

## INFORMAÇÕES GERAIS DO TRABALHO

**Título do Trabalho:** Implementação de Big Data para Processamento de Dados de Alunos Egressos em Redes Sociais por meio de Recursos de Computação em Nuvem

**Autores:** Breno Rossi Lopes, Guilherme Quadros Werner, Moisés Henrique Ramos Pereira

**Palavras-chave:** Redes Sociais, Big Data, Alunos Egressos, Alunos Evadidos, Análise de Sentimentos.

**Campus:** Ribeirão das Neves

**Área do Conhecimento (CNPq):** Ciência da Computação/ Metodologia e Técnicas da Computação

### RESUMO

Este projeto de pesquisa faz parte de uma iniciativa para realizar a modelagem e a implementação de um Big Data contendo grandes volumes de dados referentes às informações de alunos do Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais (IFMG). Muito embora a atuação *in loco* sobre as necessidades dos alunos dentro das instituições seja extremamente importante para potencializar sua motivação em torno da atmosfera acadêmica, observa-se ainda uma grande ausência de esforços que permitam o acompanhamento da vivência profissional desses alunos depois de formados, quais as suas necessidades acadêmicas complementares a sua formação e, principalmente, qual a influência da comunidade científica e comercial sobre as escolhas desses alunos fora dos âmbitos da instituição. A escolha da fonte de dados para tais abordagens e o seu devido gerenciamento podem possibilitar a descoberta de padrões de informação valiosos para o ganho competitivo das instituições. Neste contexto, este projeto propõe a implementação de um Big Data em ambientes de computação em nuvem para processamento e tratamento dos dados dos alunos egressos obtidos a partir de redes sociais. Como primeiro passo, uma aplicação foi desenvolvida para extrair os dados dos usuários do Facebook que curtiram e que avaliaram a página oficial do IFMG nessa rede social. Sobre a base de dados coletada, experimentos foram realizados a fim de analisar estatisticamente essas avaliações e promover a caracterização do público que curtiu a página. Além disso, essas informações foram cruzadas com as bases de dados dos alunos ativos na instituição, dos alunos egressos, bem como com a base de dados de ex-alunos gerada pelo levantamento da Comissão de Combate à Evasão lotada no IFMG *Campus* Ribeirão das Neves. O arcabouço pretende manter essas informações em um grande repositório de dados que permitirá a execução de outros projetos de pesquisa para analisar esse tipo de informação pela ótica da mineração de dados, análise de sentimentos, construção de grafos sobre as conexões entre os dados de usuários e a sua trajetória profissional, dentre outros estudos. A abordagem proposta consistirá em um dos recursos do Centro de Pesquisas em Mineração de Dados (CEPEMIND) em fase de elaboração no IFMG *Campus* Ribeirão das Neves.

### INTRODUÇÃO:

Este projeto de pesquisa faz parte de uma iniciativa dos professores de Informática do campus Ribeirão das Neves para realizar a modelagem e a implementação de um Big Data contendo grandes volumes de dados referentes às informações de alunos matriculados, egressos e evadidos do Instituto Federal de Educação, Ciência e Tecnologia de Minas Gerais (IFMG). Muito embora a atuação *in loco* dos profissionais de educação sobre as necessidades dos alunos dentro das instituições seja extremamente importante para mapear o desempenho acadêmico deles, observa-se ainda a ausência de esforços para promover a educação continuada a fim de acompanhar a vivência profissional desses alunos depois de formados e descobrir quais as suas necessidades acadêmicas para formação complementar e qual a influência da comunidade científica sobre as escolhas desses alunos fora dos âmbitos da instituição no mercado de trabalho. A escolha das fontes de dados e o seu devido gerenciamento podem possibilitar a descoberta de padrões de informação valiosos para o ganho competitivo das instituições. Um dos grandes desafios é integrar as diferentes bases de dados dos sistemas típicos para o gerenciamento das informações sobre os alunos, pois cada base de dados demanda configurações diferentes, exigindo a implementação de armazém de dados e aplicação de técnicas de inteligência de negócio (LAUDON; LAUDON, 2014).

Uma forma emergente para obter grandes volumes de dados é por meio das redes sociais que assumiram uma nova importância a partir da perspectiva denominada de Análise de Redes Sociais, área derivada das ciências sociais e das ciências exatas. Dentre as diversas redes sociais existentes, destacam-se as redes sociais LinkedIn e Facebook no que se refere à disseminação de informações pelos usuários e ao grande potencial para desenvolver pesquisas sobre mineração de dados por meio de técnicas de análise de sentimentos, aprendizado de máquina e geração de relatórios analíticos.

