DIGITAL REPOSITORY FOR PERIODICAL COLLECTION. "MONITORUL" COLLECTION CASE STUDY

Daniela Druguş[*1], Marius Stoianovici[2], Bogdan Alexandrescu[3], Angela Repanovici[4]

1 UMF Gr.T Popa, Iaşi, corresponding author, drugus_daniela@yahoo.com

[2]Universitatea Transilvania din Braşov, Medical Engineering Department

3 Social assistance and cultural centre, Brasov, Romania

[4] Universitatea Transilvania din Braşov, Medical Engineering Department

ABSTRACT:

Europe's cultural heritage, which includes printed materials (books, newspapers, magazines), photos, museum artefacts, archived documents, audio and audio-video recordings, monuments, and archaeological sites (further on referred as „cultural content"), represents one of the key areas of the digital agenda.

Cultural content online access will allow European users wide access and use for recreational, learning or professional purposes.

Periodicals Collection Digital Repository Web Application Architecture is presented. The developed application namely the Monitorul Express digital archive/repository, offering word search option, incorporating the latest script software, allowing data preservation, dynamic and efficient server accessing, and special libraries facilitating special effects was build and can be used for persons with NEEDS. There are studied the application efficiency from the usage resources point of view and cost support effectiveness. This represents two key components to a web project development.

KEYWORDS: software application, documents management, newspaper industry, digital repository, scanning, digital

INTRODUCTION

The web application for creating the digital archive of periodicals collection presents the application, namely the digital archive of Monitorul Expres newspaper, an archive offering word search options, created with the latest script software, allowing data preservation, dynamic and economic server accessing, and special libraries which allow special effects. In the same time, the efficiency of the application in terms of used resources and of the necessary supporting costs are two key components in development of a web project.

STRUCTURE OF THE APPLICATION

The application was created using PHP programming language and MySQL database system. We also used Yii framework. Yii (or Yes IT Is) is a modern PHP framework, known for its high performances in processing the requests sent to server, therefore a critical tool for all web programmers. [1]

Yii is based on **MVC** architecture (Model - View - Controller) [Fig.1], frequently met in several informatics projects, which proved effective in time. [3] MVC consists in dividing the

---

[*] Corresponding author. Tel.:
E-mail address: drugus_daniela@yahoo.com

structure of the classes and proper code, together with dividing the logic of the application in these three components:

- model – represents the entities of the application and processes the information in the database;
- view – types the data and the interface;
- controller – controls the communication between view, user and model, that is the database.
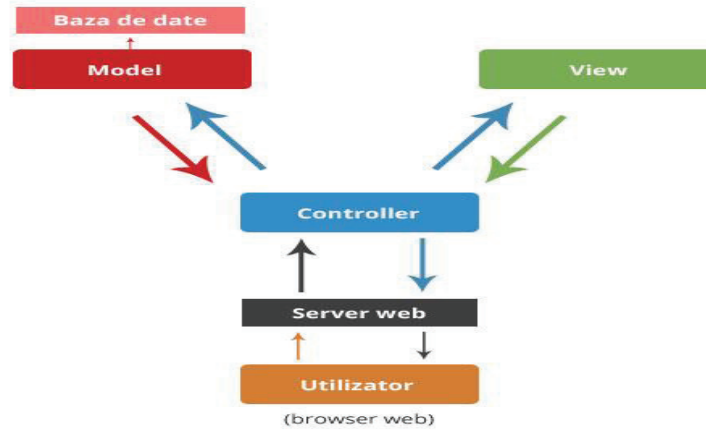

Figure 1 *MVC architecture of a web application*

APPLICATION INTERFACE

The application can be accessed through an internet browser. The browser allows the access to the digital archive based on authentication. There are two types of users:

- plain user – he/she can view the archive and can use filters or searches;
- admin – he/she can also manage the database, adding new documents to the archive.

UML diagram in figure 2 presents the action which can be done by the users of the system. [4]
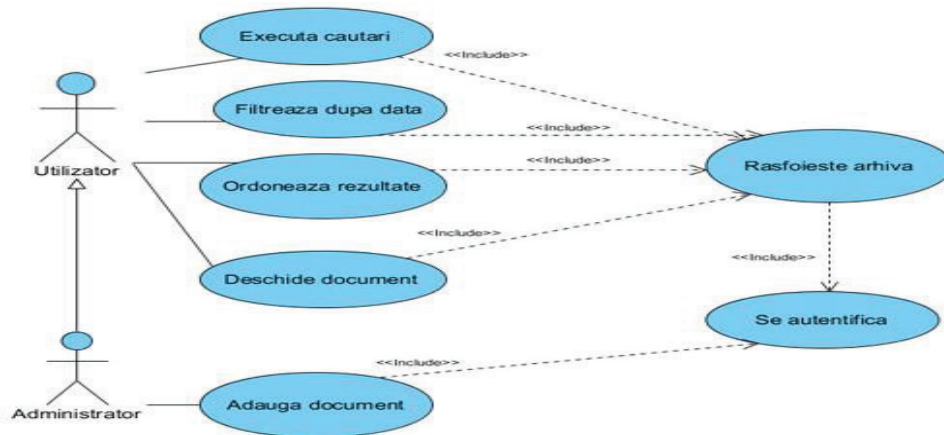

