Projektkönyvtár, tudástár – avagy a best-of meghaladásának kísérlete

(Project library, knowledge store – or better than the bests before)

Pitlik László

Kivonat: A projektek belső és külső kényszerek által hajtva dokumentumokat termelnek. Ezek fizikai tárolásának mikéntje alapvetően lényegtelen, ha a dokumentumok leírása, csoportokba sorolása a lehető legsokszínűbb. A minimális leírást a már praktizált megoldások eredője definiálja. Ennek meghaladása az alapvetően tartalom-orientált csoportok/csoport-hierarchiák képzésével és az objektumok minden létező státuszváltozó szerinti besorolásával oldható meg. Ha minden projekt hasonlóképpen viselkedik és minden korábbi projekt minden későbbi számára benchmark-ként szolgál, akkor hosszabb távon a státuszváltozók katalogizálásával egyre kisebb ütemben képződik új attribútum, vagyis egyre szabvány-közelibb lesz az objektumok leírása, ami által az éppen szükséges dokumentumok fellelésének ideje szinte bármilyen cél mellett egyre kevesebb erőforrást igényel bárki számára.

Kulcsszavak: szabványosítás, benchmarking, keresés-optimalizálás

Abstract: Projects produce documents because of internal and/or external force fields. The physical storage of documents as such is not relevant, if the descriptions, grouping possibilities of documents are ideal. The minimal number of descriptive attributes can be seen as given based on best practices. The minimal set of attributes should always be exceeded in a new project - involving content-oriented grouping (hierarchy of groups) for building new status variables. If all projects work in the above mentioned way (using each other as a kind of benchmark – needing to exceed), then a catalogue of the rational attributes will be completed step by step in future and exploring new attributes will be hardly impossible. The quasi standardized catalogue of the potential attributes makes possible to minimize search resources for quasi arbitrary users.

Keywords: standardization, benchmarking, SEO

# Bevezetés

A probléma stratégiai alapjait a tanulmány első része már letette [[1]](#footnote-1), vagyis itt és most operatív lépések és benchmark-ok bemutatása következik annak érdekében, hogy egy KÖFOP projektpár keretében keletkezett publikus anyagok (<http://miau.gau.hu/miau/kofop/?C=M;O=D>) és példaértékűen egyes nem (vagy nem direkt módon) publikus, de tulajdonképpen védelmet nem igénylő állományok kezelése tételesen bemutatásra kerülhessen.

Az alapozó dokumentum kivonatában is jelzettek szerint, a fizikai helye a konkrétan kezelendő objektumoknak nem lényeges. A symlink-jellegű támogatás, vagyis egyes publikációk konkrét kiadványok (pl. MIAU-No.\*) konkrét fizikai helyéhez való rendelése (vö. <http://miau.gau.hu/miau2009/index.php3?x=e0&string=balogh.anik>.) ismét csak nem oldja meg a problémákat, vagyis a tetszőlegesen sok kapcsolati forma leképezését. Egy összetett kereső egy megfelelő státuszváltozó-állomány esetén röptében képes lehet sokféle összefüggést kezelni, de ilyen esetekben a keresés általában egyetlen egy OAM-ra korlátozódik, ahol egy sor egy objektum és minden attribútum egy-egy oszlop. Az ilyen struktúra kényelmes az SQL-írás kapcsán, de mivel nem minden attribútumnak kell kényszerűen minden objektumra vonatkoznia (pl. frame/sec adata nincs minden objektumnak, csak a mozgóképeknek). Emellett egyes attribútumok mentén nem csak egyetlen egy egyedi adat értelmeződik, hanem tetszőlegesen sok (pl. mellékletek száma, mint attribútum még egy adat, de a mellékletek URL-jei, ID-adatai már egy-egy alapobjektum esetén korlátlanul sok attribútumot igényelnének egy 2D-s OAM-ban – ahol nem mellesleg a legtöbb objektumnak vélhetően melléklete sem lenne). A miau/kofop-utalást tartalmazó fenti URL-ben megadott paraméterek láttán még a LINUX/web-folderek kapcsán az is elérhető, hogy időrendi sorrendben legyenek kilistázva fentről le, vagy fordítva, de ez az időpecsét már az utolsó mentés és nem a keletkezés dátumát kezeli. Summa summarum: a több-dimenzió leírórendszert úgy kell kialakítani, hogy abban quasi minden ismert (pl. könyvtári) attribútum helyet kapjon és EZEN FELÜL minden más, ami az érintett objektumok tartalmából, specialitásaiból következik.

