Metodologija stručnog i naučnog rada DOI: N/A

Pristup podacima sa socijalnih mreža putem javnih API-ja

na primeru Facebook-a

Nikola Divić 1120/2013

Matematički fakultet
Studentski trg 16, 11000 Beograd
mi10131@alas.matf.bg.ac.rs

Sažetak: Kratak pregled mogućnosti pristupa jednom ogromnom skupu informacija.

Ključne reči: API, Facebook, OAuth.

1. Uvod

Prema nezvaničnoj statistici sprovedenoj 2012. godine 58% od ukupne internet populacije koristi (ili barem ima otvoren profil) na nekoj od mnogobrojnih socijalnih mreža (10) . Po istraživanju, a i iz iskustva, znamo da prednjače četiri najpopularnije mreže: Facebook, LinkedIn, Twitter i Google+, gde Facebook ubedljivo vodi sa udelom od 56%.[1]

Ogromna količina informacija, linkova i multimedijalnog sadržaja od strane stotina miliona korisnika biva svakodnevno podeljena i sačuvana u "oblaku". Facebook-ovo rukovodstvo, ako im je za verovati, se hvali cifrom od 757 miliona aktivnih korisnika dnevno. (11)

Na taj način deo svih tih podataka je dostupan i "programski", razvijaocima aplikacija čija bi se funkcionalnost temeljila na ovoj ogromnoj bazi znanja. Jasno je da postoje ograničenja radi očuvanja privatnosti i prava korisnika, barem sa ove strane. Neću se baviti time šta velike kompanije rade sa našim podacima iza zatvorenih vrata. U navedenom kontekstu može se reći da sve od navedenih velikih socijalnih mreža čuvaju naše podatke od neželjenog pristupa od strane šire populacije koristeći različite mehanizme autorizacije o kojima će biti reči kasnije.

Ova vrsta otvorenosti podržava koncept semantičkog veba (Semantic web), transformisanje postojećeg veba koji sačinjavaju većinom nestruktuirani ili slabo struktuirani dokumenti u mrežu podataka ("Web of data") čija je priroda takva da podaci mogu lako da se razmenjuju između različitih aplikacija.

Ovaj rad je koncipiran tako da će prvo biti izložene neke od mogućnosti upotrebe ovakvih podataka zatim će biti ukratko objašnjena opšta terminologija relevantna za temu da bi na kraju bila predstavljena konkretna upotreba API-ja na primeru Facebook-a, pošto je mehanizam pristupa jako sličan i za ostale socijalne mreže. Kao zaključak biće napisano par rečenica o privatnosti.

[1] Pretpostavljam da se u zadnje dve godine statistika promenila, i da je Twitter preuzeo deo FB-ovih korisnika

Statistike [2]

Socijalna mreža	Aktivnih korisnika	Procenat (2012)
Facebook	1,23 milijarde	56%
LinkedIn	200 miliona	14%
Twitter	500 miliona	11%
Google+	540 miliona	9%
Ostalo	-	10%

2. Upotrebna vrednost

Od kad su velike socijalne mreže "otvorile vrata" ka svojim bazama podataka nastao je veliki broj web strana, desktop i mobilnih aplikacija koje svoju celokupnu funkcionalnost, ili makar deo nje, zasnivaju na podacima o korisnicima socijalnih mreža i njihovim aktivnostima.

Proučavanje nekog određenog API-ja zahteva uloženo vreme ali je često upotrebna vrednost rezultata ogromna. Mogućnosti su proporcionalne kreativnosti razvijaoca. Neću se preterano baviti time šta sve postoji, jer bi mi to oduzelo previše vremena i nije u domenu rada, a neću ni izlagati moje ideje, jer da ih imam prvo bih ih realizovao. Pokušaću da samo što bolje oslikam situaciju.

Autor je primetio da postoji više aplikacija na Facebook-u za jednostavnu obradu, dodavanje efekata i druge različite modifikacije fotografija koje korisnik ima na svom profilu. [3] Pošto su fotografije jedan od dominantnih sadržaja koje ljudi postavljaju, očigledna je prilika da se ljudima pruži jedna takva usluga, da lakše na istom mestu i u isto vreme mogu da urade više stvari. Jedan takav primer je `https://apps.facebook.com/picmonkey/`.

Popularne su aplikacije koje koristeći istoriju korisnikovih aktivnosti mogu da predlože novi sadržaj koji je u skladu sa korisnikovim interesovanjima. Naravno da bi privatnost bila očuvana sam korisnik mora da potvrdi da dozvoljava dotičnoj aplikaciji da se tim informacijama služi. Iskustvo govori da često ljudi koji se najviše žale na ugrožavanje njihove privatnosti ne čitaju tekst obaveštenja pre nego što potvrde da prihvataju da svoju "on-line" personu podele sa nepouzdanim aplikacijama od sumnjivih autora.

