A Coursera aranykora, avagy az önkorlátozó mesterséges intelligencia

(Golden Age for Coursera – or self-limitation of artificial intelligence)

Pitlik László, Pitlik Marcell - MY-X team

Kivonat: A cikk egy big-data-mesterséges-intelligencia-témakörben folyó tanfolyam vizsgát helyettesítő dokumentuma számára készített demo, ahol a tanfolyami információk megértettségét és alkalmazhatóságát valós adatvagyonon, valós kérdésekre adott minél inkább hiteles válaszok formájában kellett a kiképzendőnek úgy bizonyítania, hogy a tanfolyam-vezetővel együtt képes legyen publikálni a cég és/vagy személy profiljának megfelelő jelenségkört érintve. A választott téma egy oktatást szervező, oktatást végző cég esetén az MOOC-szolgáltatók közül a Coursera 2004-2021 közötti története alapján az aranykor feltárása volt. Az elemzés egyik részeredménye egy, a mesterséges intelligencia önkorlátozó képességeit demonstrálni képes adatvagyon felismerése és az önkorlátozás mértékének szabályozása volt.

Kulcsszavak: MOOC, távoktatás, függvény-szimmetria, több-adat-nem-jobb-modell-eset, robotszem

Abstract: The paper is a demonstration material for a (enterprise-oriented/inner) course about big-data&data-mining. This kind of publication will be expected from each course-member as a kind of evidence proof concerning the needed competence-change based on the course. The publication demonstrates that a real (big) data source (like Google Trends) can be involved to model real questions (like existence of golden ages based on robot-eyes) in a step-by-step better level (see parallel models). The course-member will create an own publication in cooperation with the conductor of the course. The potential themes should have direct connection to the enterprise (keyword: education) and/or private preferences (like brand-awareness in case of beauty products, autos, IT-brands – see. Laptops/smartphones, etc.). Parallel to Coursera, it is possible to choose other MOOC-providers where the co-author has to check each details of this publication too. The real result of this publication where no real golden age could be derived for Coursera is: what kind of input data constellation can lead to massive ratio of invalidity.

Keywords: MOOC, e-learning, symmetry of functions, more-data-less-robust-model-case, robot-eye

# Bevezetés

A cikk egy mesterséges-intelligencia+big-data kurzus kapcsán azt hivatott demonstrálni, hogy a tanfolyamrésztvevő szerző a kurzusvezetővel együtt képes pl. TDK/szakdolgozat-jellegű (ezen dokumentumok magját jelentő) publikáció készítésére a tananyag alapján.

Előzmények:

* <https://miau.my-x.hu/miau2009/index.php3?x=e0&string=mooc>
* <https://miau.my-x.hu/miau2009/index.php3?x=e0&string=golden>
* <https://miau.my-x.hu/miau/273/oil1.xlsx>
* <https://miau.my-x.hu/miau/273/sap_oracle_ms_dell_samsung.xlsx>
* <https://miau.my-x.hu/miau/273/otdk_2021_abstract_standards.docx>
* <https://miau.my-x.hu/miau/273/Naiv_optimalizalt_verziok2.docx>
* …

# Adatvagyon

Az elemzés adatvagyonának keletkezési lépéseit a melléklet mutatja be. A MOOC-szolgáltatók listáját pl. a mooc-list-szerveren végrehajtott mesterséges intelligencia kulcsszót érintő keresés adta meg. Ezen szolgáltatók/projektek közül véletlenszerűen 17 MOOC-szolgáltatás/szolgálató került kiválasztásra és a Google-Trends rendszerén keresztül adatokká konvertálásra - részletek <https://miau.my-x.hu/miau/273/coursera_golden-age%20(1).xlsx>

Az XLS alapján kiemelendő, hogy ugyanazon kulcsszavak lekérdezése (a korábbi hasonló adatátvételi kísérletekhez hasonlóan ismételten) nem vezetett azonos eredményre (vö. multitimeline (29) vs. (28)). Az idősorok vizuális élménye (vö. OpenSAP az emberi szem számára nagyon hasonló, de ez nem jelenti azt, hogy a robotok mintázatkeresése számára is ez igaz kellene, hogy legyen.

A vizsgált kulcsszavakat (X1, …, X16 ill. Y= Coursera)a melléklet tartalmazza.

# Elemzések

Az aranykor-elemzéshez, vagyis, amikor azt keressük, hogy valamilyen keretfeltételek függvényében adott idősorra vonatkozóan mikor alakulnak ki (tartós) norma alatti és norma feletti állapotok, vagyis, mikor lehet jó és rossz időszakokról beszélni (ill. vajon az adott jelenség idősoros történetében vannak-e az elemzéstől függetlenül is létező feljegyzések arra vonatkozóan, milyen okok miatt alakult jól és/vagy rosszul adott jelenség adott időszaka), minden adat adott (vö. Google Trends, mint potenciális big data forrás). Az idősoros adatokat az aranykor-elemzés érdekében nem kell egymáshoz igazítani nagyságrendileg, vagyis az OpenSAP-s kapcsolóra formálisan nem lett volna szükség az elsődleges elemzési cél eléréséhez. Ha az egyes MOOC-szolgáltatók arányai is fontosak lennének, akkor nem lehetne úm. megúszni, hogy az OpenSAP által garantált arányosítási erőtér információtartalmát is kiaknázzuk (vö. OpenSAP átlaga ott, ahol a Udemy a 100%-os kulcsszó és ott, ahol az OpenSAP a 100%-ot tartalmazó kulcsszó).

