

Az erdélyi múlt digitalizálása: kihívások és sikerek

Digitizing the Transylvanian Past: Challenges and Successes

Ferenczi Szilárd 
szilard.ferenczi@gmail.com
Iskola Alapítvány, Digitéka
projektvezető, BBTE Kolozsvár
társult oktató, történész

Beérkezett: 2025.11.17.
Elfogadva: 2025.11.17.
Publikálva: 2025.12.22.

Az Erdélyi Digitális Tudománytár (Digitéka)¹ 2018 végén indult, a kolozsvári Iskola Alapítvány kezdeményezte és működteti azzal a céllal, hogy teljeskörűen elérhetővé tegye az erdélyi magyar sajtóállományt. Közzolgálatot betöltve erdélyi magyar nyelvű digitalizált sajtótermékeket, könyveket és adatbázisokat tesz ingyenesen elérhetővé. Tartalmát kezdetben az Arcanummal és az Erdélyi Magyar Adatbankkal együttműködésben gyarapította, 2022-től viszont már saját szkennelő berendezés használatával és a Kolozsvári Központi Egyetemi Könyvtárral² kötött megállapodás keretében fejleszti. A platform a saját digitalizált anyagok mellett hiperhivatkozásokkal más erdélyi digitalizált tartalmakhoz is hozzáférést biztosít, gyűjtőkönyvtári jelleggel. Működtetői 2025-re részben mesterséges intelligencia segítségével automatizálták a munkafolyamatot, jelentősen felgyorsítva a tartalomfeldolgozást, és évi mintegy 250 ezer oldal digitalizálására építették ki a kapacitást, a honlapon jelenleg 1 701 403 oldal digitalizált sajtó és könyv érhető el. A honlapnak havi 300-500 visszatérő és 2000-3000 egyszeri felhasználója van, legtöbbször Romániából és Magyarországról.

digitalizálás, erdélyi magyar sajtó, virtuális könyvtár, intézményi együttműködések, értékmentés

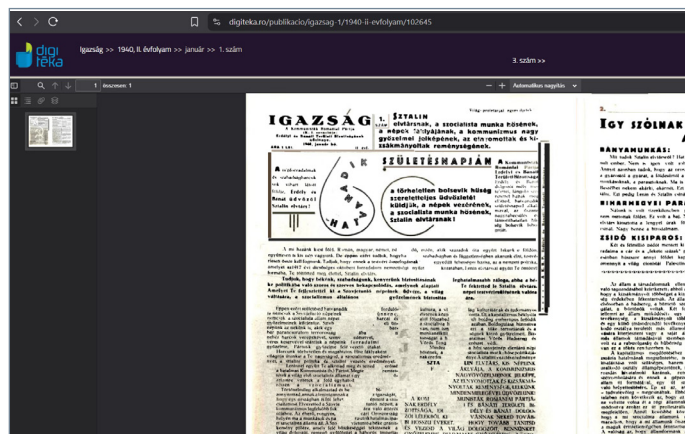
Digitéka was launched at the end of 2018, initiated and operated by the School Foundation in Cluj-Napoca/Kolozsvár. It provides free access to digitized press materials, books, and databases in Hungarian from Transylvania, and its goal is to make the entire Transylvanian Hungarian press collection fully accessible. Initially, it expanded its content in cooperation with Arcanum and the Transylvanian Databank, but since 2022 it has been developing through its own scanning equipment and by striking an agreement with the Central University Library of Cluj. The platform contains not only its own digitized materials but also provides access to other Transylvanian digitized content through hyperlinks, functioning as a collecting library. By 2025, it has automated the workflow partly with the help of artificial intelligence, significantly accelerating content processing. It has built up a capacity to digitize approximately 250,000 pages per year. The website has 300-500 returning, and 2-3,000 one-time users per month, primarily from Romania and Hungary.

digitization, transylvanian hungarian press, virtual library, institutional collaboration, heritage preservation

Cite as / Így hivatkozd: Ferenczi, Sz. (2025) Az erdélyi múlt digitalizálása: kihívások és sikerek / Digitizing the Transylvanian Past: Challenges and Successes, Central European Library and Information Science Review (CELISR), 2(4), p. 481–489.

