

dhupla

digital humanities platform / digitális bölcsészeti platform

szövegkiadások entitástár kreatív műhely a dhupláról

Keresés

Digitális forráskiadás

Elkezdődött Keresztury Dezső levelezésének közzététele!

A nemzeti könyvtár legterjedelmesebb levelezése az alkotó közel 70 évnél aktív munkásságát és rendkívül összetett kapcsolatrendszerét mutatja meg.

A korpusz elsőként kiemelt tematikus egysége Keresztury Egy József festővel folytatott levelezése, amely elérhetővé vált felületünkön.



A megújulás útján A nemzeti könyvtár tartalomszolgáltatásai 2.

A közgyűjtemények digitális szolgáltatásainak fejlődése egyre inkább túlmutat a klasszikus dokumentumalapú hozzáféréseken, és olyan komplex rendszerek irányába mozdul, amelyek a tartalmak értelmezett, újrafelhasználható és kutatható formában való publikálását célozzák.

E szemlélet egyik legjelentősebb hazai példája az OSZK Digitális Bölcsészeti Központja (DBK) által fejlesztett Digitális Bölcsészeti Platform, a dhUpla.

A platform célja, hogy a közgyűjtemények kéziratos forrásait korszerű, tudományos és ismeretterjesztő célokat egyaránt szolgáló formában tegye elérhetővé.

A szolgáltatás jelenleg négy fő komponensre épül. A szövegkiadások felület biztosítja az egyes dokumentumgyűjtemények tudományos igényű digitális publikálását. Az oldal jelenleg három fő típusba sorolja ezeket: digitális forráskiadások, például Keresztury Dezső levelezése, Móricz Zsigmond Tükör című naplói, A Tett folyóirat megjelent számai, a OSZK Vershaza kiállítás anyagai; digitális kritikai kiadások, például Móricz Zsigmond levelezésének részletes, jegyzetelt változata; digitális népszerűsítő kiadás Gróf Teleki László tanácsadása a nevelésről művéhez.

A szolgáltatásban minden szövegkiadás azonos szerkezettel jelenik meg, ahol egy részletes ismertető mutatja be a kézirat jellegét, történeti kontextusát, valamint az átíráshoz szükséges információkat. Egy gyűjteményt a kiadás szempontjából összetartozó digitális objektumok összessége jelenti. Ezek egy listában érhetők el, amelyeket a felhasználó több szempont alapján rendezhet, kereshet vagy szűrhet facetták alkalmazásával. Az egyes dokumentumok megnyitáskor az adott rekord metaadatai egy külön ablakban jeleníthetők meg, mint ahogy a dokumentumhoz kapcsolódó megjegyzések is. Továbbá a

Vershaza

Kölcsey Ferenc: Himnusz – kézirat (1823-01-22)

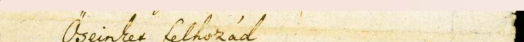
Metaadatok Megjegyzések



Himnusz, a Magyar népvilágos szavazásból.
Cseke. Január. 22d.
1823

Isten áld meg a Magyar
jó kedvű, bölcsőt,
Nyílt feléje védő kart
Ha küzd ellenséggel,
Bál sors a' kit régen tép
Mert rá van asztalodt

Összített felhőzár
Karpat szent bérceire,
Altalad nyert szép hazát
Benedegúzok vére.
"S merre zúgnak hajjai



Kiss József levelezése

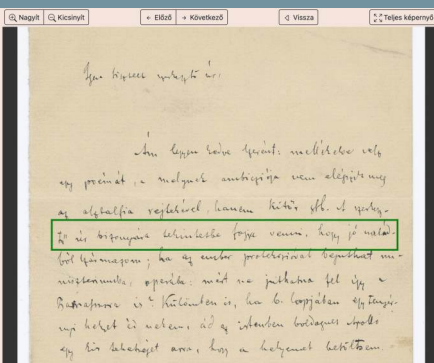
Ignotus – Kiss József (1890-07-10)

Metaadatok



Budapest, 1890. július 10^{én}.

