

AZ ADATKOR HAJNALÁN

Syi

i@syi.hu

DOI: 10.20520/JEL-KEP.2018.1.21

Absztrakt

Tanulmányomban az adatrobbanás jelenségén belül vizsgálom meg azt, hogy a régi és új korokban, a különböző tudásterületeken, a különböző információs tartományokban milyen információmennyiségekről beszélhetünk, milyen tendenciákat figyelhetünk meg, hogyan viszonyul az emberiség által a predigitális korban felhalmozott tudás mennyisége a kortárs digitális információtermelés nagyságrendjéhez képest. A cikk elején felvázolok egy információtipológiát, majd adatokat keresek az egyes információs tartományok méretére vonatkozóan. A számokból kiderül, hogy az adatrobbanás nem a kanonizált és hagyományos tudásfelhalmozási területeken jön létre, hanem sokkal inkább a laikus tartalmak tömeges keletkezésében, az információfogyasztás mérhetővé tételében, valamint az emberek, gépek, szenzorok által gyűjtött, épített adatbázisok elterjedésében érhető tetten.

Kulcsszavak

metaadat, kulturális kánon, algoritmus

AT THE DAWN OF THE DATA AGE

Syi

Abstract

In this article, I examine the phenomenon of the data explosion in terms of how much information is available in the old and new ages, in the different areas of knowledge, in the different information domains, what tendencies can be observed, how much is the amount of knowledge accumulated by mankind in the pre-digital age compared to the magnitude of the contemporary digital information production. At the beginning of the article, I outline an information typology, then I present data for the size of the different information domains. The figures show that the data explosion is not fed by the digitization of canonical, traditional old media world, but rather it consists of the mass generation of lay content, the extra large traffic of information consumption, and the world-wide spread of databases collected by people, machines, and sensors.

Keywords

data explosion, information typology, quantity of information, content, user generated content metadata, cultural canon, algorithm

AZ ADATKOR HAJNALÁN

Syi

Az információs társadalom vagy tudástársadalom fogalma azt a minőséget ragadja meg, hogy az információ, a tudás kiemelt jelentőségűvé válik a társadalmak szervezése során. Természetesen az információ, a tudás mindig kiemelt jelentőségű volt, így a különbség inkább az információ, a tudás minőségében és mennyiségében keresendő. A tisztázás érdekében jó lenne tudni azt is, mit jelent a tudás vagy az információ fogalma, milyen információ- és tudástípusokról beszélhetünk, hogy állapíthatjuk meg ezek mennyiségét, minőségét. Ezen tanulmány fókuszában az adatrobbanás jelensége áll, ezért a tudást és az információt összemossa kezelem, és mélyebb elemzés helyett csak néhány szempontot vetek fel, amelyek mentén kategorizálni lehet e fogalmakat.

Kiindulásként álljon itt egy erős tézis: állításom szerint az ember minden korban, minden kultúrában nagyságrendileg ugyanannyi információt fogadott be a különböző érzékszervi csatornáin keresztül, és a technikai újítások nem változtattak radikálisan a befogadott információk mennyiségén. Természetesen mindig volt, van, lesz különbség a kibocsátott/befogadott információ mennyiségében és minőségében a különböző emberek és helyzetek, a különböző korok és kultúrák esetében, de azt állítom, hogy a különbségek sokkal inkább az információ minőségében érhetők tetten, mint azok mennyiségében. Adja magát az ellenvetés, hogy az ingerdús és ingermentes végpontok között nagyon eltérő környezetek létezhetnek, amelyek radikálisan eltérő mennyiségű információ befogadását teszik lehetővé. Ez természetesen igaz. Egy nagyvárosi környezetben élő ember nyilván jóval több információval szembesül, mint a pusztában ballagó juhász, a tengeren hajózó matróz vagy a várbörtön sötétjében senyvedő rab. De ezek nem okoznak lényegi különbséget az információ mennyiségében. Azt az embert, aki egész nap színes-szagos hollywoodi filmeket néz, egységnyi idő alatt több impulzus, információ éri, mint azt, aki magányosan barangolva a táj alig változó audiovizuális ingereit fogadja be, de a tévénéző ember akkor sem tud kétszer-ötször-tízszer annyi időt eltölteni a befogadással, mint a pusztában ballagó ember, tehát a befogadott információ mennyisége nem térhet el drasztikus mértékben egymástól a különböző helyzetekben.

Az idők során nem annyira az ember által befogadott és memorizált információ mennyisége változott meg, hanem inkább a különböző médiumokon rögzített tudás mennyisége és minősége. Az emberiség tudása kétféleképpen bővíthet: az emberek által memorizált tudás, illetve a különféle hordozókon, médiumokon tárolt tudás növekedése révén. A digitális hálózat elterjedésével azonban a helyzet gyökeresen megváltozik. A digitális eszköz leegyszerűsíti (értsd: a felhasználói végpontokba helyezi) a tartalmak másolását (sokszorosítását), tehát radikálisan átalakítja az információs termék gyártásának (termelésének) logikáját, míg a hálózati

hozzáférés leegyszerűsíti (értsd: a felhasználói végpontokba helyezi) a tartalmak iránti fogyasztási igény kielégítését, tehát radikálisan átalakítja a terjesztés (kereskedelem) logikáját is.¹

A médiumra rögzített tudás nagyságát és jelentőségét tovább növeli az a tény, hogy már a gép is létrehozhat új információt, nem csak az ember. Ha szenzorokat helyezünk el a tér különböző pontjain, amelyek ilyen-olyan adatokat gyűjtenek, akkor elképesztő mennyiségű adathalmazokat kapunk, amelyek csak azért „látszanak”, mert rögzítjük őket. Vegyük a hőmérséklet mérésének példáját. A mérőpontokon gyűjtött hőmérsékleti adatokat rögzítik és beküldik az adatelemző központokba. Az ember azonban mindig is képes volt „mérni”, érzékelni a hőmérsékletet, rendelkezett valamilyen tudással arról, hogy a tartózkodási helyén milyen meleg van, csak ezt az adatot nem rögzítette. A digitális fényképezés elterjedésével bármikor lehet fényképet készíteni, majd azt egy közösségi felületre feltölteni, így az analóg világban készült fényképek nagyságrendjéhez képest nyilván adatrobbanásról beszélhetünk, viszont a befogadó ember oldaláról nézve megint igaz a fent megfogalmazott tézis: minden korban körülbelül ugyanannyi vizuális élményt vagyunk képesek befogadni, csak a vizuális élmények kiváltói, hordozói változtak a múltban, illetve a jelenben és a jövőben.

