

Google tárgyszó találatok idősorain végzett elemzések

Farkas Zoltán, Ürmös Antal, Nemcsics Ákos

Mikroelektronikai és Technológia Intézet, Óbudai Egyetem, 1084 Budapest,
Tavaszmező utca 17, Magyarország
e-mail: farkas.zoltan@kvk.uni-obuda.hu, urmos.antal@kvk.uni-obuda.hu,
nemcsics.akos@kvk.uni-obuda.hu

Abstract: Mojzes Imre és e cikk szerzője 2009-ben publikálták a negyedik passzív áramkörü elem, a memrisztor felfedezése után az ezzel a tárgyszóval indított Google keresések találati számának idősoráról írt cikküket. Ebben a korábbi cikkben szó esett a „google dance”-ről, a keresőmotor robot ágensei által bejárt és indexelt internetes oldalakat tartalmazó adatbázis konszolidációjakor fellépő találatszám ingadozásról. Jelen cikk tárgya számos tárgyszóra történt Google keresések találati szám idősoroknak elemzése. A tárgyszavak a koronavírussal, az értéktőzsdei szlenggel és a magfizikából ismert erős nukleáris kölcsönhatást közvetítő, újonnan felfedezett odderon gluonnal kapcsolatos kifejezések. Az idősorok elemzését alapvető statisztikai eszköztárral (átlag és szórás), valamint a korrelációszámításból származtatott technikákkal végeztük el. Előzetes eredményeink szerint a koronavírussal kapcsolatos idősorokban enyhe emelkedés és csökkenés mutatható ki. Az odderon gluonnal kapcsolatos idősorokban egy tízszeres százszoros nagyságrendbe eső változás mutatható ki a „google dance”-szel összefüggésben.

Keywords: google; timeseries; coronavirus; covid-19; koronavírus.

1. Bevezetés

Az interneten fellelhető adatmennyiség gigászi feladat elé állítja az abban eligazodni kívánó felhasználót. A probléma kezelhetővé tételére még a 90-es években létrejöttek az első internetes keresőmotorok, ezek között az azóta vezető szerepre szert tett Google. Anegyedik passzív áramkörü elem, a memrisztor megvalósítását követően a kétezres években publikáltunk egy cikket a memrisztor Google találatainak idősor-elemzéséről [1]. Kutatócsoportunk az elmúlt években további Google keresőkifejezések idősorait hozta létre jellemzően napi több, rendszertelen időközben végzett mintavételekkel. A kifejezések jellemzően angol nyelvűek voltak, kivéve a „koronavírus” szót. Ezek:

A koronavírussal kapcsolatos kifejezések: „coronavirus”, „covid-19”, „koronavírus” voltak.

Az angol nyelvű tőzsdei szlengben használatos „bull market” (bika, azaz emelkedő árfolyamokkal jellemezhető piac), „bear market” (medve, azaz csökkenő árfolyamokkal jellemezhető piac) valamint az általánosan elterjedt „market crash” (tőzsdei összeomlás) kifejezés

A nemrég felfedezett, nukleáris erős kölcsönhatás közvetítésében résztvevő, femtométer nagyságrendű méretű „odderon” gluonnal kapcsolatos kifejezések: „odderon”, helyesírási hibával: „oderon” valamint az „, „odderon” gluon” és a „, „oderon” gluon” kifejezés.

A legkorábbi a koronavírussal kapcsolatos idősor, ez 2020 március végétől kezdődik. Egy lehetséges felosztás szerint a koronavírussal és az odderonnal kapcsolatos idősor az adott dolog megjelenéséhez közeli időpontban indul, a tőzsdei kifejezések idősora az adott kifejezések elterjedt használata idején veszi kezdetét. Egy másik megközelítés szerint az odderon egyértelműen tudományos kontextusban szerepel, a koronavírus „blended”, azaz kevert, tudományos és nem tudományos kontextusban szerepel, a tőzsdei fogalmak egyértelműen nem tudományos kontextusban szerepelnek (amennyiben eltekintünk a pénzügytudomány esetleges érintettségétől).

