

# Nacionālās digitālās bibliotēkas kolekcija "Periodika"

2008.gada 2.jūlijā Latvijas Nacionālajā bibliotēkā (LNB) tika atklāta jauna Latvijas Nacionālās digitālās bibliotēkas "Letonica" (LNDB) kolekcija "Periodika". Pēc dažiem mēnešiem — 12.novembrī, noslēdzoties projekta otrajam posmam, interesenti tika iepazīstināti ar digitālās kolekcijas turpinājumu "Periodika.lv 2.0". Digitalizācijas projekts "Periodika.lv" uzsākts 2007.gada decembrī un tā mērķis ir padarīt digitāli pieejamu visu Latvijas nacionālo periodiku, kopā aptuveni 6 miljonus periodisko izdevumu lappušu.

Laikrakstu digitalizācijas aizsākums LNB gan datējams jau ar 1999.gadu, kad projekta "Mantojums — 1: Latvijas periodisko izdevumu (1822–1940) saglabāšana" ietvaros tika izveidota pirmā digitalizēto avīžu kolekcija ([http://old.lnb.lv/digitala\\_biblioteka/Laikraksti/index.htm](http://old.lnb.lv/digitala_biblioteka/Laikraksti/index.htm)). Tajā hronoloģiskā secībā pieejami 100 laikrakstu nosaukumu pilnteksti PDF formātā, kas veidoti no oriģinālu attēlu datnēm (TIFF formātā), tādēļ iespējama vien aptuveni 300 000 lappušu apjomīgā laikrakstu krājuma pārlūkošana, "šķirstot" to lapu pa lapai. Projekta galvenais mērķis bija nodrošināt izdevumu saglabāšanu, izveidojot oriģinālu digitālās kopijas. Zemās kvalitātes papīra vai pārāk liela nolietojuma dēļ liela daļa LNB laikrakstu krājuma bija kritiskā stāvoklī un vairs netika izsniegti lasītājiem, līdz 1945.gadam izdotās avīzes bija aizliegtas kopēt. Izveidotās digitālās kopijas tagad bieži vien ir vienīgās, ko lasītājam var piedāvāt. Periodikas digitalizācija nepieciešama arī tās ierobežotās pieejamības dēļ — senos laikrakstus glabā vien lielākās zinātniskās bibliotēkas, taču arī tur reti pieejami pilni izdevumu komplekti — trūkst atsevišķu numuru, lapu, tās ir sabojātas. Arī LNB laikrakstu krājums nav pilnīgs, tādēļ digitalizējamo izdevumu komplektēšanā LNB sadarbojas ar LNDB galveno partneri — Latvijas Akadēmisko bibliotēku.

Šobrīd digitālā kolekcija "Periodika" aptver vairāk nekā 350 tūkstošus lappušu (aptuveni 2 miljoni rakstu) no 41 preses izdevuma latviešu, krievu un vācu valodā no 1895.gada līdz 1957. gadam. Tie ir tādi kultūrvēsturiski nozīmīgi laikraksti kā "Jaunākās Ziņas", "Dienas Lapa", "Cīņa", "Tēvija", "Valdības Vēstnesis", "Padomju Latvija", "Rigasche Rundschau", arī reģionālie laikraksti "Kurzemes Vārds", "Brīvā Venta", "Zemgales Balss" un citi. Pieejamas arī Latvijas Republikas Saeimas stenogrammas, Izglītības ministrijas un Iekšlietu ministrijas mēnešraksti. "Periodika.lv 2.0" kolekcija papildināta arī ar laikraksta "Literatūra un Māksla" numuriem (1945–1970) un žurnālu "Economists" (1925–1931).

