DocFlow: An Integrated Document Workflow for Business Process Management

Boonsit Yimwadsana, Chalalai Chaihirunkarn,
Apichaya Jaichoom, Apichaya Thawornchak
Faculty of Information and Communication Technology, Mahidol University
999 Phuttamonthon 4 Road, Salaya, Phuttamonthon
Nakhon Pathom 73170, Thailand
{itbyw,itcch}@mahidol.ac.th,
picha_nat@hotmail.com, apichayat@hotmail.com

ABSTRACT

Document management and workflow management systems have been widely used in large business enterprises to improve productivity. However, they still do not gain large acceptance in small and medium-sized businesses due to their cost and complexity. In addition, document management and workflow management systems are often used separately because they solve different problems. Only some part of document management systems should be tied together with workflow management systems. However, in most business environment, documents actually flow according to workflow definitions. Our work, thus, combines the two concepts together and simplifies the management of both document and workflow to fit business users. Our application, DocFlow, is designed with simplicity in mind while still maintaining necessary workflow and document management standard features with security. Approval mechanism is naturally included in the workflow, and the approval can be performed by a group of actors such that only one of the team members is sufficient to make the group's decision. A case study of news publishing process is shown to demonstrate how DocFlow can be used to create a workflow that fits the news publishing process.

KEYWORDS

Document Management, Workflow Management, Process Management, DocFlow, Electronic Business

1 INTRODUCTION

Today's business organizations must make rapid decision and execute operations quickly in order to cope with global competition and the ever-changing business environment. However, most decisions and operations cannot be made and executed easily by one person. A teamwork consisting of many people from different departments within an organization is usually required. For instance, in order to purchase raw materials for the production of a plastic bottle, the engineering department must give the procurement office the specification of the plastic beats that the engineering department needs. Then, the procurement office will have to work with the accounting department to find out whether the engineering department has sufficient budgets to purchase. Once the decision to purchase the raw materials is approved, the procurement office selects the most suitable vendor and makes a purchase. The raw materials will then be received and stored in the engineering department's warehouse where raw materials will be stored. Finally, the financial department

will verify the raw materials and send the confirmation of the receipt to the supplier. In this example, the engineering, procurement and financial departments are involved in this purchase process.

In most business scenarios, there will be information exchange between two communicating divisions. Documents are used to stored necessary information as containers waiting to be exchanged between two communicating divisions. Processing and approval of information must be made before new information will be entered into another set of documents to be sent to the next division responsible for the operation.

According to the information exchange nature of collaborative work, a large number of documents are generated until the completion of an operation. Often times, documents must be reviewed by a number of individuals (usually managers) before they can be approved and forwarded to the next department. This process can take a long time and involve many individuals. Different revisions of documents revised by many individuals can create confusions on document ownership and versions. Due to today business environment, an individual does not usually focus on one single task, but instead is involved in different tasks and projects from within a single or several departments as a part of organizational integration effort. It is possible that the individual puts the task into his low priority list and completely forget about it. In order to keep track of the number of documents each individual has to process, he/she can put documents information in a database so that he/she can easily manage his/her work. It would be best if a document

database is created for the whole organization in order to help individuals to review and approve documents later.

The document database is one of the earliest applications of information technology. Documents are transformed from paper form to electronic form. Despite the convenience and benefits of electronic documents, document management software or concept is one of the least deployed solutions in businesses. This is largely because of the lack of IT knowledge in most employees in non-IT-related departments. Proper file and folder management help company staffs organize documents so that they can work with and review documents in a repository efficiently to reduce operation costs and speed up market response [20]. When many staffs have to work together as a team or work with other staffs spanning different departments, a shared document repository is needed. Hence, a standard method for organizing documents must be defined within the organization. Different types of work environment have different standards. Common concepts of document and file storage management for efficient and effective information retrieval can be introduced and enforced by using various document management systems such as [1,3-5] and they have been widely accepted in various industries.

