

Дигитална колекция "Краезнание" на Регионална библиотека "Димитър Талев" – Благоевград – история, настояще, бъдеще

Явор Василев

Увод

Съгласно общоприетите дефиниции за дигитализация на информация, дигитализация (цифровизация) е процес на преобразуване на информация в цифров формат (т.е. електронен), в който информацията е организирана в битове.¹ В съвременната практика, дигитализираните данни са под формата на двоични числа, които улесняват компютърна обработка и други операции, но строго погледнато, това просто означава превръщането на аналоговия изходен материал в цифров формат на десетичната или всяка друга приложима бройна система.

Етапите на цифровизация в РБ „Д. Талев“ условно разделяме на 2 части: ретроконверсия на традиционните каталози и дигитализация на съдържание. Това са взаимосвързани неща и за пълнота на цифровизацията е необходимо и двата елемента да са коректно направени.

Ретроконверсия на традиционните каталози

Както в повечето регионални библиотеки, стартирането на процеса на автоматизация в Регионална библиотека „Димитър Талев“ – Благоевград започва още в началото на 90-те години със стартиране на конверсията на каталозите от хартиен в цифров вид. В момента ретроконверсията на традиционните каталози е завършена над 90% и те са достъпни онлайн и под формата на уеб бази данни в MySQL формат, а търсенето се осъществява посредством HTML (HyperText Markup Language) форми с PHP (Hypertext Preprocessor) скриптове.

Важно е да отбележим три неща:

1. Съвременните технологии – SQL (Structured Query Language – Език за структурирани запитвания), PHP, нает хостинг с професионален бекъп на 24 часа, поддръжка на архиви в облака и защита от DDoS (Distributed denial-of-service attack – дистрибутирана атака за отказ на услуга), нелегитимни и недобронамерени посетители, които са световен стандарт, са приложени в собствена разработка на библиотечен софтуер още от 2007 г.
2. Още тогава се заложи на user-friendly (лесен за употреба) интерфейс, като ясно се обозначиха конкретните елементи на библиографското описание като авторска отговорност, заглавие, година на издаване, издателство и др. с цел потребителят да е максимално улеснен какво, къде и как да търси.
3. Процесът на автоматизация беше извършен и продължава да се извършва със собствени средства на библиотеката и усилия на колектива, без помощ по европейски програми и от спомоществатели.

При разработването на библиотечния софтуер целта беше максимално приближаване до Google интерфейса – тоест търсенето да не е претрупано с опции, а да е ясно и просто за потребителя. За успеха на тази стратегия говори анализа на лог файловете на уеб сървъра, които дават детайлна и информативна статистика за ръст на онлайн потреблението и устойчив

¹ Дигитализация [Онлайн]. Достъпно на:

<https://bg.wikipedia.org/wiki/%D0%94%D0%B8%D0%B3%D0%B8%D1%82%D0%B0%D0%BB%D0%B8%D0%B7%D0%B0%D1%86%D0%B8%D1%8F>

растеж на онлайн потребителите на справочни и други услуги, както и читателските предпочитания при търсенето.

Налични са 10 MySQL бази данни/библиографски описания на книги, периодика, статии от периодичния печат, краеведски каталози и др./, в които търсенето се осъществява за части от секундата и са предназначени за обслужване на десетки хиляди заявки за търсене едновременно. Към 13 октомври 2020 г. сумарният обем на библиографските записи в 10-те онлайн бази данни е 1 172 738.

Дигитализация на съдържание

След като процеса на ретроконверсия на традиционните каталози надвиши 90% се премина към дигитализация на съдържание. Въпросите, които стояха през 2016 г., бяха какво да се дигитализира, как да се дигитализира и от кого да се дигитализира.

Експертното решение беше да започнем с уникалния фонд от краеведска литература, който се съхранява в отдел „Краезнание“. В подръчния фонд на отдела се съхраняват сборници, книги и периодика от началото на XX век до 90-те години на XX век. За процеса на сканиране на библиотечни документи до формат А3 избрахме скенер Fujitsu ScanSnap SV600 и съответния сканиращ софтуер. Той е от типа безконтактни скенери, с LED осветление и оптична резолюция 600 на 600 dpi на зрителното поле на скенера.²

Остана за решаване въпросът кой ще дигитализира фонда. Поради изключителната заетост на колегите няхахме свободни библиотечни специалисти, които да извършват тази отговорна дейност за опазване на културното наследство. За да можем да съчетаем професионализъм при попълване на базите данни с дигитализирани обекти и в същото време да можем бързо да придвижим процеса на цифровизация, използвахме нестандартен метод.