Autortiesību ierobežojumu dēļ publiski tiešsaistē tīmekļa vietnē [www.periodika.lv](http://www.periodika.lv) vai LNDB tīmekļa vietnē [www.lnb.lv](http://www.lnb.lv)



Kolekcijas "Periodika" vadītājs Aigars Staks

[lnb.lv/lv/digitala-biblioteka](http://lnb.lv/lv/digitala-biblioteka) sadaļā "Laikraksti" (<http://www.lnb.lv/lv/digitala-biblioteka/laikraksti>) brīvi pieejama aptuveni 1/3 daļa kolekcijas — preses izdevumi, kas vecāki par 70 gadiem (t.i., iznākuši līdz 1938.gada 1.jūlijam). Taču, ja līdz šim ar pārējo kolekcijas daļu (tīmekļa vietnē [www.periodika.lnb.lv](http://www.periodika.lnb.lv)) varēja iepazīties vien LNB, tad līdz ar "Periodika.lv 2.0" izveidi kopš 2008.gada novembra autortiesiski aizsargātie raksti pieejami "Gaismas tīkla" bibliotēku, kā arī Latvijas Akadēmiskās bibliotēkas telpās.

Izmantojot korporācijas Olive Software (ASV) programmatūru Olive ActivePaper Archive, kolekcijā "Periodika" veikta optiskā rakstzīmju atpazīšana (Optical character recognition, OCR) un teksta indeksācija, tādēļ informāciju vēsturisko laikrakstu pilntekstos iespējams meklēt pēc atslēgvārdiem, turklāt ne vien visā tekstā, bet arī atsevišķos tā "segmentos", piemēram, tikai reklāmās vai attēlu parakstos u.c. Meklējami ir arī tie vārdi, kas sadalīti, pārnesot jaunā rindā. Kopumā teksta atpazīšanas kvalitāte ir augsta: 85–99,9% atkarībā no oriģināla tehniskā stāvokļa. Lai arī pirms digitalizēšanas materiāli tiek izlīdzināti, attīrīti un restaurēti, dažu oriģinālu kvalitāte ir tik slikta, ka nav iespējama absolūti nekļūdīga atpazīšana. Lai rezultātu uzlabotu, papildus programmatūras veikumam notiek manuāla atpazītā teksta rediģēšana un kļūdu labošana.

Lasītājam tiek piedāvāta vienkārša un izvērsta meklēšana, kā arī pārlūkošana pēc izdevumu nosaukumiem un iznākšanas datumiem. Jau tagad iespējama meklēšana arī vārdu locījumos, taču meklēšanas sistēmu plānots attīstīt un ar laiku piedāvāt arī meklēšanu pēc sinonīmiem un citas iespējas.

Olive Software programmatūra Olive ActivePaper Archive nodrošina ērtu saskarni. Atrastos rezultātus iespējams skatīt dažādos formātos (izdevuma oriģinālskats — attēls, PDF, HTML, kā arī teksta formā), sakārtot, kopēt tekstu vai tā fragmentus, lejupeļādēt savā datorā (PDF formātā), nosūtīt uz e-pastu, izdrukāt u.c. Ja raksts sākas vienā lapā un beidzas citā, laikrakstu pārlūks automātiski savieno daļas vienotā tekstā. Lai atvieglotu orientēšanos, meklētie atslēgvārdi tekstos tiek iekrāsoti. LNDB izstrādātāji plāno veikt tālāku projekta saskarnes uzlabošanu. Uzsākts darbs pie digitālās kolekcijas izmantošanas pamācību izstrādes.

Kolekcijas jaunā versija "Periodika.lv 2.0" piedāvā arī citus nozīmīgus uzlabojumus un jauninājumus. Tā kā "Periodikā" iekļauti izdevumi svešvalodās (angļu, franču, krievu, vācu u.c.), kā arī ņemot vērā to, ka LNDB ir Eiropas digitālās bibliotēkas *Europeana* sastāvdaļa, izveidota saskarne arī angļu valodā, kas sekmē kolekcijas starptautisku izmantojamību. Izveidots arī diskusiju forums, kurā savu viedokli paust acināts ikviens interesents. Tajā iespējams apspriest projekta norisi, uzdot jautājumus, izteikt ierosinājumus un vēlmes, ziņot par problēmām vai kļūdām (piemēram, neprecizitātēm izdevumu segmentēšanā) utt. Forumā jau aizsākušās diskusijas par kolekcijas saturu un tehnisko risinājumu, piemēram, tiek diskutēts par tēmu "Avīzes, kuras būtu vērts iekļaut "Periodika.lv"". Diskusiju forumā publicētā informācija tiek izmantota, lai pilnveidotu digitālo kolekciju. Uz uzdotajiem jautājumiem foruma lietotāji saņem izsmeltojas atbildes.

