

Automatic content filtering and publishing by email for the UP Open University Doctor of Communication Programme learning management system

Ria Mae H. Borromeo

University of the Philippines Open University, Philippines

Abstract

The DComm website is an electronic learning environment using the Joomla Content Management Framework (CMF), designed specifically for the Doctor of Communication Programme offered by the University of the Philippines Open University (UPOU). Previously, when a faculty member or staff wanted to publish content on the website, content was first emailed to the web administrator, who uploaded it manually. While this method ensured that proper content was published on the site, publishing was delayed at times due to the unavailability of the web administrator. To enable DComm faculty members and staff to directly publish content without having to study web publishing or going through the web administrator, publishing content through email was enabled using Post-by-Email, an extension for Joomla that allows such functionality. The resulting website now allows authorised users to publish information in a particular section of the site by emailing content to a specific email address. However, without quality control over content that gets published, this practice raises the possibility of inappropriate content publishing. To address this issue, a strategy for content filtering and publishing that uses existing technologies such as email filters and SpamAssassin, an open source mail filter based on content matching rules, was designed and implemented. The result is a learning environment where faculty members and staff can automatically publish filtered content through email, making it immediately accessible to the students, ultimately providing for a more dynamic learning environment.

Introduction

A Learning Management System (LMS) is an online system that allows information sharing and online collaboration among users. Several LMS and e-learning systems are available in the market today such as *ATutor*, *Moodle*, *WebCT*, *Learn.com*, *Joomla LMS*, *Krawler LMS* and *Blackboard* (Al-Busaidi & A-Shihi, 2010). The main LMS of the University of the Philippines Open University, called *MyPortal*, is based on *Moodle*. *Moodle* is a software package for producing Internet-based courses and web sites. In *MyPortal*, every course offered by the university has a virtual classroom where teachers can deliver content to students and assess learning using assignments and quizzes. Moreover, both students and teachers can use activity modules such as forums, databases and wikis.

While *MyPortal* caters to the needs of individual courses, it does not provide collective support for an entire degree programme. For example, if there is an announcement for the Doctor of Communication (DComm) Programme and the teacher posts such an announcement in *MyPortal*'s virtual classroom, the announcement can only be seen by students enrolled in the course and not by all students of the DComm programme. The DComm website was developed to provide more opportunities for collective and collaborative learning among all DComm students, faculty and staff.

The DComm website is made using *Joomla*, an open-source content management system (CMS), which can be used to easily create and build a variety of websites and web-enabled applications. The website contains detailed information about the DComm Programme, learning resources such as videos, academic papers, links to references, a section where all registered members can contribute and interact through forums, and a page where official announcements from the faculty and staff are posted. It is accessible to the public but some features, such as posting in forums, require user authentication.

Currently, all DComm students, faculty and staff are registered members of the website. This enables them to access all resources in the website and contribute in the forums and chat box. Aside from that, all other information must go through the web administrator to get published in the website. Since content publishing is dependent on the availability of the web administrator, delays in content publishing were encountered. To avoid delays and automate the publishing process, the publishing of content through email was enabled using *Post-by-Email*, an extension for *Joomla* that allows such functionality. This functionality enables authorized users to publish information in a particular section of the site by emailing content to a specific email address.

The implementation of such functionality, however, raises the risk of inappropriate content publishing if there is no quality control over the content that gets published. To address this issue, *SpamAssassin*, an open source email filter based on content matching rules, was used to automatically filter unsolicited bulk email (UBE), more commonly known as spam, from the mailbox. Moreover, an additional filter based on a blacklist of content classified as inappropriate was also created.

The result is a learning environment for the DComm Programme where faculty and staff can directly publish automatically filtered content through email, making it immediately accessible to the students.

In summary, the following are the contributions of this paper: the development of an LMS for the DComm Programme, the enhancement of this LMS by allowing automatic content publishing through email, and the development of a filtering strategy for the content from email.