Ahhoz, hogy lássuk, miként kezeli ma a könyvtári/akadémiai szféra a (tudományos) műveket következzék egy lista, ill. példa-állományok a mai gyakorlatot megtestesítő megoldásokról, vagyis a benchmarkok-ról:

# Benchmark-nézetek

A példák generálása érdekében a szerző saját adatvagyonára támaszkodva alakított ki demo-nézeteket:

* szaklapnézet: pl. [http://miau.gau.hu/miau2009/index.php3?x=miau128&where[indexkod]=miau234](http://miau.gau.hu/miau2009/index.php3?x=miau128&where%5bindexkod%5d=miau234)
* egyedi szaklap-adatlap:
* pl. [http://miau.gau.hu/miau2009/adatlap.php3?where[azonosito]=24144&mod=l2003](http://miau.gau.hu/miau2009/adatlap.php3?where%5bazonosito%5d=24144&mod=l2003)
* MTMT-nézet (nyomtatóbarát): <http://miau.gau.hu/miau/mtmt/20180328/pitlik%20laszlo%20munkassaga.html>
* MTMT-nézet (web): <http://miau.gau.hu/miau/mtmt/20180328/Pitlik%20Laszlo%20munkassaga.html>
* MTMT-nézet (CSV): <http://miau.gau.hu/miau/mtmt/20180328/mycite.csv>
* MTMT-nézet (XML): <http://miau.gau.hu/miau/mtmt/20180328/mycite.xml>
* MTMT-nézet (XML2XLS): <http://miau.gau.hu/miau/mtmt/20180328/xml-bol.xlsx>
* MTMT (RIS – ékezettel és anelkul): <http://miau.gau.hu/miau/mtmt/20180328/mycite.ris>, ill. <http://miau.gau.hu/miau/mtmt/20180328/mycite%20(1).ris>
* MTMT (idézők-RIS: ékezettel és anélkül): <http://miau.gau.hu/miau/mtmt/20180328/mycite_citing.ris>, ill. <http://miau.gau.hu/miau/mtmt/20180328/mycite_citing%20(1).ris>
* MTMT (XML-validációs formátum): <http://miau.gau.hu/miau/mtmt/20180328/mycite%20(1).xml>
* MATARKA-nézet (RIS-output): <http://miau.gau.hu/miau/189/matarka/lista_233-234-235.ris>
* MATARKA-nézet (txt-input): <http://miau.gau.hu/miau/189/matarka/matarka_miau_2018_1-3.txt>
* MATARKA-nézet (keresés/kosár): <http://www.matarka.hu/index.php#reszletes_kereses>, ill. <http://www.matarka.hu/index.php#egyszeru_kereses>, valamint <http://www.matarka.hu/indexkereses.php>
* MIAU2MATARKA (konverzió): <http://miau.gau.hu/miau/189/matarka/matarka_miau_2018_1-3.xlsm>
* EHM-kereső: <http://ehm.ek.szte.hu/ehm?p=0>
* Google-Scholar-nézet: <https://scholar.google.com/citations?user=xrx_gEIAAAAJ&hl=en>
* Research-Gate-nézet: <https://www.researchgate.net/profile/Laszlo_Pitlik>
* Academia.edu-nézet: <https://szie.academia.edu/LaszloPitlik>
* MEK-EPA-DKA közös kereső: <http://mek.oszk.hu/kozos-kereses.mhtml?dc_creator=ger%C5%91+p%C3%A9ter&dc_title=&dc_subject=&mehet=%C3%9AJ+KERES%C3%89S>

További részletek (pl. hibalisták, állapot-jelentések, idéző-gyűjtemény):

* <http://miau.gau.hu/miau/mtmt/20180328/?C=M;O=D>
* <http://miau.gau.hu/miau/mtmt/>
* <http://miau.gau.hu/miau/citations>

# Potenciál és kritika

Az alábbi dokumentumok részletes áttekintést adnak az MTMT által kialakított rendszer potenciáljáról és kockázatairól:

* <http://miau.gau.hu/miau/208/20151120_v2.doc> (a kánon alacsony és magas komplexitású értelmezései a tudománymetria miértjeiről és hogyanjairól)
* <http://miau.gau.hu/miau/210/mtmt_big_data_v1.doc> (MTMT-s adatok felhasználása)
* <http://miau.gau.hu/miau/209/mtmt_types.doc> (MTMT-s fogalmak értelmezési hiányosságai)
* <http://miau.gau.hu/miau/218/decens_docensek.docx> (MTMT-s adatok felhasználása)

Mint a felsorolt művek alapján látható, az objektumok leírását és értékelését/értelmezését két önálló szakmai kihívásként kell értelmezni.