Figure 2 *UML Diagram*

Utilizator = user
Administrator = admin
Executa cautari = performs searches

Filtreaza dupa data = filters by date
Ordoneza rezultatele = sorts the results
Deschide document = opens file
Adauga document = adds document
Rasfoieste arhiva = browses archive
Se autentifica = authenticates
Include = includes

The user of the application can therefore:
- enter words or keywords in order to filter the results;
- filter by certain publishing date for periodicals collection;
- sort results by publishing date, ascending or descending;
- access the digital version of each edition.

All these actions are allowed only after the user's authentication in the system. The recognition of the type of user is automatic, while the interface is adjusted to the user's right of execution. The authentication screen is presented in figure 3, where entering a valid user and password is needed. [5]

Once the authentication process is successfully executed, it is valid during a PHP session, more exactly during our interaction with the application. The user is still greeted with the main interface, in which he/she has access to all the above mentioned actions, and may view the complete list of digital achieve in the database. One may notice that the results are paginated, grouped in ten, avoiding this way the overloading of the web page and the blocking of the user's system.
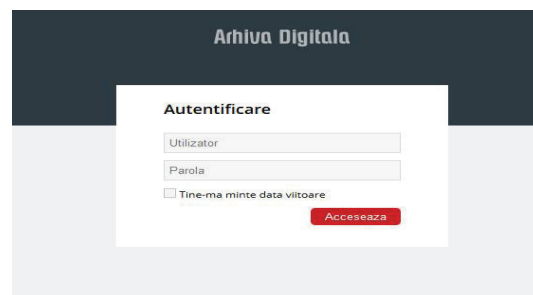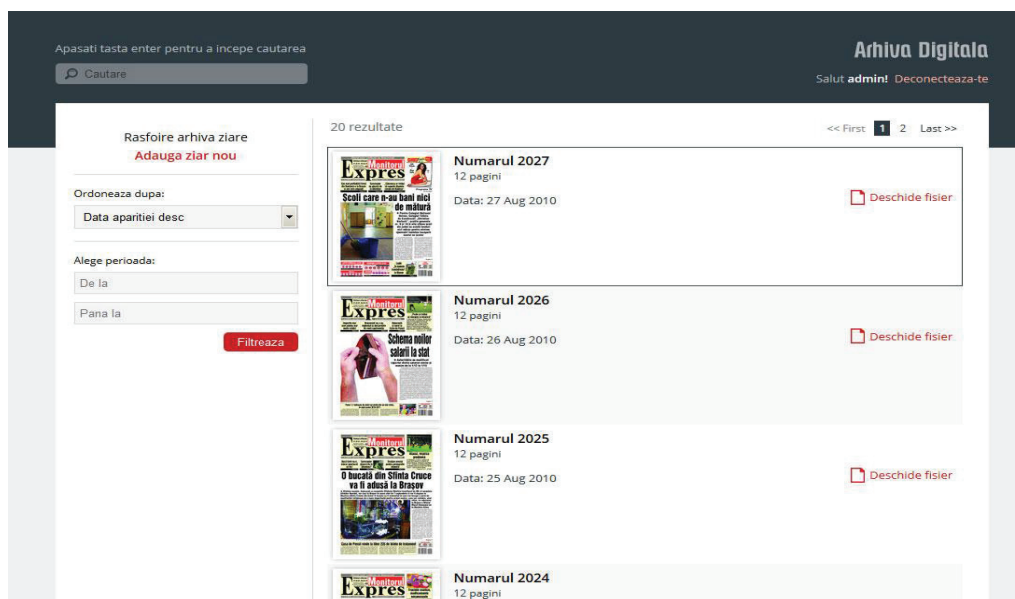

Figure 3 *Authentication screen*

In the upper right, the application presents the confirmation of the fact that it recognized and

authenticated the user and it also offers the option of disconnecting from the system. It will lead to closing the current session and redirecting the internet browser to the authentication page. To access the digital documents, the users will click the "Deschide fişier" button near each edition, or click on their image, represented by a picture of the first page. The design of the interface is presented in figure 4.[7]

Figure 4 *Application interface*

In order to perform a search, the words to be searched are entered in the upper left box and the ENTER key is pressed. This way a request will be sent to server, which will use FULL-SEARCH index to filter the column content in the database and will return the corresponding data. The total number of results will be displayed in the interface, in the right upper side of the screen.