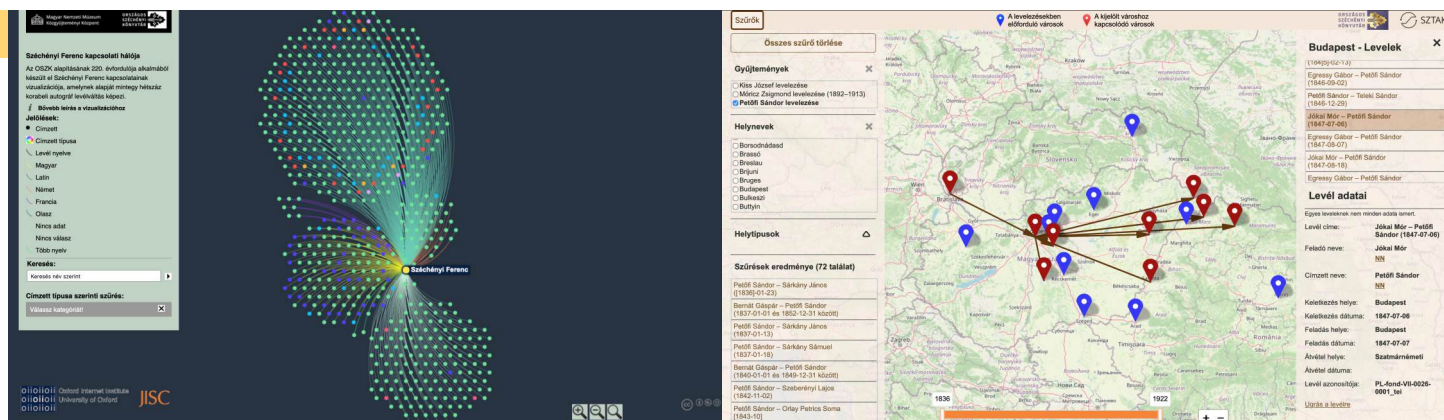
Igen tisztelt szerkesztő úr,
Am legyen kedve szerént: mellékelve veszek egy poémát, amelynek ambíciója nem elégszik meg az asztalra rejtéssel, hanem kitor stb. A szerkesztő úr bizonyára tekintetbe fogja venni, hogy jó család-ból származom; ha az ember protekcióval bejuthat m^l niszteriumba¹⁾, operába: miért ne juthatna fel így a Parnassusra is? Különben²⁾ is, ha b. lapjában egy ténylegnyi helyet ad nekem, ad az istenben boldogult Apolló egy kis tehetséget arra, hogy a helyemet betöltsöm.
Más kéresem is van a szerkesztő úrhoz, a ki-



A platform célja, hogy a közgyűjtemények kéziratos forrásait korszerű, tudományos és ismeretterjesztő célokat egyaránt szolgáló formában tegye elérhetővé

felületen lehetőség van a faksimile és az átírt szöveg egymás melletti megtekintésére, olvasására, egyes gyűjteményeknél a soronkénti összehasonlításra is. Utóbbi esetben az átírt kijelölt sora felett jelenik meg az eredeti kézirat képe, amit egy, a mesterséges intelligencián alapuló megoldás tesz lehetővé.

Az entitástár a szövegkiadásokhoz kapcsolódó névlem-információkat tartalmazza, amely nem csupán egy névmutató, ugyanis az azonosított névlemek



(személynevek, intézménynevek, helynevek, műcímek) hivatkozási pontokká is válnak, amelyek segítenek eligazodni a korpuszban, és lehetőséget adnak a külső névterekhez való kapcsolódásra is. Valamint a szövegkiadásokhoz hasonlóan itt is lehetőség van szabadszavas keresésre és típusok szerinti szűrésre.

A kreatív oldal olyan, az egyes szövegkiadásokhoz szorosan kapcsolódó anyagokat gyűjti egybe, amelyek kiegészítik a digitális szövegkiadásokat. Részben a digitális szöveg előállításánál születnek, részben pedig a kiadások készítésével párhuzamosan zajló kutatómunka eredményei. Minden ilyen jellegű tartalom rövid ismertetőt és gyakorlati útmutatót is tartalmaz. A felületen többféle adatvizualizáció érhető el, például térképes megjelenítés, kapcsolati háló, szófelhők, stilometriai elemzések, valamint egy virtuális szabadulószooba.

