DETERMINING CRITICAL CONSTRUCTION DOCUMENTS THROUGH KNOWLEDGE DISCOVERY

*A. J. Antony Chettupuzha[1], Dr Carl T. Haas[1], Joel Gray[2] and Ray Simonson[2]

[1]*University of Waterloo*

200 University AvenueWest,

Waterloo, Ontario, Canada N2L3G1

*(*Corresponding author: ajantony@uwaterloo.ca)*

[2]*Coreworx Inc.,*

22 Frederick Street, Suite 800

Kitchener, Ontario, Canada N2H6M6

ABSTRACT

During a construction project, it is important to ensure that accurate and pertinent knowledge is delivered on time to appropriate personnel. The criticality of documents exchanged or referred to by various parties, evolves over the course of the project lifecycle. Determining the criticality of documents at different stages of the project can aid companies with managing the flow of information in an organized manner, while providing for the detection of potentially disruptive, erroneous material that could result in delays and costs. Therefore, it is crucial to determine documents which have the potential to disrupt a process (by virtue of being incorrect versions etc.) so that they may be isolated or prevented from entering or interfering with the active flow of information in a project, while also ensuring standard revisions of documents are recognized and accessed accordingly. Electronic Product and Process Management [EPPM] systems provide the capability to establish and map information flow between different parties in a construction project. Analyzing an EPPM system may yield information about the criticality of certain documents within an activity, as well as isolate characteristics peculiar to documents which can assist in risk mitigation. We propose methodology that identifies such critical documents.

KEYWORDS

Construction documents, knowledge discovery, information systems, critical documents

INTRODUCTION

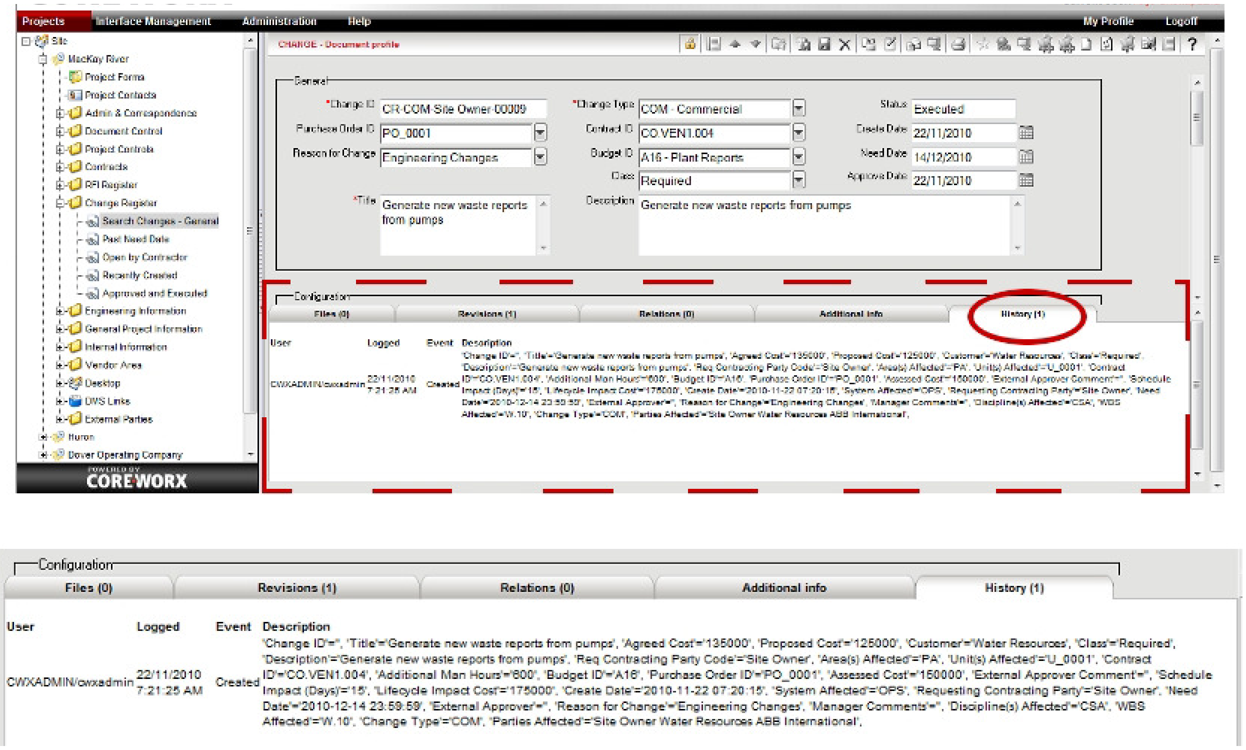
For large complex construction projects, there is a high volume of documentation and associated electronic files generated that must be transferred between all parties involved in the project in an efficient and timely manner. Delays in the transfer of documents can affect procurement, operations and schedules and therefore adversely affect project costs. A central repository of the entire collection of documents generated and related transactional records can provide a valuable resource to construction project owners. A structured document management system, such as an Electronic Product and Process Management [EPPM] system, which imbibes process oriented workflows further extends the capability of managing the massive flow of information over a project lifecycle.

