



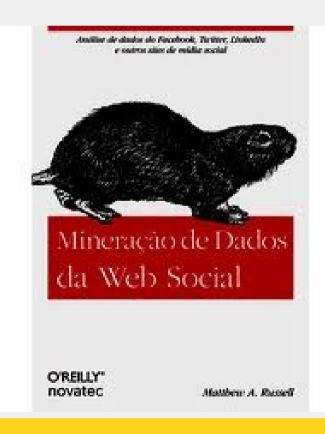
Gmail e Facebook Mineração de Dados da Web Social

Luiz Arthur F. Santos



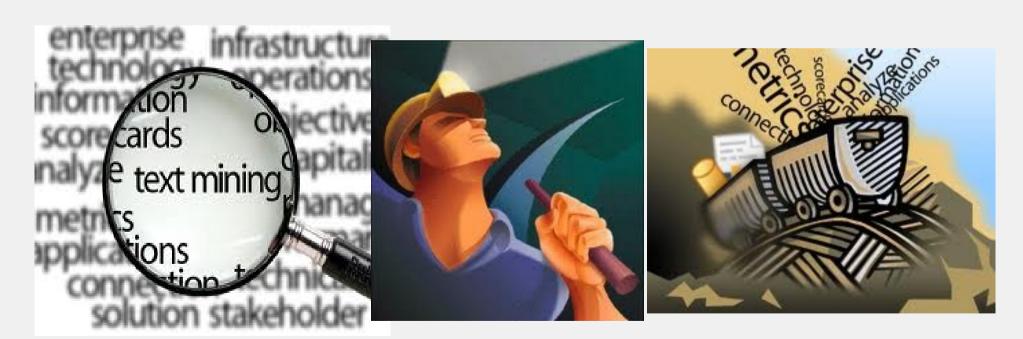
Mineração de dados:

Mineração de Dados é o processo de explorar grandes quantidades de dados à **procura de padrões** consistentes, como **regras** de associação ou **sequências** temporais, para detectar **relacionamentos** sistemáticos entre variáveis, detectando assim novos subconjuntos de dados **auxiliando na descoberta de conhecimento** (RUSSELL, 2011; WIKIPÉDIA, 2012a).



Mineração de Textos:

Text mining ou mineração de textos é uma área que tem como principal objetivo extrair conhecimento implícito de grandes quantidades de textos escritos em linguagem natural (SANTOS,2010).



Redes Sociais:

Uma rede social consiste de um **conjunto finito de atores e as** relações definidas entre eles (MAGDALENO, 2010).

"A **Web** é mais uma **criação social** do que técnica...Nela, nos reunimos em famílias, associações e empresas ..."

Tim Berners-Lee

Alguns exemplos de redes sociais são:

Rede de relações de amizade;

Casamentos entre famílias;

Comunidades de negócio;

• Rede de contatos **sexuais**;

etc...



Mineração de dados nas Web Social:

- Antigamente somente empresas tinham acesso a grandes volumes de dados;
- **Agora** praticamente **todo mundo** com acesso a Internet e consequentemente acesso a uma quantidade de informação jamais vista;
- Hoje este acesso a informação pode sugerir tendências, desvendar particularidades dela e de seu meio ambiente e permiti ações melhor informadas.





Google:

Google é uma empresa multinacional de serviços online, **fundada em 1998**.



Criada por Larry Page e Sergey Brin enquanto estavam frequentando a Universidade Stanford como estudantes de doutorado.

A missão declarada da empresa desde o início foi "**organizar a informação mundial e torná-la universalmente acessível e útil**" (WIKIPÉDIA, 2012b).

Google e suas ferramentas:

O Google hoje, não é apenas um buscador de conteúdos na Web, mas sim uma empresa que fornecem produtos em várias áreas, sendo os **principais produtos**:

- YouTube;
- Orkut;
- Google+;
- Google Chrome;
- Gtalk;
- Buzz;
- Google Docs;
- Picasa;
- Google Maps;
- Latitude;
- Google Tradutor;
- Google Agenda;
- Android;
- etc...







