Advances in Wireless and Mobile Communications.
ISSN 0973-6972 Volume 13, Number 1 (2020), pp. 35-44

http://www.ripublication.com

Multi-format converter of course materials in the New Generation of Digital Open Universities (DOUNG)

Boubacar Tawayé Abdoul Aziz, Mahamadou Issoufou Tiado, Ibrahim Ganaou Noura, Col. Harouna Gazobi Souleymane, Hamani Mounkaila Mahamadou

Department of Mathematics and Computer Science, University of Abdou Moumouni, BP 10662 Niamey – Niger

Abstract

An innovative model is built in document [1] that leads to the advent of a New Generation of Digital Open Universities (DOUNG). The model uses hybrid backbone that initially associates the Internet and GSM with an extension in [2] involving a Wi-Fi antenna and an ad hoc network deriving from the Ethernet network. This multi-support set serves as a communication channel between teacher and learners. It allows several software programs to operate, including those that guarantee the use of devices severely limited in terms of processing and display. The advantage of these devices is further consolidated if prior treatment is applied to the course materials. The idea developed in this paper is to allow the DOUNG to automatically convert the various course formats it offers into the HTML (Hyper Tex Markup Language) standard operating in offline mode and easier to process and display. A conversion mechanism is proposed here as a contribution.

Keywords: Distance learning, open digital universities, HTML, file format conversion

I. Introduction

The scientific research fields covers in an innovative approach the problem of distance education described in document [1]. In this model, the transmission channel reflects a juxtaposition of various networks with new services including m-learning

[3][4][5] and cloud computing. In addition, the interconnection between wired, telecommunications, Wi-Fi and ad hoc networks poses a real Quality of Service (QoS) problem, partly reflected by the need of content harmonization for the smooth and transparent operation of the service, regardless of the platforms. The DOUNG model leads to the production of information flows in numerous formats. It therefore becomes important to consider the method of processing these flows produced and/or stored in the course warehouse to make smooth this service operation by bringing together the different formats in a single standardized one including that of the web.

Indeed, the challenges posed by the diversity of the DOUNG environments are numerous and most ambitious. It thus became necessary to lead to its success, the basic idea supported through the exploitation of this complex architecture which tends to maximize the coverage of geographic areas with limited material resources. In these conditions of a complex and widely open network architecture, it is important to guarantee optimal harmonization of the learning content provided in correlation with their permanent availability. Steps for this harmonization are proposed in this paper.

II. Reminder on the basic architecture of the DOUNG

The backbone of the DOUNG is a set composed of several n-tier architectures. It was first developed by combining the Internet and GSM in document [1]. Considering the possible exploitation of learners devices, an extension was designed in [2] involving a Wi-Fi antenna and an ad hoc network as an extension of the Ethernet network [6]. The global network of the DOUNG is therefore a multi-support system serving as a communication channel between teacher and learners by allowing several categories of software to operate. In the classification of the implementation tools already carried out, the DOUNG deploys a set of servers (Web, multimedia, FTP, messaging) categorized into "production", "provision" of teaching and "learner-teacher exchange". It is the teacher's initiative to make the course content available in specific media via the deployed servers. To complete the description of this client/server operating model, learners use "client" software also classified into "recovery", "visualization" of lessons and "provision" of assessments. The exploitation phase authorizes the learners to access the DOUNG services using client software (Ftp, Web, Messaging, Multimedia) via the interconnected networks. As such, a main question consists in exploiting the architecture to seek for the harmonization of course content in order to ensure a good level of QoS.

III. Application tools and support protocols of the DOUNG

The figure below describes the overall architecture of the DOUNG network with the required software environments, before its description that follows.

