

Biljana Stojanović

Faculty of Mathematics, University of Belgrade

DIGITIZING THE ARCHIVE OF STUDENT SERVICES OF THE FACULTY OF MATHEMATICS

Abstract. The project of “Digitizing the Archive of Student Services of the Faculty of Mathematics” was initiated in 2004 with the aim of converting the old archive materials to electronic form with the ability to search and print data. In the middle of 2016 there was no significant progress in the work, and the main reason was the lack of human resources. The plan is to reactivate and complete the project, and connect it with the existing information system of student services, StudInfo. This paper presents a brief overview of the content of the archive and describes the necessary steps for the digitizing of archive documentation, storing and organizing documents in the appropriate database that can be searched according to relevant criteria.

Keywords. Archive, Student Services, Digitizing, Scanning, Searching, StudInfo.

Introduction

The Archive of Student Services of the Faculty of Mathematics preserves documented data in various forms, such as data in paper and electronic form, photos, and more. The Faculty is required to maintain complete documentation for students in student dossiers from the time of enrollment through to graduation. Given that the Faculty of Mathematics in the past enrolled on average four hundred students per year, the archive material is quite extensive. Student dossiers include variety of documents such as personal data, information on grades and courses certificates, test applications, degrees, certificates or awards, diplomas, payment receipts, etc. Within the archive there are also the registers of births, decisions regarding the student appeals and various other documents.

One part of the paper documentation has been damaged due to inadequate storage conditions. In order to prevent further damage, certain precautions should be taken so as to provide adequate facilities for its preservation. Today, this only applies to the one part of the student dossiers. The main reason is the fact that the documentation was recently physically moved to another Faculty’s location, and dossiers have not yet been fully distributed and classified by the appropriate academic year, and, within the same year, alphabetically. All paper documentation should be organized in a similar way (Figure 1).

Archival paper documentation still needs to be formally available for the issuance of various types of certificates or copies of documents on demand. Therefore, its proper storage and maintenance is necessary. But handling of voluminous paper documentation is not simple, and thus can be a problem. Issuing copies of documents on one's request is a process that can take some time. First, it is necessary to search the archives in order to find a particular physical dossier. Without key information such as exact academic year, the search process could be a time consuming task. Then, additional time would be needed to make the required copies. Every time a dossier is handled, there is a risk of the physical material being damaged or misplaced. The main disadvantages of this approach are large storage costs, the complex administrative procedures and inefficiency.

Electronic form of data storage, therefore, appears as a necessary procedure. This includes scanning and digitizing of student dossiers, as well as the registers of births and plans and programs of academic studies. Since the paper documentation is in various degrees of preservation and quality, scanning and digitizing would preserve it (in the present condition)

and enable its usability in the future. The need for automated student dossiers is significant. The student dossiers (for all students regardless of the year of enrollment) should be accessible, and at the same time secured, allowing access to the dossier, while protecting highly confidential student information. Digitizing of physical dossiers improves the organization of documents and data, managing and handling of data and their search. This approach enables effective strategy for saving space, increasing access, reducing retrieval time, or creating copies. The documentation thus becomes available in digital form and access to the requested information is provided, as well as the ability to send data on one's request.



Figure 1: The organized part of the student dossiers from the Archive of Student Services

1. The Archive of Student Services

The percentage of paper documentation increases with time going back into the past. For students enrolled before 1990 a very small part of the documentation exists in electronic form. After 1990, complete documentation was still kept in paper form, while part of the documentation that related to exams (test lists with grades obtained in exams) was kept in electronic form. In the past ten years, the largest part of the documents has been kept in electronic form. Paper form is retained for documents that, amongst others, include the certificates and diplomas from secondary schools and excerpts from birth registries. The Archive of Student Services preserves student dossiers since 1920. Different names of the Faculty can be seen on them: from its founding in 1873 until 1947 it was the Faculty of Philosophy (Department of Mathematics and Natural Science), from 1947 to 1995, the Faculty of Natural Sciences and Mathematics, and since 1995 until today, the Faculty of Mathematics. Depending on the year of enrollment, student dossiers contained documents of different types, shapes and formats (Figure 2). The document with the same purpose had a different look and content in different periods of time. Format and quality of paper were also different. Thus, from the time of the Faculty of Philosophy we can find a variety of documents in the student dossiers: enrollment paper (Figure 3) and the appropriate application form, application forms for the semester admissions (Figure 4), applications for graduate exams (Figure 5) and following graduation exams records, application for taking special exams and the accompanying records, different certificates, testimony on the passing of examinations (matriculation examination as a precondition for admission to the Faculty) (Figure 6), birth certificate, student appeals (hand-written in ink or typewritten), survey sheets, student writing assignments from written examinations (Figure 5), etc. At the time of the Faculty of Natural Sciences and Mathematics, a student card (Figure 7) was in use.