GMail:

O Gmail também conhecido como Google Mail é um serviço gratuito de webmail **criado** pelo Google **em 2004**.



O **Gmail pode ser uma mina de ouro das informações**, já que normalmente temos inúmeros textos em nossa caixa de entrada.

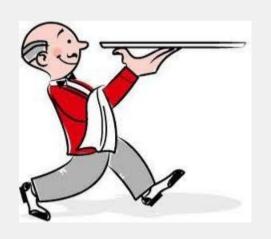
Para analisar e-mail **podemos focar** primeiramente **na estrutura dos dados das mensagens** de e-mail, tal como as **partes que compõem um e-mail**.

Como um aplicativo pode acessar o GMail?

Para acessarmos nossos dados no GMail utilizamos um **usuário e senha**...

Então, caso uma aplicação queira acessar meus dados **devo passar meu usuário e senha** para ela?

Compartilhar usuários e senha é seguro?







É claro que **não devemos compartilhar usuário e senha** com ninguém, isto representa um sério risco a segurança!

Como um aplicativo pode acessar o GMail?

Mesmo assim as vezes queremos compartilhar dados que exigem algum tipo de autenticação.

Usuários usam um site para suas fotos, outro para vídeos, outro para e-mail e por aí vai. Sites precisam acessar os recursos do usuário de outros sites e estes muitas vezes são protegidos.

Usuário + Senha (OpenID):

- Não compartilhável;
- Difícil de revogar;



OAuth!

Oauth (Open Authorization) permite que você compartilhe seus recursos privados (fotos, vídeos, lista de contatos, contas de banco) armazenadas em um site com outro site sem ter de dar seu usuário e senha (RUSSELL, 2011; KAMAKURA, 2012).

Hoje o OAuth está na **versão 2.0** e está **se tornando** uma **unanimidade** entre os aplicativos de redes sociais.

OAuth:

- Utiliza Tokens;



- Concede acesso a um site específico para recursos específicos e por uma duração definida;
- Pode ser completamente transparente para os usuários;

<u>Um pouco mais de detalhes do OAuth 2?</u>

- 1. O usuário final quer autorizar uma aplicação (o cliente) a acessar alguns de seus dados (um escopo), gerenciados por um *Service* (o proprietário do recurso;
- 2. Em vez de pedir a senha do usuário final, o cliente redireciona-o ao proprietário do recurso e o **usuário final autoriza um escopo** para o cliente diretamente com o proprietário;
- 3. Presumindo que o usuário final autorize o acesso do cliente, este cliente é notificado e **recebe um código de autorização** confirmando que o usuário final o autorizou a acessar o escopo;
- 4. O cliente apresenta ao proprietário do recurso o código de autorização que acabou de receber acompanhado do identificador e do correspondente segredo do cliente e recebe de volta um *token* de acesso. Isto tudo é que autoriza o proprietário a acessar o recurso;
- 5. O cliente utiliza o token para fazer solicitações em nome do usuário final até que o token seja revogado ou expire.

<u>Um pouco mais de detalhes do OAuth 2?</u>



O OAuth 2.0 ainda é algo relativamente novo (RUSSELL, 2011).

Em **2010**, a Google anunciou o **uso do XOAuth para acessar o Gmail** através do IMAP e SMTP.

