

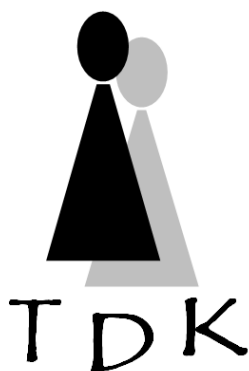


ÚJ SZÉCHENYI TERV

ANONIMITÁS AZ INTERNETEN

Tudományos Diákköri Dolgozat

Konzulens: Balogh Zoltán



Készítette: Oroszi Róbert
Gazdálkodástudományi kar
Mesterképzés
Gazdaságinformatikus szak
I. évfolyam

2012. március 26.

A BCE Közgáz Campus Tudományos Diákköri Konferenciáját a TÁMOP-4.2.2/B-10/1-2010- 0023 azonosítójú “A tudományos képzés műhelyeinek átfogó fejlesztése a Budapesti Corvinus Egyetemen” című projektje támogatja.



MAGYARORSZÁG MEGÚJUL



A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Tartalomjegyzék

1. Bevezetés	4
1.1. Anonimitás felhasználói oldalról	6
1.2. Anonimitás üzleti oldalról	6
1.3. A téma létjogosultsága, megéri vele foglalkozni?	6
2. Technológia	7
2.1. Mit lehet mérni?	8
2.2. Mit érdemes mérni?	12
3. Biztonsági	14
4. Üzleti	15
5. Összefoglalás	16

Rövidítésjegyzék

ActiveX

ActiveX

Browser Detection

Br.Det.

Feature Detection

Feat. Detection

HTML5

HTML5 egy webes szabványgyűjtemény, amelyet a webes fejlesztők, a böngészők készítői állítanak össze, és amely alapján adaptálják az új funkciókat

IndexedDB - Indexed Database API IndexedDB

JavaScript

JavaScript no mi ez lurkó?

webStorage

webStorage

1. fejezet

Bevezetés

Az internet penetráció növekedése (TODO hivatkozás) és az információs társadalom fejlődése miatt a felhasználók egyre több szolgáltatásból választják az online verziót, egyre több időt töltenek (TODO hivatkozás erről?) az internetre csatlakozva.

Az online szolgáltatások széles választéka, és az azokat használó felhasználók adatainak eltulajdonításáról (banki adatok, számlaszámok) rengeteg cikk, tudományos munka készült már, ez a dolgozat azt szeretné bemutatni, hogy az interneten tevékenykedve milyen, szinte láthatatlan nyomokat hagynak maguk után a felhasználók, ezeket kik és hogyan használják ki.

A dolgozat bemutatja, mind felhasználói, mind szolgáltatói szemszögből mire kell odafigyelni (a felhasználóknak mit érdemes elrejteni, a szolgáltatóknak mit érdemes monitorozni), hogy a lehető legkevesebb vagy éppen a legtöbb információhoz cseréljen gazdát.

A dolgozat végén pedig, egy, a felhasználói profil összeállítására alkalmas szoftver tervezését és megvalósítását ismerheti meg az olvasó.

1.1. Anonimitás felhasználói oldalról

Mivel az internet Magyarországon a rendszerváltás után jelent (TODO első szolgáltató), ezért a felhasználók tudása, és oktatása nem fejlődött az internet sebességével. Ezért sajnos az internetet böngésző felhasználók gyakran nincsenek tisztában, hogy milyen sok mindent elárulnak magukról egy-egy kattintással, elfogadnak olyan kéréseket, amelyeket el sem olvasnak, illetve megbíznak a weboldalakban.

Természetesen léteznek weboldalak, amelyek még a gyakorlott, az anonimitással teljesen tisztában lévő haladó felhasználókat is csapdába csalják.

A dolgozat megpróbál rámutatni, azokra biztonsági szempontokra, amelyeket szem előtt tartva a felhasználót sokkal kevesebbet fog elárulni magáról a böngészése során. Többek között a következő témákat érintve: privát böngészés, külső sütik, HTML5-, RIA veszélyei és lehetőségei.

1.2. Anonimitás üzleti oldalról

Az üzleti oldal természetesen teljesen más oldalról közelít az anonimitáshoz, egy weboldalnak tudnia kell monitorozni a felhasználóit, egy hírportál esetén releváns, célzott reklámokat kell tudni megjeleníteni, amelyhez szükséges egy minél pontosabb felhasználói profil felállítása.

