

# A Stage-Activity Process Model Facilitating Workflow Management for Web Publishing

*Jiannong Cao, Catherine Chan, and Keith Chan*

*Internet Computing and E-Commerce Lab  
Department of Computing  
Hong Kong Polytechnic University  
Hung Hom, Kowloon, Hong Kong*

## Abstract

*In this paper we describe workflow analysis for web publishing based on a general stage-activity process model. Using a case study approach, existing web publishing methods used by major Hong Kong ICP companies have been evaluated. Based on the evaluation results and our observations, issues and problems of the existing methods are identified. A general process model, called the stage-activity model, is proposed which facilitates the development of a framework for automating the Web publishing process, incorporating workflow management. The results and outcome of this research contribute to a clear understanding of what challenges ICPs currently face and a solution to building an automating web publishing system. The framework can also be used to evaluate and select web publishing tools in the market.*

## 1. Introduction

The continuous advance in the Internet technology and related services has created many opportunities for business organizations and companies to conduct electronic business on the Internet. There have been a large number of Internet companies appearing in the market place, offering a variety of services in different areas of E-Commerce, such as B2C, B2B, and C2C. Among the most commonly focused segments under the B2C arena is Internet Content Provider (ICP), whose main service is concerned with web publishing - providing on their web site quality contents, including news, stories, articles, etc., on a timely basis [4, 6]. Web publishing has been the core business of ICP (Internet Content Provider) companies. It is also a major functional component in other internet industry sectors, which require publishing contents / information on the Web. As a result, web publishing is vital to Internet companies operating in all areas, although each with a different degree of significance.

For Internet content, as for most information goods, there is a large cost of developing and providing the

content. Furthermore, efforts are needed to ensure the quality of the content provided. Therefore, it is usually required to have a team of producers working together during the whole web publishing process, conducting quality content development and management, collaboration, and control over the site design. However, not many ICPs in Hong Kong are able to manage and publish their content in an efficient way. To certain extent, this is also true for ICPs worldwide. Although there has been on-going developments on web publishing, up to our knowledge, very little work has been done in the area of process models and methodologies for web publishing. Web publishing remains one of the most chaotic areas [2, 3]. Most existing work on web publishing have been focused on formats, markup languages, HTTP servers, browsers, and other technical issues which are of concern to web authors. There is no systematic approach and support to facilitate the coordination of collaborative work amongst the web publishing team, and a lot of the technical implementation was done on an ad hoc basis.

In this paper, we propose a general framework based on workflow management that helps Internet companies, ICPs in particular, to streamline their Web publishing operations. Using a case study approach, existing web publishing process used in major Hong Kong ICP (Internet Content Provider) companies have been evaluated. Based on the evaluation results and our observations, the issues and problems of the existing methods are identified, and then a stage-activity process model is proposed. To develop a process without thinking about how it will be automated is academic [16]. Therefore, we have developed a framework based on the proposed process model which facilitates the automation of the web publishing process. The framework is based on web publishing workflow analysis and management incorporating the use of a computer-based automating Web publishing system.

Section 2 provides the background of our study and describes the related works. Section 3 reports our case studies for evaluating web publishing methods currently adopted by major Hong Kong ICP companies. Section 4 proposes a framework for streamline and automating Web

publishing operations. is proposed, including the stage-activity process model, a computer-based system for automating the web publishing process., and the architecture of an automating web publishing system, as well as the workflow management model incorporating the use of the automating system. Finally, Section 5 concludes the paper and describes our future work.

## 2. Background and related work

In this paper, we refer to Internet Service Providers (ISPs) that primarily provide content as Internet Content Providers (ICPs). ICPs are one of the major parts in the Internet economy model [4]. They collect, process, organize, index, update, and provide the data in a format usable by customers. An ICP can have a vertical focus or a horizontal focus. A vertical ICP provides content pertaining to a particular industry or area, e.g., a financial site providing contents on financial news, stock quote, and so on. A horizontal ICP provides contents on a wider range of topics, e.g., a community site focusing on providing a variety of contents as in daily news, horoscope, city events, and so on.

