



TRABALHO DE GRADUAÇÃO

**A Systematic Mapping Study:
Qualidade de Serviço
em Computação Orientada a Serviços**

Danilo Filgueira Mendonça

Brasília, Março de 2012

UNIVERSIDADE DE BRASÍLIA

FACULDADE DE TECNOLOGIA

UNIVERSIDADE DE BRASÍLIA
Faculdade de Tecnologia

TRABALHO DE GRADUAÇÃO

A Systematic Mapping Study: Qualidade de Serviço em Computação Orientada a Serviços

Danilo Filgueira Mendonça

Relatório submetido ao Departamento de Engenharia Elétrica
como requisito parcial para obtenção do grau de
Engenheiro de Redes de Comunicação

Banca Examinadora

Profa. Dra. Genáina Nunes Rodrigues , _____
CIC/UnB
(Orientadora)

FICHA CATALOGRÁFICA

MENDONCA, D. F.. A Systematic Mapping Study:
QOS em Computação Orientada a Serviços [Distrito Federal] 2012.
v, 52p. (ENE/FT/UnB, Engenheiro de Redes de Comunicação, 2012)
Monografia de Graduação - Universidade de Brasília. Faculdade de Tecnologia.
Departamento de Engenharia Elétrica.

1. Computação Orientada a Serviços	2. SOC
3. Arquitetura Orientada a Serviços	4. SOA
5. Qualidade de Serviços	6. QOS
I. ENE/FT/UnB	II. Título (série)

REFERÊNCIA BIBLIOGRÁFICA

MENDONCA, D. F. e (2012). A Systematic Mapping Study: QOS em Computação Orientada a Serviços Monografia de Graduação, Publicação ENE 01/2012, Departamento de Engenharia Elétrica, Universidade de Brasília, Brasília, DF, 52p.

CESSÃO DE DIREITOS

NOMES DOS AUTORES: Danilo Filgueira Mendonça e .

TÍTULO: A Systematic Mapping Study: QOS em Computação Orientada a Serviços

GRAU / ANO: Engenheiro de Redes de Comunicação / 2012.

É concedida à Universidade de Brasília permissão para reproduzir cópias desta monografia de graduação e para emprestar ou vender tais cópias somente para propósitos acadêmicos e científicos. Os autores reservam outros direitos de publicação e nenhuma parte desta monografia de graduação pode ser reproduzida sem a autorização por escrito dos autores.

Danilo Filgueira Mendonça
SQN 303 BL F APTO 605 - ASA NORTE
CEP 70735-060 - Brasília - DF - Brasil.

Dedicatória

Às manhãs de meu avô.

Danilo Filgueira Mendonça

Agradecimentos

Danilo Filgueira Mendonça

RESUMO

A Computação Orientada a Serviços ressurgiu com o intuito de prover maior eficiência à produção e consumo de recursos computacionais, especialmente os softwares, que passam a compor unidades coesas e granulares de lógica capazes de se intercomunicarem, aumentando o reuso, a agilidade e o retorno de investimento. Dada a incerteza sobre o estado da pesquisa relacionada ao tema, detectou-se a necessidade de um mapeamento de publicações, indicando quais tipos e quais subáreas receberam maior foco em detrimento daquelas que receberam menor foco, determinando também seus benefícios e limitações. A partir de um processo sistemático de mapeamento de estudos, ou Systematic Mapping Study, que envolve a busca por publicações em fóruns de interesse de modo a classificá-las, avaliamos artigos que tratam de QOS em SOC para enfim responder às questões de pesquisa que orientam e motivam o presente estudo. O resultado foi um mapeamento amplo que classifica XXX publicações e que permite uma melhor compreensão dos esforços de pesquisa dentro da área de interesse.

O trabalho é dividido em 5 capítulos: Introdução, Referência Teórica, Abordagem, Discussão, Conclusão e Trabalhos Futuros. O capítulo de Introdução trata da área de interesse e da metodologia utilizada para o mapeamento. Aborda-se também os objetivos que são divididos em objetivos principais, gerais, específicos e a longo prazo do trabalho. Neste capítulo também apresentam-se as motivações e é mostrada a devida importância do trabalho. Ainda nesse capítulo são descritos trabalhos relacionados. No capítulo de Referencial Teórico são apresentados os principais conceitos necessários para o entendimento do processo de mapeamento sistemático de estudos, além daqueles que envolvem os conceitos de QOS em SOC. No capítulo de Abordagem descreve-se o processo de mapeamento, desde a busca, o processo de inclusão e exclusão de publicações e a classificação dos mesmos. No penúltimo capítulo, Discussão, são colocadas considerações à respeito das questões de pesquisa que motivaram esse mapeamento. Na Conclusão é feito um balanço dos objetivos e a visão final sobre os resultados obtidos, assim como a expectativa para trabalhos futuros.

ABSTRACT

English version

The work here presented shows how to build applications that perform data collection and analysis of virtual social networks, focusing on Facebook. They are also summarized ways of carrying out the processes of collecting, analyzing and building applications and presented the concepts involved in understanding these processes.

The work is divided into five chapters: Introduction, Theoretical Background, Implementation, Case Studies and Conclusion. In the Introduction chapter, all chapters of this work are discussed superficially. It also addresses the goals are divided into primary objectives, general, specific and long-term work. This chapter also presents the motivations that are all justified and is shown the importance of proper work. Although this chapter describes related work. Theoretical Background In chapter presents the main concepts of work that are necessary for understanding the ensuing chapters. Implementation is presented in chapter created the application, what it does and how. In addition to showing the reader the aspects involved in software development. In the penultimate chapter, case study presents the analysis performed from the saved data categorization. In the last chapter, Conclusion, reflects on the work and the achievement of objectives, addressing some of the content described in all chapters of the work. Then we have four annexes relating to the Implementation Chapter.

SUMÁRIO

1	INTRODUÇÃO	1
1.1	MOTIVAÇÃO	9
1.2	TRABALHOS RELACIONADOS	10
1.3	OBJETIVOS	13
1.4	ORGANIZAÇÃO DO TRABALHO	13
2	EMBASAMENTO TEÓRICO	14
2.1	REDES SOCIAIS	14
2.2	REDES COMPLEXAS	15
2.3	TÉCNICAS DE COLETA E ANÁLISE DE DADOS	17
2.4	FACEBOOK	18
2.5	FACEBOOK DEVELOPER PLATFORM	22
3	IMPLEMENTAÇÃO	29
3.1	DESCRIÇÃO GERAL DA IMPLEMENTAÇÃO	29
3.2	TECNOLOGIAS UTILIZADAS	29
3.2.1	TECNOLOGIAS DO FACEBOOK	30
3.2.2	TECNOLOGIAS GERAIS: LINGUAGENS DE PROGRAMAÇÃO, BIBLIOTECAS E IDES	30
3.2.3	SERVIDOR DE HOSPEDAGEM E DE DOMÍNIO	31
3.3	REGRAS DE NEGÓCIO, ANÁLISE DE REQUISITOS E METODOLOGIAS	31
3.3.1	ELABORAÇÃO DA REGRA DE NEGÓCIO E ANÁLISE DE REQUISITOS	32
3.3.2	REGRAS DE NEGÓCIO: COLETAR E ANALISAR O QUE?	33
3.3.3	PRINCIPAIS REQUISITO PARA COLETA E ANÁLISE	34
3.4	MINERAÇÃO DE DADOS	35
3.5	METODOLOGIAS, PADRÕES E TESTES DO PROJETO	36
3.6	ESTRATÉGIAS PARA ATINGIR PÚBLICO ALVO	36
4	ESTUDO DE CASO	38
4.1	DISCUSSÃO DE RESULTADOS	38
5	CONCLUSÃO	41
5.1	SUGESTÃO DE MELHORIAS	42
	REFERÊNCIAS	43
6	ANEXOS	44
6.1	ANEXO I - ESTRUTURA DE ARQUIVOS DO PROJETO IMPLEMENTADO	44
6.2	ANEXO II - FORMAS DE ACESSO E CONFIGURAÇÃO DO APLICATIVO NO FACEBOOK	44
6.3	ANEXO III - CD COM ARQUIVOS DO PROJETO IMPLEMENTADO	44
	ANEXOS	45

LISTA DE FIGURAS

1.1	Princípios relacionados ao SOA e seus benefícios	2
1.2	Ranking Países (população) e Redes Sociais Online (usuários).....	3
1.3	Os nove sites mais populares em audiência da Web	4
1.4	Os nove sites mais populares em audiência no Brasil	5
1.5	Comunidade da Universidade de Brasília na RSO Orkut.....	6
1.6	Comunidade "Professores UnB - Avaliação"na RSO Orkut	7
1.7	Perfil da Universidade de Brasília no Facebook	7
1.8	Página da Universidade de Brasília no Facebook.....	8
1.9	Perfil do Diretório Central dos Estudantes da Universidade de Brasília no Facebook.....	8
1.10	Entidades da UnB mostradas no Perfil do DCE UnB no Facebook	9
2.1	Algumas das principais RSO categorizadas.....	14
2.2	Similaridades entre Redes e Grafos	16
2.3	Esquema de um Agregador de RSOs em que Usuário se Conecta a Múltiplas RSO Simul- taneamente	18
2.4	Propriedades e conexões de Usuários e Páginas	19
2.5	Página Inicial da RSO Facebook	20
2.6	Página Inicial da RSO Facebook	20
2.7	Gráfico Social do Facebook.....	22
2.8	Resumo da Plataforma do Facebook.....	23
2.9	Plugins Sociais da Folha de São Paulo	23
2.10	Diálogo de Autorização.....	24
2.11	Permissões do usuário e de amigos	25
2.12	Permissões do usuário e de amigos	26
2.13	Fluxo de Coleta de Informações de Usuários do Facebook	27
2.14	Exemplo de Resposta de Requisição da Graph API	28

NOTAÇÕES

Siglas

RSO	<i>Redes Sociais Online</i>
OSN	<i>Online Social Networks</i>
WEB	<i>Sistema de Documentos de Hipertexto Interligados que podem ser executados via Internet.</i>

1 INTRODUÇÃO

A tecnologia da informação já se consolidou como ferramenta essencial para as instituições, afetando rotinas, ações, estruturas e modelos de negócios. Os altos custos associados ao consumo e provimento de recursos de TI e a necessidade de uma maior agilidade para responder às mudanças casam-se às propostas da computação orientada a serviços, ou SOC, cujo paradigma de design é composto por princípios que, quando absorvidos, criam condições para o cumprimento dos objetivos da computação orientada a serviços, entre eles o aumento do retorno de investimento, da agilidade, da reusabilidade e da interoperabilidade, assim como um maior alinhamento entre TI e modelo de negócios.

O paradigma de orientação a serviços, tal qual outros paradigmas, possui um design próprio que norteia a configuração do suprimento e consumo dos recursos de TI. Apesar do sucesso de outros paradigmas, por exemplo a orientação a objetos, nota-se que problemas de ordem superior, ou seja, da eficiência, agilidade e retorno de investimento na composição de aplicações, permaneceram. Assim, aplicações silos carecem de interoperabilidade, sendo difícil ou inviável seu reuso, o que aumenta o custo de desenvolvimento de novas aplicações a longo prazo e reduz drasticamente a agilidade ou tempo de resposta às novas demandas, sobretudo quando os modelos de negócio exigem maior flexibilidade. Dessa forma, os benefícios idealizados pelo SOC fizeram desse paradigma um foco de estudos em diversos centros acadêmicos no mundo, sobretudo à partir do surgimento da arquitetura orientada a serviços, ou SOA. Também é notável a adesão de grandes vendedores da indústria de TI à comercialização de SOA. Esses foram responsáveis pela difusão, porém também pela vulgarização do conceito de SOA, ao ponto do mesmo ter se tornado ambíguo e com significados diversos.

O princípio fundamental do design estabelecido pelo SOC está na criação de unidades lógicas coesas e granulares, denominadas serviços. Assim, em vez de aplicações isoladas que trazem consigo todas as funcionalidades, são criadas unidades coesas e de baixo acoplamento cujo acesso é feito por uma interface que encapsula a lógica e cujo funcionamento é definido num contrato. Com isso, é possível realizar uma composição de serviços que irá servir uma ou, idealmente, várias aplicações por meio da orquestração desses serviços.

Com base nos princípios de SOC, criou-se uma arquitetura responsável por guiar o processo de adoção desse paradigma. O SOA, ou *Service Oriented Architecture*, define uma série de processos, papéis, regras, modelos e camadas, entre outros componentes da arquitetura, propondo transformações importantes

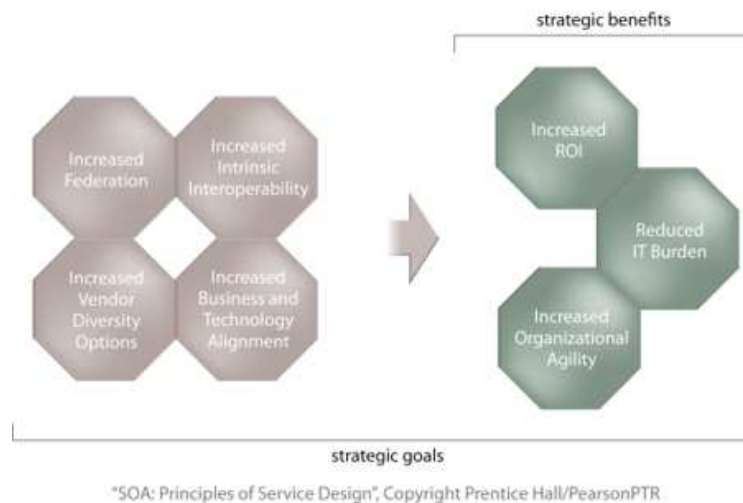


Figura 1.1: Princípios relacionados ao SOA e seus benefícios

no modus operandi no setor de tecnologia da informação. Dada sua abrangência e complexidade de fundamentos, o SOA é protagonista no avanço do SOC, tendo seu conceito se difundido de forma mais ampla que o primeiro, haja vista que recebeu grande atenção do meio acadêmico, como ilustra nosso mapeamento de estudos, além de ter sido um produto de destaque para grandes vendedores no mercado de TI. Entre as prioridades do SOA apontada pelo *SOA Manifesto* estão: o valor de negócio sobre a estratégia técnica adotada, os objetivos estratégicos sobre os benefícios específicos de projeto, a interoperabilidade intrínscita sobre as integrações customizadas, os serviços compartilhados sobre implementações com propósito específico, a flexibilidade sobre otimização e o refinamento evolutivo sobre a perfeição inicial [3]

O ambiente proposto pelas arquiteturas orientadas a serviço está sujeito a condições diferentes daquelas já estudadas e conhecidas em outros paradigmas. Por exemplo, o ambiente distribuído em que compõe-se uma aplicação a partir de unidades coesas acessadas por uma camada de serviços intermediária tem variáveis que dificultam a análise de desempenho e disponibilidade, entre outros parâmetros de qualidade de serviço. Assim, a adoção de SOA rumo à orientação a serviços é uma decisão que requer planejamento, análise de riscos e de parâmetros diversos conforme o modelo de negócios atendido. Nesse sentido o presente estudo visa mapear publicações cujo foco é QOS em SOC, de forma a reduzir a incerteza quanto ao estado dessa tecnologia e quanto aos desafios e avanços já conquistados.