A nyers Google-Trends-idősorokat csak önmagukban kell rangsorolni 1 és 2007 (hónap, mint objektum) között (vö.X1, …, X16). S a Coursera Google-Trends-adatainak 1000-szeresét kell elemzés technikai okok miatt Y-nak tekinteni (vö. <https://miau.my-x.hu/myx-free/> - COCO STD: <https://miau.my-x.hu/myx-free/coco/index.html>).

# Eredmények

A nyers eredmények értelmében (XLS: modellek munkalap):

* A tények és a becslések összege jelentősen eltér egymástól (vö. 168313 egység = 0.954% tény/becslés alapon).
* A korreláció 0.96-os értékkel látszólag magas.
* A 95%-os korrekció nyomán a 207 idősáv-objektumból 199 validitása azonos csak.
* Az inverz modell és a direkt modell korrekciók utáni 137:70 invalid:valid arányt produkál, ahol invalid minden idősáv becslése, ahol az inverz inputok utáni outputok nem inverzió a direkt inputokhoz tartozó direkt outputoknak (vö. függvény-szimmetria-elvárás, mint a mesterséges intelligencia beépített önellenőrzési rétege).

A Google-Trends nyers adatok 1000000-szoros nézete, mint Y azonos eredményekre vezet, vagyis a nagyságrendi mozgástér-növekmény nem generál minőségi átcsapásokat.

A nyers Google-Trends-adatok 1000-szerese + egy 10000 pontos eltolás eredménye már minőségi változásokra mutat rá (vö. XLS – parallel models-munkalap).

A nyers Google-Trends-adatok 1000-szerese + egy 100000 pontos eltolás eredménye további minőségi változásokra mutat rá (vö. XLS – parallel models2-munkalap):

* Korreláció nő: 0.98
* Tényösszeg vs. becslésösszeg eltérés csökken: 0.999%
* A 207-ből 206 következmény azonos a validitás kapcsán.
* A 207 esetből azonban korrekciókkal 67, anélkül 68 idősáv-becslés valid.

Az 1. ábra az ún. alapmodell karakterisztikáját mutatja, míg a 2. ábra az 2. parallel modell jellegzetességeit engedi értelmezni:

1. Ábra: Nyers Y \* 1000 (forrás: saját ábrázolás) – ahol a bal oldali és a jobb oldali Y tengelyek a tényleges Google-Trends érdeklődések mértékétől való becslés-eltérést jelzik a tényt mindenkor 100%-ként értelmezve
2. Ábra: Nyers Y \* 1000 + 100000 (forrás: saját ábrázolás) – ahol a bal oldali és a jobb oldali Y tengelyek a tényleges Google-Trends érdeklődések mértékétől való becslés-eltérést jelzik a tényt mindenkor 100%-ként értelmezve

Mint az 1-2. ábra összevetése alapján látható:

* A modellpár felső tagja a durván az idősor feléig ható pangás utáni (felszálló) időszakot jelzi zömmel invalidnak (vö. kék és piros mozgó-átlagok viszonya, ahol a piros görbe ott nulla, ahol nincs valid tény vs. becslés eltérés).
* Az alsó ábra (pontosabb modell) esete ennek ellenkezője – egyébként az elvárásoknak megfelelően – hiszen a felszálló ág előtti pangás antagonizmus-kockázata nagyobb, mint a karakteres felszállóág esetében, ugyanis a quasi azonos Y-okat sorozatban produkáló pangó ág inputoldala tetszőlegesen impulzív lehet, ami ritkán illik, hogy következmény-azonosságot eredményezzen.
* A két ábra alapján a következő modell értelemszerűen már csak a felszálló ággal (97. idősávtól a 207. idősávig tartó szakasszal) számol:

1. Ábra: A csökkentett objektum-készletű modell (forrás: saját ábrázolás) – ahol a bal oldali és a jobb oldali Y tengelyek a tényleges Google-Trends érdeklődések mértékétől való becslés-eltérést jelzik a tényt mindenkor 100%-ként értelmezve

A 3. ábra karakterisztikája sem ideális még, de

* A tényösszeg-vs-becslésösszeg arány: 1.00022 (amely quasi azonosság kapcsán minőségi átcsapásról beszélni nincs értelme a korrekciós tényező érdemi nem-létezése okán).
* A korreláció: 0.969
* A 111 megmaradt objektumból 54 valid 57 invalid (de ebből 16 1%-on belüli eltéréssel).

# Konklúziók

A 3. ábra mögött is még mindig jelentős (50+%-os) invaliditás áll. Vagyis a Google-Trends 16+1 idősorának 207 eleme a Coursera kapcsán akár véletlenszerűnek is tekinthető.

Így klasszikus/legitim/valid aranykorról csak 2019-re vonatkozóan lehet beszélni, mert a 2016 végi fellángolás és a 2021-es fellángolás csak kérészéletű…

A Coursera mellékletben látható Google-Trends-karakterisztikája két nagy hullámot mutat, melyek a többi MOOC-szolgáltatóval (versenytárssal) összevetve lényegében érdemi aranykort nem engednek felismerni a robotszem számára.

