

A felhasználói viselkedés vizsgálata (1. rész)

Bevezetés, avagy mire jók a Nagy Testvér módszerei?

Kémkedni ugyebár nem szép dolog, tehát egyáltalán miért figyeljük meg a felhasználókat? Napjainkban a világháló (World Wide Web) egyik legjellemzőbb problémája az információbőség. A felhasználók ugyanakkor a lehető legegyszerűbben szeretnének hozzájutni az általuk igényelt tudáshoz. A legtöbb honlapot a megrendelő, vagyis egy cég, egy intézmény, vagy egy kormányzati szerv igénye szerint készítik el. Gyakori ugyanakkor, hogy a megrendelő által megfogalmazott elvárások nem igazodnak a végfelhasználók, külső szemlélők elvárásaihoz. Általában nem ezeket felmérve tervezik meg a portálokat. Ez pedig azt jelenti, hogy bár a honlap híven tükrözi az adott szervezet belső struktúráját, egy új látogató számára ez a rendszer tökéletesen ismeretlen, és így az információ elérése számára igen nehézkes lesz.

Persze a látogatók igényei is folyamatosan változnak, s ilyenkor egy jó honlapnak ezt követnie kell. Ehhez pedig szükség van a látogatók viselkedésének, szokásainak folyamatos figyelésére, az adatok elemzésére, valamint a kapott eredmények „visszaforgatása”, viselkedési modellek alkotására.

Nézzünk egy példát!

Képzeld el, hogy van egy a cégünk termékeit bemutató portálunk. A termékszerkezetünk olyan, hogy az egyes árucikkek önállóan is „megállják a helyüket”, de együtt is jól használhatóak. Tegyük fel mondjuk, hogy asztalt és széket gyártunk.

Az első kérdés, ami eszünkbe jut nyilván az lesz, hogy melyek a leggyakrabban megtekintett termékleírások, vagyis mit keresnek leginkább a látogatóink. Az ilyen kérdésekre egészen egyszerűen választ kaphatunk, ha a webkiszolgáló eseménynaplóján lefuttatunk egy keresőprogramot. Rögtön látszik, hogy az asztal terméklapját százszor töltötték le, míg a székét ezerszer. Ez azonban csak az egyik lehetséges kérdés és az egyik lehetséges válasz. Az így kapott számszerű adatokból ugyanis nem lehet például olyan következtetéseket levonni, amelyek megkönnyítenék annak eldöntését, hogy egy adott termék oldalán milyen más áruinkra utaló szöveget célszerű elhelyezni. Ehhez „rejtett összefüggéseket” kell felfedezni. Ha tudjuk, hogy a zöld asztalt kereső látogatóink nagy része megnézi a zöld székek kínálatát is, akkor az asztal terméklapján egy közvetlen utalást elhelyezve a zöld székeinkre várhatóan jelentős mértékben megnő azok terméklapjainak forgalma. Ráadásul ezzel a könnyítéssel a vásárlóknak is örömet szerezünk, hiszen nem kellett keresgélniük.

Természetesen ennél sokkal összetettebb összefüggésekre is kíváncsiak lehetünk. Érdekes, a teljesítmény optimalizálása végett vagy marketingstratégiai célokra felhasználható adatokat nyerhetünk például az egyes oldalak közötti „átktintási arányokból”, az egyes lapokon eltöltött időkből, vagy a leggyakoribb útvonalakból.

Az összefüggések és számszerű adatok folyamatos figyelésével idejében felfigyelhetünk és felkészülhetünk a látogatók viselkedésének változására, vagy lemérhetjük egy-egy reklámkampány, vagy a marketingmunka hatékonyságát is.

Honnan és hogyan lesz adatunk?

A felhasználói viselkedés modellezése alapvetően két részből áll: az adatgyűjtésből és az elemzésből. A gyűjtés során tetszőleges eljárással olyan adatokat akarunk előállítani, amelyek kellően részletesek ahhoz, hogy az elemzést végző programok használható modelleket alkothassanak belőlük. Mivel csak az „átlagos látogató” paramétereire vagyunk kíváncsiak, azaz nincs külön-külön szükségünk minden egyes látogató viselkedésének a leírására, a megszemélyesítést lehetővé tévő adatokat csak az egyes felhasználók elválasztására használjuk. (Ez tipikusan hálózati címet és felhasználói nevet jelenti.) A kiszolgáló bizonyos beállításaival megoldható a személyes adatok teljes kizárása is.