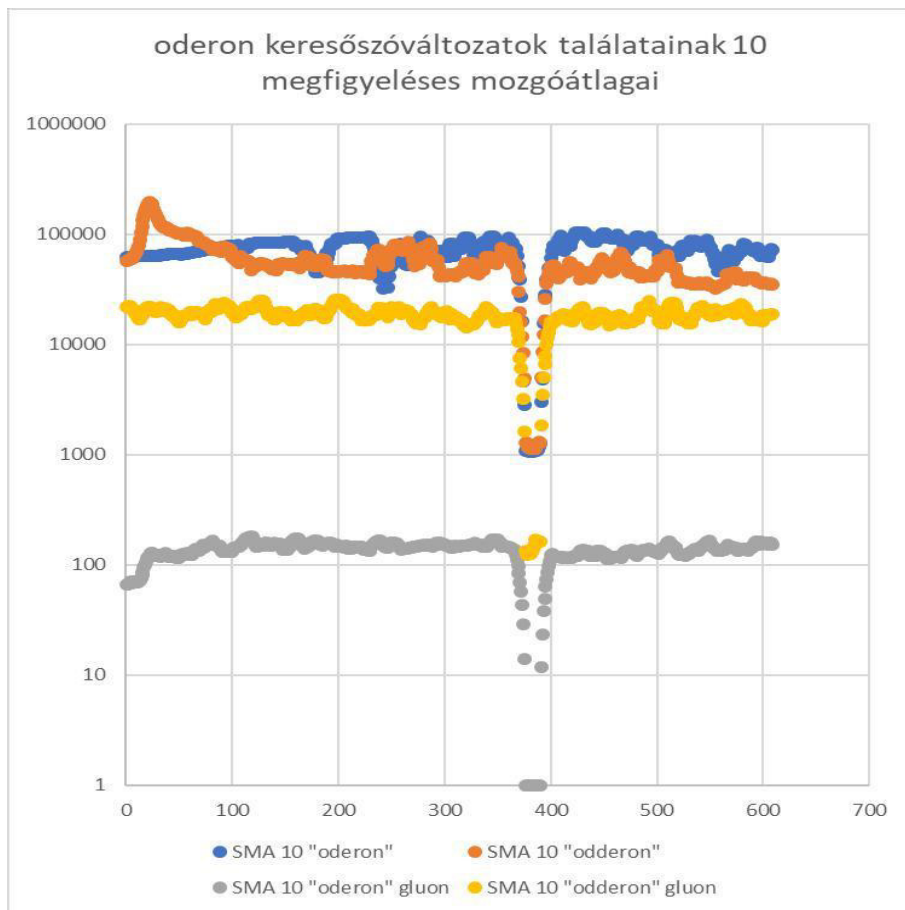
Mozgóátlag vizsgálatok

Mivel a találati számok, mint a legtöbb sztochasztikus idősor, elég nagy szórást mutatnak („ugrálnak”), az egyik legegyszerűbb technika, hogy értékelhető diagramokon jelenítsük meg azokat, a mozgóátlagok képzése volt. Az egyszerű mozgóátlag (angolul: simple moving average) egyszerűen adott megfigyeléstől visszatekintve meghatározott számú megfigyelés értékeinek átlagát jelenti [2] [3]. Ebben a fejezetben 10 megfigyelésre vonatkozó mozgóátlag értékek alakulását elemezzük.

Az „oderon” és „odderon” keresőkifejezések egymáshoz képest többször helyet cseréltek. Az „oderon” nem csupán az „odderon” szubatomi részecske hibás leírása, hanem cseh és lengyel helységnevé, különböző műszaki és kulturális jelenségek, tárgyak, személyek megjelölése. A vizsgált időszakban az „odderon” gluon és az „oderon” gluon keresőkifejezésekre adott találatok száma jelentősen eltért, az előbbire jól kivehetően két decimális nagyságrenddel több találat adódott. Nagyjából a 250. mozgóátlag értékig az „odderon” és az „odderon” keresőkifejezések találati számainak tíz megfigyelésre vonatkozó mozgóátlaga viszonylag, lassan, trendszerűen változott, ettől kezdve a mozgóátlagok diagramjain erőteljesebb rövidtávú ingadozás megfigyelhető.

Szembeötlő, hogy a Google mind a négy keresőkifejezésre lecsökkent találatsszámot adott, nagyjából a 375. mozgóátlag környékén. A lecsökkent találati számok 2021. aug. 08-ától 2021. aug. 12-éig álltak fenn. A szóbanforgó

