



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 | 1 |
| 1.2 | OBJETIVOS | 2 |
| 1.2.1 | OBJETIVOS GERAIS | 2 |
| 1.2.2 | OBJETIVOS ESPECÍFICOS..... | 2 |
| 1.3 | MOTIVAÇÃO DO FRANGO..... | 7 |
| 1.4 | TRABALHOS RELACIONADOS | 9 |
| 1.5 | ORGANIZAÇÃO DO TRABALHO | 11 |
| 2 | EMBASAMENTO TEÓRICO | 12 |
| 2.1 | SERVICE ORIENTED COMPUTING | 12 |
| 2.2 | SERVICE ORIENTATION HISTORY..... | 13 |
| 2.3 | SERVICE ORIENTED ARCHITECTURE | 15 |
| 3 | IMPLEMENTAÇÃO | 18 |
| 3.1 | DESCRIÇÃO GERAL DA IMPLEMENTAÇÃO | 18 |
| 3.2 | TECNOLOGIAS UTILIZADAS | 18 |
| 3.2.1 | TECNOLOGIAS DO FACEBOOK | 19 |
| 3.2.2 | TECNOLOGIAS GERAIS: LINGUAGENS DE PROGRAMAÇÃO, BIBLIOTECAS E IDES | 19 |
| 3.2.3 | SERVIDOR DE HOSPEDAGEM E DE DOMÍNIO | 20 |
| 3.3 | REGRAS DE NEGÓCIO, ANÁLISE DE REQUISITOS E METODOLOGIAS..... | 20 |
| 3.3.1 | ELABORAÇÃO DA REGRA DE NEGÓCIO E ANÁLISE DE REQUISITOS..... | 21 |
| 3.3.2 | REGRAS DE NEGÓCIO: COLETAR E ANALISAR O QUE?..... | 22 |
| 3.3.3 | PRINCIPAIS REQUISITO PARA COLETA E ANÁLISE..... | 23 |
| 3.4 | MINERAÇÃO DE DADOS | 24 |
| 3.5 | METODOLOGIAS, PADRÕES E TESTES DO PROJETO..... | 25 |
| 3.6 | ESTRATÉGIAS PARA ATINGIR PÚBLICO ALVO | 25 |
| 4 | ESTUDO DE CASO | 27 |
| 4.1 | DISCUSSÃO DE RESULTADOS..... | 27 |
| 5 | CONCLUSÃO | 30 |
| 5.1 | SUGESTÃO DE MELHORIAS | 31 |
| | REFERÊNCIAS..... | 32 |
| 6 | ANEXOS..... | 33 |
| 6.1 | ANEXO I - ESTRUTURA DE ARQUIVOS DO PROJETO IMPLEMENTADO | 33 |
| 6.2 | ANEXO II - FORMAS DE ACESSO E CONFIGURAÇÃO DO APLICATIVO NO FACE-BOOK..... | 33 |
| 6.3 | ANEXO III - CD COM ARQUIVOS DO PROJETO IMPLEMENTADO | 33 |
| | ANEXOS | 34 |

LISTA DE FIGURAS

| | | |
|-----|---|----|
| 1.1 | Ranking Países (população) e Redes Sociais Online (usuários)..... | 2 |
| 1.2 | Os nove sites mais populares em audiência no Brasil | 3 |
| 1.3 | Comunidade da Universidade de Brasília na RSO Orkut | 4 |
| 1.4 | Comunidade "Professores UnB - Avaliação" na RSO Orkut | 5 |
| 1.5 | Perfil da Universidade de Brasília no Facebook | 5 |
| 1.6 | Página da Universidade de Brasília no Facebook..... | 6 |
| 1.7 | Perfil do Diretório Central dos Estudantes da Universidade de Brasília no Facebook..... | 7 |
| 1.8 | Entidades da UnB mostradas no Perfil do DCE UnB no Facebook | 8 |
| 2.1 | Princípios relacionados ao SOA e seus benefícios | 12 |
| 2.2 | Comparação entre atividades típicas de componentes e serviços | 14 |

