

Az IMBITON projekt

Egy alkalmazás fejlesztésének tapasztalatai, gondolatok a továbblépés lehetőségeiről

- Média Laboratórium Kft. -

Az ÉMOP 1.3.1.-2012 kódszámú pályázat keretében támogatást nyert egy olyan informatikai fejlesztési projekt, amely hiánypótló a magyar piacon. A projekt eredményeképpen létrejött IMBITON alkalmazás az egyre bővülő, és növekedésével párhuzamosan egyre inkább áttekinthetetlen médiatérben való eligazodásban segíti a felhasználókat. Olyan, ha tetszik, médiafigyelő szolgáltatást ajánl ügyfeleinek, amely a számukra fontos témák – termékek, események, szolgáltatások, stb.- online médiumokban való terjedéséről ad számot.

Az Imbiton egy olyan program, amely lehetőséget nyújt az érdeklődés középpontjában álló „entitás” megítélésének követésére az internet segítségével. Mivel számos ilyen „entitás” értékelhető pl.: személy, jelenség, termék, szolgáltatás, cég...stb. ezért a későbbiekben ezekre csak termék/szolgáltatásként fogok hivatkozni. Tehát az Imbiton lehetőséget nyújt cégek számára, hogy egy új termék/szolgáltatás megítélését nyomon követhessék annak életciklusának megfelelően. A termék/szolgáltatás piaci megjelenése előtt a potenciális vásárlók véleménye és érdeklődése különböző internetes forrásokban már megjelenik, és ezek az Imbiton segítségével értékelhetővé válnak. A piacra kerüléssel az internetes forrásokból a vevői elégedettség és a termék/szolgáltatás egyes részei iránt támasztott követelmények vagy elégedetlenségek már a megjelenést követően értékelhetővé válnak. Az Imbiton nem csak a vevői véleményeket értékeli, hanem a források minőségét is pl.: internetes hírportál, blog post, tudományos cikk, forum. Ezáltal a cég figyelembe veheti, hogy számára mely információ forrás jelent prioritást és melyek azok, amelyek véleményformáló ereje nem annyira jelentős.

Problémák és tapasztalatok az alkalmazással kapcsolatban

Az értékelés tárgyának megismerése

Az első fontos tapasztalat, amelyet a munka kezdetén szereztünk, hogy amíg nem ismeri az értékelő a találatokat, addig nehezebben megy az értékelés is. Ahogy megismeri a terméket/szolgáltatást, a róla szóló hírek és benyomások is könnyebben az értékelő figyelmi fókuszába kerülnek az értékelendő szövegben. Ezáltal pontosabb és gyorsabb is lesz az értékelés.

Értékelés

A program a megítélést tekintve csak pozitív/semleges/negatív választást tesz lehetővé. A találat azonban nem ennyire egyértelmű. Gyakran mind pozitív és mind negatív vonatkozásokat is tartalmaz, ezeket semlegesként szoktam megítélni. Ezen felül a megítélés eléggé szubjektív. Természetesen vannak könnyen kialakuló vélemények, de vannak olyan találatok (főleg hírportáloknál), amikor a szöveg írója nem egyértelműen fejezi ki a véleményét, de lehet róla megérzésünk, hogy burkoltan a szövegben mégis csak pozitívan vagy negatívan áll hozzá a termékhez/szolgáltatáshoz.

Gyakran előfordul, hogy nem a termékre/szolgáltatásra nyilatkoznak negatívan, de egy negatív hírben hozzák szóba ezeket, ilyenkor nehéz megítélni, hiszen negatív a hír, de azért mégsem a termékre/szolgáltatásra vonatkozik közvetlenül. Pl.: A FireEye esetében találkoztam olyan példával amikor a FireEye egy alapos, mindre kiterjedő Reportot hozott nyilvánosságra, amelyben feltárta, hogy az internetes hacker támadások milyen jellegűek, milyen célpontokat szeretnek előszeretettel megtámadni és a támadás forrása hova tehető. A Reportban közölték, hogy a támadásoknak gyakran lehet politikai vonatkozása is, és a támadások a jövőben felszaporodhatnak. Ez negatív megítélésű, de mégsem közvetlenül a FireEye-ra vonatkozóan, azonban ha közvetetten nézzük, akkor a FireEye biztonsági programja lehet, hogy nem nyújt megfelelő védelmet az ilyen támadások ellen.

