

## ÁLOM VAGY VALÓSÁG?

– egy (fél)automatikusan működő bibliográfiai adatbázis felé –

Kollár István

az MTA doktora, egyetemi tanár,  
Budapesti Műszaki és Gazdaságtudományi Egyetem  
Méréstechnika és Információs Rendszerek Tanszék

kollar@mit.bme.hu

Dr. Kovács János éppen OTKA-pályázatot nyújt be. Ehhez a *Magyar Tudományos Művek Tárában* (MTMT) frissítenie kell az adatait. Kérésére titkárnoje a legutolsó feltöltés óta megjelent publikációinak teljes szövegét feltölti az intézményi repozitóriumba,<sup>1</sup> és ezzel automatikusan az MTMT-be is betölti őket. Egy Word-állományból egy-két gombnyomással áttölti még azoknak az újabb közleményeknek az adatait, melyek nem szerepelnek egyik repozitóriumban vagy adatbázisban sem (például bizonyos könyvfejezetek), majd számítógépén belép az MTMT-be, megnyomja a „szinkronizálás” nyomógombot, és elmegy kávézni. A program elindul: a még nem teljes adatú közlemények adatait kiegészíti a *Web of Science/Scopus/Pubmed/IEEE Xplore/MathSciNet/Matarka* stb. adatbázisok adataival, és ahol lehet, összeköti a tételeket az adatbázis-tételekkel, a könyvekhez és interneten elérhető könyvfejezetekhez megkeresi a könyvtári kapcsolatokat (OSZK/BME-OMIKK/Library of Congress/British Library stb.), pótolja az ISBN-számokat, a szabadal-

makat összeköti a Magyar Szabadalmi Hivatal adatbázisával, majd az idézőket is tartalmazó adatbázisokból automatikusan letölti a publikációkhoz az ismert idézőket is. Ezeket automatikusan megjelöli (függő/független), majd lefuttatja az automatikus hitelesítést mind a közleményekre, mind az idézőkre, mely ellenőrzi és megjelöli, mely adatbázisokban található meg és azonosak-e az illető cikk adatai, és hogyan található meg a cikk teljes, publikált szövege (Digital Object Identifier – DOI).

A titkárnoje visszatérve látja, hogy a program készen van, és szól dr. Kovács Jánosnak, aki megszemléli listáját és a részletes tudományometriai adatainak táblázatát. Még nem egészen elégedett, ezért elindítja a Google Scholar idézőkeresőt. A program egyenként, a meglévő idézőlistába illesztve felajánlja az interneten újonnan talált idézőket: bevigye-e vagy elvesse-e ezeket. Az idegen nyelvű adatokat (kínai, japán idézők) angolra átírva is megmutatja, hogy dönteni lehessen, majd eredeti nyelven menti őket, az angol átírással kiegészítve.

A lista másfél óra múlva készen áll pályázatbeadásra. Egy gombnyomással áttölti az adatokat az OTKA rendszerébe, és elégedet-

ten visszatér kutatómunkájához. Az elkövetkező fél évben nem kell a listájával foglalkoznia, de ha mégis fontos lenne, a „szinkronizálás” gombot titkárnoje bármikor újra meg tudja nyomni... az internet pedig „neki dolgozik”: cikkei a honlapjáról kiindulva teljes szöveggel elérhetőek, az érdeklődő kutatók könnyen letöltik és elolvassák. A webes elérhetőség miatt eredményei, gondolatai hamar bejutnak a nemzetközi vérkeringésbe, nemcsak használni, hanem hivatkozni is fogják őket, úgyhogy amikor majd éves intézeti jelentésre, kutatási beszámolóra, vagy akadémikus-jelölésre kerül sor, csak a titkárnojenek kell ismét megnyomnia a gombot, és begyűjteni az újabb idézőket...

Képzeldés, vagy kézzelfogható valóság? Mindez még álom, de ma már talán nem is lenne lehetetlen. Ez a cikk azt írja le, milyen elvek alapján lehet mindezt megvalósítani.

1. Ez volt régen...

A világ alaposan megváltozott azóta, hogy az MTA 2003-ban használatba vette a *Köztestületi Publikációs Adattár* (KPA, 2009). Akkoriban az egyetlen elektronikus bibliometriai adatforrás a CD-n elérhető *Science Citation Index* (SCI) és weben is elérhetővé változta, a *Web of Science* (WoS) volt – amit abból nem lehetett elektronikusan előbányászni, azt bizony kézzel kellett bevinni. További közkeletű forrás a CiteSeer volt, az interneten szüretelő automata és mesterséges intelligencia alapú adatfelismerő, mely a webre kitett teljes szövegű cikkek adatait szedte össze hihetetlen teljesítményekkel – képes volt szkennelt oldalak alapján szerzőket, közleménycímeket és adatokat adatbázisba gyűjteni (hibákat azért ejtett bennük...). Ezt szorította ki később (2004-től) a Google Scholar, mely ma már folyamatosan gyűjti a weben látható

adatokat – igaz, hogy a korábbi tanulságok alapján szöveges weboldalak alapján szüretel, és ezért a tartalma kevésbé hibás.

