



JÖVŐKÉPESSÉG OBSZERVATÓRIUM  
MOME ALAPÍTVÁNY 2024

# A mérőrendszerek fogalmi fejlődésének vizsgálata önszerveződő térképek segítségével

Műhelytanulmány

A mérőrendszerek fogalmi fejlődésének vizsgálata  
önszerveződő térképek segítségével – Műhelytanulmány

**Szerző**

Nádai László

**Kiadó**

Future Potentials Observatory  
Jövőkéesség Obszervatórium Nonprofit Korlátolt Felelősségű Társaság  
1121 Budapest, Zugligeti út 9-25  
Budapest, 2024

**ISBN** 978-000-0000-00-X

**Köszönetnyilvánítás**

A kutatást a Moholy-Nagy Művészeti Egyetem Alapítvány  
Jövőkéesség Obszervatóriuma támogatta.

# Tartalom

Vezetői összefoglaló	4
Executive Summary	5
A fogalmi keretrendszerek jellemzői	6
A fogalmi változás mérése	10
Kulcsszókeresés a számítógépes nyelvészetben	13
A Kohonen-féle önszerveződő térképek	16
Az EIS fogalmi fejlődése a kulcsszavak alapján	19
A GII fogalmi fejlődése a kulcsszavak alapján	25
A fogalmi változás mértéke: a kölcsönös információ	29
Validálás: a digitalizáció fogalmának változása	31
Esettanulmányok	36
Ajánlott irodalom	38

# Vezetői összefoglaló

**A tanulmány a mérőrendszerek fogalmi fejlődésének, a változás dinamikájának leírását lehetővé tevő, tudományosan megalapozott módszertan kidolgozására tesz kísérletet. A környezeti feltételek megváltozásához történő adaptációkat jellemeztes esetek bemutatásán keresztül tárja az olvasók elé.**

Kiindulásként ismertetjük a mérőrendszerekkel kapcsolatos diskurzusban alkalmazott különböző típusú *fogalmi kereteket*. Tárgyaljuk a típusok néhány fontos jellemzőjét, és példákon keresztül bemutatjuk, hogy a különböző keretrendszerek mindegyike releváns lehet a rangsorokról való gondolkodás szempontjából.

A fogalmi *változás mérése* többféle adatforrás felhasználásával, kvalitatív és kvantitatív módszerek kombinálásával történik. A kontextustól, a céloktól és az erőforrásoktól függően a kutatók választhatják e módszerek egyikét vagy kombinációját annak felmérésére, hogy a fogalmi megértés változott-e és hogyan.

A kulcsfogalmak manuális módszerekkel történő keresése lassú, költséges és hibákkal teli folyamat. Az utóbbi években számos automatikus, gyors kulcsszó-kinyerési technikát dolgoztak ki a kulcsszavak nagy mennyiségű adatból történő kinyerésére. Ezek a *kulcsszavak* segítenek a témák és a meghatározó szövegrészek azonosításában, valamint a jelentésbeli hálózatok feltérképezésében.

A Kohonen-féle *önszerveződő térképek* a gépi tanulásban és adatelemzésben használt mesterséges neurális hálózatok egy bizonyos típusa, amelyet a sokdimenziós adatok alacsonyabb dimenziós térben történő klaszterezésére és megjelenítésére használnak. A K-SOM eljárással elrendezzük a síkon a két nagy, nemzetközileg ismert és elismert innovációs rangsor, avagy mérőrendszer, az Európai Innovációs Eredménytábla és a Globális Innovációs Index fogalmi kereteit definiáló kulcskifejezéseket, és meghatározzuk az egymással összefüggő fogalmak csoportjait.

Miután a kulcsszavak segítségével meghatároztuk a mérőrendszerek fogalmi kereteit, és a relatív gyakoriságok időbeli változása alapján, az önszerveződő térképek felhasználásával vizualizáltuk a fogalmi rendszerek fejlődését, már csak az marad hátra, hogy javaslatot tegyünk a változás gyorsaságának mérésére. Ehhez az információelméletből ismert *kölcsönös információ* fogalmát vesszük igénybe.

A fogalmi változás sebességének jellemzésére bevezetett kölcsönös információ értékét végül egy teljesen másik területre, a *digitalizációra* kidolgozott mérőrendszerekre számoljuk ki, így ellenőrizve a módszerünk érvényességét.

# Executive Summary

**The paper attempts to develop a scientifically sound methodology to describe the conceptual development of measurement systems and the dynamics of change. The adaptations to changing environmental conditions are presented through typical case studies.**

We start by describing the different types of *conceptual frameworks* used in the discourse on measurement systems. We discuss some important features of each type and show through examples that each of the different frameworks can be relevant for thinking about rankings.

*Conceptual change* is measured using multiple data sources, combining qualitative and quantitative methods. Depending on context, objectives and resources, researchers may choose one or a combination of these methods to assess whether and how conceptual understanding has changed.

Searching for key concepts using manual methods is a slow, costly and error-prone process. In recent years, several automatic, fast keyword retrieval techniques have been developed to extract keywords from large amounts of data. These *keywords* help to identify topics and key phrases and to map meaning networks.

*Kohonen Self-Organising Maps* are a particular type of artificial neural network used in machine learning and data analysis to cluster and visualise multidimensional data in a lower dimensional space. The K-SOM process is used to arrange on the plane the key terms defining the conceptual frameworks of two major internationally known and recognised innovation ranking or measurement systems, the European Innovation Scoreboard and the Global Innovation Index, and to define groups of interrelated concepts.

Having defined the conceptual framework of the measurement systems using the keywords and visualised the evolution of the conceptual systems based on the change in relative frequencies over time using self-organising maps, it remains to propose a measure of the speed of change. To do this, we make use of the concept of *mutual information*, familiar from information theory.

Finally, the value of mutual information introduced to characterize the speed of conceptual change is calculated for measurement systems developed for a completely different domain, *digitalization*, thus validating of our method.

# A fogalmi keretrendszerek jellemzői

**A fejezet célja, hogy ismertesse a mérőrendszerekkel kapcsolatban alkalmazott különböző típusú fogalmi kereteket. Tárgyaljuk a típusok néhány fontos jellemzőjét, és példákon keresztül bemutatjuk, hogy a különböző keretrendszerek mindegyike releváns lehet a rangsorkról való gondolkodás szempontjából.**

Bármely kutatási terület áttekintéséhez, vizsgálatához és értékeléséhez explicit fogalmi keret van szükség. A fogalmi keret az ötletek rendszerezésének, a kutatás orientálásának és annak biztosításának alapvető eszköze, hogy az elméletek és a megállapítások összekapcsolódjanak és értelmet nyerjenek. Ez nyilvánvalóan ugyanúgy igaz a mérőrendszerekre, mint más diszciplínákra. A rangsorkészítést az különbözteti meg sok más szakmai területtől, hogy nem igazán történetek összehangolt, tudatos erőfeszítések a fogalmi keretek meghatározására és egyértelművé tételére.

A fogalmi keretrendszerek strukturált megközelítést biztosítanak az összetett jelenségek megértéséhez és vizsgálatához, különösen a tudományos kutatás, az elméletalkotás, és a professzionális gyakorlat területén. Segítenek tisztázni a kulcsfontosságú fogalmak és változók közötti kapcsolatokat az adott szakterületen belül. Feltérképezik, hogy a különböző tényezők hogyan kapcsolódnak egymáshoz és hogyan befolyásolják egymást, segítve a kutatókat és a nem szakavatott olvasókat a dinamika megértésében.

A kulcsfogalmak és változók egyértelmű definiálása irányítja a kutatókat és az eredményeket felhasználó döntéshozókat a megfelelő módszerek, eszközök és adatgyűjtési stratégiák kiválasztásában. Ez biztosítja, hogy a munka fókuszált és elméletileg megalapozott legyen. A fogalmi keretrendszer ugyanakkor a kutatók és a szakemberek közös nyelveként is szolgál. Olyan vizuális vagy elméleti ábrázolást biztosít, amely könnyen kommunikálható és érthető, lehetővé téve a célirányos és következetes beszélgetéseket.

Az *1. táblázat* a fogalmi keretek három típusát és jellemzőiket mutatja be.<sup>1</sup> Az első a **taxonómia**, amelyik egy olyan verbális leírás, amely a jelenségeket osztályokba sorolja. A kapcsolatok nyilvánvalóak, amennyiben az egy osztályon belüli elemek hasonlóak; de az osztályok közötti kapcsolatok vagy gyengék, vagy nem léteznek. A leírt jelenségek köre lehet szűk, de gyakran tág is. A jelenségekre vonatkozó bizonyítékok lehetnek közvetlen tapasztalatból származó, vagy logikai érvelésből, illetve empirikusan kifejlesztett bizonyítékok. Vagyis a forrás – tudományos szempontból – nélkülözi a szigorúságot. A taxonómiákra példaként említjük az Európai Innovációs Eredménytábla fogalmi keretrendszerét<sup>2</sup>, amelyet az *1. ábrán* mutatunk be.

A fogalmi keretrendszerek második típusa a **vizuális reprezentáció**, amely képet ad a jelenségekről. Az osztályok közötti kapcsolatok jelennek meg, míg a taxonómiában általában nincsenek ilyen kapcsolatok. A bemutatott jelenségek lehetnek olyan tágak, mint a teljes szakterület, vagy olyan szűkek, mint egyetlen fogalmi összefüggés. A bizonyítékoknak mindenképpen logikusnak kell lenniük, és rendelkezhetnek empirikus alátámasztással. Ez származhat szakértői véleményből vagy kutatásból. Tudományos szempontból a bizonyítékok lehetnek szigorúak, vagy kevésbé

<sup>1</sup> Emans, R. (1970). A Schema for the Classification of Conceptual Frameworks Involving Reading. *Journal of Reading Behavior*, 3(2), 15-21. <https://doi.org/10.1080/10862967009546931>

<sup>2</sup> [https://research-and-innovation.ec.europa.eu/document/download/074d5495-433a-440f-bcf9-dc620fce7af1\\_en?filename=ec\\_rtd\\_eis-2024-methodology-report.pdf](https://research-and-innovation.ec.europa.eu/document/download/074d5495-433a-440f-bcf9-dc620fce7af1_en?filename=ec_rtd_eis-2024-methodology-report.pdf)

szigorúak. A vizuális ábrázolásra jó példa a Globális Innovációs Index fogalmi keretrendszere<sup>3</sup>, amelyet a 2. ábrán láthatunk.

A fogalmi keretek harmadik típusa a **matematikai leírás**, amelyben a jelenségek valamilyen matematikai egyenletben fejezhetők ki, ami természetesen nem zárja ki a szóbeli leírás és a képi ábrázolás lehetőségét. A jelenségek közötti kapcsolatokat számszerűsítik, és az egyes jelenségeknek meghatározott súlyokat adnak; ez egyértelműen megkülönbözteti a fogalmi keretrendszer e típusát a vizuális ábrázolástól, amely csak azt mutatja, hogy létezik kapcsolat, de annak mértékét nem; és a taxonómiától, amely nem mutat semmilyen kapcsolatot a bemutatott osztályok között. Az ábrázolt jelenségek általában szűknek mondhatók a többi típushoz képest, de ez változó. A kutatásból származó empirikus bizonyítékokra szükség van; de a jelenségek logikai magyarázatát (értelmezését) nem feltétlenül kapjuk meg, mivel a matematikai keretek azt írják le, ami van, nem pedig azt, hogy miért. Az alátámasztó bizonyítékoknak természetüknél fogva szigorúnak kell lenniük. Ez magyarázza, hogy jelenleg a matematikai leírások általában szűk körűek, mivel csak empirikusan gyűjthető bizonyítékokat tartalmaznak. Ilyen típusú fogalmi keretrendszerre teszünk javaslatot ebben a tanulmányban.

Szeretnénk egyértelművé tenni, hogy a háromféle fogalmi keretrendszer között nincs merev lehatárolás. Egy matematikai leírás lehet vizuálisan ábrázolt vagy verbálisan leírt. Hasonlóképpen, egy vizuális ábrázolás leírható verbálisan; és egy taxonómia képe megrajzolható, bár a különböző osztályok közötti kapcsolat nem mindig látható egyértelműen. Egy taxonómia vagy vizuális reprezentáció végül matematikai leírássá válhat, ha a megfelelő empirikus bizonyítékokat összegyűjtik és elemzik.

1. táblázat. A fogalmi keretrendszerek típusai és jellemzői

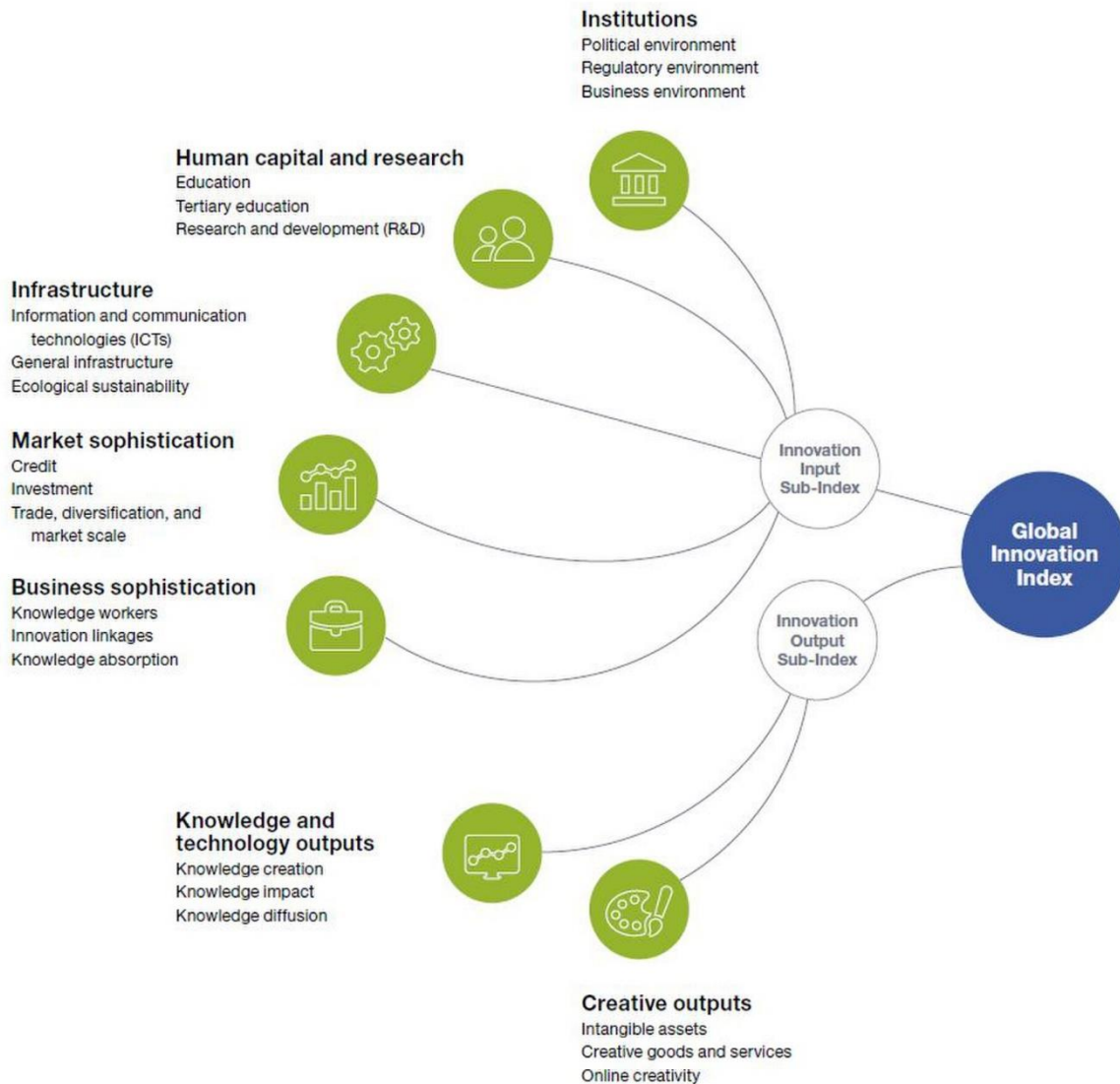
Jellemzők	Taxonómia	Vizuális ábrázolás	Matematikai leírás
<b>Előadásmód</b>	Szóbeli	Képi	Egyenletek
<b>Kapcsolatok</b>	Osztályokon belüli kapcsolatok; nem az osztályok közötti	Az osztályok közötti kapcsolatok nyilvánvalóak	A kapcsolatok számszerűsítettek
<b>Hatókör</b>	Szűk vagy tág; gyakran tág	Szűk vagy tág	Valószínűleg szűk körű
<b>Igazolás</b>	Közvetlen tapasztalat; amely lehet nem logikus vagy nem empirikus is	Logikus; lehet empirikus	Empirikus; nem feltétlenül logikus
<b>Az igazolás alapja</b>	Gyakorlat, szakértői vélemény, kutatás	Szakértői vélemény, kutatás	Kutatás
<b>Az igazolás szigorúsága</b>	Gyakran nem túl szigorú	Lehet, hogy szigorú, de lehet, hogy nem	Kötelezően szigorú

