

Quem procura acha

Centro de Engenharia do Google
no Brasil responde por mudanças
relevantes no sistema de busca

Yuri Vasconcelos



A partir da esquerda,
Berthier Ribeiro-Neto,
Bruno Póssas, Paulo
Golgher, Bruno
Fonseca e Hugo
Santana (em pé)



EMPRESA

GOOGLE

Centro de
Engenharia para
a América Latina
Belo Horizonte, MG

Nº de funcionários
500

Principais produtos
Buscador na internet,
Gmail e rede social G+

Duas piscadas de olho – tempo que corresponde a um quarto de segundo –, esse é o intervalo médio que o Google, principal ferramenta de busca na internet no mundo, leva para fornecer uma resposta ao internauta. São 100 bilhões de consultas realizadas por mês no buscador – uma média de 3,3 bilhões de pesquisas por dia, 137,5 milhões por hora e incríveis 2,3 milhões por minuto. Mais de 20 bilhões de endereços na *web* são analisados a cada 24 horas pelo Google, que depara com 500 milhões de procuras diárias inéditas, ou seja, que nunca haviam sido feitas antes. Aqui no Brasil o *site* detém 91% de participação no mercado de buscas pela internet. O que poucos sabem é que, para dar conta de tanta informação, a empresa de tecnologia com sede em Mountain View, uma das maiores cidades do Vale do Silício, na Califórnia, nos Estados Unidos, conta com o talento de uma equipe de pesquisadores brasileiros no Centro de Engenharia do Google para a América Latina, em Belo Horizonte, Minas Gerais.

Localizada em um prédio da região central da capital mineira, a unidade foi criada em 2005 e hoje é um dos mais importantes entre os cerca de 30 centros de pesquisa e desenvolvimento da empresa espalhados por cidades como Nova York (Estados Unidos), Zurique (Suíça), Tóquio (Japão), e Bangalore (Índia). “Cem por cento dos resultados de busca realizados globalmente, a cada dia, são melhores, em termos de relevância para a consulta, devido a projetos desenvolvidos pelo time de BH”, diz o cientista da computação Berthier Ribeiro-Neto, diretor de engenharia do Google e um dos líderes da equipe brasileira. “Somos responsáveis pela segunda mudança mais relevante na melhoria de busca da história do Google. Além disso, cinco dos 30 principais projetos experimentais com usuários do buscador saíram do nosso escritório”, diz o pesquisador, de 53 anos. Essa segunda inovação mais importante – que não pode ser descrita em detalhes por ser uma informação sigilosa – está relacionada a dois problemas fundamentais dos mecanismos de busca: compreender com precisão o que o usuário está expressando na consulta e entender o que cada um dos documentos na *web* quer dizer. É o casamento desses dois “entendimentos” que, ao final, tornam a informação apresentada pelo buscador mais próxima daquela que o internauta quer encontrar.

A principal área de atuação dos pesquisadores de BH é o ranqueamento – *core ranking*, em inglês

–, que é a ordem em que os *links* são apresentados na página de resultados. “O foco do nosso grupo é a qualidade de busca. Trabalhamos para garantir que o resultado da consulta do usuário seja o melhor possível, e que a primeira resposta do *ranking* apresentado pelo Google atenda de fato o que o internauta está procurando”, diz o cientista da computação Hugo Pimentel de Santana, de 32 anos. “Um dos nossos maiores esforços é entender que certas consultas escritas de forma diferente representam a mesma intenção do usuário”, diz ele. Formado pela Universidade Federal de Pernambuco (UFPE) e com mestrado na mesma instituição na área de inteligência artificial, Hugo é engenheiro de *software* do Google há seis anos e lidera uma equipe de 16 pessoas do time de *ranking*. Ele viaja de duas a três vezes por ano para Mountain View a fim de participar de treinamentos e encontros com a equipe global da empresa. A sede abriga o principal centro da máquina de busca do Google.