There are ICPs that adopts a full cyber marketing approach and whose core business is web publishing. Such an ICP develops and maintains a quality Web site, where the production of the contents starts from within the company. Web contents are produced from scratch by using various languages and authoring tools. There are also ICPs, such as a newspaper company publishing its newspaper through both traditional print media and on the web, that adopt a partial cyber marketing approach, where its web publishing business is only an extension of its core business in other industry. The production of the contents does not start from the ICP division of the company since the contents originate from its core business. In general, the web publishing methods used by these ICPs are specific to the industry of its core business.; the web publishing process starts from transferring the contents that make up of the printed products to their web site.

Related work can be classified into three main categories: electronic document management [8, 15], tools and techniques [2, 9, 13, 14], and underlying process models [3, 10]. Although all of them provide insights and may also be applied to our research, in this paper we focus on the development of web publishing process models and systems. More specifically, we focus on identifying the activities in web publishing process and the key components that are required for an integrated web publishing system that support workflow management.

A process defines a description and ordering of work activities across time and space that is designed to yield specific products or services while ensuring the organization's overall objectives [12]. It provides a conceptual basis for the integration and coordination of distributed resources, tasks, and individuals [5]. The web publishing process aims to achieve the objective of delivering *timely* and *accurate* contents using an

integrated and coordinated system involving both people and technology. Timely contents, in this context, means up-to-date contents, where the status of contents is defined according to the site policy. Accurate contents refers to contents free from errors such as broken or wrong links, missing or wrong images, wrong aggregation of materials, wrong data / materials and distorted layouts, etc. Furthermore, investment on human resources for maintaining an accurate and timely site is usually large, which raises a continuous problem for management. Schedule delays and long development time are frequently encountered. Finally, factors such as lack of involvement of technically skilled people, inappropriate tools, and uncertainty on how WWW technology can be integrated with the existing organization business model also contribute to the difficulty of creation and maintenance of electronic document [8, 11].

A general model of online publishing focusing on the sequence of major activities involved has been proposed by David G. Green [10]. Five steps, namely *submission*, *acquisition*, *quality control*, *publication* and *delivery*, have been identified as the main steps involved in web publishing process. The model, however, does not include the stage required prior to submission, which includes activities covering production work by web editors and web designers to produce the contents / materials needed. As pointed out by Seesing [11], there are lots of pitfalls in the content production stage, and having the right process workflow plan for this stage requires much thoughts and attention. As we will see, bottlenecks and needs for additional resources often occur in this part of the web publishing process, which results in overall schedule delays and budget overruns. The lack of integration of the production stage into the process model makes the total control of web publishing difficult.

Unlike client / server applications, web publishing applications are intended not only for experts in a given domain, but are also for the general public, including partners, knowledge workers and clients. This means new flows of information and new requirements [2]. The web publishing process model proposed in this paper is obtained using a workflow analysis approach and thus facilitates workflow management for web publishing. In general, a workflow management system has been considered as a computer automated infrastructure where a group of people participate together to achieve a common goal following some predefined rules and task assignments [5]. In the context of this paper, we aim at identifying the key activities carried out by the respective individuals / entities at each stage and proposing a framework for coordinating the activities.

## 3. An evaluation of web publishing methods used by Hong Kong ICPs

In this section, we describe case studies of three major Hong Kong ICPs. The main purpose of the studies is to evaluate the web publishing methods currently used by these ICP companies so that we can identify the key issues

and problems in existing practice and propose potential areas for improvement. The results of the evaluation study and our observations derived from the results are discussed, which provide the basis of analyzing the web publishing workflow and developing the stage-activity process model.