Sokkal értékesebbé vált tehát az a melléktermék, mely rámutat arra, milyen invaliditás-arányt okozhat az antagonizmusokkal terhes pangó időszak, ill. a Google-Trends-adatok első és többedik lekérdezése kapcsán fennálló közelítő „mérés” véletlenszerűségre gyakorolt hatása…

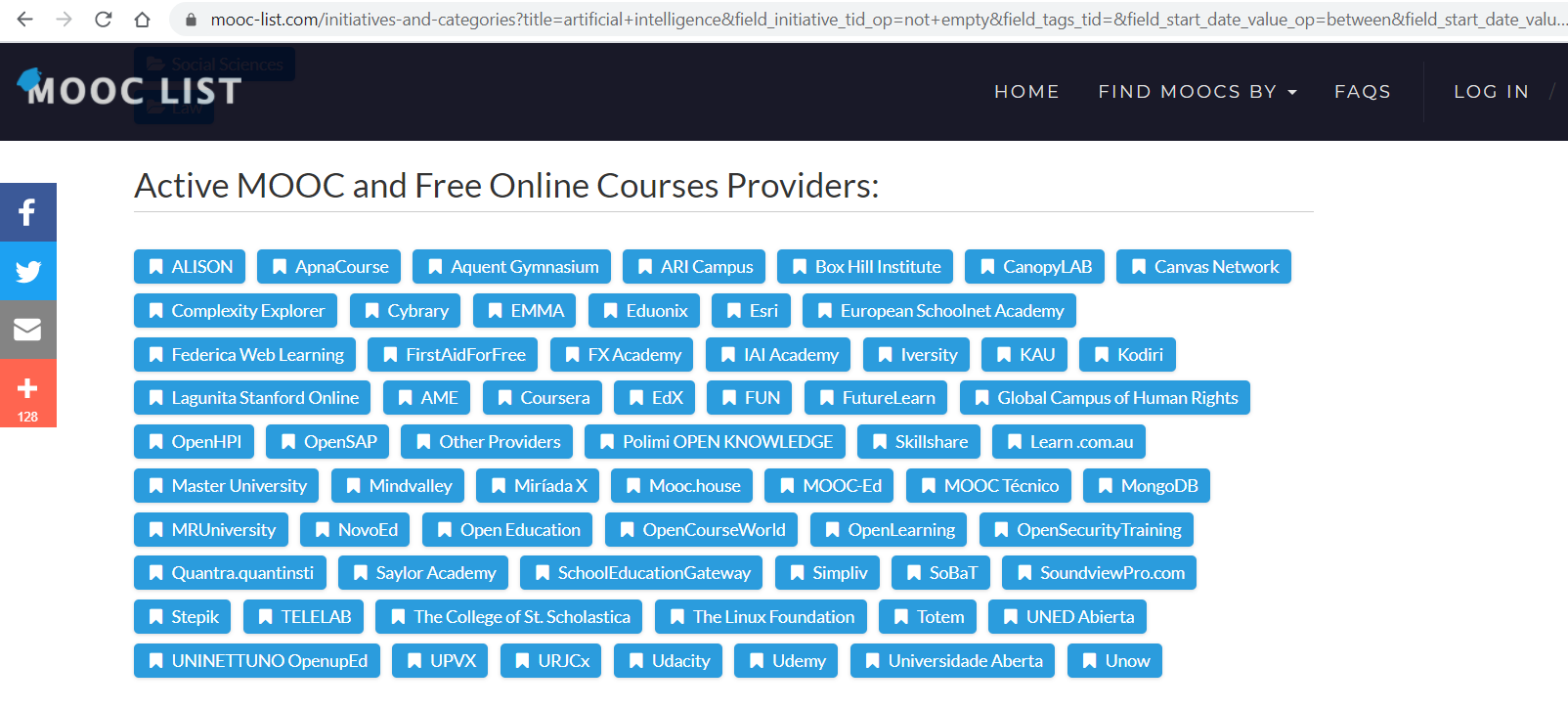
Nem mellesleg az invaliditás oka vélelmezhetően maga a Google-Trends adatbázis, mely egyrészt minden egyes ismételt lekérdezéssel újabb és újabb számmisztikákat adhat át. Másrészt: az emberi tranzakció szintű érdeklődések témakörökhöz, kulcsszavakhoz kötése is lehet önkényes (vö. igaz-e, hogy kisebb lesz az invaliditási arány, ha csak a keresett kifejezést vagy témakört/cégek választunk – ill. éppen, hogy nagyobb, mert a keresett kifejezés mögött bármi megbújhat = nő a véletlenszámgenerálási jelleg). S végül maga az emberi érdeklődés, mint olyan, még a nagy számok törvényszerűsége esetén sem tűnik feltétlenül a véletlentől jelentősen különbözni akarónak?!