Kategóriák összemosódása

A kategóriák gyakori összemosódása szintén problémát jelent. A legszembeötlőbbben a News Portal és Blog Post esetén figyelehető meg. Ilyen odalacról a következő tapasztalatokat gyűjtöttem:

- . (i) Az igényes blogoknak ma már nagyon stílusos kinézete van. Az ilyen blogok gyakran reflektálnak napi hírekre, politikára, ezért nehezen felismerhető, hogy igazából blogról van-e szó.
- . (ii) A blogok gyakran nem viselik nevükben a „blog” kifejezést.
- . (iii) Ahogy a news portálok esetén is, a blogoknál is megfigyelhető, hogy többen szerkesztik, így nem egy személy képviseli a blogot. Gyakran előfordul, hogy igen tájékozott és képzett bloggerek működtetik az oldalt, így igazán hivatalos információ(nak tűnik) amit közzétesznek, sokszor megfelelő hivatkozásokkal társítva.
- . (iv) A news portálok próbálnak többféle igényt kielégíteni, ezért a portál gyakran tartalmaz többféle blogot is. Ez sokszor nagyobb részegységgé válik, így már a főmenü is tartalmaz „blog” fület.
- . (v) Személy szerint a blog egyik megkülönböztető jellegzetességének tartom, hogy a szerző a bejegyzésben gyakran első szám első személyben ír, illetve a bejegyzés egyértelműen tükrözi az író véleményét.
- . (vi) Azonban a news portálok esetén is gyakran előfordul, hogy a hír

szerkesztője nem feltétlenül objektív nézőpontból tudósít, így a személyes véleménye is előtérbe kerül.

Olyan kisebb átfedésekkel is találkoztunk, amikor például a blog postokhoz vagy news portálokhoz értékelő egységet társítanak (rating portal jellegzetességei bukkanak fel). Ezzel főleg termékek (Túró Rudi, Loving Earth) értékelésénél találkoztunk.

Az ilyen átfedések elég komoly formát öltenek akkor, amikor például a Secret Life of Walter Mitty esetében egy forumnak álcáztak torrent oldalra találunk. Az oldal linkje és címe is egy forumra utal, azonban megnyitva sorra ajánlja a letöltési lehetőségeket.

Ha terméket/szolgáltatást értékesítő cég oldalára kattintunk, akkor előfordul, hogy az oldalon található egy news portálhoz hasonló rész. Az oldal nagy része a termékkel/szolgáltatással kapcsolatos, de az ilyen cégek gyakran működtetnek egy kisebb hírportált az oldalukon. Ezek a hírportálok az adott termékkel/szolgáltatással kapcsolatos hírekkel foglalkoznak, esetleg a termék/szolgáltatás típusától függő szűk piaci területtel.

Az előző példákból látható, hogy a kategóriák összemosódása bármely kiválasztott találat esetén előfordulhat. Bár a kategóriákat az Imbiton segítségével mi választhatjuk meg, de a valóságban gyakran előfordul ezek keveredése.

Fontosabb véleményformáló kategóriák

A social mediát (facebook, instagram, pinterest) nem találatunk nagyon hangsúlyosnak a véleményformálásban. Az ilyen találatok általában az adott termék/szolgáltatás oldalát tartalmazzák. Tehát a cég által létrehozott oldalról van szó. Azok, akik like-olják az oldalt, és így napi szinten tudják követni azt, mondhatjuk, hogy nagy százalékban pozitív véleménnyel vannak az adott termékről/szolgáltatásról. Így az ilyen oldalt követők véleménye már előzetesen kialakult.