Conforme será apresentado posteriormente nesse trabalho, o número de publicações que trazem como foco de estudo o SOC cresceu sensivelmente, sobretudo à partir de 2005 (REFERÊNCIA), gerando um cenário pouco claro a respeito dos avanços e principais desafios relacionados a esse paradigma. Portanto, um trabalho de mapeamento permite organizar e classificar as publicações a partir de determinadas facetas, gerando embasamento a quem deseja iniciar ou avançar nos estudos da área mapeada.

O *Systematic Mapping Study* propõem uma abordagem que preza pela amplitude em detrimento da profundidade de análise. Com isso, é possível incluir um número mais significativo de publicações e um escopo de estudos mais amplo que aquele obtido pelo *Systematic Literature Review*, tendo esse último uma abordagem que difere-se pela maior precisão e riqueza de detalhes em sua classificação, podendo ser aplicado num possível trabalho futuro, cujo enfoque seriam subáreas de interesse identificadas pelo primeiro processo de mapeamento aqui apresentado.

 1. China	1,336,450,000	 6. Twitter	200,000,000
 2. India	1,178,436,000	 7. Brazil	192,651,000
 3. Facebook	750,000,000	 8. Pakistan	169,010,500
 4. United States	308,898,000	 9. Bangladesh	162,221,000
 5. Indonesia	231,369,500	 10. Nigeria	154,729,000

Figura 1.2: Ranking Países (população) e Redes Sociais Online (usuários)

Nos Estados Unidos, o uso de *mídias sociais*, que são estratégias e canais para difusão de conteúdo em RSOs, é mais popular que o email, segundo *Nielsen Online*. Segundo o site www.alexacom.com, dos nove sites mais *tops* da Web quatro são RSOs, que estão circuladas na Figura 1.2. Neste site o parâmetro que mede quão *top* um site é uma combinação da média de visitantes diários e a quantidade de visualizações de páginas do mês passado. Este parâmetro é também comumente chamado de audiência. No Brasil, a proporção de RSOs entre os nove sites mais *top* é igual a proporção mundial, mostrada anteriormente. Confira na figura 1.3. É interessante notar também que a RSO Facebook ultrapassa números da RSO Orkut nesta análise.

A RSO Facebook armazena mais de 1.5 PB de fotos compartilhadas. A RSO YouTube armazena por dia o equivalente a um ano de conteúdo televisionado pelas principais emissoras dos Estados Unidos. De acordo com o *Global Internet Phenomena Report* (Relatório Global do Fenômeno da Internet, em português), o volume de tráfego gerado por RSOs vem crescendo, principalmente na América Latina onde representa 13,7% do volume total de tráfego [1]. No Brasil, as principais RSOs existentes são a Orkut, com cerca de 27 milhões de usuários ativos, e a Facebook, com cerca de 30 milhões de usuários ativos. Na Universidade de Brasília, o Orkut é lembrado principalmente pela comunidade "UnB - Universidade de Brasília" com 36.327 membros mostrada na figura XX pela comunidade "UnB é Hogwarts" com 5.481 membros. Outras comunidades relevantes, pela quantidade de membros, são "UnB/Gama - Engenharia" (1.760 membros) e "UnB Ceilândia" (1.260 membros).


 Top Sites The top 500 sites on the web. ?	
1	Google <small>google.com</small> Enables users to search the world's information, including webpages, images, and videos. Offers... More ★★★★★ Search Analytics ► Audience ►
2	Facebook <small>facebook.com</small> A social utility that connects people, to keep up with friends, upload photos, share links and ... More ★★★★★ Search Analytics ► Audience ►
3	YouTube - Broadcast yourself <small>youtube.com</small> YouTube is a way to get your videos to the people who matter to you. Upload, tag and share your... More ★★★★★ Search Analytics ► Audience ►
4	Yahoo! <small>yahoo.com</small> A major internet portal and service provider offering search results, customizable content, cha... More ★★★★★ Search Analytics ► Audience ►
5	Baidu.com <small>baidu.com</small> The leading Chinese language search engine, provides "simple and reliable" search exp... More ★★★★★ Search Analytics ► Audience ►
6	Blogger.com <small>blogspot.com</small> Free, automated weblog publishing tool that sends updates to a site via FTP. ★★★★★ Search Analytics ► Audience ►
7	Wikipedia <small>wikipedia.org</small> A free encyclopedia built collaboratively using wiki software. (Creative Commons Attribution-Sh... More ★★★★★ Search Analytics ► Audience ►
8	Windows Live <small>live.com</small> Search engine from Microsoft. ★★★★★ Search Analytics ► Audience ►
9	Twitter <small>twitter.com</small> Social networking and microblogging service utilising instant messaging, SMS or a web interface.

Figura 1.3: Os nove sites mais populares em audiência da Web




 Top Sites in Brazil  The top 500 sites in Brazil. 	
1	Google Brasil <small>google.com.br</small> Buscador que enfoca sus resultados para este país y a nivel internacional tanto en portugués co... More ★★★★★ Search Analytics ▶ Audience ▶
2	Google <small>google.com</small> Enables users to search the world's information, including webpages, images, and videos. Offers... More ★★★★★ Search Analytics ▶ Audience ▶
3	Facebook <small>facebook.com</small> A social utility that connects people, to keep up with friends, upload photos, share links and ... More ★★★★★ Search Analytics ▶ Audience ▶
4	YouTube - Broadcast yourself <small>youtube.com</small> YouTube is a way to get your videos to the people who matter to you. Upload, tag and share your... More ★★★★★ Search Analytics ▶ Audience ▶
5	Universo Online <small>uol.com.br</small> Acesso à internet e serviços agregados. ★★★★★ Search Analytics ▶ Audience ▶
6	Windows Live <small>live.com</small> Search engine from Microsoft. ★★★★★ Search Analytics ▶ Audience ▶
7	Globo.com - Absolutamente tudo sobre esportes, notícias, entretenimento e vídeos <small>globo.com</small> ★★★★★ Search Analytics ▶ Audience ▶
8	orkut.com.br <small>orkut.com.br</small> ★★★★★ Search Analytics ▶ Audience ▶
9	Blogger.com <small>blogspot.com</small>

Figura 1.4: Os nove sites mais populares em audiência no Brasil

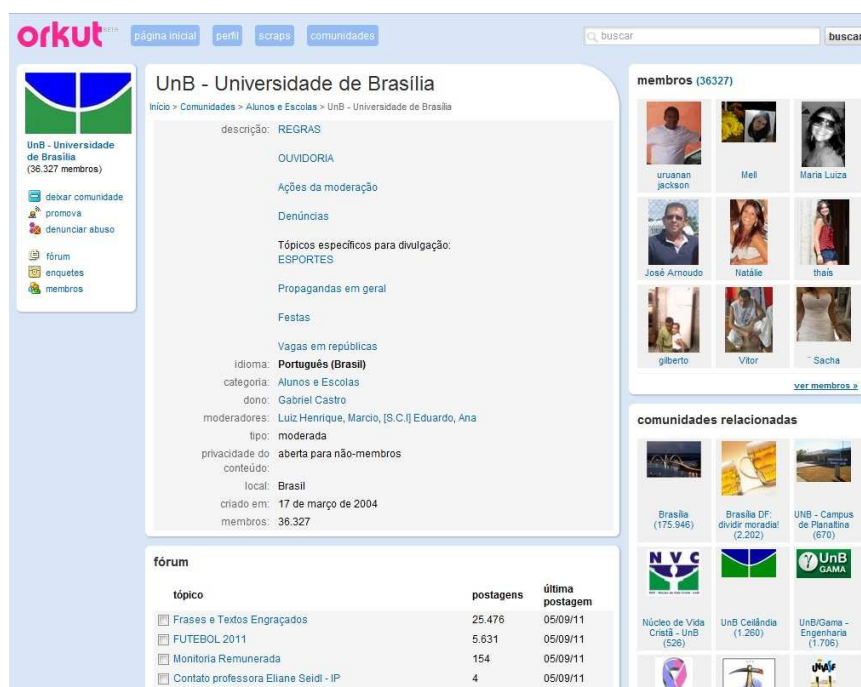


Figura 1.5: Comunidade da Universidade de Brasília na RSO Orkut

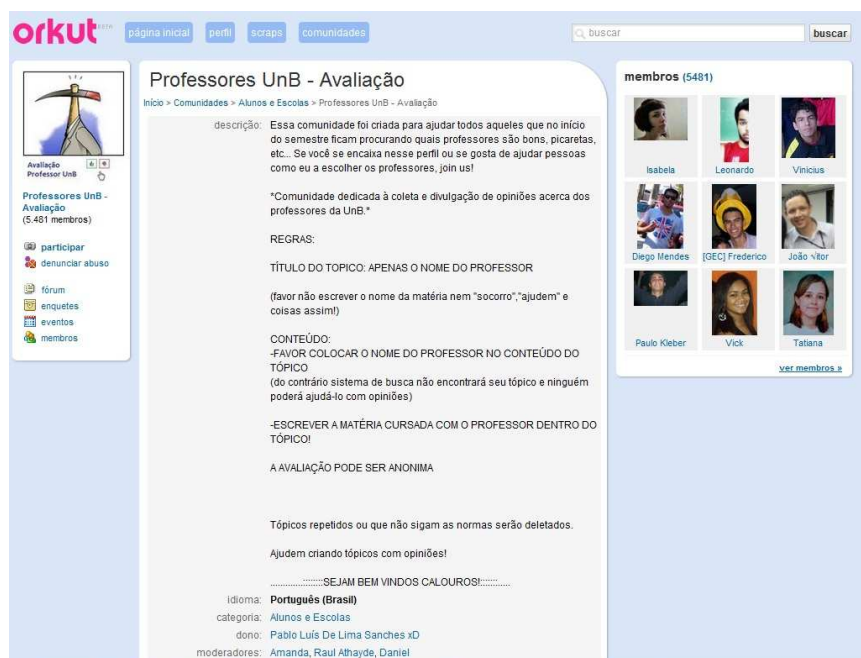


Figura 1.6: Comunidade "Professores UnB - Avaliação" na RSO Orkut

Os estudantes da Universidade de Brasília tem forte presença na RSO Facebook também. O perfil da Universidade de Brasília no Facebook (<http://www.facebook.com/unibrasilia>), mostrado na figura 1.4, conta com 4.998 amigos, dos quais a maioria tem a página da Universidade de Brasília (<http://www.facebook.com/pages/Unb-de-Brasilia/113901608620360>), mostrada figura 1.5 setada na parte de educação do seu perfil. Estes usuários representam graduandos ou graduados da Universidade de Brasília.

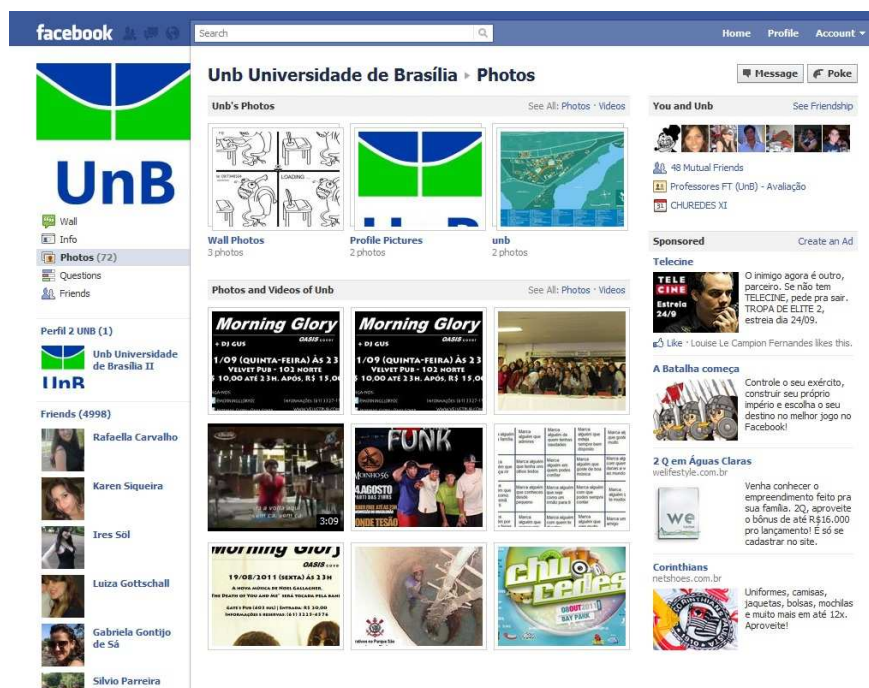


Figura 1.7: Perfil da Universidade de Brasília no Facebook

E Mais: O perfil do Diretório Central dos Estudantes da UnB (<http://www.facebook.com/dceunb>), mostrado na figura 1.6, tem cerca de 3490 amigos, dos quais a maioria tem a página da Universidade de Brasília setada na parte de educação do seu perfil. Neste perfil, também são mostradas alguns perfis de Entidades da UnB como as Empresas Juniores: ADM Consultoria Empresarial (Curso de Administração), Strategos (Curso de Ciências Políticas), Socius Consultoria Jr. (Curso de Ciências Sociais); os Centro Acadêmicos: CaAdm (Curso de Administração), CaEco (Curso de Economia), Calet (Curso de Letras), Capol (Curso de Ciências Políticas); e outras entidades como CDT (Centro de Desenvolvimento Tecnológico), UAB (Universidade Aberta) e Petenc (PET do curso de Engenharia Civil) mostrados na figura 1.7.