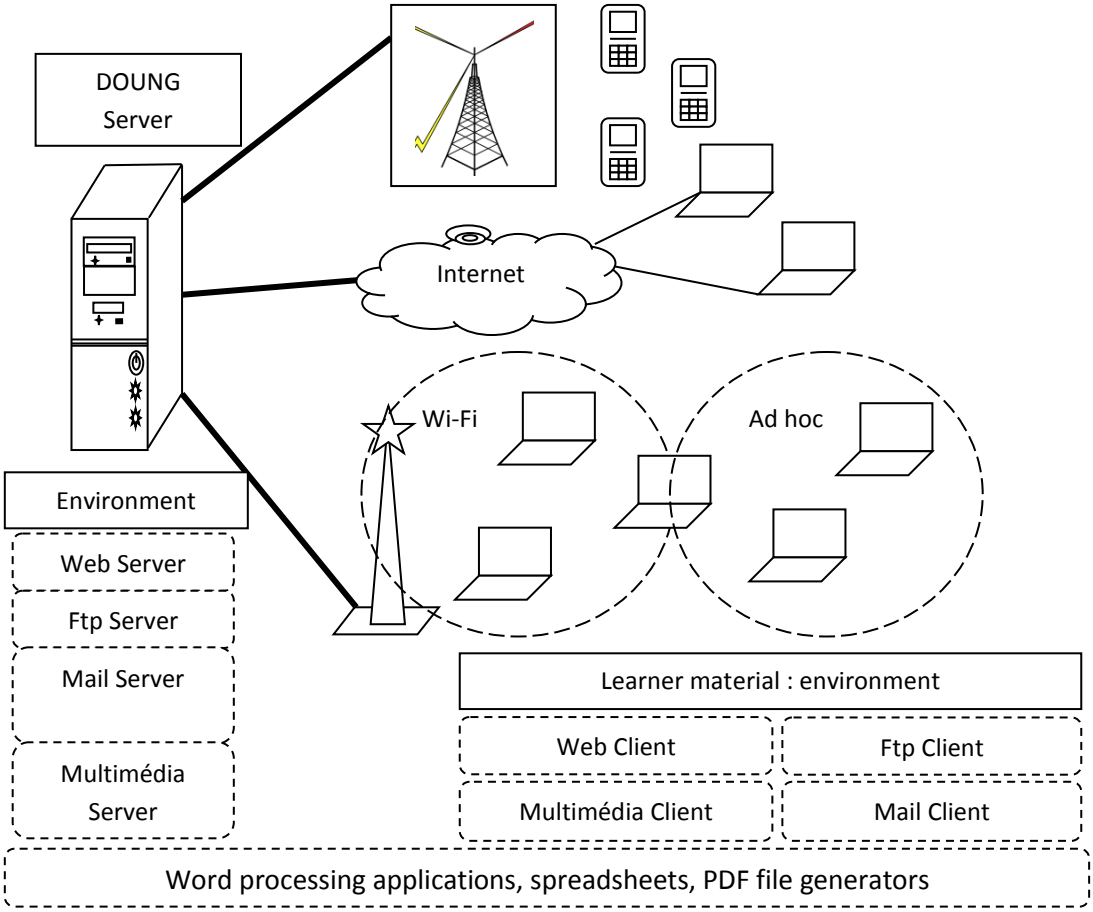

Figure 1: DOUNG software environment

To complete the previous, category (1) of the teaching dispensing software groups multimedia applications including that of the capture and recording of the audio/video stream. This category extends to word processing applications, spreadsheets, PDF file generators, etc. Category (2) consists of an instant transmission media server and an FTP server. Category (3) groups applications for retrieving lessons. It is implemented with instant viewing multimedia clients and FTP clients. The applications classified in category (4) group together multimedia clients for deferred visualization, word processors, spreadsheets and PDF file viewers. The fifth (5) category is that of learner-teacher exchange with instant information transmission and extend to Internet telephony with e-mail service. The web server is adapted to this context for its multi-functionality integrating the course provide in the page service, the access to the multimedia content, the FTP service, the archives research …

Given the interactions established between the learners and the DOUNG materials through open networks including the Internet, GSM, Wireless and ad hoc extension, the problem of securing these exchanges becomes important. Its resolution is further

facilitated using security devices and the software tools. This is the example of the HTTPS (HyperText Transport Protocol Secure) and FTPS (File Transfer Protocol Secure) [7] protocols of the Web environment based on the SSL (Secure Sockets Layer) protocol [8].

IV. Summary of the content formats produced by the DOUNG

The synthesis of the various lessons produced by the DOUNG leads to multiple formats which include multimedia streams with immediate transmission, audio/video files, processed/unprocessed text or pdf files, files of proprietary formats like those of the office suite. We ultimately consider these formats as inputs that enrich the course warehouse from which our converter can generate for each identified case, a standard format file.

