

Artigo

Os mecanismos de busca comunicacionais: da era pré-web a era pré-Google

Pedro Henrique Toth*

Resumo

O século XXI é sem sombra de dúvidas o século da Internet. Em praticamente todas as tarefas executadas pelo ser humano é possível encontrar vestígios de conectividade com a grande rede em quase todas as fatias da população mundial. Inclusive tarefas que até um momento atrás dependiam de simples empenho físico por parte das pessoas agora carregam traços tecnológicos conectados, como no caso dos esportes. E dentre todas as maravilhas e praticidades que a Internet trouxe com sua evolução, surgiram os mecanismos de busca. Surgiram e instalaram-se em nossas vidas, tornando-se, nos dias de hoje, um dos maiores geradores de tráfego na Internet. Dos usuários mais experientes aos mais leigos, mecanismos de busca como o Google são atualmente os grandes mananciais do conhecimento, direcionando o usuário ao *website* que melhor atenda a sua expectativa de busca. Sob este olhar torna-se interessante analisar e descobrir quais foram os caminhos tomados pelas diversas tecnologias criadas que, em determinando momento histórico, se uniram para oferecer os serviços, hoje, presentes.

Palavras-chave

Mecanismos de busca. Conteúdo comunicacional. Redes telemáticas. Web.

Abstract

The XXI century is undoubtedly the Internet century. In virtually all tasks performed by humans it is possible to trace connectivity with the large network, in almost all of the world's population. Even tasks that until a moment ago relied on simple physical commitment now carry technological features within it, as in the case of sports. And of all the wonders and practicalities that the Internet has brought with its evolution, came the search engines. They appeared and settled in our lives and have become one of today's major Internet traffic generators. From the most experienced to the most lay users, search engines like Google are currently the major sources of knowledge, directing the user to the website that best meets their search expectation. Under this light it becomes interesting to analyze and find out what were the paths taken by the various technologies created, which in a certain historical moment have come together to offer the services present today.

Keywords

Search engine. Communicational content. Telematics network. Web.

* Tecnólogo em redes de computadores. Mestrando em Comunicação Social pela Universidade Metodista de São Paulo (Umesp). Docente no curso de Gestão de Tecnologia da Informação da Universidade Metodista de São Paulo (Umesp). Membro do grupo de pesquisa ComTec (Comunicação e Tecnologias Digitais). E-mail: pedro.toth@metodista.br. Currículo Lattes: <http://lattes.cnpq.br/6991754397400831>.

Sobre o crescimento da Internet

Vivemos em um mundo conectado. Isso já não é segredo para grande parte da população mundial. Segundo notícia publicada em dezembro de 2014 no jornal O Dia¹, os brasileiros gastam em média quatro horas e cinquenta e nove minutos por dia conectados na Internet durante a semana. E esses números tendem a aumentar. Uma notícia publicada na Folha², em junho de 2014, apontou que, em 2013, aproximadamente 31% da população do país acessava a internet por meio de telefones celulares. A mesma notícia revelou que, naquele ano, aproximadamente 51% dos brasileiros já possuíam acesso à Internet. Essas informações são de caráter promissor, pois em 2015 foram comemorados os 20 anos da Internet comercial no Brasil³ e para um país que possui tantas divergências geográficas, políticas e sociais, chegar na marca de mais da metade da população com acesso à rede é motivo de comemoração. Claro que estar conectado não representa qualidade de vida, este aspecto não sinaliza que o país está com boa qualidade na educação, saúde, transportes e etc. Mas não cabe a este texto adentrar nestas questões.

Outro fator que mudou, desde a criação da Internet até os dias atuais, foi a porcentagem de atenção que dedicamos as diferentes ferramentas presentes na web. Em um momento antes da Internet, era possível apenas se conectar às conhecidas BBS (*Bulletin Board System*). Dependendo do provedor ao qual o usuário utilizava, algumas poucas opções de conteúdo eram oferecidas: notícias, previsão do tempo, cotações da bolsa e afins, o que tornava o computador uma ferramenta de entretenimento e informação limitada. Com o surgimento da web de Tim Berners-Lee, em 1991, um leque imenso de oportunidades foi aberto. Em um primeiro momento foi possível a criação de páginas estáticas utilizando-se da linguagem HTML (*Hyper-Text Markup Language* ou Linguagem de Marcação de Hipertexto), onde era possível inserir hiperlinks em um documento⁴, imagens, áudio e também uma melhor formatação do texto. Com o passar dos anos novas linguagens de programação voltadas para a web surgiram. E com elas começou algo que até então não existia: a interação dos usuários com o website. Graças a esta interação que serviços como Facebook, Twitter e WhatsApp puderam existir, aumentando a permanência dos usuários na rede.

