BŐGEL GYÖRGY

A BIG DATA ÖKOSZISZTÉMÁJA

TARTALOM

ÁBRÁK JEGYZÉKE

TÁBLÁZATOK JEGYZÉKE

AJÁNLÁS

Az informatika világában nem értelmezhető a „követő üzemmód”: gyakorlatilag ugyanolyan eszközökkel, forrásokkal dolgozhat Magyarországon is bárki, mint a világ legfejlettebb országaiban. Sőt mi több, olcsó és felhasználóbarát eszközeivel pont az informatika teremti meg ezt a lehetőséget számos más szakma, illetve iparág számára, amiről mindenki közvetlen, személyes tapasztalatokat is szerezhet. A hasonlóságok ellenére a felhasználók mégsem egyformák: ami valakinek az izgalmas jövő, az másnak már az unalmas múlt.

Akik az informatika világában élnek, évente kapják fel a fejüket világrengető új trendekre, ismernek meg új fogalmakat, rövidítéseket, és habitusuktól függően tarthatják egyiket-másikat forradalminak, valódi áttörésnek. Ha távolabbról szemléljük az eseményeket, rájövünk, hogy bár ebben az iparágban a fejlődés valóban rendkívül gyors, ugyanaz történik itt is, mint minden technológiai területen: az újdonságok fokozatosan beépülnek a hétköznapi életbe, a termékek és szolgáltatások idővel olyan emberek számára is elérhetővé válnak, akik nincsenek tisztában a működésükkel, de a maguk helyén és szintjén használják azokat, és jól elboldogulnak velük. Nem kell minden sofőrnek autószerelőnek vagy pláne gépészmérnöknek is lennie egyben, ahogy a számítógép vagy okostelefon használója sem feltétlenül képzett informatikus.

Mindannyiunk érdeke, fejlődésünk záloga, hogy Magyarországon minél többen legyenek a digitális képességek birtokában. Ez alatt elsősorban azt értem, hogy munkájuk, tanulásuk eredményesebb, ha informatikai eszközöket használnak, segítségükkel könnyen és gyorsan tudnak hasznos információkhoz jutni vagy információt előállítani.

Körülöttünk elképesztő ütemben nő az adatok mennyisége: megállíthatatlanul terjednek az „adatosítás” eszközei és módszerei a mezőgazdaságban, az iparban, a közlekedésben, az egészségügyben, az energetikában – gyakorlatilag mindenhol. Az így keletkezett adatmennyiség szintetizálásához, elemzéséhez, gyakorlati hasznosításához az kell, hogy ne csak az informatikusok (jelentsen bármit ez a megnevezés) értsenek ehhez, hanem saját szakterületén mindenki meg tudja oldani ezeket a feladatokat.

Ez a könyv elsősorban a jelen és a jövő különböző területeken, szakmákban tevékenykedő felhasználóihoz szól közérthető nyelven és olvasmányosan; azokhoz, akiket nem maga a gép érdekel, hanem dolgozni szeretnének, a munkájukat akarják végezni, de azt hatékonyan. Orvosokhoz, jogászokhoz, agrármérnökökhöz, közlekedésszervezőkhöz, logisztikai szakemberekhez beszél – mindenkihez, aki meg akarja érteni az új lehetőségeket, és élni szeretne az új eszközökkel, hogy eredményesebb, sikeresebb legyen. Az informatikai iparban dolgozókat gyakran frusztrálja, hogy nem tudnak hidat verni saját szakmájuk és a felhasználók között; e könyv nekik is segít megérteni az üzleti problémákat, fejleszti az absztrakciós képességüket, gyarapítja a felhasználókkal való kommunikációhoz szükséges szakmai szókincsüket.

Sokaknak ajánlom tehát ezt a könyvet, kötelezővé viszont a pályaválasztás előtt állóknak, szüleiknek és tanáraiknak tenném, hogy egyértelmű legyen számukra, milyen képességeket kell most megszerezni ahhoz, hogy valaki öt, tíz, vagy akár húsz év múlva is értékes tudással rendelkezzen.

