Švástová, Pavla

Aspekty digitalizace: pořadí čtení článků v digitalizovaných starých novinách

ProInflow. 2013, vol. 5, iss. 2, pp. 4-14

ISSN 1804-2406

ASPEKTY DIGITALIZACE: POŘADÍ ČTENÍ ČLÁNKŮ V DIGITALIZOVANÝCH STARÝCH NOVINÁCH

Aspect of Digitization: The Reading order in Digitized Old Newspapers

Pavla Švástová

Moravská zemská knihovna Brno

Recenzenti:
PhDr. Libuše Foberová, PhD.
Mgr. Martin Šuraba

Abstrakt:
Výtisk novin je zaplněn různorodými informacemi sdruženými do článků, které jsou doplněny fotografiemi či obrázky a obklopeny reklamními sděleními. Jednotlivé stránky obsahují množství odkazů a grafických prvků, které čtenáře intuitivně navádějí k pokračování článku či příslušné ilustraci. Při digitalizaci je zachycen obraz strany, který toto rozvržení zachovává, ale díky novým technologiím a metadatovým formátům jako METS a ALTO je možné jít více do hloubky a extrahovat z novin informace na úrovni článků. To vyžaduje noviny digitálně "rozstříhat" na jednotlivé zóny a ty potom logicky provázat tak, že mohou být prezentovány sdruženě jako samostatný dokument, přestože článek v papírové verzi je obklopen dalšími články a vytištěn na několika stranách. Odborníky v ČR byla vytvořena specifikace profilů formátů METS a ALTO, která řeší logickou strukturu novin, jejich navázání na plný text včetně pořadí čtení jednotlivých segmentů tvořících článek. Tato specifikace zaujala i odborníky z Library of Congress, která je správcem většiny knihovnických metadatových standardů včetně standardů METS a ALTO. V současnosti se pracuje na oficiálním METS profilu.

Klíčová slova: *digitalizace starých novin, METS, ALTO, OCR, metadata*

Abstract:
The issue of the newspaper is filled with diverse information associated to the articles, which are supplemented by photographs or images and surrounded by advertisments. Newspaper page contains a number of references and graphic elements that intuitively guides the reader to continuation of the article or illustration. Digitized image preserves the layout of the page but thanks to new technologies and metadata formats such as METS and ALTO we can go into more depth and extract information from newspapers at the level of articles. This requires newspapers digitally "reshape" to the individual zones and then create a logical structure that can be presented together as a separate document, although an article in the paper version is surrounded by other articles and contained in a several pages. Experts in the Czech Republic created a specification of METS and ALTO format profiles that contains the logical structure of newspapers, their binding to the fulltext and the reading order of the segments forming the article. Experts from the Library of Congress, which manages most of library metadata standards including METS and ALTO, are interested in our solution. Now we are working on the official METS profile.

Keywords: *digitization of old newspapers, METS, ALTO, OCR, metadata*

Úvod

Digitalizace starých novin se provádí především z důvodu ochrany kulturního dědictví a zachování autentického vnímání historických událostí a kulturního dění. Noviny jako věc, která je aktuální jediný den, se tiskne na nekvalitní novinový papír, který je ohrožen degradací z důvodu kyselosti. Vzhledem k tomu, že pro historiky, genealogy i další badatele jsou noviny doslova pokladem, jejich časté užívání jen zhoršuje jejich stav. Proto se již od konce 60. let 20. století používají na území Česka prostředky pro ochranné reformátování.[1] Dříve představovala záchrannou techniku metoda mikrofilmování, dnes se převádějí papírové dokumenty či dříve vyrobené mikrofilmy (z důvodu velkého poškození papírové předlohy) do digitální podoby.

Digitalizace, pokud je provedeno kvalitní OCR[2], znamená pro čtenáře novin velkou přidanou hodnotu díky možnosti fulltextového vyhledávání. Jednotlivé stránky periodika obsahují mnoho různorodých informací od různých typů článků přes ilustrační obrázky, reklamní sdělení či inzerci. Úpravou formátu OCR z prostého txt na xml lze definovat jednotlivé zóny na stránce novin a přiřazovat jim další popis. Na základě toho lze redefinovat původní podobu a např. vyčlenit novinové články, které se nacházejí na více stranách, sdružit články od stejného autora do jediného dokumentu či "vystříhat" obrázky. Možností je mnoho.