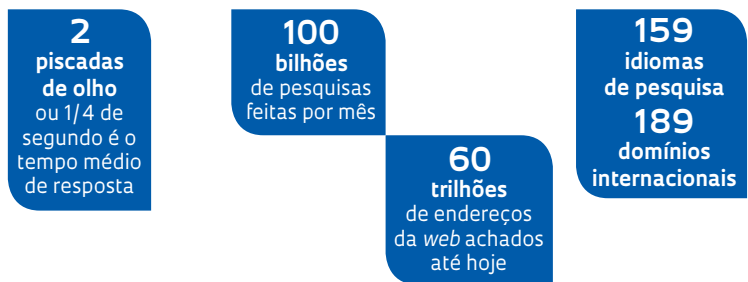
O desafio para melhorar o *core ranking* é grande e cada sugestão passa por um longo e criterioso processo de avaliação. Apenas em 2012, 118.812 ideias para tornar a ferramenta de *ranking* mais eficaz foram apresentadas pelos engenheiros da empresa no mundo. Desse total, menos de 10% (10.391) receberam análises de um grupo de usuários contratados pelo Google, os *raters*. Cerca de 30% dessas implementações não vingaram e 7.018 foram para a fase seguinte, de implantação parcial, quando passaram por avaliações de grupos de usuários reais. Ao final do processo, somente 665 mudanças foram aprovadas e incorporadas ao engenho de busca do Google.

Um indicativo da relevância do Centro de Engenharia de BH é que alguns de seus pesquisadores têm acesso irrestrito ao algoritmo de busca do Google, a enorme sequência numérica que faz

Por dentro do Google

Conheça detalhes sobre a gigante americana de tecnologia dona da principal ferramenta de busca na internet do mundo

MECANISMO DE BUSCA



RESULTADOS FINANCEIROS GLOBAIS DA EMPRESA



*FONTE WALL STREET JOURNAL
**ENTRE JAN E SET/2013
***3º TRIMESTRE DE 2013

CENTRO DE ENGENHARIA DO GOOGLE EM BELO HORIZONTE



o mecanismo funcionar. O algoritmo é um código altamente confidencial e tem para a empresa a mesma importância da fórmula da Coca-Cola para o fabricante do refrigerante – ele é a base de seu sucesso. “São poucos os grupos de fora da sede em Mountain View que trabalham na melhoria do algoritmo de busca”, diz Bruno Pôssas. Aos 36 anos, com graduação, mestrado e doutorado em ciência da computação pela Universidade Federal de Minas Gerais (UFMG), Pôssas é o responsável por todas as melhorias propostas pela equipe de Belo Horizonte no algoritmo do Google.

MÁQUINA DE BUSCA

Para Pôssas, o conhecimento dos pesquisadores brasileiros numa área teórica de fundamental importância para a construção de máquinas de busca, conhecida como “recuperação de informação” (*information retrieval*, em inglês), explica a boa reputação que a unidade mineira goza junto ao escritório central no Vale do Silício. “O Google começou em BH com um grupo muito bom em recuperação de informação, que era reconhecido pela comunidade científica internacional pela qualidade dos artigos que publicava”, afirma Pôssas. Ele se refere à equipe que criou no ano 2000 a empresa mineira Akwan Information Technologies, adquirida cinco anos depois pelo Google para ser transformada em seu centro de pesquisa no país.

A mineira Akwan foi a primeira empresa adquirida pelo Google fora dos Estados Unidos

A Akwan era dona de um engenho de busca centrado na *web* brasileira, chamado TodoBr, que tinha sido desenvolvido por um grupo de professores do Departamento de Ciência da Computação da UFMG. “O TodoBr tinha uma qualidade muito melhor do que a busca do Google para o Brasil na época. Em pouco tempo, nosso buscador explodiu e decidimos criar uma empresa”, recorda-se Ribeiro-Neto, um dos seis fundadores da Akwan – os outros foram os professores Nívio Ziviani, Alberto Laender e Ivan Moura Campos, todos da UFMG, e os investidores de mercado Guilherme Emrich e Marcus Regueira. Com o crescimento da empresa, Ribeiro-Neto decidiu licenciar-se da universidade para tocar o dia a dia do negócio.