A fenti adatok minőségét azonban nehéz volt szavatolni. Akárhogy is történt a keresés és letöltés, ez nehezen volt reprodukálható (a *Web of Science* keresési profil ma sem menthető/tölthető Magyarországról), utólag nemigen volt ellenőrizhető, és az eredmény nem is volt egyszerűen összevethető a KPA tartalmával. A CiteSeer és a Google Scholar pedig veszedelmes fegyverek: egyrészt hihetetlenül hatékonyak, másrészt természetüknél fogva egyes minőségű adatokat tartalmaznak. Néha nehéz biztosan megállapítani, hogy egy Google Scholarban látható idéző pontosan hol jelent meg, és tényleg idéző-e...

Visszatérve a KPA-hoz, az egykori, nagyrészt kézi és időigényes adatbevitel és programfejlesztéskor természetszerűleg előforduló néhány programhiba több felhasználó ellenállását is kiváltotta. A bevitel sok munkát jelentett, és a lényegében ellenőrizetlen adatbevitel hitelessége sok kételyt támasztott. Hogyan is lehetett volna jobb? Hiába vállalkoztak könyvtárosok a listák ellenőrzésére, az SCI-beli idézők kibányászására és bevitelére, a kézi bevitel hibalehetősége nagy, az „ellenőrzés” pedig megfelelő adatforrások nélkül formális jegyek alapján történő láttamozás. Ez történik ma is a KPA-ban: a doktori értekezést benyújtók és az akadémikusjelöltek listáit a KSZI munkatársai formális (konzisztencia, teljes adatok) és ésszerűségi szempontból nézik végig, ellenőrzik a besorolásokat, és az egyes tételeket *láttamozzák*. Az „ellenőrzést” vagy „jóváhagyást” nem vállalhatják fel, mert az hosszadalmas adatellenőrzéseket is jelentene. Részletes számítógépes keresések híján sokszor a könyvtárosok, az előterjesztők és a bírálók sem veszik és nem is vehetik észre,

<sup>1</sup> Repozitórium: publikációk teljes szövegét (és esetleg további szerzői anyagokat) tartalmazó, intézményi vagy tematikus jellegű archívum.

- ha egy megadott folyóiratcikk nem is jelent meg, vagy nem úgy, csak az adatai vannak betöltve,
- ha egy megadott könyv nem könyvként, csak belső kiadványként jelent meg,
- ha egy ISBN-számmal is ellátott kötet igazából nincs is kiadva és terjesztve,
- ha egy „impaktfaktoros szakcikk”-nek beírt közlemény igazából nem szabályos folyóiratcikk, hanem egy előadás bő kivonata,
- ha egy megadott, létező közlemény tévesen került ide, az adott szerző neve nem is szerepel az eredetiben,
- ha a megadott idéző az adott művet nem is idézi.

Félreértés ne essék, nem szándékosan elkövetett hibákról beszélünk, hanem a nagyrészt véletlen, de esetleg komoly tévedésekről. Az SCI-ből (WoS-ból) áttöltött idézők például általában cikkenként csoportosított állományokban vannak, de csak az adatbevitő gondosságán múlik, hogy jó közleményhez tölti-e őket. Ha téved, és később ismételt beolvasással pótolja a hiányzókat, akkor a teljes csoport továbbra is megjelenik a másik (korábban tévesen kiválasztott) közleménynél is. A könyvtáros ezt „láttamozza”, és bezárult a kör. Az MTMT-ben immár ott vannak a hibás adatok.

Ha azt tűzzük ki célul, hogy a Magyar Tudományos Művek Tára hiteles legyen, akkor ez sokkal többet igényel, mint egy láttamozás. A „közhitelességet” nem tűzhetjük ki célul, hiszen közhiteles adatbázisba változást csak az információ-jogszabály által feljogosított létrehozója jegyezhetne be, az MTMT-be viszont a szerző, illetve megbízottja vihet be adatokat. Az adatbevitel lépéseit azonban automatizálhatjuk, véletlen tévedésektől megvédhetjük, az eredményeket pedig ellenőriz-

hetjük és hitelesíthetjük. A kérdés az, hogy minderre van-e elvi lehetőség, és ha igen, van-e erre gyakorlati megoldás is?