<sup>3</sup> [https://www.wipo.int/edocs/pubdocs/en/wipo\\_pub\\_gii\\_2020-appendix1.pdf](https://www.wipo.int/edocs/pubdocs/en/wipo_pub_gii_2020-appendix1.pdf)

<p><b>Keretfeltételek</b></p> <p><b>Emberi erőforrások</b></p> <p><b>1.1.1.</b> A természettudományok, technológia, mérnöki tudományok és matematika (STEM) területén szerzett új doktori fokozatok 1000 fő 25-34 éves népességre vetítve</p> <p><b>1.1.2.</b> A felsőfokú végzettséggel rendelkező 25-34 éves népesség százalékos aránya</p> <p><b>1.1.3.</b> Az egész életen át tartó tanulásban részt vevő 25-64 éves népesség százalékos aránya</p> <p><b>Vonzó kutatási rendszerek</b></p> <p><b>1.2.1.</b> Nemzetközi társszerzővel írt tudományos publikációk egymillió lakosra vetítve</p> <p><b>1.2.2.</b> A világszerte a legtöbbet idézett publikációk top-10%-ába tartozó tudományos publikációk az ország összes tudományos publikációjának százalékában kifejezve</p> <p><b>1.2.3.</b> Külföldi doktoranduszok aránya az összes doktorandusz százalékában</p> <p><b>Digitalizáció</b></p> <p><b>1.3.1.</b> Szélessávú internet elterjedtsége a vállalkozásoknál</p> <p><b>1.3.2.</b> Az alapszintűnél magasabb általános digitális készségekkel rendelkező egyének (%-ban)</p>	<p><b>Beruházások</b></p> <p><b>Pénzügyek és támogatás</b></p> <p><b>2.1.1.</b> A közszféra K+F-kiadásai (a GDP százalékában)</p> <p><b>2.1.2.</b> Kockázati tőkekiadások (a GDP százalékában)</p> <p><b>2.1.3.</b> Közvetlen kormányzati finanszírozás és kormányzati adótámogatás a vállalkozások K+F-jéhez (a GDP százalékában)</p> <p><b>Céges befektetések</b></p> <p><b>2.2.1.</b> K+F-kiadások az üzleti szektorban (a GDP százalékában)</p> <p><b>2.2.2.</b> Nem K+F innovációs kiadások (a forgalom százalékában)</p> <p><b>2.2.3.</b> Innovációs kiadások egy innovatív vállalkozásban foglalkoztatott személyre vetítve</p> <p><b>Információs technológiák használata</b></p> <p><b>2.3.1.</b> A személyzetük IKT-ismereteinek fejlesztését vagy korszerűsítését célzó képzést nyújtó vállalkozások</p> <p><b>2.3.2.</b> IKT-szakemberek (az összes foglalkoztatott százalékában)</p>
<p><b>Innovációs tevékenységek</b></p> <p><b>Innovátorok</b></p> <p><b>3.1.1.</b> Termékinnovációkat bevezető kkv-k (a kkv-k százalékában)</p> <p><b>3.1.2.</b> Üzleti folyamatinnovációkat bevezető kkv-k (a kkv-k százalékában)</p> <p><b>Kapcsolódási pontok</b></p> <p><b>3.2.1.</b> Másokkal együttműködő innovatív kkv-k (a kkv-k százalékában)</p> <p><b>3.2.2.</b> Köz- és magánszféra közös publikációi egymillió lakosra vetítve</p> <p><b>3.2.3.</b> A tudományos és technológiai humán erőforrás munkahelyek közötti mobilitása</p> <p><b>Szellemi javak</b></p> <p><b>3.3.1.</b> PCT-szabadalmi bejelentések egymilliárd GDP-re vetítve (PPS-ben)</p> <p><b>3.3.2.</b> Védjegybejelentések egymilliárd GDP-re vetítve (PPS-ben)</p> <p><b>3.3.3.</b> Formatervezési mintaoltalom bejelentések egymilliárd GDP-re vetítve (PPS-ben)</p>	<p><b>Hatások</b></p> <p><b>Foglalkoztatási hatások</b></p> <p><b>4.1.1.</b> Foglalkoztatás tudásintenzív tevékenységekben (a teljes foglalkoztatás százalékos aránya)</p> <p><b>4.1.2.</b> Foglalkoztatás az innovatív vállalkozásokban</p> <p><b>Értékesítési hatás</b></p> <p><b>4.2.1.</b> A közepes- és csúcstechnológiai termékek exportjának aránya a teljes termékkivitelben</p> <p><b>4.2.2.</b> Tudásintenzív szolgáltatások exportja az összes szolgáltatásexport százalékában</p> <p><b>4.2.3.</b> Az újonnan piacra kerülő és újonnan a vállalkozás számára kifejlesztett innovációk értékesítése a forgalom százalékában</p> <p><b>Környezeti fenntarthatóság</b></p> <p><b>4.3.1.</b> Az erőforrások termelékenysége</p> <p><b>4.3.2.</b> A PM2,5 finom részecskék levegőbe történő kibocsátása az iparban</p> <p><b>4.3.3.</b> A környezetvédelemmel kapcsolatos technológiák fejlesztése, az összes technológia százalékában</p>

1. ábra. Az Európai Innovációs Eredménytábla fogalmi keretrendszere





Source: Global Innovation Index 2021, WIPO.

## 2. ábra. A Globális Innovációs Index fogalmi keretrendszere

Hangsúlyozzuk, hogy egyik típusú fogalmi keretrendszer sem jobb eredendően, mint a többiek. A matematikai leírások számszerűsített jellemzői nagyobb pontosságot tesznek lehetővé a vizsgálatok során, ez azonban azokra a tényezőkre korlátozódik, amelyekre az adatgyűjtés eszközeit kidolgozták. Előfordulhat, hogy más, és talán fontosabb tényezőkkel kapcsolatosan nem állnak rendelkezésre adataik. A kutatás hipotézisei is nyilván arra korlátozódnak, ami matematikai összefüggésekkel kifejezhető. Ahogyan az ilyen függvényekre vonatkozó adatgyűjtés és adatfeldolgozás eszköztára fejlődik, a matematikai leírások egyre érvényesebbé válnak. Addig is a taxonómiák és a vizuális ábrázolások fontos eszközként szolgálnak a kutatással kapcsolatos gondolkodás irányításához.

## A fogalmi változás mérése

**A fogalmi változás mérése többféle adatforrás felhasználásával, kvalitatív és kvantitatív módszerek kombinálásával történik. A kontextustól, a céloktól és az erőforrásoktól függően a kutatók választhatják e módszerek egyikét vagy kombinációját annak felmérésére, hogy a fogalmi megértés változott-e és hogyan.**

A fogalmi változás az egyén – vagy egy közösség – adott fogalom megértésében vagy mentális modellezésében bekövetkező változását jelenti, amely gyakran a korábbi ismeretek átstrukturálásával, vagy új, pontosabb, esetleg átfogóbb elképzelésekkel való helyettesítésével jár. A fogalmi változás mérése bonyolult feladat, mivel magában foglalja a tanulás kognitív és affektív aspektusainak értékelését is.

A fogalmi változás azonosításának és mérésének komoly szakirodalma<sup>4</sup> van, elsősorban a pedagógia és kognitív pszichológia tudományterületéhez kapcsolódóan. A vizsgálatok leggyakrabban annak megértésére összpontosítanak, hogy a tanulók hogyan győzik le a téves elképzeléseiket, hogyan integrálják az új információkat, és hogyan alakítják át mentális modelljeiket. Az alábbiakban néhány, a fogalmi változás mérésére általánosan használt módszert és eszközt mutatunk be, külön kiemelve azokat a jó gyakorlatokat, amelyek a mérőrendszerek fogalmi változásának elemzésébe átemelhetők.

**Előtte-utána típusú tesztek.** Ezek olyan strukturált tesztek vagy kérdőívek, amelyek célja, hogy felmérjék a tanulók bizonyos fogalmakra vonatkozó megértését egy tanulási tapasztalat előtt és után. Ha arra kérjük a tanulókat, hogy magyarázzák meg érvelésüket, vagy írják le a fogalmakat a saját szavaikkal, akkor mélyebb betekintést nyerhetünk a fogalmi változásaikról. Ennek speciális eseteként *konceptiótérképeket (concept map<sup>5</sup>)* rajzolhatunk, amelyek azt ábrázolják, hogyan szerveződnek és kapcsolódnak a fogalmak a tanuló fejében a tanulás előtt és után. A fogalmi térkép szerkezetében vagy tartalmában bekövetkezett változás fogalmi változást jelezhet.

Jelen tanulmányban pontosan erre teszünk kísérletet: mesterséges neurális hálózatok alkalmazásával elkészítjük az egyes mérőrendszerek fogalmi térképét, majd megfigyeljük, hogy hogyan változott / fejlődött mindez egy hosszabb időszakon keresztül.

**Félig strukturált interjúk.** A tanulókkal vagy résztvevőkkel folytatott interjúk készítése, különösen az oktatás előtt és után, lehetővé teszi a kutatók számára, hogy feltárják a fogalmi változás mélységét és jellegét. A mélyreható beszélgetésekből kiderülhet, hogyan változott a tanulók megértése, és milyen tényezők járultak hozzá a változáshoz.

**Viselkedés megfigyelése.** A tanárok vagy kutatók megfigyelhetik a tanulókat, amint problémamegoldó feladatokban vagy megbeszélésekben vesznek részt. A tanulók feladatmegközelítési vagy gondolatmegfogalmazási módjában bekövetkező változások a fogalmi változást tükrözhetik. A megbeszélések vagy csoportmunkák elemzése megmutathatja, hogy a tanulók új terminológiát használnak-e, új ötleteket alkalmaznak-e, vagy kifinomultabb érvelésben vesznek-e részt.

---

<sup>4</sup> Posner, George J., et al. „Accommodation of a scientific conception: Toward a theory of conceptual change.” *Science education*, 66.2 (1982): 211-227.

<sup>5</sup> William M.K. Trochim, „An introduction to concept mapping for planning and evaluation,” *Evaluation and Program Planning*, 12.1 (1989): 1-16, ISSN 0149-7189, [https://doi.org/10.1016/0149-7189\(89\)90016-5](https://doi.org/10.1016/0149-7189(89)90016-5).

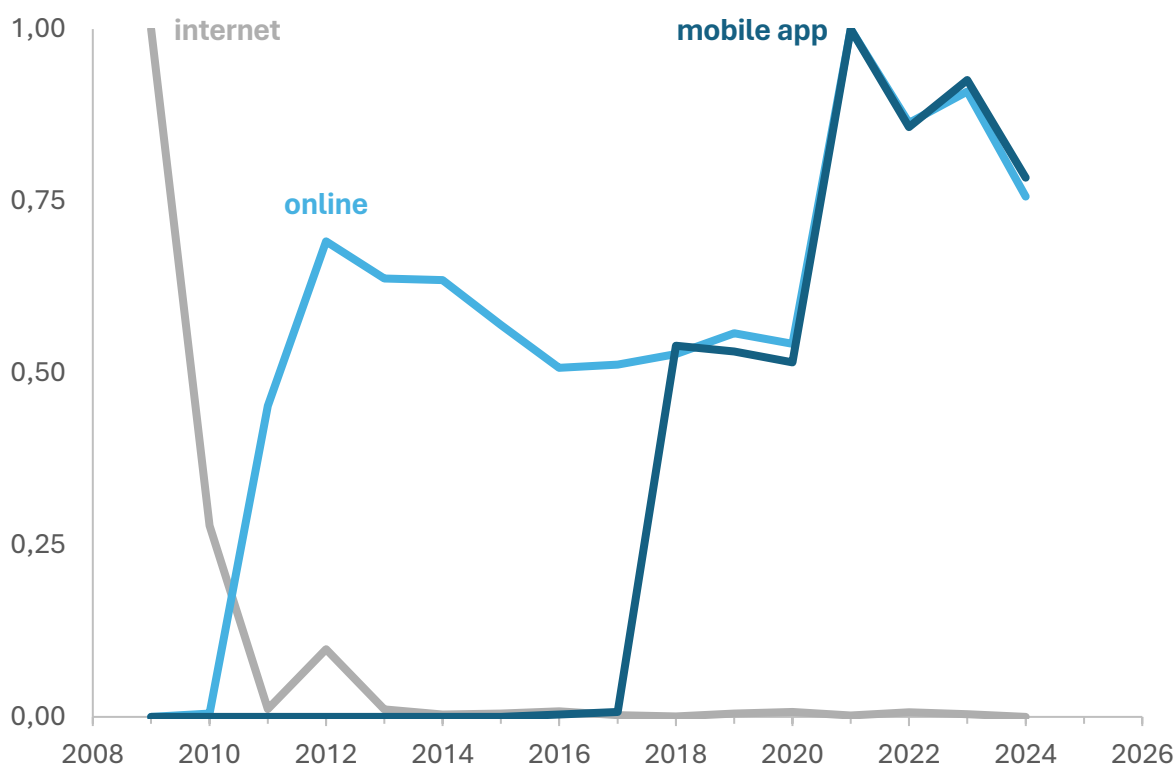
**Félreértésdiagnosztika.** A fogalmi változás mérésének hatékony módja az egy adott területen belüli gyakori téves elképzelések azonosítása és diagnosztikai eszközök használata annak nyomon követésére, hogy e téves elképzelések helyébe pontosabb fogalmak lépnek-e.

**Problémamegoldás írásban.** Az írásbeli magyarázatok tartalmának és szerkezetének elemzése a tanulási esemény előtt és után feltárhatja, hogyan változtak a tanulók elképzelései és megértései. Ha nyomon követjük, hogy a tanulók hogyan oldanak meg bizonyos típusú problémákat, az jelezheti, hogy változott-e a mögöttes fogalmi megértésük, különösen, ha az oktatás után kifinomultabb stratégiákat használnak.

**Longitudinális vizsgálatok.** A koncepcionális változás gyakran lassú, fokozatos folyamat, ezért a hosszabb időszakon átívelő változások mérése teljesebb képet adhat. A longitudinális vizsgálatok során több időpontban gyűjtenek adatokat, hogy megnézzék, hogyan fejlődik a megértés.

**Tudásstruktúrák és kognitív modellek.** Az olyan fejlett technikák, mint a kognitív modellezés (pl. számítógépes szimulációk vagy szakértői rendszerek segítségével) nyomon követhetik a mentális modellek időbeli változásait. Ezek a modellek azt próbálják megragadni, hogy az egyén hogyan szerzi meg és szervezi a tudást.

**Digitális eszközök és tanulási analitika.** A digitális tanulási eszközökkel az online platformokról származó adatok felhasználhatók a részvétel és a teljesítmény mintáinak nyomon követésére, lehetővé téve a fogalmi megértésben bekövetkező változások mérését. Ez magában foglalhatja a válaszadási minták, az egyes fogalmakra fordított idő, vagy a tanulók által elkövetett hibák típusainak elemzését.



3. ábra. Három, a digitalizációhoz kötődő fogalom fontosságának időbeli változása a Globális Innovációs Index 2009-2024 között megjelent köteteiben

Lássuk most egy konkrét példán keresztül, hogyan érvényesülnek ezek az elvek a gyakorlatban! A 3. ábrán három, a digitalizációhoz kötődő fogalom fontosságának időbeli változását látjuk a *Globális Innovációs Index* 2009 és 2024 között eltelt bő másfél évtizedben megjelent kiadásai-ban. (Azt, hogy pontosan mit is jelent ez a „fontosság”, és hogyan lehet kiszámítani, a következő fejezetben tárgyaljuk részletesen.)