The World Wide Web is a document management platform that can be used to provide a common area for individuals to gain access and share documents from anywhere in the world at any time. In particular hypertext helps alleviate various issues of document organization and information retrieval. Documents can be stored as files in a file system

along with their relationship. The success of hypertext can easily be seen from the success of the World Wide Web today. However, posting files online in the Internet or Intranet has a few problems. Not all staffs know how to or want to put information or documents on websites, and they usually do not have access to the company's web server due to security reason. In addition, enforcing user access control and permission cannot be done easily. There are a number of websites that provide online services (cloud services) that allow members to post and share information on the websites such as Wikipedia [6] and Google Docs [7]. However, using these services lock users into the services of the companies. In order to start sharing documents and manage documents, one must register an account at a website providing the document management service, and place documents in the cloud. This usually violates typical business policy which requires that all documents must be kept private inside the company.

To accommodate a business policy on document privacy, documents must be kept inside the company. Shared file and folder repositories and document management systems should be deployed within a local area network to manage documents [19]. In addition, in a typical work environment, several people work with several version of documents that are revised by many people. This creates confusion on which version to use at the end. Several file and folder names can be created in order to reduce this confusion. However, this results in unnecessary files and folders which waste a lot of memory and creates confusion. In addition, sharing files and folders require careful monitoring of

access control and file organization control at the server side which is not practical in an environment that has a large number of users.

Document management systems do not address how documents travel from the starting of an operation to the end of operation. The software system that describes the flow of documents are called workflow management systems [14,17,18] which is tightly related to business process management. Workflow management systems help business users define and draw workflows such that business users can keep track of business operations. Workflow management systems have recently become one of the most important tools used in business today. They are the most important component of business process management effort. Various workflow information systems are proposed to make flow designation easier and more effective. Widely accepted workflow management systems are now developed and supported by companies offering solutions to enterprises such as IBM, SAP and Microsoft [9-11].

In short, document management system focuses on the management of electronic documents such as indexing and retrieving of documents [21]. Some of them may have version control and concurrency control built in. Workflow management system focuses on the transformation of business processes to workflow specification [17-18]. Monique [15] discussed the differences between document management software and workflow management software, and asserted that a business must clearly identify its requirements and choose which software to use.

In many business environments, document and workflow management systems are used separately. Workflow management systems are often used to define how divisions communicate systematically through task assignments and document flow assignments [18], while document management systems are used to manage document storages. When the two concepts are not combined, a staff must first search for documents from document management system, and put them into workflow management systems in order for the document to reach the decision makers.

Our work focuses on integrating document management system together with workflow management system in order to reduce the problem of document management in workflow management system and workflow support in document management system. We propose a model of document workflow management system that combines features of document management system and workflow management system together. Currently, there are solutions that integrate document management software and workflow management software together such as [1,2] and ERP systems such as [11]. However, most solutions force users to switch to the solutions' document management systems which are often restrictive instead of allowing the users to use their favorite document management system that support popular document processing software such as Microsoft Word. In addition, the deployment of ERP systems requires initial complex customized configurations to be performed in order to support the business environment [16].

2 DOCFLOW: A DOCUMENT WORKFLOW MANAGEMENT SYSTEM

DocFlow is a document workflow management system that combines basic concept of document management system and workflow management system together to help business manage its documents, tasks, and approval process under their business process management scheme. DocFlow system provides storage repository and document retrieval, versioning, security and workflow features which are detailed as follows:

2.1 Document Storage Repository

DocFlow stores documents locally in file system normally supported by local filesystem in a server or Storage Area Network (SAN). When files are uploaded to the system, metadata of the documents, such as filenames, keywords, and dates, can be specified by the users and stored separately in DocFlow's document metadata database. A major feature of DocFlow is the support for various document formats. The storage repository will store documents in the original forms entered by the users. This is to provide support for different document formats that users would use. In Thailand, most organizations use Microsoft Office applications such as Microsoft Word, Microsoft Excel, Microsoft PowerPoint and Microsoft Visio to create documents. Other formats such as image- and vector-based documents (Adobe PDF, postscript, and JPEG), and archive-based documents such as (ZIP, GZIP, and RAR) documents are also supported. DocFlow refrains from limiting document format in order to

seamlessly integrate with other document processing software. The filesystem and metadata database are designed to allow documents to be linked to the workflow created separately to reduce the number of documents that may be a duplicate in different workflows.

