

A Magyar Tudományos Művek Tára 2.0 verziójának fejlesztése

Micsik András, Pataki Balázs, Tóth Zoltán, Pallinger Péter, Prunk-Éger Edgár, Mátételki Péter, Turbucz Sándor, Felker Tamás, Frenzl Péter, Györök Attila, Hermann Zoltán, Naszradi Mihály, Nász Balázs, Pintér Gergő, Detvay László, Kovács László

MTA SZTAKI Elosztott Rendszerek Osztály, 1111 Budapest, Lágymányosi utca 11.

Kivonat: *A cikk az MTA KIK megbízásából az MTA SZTAKI-ban készülő új MTMT2 rendszer tervezési elveit és architekturális újdonságait mutatja be röviden, kitérve a bibliográfia feldolgozási modell változásaira is. Az új szoftver folyamatosan fejlődő és elterjedt környezetekre épült Java és Javascript alapokon. Bemutatjuk a felhasznált szoftverkomponenseket, a fejlesztés módszertanát, a platform főbb részeit, a felhasználói felületeket és külső kapcsolódási pontokat.*

Bevezetés

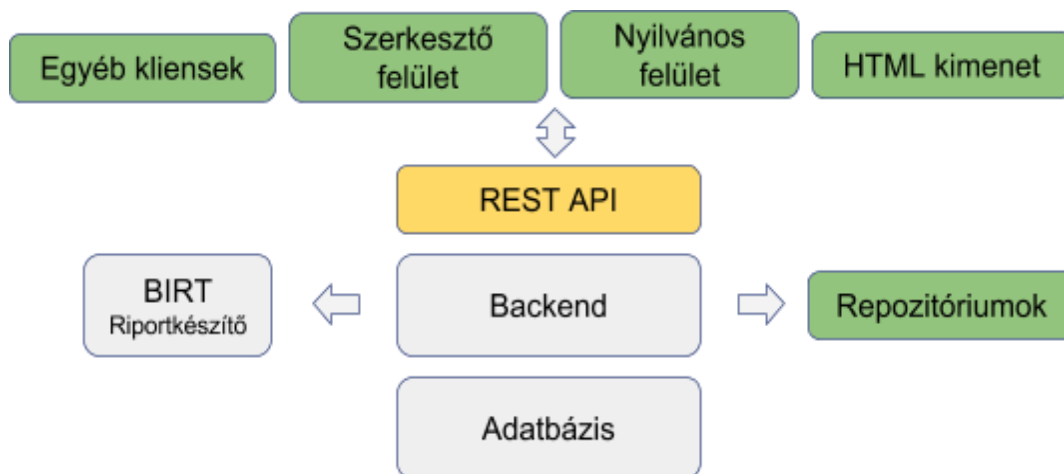
A hazai tudományos kutatás eredményeinek hiteles nyilvántartására és bemutatására létrehozott Magyar Tudományos Művek Tára (MTMT) [1] új szoftverrendszerének fejlesztésére 2014-ben az MTA SZTAKI Elosztott Rendszerek Osztály kapott megbízást. Az MTMT a magyar akadémiai és felsőoktatási szféra országos regisztere, és mint ilyen majdnem 50 ezer felhasználót szolgál ki, valamint statisztikai adatokkal látja el a tudománnyal foglalkozó különféle állami és egyéb intézményeket. A publikációs rekordok adattartalmának megbízhatóságát, magas minőségét továbbra is az országosan megszervezett humán kurátori hálózat (MTMT adminisztrátorok szervezete) biztosítja.

Az MTMT2 rendszer fejlesztése több korábbi felhasználói igényfelmérés és egy részletesen kidolgozott követelményrendszer alapján kezdődött. Az MTA SZTAKI Elosztott Rendszerek Osztálya rendszertervvel pályázott az MTMT2 kivitelezésére, melyet egy adminisztrátori fórumon be is mutatott. A pályázatokat szakmai zsűri bírálta el, mely alapján megkaptuk a megbízást, és elkezdődött a fejlesztés. A fejlesztéssel párhuzamosan a már elkészült kódot egy független tesztelő társaság folyamatosan ellenőrizte funkcionalitás és kódminőség szempontjából [2].