Megfigyelhető, hogy a 2010-es évek elején az „internet” fogalmát felváltotta az „online” jelenlét fogalma: a GII kontextusában ez arra utal, hogy már nem az számított az innovativitás mérőfoká-nak, hogy valaki (fizikailag) hozzáfér-e a világhálóhoz, hanem hogy mennyire aktívan van jelen a virtuális térben. 2017-től újabb változásnak lehettünk tanúi: mivel a tartalomfogyasztás elsődle-ges platformjaivá ekkora a mobiltelefonok váltak, ezért az „online” jelenlét fogalmilag azonosult a „mobile applikációval”, és innentől a két idősor lefutása lényegében azonos. Habár ez már nem fogalmi változás, azért mindenképpen érdekes megfigyelni, hogy a COVID világjárvány kö-vetkezményeként kialakult társadalmi-gazdasági változások hatása is világosan kirajzolódik az ábrán: 2020-ban csaknem a duplájára növekedett a digitalizációs megoldások fontossága.

\* \* \*

Innentől kezdve már „csak” az a kérdés, hogy hogyan tudjuk kiválasztani azokat a kulcsfontos-ságú fogalmakat, amelyek időbeli hangsúlyeltolódása jó eséllyel kirajzolja a mérőrendszer fo-galmi kereteinek fejlődését. A terület szakértőinek megkérdezése (a szakirodalom áttekintése) elsőre jó ötletnek tűnik, hiszen ők pontosan ismerik a tudományterületen lezajlott koncepcioná-lis változásokat. Egy gyors hipotézisvizsgálat azonban azt mutatja, hogy a szakmai-tudományos megértés fejlődése egyáltalán nem mindig, és nem kauzálisan érhető tetten magukban a mérő-rendszerekben. Ennek az a legfőbb oka, hogy a mérőrendszerek elméleti háttérét ugyan a szak-mai közösség dolgozza ki, de az eredményeket elsősorban nem szakmai célközönség, hanem szövetségi intézmények, kormányzati szereplők, vállalatvezetők, regionális, országos és nem-zetközi szervezetek, befektetők számára prezentálják, akik összehasonlítás és döntéselőkészí-tés céljából használják azokat.

Ezért magukat a mérőrendszereket, egészen pontosan az évről-évre kiadott *jelentések szövegét* fogjuk alapul venni, és azokból kiindulva határozzuk meg a kulcsfontosságú fogalmakat, ahogy azt a következő fejezetben részletesen bemutatjuk.

# Kulcsszókeresés a számítógépes nyelvészetben

**A kulcsszavak manuális módszerekkel történő keresése lassú, költséges és hibákkal teli folyamat. Az utóbbi években számos automatikus, gyors kulcsszó-kinyerési technikát dolgoztak ki a kulcsszavak nagy mennyiségű adatból történő kinyerésére. Ezek a kulcsszavak segítenek a témák és a meghatározó szövegrészek azonosításában, valamint a jelentésbeli hálózatok feltérképezésében.**

A korpusznyelvészetben a *kulcsszavak* fogalmát olyan szavakra (illetve néha többszavas egységekre, szemantikai kategóriákra, vagy akár lexiko-grammatikai szerkezetekre) használják, amelyek szokatlan, általában nagyobb gyakorisággal fordulnak elő egy adott szövegben vagy szöveggyűjteményben, azaz korpuszban. A kulcsszavak a célszöveg vagy korpusz jellegzetes szókincsét képviselik, és így számos alkalmazásuk lehetséges a korpusznyelvészetben, a digitális bölcsészettudományokban, és a számítógépes társadalomtudományokban. Megragadhatják a szöveg lényegi mondanivalóját, a műfaj vagy szakterület terminológiáját, az irodalmi stílus fontos aspektusait, a nyelvi és kulturális különbségeket, betekintést nyújtanak a történelmi perspektívákba, és alapot nyújtanak a szöveggyűjtemények hasonlóságának méréséhez. A kulcsszavak a korpuszalapú diskurzuselemzés fontos kiindulópontját is jelentik, ahol a kulcsszavak manuálisan, az emberi megértés irányításával alkotott klaszterei központi témákat, szereplőket, metaforákat és kereteket képviselnek.

A kulcsszókeresés a gyakorlatban általában a *célkorpusz* és egy *referencia-korpusz* közötti statisztikai eltérések összehasonlításával zajlik. Különböző kutatási kérdésekre válaszolhatunk a *T* cél- (target) és az *R* referenciakorpusz alkalmas megválasztásával:

- Egyetlen *T* szöveget hasonlítunk össze az *R* szöveggyűjteménnyel → lényegkiemelés;
- *T* és *R* különböző szakmai vagy tudományos álláspontot képviselő, azonos témájú folyóiratcikkek gyűjteményei → kontrasztív keretezések; vagy
- egy adott szakterület *T* szövegeit vetjük össze egy nagy, általános nyelvű *R* referenciakorpuszsal → terminológia.

A kulcsszóelemzés a relatív gyakoriságok összehasonlításán alapul: minden egyes kulcsszójelölt szó  $f_1$  gyakoriságát az  $n_1$  darab tokenből álló *T* célkorpuszban az  $n_2$  darab tokenből álló *R* referencia-korpusz  $f_2$  gyakoriságával hasonlítjuk össze. Egy jelölt akkor tekinthető (pozitív) kulcsszónak, ha a *T*-beli relatív gyakorisága  $p_1 = f_1 / n_1$  lényegesen nagyobb, mint az *R*-beli relatív gyakorisága  $p_2 = f_2 / n_2$ .

Az összehasonlítás számszerűsítésére számos kulcsszó-mértéket javasoltak, amelyek alapul szolgálnak a jelöltek rangsorolásához és/vagy a határértékek meghatározásához. Ezek három fő csoportját lehet megkülönböztetni:

- A **hipotézisvizsgálatokon** alapuló mértékek a  $p_1$  és  $p_2$  közötti statisztikailag szignifikáns különbség megállapítására helyezik a hangsúlyt. A legszélesebb körben használt mérőszámok a  $\chi^2$  és a *log-likelihood* függvény<sup>6</sup>. Ezek a mértékek a nagy gyakoriságú kulcsszavak irányába torzítanak, és gyakran a funkciószavakat és más, nem specifikus szavakat is kulcsszóként azonosítanak, ezért alkalmazásuk során szakértői felülvizsgálatot igényelnek.

---

<sup>6</sup> Dunning, Ted. „Accurate methods for the statistics of surprise and coincidence.” *Computational linguistics*, 19.1 (1994): 61-74.

- A **hatás-méret** (effect size) ehelyett arra összpontosít, hogy egy kulcsszó-jelölt felbukása hányszor gyakoribb a *T*-ben, mint az *R*-ben. A legintuitívabb ilyen mérték a *relatív kockázat*  $r = p_1 / p_2$ , más néven  $\text{LogRatio} = \log_2 r$ . Ezek a mértékek az alacsony gyakoriságú kulcsszavak felé torzítanak, ezért gyakran kombinálják őket egy további szignifikancia-szűrővel (jellemzően a log-likelihood alapján).
- Számos publikáció vagy szoftver kínál különböző **heurisztikus** mértékeket, melyek azonban általában nélkülözik a szigorú statisztikai megalapozást. Ezek vonzereje általában az, hogy könnyen kiszámíthatók, és intuitíven megragadhatók.

A kulcsszóelemzés régóta alkalmazott, jól bevált eljárás, és számos szabványos korpusznyelvézeti szoftvercsomag támogatja az alkalmazását, amelyek közül a tanulmány elkészítése során mi az AntConc<sup>7</sup> programot használtuk. A kulcsszavak keresése során a következő lépéseket hajtottuk végre:

2. táblázat. Az Európai Innovációs Eredménytábla 2010-2024 között megjelent kiadásainak szövege alapján meghatározott kulcsszavak relatív gyakorisága

	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024
<b>Tokens</b>	<b>16607</b>	<b>27144</b>	<b>17839</b>	<b>15882</b>	<b>21476</b>	<b>23788</b>	<b>27958</b>	<b>40065</b>	<b>42884</b>	<b>41251</b>	<b>41963</b>	<b>42877</b>	<b>43349</b>	<b>45855</b>	<b>50682</b>
broadband	0,069	0,000	0,000	0,000	0,000	0,000	0,000	0,687	0,655	0,736	0,779	0,762	0,859	1,000	0,939
business	0,086	0,375	0,208	0,110	0,105	0,000	0,127	0,582	0,418	0,410	0,491	0,931	1,000	0,951	0,694
co	0,483	1,000	0,000	0,372	0,508	0,456	0,476	0,487	0,471	0,520	0,559	0,316	0,712	0,599	0,455
collaborat*	0,309	1,000	0,000	0,236	0,321	0,128	0,037	0,336	0,369	0,327	0,450	0,285	0,619	0,560	0,684
community	0,927	1,000	0,047	0,346	0,567	0,391	0,387	0,030	0,016	0,005	0,015	0,017	0,016	0,001	0,000
count*	1,000	0,919	0,000	0,931	0,844	0,816	0,576	0,181	0,757	0,612	0,747	0,084	0,266	0,378	0,423
design[,s]	0,994	1,000	0,000	0,254	0,569	0,424	0,451	0,432	0,405	0,456	0,515	0,457	0,711	0,759	0,671
digital*	0,000	0,000	0,136	0,000	0,000	0,000	0,000	0,035	0,057	0,025	0,041	0,948	0,898	1,000	0,960
doctor*	0,606	1,000	0,000	0,319	0,670	0,428	0,463	0,391	0,343	0,403	0,420	0,374	0,587	0,598	0,731
employ*	0,214	0,252	0,088	0,000	0,084	0,048	0,049	0,988	0,773	0,763	0,868	1,000	0,854	0,839	0,784
enterprise*	0,026	0,000	0,027	0,036	0,023	0,020	0,009	1,000	0,676	0,838	0,861	0,890	0,518	0,480	0,471
entrepren*	0,601	0,646	0,000	0,284	0,084	0,214	0,125	0,270	0,986	0,936	1,000	0,676	0,218	0,192	0,181
environment*	0,019	0,000	0,183	0,164	0,076	0,059	0,106	0,407	0,395	0,257	0,426	0,996	0,995	1,000	0,931
[eu,europ*]	0,568	0,265	0,000	0,364	0,857	0,760	0,703	0,730	1,000	0,590	0,846	0,394	0,613	0,807	0,560
expenditure*	0,559	0,965	0,000	0,240	0,408	0,351	0,299	0,469	0,546	0,597	0,703	0,513	1,000	0,943	0,685
export*	0,888	1,000	0,000	0,385	0,390	0,408	0,435	0,391	0,354	0,388	0,428	0,366	0,616	0,623	0,453
financ*	0,845	1,000	0,000	0,431	0,130	0,248	0,229	0,460	0,392	0,391	0,384	0,264	0,283	0,297	0,298
firm*	0,773	1,000	0,108	0,545	0,228	0,250	0,184	0,298	0,296	0,311	0,319	0,000	0,049	0,069	0,201
gdp	0,217	0,078	0,112	0,059	0,000	0,124	0,139	0,763	0,902	0,974	1,000	0,528	0,190	0,147	0,097
government*	0,054	0,000	0,012	0,057	0,021	0,017	0,034	0,024	0,051	0,054	0,168	0,651	0,985	1,000	0,858
graduates	0,543	1,000	0,000	0,590	0,850	0,556	0,649	0,538	0,479	0,539	0,521	0,510	0,734	0,737	0,625
growth	1,000	0,726	0,008	0,688	0,611	0,360	0,378	0,122	0,160	0,207	0,135	0,126	0,000	0,018	0,094
ict	0,000	0,018	0,125	0,016	0,012	0,010	0,009	0,328	0,289	0,324	0,331	0,745	1,000	0,918	0,802
impact*	0,043	0,002	0,023	0,047	0,000	0,368	0,164	1,000	0,863	0,626	0,972	0,788	0,897	0,887	0,851
indicator*	1,000	0,421	0,188	0,484	0,718	0,832	0,795	0,471	0,318	0,175	0,224	0,382	0,124	0,056	0,000
innovation*	0,754	0,380	0,387	0,866	1,000	0,905	0,471	0,278	0,409	0,333	0,373	0,288	0,000	0,119	0,625
innovator*	0,262	0,120	0,043	0,270	0,110	0,000	0,063	0,085	0,026	0,067	0,073	1,000	0,610	0,660	0,584
intellectual	1,000	0,865	0,000	0,417	0,112	0,283	0,205	0,380	0,332	0,346	0,340	0,266	0,224	0,240	0,250
knowledge	0,715	1,000	0,000	0,201	0,254	0,133	0,216	0,418	0,418	0,419	0,448	0,388	0,393	0,395	0,367
lifelong	0,000	0,000	0,062	0,000	0,000	0,000	0,000	0,703	0,580	0,696	0,737	0,618	0,854	1,000	0,828
mobility	0,000	0,032	0,000	0,027	0,020	0,018	0,015	0,021	0,010	0,010	0,020	0,610	0,979	1,000	0,778
patent	0,667	1,000	0,000	0,292	0,452	0,538	0,484	0,128	0,093	0,106	0,099	0,077	0,134	0,132	0,110
performance*	0,295	0,136	0,000	0,499	0,813	1,000	0,742	0,471	0,536	0,179	0,411	0,300	0,813	0,778	0,408
population	0,269	0,353	0,000	0,072	0,155	0,112	0,121	1,000	0,758	0,897	0,843	0,693	0,372	0,366	0,489
private	0,441	1,000	0,004	0,177	0,113	0,185	0,319	0,558	0,497	0,655	0,751	0,000	0,768	0,553	0,391
product[,s]	0,537	1,000	0,000	0,348	0,256	0,394	0,381	0,531	0,716	0,685	0,827	0,966	0,851	0,869	0,670
public	0,706	1,000	0,000	0,354	0,266	0,262	0,332	0,639	0,444	0,510	0,564	0,238	0,609	0,553	0,442
publication*	0,557	1,000	0,000	0,433	0,506	0,436	0,386	0,333	0,309	0,360	0,425	0,307	0,638	0,553	0,453
R D	0,517	1,000	0,000	0,250	0,391	0,309	0,436	0,872	0,678	0,775	0,830	0,618	0,842	0,901	0,689
sales	1,000	0,995	0,000	0,089	0,375	0,246	0,241	0,944	0,916	0,695	0,992	0,776	0,982	0,993	0,855
scientific	0,485	1,000	0,000	0,224	0,715	0,590	0,447	0,265	0,229	0,252	0,305	0,225	0,459	0,386	0,586
service*	1,000	0,878	0,000	0,151	0,207	0,019	0,113	0,891	0,817	0,829	0,874	0,676	0,399	0,404	0,304
sme*	0,659	1,000	0,018	0,065	0,014	0,032	0,000	0,309	0,328	0,310	0,415	0,144	0,210	0,186	0,322
state*	0,225	0,026	0,000	0,430	0,639	0,768	0,864	0,295	0,601	0,341	0,501	0,445	1,000	0,652	0,844
sustainab*	0,000	0,049	0,150	0,084	0,062	0,028	0,048	0,000	0,000	0,032	0,016	1,000	0,865	0,964	0,899
tech	0,747	1,000	0,015	0,203	0,129	0,279	0,292	0,671	0,960	0,989	0,727	0,849	0,575	0,621	0,000
technolog*	0,000	0,174	0,313	0,252	0,166	0,169	0,251	0,038	0,060	0,069	0,251	0,630	0,768	0,804	1,000
tertiary	0,670	1,000	0,000	0,317	0,367	0,302	0,464	0,330	0,458	0,504	0,500	0,285	0,434	0,456	0,349
trademark*	0,000	1,000	0,014	0,380	0,632	0,412	0,360	0,508	0,416	0,475	0,485	0,416	0,562	0,526	0,481
venture	0,648	1,000	0,000	0,427	0,597	0,692	0,652	0,381	0,383	0,459	0,451	0,441	0,698	0,699	0,479

<sup>7</sup> <https://www.laurenceanthony.net/software/antconcc/>

- (1) A célszövegek meghatározása, a komplex fájlok, jellemzően PDF dokumentumok egyszerű szöveggé (.txt) konvertálása;
- (2) Referenciaszöveg meghatározása (BAWE, British Academic Written English<sup>8</sup>);
- (3) A kulcsszójelöltek megkeresése statisztikai alapon (log-likelihood);
- (4) A kapott jelöltlista manuális (szakértői) ellenőrzése, a funkciósavak és a tartalmas szavak elkülönítése érdekében;
- (5) Az összetartozó fogalmak csoportosítása (pl. EU, Europe, Europa, european), reguláris keresőkifejezések alkotása;
- (6) Az előfordulások megszámlálása (raw count);
- (7) A relatív gyakoriságok kiszámítása (raw count / total count);
- (8) Skálázás kulcsszavanként (soronként) a [0;1] tartományba, tehát a kulcsszavakat a vizsgálat szempontjából egyenlő fontosságúnak tekintjük.