## 2. Mit tehet egy könyvtáros?

Ha egy könyvtáros azt a feladatot kapná, hogy gondosan ellenőrizze egy szerző közleményeinek és idézőinek listáját, vajon mit tehetne? Ha ezt végiggondoljuk, a számítógéppel mindössze ezeket a lépéseket kell követnünk, vagy megfelelő számítógépes eljárással helyettesítenünk. A számítógép nem okosabb, mint egy ember, de a jól algoritmizálható feladatokat programozható módon gyorsan és precízen elvégzi.

### 2.1 Felvitel

Az első lépés a felvitel. Adatbázisunknak – amennyire lehet – védettnek kell lennie az emberi hibáktól. A kézi „pötyögős” bevitel helyett ma már, ahol csak lehet, állományfeltöltést ajánl fel az MTMT. Ez máris sokat segít, legalábbis, ha az adatfelvitő személy követi a tanácsokat. Ezen felül néhány egyszerű ellenőrzéssel (import, duplumok felismerése) fájdalommentesen javítani lehet a beviendő és a már bevitt adatok minőségét.

### 2.2 Közlemények adatainak hitelesítése

A már bevitt adatoknál a közleményeket érdemes tételesen ellenőrizni. Ma a könyvtár már nem egyszerűen papír-alapú könyvek és folyóiratok gyűjteménye: adatbázis-hozzáféréseket, elektronikus folyóirat-előfizetéseket, repozitóriumi keresőket is tartalmaz. A könyvtáros tehát nemcsak a helyi könyvtárat, hanem mindinkább a világhálót használja. Legjobb lenne a forrás közvetlen ellenőrzése a művek alapján. Ezek fizikai ellenőrzése (autopszia) azonban rengeteg munkát jelentene. Szerencsére sokuk elérhető az interneten

is: folyóiratok online változata, konferencia-kiadványok a weben stb. Ezekkel a hitelesítés megtehető. Ha ezek nem férhetők hozzá egyszerűen (az interneten sem), következnek a *elsődleges bibliográfiai források*: az ellenőrzött bevitelű adatbázisok és repozitóriumok, és könyvtári katalógusok. Ezek részben bibliográfiai adatokat tartalmaznak, részben teljes közleményszövegeket. Végül következnek *másodlagos forrásként* a „szület-alapú” adatbázisok: Google Scholar, Scirus és a nem szigorúan ellenőrzött feltöltésű repozitóriumok. Ezek nem teljesen megbízhatóak, hiszen ha például szélső esetben jól álcázott áldalakat tesz ki valaki a webre, akkor becsaphatja őket, és a bibliográfiai adataik sem ellenőrzöttek, de sokszor segítenek például a teljes közleményszövegeket megtalálni.

Mit tehet tehát a könyvtáros? Megkeresi a közleményt vagy adatainak legmegbízhatóbb forrását, és összeveti ezt a lista adott tételével. Ha minden stimmel, most már „hitelesíti”, és ennek tényét eltárolja (például az MTMT-ben): ki, mikor, mivel hasonlította össze, és találta egyezőnek az adattár egy tételét. Ha eltérést talál, akkor megkeresi az okot, és ha kell, az MTMT-ben javít, majd utána hitelesít. Ez bizalmi kérdés is: a hitelesítés konkrét, egyenkénti összehasonlítást jelent, és a könyvtáros személyében jóatáll azért, hogy látta és egyezőnek találta az adatokat.

Kalkuláljunk egy kicsit: ha a könyvtáros mindezt végrehajtotta, akkor legalábbis a közlemények fontosabb darabjait látta és hitelesítette. Egy ilyen keresés és összehasonlítás átlagosan legalább tizenöt percre becsülhető tételenként, és sajnos a meg nem talált tételekre is rá kell szánni az időt. A KPA jelenlegi 164 ezer közleményének hitelesítése kézzel ezek szerint 5100 munkanap (húsz emberév) lenne... ez tarthatatlanul sok. Ha csak a 248

MTA tagjelölt listáját néznénk, átlagosan százötven tétellel, akkor is 1160 munkanapot (4,5 emberév) kapnánk. Legyünk reálisak. Nincs ennyi könyvtárosunk/pénzünk/munkaidőnk. Számítógépre van tehát szükség.