2.2 Versioning

Simple documents versioning are supported in order to keep the history of the documents. DocFlow allows users to create different versions of the documents so that they can retrieve previous versions of the documents and continue working from a selected milestone. Versioning can also help users to create documents that are the same kind but use in different purpose or occasions. Users can define a set of documents under the same general target content and purpose type and associate them with versions.

Since DocFlow supports teamwork function. Document inconsistency or conflict can occur when several individuals in a group edit the same documents at the same time and upload their own versions to the system. Thus, the system is designed with a simple document state management such that when an individual downloads documents from DocFlow, DocFlow will notify all members in the group responsible to process the documents that the documents are being edited by the individual. DocFlow does not allow other members of the group to upload new version of the locked documents until the individual unlock the documents by uploading new versions of the documents back to DocFlow. This is to prevent content conflicts since

DocFlow does not have automated content merging capability found in advanced and specialized version control system software such as subversion [8]. During the time that the documents are locked, other group members can still download other versions of the documents except the ones that are locked. A newly uploaded document will be assigned a new version by default. It is the responsibility of the document uploader to specify in the version note that the new version of the document is an update from which version specifically.

2.3 Security

All organizations must protect their information in order to retain trade secrets and company internal information. Hence, access control and encryption must be used with documents and users. Access control designations should be kept in a separate location in the database based on standard access control policy [13] in order to implement authorization policy. A user can grant read-only, full access, or no access to another user or group based on his preference.

The integrity policy is implemented using Public Key Cryptography through the use of document data encryption and digital signing. For document encryption, we use symmetric key cryptography where the key are randomly and uniquely created for each document. To protect the symmetric key, public key cryptography is used. When a user uploads a document, each document is encrypted using a symmetric key (secret key). The symmetric key is encrypted using the document owner's public key, and stored

in a key store database table along with other encrypted secret keys using the (document ID, user ID) association.

When the document owner gives permission to a user to access the file, the symmetric key generated for this document is selected from the secret key database and decrypted using the document owner's private key. The owner's private key is protected by another password and stored either on the user's USB key drive or on the user's computer. The secret (symmetric) key will be encrypted using the target user's public key and stored in the key store database table. This security mechanism is designed with the standard security encapsulation concept. The complexity of security message communications is hidden from the users as much as possible. The document encryption mechanism is shown in Figure 1.
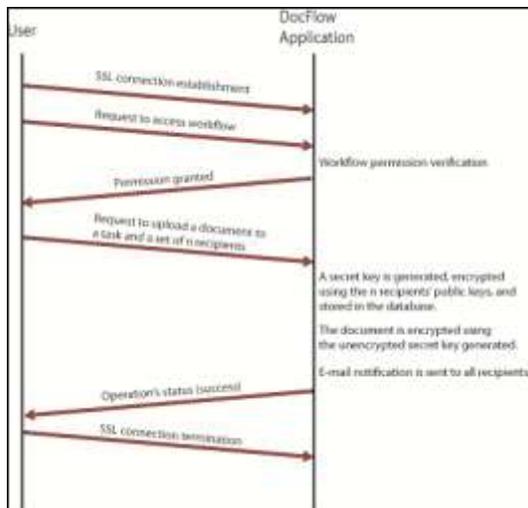

Figure 1. Encryption mechanism of DocFlow

2.4 Workflow

The workflow model of DocFlow system is based entirely on resource flow perspective [22]. A resource flow perspective defines workflow as a ternary relationship between tasks, actors and roles. A task is defined as a pair of document producing and consumption point. Each task involves the data that flow between a producer and a consumer. Normally, a task acts like a conditional point to determine whether the documents can go forward in one direction or another direction. Each task can be associated with only one individual or multiple individuals who will make conditional decision.