“Enfrentamos uma dificuldade inicial por não conseguir financiamento a baixo custo. Batemos na porta do BNDES [Banco Nacional de Desenvolvimento Econômico e Social] e recebemos um ‘não’. Sobrevivemos vendendo soluções para o mercado corporativo de São Paulo, até que, no final de 2004, um colega fez a intermediação com um vice-presidente de engenharia do Google. Em poucos meses fechamos negócio”, diz Ribeiro-Neto. O objetivo do Google ao comprar a Akwan, segundo Ribeiro-Neto, era construir um centro de P&D no país. Quando os norte-americanos adquiriram a empresa brasileira cancelaram todos os seus contratos, mas mantiveram os funcionários. “O Google percebeu que aquele grupo de acadêmicos havia feito uma ferramenta na fronteira do conhecimento. Tínhamos ideias próprias de como lidar com o problema dos mecanismos de busca. Fomos a primeira aquisição global da empresa fora dos Estados Unidos”, diz Ribeiro-Neto, destacando que “o foco do trabalho do grupo sempre foi o desenvolvimento de inovações globais”. O pesquisador foi o único dos seis fundadores da Akwan que permaneceu no Google.

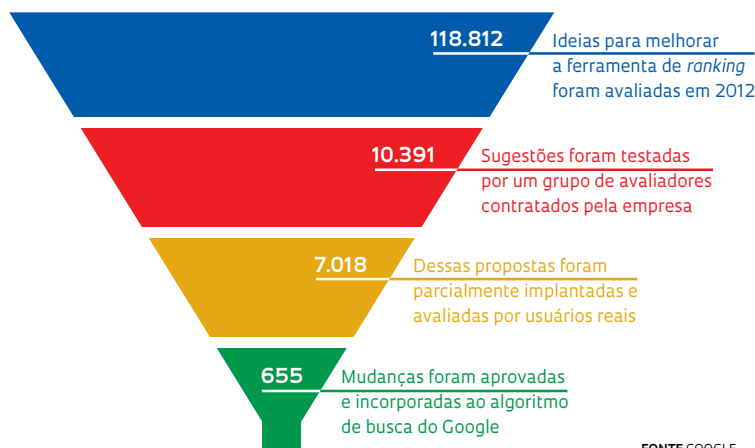
Desde que iniciou suas operações em 2005, o Centro de Engenharia já recebeu investimentos superiores a US\$ 150 milhões. Atualmente trabalham nele por volta de 100 engenheiros, sendo que 75% têm mestrado ou doutorado em ciência da computação. A maioria dos pesquisadores é brasileira, mas há também profissionais de outros países, entre eles Estados Unidos, Índia, Chile, Colômbia e Venezuela. “Buscamos engenheiros com boa formação técnica, pró-ativos, criativos e com iniciativa própria”, diz Ribeiro-Neto, que é coautor do livro *Modern information retrie-*

Descontração e trabalho nas unidades de Minas Gerais e São Paulo



O funil da inovação

O longo caminho que uma ideia percorre até ser incorporada ao engenho de busca do Google



FONTE: GOOGLE

val. Publicada originalmente em 1999, a obra é importante entre cientistas da computação e foi fonte de consulta dos fundadores do Google, Larry Page e Sergey Brin, durante a pós-graduação na Universidade Stanford, quando desenvolveram o projeto que deu origem à empresa de busca.

No Brasil são mais de 500 funcionários, divididos entre o centro de P&D de Belo Horizonte e o escritório central de São Paulo. Os funcionários desfrutam de um ambiente descontraído com espaços de lazer dedicados ao ócio criativo, equipados com redes para balançar, espreguiçadeiras, mesas de sinuca, *videogame*, gibiteca e pufes para relaxar em forma das teclas *ctrl*, *esc*, *alt* e *del*.