A fejlesztés során agilis szoftverfejlesztési módszertant alkalmaztunk, a SCRUM alapelveit követve alakítottuk ki a konkrét projektekre szabott variánst. Pár hetes fejlesztési szakaszonként (sprint) határoztuk meg a tervezett előrehaladást, illetve ellenőriztük az elkészült funkciókat. A fejlesztés során mindig volt egy prioritizált lista a tennivalókról, és egy demó rendszer, amely az aktuális állapotot mutatta. A kiindulási specifikáció frissítésre és átdolgozásra szorult, melyet a fejlesztéssel párhuzamosan meg tudtunk oldani. A megrendelő a tervezett és megvalósított funkciókat saját szakmai fórumain megvitatta, és az eredményeket Makara Gábor professzor rendezte össze és kommunikálta a fejlesztői csapattal, melyet Micsik András vezetett. A tervezés és megvalósítás során így lépésről lépésre haladva, organikusan fejlődött a rendszer, melynek jelenleg végső tesztelése zajlik, és bevezetése 2016 nyárra várható.

Architektúra

Az MTMT2 szoftver architektúráját az 1. ábra foglalja össze. A rendszer középpontjában az új API áll, amely sokféle külső rendszerrel, felülettel való integrálást tesz lehetővé. A rendszer felépítése egy hagyományos többrétegű architektúra, melynek részei; adattároló réteg, logikai (köztes, feldolgozó) réteg, és számos lehetőség a megjelenítő rétegben, melyek egy API-n keresztül kommunikálnak a logikai réteggel. Bár az ábrán ez külön nincs jelölve, a rendszer többféle módon skálázható, mivel a logikai és tároló réteg külön virtuális gépre telepíthető, illetve sokszorozható. A logikai réteg több, külön szabályozható gyorsítótárral is rendelkezik a teljesítményjavítás céljából.



1. ábra: Az MTMT2 szoftver architektúra főbb egységei

Felhasználói felület

A rendszer kétféle felületen (GUI) érhető el: a nyilvános felületen és a szerkesztői felületen.

A nyilvános felület célja az MTMT-ben gyűjtött adatok közreadása, a tájékoztatás. A főbb funkciók itt az intézmények és szerzők közleményeinek böngészése, keresése. A felület reszponzív web dizájnnal készült, így mobilon, tableten is jól használható. Az intézmények és szerzők tudományos tevékenységének összefoglalása néhány könnyen áttekinthető grafikonnal történik. A közleménylisták megoszthatóak és nyomtathatóak. Minden szerzőnek, intézménynek, részlegnek van saját oldala, amely egy rövid URL-lel behívható. A felület tehát arra fókuszál, hogy gyorsan és hatékonyan lehessen az MTMT-ben navigálni, és áttekintő információhoz jutni.

A szerkesztői felület az, ahol a szerzők saját, az adminisztrátorok pedig intézményük munkásságát gondozzák. Ennek megvalósítása egy úgynevezett single-page app, amely egy böngészőablakban megjelenő telepített alkalmazáshoz hasonlít. A felület egységes mind szerzők, mind adminisztrátorok számára, csak a végrehajtható műveletek különböznek. Ellentétben a nyilvános felülettel, a szerkesztői felület már hardverigényesebb; a szerkesztői munka kényelmes elvégzéséhez legalább 1024 pixel széles képernyő javasolt, és a webapplikációs felület meghajtása az elavult noteszgépeken már lassulást okozhat. A felület ismétlődő elemekből építkezik, mint például táblázat, eszköztár vagy űrlap, melyeket hasonló módon használhatunk különböző helyzetekben. A szerkesztői tevékenység alapciklusa így épül fel:

- kiválasztjuk a kezelni kívánt objektumtípust,
- kikeressük a szükséges objektumo(ka)t és kiválasztjuk őket a találati listából,
- elindítjuk a kívánt műveletet,
- módosításokat végzünk az űrlapo(ko)n,
- befejezzük a műveletet, és ezzel elmentjük az eredményt.

Jövőbeni cél, hogy a munkakörnyezet minél inkább személyre szabható legyen, jelenleg a táblázatokban megjelenített oszlopok sorrendje és a gyakori keresések menthetőek el.

Szerver

A szerver három fő komponensre oszlik: az adatbázisra, a webes API-ra és az API műveleteket megvalósító backend komponensre. A backend Java Spring platformra íródott, amely adatkezelési, biztonsági és REST API moduljai révén egy integrált keretet adott a fejlesztéshez. A backend és az adatbázis között Hibernate objektum-relációs megfeleltetést (ORM) alkalmaztunk. Az adatbázis PostgreSQL, melyhez a PostgreSQL saját teljes-szoveges indexelő bővítményét is felhasználtuk. A felvázolt fejlesztési környezet egy elterjedt, bevált és robusztus megoldás, mely ugyanakkor elég rugalmas az MTMT speciális igényeinek kifejlesztéséhez.