# Források

…szövegközben…

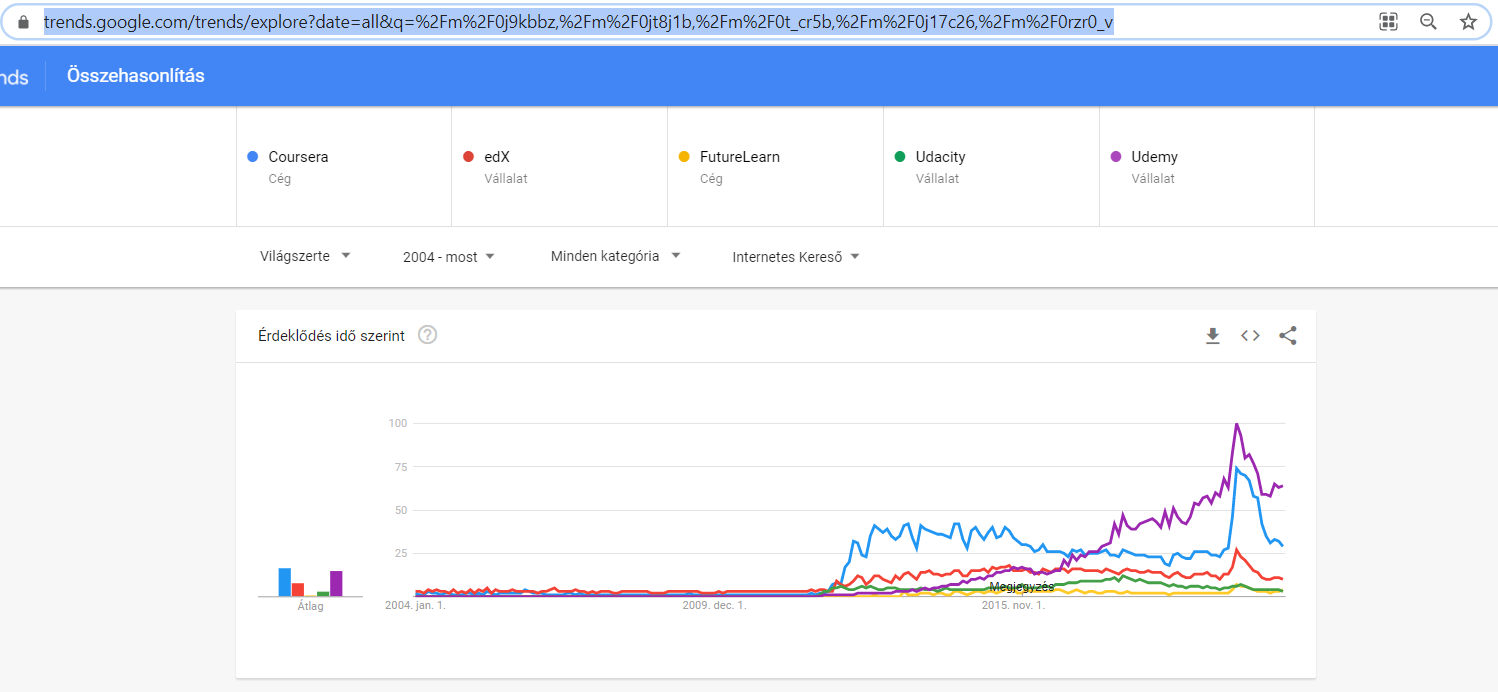
# Mellékletek

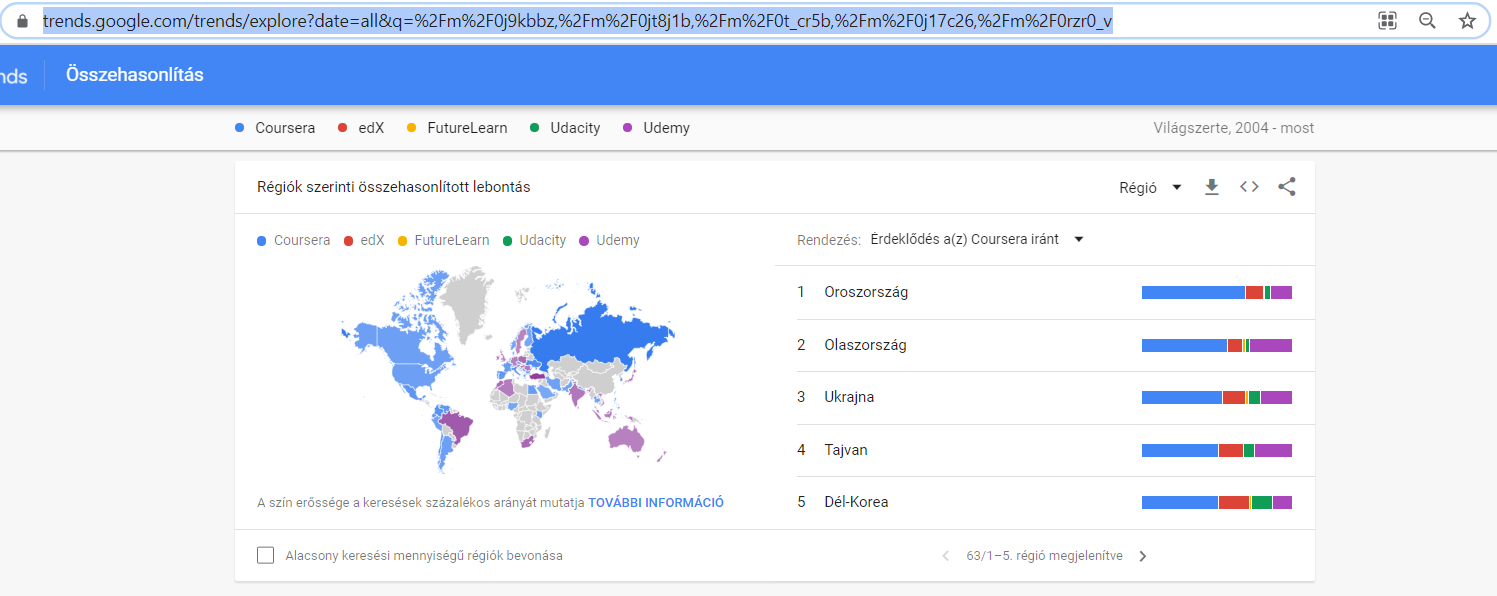
<https://www.mooc-list.com/initiatives-and-categories?title=artificial+intelligence&field_initiative_tid_op=not+empty&field_tags_tid=&field_start_date_value_op=between&field_start_date_value%5Bvalue%5D%5Bdate%5D=&field_start_date_value%5Bmin%5D%5Bdate%5D=&field_start_date_value%5Bmax%5D%5Bdate%5D=&sort_by=field_start_date_value&sort_order=DESC>



