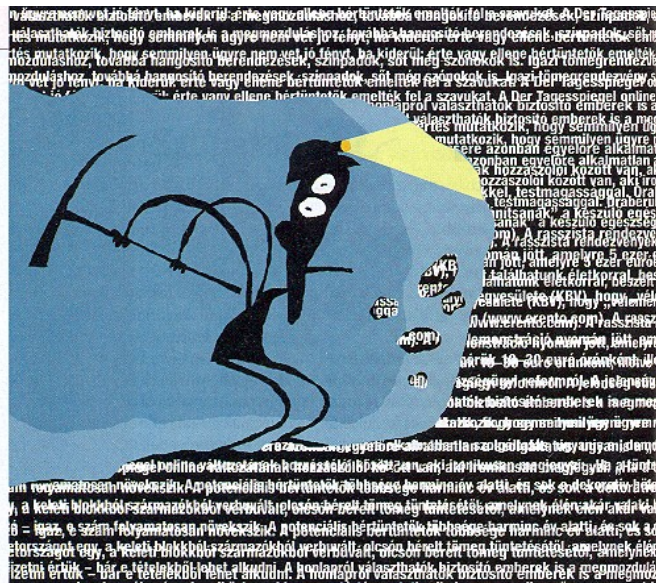


**INFOTECH**

**A rejtett összefüggéseket feltáró szövegbányászat pénzben mérhető hasznot kínál a cégeknek.**

# Sorok között

**A**kár regényt is képes írni a Mexica nevű program. Készítője, aki a szoftver működését nemrég publikálta egy tudományos folyóiratban, úgy okoskodott, hogy a szereplők érzelmi viszonyai és viszonyosságai jól kiszámítható fordulatokat követnek a puhakötésű irodalmi művekben. Habár a szoftverrel írott



mű nem tör irodalmi Nobel-díjra, érdekes szemléltetése annak, hogy a számítógépek képesek a szöveget többnek látni, mint csupán karakterek láncolatának.

Az persze aligha várható, hogy a számítógéppel írott lektűrök köré komoly iparág szerveződjön, a tárolt adatokból történő információ kinyerésének kihasználása

viszont pénzben mérhető előnnyel kecsegteti a cégeket. Különösen ott, ahol az adatokat szinte kizárólag elektronikus és túlnyomórészt strukturálatlan szövegek (dokumentumok, e-mailek) formájában tárolják. „Üzleti kockázatot is jelent, sikerül-e megtalálni a rendelkezésre álló adatokban a bennük rejlő információt” – hívja fel a figyelmet Vadász Pál, a Montana Zrt. elnök-vezérigazgatója.

A szokásos, karakterláncok egyezésén alapuló keresésnél jóval többről van szó. A szövegbányászat – az elsősorban számzerű kapcsolatokat vizsgáló adatbányászathoz hasonlóan – erősen támaszkodik a statisztikai összefüggésekre, de a mesterséges intelligencia területéről ismert módszereket is használ. Ilyen például az úgynevezett tudásfa kiépítése, ami tartalmazhat például valószínűségi alapokon nyugvó szinonimálancokat és ellentétpárokat. „Ha például a máteszalkai rendőr azt mondja, hogy a tettes koponyán ütötte az áldozatát, a zalakarosi rendőr pedig úgy veszi fel a jegyzőkönyvet, hogy a gyanúsított fejcsével fejbecsapta a szerencsétlent, akkor a klasszikus karakterláncokon alapuló keresés nem veszi észre a hasonlóságot a két eset között, a szinonimákat használó indexelés viszont igen” – hoz példát egy rendőrségi projektből Vadász Pál.

**Bányamunkák**

*A szövegbányászat olyan informatikai eszközök, módszerek összefoglaló neve, amelyek egyszerűbbé, érthetőbbé, illetve láthatóvá teszik a rejtett, vagy fáradtságosan előkereshető szövegösszefüggéseket, információkat. Ezt jellemzően a vállalati adatvagyonon, illetve weblapok jól megcélzott halmazaán végzik. Néhány alkalmazási példa:*

**WHIRLPOOL.** A háztartási gépeket gyártó cég nem tartotta elégségesnek a termékreklamációról szóló matematikai, statisztikai adatok szoftveres, adatbányászati vizsgálatát. Az óriáscég ma szövegbányászati alkalmazást használ az ügyfelek viselkedésének mélyebb megértéséhez, a garanciához kötődő információk elemzésére.

**HP.** A társaság a vevők véleményének kiértékelésére használja az alkalmazást: a vásárlási szokások terén olyan árnyalt információk után kutat, amelyek nem olvashatók ki más jellegű kutatásokból.

**ITDH.** Komplex informatikai fejlesztést valósított meg a kht. tavaly az EU strukturális alapjából pályázaton elnyert forrásból, amelynek fő célja a cég „kollektív memóriájának” megteremtése volt. Új dokumentumkezelő, ügyfélkapcsolat-kezelő szoftvert és kereső rendszert vezettek be.

**MONTANA.** Saját szövegbányászati rendszerével végzi a cégről, a konkurenciáról szóló sajtómegjelenések, vagy épp a humán erőforrás-piacról szóló hírek vizsgálatát.

