

A tudományos teljesítmény referenciaadatbázis alapú meghatározása pályázati döntéselőkészítés támogatásához

Tartalomjegyzék

Vezetői összefoglaló	2
Referencia adatbázis a IV, V, VI, VII, VIII, X, és XI tudományterület esetén	3
Referencia adatbázis a I, II, III, és IX tudományterület esetén	4
Számolt tudománymetriai mutatók IV, V, VI, VII, VIII, X, és XI. tudomány területek esetén	5
Számolt tudománymetriai mutatók I, II, III, és IX tudományterületek esetén	5
OTKA kutatók statisztikai elemzése.....	6
OTKA zsűrin belüli elemzések	6
OTKA honlap scientometriai mutatók	7
GYAKRAN FELTETT KÉRDÉSEK	7
Mi történik osztott első/utolsó szerzőség esetén?	7
Miért lógnak ki egyes kutatók a grafikonból?.....	7
Figyelembe vesszük a PhD megszerzésének évét?	8
Miért nem csak az elmúlt 10 év közleményeit vizsgáljuk?	8
Miért nem a D1-es cikkek számát alkalmazzuk a kiválóság mérésére?	8
Figyelembe vesszük gyermekek nevelését?	9
Mi történik egyező értékek esetén?.....	9
Mikor lesz frissítve az adatbázis?	9
Mit mutat a Database fül?	9
Mi történik 0-ás érték esetén?.....	10
A duplázás mindenképpen kétszeres növelést jelent?	10

Vezetői összefoglaló

Egy adott OTKA felhívás különböző tudományterületeket foglal magában, amelyekre különböző korú és különböző tudományos teljesítményű kutatók pályáznak. A pályázatok bírálata során fontos, hogy egy olyan módszer álljon rendelkezésre, amellyel az egyes pályázókat és az egyes tudományterületeket egymással össze lehet hasonlítani.

A döntéstámogató rendszer kifejlesztésének tudományos hátterét az adja, hogy irodalmi referenciák alapján a korábban nagyobb tudományos teljesítményt felmutató kutatók a későbbiekben is nagyobb publikációs teljesítményt fognak elérni.

Az adott tudományterületen belüli összehasonlítás célja a pályázók rangsorolása a támogatási sorrend felállításához, míg a tudományterületek összehasonlítása során a cél, hogy ki lehessen szűrni, hogy az adott területre pályázók a jelen felhívásra pályázatot nem benyújtó kutatóknál jobbak vagy rosszabbak. Ez az összehasonlítás lehetővé teszi, hogy egy olyan tudományterület, ahova egy adott évben döntően kiemelkedő kutatók pályáznak, ne ugyanakkora relatív súlyt kapjon, mint egy olyan terület, ahova csak az átlagnál rosszabb kutatók pályáznak. Mindkét esetben az egyszerű numerikus rangsorolás gátja, hogy az egyes pályázók életkora eltérő.

<p>Az elemzés célja, hogy mind a pályázókat, mind a tudományterületeket egymással objektív módon hasonlítsuk össze azáltal, hogy minden egyes kutatót egy közös referenciaadatbázishoz hasonlítunk, amely a hasonló korú, és egyező tudományterületen aktív magyar kutatókat tartalmazza.</p>

Referencia adatbázis a IV, V, VI, VII, VIII, X, és XI tudományterület esetén

Az MTMT online rendszeréből (<https://www.mtmt.hu>) letöltöttük

- az összes MTA doktora,
- az összes akadémikus,
- az összes Lendület-kutató,
- és az összes 2006 utáni OTKA pályázó valamennyi közleményét, és a közleményekre kapott valamennyi citációt. Az adatbázis tehát minden OTKA pályázót tartalmaz, azt is, aki például nem nyert egyszer sem, csak beadott pályázatot.
- további 10 ezer olyan kutatót, ahol ismert volt a tudományterület és az MTMT azonosító

Az utolsó letöltés dátumát az „MTMT cutoff date:” mutatja a honlap baloldali menüsávjában.

Az idézettség számítása során kizárólag a független idézőket vettük figyelembe, vagyis olyan idézőket, ahol az idézett és az idéző közleménynek nincs közös szerzője. Emellett az idéző közleménynek a következő négy kategória egyikéből kell származnia:

- szakcikk,
- összefoglaló cikk,
- sokszerzős vagy csoportos szerzőségű szakcikk,
- rövid közlemény.

