

AUTOMATIZATION OF COLLECTION AND USE OF THE INFORMATION FROM OFFICIAL WEBSITES OF EDUCATIONAL ORGANIZATIONS

Mikhail Petropavlovskiy¹ and Olga Nefedova²

¹Doctor of Engineering Science, Mari State University, Russia, mv2006-68@mail.ru

²Candidate of Engineering Science, Mari State University, Russia, olga-5@inbox.ru

Abstract

The article considers the approaches to processing automatization of the information published on official websites of educational organizations in the Internet. It analyzes the opportunities of using this information during external review as a supportive means for making a decision on the information credibility and evaluation of the accreditation indicators.

Keywords: information transparency, html-tag, data collection.

1. INTRODUCTION

The information transparency of educational organizations in the Russian Federation as well as specific regulatory requirements to the information transparency are stipulated by the Federal Law "On Education in the Russian Federation". The procedure for publication and update of the information on official websites of educational organizations is determined by the Decree of the Government of the Russian Federation of 10.07.2013 № 582. In the context of a large higher education system in Russia (over 900 higher education institutions with more than 1000 branches) the issue of automatic evaluation of each educational organization's transparency is especially pressing.

2. MATERIALS AND STUDY METHODS

The order of the Federal Service for Supervision in Education and Science (Rosobrnadzor) of 29.05.2014 № 785 establishes the structure of a special section "Information about the educational organization" which educational organizations are required to display and update on their websites, as well as technical specifications for the published information (navigation mechanism, valid file formats, etc.).

The section "Information about the educational organization" contains information grouped by thematic subsections:

- 1) key information;
- 2) structure and administration of the educational organization;
- 3) documents;
- 4) education;
- 5) educational standards;
- 6) administration, teaching (academic) staff;
- 7) logistics and equipment of the educational process;
- 8) scholarships and other types of financial support;
- 9) paid educational services;
- 10) financial and operational activities;
- 11) vacancies for admission (transfer).

In accordance with the requirements, all pages of the official website, included in the section "Information about the educational organization", contain a special html-tag with microdata allowing for unique identification of the information which is the subject to mandatory publication. The availability of such a microdata makes it possible to automate the procedure of data collection from official websites of HEIs. A detailed description of the special microdata allowing for data collection from the site is contained in the "Methodological recommendations for presentation of the information about the educational organization in open sources with due consideration of the legislation requirements" (Methodological recommendations) which are approved by Rosobrnadzor. Methodological recommendations contain descriptions of links, special microdata and vocabularies.

The regulated html microdata on the pages of the special section allows the web crawler ("Spider") which is algorithmically developed and implemented in software to perform automatic scanning of websites in search of information and further saving the identified structured data from the marked-up fields into the data warehouses (Petropavlovskiy, 2016, p. 127).

The web crawler is a software program that visits the web pages of educational organizations like an ordinary Internet browser. Upon successful receipt of the web page, the crawler saves it in a new set of data and tries to analyze it. As a result, there appears a dataset on an educational organization containing copies of web pages with their URL-addresses and the structured data itself.

The saved pages are analyzed as follows: the program analyzes information recursively in all page tags. If an attribute is found, the tag compares it with a vocabulary containing the attributes described in the Methodological recommendations. In case of a positive result, the data is written in pairs (attribute name - value). The collection and analysis result is visualized. Fig. 1 shows an example of displaying the successfully analyzed data in the form of "attribute-> value".

The process of data collection from all educational organizations' websites takes a bit more than 3 hours. Software use experience has shown that currently 577 out of 977 analyzed HEIs' official websites have the recommended html page format.

The implementation of this approach is based on the use of the software product "Data collection module for a special section of the educational organization website" (Petropavlovskiy, 2016, p. 129) allowing for data input, modification of the information presented in the special section for its update, as well as automatic generation of html-pages of the special section "Information about the educational organization" in compliance with the legislation requirements (Fig. 2).