Dentre todas as ferramentas existentes, desde antes da criação da web até os dias de hoje, os mecanismos de busca tornaram-se ferramentas tão úteis que, sob um olhar crítico, são indispensáveis e vitais tanto quanto a luz elétrica é. A evolução dessas ferramentas ocorreu de maneira tão esplendorosa que nos dias de hoje é possível fazer pesquisas sem resultado definitivo, sem indexação prévia, como por exemplo: Qual será a temperatura amanhã em São Paulo? Qual foi o último filme estrelado por Adam Sandler? Um dólar equivale a quantos reais? Dados do site Internet Live Stats⁵ mostram que, a cada segundo, aproximadamente 40.000 pesquisas são feitas ao redor do globo, gerando 3.5 bilhões de buscas por dia e 1.2 trilhões de buscas por ano. E esses números crescem a cada momento. O mesmo site mostra que, desde o ano 2000, o crescimento das buscas se mantém constante, o que pode ser explicado pelo au-

1 Disponível em: <<http://odia.ig.com.br/noticia/brasil/2014-12-20/brasil-brasileiros-ja-gastam-mais-tempo-na-internet-que-assistindo-a-tv.html>>. Acesso em: 18 ago. 2015.

2 Disponível em: <<http://www1.folha.uol.com.br/tec/2014/06/1476690-numero-de-brasileiros-que-usa-a-internet-pelo-celular-mais-que-dobra-em-dois-anos-diz-pesquisa.shtml>>. Acesso em: 18 ago. 2015.

3 Disponível em: <<http://blogs.estado.com.br/link/internet-brasileira-completa-20-anos/>>. Acesso em: 19 ago. 2015.

4 Hiperlinks são pontos no documento que permitem a navegação para outros documentos.

5 Disponível em: <<http://www.internetlivestats.com/one-second/#google-band>>. Acesso em: 18 ago. 2015.

mento de usuários acessando a Internet ao redor do mundo. É importante aos pesquisadores e profissionais da área de comunicação social entenderem que todo resultado de uma pesquisa feita através destes mecanismos é um conteúdo comunicacional, merecendo um olhar detalhado sobre o tema. Para tal, é importante resgatar a história dos mecanismos de busca, incluindo todo o pano de fundo tecnológico presente, a fim de se compreender melhor suas motivações e objetivos. Este resgate histórico mostrará quais foram as mudanças que ocorreram nos processos comunicacionais e como a interação homem-máquina e também a interação do homem-homem mudou nos últimos dois séculos..

A fase pré-web

A comunicação à distância tornou-se a obsessão de cientistas e pesquisadores do passado. O primeiro grande invento comunicacional que merece um enfoque inicial foi o telégrafo, criado por Claude Chappe em 1791. Este primeiro modelo de telégrafo difundiu-se rapidamente por grande parte da Europa e era constituído de painéis com braços móveis, telescópios, relógios e livros de códigos. Cada posição dos braços móveis representava uma letra ou número. O grande inconveniente desta primeira versão era que a comunicação só se dava em dias de tempo aberto, pois os nevoeiros e neblina atrapalhavam a visualização dos sinais. A figura de Napoleão Bonaparte foi importante para a expansão das linhas telegráficas na França, pois ele acreditava no poder, no sentido dominador da palavra, que o telégrafo possuía.

A partir do primeiro modelo de telégrafo diversos experimentos começaram a ser feitos para a construção de um novo modelo, porém utilizando-se da eletricidade para o envio e recebimento das informações. De todos os experimentos criados, foi o de Samuel Morse que sobressaiu, pois trabalhava na simplicidade. Ele utilizava um sistema tecnologicamente simples, apenas com o uso de relés, eletroímãs, fios e baterias, para o envio e recebimento das informações. Ele também empregava uma nova simbologia composta de pontos e traços, com sinais mais curtos e mais longos, que ficou conhecida como código Morse. A primeira mensagem transmitida pelo telégrafo elétrico foi em maio de 1844, quando a mensagem “Que obra fez Deus!” foi enviada pela linha Baltimore-Washington.

O telégrafo elétrico logo popularizou-se pelos Estados Unidos, Europa e América Latina. Seus diferenciais frente aos métodos de comunicação da época eram a velocidade e os custos reduzidos. Isto levou tradicionais serviços de comunicação, como o conhecido *Pony Express*, onde as mensagens eram enviadas a cavalo, a terem seus fins decretados. Com este novo meio de comunicação, as barreiras geográficas (continentais e intercontinentais) começaram a ser quebradas e o conceito de comunicação instantânea começou a ser introduzido em nossa sociedade. O telégrafo ganhou maior poder quando o primeiro cabo submarino foi instalado, interligando os Estados Unidos à Inglaterra. Por essa linha intercontinental começaram a ser

enviadas diariamente informações sobre os valores das ações nas bolsas de valores dos dois países, “padronizando” o valor de venda dos produtos nos dois continentes. Durante anos as linhas telegráficas levaram mensagens de carinho e amor de parentes separados por milhares de quilômetros, como também transmitiram em alta velocidade as notícias trazidas pelas grandes guerras que aconteceriam anos mais tarde.