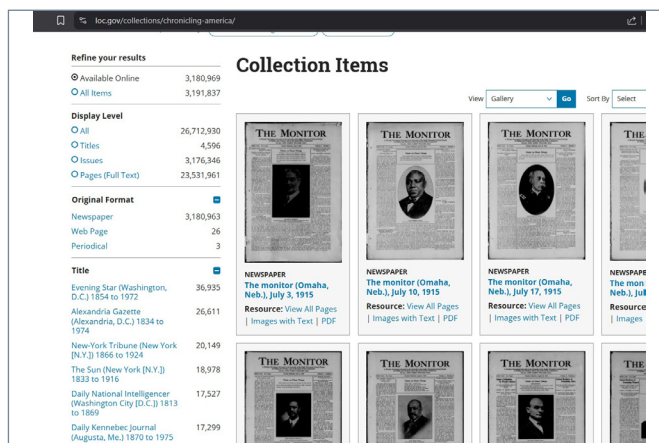
ket, kiállításokat vállalt és bonyolított le. A program keretében Digitéka névvel sajtó- és könyvdigitalizálást is indítottunk, amelynek vezetésével engem bíztak meg, és amelynek folytatása mellett döntöttünk 2022-ben.

Alapítványunk saját gyűjteménnyel nem rendelkezett, ezért első lépésként együttműködést kötöttünk az Arcanummal, az Adatbankkal és a Minerva Művelődési Egyesülettel. A Digitéka kezdeti elérhető tartalmai között volt a Minervával közösen digitalizált Szabadság⁹, Világosság¹⁰, Igazság¹¹ és a Romániai Magyar Szó/Előre¹² hírlapok.



2. ábra: Az 1939–40-ben illegálisban, Kolozsváron megjelenő kommunista Igazság ekkor még alkalmi lap (1945-től napilap) egyik ritka száma, 1940. január

Az erdélyi magyar sajtóállomány anyaga 200-230 évet ölel fel, amelyhez ha hozzáadjuk a szépirodalmi és tudományos könyveket is, valamint az egyéb nyomtatott és bekötött anyagokat (kalendáriumok, évkönyvek, jelentések, értesítők, aprónyomtatványok, szabályrendeletek, városi, megyei és egyéb kimutatók stb.), a digitalizálandó mennyiség több tíz millió oldalt tenne ki. A feladat nagyságára való tekintettel, valamint a várható kutatói igényekre gondolva a teljes digitalizáció helyett csapatunk úgy döntött, hogy a saját anyagainkon túl hiperhivatkozással más honlapok erdélyi tartalmához is hozzáférést kínálunk a Digitékán. Ez a funkció egybegyűjti a világ számos honlapján (a MEK, az EPA, a Hungaricana, az Arcanum, a magyarországi egyetemi és akadémiai repozitóriumok, a Kolozsvári Központi Egyetemi Könyvtár digitális gyűjteménye, más romániai, amerikai, osztrák stb.) tárolt erdélyi magyar nyelvű anyagokat.¹³ A kialakult gyűjtőkönyvtári jelleg hátránya, hogy a honlapon történő kereséskor az eredmények között ezek a tartalmak nem jelennek meg.

