

Adatbányászat és a közösségi hálózatok

***„Beírtak engem mindenféle Könyvbe
és minden módon számon tartanak.
Porzó-szagú, sötét hivatalokban
énrólam is szól egy agg-szürke lap.
Ó, fogcsikorgatás. Ó, megalázás,
hogyan rab vagyok és nem vagyok szabad.
Nem az enyém már a kezem, a lábam,
és a fejem, az is csak egy adat.”***

- KOSZTOLÁNY DEZSŐ

TARTALOM

Bevezető	2
Adatbányászatról általánosságban	3
Adatbányászat, üzleti analitika	4
Új trend kezd kibontakozni	4
Marketing, és a közösségi média	5
A közösségi adatbányászat	5
Társadalomtudomány és a közösségi hálózatok	6
Profil.....	7
Adatgyűjtés, NSA és ami mögötte van	8
Mire használhatók az összegyűjtött anyagok?.....	9

BEVEZETŐ

Dolgozatom témájául, a közösségi hálózatok és az adatbányászat kapcsolatát választottam. Véleményem szerint az ezen hálózatok működésének kiaknázása, a "Big Data", adatbányászat és ezek elemzése nélkül nem lenne megfelelő. Nagyon sok területre kihatással van a közösségi hálózatokból nyert adatok elemzése. A médiától kezdve a marketingen át, egészen az oly vitatott megfigyelési botrányokig, ami manapság eléggé előtérbe került, hogy jogosan vagy alaptalanul azt mindenki döntse el maga. Ha ezeket együtt társadalomtudományi szempontból is vizsgálánk, érdekes dolgokra bukkanhatunk. Dolgozatomban ezeket a területeket fogom érinteni, kisebb-nagyobb kitérőkkel.

ADATBÁNYÁSZATRÓL ÁLTALÁNOSÁGBAN

Az adatbányászat a nagy mennyiségű adatokban rejlő információk fél-automatikus feltárása különféle algoritmusok alkalmazásával. A leggyakrabban céljának az érdekes, értékes, értelmes összefüggések keresését értjük nagy adathalmazokban. A megfogalmazásokban szereplő szavaknak különös súlya van és egyik sem elhagyható (ahogyan sokan teszik). Az adatbányászatnak mindig üzleti célúnak és szemléletűnek kell lennie, értékelni kell minden esetben, hogy egy-egy feltárt összefüggés hordoz-e értéket potenciálisan a szervezet számára. Végezetül magyarázhatónak kell lennie az összefüggésnek, azaz fel kell tárni azt, hogy miért és hogyan jutottunk egy következtetésre, nem elegendő elfogadni az eredményeket.

Az adatbányászat egy olyan módszertan, amely adatbázisokból olyan implicit és rejtett információkat, összefüggéseket, mintázatokat nyer ki, amelyek a gyakorlatban is hasznosíthatók. Az adatbányászat felhasználja a tudomány több ágának eredményeit: a matematikán belül a statisztikát, számítástudományon belül a mesterséges intelligenciát, valamint az információtechnológia eszközeit. Szubjektív vagy nem egyértelműen eldönthető, esetleg vitatható dolgokat az adatbányászati módszerek legfeljebb közelítéssel tudnak megoldani, de a közelítés pontossága, jósága sosem tárható fel, azaz használhatatlan eredményeket kapnánk. A megfogalmazás másik erőssége, hogy ebbe az adatbányászat számos ága belefér, így pl. szövegbányászat és a génkutatás is. Szűk értelemben azonban adatbányászat alatt a strukturált adatbázisokban tárolható adatokon értelmezett összefüggés-kereső tevékenységeket értjük.

ADATBÁNYÁSZAT, ÜZLETI ANALITIKA

Itbusiness.hu, 2012. januári cikke alapján

Az elemzőeszközöket a vállalati vezetők leghatékonyabb fegyverei között tartják számon. Manapság különösen az előrejelzésre képes algoritmusokat becsülik sokra a bizonytalan gazdasági környezetben. Nem felejthetjük el, hogy igen veszélyes terület is tud lenni, ugyanis egy rossz döntéssel jól menő cégek mehetnek tönkre, pillanatok alatt. A közösségi hálózatok térnyerésével egy közzétett információ pillanatok alatt elterjed. Hogy mennyire képes a közösségi média befolyásolni például pénzügyi világot, még nem ismert.