Posteriormente, em 1856, Antonio Meucci, um italiano que estabeleceu residência nos Estados Unidos, deu outro passo importante para a evolução dos processos comunicacionais: a criação do primeiro modelo de aparelho telefônico, o “telettrofono”. Este primeiro modelo de aparelho foi construído com o intuito de ajudar Meucci a comunicar-se de sua oficina com sua esposa, que sofria de reumatismo e não conseguia mais sair de sua cama. Meucci conseguiu a patente provisória do aparelho, porém, por conta dos problemas financeiros, não conseguiu a patente final que, num segundo momento, foi requerida por seu colega de trabalho, Alexander Graham Bell, que rebatizou o aparelho com o nome de telefone. Em 1877 foi fundada a primeira companhia telefônica, a Bell Telephone Company (que anos mais tarde ficaria conhecida por AT&T), criada pelo sogro de Gram Bell, Gardiner Greene Hubbard. De forma tardia, em junho de 2002⁶, o congresso norte-americano reconheceu Meucci como o verdadeiro inventor do telefone.

Em comparação com o telégrafo elétrico, o telefone eliminou a necessidade de um segundo código no processo comunicacional, reduzindo desta maneira a probabilidade de erros, graças à redundância presente no processo de comunicação oral.

6 Disponível em:
<<http://www.theguardian.com/world/2002/jun/17/humanities.international-education-news>>. Acesso em: 18 ago. 2015.

A redundância — ineficiente por definição — serve como antídoto para a confusão. Ela proporciona segundas chances. Toda linguagem natural contém algum tipo de redundância. É por isso que as pessoas são capazes de compreender textos repletos de erros e também entender conversas num ambiente barulhento. (GLEICK, 2011, p.23)

Em uma mesma linha de raciocínio, Charle Seife afirma que:

A redundância são as pistas a mais numa frase ou mensagem que permitem que se compreenda o sentido mesmo quando a mensagem está um tanto truncada. [...] O sentido de uma mensagem continua inalterado mesmo se partes dela forem removidas. Essa é a essência da redundância. (SEIFE, 2010, p. 17)

Outro fator importante é que por mais que a barreira geográfica e o modo de comunicação estivessem sendo alterados, o telégrafo elétrico e o telefone iniciaram consigo um segundo processo que perdura até os dias atuais: o controle e monitoramento das comunicações. Isto se tornou mais claro aos olhos de todos quando Edward Snowden, ex-agente da NSA (*National Security Agency* ou Agência Nacional de Segurança Norte-Americana) publicou documentos

<http://www.revista.teccog.net>



confidenciais no website WikiLeaks⁷. Dentre os diversos documentos, dois deles apontavam para o “grampo” que estaria sendo feito no telefone celular da chanceler alemã Angela Merkel⁸ e outro sobre o monitoramento da conta de e-mail da presidente brasileira Dilma Rousseff⁹. Se poderosas chefes de governo tiveram sua privacidade comprometida, cristaliza-se a ideia de que toda população, de qualquer local do globo que utiliza sistemas telemáticos e computacionais, pode, de alguma maneira, estar sendo monitorada.

A criação do telefone rompeu as barreiras geográfico-comunicacionais do mundo. Ele popularizou-se rapidamente ao redor do globo, o que gerou a necessidade de estudos científicos focados sobre o tema, que resultaram no desenvolvimento de inúmeras tecnologias para a transmissão da voz e, em seguida, de dados. A tecnologia trazida pelo telefone aproveitou grande parte da base tecnológica que havia sido criada pelo telégrafo de Morse. Posteriormente, dentro da AT&T (antiga Bell Labs) surgiu a figura icônica de Claude Shannon, que revolucionou o campo da comunicação digital. Com o auxílio de Warren Weaver, Shannon escreveu a Teoria Matemática da Comunicação, que é conhecida por muitos como Magna Carta da comunicação e é um dos modelos mais utilizados para descrever o processo comunicacional até os dias de hoje. A Teoria Matemática da Comunicação é tão importante que, sem ela, diversos dispositivos do cotidiano como smartphones, tablets, notebooks e até mesmo a Internet possivelmente não seriam da maneira como conhecemos hoje (ou quem sabe nem existiriam).