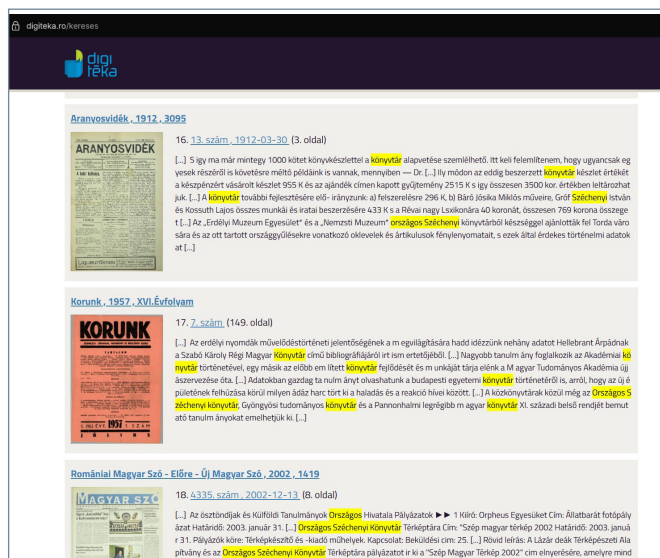


3. ábra: A Digitéka Külső hivatkozások kategóriája a Sajtó aloldalon

„Az erdélyi magyar sajtóállomány anyaga 200-230 évet ölel fel ...”

3. Keresés

Az oldal indulásakor egy olyan műszaki megoldást használtunk, amely a tartalmat jól megjelenítette, de a pontos kifejezésre történő keresésre mindössze ropant időigényes megoldást tudott felkínálni. A rendszerre néhány év alatt – s a keresés lassú természetére is – egyre inkább fogyatékként tekintettünk. 2021-ben váltottunk egy új, integratívabb felületre, amelybe beépítettük az ElasticSearch¹⁴ nyílt forráskódú keresőmotort is, amely a keresési eredmények listázásakor a lapszámok címloldalát és a szöveggörnyezetet is megjeleníti.



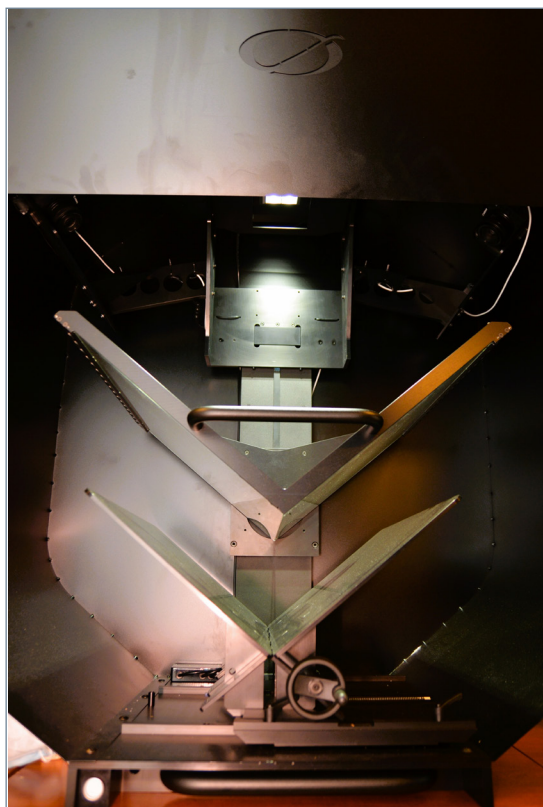
4. ábra: Az „országos széchényi könyvtár” (így!) keresőkifejezés találatai közül három, a lapszám címloldalát és a szöveggörnyezet megjelenítésével

A számítástechnikai újítással egy időben, a gyerekcipő lerúgva két másik lényeges újítást is végrehajtottunk: a 2022. év elején megvásároltunk egy Qidenus Smart 4.0¹⁵ (A1-es modell) szkennelő berendezést, továbbá a raktár előzetes felmérését követően a Kolozsvári Központi Egyetemi Könyvtárral kétoldalú megállapodást kötöttünk 880 000 oldal sajtótermék digitalizálására, amelyet mindkét fél megjeleníthet a saját honlapján. A címlajstrom összeállításánál a szűrés több szempont szerint történt: elsősorban olyan erdélyi lapokat kerestem, amelyekkel a könyvtár rendelkezik és fontosságuk sajtótörténeti, nyilvánossági, a maguk idejében közösségerősítő megfontolásokból jelentős (ez utóbbi szempont főként a két világháború közötti lapokra érvényes), másrészt figyelembe vettem a szórását is, és minél átfogóbb lefedettségre törekedtem mind a műfajt (politikai-közéleti, irodalmi, szórakoztató és szaksajtó), mind a régiókat illetően, harmadrészt azokat a címeket válogattam össze, amelyek nem voltak még online hozzáférhetőek. Ez utóbbi szempont annyiban csorbult az elmúlt években, hogy az Arcanum és az OSZK a maguk nagyüzemi képességeivel egyik-másik címmel megelőztek, oda kikerültek, mi pedig a magunk szerény tempójával ezután tesszük közzé. A „megelőzést” nem a „verseny” értelemben használom, örülök annak, hogy vannak sorozatok, amelyek több példányban fellelhetőek a hálón, így nagyobb esélyt látok a sajtótermékek online túlélésére.