A kulcsszókeresés eredménye 2. táblázathoz hasonló, amelynek sorai a kulcsszavaknak, oszlopai a korpuszban lévő dokumentumoknak felelnek meg, és a cellák a megfelelő kulcsszó adott dokumentumban való előfordulásának [0;1] közé skálázott valószínűségét adják meg.

A kulcsszavak azonosítása, és a relatív gyakoriságukkal számszerűsített „fontosság” megállapítása után egy vizuális megközelítést mutatunk be a különböző korpuszok empirikus tulajdonságainak feltárására, mely az ún. Kohonen-féle önszerveződő térképek felhasználásán alapul.

---

<sup>8</sup> <https://llds.ling-phil.ox.ac.uk/llds/xmlui/handle/20.500.14106/2539>



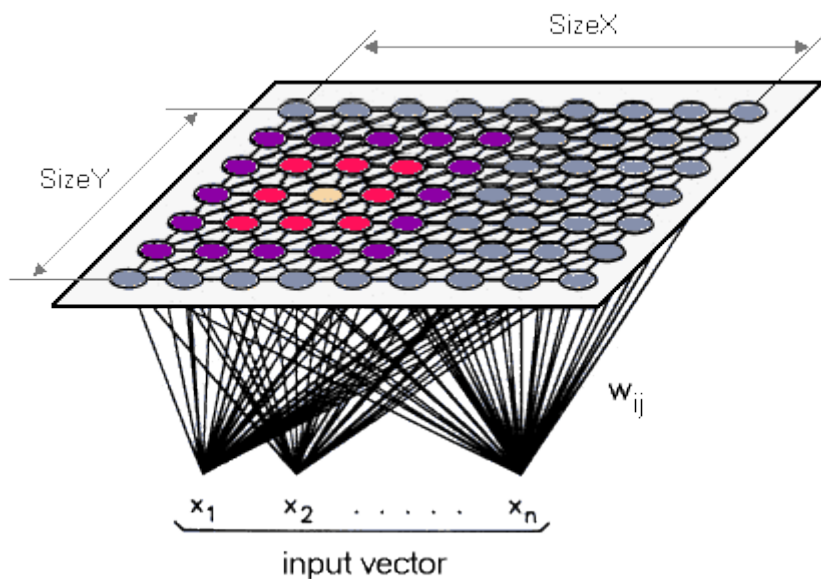
## A Kohonen-féle önszerveződő térképek

A Kohonen-féle önszerveződő térképek<sup>9</sup> (Self-Organizing Maps, SOM) a gépi tanulásban és adatelemzésben használt mesterséges neurális hálózatok egy bizonyos típusa. A K-SOM algoritmus olyan felügyelet nélküli tanulási technika, amelyet a sokdimenziós adatok alacsonyabb dimenziós térben történő klaszterezésére és megjelenítésére használnak.

A K-SOM a neurális hálózatok azon csoportjába tartozik, amelyet „versengés” felhasználásával tanítanak. A kompetitív tanulás a hálózat neuronjai közötti versenyen alapul, ahol minden egyes neuron megpróbál a legaktívabb, azaz „győztes” neuronná válni egy meghatározott bemenetre adott válaszként. A versengő tanulás számos feladatra jól használható, például mintázat-felismerésre, osztályozásra, és jellemző-kinyerésre.

### A K-SOM architektúrája

A Kohonen-féle önszerveződő térképek architektúrája kétdimenziós rácsba rendezett neuronok hálózatából áll. A rács minden egyes neuronja közvetlenül kapcsolódik a bemeneti réteghez, és megkapja a bemenő jeleket. A rácsban lévő neuronok úgy vannak elrendezve, hogy megőrzik a bemeneti tér topológiáját, ami azt jelenti, hogy a rácsban lévő szomszédos neuronok nagyobb valószínűséggel reagálnak hasonló bemeneti adatokra. A kapcsolatok súlyai az adott bemeneti adatok függvényében frissülnek.



4. ábra. A Kohonen-féle önszerveződő térképek architektúrája

<sup>9</sup> T. Kohonen, "The self-organizing map," in Proceedings of the IEEE, vol. 78, no. 9, pp. 1464-1480, Sept. 1990, doi: 10.1109/5.58325.



## A K-SOM algoritmus

A Kohonen SOM algoritmus a következő lépésekben foglalható össze:

1. A rácsban lévő neuronok súlyainak inicializálása véletlenszerű értékekkel.
2. Megmutatunk egy bemeneti vektort a hálózatnak.
3. Kiszámítjuk a rács minden egyes neuronjának aktiválási szintjét, a bemeneti adatra adott válaszként.
4. Kiválasztjuk a legmagasabb aktivációs szinttel rendelkező neuront „nyertes” neuronnak.
5. Frissítjük a nyertes neuron és a szomszédjai súlyait egy tanulási ráta és egy szomszédsági függvény (ún. „mexikói kalap”) segítségével, amely a nyertes neurontól való távolsággal csökken.
6. A 2-5. lépést előre meghatározott számú alkalommal, vagy a konvergencia eléréséig ismételjük.

A rácsban lévő neuronok eredő súlyvektorai a sokdimenziós bemeneti adatok alacsonyabb dimenziós reprezentációjaként ábrázolhatók.

## A K-SOM előnyei

A Kohonen-féle önszerveződő térképek számos olyan előnnyel rendelkeznek, amelyek hasznossá teszik őket a különféle alkalmazásokban, többek között:

**Nemlineáris dimenziócsökkentés:** A K-SOM-ok felhasználhatók a sokdimenziós adatok alacsony dimenziós térben történő ábrázolására, miközben megőrzik az adatpontok közötti topológiai kapcsolatokat. Ez segíthet feltárni az adatok mögött rejlő mintákat és struktúrákat, amelyek a sokdimenziós térben nem feltétlenül látszanak.

**Felügyelet nélküli tanulás:** A K-SOM egyfajta felügyelet nélküli tanulási technika, ami azt jelenti, hogy a tanításhoz nincs szükség előre felcímkézett adatokra. Ez olyan feladatoknál teszi őket hasznossá, ahol címkézett adatok nem állnak rendelkezésre, vagy túl drága a beszerzésük.

**Osztályozás és vizualizáció:** A K-SOM-ok remekül használhatók komplex adatok osztályozására és vizualizálására. Az eredményül kapott alacsony dimenziós reprezentáció megkönnyíti az eredeti adatokban található klaszterek és minták azonosítását, ami hasznos segítség lehet a feltáró adatelemzéshez és az adatbányászathoz.

**Robusztusság a zajjal szemben:** A K-SOM-ok viszonylag robusztusok a zajjal szemben, és akkor is jól működnek, ha a bemeneti adatok bizonyos szintű zajt vagy hibát tartalmaznak.

**Könnyű értelmezés:** A K-SOM kimenete könnyen vizualizálható és értelmezhető, ami hasznos lehet az adatokban lévő tendenciák és minták azonosításához, valamint az eredmények mások felé történő kommunikálása során.

**Rugalmasság:** A K-SOM-ok az adattípusok széles köréhez igazíthatók, beleértve a folytonos, diszkrét és kategorikus adatokat.

## A K-SOM hátrányai

Bár a Kohonen-féle önszerveződő térképek alkalmazása számos előnnyel bír, a módszer használatának vannak bizonyos korlátai, sőt hátrányai is, úgy mint:

**Érzékenység a kezdeti feltételekre:** A K-SOM teljesítménye érzékeny lehet a hálózat kezdeti feltételeire, például a hálózat neuronjainak kezdeti súlyaira. Ez azt jelenti, hogy a különböző inicializálások eltérő végső megoldásokat eredményezhetnek, és szükség lehet az algoritmus többszöri futtatására a stabil megoldás eléréséhez.

**Számítási komplexitás:** A K-SOM-ok számítási komplexitása magas lehet, különösen nagy adathalmazok és összetett hálózati architektúrák esetén. Ez idő- és számításigényessé teheti a hálózat betanítását és tesztelését.

**Az optimális hálózatméret meghatározásának nehézségei:** Az optimális hálózatméret, vagyis a hálózatban lévő neuronok számának meghatározása nehéz lehet, és gyakran csak számos próbálkozás után sikerül. A túl kevés neuron használata a bemeneti adatok rossz reprezentációját eredményezheti, míg a túl sok neuron használata túlillesztéshez vezethet.

**Használatuk alacsony dimenziós adatokra korlátozódik:** A K-SOM-okat jellemzően sokdimenziós adatok dimenziócsökkentésére használják. A K-SOM-ok teljesítménye azonban romolhat a bemeneti adatok dimenziójának növekedésével, így kevésbé hatékonyak a nagyon nagy dimenziós adathalmazok esetében.

**Korlátozott értelmezhetőség:** Habár a K-SOM kimenete könnyen vizualizálható, a kapott osztályok vagy minták értelmezése, az adatokra vonatkozó „jelentésük” meghatározása nehézségekbe ütközhet. A klaszterek vagy minták jelentése nem mindig egyértelmű, és szükség lehet a K-SOM-ok más technikákkal való kombinálására az adatok mélyebb megértéséhez.

A Kohonen-féle önszerveződő térképeket sikerrel alkalmazták kép- és jelfeldolgozásra, szöveg- és adatbányászat során, valamint a bioinformatika területén. Segítségükkel sikerrel analizáltak génkifejeződési adatokat, azonosítottak szociális hálózatokban rejlő mintázatokat, és osztályoztak vizuális jellemzők alapján fényképeket.<sup>10</sup>

---

<sup>10</sup> Teuvo Kohonen, Essentials of the self-organizing map, *Neural Networks*, Volume 37, 2013, Pages 52-65, ISSN 0893-6080, <https://doi.org/10.1016/j.neunet.2012.09.018>.

## Az EIS fogalmi fejlődése a kulcsszavak alapján

Az előző fejezetben megismert önszerveződő térképek módszerével elrendezzük a síkon az EIS 50 fontos kulcskifejezését. Az algoritmus által megrajzolt „domborzati térképen” grafikai transzformációkat (küszöbérték<sup>11</sup> meghatározása, színezése hasonlóság alapján) hajtottunk végre, hogy kiemeljük az egymással összefüggő indikátorcsoportok mintázatát.

A korábbiakban már többször használtuk mondanivalónk illusztrálására az Európai Innovációs Eredménytáblát, mint a nemzetközi mérőrendszerek egyik eminens példáját. Mostani elemzésünk kiindulópontja az EIS honlapján<sup>12</sup> elérhető 2010 és 2024 közötti jelentések PDF formátumú szövege. A megosztásra és nyomtatásra optimalizált, gazdagon illusztrált fájlokat egyszerű szövegfájlá (.txt) konvertáltuk, és lefuttattuk az automatikus kulcsszókinyerési algoritmust.

Referenciaszöveggként a már említett BAWE (British Academic Written English) korpuszt választottuk, és arra kerestük a választ, hogy mennyiben különbözik az *EIS specifikus terminológiája* az általános, egyetemi műveltséggel rendelkező angol anyanyelvűek által megírt szövegektől. Egy további ellenőrzést is végeztünk: referenciaszöveggként ezúttal a teljes 2010-24 közötti EIS korpuszt választottuk, célszöveggként pedig az egyes, 2010, 2011, ..., 2024-es kiadások szövegét; ezzel arra kérdeztünk rá, hogy vannak-e specifikus, csak az egyes évfolyamokra jellemző fogalmak? Mindennek eredménye a 2. táblázatban korábban már látott kulcsszó- és gyakoriság-lista.



5. ábra. Az EIS 2010-2024 kötetekre jellemző kulcsszavak alapján megrajzolt Kohonen-térkép

<sup>11</sup> A „küszöb” a grafikai programok egyik alapvető funkciója, amely lehetővé teszi a felhasználók számára, hogy a képeket nagy kontrasztú, fekete-fehér változatokká alakítsák át. Egy adott szintet küszöbértékként kiválasztva a program a küszöbértéknél világosabb pixeleket fehérré, a sötétebbeket feketévé alakítja át.

<sup>12</sup> [https://research-and-innovation.ec.europa.eu/statistics/performance-indicators/european-innovation-scoreboard\\_en](https://research-and-innovation.ec.europa.eu/statistics/performance-indicators/european-innovation-scoreboard_en)

Lefuttatva a K-SOM algoritmust a 5. ábrán látható „térkép” rajzolódik ki. Az olvashatóság érdekében a kulcsszavak reguláris keresőkifejezését az egyik – általában a leggyakoribb – változatukkal helyettesítettük. Ahogy azt már korábban elmondtuk, a kialakuló kép emlékeztet a topográfiai térképekről ismert ábrázoláshoz, ahol az egymáshoz földrajzi közelségben lévő helyek (a zölddel színezett „síkságok”) geográfiai jellemzői általában hasonlóak, és a kvázi-homogén területek határát az ún. vízválasztó, vagy éghajlatválasztó „hegységek” (fehérral színezve) jelölik ki.

Ezzel analóg módon, az EIS „kulcsszó-térképén” úgy helyezkedik el a 50 *terminus technicus*, hogy az egymáshoz közel lévő kifejezések által reprezentált fogalmak – statisztikai értelemben – minél erősebb összefüggést mutassanak, és ezáltal kialakuljon az együtt változó fogalmi keret-  
ezések mintázata. Az ábrát úgy kell néznünk, mintha a mintázat minden irányban folytatódna, tehát az alul és felül, illetve a jobb- és balszélen elhelyezkedő pontok egymással szomszédosnak tekintendők. (Vagyis egy *tórusz* kiterített palástjáról van szó.)

Annak érdekében, hogy az emberi szem számára jobban kirajzolódjanak az összefüggő csoportok, színezzük ki a térképet a kulcsfogalmak egymáshoz való hasonlóságának mértéke szerint, és ekkor a 6. ábrán látható szabálytalan alakzatokat kapjuk. Az egy „krumplibá” eső kulcsszavak között összefüggés (korreláció) mérhető, ami annál erősebb, minél jobban hasonlít egymáshoz a színük. Hangsúlyozni szeretnénk, hogy ezek a (statisztikai) kapcsolatok az Európai Innovációs Eredménytábla 2010-24 közötti kiadásainak kulcsszó-gyakoriságán alapulnak. *Más szerzők által jegyzett szövegek másik részhalmaza, más időszakban más fogalmi kereteket mutathat!*

A térkép segítségével az 50 kulcsszó öt nagy csoportba szervezhető: ezek az innováció tudásalapú, politikai, értéklánc, vállalkozói, és posztmodern megközelítéséhez kapcsolódó mutatók. A csoportokat nem éles határvonalak választják el egymástól, néhány fogalom több gondolatkörbe is besorolható, az elnevezéseket a szakterület terminológiája alapján választottuk.