Analiza socijalnih medija se definiše kao prikupljanje i analiza informacija sa socijalnih mreža (i drugih veb strana sa socijalnom komponentom) radi efikasnijeg poslovnog odlučivanja. Postoje softverski alati za analizu nestruktuiranih podataka kao što su Facebook statusi ili Tweet-ovi. U kontekstu poslovanja i ovakve analize često su korisni i alati koji mogu da agregiraju podatke iz više izvora na jednom mestu, na primer prikupljanje podataka sa Google Analytics servisa, Twitter-a, email-ova i Facebook-a može stvoriti jednu bogatu bazu znanja koju može pospešiti buduću marketinšku kampanju ili uopšteno funkcionisanje preduzeća.

[2] Procenti i brojevi korisnika se ne poklapaju zbog nepoklapanja zvaničnih izjava i statističkih istraživanja

[3] Ne mislim na Instagram, pošto je on prvenstveno samostalna aplikacija koja ima mogućnost postavljanja slika slikanih mobilnim telefonom na socijalne mreže.

Postoji i primeri zabavnog sadržaja koje koristi nešto od vaših ličnih podataka da bi izgradilo jedno "personalizovano" iskustvo kakvo nije uobičajeno i ne bi bilo moguće da ovakve informacije nisu na raspolaganju. Kao zanimljiv primer, iako rizikujem da ispadnem neozbiljan, naveo bih `http://www.takethislollipop.com/`. Pogledati na sopstvenu odgovornost. :)

Zaključak je da u ovom domenu nema šablona, već postoje samo dobre ideje.

3. Terminologija

3.1. Šta je to Web API?

Opšti pojam API (application programming interface) u računarstvu se može definisati kao opis metoda, funkcija ili rutina za interakciju sa nekom softverskom komponentom ili sistemom.

U kontekstu veb programiranja API se najčešće definiše kao skup HTTP zahteva, uz definiciju strukture odgovora, koji su najčešće u XML (Extensible Markup Language) ili JSON (JavaScript Object Notation) formatu. Uprošćeno, definicija API-ja jedne veb aplikacije nam govori kako da formulišemo upite koje prosleđujemo pomoću HTTP GET i POST metoda toj aplikaciji i u kom formatu će nam biti vraćen odgovor. Zahvaljujući ovom mehanizmu možemo da postavljamo upite nad podacima koji su nam raspoloživi na socijalnim mrežama i da dobijemo dobro formatirane odgovore.

Na najjednostavniji način funkcionalnost API-ja možemo isprobati ručno, unoseći u polje za adresu u browser adresu oblika `VEB_ADRESA_APIJA/UPIT` što nema neki praktični značaj. Isprobati na primeru `http://graph.facebook.com/facebookDevelopers` , kao odgovor dobija se JSON sa poljima definisanim za Facebook developers stranu, koja je javna i nije potreban token za pristup (biće reči o tokenima dalje u tekstu) da bi se video njen sadržaj.

Praktična upotreba ovoga je što ovako nešto možemo uraditi programski šaljući HTTP GET i POST zahteve i onda vršiti potrebnu obradu nad rezultatima.

3.2. JSON

JavaScript Object Notation (8) je popularan i jednostavan format za razmenu podataka. Sintaksa je zasnovana na podskupu programskog jezika JavaScript i jednostavna je za ljudsku upotrebu kao i za računarsko parsiranje. Ova notacija je u potpunosti nezavisna od programskih jezika ali koristi konvencije koje su specifične za C-familiju jezika.

JSON prepoznaje dve osnovne strukture:

- Skup parova ključ-vrednost, struktura koja se u objektno orijentisanoj paradigmi može poistovetiti sa objektom, slogom, strukturom, heš tabelom ili asocijativnim nizom.
- Uređena lista vrednosti, struktura koja se u objektno orijentisanoj paradigmi može poistovetiti sa nizom, vektorom, listom ili sekvencom.