V tomto článku bych ráda hlouběji popsala základní informace, které souvisí s nadstandardním metadatových popisem. První část velice stručně přiblíží problematiku grafické úpravy novin a typografie, druhá část popíše všechny kroky digitalizace, které je třeba udělat, než se dokument dostane ke koncovému uživateli. Poslední část se potom bude podrobněji věnovat potřebám metadatového popisu, metadatovým standardům METS a ALTO, které určují strukturu digitalizovaného dokumentu, a jejich konkrétnímu využití v českém prostředí.

1. Problematika grafické úpravy novin

Noviny patří mezi periodické čili pravidelně vycházející publikace, které mají specifickou grafickou úpravu, obsahují typické novinářské či literární útvary a přinášejí čtenáři důležité aktuální informace o dění ve společnosti. Design novin klade hlavní důraz na snadnou čitelnost, přehlednost a vizuální přitažlivost.[3] Existují definované zásady, na které by měl grafik dbát. První zásadou je, že obsah článku je určující pro umístění i úpravu (výběr fontu, zvýraznění, velikost titulku apod.), proto nejdůležitější zprávy hledáme na titulní straně s výrazným titulkem, fejeton či povídku

[1] Ochranné reformátování. *Národní knihovna České republiky* [online]. 01.12.2012 [cit. 2013-07-01]. Dostupné z: http://wwwold.nkp.cz/pages/page.php3?page=weba_reform.htm
[2] OCR = Optical Character Recognition
[3] GARCIA, M.: Pure Design [online]. Miller Media, 2002 [cit. 2013-04-12]. ISBN 0-9724696-0-5. Dostupné z: http://issuu.com/mariogarcia/docs/mario_garcia_pure_design

označíme kurzívou, stránka s názory a komentáři je snadno vizuálně odlišitelná atd. Druhá zásada říká, že grafická úprava má korespondovat s účelem a společenským dosahem novin. Na pultu novinového stánku tak jednoduše rozeznáme zpravodajské deníky od bulvárních plátků.[4]

Základní rozvržení novinové stránky obsahuje okraj, sloupce a mezery mezi sloupci. Z typografického hlediska rozlišujeme v novinách titulky, podtitulky, perex, sloupce textu, ilustrace, popisky pod ilustrace, navigační prvky jako vodící lišty či rámečky a nedílnou součástí novin je také inzerce. Pro vzhled novin je nesmírně důležitá charakteristická hlavička, která obsahuje název novin, datum, číslo vydání a ročníku apod. Všechny tyto prvky slouží primárně k orientaci čtenáře v textu, zvýraznění podstatných informací a jejich správnou kombinací docílíme vizuální vyváženosti.

2. Digitalizace starých novin

2.1 Příprava dokumentů

Digitalizace či obecně reformátování starých novin[5] začíná poměrně náročnou přípravou. Příprava zahrnuje v první řadě pečlivé naplánování všech níže popsaných procesů, kalkulace nákladů a náročnosti na personální obsazení. Následuje kompletace celého periodika. To obnáší kontrolu desítek až stovek svazků číslo po čísle a následné shánění chybějících kusů. V případě mikrofilmování, které se často provádí současně s digitalizací i v dnešní době, to obnáší mít řady periodika správně seřazené předtím, protože nelze čísla na mikrofilmu jednoduše přeskládat. Kompletace může trvat i roky, ale není vhodné tuto fázi podcenit.

V některých případech je do fáze přípravy zahrnuta i konzervační a restaurátorská činnost. Ta se může lišit v rozsahu i metodologii jednotlivých činností, ale můžeme si ji zjednodušeně představit jako souhrn následujících činností: rozřezání svazků na listy, jejich odkyselení, vyrovnání, zrekonstruování chybějících částí listů a vylisování. Takto upravené stránky jsou samozřejmě pro digitalizaci vhodnější. Zrestaurované stránky jsou často po digitalizaci zakonzervovány v nesvázané podobě a dále již nepůjčovány.