# Attribútumok

A dokumentumok leírása mindenkor önkényes, vagyis a leíró attribútumok száma és jellege korlátlan lehet. A GDPR elvek a személyes adatkezelés esetén a célirányosságot és a visszafogottságot várja el. Ezzel szemben a közérdekű adatok esetében a minél nagyobb részletesség a mértékadó:

## MTMT

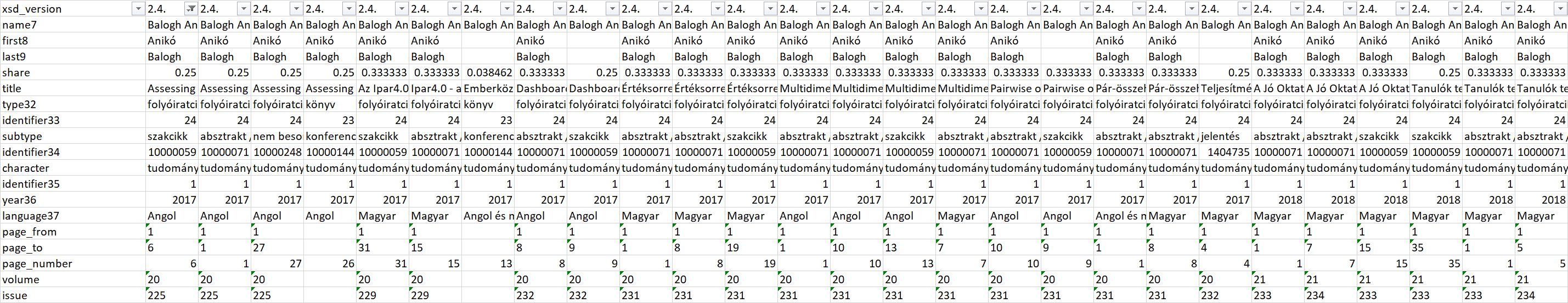
Az alábbi lista a KÖFOP projektek keretében keletkezett dokumentumok MTMT-ben való tárolása nyomán kezelt attribútumait (210) mutatja be a publikus mezőnevek szintjén, melyek értelmező feloldása nem áll rendelkezésre. Az egyes mezők értelmét a bevitt adatvagyon alapján lehet feltárni külső szemlélőként: pl. <http://miau.gau.hu/miau/mtmt/20180328/kofop-demo.xlsx>

xsd\_version identifier status language first last comment identifier2 type name year event date time login timestamp first3 last4 comment5 identifier6 name7 first8 last9 share comment10 identifier11 type12 for institute identifier13 parent direct abbreviation name14 first15 last16 share17 comment18 identifier19 type20 for21 institute22 identifier23 parent24 direct25 abbreviation26 title identifier27 type28 name29 public identifier30 validator date31 type32 identifier33 subtype identifier34 character identifier35 free\_field num year36 from to language37 page\_from page\_to page\_number type38 title39 shorttitle peerreviewed published\_abroad year40 from41 to42 language43 city country publisher city44 country45 publisher-country city46 country47 identifier48 type49 name50 type51 identifier52 subtype53 identifier54 character55 identifier56 name57 share58 first59 last60 identifier61 type62 comment63 institute64 identifier65 parent66 direct67 abbreviation68 name69 share70 first71 last72 identifier73 type74 comment75 volume\_number page\_number76 impactfactor type77 volume\_number78 title\_series identifier79 type80 title\_series81 subtitle publisher82 city83 country84 publisher-country85 city86 country87 media volume issue administrative\_comment comment88 city89 country90 participation organizer city91 country92 abstract for93 institute94 identifier95 parent96 share97 direct98 abbreviation99 status100 name101 year102 first103 last104 comment105 identifier106 type107 event108 date109 time110 login111 independent independent-ok language112 foreign\_only published\_abroad113 identifier114 type115 name116 identifier117 title118 comment119 context year120 from121 to122 name123 name124 page\_from125 page\_to126 page\_number127 type128 identifier129 subtype130 identifier131 character132 identifier133 references subtitle134 publisher135 city136 type137 title138 identifier139 type140 type141 title\_series142 title\_series143 volume144 free\_field145 num146 issue147 city148 country149 timestamp150 first151 last152 comment153 identifier154 organizer155

A nem (kellően) beszédes mezőnevek és a közérdekű adatvagyon leírásának nem közérdekű volta komoly értelmezési kockázatok forrása, ahogy a teljesnek vélt/ígért adatvagyon kivonat matarka-mentessége is felveti az MTMT-s adatkezelés esetlegességeit.