7 Disponível em: <<https://wikileaks.org/>>. Acesso em: 18 ago. 2015.

8 Disponível em: <http://www.dia-riodepernambuco.com.br/app/noticia/mundo/2015/07/02/interna_mundo,584294/alemanha-grampo-da-nsa-foi-muito-alem-do-celular-de-merkel.shtml>. Acesso em: 18 ago. 2015.

9 Disponível em: <<http://g1.globo.com/politica/noticia/2013/09/documentos-da-nsa-apontam-dilma-rousseff-como-alvo-de-espionagem.html>>. Acesso em: 18 ago. 2015.

A era das máquinas de calcular

Avançando 89 anos após a criação do telefone, em 1945 foi concluído o projeto do primeiro computador eletrônico, que foi batizado de Eniac (*Electronic Numerical Integrator And Computer* ou Computador e Integrador Numérico Eletrônico). Ocupando uma área de 25 metros na faculdade de engenharia elétrica da Pensilvânia, o Eniac era composto basicamente por válvulas termiônicas, fios e relés que haviam sido soldados à mão. A principal motivação do Eniac era resolver cálculos matemáticos com números de até dez dígitos decimais, sendo utilizado principalmente na época pelo exército norte-americano no cálculo de tabelas de disparo de artilharia. É válido observar que o Eniac foi uma evolução de diversos projetos e conceitos de máquinas de calcular criados no passado. Dentre os cientistas que vislumbravam efetuar cálculos com maior facilidade, vale ressaltar o brilhante nome de Charles Babbage, projetor do primeiro computador de uso pessoal, que ficou conhecido como máquina analítica. Um fato histórico importante sobre a máquina analítica de Babbage foi que, neste momento, surgiu, de maneira mais clara, o conceito de *hardware* e *software*, pois havia a necessidade de programar esta máquina para obter os resultados esperados.

Ainda em 1945, Vannevar Bush, engenheiro elétrico por formação, publicou um artigo nomeado “As We May Think” na revista *The Atlantic*. Neste artigo, Bush clamava que os cien-

tistas trabalhassem em conjunto a fim de criar um corpo de conhecimento para toda humanidade. Sob o seu pensamento visionário, Bush propunha um modelo de máquina que fosse capaz de trabalhar da mesma maneira que o cérebro humano, funcionando através do uso de associações, ou em outras palavras, de *links*. Esta máquina teórica recebeu o nome de Memex. Uma das grandes revoluções trazidas pelo Memex era o uso de uma memória estendida e do hipertexto, tão comuns nos tempos atuais.

De 1945 até hoje, a história do computador foi subdividida em diversas linhas, cada uma com um enfoque especial em algum tipo de tecnologia ou *software* específico, impossibilitando este texto de cobrir toda a história. O ponto importante a ser observado é que o desejo do homem de facilitar tarefas corriqueiras foi uma das importantes alavancas do processo de evolução dos computadores e, dentro desta evolução, a cada nova tecnologia que surgia, a interação do homem com a máquina se modificava. Dos cartões perfurados aos dispositivos com tela sensível ao toque, a lógica empregada pelo homem, para a utilização dessas máquinas de cálculo e, posteriormente, do computador teve de ser adaptada, alterando seu modo de pensar e interagir com elas.

O Pai dos mecanismos de busca

Considerado como o pai dos mecanismos de busca, Gerard Salton (1927 - 1995) era professor de ciências da computação na universidade de Cornell, Estados Unidos. Ele focou suas pesquisas nos mecanismos de busca computacionais existentes. Líder de um grupo de estudos na universidade, Salton foi o mentor do projeto Smart (System for the Mechanical Analysis and Retrieval of Text ou Sistema de Análise Mecânica e Recuperação de Texto). A ideia do Smart não era ser mais um mecanismo de busca, mas sim um sistema que comparava os diversos mecanismos de busca existentes na época, medindo a performance de cada um deles em diversos contextos computacionais e apontando o melhor método para cada caso. Salton ficou conhecido por desenvolver o *Vector Space Model* (Modelo de Espaço Vetorial), base dos sistemas de indexação de conteúdo atuais. Sobre o tema, Salton publicou diversos livros, sendo o mais famoso “*A Theory of Indexing*”.

Hipertexto e Hipermídia

Hipertexto e hipermídia são termos que foram cunhados por Theodor Holm Nelson (ou apenas Ted Nelson) em 1963. Estes termos descrevem documentos que podem conter, dentro de si, imagens e áudio, onde determinadas palavras (*links*) conectam o documento em questão a outros documentos relacionados. Filósofo por formação e professor convidado da universi-

dade de Oxford, Ted Nelson foi o criador do projeto Xanadu em 1960. O objetivo deste projeto era a criação de uma rede de computadores que possuísse uma interface de uso simples, construída por uma linguagem de programação operável por qualquer ser humano e, principalmente, um sistema no qual o link entre as páginas nunca perdesse as referências.