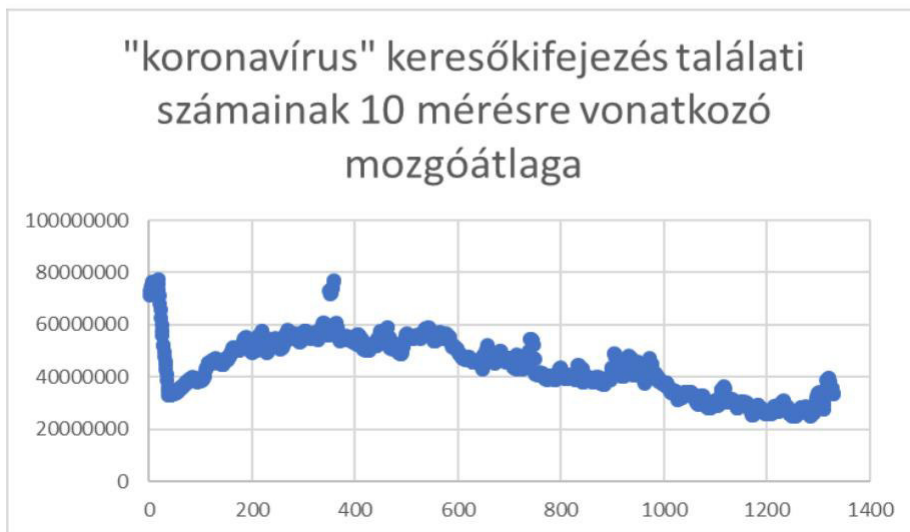
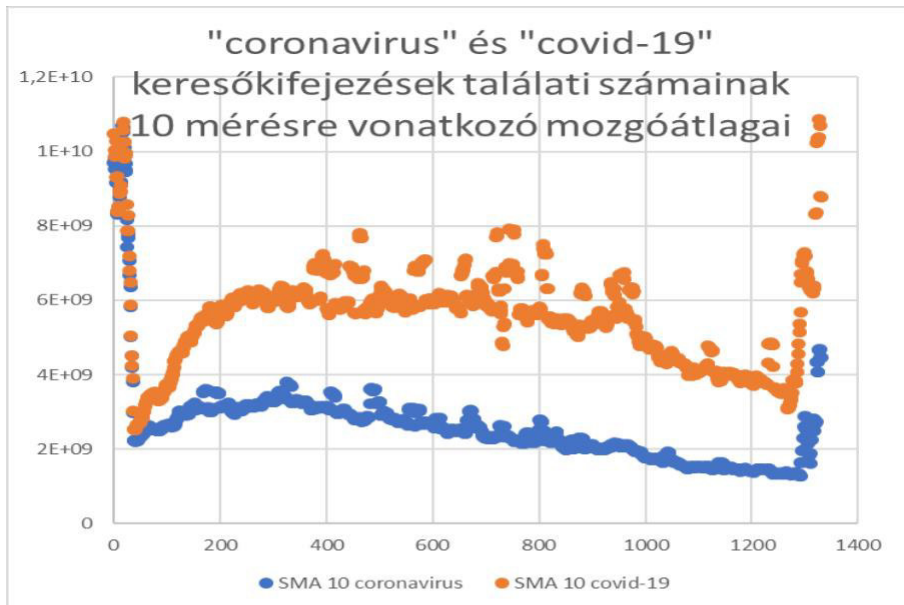
keresőkifejezésekre nézve szinkronizált csökkenés és emelkedés műszaki oka számunkra ismeretlen.



1. ábra

Függőleges tengely: megfigyelések tíz megfigyelésre vonatkozó mozgóátlaga, vízszintes tengely: mozgóátlag sorszáma

Ami a koronavírus különböző nyelvi mutációira vonatkozó keresőkérdést illeti, a találati számok 10 mérést figyelembe vevő mozgóátlagainak alakulását a 2. a., és b., ábrán láthatjuk.