A rede social LinkedIn é voltada para perfis profissionais e o relacionamento entre profissionais e empresas, contando atualmente com cerca de 300 milhões de usuários em mais de 200 países e representa uma das principais formas de se encontrar novas oportunidades profissionais, além de ser usada em processos de recrutamento por empresas (LINKEDIN, 2014). Já o Facebook vem se mostrando como a principal rede social para interação de ideias e assuntos mais corriqueiros e de interesse comum entre os alunos (PATRÍCIO; GONÇALVES, 2010). Além disso, essas redes sociais fortalecem o processo de aprendizagem em equipe e apoiam ações de inovação e aprendizagem, pois se tratam de espaços virtuais de compartilhamento e distribuição de informação e conhecimento, sendo então utilizadas nos processos educativos e de acompanhamento profissional (MARCON; MACHADO; CARVALHO, 2012; STRÖELE; ZIMBRÃO; SOUZA, 2013).

Pela grande quantidade de informação que pode ser extraída dessas redes, além das possibilidades que elas oferecem para captar a percepção dos alunos sobre a instituição, o que irá influenciar em sua trajetória acadêmica e profissional, mostra-se necessário uma etapa de pré-processamento para o tratamento dos respectivos dados por meio de técnicas de modelagem matemática. Sobre os dados tratados, mostra-se necessário processar técnicas de análise de sentimentos e de mineração de dados para a descoberta de padrões que possam propiciar à instituição promover ações e estratégias de marketing de forma mais assertiva para um público-alvo em potencial, tanto sobre os cursos atuais ou para a captação de novos cursos a fim de promover a educação continuada.

Diante da atual necessidade do mercado de educação superior e da grande massa de informações disponíveis todos os dias na rede, vislumbra-se a possibilidade de obtenção de informações do perfil dos alunos egressos e evadidos do IFMG a fim de verificar a situação deles no mercado de trabalho. Desta forma, tem-se por objetivo, a partir de um Big Data para processamento de informações de redes sociais, propor um estudo que avalie a relação entre alunos no mercado de trabalho, se estão atuando ou não nas suas áreas de formação, quais setores são os maiores empregadores desses egressos e qual a expectativa de inserção no mercado para futuros egressos, bem como promover uma análise da projeção da eficiência do ferramental teórico oferecido pela instituição. Para isso, é fundamental que a abordagem proposta seja capaz de processar e tratar, de forma assertiva, grandes volumes de dados em ambientes de Computação em Nuvem a fim de extrair as informações necessárias para posterior análise e mineração desses dados.

Neste contexto, como primeiro passo, uma aplicação foi desenvolvida para extrair os dados dos 53 mil usuários do Facebook que curtiram e que avaliaram a página oficial do IFMG. Sobre a base de dados coletada, experimentos foram realizados a fim de analisar estatisticamente essas avaliações, incluindo a análise de sentimentos sobre os comentários dos posts associados, se existirem, e promover a caracterização do público que curtiu a página. Além disso, essas informações foram cruzadas com as bases

de dados dos alunos atuais matriculados na instituição, dos alunos egressos, bem como com a base de dados dos alunos evadidos gerada pelo levantamento da Comissão de Combate à Evasão lotada no IFMG *Campus* Ribeirão das Neves.

## METODOLOGIA:

Esta seção descreve sucintamente a abordagem proposta para analisar os dados dos alunos atualmente matriculados, egressos ou evadidos do IFMG *Campus* Ribeirão das Neves. Esta abordagem é dividida em cinco etapas principais, como ilustrado na Figura 1.

A Etapa 1 é responsável por extrair os dados de todos os usuários que curtiram e avaliaram a página oficial do IFMG existente no Facebook por meio da biblioteca *BeautifulSoup* em linguagem de programação *Python*, gerando uma base contendo o nome de quem avaliou a página, o link do perfil, a idade, a quantidade de amigos, o histórico de escolas e empregos, a cidade onde mora e as suas páginas favoritas. A Etapa 2 verifica quais os usuários listados na etapa anterior são atualmente alunos matriculados na instituição. Na Etapa 3, a abordagem proposta analisa se existem alunos egressos entre os usuários que curtiram a página, qual o ano de formação e em qual curso. A Etapa 4 verifica se os alunos evadidos estão estudando em outra instituição atualmente e qual era o curso que faziam no IFMG. A Etapa 5, por sua vez, realiza o levantamento estatístico das informações requeridas sobre cada base de dados e utiliza as técnicas de análise de sentimentos implementadas no sistema Web iFeel em Araújo *et al.* (2014) e no método UniLex proposto por Souza, Pereira e Dalip (2017) sobre os comentários dos alunos, principalmente quando compartilham alguma *post* da página. Finalmente, a Etapa 6 consiste em gerar os relatórios analíticos sobre as informações demandadas pelos profissionais de educação na etapa anterior.