Az idézettség – és később a H-index – értékének a meghatározásánál az ezektől eltérő kategóriák (pld. disszertáció, konferencia absztrakt) nincsenek beszámítva.

Az egyező nevű kutatók esetén az MTMT azonosító alapján történt az azonosítás. Az OTKA kutatók vizsgálata során a referencia adatok meghatározásakor azokat a kutatókat nem vettük figyelembe, akik 2017 óta nem frissítették az MTMT fiókjukat.

Az összes kutatóra letöltöttük a születési évet a doktori.hu oldalról, illetve felhasználtuk az OTKA pályázatban megadott születési évet.

A kutatók tudományterületi besorolása köztestületi tagok esetén az MTA honlapján szereplő beosztás alapján, a Lendület pályázók esetén a téma alapján, az OTKA pályázók esetén pedig a választott zsűri alapján történt meg.

Az összes folyóirathoz hozzárendeltük a tudományterületen belüli rangot a SCImago adatbázis felhasználásával (<http://www.scimagojr.com>). Az elemzések során kizárólag a Q1-es (felső kvartilis) rangú folyóiratokban megjelent közleményeket vettük figyelembe, mivel csak ezek teljesítik a „tudományos kiválóság” feltételét. A Q rangok csak az 1999-es és későbbi közleményekre érhetőek el.

Referencia adatbázis a I, II, III, és IX tudományterület esetén

Az MTMT online rendszeréből (<https://www.mtmt.hu>) letöltöttük

- az összes MTA doktora,
- az összes akadémikus,
- az összes Lendület-kutató,
- és az összes 2006 utáni OTKA pályázó valamennyi közleményét, és a közleményekre kapott valamennyi citációt. Az adatbázis tehát minden OTKA pályázót tartalmaz, azt is, aki például nem nyert egyszer sem, csak beadott pályázatot.
- további 10 ezer olyan kutatót, ahol ismert volt a tudományterület és az MTMT azonosító

Az utolsó letöltés dátumát az „MTMT cutoff date:” mutatja a honlap baloldali menüsávjában.

Az idézettség számítása során kizárólag a független idézőket vettük figyelembe, vagyis olyan idézőket, ahol az idézett és az idéző közleménynek nincs közös szerzője.

Az idézettség – és később a H-index – értékének a meghatározásánál az összes tudományos közleményt figyelembe vettük.

Az egyező nevű kutatók esetén az MTMT azonosító alapján történt az azonosítás. Az OTKA kutatók vizsgálata során a referencia adatok meghatározásakor azokat a kutatókat nem vettük figyelembe, akik 2017 óta nem frissítették az MTMT fiókjukat.

Az összes kutatóra letöltöttük a születési évet a doktori.hu oldalról, illetve felhasználtuk az OTKA pályázatban megadott születési évet.

A kutatók tudományterületi besorolása köztestületi tagok esetén az MTA honlapján szereplő beosztás alapján, a Lendület pályázók esetén a téma alapján, az OTKA pályázók esetén pedig a választott zsűri alapján történt meg.

Számolt tudományometriai mutatók IV, V, VI, VII, VIII, X, és XI.

tudományterületek esetén

Minden kutató minden életkorára kiszámoltunk három paramétert, amelyek:

1. **A kutató H-indexe**, amelynek számítása során az összes közleményre kapott valamennyi független citációt figyelembe vettünk. Egy n-es H-index azt mutatja meg, hogy a kutatónak n darab olyan közleménye van, amelyre n idézettség érkezett. A H-index a kutató eddigi teljes életművének hatását mutatja meg.
2. **A kutató összesített éves független idézettsége**. Ennek számítása során az utolsó teljese év értékével számoltunk. Az első/utolsó/levelező szerzős szerzős közlemények idézettségét duplán számoltuk. Az idézettség megmutatja, mekkora a kutató eddigi összesített tudományos teljesítményének a hatása a jelenben.
3. **A kutató Q1-es közleményeinek száma az elmúlt 5 évben**. Az élettudományok és élettelen természettudományok esetén ezen mutató csak az első-, a levelező- és az utolsó-szerzős közlemények figyelembevételével került kiszámításra. A Q2/Q3/Q4-es rangú folyóiratokban megjelent közleményeket nem vettük figyelembe. Az elmúlt 5 évben közölt minőségi publikációk száma azt mutatja meg, hogy milyen a kutató jelenlegi publikációs aktivitása.