6. ábra. Az EIS 2010-2024 fogalmi keretrendszerének vizualizációja

(1) Az első nagy csoportba a **tudásalapú** gazdaság eszmeiségét tükröző fogalmak tartoznak, miszerint a növekedés kulcsa az intellektuális és pénzügyi tőke társasági formában történő együttműködése, a szellemi (és gazdasági) jogvédelem által szabályozott keretek között:

collaborating, knowledge, SMEs, finance, firm, intellectual (capital),  
patent, community, growth

(2) A másodikba az Európai Unió **politikai** szerveződéséhez, illetve általában magához a teljesítményértékeléshez kapcsolódó fogalmakat sorolta az algoritmus:

Europe, state, country  
innovation, performance, indicator

(3) A harmadik nagy csoportba az innováció **értéklánc** típusú megközelítéséhez (felsőoktatás → tudomány → akadémia-ipar együttműködés → szellemi jogvédelem → gazdasági hasznosítás), fogalmi keretezéséhez illeszkedő kulcsszavak kerültek, úgy mint:

tertiary, doctorate, graduates  
scientific, publications  
public, private, co-(publications)  
design, trademark  
export, venture

(4) Az innováció **vállalkozói** szemléletét, a kutatás-fejlesztési eredmények technológiai vállalkozások általi, termékek és szolgáltatások formájában történő értékesítését, tehát elsősorban a gazdasági hasznosítását kifejező kulcsszavak tartoznak a negyedik nagy csoportba:

GDP, population, enterprise, entrepreneurship  
expenditure, product, R&D, sales  
high-tech, services

(5) Végezetül az ötödik nagy csoportot azok a **posztmodern** fogalmak alkotják, melyek arra reflektálnak, hogy a globális klímaváltozás, a digitalizáció, és úgy általában a technológiai fejlődés alapjaiban rendezik át a munkaerőpiacot, és a társadalmi hatások mitigációja a tiszta piaci mechanizmusok felülbírálatát, erősödő kormányzati szerepvállalást feltételez:

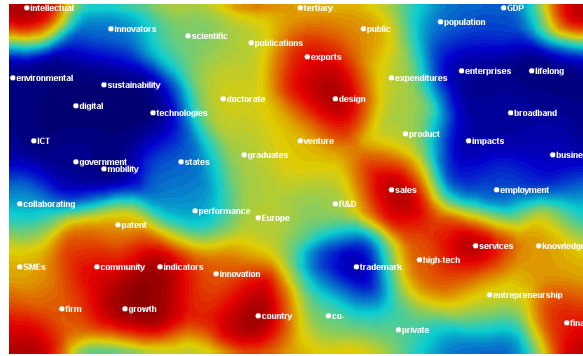
innovator, sustainability, digital, technologies, government, mobility  
business, environmental, ICT  
lifelong, broadband, impacts, employment

\* \* \*

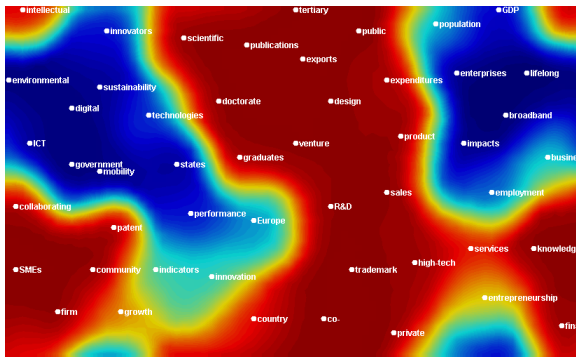
Az önszerveződő térképek kiválóan alkalmasak arra, hogy segítségükkel az Európai Innovációs Eredménytábla **fogalmi rendszerének időbeli fejlődését** látványosan megjelenítsük. Ehhez vezük újra elő a 6. *ábrán* bemutatott térképet, tehát az innovációhoz kötődő 50 kulcsszó olyan síkbeli elrendezését, ahol az egymáshoz közeli fogalmak jelentésbeli kapcsolatban állnak egymással, és az összefüggés annál erősebb, minél közelebb vannak egymáshoz a pontok. Azonban a korábbiaktól eltérően most más módon színezzük ki a térképet!

A színezés alapja ezúttal az, hogy az EIS egy-egy kiadásának szövegében mennyire gyakoriak bizonyos kulcsszavak, tehát egyfajta *hőtérképet* rajzolunk, ahol **meleg színekkel** jelöljük azokat a fogalmakat, melyek az adott év diskurzusában hangsúlyosan szerepelnek, míg **hideg színekkel** azokat, amelyekről kevesebb szó esik. A szavak síkbeli pozíciója megegyezik az eddigiekkel, így egymással vizuálisan könnyen összevethető a különböző évek kulcsszó-profilja.

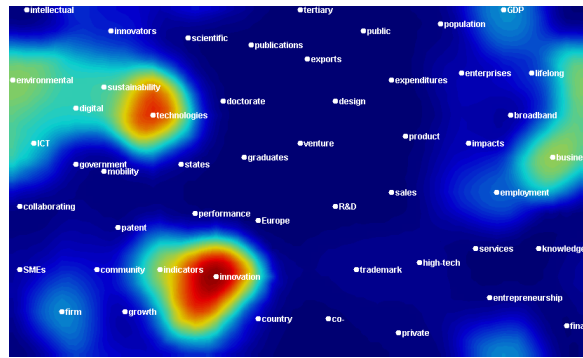




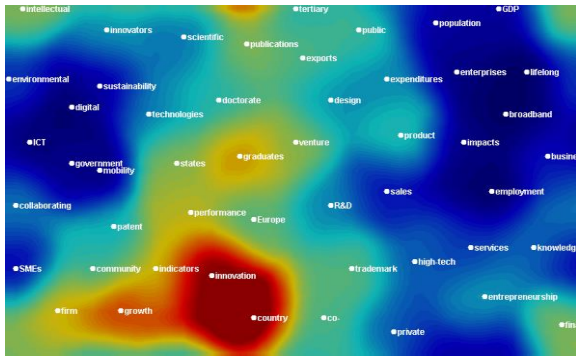
*EIS 2010*



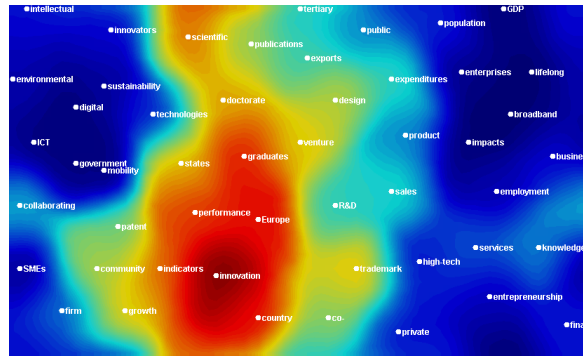
*EIS 2011*



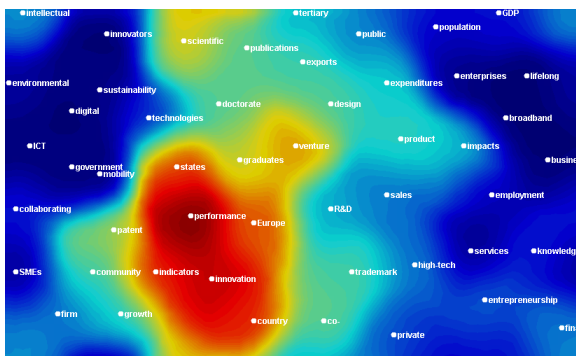
*RIS 2012*



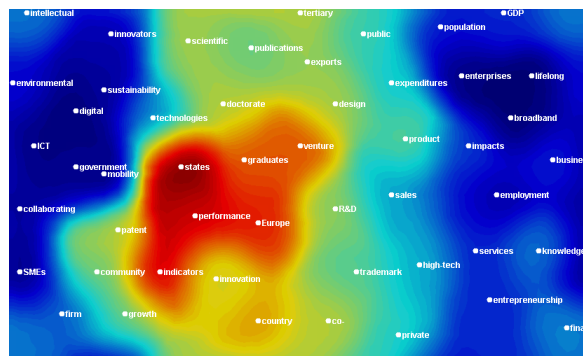
*EIS 2013*



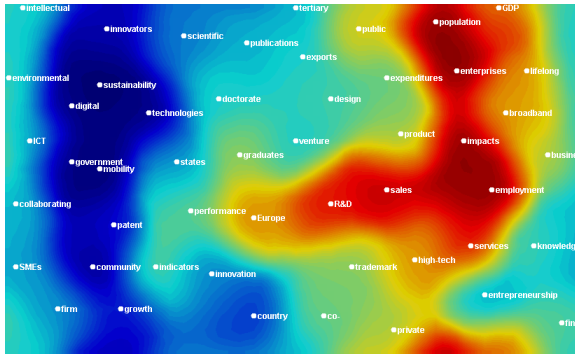
*EIS 2014*



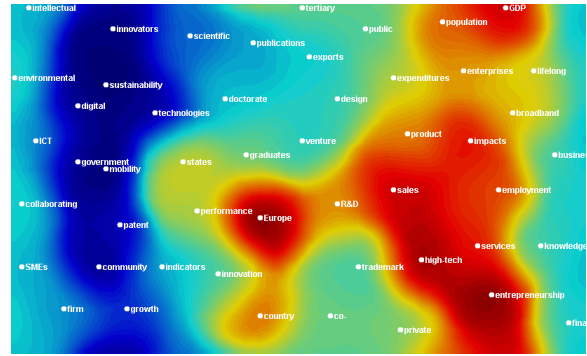
*EIS 2015*



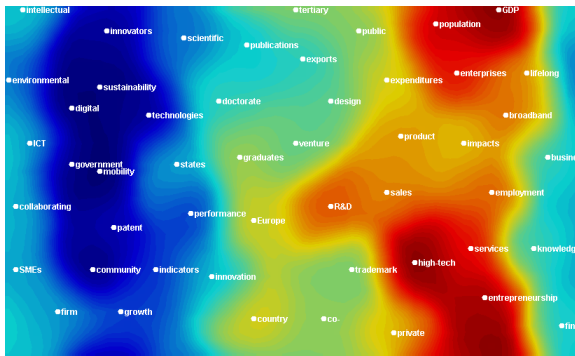
*EIS 2016*



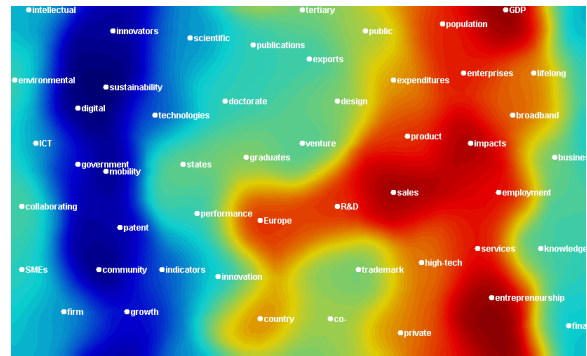
EIS 2017



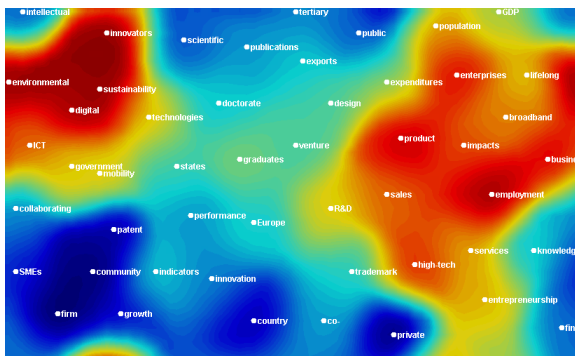
EIS 2018



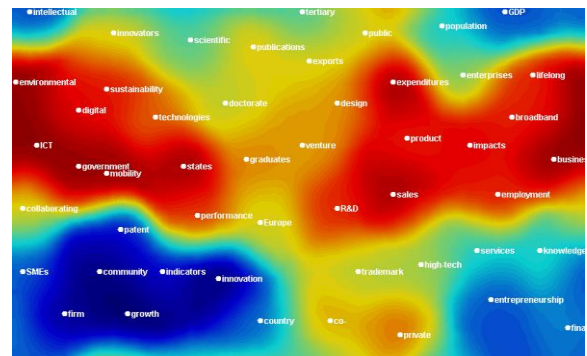
EIS 2019



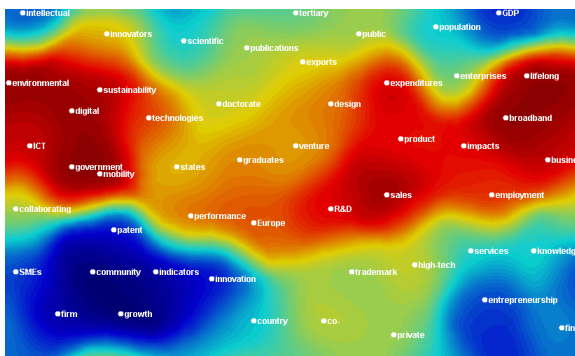
EIS 2020



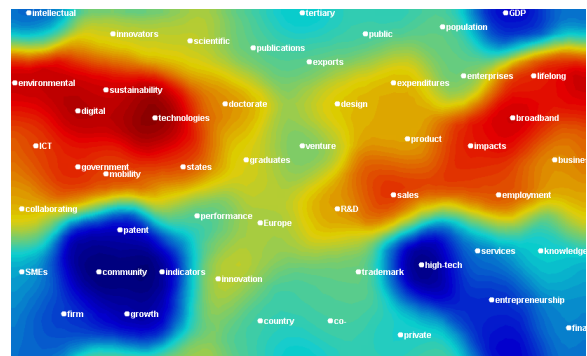
EIS 2021



EIS 2022



EIS 2023



EIS 2024

7. ábra. Az Európai Innovációs Eredménytábla fogalmi fejlődésének vizualizációja

A 7. ábrán jól megfigyelhető, hogy az Európai Innovációs Eredménytábla a 2010 és 2024 közötti időszakban jelentős fogalmi változáson (fejlődésen) ment keresztül. A korai évek az „útkeresésről”, a széles spektrumú gondolkodásról, a közös nyelv kialakításáról szóltak, 2013 után viszont négyéves periódusokban, ismétlődő jelleggel átrendeződtek a fogalmi keretek:

- 2010-2012: az útkeresés időszaka, hagyományos kérdésseltevések, az innováció szinte mindennel aspektusával megpróbál foglalkozni (a 2012 átmeneti év, EIS helyett RIS van);
- 2013-2016: értéklánc-alapú megközelítés, a közös európai policy és a teljesítményértékelés kiemelt fontossága;
- 2017-2020: hangsúlyeltolódás a vállalkozói szemlélet irányába;
- 2021-2024: posztmodern kérdésseltevések, tágabb társadalmi és környezeti keretezés.

Az elmúlt másfél évtized tapasztalatai alapján – bár teljes bizonyossággal nem jelenthetjük ki – az várható, hogy 2025-ben egy újabb paradigmaváltásnak lehetünk tanúi az Európai Innovációs Eredménytábla fogalmi keretrendszerében, amely feltételezésünket a 2024. évi EIS jelentés 14. oldalán található következő mondat is alátámasztja: *„A revision of the EIS indicators and methodology will start in Autumn 2024 and changes agreed with the Member States will be applied for the EIS 2025.”*



## A GII fogalmi fejlődése a kulcsszavak alapján

Vizsgáljuk most meg a másik nagy, nemzetközileg ismert és elismert innovációs rangsor, avagy mérőrendszer, a Globális Innovációs Index fogalmi kereteinek fejlődését.

Elemzésünk kiindulópontja ezúttal a *World Intellectual Property Organization (WIPO)* honlapján<sup>13</sup> elérhető 2009 és 2024 közötti GII jelentések PDF formátumú szövege. A megosztásra és nyomtatásra optimalizált, gazdagon illusztrált fájlokat most is egyszerű szövegfájlá (.txt) konvertáltuk, és lefuttattuk az automatikus kulcsszókinyerési algoritmust. Referenciaszöveggént újra a jól bevált BAWE (British Academic Written English) korpuszt választottuk, vagyis arra kerestük a választ, hogy mennyiben különbözik az *GII specifikus terminológiája* az általános, egyetemi műveltséggel rendelkező angol anyanyelvűek által megírt szövegektől. Azt is ellenőriztünk, hogy a 2009, 2010, ..., 2024-es kiadások szövegében vannak-e specifikus, csak az egyes évfolyamokra jellemző fogalmak.

Ha a kapott kulcsszó- és gyakoriság-listát inputként használjuk a K-SOM algoritmusban, akkor a 8. ábrán látható „térkép” rajzolódik ki. Összevetve ezt az EIS 6. ábrán vizualizált koncepcionális keretrendszerével, azonnal szembeötlik, hogy most jóval kevesebb nagy klasztert látunk, viszont sokkal több a magányos, másokhoz kevésbé kapcsolódó kifejezés. Konkrétan két csoport azonosítható, míg a „magányosság” nem azt jelenti, hogy az adott kulcsszónak ne lenne kapcsolata a többi fogalommal, hanem csak annyit, hogy semelyik másik fogalomhoz nem kötődik jobban – a GII jelentések szövegében (!) – mint a többi kifejezéshez.