A digitalizálás annyiban van hatással az információ és az ember kapcsolatára, hogy sokkal könnyebb információt rögzíteni és sokszorosítani (terjeszteni), de ez nem elsősorban az egyes ember által befogadható információ mennyiségére, hanem annak minőségére vonatkozik. Az az új lehetőség, hogy a hálózaton keresztül hozzáférhetővé válik az emberiség által eddig felhalmozott tudás, a befogadás oldaláról tekintve megint csak nem mennyiségi, hanem minőségi előrelépést jelent.²

A továbblépéshez az információ fogalmának tipizálására van szükség. Az információ többféle típusáról beszélhetünk aszerint, hogy milyen érzékszervünkön keresztül fogadjuk be, illetve nyelvi vagy nem-nyelvi kommunikációról van-e szó. Ha csak a vizuális és auditív csatornákon keresztül érkező információkra szűkítjük a vizsgálódás fókuszát, akkor az alábbi típusokat kapjuk (Syi 2007, 2017a).

	auditív jel	vizuális jel
nem-nyelvi jel	zene, zörej	mozgóképek, állókép
nyelvi jel	beszéd	írás, jelnyelvelés

További tipizálási lehetőséget jelent, hogy az információ rögzített-e, vagy sem. Egy nyelvi üzenetet kifejezhetünk jelnyelvi mutogatással és elmondhatjuk szóban is. Ekkor az üzenet nem marad fenn az időben, azt csak a kibocsátásával egy időben fogadhatják be mások. Ha viszont rögzítjük a tartalmat (leírjuk papírra, felvesszünk diktafonnal, kamerával), akkor nem vagyunk a kibocsátás pillanatához kötve, időben később is hozzá lehet férni a tartalomhoz. Mind a négy fenti helyzetben (a zene, a kép, a beszéd vagy az írás esetében) beszélhetünk rögzített és nem rögzített kommunikációs formáról.

¹ Az első változás fontos következménye a szerzői jog átalakulása, a másodiké a gépi navigációs technológiák elterjedése.

² A szenzorforradalom hatására valóban olyan új információs réteg keletkezik, amelyre korábban nem volt példa. Viszont az így keletkező adatokat nem is emberek „olvassák”, hanem gépek.

	auditív		vizuális	
	nem-nyelvi	nyelvi	nem-nyelvi	nyelvi
nem-rögzített	élő zene	élő beszéd	táncszínház	jelnyelvelés
rögzített	rögzített zene	rögzített beszéd	fénykép, mozifilm	írás

Végül a nyelvi információ fogalmát érdemes még egy szempont szerint alábontani, hogy megragadható legyen a digitális korszak talán legfontosabb vonása. Ferdinand de Saussure nyomán a nyelvi megnyilatkozásainkat két dimenzió mentén értelmezhetjük (Saussure 1998). A szintagmatikus tengely mentén az akusztikus és/vagy vizuális térben jelenlevő, látható, hallható, szintaktikailag jól-formált mondatok interpretációját végezhetjük el, míg az asszociatív/paradigmatikus dimenzióban a kommunikatív térben nem jelenlevő (az akusztikus térben nem hallható, a vizuális térben nem látható), de a fejünkben meglevő, szemantikailag rendezett (vagy rendezhető) tudást mozgósíthatjuk. Másként fogalmazva: a világról szóló tudásunkat nyelvi mintázatok, sémák alapján általánosíthatjuk, amelynek kifejezésére nem alkalmas a narratív logika, a mondatok lineáris, szekvenciális sorba rendezése, amely a nyelv szintagmatikus dimenzióját jellemi. Ehhez egy sajátos grafonyelvi technikára van szükség (Goody 1998), amely csak írásban lehetséges, amelynek ősképe a táblázat, a modern formája pedig az adatbázis. Ez a technika, formai értelemben, biztosítja a mondatainkba foglalt tudás két- vagy többdimenziós elrendezését azáltal, hogy egyszerre teszi lehetővé a nyelvi üzenetek sorok és oszlopok szerinti olvasását, tartalmi szempontból pedig lehetőséget ad a világról való egyedi tudásunk általánosítására az osztályozási műveleteken keresztül (Syi2017b). Az adatbázis is rögzített, nyelvi információ, de strukturált szöveg, ami annyit jelent, hogy egyfelől a mondatok komponensei – osztályozási szempontok szerint – szegmentálva, tipizálva vannak, másfelől minden mondat új sorban (de legalábbis a többi mondattól egyértelműen elválasztva) van rögzítve. A beszéd – ebben az értelemben – nem strukturált, és az írás is csak akkor és úgy válik azzá, ha rendezett formába alakítjuk át a szöveget. Ehhez a szöveg önmagában nem elegendő, további információra van szükség, amely a szöveg belső tagolását, elrendezését adja meg. Ezt nevezhetjük sémának, ami nélkül a táblázat, adatbázis értelmezhetetlen lenne. Azt már említettem, hogy a szemantikailag jól-strukturált mondatok egyik előnye az, hogy megadják az általánosítás – és azon keresztül az új, érvényes tudás megszerzésének – lehetőségét. A strukturálva osztályozás másik előnye az, hogy az ilyen mondatok semantikusságukból fakadó „egyszerűségük” révén könnyen kezelhetők, ami – idővel – megteremti annak lehetőségét, hogy szenzorokkal gyűjtsük, gépekkel növeljük a világról szóló, sémákba írható egyedi állítások számosságát, ezáltal növeljük a tudásunk mennyiségi, statisztikai értelemben vett megbízhatóságát. Van tehát az egyszerű, strukturálatlan, rögzített, nyelvi információ, amit szövegnek szokás nevezni, és van a rendezett, strukturált szöveg, ami az adatbázis alapját képezi. Ezen kiegészítéssel immár rendelkezésünkre áll egy információtipológia, amelynek segítségével pontosabb állításokat fogalmazhatunk meg.

	auditív		vizuális	
	nem-nyelvi	nyelvi	nem-nyelvi	nyelvi
nem-rögzített	élő zene	élő beszéd	táncszínház	jelnyelvelés
rögzített	rögzített zene	rögzített beszéd	fénykép, mozifilm	írás
				sima szöveg

Az információ tipizálási lehetőségeinek áttekintése után nézzük meg, mit tudunk mondani az egyes információtípusokat jellemző méretekről.