Программа автоматизированного сбора сведений (01:59:52)

ОСНОВНЫЕ СВЕДЕНИЯ

СОХРАНЕННАЯ СТРАНИЦА

regdate->23.12.1993

<http://obrnadzor.gov.ru/microformats/uchredlaw>

nameuchred-><http://mon.gov.ru/>

fullnameuchred->Ливанов Дмитрий Викторович

addressuchred->125993, г. Москва, ул. Тверская, д. 11.

teluchred->Тел.: +7 (499) 237-22-63, +7 (499) 237-54-64

mailuchred-><mailto:info@mon.gov.ru>

websiteuchred-><http://минобрнауки.рф>

address->656038, Сибирский федеральный округ, Алтайский край, г. Барнаул, проспект Ленина, д. 46

addressfil->659305, Сибирский федеральный округ, Алтайский край, г. Бийск, ул. Трофимова, д. 27.

addressfil->658207, Сибирский федеральный округ, Алтайский край, г. Рубцовск, ул. Тракторная, 2/6.

worktime->Учебный год начинается с 1 сентября (или первый рабочий день сентября). Учебные занятия проводятся по графику шестидневн
расписанием учебных групп. Расписание звонков: 1 пара 8:15 – 9:45 2 пара 9:55 – 11:25 3 пара 11:35 – 13:05 4 пара 13:35 – 15:05 5 пара 15:15 –
– 21:45 Для сотрудников, не связанных с учебным процессом — пятидневная рабочая неделя с 8.00 до 17.00 Для профессорско-преподавател
учебным процессом — шестидневная рабочая неделя

telephone->+7 (3852) 290-706 — справочная

telephone->+7 (3852) 290-710 — приемная ректора

telephone->+7 (3852) 367-864 — факс

e-mail-><mailto:elvisadoo@mail.ru>

e-mail-><mailto:altgtu@list.ru>

address->Адрес: 656038, Российская Федерация, Алтайский край, г. Барнаул, пр. Ленина, 46

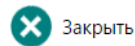


Fig. 1. Analyzed website pages, displaying the pairs “attribute – value” for the subsection “Key information”. The picture presents the content of the saved page of one of the HEI’s official web site in the form of a list of attributes (e.g. “nameuchred”, “fullnameuchred”, “addressuchred”, and etc.) and values for each attribute (after symbols “->”)