Three major Hong Kong ICP companies have been evaluated. The first case is the on-line division of a major directory company. The company has an existing core business in print media and is in the third year of running the online business. In the very beginning, the directory site was developed for the purpose of offering only online search function of the directory listing. With the recent take off in the internet industry, the company keeps on adding new features and channels to the site, including a search engine for advertising listing of specific industries. Up till now, there is no Internet technology expertise in the online division to support the team. Most of the technical development and web publishing tasks are outsourced to major IT vendors and supported and maintained by the division. The second case is the portal group of the ISP division of a major Telecommunication company, which has an existing core business in providing mobile communication services and is in the second year of operation in the ISP business. The portal group of the ISP division focuses on building and maintaining the portal site of the ISP division. It has a business model of an ICP, that is, it provides quality contents to attract traffic to the site and monetizes the traffic of the site by revenue generated from online advertisements. In order to make the site attractive, more and more new channels with many sub-channels are being added to the site, featuring contents on travel, IT, entertainment, finance, E-commerce, etc. The company's initial strategy for building and maintaining the site was to leverage the resources from the existing core business. This resulted in the lack of Internet related technical expertise in the group. Most members of the content development team are hired from outside. Outsourcing as in the form of using free lance editor for content development is also frequently used. The last case is a pure ICP company whose main business focuses on producing quality web site and making revenue out of online advertisement. The company is in the second year of operation at the time of interview. It adopts two different approaches in developing contents. Its Web site contains many different channels and sub-channels. In particular, there is a news channel, which is updated daily and produces about 30 news per day. The other channels, including e-commerce, business computing, and games review, are produced and updated less frequently.

Using an extended model of the web publishing process based on Green's work [10], we divide the web publishing operations into six stages:

- (a) Production - creation of the contents by web editors and web designers using web publishing tools and the development of templates that hold these contents.
- (b) Submission - submission of the produced contents for approval and publish.
- (c) Acquisition - reception & retrieval of the submitted contents.
- (d) Quality control - quality check on the contents as well as approval of the contents for advancing to the next stage.
- (e) Publication - publication of the approved contents to the preview platform for preview
- (f) Delivery - final release of the approved contents for public access.

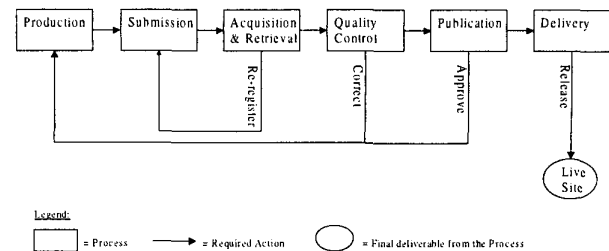


Figure 1: Extended Process Model of Web Publishing

As we can see from Figure 1, according to the model, the web publishing process follows a linear but iterative model. Each stage carried out one after the other but, in the case where the produced contents do not pass the quality control, the process iterates back to previous stages to fine tune the activities made. This is common in web publishing as most contents are developed through a series of fine-tuning and approval activities. The iteration from Acquisition & Retrieval stage back to Submission stage is expected to happen only when system error occurs during acquisition of contents.

As a common finding for all of the three cases, content development and content / site management involve the following areas of expertise:

- Web Designer – responsible for the production of graphical contents
- Web Editor – responsible for the production of textual contents
- Content Manager – responsible for overseeing all of content development activities of the site
- Web Specialist / Web Master – responsible for technical support to content development team, programming markup language to facilitate content development, and managing the contents in the areas of uploading, downloading, updating, backup, and archiving the contents

Staffs involved in each stage of the web publishing process were interviewed. A well-prepared questionnaire was used to conduct the interviews (see appendix). There are two main aspects in the evaluation study:

1. The workflow of the web publishing process which focuses on the major phases and the activities conducted by members in the team;
2. The mechanisms that are used for facilitating and integrating each phase in the web publishing process.

The questionnaire was designed to facilitate the identification of key issues related to the two components. It consists of two main parts, namely *content development*

and *content/site management*, which contain a total of 49 questions covering operations, entities, activities, and management issues involved in different stages of web publishing. Through the case studies, it was found that the underlining web publishing process adopted by each company matches quite closely with the six stages in Figure 1, although there was no such a clear cut between the six stages, and the ways to carry out the activities within these companies were quite different. Problems like inappropriate segregation of activities, non cohesive and non structured web publishing process, lack of collaborative platform to facilitate the whole process, and system developed for inappropriate user group were found as the major concerns. There are two major reasons: the web publishing methods adopted by the ICP companies are still very primitive and, as a consequence, planning and management are difficult.