Traditional construction information systems do not explicitly store information related to how documents are involved in construction activities. However in an EPPM system, workflows are employed to automate construction activities. Workflows include information relating to the order in which processes must be executed, the actors who must execute them and the resource requirements for a particular activity. Information about the processes includes information of when documents are to be attached to an activity. In an ideal EPPM system, information relating to when a document should be accessed and every single instance that it is accessed is also recorded.

The nature of an EPPM system, as a repository for process oriented information flow, puts it in a unique position to exploit the extraction and storage of inherent knowledge encapsulated within documents and their movement over the project life cycle. During a construction project, it is important to ensure that accurate and pertinent knowledge is delivered on time to appropriate personnel. The criticality of documents exchanged or referred to by various parties, evolves over the course of the project lifecycle. For example, it is not uncommon for specifications to undergo revisions from the design phase to implementation. Referencing an incorrect version of a specification during the construction phase may directly affect activities at the site. Determining the criticality of documents at different stages of the project can aid companies with managing the flow of information in an organized manner, while providing for the detection of potentially disruptive, erroneous material that could result in delays and costs.

BACKGROUND

In EPPM systems, documents are divided into distinct classes based upon the activity they are modeled around. Examples of typical document classes are Requests for Information (RFI), Change Requests (CRs) and Approvals. Customized templates are designed for each class containing specific meta-data fields that all documents within a class will require. These customized document templates are referred to as document profiles. In addition, usage characteristics such as when the document was changed recently, who made the changes etc., are recorded. Storing usage characteristics are important for auditing and litigation purposes and are usually tracked in history revisions of document profile instances in traditional EPPM systems as shown in Figure 1 below.



CHANGE - Document profile

General

*Change ID: CR-COM-Site Owner-00009
Purchase Order ID: PO_0001
Reason for Change: Engineering Changes
*Title: Generate new waste reports from pumps
*Change Type: COM - Commercial
Contract ID: CO.VEN1.004
Budget ID: A16 - Plant Reports
Class: Required
Description: Generate new waste reports from pumps
Status: Executed
Create Date: 22/11/2010
Need Date: 14/12/2010
Approve Date: 22/11/2010

Configuration

User	Logged	Event	Description
CWXADMIN/cwxadmin	22/11/2010 7:21:25 AM	Created	'Change ID'='', 'Title'='Generate new waste reports from pumps', 'Agreed Cost'='135000', 'Proposed Cost'='125000', 'Customer'='Water Resources', 'Class'='Required', 'Description'='Generate new waste reports from pumps', 'Req Contracting Party Code'='Site Owner', 'Area(s) Affected'='PA', 'Unit(s) Affected'='U_0001', 'Contract ID'='CO.VEN1.004', 'Additional Man Hours'='600', 'Budget ID'='A16', 'Purchase Order ID'='PO_0001', 'Assessed Cost'='150000', 'External Approver Comment'='', 'Schedule Impact (Days)'='15', 'Lifecycle Impact Cost'='175000', 'Create Date'='2010-11-22 07:20:15', 'System Affected'='OPS', 'Requesting Contracting Party'='Site Owner', 'Need Date'='2010-12-14 23:59:59', 'External Approver'='', 'Reason for Change'='Engineering Changes', 'Manager Comments'='', 'Discipline(s) Affected'='CSA', 'WBS Affected'='W.10', 'Change Type'='COM', 'Parties Affected'='Site Owner Water Resources ABB International',