Major Gábor
főtitkár
Informatikai, Távközlési és Elektronikai
Vállalkozások Szövetsége

BEVEZETÉS

Egy digitális marketinggel foglalkozó multinacionális cég 2014-ben közel 800 ezer változatban küldött ki egy hirdetést ügyfele vásárlóinak. A variánsokat speciális algoritmussal készítették abból a célból, hogy minél pontosabban célba találjanak, minél jobban felkeltsék az emberek érdeklődését.

Ezt nyilván csak az tudja megcsinálni, aki jól ismeri a megcélzott ügyfélkört, képes azt nem tömegként, hanem egyénekből, egyéniségekből álló tarka csoportként kezelni. A reklám testreszabásához, a „személyes marketinghez" tehát adatokra van szükség, és persze az említett okos algoritmusra, ami jelzi, hogy ki mire fog várhatóan reagálni, mivel lehet hatni rá. 800 ezer változat ki tudja hány vásárlónak, akikről rendezett adatokból álló profilok készülnek – elképzelhetjük az adatbázis nagyságát és a feldolgozáshoz szükséges analitikai feladat bonyolultságát. Azt is biztosra vehetjük, hogy ezek a hirdetésvariációk nem újságokban vagy plakátokon jelentek meg, hanem minden bizonnyal azokon a képernyőkön, amelyeket a megcélzott személyek nézni szoktak, méghozzá lehetőleg a megfelelő helyen és a megfelelő időpontban. Azt sem árt tudni, hogy ezeket az „impressziók"-nak nevezett „megfelelő helyeket és időpontokat" manapság elektronikus aukciókon értékesítik a reklámozóknak, méghozzá másodpercenként sok milliót.

Villámgyorsan kell okosan dönteni, olyan gyorsan, hogy arra már csak valamilyen automatizált rendszer képes.

Ez bizony nem a marketing és a reklám megszokott világa. A gyakorlat évtizedeken át az volt, hogy egy termékhez plakátok, újsághirdetések, reklámfilmek készültek, megjelentek valahol, majd a reklámozók figyelték, mi történik, megmozdul-e a piac. A testreszabás legfeljebb néhány szegmens (fiatalok-öregek, nők-férfiak, modernek-konzervatívok stb.) megkülönböztetését jelentette, a piackutatás kérdőívekkel, mintavételes eljárásokkal történt. Ezt a hagyományos világot alaposan felforgatta az infokommunikációs technológia fejlődése, az internet terjedése, a megállíthatatlan hálózatosodás, a közösségi média, az elektronikus kereskedelem, a minden zsebben ott lapuló mobiltelefon. A piaci statisztikák mindenütt az internetes reklámok előretörését jelzik: sok helyen ez a reklámtorta egyetlen növekvő, méghozzá gyorsan növekvő szelete.

A marketingipar tehát megérkezett az adatrobbanás, a Big Data világába. Az iparág átalakul: új módszerek és megoldások jelennek meg, megváltoznak a verseny szabályai, megváltozik a versenymezőny. Ez az átrendeződés bizonyára nem fájdalommentes, hiszen ugyanazt vagy egy éppenséggel kisebb tortát másképpen kell elosztani, és egy ilyen játszmában nem nyerhet mindenki.

A marketing- és reklámipar átalakulása karakteres példa, de nem az egyetlen. Az adatrobbanás hatása szinte minden iparágat érint vagy érinteni fog, egyeseket jobban, másokat kevésbé, lesz, akit gyorsan, másokat lassabban, lépésről lépésre haladva. De nemcsak iparágakról van szó, hanem foglalkozásokról, tevékenységi körökről, tudományágakról, politikáról, közéletről, mindenféle rendszerekről, és persze emberekről, állásokról, a munkáról és a magánéletről is.