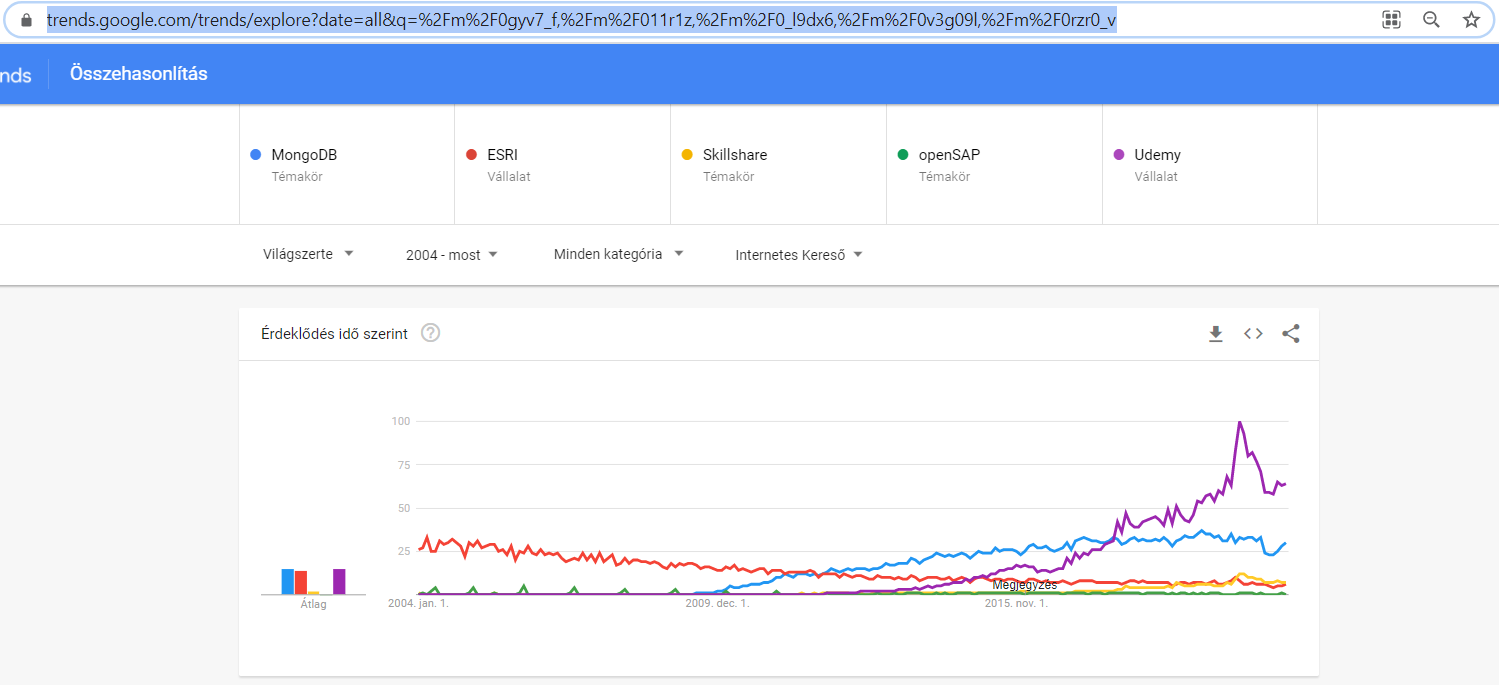
A következő ábrákon minden Y tengely %-ként értelmezendő – s ezt a Google-nak magának is fel kellett volna tüntetnie, mert az Y tengely jelentése a vizsgált univerzumban legnagyobb érdeklődés=100(%)