Configuration

User	Logged	Event	Description
CWXADMIN/cwxadmin	22/11/2010 7:21:25 AM	Created	'Change ID'='', 'Title'='Generate new waste reports from pumps', 'Agreed Cost'='135000', 'Proposed Cost'='125000', 'Customer'='Water Resources', 'Class'='Required', 'Description'='Generate new waste reports from pumps', 'Req Contracting Party Code'='Site Owner', 'Area(s) Affected'='PA', 'Unit(s) Affected'='U_0001', 'Contract ID'='CO.VEN1.004', 'Additional Man Hours'='600', 'Budget ID'='A16', 'Purchase Order ID'='PO_0001', 'Assessed Cost'='150000', 'External Approver Comment'='', 'Schedule Impact (Days)'='15', 'Lifecycle Impact Cost'='175000', 'Create Date'='2010-11-22 07:20:15', 'System Affected'='OPS', 'Requesting Contracting Party'='Site Owner', 'Need Date'='2010-12-14 23:59:59', 'External Approver'='', 'Reason for Change'='Engineering Changes', 'Manager Comments'='', 'Discipline(s) Affected'='CSA', 'WBS Affected'='W.10', 'Change Type'='COM', 'Parties Affected'='Site Owner Water Resources ABB International',

Figure 1 – A history tab where revision changes are tracked in audit reports within traditional EPPM systems

In analyzing information systems, there are several different sources that may be mined depending on the type of knowledge that is being sought. Often in web-based systems, there are three main sources; (i) the content of a webpage, (ii) the structure of a web-page and (iii) the usage and system logs pertaining to that webpage (Dunham, M., 2003). Suitable mining methods are adopted based upon the source of data. Mining knowledge from content often employs techniques such as text data mining and natural language processing so as to be able to associate keywords with intelligent contextual understanding of what the document describes.

Mining knowledge from the structure, usage and systems logs may be attained from workflow histories. Zhao (1998) contends that this knowledge exists within the workflow models, the history of executed instances of a workflow as well as external vaults such as databases and decision support systems. Aalst et al (2003) have described how data mining techniques may be applied to a database containing the transactions of workflow instances to assist with analysis of workflow functioning. Leymann and Roller (Leymann and Roller, 2000) also observe that data mining permits meticulous analysis of processes which could lead to the recognition risk inducing factors.

In the AEC sector, the common approach to analyzing documents has built upon these techniques. El Gohary (2008) developed ontologies to enable semantic interpretation of infrastructure documents. Caldas and Soibelman (2003) automated the classification of construction documents hierarchically. Research in this area has usually been confined to understanding the content of documents, which pertains to the first source of data mentioned above.

However research about the importance of a document and the evolutionary characteristics that determine its criticality as the project progresses has not been explored in the context of mining. The content of a document in this regard ceases to be as important as the structure, or the relations and links between documents in the overall context of an information system in a construction project, and the usage characteristics associated with the document. Such information will enable rapid assessment of wasteful access and distribution of a document without computationally expensive mining of the content of a document. As such in an ideal EPPM system, it shall be possible to obtain critical insight of a document's importance by analyzing associated workflow and usage statistic histories from stored document meta-data.

DETERMINING CRITICAL CONSTRUCTION DOCUMENTS

An instance, or an executed implementation, of a document profile may contain attachments pertinent to that particular activity. For example, an RFI may contain a reference document or a design drawing. Due to confidentiality agreements between contracted parties, the content of these attachments were not accessible and therefore intelligent contextual and semantic information from these attachments was not considered for this research initiative. Information obtained from metadata and usage characteristics was deemed to be sufficient.

Interviews were conducted with EPPM system development consultants and construction clients to determine criteria that establish the criticality of a document at a particular phase of a construction project. Traditional EPPM systems do not store all associated access related to documents explicitly in databases, therefore for further considerations in this paper we consider the implications of using an ideal EPPM system wherein such information is easily attainable. Based on the feedback obtained from these experts, a list of factors was created, which are detailed below:

1. Relations between documents

Within an EPPM system, it is possible to track internal relationships between documents. If a document is initially included in a workflow, when another document is added a one-to-one relationship between the documents is maintained within the system's database. At a system level, if all the relationships between documents are analyzed, it is possible to determine secondary and lower order relationships between documents. Figure 2 below depicts how it documents within an EPPM system may be mapped in terms of relations they have with one another.