Related works

The DComm website was created to satisfy several needs that are not met by *MyPortal*, the main LMS of UPOU, which has been using *Moodle* since 2007. However, *Moodle* provides little flexibility when it comes to collaborative learning. In *Moodle*, the compartmentalized nature of courses within the system inhibits the sharing of learning objects and resources among the users (Librero, 2011). Due to these limitations, enhancements to the LMS have been made. In 2011, Librero expanded the functionalities of *MyPortal* by installing modules and plugins within *Moodle* and integrated external web applications such as conferencing, blogging, e-portfolios and rich media content management through single sign-on (SSO). The resulting LMS was used in two courses for the Bachelor of Arts in Multimedia Studies programme.

According to the article in the *Moodle* tracker entitled *Posting by Email*, the publishing of content within *MyPortal* through email is a feature that is not yet supported by *Moodle*. However, posting content through email is not a new feature on the web. It is a feature readily available in popular blog hosting sites such as *Blogger*, *Wordpress* and *Tumblr*. It is also available in social networking sites such as *Facebook* and *Twitter*. These sites provide a special email address where posts to be published can be sent.

In *Facebook*, aside from having an email address where posts can be sent, notifications received through email can be replied to and the reply is automatically posted on *Facebook*. Another service called *send2page* allows publishing of website content through email. Like the sites mentioned earlier, *send2page* provides a special email address where updates to a specific area of the registered website can be sent. For the *Joomla* CMS, which powers the DComm website, there are several extensions that implement publishing content through email such as *Powermail* and *Post-by-Email*. These extensions access a specified email account and then publish the messages received from the inbox depending on the rules created by the web administrator.

The additional feature of content publishing through email introduces the email spam-filtering problem to the LMS. A lot of research effort has been devoted to this problem. In 2011, Dasgupta, Guervich, and Punera proposed a method for enhanced email spam filtering through combining similarity graphs. Based on their experiments on real-world email messages, the method was shown to lead to significant improvements compared to two baselines in previous research. Currently, there are commercial and open-source out-of-the-box spam filtering software available, which can be installed and integrated to mail servers.

In this study, *Joomla* was used to implement the LMS for the DComm Programme, *Post-by-Email* was the *Joomla* extension used to implement automatic content publishing through email, and *SpamAssassin*, an open source anti-spam software, was used for email spam filtering.

