



## A Google Scholar helye a társadalomtudományokban

### Mi a Google Scholar?

A *Google Scholar* (<http://scholar.google.com>) 2004. novemberi megjelenése idején a világ legnagyobb tudományos könyvtárának kiáltották ki. A még ma is béta-verzióban futó költségmentes keresőmotor olyan dokumentumokat is megtalál, amelyeket más keresők nem. A Scholar egyik erőssége éppen az, hogy végigpásztazza a láthatatlan webet, a háló szabadon nem elérhető részét, és onnan is megjeleníti a releváns találatokat. Tudományági szempontból a Google Scholar hatóköre legátfogóbb a tudományos, technikai, orvosi területen, az ún. STM diszciplínákban.

### Google Scholar: a hagyományos könyvtári források bitorlója?

Ellentmondásos kép alakult ki a Scholarról. Vannak, akik dicsérik, vannak, akik kritikával illetik. *Jacsó Péter* a *Peter's Digital Reference Shelf*ben rámutat arra, hogy a többi adatbázissal szemben a Scholar tárgyszavazása nem ellenőrzött. A szerzők neve, a folyóiratok címe különböző írásmóddal szerepel. Az összetett keresésben a kiadványra keresés csak egyfajta szűkítést enged, csonkolni és kifejezésre keresni itt nem lehet. Jacsó azt is kifogásolja, hogy a keresés eredményét nem lehet rendezni, letölteni, vagy e-mailben elküldeni, mint ahogy ez a fizetős adatbázisok esetében természetes.

Fontos a Google Scholar és az előfizetéses adatbázisok összevetése tartalmi szempontból. Az eddigi kisebb terjedelmű tesztek alapján a Scholarnál jobb bizonyítványt kapott az előfizetéses multidiszciplináris citációs adatbázis, a *Web of Science (ISI)*, a *Scopus*, az Innovative Interfaces Inc. által működtetett *Metafind*, a *PsycINFO* és az *ATLA*, valamint az Elsevier *Scirus* nevű tudományos keresőmotorja.

### Elemzés

A két szerző munkahelyén, a *Dél-kaliforniai Egyetem (USC)* könyvtárában megvizsgálta, hogy a Google Scholar beiktatható-e az oktatásba és a könyvtári referenzforrások közé. Közismert tény, hogy az USC diákjai széles körben használják, mert ismerős felületen könnyű keresést kínál. A hallgatók a Scholar esetében általában nem fordulnak a könyvtárhoz, hanem maguk használják, ezért nem könnyű megtudni, hogy milyen eredménnyel keresnek a Scholarban, mennyire felelnek meg a keresőkérdésre kapott válaszok. Ha a Google Scholar beválik, a felsőoktatási intézményeknek nem kell előfizetniük a drága adatbázisokra.

### Módszer

Saját vizsgálatok érdekében szükség volt a Google Scholar tüzetes vizsgálatára, ezért összehasonlították azt elismert, előfizetéses társadalomtudományi adatbázisokkal. Erre a célra a *PsycINFO-t*, az *ERIC-et* és a *Social Science Citation Indexet (SSCI)* választották. A keresendő tárgyszó a „homeschooling” (magántanulás) volt, mert annak pedagógiai és pszichológiai vonatkozása is van. A cél a Google Scholar és az előfizetéses adatbázisok eredményének összehasonlítása volt. Az ehhez végzett keresésben a tárgyszót mindkét írásmóddal („homeschooling” és „home schooling”) megadták. A Google Scholar és a három adatbázis eredményének első száz tételét kinyomtatták, majd a kiválasztott elemeket rögzítették egy Excel-táblázatban: a kiadvány típusa (könyv, folyóiratcikk stb.), lektorált vagy nem lektorált címek; a kereső szó címben, tartalmi összefoglalóban, a teljes szövegben vagy deskriptorként szerepel; a publikálás dátuma, valamint az átfedések a Google Scholar és a három előfizetéses adatbázis eredménye között. Egy saját készítésű, tízpontos értékelési listát alkalmaztak a relevancia további mérésére. Tíz pontot ért, ha a publikáció címében előfordult a

„homeschool”, a „homeschooling” vagy a „home-schooled”; kilenc pontot, ha a címben nem, csak az összefoglalóban vagy a deskriptorok között. Ha csak a szövegben fordult elő, az két pontot kapott. Az eredményt az 1. táblázat mutatja.