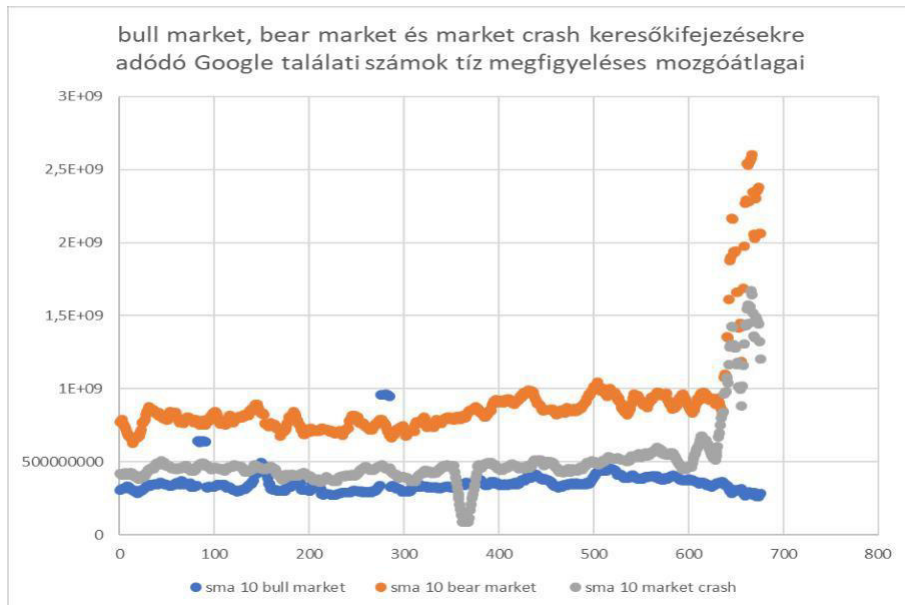


2 a., és b., ábra

Függőleges tengely: megfigyelések tíz megfigyelésre vonatkozó mozgóátlaga, vízszintes tengely: mozgóátlag sorszáma

A 2. a., és b., ábrákon egyértelmű hosszótávú csökkenés látható, amelyet az idegennyelvű keresőkifejezések találati számainak megugrása követ 2021. november elejétől. A találatok mozgóátlagai megugranak az adatok rögzítésének

kezdeten (ne feledjük, hogy ez az időszak 2020 március legvégére és április elejére esik, amikor a globális figyelem ráirányult a pandémiára), majd egy visszaesést követően magas szinten stabilizálódnak, ahonnan a megfigyelési időszak második felében lassan csökkennek. Legtovább a „covid-19” keresőkifejezés találati számainak mozgóátlaga maradt a magas érték közelében. Naptári dátumra lefordítva a „magas” értékek a koronavírus magyar és angol formájának idősorában 2020. júniusától 2020 októberéig álltak fenn. A „covid-19” találati listáján 2021. áprilisától szembeötlő a csökkenés.



3. ábra

Függőleges tengely: megfigyelések tíz megfigyelésre vonatkozó mozgóátlaga, vízszintes tengely: mozgóátlag sorszáma

A 3. ábrán az értékpapírpiacon kapcsolatos kifejezések találati szám mozgóátlag idősorának grafikonját láthatjuk. Az áremelkedéssel szinoním „bikapiac” (bull market) találati számai tendenciájukban alacsonyabbak, mint az árcsökkenéssel szinoním „medvepiac” (bear market) találati számai. A tőzsdei összeomlást kifejező (market crash) kifejezés találati számai a két előző kategória között alakulnak. Mind a három kifejezés találati számainak mozgóátlagaiban növekedés figyelhető meg a 300. megfigyelést követően. Ez naptár szerint 2021. június 16-ának felel meg. A market crash 300. és 400. adatok közt levő visszaesése 2021. júliusára tehető. A bear market és market crash kifejezések megfigyelési időszak vége felé történő felfutása 2021. novemberében történt.

Talán nem értelmetlen a növekvő tendenciát összevetni az Amerikai Egyesült Államokbeli tőkepiac történéseivel. Az amerikai tőkepiaci forgalom három legfontosabb indexe a Standard and Poor 500 (S&P 500), a Dow Jones Industrial

Average (DJIA), valamint a NASDAQ Composite. Az S&P 500 egy átfogó nagyvállalati index, a DJIA a harminc vezető iparvállalat egy-egy részvényéből álló részvénycsomag árfolyama, a NASDAQ index pedig a vezető technológiai vállalatok árfolyamát sűríti egyetlen számmá.

A június közepe az S&P 500 tekintetében egy többhónapos árfolyamnövekedési periódus közepére esik, a DJIA erős növekedés után oldalazó mozgásba kezdett, a NASDAQ index pedig éppen fordítva, egy hosszabb oldalazás után indult növekedésnek ebben az időszakban.

Összefoglalás

A vizsgált tíz idősorban érdekes anomáliákat találtunk. Erre az egyik példa az odderonnal kapcsolatos kulcsszavak találati számainak egyidejű visszaesése. A másik példa, hogy a tízből öt idősor markánsan felugrott 2021. novemberében. Ez megelőzte a koronavírus omikron változatának megjelenését.

Irodalomjegyzék

- [1]. I.Mojzes, Z. B. Farkas: The speed of dissemination of information about the realization of the fourth passive electronic circuit element measured by Google hits. SCIENTOMETRICS 81: 3 pp. 699-702., 4 p. (2009)
- [2]. <https://www.investopedia.com/ask/answers/071414/whats-difference-between-moving-average-and-weighted-moving-average.asp> (as seen on 08th december 2021 9:34 UTC+1)
- [3]. https://school.stockcharts.com/doku.php?id=technical_indicators:moving_averages (as seen on 08th december 2021 9:36 UTC+1)