Figura 1.8: Página da Universidade de Brasília no Facebook

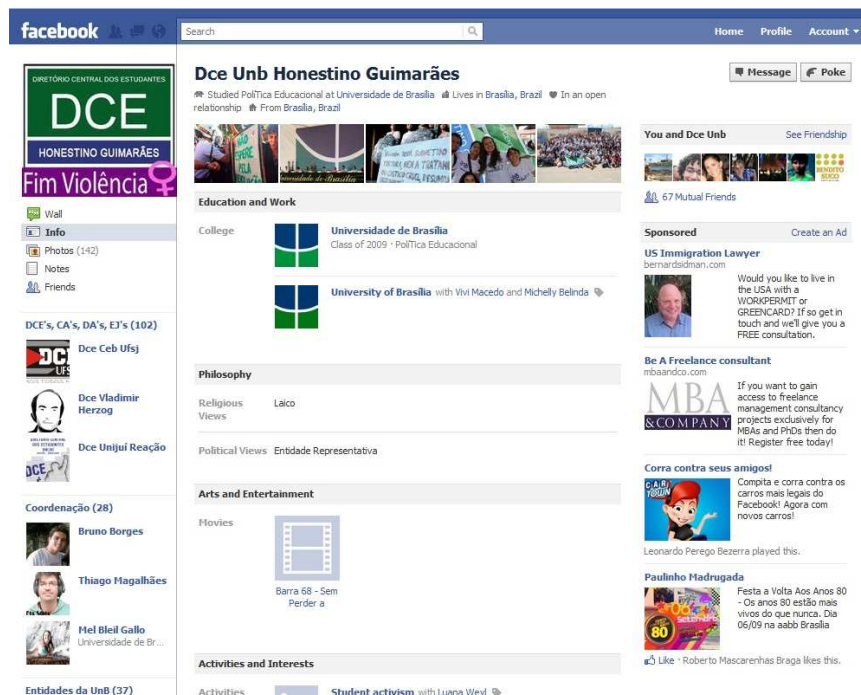


Figura 1.9: Perfil do Diretório Central dos Estudantes da Universidade de Brasília no Facebook

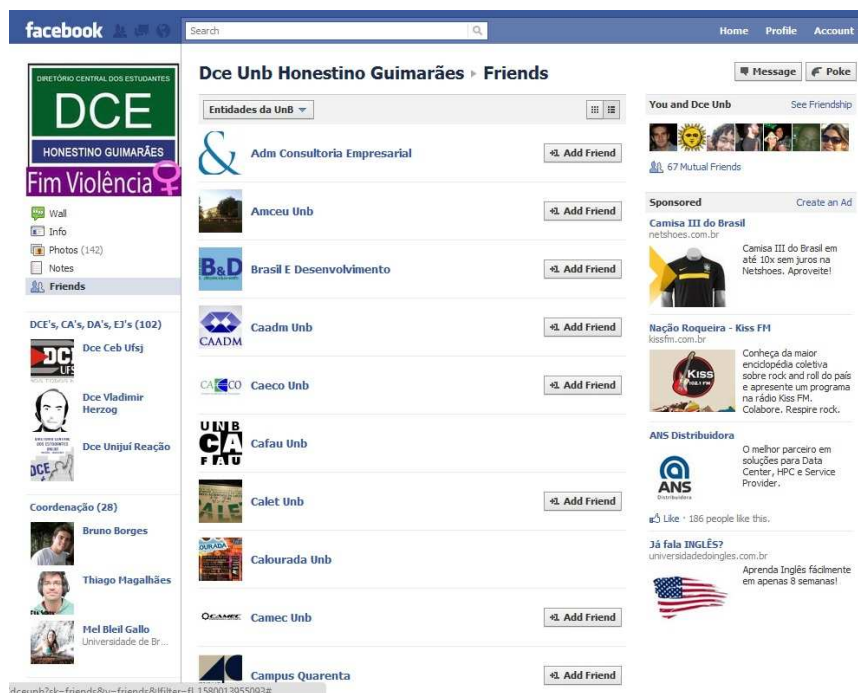


Figura 1.10: Entidades da UnB mostradas no Perfil do DCE UnB no Facebook

1.1 MOTIVAÇÃO

De acordo com as estatísticas apresentadas acima, nota-se a grande relevância de RSOs ao redor do mundo e na organização educacional do autor deste trabalho: a Universidade de Brasília. Apesar de tanta relevância os estudos sobre RSOs ainda estão na sua infância, segundo Fabrício Benevuto professor da UFOP [1].

As principais motivações para se estudar RSOs são relativas as oportunidades comerciais, as pesquisas sociológicas, as sugestões de melhorias nos sistemas atuais. Segundo o site Mashable, serão investidos \$3,8 em propagandas dentro de RSOs no ano de 2011 nos EUA. Em 2009, a RSO Facebook teve lucro de \$200 milhões e a tendência é melhorar. Segundo o site <http://www.syscomminternational.com>, mais de 60% das empresas no Reino Unido estão planejando fazer campanhas para aumentar sua presença em RSOs. Tradicionalmente, o meio de pesquisa sociológica é a coleta de informações em redes sociais reais é através de entrevistas com formulários. As RSOs são facilitadoras de estudos sociológicos, já que o espaço amostral de pessoas (usuários) a serem "entrevistada" é maior e os dados podem ser coletados de forma mais fácil e precisa. Pode-se utilizar métodos de coleta e análise de dados em RSOs para realização destas pesquisas. As possíveis melhorias nos sistemas atuais são relacionadas ao design, a usabilidade, a segurança, *spamming*, a vulnerabilidades a novas tendências (*Hot today, Gone tomorrow* [2]), a otimização do volume de tráfego gerado por RSOs em uma rede. Percebe-se então, várias sub-áreas de estudo derivadas do tema

Redes Sociais Online. Na próxima seção são apresentados alguns trabalhos relacionados. Os dois últimos trabalhos descritos são destaques motivacionais uma vez que são bastantes similares a este, com exceção de que nestes trabalhos os dados não precisam ser coletados, mas somente analisados. Estes trabalhos relacionados serviram de grande motivação e inspiração para o desenvolvimento deste trabalho, já que mostraram mais ainda a relevância do tema do trabalho e serviram como guia de estudos ao autor deste trabalho. A principal motivação para o desenvolvimento da implementação deste trabalho foi a oportunidade de se explorar e mapear a presença da Universidade de Brasília na Facebook. Dessa maneira, a partir da análise dos dados coletados poder auxiliar as diversas entidades da UnB na tomada de decisões. Trata-se de uma motivação altruísta, uma vez que o autor deste trabalho poderá contribuir de alguma forma para a universidade que ele estudou. Esta motivação levou a construção coletor e analisador de informações dos alunos e ex-alunos da Universidade de Brasília (UnB) presentes na RSO Facebook. Este aplicativo faz jus ao tema principal do trabalho. A coleta de dados é feita através da Graph API do Facebook Developer Platform, uma vez que os outros métodos de coleta seriam mais trabalhosos de serem implementados. Após a coleta, os dados são analisados de modo a descobrir padrões.

1.2 TRABALHOS RELACIONADOS

A seguir, são apresentados trabalhos importantes relacionados a área geral de RSOs, seguidos por trabalhos relacionadas as sub-áreas de RSOs citadas anteriormente com foco em coleta e análise de dados. Por último, são apresentados trabalhos relacionados a coleta e análise de dados com algum foco sociológico. No artigo *The Future of Social Networks on the Internet - The Need for Semantics*, John Breslin e Stefan Decker resumizam algumas características inerentes de RSOs e essenciais para o bom funcionamento destas como os objetos centrados em sociabilidade. Aborda-se também sobre a necessidade de semântica em RSOs, tema foco do artigo. No artigo *A General Architecture of Mobile Social Network Services*, Yao-Jen Chang et al. abordam as possibilidades de serviços de redes sociais em dispositivos móveis (Em inglês, MSNSs - Mobile Social Network Services), a arquitetura necessária para o funcionamento deste tipo de rede social e as oportunidades e utilidades desta tecnologia. No artigo *Professional Networking on the Internet*, B. A. Lloyd realiza uma pesquisa sobre as RSOs profissionais como Xing e LinkedIn. São mostradas funcionalidades gerais deste tipo de RSO como a procura de empregos e a seleção de profissionais qualificados por empresas ou agências de RH. No artigo *An Architecture for Applying Social Networking to Business*, Dr. Andres Fortino e Aparna Nayak apresentam os serviços de RSOs úteis

para o ambiente empresarial como ferramentas colaborativas, mensagens instantâneas entre outros. São também expostas precauções no uso destes serviço para evitar falhas de segurança e o mal uso das RSOs. A seguir, são apresentados alguns trabalhos relacionados a coleta e análise de dados em RSOs. No artigo *Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors*, Takehi Sakaki et al. mostram utilidades e vantagens do uso da RSO Twitter para detecção em tempo real de terremotos. Mostra-se um caso de estudo no Japão em que foi construído um sistema que coleta e analisa (filtra) tweets, detectando se houve algum terremoto. Este sistema notifica usuários cadastrados via email. É mostrado ainda que as informações chegam a estes usuários de forma muito mais rápida que notícias televisionadas pela JMA (*Japan Meteorological Agency*). No artigo *Measurement and Analysis of Online Social Networks*, Alan Mislove et al. análise de grande quantidade de dados coletados das RSO Flickr, LiveJournal, Orkut e YouTube. São feitas análises com objetivo de medir a quantidade de dados coletados e entender algumas propriedades da estrutura estática destas RSOs. Aborda-se os desafios enfrentados como a lentidão do processo de coleta e a dificuldade de se coletar completamente todas informações de uma RSO. Também são abordados dados sobre a infraestrutura de coleta e os resultados e conclusões como a comprovação de propriedades *small-world*. No artigo *Detecting Spammers on Twitter*, Benevuto et al. mostram maneiras que *spam* são feitos na RSO Twitter e criam um método de detecção de *spammers*. O tema de estudo principal deste trabalho é a coleta e análise de dados com foco sociológico, uma vez que serão coletados e analisados dados de forma a descobrir um padrão de perfis de usuários. Dessa maneira, serão apresentados a seguir trabalhos relacionados a este tema. No Artigo *Characterizing User Behavior in Online Social Networks*, Benevuto et al. caracterizam o comportamento de usuários em cinco RSOs, incluindo o Orkut, a mais caracterizada. Os dados são coletados através de um agregador de RSOs e são feitas algumas caracterizações como frequência de acesso a estas redes sociais, tempo que usuário passa na RSO, duração da sessão por RSO, número de usuários online em determinados momentos, entre outros. O estudo também confirma características de redes sociais online de amigos: *small-world*, *power-law* e *scale-free* Estas características remontam ao campo de estudo de redes complexas inerentes a redes sociais (online). Este campo de estudo é usado no estudos de RSOs como forma de analisar vários fenômenos como propagação de informação. Faz-se isto através de propriedades estatísticas e métricas que serão abordados no capítulo Embasamento teórico. Este artigo trás contribuições para melhora do design do sistemas atuais, deixa oportunidade de estudos sociológicos relacionadas a área de marketing viral e mais: o entendimento do comportamento do usuário e consequentemente do tráfego gerado permite remodelar a redes de comunicação de uma organização de forma a otimizar o tráfego. No artigo *Understanding Online Social Networks Usage from a Network Perspective*, Fabian Schneider et

al. fazem um estudo similar ao citado anteriormente nas RSOs Facebook, StudiVZ, LinkedIN e HI5, a partir da reconstrução de transações e sessões utilizando coletores localizados em servidores proxys. No artigo *Measuring User Influence in Twitter: The Million Follower Fallacy*, M. Cha et al. medem a influência de usuários no Twitter analisando se a influência de um usuário se mantém através de tópicos diferentes e quais comportamentos fazem do usuário comum um usuário influente. Algumas medidas de usuários como grau de entrada(*indegree*), quantidade de retweets, menções e tweets são coletados e analisados. A partir da análise do dados coletados descobriu-se que usuários não ganham influência espontaneamente ou de forma acidental, mas através de combinados esforços como abordar sobre um único tópico no conteúdo dos *tweets*. No artigo *Corporate espionage via social networks*, Zeljka Zorz descreve o *talk Mapping an Organization's DNA Using Social Media* de Abhilash Sonwane na RSA Conference de 2011 em São Francisco. Abhilash Sonwane demonstrou que coletando-se e analisando-se informações reunidas no Facebook, Twitter e LinkedIN e outras RSOs é possível obter diversas informações sobre pequenas e médias empresas localizadas ao redor do mundo. As informações coletadas são relacionadas a quem tomam as decisões, como as atividades são coordenadas, como conhecimento é transferido, quais os objetivos e incentivos de funcionários, a estrutura geral de cargos da empresa entre outras informações. Dessa forma, descobriu-se o *O DNA organizacional* completo, relacionado a todas as informações descritas anteriormente, da metade das empresas analisadas. No artigo *On the Evolution of User Interactions in Facebook B.*, B. Viswanath et al. medem a força de "links sociais", ou seja, medem quão interagida é um determinado grupo de usuários da RSO Facebook a partir de várias análises de dois usuários desta RSO através da coleta e análise de dados destes usuários. No artigo *Comparing Community Structure to Characteristics in Online Collegiate Social Networks*, Amanda L. Traud et al. examinam o gráfico social de usuários da RSO Facebook que são estudantes de uma das cinco Universidades analisadas. Não houve necessidade de coleta dos dados estudados, já que estes foram recebidos por um funcionário do Facebook. Analisa-se a estrutura da rede de cada universidade, incluindo como os alunos destas universidades interagem entre si e com outras universidades. No artigo *Tastes, ties, and time: A new social network dataset using Facebook.com*, Kevin Lewi et al. fazem uma análise de dados públicos de estudantes da RSO Facebook classificando-os em subgrupos definidos por gênero sexual, raça/etnia, atividade online e *status* socioeconômico. As conclusões do trabalho mostram potenciais científicos e pedagógicos desta RSO e provê uma base para futuras análises.