É correto afirmar que a web criada por Tim Berners-Lee possui forte semelhança com o projeto Xanadu, todavia, todos os pontos que Ted Nelson abominava foram embarcados nela. Por conta disso, Ted Nelson criticou por diversas vezes o projeto da web^{10 11}. Por mais que seu projeto não tenha tido sucesso, Ted Nelson conseguiu introduzir o conceito de navegação entre documentos multimídia, que perdura até os dias de hoje. Vale ressaltar que o modelo sugerido pelo projeto Xanadu simplificou parte do processo de interação e, conseqüentemente, a comunicação do homem com a máquina. A evolução tecnológica atual está caminhando para um sentido onde todos os aplicativos (ou *apps* no jargão popular) serão acessados via Internet e não mais instalados nos computadores e dispositivos móveis. Nesta linha de pensamento, o uso do computador (ou demais dispositivos) tende a se tornar cada vez mais simples e acessível a todas as pessoas.

10 Disponível em: <[http://www.cio.com.au/article/383395/hypertext_creator_says_structure_world_wide_web_completely_wrong_/> . Acesso em: 18 ago. 2015.](http://www.cio.com.au/article/383395/hypertext_creator_says_structure_world_wide_web_completely_wrong/)

11 Disponível em: <<http://www.economist.com/node/442985>>. Acesso em: 18 ago. 2015.

A irmã mais velha da Internet

Se a rede telegráfica é considerada a mãe de todas as redes, pode-se afirmar que a Arpanet (*Advanced Research Projects Agency Network* ou Agência de Projetos de Pesquisa Avançada de Defesa) é a irmã mais velha da Internet. Criada pelo departamento de defesa dos Estados Unidos em 1969, o projeto tinha por objetivo inicial interligar diferentes bases militares americanas. Sua estrutura foi pensada para, caso algum trecho da comunicação fosse interrompido por qualquer motivo (como por consequência de um bombardeio), que a comunicação se restabelecesse tomando um outro caminho (rota). Diferente das redes telefônicas, onde a comunicação é feita por circuitos - a conexão entre os dois pontos usa uma rota fixa de comunicação -, a Arpanet criou o padrão de comunicação que ficou conhecido por comutação de pacotes, que divide a informação em pequenos pacotes e os envia por diferentes caminhos para o destinatário.

Outro fator importante que a Arpanet queria resolver era padronizar os protocolos utilizados nas comunicações entre computadores ligados em rede. No passado, os fabricantes de computador (Dell, Apple, IBM, etc.) criavam suas próprias interfaces de conexão de rede e com elas eram criados também protocolos de comunicação (padrões de *software* para a comunicação de computadores em rede). Sob este panorama, a conexão de computadores de diferentes fabricantes tornava-se uma tarefa árdua. Para solucionar este problema, Vinton Gray Cerf e Robert Elliot Kahn juntaram forças e criaram o protocolo que posteriormente se popularizaria no mundo todo, o TCP/IP. Este protocolo é dividido em duas partes: o IP (*Internet Protocol* ou

Protocolo de Internet), que é responsável por endereçar todos os dispositivos dentro de uma rede (uma espécie de RG de cada computador) e o TCP (*Transmission Control Protocol* ou Protocolo de Controle de Transmissão), o protocolo responsável por garantir¹² que a informação chegue ao seu destino.

É válida a observação que, por ter suas bases dentro do exército, a Arpanet (e na sequência a Internet) foi criada de maneira que toda e qualquer comunicação executada dentro da rede pode ser facilmente monitorada e capturada, não demandando nenhum tipo “poder especial” para tal. Neste ponto cria-se o vínculo com os casos de escândalos de espionagem já citados. Ao fazer parte da rede (neste caso especificamente da Internet), o usuário torna-se um alvo passível de espionagem por parte da sua operadora de telecomunicações, pelo seu governo e, por fim, pelo governo norte-americano¹³. Todavia, sob o olhar dos pesquisadores que tiveram suas universidades interconectadas pela Arpanet, o fato de poderem ter seus dados monitorados nunca foi uma dor de cabeça. Muito pelo contrário, com a possibilidade de poder compartilhar seus trabalhos com demais cientistas, a Arpanet acabou aproximando a comunidade de pesquisadores da época e acelerando uma parte do desenvolvimento científico, legado que a Internet traz consigo até os dias atuais.

