

ՉՈՒԼՅԱՆ ԱՆՆԱ

Հայաստանի ազգային գրադարան

ՄԱՐԿՈՍՅԱՆ ԱՆԻ

Հայաստանում ֆրանսիական համալսարան

**«ՄԵԴԻԱ ԱՐԽԻՎ»՝ ՆՈՐ ՔԱՅԼ ԴԵՊԻ ԱԶԳԱՅԻՆ
ԹՎԱՅԻՆ ԺԱՌԱՆԳՈՒԹՅԱՆ ՊԱՀՊԱՆՈՒՄ**

Բանալի բառեր՝ Թվային ժառանգություն, Մեդիա արխիվ, պարտադիր էլեկտրոնային օրինակ, տեղեկատվության պահպանում, մետատվյալներ, տվյալների հավաքագրում:

Keywords: digital heritage, Media Archive, Electronic Legal Deposit, information preservation, metadata, data collection.

Հայաստանի ազգային գրադարանի նոր գործիքի ստեղծման և կիրառման մասին

Ժամանակակից տեղեկատվական դարաշրջանում, երբ համացանցում ստեղծվող բովանդակությունը ծավալով գերազանցում է ավանդական տպագիր նյութերին, ազգային գրադարանների առջև ծառացել է նոր և հրատապ խնդիր՝ թվային ժառանգության պահպանություն: «Փաստաթղթերի (վավերագրերի) պարտադիր օրինակի մասին» Հայաստանի Հանրապետության օրենքը կարգավորում է բոլոր հրատարակվող տպագիր և որոշ թվային նյութերի օրինակների պարտադիր տրամադրումը որոշված կազմակերպություններին՝ հիմնականում Հայաստանի ազգային գրադարանին: Պարտադիր օրինակի հավաքագրման նպատակն է՝

1. Ազգային գրադարանային հավաքածուի համալրումը
2. տպագիր և թվային ժառանգության պահպանումը և հասանելիությունը
3. գրականության, մամուլի և տեղեկատվության պատմական հիշողության պահպանում:

Մինչև օրս օրենքը հիմնականում վերաբերում է տպագիր նյութերին, ինչը փաստում է ժամանակակից հրատարակչական ոլորտի թվային արտադրանքի հետ անհամապատասխանության և օրենքի ապագա փոփոխությունների անհրաժեշտության մասին: Թվային (էլեկտրոնային) հրապարակումների համար չկա լիարժեք իրավական կարգավորում: Անհրաժեշտ է օրենքի լրացում՝ ապահովելու էլեկտրոնային նյութերի հավաքագրման և պահպանման իրավական կողմը:

Պարտադիր էլեկտրոնային օրինակը (e-legal deposit) գրադարանային գործի ժամանակակից ուղղություն է, որը համապատասխանում է տեխնոլոգիական դարաշրջանի մարտահրավերներին: Եթե ավանդաբար օրենքով պարտադիր պատճենը վերաբերում էր

միայն տպագիր գրականությանը՝ գիրք, պարբերական, թերթ, ապա թվային դարաշրջանում սա այլևս բավարար չէ ազգային ժառանգության ամբողջական պահպանման համար: Էլեկտրոնային պարտադիր օրինակը թույլ է տալիս հավաքել և պահել բոլոր այն նյութերը, որոնք հրապարակվում են միայն թվային ձևաչափով՝ առանց տպագրման:

Դրանք են՝

- առցանց լրատվական կայքերի բովանդակությունը
- թվային ամսագրեր և էլեկտրոնային թերթեր
- կայքերում տեղադրված հետազոտական կամ մշակութային նյութեր
- բլոգներ և առցանց նյութերի այլ ձևաչափեր:

Այս բովանդակությունը հաճախ անհետանում է՝ կապված կայքերի ժամանակավոր բնույթի, հղումների փոփոխության կամ տվյալների թարմացման հետ: Ահա թե ինչու ազգային գրադարաններին անհրաժեշտ է ունենալ համապատասխան տեխնիկական համակարգեր՝ այս արժեքավոր տեղեկատվությունը արխիվացնելու համար:

Հայաստանի ազգային գրադարանը, գիտակցելով այս իրողության կարևորությունը, առաջին անգամ ձեռնամուխ է եղել մշակել և կիրառել հարթակ, որը թույլ է տալիս գրադարանին համալրել իր ֆոնդը էլեկտրոնային մեդիա նյութերով՝ ծառայելով որպես հիմք հետագա պարտադիր էլեկտրոնային օրինակի (e-legal deposit) հավաքագրման համար:

Հայաստանի ազգային գրադարանում «Մեդիա արխիվ»-ը առաջին գործիքն է, որը ստեղծվել է e-legal deposit-ի գաղափարը կյանքի կոչելու համար:

«Մեդիա արխիվ»-ը՝ համակարգչային ծրագիր և գրադարանավարների համար նախատեսված կառավարման հարթակ է, որը մշակվել է Հայաստանում ֆրանսիական համալսարանի (ՀՖՀՀ) երիտասարդ մասնագետի կողմից՝ որպես պրակտիկայի շրջանակում իրականացված նախաձեռնություն: Այս գործիքը մշակվել է Հայաստանի ազգային գրադարանի կարիքներին համապատասխան և ուղղված է մի քանի առանցքային խնդիրների լուծմանը.

- **Թվային մամուլի կուտակում.** հայկական կայքերն ու առցանց մեդիաները մեծ ծավալով հրապարակում են նյութեր, որոնք, չունենալով ֆիզիկական կրիչ, կարող են անվերադարձ կորսվել:
- **Թվային ժառանգության համակարգային պահպանում.** մինչ այժմ գրադարանն ունեցել է սահմանափակ հնարավորություններ համացանցային աղբյուրների գտման և պահպանման համար:
- **Օրենքով պարտադիր պատճենի թվային բաղադրիչի բացը.** մեդիա արխիվը կարող է դառնալ գործնական հիմք էլեկտրոնային պարտադիր պատճենի իրականացման համար:

«Մեդիա արխիվ»-ը թույլ է տալիս գրադարանին ավտոմատացված կերպով հավաքել հայաստանյան մեդիա կայքերի բովանդակությունը՝ պահպանելով նորությունները, հոդվածները և այլ մեդիա նյութեր: Այն առաջարկում է հետևյալ հնարավորությունները.

1. Հավաքագրում. հարթակն իրականացնում է առցանց լրատվամիջոցներից և թիրախային կայքերի բովանդակության ավտոմատացված զետեղում (web scraping) սահման-

ված պարբերականությամբ՝ նախապես նշված դաշտերից (օրինակ՝ վերնագիր, բովանդակություն, հեղինակ, ամսաթիվ): Մա հնարավոր է դարձնում գրադարանին՝ առանց ձեռքով մուտքագրման, ավտոմատ կերպով հավաքել համապարփակ տեղեկատվություն՝ ստանալով բազմալեզու ու բազմաբովանդակ բովանդակության շտեմարան:

2. Արխիվացում. հավաքագրված նյութերը պահվում են գրադարանի ներքին սերվերներում, համապատասխան մետատվյալներով՝ ապագայում որոնման, ներկայացման և վերամշակման նպատակով: Ներառվում են նաև տվյալների պահպանումն ըստ ISO ստանդարտների՝ իրավական պաշտպանության ու վավերականության ապահովման նպատակով:

3. Մատչելիություն գրադարանավարների համար. գրադարանավարի համար նախատեսված է պարզ և կառավարելի հարթակ՝ հավաքագրված նյութերի դիտման, խմբագրման և որոնման համար: Հարթակը թույլ է տալիս.

- վերահսկել հավաքագրման գործընթացը,
- պահպանել բնօրինակ աղբյուրների URL-ները և ստեղծման ամսաթվերը,
- խմբագրել և կարգավորել մետատվյալները,
- դիտել տարբեր ժամանակաշրջանների հավաքածուները,
- ստեղծել գրառումներ՝ MARC կամ այլ գրադարանային ստանդարտներով:

Գրադարանի տարբեր բաժինների աշխատակիցները կօգտագործեն հարթակը՝ պարբերաբար համալրելու, վերահսկելու և պահպանելու էլեկտրոնային աղբյուրներից հավաքված նյութերը: Նյութերը կմշակվեն՝ դասակարգելով ըստ ամսաթվի, աղբյուրի, թեմատիկայի/խորագրի և պահպանման նպատակով՝ ըստ ձևաչափերի՝ համաձայն գրադարանի թվային արխիվի պահանջների:

«Մեդիա արխիվ»-ը ոչ միայն տեխնոլոգիական լուծում է, այլև հայկական թվային ժառանգության պահպանման հենք, որը ծառայում է Գրադարանի զարգացման հայեցակարգի մի շարք ռազմավարական նպատակների՝ նպաստելով ոչ միայն գրադարանի աշխատանքին, այլև՝