Kolekcijā "Periodika" iekļauta arī projekta "Mantojums — 1: Latvijas periodisko izdevumu (1822–1940) saglabāšana" ietvaros izveidotās digitalizēto avīžu kolekcijas antikvā drukātā daļa, bet laikraksti, kas iespiesti vecajā drukā jeb gotikā (tika izdoti līdz 20.gadsimta 30.gadu vidum), kolekcijai tiks pievienoti pēc vecās drukas atpazīšanas programmas izstrādes. Tā kā izmēģinājumu projektā izrādījās, ka latviešu gotiku, kam raksturīga liela burtveidolu dažādība, atšķirībā no, piemēram, vācu vai krievu gotikas, kvalitatīvi atpazīt nespēj neviena no tirgū piedāvātajām gatavajām programmatūrām un programmu izstrādātāji latviešu valodas mazās izplatības dēļ šo funkcionalitāti, visticamāk, neiekļautu arī nākamajās programmas standarta versijās, LNB izsludināja iepirkuma konkursu tās speciālai izstrādei. Konkursā uzvarēja starptautiska kompānija *ABBYY*, kas piedāvā programmatūru *ABBYY FineReader*, kam, pamatojoties uz LNB iesniegtajiem 400 paraugiem, tiek veidots īpašs latviešu gotikas atpazīšanas modulis. Jaunā versija būs gatava 2009.gada pavasarī. Viens no konkursa nosacījumiem bija iespēja jaunizstrādāto versiju, atšķirībā no kompānijas standarta komercproduktiem (piemēram, vecās drukas atpazīšanas programmas *ABBYY FineReader XIX* licence paredzēta 10 000 lappusēm) izmantot neierobežota skaita lappušu apstrādei. Izstrādāto produktu kompānija *ABBYY* izplatīs arī tirgū un to varēs iegādāties visi interesenti.

Lai nodrošinātu meklēšanas iespēju atpazītajā vecās drukas tekstā, tas *pārtulkojams* mūsdienu cilvēkam un arī meklēšanas programmai saprotamā valodā, tādēļ atpazītajam tekstam piesaistāmas vārdnīcas. Sadarbībā ar Latvijas Universitātes Matemātikas un informātikas institūta (LU MII) Mākslīgā intelekta laboratoriju noris darbs pie vēsturisku tekstu apstrādes programmas izveides. Veicot vecās drukas tekstu atpazīšanu, sākotnēji tiek iegūts *OCR* kļūdas saturošs izejmateriāls. Piemēram, atpazīšanas programmas bieži vien jauc "s", "z" un "f" burtveidolus, kas savstarpēji ir ļoti līdzīgi, līdz ar to vārdu "saskatīt" varētu atpazīt kā "fafkaītīt". Balstoties uz dažādiem latviešu valodas vārdu veidošanas likumiem, šādas kļūdas var automātiski labot. Kad tas paveikts, jāveic vārdu ortogrāfijas "modernizēšana", pārakstot tos atbilstoši mūsdienu valodas likumiem. Piemēram, vārdu "wehsture" pārveidojot par "vēsture". Visbeidzot nepieciešama arī sinonīmu iekļaušana, tāpēc LU MII strādā pie vecvārdiem bagātās Kārļa Milēnbaha un Jāņa Endzelīna "Latviešu valodas vārdnīcas" (<http://www.aialab.lv>).