Az adatgyűjtést végző kódot beépíthetjük a honlap programjába, futtathatjuk a kiszolgálón vagy az ügyfélen. Utóbbi esetben lehet ez egy böngésző-kiterjesztés vagy különálló program is. A honlap programjának módosítása munka- és költségigényes, az ügyfél oldalán végzett megfigyelés viszont adatvédelmi problémákat vet fel. A kiszolgálón végzett adatgyűjtés ezzel szemben a honlaptól és az ügyfélől is függetleníthető, de ennek a módszernek is megvannak a maga problémái. Az egyes módszerek előnyeit és hátrányait a sorozat következő tagjában fogom részletesen elemezni.

Az összegyűjtött adatok sok felesleges információt is tartalmazhatnak, ezért az adatsorokat minden esetben az adott portálra jellemző módon meg kell tisztítani. Mivel a valóságos információt néha több bejegyzés összefüggése (egyidejű megjelenése) jelenti, szükség lehet a honlapra jellemző kapcsolatok és összefüggések alapján egy további előfeldolgozási lépésre is. Előfordulhat aztán, hogy több, látszólag eltérő bejegyzés valójában azonos oldalt jelent. Az ilyen esetek kezelésére is be kell építeni a feldolgozóba egy szűrő és átalakító alrendszer.

Elemzés

A modellalkotásnak igazodnia kell a felhasználók viselkedésére vonatkozó adatok begyűjtésének módszeréhez. Az elkészülő modelleket alapvetően befolyásolja, hogy az egyes felhasználókat milyen módon sikerült megkülönböztetni, ez pedig erősen függ az adatgyűjtési módszertől. Akár az ügyfél oldalán gyűjtjük az adatokat, akár a a portál motorjába építettük be a szükséges kódot, a látogatók elkülönítése tulajdonképpen egész egyszerűen megoldható, hiszen csak arra kell figyelni, hogy az adott programban nyilvánartsuk, hogy most egy másik felhasználó használja a portált. A portálba beépített módszer esetén ez egy beléptetési (regisztrációs) folyamattal könnyedén megoldható, ügyféloldali gyűjtés esetén pedig szintén a programba épített azonosítási folyamat biztosítja a megfelelő elkülönítést.

A kiszolgálón begyűjtött adatok alapján általában jóval nehezebb a felhasználókat megkülönböztetni. Ennek az az oka, hogy a webkiszolgáló alapesetben nem tárol semmiféle, az egyes felhasználók elkülönítésére alkalmas adatot. További nehézség, hogy az esetleges megkülönböztetés alapjául szolgáló adatok (IP cím, felhasználói név, dátum) felhasználhatóságát az átmeneti táruk és tűzfalak is jelentősen rontják, hiszen ezek egy adott alhálózat valamennyi tagjának kéréseit azonos adatokkal küldik el a webkiszolgálóhoz.

A látogatók megkülönböztetését a kiszolgáló oldalán jelentősen megkönnyíti, ha egy megfelelően megírt modul segítségével minden egyes látogatóhoz egyedi munkamenetet és egyedi azonosítót rendelünk. Erről részletesen a következő cikkben, az adatgyűjtési módszerek ismertetése kapcsán fogok írni.

Röviden tehát, ha csak a legegyszerűbb, a kiszolgáló oldalán az eseménynaplóból kigyűjtött adatokat tudjuk használni, az a legfontosabb, hogy valamilyen módon az egyes látogatókat el tudjuk különíteni. Pontosabban fogalmazva azt kell megoldanunk, hogy a független oldallekéréseket valamilyen módon egyedi látogatói szakaszokba rendezzük. A módszer lényege, hogy az eseménynapló bejegyzéseiben szereplő IP cím, időbélyeg és minden egyéb, a felhasználóra jellemző adat (ilyen például a böngésző típusa) segítségével az egy meghatározott időtartamba eső, azonos helyről érkező kéréseket egy felhasználó látogatásának tekintjük. A megfelelő időtartam megválasztása nem könnyű. Érdekes több különböző intervallumhosszal elkészíteni a modelleket, hogy a látogatók viselkedési mintázatának megfelelőt kiválaszthassuk. Általánosságban elmondható, hogy a harminc perces intervallum a legjellemzőbb, tehát ezt tekinthetjük kiindulási alapnak. Egy rövid híreket közlő oldal ettől természetesen lefelé fog eltérni (gyorsabb az információ átadása), míg egy hosszabb publikációkat ismertető honlap felfelé növelheti az időkeretet, hiszen egy cikk végigolvasása hosszadalmas is lehet.