Az IDC (International Data Corporation) előrejelzése alapján, a vállalkozások 2015-re kb. 120 milliárd dollárt fognak költeni hardverre, szoftverre és szolgáltatások beszerzésére, melyekkel jobban kihasználják a nagy adatban (Big Data) rejlő lehetőségeket, és hogy pontosabb elemzéseket készíthessenek, hatékonyabban menedzseljék pénzügyi műveleteiket. Ehhez a prediktív analitikát használják. Hogy mi is az a prediktív analitika? Az adatbányászat és a matematikai elemzés eszközeivel dolgozik, hogy összefüggésekre leljen az adattömegben rejlő mintázatok között, és előre jelezze az üzleti trendeket például a fogyasztói aktivitásban. Azonban az eredményei nem valósidejűek.

ÚJ TREND KEZD KIBONTAKOZNI

Szintén az IDC előrejelzése erősíti meg, hogy a hálózatelemzés, a közösségi hálózatok terjedésével a „social decision”, a tudásmenedzsment és a BI¹-megoldások alkalmazása a mobileszközökről azok az irányok, amelyek új utat mutatnak. Alkalmazási fókuszterület továbbra is

¹ BI: Business Intelligence, üzleti intelligencia

a kockázatkezelés, a teljesítménymenedzsment, valamint a marketingmenedzsment.

A prediktív analitika és a vizuális adatfelderítés a legnépszerűbb döntéstámogatási módszerek között foglal helyet. Szakértők állítása szerint az elmúlt években legfőképp az adatgyűjtésen volt a hangsúly, majd a közösségi hálózatok terjedésével az adatok kinyerése és részletes elemzése került előtérbe.

Offline elemzésről jobban megéri áttérni az online elemzési formára, ugyanis sokkal nagyobb potenciál rejlik a valós idejű elemzésekben. A teljes szöveg-, videó- és hangbányászat mellett egyszerre komplexebb adattömeget vizsgálhatunk. A minél pontosabb jóslatok elkészítéséhez egyre fontosabbá válik a szimuláció és az előrejelzés. Vajon képesek leszünk-e az adatbázisokba gyűjtött adatok elemzésével a "jövőbe látni" vagy következtetéseket levonni?

MARKETING ÉS A KÖZÖSSÉGI MÉDIA

Socialmediadata.com, Adam, 2013. januári bejegyzése lapján

A vállalkozások többségének elengedhetetlen, hogy jelen legyen az interneten. Alapvető fontosságú számukra, hogy használják a közösségi médiát, ami változatos és nagy mennyiségű adatot biztosít ahhoz, hogy saját termékeik és szolgáltatásaik felé irányítsák a felhasználókat, és ily módon építsék a márka identitását.

A KÖZÖSSÉGI ADATBÁNYÁSZAT

A közösségi média adatbányászat fontos kutatási területté vált, ami hatalmas mennyiségű adatot kezel. Ezekből az adatokból kinyert minták segítségével sok problémára megoldást találhatunk. A „social engineer” vagy a vezető adatelemző (chief data scientist) ezeknek az

adatoknak az elemzésével segíti a vállalkozások intelligens fejlődését és különféle hatékony üzleti tervek kidolgozását – mindezeket a társadalmi-háló elemzésével, ahol fórumok, blogok, videó megosztó portálok, közösségi oldalak tartalmát vizsgálják. Nem csak arra kéne összpontosítani, hogy a felhasználók hogyan használják ezeket az oldalakat, milyen termékeket, szolgáltatásokat használnak, hanem hogy hogyan reagálnak jelentős eseményekre, és mi befolyásolja döntésüket. Ezeknek az adatoknak az összességéből többféle profil is felállítható. Bizonyos szempontokat figyelembe véve ezek nem csak a marketingeseknek lehet hasznos, hanem választások előtt a politikai pártoknak vagy éppen a titkosszolgálatoknak.