1.3 OBJETIVOS

Nesta seção são apresentados os objetivos gerais e específicos do trabalho. O principal objetivo deste trabalho é mostrar as utilidades e vantagens de métodos de coleta e análise de dados em RSOs. A utilidade destes métodos é comprovada através da análise dos dados coletados (*Data Mining*). As vantagens destes métodos em RSO são relacionadas a facilidade da coleta de grandes quantidades de dados. Este objetivo será alcançado por meio da implementação deste trabalho, um aplicativo descrito anteriormente na seção Motivação deste capítulo. Espera-se também, que a teoria, a implementação e os resultados deste trabalho incentivem os estudantes e professores da Universidade de Brasília a iniciar estudos sobre redes sociais nas suas respectivas áreas de interesse, como por exemplo marketing, ferramentas educacionais, *Business Intelligence* etc. Um dos objetivos específicos deste trabalho é mostrar alguns detalhes da construção de aplicativos de coleta e análise de dados em RSOs, com foco no Facebook. Outro objetivo específico a longo prazo é mapear de forma plena a presença online da UnB na RSO Facebook. Estes objetivos serão atingidos seguindo a estrutura do trabalho abordada na próxima seção.

1.4 ORGANIZAÇÃO DO TRABALHO

Neste trabalho descreveremos características de RSO em geral e da RSO Facebook em específico, apresentaremos conceitos de redes complexas, resumiremos as principais técnicas de coleta e análise de dados em RSOs entre outros no próximo capítulo, Embasamento Teórico. Logo após no capítulo Implementação, descreveremos o Aplicativo Desenvolvido, Tecnologias Utilizadas, Metodologia de Desenvolvimento e Gerenciamento, Estratégias para atingir o Público Alvo No aplicativo no capítulo Implementação. Seguindo pelo capítulo Resultados. Enfim, chegamos a uma conclusão e propomos algumas melhorias para este trabalho, no capítulo conclusão.

2 EMBASAMENTO TEÓRICO

Aqui serão apresentados conceitos sobre a rede social facebook, facebook platform detalhes da API facebook. Os conceitos apresentados aqui são relacionados com os trabalhos de Benevenuto [1] [2] [?].

2.1 REDES SOCIAIS

O termo original de redes sociais é ligado a área da sociologia: "conjunto de relações sociais entre um conjunto de atores e também entre os próprios atores. Designa ainda os movimentos pouco institucionalizados, reunindo indivíduos ou grupos numa associação cujos limites são variáveis e sujeitos a reinterpretações" (COLONOMOS, 1995). Atualmente, o termo redes sociais é articulado a tecnologia de informação em referência às redes sociais online (RSO). As RSO também são chamadas de redes sociais virtuais. Em inglês, dois termos bastante utilizados são SNS (*Social Networking Sites*) e OSN (*Online Social Networks*). Adotaremos o seguinte conceito de RSO: um serviço web que dá a um indivíduo a possibilidade de construir um perfil público ou semi-público; articular uma lista de conexão com outros usuários com os quais este indivíduo compartilha conexões; visualizar e percorrer suas listas de conexões, assim como as listas de conexões de outros usuários que o indivíduo esteja conectado ou do sistema, caso estas listas de conexões sejam públicas. [] Existem várias categorias de redes sociais. Na figura XX, são apresentadas algumas das principais RSO de acordo com sua categoria. Segundo Benevenuto, as funcionalidades

Nome	Propósito	URL
Orkut	Amizades	http://www.orkut.com
Facebook	Amizades	http://www.facebook.com
MySpace	Amizades	http://www.myspace.com
Hi5	Amizades	http://www.hi5.com
LinkedIn	Profissionais	http://www.linkedin.com
YouTube	Compartilhamento de vídeos	http://www.youtube.com
Flickr	Compartilhamento de fotos	http://www.flickr.com
LiveJournal	Blogs e diários	http://www.livejournal.com
Digg	Compartilhamento de (<i>bookmarks</i>)	http://digg.com
Twitter	Troca de mensagens curtas	http://twitter.com
LastFM	Compartilhamento de rádio/músicas	http://www.last.fm

Tabela 2.1. Algumas Redes Sociais Online Populares

Figura 2.1: Algumas das principais RSO categorizadas

mais relevantes oferecidas pelas principais RSO atuais são as seguintes

- Perfis de Usuários: contêm informações como nome, idade, gênero sexual, localização e uma foto de perfil.
- Atualizações: formas efetivas de ajudar usuários a descobrir conteúdo compartilhado por outros usuários na RSO. Dessa maneira, o usuário não vai atrás da informação, é a informação que vai até ele de forma personalizada.
- Comentários: como o próprio nome diz são comentários que podem ser feitos relacionados ao conteúdo compartilhado por outros usuários na RSO.
- Avaliações: formas de avaliar o conteúdo compartilhado por outros usuários. Tem diferentes níveis de granularidade e formas dependendo da RSO. Como exemplos, na Facebook tem-se o botão "curtir" e na Youtube tem-se um sistema de avaliação binária "gostei deste" e "não gostei deste".
- Lista de favoritos: maneira de organizar o objetos sociais (todos tipo de conteúdo compartilhável da web). São úteis para recomendações sociais. Na Facebook tem-se "fan pages", na Youtube e no Flickr tem-se vídeos e fotos favoritas, respectivamente.
- Lista de mais populares: lista de conteúdos, objetos sociais, usuários mais populares da RSO.
- Metadados: elementos essenciais para identificação de conteúdo, objetos sociais, usuários. O uso de metadados mais comum é relativo a marcações de usuários na Facebook ou em um comentário.

2.2 REDES COMPLEXAS

A base para o entendimento de RSOs no que diz respeito a identificação de padrões e mensuração de dados está presente na Teoria de Redes Complexas. Esta teoria provê uma forma de modelar a natureza onde as propriedades de um elemento são resumidas às conexões que ele estabelece com outros elementos do mesmo sistema. De acordo com esta teoria, uma rede é um conjunto de elementos que são chamados de nós ou vértices e das conexões entre estes elementos, chamados de arestas. Assim, a estrutura de uma rede pode ser modelado por um grafo, como mostrado na figura XX. Este grafo pode ser caracterizado a partir de diversas métricas. A seguir, são apresentadas as propriedades estatísticas e métricas em redes complexas. Estas são utilizadas para analisar e classificar redes complexas.

- Grau dos vértices: quantidade de ligações que apontam para um dado nó. Em RSO, representam o número de amigos ou seguidores de uma pessoa. Uma característica importante de uma rede é a

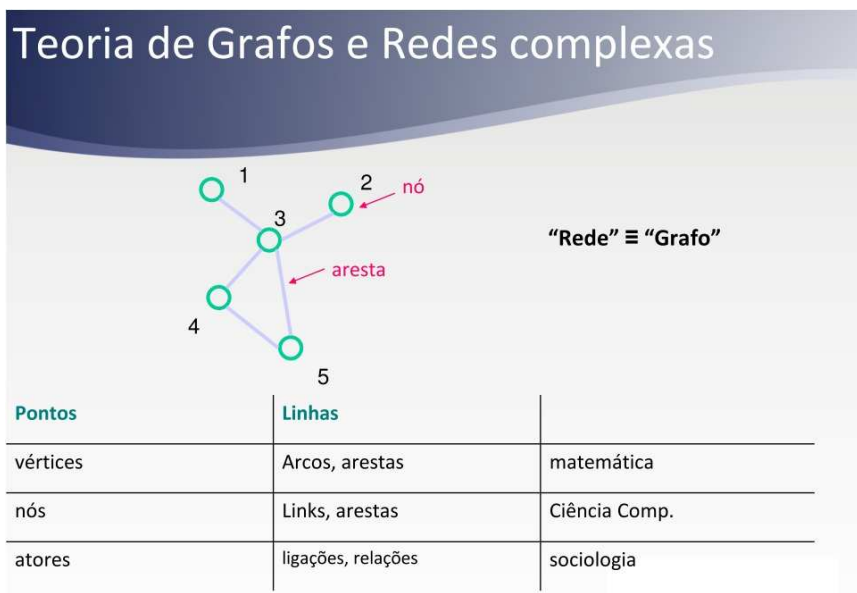


Figura 2.2: Similaridades entre Redes e Grafos

distribuição dos graus dos nós desta.

- Coeficiente de agrupamento: também chamado de coeficiente de aglomeração ou *clusterizing*, é a razão entre a quantidade de arestas existentes entre os vizinhos de um nó e número máximo de arestas possíveis entre estes vizinhos. O coeficiente de agrupamento global de uma rede é a soma de dos coeficientes de agrupamento de cada nó.
- Componentes: um conjunto de nodos em grafo, onde cada nodo possui um caminho para todos os outros nodos deste conjunto. Um componente é chamado de fortemente conectado (*SCC - Strongly Connect Component*, em inglês) quando cada nó dentro do componente pode ser alcançado de outro nó do componente seguindo arestas orientadas. Um componente é chamado de fracamente conectado (*WCC - Weakly connected components*) quando cada nó pode ser alcançado a partir de qualquer outro nó seguindo arestas em qualquer direção.
- Distância Média: é o número médio de arestas em todos os caminhos mínimos existentes entre todos os pares de nodos do grafo. Entende-se por caminhos mínimos a menor distância entre dois nós.
- Diâmetro: a distância do maior caminho mínimo existente ou a maior distância geodésica no grafo.
- Propriedades Assortivas: típicas de redes mensuradas pela tendência que nós com um certa quantidade de graus de uma rede tende a se conectar outros nodos com certa quantidade de nós. Geralmente, esta propriedade é estimada a partir da função $k_{nn}(k)$ que é a média do grau dos vizinhos dos nós de grau k .

- *betweenness*: grau de centralidade entre nodos ou arestas definido como o número de caminhos mínimos entre todos os pares de nodos em um grafo que passam por estes nodos ou arestas.
- Reciprocidade: de forma simples, é a porcentagem dos nodos apontados por i que apontam para este.
- PageRank: é basicamente um algoritmo iterativo que assinala um peso numérico para cada nodo com o propósito de estimar sua importância relativa no grafo.

Tipos de Redes

- Small-World: se esta tiver duas propriedades básicas que são coeficientes de agrupamento alto e diâmetro pequeno.
- Power-Law: redes cujas as distribuições dos graus dos nodos seguem uma lei de potência.
- Livre Escala: uma classe de redes que seguem leis de potência caracterizadas pela seguinte propriedade: nodos de grau alto tendem a se conectar a outros nodos de grau alto, ou seja, apresentam propriedades assortivas.

2.3 TÉCNICAS DE COLETA E ANÁLISE DE DADOS

É possível se obter dados de usuário em rede sociais de três maneiras. Uma das maneiras é via agregadores de tráfego, ou seja, proxies ou websites centralizadores de redes sociais. No primeiro caso, coletam-se os dados que passam por um provedor de serviços Internet (ISP) e filtrar as requisições que correspondem a acessos às redes sociais online. Aqui pode-se obter características do tráfego como volume gerado, das sessões dos usuários, os tipos de conteúdo transferido em cada sessão e o comportamento de usuários em redes sociais, a partir dos dados de navegação dos usuários. No segundo caso, coletam-se os dados que por meio de websites centralizadores (agregadores) de redes sociais. Este websites são sistemas que auxiliam usuários a gerenciar contas de diversas redes sociais. Dessa maneira, ao se conectar a estes sistemas um usuário pode acessar suas contas através de uma única interface, assim não precisando se conectar em todos os sites de redes sociais separadamente[]. Na figura XX, temos o esquema de um agregador de RSOs. Utilizando este método obtém-se os mesmos tipos de dados que utilizando um servidor proxy. Obtém-se os dados de usuários diretamente por meio de redes sociais online, quando estes dados são públicos, portanto passíveis de coleta ou quando estes são fornecidos de forma anonimizada por uma rede social. O segundo caso é raro, servidores de RSO dificilmente disponibilizam seus dados.

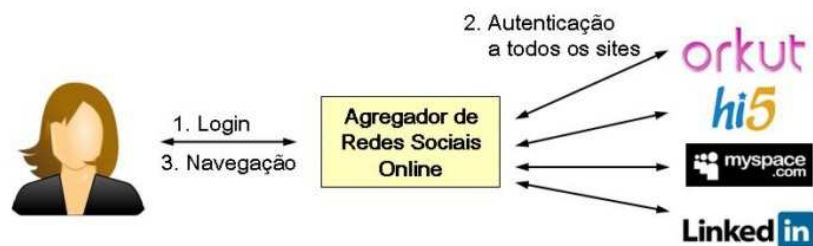


Figura 2.3: Esquema de um Agregador de RSOs em que Usuário se Conecta a Múltiplas RSO Simultaneamente

Logo, recorre-se aos *crawlers* ou *robots* que são ferramentas automáticas que coletam informações públicas de usuários de forma sistemática. Dificilmente coleta-se os dados de todos usuários da RSO. Isto ocorre devido a lentidão no processamento e armazenamento destes dados e na dificuldade de varrer toda a rede, ou seja, encontrar todos os usuários da rede. Então realiza-se coletas por amostragem. Como dito, anteriormente este tipo de coleta exige grande quantidade de processamento e espaço de armazenamento. Daí, tem-se a necessidade de utilizar coletores distribuídos em diferentes máquinas. Este tipo de utilização evita que as redes sociais interpretem a coleta de dados públicos como um ataque a seus servidores. Os dados públicos em redes sociais são atualizados assim que usuários publicam estes dados, portanto os dados em redes sociais são passíveis de serem coletados em tempo real. O Twitter por exemplo disponibiliza inclusive uma API que facilita a coleta de *tweets* públicos. A coleta de dados também pode se dar com o uso de aplicações de terceiros, ou seja, dados que foram coletados via aplicativos de RSOs. Estes aplicativos geralmente utilizam alguma API provida pela RSO que exige que o usuário autorize a coleta. Abordaremos mais sobre aplicativos de terceiros na seção Facebook Platform. Depois da coleta dos dados estes devem ser tratados para se realizar várias análises diferentes em cima destes. As principais análises podem ser categorizadas em identificação de padrões nos dados do perfil usuários, identificação em cima do conteúdo que o usuário gera na RSO e identificação de padrões de comportamento dos usuários. Estas análises consistem basicamente na filtragem dos dados coletados. Neste trabalho fazer-se-á análises com foco na identificação de padrões nos dados do perfil de usuários da RSO Facebook utilizando um aplicações de terceiros.