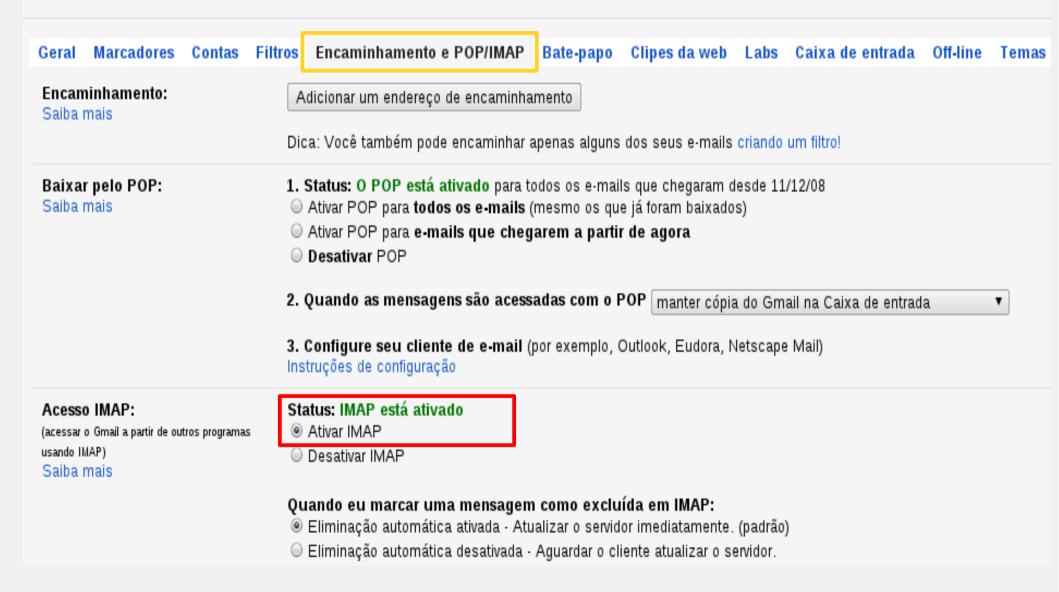
Vamos falar brevemente do XOauth, que é uma implementação particular do OAuth feita pelo Google.

Via aplicação de terceiros podemos acessar o Gmail, assim:

1. Nas configurações do Gmail (no site www.gmail.com) **habilitamos o IMAP** na opção de "Encaminhamento e POP/IMAP";

Habilitando o IMAP do Gmail:

Configurações



Acesse o link:

google-mail-xoauth-tools

Faça o download do **xoauth.py** (mesmo utilizando Java o Google recomenda a obtenção do *token* usando este *script* Python).

Seguindo as instruções do *script* gere um *token* OAuth e o segredo para um usuário anonimo (*anonymous*).

Então execute:

\$ python xoauth.py --generate_oauth_token --user=xoauth@gmail.com

xoauth.py:74: DeprecationWarning: the sha module is deprecated; use the hashlib module instead

import sha

oauth_token_secret: KNBCl_QQWn6BScygEw_GlWv0 oauth_token: 4/WwRMrqo8N12khzEU71Zlo1y1vv5o

oauth_callback_confirmed: true

To authorize token, visit this url and follow the directions to generate a verification code: https://www.google.com/accounts/OAuthAuthorizeToken?oauth_token=4%2FWwRMrqo8N12khzEU71ZIo1y1vv5o

Enter verification code:

Devemos copiar o link da penúltima linha em um navegador para autorizar o software.



O site anonymous está solicitando acesso à sua Conta do Google para o(s) produto(s) listado(s) a seguir.



O Google não é afiliado ao anonymous e recomendamos que você conceda acesso somente se confiar no site.

Se você conceder o acesso, poderá revogá-lo a qualquer momento em "Minha conta". anonymous não terá acesso à sua senha nem a qualquer informação pessoal da sua Conta do Google. Saiba mais

Leste site não foi registrado no Google para estabelecer uma conexão segura para solicitações de autorização. Recomendamos que você negue o acesso, a menos que confie no site.

Conceder acesso

Negar acesso

Se você não for luiz.arthur.feitosa.santos@gmail.com, você pode <u>fazer login como outro usuário</u>.



Você concedeu o acesso de **anonymous** à sua conta do Google. Você poderá revogá-lo a qualquer momento em "Minha conta". Para concluir o processo, forneça ao anonymous este código de verificação: **3qbhewcCSmeQWL_5XLF**Saiba mais

Copie o código enviado pelo Google para o script:

\$ python xoauth.py --generate_oauth_token --user=xoauth@gmail.com xoauth.py:74: DeprecationWarning: the sha module is deprecated; use the hashlib module instead

import sha

oauth_token_secret: KNBCl_QQWn6BScygEw_GlWv0 oauth_token: 4/WwRMrgo8N12khzEU71Zlo1y1vv5o

courts collegely confirmed: true

oauth_callback_confirmed: true

To authorize token, visit this url and follow the directions to generate a verification code:

https://www.google.com/accounts/OAuthAuthorizeToken?