The DComm website

Sitemap

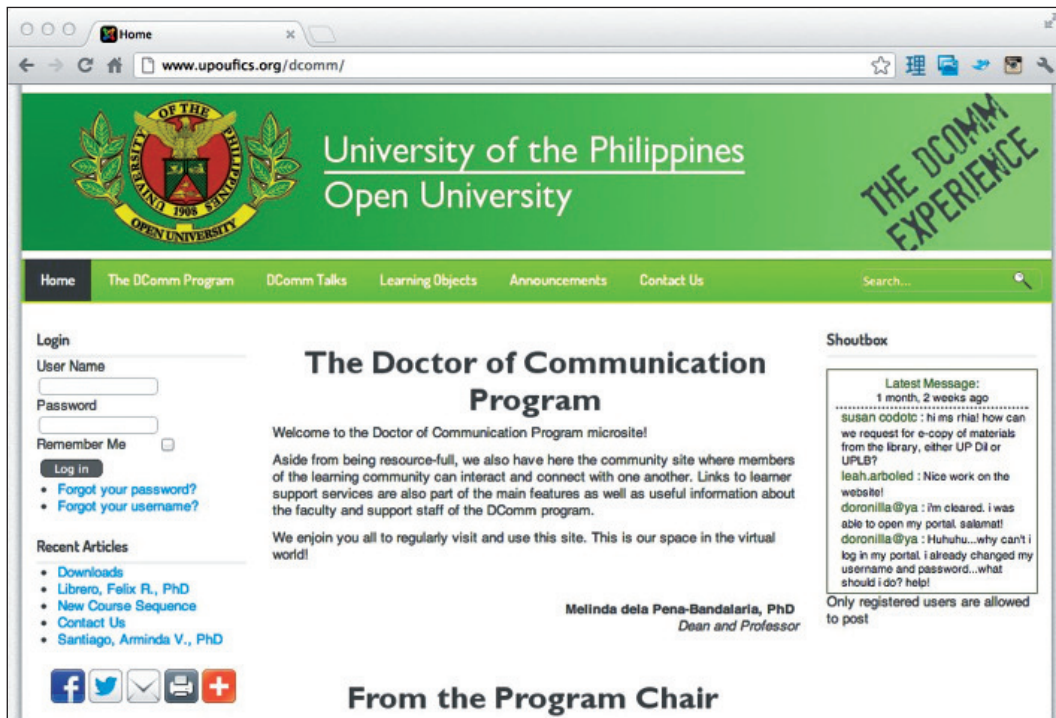


Figure 1 The DComm website homepage

The DComm website is divided into six sections: the homepage, the DComm Programme information page, *DComm Talks*, *Learning Objects*, *Announcements* and *Contact Us*. The homepage contains an introduction to the site. The DComm Programme information page contains details about the admission requirements, academic programme requirements, and the DComm faculty. The *DComm Talks* page contains 10–15 minute videos on various topics related to the content, experience, and goals of the DComm programme. The *Learning Objects* page contains links to student papers, DComm forms, and a link to the File Uploader where students can upload their work. The *Announcements* section, which is the most frequently updated section, contains announcements from DComm faculty and/or staff. Lastly, the *Contact Us* page contains information on how to communicate with the DComm Programme staff. In all of the pages, a chat box where registered users can leave messages, links to articles recently published in the site, and a login box that becomes a user menu once the user is authenticated are visible.

Content publishing procedure

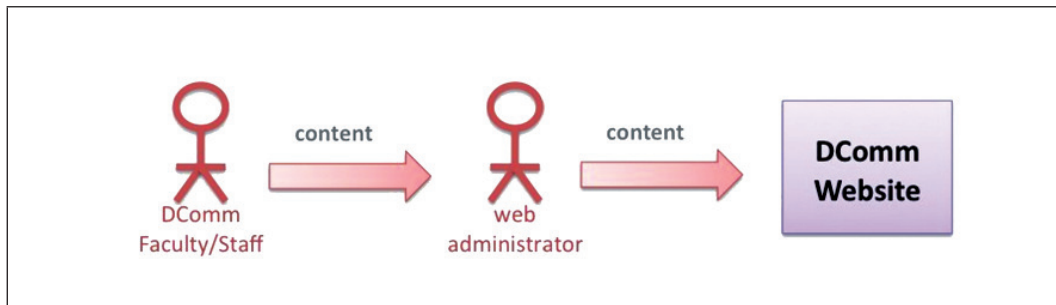


Figure 2 The content publishing process

To publish content in the DComm website, the faculty or staff member first sends the content to be published to the web administrator. The web administrator then manually posts the content on the website. This method ensures that appropriate content is published in the website. However, since it is a manual process, it relies highly on the availability of the web administrator.

Automatic content publishing and filtering by email

In this study, the functionality of publishing content in the DComm website by sending an email to a particular address was enabled. The process by which the email sent by a DComm faculty or staff member is published is illustrated in **Figure 3**.