Em um artigo sobre redes de computadores publicado na revista *Scientific American* em setembro de 1991, Robert Kahn cunha um termo que em um momento mais tarde se tornaria a segunda base¹⁴ para os mecanismos de busca: o *knowbot*. Na visão de Kahn, os *knowbots* seriam programas de computadores que navegariam pelas redes e executariam diversas funções. Em um exemplo apresentado na revista, é proposto que um *knowbot* seria capaz de procurar por veículos em determinadas faixas de preço com determinados acessórios (algo muito parecido com os comparadores de preço existentes atualmente). O projeto dos *knowbots* foi visto com bons olhos e um grupo¹⁵ de pesquisa no CNRI¹⁶ (*Corporation for National Research Initiatives* ou Corporação para Iniciativas Nacionais de Pesquisa) foi criado.

O primeiro dos buscadores

Antes mesmo do surgimento da web de Tim Berners-Lee, em 1990 foi criado o primeiro sistema de busca da Internet, nomeado Archie. Criado por Alan Emtage, estudante de ciências da computação na Universidade McGill de Montreal, o Archie era um sistema que executava varreduras em servidores de FTP¹⁷ abertos na Internet. Os resultados das varreduras eram salvos em bancos de dados que, mais tarde, eram indexados¹⁸. Desta maneira, os usuários podiam através do *prompt* de comando de seus computadores localizar “com maior facilidade” arquivos em servidores abertos.

No momento histórico em que o Archie foi criado, a Internet era composta em sua maio-

12 Desde que exista a possibilidade para que isso aconteça. O TCP fará sempre o maior esforço possível para a entrega do pacote (best effort).

13 Isso se dá, pois, parte da estrutura do DNS (Domain Name System ou Sistema de Nomes de Domínio, uma espécie de lista telefônica da Internet) está localizada nos Estados Unidos, o que torna passível o monitoramento de qualquer usuário.

14 Tomando que o Modelo de Espaço Vetorial criado por Gerard Salton é a primeira base dos mecanismos de busca modernos.

15 Disponível em: <<http://www.cnri.reston.va.us/home/koe/>>. Acesso em: 18 ago. 2015.

16 Organização norte-americana sem fins lucrativos que tem pesquisas voltadas ao desenvolvimento da infraestrutura de comunicação americana.

17 FTP (File Transfer Protocol ou Protocolo de Transferência de Arquivo) é um protocolo que tem por objetivo executar a transferência de arquivos entre computadores/servidores que estão em uma mesma rede ou ainda em redes geograficamente separadas. Este protocolo foi criado em outubro de 1985 e é utilizado até os dias atuais.

18 Índices em bancos de dados são mecanismos utilizados para acelerar as buscas nos mesmos, sendo invisíveis aos usuários finais.

ria por universidades, colégios e por centros educacionais que possuíam grupos de pesquisa de diferentes áreas. Em um artigo¹⁹ publicado no mesmo ano do lançamento do Archie, estimava-se que existiam, naquela época, cerca de 800 servidores de FTP abertos na Internet, com um total de 50 gigabytes em arquivos disponíveis para *download* (praticamente o mesmo tamanho dos HDs encontrados nos computadores pessoais da atualidade). Levando em consideração que a maioria destes arquivos eram relacionados a trabalhos científicos, o Archie apareceu como uma ferramenta para auxiliar os pesquisadores, não possuindo nenhum fim comercial e tão pouco a ideia de monitoramento e controle.

19 Disponível em: <<http://www1.chapman.edu/gopherdata/archives/Internet%20Information/whatis.archie>>. Acesso em: 18 ago.2015.

Os clones do Archie

Não demorou muito para que a ideia de Emtage se popularizasse no meio acadêmico. Pouco tempo após seu lançamento, projetos similares ao Archie surgiram na Internet. Dentre eles, destacaram-se os projetos chamados Jughead e Veronica. Em linhas gerais, ambos operavam de maneira similar ao Archie, todavia, o projeto Veronica indexava somente arquivos de texto simples e o projeto Jughead utilizava o protocolo Gopher²⁰ ao invés do protocolo FTP. Não cabe aqui adentrar a fundo na história do projeto Jughead e do projeto Veronica pelo fato de que estes projetos tiveram seu fim anunciado (juntamente com o Archie) com o surgimento da web. Entretanto, é importante ressaltar que no campo das ciências duras, quando um projeto específico ganha destaque, é normal que demais pesquisadores voltem seus olhares para ele, apontando benefícios e trazendo aprimoramentos.

20 O protocolo Gopher pode ser considerado o irmão mais novo da web. A navegação presente no Gopher assemelha-se muito à navegação da web por conta dos hyperlinks, todavia não existia o conceito de página. O usuário navegava somente pelos diretórios do servidor e acessava os arquivos de seu interesse.