A predigitális kornak két fontos jellegzetessége van. Egyfelől ekkor még a nem-rögzített információ a domináns (mennyiségét, használati idejét tekintve), mert a rögzítés, sokszorosítás és terjesztés technikái csak kevesek számára állnak rendelkezésre. A hétköznapi beszélgetéseken keresztül áramló nyelvi információhoz képest sokáig elhanyagolható mennyiségű az írásban rögzített információ. Másfelől viszont már a predigitális korra is az a jellemző, hogy a médiumokon tárolt információ méretéhez képest jóval nagyobb a médiafogyasztás során befogadott információ mennyisége. A hordozókra egyszeri alkalommal rögzített információ mennyisége azonban mindig is elhanyagolható volt ahhoz képest, hogy rögzített zenét, képeket, képfolyamokat hányszor hallgatták-nézték meg a zene-, kép- és filmrajongók. Épp ez a rögzítés lényege: az egyszeri kommunikációs aktus (a zeneszám, beszéd, álló- vagy mozgókép rögzítése) eredményét később elvileg végtelenszer lehet befogadni, főleg akkor, ha az információhordozót sokszorosítjuk és a térben terjesztjük, ezáltal növelve a potenciális hozzáférők számát. A tudás megőrzésében, fenntartásában, felhalmozásában, átadásában az írásbeliség, a hangrögzítés, a képrögzítés (a fénykép és a film) technikái hozták a minőségi változást, de a mennyiségeket tekintve mindig is hatalmas aránytalanság állt fenn a rögzített információ kibocsátási és befogadási oldalát jellemző nagyságrendek között.

Ahhoz, hogy haladjunk a mérhetőség felé, további megkülönböztetésekre van szükségünk. A rögzítés eredménye mindig valamilyen *dokumentum*, amely a *hordozó* mint közvetítő (médium), és a rögzített információ mint *tartalom* együttese. Ugyanazt a tartalmat rögzíthetjük többfajta hordozóra (a *Star Wars* negyedik részét mozifilmre, videókazettára, DVD-lemezre, hálózati szerverekre stb.), mindig más dokumentumot kapunk. Viszont ha egyszer elkezdünk dokumentumokat létrehozni (termelni), akkor valahogyan biztosítani kell azt is, hogy a dokumentumok tartalmát újra és újra be lehessen fogadni, vagyis a dokumentumok tárolását és visszakereshetőségét is meg kell oldani, amit a dokumentumokat leíró információval biztosíthatunk. A dokumentum(on tárolt információ) megkeresését segítő információ már másodrendű fogalom, hiszen az információról hordoz információt, ezért nevezik metainformációnak vagy metaadatnak.³ A dokumentumokra rögzített tartalmak elhelyezését, megőrzését, visszakereshetőségét biztosító technikákat, megoldásokat, intézményeket legáltalánosabb szinten archívumként definiálhatjuk.

Első lépésben az információ fogalmát feloszthatjuk a *tartalom* és a tartalmat leíró *metaadat* kettősségére.⁴ Tartalomnak minősíthetünk minden emberi kommunikáció eredményét, akár rögzítjük azokat, akár nem. Tartalom lehet a kamerák vagy hangrögzítők által készített kép- vagy hangfelvétel, és a szenzorok által gyűjtött bármilyen adat is. Amikor a tartalmat rögzítjük, akkor *dokumentumot* hozunk létre (amely tehát a tartalom és hordozó egysége). A metaadat olyan információt jelent, amely segít a tartalom (dokumentum) megtalálásában az archívumon belül. A tartalom fogalma elsőrendű, a metaadat fogalma másodrendű. A metaadat fogalmát tovább bonthatjuk *leíró* és *adminisztratív* metaadatokra. A leíró jellegű metaadatok magára a dokumentumra, annak létrehozására vonatkoznak. Ide tartoznak a *tartalmi*, a *formai* és a *strukturális* metaadatok.

³ Pontosabb lenne a metainformáció helyett metaadat fogalmát használni, de e különbség kifejtésére most nincs mód.

⁴ Ezen a ponton megmutatkozik egy kis terminológiai pongyolaság: első lépésben a tartalom és az információ fogalmát egymással felcserélhetőnek tételeztem, miközben a tartalom és metaadat szembeállításával a tartalom terjedelme szűkül, hiszen nem foglalja magába a metaadatokat, holott az is információ.

A *formai metaadatokkal* írhatók le a dokumentumok létrehozásával, hordozójával, formátumával kapcsolatos információk (cím, szerző, kiadó, kiadás helye, éve, hordozó típusa, mérete stb.). Ezek azonban még semmit sem árulnak el a dokumentumban tárolt tartalomról.

A *tartalmi leíró metaadatok* a tartalom leírásában segítenek, ezek lehetnek rövid írásos összefoglalók, mozielőzetesek, hangmontázsok vagy megfelelő tartalmi címkék, tárgyszavak is.

A *strukturális metaadatok* az összetett dokumentumok esetében használatosak, amikor egy dokumentum részdokumentumokra bontható (ilyen például egy filmen belül felhangzó zenemű, egy multimédia jellegű alkotás valamely komponense vagy a legtöbb webes dokumentum, amely egyaránt tartalmaz szöveget, képet, esetleg beágyazott videót, hangállományt). A strukturális metaadatok segítségével lehet a részdokumentumok egész dokumentumon belüli helyét meghatározni.

Az *adminisztratív metaadat* a dokumentumok használatával, adminisztrálásával kapcsolatos információt hordozza. Ilyen például a raktári jelzet, amely segít megtalálni egy könyvet a könyvtárban, egy filmtekercset a filmtárban, de a könyvkölcsönzés során felvett adatok is ide tartoznak (ki, melyik könyvet, mikor kölcsönözte ki), és természetesen adminisztratív metaadat a webes felhasználói interakciókat (letöltések, meghallgatások, megtekintések, olvasások, futtatások, hozzászólások, értékelések, lájkolások, kommentek, szavazások, vásárlások stb. számát) nyomon követő *forgalmi adatok* teljes skálája is.

Korábban is megkísérelték megbecsülni az információ mennyiségét, de a digitális korszak tette igazán láthatóvá azt, hogy a különböző üzenettípusok esetében milyen mennyiségeket kell kezelünk. Azt már a digitális korszak első időszakában is lehetett tudni, hogy a szöveg, kép, hang és mozgóképek között milyen nagyságrendi különbségek vannak. Egy könyvespolcnyi könyv teljes állománya néhány megabájtot foglal, egy órányi zene 6-700 megabájtot tesz ki, ugyanennyit jelent – felbontástól függően – pár száz digitális fotó, míg egy jobb felbontású film mérete már gigabájtokban mérhető.

