

UNIVERSIDADE FEEVALE

EDUARDO HENRIQUE KASPER

ANÁLISE AUTOMATIZADA DE HABILIDADES DE
DESENVOLVEDORES BASEADA EM CONTRIBUIÇÕES NO
GITHUB

(Título Provisório)

Anteprojeto de Trabalho de Conclusão

Novo Hamburgo
2016

EDUARDO HENRIQUE KASPER

ANÁLISE AUTOMATIZADA DE HABILIDADES DE
DESENVOLVEDORES BASEADA EM CONTRIBUIÇÕES NO
GITHUB

(Título Provisório)

Anteprojeto de Trabalho de Conclusão de
Curso, apresentado como requisito parcial
à obtenção do grau de Bacharel em
Sistemas de Informação pela
Universidade Feevale

Orientadora: Marta Rosecler Bez

Co-orientador: Juliano Varela de Carvalho

Novo Hamburgo
2016

RESUMO

O mercado de trabalho de Tecnologia da Informação (TI) tem mudado muito nos últimos anos, principalmente no que diz respeito ao setor de engenharia e desenvolvimento de *software*. Os grandes investimentos feitos no setor, decorrentes do avanço sistemático e acelerado de tecnologias, propiciam que mais empresas desenvolvam novas e melhores soluções. Entretanto, devido à vasta sorte de tecnologias está cada vez mais difícil encontrar profissionais que se adequem às necessidades tecnológicas da empresa. O Github surge, neste contexto, como uma plataforma para desenvolvedores de *software* que possibilita o compartilhamento rápido e fácil de projetos criados por eles e fonte de pesquisa para entender suas qualidades técnicas. Além disso, a plataforma funciona como uma rede social, onde seus atores, neste caso, sendo primariamente desenvolvedores, interagem e trocam informações. Como qualquer rede social moderna, o sistema gera um volume imenso de dados a respeito de seus participantes, como o código e a linguagem de programação em que trabalham. Utilizando a API pública do Github é possível extrair e analisar dados de forma a identificar ecossistemas e habilidades que os programadores possuem. Desta forma, seria possível facilitar o processo de seleção de empresas quando estas necessitam de uma qualidade específica em um profissional. Portanto, este trabalho tem por objetivo desenvolver um protótipo onde, ao informar o identificador de um usuário, interessados verão uma análise do perfil do profissional baseado em suas contribuições públicas no Github.

Palavras-chave: Github. Mercado de TI. Análise de Perfis. Seleção e Recrutamento.

SUMÁRIO

MOTIVAÇÃO	5
OBJETIVOS	8
METODOLOGIA	9
CRONOGRAMA.....	11
BIBLIOGRAFIA	12

MOTIVAÇÃO

O advento da TI gerou diversas transformações no mundo contemporâneo, trazendo consigo grandes benesses para a sociedade em geral. Buscar profissionais qualificados tem se apresentado como um grande problema para as empresas, uma vez que as habilidades necessárias para completar projetos de software são cada vez mais especializadas e escassas. Uma estimativa realizada pela Associação Brasileira de Empresas de Tecnologia da Informação e Comunicação (Brasscom), divulgada pela Folha de São Paulo, diz que 2014 teve um déficit de 45 mil profissionais (FOLHA DE SÃO PAULO, 2014). Em reportagem de fevereiro de 2016, o Jornal da Globo informou que no Brasil existem 50 mil postos de trabalho que buscam profissionais qualificados (JORNAL DA GLOBO, 2016). É possível, porém, criar maneiras de, utilizando informações disponíveis publicamente, obter mais informações a respeito de prováveis candidatos à vagas disponíveis.

O mercado brasileiro de TI, segundo a Softex (2014), deve ter um déficit de aproximadamente 408 mil profissionais até 2022. Em 2009, segundo Macedo (2011), o setor de software cresceu 4%, mesmo em um contexto de crise, como o de 2008. Matéria da revista Exame (2014) corrobora com os dados, informando que a Associação Brasileira de Empresas de Software (ABES) divulgou pesquisa evidenciando que no ano de 2013 o Brasil obteve um investimento de 15,4% a mais do que em 2012 em desenvolvimento de programas. Em 2014, a ABES apresentou um relatório informando que o setor de software cresceu 12,8% sobre 2013. Neste cenário, é cada vez mais difícil para as empresas encontrarem profissionais com as habilidades necessárias. Segundo matéria publicada pelo portal de notícias Globo News (2015) empresas estão até mesmo treinando profissionais por não encontrarem os perfis desejados no mercado de trabalho.