### 2.3 Az idéző közlemények adatainak hitelesítése

Az idézőket vizsgálva még aggasztóbb a helyzet. A könyvtárosnak ezeknél nemcsak a közlemények létét és adatait kell ellenőriznie, hanem azt is, hogy ténylegesen hivatkoznak-e az adott közleményre. Vagyis kicsivel még több a feladata... ezen részben segít, hogy az importálás során az irodalomjegyzék – amennyiben az importált állományban benne van – manipulálhatatlanul importálódik a KPA-ba, vagyis a „tényleg idézi?” kérdésre az importált irodalomjegyzék alapján – legalábbis, ha az hibátlan – gyorsan válaszolni lehet. A többi azonban ugyanolyan, mint a közleményeknél... az idézők száma azonban négyszer akkora, mint a közleményeké (662 ezer darab → 82 emberév). Egyetlen szerencsénk van: az idézők sokkal gyakrabban származnak adatbázisból, mint a közlemények, vagyis az idéző közlemény valamely adatbázisban való megtalálására jóval nagyobb az esély. De kezel akkor sem reális egy részét sem ellenőrizni. Számítógépes megoldásra van szükség.

### 3. Mire képes egy megfelelően programozott számítógép?

A fenti becslésekből az következik, hogy csak számítógéppel van esélyünk elfogadható mennyiségű rekordot hitelesíteni. A kérdés most már csak az, hogy képes-e a számítógép a könyvtárost ebben helyettesíteni? A válasz szerencsére igen.

Természetesen a mechanikus szimuláció (azt szimuláljuk, ahogy a könyvtáros rápillant

va a közleményre és az adatbázis-oldalra összeveti ezeket az MTMT tételeivel) nem járható út. A számítógép így nem tud versenyezni az emberi elmével. Ugyanakkor a talált adatokat letöltve vagy az oldalt elemezve mégis összevethető a tartalmak. Ennek két kulcsképe van:

- az MTMT-oldalnak megfelelő adatbázis-oldal (az adott publikációt önállóan mutató oldal egy adatbázisban) azonosítása,
- összevetés.

### 3.1 A közlemény (idéző) adatbázis-oldalának azonosítása

Ha – ahol lehet – adatbázisokból származó állományok importjával töltjük be az adatokat, akkor az első probléma könnyen megoldható: az adatbázis-exportok általában tartalmaznak az egyes rekordok azonosítóit is. Ha az MTMT majd tárolja ezeket, akkor az adatbázis neve és az azonosító alapján az egyes rekordok máris egyértelműen azonosíthatók lesznek.

Nehezebb a helyzet a régebbi, azonosító nem tartalmazó rekordokkal. Ilyenkor a kitöltött MTMT-adatmezők felhasználásával kell keresni az egyes adatbázisok keresőiben vagy a metakeresőkben. Szerencsére ez is automatizálható: például már a teljes cím, az év és az első szerző is általában egyértelműen azonosítja a közleményt. Ugyanígy megoldható

a könyvek könyvtári katalógusokban való automatikus megkeresése, vagy a szabadalmaknak a Magyar Szabadalmi Hivatal adatbázisában való megkeresése is. Kockázat nincs: ha valamit nem találunk meg, legfeljebb nem tudjuk megállapítani az azonosítóját. De ha megvan, akkor csak rajtunk (az MTMT-n) múlik, hogy ezeket tároljuk és használjuk-e.

Ha az oldalt megtaláltuk, gyakran megkapjuk a DOI-t (Digital Object Identifier) is. Ez a teljes szövegnek és a közlemény adatainak az azonosítója, amely internetes megnyitáshoz is használható. Ha ez is megvan, akkor ezzel még egy ellenőrzési lehetőséget találtunk, és ráadásul az MTMT-ből elérhetővé is tudjuk tenni a teljes szöveget. (Természetesen csak akkor nyitható meg, ha az oldal publikus – például *open access* cikk –, vagy az olvasónak van jogosultsága a megnyitáshoz.)

#### 3.1.1 Magyarországi korlátok

A fentiek jól megtervezhetők és kivitelezhetők, de ehhez a magyarországi adatbázis-hozzáféréseket is bővíteni kell. Nem közismert például, hogy a magyarországi *Web of Science* előfizetés (Elektronikus Információszolgáltatás – EISZ) nem teljes az egyéni felhasználók számára, így az MTMT-szerver számára sem. Íme egy illusztratív példa: egy publikáció URL-je: <http://gateway.isiknowledge.com/gateway/Gateway.cgi?&GWVersion=2&Src>

**Citing Articles** Title: Congestion control in high-speed communication networks using the Smith principle  
 Author(s): Mascolo, S  
 Source: AUTOMATICA Volume: 35 Issue: 12 Pages: 1921-1935 Published: DEC 1999  
 Refined by: [excluding] Publication Years=( 2001 )  
 Timespan=All Years. Databases=SCI-EXPANDED, A&HCI, SSCI.