# Aktív MTMT attribútumok

A teljesnek sugallt MTMT-XML az alábbi attribútumokat kezeli (vö. kofop-demo.xls):



Ebből a listából pl. a tényleges (full text URL) elérhetőség és pl. a MATARKA-ID hiányzik.

## ID-típusú attribútumok

* MTMT-ID (minden esetben)
* MATARKA-ID (minden esetben – s ennek ellenére a kofop-demo.xlsx nem tartalmazza a matarka szót – vö. melléklet, ahol matarka id csak abban az esetben releváns, ahol a dokumentum újságcikként jelenik meg, ami nem triviális pl. háttér-állományok kapcsán)
* DOI (comment88 – 2 rekord a kofop-demo.xlsx-ben)
* MIAU (quasi-DOI – minden rekord esetében – pl. [http://miau.gau.hu/miau2009/adatlap.php3?where[azonosito]=24144&mod=l2003](http://miau.gau.hu/miau2009/adatlap.php3?where%5bazonosito%5d=24144&mod=l2003))
* ResearchGate-ID: pl. <https://www.researchgate.net/publication/320125911_Dashboards_and_OLAP_services_in_LLL_and_distance_learning_processes_or_experiences_about_log-based_learning-and_education-management?ev=prf_ov_fet_res&_iepl%5BviewId%5D=WZKCCnyUHtmpNPh27xSxnR3v0PaMrGsw2J2x&_iepl%5Bcontexts%5D%5B0%5D=prfhpi&_iepl%5Bdata%5D%5BstandardItemCount%5D=3&_iepl%5Bdata%5D%5BuserSelectedItemCount%5D=0&_iepl%5Bdata%5D%5BtopHighlightCount%5D=2&_iepl%5Bdata%5D%5BtopHighlightIndex%5D=2&_iepl%5Bdata%5D%5BfeaturedItem2of2%5D=1&_iepl%5BtargetEntityId%5D=PB%3A320125911&_iepl%5BinteractionType%5D=publicationTitle> – RÖVIDEN: <https://www.researchgate.net/publication/320125911>
* Google Scholar ID: pl. <https://scholar.google.hu/scholar?oi=bibs&cluster=18264402803583623566&btnI=1&hl=hu>

Az ID típusú nyilvántartás akkor értékes, ha DOI-szerű, azaz állandó elérést biztosító. A DOI rendszer egy fajta ID-monopolizálást biztosíthat világszerte, de ettől függetlenül bármely szolgáltató garanciákat vállalhat egy URL fenntartására.

A teljes körűnek ígért MTMT-XML nem tartalmaz (DOI-szerű) URL-adatokat, csak akkor, ha ezek megjegyzésként kerültek tárolásra.

# Dokumentumok fizikai leírása

Az MTMT és a MATARKA nem várja el a digitális objektumok fizikai paramétereinek leírását, de maguk a file-szerverek, ahol ezek hozzáférhetők, ezzel mindenképpen számolnak: pl.

* Méret
* Kiterjesztés
* Kiterjesztés-variánsok (pdf, doc, docx, rft, stb.)

A MIAÚ adattárolási rendje kiterjed a fájlok fizikai leírására: <http://miau.gau.hu/admin/clone.php3>

Az űrlap teteje

**Új dokumentumok feltöltése**

Telepítve: 2004.12.06.

|  |  |
| --- | --- |
| A dokumentum alapadatai: | |
| **Kód:** |  |
| **Szerzõ:** |  |
| **Szerzõtársak:** |  |
| **Szerkesztõ:** |  |
| **Intézményi adatok:** |  |
| **Lektor:** |  |
| **Címe:** |  |
| **Megjelenés éve:** |  |
| **Kiadvány:** |  |
| **Sorozat:** |  |
| **ISBN/ISSN:** |  |
| **Nyelv:** |  |
| **Kulcsszavak:** |  |
| Ha elektronikus dokumentum, akkor | |
| **Méret (Byte):** |  |
| **Formátum:** |  |
| **URL:** |  |
| A dokumentum csoportba sorolása: | |
| **Fõcsoport:** |  |
| **Alcsoport:** |  |



Az űrlap alja

[**Admin**](http://miau.gau.hu/admin/index.html)

Mivel maga az adatbeviteli űrlap csak jelszóval érhető el, így ennek tételes bemutatása itt és most elkerülhetetlen.