- գիտահետազոտական և մեդիա ուսումնասիրությունների համար աղբյուրների ստեղծմանը,
- մշակութային ժառանգության վերականգնման աշխատանքներին,
- ազգային թվային ժառանգության հավաքագրմանը և պահպանմանը՝ ապագա սերունդների համար,
- իրավական դաշտի լրացմանը՝ նախապատրաստելով տեխնոլոգիական հենք e-legal deposit-ի (Ազգային գրադարանային էլեկտրոնային հավաքածու) համար,
- գրադարանավարների մասնագիտական կարողությունների և հմտությունների զարգացմանը՝ նորարար գործիքների կիրառմամբ,
- մատենագիտական գրառումների արագ և ճշգրիտ ստեղծմանը:

Հայաստանի տեղեկատվական դաշտը հատկապես վերջին տարիներին զգալիորեն տեղափոխվել է առցանց հարթակներ: Այսօր ավելի շատ տեղեկություն է ստեղծվում և տարածվում համացանցում, քան թղթային ձևաչափով: Այս նյութերի բացակայությունը գրադարանի ազգային հավաքածուում ստեղծում է «դատարկ տարածք» ապագա

սերունդների համար, ովքեր կուզենան ուսումնասիրել այսօրվա մամուլը, հանրային քննարկումները, իրադարձությունները և հանրային տրամադրությունները:

Նախագծի հաջորդ փուլում հնարավոր է համակարգի ինտեգրումը գրադարանի գրացուցակային համակարգերի հետ, որոնման գործիքների բարելավում, ինչպես նաև՝ մեդիա արխիվի հանրային տեսանելիության ապահովում՝ հետազոտողներին ու ընթերցողներին մատչելի դարձնելու համար:

Նախաձեռնության տեխնիկական իրականացում

«Մեդիա արխիվ» հարթակը նախագծվել և իրականացվել է՝ ամբողջությամբ համապատասխանեցված Հայաստանի ազգային գրադարանի թվային ժառանգության կառավարման ռազմավարական նպատակներին: Համակարգը միավորում է ժամանակակից ծրագրային գործիքներ, ավտոմատացման տրամաբանություն և տվյալների կազմակերպման հստակ կառուցվածք, ինչը հնարավորություն է տալիս ամբողջականորեն թվայնացնել ու պահպանել առցանց մեդիա բովանդակությունը՝ առանց մարդկային միջամտության:

Հարթակի կառուցվածքը բաղկացած է մի քանի փոխկապակցված տեխնիկական բաղադրիչներից՝ սկսած տվյալների ավտոմատ հավաքագրումից, մինչև տվյալների վերամշակում և ինտերակտիվ վեբ ինտերֆեյս՝ նախատեսված գրադարանավարների և հարակից ոլորտների մասնագետների համար:

Համակարգի աշխատանքի քայլային նկարագրություն

1. Տվյալների ավտոմատ հավաքագրում (Web Scraping)

Համակարգի առաջին և առանցքային փուլը մեդիա բովանդակության ավտոմատացված հավաքագրումն է: Այս նպատակով մշակվել են մի քանի կայքերի համար հարմարեցված web scraper-ներ՝ կառուցված Python ծրագրավորման լեզվով: Օգտագործելով requests գրադարանը՝ համակարգը հաղորդակցվում է կայքերի հետ և բեռնում էջերի HTML կառուցվածքը, որն այնուհետև վերլուծվում է BeautifulSoup գրադարանի միջոցով՝ որոնելով նախապես նշված դաշտերը՝ վերնագիր, բովանդակություն, հրապարակման ամսաթիվ, աղբյուրային հղում, հեղինակային տողեր և YouTube տեսանյութի հղումներ: Scraper-ները աշխատում են պարբերական ռեժիմով՝ ըստ նախապես սահմանված ժամանակացույցի:

2. Տվյալների վերամշակում և ֆիլտրում

Հավաքագրված հոդվածները նախքան պահպանման փուլ տեղափոխվելը անցնում են վերլուծության և մաքրման գործընթացներ: Այս փուլում իրականացվում է՝

- կրկնվող գրառումների հեռացում

- ոչ լիարժեք կամ դատարկ բովանդակության ֆիլտրում
- ամսաթվերի միասնականացում ISO 8601 ստանդարտի համաձայն՝ հետագա որոնման դյուրացման նպատակով
- հոդվածների դասակարգում ըստ հիմնական տեղեկատվական հատկանիշների:
Այս բոլոր քայլերն իրականացվում են pandas գրադարանի միջոցով՝ ապահովելով տվյալների կառուցվածքային ամբողջականություն:

3. Պահպանման կառուցվածք

Մաքրված և ստանդարտացված տվյալները սկզբում պահպանվում են CSV ֆորմատով՝ ապահովելու համար պարզ և ընթերցելի տվյալների բազա: Նույն տվյալները հետագայում վերափոխվում են JSON ֆորմատով, որն ավելի հարմար է վեբ ինտերֆեյսի միջոցով ներկայացման և բեռնման համար: Հաշվի առնելով հավաքագրված տվյալների ծավալի աճը՝ նախատեսվում է համակարգի տվյալների կառավարման բաղադրիչի տեղափոխումը PostgreSQL տվյալների բազայի, ինչը թույլ կտա բարելավել արդյունավետությունը, որոնման արագությունը և պահպանման հուսալիությունը:

4. Վեբ ինտերֆեյս և հասանելիություն

Հարթակի ինտերֆեյսը կառուցված է Flask web framework-ի միջոցով՝ որպես backend լուծում, որն ապահովում է տվյալների մատչելիությունը frontend-ի համար: Տվյալների ցուցադրությունն իրականացվում է HTML, CSS և JavaScript-ի համադրությամբ՝ ապահովելով պարզ, ինտուիտիվ և հարմարավետ միջավայր գրադարանավարների և աշխատակիցների համար: Հարթակը հնարավորություն է տալիս՝

- գտել նորությունները ըստ ամսաթվի, աղբյուրի կամ խորագրի
- խմբավորել տարբեր աղբյուրներից ստացված հոդվածները՝ ապահովելով բազմազանություն
- առաջնահերթ ցուցադրել վերջին ավելացված հոդվածները
- դիտել սկզբնաղբյուրի հղումները և համապատասխան մետատվյալները:

Հարթակը մշակվել է այնպես, որ չպահանջի տեխնիկական խոր գիտելիքներ օգտագործողից, ինչը դարձնում է այն լիարժեք գործիք գրադարանային աշխատակիցների համար:

Տեխնիկական բարելավման առաջարկներ

Առաջիկա փուլում նախատեսվում է հարթակի հնարավորությունների ընդլայնում՝ ներառելով OCR (Optical Character Recognition) տեխնոլոգիա: Սրանով կդառնա հնարավոր ոչ միայն առցանց հրապարակված հոդվածների հավաքագրումը, այլև սկանավորված թղթային նյութերի ավտոմատ ընթերցումը և թվայնացումը: OCR տեխնոլոգիայի ինտեգրումը թույլ կտա՝

- գրադարանը թվայնացնի նախկինում միայն թղթային եղած արխիվային նյութերը,

- դրանք վերածի որոնելի տեքստային բազայի,
- միացնի մեկ միասնական հարթակում՝ ինչպես թվային, այնպես էլ թվայնացված ժառանգությունը:

Տեխնիկական բարելավումը էականորեն կավելացնի գրադարանի թվային հավաքածուի ծավալը՝ դարձնելով այն համապարփակ և մատչելի ապագա սերունդների ու հետազոտողների համար:

«Մեդիա արխիվ»-ի ստեղծումը Հայաստանի ազգային գրադարանի համար կարևոր քայլ է: Այն դառնում է առաջին գործիքը, որը թույլ է տալիս համակարգված ձևով հավաքել և արխիվացնել առցանց մեդիա բովանդակությունը՝ որպես ազգային հիշողության կարևոր բաղադրիչ: Մենք համոզված ենք, որ այս նախաձեռնությունը երկարաժամկետ ազդեցություն կունենա ոչ միայն գրադարանի զարգացման, այլև երկրի թվային մշակութային ժառանգության պահպանման վրա:

SUMMARY

The National Library of Armenia (NLA) has launched Media Archive, a pioneering digital platform designed to collect and preserve online media content as part of Armenia's national heritage. Developed by a young professional of Université Française en Arménie (UFAR), the system automates the collection of content from Armenian news websites, capturing articles, blogs, and other born-digital materials that are at risk of disappearing. The platform addresses a critical gap in Armenia's legal deposit system, which currently applies mostly to printed materials.

Using web scraping technologies, data is collected, cleaned, and archived with metadata, ensuring long-term access and legal authenticity. A user-friendly interface enables librarians to manage and search the archive effectively. Future enhancements include OCR integration for digitized historical materials and catalog system integration.

Media Archive establishes the technological foundation for an electronic legal deposit system and reflects NLA's commitment to safeguarding Armenia's digital cultural memory for future generations and supporting scientific research and cultural preservation.