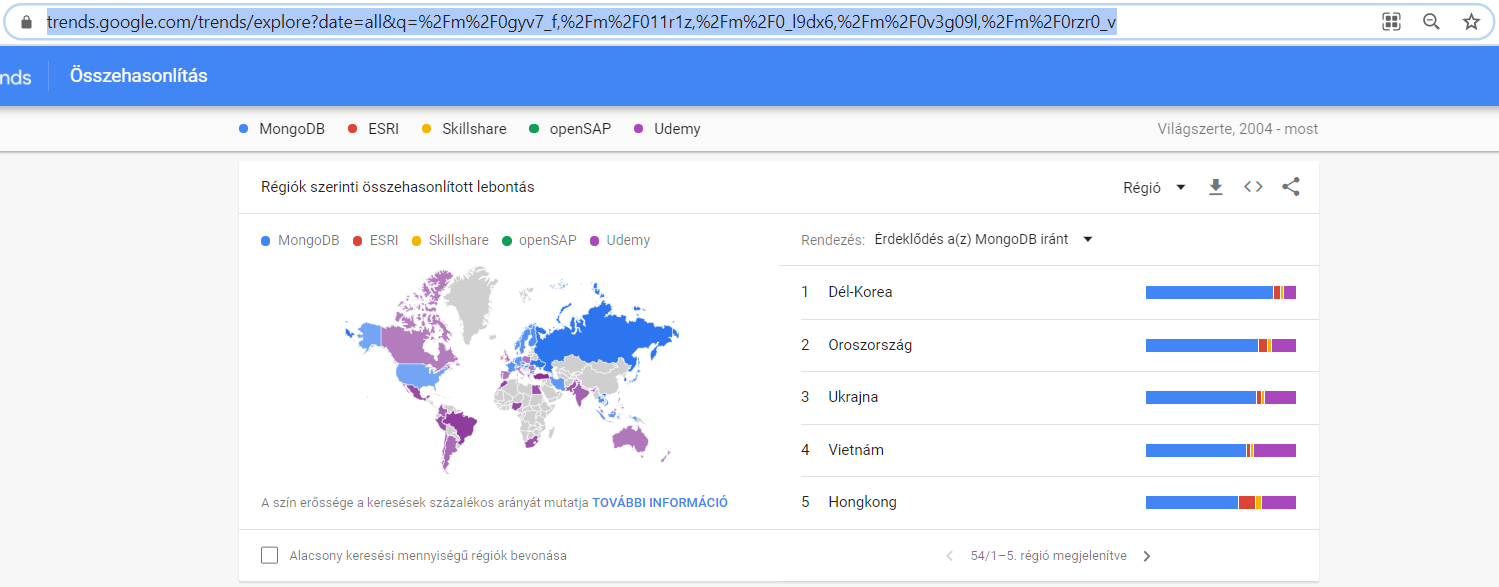
<https://trends.google.com/trends/explore?date=all&q=%2Fm%2F0j9kbbz,%2Fm%2F0jt8j1b,%2Fm%2F0t_cr5b,%2Fm%2F0j17c26,%2Fm%2F0rzr0_v>