4. Eszköztár

A kolozsvári egyetemi könyvtár digitalizáló osztálya – ahol öt-hat állandó munkatárs dolgozik – a 2000-es évektől kezdődően a román periodikák és könyvek mellett számos magyar nyelvű anyagot feldolgozott és megosztott a honlapján.¹⁶

„... elsősorban olyan erdélyi lapokat kerestem, amelyekkel a könyvtár rendelkezik és fontosságuk sajtótörténeti, nyilvánossági, a maguk idejében közösségerősítő megfontolásokból jelentős ...”



5. ábra: A Qidenus Smart 4.0 készülék a könyvtárban

A Qidenus szkennert üzembehelyezésétől Bodnár Róbert osztályvezető felügyelete alatt két munkatárs használja. A létrejött JPG-fájlokon lefuttatjuk az optikai karakterfelismerést (OCR), a létrejött PDF-fájlok kerülnek a tömeges fájl-fogadásra és megjelenítésre is alkalmas Digitéka és a könyvtár oldalára. A munka a kezdeti lassú indulást követően felgyorsult. Napjainkra éves átlagban 250 ezer oldalnyi sajtótermék szkennelését végezhetjük el, de különféle okok miatt a PDF-feldolgozás tekintetében nem tudtunk lépést tartani. A szkennelés kiterjesztése elhozta azt az elkerülhetetlen állapotot, hogy jóval nagyobb JPG-állomány gyűlt össze 2024 év végére, mint amennyi PDF-et adott időkeretben elkészíthettünk, és a szerezhető támogatások¹⁷ szűkebb köre is immár kevésbé tette lehetővé külső megbízottak bevonását. Segítséget jelent, hogy a kolozsvári Babeş–Bolyai Tudományegyetem szakmai gyakorlatot végző történészhallgatóit be tudtuk vonni kisebb feladatokkal, de így sem lehetett érdemben lépést tartani a szkennelés ütemével.

2024 végén sürgőssé vált egy tömeges OCR-ezésre alkalmas megoldás. A megoldást az ELTE Digitális Örökség Nemzeti Laboratóriumával (DH-LAB)¹⁸ kötött partneri szerződés jelentette. Az együttműködés során a JPG-fájlokon az erre is alkalmas mesterségesintelligencia-alapú fejlesztésükkel tömeges OCR-ezést tudnak futtatni, az így nyert nagy méretű fájlokat már csak darabolni (PDFSmart Visual)¹⁹ és ellenőrizni kell (kézzel, egyenként). A munkafolyamat még így is viszonylag lassúnak bizonyult, különösen az átnevezés, ezért 2025 októberében a Google-fejlesztésű Gemini AI²⁰ bevonásával automatizált átnevező Python-alapú programot készítettünk, valamint a Google Console API²¹ integrálásával 85-88 százalékos pontosságú eredményt elérve, immár gyorsabb ütemben folyik a munka. Időnként, kisebb oldalmennyiségű hiánypótlások esetében még manuálisan futtatunk OCR-ezést is az ABBYY Finereader 15²² eszközzel.

„Napjainkra éves átlagban 250 ezer oldalnyi sajtótermék szkennelését végezhetjük el ...”