УНИВЕРЗИТЕТ У БЕОГРАДУ

Копила предмета

Делатност: *Физика математичке* факултета
 Припадајуће се за полагање посебног испита из: *Физике*
Математичке физике
 којег сам предмет слушао: _____ семестра 1955/56 школске године
 код постављача: *Др. Р. Вукотића*

Испит ми је: *Положен*
 Уписану у којој има оспрета потребна број семестра из овог предмета поднашим из уписа: _____
 С. Ф. - С. Н. *Физички факултет*
 Београд, 1955 год.

✓

Отсек група: _____ Досије № *2617*
 Семестар: *Математичка*
 Кандидат: *Филиповић Олга*

Предмет: *Увод у математичку*
физичку
 Заведено у деловодном протоколу
 факултета дана _____ месеца
 1955 год.
 Бр. _____

Кандидат је данас полагао посебни
 испит из *увода у математичку*
физичку
 и показао овај успех: *мире година*
 на писменом испиту: _____
 на усменом испиту: _____
 Београд, 20-11 1955 год.
 Испитна комисија: _____

Figure 8: Exam application (1955, Faculty of Natural Sciences and Mathematics)

2.1. Scanning and storing of materials. The scanning of the entire material would first involve appropriate technical preparation of documents prior to scanning, which may include physical correction of documents and technical preparation after scanning (returning to archival state). Special attention should be paid to scan sensitive and/or damaged documents (eg. documents from period 1920-1947). Scanned documents can be saved as a digital image with compression (eg. in TIFF or JPEG format) or as documents in PDF format, which also provides high-quality compressed content presentation.

More advanced option that could be considered is to recognize text in scanned documents or, in other words, converting scanned documents into text format documents with the possibility of editing content. It can be realized by using automated OCR (Optical Character Recognition) processing. Software tool for OCR (such as ABBYY FineReader [2], currently one of the most widely used software of that kind) has the ability of converting scanned paper documents, digital images of texts, and image-only PDF documents into searchable PDF or other document text format. This enables text editing of documents and eliminates the need for retyping. Given that a large percentage of documents contains handwritten text in Cyrillic (usually in cursive), the percentage of recognized text may vary. Although software tools for OCR have support for different languages (ABBYY FineReader supports up to 190 languages), after text recognition, it would be necessary to manually make corrections of errors or to amend the text that is not recognized.

The end result of this phase of processing the scanned material should be text format documents that contain data extracted from the resulting documents of the OCR processing, and which should be generated automatically. These data should include everything relevant information that exists in the documentation and should be subsequently entered into the database.

Before entering data into the database, all scanned documentation from a single dossier should be organized and saved in one folder. In order to enable automatic data entry into the database, it would be necessary to additionally prescribe or standardize the ways of naming scanned documents and any supporting documents in text format, folders where scanned documentation will be preserved and documents with metadata. It would also be important to standardize the format of content of metadata files, ie. to define which basic data can be considered as relevant metadata. After scanning and storing all available data, it is necessary to carry out the verification of the authenticity of the material.

It can be roughly estimated that, depending on the volume of the documentation, the scan should not take longer than 30 minutes per student dossier, ie. that one person can scan an average of 20 dossiers per day (for eight hours of working time). This would mean that the scanning could be completed in approximately five man-years. The overall processing time could be significantly reduced by working in parallel on several scanners or by purchasing better equipment and hiring a larger number of staff. Moreover, this kind of work optimization is necessary, as it would need much more time for processing the scanned material, entering metadata and final verification of stored documents. This procedure can significantly reduce storage costs and improve retrieval efficiency by enabling searching capabilities that are not possible in a paper form.