Esettanulmány (New Look): Néhányan azt mondják, eldobható; mások szerint "trendi", de egy dolog biztos – napjaink legújabb divatja a gyors megtérülés szerint él vagy hal. Hat hétig várni egy felmérés eredményére, nem lehet opció. Elég kitenni valamit a közösségi hálózatra, megosztani, twittelni róla. Pillanatok alatt elterjed és az emberek véleményük lesz róla. Ezután nincs más dolgunk, mint kielemezni ezeket az adatokat.

TÁRSDALOMTUDOMÁNY ÉS A KÖZÖSSÉGI HÁLÓZATOK

Az adatok halmaza mindig valós cselekvéseket fed le, azok elemzése alapján pedig sokat tudunk meg a cselekvőről. Megtudhatjuk például, olyan ember vagy-e, aki visszafizeti a hiteleit, ahogy azt is, hajlamos vagy-e a cukorbetegségre. Minderre az adatok azért képesek, mert a társadalmi környezet jelentősen befolyásolja, hogy ki is vagy. Ha látom, hogyan viselkedsz egy adott helyzetben, akkor hozzáképelem, mit tennél egy másik szituációban – annak alapján, hogy melyik társadalmi csoport tagja vagy. Rengeteget lehet így megtudni az emberekről, még akkor is, ha ezek nem szó szerint jelennek meg az adatokban. Az emberek viselkedését nagyon meghatározza társadalmi környezetük, az,

hogyan tartanak helyes viselkedésnek, és hogy mi mindent tanulnak el egymástól. Így az adatok sokaságának elemzése elsősorban a kapcsolatok megtalálásáról szól. Adott emberek kapcsolatairól és a viselkedési módok közötti összefüggésekről.

PROFIL

Origó, Simon Tamás, 2013. októberi cikke nyomán

A morzsák, amit magad után hagysz az online világban, az egész életedet elmeséli, megmutatja aktuális választásaidat. Ez alapjaiban más, mint amit a Facebook-ra kiírsz. Oda ugyanis azt írod ki, milyen lenni szeretnél, amit el szeretnél hitetni magadról az emberekkel. Azt, hogy valójában ki vagy, sokkal inkább megmutatja, hogy hol időzöl, vagy mit vásárolsz. Amellett se mehetünk el szó nélkül, amivel a facebook kísérletezik mostanában. Tervezik bevezetni azt az új funkciót, amivel figyelik a felhasználó egérmozdulatait, hogy éppen a hirdetések felett van-e vagy csak üzenetet ír, netalántán egy cég frissen posztolt termékein akadt meg a szeme.

Kosztolányi Dezső írta az alábbi sorokat, „Beírtak engem mindenféle könyvbe” c. versében. Vajon mit írna ma a költő? Nem hiszem, hogy nagyon mást, ugyanis ezek a sorok máig nem veszítették el aktualitásukat, sőt. Felhasználói szokásaink, mobilos és netes aktivitásunk

követéséből és elemzéséből lassan tényleg adatokká válunk, amelyeket fel lehet használni társadalmi folyamatok modellezésében, trendek megjóslásában esetleg megfigyelésekhez. Az adatelemzés lehetővé te-

**„Beírtak engem mindenféle Könyvbe
és minden módon számon tartanak.
Porzó-szagú, sötét hivatalokban
énrólam is szól egy agg-szürke lap.
Ó, fogcsikorgatás. Ó, megalázás,
hogyan rab vagyok és nem vagyok szabad.
Nem az enyém már a kezem, a lábam,
és a fejem, az is csak egy adat.”**

szi például olyan csoportok és összefüggések felfedezését, amelyekre egyébként nem derülne fény. A társadalomtudósok szerint ezeknek a rejtett, úgynevezett niche-csoportoknak a felfedezése lehet az új megközelítés egyik legnagyobb haszna. Az adatok társadalomtudományi alkalmazásánál azonban van egy komoly probléma: az adatok jelentős része már cégektől (például Google, Facebook) vagy állami szervektől származik. A Big Datában a cégek is egyre nagyobb lehetőségeket látnak. Egyik konferencián a Microsoft munkatársa elmondta: ők a közösségi hálón található történeteket elemzik. Hangulatokra, trendekre, véleményvezérekre vadásznak, az eredményeket aztán a marketingkampányokban szeretnék felhasználni. Az előadó felhívta a figyelmet arra, hogy ugyanez a módszer a választási kampány alatt is alkalmazható. Becslések szerint 2020-ra körülbelül 40 000 petabájtnyi tudományos adat halmozódik fel, de ezzel együtt is valószínű, hogy unokáink csak nevetnek majd azon, hogy Big Datának neveztük a mostani évtizedeket.