V. Problem and justification of the converter development

The wide range of the DOUNG implementation tools brings to the need of developing the multi-format converter at the level of digital support for lessons content. Since these lessons can be provided in multiple forms, the nature of this input data is considered as a constraint linked to the crucial problem of providing adequate quality of service. For example, the cell phones or other low capacity devices pose a traffic bottleneck problem. Indeed, the speed of transmissions is necessarily slowed down by living the backbone and crossing the extended networks. The backbone ensures faster service by various wired channels. The other networks using wireless and telecommunication means including sometimes much more radio waves are having slow and diminished capacities. As a result, bandwidth consumption is further increased with a concentration of traffic on the gateways likely to be constantly in an emergency. Hence, the need imposes to find an appropriate solution that considers the convergence between the multiplicities of course materials formats and the weakness of certain devices environment posing the problem of resources optimization. One of the axes of the contribution expected in this paper therefore lies in the conversion of all the input formats constituted by the course materials to the HTML standard on output. This HTML format is light enough to be easily transportable in networks in offline mode and constitutes a standard adapted to the low display capacity of certain devices. The impossibility of repatriating certain parts of a course such as audio or video can be remedied by replacing them with text provided in HTML tags for example. A simple link can help to find the transcription of an audio or video file in text format easier for the learner to transfer, display and read. Ultimately, our converter is intended to be produced in a cross-platform executable version which accompanies the DOUNG web server. To achieve this, we offer in this paper a description of the overall process and use a decomposition approach.

VI. Process and basic principle of the multi-format converter

VI.1. Input formats

The DOUNG offers distance education services in a variety of formats. Through the development of the multi-format converter, we propose to optimize the process that takes as input the course files produced in the following types and extensions. The possibility remains in the pre-converter directives of our conversion engine to include all recent versions which produce new extensions. These formats are:

Text with the extension ".txt"

Web page with the extensions ".html", ".htm" or ".mhtml"

Ms-Word with the extensions ".doc" or ".docx",

Ms-Excel with the extensions ".xls" or ".xlsx",

Poscript with the extension ".pdf",

Images with the extensions ".jpg", ".jpeg", ".png", ".gif", or ".bmp",

Sound with the extension ".wav",

Video with the extensions ".avi", ".mov" or ".mpeg".

The converter engine operates by taking the various contents and converts them into HTML. The main difficulty lies specifically in accessing this content since most of the formats are sometimes practically inaccessible (PDF for example). Many other formats contain directives that frame the data and only their producing or reading software can execute them and display the content appropriately in a user-understandable manner. An image or Ms-Word file open with a text editor displays only a series of inconsistent characters nested with "skulls". Therefore, the problem highlighted in this paper leads to find a method that allow to build a final web page output with this multiplicity of input formats. If not, it will be necessary to install multitude course viewing software on learner devices. However, this option comes up against the constraining speed in terms of bandwidth and costs for some environments (GSM, ad hoc, Wireless), as well as the storage, processing and display capacity of equipment (cell phone) strongly limited and constituting an additional barrier.

VI.2. The modular approach

Automating any process by using the computer poses a problem with multiple degrees of complexity. Therefore the decomposition approach implemented through the technique of modular programming allows to resolve the constraints posed by problems of various natures and the present case of our multi-format converter is no exception. By using this technique, the objective is to facilitate the source code production of the converter engine. In the philosophy of this modular approach, the decomposition into several simplified sub-problems leads to their separate resolution and their assembly allow to solve the initial problem.

VI.3. Organization of input data

We use a warehouse implemented by a global directory divided in sub-directories for each course. Each sub-directory contains a list of files whose names are by convention chronological numbers starting from "1" regardless of their format. These numbers give the order in which the files appear in the course. For example, a first Word document can start a course followed by Excel content, the third being a jpeg image, the fourth a text file and finally the fifth an existing web page. Therefore, the content of this warehouse is as follows:

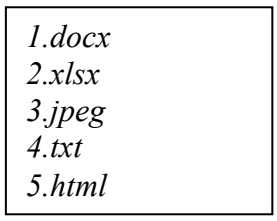

Figure 2: Example of course files deposited in the warehouse

VI.4. Description of the conversion process

Starting the conversion process requires reading the content of the course directory (CD) as input with files having names represented by chronological numbers and extensions giving their format. These chronological numbers indicate their appearance order in the course. The content of the CD is retrieved for the converter engine by using the operating system primitives. These commands are classified in the directives of our pre-converter and available according to the compilation environments. For example in the interactive interface, the redirection option allows to recover the list of a CD files in an output file as result. The access to the file system structures also allows this extraction to be performed, like many other utilities or advanced languages. The converter engine therefore takes the result file (RF) as input and processes each line according to the file extension. The following rules are applied for producing the Global Web Page called PWG. Each rule is a full-fledged module to be developed by applying the decomposition approach:

Rule 1: for the extension ".txt", the file is read and browsed; its words and paragraphs are stored considering the newlines and the end of line character. Each paragraph is inserted directly into the PWG page using either the HTML text formatting tags (P, BR, Font, etc.),

Rule 2: for images, we consider the extensions: ".jpg", ".jpeg", ".png", ".gif", ".bmp". They are inserted in PWG with appropriate tag, example <IMG SRC = "FileName.extension">. A descriptive text or a hypertext link to this text is offered in case the image cannot be acceded,