---

Results: 54 Page 1 of 6 Go

### 2. ábra • A 98-as számra kattintva megjelenő idézők

Auth=CustomerName&SrcApp=CustomerName&DestLinkType=FullRecord&KeyUT=000083587700005&DestApp=WOS

Erre kattintva az 1. ábra jelenik meg a képernyőn.

Ha itt a 98-ra rákattintunk, 98 idéző listáját kellene látnunk. Ezzel szemben azonban a 2. ábrán látható eredmény jelenik meg.

98 helyett 54... Az eltérés oka egyszerű: a magyar EISZ *Web of Science* előfizetése nem tartalmazza a konferenciákat (és nem tartalmazza az 1975 előtti adatokat sem, de ez utóbbi itt nem baj). A 98 darab mind valódi idézet, tehát a szerző joggal megadhatná őket az MTMT-ben, de nem látja, és nem tudja letölteni sem. Ha tudna is róluk, hitelesíteni akkor sem tudná... Magyarországon most egyáltalán nincs is publikus terminál, ahonnan megnézhetné ezeket. Ha azonban külföldről megpróbálja, akkor megnézheti és le is töltheti őket (megkérdeztem ismerőseimet Brüsszelben, Delftben és Oxfordban: mindhárman látták az összeset, mind a 98-at).

Ha a magyar kutatókat adataik feltöltésére akarjuk rávenni, akkor lehetővé kell tenni számukra a teljes hozzáférést. Az is megoldás, ha ehhez nem országos párhuzamos elérést biztosítunk minden kutató számítógépén, mint jelenleg, hanem néhány, távolról is használható számítógépen tesszük elérhetővé a teljes adatbázist, bár hozzá kell tenni, hogy ha az EISZ valóban mindenkinek fontos adatbázist fizet elő a *Web of Science* előfizetéssel, és a kutatók használják is kutatásra, akkor most éppen az információ legfrissebb részétől vannak elzárva.

Ugyanilyen rossz a helyzet akkor, ha a szerző – kapcsolatait igénybe véve – szíveségből letölteti adatait külföldről. Az MTMT-be így betöltött idézők természetesen „SCI”-idézők, de ha bármelyik MTMT-adminisztrátor ellenőrzi őket a WoS-linken (például: <http://gateway.isiknowledge.com/gateway/Gateway.cgi?&GWVersion=2&SrcAuth=CustomerName&SrcApp=CustomerName&DestLinkType=FullRecord&KeyUT=0002>)

Congestion control in high-speed communication networks using the Smith principle

Full Text Print E-mail Add to Marked List Save to EndNote Web Save to EndNote, RefMan, ProCite  
more options

Author(s): Mascolo S

Source: AUTOMATICA Volume: 35 Issue: 12 Pages: 1921-1935 Published: DEC 1999

Times Cited: 98 References: 39 Citation Map

### 1. ábra • *Web of Science* példa

Your search (shown below) found no records.

Search hints:

- Check the spelling of your search terms.
- Check the syntax of your search query. Compare your query to the search examples on the search page and in Help.
- If you are searching for an author, look up the name in the author index. Remember to enter last name followed by initial(s). For example, **Smith A.**
- Use a wildcard to find plurals and word variants. For example, replace **graphite nanofiber** with **graph\* nanofib\***.
- Search for multiple terms that describe the same concept. For example, **cell\* phone\* OR mobile phone\***.
- Change the timespan setting if you did not search all available data. For example, if you searched five years of data, increase the timespan to ten years.

### 3. ábra • Magyarországról nem látható rekord „találati információja”



63416500123&DestApp=WOS), a 3. ábrát kapja a böngészőben.

Vagyis azt gondolhatja, hogy „a szerző által megadott adat helytelen, mert az idéző nem is létezik a *Web of Science*-ben”. És törlő, vagy rosszabb esetben kénytelen például az MTA Doktori Tanácsának jelezni a „nem valós” adatot vagy adatokat ... ennek elkerülésére az MTMT-adminisztrátorok (intézményi adminisztrátorok) számára muszáj lesz teljes hozzáférést biztosítani, legalább az MTMT-szerveren keresztül.

Hasonló a helyzet több fontos adatbázissal. Noha van olyan egyetem, ahol a *Scopus* saját forrásból előfizetve elérhető (például SOTE, DE, PE), az idézeteket (irodalomjegyzékeket) tartalmazó rekordokat az MTMT még hitelesítési célból sem tudja megnyitni, az azonosító ismeretében sem. Ez nem is csak rajtunk múlik, hanem a *Scopus* egyelőre kizárólag csak „campus licence”-eket támogató üzletpolitikáján. Vagy az *IEEE Xplore* adatbázisa kereshető ugyan a BME–OMIKK weboldaláról, de az előfizetésben a letöltés, illetve a teljes cikkek is csak az elmúlt tizenkét évre érhetőek el – ez még a gyorsan változó villamosmérnöki szakterületen is kevés.