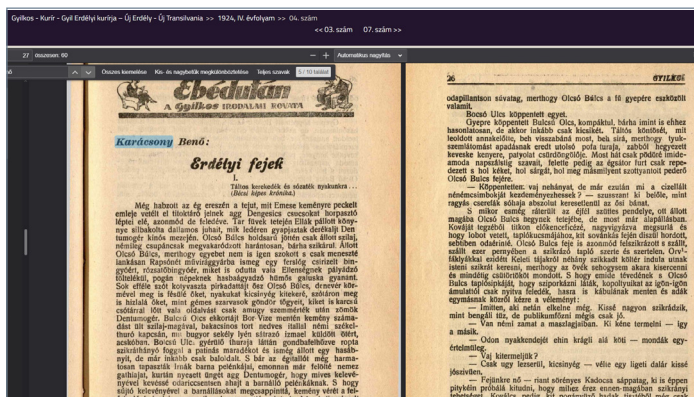
6. ábra: Optikai karakterfelismerés futtatása a kolozsvári Ujság napilap 1904-es évfolyamán az ABBYY Finereaderrel

Ezt az új és egyelőre akadálymentesen működő tartalomgyarapítási tempót kívánjuk tartani a továbbiakban, amelyek egyik feltétele, hogy a könyvtárral meg tudjuk hosszabbítani az együttműködést. Hosszú távú terveink között szerepel újabb címek és sorozatok bevonása, így idővel a teljes erdélyi sajtóállomány elérhetővé válna.

5. Tartalmi megfontolások és visszakapcsolás

A digitalizálási munka során két történelmi korszak, összesen 80 év sajtóanyaga kap kiemelkedő figyelmet: a dualizmus kora (1867–1918) és a két világháború közötti negyedszázadnak az átmenet éveivel kiegészített időszaka (1919–1948). A választásunkat indokolta, hogy ez az a két időszak, amelyben kibontakozott a regionális, helyi híreket közlő, nagy műfaji változatosságú hírlapírás, online elérésük azonban csekély volt. A nagyváradi Szabadságnál²³ és a Nagyvárad Naplónál dolgozott a XIX. és XX. század fordulóján Ady Endre, Nagy Endre, Krúdy Gyula és Bíró Lajos, a Gyulafehérvári Hírlapban közölte első novelláit Karácsony Benő 1908–1909-ben, és Désen, a Vármegyei Közéletnél volt húszévesen szépirodalmi munkatárs Harsányi Zsolt.

Az 1948 utáni időkből a Román Kommunista Párt központi és megyei szervezetei tartottak fenn egy-egy magyar nyelvű lapot Bukarestben és a magyar lakossággal is rendelkező megyékben, ezeknek országos szórását biztosítottak, de tartalmukban a helyi hírek jelentősége eltörpült a nagy felületeket elfoglaló központi hírek és a propaganda mellett. Ebből az időszakból első helyen az irodalmi, közművelő lapok jelentik a valódi értéket.



7. ábra: Karácsony Benő abszurd „novellája” az 1924. január 22-ikei (IV. évf., 4. szám) Gyilkos-ban (kolozsvári, Győri Illés István lapja). Az Erdélyi fejek című novella annyiban ritkaság, hogy a szerző életében és post-mortem megjelent köteteiből egyaránt hiányzik

„A digitalizálási munka során két történelmi korszak, összesen 80 év sajtóanyaga kap kiemelkedő figyelmet: a dualizmus kora (1867–1918) és a két világháború közötti negyedszázadnak az átmenet éveivel kiegészített időszaka (1919–1948).”