Ha sikeresen előállítottuk a felhasználókhöz rendelhető egyedi látogatói szakaszokat, máris készen állunk arra, hogy elkészítsük a modelleket. A modellkészítés során – amint azt már említettem – átlagosnak tekinthető látogatói viselkedésmintákat keressük. Ilyenek például a jellemző, sokszor előforduló oldalletöltési szekvenciák, vagy az oldalak bizonyos sorrendiségének gyakorisága. Ezek fontos jellemzők lehetnek egy portál üzemeltetése során.

Modellek

Alapvetően 5 modellt fogunk használni, de ezek természetesen az igényeknek megfelelően bővíthetők, illetve átalakíthatók. A legfontosabb modell, amelyet a többi modell elkészítésénél is felhasználunk a *felhasználói útvonalakat* adja meg. Ennél összegyűjtjük minden egyes látogatók útvonalait, és azok gyakoriságát vizsgáljuk. A kapott adatok fontos információval szolgálnak a portálon belüli mozgásokról. Megtudhatjuk például, hogy mely információk keresettek, kedveltek, illetve hogy egy adott helyre honnan érkezett a látogató, és merre távozott onnan.

A *„gyakori látogatói útvonalak”* modell a felhasználói útvonalakat aszerint vizsgálja, hogy az összes különböző bejárás hány felhasználónál található meg. Ennek a vizsgálatnak az eredményéből tehát azt tudhatjuk meg, hogy hány felhasználó járta végig weboldalunk pontosan ugyanazt a sorozatát. Az *„azonos weboldalak”* modell a *„gyakori látogatói útvonalak”* egy egyszerűsített változata. A különbség mindössze annyi, hogy ennél a modellnél nem számít az egyes meglátogatott honlapok egymáshoz képesti sorrendje, vagyis az eredmény az azonos oldalakat meglátogató felhasználók számát adja meg.

A későbbi felhasználás szempontjából fontos információ lehet az is, hogy az egyes weboldalakra honnan érkeztek a látogatók és merre mentek tovább. Ez a modell a *„be- és kilépő oldalak”* nevet viseli. Ezekből az adatokból következtetni lehet az egyes weblapok szerepére. Ha egy adott honlapra a látogatók többsége egy vagy néhány másik weblapról érkezett (ezt mutatja a belépő statisztika), akkor az a két weboldal összevonható, vagy akár helyettesíthető is a harmadikkal, hiszen a többség végül úgyis oda érkezett meg. A kilépő statisztika ennek a másik irányból történő megközelítése: ha egy weblapról meg lehet mondani, hogy melyik másik weboldalra léptek tovább róla a látogatók, akkor tisztázható az adott honlap szerepe. A két módszer nem feltétlenül ad azonos eredményt, hiszen a statisztikáknak mindig van hibája, illetve a megközelítés irány is más ebben az esetben. Az *„átkattintási, átlépési idők”* modellje a felhasználói útvonalakban összegyűjtött adatokra támaszkodik. Segítségével előállíthatók az egyes felhasználók útvonalaira jellemző karakterisztikus váltási idők, vagyis azok az időtartamok, amelyek az egyes weboldalak lekérése között elteltek. Kiszámítható a felhasználó által összesen a portálon eltöltött idő, illetve az egyes oldalak közötti jellemző átkattintási idők. Az adott tartalom, illetve az azt birtokló vállalkozás ismeretében mindezek az adatok egészen érdekes információkat tudnak szolgáltatni mind műszaki, mind marketing szempontból.

A következő rész tartalmából

A következő cikkben részletesen ismertetem az egyes adatgyűjtési módszerek előnyeit és hátrányait, majd egy saját fejlesztésű szkripten keresztül bemutatom a modellek elkészítésének menetét és lehetőségeit.



Beszédes Balázs (beszedes@ei.hu)

24 éves, az e-Média Informatikánál mérnök-informatikus. Hobbija a kerékpározás és a kirándulás.