Kolekcijas "Periodika" atklāšanas plakāts

*lv/MEV*) piesaistes kolekcijai. Lai uzlabotu meklēšanu kolekcijā, plānota arī ģeotoponīmu (vietvārdu), sinonīmu, svešvārdu, autoru pseidonīmu u.c. vārdnīcu un tezauru piesaiste. Īpaši svarīga vārdnīcu piesaiste ir tieši senajiem tekstiem, jo mūsdienu lasītājs tajos rodamos vecvārdus kā atslēgvārdus, visticamāk, nelietos.<sup>2</sup>

Periodikas tekstu digitalizēšana un apstrāde ir ne vien kultūrvēsturiski, bet arī zinātniski nozīmīga — iegūtais teksta masīvs tiks iekļauts, piemēram, latviešu valodas korpusā.

Kaut arī kultūras iestāžu radīto produktu reklamēšanai netiek rīkotas tik vērienīgas publicitātes kampaņas kā komerciāliem risinājumiem, LNDB kolekcija "Periodika" īsā laikā guvusi plašu rezonansi sabiedrībā. Nedaudzās ziņas par kolekciju tiek pārpublicētas dažādos medijos, lietotāji to iesaka cits citam dažādās tīmekļa vietnēs.

Kolekcijas izmantojamības statistika rāda, ka vidēji to apmeklē 150 cilvēki dienā. Turklāt lasītāju interese nav uzskatāma par mirkļa ziņkāri — vidēji viena apmeklējuma laikā tiek aplūkotas apmēram 34 lapas, vidējais apmeklējuma ilgums — 17 minūtes. Kolekcijai ir arī savi rekordi — garākais apmeklējums ildzis 5 stundas un 45 minūtes, maksimālais apskatīto lapu skaits — 345. Starp apmeklētājiem ir gan mājas interneta lietotāji, LNB un citas bibliotēkas, augstskolas, bankas, ministrijas, Saeima, rajonu padomes, ziņu aģentūra "LETA", arī Rīgas Nacionālais zoologiskais dārzs utt., kā arī ārzemju apmeklētāji.

## Ieceres

Digitalizācijas projekta "Periodika.lv" tuvākā nākotnes iecere ir digitalizēt un kolekcijai pievienot aptuveni 2 miljonus laikrakstu un žurnālu lappušu, galvenokārt visus (toskait Latvijā izdotos nacionālo minoritāšu — poļu, vācu, krievu, ebreju u.c.) preses izdevumus, kuru saturs nav atspoguļots Augusta Ģintera retrospektīvajos un kārtējos analītiskās bibliogrāfijas rādītājos "Latviešu

zinātne un literatūra” un “Latvijas zinātne un literatūra”, jo par tiem nav pieejama pat bibliogrāfiskā informācija.

Projekta turpinājumā plānots kolekcijai pievienot arī mākslas žurnālus, kas izdoti līdz 1945.gadam, tajā skaitā slavenu “Ilustrēto Žurnālu”, kura redakcijā strādājuši Jānis Roberts Tillbergs, Kārlis Skalbe un citi ievērojami latviešu kultūras darbinieki, un humora un satīras žurnālu “Ho-ho”, kurā darbojās Uga Skulme, Ludolfs Liberts, Romans Suta, Aleksandra Beļcova-Suta u.c. pazīstami latviešu mākslinieki. Plānots digitalizēt arī citus interesantus materiālus, piemēram, Upsalas Universitātes dāvinājumu LNB — līdz šim maz pētītos, Latvijā faktiski nepieejamos Otrā pasaules kara laikā izdotos bēgļu nometņu izdevumus, tajā skaitā — Gēsthahtes (Vācija) “Libekas Vēstnesi”, Vestfāles Halles nometnē (Vācija) izdoto “Mītnes Viesi” u.c.