Companies in cases 1 and 2 actually adopt a fully manual approach to coordinate web publishing activities. The company in the third case adopts a fully manual approach to publish almost all the channels and sub-channels of the site except the news channel. For publishing the news channel, which demands high volume of contents being delivered every day, a more automated approach is used. The content / site management team developed a small-scale news publishing system, called News Admin System, to support the content development phase of publishing news articles. The system has a content management tool by which web editor / free lancer can produce news articles using templates and then store them into a single database. However, there are many limitations. There is no access control over the system and no version control, and thus no way to verify who produced the articles. Also, there is no support for the maintenance of quality and integrity of the contents being produced and stored in the database. More importantly, there is no integration between this system and the rest of the publishing process for other channels. Coordination of activities involved in content development and content / site management remains to be manual-driven either over the phone or via email.

As we observed, under the manual methods, most of the activities are done separately by different entities at different stages. There is no systematic approach and support to facilitate the coordination of collaborative work amongst the web publishing team. For example, often, editors and graphic designers use different design and authoring tools for the production of the contents. Submission and approval of the produced contents are done manually rather than via an integrated web publishing system. Editors and operation managers rely on technical teams for uploading and maintaining the contents of the site. There is no standard operational procedures or schedules in operating and maintaining the site. Most of the time, content / site management is done on an ad hoc basis either by internal team or outsourced

vendor. Furthermore, there is no centralized repository for storing submitted contents, which is essential for automating the process [7]. Consequently, a lot of efforts are required and a large amount of time spent on manually coordinating the activities of different people and integrating the results from different stages. Integration works done manually by web specialist / web master are very repetitive and tedious and utilize a lot of technical resources which can be better used elsewhere. Moreover, contents are often approved at the last time, which puts a lot of stress on the team for maintaining updated contents on the site. Under the situation where content development team uses free lancers for content development, it makes the process even more complicated and time consuming. As this practice accumulates, it demoralizes the team spirit and results in high turnover of staff in these companies. As a matter of fact, most ICP companies face this same situation.

#### **4. The Stage-Activity process model and workflow management for web publishing**

The problems that we have identified in the existing web publishing methods urge the companies to seek solutions to building a web publishing system to manage the site. The availability of an integrated web publishing system can also help to control and maintain high quality contents on the site.

First, a general process model has been developed based on the extended web publishing process model and the results of the case studies. The so-called stage activity model is obtained by analyzing the workflow in a streamlined web publishing process, which facilitates the automation of web publishing process. Since publishing a live site involves many people in different roles cooperating to produce contents that need to pass several quality control points at various stages, use of a workflow management based approach for facilitating quality control and communication activities should greatly improve the productivity of the web publishing process.

As illustrated in Figure 1, the tasks involved in web publishing range from the initial stage of production to the final stage of uploading and maintenance of the site. When the activities of each stage are carried out smoothly, the whole process reflects a linear model, but, in practice, the process usually proceeds following a sequence of iterations of the activities. In the stage activity model proposed in this paper, activities are organized into the six stages. Table 1 defines the functions at each stage and shows the activities conducted at each stage.

Second, given the stage-activity model, a framework can be developed that provides a basis for building an automating system to streamline the process and increase the productivity of content development and content / site management teams. It provides user-friendly tools