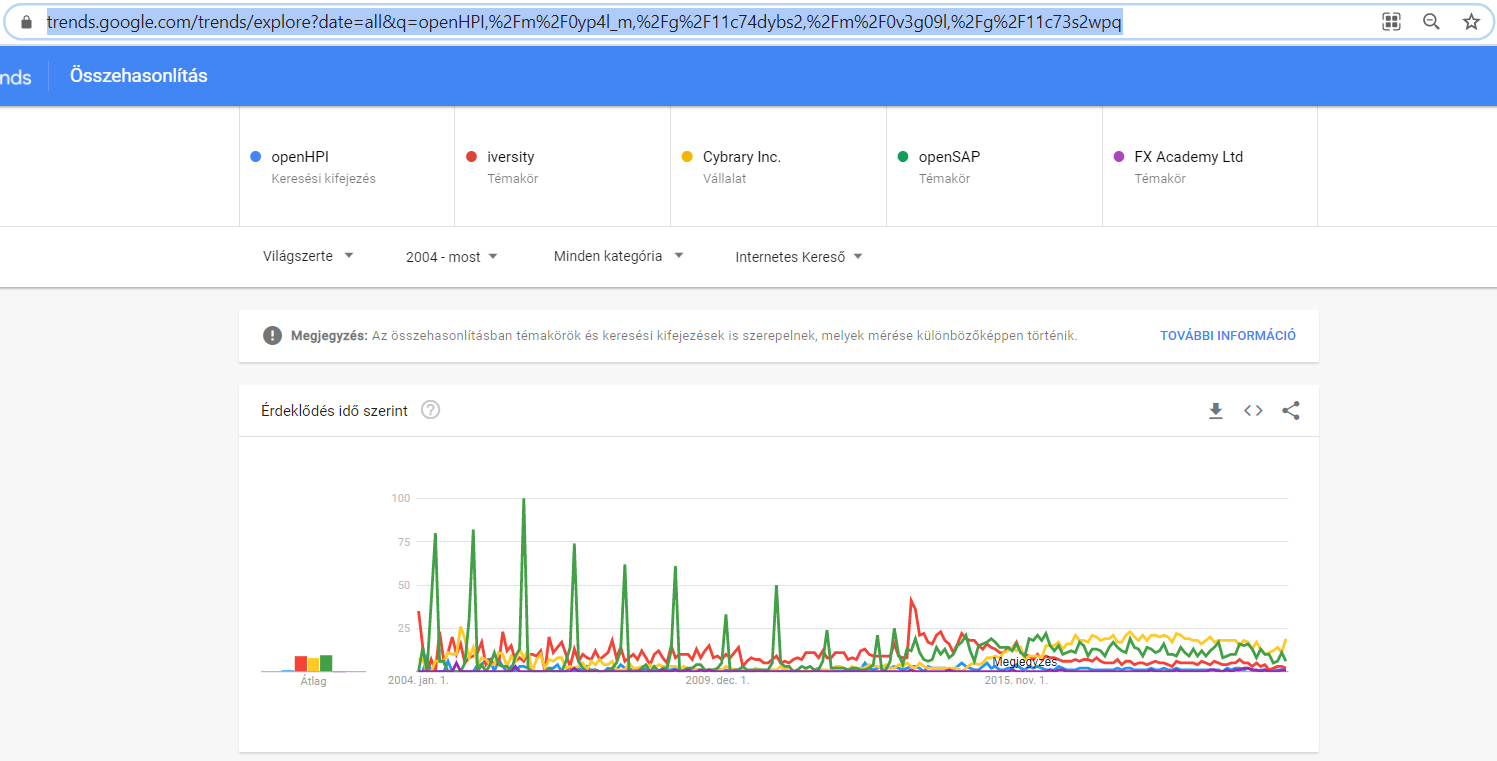


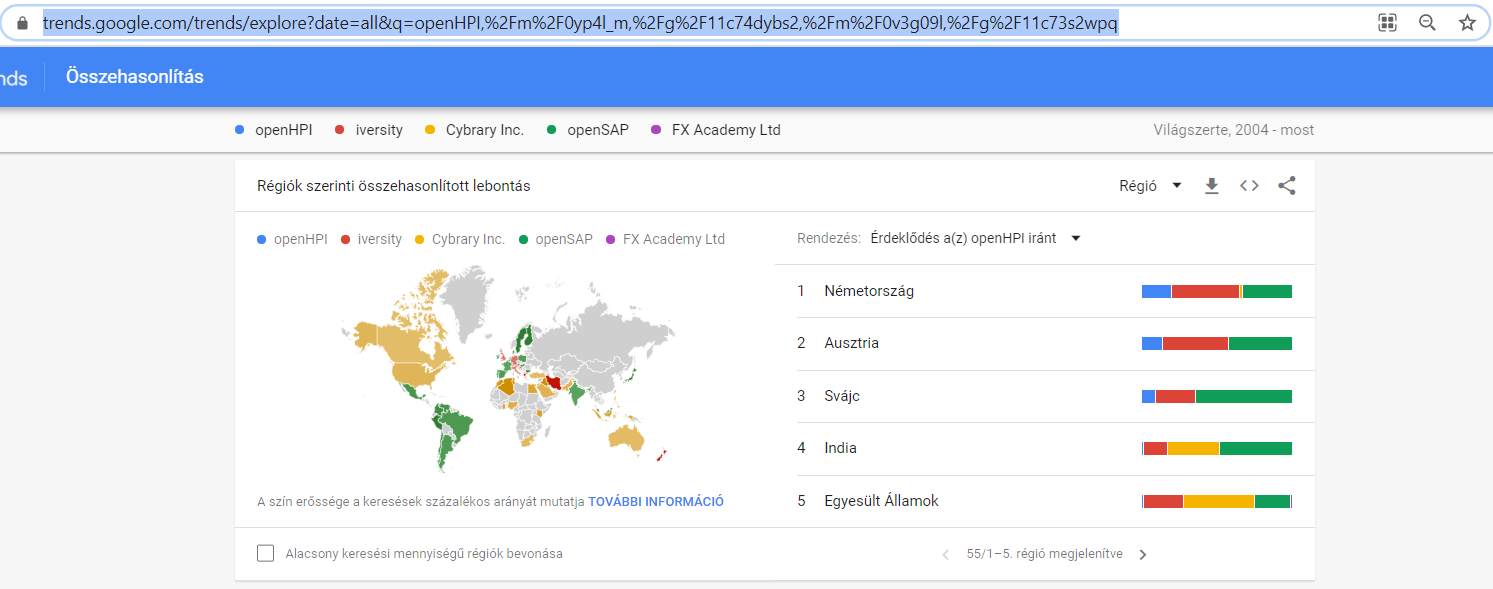
<https://trends.google.com/trends/explore?date=all&q=%2Fm%2F0gyv7_f,%2Fm%2F011r1z,%2Fm%2F0_l9dx6,%2Fm%2F0v3g09l,%2Fm%2F0rzr0_v>