A műhely fölön található meg az ún. dHUpa-specifikáció, amely a magyar nyelvű kéziratok forrásainak jelölő nyelvi kódolásához tartalmaz ajánlásokat, egyrészt a dHUpán való publikáláshoz, másrészt pedig ezen dokumentumok feldolgozásához nyújt segítséget, valamint itt lehet tájékozódni Magyarországi első e-mail-archíváló projektjéről és további kutatási érdekességekről.

A platform háttérében egy szerkesztőségi rendszer működik, amely támogatja a szövegek szerkesztését és jelölőnyelvi kódolását. A feldolgozó munka első lépéseként a kéziratok digitális másolatai a Transkribus szoftverbe kerülnek feltöltésre, amely alkalmas a kéziratok képi szegmentálás-

ra, mesterségesintelligencia-alapú kézirás-felismerésre és annotálásra. Továbbá lehetőséget ad saját kézirás-felismerő modell építésére is. (Az első publikus, magyar nyelvű számítógépes kézirás-felismerő modellt a nemzeti könyvtár tette közé 2022-ben.) Az így előállt átiratok exportja TEI XML formátumban történik, amikhez az Oxygen XML Editor segítségével további metaadatok és annotációk kerülnek felvételre, valamint adatgazdagításokkal egészül ki, amit a mesterséges intelligencián alapuló automatizált műveletek is segítenek, mint a már említett névlem-felismerés és beépített névlem-azonosítás (entitástár). Az Oxygen környezethez a DBK specifikációján alapuló keretrendszer (framework) is készült. Ez egy olyan szerkesztői felületet biztosít, amely mindenki számára könnyen használható a megadott jelölőnyelvben való jártasság nélkül is. Ez teszi lehetővé, hogy más közgyűjtemények vagy egyéni kutatók számára is elérhető legyen ezen publikálási platform.

A digitális bölcsészeti munka azonban a szövegek gondozásán túl jelent meg az előállt adatok közötti összefüggések feltárását, láthatóvá és kutathatóvá tételét is. Ezek közül kiemelendő a térképes vizualizáció, a nemzeti könyvtár és a HUN-REN SZTAKI közös fejlesztése, amely ezt szolgálja különösen a levelezési irányok földrajzi és időbeli megjelenítésével. Jelenleg három jelentős szerző (Móricz Zsigmond, Kiss József és Petőfi Sándor) levelezése szerepel a felületen. A térképes kereső lehetővé teszi a szerzői gyűjteményekben való keresést és az egyes levelekhez kapcsolódó helynevek, típusok (keletkezés, feladás, kézbesítés),

időintervallumok szerinti szűrést is. A vizualizációs modul adatainak előállítását automatikus metaadat-kinyeréssel történik a framework segítségével, amely által generált táblázat a levelek keletkezési, postázási és kézbesítési adatait tartalmazza, majd az ezt követő munkafolyamat során az itt megtalálható helynév entitásokhoz koordináták is párosítva lesznek. Az így megjelenített adatok nemcsak a dokumentum kontextusát gazdagítják, hanem olyan kapcsolatrendszereket is láthatóvá tesznek, amelyek a források egyedi vizsgálata során rejtve maradtak. Valamint a helynevek és levelek közötti hiperhivatkozások lehetővé teszik a közvetlen átjárást a térképes felület és a dHUpa-n közzétett levélszöveg között, így a kéziratok valóban „bejárhatóvá” válnak.

A dHUpa a szolgáltatások egyik legösszetettebb és leginnovatívabb példája. A platform egyrészt magas szintű tudományos igényekkel teszi elérhetővé a történeti forrásokat, másrészt lehetővé teszi azok különböző szintű felhasználását oktatási, kutatási és kreatív célokra egyaránt. A kialakított szerkesztőségi környezet, a technológiai háttér, a szemantikus gazdagítás és a vizualizáció integrált rendszere példamutató modellként szolgálhat más közgyűjteményi szolgáltatások számára is.

A sorozat első része ide kattintva érhető el.

