

Small Data fogalomkörébe tartozik, mivel kérdőíves felmérés keretében, előre meghatározott kérdésekre adott, strukturáltan rögzített és tárolt adatokról van szó.

A statisztikai hivatalok is foglalkoznak a Big Data kategóriájú adatok elemzésével. Néhány érdekes példa: a mobiltelefonos cellainformációk felhasználása a lakosság ingázási szokásainak, illetve a turizmussal összefüggő mozgások feltérképezésére, vagy egy másik eset: az internetes árak vizsgálata a fogyasztói árindex leképezése érdekében. A *Központi Statisztikai Hivatal* is elemzi a turizmussal összefüggő adatokat, illetve a kiskereskedelmi forgalomra vonatkozó becsléseket végez az online pénztárgépek által küldött adatokból. És ahogy minden más szakterületen, úgy a statisztikában is új módszerek, újfajta algoritmusok kellenek, mivel a hagyományos statisztikai szoftverekkel és adatbázis-struktúrákkal nem lehet a Big Data adathalmazt kezelni. [1]

A Big Data jellemzői

A generált és tárolt adatok *menyisége* alapvetően meghatározza, hogy milyen következtetéseket lehet a vizsgálat során a rendelkezésre álló adatokból levonni.

A *változatosság* a különböző forrásokból származó, eltérő típusú, strukturált vagy strukturálatlan adatokat jellemzi. Az adatok származhatnak szöveges, képi, hang-, videó- és logfájlokból, továbbá ki nyerhetők a szenzorokból, routerekből, egyéb okos eszközökből stb.

A *változatosság, változékonyság* az adatok konzisztenciáját mutatja; az inkonzisztens adathalmaz akadályozhatja a halmaz kezelését és feldolgozását.

Az adatok *megbízhatósága, minősége* fontos tényező az elemzés pontosságát illetően.

Az egyre nagyobb *sebesség* az adatok létrejöttének, feldolgozásának és értelmezésének a gyorsaságát jelzi. Az adatfeldolgozásnál ma már nem a processzor sebessége jelenti a kihívást: az a kérdés, milyen sebességgel lehet az adatokat a feldolgozó egységhez eljuttatni. A mai szupergyors processzorok az idő nagy részét tétlenül, az adatokra várva töltik. A Big Data fejlesztőinek egyik fontos feladata az adatátadás optimalizálása, vagyis, hogy minél ritkábban kelljen mozgatni az adatokat a gyors és a lassú memóriák között. Különösen fontos ez, ha több, egymástól távol lévő számítógépeken dolgoznak egyszerre, mert ez esetben még a nagy sávszélességet is folyamatosan biztosítani kell.

A Big Data rendszerek működéséhez mindenképpen szükség van a hálózatok és az érzékelők (szenzorok)

közötti összeköttetésre, a felhőtechnológia alkalmazására, speciális adatfeldolgozó szoftverekre.

Az adatok forrása

Az adatok különböző rendszerekből – vagy a kereskedelmi folyamatokból, tranzakciókból, vagy a gépi feldolgozásból, vagy az emberek közötti kommunikációból – származhatnak. A leggyakrabban előforduló típusok az alábbiak:

- *kereskedelmi tranzakciók* (pl. bankkártyás fizetés, online/mobil banki műveletek);
- *adminisztratív ügyintézés* (pl. elektronikus adózás, biztosítási ügyek, egészségügyi vizsgálatok);
- *nyomkövető eszközök* (pl. GPS, mobil cellainformációk, Elektronikus Közútiáruforgalom-ellenőrző Rendszer – EKÁER);
- *szenzorok* (itt főleg az IP címmel rendelkező szenzorokra, vagyis az IoT [2] eszközökre kell gondolni – pl. okosautókra, okostelefonokra, intelligens háztartási eszközökre);
- *online cselekmények* (pl. keresés, látogatottság);
- *közösségi média műveletek* (pl. Facebook, Twitter, LinkedIn, Tumblr stb. – blogok, kommentek, posztok, fotók, videók).

A budapestiek számára jól ismert *BKK Futár* rendszer jó példa a Big Data alkalmazásra. Az adatok az éppen forgalomban lévő – napközben átlag 3000 – jármű fedélzeti GPS berendezéséből érkeznek a jármű pillanatnyi helyzetéről. A járművek mozgását figyelik továbbá az útkereszteszédésekben lévő érzékelők. Ez a rengeteg valós idejű adat jelenik meg a belső kerületek megállóiban lévő több száz kijelzőn, és ugyanezek láthatók az okostelefonokra letölthető mobil applikációban is.

A dolgok internete (Internet of Things – IoT)

A Big Data témakörében mindenképpen meg kell említeni az internetre kapcsolódó okoseszközöket, amelyek a világon keletkező összes adatnak mind nagyobb, és egyre növekvő hányadát adják. A nagy gyártási sorozatoknak köszönhetően a számítástechnikai eszközök, továbbá az érzékelők ára számottevően csökkent, emellett olcsóbbá vált a vezeték nélküli technológia. További fontos adalék az IoT robbanásszerű gyarodásához, hogy végre rendelkezésre áll egy új címtartomány, az IPv6, amely lehetővé teszi sok-sok milliárd eszköz és alkatrész internet-azonosítóval való ellátását, és egyúttal hálózati elérhetővé tételét.