Méretüket tekintve a leíró metaadatok mennyisége elhanyagolható a dokumentumokban rögzített tartalmak információmennyiségéhez képest. Egy dokumentumhoz akár sok száz formai metaadat is kapcsolható, de ezek információmennyisége minimális. A tartalmi leíró és a strukturális leíró metaadatok darabszáma nagyjából azonos nagyságrendben van a dokumentumok számosságához képest, és a mennyiségük szintúgy elhanyagolható a tartalom egészéhez viszonyítva. Más a helyzet azonban az adminisztratív vagy forgalmi metaadatokkal. A digitális kor előtt nagyon munkaigényes volt az adatok gyűjtése, ezért csak a kritikus pontokon foglalkoztak vele, a digitális világban azonban a gépek minden forgalmi, használati adatot könnyen, komolyabb korlátok nélkül képesek gyűjteni, tárolni, és ez teljesen új helyzetet teremt. Amíg a leíró metaadatok valahogyan arányosak a dokumentumok számával, addig a forgalmi adatok a felhasználók, befogadók és dokumentumok számának szorzatával arányosak. Ez pedig már könnyen adatrobbanáshoz vezethet.

Adatrobbanás

Eric Schmidt, a Google vezére 2010-ben azt nyilatkozta, hogy már két naponként állítunk elő annyi információt, amennyit 2003-ig az emberiség egész történelme során megtermelt (Schmidt 2010). Schmidt a 2003-ig előállított adatmennyiséget 5 exabájtra becsülte, és a kortárs adattermelés domináns részeként a felhasználói tartalmakat jelölte meg.⁵ Az IDC egy 2010-es jelentésében (IDC 2010) a digitális forgalom, azaz a létrehozott és másolt digitális dokumentumok forgalmának nagyságát a 2010-es évre 1,2 zettabájtra becsülte, ami 6,8 exabájtot jelent

⁵ Az információ nagyságának megállapítására a következő – egymást ezerszeres szorzóval követő – mértékegységeket különböztetik el: megabájt, gigabájt, terabájt, petabájt, exabájt, zettabájt.

kétnaponként, és ez a szám nagyon közel van ahhoz, amit Eric Schmidt állított a világ digitális termelésének nagyságrendjére vonatkozóan. 2009-ben Marissa Mayer, a Google egyik vezetője tárta fel, hogy amíg 2002-ben 5 exabájt volt a Google keresők által „látott” digitális adatvagyron, addig 2009-re ez a szám 281 exabájtra emelkedett (Mayer 2009)⁶. Még ha vannak is, akik kritizálják az effajta mennyiségi becslések pontosságát (Moore 2011), az egyes időszakok közötti nagyságrendi különbség nem igazán vitatható. Az már inkább kérdéses, vajon a predigitális és digitális kor adatmennyiségei közti drámai különbségről szóló tézis is elfogadható-e.⁷

Az adatrobbanás vizsgálatának során a legfontosabb kérdés, hogy vajon mekkora lehet a predigitális korban felhalmozott tudás mennyisége, és ez hogyan viszonyul a kortárs tudás-termelés nagyságrendjeihez. Egy-két területen érdemes próbát tenni arra vonatkozóan, mennyi lehet az emberiség által felhalmozott kulturális adatvagyron mennyisége. Elsősorban a kulturális kánonok mentén működő közgyűjteményekben tárolt dokumentumok számosságát kell megbecsülni.

Mennyi lehet az eddig világszerte kiadott könyvek száma? A Google Books projekt célja az emberiség teljes könyvállományának digitalizálása. A projekt keretében 2010-ben 130 millióra becsülték az összes könyv számát (az ISBN-számok és a legnagyobb nemzetközi katalógusok alapján kalkuláltak). Az OSZK katalógusa nagyjából 2 millió rekordból áll, amiben persze nem csak magyar művek szerepelnek. Ha a digitális könyvek átlagos méretét egy gigabájtnak vesszük (ami jelentős felső becslés), akkor a teljes digitális könyvállomány mérete 130 petabájt lehet.

A fényképek mennyiségére Jeremy Good végzett becslést 2011-ben (Good 2011), és már akkor is döbbenetes számok jöttek ki az összehasonlításokból. Good szerint az analóg fotózás 2000-re érte el a csúcst. Ebben az évben 85 milliárd fényképet csináltak, ám ettől kezdve rohamosan csökkenni kezdett a hagyományos gépekkel készült képek száma, míg a digitális fotóké drámai gyorsasággal emelkedett. 2011-ben 360 milliárd digitális képre becsülték a képvgayron nagyságát, miközben a Library of Congress képgyűjteményéhez képest a Facebookra feltöltött képek nagyságrendjét tízezerszeresnek vélték. Azóta ez az különbség csak nőtt. Az Instagram fotómegosztó oldalra naponta átlagosan 52 millió fényképet töltenek fel, és 2017-ig közel összesen közel 35 milliárd fotó került fel (Instagram 2017). Évről évre és becslésről becslésre eltérő nagyságokat közölnek a Facebookra képforgalmára vonatkozóan, de a számok a 2013–2017 közti időszakban már a 200-350 közti millió napi új képfeltöltésről szóltak. Nem igazán valószínű, hogy az analóg korszak fényképei jelentős arányban átkerülnek a digitális térbe, de ha a csúcst jelentő, 2000-es, 85 milliárdos analóg fényképvagyont teljes egészében digitalizálnánk, talán elérhetnénk a zettabájtos határt. Ez azonban biztos eltörlül ahhoz a mennyiséghez képest, amely a digitális fényképek világban keletkezik folyamatosan (az éves új termés már a jelenben is a zettabájtos nagyságrendben mozog).

Mennyi mozgóképes dokumentum áll az emberiség, a magyar kultúra rendelkezésére? A legnagyobb mozgóképes online adatbázisnak, az IMDb-nek 2018 elején 4,7 millió címről volt adata, ebből sorozat-epizód több, mint 3 milliót, mozifilm 467 ezret tett ki (IMDb 2018)⁸.

⁶ Mayer szerint az adatsnövekedés mértéke meghaladja a Moore-törvény növekedési arányát (Mayer 2009). A Moore-törvény szerint az integrált áramkörökön elhelyezett tranzistorok sűrűsége (ezáltal a számítógépek számítási kapacitása, gyorsasága) 1,5-2 évente megduplázódik.

⁷ A *How much information* projekt keretében két menetben kísérletet tettek arra, hogy megbecsüljék azt, hogy a különböző médiumokra rögzített információtipusokból mennyi adat keletkezhetett az idők során az emberiség történelmében (HMI 2003, Bohn & Short 2012).

⁸ Például a bollywoodi filmgyártás nagy része 2012-ben még nem szerepelt az IMDb-ben (Jameson 2012).