The design related to the selecting way of the starting period to be browsed, using Yii facilities is presented in figure 5. The starting period is chosen as 10.06.2009. One may notice that a calendar is displayed, and the user can select the starting year and date of the browsing period. [6]
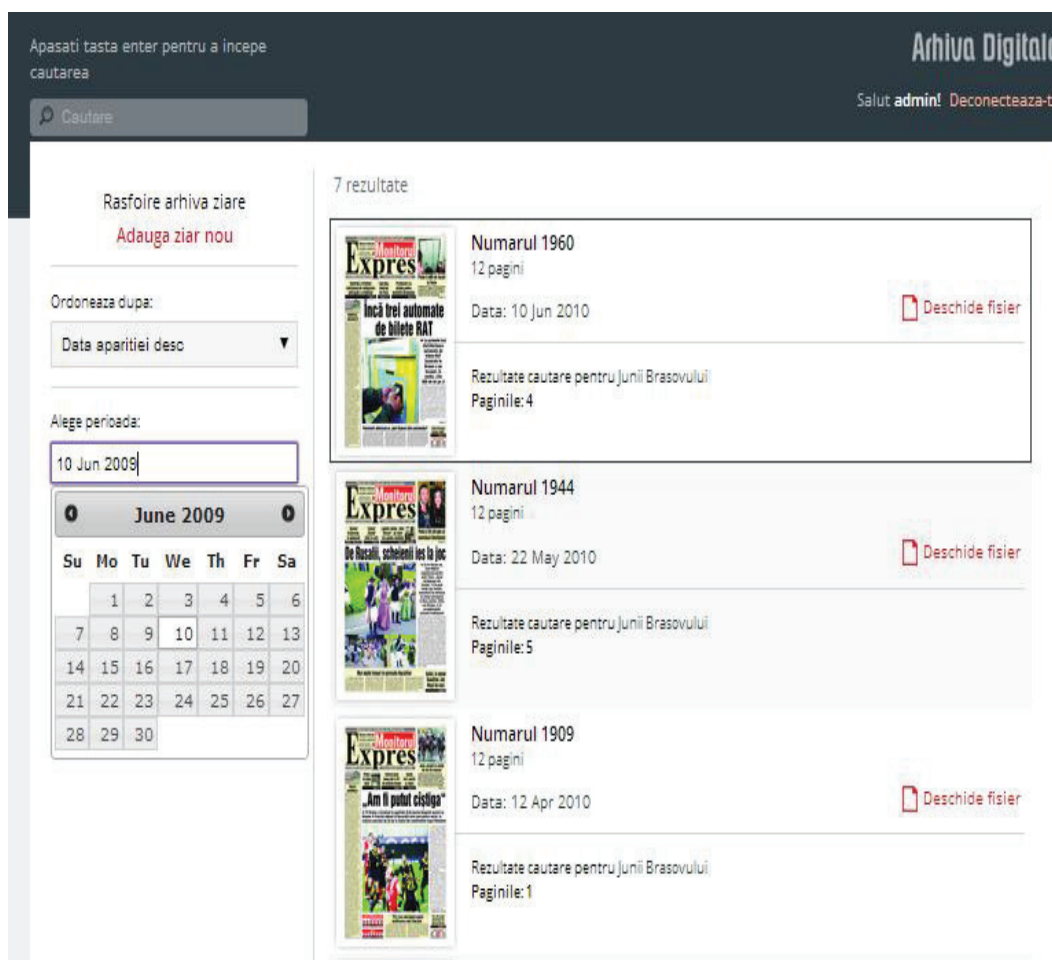