Fontosabbak a news portálok és blog postok szerepe, amelyek nagyobb hatással vannak a termékről/szolgáltatásról kialakuló véleményünre. A news portálok próbálnak objektív nézőpontot közvetíteni, a blog postok ezzel szemben már eléggé szubjektív és erős véleményű bejegyzéseket tesznek közzé.

További tapasztalatok

Egyes termékeknél megfigyelhető, hogy egy nagyobb hír vagy bejegyzés frappáns szövege a találatok nagy részénél ismét megtalálható. A Copenhagen Wheel esetében sokszor találkozunk ugyanazzal a szövegrészlettel, amelyben bemutatják a termék tulajdonságait. Mivel a blogokban nem nagyon volt hajlandóság az átfogalmazásra, ezért szép számban ismétlődött ugyanaz a szövegrészlet.

II. További fejlesztési irányok

A kifejlesztett alkalmazás további fejlesztési irányokat is magában rejt. Ennek következtében folyamatos innovációról beszélhetünk az IMBITON esetében. A teljesség igénye nélkül felsorolunk néhány lehetséges továbblépési irányt.

További termék ötletek...

A vízió lényege, hogy egy személyről vagy cégről, esetleg egy cég adott termékéről (továbbiakban: *entitás*) szóló állítások vagy vélemények megjelenését az interneten egy rendszer (1) észrevegye, (2) a terjedését figyelje, (3) esetlegesen az adott vélemény jövőbeli terjedését is előrejelezze. □A nagy webes keresőmotorok keresésenként néhány dollárcentért rendelkezésre bocsájták az API-jukat. Ezek az API-kon keresztül megfelelően paraméterezett keresésekkel felderíthetők az adott entitás említései. Ehhez előzetesen elő kell állítani az entitáshoz tartozó keresőkifejezéseket. Természetesen az entitások említéseiben a változást kell követni, tehát az első lekérdezés eredményét eltárolni, majd az újdonságokat észlelni. □A keresőmotorok programozói felülete csak az URL-eket adja meg, ahol a keresőszavak előfordulnak, a saját belső rendező algoritmusának megfelelően sorrendezve. Az URL-eket le kell tölteni, és feldolgozni, úgy, hogy gépileg is értelmezhető plain text keletkezzen belőlük, másrészt a weboldalak képe grafikusán is előállítható, ugyanúgy ahogyan a böngészőben megjelenik, egy emberi olvasó részére. A letöltött weboldalokról gépi vagy emberi módszerrel el kell dönteni, hogy valóban az adott entitásról szól-e. Ezt az információt (akár valódi, akár fals találatról van szó) el kell tárolni, hogy a következő keresésekkor az adott oldal újra feldolgozása elkerülhető legyen.

Termék ötlet #1: Az eddig leírt funkcionalitás önmagában is termék lehet. A rendszer szakadatlanul figyeli a nagyobb böngészőket, és meg tudja mondani, hogy hányadik helyen szerepel az entitás különféle keresőszavak megadása mellett, és ez a trend hogyan változik. Az entitás lehet akár egy weblap is, így azt kapja meg az ügyfél, hogy hogyan áll a keresőbeli optimalizálás (SEO) terén.

Példa ügyfél történet: Az ügyfél az XY nyelviskola, amely angol nyelvtanfolyamot hirdet Lágymányoson. Ez lesz az entitás, amelyhez az xnyelv.hu/angol weboldal tartozik. Az ügyfél kérdése az, hogy az