A magyar filmtörténeti adatbázisban 20 ezer film adatai vannak feldolgozva, amelyek között sok a filmhíradó és a dokumentumfilm, viszont nincsenek köztük tévéfilmek. Ha felső becslést akarunk adni, akkor 50 ezerre tehetjük a tévében, moziban bemutatott magyar filmek számát, amiből 2-3 ezer körülit tesznek ki a játékfilmek. Ha egy filmet átlagosan (és megint jelentősen felülbecsülve) 100 gigabájtos nagyságúnak tekintünk, akkor a világ 5 millió filmjének összemennyisége nem éri el az exabájtos nagyságrendet, a magyar filmtermés mérete a petabájtos sáv alsó részében lehet.

A hivatalos 2018-as adatok szerint minden percben 300 órányi videót töltenek fel a YouTube-ra (YouTube 2018), és legalább 5 milliárd videót néznek meg naponta a látogatók. A tízezer legnépszerűbb videó több, mint egy milliárd megtekintést generál (nyilván ezen a területen is működik a hosszúfarok hatása). A videómegosztóra feltöltött mozgófilmek számáról nincs adatunk, és azt sem tudjuk, hogy milyen lehet a professzionális és amatőr tartalmak aránya, de elég valószínű, hogy egyre elhanyagolhatóbb nagyságrendet képviselnek a kanonizált videótartalmak a felhasználói tartalomhoz képest. A folyamatos feltöltések eredményeként már jelenleg is bőven az exabájtos határ fölött lehet az éves növekmény, és nyilván gyorsan el fogja érni a zettabájtos nagyságrendet.⁹

2013-ban az Apple iTunes zeneáruház katalógusában 26 millió zenemű szerepelt (Apple 2013), a Tech Advisor felmérése alapján a népszerű zeneszolgáltatók (Spotify, Deezer, Amazon, Apple) 30-40 millió zenezámból álló készlettel rendelkeznek (Techadvisor 2018). Ezzel a dokumentumszámmal (mondjuk, egy gigabájtos átlagos nagyságot feltételezve) a petabájtos nagyságrendbe lehet belépni, de feljebb aligha. A zene világa eltér az álló- és mozgóképek világától: más megoszlást feltételezhetünk mind a laikus kontra professzionális tartalom, mind a predigitális kontra digitális korszakban keletkezett zenezámok arányai között. A zene területén kisebb lehet a különbség a predigitális és digitális nagyságrendek között egyszerűen azért, mert a digitális eszközök segítségével laikusként is tetszőleges mértékben lehet álló- és mozgóképeket előállítani és megosztani, míg zene esetében ez nincs így. Azért, mert jóval kevesebben tudnak hangszeren játszani, mint ahányan meg tudnak nyomni egy gombot egy képrögzítő eszközön.

Érdemes figyelembe vennünk még egy megkülönböztetést, amelynek segítségével az adatbázisokban tárolt információk nagyságrendjéről is képet kaphatunk. A hálózatról szóló diskurzusban már a 2000-es évek elején megjelent a mélyweb, illetve a felszíni web (deep web vs. surface web) fogalompár annak a két világnak az elkülönítésére, amelyet a keresőmotorok képesek bejárni, leindexelni és megmérni, illetve amelyet nem. Michael K. Bergman az előbbi tartományt nevezte el *felszíni webnek*, az utóbbit pedig *mélywebnek* (Bergman 2001). A keresőmotorok kétféle akadály miatt nem képesek egy weboldal tartalmát feltérképezni: vagy azért, mert az oldal jelszóval védett, vagy azért, mert adatbázisokból kiszolgált, dinamikus oldalhoz jutnak el, ahol értelmetlen egyetlen nézetben létező tartalomról beszélni, így értelmetlen az adatbázis méretét is kiszámolni. A mélyweb meghatározásában rejlő kettősség is hozzájárult ahhoz, hogy a róla szóló értekezésekbe később belekeverjék a *sötétweb* (darkweb) jelenségét, amely a nyilvánosság elől elzárt, sokszor illegális tevékenységekkel kapcsolódott össze. Fogalmilag nem szerencsés a mélyweb és a darkweb összemossa, de ennek a tanulmánynak nem ez a témája. Azt azonban nyugodtan kijelenthetjük, hogy az utóbbi mérete elhanyagolható az előbbi nagyságához képest. A mélyweb jelenség másik szegmense (a fogalom másik értelmezése) egyértelműen az adatbázisok világához köthető. Már Bergman is azt emelte ki, amit azóta mások is megerősítettek, hogy a mélyweb nagysága messze meghaladja a felszíni web méretét. Széles körben idézett az a számpár, amely a mélyweb 96%-os arányát állítja szembe a felszíni web 4%-os arányával (Lewis 2012). A mélywebben található adatbá-

⁹ Könnyen elképzelhetőnek tartom, hogy már régesrég elérte.

zisok közül példaként említhetem a Wal-Mart adattárházát, amelyben 1999-ben 1 petabájt információt tároltak, ami 2012-re felment 2,5 petabájtra (McNulty 2014), vagy a környezeti, időjárás adatokat gyűjtő amerikai ügynökség NOAA-adatbázisát, amelynek mérete a 2000-es évek elején még csak a fél petabájt környékén mozgott, 2018-ra viszont már meghaladta 20 petabájtot (NOAA 2018).¹⁰ Az ilyen adatbázisokból rengeteg van.

Eddig többnyire azokról a mennyiségekről beszéltem, amelyeket a médiumokon tárolt dokumentumok méretére vonatkoztak. Van azonban egy másik fontos kérdés is: amikor a hálózaton keresztül a felhasználók fogyasztják a médiatartalmakat, milyen nagyságrendű adatforgalmat generálnak vele. Korábban már idéztem, hogy a digitális forgalom nagyságát 2010-ben 1,2 zettabájtra becsülték (IDC 2010). A Cisco nevű cég gyártja a hálózati forgalmat menedzselő eszközök nagy részét, és az eszközökön mért adatokra támaszkodva a cég folyamatosan ad ki forgalmi előrejelzéseket. A 2017-es kiadványukban az olvasható, hogy 2016-ban az globális IP-forgalom nagysága 1,2 zettabájt volt, míg 2021-re 3,3 zettabájtos forgalmat jósoltak (Cisco 2017). Érdekes még azt is kiemelni, hogy a Cisco szerint 2021-re a videótartalmak a teljes hálózati forgalom 82 százalékát fogják kitenni a 2016-os 73 százalékhoz képest.¹¹

Ideje összefoglalni az eddigieket! Lehet vitatni, vagy nem elég megalapozottnak tekinteni a fenti becsléseket. Megengedem, hogy ezek a becslések akár több nagyságrendet is tévedhetnek. De azt nem lehet vitatni, hogy a kanonizált kulturális adatvagyon nagyságára vonatkozó számok biztosan és egyre nagyobb mértékben elmaradnak egyfelől a jelen- és jövőbeni hálózati forgalom, másfelől a vizuális felhasználói tartalmak nagyságrendjétől.