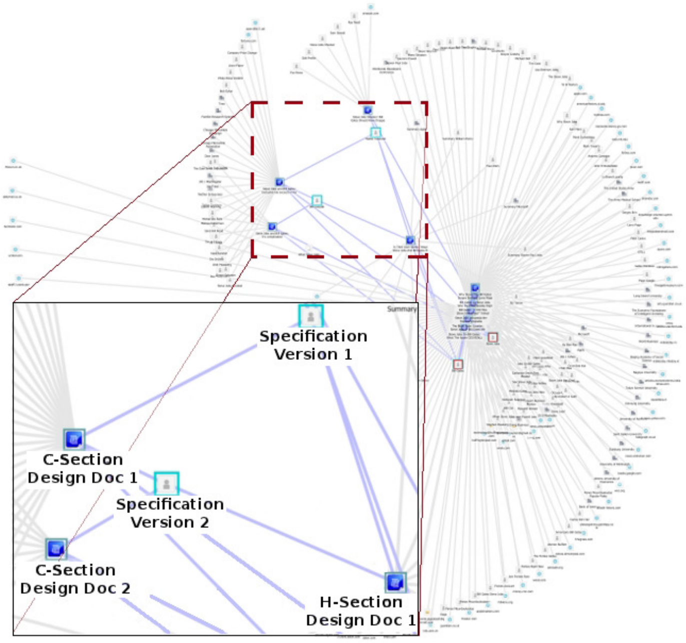


Figure 2 - An example illustration of how document relations may be mapped in an EPPM system

For example if document A has a one-to-one relationship with document B, and document B has a one-to-one relationship with document C, then we may denote a secondary relationship between document A and document C. Mapping all the relationships between documents results in the creation of a document density map, wherein it is possible to identify critical documents based on the number of important relations they have.

2. Anticipated document flow versus actual flow

The estimated flow of documents for projects is a crucial indication of the amount of information that is expected to be exchanged over the lifecycle of a project. While it is common for such an estimate to evolve with the duration of a project, there are reasonable ranges and bounds for which certain documents are expected to flow through the system. If a document begins to flow with greater than expected frequency, it might indicate that a problem has arisen and several different parties are searching for the same document which provides relevant information germane to the issue at hand. Lack of flow as compared with what is expected on the other hand might indicate that employees do not consider the document to be of significant importance and as such if a change is registered this might not propagate to all parties.

3. Document expected idle time versus actual idle time

The flow of documents or electronic products within an EPPMS might not necessarily be continuous, but instead depends on when certain workflow instances are active, or in other words, during the execution of activities or tasks. There are several instances where a document is created for a one-off instance of a workflow and is not expected to be accessed again. There are sets of documents which might be accessed at regular intervals over the lifecycle of a project. As such there are periods of time wherein the state of document can be assumed to be idle, and other times where the state of document or electronic product is active or in-use.

This in turn reveals that there are periods during which a document is not expected to be used, and if it is being accessed then this could indicate a change or deviation, either hard coded or unrecorded, in the workflow, potential access of the wrong document or unauthorized access. Further, there could be documents which are deemed relatively unimportant at the start of a project, but which become increasingly referenced as the project progresses and are hence far less idle than initially anticipated. These documents might then be recognized as critical to the functioning of a certain task. Thus, understanding how often a document's state changes from an expected idle state can be an important indicator of either a potential disruption or the increasing importance of the document in question.

4. Document access activity

Related to the idle time of a document is the access of a document. While the change of state of a document from idle to active implies access to the document, there are several additional facets of document access that make incorporating it as an additional distinct factor for identifying critically problematic documents essential. For example, document state change records when a document has been accessed but does not take into account, who has accessed the document, why they have accessed it or how they accessed it.

Users within an EPPM system have distinct roles based upon the responsibilities they are permitted to perform. These roles and permissions are stored within a RACI (responsible, accountable, consulted, and informed) matrix. The frequency of access from users with specific roles may be an indication of the importance that the document has at specific stages of the project. Noting when a document was accessed may also be an indication as to why the document was accessed. For example if a specification document is referred to multiple times over the entire project lifecycle this is an indication that this is a crucial document to the project's success. Unique to an EPPM system, is the ability to track how a document was accessed. Users have the option of "bookmarking", "favoriting", "saving" and "checking out" a document to their personal profile. This changes the way they a document is accessed on multiple occasions and provides further indication for how relatively important a user believes a document is.