- “Angol nyelvtanfolyam Lágymányos(on)”
- “Angol tanfolyam Budapest (en)”
- “Angol tanfolyam olcsón” , stb

keresőkifejezésekre hányadik helyen fordul elő, és ez hogyan változik (például egy reklámkampány hatására). Erről napi jelentést is kaphat. A keresőkifejezések kitalálása az adott entitáshoz önmagában is egy tanácsadás jellegű termék opció

lehet. Opció lehet az is, hogy a konkurencia entitásait is figyeljük és egymáshoz hasonlítsuk grafikonokon. Még egyszerűbb opció weboldalak egy megadott körének figyelése (konkurencia szemmel tartása), amennyiben változik a weboldal képeinek lementése és elküldése. Külön opció lehet mindennek excel-be betölthető CSV-ben, egy adott intervallumra való letöltése is, amelyet az ügyfél például korrelálhat kereslet vagy bevétel információkkal. Az árképzéshez paraméterek lehetnek: keresőkifejezések száma, vizsgálat gyakorisága. A folyamat következő lépése a letöltött weboldalak tartalmi feldolgozása. Itt géppel támogatott emberi feldolgozásról van szó. A feldolgozás tipikusan arra irányul, hogy negatív vagy pozitív véleményt fogalmazzanak meg az entitásról, illetve arra, hogy egy adott oldalon talált üzenet megegyezik-e egy korábban látott üzenettel.

Termék ötlet #2: Egy nagyobb ügyfél, (pl. BMW, EOn, MTelekom, stb) entitásait annyit említik az interneten, hogy ha a változásokról jelentést küldenénk, akkor az túl gyakori lenne. Így eladható, hogy csak bizonyos, kiemelkedően pozitív/negatív említések, vagy csak olyan üzenetek, amelyek ismétlődnek (tehát trend van kialakulóban) kerüljenek jelentésre. Mivel ebben a munkafázisban mindenképp részt vesz ember, opció lehet például összefoglaló iratása is. Mi több, speciális utasításokat is adhat az emberi feldolgozók számára az ügyfél arról, hogy pontosan mi érdekl. Hosszabb távon a rendszerben felhalmozódhat annyi történeti adat az egyes termékekből, hogy modellt tudunk felállítani a különféle üzenetek terjedésére, ezáltal előrejelzést, korai figyelmeztetést adva az ügyfélnek.

Termék ötlet #3: a figyelmeztetésekhez előrejelzést is csatolunk, amely megbecsüli, hogy milyen valószínűséggel és milyen időbeli lefutással következhetnek be különféle folytatások.

Termék ötlet #4: Adja magát az ötlet, hogy a figyelési rendszert egészítsük ki beavatkozási rendszerrel: amennyiben az ügyfél erre részletes megbízást ad egy munkatárs hozzászólhat a fórumokhoz, információt adva, reklámozva, védve a mundért, stb.

A továbbiakban csak a Termék ötlet #1 és #2-t fogjuk kidolgozni, a #3, #4 jövőbeli terv lehet.

Kockázatok #1 és #2-vel kapcsolatban:

- Könnyebben és nehezebben kulcsszavazható entítások. Bizonyos entítások neve □ annyira egyedi, hogy triviálisan könnyű keresni őket. Más entítások neve annyira nem-egyedi lehet, ami nagyon megdrágíthatja, szélsőséges esetben meg is hiúsíthatja a figyelésüket.

- Túl késői figyelmeztetés: a keresőmotorok nem egyforma

gyakorisággal keresik fel a weboldalakat. Egy gyakran frissülő, nagy forgalmú híroldalt akár percenként, egy mellékes, kis forgalmú oldalakat akár csak 3 hetente. Erre megoldás lehet, hogy bizonyos oldalakat maga a rendszer szed le időnként (pl. konkurrencia weboldalai, akciós ajánlataik figyelése), de nyilvánvaló, hogy ezen oldalak köre nem lehet túl nagy.

- *Túl drága keresések:* amennyiben az entitás a kereséső találati listájának pl. 150000-ik helye környékén helyezkedik el, a keresés igencsak drága, esetleg kivitelezhetetlen lehet (mert pl. a keresőmotor 50-100 találatonként felszámol egy összeget), ráadásul értelme sincs sok, hiszen jól ismert, hogy az ügyfelek csak max 2-3 lapot néznek át a találati listában.