A MIAÚ adatbeviteli űrlapból a MATARKA csak a szerző nevét, a mű címét, a kiadót, ill. a folyóirat számát, évfolyamát várja el: vö. <http://miau.gau.hu/miau/mtmt/?C=M;O=D>

A fenti adatbeviteli űrlap alján a MIAÚ már az OSIRIS-hez (<http://miau.gau.hu/osiris>) hasonló csoportosításokat is támogat, melyek őse egy html-keretrendszer volt: <http://miau.gau.hu/sitemap/index.html>

# Speciális attribútumok

A projektdokumentumok tartalmának sokszínű leírásához az alábbi mintázat mindenképpen hasznos illene, hogy legyen – különösen, ha ezen státuszváltozók nem csak kézzel kerülhetnek hozzárendelésre egy-egy objektumhoz:

* Projekt-kulcsszavak és kulcsszó-hierarchiák: a sikeres projekttervben leírt és az elszámolás szempontjából fontos kulcsszavakat a dokumentumok kulcsszavai és vagy teljes szövege kapcsán vezetni érdemes, így ezek szövegbányászati alapon feltárhatók és az adott kulcsszavakhoz való kötődés státuszváltozói automatikusan igenre állíthatók (pl. a <http://miau.gau.hu/miau/kofop> alatt lévő összes dokumentum lehetne kézzel és/vagy szövegbányászati alapon az adatelemzés, a big data, stb. kulcsszavakhoz kapcsolva). Értelemszerűen azok a kulcsszavak a „jók”, melyek nem minden dokumentum kapcsán relevánsak, hiszen ami mindenhez kapcsolódik, az semmilyen szűrésre nem ad módot. A kulcsszavak hierarchiájának mibenléte nem része a leírásnak direkt módon. Egyszerre több kulcsszó-hierarchia is létezhet párhuzamosan. A hierarchikus kapcsolatok (csoport/főcsoport-bontás - öröklődés) programozási és nem leírási probléma.
* KNUTH/knut: 1 knut az a tudás, mely már forráskódba került átfordításra a szómágia keretei közül tudásmérnöki munkával kiragadva. Ez a dokumentum nem éri el az 1 knut értéket, vagyis ez a dokumentum ebben a formájában alapvetően még szómágia, mert nincs direkt forráskód-kötődése még az automatizálási utalások kapcsán sem (pl. azon szövegbányászati kódrészlet nem része a mellékletnek, mely alapján adott dokumentumhoz adott kulcsszó esetében a státuszváltozó értékét a forráskód vezetné le és állítaná be a leíró adatbázisban). Az 1 knut értéket a szöveges dokumentumok ritkán képesek elérni, hacsak nem egy szabályrendszer írott változatát tartalmazzák – de még a jogszabályok sem igazán tekinthetők knut-ban kifejezhető értéknek, ha az input-fogalmak maguk is absztrakciók és nem mérések. Egy-egy XLS, mely az alapadatoktól a végeredményekig minden lépést reprodukálható módon tartalmaz, leéri az 1 knut értéket, vagyis ezek forráskódba forgatása már triviális programozói feladat illik, hogy legyen: vö. <http://miau.gau.hu/miau2009/index.php3?x=e0&string=without.magic.of.words>
* Mellékletek: A dokumentum-leírás speciális nézete a mellékletek kezelése, ahol is a mellékletek egyszerű esetben egy rétegben kapcsolódnak egy dokumentumhoz akármilyen mennyiségben, de komplex esetben a mellékleteknek is lehet melléklete, vagyis a kulcsszó-hierarchia logikája alakul ki URL-adatokra.
* Többnyelvűség: Amennyiben egy-egy dokumentumnak több nyelven is vannak részletei (pl. kivonatok), akkor ezeket lehet mellékletnek tekinteni, amennyiben pl. egy pdf-en belül lehet #search, és/vagy #page hivatkozásokat megadni. Vagy ezen idegen nyelvű részleteket önálló objektumként is le kell írni és a fődokumentumhoz egy fajta mellékletként kell csatolni, ahol a mellékletek fajtái is önálló paramétert jelentenek (pl. klasszikus melléklet, speciális melléklet: pl. kivonat, stb.).
* Kategóriák: Az MTMT mintájára elvileg tetszőlegesen sok és hierarchikus dokumentum-kategória hozható létre, de ezek csak akkor igazán értékesek, ha a kategóriába sorolás automatizálható teljesen és hibátlanul. A manuális=önkényes besorolásnak csak tájékoztató hatása lehet. A besorolást lehet tetszőlegesen sok érintett által is elvégezni engedni (pl. szerzők – külön-külön is, adminisztrátorok, olvasók, stb.), s önálló attribútummá válik innentől a besorolás homogenitás, időbeli szórása.
* Használati adatok: A dokumentumok létezése kapcsán ideális esetben ezek használati log-jai is attribútummá válhatnak, melyek tartalma ráadásul dinamikusan alakul: pl. letöltések száma.
* Fórumok: Amennyiben a dokumentum kapcsán moderált/nem-moderált formában fórumok engedélyezettek, akkor a dokumentum (vö. youtube-objektum) része a fórum bejegyzés-sorozata is, mely a letöltéshez (használathoz) hasonlóan önálló statisztikai kihívásként számos, dinamikus tartalmú attribútumot termel (pl. hozzászólások száma, egy hozzászólás által generált hozzászólások átlagos száma, stb.).
* Hitelesség: A dokumentumok egy része (pl. szakértői rendszerek) kapcsán eleve számos statisztikázható (statikus) attribútum létezik: pl. a dokumentum által kezelt objektumok, attribútumok száma, a kombinatorikai tér mérete (vö. <http://miau.gau.hu/miau/kofop/AVG_KK_v1.docx>), mely megközelítés már átvezet az adatvagyon-gazdálkodás általános elveihez.
* Érvényesség: A dokumentumokban (pl. jogszabályokban) szereplő üzeneteknek lehet hatálya, vagyis külön státuszváltozó a dokumentumban lévő tudás érvényessége, hatályossága.
* Verziószám: Különösen a statisztikai jelentések kapcsán a verziók nem csak azt jelentik, hogy az idő múlásával változhat a tartalom, hanem azt is, hogy az adatközlés milyen kiérleltségű (pl. gyorsjelentés, jelentés, revidiált jelentés).