Entre os segmentos de TI, o desenvolvimento de *software* é o que mais se destaca no Brasil e é preciso entender como estes profissionais são inseridos nas empresas. As inovações tecnológicas e a globalização se mostram como razão de grandes transformações no mercado e, por consequência, nos seus atores. Se por um lado esta nova economia traz mais possibilidades por estar em constante desenvolvimento, é preciso que os interessados em trabalhar no setor busquem capacitação constantemente (MACEDO, 2011). Neste cenário, para se destacar e demonstrar suas competências, desenvolvedores de *software* têm utilizado redes sociais específicas para eles.

Segundo Acioli (2007), redes sociais têm sua origem na área das Ciências Sociais onde se definem como uma conexão sem limites entre atores com vínculos entre si, formando grupos entre suas associações. O Github¹ apresenta-se hoje como uma rede social específica para desenvolvedores, possibilitando o compartilhamento fácil de código entre seus usuários. Este tipo de plataforma permite que programadores criem novos projetos de código aberto e os compartilhem de forma fácil e rápida com qualquer pessoa interessada, utilizando um software de controle de versão chamado Git² (THUNG et al., 2013). Dabbish et al. (2012) elucidam que um colaborador fica ciente das interações e contribuições de outros colaboradores devido às grandes oportunidades proporcionadas pela plataforma. Pode-se definir, portanto, que o Github, tendo ligações entre seus participantes e possibilitando suas integrações, é uma rede social voltada a colaboração de código de programação.

Nesta rede social, são permitidas conexões entre desenvolvedores onde, muitas vezes, centenas deles podem contribuir para um projeto sem se conhecer (THUNG et al., 2013). No sistema, usuários podem ter repositórios (também chamados de projetos) públicos ilimitados. Em repositórios públicos o código produzido é disponibilizado a todos, bem como todas as contribuições feitas a eles. Além disto, qualquer ator integrante pode enviar melhorias de código, fazer comentários, solicitações de correções e alterações, assistir mudanças e demonstrar interesse. Esta facilidade, junto com as ferramentas disponíveis, permitem a disseminação de projetos de código aberto, por pessoas e até empresas renomadas, como a NASA, o Google e o Facebook.

Todos estes dados são disponibilizados ao público através de uma API (*Application Programming Interface*) utilizando chamadas REST (*Representational State Transfer*). Neste caso é possível obter dados sobre seus usuários, como número de contribuições feitas a um projeto, projetos pertencentes à determinados usuários, linguagens de programação utilizadas, conteúdos de projetos, projetos os quais o usuário demonstrou interesse, entre outras informações relevantes e possíveis de serem analisadas para entender melhor o perfil dos mesmos. Debbish et al. (2012) mencionam que as pessoas tendem a deduzir habilidades técnicas de profissionais da computação baseando-se em suas contribuições no Github. Na verdade, os dados disponíveis já foram analisados por diversos pesquisadores ao redor do mundo. Lima, Rossi e Musoeli (2014) utilizam o projeto Github Archive³ para analisar e

¹ <https://github.com/>

² <http://git-scm.com/>

³ <https://www.githubarchive.org>

caracterizar contribuições baseadas em relações sociais entre os participantes. Este é apenas um exemplo da utilização da plataforma para os fins mencionados.

Debbish et al. (2012) fazem importantes aferições a respeito das contribuições de desenvolvedores e a possibilidade de análise de seus proprietários. Segundo os autores, é possível extrair importantes informações do *feed* de atividades de programadores e estimar suas competências técnicas e reputações devido ao modo social no qual a internet tem se tornado nos últimos 15 anos.