K přípravným pracem patří též kontrola bibliografických metadat v knihovním systému, zjištění, zda již dokument nebyl digitalizován, aby se zamezilo duplicitám a definování činností veškerých zodpovědných osob, které zajišťují pohyb dokumentu po pracovišti či pracovištích, pokud je digitalizace prováděna externě.[6]

2.2 Skenování a úprava obrazů

Vzhledem k formátu novin je nutné pro digitalizaci použít velkoformátové skenery velikosti minimálně A3, pokud skenujeme jednostránkově a velikosti A2, pokud skenujeme dvoustránkově. V případě starých novin nelze vzhledem k fyzickému stavu materiálu vždy využít robotické skenery.

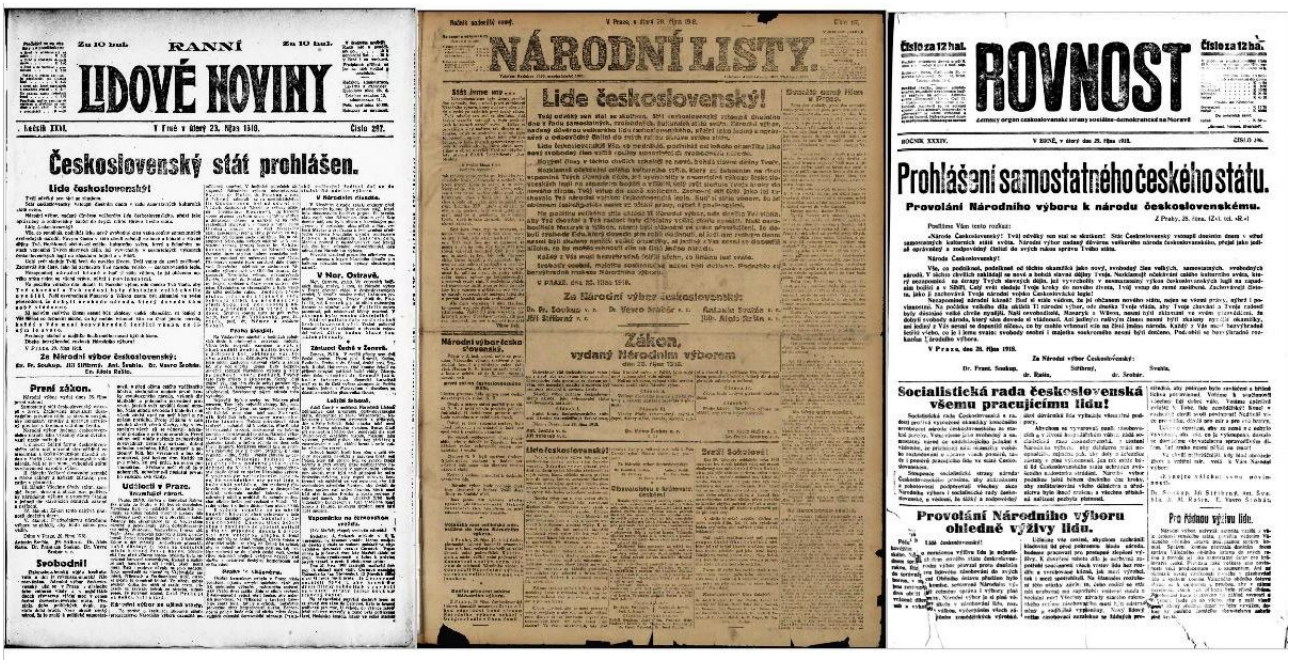
[4] BURIÁNEK, Zdeněk. ABC o grafické úpravě novin a časopisů. 1. vyd. Praha: Orbis, 1960. 222 s.

[5] Staré nebo také historické noviny (v angličtině historical newspapers) označují takové noviny, které jsou ohrožené degradací papíru.

[6] Digitalizace v projektu NDK. *Národní digitální knihovna* [online]. 09.01.2012 [cit. 2013-07-01]. Dostupné z: http://www.ndk.cz/digitalizace

V první řadě je nutné posoudit stav každého titulu novin jednotlivě až na úroveň svazku[7] a na základě toho určit typ skeneru - robotický nebo ruční skener.