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

1.1 MOTIVAÇÃO

Services, SOA, Web Services, SOAP, REST, Service Orientation, Cloud, SaaS. Nos últimos anos esses e outros acrônimos tornaram-se frequentes na tecnologia da informação. O surgimento de um novo paradigma, impulsionado pelo amadurecimento da internet e pela proximidade crescente entre negócios e TI, criou novos caminhos e oportunidades para trabalhos de desenvolvimento e pesquisa. Nesse sentido, um grande número de estudos foram e vem sendo conduzidos com foco nos diversos aspectos da computação orientada a serviços, tais quais arquitetura, modelos, métodos, processos, ferramentas diversas, frameworks, métricas, problemas solucionados e ainda vigentes. Dessa forma, a intenção daqueles interessados em iniciar suas atividades na área fica comprometida pela dificuldade em obter informações claras sobre o atual estado da arte, os desafios, os temas mais abordados e aqueles com deficit em pesquisas. Esses dados são cruciais para que esforços sejam bem direcionados e para que a ciência caminhe com cooperação e eficiência.

Um Mapping Study visa classificar de forma sistemática e ampla um conjunto de estudos. Dada a grande quantidade de publicações no escopo da orientação a serviços, sua metodologia ágil e que permite a análise de um maior número de estudos [1] justifica sua escolha em detrimento de outras metodologias, sendo a mais conhecida o Systematic Review. Essa última exige uma análise minuciosa e detalhada de cada publicação, o que requer um esforço considerável e inviabiliza a inclusão de um grande número de publicações num quadro de poucos pesquisadores. Assim, dados os fatos citados e o interesse em obter uma classificação ampla e significativa da ciência relacionada à orientação a serviços, este trabalho de conclusão de curso em Engenharia de Redes de Comunicação realiza um Mapping Study com foco em orientação a serviços. Mais especificamente e de forma a preservar um maior grau de coesão nos resultados, nosso mapeamento irá focar na questão de qualidade, sobretudo a qualidade de serviços, termo aqui empregado de forma literal e posterior ao termo QoS, uma vez que os principais agentes do paradigma em questão são, coincidentemente, denominados serviços.

Sobre a motivação pessoal desse trabalho é preciso destacar meu interesse na Engenharia de Software, principalmente em seu aspecto distribuído, uma vez que possibilita novos patamares de interoperabilidade, aumentando a gama de soluções da tecnologia da informação e tomando proveito do avanço nas tecnologias

de comunicação de dados, o que envolve questões próprias da Engenharia de Redes. Assim, vejo justificada a iniciativa desse trabalho, que mesmo em sua característica acadêmica *latu sensu*, irá trazer contribuições para ambas as áreas.

1.2 OBJETIVOS

1.2.1 Objetivos Gerais

Os objetivos gerais de um Mapping Study incluem u

1.2.2 Objetivos Específicos

| | | | |
|---|---------------|--|-------------|
|  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 1.1: 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 por-




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

Figura 1.2: Os nove sites mais populares em audiência no Brasil

tuês), 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).

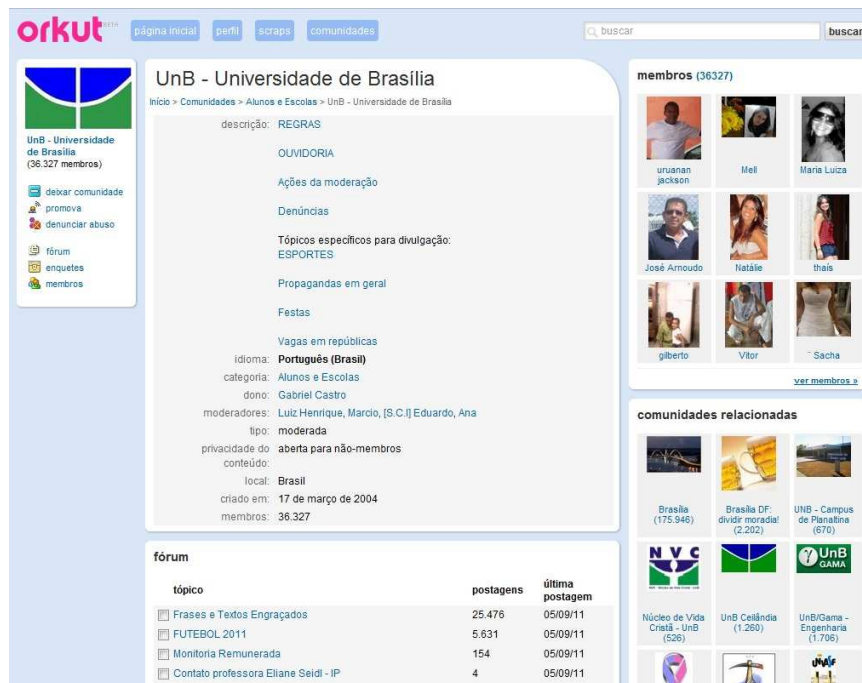


Figura 1.3: Comunidade da Universidade de Brasília na RSO Orkut

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/unibrasil>), 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.