### 3.2 Összevetés

#### 3.2.1 Közlemények és idézők bibliográfiai adatainak verifikálása

Az azonosító ismeretében ellenőrizhetjük (verifikálhatjuk) az MTMT-oldalt (4. ábra)

<p>I Kollár, J J Improved Determination of the Best Fitting Sine Wave in ADC Testing. <b>IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT</b> 54:(5) pp. 1978-1983. (2005) IF: 0.665 WoS link Teljes dokumentum DOI: 10.1109/TIM.2005.855082 Folyóiratcikk/Szakcikk/Tudományos</p> <p>Scopus link, Google scholar link</p> <p>Irodalomjegyzék Független idéző: 21 Független idéző: 4 Összesen: 25</p>
---

#### 4. ábra • Egy MTMT-rekord

a *Web of Science* oldal alapján (5. ábra). Ez azt jelenti, hogy ismét a könyvtáros viselkedését szimulálhatjuk úgy, hogy vesszük az egyes *mezőtartalmakat*, és összevetjük az internetes adatbázisoldal *tartalmával*. Technikailag ez megtehető a látható oldal mögötti adatok segítségével. Sőt, még ennél is jobban megközelítjük a hozzáértő ember viselkedését, ha az emberi felhasználó számára is elfogadható kis eltéréseket megengedünk: ha a fő adatok stimmelnek, akkor bizonyos gyakori gépelési hibákat is tolerálhatunk: betűk felcserélése, melléütés, sőt felismerhetünk rövidítéseket is. Ha pedig nem tudjuk az adatokat verifikálni, akkor a könyvtáros számára jelezhetünk: nézze meg kézzel, mi az az eltérés, amit a számítógép már nem tud kezelni.

#### 3.2.2 Idézők verifikálása

Az idézők adatai mellett fontos annak ellenőrzése is, hogy az idézők ténylegesen idézik-e az adott közleményt. Ez nehéz feladat, mert az irodalomjegyzék-tételek gyakran nem pontosak, sőt, az adatbázisokba sokszor tovább rövidítve vagy elírással kerültek be. Ugyanakkor a számítógépes program képes néhány olyan lépésre is, amelyek kézzel eléggé reménytelenek lennének, legalábbis nagyszámú tétel esetén.

Az egyes közlemények idézőinek letöltésekor a számítógép képes megjegyezni, melyik tételhez tartoznak az idézők (az adatbázisokból mentett állományokban sajnos

#### Improved determination of the best fitting sine wave in ADC testing

Full Text Print E-mail Add to Marked List Save to EndNote Web Save to EndNote, RefMan, ProCite more options

Author(s): Kollar I, Blair JJ

Source: IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT Volume: 54 Issue: 5 Pages: 1978-1983 Published: OCT 2005

Times Cited: 8 References: 10 Citation Map

**Abstract:** The sine wave test of an analog-to-digital converter (ADC) means to excite the ADC with a pure sine wave, look for the sine wave which best fits the output in least squares (LS) sense, and analyze the difference. This is described in the IEEE standards 1241-2000 and 1057-1994. Least squares is the "best" fitting method most of us can imagine, and it yields very good results indeed. Its known properties are achieved when the error (the deviation of the samples from the true sine wave) is random, white (the error samples are all independent), with zero mean Gaussian distribution. Then, the LS fit coincides with the maximum likelihood estimate of the parameters. However, in sine wave testing of ADCs, these assumptions are far from being true. The quantization error is partly deterministic, and the sample values are strongly interdependent. For sine waves covering less than, say, 20 quantum levels, this makes the sine wave fit worse than expected, and since small changes in the sine wave affect the residuals significantly, especially close to the peaks, ADC error analysis may become misleading. Processing of the residuals [e.g., the calculation of the effective number of bits, (ENOB)] can exhibit serious errors. This paper describes this phenomenon, analyzes its consequences, and suggests modified processing of samples and residuals to reduce the errors to negligible level.