2.4 FACEBOOK

A Facebook é a maior RSO do mundo com presença em todos os continentes contando com mais de 750.000.000 de usuários ativos. Todos usuários passam cerca de 700 bilhões de minutos por mês no

Facebook, estão conectados 900 milhões de objetos sociais (páginas, eventos, grupos, comunidades) e compartilham mais de 30 bilhões de partes de conteúdo (links, notícias, blogs, anotações, álbuns etc). Cada usuário tem cerca de 130 amigos, está conectado a 80 pges (página, grupo ou evento), cria 90 "pedaços de conteúdo" ("curtidas", comentários, compartilhamento de conteúdo, etc). A Facebook é composta por objetos sociais que são nodos que representam páginas, usuários, conteúdo compartilhado, fotos, álbuns etc. Os dois primeiros objetos sociais citados são os principais pois representam pessoas e objetos do mundo real. Estes dois objetos sociais contém propriedades e conexões que são apresentadas na figura XX. As propriedades são informações presentes no perfil de um usuário ou página. As conexões são

Users		Pages	
Properties	Connections	Properties	Connections
Name	Feed	Name	Feed
Picture	Friends	Picture	Tagged
Locale	Likes	Category	
About	Activities		
Bio	Interests		
Hometown	Music		
Location	Books		
Website	Checkins		
Work	Permissions		
Link			
...etc			

Figura 2.4: Propriedades e conexões de Usuários e Páginas

relativas ao conteúdo que estes interagem. A partir das propriedades e conexões citadas na figura, nota-se que usuários geralmente representam pessoas reais e páginas geralmente são objetos do mundo real, como filmes e livros. Uma página do Facebook pode representar um gosto do usuário, como uma música ou show televisivo, a universidade que ele frequenta, os aplicativos que ele tem instalado, o time que ele torce, um interesse ou hobby do usuário etc. Tudo isto é classificado de acordo com a categoria da página. É a partir desta categorização e das conexões que um usuário tem com estas páginas que chegam aos usuários propagandas segmentadas.

Depois que um usuário se cadastra e se loga na Facebook, ele é redirecionado para a tela inicial, cuja a URL é www.facebook.com, que é mostrada na figura XX. O *layout* da página inicial é dividido em áreas que permitem compartilhar conteúdo, visualizar todos os seus amigos, seus grupos, seus amigos aniversariantes do dia, o que seus amigos compartilham, etc. A parte que mostra os pedaços de conteúdo compartilhado por seus amigos é chamada de mural.

Na página do usuário, cuja a URL é <http://www.facebook.com/ppmendes>, que é mostrada na figura



Figura 2.5: Página Inicial da RSO Facebook

XX. Nota-se um layout parecido com o da página inicial, uma vez que toda estrutura da Facebook segue



Figura 2.6: Página Inicial da RSO Facebook

um modelo(*template*) padrão. Nesta página o usuário é capaz de visualizar e organizar o conteúdo do seu próprio perfil, como suas fotos, todos os pedaços de conteúdos mais populares da que este compartilhou ou que amigos compartilharam no seu mural. A parte que contém todos os pedaços de conteúdo do usuário Os tipos de conteúdo que um usuário pode compartilhar se resume a caracteres, links, fotos e vídeos. Pedaços de conteúdo podem ser compartilhados através dos canais de comunicação que são o mural, mural de amigos, mensagens privadas, páginas e grupos. Outro canal de comunicação é relativo as interações entre o conteúdo compartilhado, ou seja, "curtidas"e comentários em cima do conteúdo. Nota-se na figura XX, que a Facebook se encaixa na categoria "redes sociais de amigos". A Facebook tem todas as funcionalidades relevantes oferecidas pelas principais RSO atuais, citadas anteriormente na seção Redes Sociais, com

exceção de Lista dos mais populares. A seguir complementar-se-á a descrição destas funcionalidades com foco no Facebook.

- Perfis de Usuários: localização da moradia atual, localização de nascimento (naturalidade), breve descrição sobre o usuário ("sobre mim), línguas que falam, informações relativas ao trabalho e educação, estado civil, família, visão política e religiosa, músicas, livros, filmes, shows televisivos, esportes e atividades e interesses.
- Atualizações: se resumem basicamente ao mural e as notificações do usuário. O mural é dividido em pedaços de conteúdo compartilhados mais populares e mais recentes. As notificações são três basicamente: solicitação(ões) de amizade(s), mensagem(s) privada(s) recebida(s) e notificações gerais. As notificações gerais são relacionadas a eventos que o usuário foi convidado, a conteúdo publicado com marcações do usuário, a conteúdo publicada em um grupo que o usuário participa e a conteúdo publicado no seu mural.
- Comentários: são opções presentes em todos os conteúdos compartilhadas que o usuário tem a permissão de comentar.
- Avaliações: o botão "curtir" ou "recomendar" são formas de um usuário avaliar o conteúdo compartilhado dentro da Facebook ou compartilhar um conteúdo disponível na web de forma já avaliada para dentro da Facebook.
- Lista de favoritos: é relacionada principalmente as páginas do facebook que geralmente representam objetos reais. Quando um usuário avalia uma página da Facebook clicando no botão "curtir" este vira fã da página e esta passa a fazer parte do perfil deste usuário. As páginas da Facebook geralmente representam os principais gostos do usuário, atividades e interesses, organizações de trabalho ou educacionais. Assim, a lista de favoritos do usuário diz muito sobre sua vida.
- Metadados: são chamadas de marcações (*tags*, em inglês) na Facebook. O uso mais comum de metadados é relativo a marcações dos usuários em conteúdos compartilhados, como comentários, publicações e principalmente fotos.

Outro conceito importante relacionado ao Facebook é o *social graph* (gráfico social). O gráfico social é a base do Facebook que é composto por pessoas e todas as conexões que elas tem com tudo que elas se importam, ou seja, as propriedades e conexões de usuários e páginas como mostrado na figura 2.3. Neste gráfico os usuários e objetos sociais como fotos, publicações, páginas entre outros estão representando

por pontos (nós) na figura XX. Outro conceito interessante também é de Open Graph Protocol que permite

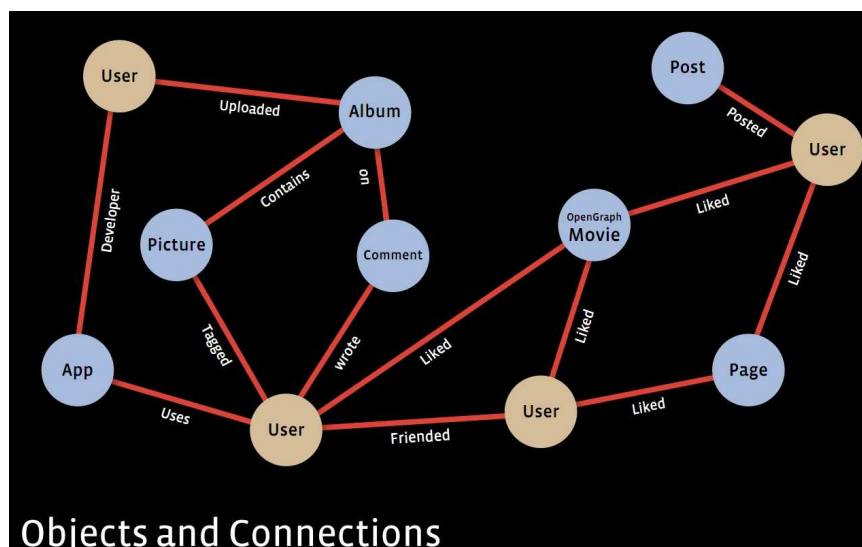


Figura 2.7: Gráfico Social do Facebook

ao usuário integrar um conteúdo web ao seu gráfico social. Este protocolo foi feito para páginas web que representam perfis de coisas do mundo real, como filmes, esportes, músicas, restaurantes, etc. Assim, quando um usuário clica no botão "curtir" de um conteúdo web, uma conexão é feita entre o conteúdo clicado e o usuário. Logo em seguida, este conteúdo web aparecerá nas informações do usuário no seu perfil do Facebook.

Para quem ainda não conhece a RSO Facebook, um guia geral para iniciantes ou pra quem quer fazer melhor uso da ferramenta pode ser encontrado no portal da Ig[.].

2.5 FACEBOOK DEVELOPER PLATFORM

Segundo o site de desenvolvedores do Facebook (<http://developer.facebook.com>), a *Facebook Platform* (Plataforma do Facebook, em português) é uma extensão do Facebook que tem como missão dar as pessoas o poder para compartilhar e fazer o mundo mais aberto e conectado. Uma definição que lembra uma das principais aplicações da web semântica (extensão da Web atual que permitirá aos computadores e humanos trabalharem em cooperação) que é a integração dos diversos e independentes "depósitos de dados" da internet numa aplicação coerente. Existem empreendedores e desenvolvedores espalhados por 190 países utilizando a Facebook Platform. A *Facebook Platform* está toda documentada no site de desenvolvedores do Facebook e tem três frentes que são websites, dispositivos móveis e aplicativos, sendo o último o foco deste trabalho. Esta plataforma pode ser resumida conforme mostrado na figura XX. A



Figura 2.8: Resumo da Plataforma do Facebook

frente de websites é a responsável por tornar um pedaço de conteúdo web, por exemplo um website, mais social integrando-o com o Facebook por meio de *plugins* sociais como botão "curtir", botão enviar, botão de login, plugin de registro entre outros. Existem 250 milhões de usuários ao redor do mundo que interagem por meio de plugins sociais com websites externos (ao Facebook), através do *Open Graph Protocol*. Existem mais de 2,5 milhões de sites integrados com o Facebook desde Abril de 2010 quando estes plugins foram lançados. Na figura abaixo são mostrados dois plugins do site da Folha de São Paulo [] A



Figura 2.9: Plugins Sociais da Folha de São Paulo

frente de dispositivos móveis provê funcionalidades do Facebook em dispositivos móveis. A *Facebook Platform* provê documentação para o desenvolvimento em iOS, Android e *mobile web* (acessada em navegadores de dispositivos móveis). Mais de 200 milhões de usuários acessam o Facebook todo mês através

dos seus dispositivos móveis. A frente de aplicativos como o próprio nome diz é relacionada a construção de aplicativos para o Facebook. Os aplicativos são componentes importantes do Facebook pois permitem a implementação de várias funcionalidades. Atualmente, existem 200 milhões de usuários mensais de aplicativos. A definição de aplicativo remete a softwares desenvolvidos para realizar tarefas específicas. Os aplicativos do Facebook permitem a integração com os objetos sociais do Facebook. Os aplicativos com maior número de usuários mensais são jogos (*gaming*). Existem também os aplicativos de utilidade social como por exemplo o aplicativo Friendmatrix (www.friendmatrix.co) que analisa o perfil de um usuário e gera uma foto formada pelo conjunto de fotos de perfis dos amigos deste usuários, sendo que os amigos que o usuário mais interagi, chamados de melhores amigos, ocupam maiores espaços na foto gerada. Todos os aplicativos coletam dados do perfil de usuários para gerar algum tipo de experiência social. Todavia, é necessário que o usuário permita que o aplicativo colete suas informações a partir de um diálogo de autorização (*oauth dialog*, em inglês) que é mostrado na figura XX, como exemplo. A especificação de quais permissões serão requisitadas ao usuário no diálogo de autorização são setadas no código do aplicativo. Neste diálogo são mostrados todos os dados do usuário que o aplicativo da Facebook pode coletar do usuário, ou seja, são setados no código do aplicativo todas as setenta permissões existentes. Após aceitar a coleta de dados, o aplicativo é instalado. As permissões necessárias para a coleta de dados dos usuários

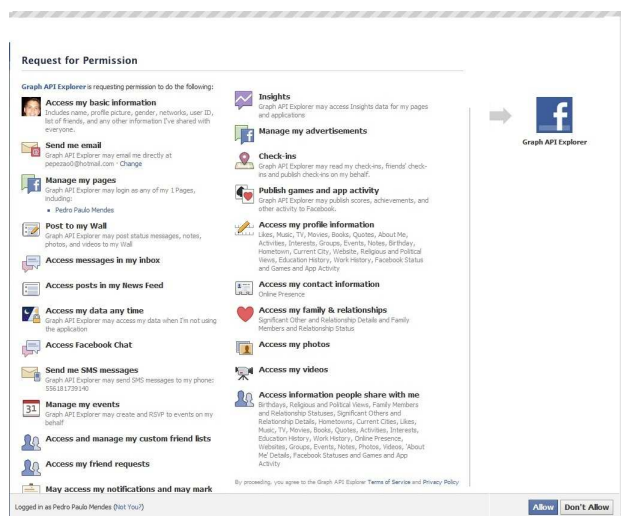


Figura 2.10: Diálogo de Autorização

podem ser divididas em três categorias: permissões do usuário, permissões de amigos e permissões estendidas. As permissões do usuário permitem que um aplicativo colete as informações relativas ao usuário que o instalou. As permissões de amigos permitem que um aplicativo colete informações relativas aos amigos do usuário que instalou o aplicativo. As permissões estendidas permitem a implementação de funcionalidades como leitura da caixa de entrada (*mailbox*), publicação de conteúdo, integração com bate-papo,

gerenciamento de páginas etc. As principais permissões do usuário e de amigos são resumidas na figura 2.10 e as principais permissões extendidas são resumidas na figura 2.11.