Mindez azért fontos, mert ennek tudatában érdemes értékelni azokat a mostanában felbukkanó új jelenségeket és fogalmakat, amelyek mind valahogyan az adatrobbanás köré építhetők (big data, dolgok internetje, szenzorforradalom, adattudomány, computational social science stb.). Milyen információ robban? Beszélhetünk egyfajta kulturális adatrobbanásról, hiszen a tartalom előállításában, mennyiségében is hatalmas bővülés figyelhető meg. Ennek elsősorban a felhasználói tartalom megjelenése az oka, de ennek a jelenségnek nem a nagyságrendi bővülés az igazi lényege, hanem a kánonok átalakulása. Erre a problémára később még visszatérek. Korábban említettem, hogy méretüket tekintve a leíró metaadatok mennyisége elhanyagolható a dokumentumokban rögzített tartalmak információ-mennyiségéhez képest. Nem így van azonban a forgalmi-fogyasztási metaadatokkal, amelyek nagyságrendje a hálózati felhasználók számának, aktivitásának növekedésének köszönhetően exponenciális arányban növekszik. A kulturális adatok világán belül ezeknek a használati-forgalmi metaadatoknak a mennyisége robban! De talán még ennél is fontosabb, hogy ugyanilyen mértékben növekszik a médiafogyasztási forgalom is. Ennek az okát már említettem. A digitális hálózati kommunikáció a sokszorosítás feladatát a felhasználói végpontokba helyezi át. A digitális környezetben a sokszorosítás (határ)költsége gyakorlatilag nulla, a felhasználók egyetlen kattintással másolatokat készítenek, és a dokumentumok méretéhez képest nagyságrendekkel nagyon forgalmat generálnak a sokszoros letöltések révén.

A hálózati forgalom nagy része a felkínált tartalmak (sokszor parazita reklámtartalmak) eléréséből, fogyasztásából ered. A forgalmi, fogyasztási, használati adatok gyűjtése, feldolgozása, az ezek alapján készített felhasználói profilok készítése, és mindezeknek a tartalom-

¹⁰ A NOAA-adatok szabadon letölthetők. Sokatmondó, hogy a felhasználók által végzett letöltések össz mennyisége 2005-ben 46 terabájt, 2011-ben 1,3 petabájt, 2012-ben pedig 1,9 petabájt volt (NCDC 2012).

¹¹ Régóta nyilvánvaló, hogy a hálózati forgalom döntő hányadát a videótartalmak adják. Ezt elsősorban azzal magyarázhatjuk, hogy a mozgóképes dokumentumok mérete nagyságrendekkel nagyobb bármely más dokumentumtípushoz képest. Az irántuk megnyilvánuló igény pedig a képek, mozgóképek informatívitásából, illetve a mozgóképek időbeliségén alapuló figyelemkontrollból fakad.

előállítási és tartalomelosztási folyamatokba történő visszacsatolása elsősorban kereskedelmi, marketing célokat szolgál. Ez nyilván hasznos lehet a hálózati óriáscégek számára és mindazoknak, akik valahogyan az új digitális ökoszisztéma részeivé válnak, de kultúra- és társadalomszervezési szempontból – legalább is számomra – nem annyira releváns.

Természetesen a digitális világ nem csak a hálózaton keresztül elérhető, kulturális dokumentumokból, illetve az ezek fogyasztása során keletkező forgalmi adatokból áll. Az előbbiek nagyrészt a felszíni web részét képezik, utóbbiak már a mélyweb világába tartoznak. A mélywebben belül azonban rengeteg olyan adatbázis létezik, amelynek nem marketing értéke van, hanem más – sokszor tudományos – célokból építik őket. A big data címke ezekre is vonatkoztatható. Még a dokumentumok világához, tehát a felszíni web területéhez köthetők azok a kezdeményezések, amelyek a weben fellelhető szövegtenger újfajta elemzését végzik. A felszíni weben található szövegek, audiovizuális állományok alapján építhető hatalmas korpuszokban a nyelvhasználat szabályszerűségeit lehet nyelvtechnológiai eszközökkel feltárni (ott magának a szövegnek a létrehozásában tetten érhető nyelvhasználati szokások, konvenciók jelentik a nyelvhasználatot). Ezek a fejlesztések járulhatnak hozzá a teljes körű és magas színvonalú nyelvtechnológiai képességek eléréséhez (gépi fordítókhoz, szövegértelmezéshez, beszédfelismeréshez, beszédszintézishez stb.).

Az adatrobbanás jelensége más információs tartományokban is megfigyelhető – nem függetlenül a szenzorforradalom kialakulásától, a mindenütt jelenlevő számítástechnika, a mediatizált terek elterjedésétől. Az egészségügyi adatok gyűjtésében és feldolgozásában már a jelenünkben látszik a változás. A testünk működésével kapcsolatos adatokat könnyen lehet mérni, a szenzorforradalom eredményeként automatizálni, a hálózat segítségével egy helyen gyűjteni, elemezni, ami minőségi változást ígér a gondozásban, gyógyításban, kutatás-fejlesztésben. A közlekedési adatok, vagy általánosítva a mozgási adatok (az ember, a közlekedési eszközök, az állatok, a tárgyak mozgása) megint csak könnyen nyomon követhetők, gyűjthetők, elemezhetők, és ennek az adatrobbanásnak a logisztikai, termelési következményeiről nem kell sokat beszélni. Minden gépesíthető termelési folyamat a folyamatosan mért, gyűjtött, elemzett, visszacsatolt termelési adatok feldolgozásán alapult eddig is, ennek kiterjedtsége, alkalmazási mélysége változhat a jelenben, jövőben, ami kialakítja az adatvezérelt termelés új világát. A környezeti, időjárás adatok a mezőgazdaságban, a környezetvédelemben, a környezethasználatban, az energia-gazdálkodásban jelenthetnek hatalmas előrelépést. Az adattudomány további területeken ígér előrelépést, olykor talán áttörést (géntérkép elkészítése, vizuális mintázatok felismerése, manipulálása stb.).