Rule 3: For sound, the extension ".wav" is considered and the extensions ".avi", ".mpeg", ".mov" are for video. They are inserted in the PWG page using appropriate

markers including for example EMBED: <EMBED SRC = "FileName.extension">. Here also, a transcription text or a hypertext link to this text is offered in case the sound or video cannot be downloaded,

Rule 4: for already existing web pages with previous extensions, the hypertext link marker is used with the example of the inkpad "A": <A HREF = "FileName.html">. This solution is preferable to browsing the file and inserting all the text of the HTML file in the new PWG page between the <BODY> and </BODY> tags.

Rule 5: for office and pdf files, they are converted to HTML by using available Java APIs. Their HTML page is inserted in the new PWG page as previously by using ink tags instead of copying their content. In this context, the use of Java APIs is preferable to the direct HTML conversion mechanism provided by this software because of the generation of source code heavily burdened by the definitions of variables, data and the using of directives of the Javascript language. We classify this conversion process in the guidelines of our pre-converter before proposing a processing method at the following point.

VI.5. Problems with converting Office suite files

As previously stated, at least two possible methods are available for the conversion to HTML of a course in the office suite format, before its integration in the PWG page. The first conversion method consists in using the menu option offered by the software which produces the corresponding HTML file. The disadvantage of this method is in the generated web page containing many unnecessary lines for some weak learner devices. However, the set of HTML tags in the latest versions is wide enough to allow formatting the text and at the same time ensures the clarity of the content of a teaching. Therefore, the second method will be used to obtain a web page using exclusively HTML language tags without Javascript code. This method has the advantage of decomposing a file from the office suite into exploded parts according to their format: text, images, graphics, tables and other objects contained in the file.

If the first conversion method is essential in the document of the Office suite formats before 2007, the new formats facilitate the use of the second conversion method. Before 2007, the office suite document formats were private (binary) and accessible only by specific applications. This obstacle was removed with the new OpenXML format adopted giving "Office OpenXML" documents as compressed files in Zip format instead of text format. With this new format, the content of documents can therefore be viewed and understood. It constitutes a radical break with proprietary binary formats. The evolution of this option is reinforced by the adoption of the open language XML as a global storage format to be used. As a result, it is now possible for programmers to read, create, modify, display on different media, Office documents, without depending on proprietary applications, by simply using tools like XSLT, SAX or DOM, directly or through OpenXML libraries.

VI.6. Internal structure of an OpenXML file

Office OpenXML documents are compressed files in Zip format. Their content is visualized by decompression and the using of any utility recognizing this format. OpenXML assigns to each part or dismemberment, a unique name composed of the logical path from the root of the package to the file constituting the part itself. Thus, from our example, we can draw this (partial) list of names representing the global file:

\[Content_Types].xml

_reals\.reals

\word\Chapitre I.xml

\word\stylesdoc.xml

\word_rels\ Chapitre 2.xml.reals

\docCours\core1.xml

\word\ Chapitre I\image1.jpeg

Based on this organization of OpenXML formats, the Apache-POI (Poor Obfascation Implementation) and Apache-Tika projects made available new libraries to add to the basic libraries already existing in the Java environment. These libraries allow programmers to handle all types of Office suite documents with the benefit of this manipulation on all operating systems. With these new skills, our converter processing consists in browsing the XML format of the office suite file, identifying the parts that appear in chronological order and applying to each part one of the previous five rules of conversion.

VI.7. The multi-format converter: processing module of the Office suite

The handling and integration of the office suite in the new PWG page is ensured by the converter engine. Their intermediate processing involves the use of Java APIs. When opening a document in reading mode, it is possible to extract its content and translate it into HTML format by using the additional libraries of the Apache-Tika APIs and Apache-POI due to the impossibility of obtaining this result with the basic Java libraries. The process registered in the directives of our pre-converter consists in automating the download of these new libraries, in "dezziping" and adding them to those of Java, even using batch files.

The module to be developed in this part by applying the decomposition approach takes as input a document from the office suite and results in an HTML document purified of unnecessary directives and declarations. This refined document is subsequently inserted into the PWG page by means of a hypertext link created with the HTML inker according to the rule 4. For example, if the office document has the name "11.docx", this module of our converter replaces it with the produced document

"11.html". It then inserts this new document into the PWG page following the chronology of the course object files.