Žádoucí pro archivaci je skenování do formátu bezeztrátového tiffu s barevnou hloubkou 24 bitů a s rozlišením minimálně 300 DPI (u novin s drobným písmem je vhodné použít 400 DPI). Následná úprava obrazu zahrnuje otočení, ořez, vyrovnání řádků a případnou úpravu barevnosti, odstranění šumu, vylepšení kontrastu a jasu apod. Veškerá úprava je důležitá pro následnou výrobu OCR, protože čím je sken kvalitnější, tím větší je úspěšnost rozpoznání znaků. Dle národních doporučení je obraz zkonvertován do formátu JPEG2000 a to jak pro archivní kopii, tak pro uživatelskou kopii.[8] Periodika, u kterých proběhlo ochranné reformátování v minulosti pouze metodou mikrofilmování, se z důvodu náročné přípravy a stavu dokumentů skenují z mikrofilmů speciálním mikrofilmovým skenerem.



Obr. 1: Ukázka titulních stran starých novin (29.10.1918)

2.3 OCR

Fulltextové vyhledávání je největší přidanou hodnotou, kterou digitalizace tištěných dokumentů přináší. Optical Character Recognition neboli optické rozpoznání znaků je technologie, kterou lze převádět obrazová data do počítačem čitelné textové podoby. Po naskenování a úpravě jednotlivých obrázků dokumentu dokáže text přečíst člověk, ale pro počítač je to jen shluk pixelů s různou barevnou hodnotou a neliší se pro něj od obrázku, kde se žádná textová informace nenachází. Pokud však chceme v obrázku s textem vyhledávat či jeho původní znění upravovat, potřebujeme

[7] Svazek je soubor čísel periodika svázaných po učitých částech - ročník, půl ročníku ale i několik ročníků, které v knihovním katalogu tvoří jednu jednotku. Každý svazek posuzujeme zvlášť - starší ročníky bývají v horším stavu.
[8] Nové standardy digitalizace (od roku 2012). *Národní digitální knihovna* [online]. 2012, 20.2.2013 [cit. 2013-06-10]. Dostupné z: http://ndk.cz/digitalizace/nove-standardy-digitalizace-od-roku-2011

ho převést do textové podoby, kterou bude možné upravit např. jako e-mail, textový dokument nebo obsah webu.

Úspěšnost OCR je závislá na kvalitě předlohy. Nově vytištěné knihy a časopisy, jejichž obrázky jsou před OCR upraveny, mají takřka stoprocentní výsledky. Mezi dokumenty s nekvalitním výstupem OCR se řadí zejména historická periodika či rukopisná díla. Existuje řada projektů (např. Impact[9] nebo Europeana Newspapers[10]), jejichž cílem je vylepšení kvality OCR. Řešením může být např. optimalizace algoritmů provádějících rozpoznání znaků, hledání optimální úpravy obrazu před samotných vyčítáním textu, výroba specializovaných jazykových slovníků či zapojení veřejnosti do ručních oprav již vytvořeného OCR.

Výsledek OCR je zapsán v textovém souboru, který je pomocí strukturálních metadat připojen k příslušnému obrazu. Dříve byl využívám formát txt, nyní se přechází k formátu xml, konkrétně standardu ALTO, který bude podrobněji popsán níže.

2.4 Metadata

Digitální dokumenty se popisují pomocí metadat. Metadata dělíme na:

1. popisná (také bibliografická či deskriptivní)
2. technická
3. administrativní
4. strukturální

Popisná metadata digitálních dokumentů se v podstatě neliší od "klasických knihovnických" záznamů v knihovním katalogu, obsahují jmenný a věcný popis dokumentu. Při digitalizaci novin je nutné vytvořit popis nejen pro celé periodikum, ale též pro ročník, číslo a jednotlivé články. Popisná metadata k článkům obsahují minimální popis, který zahrnuje název a autora článku, žánr a jazyk, nepovinně věcný popis.

Další typy metadat jsou specifické pro digitalizované dokumenty. Technická metadata popisují vlastnosti obrazových dat (rozlišení, velikost, použité hardwarové i softwarové nástroje atd.) a patří mezi ně např. standard MIX nebo ALTO, jehož podrobnější popis bude následovat v další kapitole. Administrativní metadata řeší autorsko-právní aspekty digitalizace a lze jimi zaznamenat historii digitálního dokumentu od jeho vzniku až po zpřístupnění a archivaci. Nejvyužívanějším standardem je v současnosti PREMIS[11]. Digitalizovaným článkům by se v ideálním případě měla přiřadit informace o autorských právech, zda je článek již autorsky volný případně kdy k tomu dojde.