Az adatrobbanásnak köszönhetően megjelenő hatalmas adatmennyiségek, az adattudomány széleskörű terjedése, az adatvezérelt iparágak ki- és átalakulása egyre meghatározóbbá válik a web fejlődésében, de ezeket a paradigmátikus tudásszervezés világába tartozó jelenségeket el kell különíteni a narratív tudásszervezési logikára épülő – kulturális – dokumentumok világától, a méretek és a tartalmak vonatkozásában is. A fentebb bemutatott számszerű mennyiségi adatokkal bizonyos értelemben nem tudunk semmit kezdeni. Mert semmitmondók. Nem lehet csak azért összemosni az adatokat, mert azonos módon tudjuk mérni őket.¹² Ha a kulturális adatvagyon nagyságrendjét közvetlenül akarnánk összevetni a kulturális fogyasztás forgalmi adatainak méretével (vagy még inkább a big data világ nagyságrend-

¹² Nyilvánvaló, hogy a Claude Shannon elmélete alapján definiált információ bit-alapú mérése mellett szükség lenne más metrikára is (Shannon–Weaver 1986). Donald MacKay az információ újfajta tipizálására és mérhetőségére vonatkozóan már évtizedekkel ezelőtt kidolgozott egy elméletet, amelyben bevezette a szelektív, strukturális és metrikus információ fogalmait (MacKay 1950, 1969), de az elmélete nem került át a gyakorlatba. A tanulmányomban elemzett jelenségek alapján mondhatjuk: kár, mert igény lenne rá.

jeivel), és nem törődne a mérhető információ minőségi tartalmával, elkedvetlenítő, vagy épp ellenkezőleg, mulatságos következtetésekhez jutnánk. Ha első, durva becslésként négy nagyságrend (tízezerszeres) különbséget tételeznénk köztük,¹³ akkor ezt az arányt ennek a tanulmánynak a méretére vetítve az egész szövegből csak 4-5 betű maradhatna értékes tartalom. Nagyjából: „ennyi”.

A nyilvánosság szerkezetének átalakulása

Eddig arról beszéltem, hogy

- i) a predigitális korban a rögzített információ elenyésző méretű volt az emberek által befogadott, feldolgozott információ mennyiségéhez képest;
- ii) a digitális korszak nem általában a kulturális dokumentumok mennyiségében, mint inkább a vizuális dokumentumok számában eredményez radikális növekedést, és azon belül is elsősorban a felhasználói tartalmak bővülésében;
- iii) az adatrobbanás döntő részét a mélywebhez sorolható adatbázisok adják, amelyen belül a forgalmi adatok elsősorban marketing érdekeket szolgálnak, és ezeknek a tudományos hozadéka és érdekessége így talán kisebb, de
- iv) a big data világán belül vannak fontos, tudományosan érdekesebb adatszegmensek is;
- v) nem szabad összemosni a különböző információ típusokat azon az alapon, hogy a mennyiségüket tekintve egységesen mérhetők.

Nem volt szó azonban egy fontos változásról, amit érdemes megemlíteni. A rögzítésnek a digitális korszak előtt volt egy nagyon fontos szűrési funkciója, ami a digitális eszközök korában, úgy tűnik, mintha eltűnőben (de legalább is változóban) lenne. A predigitális környezetben a rögzítésnek és sokszorosításnak komoly belépési korlátai voltak, és ez a szűköség egyfajta kapuőri, kánonképző szerepet biztosított a rögzítő, sokszorosító technológiákat birtokló, a sokszorosítást, publikálást, hozzáférést, terjesztést szabályozó személyek, szervezetek, államok számára. Úgy tűnik, ez a kapuőri szerep megszűnőben (vagy inkább átalakulóban) van a digitális technológiák elterjedésével. Ennek a folyamatnak vannak nyilvánvaló előnyei, de vannak hátrányai is.

A kanonizált kultúra – pont a kanonizáltság révén – közös tudást teremtett a közösen befogadható dokumentumok révén. A sokszorosítás technológiái a könyvnyomtatástól kezdve jelentősen hozzájárultak mind a nemzetközi, mind a nemzeti közös tudások létrejöttéhez és fenntartásához. A tudomány megerősödéséhez, elterjedéséhez szükség volt a lokális kontextustól való elszakadásra, amit a sokszorosított könyvek (írások), a kép- és hangrögzítés technikai tettek lehetővé, míg a nemzeti kultúrák, a közös nemzettudat, nemzeti identitás, nemzeti emlékezet kialakításához, fenntartásához szükség volt a kulturális dokumentumok (szövegek, képek, zenék, mozgóképek) kanonizálására. A predigitális korszakban a rögzítés minőségi szűrőt, egyfajta belépési küszöböt jelentett (még akkor is, ha voltak művek, amelyek érdemtelenül kerültek be a kapun belülre, és voltak olyan alkotások is, amelyek pedig nem tudtak kanonizálódni, a rögzített kultúra falain kívül rekedtek, elenyésztek az időben.

¹³ A – minimum – négyszeres nagyságrendi különbség feltételezése nem légből kapott. A 2000-es évek elején egy online szerkesztőségben a látogatói aktivitást nyomon követő naplózási rendszer fejlesztésekor szembesültem először azzal, hogy különböző típusú információból mennyire eltérő mennyiségeket kell kezelnünk. Mellbevágó volt először felismerni azt a tényt, hogy miközben egy teljes szerkesztőség több évi munkája elfért egy CD-ROM-on (650 Mb), a látogatói forgalom naplózása egyetlen nap alatt nagyságrendekkel több adatot termelt (8 Gb-ot).

A kanonizálásra, az identitásképzésre, a közösségformálásra a történetmesélésen alapuló, narratív dokumentumok alkalmasak, nem a paradigmaticus tudást megtestesítő adatbázisok (Bruner 2005). Az adatrobbanás jelenségében nagy hányadot kitevő adatbázisok ebben nem segíthetnek. Az itt keletkező adatok segítik a tudásunk elmélyülését, szélesítését is, de az élmény-, érték- és tudásközösségek kialakítását, fenntartását csak a mélywebből a felszíni webbe átszivárgó tudások alapján készülő, új és – ami nagyon fontos – narratív dokumentumok tudják támogatni.