oauth_token=4%2FWwRMrqo8N12khzEU71Zlo1y1vv5o

Enter verification code: 3qbhewcCSmeQWL_5XLXXXXXX

oauth_token: 1/jbXGePhYX3zPmaLemUReSlFaly15YXJB1kxFuXXXXXX

oauth_token_secret: Rpab7rbZoFzAe76iXXXXXX

Agora geramos a *string* de autenticação com o comando:

\$ python xoauth.py -generate_xoauth_string /

- --user=luiz.arthur.feitosa.santos@gmail.com/
- --oauth_token=1/jbXGePhYX3zPmaLemUReSIFaly15YXJB1kxFuXXXXXX/
- --oauth_token_secret=Rpab7rbZoFzAe76i_9XXXXXX

xoauth.py:74: DeprecationWarning: the sha module is deprecated; use the hashlib module instead import sha signature base string:

GET&https%3A%2F%2Fmail.google.com%2Fmail%2Fb%2Fluiz.arthur.feitosa.santos %40gmail.com%2Fimap%2F&oauth_consumer_key%3Danonymous%... texto oculto

xoauth string (before base64-encoding):

GET https://mail.google.com/mail/b/luiz.arthur.feitosa.santos@gmail.com/imap/oauth_consumer_key="anonymous",oauth_nonce=... texto oculto

XOAUTH string (base64-encoded): R0VUIGh0dHBz... chave bem grande!

Acessando o Gmail via IMAP:

Para acessar o Google via Java podemos usar os softwares do seguinte link:

<u>JavaSampleCode</u>

Podemos iniciar com o xoauth-java-sample;

Também precisamos:

- JDK 1.5 ou superior;
- JavaMail API;
- Apache Ant;



Mostrar o código

O programa utilizado por estes slides para testar o acesso aos dados do Gmail podem ser acessados em:

https://github.com/luizsantos/webGmail

Facebook:

Em **2004, Mark Zuckerberg**, Dustin Moskovitz e Chris Hughes, lançaram um **site projetado para colocar os estudantes em contato uns com os outros**, a fim de compartilharem suas fotos e encontrarem novas pessoas.

Em **2005**, os estudantes em 800 redes universitárias que acessavam o **Facebook somavam mais de 5 milhões de usuários** ativos (STRICKLAND, 2012).

Em fevereiro de **2012, o Facebook tinha mais de 845 milhões de usuários** ativos.





Facebook e seus serviços:

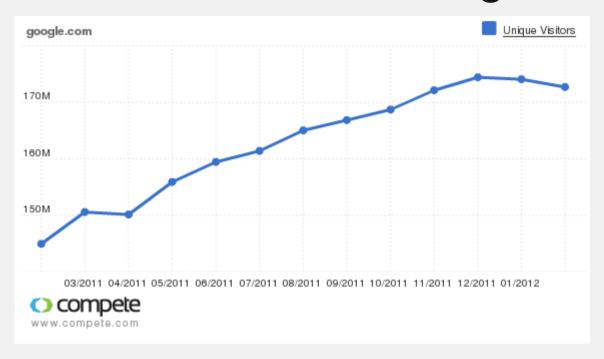
Do ponto de vista das redes sociais o **Facebook é verdadeiramente uma maravilhosa criação** capaz de tudo (RUSSELL, 2011). Considerando que seus milhões de usuários podem :

- Atualizar seus status públicos;
- Informar aos seus amigos o que estão fazendo;
- Trocar mensagens mais longas (semelhante ao e-mail);
- Participar de **chats** em tempo real;
- Organizar e compartilhar **fotos**;
- Fazer *check in em locais físicos*;
- Oferecer aplicativos e jogos;
- etc...