1. táblázat

**Tartalom, előfordulási gyakoriság és relevancia**

	Google Scholar	PsycINFO	SSCI	ERIC (alapértelmezett keresés)	ERIC (relevancia)
Tartalom	2200	105	65	602	602
Aktualitás	1983–2002	1980–2005		1999–2003	1984–2003
Relevancia (%)	9,8	7,0	9,6	8,54	9,62

**Tartalom**

A Google cég saját definíciója szerint a Google Scholar anyaga a tudomány összes területe, tudományos irodalom, lektorált művek, disszertációk, könyvek, különnyomatok, kivonatok, műszaki jelentések. Forráslista nem található, de tudni lehet, hogy az *ACM*, a *Nature*, az *IEEE*, a *PubMed* és az *Open World Cat* is partner. A táblázatra pillantva nyomban szembetűnik, hogy a Scholar 2200-as találatával szemben a többi adatbázis lényegesen szerényebb találatot produkált. De nemcsak a találatok száma figyelemre méltó, hanem a változékonysága is. A Scholar száz vizsgált tételéből negyven folyóirat (harminckettő lektorált), tizenhat könyv, tizenöt magazin, hét disszertáció, hat ERIC-dokumentum, öt újságcikk, három honlap, két konferenciakötet és egy-egy monográfia, hírlevél és kormányzati kiadvány.

A kizárólag folyóiratokat, könyveket, disszertációkat és kutatási jelentéseket feldolgozó PsycINFO lekérdezése ötvennégy folyóiratot (ebből ötvenhárom lektorált), huszonkilenc disszertációt és huszonkét könyvet eredményezett. A csak folyóiratcikkeket tartalmazó Social Science Citation Index eredménye: hatvanöt folyóirat (ebből negyvenkilenc lektorált), de közülük tizenkilenc könyvrecenzió, öt cím kétszer jelent meg, tizenegy szerkesztőségi cikk volt. Az ERIC a források szempontjából nagyobb változatosságot mutatott találati halmzában, szerepeltek folyóiratok, könyvek, disszertációk, ERIC-publikációk, kutatási jelentések és különféle egyéb dokumentumok. A nem rangsorolt, azaz alapértelmezett keresés ötvenhat ERIC-do-

kumentumot eredményezett. Ezek közül kilenc könyvként is szerepelt, harminc folyóiratcikk (huszonegy lektorált), tizenhárom magazin- és egy újságcikk volt.

Az eredmények értékelése alapján a két kutató az előzetes vizsgálatokat némileg cáfolva megállapította, hogy a Google Scholar nagyobb változatosságot mutat, mint a három díjköteles adatbázis, és nagyobb találati halmazt is kapunk. Nem bizonyított viszont a tételek tudományossági szintje és relevanciája. A minőség-ellenőrzés kevésbé szigorú volta miatt a Scholarban való keresést körültekintő módon kell végezni.

**Aktualitás**

Amint azt az 1. táblázat is mutatja, a Scholar-találatok nem a legfrissebb publikációk közül valók: 1983 és 2002 között jelentek meg, ebből 25% 2000-ben vagy később, 50% pedig 1995-ben vagy korábban. A PsycINFO-közlemények 1980 és 2005 közöttiek, 62% 2000-es keltezésű vagy későbbi, és csak 21% 1995-ös vagy korábbi. Az SSCI-eredmények 1983 és 2005 közötti publikációk, 40% 2000 utáni, és csak 37% 1995-ös vagy korábbi. Az ERIC alapértelmezett keresése 1993 és 2003 között talált irodalmat, ebből 95% 2000-ből való vagy későbbi, és csak 5% volt 1999-nél korábbi. A nem alapértelmezett keresés eredménye: a publikációk 1984 és 2003 között jelentek meg, ebből 46% 2000-es vagy későbbi, és csak 23% 1995-ös vagy korábbi.