A digitális korszakban mindenki alkotóvá válhat, az új digitális dokumentumok könnyedén kerülhetnek be a közös terekbe. Ez a változás abban az értelemben radikális, hogy ezáltal eltűnnek – a predigitális kort meghatározó – belépési korlátok. Ezzel kapcsolatban legalább két – egymással összefüggő – kérdést kell feltennünk. Egyfelől ha megszűnik a kanonizálás, vagy radikálisan csökken a jelentősége, akkor milyen technikák biztosítják majd a közös tudások, közös értékek, közös élmény, közös identitások kialakulását, fennmaradását? Ha nincs szűrés, akkor hogyan emelkedhetnek ki az értékesek az értéktelenebbek közül? Ez annál inkább válik egyre fontosabb kérdéssé, minél több tartalom érhető el bárki számára, illetve minél inkább globalizálódik a tartalomkínálat, és ezáltal minél inkább eltűnnek a helyi, nemzeti, regionális szintek, az ezeket addig menedzselő intézmények, általánosságban mondván: a közvetítők a színről. Erre persze adható az a válasz, hogy majd az algoritmusok segítenek nekünk az eligazodásban (társas navigációs technikák, közösség ajánló rendszerek stb. révén), ám ezzel kapcsolatban rögtön felmerül a második kérdés: kik határozzák meg, hogy a különböző nyilvánossági szinteken működő algoritmusoknak mi a tartalma, mi a működési módja, mik a döntési szempontjai? A predigitális korszakban a nyilvánosság különböző szinterein folyamatosan vitatták, vitathatták (még ha nem is teljesen szabadon) az elosztási és hozzáférési elvek, kánonképző mechanizmusok, megoldások célját, tartalmát, szempontjait. A nagy kérdés az, hogy a digitális világban, ahol az algoritmusok fejlesztése privát multinacionális cégek kezében van, milyen módon lehet átláthatóvá, vitathatóvá, befolyásolhatóvá tenni az általuk fejlesztett algoritmusokat, és ezáltal – végső soron – a nyilvánosság szerkezetét. A tét nem kicsi.

IRODALOM

- Apple (2013) *iTunes Store Sets New Record with 25 Billion Songs Sold*.
<https://www.apple.com/newsroom/2013/02/06iTunes-Store-Sets-New-Record-with-25-Billion-Songs-Sold/>
- Bergman, Michael K. (2001) The Deep Web: Surfacing Hidden Value. *The Journal of Electronic Publishing* Vol.7. No.1. <https://doi.org/10.3998/3336451.0007.104>
- Bohn, Roger – Short, James (2012) Measuring Consumer Information. *International Journal of Communication* 6 (2012). 980–1000.
<http://ijoc.org/index.php/ijoc/article/viewFile/1566/743>
- Bruner, Jerome (2005) *Valóságos elmék, lehetséges világok*. Budapest, Új Mandátum Könyvkiadó.
- Cisco (2017) *The Zettabyte Era: Trends and Analysis*.
<https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/vni-hyperconnectivity-wp.html>
- Good, Jonathan (2011) *How many photos have ever been taken?*
<http://blog.1000memories.com/94-number-of-photos-ever-taken-digital-and-analog-in-shoebox>
- Goody, Jack (1998) Nyelv és írás. In: Nyíri Kristóf – Szécsi Gábor (szerk.) *Szóbeliség és írásbeliség*. Budapest, Áron Kiadó. 189–221.
- HMI (2003) *How much information, 2003*.
<http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/execsum.htm>
- IDC (2010) John Gantz – David Reinsel: *The Digital Universe Decade – Are You Ready?*
<https://www.emc.com/collateral/analyst-reports/idc-digital-universe-are-you-ready.pdf>
- IMDb (2018). *IMDb statistics*. <http://www.imdb.com/pressroom/stats/>
- Instagram (2017) *Instagram Company Statistics*.
<https://www.statisticbrain.com/instagram-company-statistics/>
- Jameson, Adam D. (2012) *How Many Movies Are There?*
<http://htmlgiant.com/film/how-many-movies-are-there/>
- Lewis, Allison (2012) The Deep Web: Semantic Search Takes Innovation to New Depths. *BOSTINNO*.
<https://www.americaninno.com/boston/the-deep-web-semantic-search-takes-innovation-to-new-depths/>
- Mackay, Donald M. (1950) Quantal aspects of scientific information. *Phil. Mag.*, 41, 289.
<https://doi.org/10.1080/14786445008521798>
- Mackay, Donald M. (1969) *Information, Mechanism and Meaning*. London, MIT Press.
- Mayer, Marissa (2009) *Innovation at Google: the physics of data*.
<http://www.parc.com/event/936/innovation-at-google.html>
- McNulty, Eileen (2014) Understanding Big Data: The Seven V's. *Dataconomy*.
<http://dataconomy.com/2014/05/seven-vs-big-data/>
- Moore, Robert J. (2011) *Eric Schmidt's "5 Exabytes" Quote is a Load of Crap*.
<https://blog.rjmetrics.com/2011/02/07/eric-schmidts-5-exabytes-quote-is-a-load-of-crap/>

- NCDC (2012) *NCDC Sets New Record for Data Downloaded – Over 1.9 Petabytes!*
<https://www.ncdc.noaa.gov/news/ncdc-sets-new-record-data-downloaded-E2%80%93-over-19-petabytes>
- NOAA (2018) *National Oceanic and Atmospheric Administration. Climate.*
<https://www.nesdis.noaa.gov/content/climate>
- Saussure, Ferdinand de (1998) *Bevezetés az általános nyelvészetbe.* Budapest, Corvina Kiadó.
- Schmidt, Eric (2010) Every 2 Days We Create As Much Information As We Did Up To 2003. *Techcrunch.* <https://techcrunch.com/2010/08/04/schmidt-data/>
- Shannon, Claude E. – Weaver, Warren (1986) *A kommunikáció matematikai elmélete.* Budapest, OMIKK.
- Singleton, William T. (1981) *Management Skills.* Springer.
<https://doi.org/10.1007/978-94-010-9476-4>
- Syi (2007) *Egyben az egész.* Budapest, Typotex.
- Syi (2014) *syi.hu/cse.* Budapest, L'Harmattan – Könyvespont Kiadó.
- Syi (2017a) Interakció. In Rab Árpád (2017szerk.) *Csomópontok.* Budapest, Gondolat Könyvkiadó – Infonia – Óbudai Egyetem Digitális Kultúra és Humán Technológia Tudásközpont. 13–33.
- Syi (2017b) Olvasni az oszlopok között. In Rab Árpád (2017szerk.) *Csomópontok.* Budapest, Gondolat Könyvkiadó – Infonia – Óbudai Egyetem Digitális Kultúra és Humán Technológia Tudásközpont. 128–154.
- Techadvisor (2018) *Best music streaming services 2018.*
<https://www.techadvisor.co.uk/review/audio/best-music-streaming-services-2018-3523953/>
- YouTube (2017) *36 Mind Blowing YouTube Facts, Figures and Statistics – 2017.*
<https://fortunelords.com/youtube-statistics/>