K provázání veškerých obrazových dat a metadat se používají strukturální metadata, která kromě toho, že slouží jako kontejner pro ostatní typy metadat, dokáží vystihnout strukturu dokumentu,

[9] http://www.impact-project.eu/
[10] http://www.europeana-newspapers.eu/

[11] PREMIS (Preservation Metadata Implementation Strategies) je standard, pomocí kterého lze zapsat metadata, která napomáhají jejich dlouhodobé ochraně. Více informací naleznete na http://www.loc.gov/standards/premis/.

která je zejména u novin poměrně složitá. Všechny výše zmíněné standardy jsou spravovány Kongresovou knihovnou.[12]

2.5 Archivace a zpřístupnění

Digitální obrazy spolu s metadatovým popisem putují jako balík do digitálního archivu a do digitální knihovny. Digitální archiv by měl zajistit dlouhodobou ochranu dokumentu, digitální knihovna jeho zpřístupnění pro uživatele.

3. Standardy METS a ALTO a problematika pořadí čtení

V této kapitole jsou stručně popsány standardy METS a ALTO a jejich vzájemná provazba za účelem popisu detailní struktury novinových stránek.

3.1 METS

METS (Metadata Encoding and Transmission Schema)[13] je standard, který popisuje strukturu digitalizovaného dokumentu a připojuje k obrazovým datům popisná, administrativní a technická metadata. Je vyjádřený ve formátu xml a jeho správu a vývoj zaštiťuje Kongresová knihovna. Standard METS je velice obecný, což je jeho největší výhodou i slabinou zároveň. Výhodou proto, že jím lze popsat prakticky jakýkoliv typ dokumentu, slabinou proto, že díky této obecnosti se jeho implementace v různých digitalizačních projektech podstatně liší. Kvůli tomu vznikla řada doporučení, jak METS používat v kombinaci s dalšími standardy včetně ALTO[14] a existuje seznam oficiálních profilů, které popisují implementaci METS v různých institucích a projektech.[15]

Pro popis struktury novin s rozdělením na články byla v rámci plánování projektu Národní digitální knihovna definována podoba, která odpovídá standardu METS a je v současnosti závazná pro digitalizační projekty v České republice (např. digitalizace v rámci programu VISK podprogramu VISK 7, krajské digitalizace) a plánuje se použít také na Slovensku v digitalizačním projektu Slovenskej národnej knižnice DIGDA. Do konce roku 2013 se plánuje přeložení standardu do anglického jazyka a vytvoření oficiálního profilu.

METS dokument se skládá ze sedmi hlavních sekcí:

1. Hlavička METS - obsahuje popis samotného METS dokumentu.
2. Sekce popisných metadat - obsahuje popisná metadata ve formátu MODS[16], DC[17], MARCXML[18] apod.
3. Sekce administrativních metadat - obsahuje administrativní a technická metadata ve formátu PREMIS, MIX[19] apod.

---

[12] ŠVÁSTOVÁ, Pavla. Metadata a metadatové standardy užívané v knihovnách. 9.12.2010 [cit. 2013-06-10]. Dostupné z: http://www.slideshare.net/pavluskas/prezentace-pardubice

[13] METS: Metadata Encoding & Transmission Standard. *The Library od Congress* [online]. 2012 [cit. 2013-06-12]. Dostupné z: http://www.loc.gov/standards/mets/

[14] Další informace: http://www.loc.gov/standards/alto/techcenter/use-with-mets.php

[15] http://www.loc.gov/standards/mets/mets-profiles.html

[16] Další informace: http://www.loc.gov/standards/mods/

[17] Další informace: http://dublincore.org/documents/dces/

[18] Další informace: http://www.loc.gov/standards/marcxml/

4. Sekce souborů - obsahuje seznam souborů, které obsahuje digitalizovaný dokument, rozdělených do skupin podle jejich typu.
5. Strukturální mapa - nejdůležitější část, která vyjadřuje strukturu digitalizovaného dokumentu, propojuje skeny a metadata a ze které je odkazováno do dalších sekcí. Dělí se na fyzickou a logickou mapu.
6. Strukturální odkazy - vyjadřuje vazby mezi fyzickou a logickou strukturální mapou.
7. Sekce chování - definuje chování mezi jednotlivými objekty v METS dokumentu