8. ábra. A GII 2009-2024 fogalmi keretrendszerének vizualizációja

<sup>13</sup> <https://www.wipo.int/en/web/global-innovation-index>

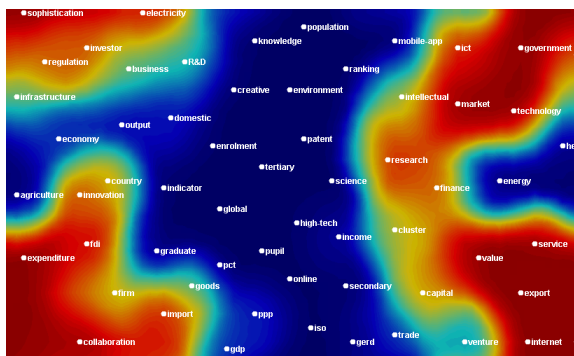
(1) Az innovációhoz a szakirodalomban **szokásos** módon társított fogalmak minden további látható elkülönülés nélkül, egy nagy csokorba szerveződnek. Szakavatott olvasóként természetesen felfedezhetünk közöttük további, részletesebb felosztást eredményező kisebb csoportosulásokat, de maga a Globális Innovációs Keretrendszer, legalábbis a szövegek tanúsága szerint, nem igazán hangsúlyozza jobban vagy kevésbé az innováció különféle aspektusait.

creative, enrolment, environment, GDP, GERD, global, goods, graduate, high-tech, import, income, ISO, knowledge, mobile-app, online, patent, PCP, population, PPP, pupil, ranking, science, secondary, tertiary, trade

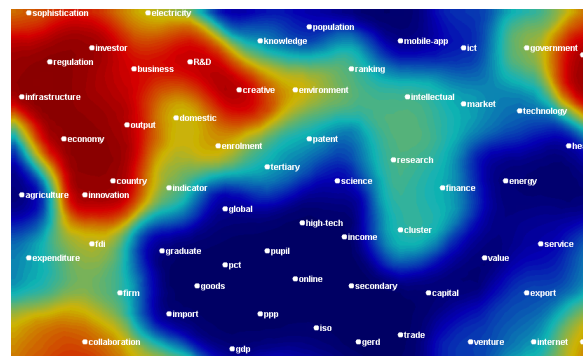
(2) A **piaci viszonyokat**, a befektetők és a kormányzat viszonyát leíró kifejezések találhatóak a másik nagy, az előzőhöz képest marginális helyzetű csoportban. Ezek különválása arra mutat rá, hogy a WIPO nézőpontjából bizonyos fogalmak különös fontossággal bírnak az innovációs ökoszisztémák jellemzése tekintetében.

business, credit, expenditure, export, government, ICT, infrastructure, internet, investor, market, quality, regulation, service, sophistication, technology, value, venture

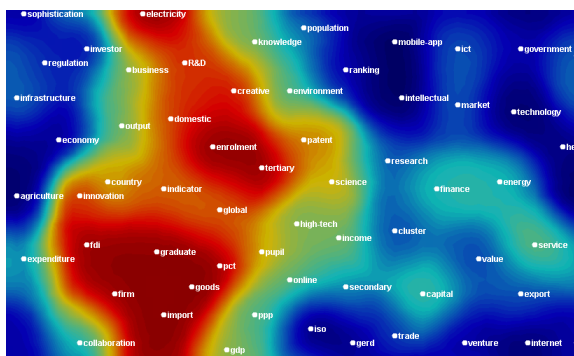
Ez a fogalmi stabilitás (vagy differenciálatlanság) jól látható akkor is, ha a Globális Innovációs Index fogalmi rendszerének **időbeli** változását – pontosabban **változatlanságát** – vizualizáljuk az önszerveződő térképek segítségével. A 9. ábra idősora világosan bemutatja, hogy két év kezdeti bizonytalanság után kialakultak a stabil fogalmi keretek, és ezután évről-évre csak ezen a gondolati rendszeren belül történtek hangsúlyeltolódások, amit a tematikus célszámok speciális terminológiája indokolt (2015: innovációpolitika, 2016: globalizáció, 2017: élelmiszer, 2018: energia, 2019: egészség, 2020: finanszírozás).



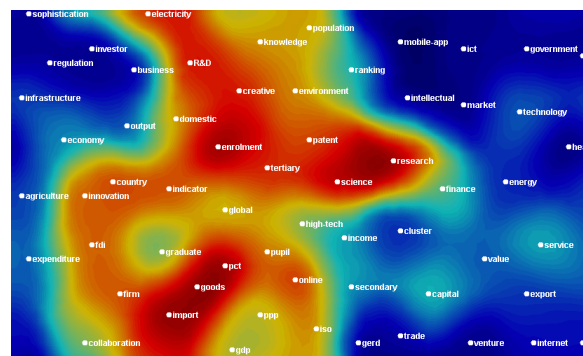
GII 2009



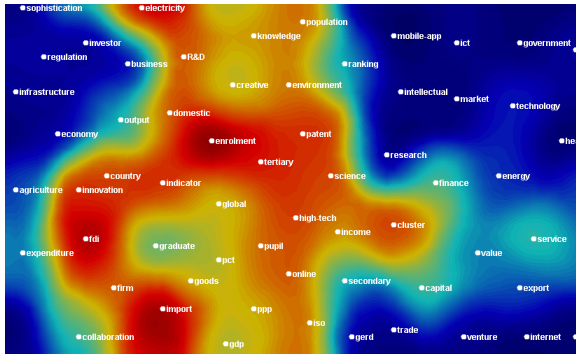
GII 2010



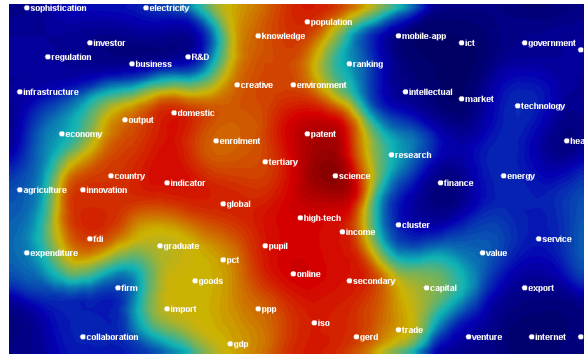
GII 2011



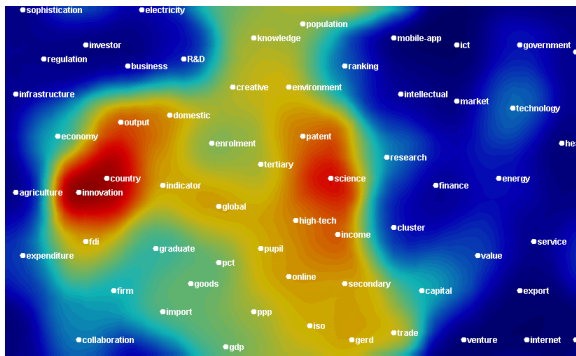
GII 2012



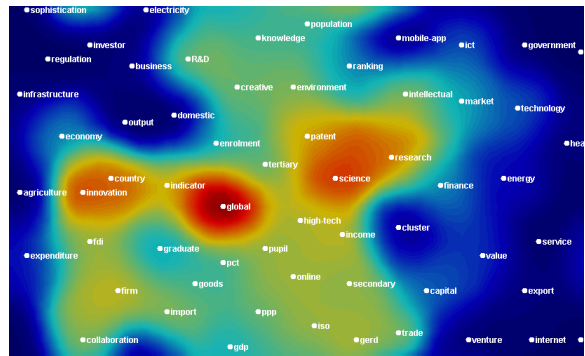
GII 2013



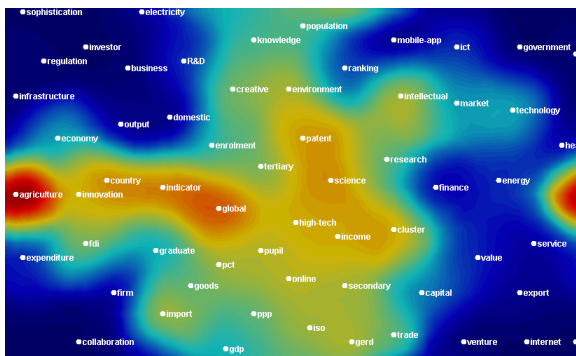
GII 2014



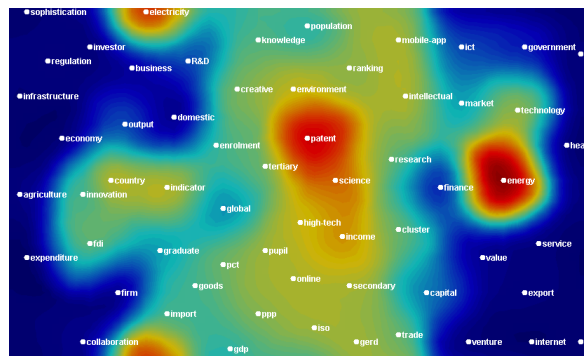
GII 2015



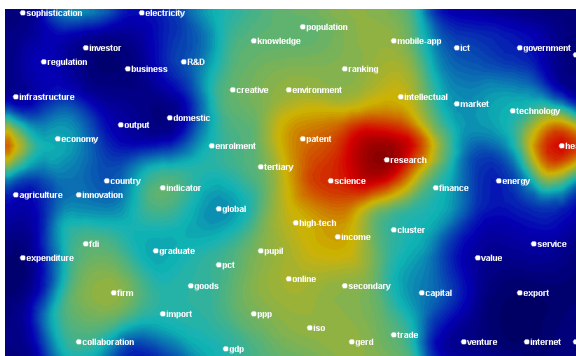
GII 2016



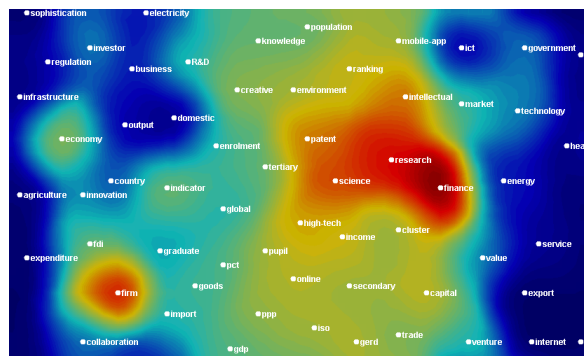
GII 2017



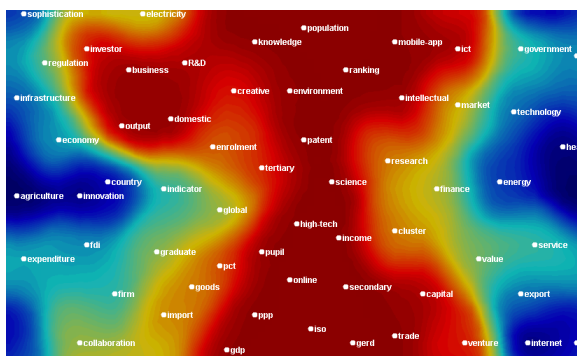
GII 2018



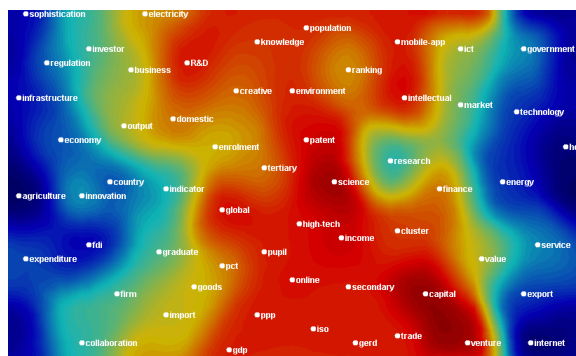
GII 2019



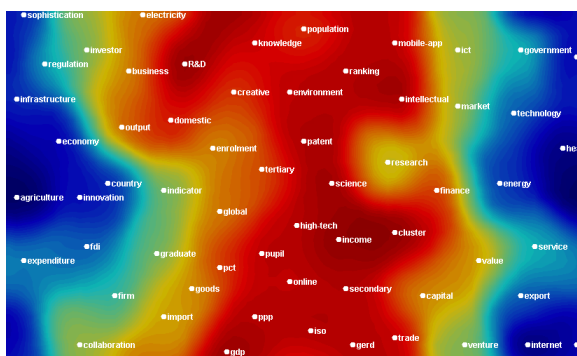
GII 2020



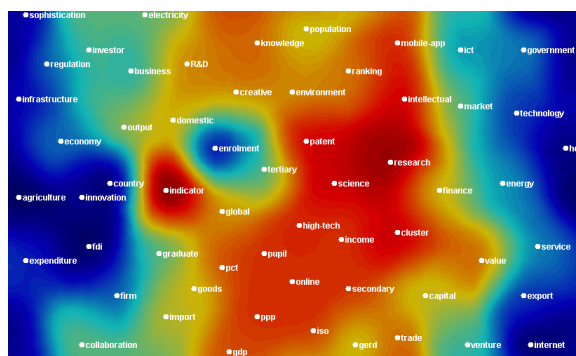
GII 2021



GII 2022



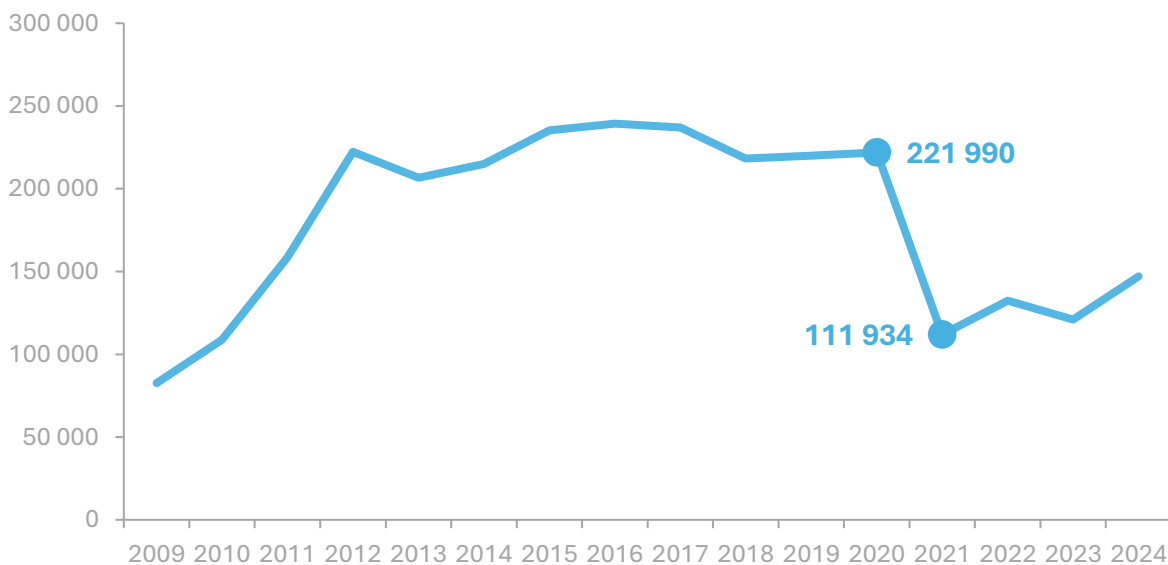
GII 2023



GII 2024

9. ábra. A Globális Innovációs Index fogalmi fejlődésének vizualizációja

2020 és 2021 között is egy éles változást láthatunk, amikor is a fogalmi keretek továbbra sem változtak meg alapvetően, viszont sokkal **koncentráltabban** jelennek meg a 2021-24 közötti években, mint előtte. Ennek az a magyarázata, hogy a jelentés terjedelme lényegében megfelelő, míg 2021-ben 221.990, addig 2022-ben csak 111.934 szóból állt összesen (10. ábra).