Plataformas colaborativas como o Github possibilitam a vasta obtenção e acesso a análises em cima dos dados proporcionados por seus atores integrantes. Aliando este fato ao grande déficit e dificuldades na área de desenvolvimento de software, é possível utilizar tecnologias ao nosso alcance para facilitar a conexão e obtenção de informação do contratante sobre o contratado, baseando-se em suas necessidades técnicas. É possível que gerentes de projetos, por exemplo, possam rastrear as habilidades de colaboradores dentro da empresa e até mesmo buscar novos colaboradores utilizando estas informações.

As referidas necessidades técnicas variam entre as empresas, por causa de diversos fatores, como o grande número de linguagens de programação, diferentes ecossistemas, necessidades de clientes, conhecimentos de colaboradores, etc. Muitas vezes, empresas necessitam de profissionais com conhecimento a respeito de ecossistemas (aqui definidos como *frameworks* e bibliotecas) específicos. Para exemplificar bibliotecas, pode-se citar JQuery⁴ ou Backbone⁵, em Javascript, ou no caso de *frameworks*, Rails para Ruby⁶ e Laravel⁷ para PHP. A necessidade de conhecer as capacidades técnicas de desenvolvedores dá-se devido a esta variedade. Assim, as contribuições públicas de usuários no Github apresentam-se como uma forma de facilitar este processo. Por isso, um dos objetivos deste trabalho é desenvolver um algoritmo para análise de repositórios públicos que identifiquem qual ecossistema eles utilizam.

Desta forma, este trabalho se focará em construir um protótipo que possibilite aos responsáveis por administrar times de desenvolvimento a buscar mais informações sobre as habilidades de profissionais, de forma a dar suporte em processos de seleção de candidatos.

⁴ <http://jquery.com>

⁵ <http://backbonejs.org>

⁶ <http://rubyonrails.org>

⁷ <http://laravel.com>

OBJETIVOS

1. OBJETIVO GERAL

Oferecer à administradores e analistas de RH um protótipo de *software* que colete informações a respeito de desenvolvedores candidatos, levando em consideração o perfil profissional desejado e as características dos profissionais que estão registradas no Github.