3.2 ALTO

ALTO (Analyzed Layout and Text Object)[20] je formát založený na xml popisující vzhled a obsah textu digitalizovaného dokumentu. Současným trendem zejména při digitalizaci periodik je přímo v procesu zpracování generovat OCR právě ve formátu ALTO. To určuje přesnou polohu a hodnotu jednotlivých částí konkrétní digitalizované strany od určení polohy a typu článků a obrázků, umístěných na stránce, až na úroveň jednotlivých slov.

Dokument ALTO se skládá ze 3 základních sekcí:

1. Popis - obsahuje informace o ALTO dokumentu jako takovém.
2. Styly - obsahuje informace o použitých fontech a dalších vlastnostech týkajících se vzhledu dokumentu.
3. Rozvržení - popisuje samotný obsah dokumentu, obsahuje znaky vyčtené OCR softwarem sdružené do slov a jejich přesné umístění na stránce dokumentu. V této sekci a jejích podsekcích je přesně definovaná oblast textu na celé straně, sloupce, odstavce, řádky, slova apod. Každý element, který popisuje konkrétní oblast na stránce, má svůj jednoznačný identifikátor a může na něj být odkazováno z METS dokumentu.

Pro ALTO platí v procesu implementace to samé jako pro METS - jedná se obecný rámec, který je možný využít více způsoby a proto se mohou výstupy v různých projektech lišit. Záleží na počáteční definici pro zhotovitele projektu. Z ALTO výstupu lze konverzí zrekonstruovat přesnou podobu stránky v textové podobě, aniž bychom měli k dispozici obrazová data. To může být výhodou např. při generování uživatelských kopií, např. pdf nebo dalších textových formátů (e-pub, mobi atd.) pro elektronické čtečky, kde upřednostňujeme malou velikost souboru.