Ez a könyv az adatokról szól: adatok gyűjtéséről, feldolgozásáról, hasznosításáról. Adatokból nincs hiány: soha nem látott tömegben keletkeznek és özönlenek mindenfelől. Adattengerben élünk, és ez a tenger egyre csak árad. Valószínűleg igazuk van azoknak, akik ezt a jelenséget a villamosításhoz hasonlítják: ahogy a múlt században az elektromos energia megjelent mindenütt, ahogy átalakította a társadalmi és a gazdasági életet, a tömegek és az egyes emberek mindennapjait, azt a módot, ahogy dolgozunk, szórakozunk, irányítjuk és szabályozzuk az életünket, gondoskodunk magunkról, érintkezünk másokkal, ugyanúgy történik most mindez az adatokkal. Adatrobbanás korában élünk, bár ez a hasonlat sántít kissé, hiszen egy robbanás pillanatok alatt zajlik le, az adatok esetében viszont hosszú folyamatról van szó, robbanásról annyiban beszélhetünk, hogy ez a folyamat pár évvel ezelőtt felgyorsult, az adatok mennyiségének, változatosságának növekedése rendkívüli sebességre kapcsolt.

Az adatokban óriási lehetőségek rejlenek, amelyeket meg kell látni és ki kell aknázni. Ezek a lehetőségek sokfélék és sokrétűek – könyvünkben számtalan példát hozunk majd fel erre. Lesznek, akik élni tudnak az új lehetőségekkel, és olyanok is, akik nem. A feladat nem könnyű, különleges felkészültséget és csapatmunkát igényel. Az adatrobbanás a lehetőségek mellett új kockázatokkal és veszélyekkel is jár, a lehetőségekkel élni és visszaélni egyaránt lehet. Bár nem válhat mindenki adatbányásszá vagy éppenséggel „adattudóssá", elemi szintű tájékozottságra szüksége lesz, hiszen ami történik, gyakorlatilag mindenkit érint.

Könyvünkkel ezt a „mindenkit" célozzuk meg: az újdonságok iránt érdeklődő embert, annak szakmájától, előzetes felkészültségétől függetlenül.

A témáról egyfelől sok hatásvadász, könnyen olvasható, de tartalmi szempontból meglehetősen sekélyes, másfelől számos nagyon igényes, de csak specialisták számára érthető munka jelenik meg; ezt a könyvet e két szélsőség közé igyekeztünk belőni.

Legfőbb mondanivalónk az, hogy az adatok önmagukban semmit sem érnek: értéküket a feldolgozás, a hasznosítás adja. Az adatrobbanás lehetőség arra, hogy gazdagabbak, hatékonyabbak, termelékenyebbek legyünk, hogy új értéket teremtsünk. Példák sorával illusztrálhatjuk, hogyan használják az adatokat költségcsökkentésre, innovációra, folyamatok felgyorsítására, kockázatok felmérésére, problémák előrejelzésére és más célokra. A technológiai lehetőségek súlyos problémákkal és feszültségekkel találkoznak: elöregedő társadalom, környezetszennyezés, vízellátási és élelmezési gondok, anyagilag fenntarthatatlan, alacsony hatékonyságú egészségügyi rendszerek, képzési és átképzési feladatok, eladósodott államok... Az adatrobbanás, az okos rendszerek segítséget adhatnak a gondok enyhítéséhez. Bizonyára olyat is látunk majd, hogy egyes elmaradott vidékeken egész fejlődési fázisokat ugranak át az új megoldások segítségével.[1] E könyv megírása idején a technológiai piacra szakosodott piacelemző és tanácsadó *IDC* cég azt jósolta, hogy a kiskereskedelmi szektorban 2014-ben 1,3 milliárd dollárt fognak analitikai szoftverekre költeni, vagyis arra, hogy hasznos következtetéseket szűrjenek le az adatokból.[2]

Súlyos globális problémák és óriási technikai lehetőségek korában élünk. Számtalan rendszert, eszközt lehet adatok és okos algoritmusok segítségével jobbá, okosabbá tenni. Az adatfeldolgozás, az analitika, az adatokra épülő döntés-előkészítés ugyanakkor nem csodaszer: a lehetőségek végesek. Ráadásul semmi sincs ingyen, a lehetőségek kihasználásához beruházásokra van szükség: gépi kapacitásokat kell vásárolni vagy bérelni, szoftvereket kell fejleszteni, szakembereket kell foglalkoztatni, szolgáltatásokat kell megvásárolni. A munkának akkor van értelme, ha ezek a befektetések és kiadások megtérülnek.