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

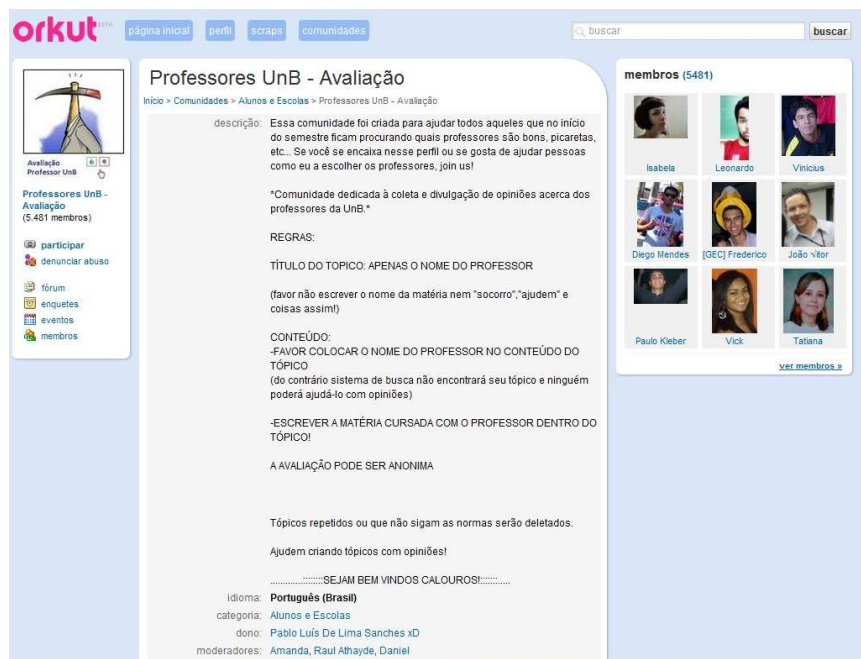


Figura 1.4: Comunidade "Professores UnB - Avaliação" na RSO Orkut



Figura 1.5: Perfil da Universidade de Brasília no Facebook

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.

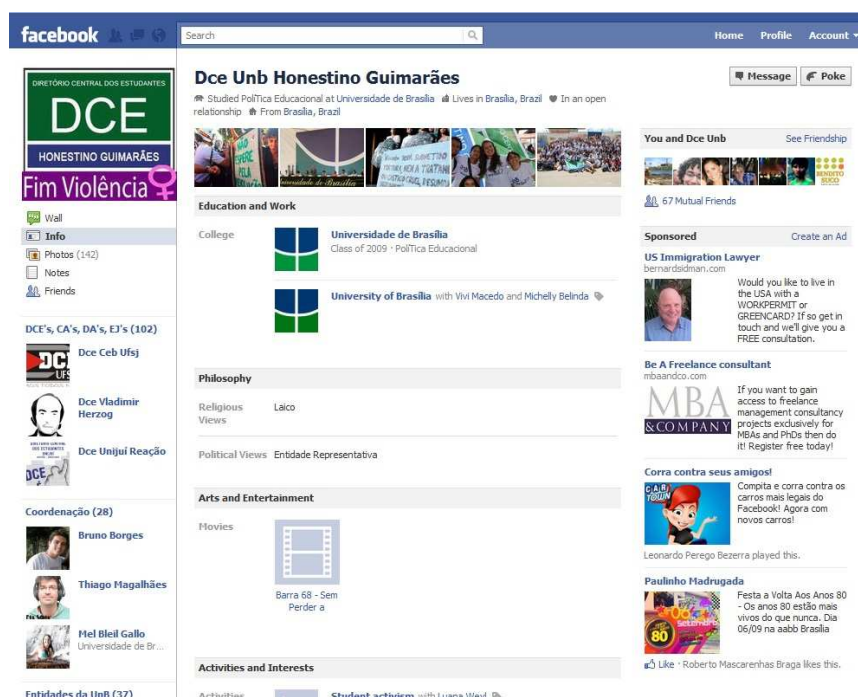


Figura 1.7: Perfil do Diretório Central dos Estudantes da Universidade de Brasília no Facebook