DocFlow needs the flow to be designed first before documents can be produced and associated with each task-to-task communication. The path containing connected producer/consumer paths defines a workflow. In other words, a workflow defines a set of tasks. Each task can be considered as a decision point where the user associated with the task makes decisions whether to direct the flow to a specific direction. A complete workflow has a start condition and an end condition. In our workflow concept, a document produced by an actor of each task is digitally encrypted and signed by the document owner using the security mechanism described earlier.

DocFlow allows documents to flow in both directions between two adjacent tasks. The reverse direction is usually used when the documents produced by a prior task are not approved by one of the actors in the current task. The disapproved documents are revised, commented and sent back to the prior task for rework. In some cases, disapproved work can be directed towards another task. All documents produced will have a new version and are digitally signed to confirm the identity of the document owner and the

integrity of the documents. Documents can only move on to the next task in the forward direction of the workflow only when one of the actors in each task approves all the documents received for the task.

DocFlow also allows the skipping of some tasks in the workflow in case there is an exception which must be authorized by the flow owner. The flow owner can modify the workflow later on such as adding new tasks to the workflow, and modifying actors of the tasks in the workflow. DocFlow will send notification e-mails to all affected DocFlow members for every change related to the workflow.

It is important that a task should be completed easily by placing documents into the task, approving or disapproving the tasks, or delegating the responsibility to another actor. DocFlow also provides a reminder service to the actors of a task when the task is active for more than a threshold value in order to make sure that a specific task must be completed within a period of time.

However, not all communication must flow through the workflow path. Actors should be reminded that workflow does not limit communication channels. Peer-to-peer communication such as e-mail should be allowed and encouraged in order to solve business issues.

3 SYSTEM ARCHITECTURE AND IMPLEMENTATION

DocFlow system is designed with three-tier architecture concept. It is implemented as a web-based system whose server-side consists of 4 major modules which are authentication, user and group management, document management and workflow management. The client-side module of the system is implemented using Adobe Flash and Adobe Flex technology while the server-side business process modules are implemented using PHP connecting to a MySQL database. Users use Web browser to access the system through https protocol. Adobe Flash and Flex technology allows simple and attractive interface. The client-side modules exchange messages with the server-side modules using web-services technology. Uploaded documents are stored in their original formats in a back-end SAN. The system architecture and details of each module are shown in Figure 2.

Figure 2. DocFlow System Architecture

4 DEPLOYING DOCFLOW IN ORGANIZATION

Since a workflow usually involves employees from various departments of an organization, coming up with a workflow is usually a daunting task. Most employees do not see the big picture of the workflow in an operation. They think that an operation is just a form and a set of documents for them that they have to send to the next department. Most workers in a department will completely forget about the documents once they are sent out of

the department.   Figure 3 shows common operations that happen in a business purchase request operation. The engineering department is only concerned with generating the "purchase request form" to the procurement department and waits for approval. The engineering department will not get involved in the "budget request" operation that the procurement department has to perform.
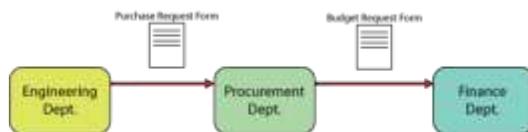

Figure 3. A typical purchase request operation of a business

Situation in Figure 3 makes workflow design process very difficult because unless the pictures like Fig. 3. are drawn, no one will have the big picture of the workflow for an operation. Each department is only responsible for its own portion of the task. Without the big picture of a workflow, different departments may not be able to create workflow in a workflow management system such as DocFlow. The organization will be unable to keep track of business operations, and the business process management effort will not be successful.