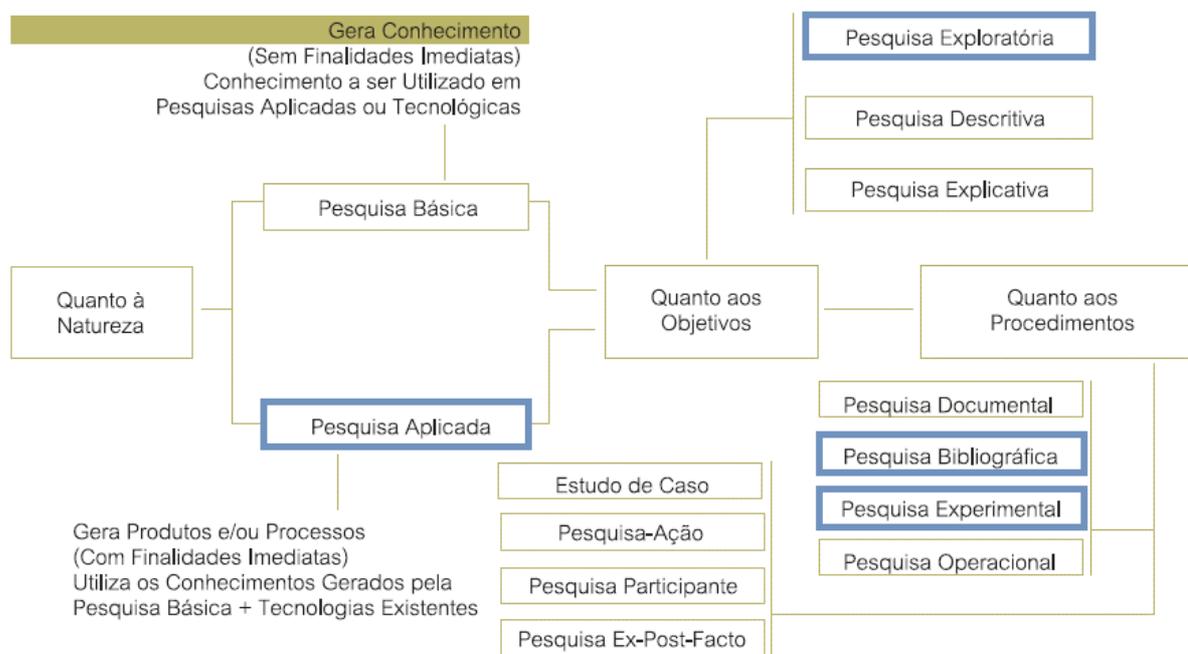
2. OBJETIVOS ESPECÍFICOS

- Explorar a API do Github;
- Definir as tecnologias para desenvolvimento do protótipo;
- Avaliar as características dos usuários que serão necessárias para o protótipo;
- Escolher o método mais eficaz para a análise de ecossistemas;
- Desenvolver um protótipo de sistema baseado em informações coletadas a partir do perfil público do usuário no Github.

METODOLOGIA

A figura a seguir denota os conceitos a serem adotados na metodologia deste trabalho.

Figura 1 - Tipos de pesquisa científica



Fonte: adaptado de Prodanov e Freitas (2013).

O presente trabalho tem por objetivo o desenvolvimento de um protótipo de *software* capaz de auxiliar líderes de projetos, administradores, analistas de RH e demais interessados, na tomada de decisão de candidatos a desenvolvedores. Considerando que o desenvolvimento do protótipo seja para solucionar um problema específico, pode-se caracterizar este trabalho como pesquisa aplicada (PRODANOV; FREITAS, 2013).

Ao buscar a definição dos conceitos envolvidos, e com o intuito de viabilizar uma solução para o problema proposto, pode-se afirmar que, quanto aos seus objetivos, este trabalho caracteriza-se como pesquisa exploratória. Os procedimentos adotados são de pesquisa bibliográfica e experimental, sendo que o primeiro se refere aos estudos dos conceitos necessários utilizados como referencial teórico para este trabalho, enquanto o segundo visa o desenvolvimento de um experimento.

O desenvolvimento do protótipo será realizado utilizando bibliotecas de integração com o *framework* Rails disponibilizado pelo Github⁸, possibilitando integração fácil e aumentando a velocidade de desenvolvimento.

⁸ <https://github.com/octokit/octokit.rb>

Durante a primeira parte da disciplina de trabalho de conclusão de curso a maior ênfase será dada para o levantamento bibliográfico e exploração da API do Github, com o objetivo de entender o que deve ser necessário para a construção e análise dos perfis de programadores. Ao fim desta primeira etapa, será necessário ter um algoritmo que possa retornar informações de repositórios, como *frameworks* e bibliotecas utilizadas. Na segunda parte da disciplina (TC2), o algoritmo será melhorado para possibilitar a busca de informações baseada em entradas do usuário e o protótipo deverá ser desenvolvido.

Durante a segunda etapa do trabalho de conclusão também será necessário validar os resultados obtidos. Para tal, será buscado um número previamente estipulado de usuários com resultados previsíveis para comparação com os resultados retornados pelo protótipo. Assim, poder-se-á validar o projeto desenvolvido com perfis de usuários reais.

CRONOGRAMA

1. TRABALHO DE CONCLUSÃO I

Etapa	Meses			
	Mar	Abr	Mai	Jun
Escrita do Anteprojeto				
Revisão do Anteprojeto				
Levantamento bibliográfico				
Estudo da API do Github				
Projeto e início da implementação de um protótipo				
Escrita do Trabalho de Conclusão I				
Revisão do Trabalho de Conclusão I				
Entrega do Trabalho de Conclusão I				

2. TRABALHO DE CONCLUSÃO II

Etapa	Meses			
	Ago	Set	Out	Nov
Escrita do texto do Trabalho de Conclusão II				
Implementação do Protótipo				
Testes do protótipo				
Validação dos resultados do protótipo				
Revisão do Trabalho de Conclusão II				
Entrega e apresentação do trabalho na Banca				

BIBLIOGRAFIA

ABES - ASSOCIAÇÃO BRASILEIRA DAS EMPRESAS DE SOFTWARE (Org.). Mercado Brasileiro de Software: panorama e tendências. **Abes Software**, São Paulo, v. 1, n. 1, p.01-24, jun. 2015. Disponível em:

<[http://central.abessoftware.com.br/Content/UploadedFiles/Arquivos/Dados 2011/ABES-Publicacao-Mercado-2015-digital.pdf](http://central.abessoftware.com.br/Content/UploadedFiles/Arquivos/Dados%2011/ABES-Publicacao-Mercado-2015-digital.pdf)>. Acesso em: 20 fev. 2016.