Table 2: The Stage-Activity Model for Web Publishing Process: Activities

<u>Stage</u>	<u>Activities</u>
<b>S1 Production</b> <ul style="list-style-type: none"> <li>Automate and streamline the production process for textual content development</li> <li>Provide input function for upload schedule and display order of the contents</li> </ul>	S1 A1 – Develop site template upon receiving job order S1 A2 – Develop textual contents upon receiving job orders S1 A3 – Develop graphical contents upon receiving job orders S1 A4 – Input upload schedule and display order of the contents
<b>S2 Submission</b> <ul style="list-style-type: none"> <li>Facilitate automatic submission process for all templates and contents</li> <li>Facilitate automatic submission of upload schedule and display order on submitted contents.</li> </ul>	S2 A1 – Submit site templates layout S2 A2 – Submit textual and graphical contents S2 A3 – Submit upload schedule and display order of submitted contents
<b>S3 Acquisition &amp; Retrieval</b> <ul style="list-style-type: none"> <li>Ensure that submitted contents are received properly and version applied</li> <li>Write (store) and read all submitted contents to and from central content repository</li> </ul>	S3 A1 – Receives submitted site templates & contents S3 A2 – Perform file version control on submitted items S3 A3 – Stores templates and contents in central content repository S3 A4 – Retrieval of submitted templates and contents from repository
<b>S4 Quality Control</b> <ul style="list-style-type: none"> <li>Coordinate all job orders and acknowledgements</li> <li>Coordinate all approval requests and ensure all submitted contents have proper approval</li> </ul>	S4 A1 – Coordinate job orders & acknowledgement for development works S4 A2 – Coordinate approval requests & status on submitted contents S4 A3 – Ensure contents are approved before advancing to publication stage S4 A4 – Coordinate adjustment requests on submitted contents S4 A5 – Coordinate and ensure final approval on preview site
<b>S5 Publication</b> <ul style="list-style-type: none"> <li>Create preview site using approved contents and its upload schedule and display order</li> </ul>	S5 A1 – Production platform receives approved contents S5 A2 – Production platform creates and display preview site with approved contents based on uploaded schedule and display order S5 A3 – Initiate final approval on preview site S5 A4 – Transfer approved preview site to live platform
<b>S6 Delivery</b> <ul style="list-style-type: none"> <li>Upload preview site to live platform upon approval signal</li> <li>Perform back up and archive functions on old contents based on pre-programmed schedule and procedures</li> </ul>	S6 A1 – Receive uploaded files from preview platform S6 A2 – Generates live site for public access  S6 A3 – Backup and archive old data based on pre-programmed schedule and procedures

for producing and managing contents, storing contents and related information in a common repository, and automating labor intensive repetitive tasks. It is also expected to increase the quality of the contents by applying access control to the system, ensuring integrity of the submitted contents, facilitating and ensuring quality control of the contents, and ensuring timely delivery and regular back up and archiving of old contents.

The proposed automating system consists of three major component subsystems, namely, *Staging System*, *Publication and Delivery System*, and *Central Archive System*. The Staging System is the core component of the proposed system that manages all activities prior to publication and delivery stage and is the main focus of the proposed system. The Publication and Delivery System handles all activities for publishing the preview site and final delivery of the live site for public access. The Central Archive System is responsible for all activities on backup and archiving of old contents and related data. Finally, the main function of the *System Administration Module* is to administer the overall administration activities of the proposed system. The main user of this module is system administrator of the proposed system. Based on

instructions from content manager, system administrator can use this module to handle the creation and maintenance of all user identification and access rights to the proposed system.

Figure 2 illustrates our workflow analysis and management model incorporating the use of the automating system. They demonstrate how different entities within content development and content / site management teams carry out the various web publishing activities using the proposed system. To validate the proposed process model, a second round of interviews have been conducted. The proposed stage-activity process model and the automating web publishing system closely match the expectation of the companies that have been interviewed. It is expected that the introduction of the automating system can help streamline the web publishing process, separate the responsibilities and concerns of team members, ease the tasks involved in different stages, facilitate the collaboration and coordination, and reduce the difficulties and cost of maintaining quality contents. Several implementation considerations have also been identified during the interview. These include ease of integration with rest of the

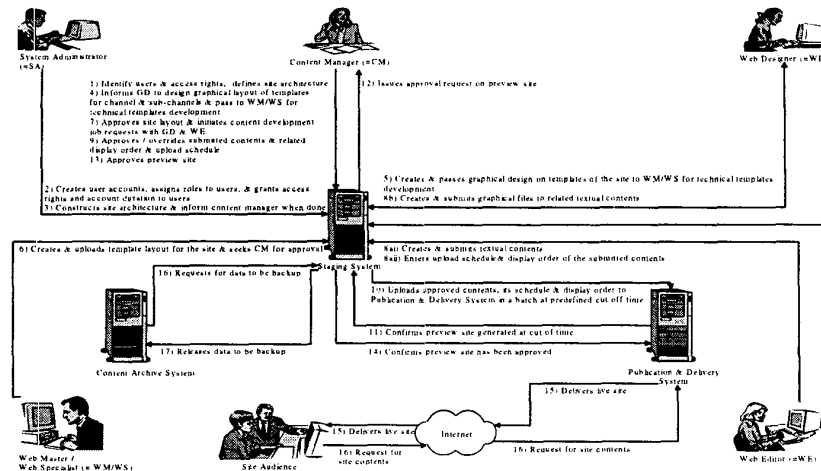


Figure 2: The Proposed Web Publishing Process Model – Workflow Analysis

technical set up of the company, organization restructuring, change over plan, and user interface issues.