3.3 Propojení METS a ALTO a problematika pořadí čtení

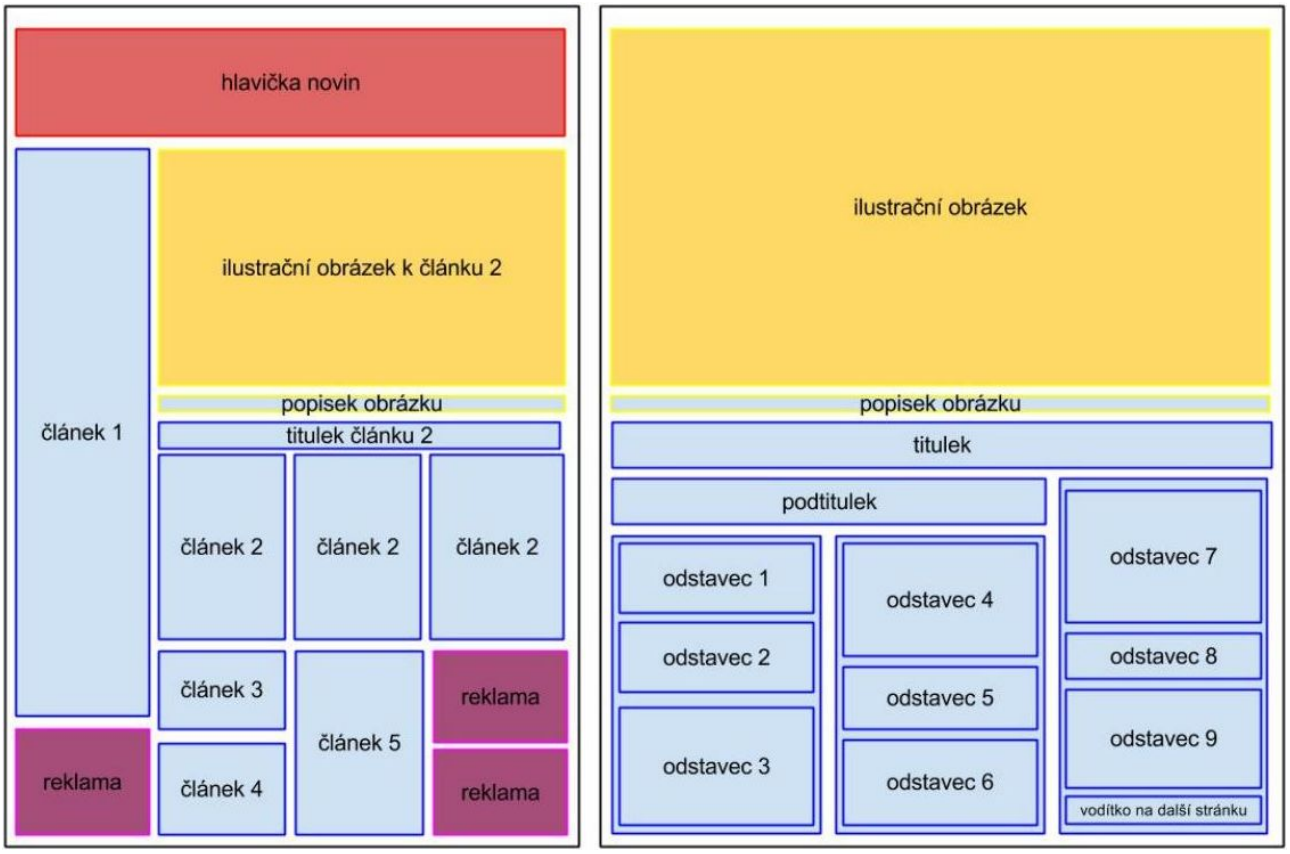
Vymezení hranice jednotlivých článků na stránce a jejich pořadí čtení není možné zcela automatizovat a z velké části je potřebný zásah člověka. Specializovaný software[21] vytvoří soubor ALTO, kde jsou vyjádřeny jednotlivé segmenty - sloupce, odstavce, titulky, ilustrace atd., ale strukturu článku v METS není možné určit se stoprocentní úspěšností a zejména u článků přesahujících několik stran, je nutná kontrola a případné propojení a určení pořadí čtení včetně napojení na popisná a administrativní metadata konkrétního článku.
Níže uvedené rozvržení stránky bude popsáno pouze symbolicky, protože podrobný popis ve struktuře xml s konkrétními příklady lze nastudovat přímo z definice českého národního standardu

[19] Další informace: http://www.loc.gov/standards/mix/
[20] ALTO: Technical Metadata for Optical Character Recognition [online]. 2013 [cit. 2013-06-12]. Dostupné z: http://www.loc.gov/standards/alto/

[21] např. ABBYY Fine Reader

na webových stránkách Národní knihovny.[22] V METS je struktura včetně popisu článků vyjádřená v logické strukturální mapě pomocí v sobě zanořených elementů s odkazy do dokumentu ALTO na jednotlivé textové bloky představující odstavce a další útvary jako např. obrázky.

Představme si stránku novin, která obsahuje několik článků, přičemž některé z nich pokračují na dalších strankách, ilustraci, která doplňuje jeden z článků a reklamní sdělení. Každý článek se skládá z titulku, podtitulku a odstavců, některé z nich zahrnují ilustrace a popisky ilustrací.



Obr. 2: Ukázka rozvržení stránky a rozvržení článku

Struktura stránky včetně pořadí čtení vyjádřená symbolicky by vypadala následovně:

1 Stránka
- 1.1 Hlavička novin
- 1.2 Článek 1
- 1.3 Článek 2 pokračující na str. 2
- 1.4 Článek 3
- 1.5 Článek 4
- 1.6 Článek 5
- 1.7 Reklama 1
- 1.8 Reklama 2

[22] Další informace: http://ndk.cz/digitalizace/nove-standardy-digitalizace-od-roku-2011/specifikace_periodika_1-4.pdf

1.9 Reklama 3

Článek 2 je z útvarů na stránce nejsložitější a navíc pokračuje na další stránce. Vyjádřený symbolicky včetně pořadí čtení by vypadal následovně:

1 Článek
1.1 Titulek
1.2 Podtitulek
1.3 Odstavec 1 (strana 1)
1.4 Odstavec 2 (strana 1)
...
1.X Odstavec 1 (strana 2)
...
1.Y Ilustrace k článku
1.Z Popisek ilustrace

Výsledná struktura článku kombinuje pozici (přesné umístění na stránce) a hodnotu (vyčtené OCR) odstavce na stránce ze souoru ALTO a logickou mapu zapsanou v souboru METS, který určuje kromě prosté struktury také vlastnosti jednotlivých odstavců a dalších segmentů na stránce. Vlastností může být typ odstavce (titulek, podtitulek, autor článku, text článku, ilustrace, popisek k ilustraci atd.) a metadata popisující článek jako celek.