Figure 5 *Selecting the starting period of newspaper browsing*

In figure 6 one can see a search/browse performed only by a certain period, chosen by the user, therefore I selected *10.06.2009 – 11.06.2009* period and 2 newspapers have been found (2 results).
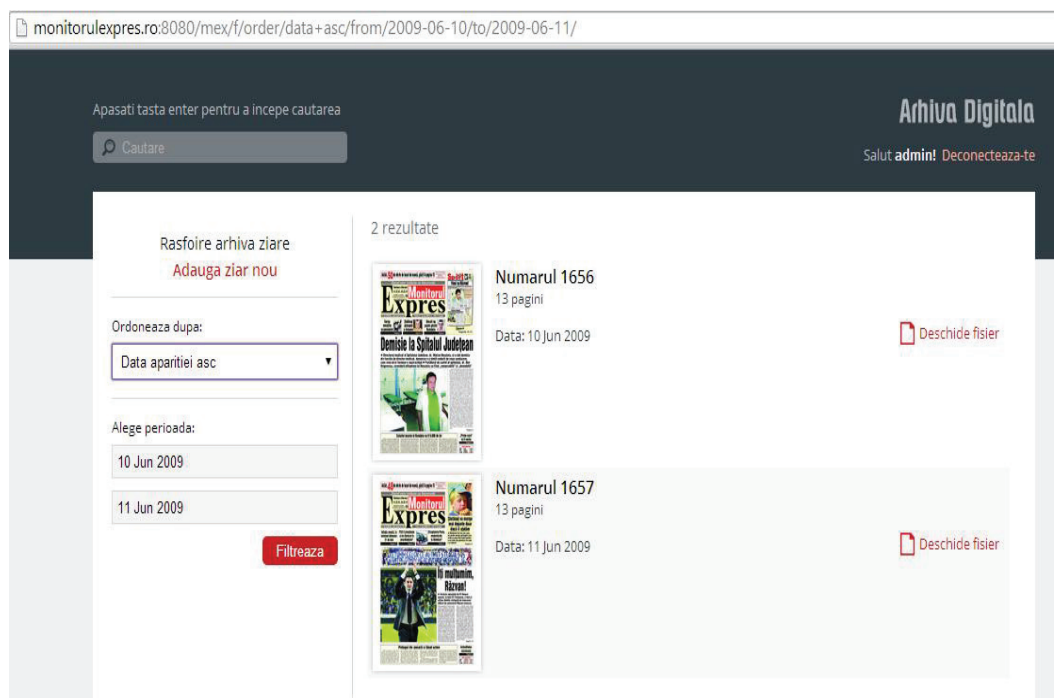

Figure 6 *Selecting the browsing period for the newspaper*

CONCLUSIONS

This software application, a platform independent from the operating system, used the latest technologies in the IT market. To create it, MVC (Model-View-Controller) architecture was used. All these technologies, together with MVC architecture, were presented in chapter 5 of the thesis. The created platform, elegant and flexible, allows the access of the digital archive to a wide audience. Database populating was done through PDF parsed data. These data were stored in an MySQL database, using a PHP library named **pdfparser**. The created database is structured in three tables, as follows: a table, named *utilizator,* containing information related to the user; a tale, named *ziar*, containing information related to the analysed newspaper; and a table, named *pagina*, linked to the table *ziar,* containing information from a certain newspaper page. In order to get a more precise search, the information stored on the individual page can be used.[9]

The structure of the created database has a field named "*conţinut*" in table *pagina*, full-text type, in which the entire contents of each reference document is available to be viewed online, printed, or downloaded. In addition to text documents, one can include images like charts, maps, photos and diagrams. A database with a full-text field allows searches by words, keywords or both.

In the web application of the project, the full-text index is used to offer users the possibility to do a rapid search for a word or keyword in one of the documents in the digital archive. The application will return the list of documents corresponding to the criterion that was inserted. It implies saving the result of a search in a binary format in cache files, and using this format, already processed for the next searches, by the the same words. The users that are about

to perform searches by the same word will get the result in cache files, the database being this way protected against a density of inquiries that might hamper its performances. [8]

Accessing the digital archive can be done through an internet browser. It can be performed based on authentication. We must emphasize the fact that this platform has two actors, namely: the plain user and the admin of the platform. Each of these actors has the right to execute certain actions to the database. The user has the right to view the archive and can execute filtering or searching actions, while the admin can administer the database adding new documents to the archive. [11]

Thus, the user can:
- enter words or keywords by which to filter the results;
- filter by a certain publishing date for periodicals collection;
- sort the results by publishing date, ascending or descending;
- access the digital document of each edition.

Display of the results of the inquiry performed by the user is clear, concise, with useful information obtained by browsing the archive. [10]

REFERENCES

[1] https://www.invata-online.ro/html-css/html-partea-2/scurt-istoric-htmlxhtml-css
[2] [accessed 24.05.2014]
[3] http://www.rasfoiesc.com/educatie/informatica/html/PROIECT-INFORMATICA-WEB-CAPITA36.php
[4] http://ro.wikipedia.org/wiki/Apache_(server)
[5] http://marianpc.eu/2011/02/26/html-5-si-css3-designul-la-un-nou-nivel/
[6] [accessed 17.05.2014]
[7] http://dpw.bistrita.ro/wp-content/uploads/jQuery_ro.pdf [accessed 14.05.2014]
[8] http://www.php.net/manual/en/intro-whatcando.php [accessed 16.05.2014]
[9] http://www2.unitbv.ro/LinkClick.aspx?fileticket=8Gsfz0SfMt0%3D&tabid=4146&language=en-US [accesed10.06.2014]
[10] http://www.icpsr.umich.edu/DP/policies/dpp-framework.html. [accessed 14.05.2014]
[11] http://www.bl.uk/aboutus/stratpolprog/ccare/introduction/digital/digpresmicro.pdf. [accessed 14.05.2014]