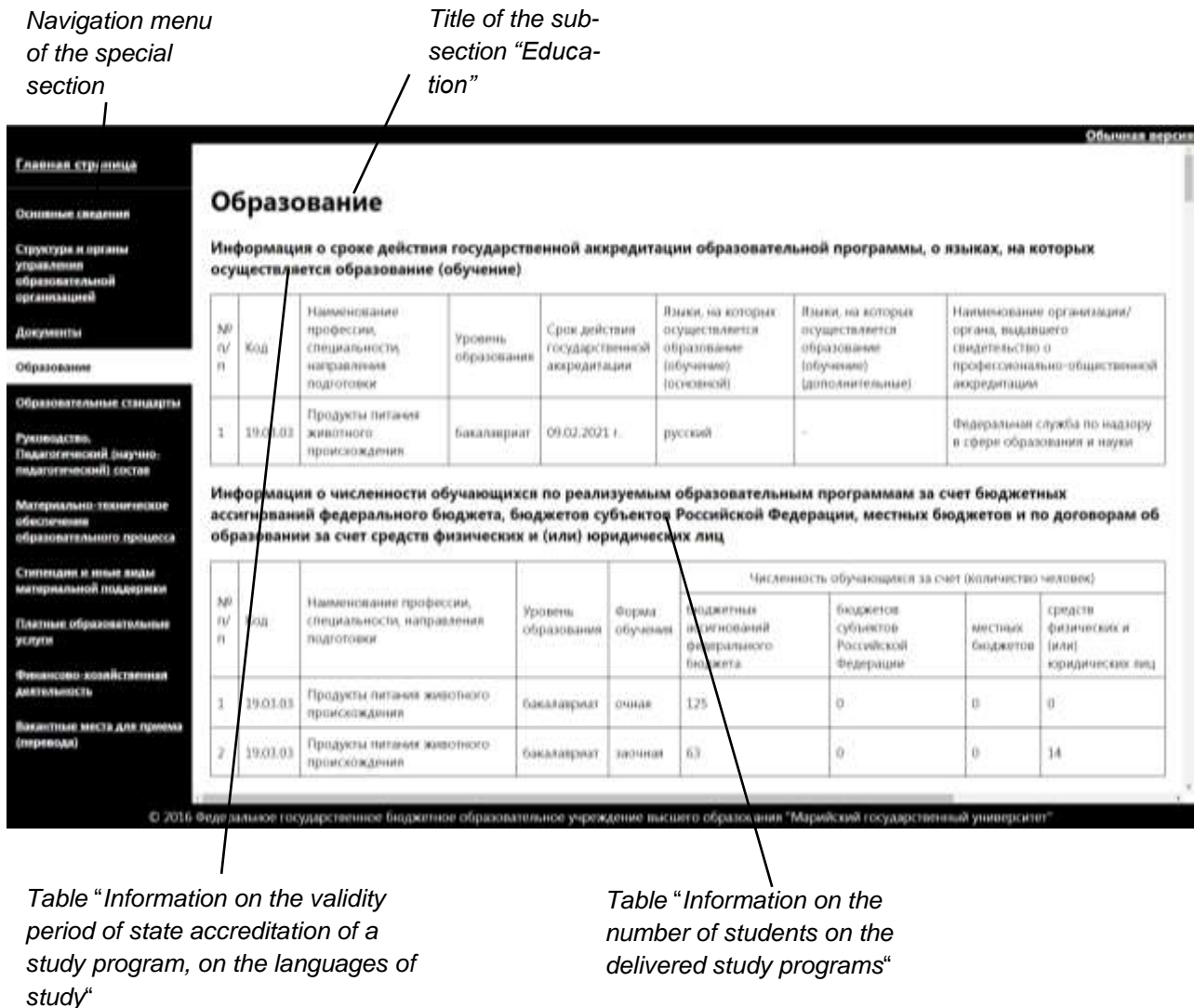


Fig. 2. Fragment of an html page, generated by "Data collection module for a special section of an educational organization website". Subsection "Education"

3. RESULTS AND DISCUSSION

Thus, the main stages of the automatic collection and use of the information from HEIs' official websites are as follows (Figure 3.):

- 1) data input by HEIs using the software "Data collection module for a special section of an educational organization website";
- 2) automatic generation of html-pages of the special section "Information about the educational organization" and their hosting on the official website in the Internet based on the input data; simultaneously the software enters the special microdata providing the automatic data collection into the generated program html-code;
- 3) automatic data collection from the marked html-pages of the special section on official websites is implemented monthly by the software module "Spider" for the information update. The collected data is integrated into the data warehouse;
- 4) use of the information during the procedure of state accreditation of educational activities: to evaluate the accreditation indicators (it is possible to evaluate over 50% of the indicators in terms of their compliance with the FSSES of HE automatically), and to analyze the credibility of the information submitted for state accreditation. Besides, the data is used to analyze the compliance of the official websites' content with the requirements when carrying out control and supervision activities and various monitoring of the higher education system.