Závěr

Možnost členit jednotlivé stránky novin na články a další útvary přináší novou přidanou hodnotu digitalizovaným dokumentům. Zpracování je sice poněkud náročnější na lidskou práci a vyžaduje pořízení specializovaného softwaru, ale na druhou stranu se tím posouvají hranice možností pro badatele i instituce, které poskytují obsah starých novin. Nespornou výhodou je možnost zpřístupnění jednotlivých článků, které jsou autorsky volné[23] i v případě, že stránka novin, na které se konkrétní článek nachází obsahuje i autorsky nepřístupné články. Doposud nebylo možné konkrétní článek automaticky oddělit od ostatních a zpřístupnit samostatně (samozřejmě pokud ho nechtěl knihovník "vystříhat" z digitalizovaného obrazu pomocí některého z grafických programů). Nově zavedené standardy tuto možnost podporují a nyní zbývá jen upravit software pro zpřístupnění, aby měl jeho uživatel možnost stáhnout si jen to, co potřebuje. Výstupem by v tomto případě byla obrazová prezentace, která cenzuruje autorsky nepřístupné články pomocí neprůhledné vrstvy nebo generuje obrazový výstup, kdy jsou jednotlivé segmenty článku sdruženy do jediného celku podle pořadí čtení.

Další možností, kterou jsem naznačila výše, je získání plného textu článku v podobě vhodné pro zařízení, kde nám již nezáleží na původních obrazových datech, ale chceme získat pouze textovou vrstvu vyčtenou pomocí OCR. Bohužel konkrétně u starých novin je výstup OCR ve velice špatné kvalitě a není možné jej bez ruční opravy použít.

[23] Pozn.: Dílo je autorsky volné po 70 letech od smrti všech autorů.

V současnosti je vyřešen základ, tedy možnost metadatového popisu a metodika propojování jednotlivých segmentů do článků a již druhým rokem vznikají data, která popsanou strukturu obsahují. Nyní je na řadě upravení nástrojů pro zpřístupnění, které umožní uživatelům zmíněné výhody využívat.

Použité zdroje

1. BURIÁNEK, Zdeněk. ABC o grafické úpravě novin a časopisů. 1. vyd. Praha: Orbis, 1960. 222 s.
2. GARCIA, M.: Pure Design [online]. Miller Media, 2002 [cit. 2013-04-12]. ISBN 0-9724696-0-5. Dostupné z: http://issuu.com/mariogarcia/docs/mario_garcia_pure_design
3. Nové standardy digitalizace (od roku 2012). *Národní digitální knihovna* [online]. 2012, 20.2.2013 [cit. 2013-06-10]. Dostupné z: http://ndk.cz/digitalizace/nove-standardy-digitalizace-od-roku-2011
4. METS: Metadata Encoding & Transmission Standard. *The Library od Congress* [online]. 2012 [cit. 2013-06-12]. Dostupné z: http://www.loc.gov/standards/mets/
5. ALTO: Technical Metadata for Optical Character Recognition [online]. 2013 [cit. 2013-06-12]. Dostupné z: http://www.loc.gov/standards/alto/
6. Ochranné reformátování. *Národní knihovna České republiky* [online]. 01.12.2012 [cit. 2013-07-01]. Dostupné z: http://wwwold.nkp.cz/pages/page.php3?page=weba_reform.htm
7. Digitalizace v projektu NDK. *Národní digitální knihovna* [online]. 09.01.2012 [cit. 2013-07-01]. Dostupné z: http://www.ndk.cz/digitalizace
8. ŠVÁSTOVÁ, Pavla. Metadata a metadatové standardy užívané v knihovnách. 9.12.2010 [cit. 2013-06-10]. Dostupné z: http://www.slideshare.net/pavluskas/prezentace-pardubice