VII. Summary of the conversion approach used

The architectural nature of the DOUNG transmission channel reflects a juxtaposition of networks of various natures to be exploited. To this factor, we must add the interconnection between conventional computer networks, wireless, ad hoc and telecommunications networks which poses a real problem of QoS. The treatment of this problem was approached in this paper with the need to harmonize the course content for the smooth and transparent operation of the DOUNG service regardless of the learner platforms. Thus, from the course warehouse, the proposed processing consists in considering and isolating the formats of the different parts of a given course. Then, specific software to be used for each is determined. The chronology of these parts is called upon to respect the model chosen at the input of our converter. Numbers are assigned in place of the alphabetic character strings to compose the names of the files to be assembled. This is the definition of the formalism at the entry.

The production of the PWG page of course assembly is obtained as a result which facilitates its monitoring and its routing in all the DOUNG environments (networks and learner devices). The proposed mechanism is based on the application of the 5 (five) basic rules previously stated. The complexity noted in this approach lies in the processing of Office suite files with a clarification linked to the use of new java libraries and the advanced exploitation of the possibilities offered by the XML format. A perspective is define through the need to set up a descriptive text for the images and above all, to develop a sound and video transcription tool for giving a special corresponding text. This perspective is more interesting as it opens the way of processing audio or video sequences to detect actors and dialogues, to isolate the accompanying music, transcribe the striking movements and sounds (squeaking tires, explosion, gunshot, etc.). If in document [9] it was a question of developing an automatic text reading tool ATRS starting from a text and offering the corresponding audio, this paper set the challenge to carry out the opposite operations and to add the processing of the video by proposing algorithms using the methods and tools available.

VIII. Conclusion

This paper is part of the contribution intended to implement the concepts developed in document [1] which defines the bases of the advent of a DOUNG. From the evolution of this new model, we situate the problem addressed for the development of the multi-format converter with the standardization of the format of the course supports. It is specifically to use the HTML standard because of the possibility offered to learners to have multiple devices from equally diverse network architecture. These devices use multimedia applications (real time and deferred time), as well as applications for transferring and viewing files according to their format, some of which have the

crucial problem of a limited processing, storage and display capacity. We therefore proposed an approach based on the implementation of 5 (five) rules which leads to many challenges to be met by scientific research in this area in order to allow the DOUNG to cover large planetary areas, starting from of a fixed location and overcoming the barriers erected by geographic distances and languages around the world.

References

[1] M. Issoufou Tiado, H. Saliah-Hassane, "Cloud-Computing based architecture for the advent of a New Generation of Digital Open Universities in m-learning", ICEER13 Proceedings, pp 572-579, July 2013, www.labader.org

[2] I. G. Noura, M. I. Tiado, H. G. Souleymane, C. I. Hussein, H. M. Mahamadou, "Quality of Service Evaluation by Simulation in the classroom ad hoc network for the New Generation of Digital Open Universities (DOUNG)", IJNCE-International Journal of Networking & Computer Engineering, July 2020

[3] N. Dolgova1, J. Larionova1, A. Shirokolobova1, "Engineering Students English Teaching in ELearning Environment", MATEC Web of Conferences 297, https://doi.org/10.1051/matecconf/201929706007, ISPCIME-2019

[4] S. K. Basak, M. Wotto, P. Be´langer, "E-learn ing, M-learning and D-learning: Conceptual definition and comparative analysis", E-Learning and Digital Media 2018, Vol. 15(4) 191–216, Universite du Quebec à Montréal (UQAM), Canada

[5] M.R.M. VeeraManickam, .M. Mohanapriya, "Research Study on Centralized E-Learning Architecture Model for Educational Institutes in INDIA: Teaching & Learning Process", MATEC Web of Conferences 2016, DOI: 10.1051/57

[6] F.Chughtai, R. UlAmin, A. S. Malik, N. Saeed, "Performance Analysis of Microsoft Network Policy Server and FreeRADIUS Authentication Systems in 802.1x based Secured Wired Ethernet using PEAP", The International Arab Journal of Information Technology, Vol. 16, No. 5, September 2019

[7] P. Gulia, Reena, "A Novel Technique of Security Improvement in Ad-hoc Network by using FTP", International Journal of Applied Engineering Research ISSN 0973 -4562 Volume 1 2, Number 1 7 (2017) pp. 6658-6662, http://www.ripublication.com

[8] Freier, P. Karlton, P. Kocher, "The Secure Sockets Layer (SSL) Protocol version 3.0", RFC 6101, August 2011

[9] M. I. Tiado, A. Idrissa, D. Karimou, "Improved Text Reading System for Digital Open Universities", IJARAI - International Journal of Advanced Research in Artificial Intelligence, October 2015, Vol 4, No 10, pp 29-34