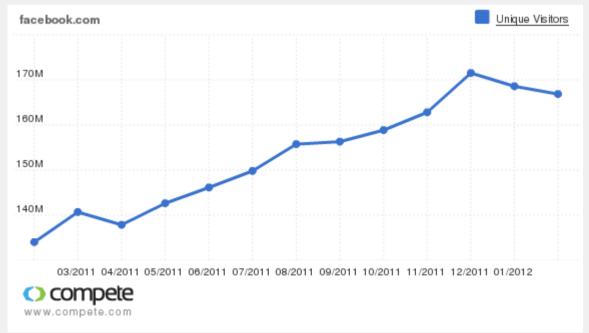


Não é surpresa que **em 2010 o Facebook tenha passado o Google em número de acessos**(PROTALINSKI, 2012).

Facebook versus Google:







Facebook
166,890,779
(acessos mensais)
2° no rank

Aplicativos no Facebook:

Existem centenas de aplicações no Facebook, mas:

- Existem muitos aplicativos de terceiros;
- Fornece acesso limitado a uma interface de programação de aplicativos (API);
- Para criar aplicativos no Facebook, primeiro você tem que adicionar o desenvolvedor de aplicativos ao seu perfil;
- Não hospeda programas de terceiros (STRICKLAND, 2012).

Para criar uma aplicação no Facebook são necessários os seguintes passos:

- Ter uma conta no Facebook (http://facebook.com);
- "Instalar" a aplicação de desenvolvimento Developer (http://www.facebook.com/developers);
- Depois disso é possível criar Aplicações acessando "Set Up New Application".



Selecione o modo como seu aplicativo se integra com Facebook Website http://localhost:8080/ Site URL: [2] App on Facebook Canvas URL: [2] http://localhost:8080/ https://localhost:8080/ Secure Canvas URL: [2] Canvas Page: [?] http://apps.facebook.com/olamundofb Mobile Web Bookmark my web app on Facebook mobile. Native iOS App Publish from my iOS app to Facebook. Native Android App Publish from my Android app to Facebook. Page Tab Build a custom tab for Facebook Pages. Salvar alterações

O Facebook utiliza o OAuth 2.0.

Depois de configurar a aplicação **serão apresentados dados sobre a nova aplicação**, principalmente dados referentes a segurança, tal como:

- Número de identificação da aplicação (ID);
- E um **segredo** (secret) da aplicação,

É importante colocar uma URL de um site que você pretende utilizar para armazenar a aplicação, pois isto é usado como parte da autenticação.

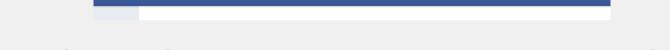
Acessar dados sem usar o token:

Vamos fazer um teste e **ver dados de usuário** usando o *token* e sem utilizar *token*. Primeiro **sem** *token*:

1 - Acesso o Facebook e pegue o nome ou ID do usuário que você quer ver os dados:

| O www.facebook.com/profile.php?id=1314662168|

facebook 🚜 🗖 🚱



2 - Depois acesso o endereço (ex. https://graph.facebook.com/1314662168):

Acessar dados usando o token:

Então, vimos os dados sem o uso do *token*, agora **vamos fazer o mesmo utilizando um** *token* **gerado via** *site***, primeiro o usuário tem que liberar o acesso pelo site**

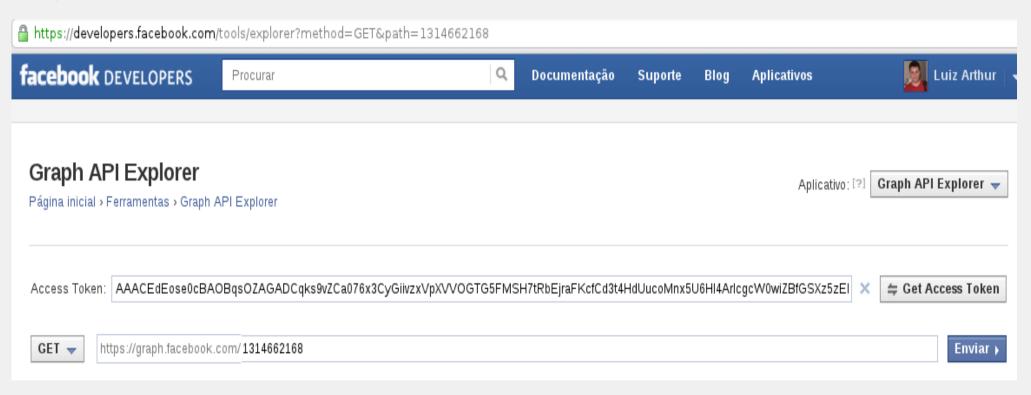
https://developers.facebook.com/tools/explorer? method=GET&path=id_do_usuario_do_face