1.3 MOTIVAÇÃO DO FRANGO

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

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 pes-

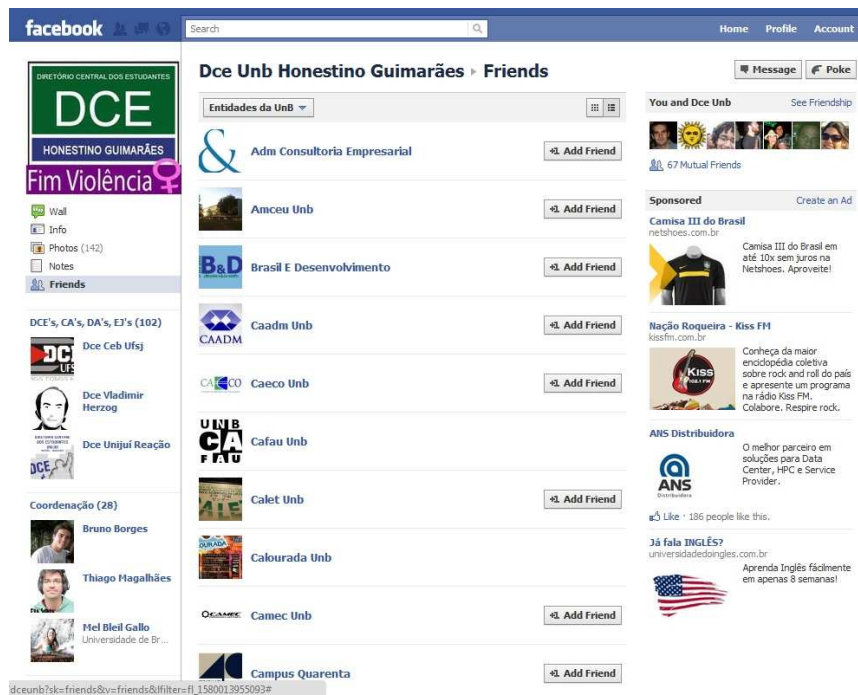


Figura 1.8: Entidades da UnB mostradas no Perfil do DCE UnB no Facebook

quisas. As possíveis melhorias nos sistemas atuais são relacionadas ao design, a usabilidade, a segurança, *spamming*, a vulnerabilidades a novas tendências (*Hot today, Gone tomorrow* [1]), 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.

1.4 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 serviços 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 rede 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 dos 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.

1.5 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

2.1 SERVICE ORIENTED COMPUTING

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 e benefícios almejados pela orientação 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.



Figura 2.1: Princípios relacionados ao SOA e seus benefícios

O paradigma de design da orientação a serviços norteia a configuração do suprimento e consumo dos recursos de TI, em especial as soluções lógicas ou softwares. Apesar do sucesso de outros paradigmas na computação, por exemplo a orientação a objetos, nota-se que problemas distintos, ou seja, da eficiência, agilidade e retorno de investimento na composição de aplicações, ainda deveriam ser atacados. Assim, as aplicações silos com pouca interoperabilidade e sem reuso passaram a ser vistas como um modelo ineficiente, visto que aumentam o custo de desenvolvimento de novas aplicações a longo prazo e reduzem

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 a essa tecnologia, fomentando tanto pesquisa quanto sua adoção em instituições públicas e privadas.

O termo SOC representa um aglomerado de conceitos relacionados à orientação a serviços [2]. Além de um paradigma de design, engloba princípios, padrões e arquitetura, notavelmente o SOA.

O princípio fundamental do design estabelecido pelo SOC está na criação de unidades lógicas coesas e granulares, denominadas serviços. Assim, em vez de aplicações silos que trazem consigo todas as funcionalidades, são criadas unidades coesas, abstratas 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 que irá servir uma ou, idealmente, várias aplicações por meio da orquestração dos serviços.