# Konklúziók

A portfólió az a fogalom, ami mentén általában a szerzők mutatják be életútjukat. Ez a formátumok és az alkotás sokfélesége kapcsán olyan komplex leíró rendszerré vált, melyet állandóan a megújuló lehetőségi térhet kell igazítani: <http://miau.gau.hu/miau/138/la138.doc>

# Mellékletek

MATARKA-azonosító vezetése az MTMT-ben:

Balogh Anikó;

Balogh Anikó (tudásmenedzsment) Pitlik László;

Pitlik László (Informatika) Schnellbach Máté; Szani Ferenc

Assessing diversity in learners background an performance

MAGYAR INTERNETES AGRÁRINFORMATIKAI ÚJSÁG (ISSN: 1419-1652) 20: (225) pp. 1-6. (2017)

(ISBN 978-⁠615-⁠5511-⁠18-⁠9)

Nyelv: Angol

Link(ek): Teljes dokumentum, Matarka

Folyóiratcikk/Szakcikk/Tudományos [3245860] Matarka (2017-07-03 18:18:32) {Vit Lukas (CEU 4)}

[Nyilvános, admin láttamozta: Vit Lukas (CEU 4) 2018.03.21. 12:05:31]

Vit Lukas (CEU 4) Külső azonosítók

Idézők

Repozitorium

ODT/⁠MAB 10 legfontosabb közlemény

\*\*\*\*\*\*\*\*\*\*\*

Balogh Anikó;

Balogh Anikó (tudásmenedzsment)

ELTE/⁠IK/⁠Programozási Nyelvek és Fordítóprogramok Tanszék Pitlik László;

Pitlik László (Informatika) Schnellbach Máté; Szani Ferenc

Assessing diversity in learners background an performance -⁠ presentation

MAGYAR INTERNETES AGRÁRINFORMATIKAI ÚJSÁG (ISSN: 1419-1652) 20: (225) pp. 1-27. (2017)

EDEN. Konferencia helye, ideje: Jönköping, Svédország: 2017.06.13-2017.06.16.

Nyelv: Angol

Link(ek): Teljes dokumentum, Matarka

Folyóiratcikk/Nem besorolt/Tudományos [3245859] Matarka (2017-07-03 18:18:32) {}

[Nyilvános, admin láttamozta: Vit Lukas (CEU 4) 2018.03.21. 10:54:29]

1. (vö. <http://miau.gau.hu//miau/236/projektkonyvtarak_v1.docx>) [↑](#footnote-ref-1)