Substitua o final da URL (id_do_usuario_do_face) pelo ID ou nome do usuário do Facebook. Exemplo de link:

https://developers.facebook.com/tools/explorer? method=GET&path=1314662168

Acessar dados usando o token:

Neste clique em "Get Access Token" e este abrirá uma janela pedindo os níveis de acesso do token.



Feito isto podemos **copiar o** *token* **e tentar acessar os dados** do usuário com o *graph* novamente, tal como anteriormente.

Acessar dados usando o token:

Exemplo: https://graph.facebook.com/YYY&access_token=XXX Onde YYY é o ID e XXX é o token gerado no site.

```
C https://graph.facebook.com/1314662168&access_token=AAACEdEose0cBAOBqsOZAGADCqks9vZCa076x3CyGiivz
 "id": "1314662168",
 "name": "Luiz Arthur",
 "first_name": "Luiz"
 "last_name": "Arthur",
"link": "http://www.facebook.com/profile.php?id=1314662168",
 "birthday": "01/01/1978",
 "location": {
    "id": "111761055506671",
    "name": "Paranava\u00ed"
},
"work": [
       "employer": {
          "id": "103127323061176",
          "name": "Universidade Tecnol\u00f3gica Federal do Paran\u00e1"
    },
{
       "employer": {
          "id": "207638419267434",
          "name": "FALS - Faculdade Le\u00e3o Sampaio"
    },
{
       "employer": {
          "id": "123691214393044",
          "name": "CESUMAR - CENTRO UNIVERSIT\u00c1RIO DE MARING\u00c1"
```

Token e permissões estendidas:

A aplicação **a principio só pode acessar dados básicos** do usuário tal como: nome, sexo e foto do perfil.

Para acessar "todos" os dados precisamos de permissões estendidas.

Repare que você dentro do Facebook pode ver detalhes sobre dados de usuários, mas sua aplicação do Facebook não tem o mesmo direito que você navegando dentro do Facebook, ou seja, a sua aplicação só pode trabalhar, a principio, com um conjunto básico de dados.

Ao tentar acessar dados sem permissão, você terá como resposta objetos em branco (nulos) e uma possível mensagem de erro.

APIs do Facebook:

- Ambiente de desenvolvimento complexo;
- Em constante evolução;
- Inúmeras políticas/controles de privacidade.

Segundo Russell 2012 o Facebook tem o **controle de privacidade mais complexo que podemos encontrar na Web**.

O Facebook suporta:

- O padrão REST;
- FQL Facebook Query Language para consultar dados;
- O Graph API;
- O Open Graph Protocol OGP.

OGP e Graph:

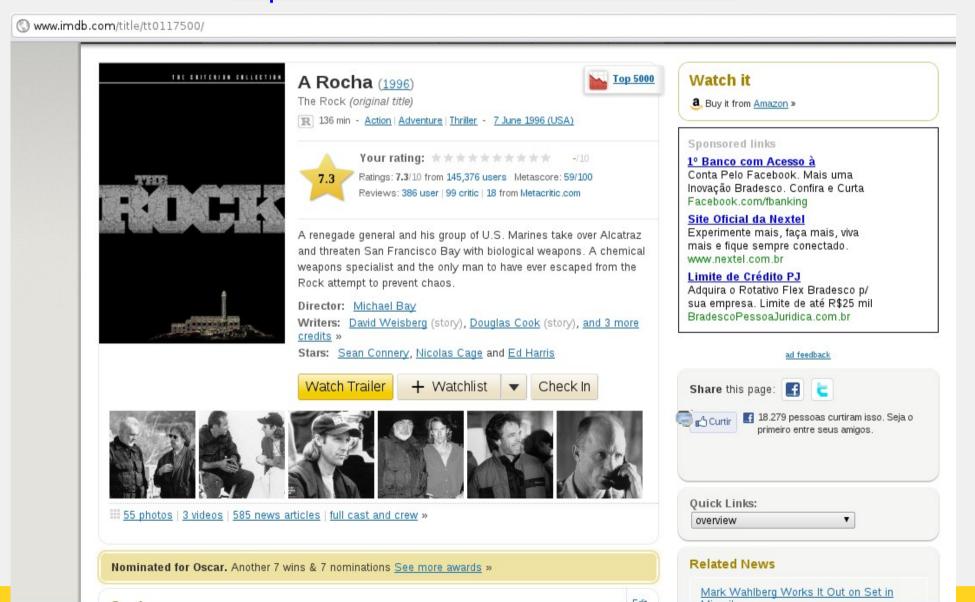
O OGP é um mecanismo que permite fazer com que qualquer página Web seja representada como um objeto.