ADATGYŰJTÉS, NSA ÉS AMI MÖGÖTTE VAN

MTI, 2013. júniusi cikke nyomán

„...Amikor feltölt vagy valamilyen más módon elküld tartalmakat a Szolgáltatásainkba, világszerte érvényes engedélyt ad a Google-nak az ilyen tartalmak felhasználására, hosztolására, tárolására, reprodukálására, módosítására, származékos művek létrehozására, megosztására, közzétételére, nyilvános előadására, nyilvános megjelenítésére és terjesztésére...”

GOOGLE ÁLTALÁNOS SZERZŐDÉSI FELTÉTELEK

A Washington Post és a Guardian cikke szerint, több nagy szolgáltató - mint például a Google vagy éppen a Facebook - hozzáférést biztosított az amerikai hírszerzésnek felhasználók adataihoz. Cégek természetesen tagadják, a titkosszolgálat pedig azzal védekezik,

hogyan csak külföldiek adatait érinti, bár azt sem tagadják, hogy „véletlenül” több millió amerikai adatai is hozzájuk kerülhetett.

A hírszerző irodák széleskörű adatbányászatot folytattak a legnagyobb internetes cégeknél, így hozzájuthattak felhasználók fotóihoz, beszélgetéseikhez, e-maileikhez. George W. Bush elnöksége alatt 2007-től férhetett hozzá az NSA² és az FBI³ a Microsoft, a Google, a Facebook, a Skype, és az Apple szervereihez. James Clapper az amerikai nemzeti hírszerzés igazgatója elismerte az adatbányászatot. Ezek már szinte tények, bár elismerni senki nem fogja. És mégis mit tehet egy átlagos felhasználó ez ellen, ha nem tetszik neki? Nagyrészt semmit, mert ezek a szolgáltatások annyira összefonódtak az internettel, hogy kikerülhetetlen a használatuk.

MIRE HASZNÁLHATÓK AZ ÖSSZEGYŰJTÖTT ANYAGOK?

Megtudhatjuk belőle, ki, mikor merre járt, kivel beszélt meg találkozót, közösségi oldalakon megjelent eseményekre hogyan reagált. Különböző adatbányászati technikák segítségével kapcsolati hálók, illetve kommunikációs minták készíthetők. Mivel az adatok konkrét személyekhez köthetők, így egy teljes profil állítható fel, emberek, csoportok életmódjáról, napi rutinjáról. Adatelemzés terén a különböző elemző technikák és a Big Data, lehetőséget teremt arra, hogy ezt a hatalmas méretű adattömeget hatékonyan feldolgozzák. Hasonló témáról beszélt Ságvári Bence a [CIO Hungary 2013 Konferencián](#), érdemes átlapozni az előadását.

² NSA: National Security Agency (Nemzetbiztonsági Ügynökség)

³ FBI: Federal Bureau of Investigation (Szövetségi Nyomozó Iroda)

IDÉZETT FORRÁSMUNKÁK

Adam. „Discover What Mining Social Media Data Means?”
Socialmediadata.com, 2013. január.

„Az NSA feltörte a Google és a Yahoo kommunikációs csatornáit is.”
MTI, 2013. június.

Bence, Ságvári. „Adatok leszünk mindannyian.” CIO Hungary Konferen-
cia, Balatonalmádi, 2013-április 18-19.

„Memória alapú alkalmazásplatformok.” itbusiness.hu, 2012. január.

Tamás, Simon. „Új korszak kezdődött a tudományban.” Origo.hu, 2013.
október.