2.2 SERVICE ORIENTATION HISTORY

Historicamente, o paradigma de orientação a serviços tem como base diversos outros paradigmas e designs. Entre os mais importantes está a arquitetura baseada em componentes. Com ela, a orientação a serviços compartilha visões, uma vez que ambas se sustentam sobre os conceitos de unidades lógicas auto-contidas, auto-descritas, modulares, encapsuladas, que fazem uso de interfaces, contratos e especificações com possibilidade de composição com outras unidades.

No entanto, serviços diferenciam-se fundamentalmente de componentes. Enquanto componentes variam entre modelos *white box*, *gray box* ou *black box*, de acordo com o nível de customização possível, os serviços são sempre hermeticamente encapsulados e distribuídos em forma de *black boxes*. Nos primeiros, foca-se na especificação técnica do conjunto de funcionalidades que desempenham, permitindo o uso das mesmas nos códigos que as invocam uma vez que os componentes tenham sido devidamente importados ou ligados à aplicação. Para os últimos, contratos são estabelecidos em conjunto às descrições das funcionalidades expostas por meio das interfaces do serviço. De fato, em muitos casos um serviço irá fazer uso de componentes em sua implementação, que por conceito independe de sua descrição, desde que mantenha a funcionalidade por ela exposta. Já a especificação de componentes pode incluir também uma definição abstrata de sua estrutura interna. [3].

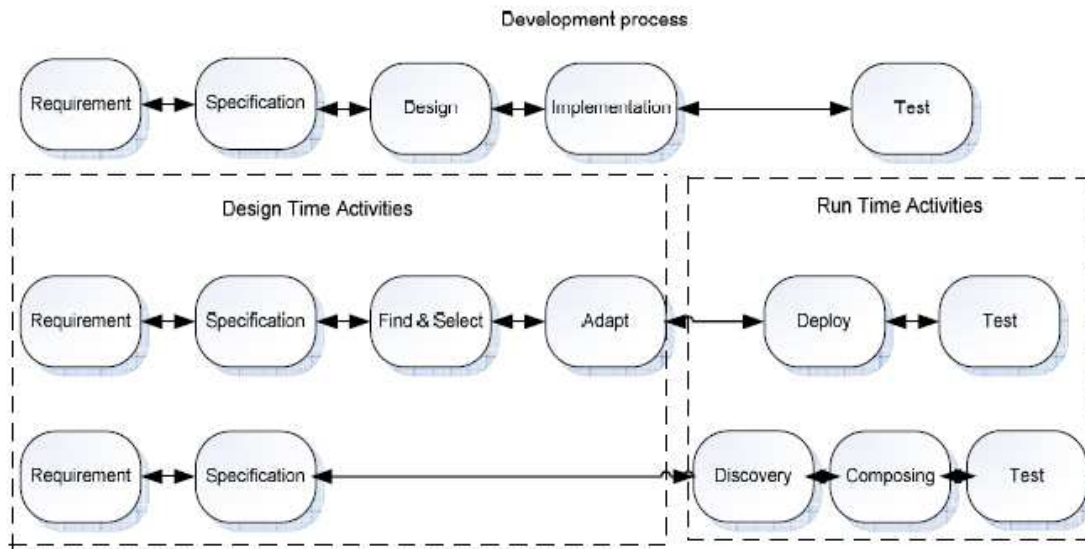


Figura 2.2: Comparação entre atividades típicas de componentes e serviços

A composição dos serviços é feita de forma dinâmica em tempo de execução, enquanto a composição de componentes é feita na fase de design ou em tempo de execução. Isso relaciona-se ao tipo de acoplamento entre provedor e consumidor, havendo diferença entre ambos. Nos componentes é frequente um alto acoplamento, de forma que um grau mais elevado de cumplicidade entre provedor e consumidor é exigida de acordo com o tipo de interação e modelo de componentes usado. Em contraste, nos serviços o baixo acoplamento entre provedor e consumidor é um princípio a ser seguido e um de maior importância no paradigma de orientação a serviços, havendo transparência no processo de troca de mensagens. Dessa forma, o provedor irá desfrutar de liberdade e flexibilidade para a escolha do design e da tecnologia para implementação, assim como o contratante irá somente se ater, funcionalmente, às descrições do serviço e também leva tais vantagens. O baixo acoplamento é viabilizado pela descrição dos serviços que possibilitam sua interação usando parâmetros pré-estabelecidos [4]. Para componentes, é possível uma otimização da performance em detrimento da flexibilidade por meio da composição na fase de design.