Document Type: Article

Language: English

**Author Keywords:** analog-to-digital converter (ADC) test; analog-to-digital converter; effective number of bits (ENOB); four-parameter method; IEEE Standard 1057-1994; IEEE Standard 1241-2000; least squares (LS); sine wave fitting; three-parameter method

**Reprint Address:** Kollar, I (reprint author), Budapest Univ Technol & Econ, Dept Measurement & Informat Syst, H-1521 Budapest, Hungary

Addresses:

1. Budapest Univ Technol & Econ, Dept Measurement & Informat Syst, H-1521 Budapest, Hungary  
2. Bechtel Nevada, Las Vegas, NV 89193 USA

**E-mail Addresses:** kollar@mit.bme.hu, blairjj@nv.doe.gov

**Publisher:** IEEE-INST ELECTRICAL ELECTRONICS ENGINEERS INC, 445 HOES LANE, PISCATAWAY, NJ 08855 USA

**Subject Category:** Engineering, Electrical & Electronic; Instruments & Instrumentation

IDS Number: 971AZ

ISSN: 0018-9456

DOI: 10.1109/TIM.2005.855082

#### 5. ábra • Hitelesítésre használható *Web of Science* oldal

nincsen erre vonatkozó adat), és az idézőket betöltheti a megfelelő tételhez, különösen, ha ezt nem részadatok, hanem azonosító alapján találja meg.

Képes a betölthető irodalomjegyzékeket pótolni, ha ezek még nincsenek betöltve.

Az ismert adatok felhasználásával megkísérülheti az irodalomjegyzékben az idéző sorok azonosítását. Mivel az MTMT-tételt ismerjük, az első szerzői név, az év és a kezdőoldal gyakran automatikusan felismerhető. Ha mégsem, még mindig felajánlhatjuk a szerzőnek, hogy jelölje meg az idéző tételt: ha ez gyorsan és egyszerűen megtehető, akkor meg is fogja tenni. Ha nem teszi, nem baj: a tétel megmarad „nem teljesen azonosított

idéző”-ként. Mindenesetre a tévesen bevitt tételek így hamar kiderülnek.

#### 4. Az idézők frissítése

Végül a számítógépes program arra is képes, amire a könyvtáros nem: az újabb idézőket az adatbázisok segítségével rendszeresen figyelheti, és automatikusan áttöltheti. Ez azt jelenti, hogy az *azonosított* közleményeknél *szerzőadatbeviteli munka nélkül* bővítheti az idézők listáját. Hasonlóan működik a *ResearcherID* is (ResearcherID, 2009): a *Web of Science*-ben a szerzői lista alapján folyamatosan gyűjti az idézeteket. Ennél azonban az MTMT jelentősen többet mutathat: az *összes* ilyen adatbázisban megjelenő idézeteket au-

tomatikusan egységes listába gyűjtheti, és kombinálva mutathatja meg. Sőt: a Google Scholarban megjelenő idézeteket is figyelni/gyűjteni lehet: ezeknél azonban általában a szerző vagy adatbevivő kézi ellenőrzése is szükséges, mert az adatok egy része elfogadhatatlanul hiányos vagy pontatlan.

Van egy további számítógépes lehetőség, mely sok szerzői munka nélkül meglehet további, másképp meg nem található, de létező hivatkozásokat. A CD-n, pendrive-on megkapott, ill. a weben megtalálható teljes szövegű közlemények egy része nincs a Google Scholarban sem indexelve (pl. olyan konferencia-kiadványok egy része, amelyeket megkaptunk konferencia-résztevőként; korlátozott hozzáférésű weboldalak, melyeket a Google nem tud indexelni, de mi hozzáférünk stb.). Itt is azt kell tenni, amit kézzel tennénk, csak gyorsabban és megbízhatóbban. Az egyes cikkek szövegét meg kell nyitni, az irodalomjegyzékben névre (esetleg csak első névre) keresni, és találat esetén az MTMT adatai segítségével felismerni, hogy a mi közleményünkről van-e szó. Ez PDF-állományokkal is megtehető (Citex, 2009). Az ilyen idézők sokszor nem hitelesíthetők, de ezek is létező idézetek, és a szerzőknek fontosak lehetnek. Természetesen nem tudunk a Google Scholarral versenyezni, de adott publikációs körben, adott rendezvény cikkei között, adott típusú közleményekben, esetleg adott szerzők és doktoranduszai cikkeiben keresve jó eséllyel találhatunk még nem ismert idézeteket. A számítógép pedig nem kér enni: ha elindítjuk a keresést, magától gyűjtöget, mi meg közben tehetjük a dolgunkat.

##### 5. Adatbázisokban nem szereplő tételek

Lelkesedésünket kissé lehűtheti a kérdés: mi lesz azokkal a tételekkel, amelyek semelyik

adatbázisban sem található meg, nemcsak a *Web of Science*-ben vagy a *Scopus*-ban nem, de még a *Matarka*-ban sem? A kézi bevitt még csak-csak kiválthatjuk állomány-előképzéssel és -beolvasással (Publex, 2009), de hogyan hitelesítsük őket?