Permissão do usuário	Permissão dos amigos	Descrição
<code>user_about_me</code>	<code>friends_about_me</code>	Provê acesso a seção "sobre mim" (descrição do usuário com propriedade <code>\textit{about_me}</code>)
<code>user_activities</code>	<code>friends_activities</code>	Provê acesso a lista de atividades (propriedade <code>\textit{activities}</code>)
<code>user_birthday</code>	<code>friends_birthday</code>	Provê acesso a data de aniversário (propriedade <code>\textit{birthday_date}</code>)
<code>user_checkins</code>	<code>friends_checkins</code>	Provê acesso aos checkins (propriedade <code>\textit{checkins}</code>), relativo a localidade do usuário
<code>user_education_history</code>	<code>friends_education_history</code>	Provê acesso a histórico educacional (propriedade <code>\textit{education}</code>)
<code>user_events</code>	<code>friends_events</code>	Provê acesso a lista de eventos (conexão <code>\textit{events}</code>)
<code>user_groups</code>	<code>friends_groups</code>	Provê acesso a lista de grupos que um usuário é membro (conexão <code>\textit{groups}</code>)
<code>user_hometown</code>	<code>friends_hometown</code>	Provê a acesso a naturalidade (propriedade <code>\textit{hometown}</code>), onde o usuário nasceu.
<code>user_interests</code>	<code>friends_interests</code>	Provê a acesso a lista de interesses (conexão <code>\textit{interests}</code>).
<code>user_likes</code>	<code>friends_likes</code>	Provê acesso a lista de todas as páginas que o usuário "curtiu" (conexão <code>\textit{likes}</code>)
<code>user_location</code>	<code>friends_location</code>	Provê acesso a localização da moradia atual do usuário (propriedade <code>\textit{location}</code>)
<code>user_notes</code>	<code>friends_notes</code>	Provê acesso as anotações dos usuários (conexão <code>\textit{notes}</code>)
<code>user_online_presence</code>	<code>friends_online_presence</code>	Provê acesso a presença online do usuário que permite saber se um usuário está online ou offline no chat do Facebook
<code>user_photos</code>	<code>friends_photos</code>	Provê acesso as fotos que o usuário fez <code>\textit{upload}</code> e as fotos que o usuário está marcado.
<code>user_relationships</code>	<code>friends_relationships</code>	Provê acesso a lista dos usuários classificados como familiares, o estado civil do usuário e com quem ele está relacionando, caso esteja.
<code>user_relationship_details</code>	<code>friends_relationship_details</code>	Provê acesso as preferências de relacionamento (sexo masculino e/ou feminino).
<code>user_videos</code>	<code>friends_videos</code>	Provê acesso aos vídeos que o usuário fez <code>\textit{upload}</code> e os vídeos que o usuário está marcado.
<code>user_website</code>	<code>friends_website</code>	Provê acesso a URL do website do usuário
<code>user_work_history</code>	<code>friends_work_history</code>	Provê acesso ao histórico de trabalho (propriedade <code>\textit{work}</code>)

Figura 2.11: Permissões do usuário e de amigos

Um detalhe do mecanismo de autorização é que para um aplicativo acessar as informações do usuário no servidor do Facebook, este precisa de um código chamado de *access_token* que especifica quais dados o aplicativo pode capturar e por quanto tempo. Este mecanismo de autorização é um padrão aberto chamado de OAuth que também é utilizado nos aplicativos da RSO Twitter. O fluxo para coleta de da-

Permissão estendida	Descrição
email	Provê acesso ao email primário do usuário na propriedade email.
read_friendlists	Provê acesso a lista de amigos de um usuário. A lista de amigos é o modo como o usuário organiza seus amigos, por exemplo: amigos do futebol, Universidade etc.
read_insights	Provê acesso de leitura aos dados de Insights (estatísticas de acesso) de páginas ou domínios que pertence a um usuário.
read_mailbox	Provê acesso de leitura as mensagens (mailbox) de um usuário.
read_requests	Provê acesso as requisições de amizade de um usuário.
read_stream	Provê acesso de leitura a todas as publicações presentes no mural de um usuário.
xmpp_login	Permite que aplicação integre com o chat do Facebook e faça login na aplicação.
ads_management	Provê a habilidade de gerenciamento de anúncios no Facebook
create_event	Permite ao aplicativo criar e modificar eventos do usuário.
manage_friendlists	Permite que o aplicativo crie e edite listas de amigos do usuário
manage_notifications	Permite que o aplicativo leia notificações e as marque como lidas.
offline_access	Permite que o aplicativo acesse os dados do usuário a qualquer hora do dia ou da noite. Dessa maneira dá-se ao aplicativo um <code>access_token</code> de longa duração.
publish_checkins	Permite que o aplicativo faça checkins pelo usuário.
publish_stream	Permite que o aplicativo faça qualquer tipo de publicação de conteúdo pelo usuário
rspv_event	Permite que o aplicativo <code>RSVP</code> qualquer evento pelo usuário.
sms	Permite que o aplicativo envie e receba mensagens privadas (mailbox) por SMS.
publish_actions	Permite que o aplicativo publique informações de propriedades <code>scores</code> e <code>achievements</code>
manage_pages	Permite que o aplicativo administre páginas pertencentes a conta de um usuário.

Figura 2.12: Permissões do usuário e de amigos

dos no facebook é mostrado na figura 2.11. Para realizar a coleta de dados o aplicativo faz uso da Graph

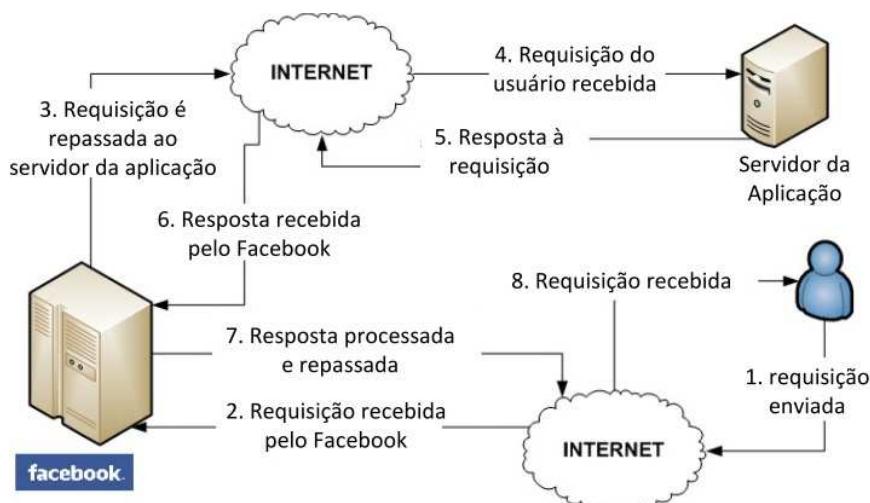


Figura 2.13: Fluxo de Coleta de Informações de Usuários do Facebook

API. Uma API(*Application programming interface*) é um conjunto de rotinas e padrões estabelecidos por um software para a utilização das suas funcionalidades por aplicativos que não pretendem envolver-se em detalhes da implementação do software, mas apenas usar seus serviços. No contexto de desenvolvimento web, as APIs são tipicamente um conjunto de tipos de requisições HTTP juntamente com suas respectivas definições de resposta e no caso do Facebook não é diferente. Portanto, a Graph API apresenta uma maneira simples e consistente de coletar propriedades e conexões de objetos sociais. Esta API faz isto representando uniformemente os objetos sociais contido no gráfico social. É importante ressaltar ainda que cada objeto tem um ID único. As informações são coletadas a partir de requisições que utilizam a URL `http://graph.facebook.com/ID_OBJETO`, sendo ID_OBJETO o identificador do objeto requerido. Após a requisição são retornados os dados do objeto pedido no formato *JSON*(JavaScript Object Name), conforme mostrado na figura 2.13. A utilização da Graph API é dispendiosa. Ocorre desperdício de banda e lentidão no processo de coleta, já que algumas vezes são necessárias várias requisições HTTP para coletar todas as informações necessárias. Além de exigir maior processamento na análise dos dados coletados, uma vez que estes dados não estão filtrados. Ameniza-se estes problemas fazendo uso da *Facebook Query Language*(FQL) que permite o uso de uma interface de estilo SQL para coletar e filtrar as informações expostas pela GRAPH API. Esta interface permite o processamento de várias consultas aos dados que se pretende coletar em uma única requisição. Assim os dados já retornam filtrados e sem necessidades de outras requisições. Para facilitar e orientar o desenvolvimento de aplicativos são providos SDKs. SDK é o acrônimo de Software Development Kit que normalmente são disponibilizados por empresas para que programadores externos tenham uma melhor integração com o software proposto. Atualmente, estão disponíveis quatro

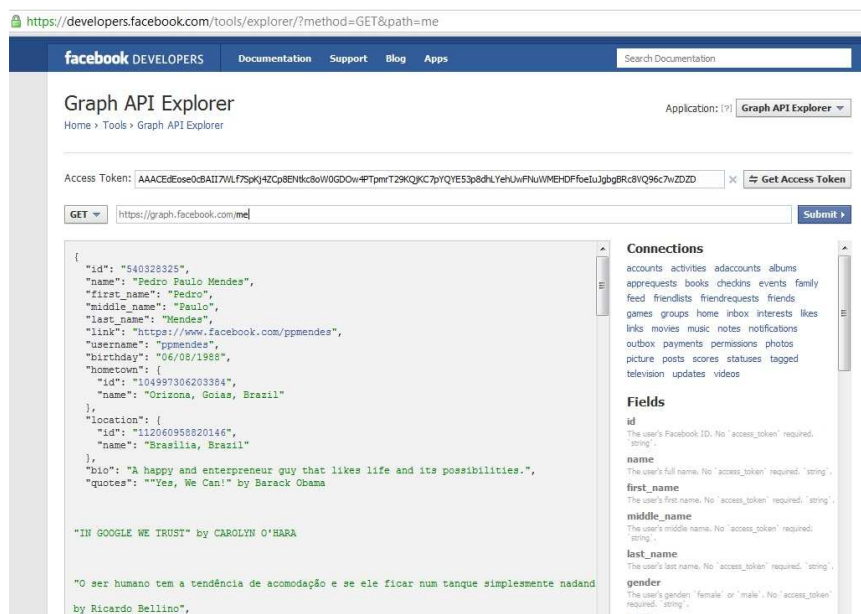


Figura 2.14: Exemplo de Resposta de Requisição da Graph API

SDKs. O javascript SDK que permite o acesso, do lado do cliente, a todas as características da Graph API e diálogos via JavaScript. O PHP SDK que contém bibliotecas que auxiliam na implementação de *login/logout* e suporte a GRAPH API. As outras duas SDKs auxiliam no desenvolvimento de aplicativos para dispositivos móveis: iOS SDK (iPhone & iPad) e Android SDK. São também providas as ferramentas que são usadas para desenvolver, testar e monitorar os aplicativos como a *Developer App* e a Graph API Explorer. A ferramenta *Developer App*, mostrada na figura XX, é usada para criar e gerenciar uma aplicação. A ferramenta Graph API Explorer testa a quais dados são coletados, a partir de quais requisições e com quais permissões. Outra ferramenta de tecnologia antiga REST API, mas que é de muita serventia é o console de teste fql.query, muito útil para testar consultas as tabelas FQL.

Expôs-se nesta seção apenas tecnologias atuais do Facebook. As tecnologias antigas serão substituídas por estas que foram expostas aqui anteriormente. Como exemplo de tecnologias antigas temos a Old REST API, a Legacy JavaScript SDK, FBML e FBJS.

Antes de se começar a fazer um aplicativo do Facebook é importante verificar se as intenções desse aplicativo são compatíveis com as Políticas de desenvolvimento de aplicativos para o Facebook [].

No próximo capítulo utilizaremos os conceitos descritos aqui para apresentar a implementação deste trabalho.

3 IMPLEMENTAÇÃO

Na implementação deste trabalho, põe-se em prática os conceitos apresentados no capítulo Embasamento Teórico, principalmente no que é relativo ao Facebook. Este aplicativo, como dito no capítulo Motivação, é de forma geral um coletor e analisador de dados da rede social online Facebook. As vantagens de se coletar dados utilizando uma API e os motivos da escolha da rede social online Facebook foram abordados no capítulo Motivação. Neste capítulo apresentar-se-á o aplicativo desenvolvido e serão abordados os principais temas relacionados ao aplicativo. Este capítulo está dividido em seis seções, abordadas na seguinte ordem: Descrição Geral da Implementação; Tecnologias Utilizadas; Regras de Negócio, Análise de Requisitos e Metodologias; Metodologias e Testes do Projeto; Estratégias para Atingir Público Alvo. Neste capítulo são também referenciados todos os anexos deste trabalho.

3.1 DESCRIÇÃO GERAL DA IMPLEMENTAÇÃO

A implementação deste trabalho é um software coletor e analisador de informações de estudantes da UnB presentes na RSO Facebook. Trata-se de um aplicativo da Facebook que se enquadra na categoria de aplicativos de utilidade social. Entende-se por "estudantes da Universidade de Brasília presentes na rede social online Facebook" usuários da rede social online Facebook que tem a página da Universidade de Brasília adicionada na parte de educação do seu perfil. Por conveniência, chamaremos de agora em diante este grupo de Rede UnB Facebook. O aplicativo funciona da seguinte maneira. As informações da Rede UnB Facebook são coletadas e salvas em um banco de dados. As informações salvas são analisadas de forma a identificar padrões através de um método de categorização (mineração de dados). Estas informações são coletadas do perfil do DCE-UnB no Facebook. As informações são coletadas por meio de um aplicativo do Facebook após devida autorização do usuário. Por final, as informações analisadas são apresentadas para o usuário final.

3.2 TECNOLOGIAS UTILIZADAS

O software foi desenvolvido com tecnologias atuais e agéis que eram dominadas pelo desenvolvedor do software, autor deste trabalho. Dessa maneira, buscou-se obter a melhor performance na coleta e análise

das informações, aumentar o tempo de vida e a escalabilidade do software. As tecnologias abordadas a seguir são divididas em Tecnologias do Facebook e Tecnologias Gerais.