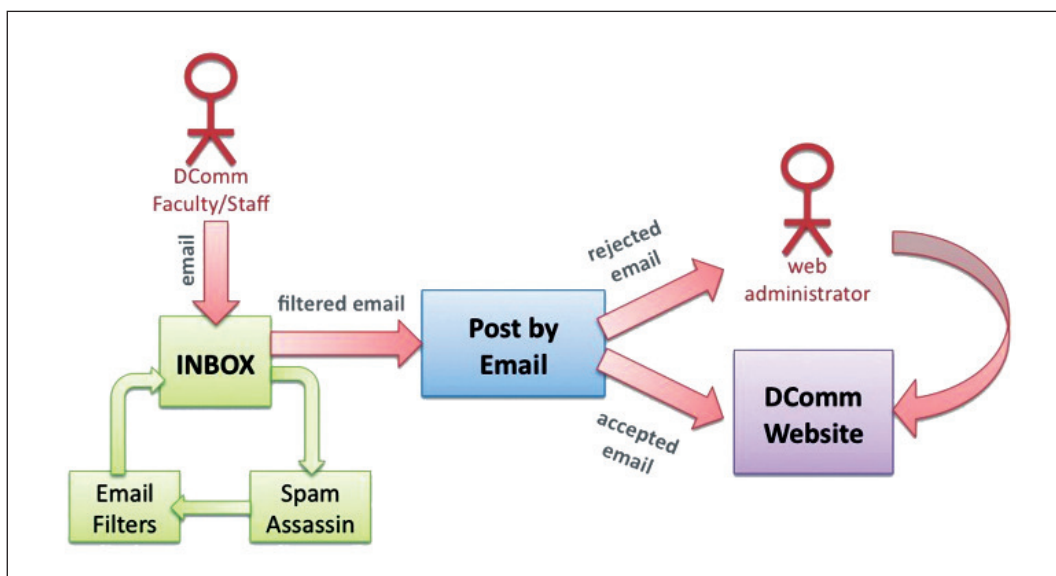


Figure 3 Proposed process of email content filtering and publishing

In this process, a dedicated mailbox receives the email sent by the faculty or staff. The *SpamAssassin* software, which runs in the mail server, automatically screens the mail received to see whether it is spam. If it is spam, it is ignored. Otherwise, it goes through another set of filters that determines to which section of the website the contents of the email are to be published. After that, the filtered mail is retrieved from the mailbox by the *Joomla* extension, *Post-by-Email*. *Post-by-Email* checks if the email address of the sender is registered in the DComm website and has permission to post such content. If the email fails to satisfy any of those conditions it is ignored. The accepted email is once again screened for blacklisted words such as vulgar words in the English and Filipino language. If the email passes the screening, it is automatically published in the DComm website. Otherwise, it is sent to the web administrator who can manually upload it based on his/her judgment.

The email format

The following email header describes the header format required for an email to be published in the website. The category, which specifies where the content will be published, should be explicitly written in the subject field.

To: dcomm@upoufics.org Subject: [Category]: [Article Title] Message: [html or plain text format]

Email filtering

Email filtering may refer to the automatic processing of email to organise it into different criteria or the use of anti-spam techniques and human intervention to prevent spam email. The proposed system has three filtering steps. The first step uses *SpamAssassin*, a mail filter that applies a diverse range of tests based on content matching rules to identify junk email. The tests are applied to email headers and content to classify email using advanced statistical methods (The Apache Spam Assassin Project, 2012). It runs on the mail server and filters spam before it reaches the inbox. *SpamAssassin* was enabled in the mail server where content to be published is sent.

The second step uses email filters, provided by the email client, to identify the section of the website to which a particular email must be published. The category in the subject field of the email is used to place the email into the appropriate section.

The third step checks the email against a list of blacklisted words. Blacklisted words are vulgar words and phrases from the Filipino and English language, which were taken from the *Banned Word List* and *Filipino Dirty Words*. This step is necessary since *SpamAssassin* may not filter emails containing these words because there are instances when these words are used in personal email exchanges. However, since the email is to be published in an academic website, these types of words should not be allowed.

Post-by-email

Post-by-Email is an extension for *Joomla* that retrieves emails from a mailbox and automatically publishes them as content in the *Joomla* website. Email content may either be in HTML or plain text format. Image attachments can also be published. In the *Post-by-Email* configuration created for the DComm website, *Post-by-Email* checks for new mail every five minutes. Email content is automatically published upon retrieval.

Results

In the initial test, four types of email were sent. The first email contained words in the blacklist. The second type did not use the correct email format. The third came from email addresses not registered in the DComm website. All were ignored by the system and were not published in the website. The fourth type had the correct format, contained no words from the blacklist, and came from a registered user. An example of this type of email can be seen in **Figure 4**. Such emails were published in the DComm website at most five minutes after they were sent. The corresponding output can be seen in **Figure 5**.