3. Database implementation

In order to have the full benefits of digital search and indexing capabilities, files need to be well organized. The scanned, adequately processed, verified and organized data should be stored in the appropriate database. Entering data into the database should be an automated procedure, and should be taken by the software application. Creating a database is a key part of the project and should be carefully implemented. If all the previous steps of data preparation were systematically implemented, this should not be a problem.

In addition to the basic information about students, the database should contain all other relevant information from the scanned documents, such as data from semester sheets and verifications of semesters, data from the exam applications, etc. The database should also include scanned PDF documents (or images) which can be later be available and obtained from the database on one's request, for the purpose of visual verification. Scanned personal photos of students should also be introduced into the database. These data can be used to bring the scanned material under a similar interface that is currently used on the StudInfo system. In this way, it becomes possible to perform searches based on the data from the scanned documents withing the database.

The database structure and layout of the software application screen for viewing and searching database content must be strictly defined in the project, in cooperation with the relevant Faculty's structures, for the purpose of proper and responsible realization of the project. The data structure of the new database should be compatible with the structure of data in the current information system StudInfo.

3.1. Information retrieval. Information retrieval from the database implies that the graphical user interface of appropriate software application that communicates with the database provides relevant information based on the input of basic data about the student (name, surname, study group, enrollment year, the index number, etc.) in the text format as the search criteria. Compatibility of the database with the data of the current information system StudInfo will allow easily information retrieval, with corrections or extensions of software related to the specifics of the material from the previous period.

The StudInfo system has the possibility to display relevant information about the courses, proffesors, and to retrieve information about the students, such as lists of enrolled and passed exams (Figure 10) and lists of enrolled modules and scheduling per groups within the semesters (Figure 11), by entering names and surnames of the students or the number of indexes (Figure 9). Extensions of the StudInfo system that could be considered may include review of the test applications, review of the enrolled and verified semesters and more.


Претраживање студената

Индекс Име и/или презиме Статус Сви активни Уреди по презимену и имену

Претходна претрага | Нова претрага

Информациони систем СтудИнфо, верзија 2016-06 © 2005-2016 Универзитет у Београду - Математички факултет

Figure 9: The screen for searching data about the student


Ниво квалификације Основне академске студије
Студијски програм Математика (1М)
Модул Статистика, актуарска и финансијска математика (1МС)
Тренутни статус Буџет

2015/2016 - Буџет, 7. октобар 2015.

Бр.	Семестар	Ознака	Назив	Наставник	ЕСПБ	Оцена
1.	1	M1.04	Алгебра 1	Александра Ерић	6	10 (91)
2.	1	09-M2.03	Анализа 2А	проф. др Дарко Милинковић	9	9 (83)
3.	1	M5.01	Вероватноћа и статистика А	проф. др Слободанка Јанковић	5	10 (91)
4.	1	M3.02	Геометрија 2	др Мирослава Антић	6	-
5.	1	09-PM03	Увод у организацију рачунара	др Јелена Граовац	5	-
6.	2	09-M2.04	Анализа 2Б	др Јелена Катић	9	-
7.	2	M5.02	Вероватноћа и статистика Б	проф. др Слободанка Јанковић	5	10 (91)
8.	2	M3.03	Геометрија 3	проф. др Мирјана Ђорић	5	-
9.	2	09-PM04	Објектно оријентисано програмирање	др Александар Картељ	5	9 (81)
10.	2	M4.01	Увод у нумеричку математику	др Александар Савић	5	9 (86)
Укупно положено:					35 / 60	9,50

2014/2015 - Буџет, 16. јул 2014. (оверена 7. октобра 2015.)