Mesmo com modelos que definem a distribuição de componentes, i.e. J2EE e CORBA, os resultados obtidos são de naturezas distintas. O processo de integração remota previsto pela distribuição de componentes mantém o aspecto citado para a sua especificação, uma vez que irá, de forma refinada, distribuir modelos de objetos, seus estados e suas propriedades, enquanto os serviços estão num patamar menos elaborado de integração, fornecendo funcionalidades coesas em alto nível no padrão requisição e resposta. Portanto, componentes e serviços se situam em domínios diferentes, assim como os problemas por eles

enfrentados e as necessidades atendidas [5].

2.3 SERVICE ORIENTED ARCHITECTURE

Esses mesmos agentes responsáveis pela comercialização de SOA, também pela vulgarização do conceito de SOA, ao ponto do mesmo ter se tornado ambíguo e com significados diversos.

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 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 da indústria 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 intrínseca sobre as integrações customizadas, os serviços compartilhados sobre implementações com propósito específico, a flexibilidade sobre otimização e o refinamento evolutivo sobre a perfeição inicial [6].

Because of the increasing agreement on the implementation and management of the functional aspects of those services, such as the adoption of WSDL for service description, SOAP for Web service messaging, or WS-BPEL for Web service composition, the interests of researchers is shifting toward the ?non-functional? or quality aspects of web-enabled software services [7].

Por qualidade de serviço ou QOS de um sistema entendem-se seus atributos ou requisitos não funcionais. Ou seja, são aspectos qualitativos de estado e não definições de funcionamento, podem ser citados a performance, a segurança, a disponibilidade, a modificabilidade, a escalabilidade, etc. Com o amadurecimento e evolução da engenharia de software, tais atributos tornaram-se cruciais para a gerência de riscos e também de custos das aplicações, participando ativamente na definição da arquitetura do sistema com peso maior que os próprios requisitos funcionais.

A performance está diretamente relacionada ao tempo, seja pelo tempo de resposta, pela quantidade de requisições por unidade de tempo ou pela capacidade de atender à requisição numa quantidade de tempo determinada[8]. Nesse sentido, o uso de serviços distribuídos acarreta prejuízos à performance inerentes

à rede utilizada, além de processamento extra devido ao tratamento e encaminhamento das mensagens entre serviços, especialmente quando trata-se de formatos baseados em texto, como o XML, que possuem tamanho maior que sua representação binária. Também irá prejudicar a performance o processo de descobrimento e alocação de serviços nas situações em que esse processo não é pré-estabelecido.

A interoperabilidade relaciona-se à capacidade dos componentes de compartilharem informação e operarem de forma acordada[8]. Dado que o ambiente SOC pressupõe a criação de serviços e, especificamente em SOA, serviços implementados com diversas tecnologias sem prejuízo à intercomunicação, verifica-se que um dos principais benefícios trazidos por essa arquitetura é para a interoperabilidade.

A segurança, como um atributo de QOS, possui alguns princípios. A autenticidade trata da capacidade do sistema em identificar a autoria da informação enviada. A integridade irá resguardar a informação tal qual foi enviada, sem alterações. A confidencialidade garante que somente atores ou entidades autorizadas tenham acesso às informações. O paradigma SOC tem desafios nesses critérios de qualidade à medida em que os serviços são distribuídos em ambientes internos restritos ou na própria internet. É imediata a percepção de que a troca de informações entre as unidades lógicas do sistema pode acarretar sérios riscos à segurança, estando os esforços disponíveis para garanti-la limitados pela necessidade de se manter a interoperabilidade e a performance no ambiente SOC, já que requerem a adoção adicional de especificações, protocolos e consequentemente uma maior redundância de informação e processamento.

A disponibilidade define o tempo em que o sistema estará operacional e portanto respondendo a requisições. Tratando-se de serviços, cada um irá possuir uma disponibilidade que irá influenciar diretamente na disponibilidade dos sistemas que dependem desses serviços, portanto o ambiente SOC possui o desafio de manter-se operacional diante de suas unidades com particularidades próprias. Como possíveis estratégias para mitigar os problemas causados por serviços inoperantes estão a replicação e o balanceamento de carga, além da troca do provedor do serviço. Nesse sentido, o monitoramento, a gerência e a seleção de serviços são processos importantes em SOC e sua automação um dos maiores desafios.