A dolgozatban bemutatásra kerülnek azok a technikák, technológia lehetőségek, melyekkel a felhasználók minél könnyebben, pontosabban beazonosíthatók. Továbbá megvizsgálásra kerül a Facebook, melynek segítségével a felhasználói profil pontosítható.

1.3. A téma létjogosultsága, megéri vele foglalkozni?

Az anonimitás olyan webes alkalmazásoknál, ahol van regisztráció - és kötelező is regisztrálni (email alkalmazások, közösségi média) - természetesen nem kap komoly hangsúlyt, ugyanis tisztában vannak a felhasználóik adataival.

Azonban online hírportáloknál, keresőmotoroknál, ahol a tartalom ingyenesen elérhető és az elsődleges bevétel a reklámokból van, ott kimondottan fontosat szerepet kap a felhasználói profilok felépítése.

Az analitika rendkívül fontos ilyen weboldalak esetén, azonban nem képes arra, hogy megmondja az oldalra látogató felhasználókból, hogy volt-e már az oldalon, illetve mennyit töltött és milyen tartalmak érdeklik.

2. fejezet

Technológia

A felhasználó azonosítása hálózati kapcsolaton keresztül történő kommunikáció során figyelhető meg leginkább, ezért a technológia áttekintés a webes technológia lehetőségeire fog fókuszálni.

A webes technikák fejlődése egy inkább biztonságossá teszi a böngészést, amellet az új funkciók bevezetésével egyre több lehetőséget kínál a felhasználók azonosítására.

A modern böngészők egyre több mérési lehetőséget kínálnak a következő pontban összefoglalásra kerül, hogy mik ezek, milyen adatokat tudhatunk meg ezek segítségével a felhasználóról. Ezután pedig a lehetőségeket kerülnek mérlegelésre technológia szempontból, hogy melyek azok az adatok, amelyek valóban információkat is tartalmaznak.

2.1. Mit lehet mérni?

Szerver oldali mérés. A böngészők már a weboldalak lekérések elküldésekor is számottevő mennyiségű adatot küldenek el a szervereknek. A 2.1 ábrán látható HTTP kérésből jól látható, hogy felhasználó milyen böngészőt használ (*User-Agent*), amely tartalmazza az operációs rendszer típusát (*Mac OS X*), a böngésző verziószámát (*Chrome 17*), a böngésző kompatibilitását (*Mozilla 5.0, AppleWebkit, Safari*). Továbbá fontos megemlíteni az fejlécben lévő sütiket (*Cookie*), amelynek segítségével, a webszerverek azonosítják a felhasználókat.

A 2.1 ábrán látható sütik kiválóan bemutatják a webes anonimitás/azonosítás fontosságát, ugyanis az `__utma` és az `__utmz` sütiket a Google Analytics webanalitikai szoftver használja felhasználók azonosítására.

A kérés fejlécéből még megállapíthatóak, olyan adatok, mint a felhasználó operációs rendszerének/böngészőjének nyelvi beállítása (*Accept-Language*) vagy éppen a használt karakterkódolás (*Accept-Encoding*).


```
Accept: text/html, application/xhtml+xml, application/xml;q
    =0.9, */*;q=0.8
Accept-Charset: UTF-8, *;q=0.5
Accept-Encoding: gzip, deflate, sdch
Accept-Language: en-US, en;q=0.8
Cache-Control: max-age=0
Connection: keep-alive
Cookie: __utma
    =176087398.927465515.1331063740.1331063740.1331063740.1;
    __utmz=176087398.1331063740.1.1.utmcsr=(direct)|utmccn
    =(direct)|utmcmd=(none);
Host: www.uni-corvinus.hu
User-Agent: Mozilla/5.0 (Macintosh; Intel Mac OS X 10_7_2)
    AppleWebKit/535.11 (KHTML, like Gecko) Chrome
    /17.0.963.79 Safari/535.11
```

2.1. ábra. Egy HTTP kérés fejléce

Kliens oldali mérés. Azonban, a legtöbb adatot a böngészők nem küldik el a kéréskor, hanem JavaScript segítségével lehet kinyerni a böngészőből. A HTML5 ajánlások bővülésével a JavaScript nyelv segítségével, egyre közelebb lehet kerülni az operációs szintű funkciókhoz, természetesen a megfelelő biztonsági korlátozások mellett.

A HTTP fejlécben látható adatok, mind elérhetőek JavaScript segítségével is (ez alól kivételt jelentenek a biztonságos címkével ellátott sütik). De milyen plusz adatok érhetőek kliensoldalról?