Egy Big Data típusú adatbázis annyit ér, amennyi hasznot hozhatnak az elemzéséből levonható következtetések. Bizonyára sokan tapasztalják, hogy egy bizonyos ponton túl fizikai vagy gazdasági korlátokba ütköznek. Szerencsére ezek a korlátok mozognak, mert a technika fejlődik, a számítógépes kapacitások pedig egyre olcsóbbak lesznek. A korlátok mozognak, kitolódnak – de léteznek, a csökkenő hozadék törvénye itt is érvényesül. Mindemellett a számok, az adatok nem mondanak el mindent: „A számok egyelőre nem tudják megragadni az élet mindennapi gazdagságát, színeit, érzéseit, titkait. Egyre nagyobb szükség van a meditatív, elmélyült, a dolgozószobák csendjében elemző társadalomtudományokra is” – mondta egy interjúban[3] *Hankiss Elemér*.

---

[1] Egyes afrikai országokban például azért is terjed gyorsan a modern infokommunikációs eszközökkel végzett egészségügyi távdiagnózis, mert egyszerűen nincs elegendő helyi specialista.

[2] *Retail, mining the store*. Bloomberg BusinessWeek, 2014. okt. 13. 54. o.

[3] *Mintha újra egy zátony felé sodródnánk*. Interjú Hankiss Elemérrel, készítette Hercsel Adél. HVG online, 2014. aug. 18., http://hvg.hu/kultura/20140818_mintha_ujra_egy_zatony_fele_sodrodnank/.

A KÖNYV FELÉPÍTÉSE

Könyvünk hat fejezetből áll. Az *1.* az adatrobbanást általában, illetve annak dimenzióit igyekszik megragadni és leírni: példák segítségével mutatjuk be, hogyan halad „a világ adatosítása", mekkora adatbázisokról és adatfeldolgozó kapacitásokról beszélünk egyáltalán, és hogy miért fontos az adatvagyon szakszerű kezelése. A *2.* fejezet a könyv legfontosabb szakasza: az adatokra épülő okos rendszerekről szól, vagyis lényegében azt tárgyalja, miként lehet hasznosítani a felhalmozott adatvagyont. Az olvasó számos példa kíséretében megismerkedhet az okos rendszerek logikai modelljével és az ahhoz tartozó tevékenységekkel.
A *3.* fejezet a modern orvosi biológia példájából kiindulva összefoglalja az adatrobbanással kapcsolatos fontosabb technológiai trendeket, az okos rendszerek építésénél használt infokommunikációs technológiákat.

A *4.* fejezet tárgya az elemzés és a modellezés. Röviden bemutatja a támogatható döntések körét, felvázolja az adatbányászat általános folyamatmodelljét, végül a teljesség igénye nélkül áttekintést ad az analitikai munka eszközeiről és a sikeresség feltételeiről. Az *5.* fejezet a veszélyekre és a kockázatokra hívja fel a figyelmet: kitér a magánélet védelmére, az emberek és a gépek versenyére, a felhalmozódó társadalmi feszültségekre. A *6.* fejezet gyorsfényképet ad a Big Data ökoszisztémáról, vagyis az új lehetőségeket kihasználó kisebb és nagyobb vállalkozásokról, a régi és az új játékosokról, az érintett, különböző szerepeket betöltő intézményekről, kutatóhelyekről, iskolákról.

Az ökoszisztéma talán legérdekesebb tagja az „adattudós": az a szakember, aki központi szerepet játszik a Big Data projektekben, okos rendszerek építésében és működtetésében, és aki iránt manapság különösen élénk kereslet mutatkozik a munkaerőpiacon.

Az adattenger „élővilágának" alaposabb elemzése több kötetet töltene meg, a könyv záró fejezetében ezért csak néhány jellegzetes példa bemutatására vállalkozunk. Ez a könyv leghosszabb szakasza, és nem véletlenül: a Big Data ökoszisztéma nyüzsög és fejlődik, új vállalkozások, termékek, szolgáltatások rajzanak ki, amelyeket nehéz rendszerezni (bár teszünk erre egy kísérletet); az idő, a természetes kiválasztódás dönti majd el, hogy mi lesz életképes, mi marad fent és