2008 körülre teszik a fordulópontot, amikortól kezdve az internetre már több eszköz kapcsolódott, mint ahány ember. Pár éven belül, 2020-ra már 50 milliárd (!) internetre kötött eszközzel számolnak a szakemberek, de a mértékadó becslések (Cisco, Gartner) is legalább 20 milliárdnyira jósolják az évtized végén a saját internetcímmel rendelkező eszközök számát. Az IoT eszközök számos előnyt kínálnak, melyek közül az életminőség javulását, az erőforrások takarékosabb felhasználását, a közösségi tudás jelentős gyarapodását emeljük ki.

Nem szabad elhallgatni azonban az óriási kockázatokot sem, amelyek között első helyen a jogosulatlan felhasználást kell említeni. A közművek elleni infrastrukturális támadások eshetősége, az egyes erőforrások fölötti irányítás megszerzése hatalmas veszélyforrás. [3]

A Big Data egyik nagy nyertese a szociálpszichológia

A Facebook, a Twitter és a többi közösségi médium nemcsak a kapcsolattartásra és az idő haszontalan eltöltésére jó. A közösségi hálón megjelenő posztok, blogok stb., illetve a keresőmotorokban megjelenő keresőszavak generálta szédületes mennyiségű adat elemzésével a kutatók jobban megérthetik az emberi gondolkodásmódot, a viselkedési mintázatokat, a kommunikációs szokásokat, továbbá vizsgálhatják a hiedelmek és az előítéletek kialakulását, képet kaphatnak a népesség egészségi állapotáról, és szinte bármiről, amelynek a kutatására korábban megközelítőleg sem kínálkozott ennyi lehetőség. „Ezelőtt nem volt még példa arra a történelem során, hogy néhány óriásvállalat – mint a Google, a Facebook, a Twitter – számára elérhető adathalmaz ekkora hatást gyakoroljon a társadalomtudományi kutatásokra.” [4]

Az adat az új olaj, vagyis a gazdaság új hajtóanyaga

Az Európai Bizottság szerint percenként 1,7 millió gigabájt adat keletkezik. Az adatipar évente 40%-kal növekszik. Európa GDP-je 2020-ra 1,8%-kal nőhet. Az adatgazdasághoz kapcsolódó munkahelyek száma néhány éven belül Európában eléri a 250 ezret. [5]

A Gartner alelnöke hasonlóképpen fogalmazott, bár ő nemcsak a Big Data, hanem az egész informatika gazdasági hajtóerejét hasonlítja a XX. század hajtóanyagához, az olajéhoz.

A Big Data az új Big Brother?

Nem szabad áztatni magunkat azzal, hogy csupa jóval kecsgetet a Big Data korszak. A leggyakrabban hangoztatott aggályok szerint az online tevékenység minden egyes elemének rögzítése és tárolása veszélyeztetheti a személyi szabadságot, visszaélésekre adhat alkalmat, megsértheti a magánélet „szentségét”. Azt csak helyeseltethetjük, ha egy bűnöző elleni bizonyítási eljárás során a hatóságok felhasználják a mobiltelefonok híváslistáit, cellainformációit, az elkövető által meglátogatott honlapok adatait, és ezekkel az adatokkal tudják bizonyítani a bűntettet. Ahhoz azonban, hogy egy esetleges bűntény esetén a bizonyításra alkalmas adatok rendelkezésre álljanak, mindenkiről minden adatot össze kell gyűjteni. Ha ezek az adatok megvannak, akkor viszont fennáll a veszélye annak, hogy illetéktelenek kezébe kerülnek – és erre számos példából értesültünk az utóbbi időben.

További gond a felejtés hiánya, mert az internet „nem felejt”: az egyszer feltöltött képek, szövegek, videók, ezek megosztása stb. megőrződik, köztük azok is, amelyekről az érintettek szeretnék, ha egyszer s mindenkorra eltűnnének a háttértárakról.

*

Mi a tanulság? Ne csak az eszközeink, mi is legyünk okosak! Először tanuljuk meg, aztán tanítsuk meg a gyerekeknek, hogyan lehet biztonságosan közlekedni a virtuális világban. Minden hálózati művelet előtt gondoljuk át: akarjuk-e, hogy azt a kattintást is megőrizze a soha nem felejtő internet a digitális ujjlenyomatunkban.

TÓSZEGI ZSUZSANNA

IRODALOM

- [1] Mag Kornélia: Big Data a hivatalos statisztikában. www.ksh.hu/docs/bemutakozas/mta/mta_sjtb/rendezvenyek/1_mag_big-data_beveteto.pps
- [2] IoT (Internet of Things) – a „dolgok internete”. Az összefoglaló elnevezés azokat a digitális eszközöket jelenti, amelyek képesek információt rögzíteni és azt a hálózaton keresztül kommunikálni.
- [3] Bagó Péter: Mit kezdjünk 50 milliárd potenciális ügyféllel? <http://dibiz.hu/mit-kezdjunk-50-milliard-potencialis-ugyfellet/>
- [4] Szepesi András: Közösségi média a tudomány szolgálatában. <http://techindex.hu/kozossegi-media-a-tudomany-szolgalataban/>
- [5] Fehér könyv a nemzeti adatpolitikáról. Budapest, Nemzeti Hírközlési és Informatikai Tanács, 2016. április. 3. o. <http://docplayer.hu/24526642-Feher-konyv-a-nemzeti-adatpolitikarol.html>

Kép forrása:

Big Data illusztráció. https://commons.wikimedia.org/wiki/User:Camelia.boban#/media/File:BigData_2267x1146_trasparent.png