5. Document revisions

It is natural for documents to undergo revisions over the course of a project. Changes to design or requirements, updating the completion of tasks, and reporting of on-site activities etc., all constitute updates made to documents. However it is crucial to ensure that only the most up-to-date or relevant document is accessed by a participant at all stages of a project. Using outdated requirements or specifications can have serious implications during the construction and execution stage of a project potentially resulting in costly delays and damages. Keeping track of version changes and ensuring that only the most recent accepted revision is accessed is therefore a vital requirement for an EPPMS.

Older versions are usually preserved for auditing purposes, and therefore are usually accessed when there is sufficient reason to investigate the evolution of a document. The availability of older versions is also useful for maintaining references to activities that might not have been important enough to be recorded. There may be instances however wherein a newer version is deliberately not accessed by the participant, say if outside of the EPPM system, a group of participants have mutually agreed upon using a previous version for the completion of a task, but wish to preserve the newest version as a draft for the next stage of a project. Or perhaps a previous version contains a set of instructions or notes which might be applicable in similar repetitive activities in a project, but which was left out in the final version of the submitted document. When a particular version is accessed on a more frequent basis at a given project time, it is identified as a *standard* version and its importance to the project is greater than previous versions.

6. Documents referenced at critical interface points

It is common for participants from different crews to communicate and transfer a high volume of documents during the execution of an activity or task. These interactions might occur at physical locations of a project, say at an interface point where multiple crews are working on completing an activity (for example, a pipeline joint) or they might be virtual in nature when say people from different regions revise stakeholder provisions or requirements. During these interactions, there will be a transfer of official documents that should usually occur as part of a workflow. EPPM systems have in-built capabilities to identify and monitor activities that occur at interface points. Different interface points have differing levels of criticality, and documents that are part of critical interface points may be considered to have a higher level of importance.

Establishing the overall criticality of a document

Scores may be given for each document based on the individual scores from each of the factors outlined above. The individual factor scores will have to be weighted. Based on interviews with EPPM consultants and clients it was learned that there is no uniform consensus on the weightage assigned to each of the above factors. Different companies place differing levels of importance on each of the above factors. For example some companies do not employ the integration of an interface point application in their EPPMS while others may not choose to record the relationships developed between documents. Allowing for flexibility in the weighting, wherein an end-user may change weights to get real-time criticality based on one configuration of weight settings, permits users to see if a particular set of documents are consistently deemed important based on differing weight permutations. Hence in the preliminary stages of this research investigation this approach shall be used, and as more data is made available from multiple projects it will be possible to determine common user-defined weight settings that may be provided as a basis for clients.

SUMMARY

It is important to identify documents that are critical to a construction project for risk management purposes and also to manage the flow of documents in an organized manner. In traditional construction information systems information relating to a document's role in work processes is not stored. Traditional EPPM systems however track this information as they incorporate workflows to manage to flow of information associated with the execution of an activity. It is possible to determine a document's criticality by studying the workflow histories, usage statistics and metadata of a document and its associated document profile in an EPPM system. A list of factors that influence the criticality of documents in an ideal EPPM system were identified based on interviews with experts from both vendor and client organizations. By incorporating flexibility in the weightage assigned to each of these factors, it is possible to identify which documents recur as being critical to a project based on different weightage permutations. Hence critical documents may be determined during different stages of a construction project, providing the capability for graceful handling of such documents.

REFERENCES

Caldas, C.H. and Soibelman, L., (2003) “Automating hierarchical document classification for construction management information systems”, Automation in Construction 12, pp. 395–406.

Dunham, M., “Data Mining Introductory and Advanced Topics”, Prentice Hall, 2003.

El-Gohary, N. 2008. “Semantic Process Modelling and Integration for Collaborative Construction and Infrastructure Development”, PhD Thesis, Department of Civil Engineering, University of Toronto, Canada.

Leymann F, Roller D., (2000) "Production Workflow: Concepts and Techniques" Prentice Hall

van der Aalst, W.M.P., van Dongen, B.F., Herbst, J., Maruster, L., Schimm, G., Weijters, A.J.M.M, (2003), “Workflow Mining: A Survey of Issues and Approaches”, Data and Knowledge Engineering 47(2), 237–267

Zhao, J.L., “Knowledge Management and Organizational Learning in Workflow Systems”, Proceedings of AIS 1998