E tételek sajnos nem hitelesíthetők számítógéppel, mert nincs mihez. A fontosabb tételeket szükség esetén az „igazi” (fizikailag megfogható) közlemények alapján egy könyvtáros hitelesítheti. Amit pedig egyáltalán nem lehet hitelesíteni, azt nem hitelesítjük. Ettől még irodalomjegyzék-tétel maradhat, csak a „hitelesség pecsétje” nélkül. Végül is nem muszáj mindent hitelesíteni, csak a fontosabb tételeket. Mindenesetre az STM- (Scientific, Technical, Medical) tudománycsoportban a közlemények többségét feltehetően hitelesíteni lehet, a kevésbé számítógépesített tudományágaknál (például társadalom- és humán tudományok) pedig csak kisebb részüket.

##### 6. Összefoglalás

Az internet mai állapota és számítógépeink lehetővé teszik, hogy az adatbevitel jelentős részét automatizáljuk, bibliográfiai adatbázisunkat megterhelő emberi munka nélkül is kibővítsük, és automatikusan ellenőrzötté tegyük. Megvannak az eszközeink: csak használnunk kell őket, és végre elfelejthetjük a kínos, hosszadalmas kézi bevittet és ellenőrzést. A szerzőket pedig így kötelezni sem kell a bevitelre, hanem csak „szórakoztatni”... De amíg nem jutunk el az automatizmusokig és a hitelesítésig, addig is feltölthetjük adatainkkal az adatbázist. Az adatok pontossága most a szerzők kizárólagos felelőssége, a „hitelesítési pecsét” majd a fejlesztés függvényében kerülhet rájuk. Az előterjesztők és a bírálók a kiszámolt fontosabb tudánymetriai számokat addig kézzel ellenőrizni tudják. A

szomszéd országok is megindultak az adatbázisok létrehozása irányába, tehát ez nem hungarikum, hanem alkalmazkodás a nemzetközi tendenciához. Magyarország ebben

a világ élvonalába ugorhatna, bár az átlaghoz képest most is jól állunk. A meglévő tudás csodákra lehet képes. Miért ne kerüljünk végre egyszer mi helyzeti előnybe?

Köszönöm a kéziratához fűzött értékes megjegyzéseket kollégáimnak, különösen Dobos Anikónak, Dobrowiecki Tadeusznak, Kmety Andreának, Kurutzné Kovács Mártának és Mészáros Tamásnak.

Kulcsszavak: *EISZ, Elektronikus Információs szolgáltatás, Web of Science, Scopus, IEEE Xplore, Magyar Tudományos Művek Tára, MTMT, Köztisztviselői Publikációs Adattár, KPA, hitelesítés, bibliográfia, frissítés, automatikus feltöltés*

##### IRODALOM ÉS ADATBÁZISCÍMEK

- Citex (2010): Citex idézetgyűjtő program. <http://www.mit.bme.hu/services/pubinfo/citex/>
- Csurcsia Péter Zoltán (2010): *Bibliográfiai adatok hitelesítése a Magyar Tudományos Művek Tárában. Diplomateru*. BME MIT.
- EISZ (2010): Elektronikus Információs szolgáltatás <http://www.eisz.hu/>
- Falagas, Matthew E. – Pitsouni, Eleni I. – Malietzis, George A. – Pappas, G. (2008): Comparison of PubMed, Scopus, Web of Science, and Google Scholar: Strengths and Weaknesses. *The FASEB Journal*. 22, Feb, 338–342. <http://www.fasebj.org/cgi/reprint/22/2/338.pdf>
- Google Scholar (2010): <http://www.scholar.google.com/>

- IEEE Xplore (2010): <http://ieeexplore.ieee.org/Xplore/dynhome.jsp> (automatikus belépés BME domainből indítva)
- KPA (2010): *Általános tájékoztató az MTA Köztisztviselői Publikációs Adattár (KPA) használatához*. [http://www.mtakoztest.hu/kpa\\_tah\\_at.htm](http://www.mtakoztest.hu/kpa_tah_at.htm)
- Matarka (2010): Magyar folyóiratok tartalomjegyzékinek kereshető adatbázisa. <http://www.matarka.hu/>
- MathSciNet (2010): <http://www.ams.org/mathscinet/>
- Publex (2010): Szöveges állományok bevitelére a BME-PA-ba. <http://www.mit.bme.hu/services/pubinfo/publex/>
- Pubmed (2010): <http://www.ncbi.nlm.nih.gov/pubmed/>
- ResearcherID (2010): <http://www.researcherid.com>
- Scopus (2010): <http://www.scopus.com/>