Megjegyzés: a "minden életkorra" történő számítás során egy például 50 éves kutató 25,26,27... 49,50-éves korára is kiszámoltuk mindegyik paramétert. Ez teszi később lehetővé, hogy egy 30 éves kutató teljesítményének elemzése esetén a referencia-adatbázisból minden kutató 30 éves kori értékét vehessük figyelembe az adott tudományterületen belül.

Számolt tudományometriai mutatók I, II, III, és IX tudományterületek esetén

Minden kutató minden életkorára kiszámoltunk három paramétert, amelyek:

4. **A kutató H-indexe**, amelynek számítása során az összes közleményre kapott valamennyi független citációt figyelembe vettünk. Egy n-es H-index azt mutatja meg, hogy a kutatónak n darab olyan közleménye van, amelyre n idézettség érkezett. A H-index a kutató eddigi teljes életművének hatását mutatja meg.
5. **A kutató összesített éves független idézettsége**. Ennek számítása során az utolsó teljes év értékével számoltunk. Az idézettség megmutatja, mekkora a kutató eddigi összesített tudományos teljesítményének a hatása a jelenben.

6. **A kutató közleményeinek száma az elmúlt 5 évben.** Az elmúlt 5 évben közölt minőségi publikációk száma azt mutatja meg, hogy milyen a kutató jelenlegi publikációs aktivitása. *Megjegyzés: a "minden életkorra" történő számítás során egy például 50 éves kutató 25,26,27... 49,50-éves korára is kiszámoltuk mindegyik paramétert. Ez teszi később lehetővé, hogy egy 30 éves kutató teljesítményének elemzése esetén a referencia-adatbázisból minden kutató 30 éves kori értékét vehessük figyelembe az adott tudományterületen belül.*

OTKA kutatók statisztikai elemzése

Első lépésként az összes 2020-as OTKA kutatóra (n=1104) megtörtént a tudománymetriai paraméterek számítása az MTMT-ből letöltött adatok alapján. Néhány kutató esetén nem lehetett kézi módszerrel sem meghatározni az MTMT azonosítót, ezeket a kutatókat kihagytuk az elemzésből.

Második lépésben a fentebb megadott három paraméter összehasonlításra került az adott tudományterületen belüli valamennyi hasonló korú kutatóval. *Tehát pld egy 34 éves fizikus idézettsége az összes 34 évnél idősebb fizikus 34-éves kori idézettségével volt összehasonlítva.* Az összehasonlítás során a kutatók egy percentilis alapú beosztást kaptak, ahol a 100-as érték azt jelenti, hogy az adott kutató a legjobb eredményt érte el (más szóval még soha senki korábban nem mutatott fel nála jobb teljesítményt), a 0-s eredmény azt jelenti, hogy a legrosszabb eredményt érte el (eddig mindenki jobb volt nála), az 50-es érték pedig egy átlagos kutatót takar. Az adott kutatók ezen értékét nevezzük a továbbiakban **„tudománymetria szerinti”** percentilisnek.

A fenti számítás után a három számolt érték átlaga, az ún. **„overall score”** is kiszámításra került a kutató által elért percentilis értékek felhasználásával. Az átlag számítása lehetővé teszi, hogy egyetlen „kompozit mutatóba” tömörítsük az egyes kutatók értékeit, és ezáltal egyetlen végső sorrendet tudjunk valamennyi kutatóra megállapítani. Az átlag számítása során az elmúlt öt évi Q1-es cikkek száma 200%-os súlyt kap, a citáció és a H-index 100%-os súlyt – ennek célja, hogy a vezető szerzős közlemények súlya nagyobb legyen.

OTKA zsűrin belüli elemzések

Összesen 27 zsűrit vettünk figyelembe. Ezek az „Állam, Jog és Politika”, az „Élettan”, a „Fizika”, a „Gazdaság”, az „Gépészmérnök”, a „Genetika, Bioinformatika”, a „Földtudományok 2”, a „Földtudományok 1”, az „Idegtudományok”, az „Immun-, Tumor- és Mikrobiológia”, az „Irodalom”, az „Informatikai”, a „Klinikai Orvostudományok”, a „Kémia 1”, és a „Kémia 2”, a

„Komplex agrártudomány”, a „Kultúra”, a „Matematika”, a „Molekuláris és Szerkezeti Biológia, Biokémia”, a „Növénytermesztés, állattenyésztés”, a „Nyelvészet”, az „Ökológia és evolúció”, a „Pszichológia”, a „Régészet”, a „Sejt- és Fejlődésbiológia”, a „Társadalom”, valamint a „Történelem”.