Látszólag az MTMT egy egyszerű adatnyilvántartó rendszer, viszont funkcionalitását több tényező is bonyolítja:

- a könyvtárak kifinomult katalogizáló rendszere, mely hagyományosan több évtizede meghatározó a közlemények regisztrálásában,
- a tudományos tevékenységek sokfélesége, az egyes tudományágak egymástól merőben eltérő teljesítménymérési szokásai és elvárásai,
- a sokféle adatkapcsolat és -formátum, melyen közlemények érkeznek és aggregált adatok távoznak a rendszerből.

Ezt szemlélteti az a tény is, hogy a rendszer adatmodellje több, mint 160 objektumtípusból áll. Az adatmodellben történt pár olyan nagyobb szabású módosítás, melyek megvalósításához a teljesen új rendszer elkészítése adott megfelelő alkalmat:

- A közleményekhez úgynevezett szerzőségi rekordok társulnak, melyek rögzítik a cikken feltüntetett szerzői nevet, az ahhoz adatbázisban hozzárendelt szerző objektumot, és a szerző egy vagy több affiliációját. Speciális szerzőség altípusok az intézményi illetve csoportos szerzőség.
- Az idézőközlemények és forrásközlemények egyformán írhatók le, a hivatkozások objektumai kapcsolják össze az idézőt a forrással. Ezáltal lehetővé válik, hogy egy közlemény egyszerre legyen forrás- és idézőközlemény, valamint az is, hogy egy idéző egy közleményen belül több helyre is hivatkozhatson.
- A közleményt leíró mezőértékek számossága egyről többre változott például nyelv, médium és kiadás helye esetén.

Pár új objektumtípus is megjelent az adatmodellben, amelyek elkülönített tárolására eddig nem volt lehetőség, például konferencia és projekt (ez utóbbi a szponzorálás, támogatás leírásakor használható fel). A cédula egy olyan objektum, amely közleményhez vagy más alapobjektumhoz társítható, és annak történetét, az azzal kapcsolatos feladatokat örökíti meg. Közlemények esetén így követhetővé

válik, hogy mikor és ki láttamozta, utána milyen azonosítókat rendeltek esetleg hozzá, vagy ki milyen repozitóriumba töltötte fel, stb.

API

Az API kialakításánál az volt a fő célunk, hogy egy általános felületen keresztül történjen minden kommunikáció a rendszerrel. Az API a REST alapelveket követi, és saját webes dokumentációs felülete van, mely tartalmazza az egyes műveletek és lehetséges paramétereik leírását.

Az API a REST esetében szokásos egyszerű URL sémát használja:

```
/api/<objektumtípus>[/<művelet>] [ ?<paraméterek> ]
```

Az URL-ben megadható paraméterek között nagyon sok az általánosan megadható, például a válasz nyelve, formátuma, részletezettsége. Listázás esetén komplex szűrőfeltételek, illetve rendezés adható meg paraméterként. Az API nyelve angol, vagyis az objektumtípusok, műveletek, a válaszban a mezőnevek mind angolul szerepelnek. Az API verziókat a válasz MIME típusába kódoljuk, így a verzióváltásról a kliens értesül, illetve jelezheti is hogy melyik API verzióra van felkészülve.

Az API használatához Java és PHP példakódot is készítettünk, mely bejelentkezés, lekérdezés és módosítás műveletein keresztül mutatja be az API használatát. Ezenkívül a Swagger segédeszköz által generált online API dokumentáció is a fejlesztők rendelkezésére áll.

Kapcsolódások

Az alábbiakban röviden áttekintjük a rendszer eddig megvalósult kapcsolódási módjait. A nyilvános és a szerkesztői felületet már korábban bemutattuk, ezek is az API-n keresztül végzik működésüket.

Intézményi kliensek: ezek a fenntartó és kifejlesztő intézményen belül működnek, így lehetőségük van arra, hogy az MTMT adatait az intézmény belső adatbázisaival összekapcsolva, fuzionálva kezeljék és jelenítsék meg. A kutatómenedzsment egy tipikus ilyen feladat, amelyhez szükség van a tudományos teljesítmény adataira, valamint a finansziális és egyéb támogatások, a folyó projektek részleteire is. Bár a CRIS (Current Research Information System) és az MTMT jellegű rendszerek közötti határvonal eléggé elmosódott, az biztos, hogy az MTMT nem foglalkozhat a közlemények születésének anyagi és erőforrásbeli hátterével.