Līdztekus projektam “Periodika.lv” plānots uzsākt arī masveida grāmatu digitalizāciju. Izmantojot ārpalikojumu, iecerēts apstrādāt aptuveni 1 miljonu lappušu jeb aptuveni 5000–10 000 izdevumu. Šobrīd grāmatas LNB tiek digitalizētas nelielos apjomos, cik tas iespējams pašu spēkiem. Lielajā grāmatu digitalizācijas projektā iesaistīsies arī Latvijas Universitātes bibliotēka, kas cīparos pirms Otrā pasaules kara izdotos Latvijas Universitātes Zinātniskos rakstus, un Latvijas Akadēmiskā bibliotēka.

Savukārt plānotajā audio digitalizācijas studijā tiks veikta LNB Mūzikas un Fonotēkas nodaļās glabāto šellakas un vinila skaņuplašu krājuma cīparošana. Plātes ne tikai digitalizēs, bet arī attīrīs no trokšņiem, kas ļaus 70 gadus vecos ierakstus klausīties to sākotnējā kvalitātē. Skaņuplašu digitali-

zācijas mērķis ir ne tikai to saglabāšana, bet arī izmantošanas iespēju paplašināšana, analogos skaņu ierakstus pārvēršot mūsdienīgajā digitālajā formātā. Vienojoties ar autortiesību turētājiem, ar laiku LNB plāno skaņuplašu digitalizācijas pakalpojumu par maksu piedāvāt arī privātpersonām.

Tuvāko gadu laikā tiks izveidots arī vienots piekļuves punkts visiem LNB (un nākotnē arī citu Latvijas atmiņas iestāžu) digitālajiem resursiem (bibliotēku katalogiem, bibliogrāfiskajām un pilnteksta datubāzēm, digitālajām kolekcijām u.c.), lai lasītājs varētu meklēt vajadzīgo informāciju vienlaicīgi visos avotos.

Vēl tālākā nākotnē plānota e-norēķinu sistēmas ieviešana, lai par nelielu samaksu ikviens no jebkuras vietas varētu piekļūt arī autortiesiski aizsargātajai LNDB daļai. LNB piedāvās iespēju elektroniski pasūtīt tieši tā izdevuma digitalizāciju, kas pasūtītājam nepieciešams.

Diemžēl ekonomiskā krīze valstī iecerēm var pārvilkt svītru — ieplānotā īstenošana atkarīga no tā, vai tiks piešķirts valsts līdzfinansējums Eiropas Reģionālās attīstības fonda (ERAF) aktivitātes “Publiskās pārvaldes elektronisko pakalpojumu un informācijas sistēmu attīstība” apakšaktivitātes “Informācijas sistēmu un elektronisko pakalpojumu attīstība” programmā iesniegtajiem LNDB projektiem.

Taču par to, ka ekonomiskā krīze pārsteigusi Latviju ne reizi vien — lasiet vēsturisko laikrakstu un žurnālu digitalizētajās lappusēs kolekcijā “Periodika”! Kas ziņa — varbūt tieši tur rodami vērtīgi padomi krīzes pārvarēšanai?



(Pēc Latvijas Nacionālās digitālās bibliotēkas “Letonica” biroja vadītājas **Ingas Grīnfeldes** stāstījuma sagatavojusi **Marlēna Krasovska**; Ligitas leviņas foto; kolāža no kolekcijas “Periodika” materiāliem)

<sup>1</sup> Plašāk lasiet: **Žogla, Artūrs**. Vecajai drukai — jauna elpa. *Latvijas Nacionālās digitālās bibliotēkas blogs* [tiešsaiste]. 2008, 8.okt. [skatīts 2009.g. 6.janv.]. Pieejams: <http://lndb.wordpress.com/2008/10/08/vecajai-drukai-jauna-elpa>

<sup>2</sup> **Žogla, Artūrs**. Kā digitālā bibliotēka plāno tērēt Eiropas naudu (II daļa) *Latvijas Nacionālās digitālās bibliotēkas blogs* [tiešsaiste]. 2008, 25.nov. [skatīts 2009.g. 6.janv.]. Pieejams: <http://lndb.wordpress.com/2008/11/25/ka-digitala-biblioteka-plano-teret-eiropas-naudu-ii-dala>