Бр.	Семестар	Ознака	Назив	Наставник	ЕСПБ	Оцена
1.	1	09-M2.01	Анализа 1А	проф. др Миодраг Матељевић	10	9 (83)
2.	1	09-C2.01	Енглески језик 1	Ирена Павловић	3	10 (97)
3.	1	09-M1.01	Линеарна алгебра А	проф. др Александар Липковски	7	10 (96)
4.	1	PM01	Програмирање 1	др Младен Николић	6	8 (73)
5.	1	M1.02	Увод у математичку логику	проф. др Милан Божић	5	10 (100)
6.	2	09-M2.02	Анализа 1Б	проф. др Мирољуб Јевтић	10	9 (81)
7.	2	09-M3.01	Геометрија 1	др Владица Андрејић	5	9 (86)
8.	2	09-C2.02	Енглески језик 2	Ирена Павловић	3	10 (94)
9.	2	09-M1.03	Линеарна алгебра Б	проф. др Александар Липковски	5	10 (95)
10.	2	PM02	Програмирање 2	др Младен Николић	6	9 (85)
Укупно положено:					60 / 60	9,40

Подаци о студенту
 Претрага студената

Информациони систем СтудИнфо, верзија 2016-06 © 2005-2016 Универзитет у Београду - Математички факултет

Figure 10: Search results for the student: the list of enrolled and passed exams

корисник: **biljana Стојановић**

Распоредивање по групама у семестру 2015/16 - 2

Ознака	Назив	Претходи ЕСПБ	ЕСПБ	Датум уписа	При упису школске
1МС.0	Статистика, актуарска и финансијска математика (основни)	0	60	16. јул 2014.	2014/2015.
1МС	Статистика, актуарска и финансијска математика	60	180	7. октобар 2015.	2015/2016.

Уписани модули

Уписани модули | Уписане године студента
 Претходна претрага | Нова претрага

Figure 11: Search results for the student: the list of enrolled modules and scheduling per groups within the semester

3.2. Pairing of appropriate plans and programs. For each student dossier it would be necessary to check whether the data on registered and passed exams correspond to the

prescribed plan and program for the corresponding academic year, ie. to carry out their pairing. The procedure will be considerably simplified if it is done after inserting data into the database. For that reason, information about the plans and programs for different academic years should also be part of the database. Lists with the names of the professors who were teaching certain courses in different academic years could be also important. Since in earlier periods there were special courses outside of the regular plan and program, which could not replace the mandatory courses but which students were able to attend, these should be taken into account during the pairing process.

Complete procedure of pairing of plans and programs would allow later use of material for various aspects of the searches from the database.

It can be roughly estimated that it would take about two more man-years for writing the programs, verifications, pairing of plans and programs, making corrections and extensions, adjusting the interface and more.

4. Conclusion

The present state of the Archive of Student Services of the Faculty of Mathematics is not satisfactory and cannot meet the requirements. Digitization is shown as a necessary procedure that should be done carefully. It requires an investment in time and money. It should be very important to carefully consider the costs and benefits of such a project before proceeding further.

Acknowledgments

This work is supported by a grant of the Serbian Ministry of education, science and technology under the project III 44006.

References

1. StudInfo: <https://hypatia.matf.bg.ac.rs:10022/StudInfo/scripts/nastavnici/prijavljivanjeFormular>
2. ABBYY FineReader: <https://www.abbyy.com/en-ee/finereader/>
3. Piorun M., Palmer L. *Digitizing Dissertations for an Institutional Repository: A Process and Cost Analysis*, J Med Libr Assoc, 96:3 (2008), 223–229. doi: [10.3163/1536-5050.96.3.008](https://doi.org/10.3163/1536-5050.96.3.008)
4. University of Oregon Applied Information Management Program, [Document Imaging Standards for the Migration of Student Records from a Paper to a Digital System](#)
5. University of Wisconsin – Milwaukee, [Document Imaging and Management of University Records](#)
6. Suleman H., Atkins A., Gonçalves M.A., France R.K., Fox E.A., Chachra V., Crowder M., Young J. *Networked Digital Library of Theses and Dissertations (NDLTD)*, <http://www.ndltd.org>, D-Lib Magazine, 7:9 (2001)
7. Okoro U.R., Awodele O., Kuyoro S.O., Adekunle Y.A, *Office of the Future: Digitizing Record-Keeping*, International Journal of Computing Academic Research (IJCAR), 2:2 (2013), 75–87
https://www.researchgate.net/publication/261588695_Office_of_the_FutureDigitizing_Record_Keeping

biljana@matf.bg.ac.rs