Böngészőképességek. A felhasználó böngészője által implementált képességek, amelyből kinyerhető, hogy a felhasználó milyen a böngészők használ. Természetesen mindezt elárulja a HTTP fejlécben található *User-Agent* is, azonban a HTTP fejléc a legtöbb böngészőben kézzel is módosítható, tehát az ilyen detektálásból (*Browser Detection*) gyakran fals-pozitív vagy fals-negatív azonosítás születhet, míg a JavaScript alapú megoldásból (*Feature Detection*) mindig pontos születik.

Bővítmények. Olyan fontos - és hosszútávon állandó - adatok is kiolvashatók a böngészőkből, mint például a felhasználó által telepített bővítmények (pluginek). Ebben a kategóriában a következő elemek fordulhatnak elő:

- **Adobe Flash**

Az Adobe Flash bővítmény pontos neve, és verziószáma, videók lejátszásához, streameléshez, animációkhoz használják.

- **Java**

A böngészőbe telepített Java bővítmény verziószáma, komplex hálózati adatfolyamok kezelésére, magas biztonsági szintet megkívánó alkalmazások (például internetbankok) futtatására használják.

- **Silverlight**

A Microsoft Silverlight bővítménye, videók lejátszásához, és adat streameléshez használatos.

- **PDF olvasó**

Böngészőbe épített PDF olvasó, ez lehet alapértelmezett böngésző része, vagy külső beépülő bővítmény is lehet (például Adobe Reader).

Képernyőadatok. A kliensoldalon hasznos adatok érhetőek JavaScript és CSS segítségével, a felhasználó által használt monitor felbontásáról, színmélységéről.

Sütik. A webes alkalmazások sütikben (cookie) tárolják a felhasználóhoz kapcsolódó információkat, ilyenek lehetnek például a bejelentkezéshez, a legutóbbi látogatáshoz tartozó adatok. Ezek böngészőkhöz vannak kötve, tehát ha a felhasználó egy másik böngészőt indít el az internetezésre használt eszközén, vagy másik eszközről éri el az oldalt akkor nem lesznek elérhetőek.

A már említett és a 2.1 ábrán látható módon a sütikben az analitikai szoftverek egyéb, a felhasználói azonosítást megkönnyítendő adatot is elhelyeznek.

A HTML5-ben már nem csak sütikben van lehetőség adatok tárolására (a sütik mérete négy kilobájtra van limitálva (<http://support.microsoft.com/kb/306070>)), hanem *webStorage*-ban is, amely már strukturáltabb, és nagyobb adathalmaz tárolását teszi lehetővé (<http://www.w3.org/TR/webstorage/>) illetve az *IndexedDB*, amely *webStorage* méretbeli lehetőségeivel bír, viszont növelési lehetőséggel (bármeddig növelhető méret(<https://developer.mozilla.org/en/IndexedDB>)) (<http://www.w3.org/TR/IndexedDB/>), és még komplexebb adatformátumok tárolására is fel van készítve.

Lokalizációs adatok. A felhasználó földrajzi pozíciójának meghatározása nem egyszerű dolog, a HTML5 lokalizációs interfészének bevezetése előtt, mindössze az ún. *fordított helymeghatározás* (reverse geocoding) létezett, amely a felhasználó IP címe alapján próbálta megállapítani a kliens pozícióját.

Azonban, míg az Egyesült Államokban az IP cím tartományok kiosztása régióként történt, addig Magyarországon nem volt ilyen szabályozás, tehát ezzel a módszer Magyarországon nem lehetséges pontosan meghatározni a felhasználó tartózkodási helyét.

Viszont az említett HTML5 lokalizációs interfésze (geoLocation) lehetővé teszi, a felhasználó pozíciójának pontosabb meghatározását, amely következőképpen történhet:

- **A környezet vezeték nélküli hálózatok segítségével**

Ebben az esetben (Google Chrome és Firefox böngésző esetén) elküldésre kerül a Google térkép és lokalizációs szervereinek a felhasználó készüléke körül elhelyezkedő privát, és publikus vezeték nélküli hálózatok azonosítója (SSID), egyéni fizikai címe (MAC cím) illetve a jel erőssége. A válaszüzenetben visszaküldésre kerülnek a szélességi és a hosszúsági fokok.

- **Mobiltelefon tornyok**

Amennyiben a felhasználó készüléke képes mobiltelefon tornyokhoz kapcsolódni (mobiltelefon, táblagép, 3G modem), akkor a tornyok pozíciója is pontosítja a pozíciót.

- **GPS**

A specifikáció lehetőséget ad arra is, hogy ha felhasználó készülékében található GPS eszköz (mobiltelefon, táblagép, de akár laptop is), akkor a kért pozíció pontosságától függően ez is használatra kerül.