A „tudománymetria szerinti” percentilisek minden kutatóra való kiszámítása után az egyes zsűriken belüli kutatókat egymással is összehasonlítottuk. Ennek során már nem vettük figyelembe az MTMT alapján számolt „nyers” értékeket (H-index, idézettség, közlemények száma), csak a percentiliseket. A kutatókat a percentilis alapján sorba rendeztük, és az ebben a rangsorban elfoglalt hely alapján minden kutató kapott egy új, **„zsűrin belüli rangot”**, amely szintén egy percentilis alapú mutató.

Végezetül a zsűrin belüli rangokat az egyszerű áttekinthetőség kedvéért kvartilisekre soroltuk be, ahol a Q1 a legjobb, és a Q4 a leroszabb kvartilist takarja.

OTKA honlap scientometriai mutatók

Összefoglalva az elvégzett elemzéseket: minden zsűrin belül az összes kutatóra kiszámoltuk valamennyi vizsgált paramétert, majd a kutatókat rangsoroltuk az általuk elért overall score alapján, és ez alapján kaptak a kutatók „zsűrin belüli rang”-ot. Az OTKA honlapon az overall score értéke **„Scientometric rank in Hungary”** néven látható, a zsűrin belüli rang pedig **„Scientometric quartile in this panel”** néven.

GYAKRAN FELTETT KÉRDÉSEK

Mi történik osztott első/utolsó szerzőség esetén?

Az első/utolsó szerzős közlemények azonosításakor az MTMT adatok alapján nem lehetséges az osztott első/utolsó szerzőséget azonosítani, ezért ezek a közlemények nem szerepelnek egy-egy kutató önálló tudományos eredményességének meghatározásában.

Miért lógnak ki egyes kutatók a grafikonból?

Egyes kutatók kilóghatnak a korábbi kutatók által megadott tartományból. Ennek oka lehet:

- Az MTMT frissítés utolsó időpontja hiányzik.
- Az aktuális letöltés időpontjában az MTMT-ben „A közlemények listája nem teljes” vagy „Ez az adatlap nem nyilvános vagy szerkesztés alatt áll” volt bejelölve.
- Az eredeti adatbázisban levő születési év eltér az OTKA adatbázisban levő születési évtől.

- Az MTMT-ből az XML letöltése sikertelen volt.

Figyelembe vesszük a PhD megszerzésének évét?

Az OTKA pályázatok bírálói között végzett felmérés eredményeképpen a PhD idejére való korrekciót kivettük a rendszerben elérhető lehetőségek között, mivel a válaszadók több, mint 90%-nem kérte ezen lehetőség fenntartását.

Miért nem csak az elmúlt 10 év közleményeit vizsgáljuk?

Az összesített publikációs aktivitás további mérése során a Q1-es cikkek súlya 200%. Ezen elemzési opció hat alternatíváját is megvizsgáltuk:

1. a H-index számítása során csak a vezető szerzős közlemények kerülnek felhasználásra;
2. a H-index számítása során csak az elmúlt 10 év közleményeit számítjuk be;
3. a H-index számítása során csak az elmúlt 10 évben megjelent vezető szerzős közlemények kerülnek felhasználásra;
4. az éves idézettsége számítása során csak a vezető szerzős közlemények kerülnek felhasználásra;
5. az éves idézettsége számítása során csak az elmúlt 10 év közleményeit számítjuk be;
6. az éves idézettsége számítása során csak az elmúlt 10 évben megjelent vezető szerzős közlemények kerülnek felhasználásra.

Ezen opciók alkalmazása esetén azonban a kutatók mérőszámai olyan mértékben lecsökkentek, hogy már nem volt lehetséges a négy kvartilisbe való beosztás. A Q1-es cikkek súlyának növelését azért választottuk, mert az nem eredményezett adatvesztést.

Miért nem a D1-es cikkek számát alkalmazzuk a kiválóság mérésére?