In order to deploy DocFlow, all the forms that each department uses must be analyzed and connected to each other in order to identify and refine business operations. Once the graph connecting departments (or tasks) are drawn, the workflow should then be associated with forms and documents that should be refined to fit the revised workflow. The forms and documents will be given permissions to be accessed by users

according to the company's security policy and business operations. Once these three steps are completed, DocFlow will be ready for business users to operate.

5 EXAMPLE

Staffs in the public relation (PR) division at the Faculty of Information and Communication Technology, Mahidol University, Thailand, usually write news and event articles to promote the faculty and the university. Normally there will be a few staffs who gather the content of the news and events in Thai language and pass it to a staff (news writer) who write each news. The news writer will forward the written news to another staff (English translator) who can translate the news from Thai to English. The news in both Thai and English will then be sent back to the news writer to make the final pass of the news before it is submitted to a group of faculty administrators (news Editor) who can approve the content of the news. The faculty administrator will then revise or comment on the news and events and send the revised document consisting of Thai and English versions back to the news writer who will make the final pass of the news.

Normally, the staffs communicate by e-mail and conversation. Since PR staffs have other responsibilities, often times the e-mails are not processed right away. There are a few times that one of the staffs forgets to take his/her responsible actions. Sometimes a staff completely forgets that there is a news article waiting for him/her to take action, and sometimes the staff forgets that he has already taken action. This delays the posting of the news update on the website and faculty newsletter.

Using DocFlow, assuming that the workflow for PR news posting is already established, the PR writer can post news article to the system and approve it so that the English translator can translate the news, view the news articles in progress in the workflow, and send news article back to the news writer to publish the news. There can be many English translators who can translate the news. However, only one English translator is sufficient to work on and approve the translated news. The workflow system for this set of tasks is depicted in Figure 4.
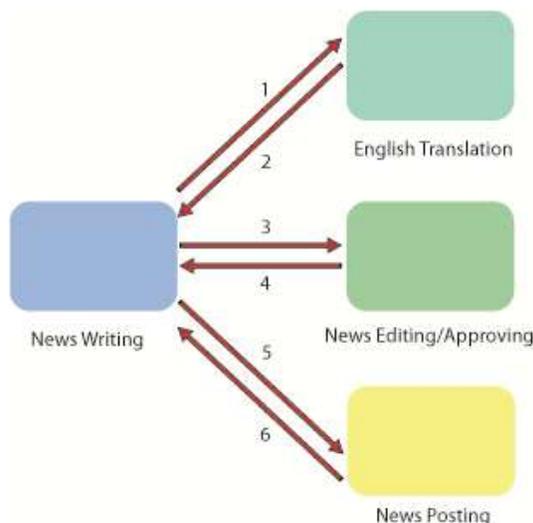

Figure 4. News Publishing Workflow at the Faculty of ICT, Mahidol University consists of four actor groups categorized by roles. A task is defined by an arrow. DocFlow allows documents to flow from an actor to another actor. The state of the workflow system changes only when an actor approves the document. This change can be forward or backward depending on the actor's approval decision.

All PR staffs involving in news publishing can login securely through https connection and take responsible actions. Other faculty staffs who have access to DocFlow cannot open news article without permission from each document creator in the PR news publishing workflow. If one of the PR staffs forgets to complete a task within 2 business days, DocFlow will send a reminder via e-mail and system message to everyone in the workflow indicating a problem in the flow. In the aspect of document management system, if the news writer would like to look for news articles related to the faculty's soccer activities happening during December 2010, he/she can use document management service of DocFlow to search for the news articles which should also be displayed in different versions in the search results. Thus, DocFlow can help make task collaboration and document management simple, organized and effective.

5 DISCUSSION AND FUTURE WORKS

DocFlow is a simple workflow solution that combines key features of document management systems together with workflow features. DocFlow also support different types of document formats. In order to use DocFlow, clear understanding of business processes must be realized first. This step is the most difficult step in most business process management effort because of inter-departmental complex communication.