Primer jednog jednostavnog JSON dokumenta:

```
{
 "first_name": "Nikola",
 "last_name": "Divic",
 "birthday": "20/01/1992",
 "parents": [
  {
       "firstName":"Dragan",
       "lastName":"Divic"
  },
  {
       "firstName":"Ljilja",
       "lastName":"Divic"
  }
  ]
}
```

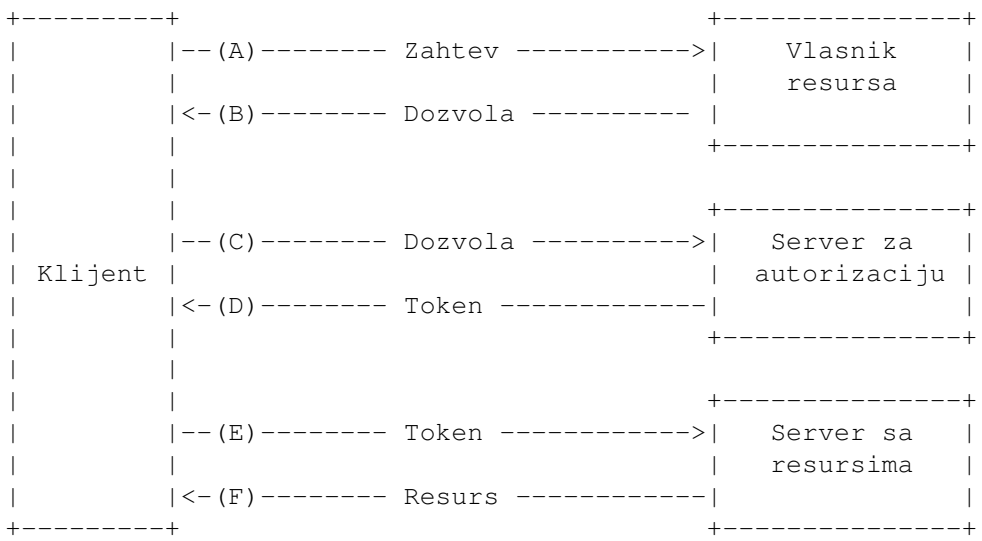
3.3. OAuth

OAuth je protokol otvorenog koda za bezbednu autorizaciju na jednostavan i standardizovan način. (5)

Sve socijalne mreže spominjane u ovom radu u pozadini koriste ovaj protokol za autorizaciju svojih korisnika i zbog toga je vredno napisati nekoliko rečenica o njemu.

OAuth omogućava korisniku (User) da dozvoli pristup njegovim privatnim podacima na jednom veb sajtu (koji se naziva Service Provider) drugom veb sajtu (koji se naziva Consumer). Ovo je problem koji OAuth efikasno rešava i u tome je ključna razlika između njega i drugog popularnog protokola OpenID (7), koji omogućava da se sa istim identitetom pristupa na više sajtova. Ovde je u fokusu deljenje pristupa podacima bez deljenja samog identiteta, tj. delova identiteta za koje želimo da ostanu tajni.

Tok protokola (6)



(A) Klijent zahteva autorizaciju od vlasnika resursa. Ovaj zahtev može biti upućen direktno vlasniku resursa (kao što je prikazano iznad) ali preporučuje se da se koristi server za autorizaciju kao posrednik.

(B) Klijent dobija dozvolu za autorizacijom (Authorization grant). Postoje četiri tipa dozvole i opisani su detaljno u OAuth 2.0 specifikaciji navedenoj u referenci.

(C) Klijent zahteva token za pristup (Access Token) tako što predaje dozvolu serveru za autorizaciju.

(D) Server za autorizaciju proverava klijentovu dozvolu, i ako je validna, izdaje mu token za pristup.

(E) Klijent zahteva zaštićeni resurs od servera sa resursima (Resource server) i autentikuje se predajući token za pristup.

(F) Server sa resursima proverava token, i ako je validan, obrađuje zahtev i prosleđuje odgovor.

Konkretni primeri navedene procedure za različite veb sajtove su obično detaljno objašnjeni u dokumentaciji njihovih API-ja i realizuju se najčešće preko dijaloga u kojima korisnik, istovremeno i vlasnik resursa, dozvoljava (ili ne dozvoljava) posredstvom servera za autorizaciju pristup svojim resursima a klijent program sa izdatom dozvolom na unapred zadat način uzima token od servera i na kraju željene resurse.

4. Facebook Graph API

Kako je navedeno na oficijalnoj facebook strani (1) **Graph API** je primarni (ali ne i jedini) način za pristup podacima unutar Facebook-ovog socijalnog grafa, strukture čiji su čvorovi korinici a ivice njihovi odnosi unutar ove socijalne mreže. On je zasnovan na HTTP-u i može biti korišćen za postavljanje upita nad podacima i unošenje novih podataka kao što su, na primer, statusi i slike.

Pošto je, kao što je navedeno, ovaj API zasnovan na HTTP protokolu može se lako isprobati na adresi `http://graph.facebook.com`.