Tim Berners-Lee e a Web

Graduado em física pela universidade de Oxford, Tim Berners-Lee foi o criador da web no ano de 1989 quando trabalhava no Cern (*European Organization for Nuclear Research* ou Organização Europeia para a Pesquisa Nuclear). A web (também conhecida por WWW - *World Wide Web* - Rede Mundial de Computadores) é um sistema de navegação em documentos hipermídia - contendo texto, imagem, som e vídeo - através de hipertexto. O intuito de Berners-Lee era criar um sistema a fim de facilitar o compartilhamento de textos entre os pesquisadores. Para tal, ele desenvolveu dois *softwares* distintos (o browser, que é um *software* cliente para a navegação de páginas e um software servidor onde ficariam hospedados os documentos); um protocolo de rede (que foi nomeado de HTTP - *Hypertext Transfer Protocol* ou Protocolo de Transferência de Hipertexto) e a linguagem de marcação HTML.

21 Disponível em: <<http://info.cern.ch/>>. Acesso em: 18 ago. 2015.

O primeiro website criado²¹ foi uma página do Cern, onde era possível obter informações sobre o funcionamento da recém criada web, como se utilizar o browser e como instalar um

servidor web. A página foi ao ar no dia 06 de agosto de 1991 e continua on-line até hoje. Ainda em 1991, Berners-Lee lançou o serviço chamado *WWW Virtual Library*²², o primeiro serviço de catálogo da Web. Neste serviço são cadastrados diversos hyperlinks relacionados às diversas áreas do conhecimento. Pode-se afirmar que na era da web este foi o primeiro serviço de busca disponível, porém, diferente do Archie, nele todos os hyperlinks eram cadastrados manualmente. Três anos mais tarde, Berners-Lee fundou o W3C (*World Wide Web Consortium* ou Consórcio da Rede Mundial de Computadores), um grupo de pesquisa com o objetivo de manter a evolução da web e sua interoperabilidade entre os diferentes sistemas operacionais. O W3C é o atual responsável por manter e atualizar a linguagem HTML.

Desde sua criação, a web sempre teve por objetivo principal o compartilhamento de informações entre as pessoas, de maneira livre e rápida. Sua história não mostra em nenhum momento o interesse ou intenção de ser uma ferramenta de controle ou manipulação. Infelizmente, como a web é apenas uma ferramenta que funciona dentro das teias da Internet, seu uso subordina-se à estrutura de controle existente, onde tudo o que é trafegado pode, através de determinadas maneiras, ser interceptado e lido, o que expõe a privacidade de todos os seus usuários. Diversos mecanismos de segurança foram criados para melhor proteger os usuários, como por exemplo o HTTPS (*Hyper Text Transfer Protocol Secure* ou Protocolo de Transferência de Hipertexto Seguro), onde os dados do usuário passam por um processo de criptografia²³. Contudo, grandes entidades, como a já citada NSA, possuem poder de processamento computacional suficiente para “quebrar” os processos criptográficos atuais²⁴. A fim de tentar burlar as organizações como a NSA, diversos projetos surgiram na Internet. Dentre eles destaca-se o conhecido TOR²⁵, o qual utiliza técnicas que, segundo os desenvolvedores, conseguem driblar as atuais tecnologias para impedir que usuários sejam rastreados. A ideia do projeto é boa e visa a segurança do usuário, porém já foi constatado que diversos serviços ilegais, como, por exemplo, o tráfico de drogas e de órgãos humanos, aderiram a essa rede colocando o sistema em cheque sob o ponto de vista ético e moral.

22 Disponível em: <http://vlib.org/>. Acesso em: 18 ago. 2015.

23 Criptografia é o conjunto de técnicas onde a informação é cifrada, ou seja, torna-se ilegível para que demais pessoas/organizações não tenham acesso a ela.

24 Disponível em: <https://tecnoblog.net/139612/nsa-protocolos-de-seguranca-internet/>. Acesso em: 18 ago. 2015.

25 Disponível em: <https://www.tor-project.org/>. Acesso em: 18 ago. 2015.

Os serviços de busca

A criação da web abriu um vasto campo de novas oportunidades com um potencial de crescimento gigantesco. Não demorou para que a ideia introduzida pelo Archie ganhasse espaço na web. Foi durante os anos de 1993 e 1994 que diversos mecanismos de busca começaram a surgir, iniciando um processo de desenvolvimento que perdura até os dias atuais. No primeiro momento, os mecanismos de busca dividiam-se em duas categorias: os serviços de diretório e os *crawlers* (rastreadores).