DocFlow lacks communication platform because it does not integrate seamlessly into e-mail communication application such as Microsoft Outlook and Horde web-based e-mail service. This can increase the work that workers have to perform on the daily basis. Today, an organization uses many types of communication channels which can be categorized by medium and application

types. The workflow and document management system should integrate common communication channels and formats together rather than create a new one because all of them are used to facilitate communications and collaboration within an organization. In addition, workflow should support team collaboration in such a way that task completion can be approved by a team consensus or decision maker.

Computer-supported task organization has been shown by many researchers and business managers that it can significantly improve the performance of workers who collaborate. Miscommunication and inefficiency can be reduced when workflows are clearly defined. Documents can be located and tracked quickly through document management system. In general, each worker is presented with a clear workbook that he shares with other workers. The workbook has clear task assignments and progress level report. However, it is not possible to put all human tasks in a computerized workbook. Modelling human tasks sometimes cannot be documented and computerized. Computerized Workflow should be used largely to help making decisions, keeping milestones of tasks, and managing documents.

6 REFERENCES

1. HP AutomateWorkflows, http://h71028.www7.hp.com/enterprise/us/en/ipg/workflow-automation.html
2. Xerox Document Management, http://www.realbusiness.com/#/document-management/service-offerings-dm
3. EMC Documentum, http://www.emc.com/domains/documentum/index.htm
4. Bonita Open Solution, http://www.bonitasoft.com
5. CuteFlow - Open Source document circulation and workflow system, http://www.cuteflow.org
6. Wikipedia, http://www.wikipedia.org
7. Google Docs, http://docs.google.com
8. Subversion, http://subversion.tigris.org
9. IBM Lotus Workflow, http://www.ibm.com/software/lotus/products/workflow
10. IBM Websphere MQ Workflow, http://www.ibm.com/software/integration/wmqwf
11. SAP ERP Operations, http://www.sap.com/solutions/business-suite/erp/operations/index.epx
12. Microsoft SharePoint, http://sharepoint.microsoft.com
13. Sandhu, R., Ferraiolo, D., Kuhn, R.: The NIST Model for Role Based Access Control: Towards a Unified Standard, In: Proceedings, 5th ACM Workshop on Role Based Access Control, pp. 47--63. Berlin (2000)
14. van der Aalst, W., van Hee, K.: Workflow Management: Models, Methods, and Systems (Cooperative Information Systems). The MIT Press (2002)
15. Attinger, M, Blurring the lines: Are document management software and automated workflow the same thing?, In: Information Management Journal, pp. 14--20 (1996)
16. Cardoso, J., Bostrom, R., Sheth, A., Workflow Management Systems and ERP Systems: Differences, Commonalities, and Applications, Information Technology and Management, 5, pp. 319--338 (2004)
17. Basu, A., Kumar, A., Research commentary: Workflow management issues in e-business, Information Systems Research, vol. 13, no. 1, pp. 1--14 (2002)
18. Stohr, E., Zhao, J., Workflow Automation: Overview and Research Issues, Information Systems Frontiers, 3, 3, pp. 281--296 (2001)
19. Harpaz, J., Securing Document Management Systems: Call for Standards, Leadership, The CPA Journal, 75, 7, pp. 11. (2005)
20. Neal, K., Driving Better Business Performance with Document Management Processes, Information Management Journal, 42, 6, pp. 48--49 (2008)
21. Paganelli, F., Pettenati, M. C., Giuli, D., A Metadata-Based Approach for Unstructured Document Management in Organizations, Information Resources Management Journal, 19, 1, pp. 1--22 (2006)

22. Wassink, I., Rauwerda, H., van der Vet, P., Breit, T., Nijholt, A., E-BioFlow: Different Perspectives on Scientific Workflows, Bioinformatic Research and Development, Communications in Computer and Information Science, vol. 13, 1, pp. 243--257 (2008)