A equipe de Belo Horizonte também se dedica a pesquisar e a desenvolver produtos para a rede social Google+, lançada em 2011 para concorrer com o Facebook, e é responsável pela gestão do Orkut, que já foi o mais popular *site* de relaciona-

mentos virtuais do país. Quando o Google comprou a Akwan, em 2005, o Brasil era o principal mercado mundial do Orkut – três anos depois, a subsidiária brasileira passou a ser a responsável global pela plataforma, que chegou a ter 30 milhões de usuários no país. “O setor que lidero tem cerca de 40 pessoas e é um dos três mais relevantes junto com os times de Mountain View e Zurique”, diz o diretor de engenharia do G+, Paulo Golgher, de 36 anos. O dia a dia dos engenheiros é criar novas funcionalidades para o Google+ e desenvolver programas que tornem a rede social mais segura e fora do alcance de *hackers*. “Projetamos sistemas automáticos para que a própria plataforma detecte ameaças e abusos, como conteúdos pornográficos, vírus e *spam*”, conta o engenheiro de *software* Bruno Maciel Fonseca, de 32 anos.

PESQUISA ACADÊMICA

Além de investir em inovações voltadas aos seus próprios produtos, o Google também financia projetos acadêmicos em universidades brasileiras. O programa Google Brazil Focused Research Grants, lançado em 2013, distribuiu cerca de R\$ 1 milhão entre cinco pesquisas de doutorado que procuram entender como as pessoas se comportam no ambiente virtual da internet. A empresa tem tradição em fomentar a pesquisa em áreas de seu interesse em instituições de ensino superior americanas e europeias, mas esta é a primeira vez que apoia projetos brasileiros. O financiamento não tem como contrapartida a cessão dos direitos de propriedade sobre as pesquisas. “No processo de seleção, enviamos convite a 25 pesquisadores e recebemos 20 propostas. Escolhemos as cinco que tinham qualidade compatível com a marca Google”, explica Ribeiro-Neto.

Um dos projetos contemplados é uma pesquisa do Instituto de Pós-Graduação e Pesquisa de Engenharia (Coppe) da Universidade Federal do Rio de Janeiro (UFRJ), cujo objetivo final é melhorar a qualidade do ensino a distância no país. O projeto pretende analisar as reações de alunos desses cursos durante as videoaulas e identificar seu nível de atenção. Liderado pelo professor Edmundo de Souza e Silva, coordenador do grupo de pesquisa da Coppe, o estudo é feito por meio de uma série de sensores e aparelhos conectados ao aluno, que fornecem informações sobre seu estado mental durante a aula a distância. Enquanto uma *webcam* filma suas expressões faciais e o tamanho da pupila, uma pulseira dotada de biossensores mede a condutividade da pele, uma faixa na cabeça capta suas ondas cerebrais e um sensor mede a movimentação do *mouse*. “O nível de condutividade da pele e a dilatação da pupila são indicadores de que a pessoa está mais ou menos atenta”, explica Silva. Segundo

INSTITUIÇÕES QUE FORMARAM OS PESQUISADORES DA EMPRESA

Berthier Ribeiro-Neto, cientista da computação, diretor de engenharia	Universidade Federal de Minas Gerais (UFMG): graduação UFMG: mestrado Universidade da Califórnia, em Los Angeles (EUA): doutorado
Bruno Pôssas, cientista da computação, engenheiro de <i>software</i> sênior	UFMG: graduação UFMG: mestrado UFMG: doutorado
Hugo Santana, cientista da computação, engenheiro de <i>software</i> sênior	Universidade Federal de Pernambuco (UFPE): graduação UFPE: mestrado
Paulo Braz Golgher, cientista da computação e diretor de engenharia do G+	UFMG: graduação UFMG: mestrado
Bruno Fonseca, cientista da computação e engenheiro de <i>software</i>	UFMG: graduação UFMG: mestrado



Experimento realizado na UFRJ analisa o nível de atenção em alunos de ensino a distância