A Kolozsváron megjelent Magyar Polgár/Kolozsvár és Ellenzék az országrész két legfontosabb lapja volt a XIX. és XX. század fordulóján, Budapesten is sokan jártatták mindkettőt, jelentős közvélemény-formáló erejüknel fogva, mivel előbbi az erdélyi Szabadelvű Párt, utóbbi a Függetlenségi és 48-as Párt erdélyi szárnyának programadó felülete volt. Akárcsak Magyarországon, Erdélyben (és a Partiumban, valamint a Bánságban) is mind a dualizmus korában, mind a két világháború között a hírlapok szerkesztőségében dolgozott az írók és költők javarésze, a kolozsváriakéban a két világháború között szinte mindenki, aki fiatalon az Erdélyi Helikon írói közösségéhez tartozott. Ezzel egy időben a Szövetkezeti Értesítő és a Szövetkezés időszaki kiadványok a kisebbségbe szorult erdélyi magyar termelők terményértékesítő hálózatát fogták össze és teremtettek számukra kommunikációs platformot, képviseletet, érdekérvényesítő közösséget és erőt az 1920–30-as években. A digitalizált újdonságokról az Iskola Alapítvány Facebook-oldalán²⁴ is beszámolunk időszakosan.



8. ábra: A kézdivásárhelyi Székelyföld hetilap feltöltéséről szóló Facebook-bejegyzés

„... főleg romániai és magyarországi felhasználók töltönek időt a honlapon, de sokan keresnek fel Észak-Amerikából, Izraelből, Szlovákiából, valamint Nyugat-Európából, és akadnak ausztrál, japán és dél-amerikai látogatók is ...”

A honlap az elmúlt hét évben egy parányi, de annál lelkesebb felhasználói kör számára jelenti dobogós felületét az online hozzáférhető erdélyi digitalizált sajtóanyagoknak. Érdeklődő, javaslatokat is megfogalmazó feedbacket kapunk, jelenleg havi átlagban 300–500 visszatérő felhasználónk van, és 2000–3000 egyszeri. A Google Analytics statisztikái szerint főleg romániai és magyarországi felhasználók töltönek időt a honlapon, de sokan keresnek fel Észak-Amerikából, Izraelből, Szlovákiából, valamint Nyugat-Európából, és akadnak ausztrál, japán és dél-amerikai látogatók is – valószínűleg magyarul értő kutatók vagy érdeklődők. T. Szabó Levente irodalomtörténész szerint a Digitéka erőssége, hogy a felhasználó a beépített szűrő szűrővel az erdélyi vidékiséghez, a régiók irodalmához és specifikus hírlapíráshoz pillanatok alatt hozzáférhet, s a főoldalra rátekintve rögtön tudja, mely lap melyik városhoz vagy tájegységhez köthető. Sokan kiemelik a felhasználóbarát szerkezetet, a gyors és akadálymentes görgethető felületet, másokat a lap történeti részletek,²⁵ a csak itt elérhető és hiánypótló tartalom ragadja meg (többen is megköszönték 2025 októberében a kézdivásárhelyi Székelyföld²⁶ hírlap feltöltését). Időszakosan kapunk felkereséseket szerkesztőségektől vagy egyes rég lezárt sorozatok jogutód őrzőitől azzal a céllal, hogy felajánlják a lapjukat digitalizálásra, illetve hiánypótlást kínálnak a már megjelent tartalomhoz.

6. Jelentőség

Az utóbbi években látványosan nő az érdeklődés a digitális bölcsészet iránt, egyre több egyetemi kurzus, konferencia és műhely foglalkozik ezzel a témával. Meghívásokat kapok, melyből egyértelműen érződik, hogy érdeklődés övezi a projektet, és a gyakorlati tapasztalatainkra is sokan kíváncsiak, de még így sem

tudom felmérni kicsi és szerény költségvetésből működő műhelyünk jelentőségét, eltúlozni pedig nincs szándékomban. Mint gyakorló kutató és hivatásszerű digitalizáló fontosnak érzem – főként erdélyi viszonylatban –, felismerem helyét és létjogát az online tudományos segédeszközök sorában, de mindössze remélni tudom, hogy parányi szerszámosládája és egyre bővülő tartalma más platformokhoz mérten kielégíti a látogatók igényeit. A kutatótársaknak és alkalmi olvasóknak üzenem: használják, nekik készül, és ha akad jobbító javaslatuk, írjanak bátran.