10. ábra. A Globális Innovációs Index köteteinek terjedelme



## A fogalmi változás mértéke: a kölcsönös információ

Miután a kulcsszavak segítségével meghatároztuk a mérőrendszerek fogalmi kereteit, és a relatív gyakoriságok időbeli változása alapján, az önszerveződő térképek felhasználásával vizualizáltuk a fogalmi rendszerek fejlődését, már csak az maradt hátra, hogy javaslatot tegyünk a változás gyorsaságának mérésére.

Karen Spärck Jones 1972-ben a *Journal of Documentation* című folyóiratban publikált egy tanulmányt<sup>14</sup>, amelyben először javasolta a kifejezésspecifikusság mérőszámaként a később **inverz dokumentumfrekvencia** néven ismertté vált mennyiséget. Az IDF-et (inverse document frequency) úgy kapjuk meg, hogy a vizsgált gyűjteményben leszámoljuk azokat a dokumentumokat, amelyek tartalmazzák a kérdéses kifejezést. Jones intuíciója az volt, hogy egy olyan keresési kifejezés, amely sok dokumentumban előfordul, nem jó megkülönböztető jegy, és ezért kisebb súlyt kell neki adni, mint egy olyan kifejezésnek, amely kevés dokumentumban fordul elő.

A heurisztikus módon javasolt mérőszám óriási lökést adott az információkeresés területének. A korábban már megismert TF-fel (term frequency, a kifejezés gyakorisága magában a dokumentumban, ebben az esetben minél több, annál jobb) párosítva szinte minden kifejezéssúlyozási sémába bekerült. Az általánosan **TF · IDF** néven ismert súlyozási sémák osztálya, amelyek az IDF-mérték és a TF-mérték valamely változatának szorzatát tartalmazzák, rendkívül robusztusnak és szinte felülmúlhatatlannak bizonyult, még az elméletileg nagyon alaposan kidolgozott modellek és algoritmusok által is.

Az IDF-ről szóló tanulmányok, a vele kapcsolatos kritikák egyik visszatérő eleme a heurisztikus jellege. Bár a gyakorlatban jól működött, elméleti megalapozása legalább három évtizeden át gondot okozott a kutatóknak, míg végül Stephen Robertsonnak<sup>15</sup> sikerült **információelméleti** igazolást találni rá.

Mind a kulcsszógyakoriság, mind az inverz dokumentumgyakoriság kifejezhető az információelmélet fogalmaival; ez segít megérteni, hogy miért van a szorzatuknak jelentése a dokumentum együttes információtartalma szempontjából. A  $p(d|t)$  eloszlásra vonatkozó jellemző feltételezés az, hogy:

$$p(d|t) = \frac{1}{|\{d \in D: t \in d\}|}$$

A  $D$  korpusz egy véletlenszerűen kiválasztott  $D$  dokumentumának feltételes entrópiája, attól függően, hogy tartalmaz-e egy adott  $t$  kifejezést (és feltételezve, hogy minden dokumentumot azonos valószínűséggel választunk ki) a következő:

$$\begin{aligned} H(D|T = t) &= - \sum_d p(d|t) \log p(d|t) = - \log \frac{1}{|\{d \in D: t \in d\}|} = \\ &= \log \frac{|\{d \in D: t \in d\}|}{|D|} + \log |D| = -idf(t) + \log |D|. \end{aligned}$$

<sup>14</sup> Karen Spärck Jones: „A statistical interpretation of term specificity and its application in retrieval.” *Journal of Documentation*, 28(1), 1972, pp. 11-21.

<sup>15</sup> Robertson, S.: „Understanding inverse document frequency: on theoretical arguments for IDF”, *Journal of Documentation*, 60(5), 2004, pp. 503-520. <https://doi.org/10.1108/00220410410560582>

Az előző egyenletben  $\mathcal{D}$  és  $\mathcal{T}$  véletlen változókat jelölnek, amelyek egy *dokumentum* vagy egy *kifejezés* „kiválasztásának” felelnek meg. A kölcsönös információ a következőképpen fejezhető ki:

$$M(\mathcal{T}; \mathcal{D}) = H(\mathcal{D}) - H(\mathcal{D}|\mathcal{T}) = \sum_t p_t (H(\mathcal{D}) - H(\mathcal{D}|W = t)) = \sum_t p_t \cdot \text{idf}(t).$$

Az utolsó lépésben a  $p_t$ -t, vagyis a  $t$  kifejezés kihúzásának feltétel nélküli valószínűségét kifejezzük a  $d$  dokumentum (véletlenszerű) kiválasztása függvényében, és így azt kapjuk, hogy:

$$M(\mathcal{T}; \mathcal{D}) = \sum_{t,d} p_{d|t} \cdot p_d \cdot \text{idf}(t) = \sum_{t,d} \text{tf}(t, d) \cdot \frac{1}{|D|} \cdot \text{idf}(t) = \frac{1}{|D|} \sum_{t,d} \text{tf}(t, d) \cdot \text{idf}(t).$$

Ez a kifejezés azt mutatja, hogy az összes lehetséges kifejezés és dokumentum párra összegezzük a  $\text{tf} \cdot \text{idf}$  szorzatokat, akkor megkapjuk a dokumentumok és kifejezések közötti **kölcsönös információt**, figyelembe véve a közös eloszlásuk összes sajátosságát. Minden egyes  $\text{tf} \cdot \text{idf}$  tehát egy-egy (kulcsszó; dokumentum) párhoz kapcsolódó *információs bitet* hordoz.

Ha a fentiekben levezetett képletbe behelyettesítjük az Európai Innovációs Eredménytábla, és a Globális Innovációs Index szövegeinek elemzése során kapott kulcsszó- és (inverz) dokumentumgyakoriságokat, akkor a 3. táblázatban látható értékek adódnak. Figyeljük meg, hogy bár a GII minden szempontból bővebb korpusznak tekinthető, mint az EIS, a benne foglalt információ mennyisége mégiscsak kevesebb! Ez egybevág azzal az intuitív érzésünkkel – amit az előző fejezetekben bemutatott vizualizációk is megerősítettek –, hogy fennállásának másfél évtizede alatt a GII kevesebbet változott, mint az EIS.

3. táblázat. A kölcsönös információ mennyisége az EIS és GII jelentések szövegében

	EIS	GII
Megjelenés éve	2010–2024	2009–2024
Kötetek száma	15 db	16 db
Kifejezések átlagos száma	33 308	179 901
Egyedi szavak száma	6 934	29 992
$\Sigma \text{TF} \cdot \text{IDF} /  D $	4 208,437	3 134,294

**Összegezve:** a kölcsönös információ fogalmát felhasználva számszerűsítettük a mérőrendszerek koncepcionális kereteinek változási (adaptációs) sebességét. A nagyobb kölcsönös információ nagyobb változékonyságot mutat, ami jelenthet jobb *adaptációs képességet*, de jelentheti a fogalmi keretek stabilitásának hiányát is. A fejlődéstörténet megmutatta, hogy evolúciós szempontból mindkét stratégia (nagy, stabil, illetve kisebb, gyorsabban mutálódó „génkészlet”) lehet sikeres, végső választ csak az idő ad arra a kérdésre, hogy melyik fogalmi rendszer alkalmazkodott jobban a környezeti feltételekhez.

## Validálás: a digitalizáció fogalmának változása

**A fogalmi változás sebességének jellemzésére bevezetett kölcsönös információ értékét ezúttal egy teljesen másik területre, a digitalizációra kidolgozott mérőrendszerekre számoljuk ki, így ellenőrizve a módszerünk érvényességét.**

Az Európai Bizottság 2014 óta követi nyomon a tagállamok digitális fejlődését a digitális gazdaság és társadalom fejlettségét mérő mutatóról (DESI) szóló jelentések révén. Az adott évi *DESI-jelentések*<sup>16</sup> főként a megelőző évi adatokon alapulnak, és nyomon követik az uniós tagállamokban a digitális területen elért eredményeket. A Covid19-világjárvány idején a tagállamok komoly előrelépést tettek digitalizációs erőfeszítéseik terén, de még mindig nehezen tudták megszüntetni a digitális készségek, a kkv-k digitális átalakulása, és a fejlett telekommunikációs hálózatok kiépítése terén fennálló hiányosságokat.

Az IMD (International Institute for Management Development) *Digitális Versenyképességi Világrangsora*<sup>17</sup> nyolcadik éve méri 67 gazdaság kapacitását és felkészültségét a digitális technológiák bevezetésére és felfedezésére, mint a gazdasági átalakulás kulcsfontosságú hajtóerejét az üzleti életben, a kormányzatban és a társadalomban. Az üzleti és kormányzati vezetők megkérdezésén, valamint a felmérésből származó válaszokon alapuló rangsor segít a kormányoknak és a vállalatoknak abban, hogy megértsék, hová összpontosítsák erőforrásaikat, és melyek lehetnek a legjobb gyakorlatok a digitális átalakulás megkezdésekor vagy továbbfejlesztésekor.

Elemzésünk kiindulópontja ezúttal a 2017 és 2024 közötti DESI és WDCR jelentések PDF formátumú szövege volt. Ebbe a nyolcéves intervallumba éppen beleesik a Covid-19 világjárvány időszaka, ami a digitalizáció folyamatát hatalmas mértékben felgyorsította, érvényességi körét korábban ismeretlen területekre is kiterjesztette, így előzetesen azt várjuk, hogy a két mérőrendszer fogalmi kereteinek változása valamilyen módon majd tükrözi mindezt.

A megosztásra és nyomtatásra optimalizált, gazdagon illusztrált fájlokat a már megszokott módon egyszerű szövegfájlra (.txt) konvertáltuk, és lefuttattuk az automatikus kulcsszókinyerési algoritmust. Referenciaszövegként újra a bevált BAWE (British Academic Written English) korpuszt választottuk, vagyis arra kerestük a választ, hogy mennyiben különbözik a *DESI és a WDCR specifikus terminológiája* az általános, egyetemi műveltséggel rendelkező angol anyanyelvűek által megírt szövegektől. Azt is ellenőriztünk, hogy a 2017, 2018, ..., 2024-es kiadások szövegében vannak-e specifikus, csakis az egyes évfolyamokra jellemző fogalmak.

Ha a kapott kulcsszó- és gyakoriság-listát inputként használjuk a K-SOM algoritmusban, akkor a 11. és 12. ábrán látható „térképek” rajzolódnak ki. Összevetve a két ábrát azonnal szembeötlik, hogy míg a *DESI fogalmi rendszere jól struktúrált*, addig a WDCR néhány kivételtől eltekintve homogén masszaként tekint a fogalmakra, ami természetesen nem azt jelenti, hogy egy-egy kulcsszónak ne lenne kapcsolata a többi fogalommal, hanem csak annyit, hogy semelyik másik fogalomhoz nem kötődik jobban – a WDCR jelentések szövegében (!) – mint a többi kifejezéshez. A 13. és 14. ábrák idősoros megjelenítés alapján pedig az érzés alakul ki bennünk, hogy a *DESI lekövette a digitalizációban bekövetkezett fogalmi változásokat*, a WDCR azonban nem (annyira).

<sup>16</sup> Digital Economy and Society Index, <https://digital-strategy.ec.europa.eu/hu/policies/desi>

<sup>17</sup> IMD World Digital Competitiveness Ranking, <https://www.imd.org/centers/wcc/world-competitiveness-center/rankings/world-digital-competitiveness-ranking/>