Projeto busca entender o que faz um vídeo postado no YouTube se tornar popular ao longo do tempo

ele, numa aula tradicional, o professor pode observar as reações dos alunos e perceber o quanto ele está atento ou desatento. Já nos cursos a distância isso é impossível. “O sistema que estamos desenvolvendo pretende ajudar a cobrir essa lacuna”, diz Silva. Durante uma videoaula, caso o sistema conclua, por meio dos dados enviados pelos aparelhos (*webcam*, pulseira e sensores), que o aluno está desatento, ele automaticamente altera o curso da aula, por exemplo, pedindo para o estudante fazer alguma tarefa ou mudando o conteúdo que está sendo exibido. Conduzida em conjunto com a professora Rosa Leão, da Coppe, o doutorando Gaspare Bruno e o mestrando Thothadri Rajesh, a pesquisa tem como alvo inicial alunos do curso de sistemas de computação do Cederj, consórcio formado por sete

instituições públicas do ensino superior do Rio de Janeiro, entre elas a UFRJ e a Universidade Federal Fluminense (UFF).

Outro projeto apoiado pelo programa do Google no país busca compreender o que faz um conteúdo postado no canal de compartilhamento de vídeos YouTube se tornar popular. “Queremos entender os vários fatores que podem afetar a audiência de um vídeo e, assim, prever sua curva de popularidade no tempo”, afirma a professora Jussara Almeida, do Departamento de Ciência da Computação da UFMG. A pesquisa faz parte do trabalho de doutorado do cientista da computação Flavio Figueiredo. A metodologia criada na UFMG é capaz de melhorar em mais de 30% a média de previsão de popularidade dos vídeos em relação à técnica mais renomada usada

para tal fim, desenvolvida por Bernardo Huberman, pesquisador do HP Labs, localizado em Palo Alto, na Califórnia. Durante a pesquisa, foram monitoradas centenas de milhares de vídeos no YouTube e coletadas informações diversas no *site*, entre elas a categoria do vídeo, a curva de visualização no tempo, desde o *upload*, e a origem dos *links* utilizados para chegar a esse vídeo.

“Ao analisar a curva de visualizações destes vídeos, vimos que um número pequeno de padrões de curvas de popularidade se repete. Percebemos que, se conseguirmos prever essa curva, podemos melhorar a predição e dizer como a popularidade de um determinado conteúdo irá evoluir ao longo do tempo”, diz Jussara. Uma das conclusões do estudo foi que a qualidade do conteúdo do vídeo nem sempre é de-

terminante para sua popularidade. Muitas vezes, o vídeo “bomba” na *web* depois que um *link* para ele foi postado em algum *site* externo, como um *blog* ou mesmo o Facebook. A compreensão dessa dinâmica pode fornecer informações importantes para anunciantes de bens e serviços na internet, além de produtores de conteúdo.

O mesmo Departamento de Ciência da Computação da UFMG teve outro projeto contemplado com os recursos do Google. O professor Marcos André Gonçalves e o doutorando Daniel Hasan utilizaram algoritmos e técnicas computacionais para aferir automaticamente a qualidade de artigos e conteúdos postados na *web* 2.0, aquela cujas páginas são criadas a partir da colaboração dos internautas. A enciclopédia virtual Wikipedia, com mais de 14 milhões de artigos, foi o foco inicial da pesquisa. “Começamos com a Wikipedia e estendemos o estudo para fóruns de pergunta e resposta”, explica Gonçalves. Para determinar o grau de confiabilidade das páginas, os pesquisadores elaboraram um conjunto de 68 critérios de qualidade, como legibilidade do texto, estrutura e organização dos artigos e o histórico de revisões dos conteúdos postados. “Criamos um aplicativo, ainda não comercial, que dá uma nota para cada um dos critérios”, diz o professor da UFMG. Dentre as várias metodologias existentes que se propõem a fazer algo similar, a projetada por ele e seu aluno fornece os melhores resultados quando submetida à experimentação. “Nossa metodologia pode funcionar como uma bússola mostrando ao internauta quais são os conteúdos da *web* com mais qualidade e credibilidade. Imaginamos que, no futuro, ela poderá ser usada para ordenar as páginas retornadas em uma busca conforme algum critério de confiabilidade”, afirma Gonçalves. ■