3.2.1 Tecnologias do Facebook

No capítulo Embasamento Teórico, mostrou-se que o Facebook está em processo de mudança de tecnologias antigas, mas ainda funcionais, para tecnologias atuais e compatíveis com o resto da web. As tecnologias escolhidas para o desenvolvimento da implementação deste trabalho foram todas atuais: GRAPH API, FQL, Javascript SDK com OAuth 2.0, PHP SDK v3.0.1. O FQL foi implementado utilizando as funções "FB.Data.query" e "query.wait", providas pela Javascript SDK e que fazem consultas via Graph API. Estas funções juntamente com a função FB.api, também provida pela Javascript SDK, foram as responsáveis pela coleta dos dados.

Durante o desenvolvimento do aplicativo; foram realizados alguns testes utilizando o Console de Teste FQL.query e a ferramenta Graph API Explorer; ambos presentes no portal de desenvolvedores do Facebook. Na subseção Testes, apresentar-se-á mais detalhes sobre a fase de testes do aplicativo.

3.2.2 Tecnologias Gerais: Linguagens de Programação, Bibliotecas e IDEs

O aplicativo do Facebook foi desenvolvido utilizando um misto de três linguagens de programação diferentes. As linguagens foram escolhidas levando em consideração três pré-requisitos estabelecidos pelo desenvolvedor do aplicativo. Os pré-requisitos foram o domínio das linguagens a serem escolhidas pelo desenvolvedor do software (autor deste trabalho) e a clareza e a quantidade de informações úteis disponíveis na documentação oficial de desenvolvedores do Facebook e na web relacionadas as linguagens a serem escolhidas. Além da existência de um SDK, relacionado as linguagens a serem escolhidas, que facilitasse a construção do aplicativo. Após serem analisados os requisitos, foram escolhidas as seguintes linguagens de programação: PHP, Javascript e Actionscript. A escolha das duas primeiras linguagens citadas são justificados pelo fato do desenvolvedor ter domínio sobre estas, a documentação das mesmas para a construção de aplicativos no Facebook é clara e a maior existente. Além disso, estas linguagens tem SDKs que agilizam a construção de aplicativos, tornam a aplicação mais escalável e fácil de ser configurada. A escolha do ActionScript, linguagem proprietária da empresa Adobe, é justificada principalmente pela capacidade de se fazer interfaces gráficas de usuários ricas de forma rápida. Esta linguagem foi utilizada junto ao IDE Flash Professional CS5.5 Tiral para desenvolver os infográficos. Nota-se que nenhuma das linguagens sozinhas

atende plenamente as necessidades de desenvolvimento do aplicativo sozinhas, por isso a necessidade das três.

A linguagem de programação PHP foi utilizada na estruturação do aplicativo, gravação dos dados coletados no banco de dados, para prover autenticação do aplicativo, cobrar permissões do usuário, para acessar as informações de *signed request* do usuário entre outras utilidades.

A linguagem de programação Javascript foi utilizada em diversas frentes do software: Interface Gráfica do Usuário, Chamada de Requisições POST e algumas funções intrínsecas da API do Facebook. A Interface Gráfica do Usuário relativa as duas abas onde serão apresentados os infográficos foi criada utilizando-se a biblioteca javascript JQuery conhecida como jquery-ui.min.js (v1.8). As chamadas de requisições POST foram feitas utilizando a biblioteca javascript JQuery conhecida como jquery.min.js (v1.4). As funções intrínsecas da API do Facebook se referem a coleta de dados dos usuários, utilização de cookies e autenticação OAuth 2.0.

Durante o processo de codificação do aplicativo foi utilizado o IDE Zend Studio Trial (v8.0.1). Este IDE foi escolhido, pois foi premiado pela Infoworld 2011 como Melhor IDE PHP.

As tecnologias utilizadas pouparam tempo de desenvolvimento do software. Mais detalhes das linguagens de programação PHP e JavaScript estão no capítulo Embasamento Teórico. A estrutura de arquivos do projeto implementado e a descrição das formas de acesso e configuração do aplicativo no Facebook está disponível no anexo I e II, respectivamente.

3.2.3 Servidor de Hospedagem e de Domínio

O aplicativo está rodando em servidor de hospedagem configurado com Apache Web Server (v2.2.+); PHP (v5.3.6+) e PostgreSQL (v9.0.4+). Trata-se de um servidor de hospedagem alugado pela empresa Locaweb. O domínio do aplicativo é www.adsum.com.br/facebook/pedro3/ está registrado junto a FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) por intermédio da empresa RegistroBR, que realiza registros de domínios para Internet no Brasil.

3.3 REGRAS DE NEGÓCIO, ANÁLISE DE REQUISITOS E METODOLOGIAS

Após a realização de uma pesquisa sobre os principais métodos de coleta e análise de dados em redes sociais abordados no capítulo Embasamento Teórico, optou-se dentre várias opções pela construção de

um aplicativo para a rede social online Facebook. A tomada desta decisão levou a realização de mais uma pesquisa sobre como desenvolver aplicativos para o Facebook que resultaram nas seguintes decisões: utilização das Tecnologias do Facebook, Linguagens de Programação, Bibliotecas, IDEs e Servidor de Hospedagem com as características citados na seção passada deste capítulo.

3.3.1 Elaboração da Regra de Negócio e Análise de Requisitos

Após a realização da primeira pesquisa citada na subseção anterior foi definida uma parte da regra de negócio do aplicativo que está descrita na seção Descrição Geral da Implementação. Logo após, foi realizada uma segunda pesquisa, também descrita na subseção anterior, que definiu alguns requisitos do projeto. Esta pesquisa foi orientada aos pré-requisitos definidos pelo desenvolvedor. A análise de requisitos foi realizada de forma bastante intuitiva. Sabia-se que seriam usadas as *APIs* do Facebook. Portanto, decidiu-se fazer uso das Tecnologias do Facebook mais atuais. Sabia-se que o aplicativo coletaria e analisaria alguns dados de usuários do Facebook pertencentes ao grupo Rede UnB Facebook, porém não se sabia de forma detalhada quais dados seriam estes e quais seriam as análises feitas com base nos mesmos. Neste momento, o processo de análise de requisitos parou e o processo de desenvolvimento das regras de negócio recomeçou. Os dados e as análises foram definidos para complementar as regras de negócio. Esta parte das regras de negócio se encontram na subseção seguinte. Logo após a complementação das regras de negócio recomeçou-se a análise de requisitos da mesma maneira. Sabia-se que o ambiente de funcionamento do aplicativo desenvolvido é a web; portanto foram escolhidas Linguagens de Programação, Bibliotecas e IDEs para este ambiente de acordo com os pré-requisitos do desenvolvedor. Sabia-se que para coletar dados de usuários do Facebook são necessárias permissões, descritas no capítulo Embasamento Teórico, portanto foram listados as permissões necessárias para a coleta dos dados necessários. E assim por diante todo o escopo do processo de análise de requisitos e das regras de negócio foi definido. Foram demandadas cerca de 30 horas para a análise de requisitos; 100 horas para a elaboração das regras de negócio e 250 horas de codificação, num total de 380 horas. Neste relatório continham as datas e os horários que as atividades relativas ao desenvolvimento do software começavam e terminavam. Estas eram separadas em três sessões: análise de requisitos, regras de negócio e codificação. Uma observação importante é que o processo de análise de requisitos demandou apenas 8% do total de horas dedicadas ao projeto. Acredita-se que isto tenha acontecido, porque as regras de negócio do aplicativo e os pré-requisitos do desenvolvedor foram bastante específicos, deixando o espaço amostral de possíveis requisitos pequeno. Outra observação importante é que no processo de desenvolvimento deste software as análises de requisitos não dependeram

do fato de as regras de negócio estarem totalmente concluídas. Notou-se assim um desenvolvimento de software iterativo.

3.3.2 Regras de Negócio: Coletar e Analisar o que?

Como dito anteriormente; a implementação deste trabalho é um aplicativo que coleta e analisa alguns dados de usuários da Rede UnB Facebook. Nesta subseção são descritos de forma detalhada quais os dados que o aplicativo coleta e analisa e assim as regras de negócio do aplicativo são totalmente abordadas.

Nesta primeira versão do aplicativo, versão beta, coletar-se-á somente os dados relativos as informações do perfil dos amigos dos usuários. Estes dados não incluem o conteúdo gerado por estes usuários nem os dados do próprio perfil dos amigos destes usuários. É importante saber que somente os dados disponíveis serão coletados, ou seja, aqueles existentes no perfil do usuário. Os dados que serão coletados são descritos abaixo e estão classificados em quatro camadas.

- Principal

- nome; uid; quantidade de amigos (grau); gênero sexual; data de nascimento; naturalidade; localidade da moradia atual; email; visão política; religião; estado civil; website;

- Trabalho

- nome(s) da(s) instituição(ões) que trabalha ou que trabalhou; localidade(s) da(s) respectiva(s) instituição(ões); cargo(s) que trabalha ou que trabalhou na(s) respectiva(s) instituição(ões); data de entrada e data de saída na(s) respectiva(s) instituição(ões); instituição que está trabalhando atualmente e desde quando

- Educação

- nome(s) da(s) instituição(ões) que estuda ou que estudou; tipo da(s) instituição(ões), por exemplo, ensino superior; Ano de entrada na(s) respectiva(s) instituição(ões) área de concentração na(s) respectiva(s) instituição(ões)

- Páginas "curtidas"

- nome(s) da(s) seguinte(s) categoria(s) página(s) curtidas: filmes, músicas, livros e show televisivos.

Como abordado no capítulo Embasamento Teórico, as páginas representam os gostos do usuário e as categorias de páginas mais populares são relacionadas filmes e livros. É importante notar nesta subseção como os dados do perfil de um usuário no Facebook são bem organizados, justificando mais uma vez a escolha desta rede social online para a coleta de dados.

Na análise de dados, descrita abaixo nesta subseção, são mostrados os tratamentos que serão feitos com os dados coletados. As conclusões alcançadas a partir destes tratamentos serão abordadas no capítulo seguinte chamado Resultados Obtidos.

Serão feitos os seguintes tratamentos de dados: Descobrir-se-á padrões de usuários da Rede UnB Facebook categorizando estes padrões segundo o método de categorização explicado no capítulo Embasamento Teórico. Assim, a partir das informações salvas descobriremos várias "subredes" de acordo com os dados do perfil do usuário. Logo, existiram várias "subredes", como por exemplo, a "subrede" de organização educacional de ensino médio, a "subrede" de organização de trabalho, "subrede" de gênero sexual, "subrede" idade e assim por diante. É importante abordar aqui que também haverá a subrede de dados não preenchidos pelo perfil da pessoa, ou seja, dados que foram salvos como uma *string* sem caracteres ou nulos.

3.3.3 Principais Requisito para Coleta e Análise

A partir das regras de negócio descritas na subseção passada foram pesquisados os requisitos para continuar a construção do aplicativo. Abordar-se-á nesta subseção estes requisitos: a escolha das permissões para a coleta dos dados e alguns detalhes de tratamento destes dados. Para a escolha das permissões, fez-se uma simples associação de quais permissões seriam necessárias para coletar todas informações. Esta associação foi feita com base na tabela de permissões presente no portal de desenvolvedores do Facebook. Abaixo são apresentadas as 17 permissões necessárias para a coleta de dados que pretendida pelo aplicativo:

- user_about_me, friends_about_me, friends_likes, friends_online_presence, friends_relationships, friends_work_history, friends_activities, friends_education_history, friends_hometown, friends_location, friends_photo_video_tags, friends_videos, friends_birthday, friends_events, friends_interests, friends_photos, friends_religion_politics, friends_website, email

Todas as permissões citadas acima referem-se a dados exclusivos do usuário. Contudo, o aplicativo desenvolvido para este trabalho pede ao usuário mais permissões. Isto é feito para que seja possível em

versões futuras do aplicativo a coleta de outros dados do usuário, sem que haja necessidade de pedir estas informações para os usuários novamente. O resto das permissões pedidas são mostradas abaixo de forma categorizada:

- Permissões para Coleta do próprio usuário:
 - user_about_me, user_likes, user_online_presence, user_relationships, user_work_history, user_activities, user_education_history, user_hometown, user_location, user_photo_video_tags, user_videos, user_birthday, user_events, user_interests, user_photos, user_religion_politics, user_website
- Permissões extendidas
 - read_stream, off-line_access, online_presence, publish_stream

Estas permissões garantem a coleta dos seguintes dados do usuário: todos os dados citados na subseção anterior, mudando apenas o fato de que agora estes são relacionados aos amigos do usuário analisado; o conteúdo gerado pelo usuário (CGU) em tempo real; a presença online. Além disso, a permissão off-line_access permite que o aplicativo acesse os dados a qualquer hora do dia e a permissão publish_stream garante que o aplicativo gere qualquer tipo conteúdo como se fosse o usuário.

Os detalhes relativos ao tratamento dos dados são abordados nas seguintes situações. Deve-se extrair do dado coletado email apenas o domínio deste email, já que será apenas este o objeto a ser analisado. Outro caso que se enquadra neste requisito é quando um dado coletado não está preenchido no perfil de um usuário, ou seja, não existe. Este dado tem valor nulo ou indefinido. Se um dado com valor nulo ou indefinido for acessado pelo aplicativo ocorre um erro e o aplicativo para de funcionar. Dessa maneira, não é continuada a coleta e os demais dados já coletados não são gravados no banco de dados. Daí, a necessidade de verificação em todos os dados colhidos se estes existem.

3.4 MINERAÇÃO DE DADOS

O processo de Mineração de Dados, em inglês *Data Mining*, consiste em 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. Dessa forma, obter a partir dos dados operativos brutos, informação útil para subsidiar a tomada de decisão nos escalões médios e altos da empresa. O método de categorização consiste em agrupar perfis, dentre todos

já salvos, de acordo com uma ou mais informações colhidas. Por exemplo: onde a maioria dos alunos da Rede UnB Facebook estudou no ensino médio. Os dados categorizadas e o conhecimento gerado a partir destes dados que serão apresentados no próximo capítulo, Resultados.