- **Fordított helymeghatározás**

Abban az esetben, ha az adott készülék sem mobiltelefon tornyokra nem képes csatlakozni, vezeték nélküli hálózat sincs a környezetében és GPS eszközzel sem rendelkezik, a fordított helymeghatározás ilyenkor is használható, hiszen internetkapcsolattal rendelkeznie kell a webes tartalmak eléréséhez.

A HTML5 lokalizációnál fontos kiemelni, hogy a böngésző alapértelmezetten nem teszi elérhetővé a felhasználó pozícióját, hanem egy megerősítést kér, amelyet elfogadva lesz csak elérhető az adat.

Egyéb mérhető adatok. Lehetőség van még a böngészőkben olyan egyéb adatok mérésére, melyek nincsenek benne a specifikációban illetve a beépülő bővítmények segítségével érhetőek el.

Ilyenek lehetnek például az Adobe Flash segítségével elérhető operációs rendszerre telepített betűtípusok vagy az adott készülékben megtalálható vagy készülékhez kapcsolt multimédiás eszközök, mint például a webkamera.

Továbbá a nem standard megoldások közé tartozik, az Internet Explorerben megtalálható ActiveX komponensek, melyekkel akár az operációs rendszer szintjén lehet futtatni nem webre fejlesztett szoftvereket, amelyek természetesen szinte mindent elérhetnek - az asztali programokhoz hasonlóan. Az ActiveX komponensek hátránya, hogy csak Internet Explorerben és Windows platformon érhetőek el.

2.2. Mit érdemes mérni?

A 2.1 pontban bemutatásra kerültek a legfontosabb mérhető adatok, azonban fontos kiemelni, hogy nem érdemes minden adattal foglalkozni, de vajon melyek azok az adatok, amelyek valójában információt is tartalmaznak?

A HTTP fejléc. A HTTP fejléc módosítása nem tartalmaz elegendő információt ahhoz, hogy érdemben megérje vele foglalkozni, és az elküldött adatok legtöbbje módosítható, korlátozható illetve ki is egészíthető a telepített böngészőkben (ez történhet kézzel vagy a bővítmények, kiegészítők által).

A böngésző neve, verziója. Mint említésre került a 2.1 bekezdéseiben a böngésző neve, kompatibilitása, verziószáma mind kliens-, mind szerveroldalról elérhető, azonban ennek mérése hosszútávon nem profitábilis. Ennek oka, hogy 2010 júliusában a Google Chrome - szakítva a korábban megszokott böngészőverziók frissítési szokásaival - havi kiadási ütemtervre váltott (<http://blog.chromium.org/2010/07/release-early-release-often.html>), majd őt követte a Mozilla is, amely pont egy évvel később hozta meg hasonló döntését (<https://blog.mozilla.com/futurereleases/2011/07/19/every-six-weeks/>).

Tehát az ilyen gyors ütemű verzióváltások mellett, ez az adat sem szolgálhat nagy segítséggel.

Sütik. A sütik (és egyéb kliensoldali adattárolási formák) kihasználása rendkívül fontos dolog, ugyanis a felhasználóról korábban eltárolt információk, kliens és szerveroldali közötti megosztása ezeken keresztül történhet meg.

Lokalizációs információk. A felhasználó pozíciójának elmentése is kiemelten fontos eleme a mérésnek. Természetesen, a minél pontosabb méréshez szükség van a felhasználó beleegyezésére, azonban ha az adott oldal képes rávenni a képernyő túloldalán helyet foglaló internetezőt, hogy erősítse meg megosztási szándékát (például nyereményjátékba való bekerülés lehetőségével), akkor mindenképpen az egyik legfontosabb információvá lép elő.

Bővítmények. Érdemes monitorozni a bővítmények verzióit, illetve a különböző bővítmények jelenlétét a böngészőkben.

A 2.1 pontban megemlítsre került, hogy a bővítményekkel olyan adatok is elérhetőek, melyek a webes kliensoldali technológiákkal (JavaScript, CSS) nem, vagy csak részlegesen. Viszont ezek a bővítmények nincsenek jelen minden böngészőben, vagy éppen minden platformon, de akkor mégis miért szükséges velük foglalkozni?

Éppen az a plusz információ teszi őket fontossá, hiszen ezen apró pluszok segítségével válik lehetővé a felhasználók beazonosítása.

3. fejezet

Biztonsági

4. fejezet

Üzleti

5. fejezet

Összefoglalás