A közvetlenebb hangnem végén, hivatalos zárszót keresve, álljon itt az a két mondat, amelyeket a Digitéka működéséről és működtetéséről krédóként megfogalmazhatok: az Iskola Alapítvány a honlap teljes tartalmát szabad felhasználással a magyar tudományosság szolgálatába állítja, a feldolgozással és közzététellel egyszersmind értékmentést is vállal. A közgyűjteményekben található, részlegesen sérült szálas anyag így nem csupán kutathatóvá válik, hanem az erdélyi magyar írott örökség is konzerválódik, a továbbiakban nem szükséges kikérni és lapozni (az állomány nem sérül tovább), s az online tér által immár bárkihez pár klikkeléssel közel kerül az erdélyi múlt, és hosszú távon a digitálisra egyre inkább átálló utókor számára válik a honlap hovatovább emlékeztető felületté.

Jegyzetek

¹ Digitéka: <https://digiteka.ro>

² Kolozsvári „Lucian Blaga” Központi Egyetemi Könyvtár: <https://bcucluj.ro>

³ Bartos Elekes Zsombor: Mappae Comitatum Transylvaniae című, az Iskola Alapítvány 2020-as kiadványának kísérő applikációja: <https://iskolaalapitvany.ro/konyvek/mappae-comitatum-transylvaniae>

⁴ Bárdi Nándor az ELTE Társadalomtudományi Kutatóközpont Kisebbségkutató Intézetének vezetője, valamint az Arcanum és az Erdélyi Adatbank tartalomfejlesztője.

⁵ Magyar Elektromos Elektronikus Könyvtár: <https://mek.oszk.hu>

⁶ Elektronikus Periodika Adatbázis/Archívum: <https://epa.oszk.hu>

⁷ Adatbank. Elérhető: <https://adatbank.ro> (Utolsó elérés: 2025. 11. 18.). Az Adatbank 2025-ben új felületet kap <https://uj.adatbank.ro> (utolsó Utolsó elérés: 2025. 11. 18.), átmenetileg Átmenetileg párhuzamosan mindkét honlap működik.

⁸ Dokumentum- és tévéfilmek: <https://ezer100.ro>; könyvkiadás: <https://iskolaalapitvany.ro/kiado>; népességi és intézményi statisztikák: <https://statistikak.erdelystat.ro>; Elmúlt jelen, avagy a magyar kisebbség az államszocializmus korszakában, könyv és vándorkiállítás: <https://www.elmultjelen.ro>; stb.

⁹ <https://digiteka.ro/publikacio/szabadsag>

¹⁰ <https://digiteka.ro/publikacio/vilagossag>

¹¹ <https://digiteka.ro/publikacio/igazsag-1>

¹² <https://digiteka.ro/publikacio/romaniai-magyar-szo-elore-uj-magyar-szo>

¹³ „Külső hivatkozások” kategória a sajtó, könyv, térkép, repertórium, adatbázis és szövegtár aloldalakon egyaránt található.

¹⁴ <https://www.elastic.co>

¹⁵ <https://qidenus.com/portfolio/qidenus-smart-book-scanner>

¹⁶ Digitális Könyvtár (Biblioteca Digitală). Elérhető: <https://dspace.bcuculuj.ro> (Utolsó elérés: 2025. 11. 18.)

¹⁷ A Digitéka működtetését az RMDSz RMDSZ és a Nemzeti Együttműködési Alap biztosítja 2022 óta.

¹⁸ <https://dh-lab.hu>

¹⁹ <https://pdfsam.org/pdfsam-visual>

²⁰ <https://gemini.google.com/app>

²¹ <https://console.cloud.google.com>

²² <https://pdf.abbyy.com/finereader-pdf>

²³ A Szabadságot még nem szereztük be, de bízunk benne, hogy rövidesen sikerül.

²⁴ <https://www.facebook.com/IskolaAlapitvany>

²⁵ Ide legtöbbször a Romániai Magyar magyar Irodalmi irodalmi Lexikon lexikon online szócikkét idézem be <https://kriterion.ro/lexikon/lexikon>, illetve, ha ott nincs, keresek más forrást, vagy magam írok rövidet.

²⁶ <https://digiteka.ro/publikacio/szekelyfold-1>