Az elemző rendszer első verziója a Q1-es közlemények helyett a D1-es közleményeket vette csak figyelembe. Az adatbázisban levő összes kutató 88%-a nem közölt 2019-ben egyetlen darab első/utolsó szerzős D1-es rangú cikket sem. Egyes területeken (pld I, II, IX) a D1-es cikkek száma olyan alacsony volt, hogy értelmes rangsor készítésére már nem volt használható. Mivel a rendszer felépítése során cél volt, hogy egy olyan referencia-keretet adjunk, amelyben a kutatók nagy részét, valamint a tudományterületeket egymással is össze tudjuk hasonlítani, ezért végül a Q1-es rangot választottuk, mint a kiválóság mérőszáma. A D1-es rangú cikkek továbbra is befolyásolják a rendszer által számolt eredményt azáltal, hogy a D1-es cikkekre több idézet érkezik, mint a Q1-es rangú közleményekre.

Figyelembe vesszük gyermekek nevelését?

A saját gyermekek nevelése női kutatók esetén több pályázati séma esetén külön szempontként jelenik meg. Ezért megvizsgáltuk, hogy hogyan tudjuk ezt plusz változóként beépíteni a rendszerbe. Sajnos azonban nem áll rendelkezése elegendő kutató esetében a gyermekek száma ahhoz, hogy referencia-értékeket lehessen számolni.

Mi történik egyező értékek esetén?

Egyes kutatók esetén az adott mérőszámok esetén lehetséges egyező érték elérése. Mivel egy relatív értéket számolunk, ezért egyező érték esetén lehet a legjobb vagy a legrosszabb értéket is referenciaként venni. Mivel sok esetben alacsony értékek jellemzőek voltak a kutatók egy jelentős hányadára, ezért az osztályozás során mindig a legrosszabb értéket vettük. Valamint, a 0-val egyező értékek esetén kizárólag a legrosszabb kategóriába osztottuk be a kutatókat függetlenül a többi kutató relatív publikációs teljesítményétől.

Mikor lesz frissítve az adatbázis?

“ MTMT cutoff date:” az MTMT-ből teljes letöltés néhány napig tart, ezen dátum az adatletöltés utolsó napja (ez utáni frissítések már nincsenek benne a tudománymetria adatbázisában). *Tervezett következő letöltés: 2021 március eleje.*

Mit mutat a Database fül?

“Database” fül: a “Science” és többi fülön mutatott aktuális adatbázis áttekintő adatai. Ezen adok “All” beállítása esetén sem egyeznek a teljes adatbázissal, ugyanis a menürendszer szűrési alapbeállításai miatt itt nem jelennek meg:

- MTA külső tagok (n=55 kutató)
- Az MTMT-ben „A közlemények listája nem teljes” vagy „Ez az adatlap nem nyilvános vagy szerkesztés alatt áll” van bejelölve.
- MTMT utolsó frissítésre való szűrés alapján:
 - Alapbeállításként 2009-07-01 előtti frissítési dátumú személyek nincsenek benne, mivel ez volt az MTMT indulásának napja (n=46 kutató).
 - Kizártuk azokat is, akiknél nincs megadva frissítési dátum az MTMT adatlapjukon, tehát nincs is mire szűrni (n=317 kutató).

Mi történik 0-ás érték esetén?

Az OTKA pályázók közül azoknál a kutatóknál, akiknek 0 a teljesítménye bármilyen mutatónál (*_num oszlopok), ott a *_rank_cohort oszlopokban is 0 szerepel és eszerint számítnak a *_rank_cohort_quartile oszlopok is. Ennek eredményeképpen a 0-s érték *az adott mutató szempontjából* mindenképpen Q4-es besorolást kap az adott kutató.

A duplázás mindenképpen kétszeres növelést jelent?

Hogy rugalmas legyen a kód, ezért egy külön állítható paraméter teszi lehetővé a "duplázás"-t futtatáskor: ha valakinek 2,345-szöröshöz lenne kedve, akkor úgy is futtathatja, de nyitva áll csökkentés is, pl. 0,565-tel, vagy a helybenhagyás 1-gyel. Jelenleg az online adatbázis ezt a paramétert 2-essel használja. Az online elérhető kezelőfelület nem tartalmazza ezen paraméter állíthatóságát, mivel ehhez a teljes adatbázisra vonatkozó kódot is változtatni kellene.

Miért nagyobb az MTMT-ben a H-indexem?

A tudománymetria rendszer a H-index számítása során csak a független idézőket veszi figyelembe, valamint a számítások csak teljes évre vonatkoznak. Emiatt például a 2020-as idézetek csak 2021 január után kerülnek bele a H-index számításába.

Osztott első/utolsó szerzőséget is figyelembe vesz a rendszer?

Igen. (Amennyiben az MTMT-ben ez szerepel.)