Portálok, országos szervezetek, mint például az NKFIH vagy az ODT (doktori.hu) átemelhetnek az MTMT-ből adatokat, például kutatók bemutatásához, vagy pályázatok értékelésének háttéranyagaként. Az MTMT2 API lekérdezéseit XML, JSON vagy HTML formátumban is lekérhetik ezek a szervezetek, akár belső felhasználásra, akár külső megjelenítésre kívánják felhasználni.

Beágyazható elemek, angol néven widget-ek állnak minden intézmény és szerző rendelkezésére, hogy az MTMT-ből származó, ellenőrzött információval gazdagíthassák oldalukat. Az előző megoldáshoz hasonlít ez is, viszont itt paraméteresen választhatóak grafikonok, közleménylista formátumok is, melyek a tudományos előéletet mutatják be, és könnyen beépíthetőek vagy hivatkozhatóak egyéni vagy intézményi honlapokról. Ennek egyik előnye a kutatói munkásság dupla vagy tripla dokumentálásának elkerülése, elég az MTMT-ben tárolt információkat gondozni.

Riportok, statisztikák készítéséhez a BIRT riportkészítő rendszert integráltuk [3]. A BIRT-nek van egy szabadon letölthető riporttervező alkalmazása, ebben készítettük el a régi MTMT rendszerben használatos közleménylisták, összefoglaló táblázatok több variánsát. Az elkészült riportsablonokat a rendszerbe betöltve megfelelő engedélyezés után le lehet futtatni az éles adatokon. Arra is lehetőség van, hogy a BIRT riportkészítő a saját gépünkön készítse el a riportot az MTMT2 API-val letöltve a szükséges adatokat.

Az **importálás** során közleményadatokat más rendszerekből lehet nagy számban áttölteni. Az automatikus import módozatot a rendszer periodikusan futtatja le, ebből a Web of Science import van tesztelés alatt. Más rendszerekből, mint például Scopus, Matarka, stb. a RIS formátumban exportált adatokat lehet az MTMT-be betölteni. Azonosító alapján az egyes cikkek metaadatait is be tudjuk tölteni CrossRef DOI alapján (Scopus és PubMed hasonló szolgáltatás is tervezett).

A **repozitóriumok** feladata a közlemények teljes szövegének illetve kísérő anyagainak tárolása. Az MTMT kiemelten kezeli az ilyen repozitórium linkeket, mivel ezek teszik igazán hasznossá a nyílt hozzáférés rendeleteket (open access mandate) és a nyílt hozzáférés mozgalmat. Ezért az MTMT-ből is fel lehet tölteni a közleményeket a kiválasztott repozitóriumokba az egységes SWORD protokoll segítségével [4]. Ilyenkor a feltöltött fájlokat a backend továbbítja a kijelölt repozitóriumnak a közlemény metaadataival együtt, ezeket a repozitórium adminisztrátorai hagyják jóvá és illesztik be a repozitóriumba.

Összegzés

Az új MTMT rendszer jóval több lehetőséget teremt más rendszerekhez való kapcsolódásra, legyen az máshol már létrehozott közleményadatok átvétele, vagy az összegyűjtött adatok újrafelhasználása, megjelenítése. Reményeink szerint így az MTMT2 körül új, nyílt szoftverkomponensek és segédeszközök jelenhetnek meg, jobban kiszolgálva a kutatók igényeit, és élénkítve a magyar tudományos élet vérkeringését.

Hivatkozások

[1] Holl András, Makara Gábor, Micsik András, Kovács László: *MTMT: The Hungarian Scientific Bibliography*. Share-PSI workshop on Uses of open data within government for innovation and efficiency, 2014, Samos, Greece

[2] Szűts Béla, Gyimóthi Zoltán: *MTMT2/MyCite2 szoftver minőség-ellenőrzése*. Networkshop 2016, Debrecen

[3] BIRT Report Designer, <http://www.eclipse.org/birt/about/designer.php>

[4] Holl András, Karácsony Gyöngyi, Harangi Gabriella, Cserbák Márton: *Teljes szövegű publikációk cseréje repozitóriumok között - a SWORD protokoll megvalósítása az MTA Könyvtárban és a Debreceni Egyetemi Könyvtárban*. Networkshop 2009, Szeged