fejlődik tovább. Az érdeklődés mindenesetre óriási, a tervek és szándékok sokfélék. Az infokommunikációs ipar olyan régi óriásai, mint például az *Oracle*, a *SAP*, a *Hewlett Packard*, az *IBM*, a *Cisco*, a *Microsoft* vagy az *EMC* sorra hirdetik meg Big Data stratégiájukat. Vállalkozások tömege bukkan fel szinte a semmiből. Különböző iparágakban tevékenykedő nagyvállalatok nyitnak saját analitikai fejlesztő központokat vagy adnak megbízásokat régi és új tanácsadó-szolgáltató cégeknek. Politikusoknak, fontos állami hivataloknak kell eldönteniük, hogy mit kezdenek az intézményeknél felhalmozott adatvagyonnal. Az Európai Unió vezetői új programokat és projekteket hirdetnek meg. Kutatási és oktatási programok indulnak mindenfelé, platformok épülnek, sorjáznak az adattudósi álláshirdetések...

A témák tárgyalásánál, amennyire lehetséges volt, gyakorlatiasságra törekedtünk. Az egyes fejezeteket úgy építettük fel, hogy az olvasó példák (esetenként párhuzamos példák) segítségével ismerkedjen meg a vizsgált jelenségekkel, és minél több valóságból vett esettel, illusztrációval találkozzon. Természetesen minden gyakorlati példánál felmerül az a lehetőség, hogy annak alanya megváltozik, átalakul, átértékelődik, éppen ezért olvasás közben nem árt utánanézni, mi is történt a kézirat lezárása óta. A Big Data ökoszisztémája rendkívül dinamikus világ: szinte percenként születnek új vállalkozások, cserélődnek a szereplők, emelkednek vagy hullanak a szerencsecsillagok; új kutatási-fejlesztési eredmények születnek, friss termékek és szolgáltatások jelennek meg a piacon, állami és szövetségi projektek indulnak... A változások szinte követhetetlenek.[4]

A könyv megírásához sokféle szakirodalmi forrást használtunk fel. A végén található irodalomjegyzék természetesen nem teljes, és biztosak vagyunk abban, hogy mire a kéziratból valódi könyv lesz, újabb művek jelennek meg, méghozzá nem kis számban, hiszen a téma iránt óriási érdeklődés mutatkozik. A szöveges anyagok mellett természetesen a filmek körében is érdemes körülnézni: a tárgyalt témákhoz remek videókat lehet találni, néhányra ezek közül hivatkozunk is a megfelelő helyeken. Nyilván az olvasó is tapasztalja, hogy az okos rendszerek világa nagyon látványos; egy jól megcsinált film gyakran sokkal többet mond egy hosszú szövegnél.

A szerzőnek sok gondot okozott egyes angol elnevezések magyar megfelelőjének megtalálása, különösen azoké, amelyek eredeti formájukban is bizonytalan, többféleképpen értelmezhető tartalmúak. A változás olyan gyors, annyi újdonság zúdul ránk, hogy a nyelv nehezen tudja követni.

---

[4] Csak egyetlen példa: e könyv megírásával egy időben jelentette be szétválását az informatikai ipar egyik óriása, a *Hewlett Packard*. A sajtónyilatkozatok szerint egyik utódcége adatközponti szolgáltatásokat fog nyújtani vállalatoknak.

Az olvasó abban a szerencsés helyzetben van, hogy a könyvben leírt okos rendszerek közül sokkal maga is találkozhat, kipróbálhatja azokat. Az említett okos háztartási eszközök, jeladó órák, karperecek, forgalomirányítási rendszerek, orvosi diagnosztikai készülékek, szenzorok, drónok, intelligens autók itt vannak körülöttünk, és ha valamelyik még nem próbálható ki, majd az lesz holnap vagy holnapután. Az érzékelhetőség és a „kipróbálhatóság" sajnos a veszélyekre és a kockázatokra is igaz: tapasztalhatjuk, miként hatolnak be a magánéletünkbe az új rendszerek, hogyan harapnak egyre nagyobbat az algoritmusok és az intelligens robotok a munkaerőpiacból, hogyan silányítják tömegmanipulációs társasjátékká a politikai életet egyes, a marketingből átvett eszközök.

Csak megismételni tudjuk: nem árt tájékozódni, az adatrobbanásról és annak következményeiről mindenkinek tudnia kell.