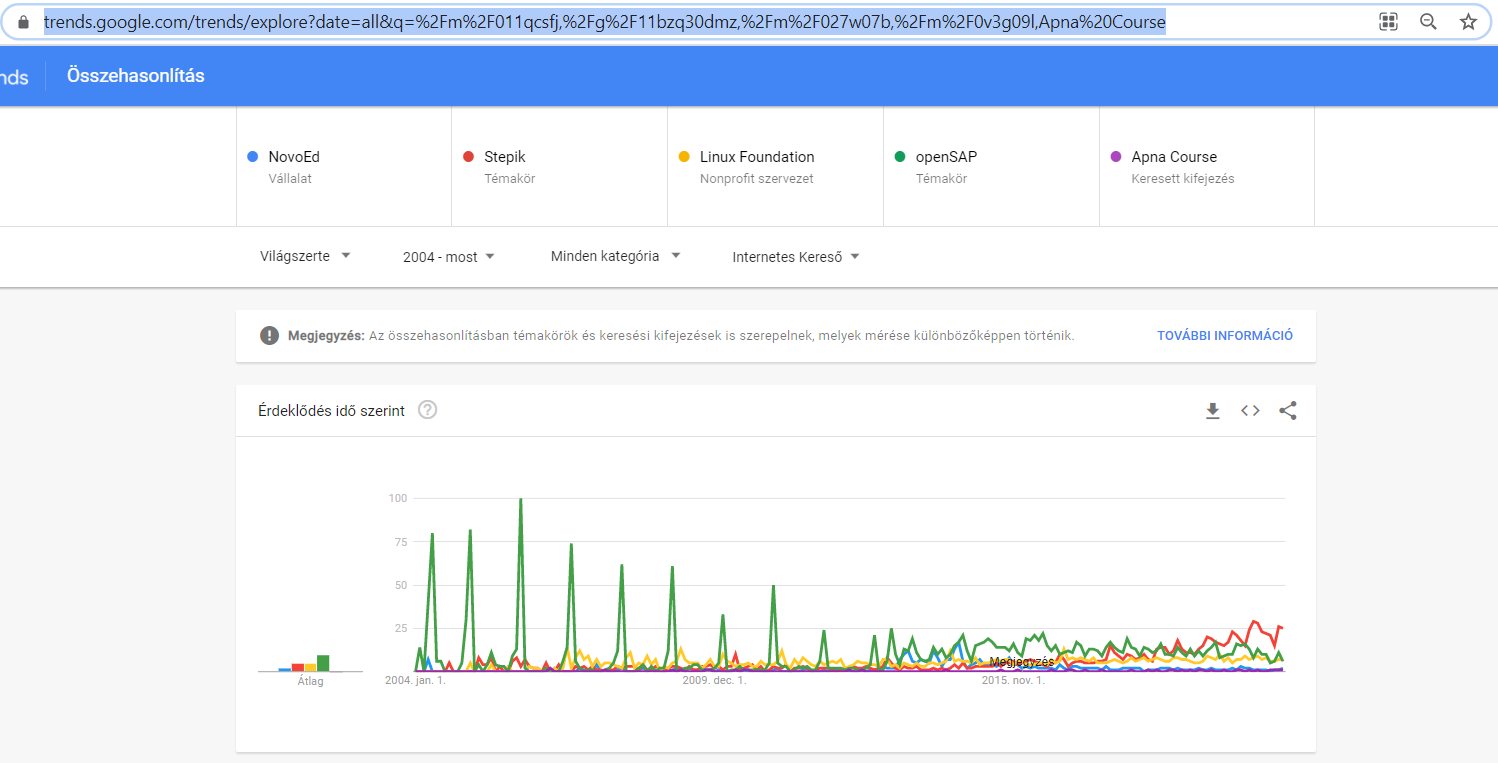


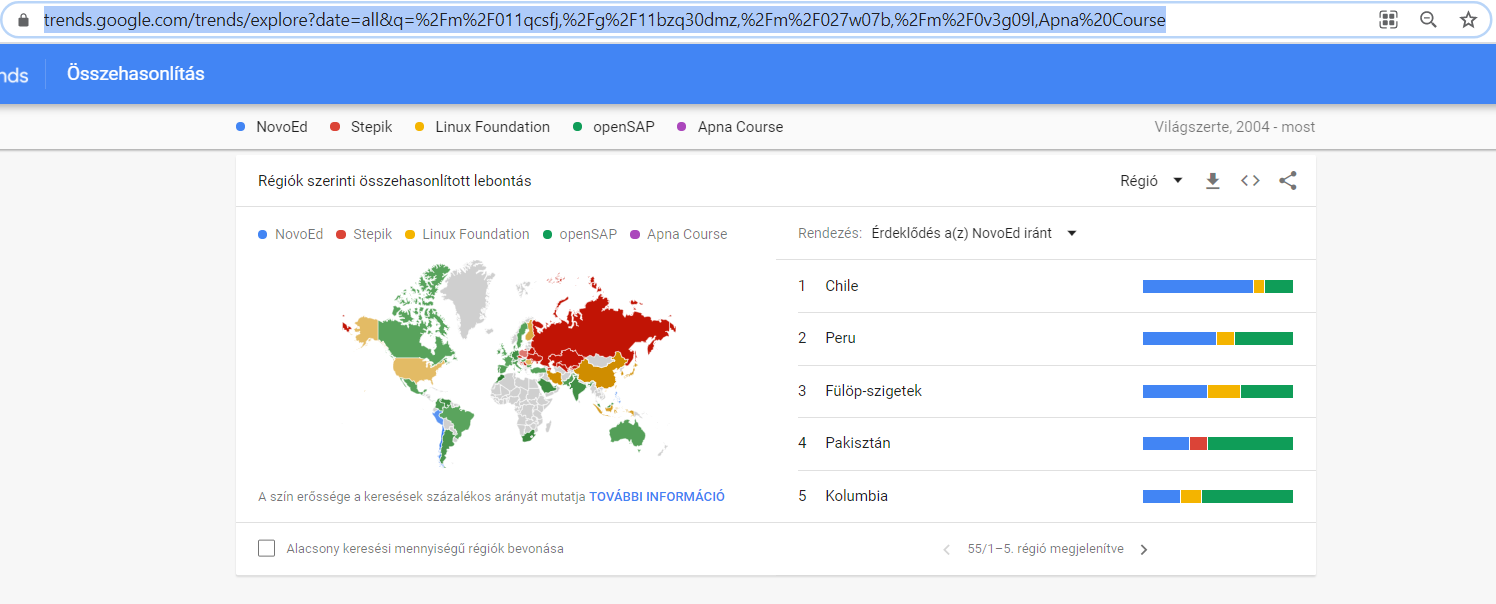
<https://trends.google.com/trends/explore?date=all&q=openHPI,%2Fm%2F0yp4l_m,%2Fg%2F11c74dybs2,%2Fm%2F0v3g09l,%2Fg%2F11c73s2wpq>





<https://trends.google.com/trends/explore?date=all&q=%2Fm%2F011qcsfj,%2Fg%2F11bzq30dmz,%2Fm%2F027w07b,%2Fm%2F0v3g09l,Apna%20Course>





<https://trends.google.com/trends/explore?date=all&q=%2Fm%2F0gyvy46>