O Ghaph API é um simples e intuitivo mecanismo para consultar os gráficos gerados pelo OGP. Cada objeto tem um tipo particular. O Graph API suporta basicamente os seguintes objetos:

Album	Application	Event
Group	Link	Note
Page	Photo	Post
Status Message	Subscription	User
Video	Checkin (checagens feitas do Facecook)	

Vamos transformar a página do IMDB referente ao filme The Rock em um objeto do OGP:

http://www.imdb.com/title/tt0117500



É só colocar o link do site desejado no no final do link http://graph.facebook.com/http://www.imdb.com/title/tt0117500/ que é o resultado do objeto gerado pelo OGP.

```
G graph.facebook.com/http://www.imdb.com/title/tt0117500/
   "id": "114324145263104",
   "name": "The Rock (1996)",
   "picture": "http://profile.ak.fbcdn.net/hprofile-ak-snc4/276757_114324145263104_1964457943_s.jpg",
   "link": "http://www.imdb.com/title/tt0117500/",
   "likes": 8630,
   "app_id": 115109575169727,
   "category": "Movie",
   "is_published": true,
   "website": "http://www.imdb.com/title/tt0117500/",
   "description": "Directed by Michael Bay. With Sean Connery, Nicolas Cage, Ed Harris, John Spencer. A renegade general
and his group of U.S. Marines take over Alcatraz and threaten San Francisco Bay with biological weapons. A chemical
weapons specialist and the only man t",
   "about": "Directed by Michael Bay. With Sean Connery, Nicolas Cage, Ed Harris, John Spencer. A renegade general and his
group of U.S. Marines take over Alcatraz and threaten San Francisco Bay with biological weapons. A chemical weapons
specialist and the only man t",
   "can_post": true,
   "talking_about_count": 4
```

O OGP não é muito utilizado atualmente, mas existe um potencial muito grande para esta ferramenta.

Podemos tornar a busca mais interessante colocando o parâmetro metadata=1, aumentando o nível de consulta.

https://graph.facebook.com/114324145263104?metadata=1

O item *metadata.connections* são ponteiros para outros nós no grafo que podemos navegar.

Podemos, por exemplo, seguir os links das fotos associadas com com o filme ou andar por links para descobrir comentários sobre o mesmo.

O resultado de The Rock com o metadata=1 será:

```
"id": "114324145263104".
 "name": "The Rock (1996)",
.... (dados omitidos)
 "metadata": {
   "connections": {
     "feed": "https://graph.facebook.com/114324145263104/feed",
     "posts": "https://graph.facebook.com/114324145263104/posts",
     "tagged": "https://graph.facebook.com/114324145263104/tagged",
     "statuses": "https://graph.facebook.com/114324145263104/statuses",
     "links": "https://graph.facebook.com/114324145263104/links",
     "notes": "https://graph.facebook.com/114324145263104/notes",
     "photos": "https://graph.facebook.com/114324145263104/photos",
     "albums": "https://graph.facebook.com/114324145263104/albums",
     "events": "https://graph.facebook.com/114324145263104/events",
     "videos": "https://graph.facebook.com/114324145263104/videos",
     "questions": "https://graph.facebook.com/114324145263104/questions",
     "coupons": "https://graph.facebook.com/114324145263104/coupons",
     "milestones": "https://graph.facebook.com/114324145263104/milestones"
   "fields": [
       "name": "id",
       "description": "The Page's ID. No access token or user `access token`. `string`."
     },
       "name": "name",
       "description": "The Page's name. No access token or user `access token`. `string`."
```

Você mesmo é um objeto. Tente visitar a mesma URL, mas troque o ID por se próprio ID.