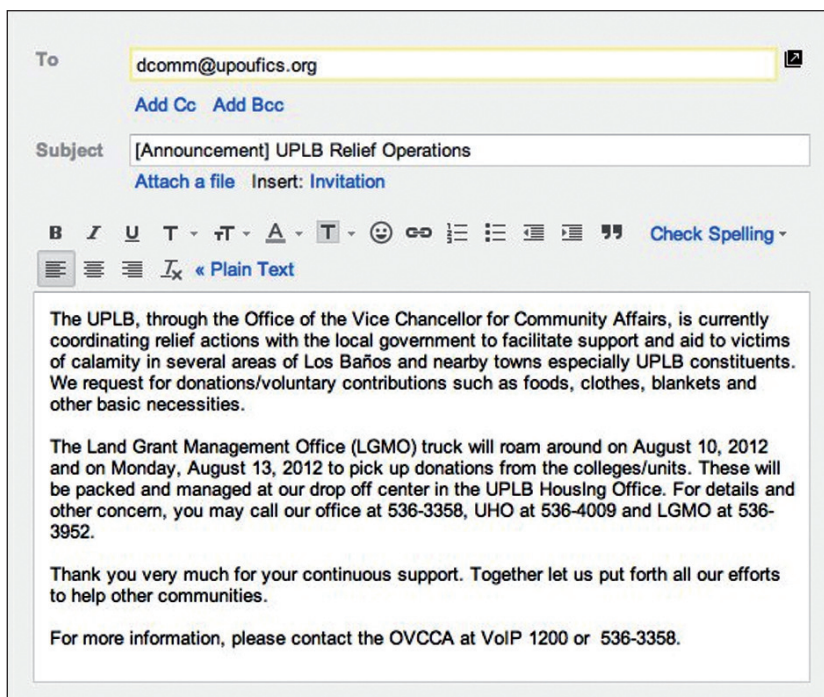


Figure 4 Sample valid email



Figure 5 Sample output of emailed content

Conclusions and future work

There is now an LMS that can provide collective support for the Doctor of Communication Programme. DComm faculty and staff members can post to this LMS by sending an email to a particular address. If the email satisfies the filtering strategy of the LMS, it is automatically published in the website. This procedure reduces the dependence of content management on the web administrator and makes content from faculty/staff members immediately accessible to the students.

In future, the filtering strategy can be enhanced. Currently, it only screens the text content. An improvement to this would be to also filter the image attachments in email. A mechanism for editing and deleting content through email would be another research direction. Lastly, another area of research would be to evaluate the effectiveness of the DComm LMS.

Ria Mae H. Borromeo (email: riamae.borromeo@upou.edu.ph) is with the Faculty of Information and Communication Studies, University of the Philippines Open University, Los Baños, Laguna, Philippines.

References

Al-Busaidi, K. A., & Al-Shihi, H. (2010). Instructors' acceptance of learning management systems: A theoretical framework. *Communications of the IBIMA*, Volume 2010. Retrieved from <http://www.ibimapublishing.com/journals/CIBIMA/2010/862128/a862128.html>

Banned Word List. (2009). Retrieved from <http://bannedwordlist.com/lists/swearWords.txt>

Bertino, E., Ferrari, E., & Perego, A. (2010). A general framework for web content filtering. *World Wide Web*, 13(3), 215 – 249. Retrieved from <http://dx.doi.org/10.1007/s11280-009-0073-5>

Dasgupta, A., Gurevich, M., & Punera, K. (2011). Enhanced email spam filtering through combining similarity graphs. *Proceedings of the fourth ACM international conference on Web search and data mining (WSDM '11)*, New York: ACM. Retrieved from <http://doi.acm.org/10.1145/1935826.1935929>

Filipino Dirty Words. (n.d.). Retrieved from <http://www.tribo.org/dirty/swear.html>

Librero, A. F. (2011). Augmenting the learning management system of the UP Open University [Abstract]. Retrieved from <http://www.allibrero.com/index.php/research/18-augmenting-the-learning-management-system-of-the-up-open-university>

Posting by Email. (2011). Retrieved from <http://tracker.moodle.org/browse/MDL-3714>

The Apache Spam Assassin Project. (2012). Retrieved from <http://wiki.apache.org/spamassassin/SpamAssassin>

What is Joomla? (2012). Retrieved from <http://www.joomla.org/about-joomla.html>