От 2014 г. в библиотеката се провеждат курсове за компютърна грамотност – MS OFFICE и компютърна грамотност за безработни, сертифицирани от МОН. В течение на 3 месеца се обучават групи безработни, които след положени изпити получават удостоверение, че съответното лице има квалификация, за да работи свободно с компютърна техника и специализирани офисни софтуерни продукти.

В партньорство с НОИ, Държавен архив, Комисията за защита от дискриминация, Областната управа и др. след подбор и интервю на кандидатстващите, положили успешно изпит и с придобит сертификат, се реализираше 3-месечен стаж по специалността в съответните институции. Използвахме този момент да подберем по-компетентните обучаеми и вземахме по 2 души годишно на стаж в библиотеката, платен от Бюрото по труда – Благоевград, като работата им беше да сканират. Те преминаваха и допълнително обучение за това как се дигитализира на съответното сканиращо устройство, а колеги от библиотеката осъществяваха пост дигитализационен контрол.

За по-добро сканиране на подвързани библиотечни документи се използваше стъклен параван, като осветяването е странично, за да се сведат до минимум отблясъците. След което образа се записва от програмата под формата на Image в единична страница от многостраничен PDF файл.

Към скенера има лицензен софтуер за OCR (Optical character recognition – оптично разпознаване на символи), но поради ниското качество на източника – пожълтяване на хартията, механични увреждания и разкъсвания, ниско качество на печата, изразяващо се в

² Таблица на препоръчителен размер за печат [онлайн]. Достъпно на: <https://www.printcenter.bg/%D0%A4%D0%B0%D0%B9%D0%BB%D0%BE%D0%B2%D0%B5+%D0%B7%D0%B0+%D0%BF%D0%B5%D1%87%D0%B0%D1%82>

избледняване на мастилото, превръщането на образа на документа във full-text (пълнен текст) е крайно затруднено. Въпреки тези затруднения колекцията е значителна – по оценки на колегите над 70% от заглавията в подръчния фонд са дигитализирани и в перспектива ще се обхванат целият подръчен фонд на отдел „Краезнание“, а след това и на другите отдели на библиотеката. По възможност ще се обхванат всички отдели на библиотеката, защото няма отдел където да няма заплашени от унищожение вследствие времето, некачествена хартия и мастило и др. библиографски документи. Цялата дигитализация е за целите на опазването и съхранението на библиотечните колекции при което строго се спазва авторското право. Много от изданията са напечатани на некачествена хартия и мастилото също е некачествено в резултат на което с времето - стават негодни за употреба. Строго е забранен достъпа до дигитализираните колекции за външни лица. Достъпа е локален, на отделен компютър, който не е в мрежа и само от оторизирано от директора на библиотеката служебно лице-библиотекар. По този начин се спазва строго авторското право и в същото време цифровите копия са достъпни, за да не се увреждат повече оригиналите.

За сканиране на по-големи обекти от формат А3 обмисляме да използваме камера на статив с резолюция на CCD (Charge-Coupled Device – елемент със зарядна връзка) матрицата не по-ниска от 20 мегапиксела, с оптично приближение не по-ниско от 20X, при използване на стъклен параван и подходящо меко осветление с цел отлична видимост и минимум отблясъци. Така генерираните образи ще бъдат конвертирани в многостраничен PDF файл.

Друг временен вариант е да се използва камерата на мобилен телефон. Например Xiaomi Mi Note 10 Pro има 108 MP камера, която използва алгоритми на AI (Artificial intelligence – изкуствен интелект), базиран на невронни мрежи с deep learning (дълбоко обучение) за автоматична професионална настройка на параметрите на камерата и съответния софтуер за сканиране от Google Play Store (онлайн магазин за софтуер), където има десетки приложения. Наличието на десетки приложения за сканиране говори, че мобилното сканиране на документи е широко разпространено.

Скенерите, опростено казано, всъщност са професионални фотоапарати със съответния софтуер, които са специално пригодени да сканират хартиени документи и да генерират висококачествени изображения в изхода си. Такова палиативно решение се налага от недостъпната в момента за нас цена на скенерите формат А2, А1 или А0. Това не са професионални решения, но във времена на пандемия, социална изолация, дистанционна работа и очаквана икономическа криза, се изискват нестандартни решения. Тези решения при добро желание могат да доведат до много бърз прогрес на дигитализацията ако не на периодични заглавия, то на книги стандартен формат и то с много добро качество. За да не забавяме дигитализацията, която за нас е много важен процес, засега ще се задоволим с такива компромисни методи. Но от това качеството на дигитализационния процес няма да пострада. Само скоростта на дигитализация ще се забави.