Currently, we are building a prototype of the proposed workflow management system using WAT, a generic workflow automation tool developed by us. A separate paper is under preparation, which describes the structure, functional components, and a client-server based system architecture of the proposed automated web publishing system [1].

## 5. Conclusions

In this paper, we have described a general process model and a workflow management based automating system for web publishing. The results and outcome of this research contribute to a clear understanding of what challenges ICPs currently face and a solution to automating the web publishing process. Using case studies we were able to identify key factors that contribute to the difficulties and problems in each stage of the web publishing process. Then, based on workflow analysis, we have proposed a general, stage-activity process model and discussed how the model can be used to facilitate the development of an integrated computer supported system for automating the web publishing process.

Our future work includes designing and building a prototypical software platform for an integrated web publishing system, which supports the proposed process model. The prototype can be used to further validate the practical usage and applicability of our findings and framework.

## References

- [1] J. Cao, C. Chan, and K. Chan, "Toward the automation of the web publishing system", under preparation.
- [2] Christine, Aceto, "Managing Web content", *Document World* 4(5), Sept / Oct 1999, p. 18-23
- [3] David M. Choy and Robert J. T. Morris, "Services and Architectures for Electronic Publishing", IBM Almaden Research Center, *IEEE*, 1996, p. 291-297

- [4] R.Dewan, M.L. Freimer, and A. Seidmann, "Internet Service Providers, Proprietary Content, and the Battle for User's Dollars", *Communications of the ACM*, Vol. 41, No. 8, Aug. 1998, pp.43-48.
- [5] Andrzej Cichocki, Abdelsalam (Sumi) Helal, Marek Rusinkiewicz, Darrell Woelk, "Workflow and Process Automation: Concepts and Technology", Boston: Kluwer Academic Publishers, 1998, p.1-2
- [6] T. Matthew Ciolek, "Today's WWW - Tomorrow's MMM?", *Computer*, Jan 1996, p. 106-108
- [7] Dzlina, Dzintars, "Publishers look to the database for efficient content repackaging", *Folio: the Magazine for Magazine Management* 28(8), Jul 1999, p. 88-89
- [8] Lars Bo Eriksen, "Digital Documents, Work and Technology, three cases of Internet News Publishing", *Proc. 13<sup>th</sup> Hawaii Int'l Conference on System Sciences*, Vol. 6, 1997, pp. 87-96
- [9] Schubert Foo, Ee-Peng Lim, "A hypermedia Database to Manage World-Wide-Web documents", *Information & Management*, vol. 31 (1997), p 235-249H
- [10] David G. Green, "A General Model for On-line Publishing", *Proceedings of AUUG'96 and Asia Pacific World Wide Web '96 Conference*, Australian Unix Users Group, Sydney, 1996, pp. 152-158
- [11] Paul R. Seesing, "Electronic Publication on the World-Wide Web: The Process and the Pitfalls", *IPCC96, IEEE*, May 1996, p. 145-155
- [12] Ian Sommerville, "Software Engineering", *Addison-Wesley Publishing Company*, Fifth Edition, 1996
- [13] Steven J. Vaughan-Nichols, "Internet Publishing Tool Proliferate", *Byte*, Mar 1995
- [14] Hakon Wium Lie, Janne Saarela, "Multipurpose Web Publishing using HTML, XML, and CSS", *Communications of the ACM* 42(10), Oct 1999, p. 95-101
- [15] "IEEE Recommended Practice for Internet Practices - Web Page Engineering - Intranet / Extranet Applications", Sponsored by the Internet Best Practices Working Group, *IEEE Computer Society*, May 1999
- [16] I. Jacobson, G. Booch, and J. Rumbaugh, "The Unified Software Development Process", *Addison Wesley*, 1999.