A Google Scholar kevésbé friss, mint a három előfizetéses adatbázis. Ez valószínűleg abból adódik, hogy az idézett műveket dolgozza fel, az idézettségre pedig a régebbi publikációknak van nagyobb esélyük. Nem lehet tudni, milyen gyakran frissíti az adatokat, de úgy tűnik, hogy rendszeretlenül.

**Relevancia**

A keresési eredmény azt mutatja, hogy a Google Scholar első 100 találatának nagy a relevanciája a tágabb tematika szempontjából, de a szűkebb, pszichológiai szempontból már kisebb. A vizsgált 100 válaszban a keresett „homeschooling” majdnem minden címben szerepelt, ezért az értékelésnél 9,8 pontot kapott. Az 1. táblázat mutatja, hogy a PsycINFO 7,0, az SSCI 9,6, az ERIC alapértelmezett 8,54, az ERIC releváns keresés 9,62 pontot kapott.

A Google Scholar a relevanciát a szöveg, a szerző, a hivatkozások száma alapján értékeli. A PsycINFO a deskriptormezőben megjelenő első nyolc deskriptor alapján, az SSCI pedig annak alapján, hogy a keresőkifejezés hányszor fordul elő a rekordban. Az ERIC a keresőszó előfordulási száma és a szövegbeli helye kombinációjával értékeli a relevanciát.

A könyvtáros szerzők értékelése szerint a PsycINFO 30%-ban *irreleváns* eredményt adott. Dátum és relevancia szerint is értékelték az adatbázisokat. A PsycINFO első húsz helyezette a relevancia alapján 4,9 pontot, a dátum szerint 5 pontot kapott. Az SSCI első húsz helyezette a dátum szerint 9,7 pontot, míg a relevancia szerint csak 9 pontot, az ERIC ugyanezen szempontok szerint 8,65 és 9,65 pontot kapott.

A vizsgálat alapján megállapítható, hogy a Google Scholar relevanciaszintje magas, bár a tudományági relevanciához a találati halmaz tárgyi rendezése javallott. Meglepő viszont a PsycINFO és az SSCI alacsonyabb relevanciaszintje.

### Átfedés

A Scholar első száz találatában a PsycINFO-val 3%, az SSCI-vel 8%, az ERIC alapértelmezett lekérdezéssel 2%, az ERIC relevancia szerinti lekérdezéssel 8% az átfedés. Ha a teljes, 2200-as halmazra számítással kiterjesztjük a vizsgálatot, akkor az átfedés a PsycINFO esetében 31%, az SSCI esetében 25%, az ERIC alapértelmezett lekérdezés esetében 17% és az ERIC relevancia szerinti lekérdezés esetében 29% (2. táblázat).

A táblázatból látható, hogy nagy átfedés csak a PsycINFO-ban volt. Általában elmondható, hogy a három adatbázis és a Google Scholar keresési eredményei jelentősen különböznek egymástól.

### További megfigyelések

Talán a minőség-ellenőrzés hiányának és egy tétel ismételt idézettségének tudható be, hogy a Google Scholarban való kereséskor ugyanazok a dokumentumok többször megjelennek egy találati halmazban. Megemlíthető még, hogy a Google Scholarban régebbi dokumentumokat és kevesebb teljes szövegű anyagot találunk, mint a három előfizetési adatbázisban.

2. táblázat

### Átfedés a Google Scholar és a három adatbázis között (%)

	PsycINFO	SSCI	ERIC (alapértelmezett lekérdezés)	ERIC (relevancia szerinti lekérdezés)
Az első 100 eredményen belül	3	8	2	8
A teljes 2200 eredményen belül	31	25	17	29
Google Scholarban (de nem a keresési eredményen belül)	46	3	10	5
Nincs a Google Scholarban	23	72	73	66

A Google Scholar jelenleg még béta-verzióban működik, ám várható, hogy a végleges verzió üzembe helyezésével jelentős keresővé válik. Jelenlegi állapotában még nem versenyezhet a három indexelt társadalomtudományi adatbázissal, de mivel a kutatók, különösen az egyetemi hallgatók szívesen használják, a könyvtárosoknak fel kell készülniük a kezeléséből.

/GARDNER, Susan–ENG, Susanna: *Gaga over Google? Scholar in the social sciences.* = Library High Tech News, 22. köt. 8. sz. 2005. p. 42–45./

(Wéber Katalin)