3.5 METODOLOGIAS, PADRÕES E TESTES DO PROJETO

O aplicativo desenvolvido não seguiu nenhum *Design Pattern*(Padrões de Projetos) como o modelo *MVC*(*Model View Controller*) e nenhuma Metodologia de Desenvolvimento de Software como *RUP* (*Rational Unified Process*) ou utilização de Diagramas *UML*, (*Unified Modeling Language*). Além disso, não foram utilizadas Metodologias de Gerenciamento de Projetos de Software como *SCRUM* ou *XP* ou Metodologias de Testes. Dessa maneira, a construção deste aplicativo não faz uso de metodologias comuns em projetos de Software atuais e que são consideradas boas práticas da Engenharia de Software. O projeto foi feito dessa maneira; pois a parte de codificação do software não é complexa nem extensa. Além disso, o software foi desenvolvido por apenas uma pessoa. Dessa maneira, não houve a necessidade de seguir metodologias comuns do desenvolvimento de software e nem houveram dificuldades de desenvolvimento resultantes desta decisão. Os testes foram feitos em paralelo ao desenvolvimento do projeto e realizados pelo próprio desenvolvedor. Como o software não tinha muitos pontos de função e realizava apenas duas tarefas complexas que são a análise e a coleta de dados. Nesta fase de desenvolvimento do projeto, não houveram grandes dificuldades nem necessidade de grandes correções ou modificações de fluxos a partir da detecção de erros.

3.6 ESTRATÉGIAS PARA ATINGIR PÚBLICO ALVO

O público alvo do projeto (alunos e ex-alunos da UnB que tem perfil na RSO Facebook) foi atingido por meio da coleta de dados de amigos do perfil do DCE-UNB, mostrado na figura 1.8, que conta com 3490 "amigos", sendo que a maioria destes "amigos" alunos e ex-alunos da UnB como será mostrado no próximo capítulo, Resultados. Foi assinado um termo de contrato entre o autor deste trabalho e o DCE-UNB afirmando que os dados não seriam vendidos ou cedidas a outros projetos que não fossem também aprovados pela instituição. Outra cláusula acorda que os dados serviriam somente para identificar padrões no público alvo do projeto e nada além disso. Os dados de todo público alvo do projeto seria viável através do acesso a conta da página da Universidade de Brasília, mostrada na figura 1.7. Isto porque toda

vez que alguém adiciona a Universidade de Brasília na parte de educação do seu perfil, esta marcação (/textitmetadados) é notificada no perfil desta página. Assim, através do número de notificações recebidas sabe-se exatamente quantas pessoas fizeram a ação descrita anteriormente. Estima-se que existam mais de 36.000 alunos da UnB representados por perfis no Facebook. Esta estimativa é baseada na quantidade de membros da comunidade da "UnB - Universidade de Brasília" no Orkut, mostrada na figura 1.4, em que a maioria dos membros era aluno ou ex-aluno da UnB. A partir dessa base, leva-se em conta que como dito no capítulo Introdução hoje o número de usuários e acessos ao Facebook são maiores do que no Orkut portanto notou-se uma forte migração de usuários da primeira RSO citada para a segunda, ainda mais considerando os alunos da Universidade de Brasília que são mais "elitizados digitalmente".

4 ESTUDO DE CASO

Nesta seção, discutiremos os dados categorizadas que foram salvos no banco de dados e descreveremos o conhecimento que pode ser gerado a partir destes dados. Estima-se que o aplicativo tenha coletado cerca de 11% do espaço amostral universal da rede UnB Facebook, segundo estimativas mostradas na sessão "Estratégias para atingir público alvo" do capítulo anterior. Esta porcentagem é bastante significativa e serve para ser aproveitada como caso de estudo deste trabalho. Em pesquisas sociológicas comuns a porcentagem de dados colhidos raras vezes atinge esta porcentagem. O que revela maior acurácia destas pesquisas e portanto possibilidades de análises mais verossímilantes. Um exemplo disso é relativo as pesquisas eleitorais que atingi somente 0,00158421% do espaço amostral universal. Estatísticas: coletou-se dados de 3.960 pessoas do total de 36.000. No total temos 40.000 informações dos perfis dos usuários classificados entre informações pessoais e páginas "curtidas", sendo 10.000 são informações não preenchidas. A partir dos dados coletados fizemos diversos tipos de análises obtivemos resultados interessantes. Nas seções deste capítulo detalhar-se-á quais foram estes resultados e análises.

4.1 DISCUSSÃO DE RESULTADOS

A seguir serão apresentados os dados categorizadas em "subredes" conforme descrito no capítulo anterior. A partir destes dados teremos subsídios para fazermos algumas análises que geraram conhecimento útil. São apresentados somente os dois dados melhor "rankeados" na "subrede" quando existe a possibilidade de se cadastrar mais de dois dados. quantidade de amigos (grau); gênero sexual; data de nascimento; naturalidade; localidade da moradia atual; email; visão política; religião; estado civil; website;

- média da quantidade de amigos: 260
- Gênero sexual: Masculino(53%) e Feminino(45%); Não-preenchido: (2%)
- Idade: 17-18(20%); 19-21(30%) ; 21-23(24,5%) ; Outros (25,5%) ;
- Trabalho: Universidade de Brasília (2%); Outros: (40%) ; Não-preenchido(54%)
- Educação: Alub (6%); Galois (3,5%); Outros(30%); Não-preenchido(60%)
- Naturalidade: Brasília (60,1%) , Nordeste (15,3%) e Outros(8,7%) ; Não-preenchido(4%)

- Visão Política: Partido dos Trabalhadores (2,7%); Outros: (2,3%); Não-preenchido(95%)
- Religião: Católico(2,1%); Outros(2,9%); Não-preenchido(92,4%)
- Moradia Atual: Brasília(90,4%); Outros(4,1%); Não-preenchido(4,5%)
- Estado civil: Solteiro (40%¹); Algum tipo de relacionamento (2%) ; Não-preenchido(56%)
- Email: gmail (50%) ; hotmail (30%); Outros(10%) ;
- Website: preenchido (1%); Não-preenchido(97%)
- Filmes: The GodFather (6%); Tropa de Elite (4%); Outros (15%); Não-preenchido(75%)
- livros: O Pequeno Príncipe (3%) ; Outros (30%); Não-preenchido(77%)
- Músicas: Black Eyed Peas (7%); Jorge Ben Jor (5%); Outros (30%) ; Não-preenchido(77%)
- Shows Televisivos: Pânico da TV (4,1%) ; Outros (6,7%); Não-preenchido(56%)

O único dado disponível na página de estatísticas oficiais do Facebook, mas que não tem correlação nenhuma com os dados coletados é a média da quantidade de amigos que na Rede UnB Facebook é de 260 e nas estatísticas oficiais do Facebook é de 130 usuários. Isto pode mostrar que o usuários desta rede fazem mais uso do Facebook que a média global de usuários. Pode-se se base dessa análise consiste que para o usuário aceitar a solicitação de amizade de uma pessoa e esta adicionar no número de amigos do seu perfil esta necessita de pelo menos fazer *login* no Facebook. Outra possível análise é que pessoas dentro da Universidade de Brasília se relacionam com mais pessoas do que a média global, o que talvez possa ser generalizado às pessoas que estudam em grandes organizações educacionais. Isto pode representar maior troca de experiências entre pessoas vivenciados durante a fase da Universidade. Pode se ainda tentar generalizar que os (ex-)alunos da UnB são mais incluídos digitalmente do que a média global. Todavia, estas suposições só poderiam ser comprovadas através de análises de CGU e modelagem do comportamento do usuário.

A disparidade de 8% a mais dos perfis pertencentes ao gênero sexual masculino pode querer dizer que usuários são mais influentes a novas tendências. Outra análise baseada no fato de a porcentagem de homens solteiros na rede da UnB ser maior que é a de mulheres solteiras em 15% pode significar que eles estão utilizando o Facebook para procurar um relacionamento.

Percebe-se também que o colégio de ensino médio mais bem rankeado na coleta é o ALUB que dessa maneira pode comprovar o slogan da rede de colégios e cursinho que é "A que mais aprova na UnB". Em

segundo lugar tem-se o colégio Galois que é mais elitizado. Supõe-se que poucas pessoas que estudaram na UnB teriam acesso, mas está bem colocado porque a maioria dos estudantes que estudaram neste colégio e fazem UnB preencheram esta informação na parte de educação do seu perfil.

A parte de trabalho foi muito "disputada" no ranking uma vez que a diferenças eram pequenas de cerca de 0.1%. Contudo notou-se que a maioria dos estudantes que fazem UnB trabalham também na própria universidade. Talvez por causa no movimento estudantil, ganhando bolsas da universidade, pbics entre outros.

A idade da maioria dos estudantes da Rede UnB Facebook é na faixa 19 a 21 anos.

Percebe-se também a presença de pessoas de outras cidades do país principalmente do nordeste e centro-oeste para UnB.

A visão política e religiosa está entre as três menos preenchidas dos perfis da Rede UnB Facebook. Supõe-se que esta rede não se importe muito ou não quer divulgar para os amigos a religião, já que na universidade de Brasília há muito diversidade e a pessoa com esta informação pode ser alvo críticas.

A informação menos preenchida da rede é websites. Supõe-se que usuários não tenham o próprio website para compartilhar suas informações ou coisas que julgam interessantes com seus amigos porque tudo está "embutido" dentro do Facebook.

O livro mais famoso é o pequeno príncipe e representado por apenas 3% quando é comparado com show televisivo a diferença 1,1%, o que pode mostrar que apesar do ambiente acadêmicos os (ex-)alunos se interessam mais por programas televisivos do que por livros.

Vale lembrar ainda que todas as análises aqui feitas representam as informações capturadas do perfil do DCE-UNB, ou seja, talvez a captura plena das informações através de outros perfil como por exemplo o perfil da Universidade de Brasília, mostrado na figura 1.6, possa mudar o *ranking* de alguns dados, principalmente aqueles que a porcentagem está mais próxima como educação de ensino médio.

5 CONCLUSÃO

Conclui-se assim as atividades necessárias para o desempenho deste trabalho. Ao fim de 1 ano de pesquisa na área macro Redes Sociais e 3 meses intensos de pesquisas e desenvolvimento da implementação, incluindo levantamentos e elaboração de texto, entende-se que os objetivos foram em boa parte alcançados.

Apesar do trabalho não coletar na sua totalidade todo o universo amostral de (ex-)alunos da Rede UnB Facebook, por não ter acesso ao perfil da página da UnB no Facebook como descrito no capítulo Embasamento Teórico, foi coletado uma porcentagem relevante para validar a pesquisa sociológica que mostrou ser muito maior que pesquisa sociológicas comuns.

O capítulo Embasamento Teórico serve de guia geral de Redes Sociais Online introduzindo os principais conceitos e a importância destas cativando o leitor de forma que este continue a leitura do trabalho. No capítulo Embasamento Teórico faz-se uma conceituação dos principais conceitos por trás das técnicas de coleta e análise de dados em redes sociais. Logo em seguida neste mesmo capítulo é feita uma sumariação das técnicas de coleta e análise de dados que serve para motivar e basear pesquisadores para futuros estudos na área.

Na parte de implementação temos uma descrição detalhada do software feito, que pode servir como base para replicação de trabalhos similares em outras organizações educacionais. Como abordou-se neste mesmo capítulo o software está disponibilizado no sistema de versionamento social GIT (Anexo III) de forma gratuita, podendo ser baixado e utilizado de forma gratuita.

No capítulo Resultados mostram-se as vantagens e as utilidades das técnicas apresentadas, motivando ainda mais futuros pesquisadores. Nota-se aqui ainda que trata-se de um trabalho original com resultados originais e únicos no Brasil.

Entende-se ainda que o trabalho aqui desenvolvido conseguiu alcançar sua motivação de ser útil para a UnB, principal do autor deste trabalho. Dado estas considerações, espera-se que este trabalho seja bem utilizado pela UnB e seja dada continuidade deste dentro da UnB.

O trabalho aqui apresentado apresenta O capítulo de resultados nos leva a crer que Conclui-se assim que os métodos de coleta e análise de dados em redes sociais virtuais tem a utilidade e a vantagens. As utilidades comprovadas foram as seguintes... As vantagens comprovadas foram as seguintes...

As atividades práticas e os resultados necessários para a comprovação desta conclusão foram demons-

tradas com detalhes. Assim, acentuamos o resultado demonstrado. Além disso, foram sugeridas melhorias para o aplicativo e formas de atingir na prática estas melhorias. para que esta comprovação fosse mais notada inicialo deste trabalho foram .

Apesar das informações coletadas serem virtuais, o impacto dos resultados destes métodos são visíveis no mundo real.

5.1 SUGESTÃO DE MELHORIAS

Com dito no capítulo Implementação a versão implementada deste projeto é a versão beta, coletando somente dados do perfil do usuário. A seguir, descrever-se-á algumas funcionalidades a serem implementadas em versões futuras:

- OLAP
- Análises de CGU e Redes Complexas
- Modelagem do comportamento do usuário

REFERÊNCIAS

- [1] BENEVENUTO, J. M. A. e. A. S. S. F. Curso - explorando redes sociais online: Da coleta e análise de grandes bases de dados às aplicações. *SBRC*, scielo, v. 8, p. 160 – 179, 08 2001. ISSN 0104-530X.
- [2] BENEVENUTO, J. M. A. e. A. S. S. F. Talk - explorando redes sociais online: Da coleta e análise de grandes bases de dados às aplicações. *SBRC*, scielo, v. 8, p. 160 – 179, 08 2001. ISSN 0104-530X.
- [3] VARIOS. *Soa Manifesto*. 2012. [Online; accessed 12-December-2011]. Disponível em: <<http://www.soa-manifesto.org/>>.

6 ANEXOS

6.1 ANEXO I - ESTRUTURA DE ARQUIVOS DO PROJETO IMPLEMENTADO

6.2 ANEXO II - FORMAS DE ACESSO E CONFIGURAÇÃO DO APLICATIVO NO FACEBOOK

6.3 ANEXO III - CD COM ARQUIVOS DO PROJETO IMPLEMENTADO