A confiabilidade é um atributo relacionado à disponibilidade e trata da capacidade do sistema em operar sem erros. Mais uma vez, no ambiente distribuído essa capacidade irá depender da confiabilidade dos serviços que compõe o sistema, mas também das mensagens enviadas e recebidas. Dessa forma, um sistema SOC confiável deverá utilizar especificações e protocolos que garantam a confiabilidade das mensagens ou fazer uso de ferramentas e implementações que cuidem desse processo.

A modificabilidade mede a capacidade do sistema de sofrer modificações com menores custos e maior velocidade [9].

O ambiente proposto pelas arquiteturas orientadas a serviço está sujeito a condições particulares diferentes daquelas já estudadas e conhecidas em outros paradigmas, havendo variáveis que elevam a complexidade da análise de parâmetros de qualidade, tanto na fase de planejamento quanto em fase de execução por meio do monitoramento e da gerência dos serviços, sendo esse um obstáculo sólido à adoção de arquiteturas como o SOA. Nesse sentido, o presente estudo visa mapear publicações cujo foco está na qualidade de serviços em SOC, contemplando cenários com ou sem o uso de SOA, proporcionando uma redução da incerteza quanto ao atual estado de desenvolvimento da ciência contribuinte ao tema abordado 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]

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 verossimilhanças. 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
- Genero 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] PETERSEN, K. et al. Systematic mapping studies in software engineering. *12th International Conference on Evaluation and Assessment in Software Engineering*, v. 17, n. 1, p. 1–10, 2007.
- [2] ERL, T. *SOA Principles of Service Design (The Prentice Hall Service-Oriented Computing Series from Thomas Erl)*. Upper Saddle River, NJ, USA: Prentice Hall PTR, 2007. ISBN 0132344823.
- [3] BREIVOLD, H. P.; LARSSON, M. Component-based and service-oriented software engineering: Key concepts and principles. In: *Proceedings of the 33rd EUROMICRO Conference on Software Engineering and Advanced Applications*. Washington, DC, USA: IEEE Computer Society, 2007. p. 13–20. ISBN 0-7695-2977-1. Disponível em: <<http://dl.acm.org/citation.cfm?id=1302497.1302990>>.
- [4] CRNKOVIC, I.; CHAUDRON, M.; LARSSON, S. Component-based development process and component lifecycle. In: *Proceedings of the International Conference on Software Engineering Advances*. Washington, DC, USA: IEEE Computer Society, 2006. p. 44–. ISBN 0-7695-2703-5. Disponível em: <<http://dl.acm.org/citation.cfm?id=1193212.1193814>>.
- [5] PATIL, S. *Integration Approaches: Web Services vs Distributed Component Models PART II*. 2003. [Online; accessed 20-February-2012]. Disponível em: <<http://soa.sys-con.com/node/39754>>.
- [6] VARIOS. *Soa Manifesto*. 2012. [Online; accessed 12-December-2011]. Disponível em: <<http://www.soa-manifesto.org/>>.
- [7] PAPAZOGLU, M. P.; HEUVEL, W.-J. Service oriented architectures: approaches, technologies and research issues. *The VLDB Journal*, Springer-Verlag New York, Inc., Secaucus, NJ, USA, v. 16, p. 389–415, July 2007. ISSN 1066-8888. Disponível em: <<http://dx.doi.org/10.1007/s00778-007-0044-3>>.
- [8] O'BRIEN, L.; MERSON, P.; BASS, L. Quality attributes for service-oriented architectures. In: *Proceedings of the International Workshop on Systems Development in SOA Environments*. Washington, DC, USA: IEEE Computer Society, 2007. (SDSOA '07), p. 3–. ISBN 0-7695-2960-7. Disponível em: <<http://dx.doi.org/10.1109/SDSOA.2007.10>>.
- [9] CLEMENTS, P.; KAZMAN, R.; KLEIN, M. *Evaluating Software Architectures: Methods and Case Studies*. [S.l.]: Addison-Wesley, 2001. ISBN 978-0-201-70482-2.

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