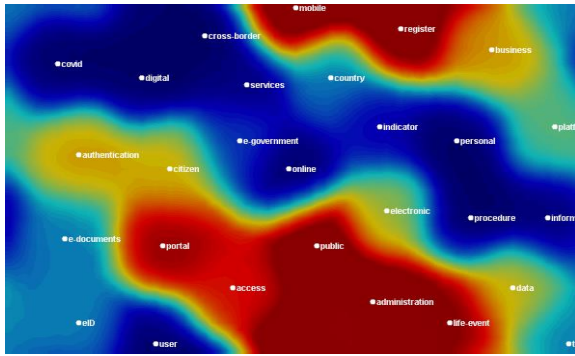


11. ábra. A DESI 2017-2024 fogalmi keretrendszerének vizualizációja

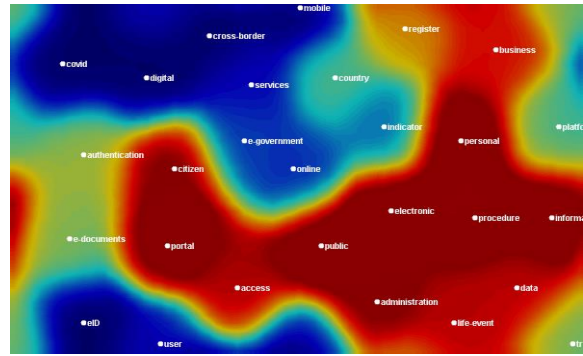


12. ábra. A WDCR 2017-2024 fogalmi keretrendszerének vizualizációja

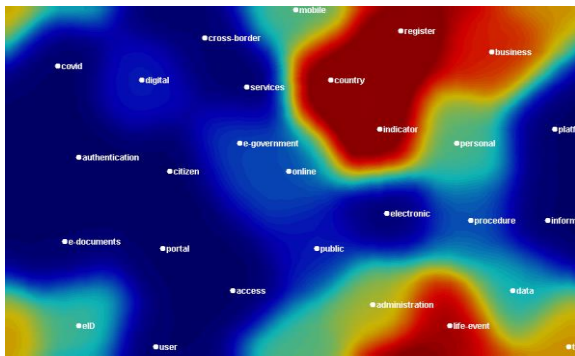




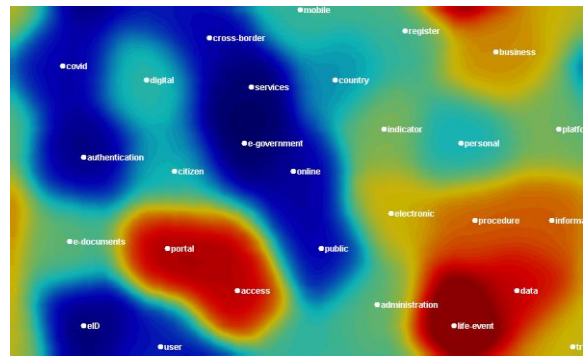
DESI 2017



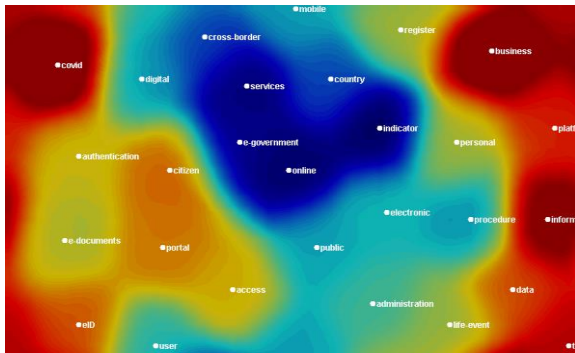
DESI 2018



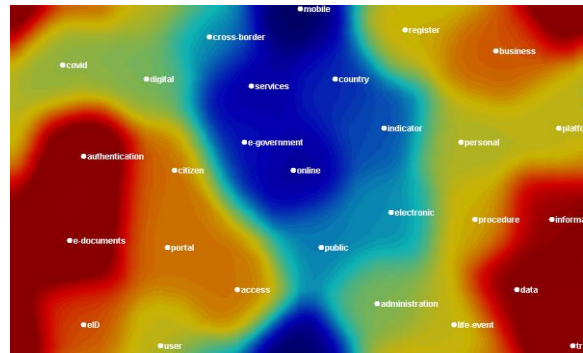
DESI 2019



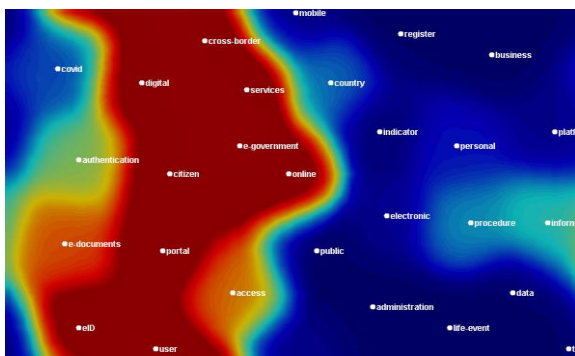
DESI 2020



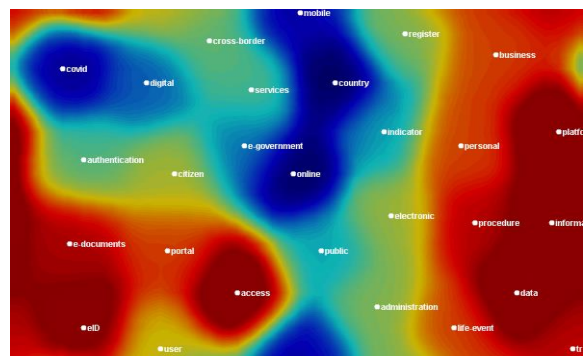
DESI 2021



DESI 2022

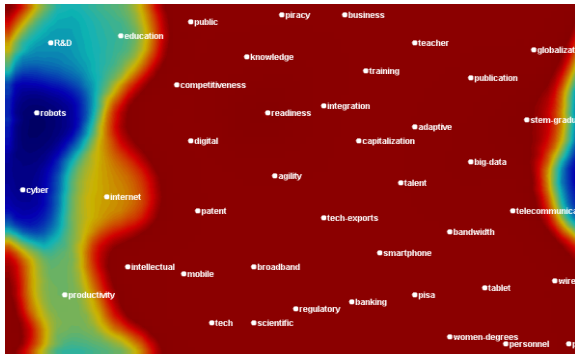


DESI 2023

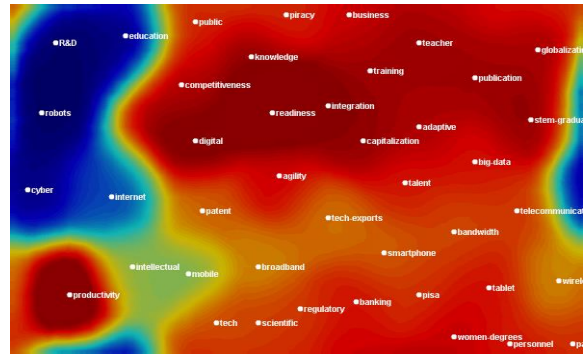


DESI 2024

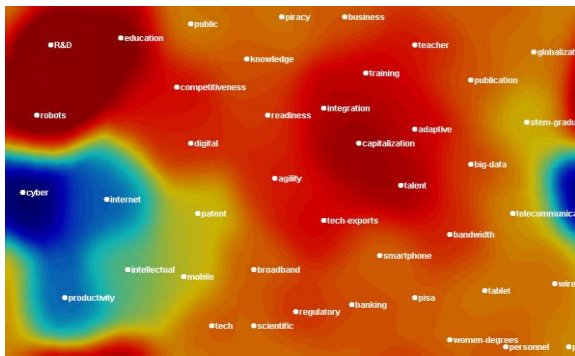
13. ábra. A Digital Economy and Society Index fogalmi fejlődésének vizualizációja



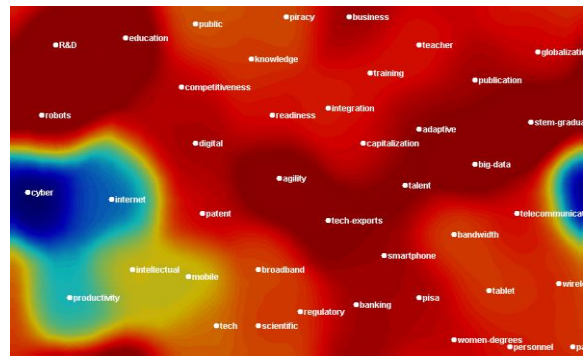
WDCR 2017



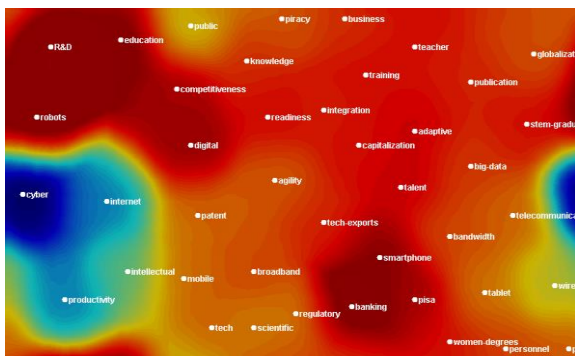
WDCR 2018



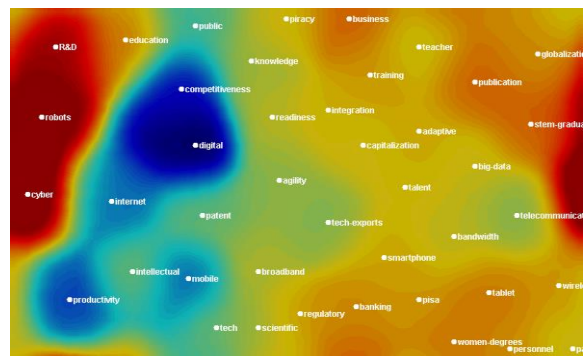
WDCR 2019



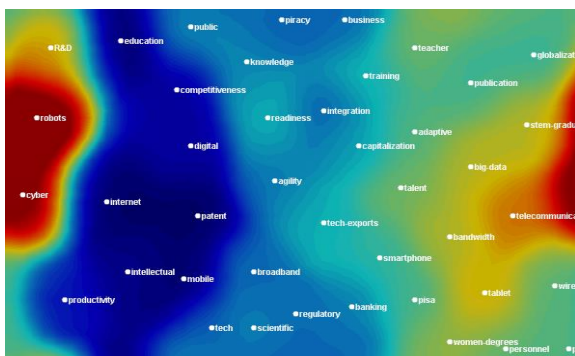
WDCR 2020



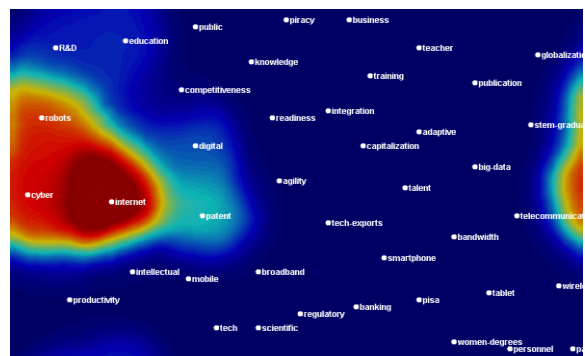
WDCR 2021



WDCR 2022



WDCR 2023



WDCR 2024

14. ábra. A World Digital Competitiveness Ranking fogalmi fejlődésének vizualizációja

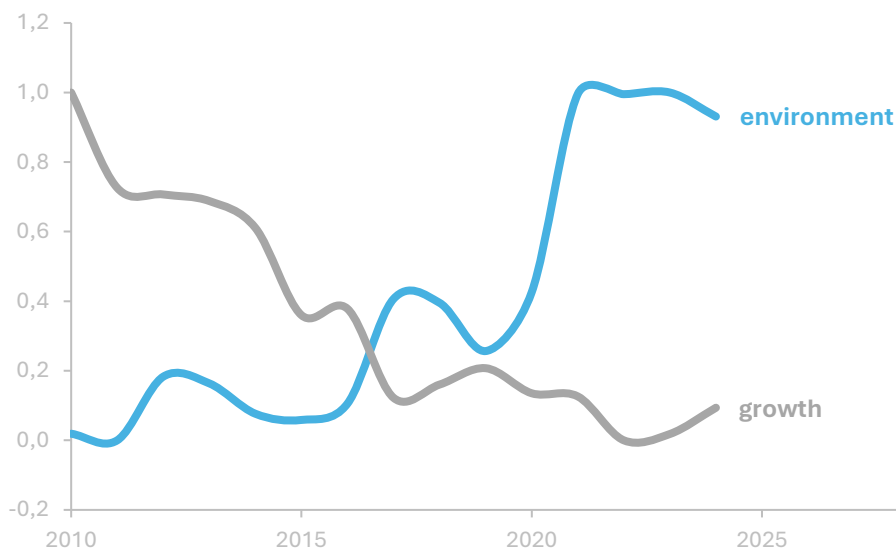
Ha a kölcsönös információ előző fejezetben megismert képletébe behelyettesítjük a DESI jelentések, és a Digitális Versenyképességi Világrangsor szövegeinek elemzése során kapott kulcszó- és (inverz) dokumentumgyakoriságokat, akkor a 4. táblázatban látható értékek adódnak. Jól megfigyelhető, hogy mennyivel nagyobb a DESI jelentésekre kiszámított kölcsönös információ értéke, mint a WDCR jelentéseké, ami egybevág a vizualizációk alapján megfogalmazódott véleményünkkel. Az már természetesen értelmezés kérdése, hogy ezt a fogalmi változást *adaptációként* vagy pedig a *stabilitás hiányaként* könyveljük el.

4. táblázat. A kölcsönös információ mennyisége az EIS és GI jelentések szövegében

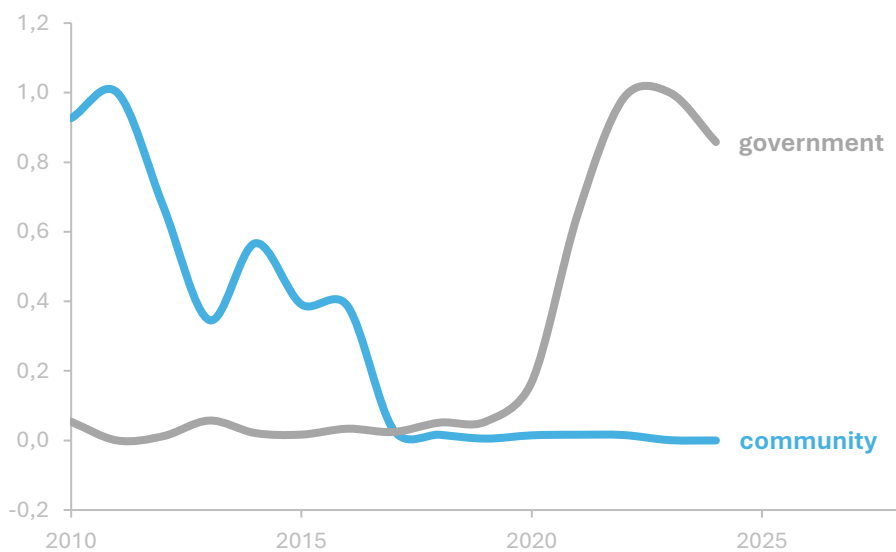
	<b>DESI</b>	<b>WDCR</b>
Megjelenés éve	2017–2024	2017–2024
Kötetek száma	8 db	8 db
Kifejezések átlagos száma	44 741	36 619
Egyedi szavak száma	10 569	6 365
$\Sigma TF \cdot IDF /  D $	<b>7 091</b>	<b>3 928,6</b>

## Esettanulmányok

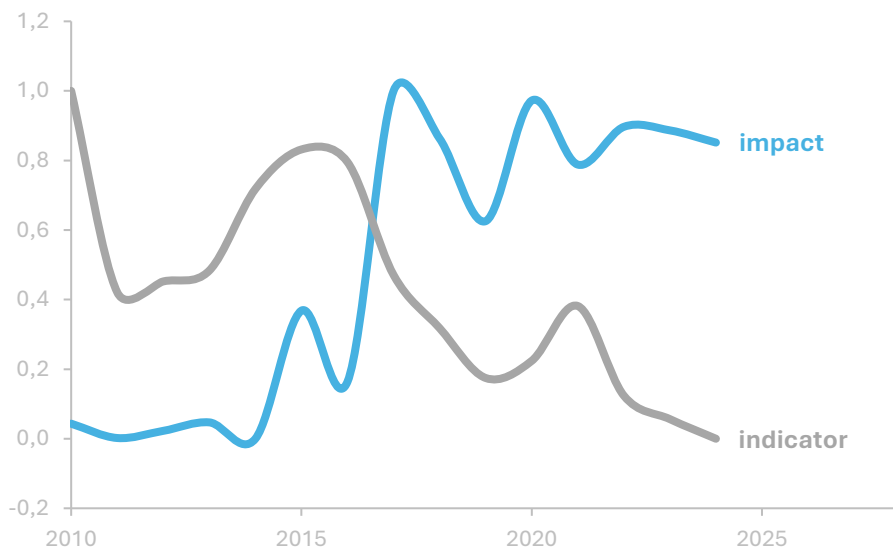
A következőkben bemutatunk néhány érdekességet a fogalmi változások világából, külön kommentár nélkül, mivel a grafikonok magukért beszélnek.



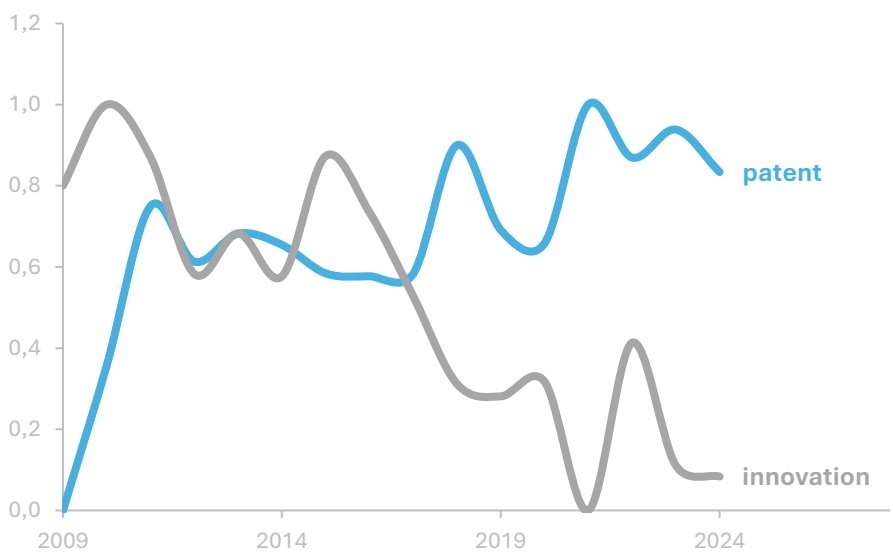
15. ábra. A növekedés és a környezet védelmének relatív fontossága az EIS szerint



16. ábra. A közösségi és kormányzati szerepvállalás relatív fontossága az EIS szerint



17. ábra. Az indikátorok és a hatások relatív fontossága az EIS szerint



18. ábra. Hogyan értelmezi az innováció fogalmát a WIPO által szerkesztett GII?

## Ajánlott irodalom

**A tanulmányban érintett témakörök, mint az innováció- és fejlesztéspolitika, a nemzetközi rangsorok, és különösen a fogalmi fejlődés iránt mélyebben érdeklődő olvasók számára az alábbi szakirodalmat javasoljuk kiindulásként.**

Tero Erkkilä (editor): *Global University Rankings. Challenges for European Higher Education*, Palgrave-Macmillan, 2013, 273 pages, ISBN 978 1 137 29686 3.

Ellen Hazelkorn, Georgiana Mihut (editors): *Research Handbook on University Rankings. Theory, Methodology, Influence and Impact*, Elgar Handbooks in Education, 2021, 584 pages, ISBN: 978 1 78897 497 4.

*Innovation Policy. A Guide for Developing Countries*, The World Bank, 2010, 436 pages, ISBN 978 0 8213 8269 1.

P.T.M. Marope, P.J. Wells, E. Hazelkorn (editors): *Rankings and Accountability in Higher Education. Uses and Misuses*, UNESCO, 2013, 296 pages, ISBN 978-92-3-001156-7.

Pléh Csaba, Lukács Ágnes: *Pszicholingvisztika 1-2. Magyar pszicholingvisztikai kézikönyv*. Akadémiai Kiadó, 2014, 1445 oldal, ISBN 978 963 05 9499 8.

Jung Cheol Shin, Robert K. Toutkoushian, Ulrich Teichler (editors): *University Rankings. Theoretical Basis, Methodology, and Impacts on Global Higher Education*. Springer, 2011, 286 pages, ISBN 978 94 007 1115 0.

Vickers, Peter, *Identifying Future-Proof Science* (Oxford, 2022; online edn, Oxford Academic, 20 Oct. 2022), <https://doi.org/10.1093/oso/9780192862730.001.0001>.

Frans A. van Vught, Frank Ziegele (editors): *Multidimensional Ranking. The Design and Development of U-Multirank*, Springer, 2012, 198 pages, ISBN 978 94 007 3004 5.