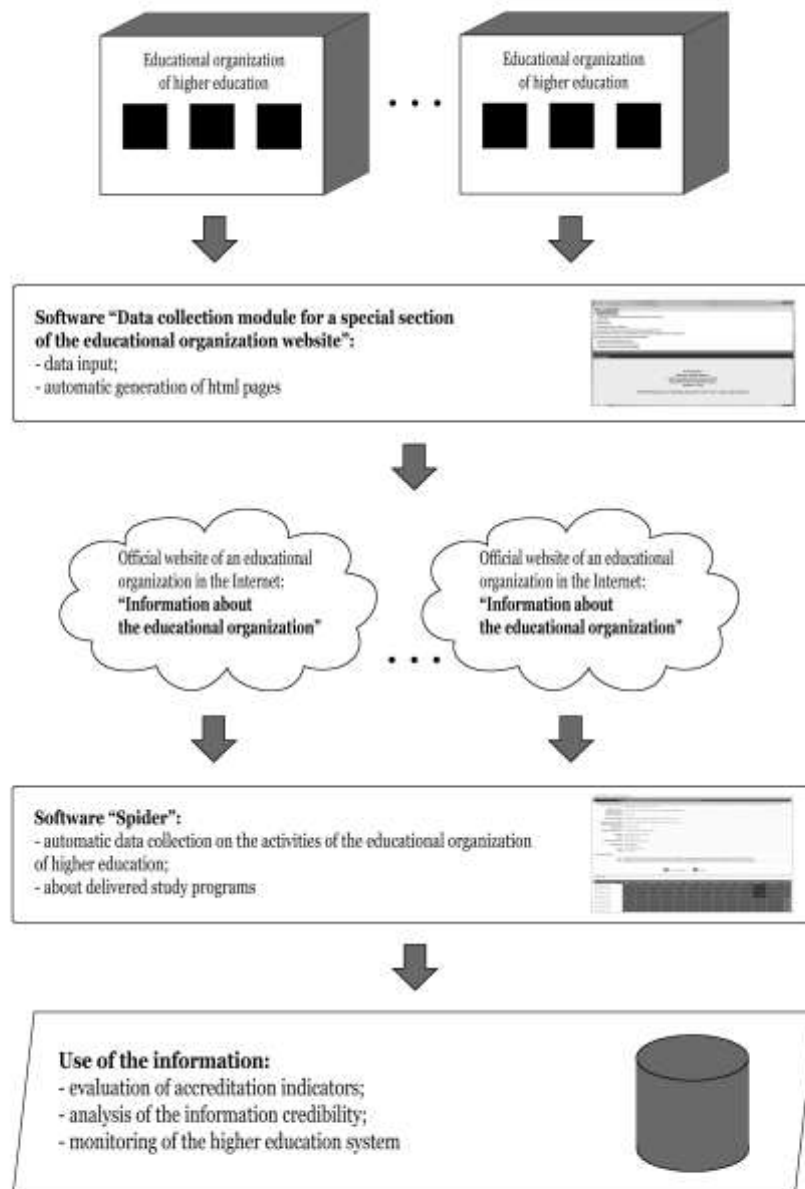


Fig. 3. Stages of the technology for using the information from HEIs' official websites during the procedure of state accreditation of educational activities

4. CONCLUSIONS

The suggested information technology of publishing, collecting and using the information from HEIs' official websites during state accreditation of educational activities has doubtless advantages over the existing methods of publishing and manual data collection, as it provides not only the information transparency of the higher education system, but also allows partial automating of the external review procedure. At the same time, more attention is be paid directly to the education quality evaluation during external review.

The developed technology can be replicated on other levels of education (secondary vocational education and training, general and pre-school education), and it can be offered for use to accreditation authorities of the Russian Federation regions.

5. ACKNOWLEDGEMENT

This research was financially supported by TEMPUS project "Development of Public Accreditation of Agricultural Programs in Russia" (543902-TEMPUS-1-2013-SK-TEMPUS-SMGR).

REFERENCE LIST

- Decree of the Government of the Russian Federation of 10.07.2013 № 582 (edit. of 20.10.2015) "On approval of the Rules of publication and update of the information about the educational organization on its official website in the information and telecommunication network Internet". URL: <http://base.consultant.ru/cons/cgi/online.cgi?req=doc;base=LAW;n=187784> (accessed date: 18.07.2016).
- Order of Rosobrnadzor of 29.05.2014 № 785 (edit. of 02.02.2016) "On approval of the requirements to the structure of the official website of an educational organization displayed in the information and telecommunication network Internet and to the format of the published information". URL: <http://base.consultant.ru/cons/cgi/online.cgi?req=doc;base=LAW;n=194674> (accessed date: 18.07.2016).
- Petropavlovskiy, M., Nefedova, O., Koshkin, A. (2016). Technology of Using the Information From Official Web Sites of HEIS During State Accreditation of Educational Activities. *Informatika i sistemy upravleniya*, 4(50). P. 124-132. URL: http://ics.khstu.ru/media/2016/N50_13_DOI.pdf. (DOI: 10.22250/isu.2016.50.124-132).