to develop and maintain adapters for all cases of data transfer between model and toolsets in the industry [18].

The semantic web approach uses ontologies to provide a way to add knowledge to unstructured data. These ontologies are typically based on Web Ontology Language (OWL) [19], which uses the Resource Description Framework (RDF) to describe the relationships between specific objects using the subject-object-predicate structure [20]. This approach has proven to be quite successful in integrating cross-domain data sources in the AEC industry [11,21–23], by allowing the description of building context data in a homogeneous fashion [11,22,24]. For example, the Sensor Ontology [25] was developed to describe sensor information including placement and measurement contexts.

The semantic web approach can be used to overcome some of the issues associated with product-centric data models such as IFC, and ontology matching approaches [26] can be used to provide greater meaning across domain-specific ontologies. The semantic web approach can be used very effectively to link contextual building data, including building geometry, material properties, as-built construction details and HVAC specifications. Time-series data can be managed very efficiently in existing database structures.

This paper presents a flexible, *as-needed* solution to information sharing in the AEC domain, whereby data is retained in the most *appropriate* format and shared as required. This hybrid approach may provide an inexpensive and effective solution to the data integration issue.

The well-recognised performance gap between measured and simulated data is a multi-faceted problem with root causes spread throughout the building lifecycle [2]. Different aspects of the performance gap have been considered in isolation, including issues around assumption, approximation, and simplification of both measurement and modelling [27]. Additionally the performance gap associated with user behaviour and occupancy patterns within buildings is difficult to qualify for a number of complex and related reasons [28]. Notably, the manual linkage of person and room information has a number of practical and accuracy constraints that permeate the building life cycle [29]. This paper is concerned specifically with enhancing the interoperability between measured and simulated performance data encountered during building operation [2].

The hybrid approach enables enhanced building performance assessment through a hybrid data architecture solution, integrating existing and available data sources in buildings using both semantic web technologies and more traditional database solutions. A new mechanism is defined to automatically prepare data streams for processing by rule-based performance definitions.

Building on previous work by Corry et al. [8] and O'Donnell et al. [9,11,21], the objective of this paper is to show how building context data can be mapped to related performance data. The proposed approach keeps time-series data within its original database, avoiding duplication while benefitting from the high efficiency of mature database platforms, especially for structured fixed data [30,31]. This approach defines an RDF syntax structure and vocabulary to represent database schemas based on semantic web technology. It also provides a framework for the access of time-series data in databases based on building geometry contextual data and SSN contextual data.

Section 2 describes the technologies used in this work. Section 3 introduces the hybrid architecture used to integrate AEC performance data with building context data. Section 4 details the syntax of the mapping description, the connection module and the mapping module. Section 5 describes two test cases for evaluation and verification of the new approach. Finally, Section 6, the conclusion, presents the main outcomes from this work.

2. Case for a hybrid solution

Augenbroe advocates a rigorous use of building performance indicators to ensure compliance between project specification and performance [32]. This type of rigour is hampered by the nature of the data capture and storage systems used currently in the AEC industry.

Due to the fragmented nature of the AEC industry, many domain-specific data models exist. The rigorous use of performance indicators requires the integration of data retained in these formats, in some manner.

The performance framework using the scenario modelling method [9] followed on from previous performance metric/indicator work by Hitchcock and Augenbroe [33–35]. Additionally, BuildingEQ developed practical measurement and metric sets for buildings [36]. Corry et al. addressed some of the limiting factors of this framework through the formalisation of a performance assessment ontology for buildings [37]. The value of the performance indicator approach is enhanced through recent research developments [11,21,22,38]. Each of these efforts identified how the successful implementation of a performance indicator approach is dependent on access to reliable, integrated, building information.

Choosing one method for the integration of such diverse data is neither feasible, nor appropriate.

2.1. Data sharing for performance evaluation

Cross-domain data sharing is not commonplace in the industry. For example, linking occupancy patterns to building operation, as illustrated in Fig. 1, would enable traditionally separate information sources to be combined and potentially offer new insights into building operation. In this example, tracking the number of members in a given part of a sports centre can aid with staff scheduling and more focused HVAC operational strategies. Similarly, linking design intent to building operation or linking simulation to building operation requires an integrated approach to data management that is not always evident.

This paper strongly advocates leaving building data in the most appropriate platform and format, only using and linking such data on an as-needed basis. The paper presents a novel architecture to allow this to take place.

Contextual information about a building can add to the effectiveness of performance data. Context information may include building geometry and HVAC descriptions as obtained from Building Information Model (BIM) formats, including IFC [39,40] and SimModel [41,42]. Sensor definitions described using the Semantic Sensor Network (SSN) ontology [43] and other soft building information [11] based on the RDF format contribute to a richer set of context information. This type of information can be described and integrated very effectively using the semantic web approach.

Time-series data mainly describes continuous records from deployed sensors and metres in buildings [44]. There are some relationships between context information and time-series data. For example, a sensor ID described using the SSN ontology may also be referred to in time-series data generated from the sensor (see Fig. 2).

Traditionally, semantically-described contextual data might be linked to time-series data through an Application Programming Interface (API),