Većina upita zahteva posedovanje *OAuth 2.0 tokena za pristup* koji aplikacija koju (fiktivno za sad) pravimo može da generiše implementirajući **Facebook login** (2).

4.1. Struktura

Na najvišem nivou Graph API se sastoji od:

- **Čvorova** (nodes) kao što su jedan specifičan korisnik, fotografija ili strana,
- **Ivica** (edges) kao što su fotografije jedne strane ili komentari na jednoj fotografiji,
- **Polja** (fields) kao što su korisnikov rođendan ili ime strane.

Get zahtevi za odgovarajuće komponente bi izgledali:

- Za čvor: **GET** graph.facebook.com/{id-čvora}
- Za ivicu: **GET** graph.facebook.com/{id-čvora}/{ime-ivice}
- Za polje: **GET** graph.facebook.com/{id-čvora}?fields={imena-polja}[4]

4.2. Jednostavan primer

Najlakši način za upoznavanje sa API-jem i takođe da se isproba iznad navedeno je korišćenje oficijalnog alata **Graph API Explorer** (3). Autor preporučuje korišćenje ovog alata u svrhu učenja, iz praktičnih razloga što jednostavno mogu da isprobaju svi primeri koji će dalje biti navedeni u tekstu.

Nakon što sam generisao token za pristup dozvoljavajući ovoj aplikaciji da pristupi delu mojih ličnih podataka, Graph API explorer je prosledio odgovor u JSON formatu za upit:

[4] Ukoliko želimo da navedemo više polja onda navodimo njihova imena odvojena zarezom

GET `/me?fields=id,name,birthday,education`[5]


```
{
  "id": "1103582308",
  "name": "Nikola Divic",
  "birthday": "01/20/1992",
  "education": [
       {
        "school": {
             "id": "110092035680748",
             "name": "X gimnazija \"Mihajlo Pupin\""
        },
        "type": "High School"
       },
       {
        "school": {
             "id": "372834456147822",
             "name": "Matematicki fakultet, Beograd"
        },
        "type": "College",
        "year": {
             "id": "142963519060927",
             "name": "2010"
        }
       }
  ]
}
```


Većina upita mora da bude potvrđena sa tokenom za pristup. U Graph API referenci (4) je navedeno koje dozvole su potrebne za pristup različitim vrstama čvorova i ivica. Explorer nam omogućava da brzo i jednostavno generišemo tokene u svrhu istraživanja API-ja. Ukoliko želimo da napišemo našu aplikaciju taj proces je nešto komplikovaniji.

4.3. Detaljnija upotreba

Zahtevi mogu biti i **ugnježdeni** (nested), tj možemo smestiti više upita u jedan poziv. Njihov oblik je:

– GET graph.facebook.com/{id-čvora}?fields={prvi-nivo}.fields({drugi-nivo})

Za ograničavanje broja entiteta koji će se nalaziti u odgovoru možemo koristiti argument *limit(n)*.

Primer - imena prva tri prijatelja u listi:

– **GET** `graph.facebook.com/me?fields=friends.limit(3).fields(name)`

[5] Čvor *me* je specijalni tačka koja se preslikava u ID korisnika ili facebook strane čiji token se trenutno koristi za pristup podacima

Pošto neki upiti mogu da proizvedu odgovore koji se sastoje nekoliko hiljada (nekad i desetina hiljada) objekata, nije praksa da se celokupan rezultat prosleđuje u jednom odgovoru, već se vrši **straničenje**, podela rezultata na više odvojenih stranica.

Straničenje može biti na bazi **kursora** gde se uz odgovor prosleđuje i polje *cursors* sa dva pod-polja *"after" : "vrednost-after"* i *"before" : "vrednost-before"*. Kada želimo da pročitamo sledeću stranu našeg rezultata napisaćemo isti upit kao i prvi put uz dodavanje argumenta `after={vrednost-after}`. Analogno važi i za prethodnu stranu. API nam to olakšava prosleđujući i polja *next* i *previous* sa celim upitima. Postoji i **vremensko** straničenje i straničenje sa **offsetom**, ali pošto su kursori najefikasnija metoda, preporučeno je da se koriste kad god je to moguće.

Primer: Odgovor od servera za upit `me/albums?fields=name&limit=3`

```
...
...
"cursors": {
        "after": "MzAwMDcwMjA5MjI2Mg==",
        "before": "MTU0MjQ2MTQ3NzE1OA=="
      },
"next": "https://graph.facebook.com/1103582308/albums?fields=name..."
  ...
```

Za sledeću stranu napisaćemo upit identičan onome u next polju.