A szövegekből kinyert információk számszerű adatokkal összekapcsolva fontos rejtett összefüggések feltárására is képesek. „Ez biztos alapot nyújthat akár termékfejlesztési, akár rövid vagy hosszú távú üzleti döntések megalapozásához” – állítja Elbert Éva, a SAS Institute Kft. marketingvezetője. A határos keresésekhez persze megfelelő nyelvi tudás is szükséges. „A szótöveken vagy összetett szavakon túl fontos lehet például a szófajok felismerése: magyar szövegekben ugyanis a főnevek jellemzően többet árulnak el a szöveg tartalmáról, mint az igék” – árul el egy szakmai fogást a Montana Zrt. első embere.

Akárcsak sok más ötlet, a szövegbányászat is szigorúan titkos eszközként indult: az 1980-as évek közepén a CIA alkalmazta a módszert a nagytömegű kémjelentések összegzésére, a technológia azonban ma már az üzleti szférában is egyre terjed. Különösen hasznos lehet a használata olyan

területeken, ahol nagy mennyiségű szöveggel kell dolgozni: ilyen lehet például a szabadalmak vizsgálata vagy a sajtófigyelés. Ez utóbbit a Montanánál saját magunk végzik szövegbányászati eszközzel, akárcsak a konkurencia vagy a humán erőforrás-piac híreinek vizsgálatát.

Magyarországon is vannak már példák nagyobb szövegbányászati projektekre. Jankó Attila, az ITDH Kht. informatikai igazgatója szerint új kereső rendszerük alkalmas előre definiált információ-források – például vállalati weblapok – rendszeres figyelésére, és ha azokon valamilyen érdekes információt publikálnak, akkor erről levelben értesíti a megrendelőt. A bevezetés az ITDH tapasztalatai szerint két hónap alatt megvalósítható, bár az összes funkció elsajátítása hosszabb időt igényel. A rendszer a technikai jellegű feladatokra fordított munkaidőt is radikálisan csökkentheti.

SIMON S. ANDRÁS

## Nekilődül a Linux?

**M**egtört a jég a nagy rendszerintegrátor cégek háza táján: immár nyíltan vállalják, hogy nem csak a Microsoft megoldásait, hanem linuxos szoftvereket is kínálnak ügyfeleiknek. Legalábbis ezt mutatja a Novell és a Synergon múlt hét végén bejelentett régiós szintű stratégiai együttműködése. Szittya Tamás, a Novell Magyarország Kft. ügyvezetője szerint már évek óta nem „vagy”-ban, hanem „és”-ben gondolkodnak a cégek, amikor szoftverrendszerek kiépítése kerül szóba. Magyarán a linuxos és a windowsos környezetet ma már jól megfér egymás mellett az összetett vállalati környezetben. Sőt, a Microsoft és a Novell tavaly novemberben bejelentett világszintű együttműködése nyomán (*Figyelő, 2006/46. szám*) a hatalmas hálózatokat üzembe állító és fejlesztő rendszerintegrátorok is sokkal bátrabban ajánlják ügyfeleiknek az olcsóbban beszerezhető, nyílt forráskódú Linux rendszereket.

A költségérzékeny államigazgatási szektorban erős Synergon részéről így érthető a megállapodás várható hozadéka. Hazánkban az Országos Igazságszolgáltatási Tanács Hivatala, a Közlekedési Főfelügye-



let és a Magyar Szabadalmi Hivatal jár elől a Linux alkalmazásában. A megoldás Európa-szerte is egyre terjed, különösen a közgazdaságban és néhány nagy cégnél (Audi, PSA Peugeot Citroën, Siemens Medical, Deutsche Bahn). Az Európai Bizottság egy friss jelentése szerint 2010-re az unió összes IT szolgáltatásának 32 százaléka nyílt forráskódú szoftverekre fog épülni. ●

## INFO.TECH

**Robbantott a Vista** a PC-piacon. A Current Analysis elemző cég adatai alapján a Vista bejelentése után, a január és február fordulójára eső héten, 67 százalékkal több számítógépet adtak el Amerikában, mint egy évvel korábban. Az idei megelőző héthez viszonyítva a növekedés 173 százalékos volt.

**Naponta négyezer milliárd** kérés végrehajtására is alkalmassá kívánják tenni 2010-re a Verisign cég a szervereit. A nagygépek a világháló legfontosabb szerverei közé tartoznak, amelyek a .com és a .net végződésű címekre érkező kéréseket irányítják.

**Érintőképernyős mobiltelefon** mutatott be a Samsung és az LG is. Mindkét kutyú – akárcsak a Sony Ericsson már kapható újdonsága – ellenele lehet a tavasszal érkező Apple mobiltelefonnak, amelynek szintén képernyőn megjelenő gombjai lesznek.

**Mintegy 200 ezer vírus** és egyéb infokártevő megjelenését jósolja az Idén Art Coviello, az RSA Security elnöke. Szerinte egy-két éven belül eltűnnek a csak egy-egy részterületre koncentrált infobiztonsági cégek.

**A valódi e-mailnek látszó** levélszemét immár Magyarországon is hódít. Ismerve, hogy egy régi ismerősként bemutatkozó „kedves” lány különböző weboldalak látogatására biztat.




**Legyen CsmagPONT Partner!**

A GLS CsmagPONT rendszerének bővítéséhez innovatív, ügyfélorientált vállalkozókat keres.

Ha az Ön üzlete frekvenciáit, jól megközelíthető helyen van és nyitott shop-in-shop jellegű üzleti potenciál átvételére, vegye fel velünk a kapcsolatot!

E-mail: [parcelshop@gls-hungary.com](mailto:parcelshop@gls-hungary.com)  
További információt a [www.gls-hungary.com](http://www.gls-hungary.com) honlapon talál.