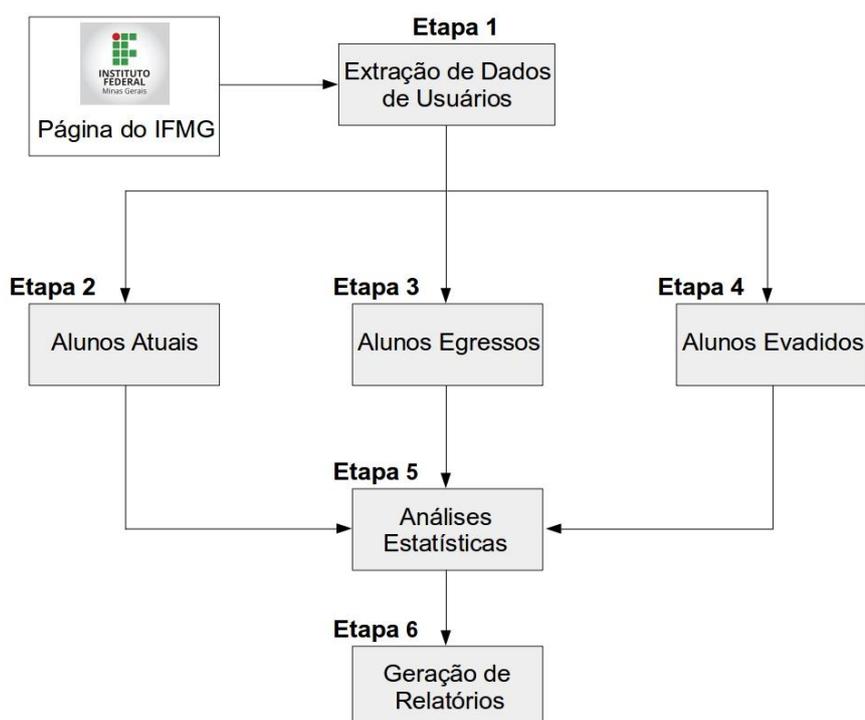


Figura 1 - Visão geral da abordagem proposta para análise dos dados dos alunos usuários da página do IFMG no Facebook.

## **RESULTADOS E DISCUSSÕES:**

Os resultados preliminares do projeto até o momento da submissão deste resumo estendido ainda não estavam disponíveis, pois os algoritmos do iFeel (ARAÚJO *et al.*, 2014) e do UniLex (SOUZA; PEREIRA; DALIP, 2017) ainda processavam os comentários, compartilhamentos e as reações dos respectivos alunos na página oficial do IFMG no Facebook. Assim, o levantamento estatístico dos alunos egressos e evadidos do *campus* está sendo gerado, pois as bases de dados foram conseguidas recentemente.

## **CONCLUSÕES:**

Mesmo a plataforma o LinkedIn disponibilizando as informações de sua base de dados por meio de softwares específicos, o acesso automatizado para *download* dos dados não é permitido em sua versão gratuita, conforme previamente estabelecido pelas Políticas de Privacidade do próprio site, impossibilitando sua inclusão na etapa inicial deste trabalho. Essa grande quantidade de informações só poderia ser acessada por meio de ferramentas pagas, o que está sendo demandado aos setores competentes da instituição. De qualquer forma, foi possível iniciar as coletas dos dados demandados por meio do Facebook, o que permitirá estudos mais aprofundados quando a integração com as demais redes sociais for realizada.

## **REFERÊNCIAS BIBLIOGRÁFICAS:**

ARAÚJO, M.; GONÇALVES, P.; CHA, M.; BENEVENUTO, F.. iFeel: A System that Compares and Combines Sentiment Analysis Methods. Proceedings of the World Wide Web Conference (WWW'14). Seoul, Korea: ACM DL, 2014.

LAUDON, K. C.; LAUDON, J. P.. Sistemas de Informação Gerenciais. 11ª edição. São Paulo. Pearson Education do Brasil, 2014.

LINKEDIN. Março 2014. LinkedIn aponta o perfil do usuário brasileiro. Disponível no link <http://www.agenciaopen.com/blog/linkedin-aponta-o-perfil-do-usuario-brasileiro/> em 30 de setembro de 2015.

MARCON, K.; MACHADO, J. B.; CARVALHO, N. J. S.. Redes sociais e arquiteturas pedagógicas: uma experiência no Facebook. Simpósio Brasileiro de Informática na Educação, 23. Rio de Janeiro: CBIE, 2012.

PATRÍCIO, M. R.; GONÇALVES, V.. Facebook: rede social educativa?. I Encontro Internacional TIC e Educação. Lisboa: Universidade de Lisboa, Instituto de Educação, 2010. 593-598.

SOUZA, K. F. de; PEREIRA, M. H. R.; DALIP, D. H.. UniLex: Método Léxico para Análise de Sentimentos Textuais sobre Conteúdo de Tweets em Português Brasileiro. Abakós, v. 5, n. 2, 2017. 79-96.

STRÖELE, V.; ZIMBRÃO, G.; SOUZA, J. M.. Análise de Redes Sociais Científicas: Modelagem Multirelacional. Brazilian Workshop on Social Network Analysis and Mining. Rio de Janeiro: UFRJ, 2012. p. 1-12.