4.4. Dobijanje tokena za pristup

Dosadašnji primeri su podrazumevali korišćenje Graph API explorer-a ali svrha toga je u tome da naučimo kako da napravimo našu aplikaciju koja će moći da iskoristi sve mogućnosti API-ja.

Jedan od većih problema u implementaciji aplikacija koje pristupaju Facebook-u jeste dobijanje tokena za pristup praćenjem **Login Flow-a**. (2). To je problem za sebe i samostalna implementacija bi zahtevala dobro poznavanje OAuth 2.0 protokola i Facebook-ovih metoda za pristup. Najsigurniji način je da, ako je to moguće, koristimo neki od gotovih SDK-ova (*Software Development Kit*) koje FB podržava.

- **Za veb aplikacije:** JavaScript i PHP SDK
- **Za mobilne aplikacije:** Android i iOS SDK
- **Za desktop aplikacije:** situacija nije tako jednostavna. Postoje neoficijalna (third-party) rešenja ali se proces većinom zasniva na integraciji internet pregledača (browsera) u aplikaciju i onda korišćenje JavaScript SDK. U tom kontekstu se pregledač često naziva WebView.

Svaki od ovih SDK-ova prati isti tok:

- Utvrdi da li je korisnik već log-ovan
- Ako nije prikaži mu Login dijalog
- Razmeni tajne kodove da bi se potvrdio identitet
- Generiši token za pristup

Token za pristup je slučajno generisana niska koja daje aplikaciji privremen i siguran pristup Facebook API-ju. Token može biti generisan za osobu, Facebook stranu ili aplikaciju u poslednjem koraku Login flow-a. SDK-ovi se brinu o generisanju i čuvanju tokena automatski. Aplikacije koje ne žele da koriste preporučene SDK-ove moraju same da implementiraju tok koji prati Login flow.

5. Osvrt na privatnost i zaključak

Ono što može predstavljati problem po privatnost je činjenica da javni API-ji popularnih socijalnih mreža integrišu strani sadržaj u svoj i dozvoljavaju nepoznatim, i potencijalno neodgovornim i nepoštenim razvijaocima, pristup korisnikovim podacima. Ovo poboljšava kvalitet samih mreža ali može predstavljati pretnju po korisnika.

Pitanje slobode korisnika je često i etičko pitanje, jer ako implementatori socijalnih mreža pruže korisniku apsolutnu slobodu izbora šta će od podataka podeliti sa nepoznatom aplikacijom, da li onda njih ne zanima da li će naivni korisnik biti prevaren? Da li je na njima odgovornost da zaštite korisnike njihovih usluga?

Pretnja po privatnost su i podaci koje dođu u ruke vlasnika aplikacija "posredno". Uzmimo za primer aplikaciju koja traži pristup stranicama koje je korisnik "lajkovao" (autor nije u mogućnosti da nađe bolji izraz). Uz takvu potvrdu aplikacija dolazi u posed korisnikovog identifikatora. Sa identifikatorom može da nađe korisnikov profil i da utvrdi ko je osoba koja koristi njihovu aplikaciju i da preuzme sve potencijalne "javne" podatke sa korisnikovog profila.

Jedan od interesantnih koncepata koje rešavaju gore navedeno je **Privacy by proxy** (9) koji aplikaciji pruža podatke posredno tako da je jako teško utvrditi ko je vlasnik preuzetih podataka.

Moderne socijalne mreže su nastale i dobile globalnu popularnost tek početkom prve decenije 21. veka i od tad se ubrzano razvijaju. Koncepti povezanosti, otvorenosti i globalizacije predstavljaju afinitete modernog društva i posvećivanje tehnologijama koje ove koncepte pospešuju za sada izgleda kao dobro uložen trud.

Bibliography

[1] Facebook graph api docs. https://developers.facebook.com/docs/graph-api/

[2] Facebook login. https://developers.facebook.com/docs/facebook-login/

[3] Graph api explorer. https://developers.facebook.com/tools/explorer

[4] Graph api reference. https://developers.facebook.com/docs/graph-api/reference/

[5] Oauth. http://oauth.net/about/

[6] Oauth 2.0 specifikacija. http://tools.ietf.org/html/rfc6749#section-1.2

[7] Openid. http://openid.net/

[8] Crockford, D.: Javascript object notation. http://www.json.org/

[9] Felt, A., Evans, D.: Privacy protection for social networking apis

[10] Media, B., Socialnomics, MacWorld: Social networking statistics. http://www.statisticbrain.com/social-networking-statistics/

[11] newsroom, F.: Oficijalna facebook statistika. http://newsroom.fb.com/company-info/