С помощта на скенера, който притежаваме, дигитализирахме преобладаващата част на подръчния фонд на отдел „Краезнание“, а именно – 498 тома с обем от 1 – 2 до над 300 страници във всеки дигитален обект във вид на неразчетени PDF файлове, а обемът на колекцията до момента надхвърля 20 GB. Идеята е, паралелно или след пълната дигитализация на фонда на отдел „Краезнание“ и след аналитично описание на всички статии в дигитализираните издания, а те както казах са от началото до 90-те години на XX век, те да бъдат въведени в база данни с изход към Европеана. Освен това тази база евентуално ще бъде налична и в интранет мрежата на библиотеката, при строго спазване на Закона за авторското право и сродните му права, където в чл. 27 се определят 70 години след смъртта на автора като срок за запазване на авторските права.³ Достъп на потребители и читатели до тези колекции

³ Lex.bg. София:Лекс БГ [онлайн]. Достъпно от :<https://lex.bg/laws/ldoc/2133094401>

ще стане след писмено разрешение на Дирекцията „Авторско право и сродни права“ към Министерство на културата. Както казах вече, дигитализацията е с цел опазване и съхранение на библиотечните колекции.

Пълният списък на дигиталната колекция „Краезнание“ с 498 обекта е наличен на сайта на библиотеката в меню Онлайн публикации. Това са някои от по-интересните заглавията, които успяхме да съхраним в цифров вид:

1. „История на Охридската архиепископия-патриаршия от падането ѝ под турците до нейното унищожение 1394 – 1767“ – издание на Македонския научен институт – 322 страници в неразчетен PDF файл.
2. „Извори за старата история на Тракия и Македония“ – издание на Българската академия на науките – 250 страници в неразчетен PDF файл.
3. Албум „Борческа Македония“ – издаден в София през 1931 г. – 177 страници в неразчетен PDF файл.
4. „Документи за противобългарските действия на сръбските и гръцки власти в Македония през 1912 – 1913 г.“ – издание на Македонския научен институт от 1929 г. - 164 страници в многостраничен неразчетен PDF файл.
5. „Показалец на печатаните през 19 век български народни песни – част II: 1861 – 1878“ – издаден в София през 1918 г. – 189 страници в неразчетен PDF файл.
6. „Будителите – исторически роман“ – публикуван в София през 1937 г. – 139 стр. в неразчетен PDF файл.
7. „Сборник Български народни песни – книжка 1“ – издаден в София през 1884 г. – 64 страници в неразчетен PDF файл.
8. „Сборник Илинден - 1903-1925 г. – книжка 4“ – издаден в София през 1925 г. – 46 страници в неразчетен PDF файл.
9. „Показалец на печатаните през 19 век български народни песни – част I: 1815-1860“ – издаден в София през 1916 г. – 162 страници в неразчетен PDF файл.
10. „Град Горна Джумая – минало и днес“ с автор Васил Шарков – издаден в София през 1929 г. – 117 страници в неразчетен PDF файл.

Остава да се дигитализират пълните течения на периодичните издания Пиринско дело, Вяра, Време, Пиринска Македония и др., които са подвързани в томове с формат по-голям от А3/А2.

В заключение може да се каже, че пълната дигитализацията на фондовете на библиотеките у нас, е безспорно бъдещето, но изхождайки от опита при посещението на библиотеките в Кралство Нидерландия през 2018 г., смятам, че по-лесно ще се приложи по отношение на периодичните издания, а книгите на хартиен носител още дълго време ще имат широка употреба именно поради трудността за опазване на авторските права в дигиталния свят. Защото 50% от пазара на книги се формира от учебните и професионални издания и при тях дигитализацията върви с впечатляващи темпове, особено в условията на пандемия, но за останалите заглавия това не е така.⁴ Самият факт, че се закупуват машини за стерилизиране на книги от библиотеките, говори показателно, че традиционните носители на информация – книгите на хартия, все още са много популярни.

⁴ Печатните медии в дигиталната епоха. София:Нов български университет. [онлайн]. Достъпно на: http://ebox.nbu.bg/mascom12/view_lesson.php?id=19