Tente primeiro sem o token:

https://graph.facebook.com/1314662168&metadata=1

Depois tente com o token:

https://graph.facebook.com/1314662168&metadata=1? &access_token=YYY

Troque YYY por seu token.

FQL:

Desenvolvedores podem usar o Facebook Query Language (FQL), que é semelhante ao Structured Query Language (SQL).

Com FQL, o desenvolvedor pode obter informação sobre qualquer usuário que use o seu aplicativo.

Os desenvolvedores podem usar essa informação como uma forma de criar públicos alvo para propaganda ou construir uma base de consumidores para um produto em particular.

O Facebook tem sido criticado pelas pessoas que acreditam que o site permite aos desenvolvedores de aplicativos coletar dados dos membros, incluindo informação pessoal e hábitos de navegação na web (STRICKLAND, 2012).

Sintaxe do FQL:

select [fields] from [table] where [conditions]

O FQL possui várias restrições, o que o torna um SQL reduzido.

Por exemplo, apenas **um nome de tabela pode ser utilizado na clausula from** e as condições que aparecem no **where são limitadas**.

APIs Java para o Facebook:

A API Java do Facebook foi descontinuada.

Existem APIs de terceiros tal como:

- Facebook-java-api que pode ser encontrado em http://code.google.com/p/facebook-java-api/;
- RestFB disponível em http://restfb.com/;



O projeto de exemplo utilizado para acessar o Facebook através do Java utilizando o RestFB pode ser encontrado em:

Visualizando Dados:

Depois de obter os dados do Facebook, devemos ter uma boa **ferramenta para visualizá-los**, um bom exemplo do que podemos fazer pode ser visto na Ferramenta **JIT**, exemplo:

http://thejit.org/static/v20/Jit/Examples/Sunburst/example2.html

Exercício 1:

Seguindo os passos anteriores e utilizando o exemplo xoauth-javasample, faça a conexão ao GMail utilizando o IMAP e mostre o Assunto de todas as suas mensagens da Caixa de entrada.

Exercício 2:

Faça uma aplicação que gere um JSON com os seguintes dados de seus amigos do Facebook: nome, sexo, data de nascimento e cidade;

Referências

RUSSELL. Matthew A. **Mineração de Dados na Web Social**. Editora Novatec. 2011.

WIKIPÉDIA. **Mineração de Dados**. Disponível em: http://pt.wikipedia.org/wiki/Minera%C3%A7%C3%A3o_de_dados Acessado em: Abril de 2012a.

SANTOS, Leandro Matioli. **Protótipo para mineração de opinião em redes sociais: estudo de casos selecionados usando o Twitter**. Universidade Federal de Lavras. 2010.

MAGDALENO, Andréa M; *et al.* **Estudo de Ferramentas de Mineração, Visualização e Análise de Redes Sociais.** COPPE/UFRJ. 2010.

WIKIPÉDIA. **GMail**. Disponível em: http://pt.wikipedia.org/wiki/Gmail Acessado em: Abril de 2012b.

KAMAKURA, Felipe. Como implementar login com Facebook (OAuth 2.0). Disponível em:

http://kamaondev.wordpress.com/2011/01/24/como-implementar-login-com-facebook-com-sado em: Abril de 2012.

46

Referências

STRICKLAND, Jonathan. **Como funciona o Facebook**. Disponível em: http://informatica.hsw.uol.com.br/facebook3.htm Acessado em: Abril de 2012.

PROTALINSKI, Emil. Facebook has over 845 million users. Disponível em: http://www.zdnet.com/blog/facebook/facebook-has-over-845-million-users/8332 Acessado em: Abril de 2012.







Perguntas?

Luiz Arthur – luizsan@ime.usp.br