ACIOLI, Sonia. **Redes Sociais e Teoria Social: Revendo os Fundamentos do Conceito.** Informação & Informação, [S.l.], v. 12, n. 1 esp, 2007. Disponível em:

<<http://www.uel.br/revistas/uel/index.php/informacao/article/view/1784>>. Acesso em: 19 fev. 2016.

CHAVES, Reinaldo. **Faltam 45 mil profissionais de TI no Brasil.** 2014. Disponível em: <<http://classificados.folha.uol.com.br/empregos/2014/06/1466085-faltam-45-mil-profissionais-de-ti-no-brasil.shtml>>. Acesso em: 21 fev. 2016.

DABBISH, Laura et al. **Social Coding in GitHub: Transparency and Collaboration in an Open Software Repository.** 2012. Curso de Ciências da Computação, School Of Computer Science And Center For The Future Of Work, Heinz College Carnegie Mellon University, Pittsburgh, 2012. Disponível em:

<https://www.cs.cmu.edu/~xia/resources/Documents/cscw2012_Github-paper-FinalVersion-1.pdf>. Acesso em: 15 fev. 2016.

FOLHA DE SÃO PAULO (Org.). **Faltam 45 mil profissionais de TI no Brasil.** Disponível em <<http://classificados.folha.uol.com.br/empregos/2014/06/1466085-faltam-45-mil-profissionais-de-ti-no-brasil.shtml>>. Acesso em: 15 mar. 2016.

GUSMÃO, Gustavo. **Mercado brasileiro de TI cresceu mais de 15% em 2013.** 2014. Disponível em: <<http://exame.abril.com.br/tecnologia/noticias/mercado-brasileiro-de-ti-cresceu-mais-de-15-em-2013>>. Acesso em: 19 fev. 2016.

JORNAL DA GLOBO (Org.). **Mercado de TI é um dos setores que não pararam de contratar no Brasil**. Disponível em <<http://g1.globo.com/jornal-da-globo/noticia/2016/02/mercado-de-ti-e-um-dos-setores-que-nao-pararam-de-contratar-no-brasil.html>>. Acesso em: 15 mar. 2016.

MACEDO, Maicon Costa Borba. **O Mercado de Trabalho em Tecnologia de Informação: a inserção profissional dos desenvolvedores de software**. 2011. 103 f. Dissertação (Mestrado) - Curso de Sociologia, Instituto de Filosofia e Ciências Humanas, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2011. Disponível em: <<https://www.lume.ufrgs.br/bitstream/handle/10183/49103/000827129.pdf?sequence=1>>. Acesso em: 20 fev. 2016.

GLOBO NEWS (Org.). **Empresas investem no treinamento de lideranças para melhorar resultados**. 2015. Disponível em: <<http://g1.globo.com/globo-news/contacorrente/noticia/2015/01/empresas-investem-no-treinamento-de-liderancas-para-melhorar-resultados.html>>. Acesso em: 20 fev. 2016.

PRODANOV, Cleber Cristiano. **Metodologia do trabalho científico: métodos e técnicas da pesquisa e do trabalho acadêmico**, 2013. Disponível em: <<http://www.feevale.br/cultura/editora-feevale/metodologia-do-trabalho-cientifico---2-edicao>>. Acesso em: 23 fev. 2016.

SOFTEX (Org.). **Relatório SOFTEX**. 2014. Disponível em: <http://www.softex.br/wp-content/uploads/2015/04/Relatorio_Anuar_2014.pdf>. Acesso em: 15 mar. 2016.

THUNG, Ferdian et al. **Network Structure of Social Coding in GitHub**. (2013). CSMR 2013: Proceedings of the 2013 17th European Conference on Software Maintenance and Reengineering: 5-8 Março 2013, Genova, Itália. Research Collection School Of Information Systems. Disponível em: <http://ink.library.smu.edu.sg/sis_research/1687>. Acesso em: 21 fev. 2016.