Os serviços de diretório ofereciam a busca por informações em bases de dados, onde as informações eram cadastradas pelos donos dos websites. Para que o site tivesse destaque na

busca, eram cadastradas diversas palavras-chave que possuíam relação com o site. Infelizmente, manter os serviços de diretório atualizados era uma tarefa árdua, que envolvia trabalho manual repetitivo. Isso gerava dois problemas: sites que nunca eram encontrados, pois não estavam cadastrados nos diretórios ou não possuíam palavras-chave suficiente para sua localização; ou ainda sites que possuíam palavras-chave em excesso e que não tinham correlação com a página encontrada. Dentre os serviços de diretório que surgiram, merecem destaque o Virtual Library, que foi anteriormente mencionado, o ElNet Galaxy, o Yahoo! Directory e o LookSmart.

Em outra linha de pesquisa e mais parecidos com o Archie, surgiram os chamados *crawlers*. Em linhas gerais, esses rastreadores são softwares que navegam nas páginas abertas presentes na web e buscam dentro das mesmas palavras-chave que, em um segundo momento, são armazenadas em um banco de dados e utilizadas na busca. Encontra-se aqui a ideia de Robert Kahn sobre os *knowbots*, que são pequenos softwares que vasculham as redes. Neste caso, informações específicas (as palavras-chave) são procuradas nos websites. A versatilidade dos *crawlers* é tão boa que foram eles os grandes responsáveis pela derrubada dos serviços de diretório. A busca na web virou algo tão importante que, na atualidade, as empresas de mídia costumam dispor de profissionais totalmente focados em uma área que ficou conhecida como SEO (*Search Engine Optimization* ou Otimização dos Motores de Busca). As equipes de SEO procuram entender como cada mecanismo de busca funciona. Com base nestes conhecimentos, eles alteram os websites das empresas onde trabalham para que cada vez mais apareçam nos primeiros resultados das buscas. Dentre os *crawlers*, merecem destaque os serviços WebCrawler, Lycos, Altavista, Yahoo! (que modernizou seu antigo serviço de diretório), Altavista e, por fim, o Google.

Como foi realçado no início deste texto, as buscas já são parte da nossa vida cotidiana. As pessoas, no dia a dia, raramente vão às bibliotecas atrás de informações, elas simplesmente pegam os computadores portáteis ou os smartphones e pesquisam o que desejam. O crescimento do volume de informações na web, juntamente com a diversidade de assuntos, aliados ao poder e velocidade dos mecanismos de busca, tornaram a busca de informações em dispositivos analógicos algo que não faz mais sentido nos tempos atuais. Isso não invalida os meios analógicos como fonte de informação e conhecimento. Mas fica evidente que é muito mais rápido e eficiente encontrar uma notícia em um buscador, por exemplo, do que folhear um jornal atrás da mesma.

Um ponto interessante a ser observado, que passa despercebido sob o olhar das pessoas mais leigas, é que a cada nova busca efetuada, mais dados sobre nossas vidas fornecemos aos serviços buscadores. E com isso as grandes empresas, por detrás desses serviços, conseguem traçar nossos perfis com maior facilidade. Com o perfil do usuário “em mãos” (classe finan-

ceira, local onde reside, time que torce, etc. – através das buscas realizadas em aplicativos de GPS²⁶) fica mais fácil ofertar serviços e produtos de forma mais assertiva. O poder de armazenamento dos perfis de usuário é tão grande que, atualmente, sistemas como o Google oferecem ao usuário a opção de ver todas as suas buscas realizadas até o presente momento. Se antes já era possível rastrear o usuário que utilizava a Internet, agora as empresas possuem um cadastro completo sobre seus gostos, nível de estudo e possíveis intensões de compra. Até que ponto é interessante que grandes corporações como Google, Microsoft e Facebook²⁷ saibam tanto de nossas vidas? É possível, nos dias de hoje, ser um anônimo em uma rede que é vigiada 24 horas por dia? Será que “a rede” sabe mais de nós do que nós mesmos? E a pior de todas as perguntas: conseguimos viver sem os mecanismos de busca?

26 Global Positioning System - Sistema de Posicionamento Global.

27 Dentro do Facebook e outras redes sociais buscas são efetuadas o tempo todo.

Considerações finais

Observa-se no decorrer da história, que tange a evolução das consideradas ciências duras, que praticamente todas as tecnologias tiveram sua criação com fins bem definidos, como foi o caso da Arpanet. Entretanto, cristaliza-se a ideia de que tecnologias como a dos mecanismos de busca tiveram nos últimos anos seu objetivo inicial distorcido, tornando-se verdadeiras minas de informação para empresas que detém o serviço de busca. Fica evidente que os mecanismos de busca têm o seu valor enquanto serviço para a população mundial, ajudando no dia a dia das pessoas e facilitando nossas vidas de um modo geral. Mas por outro lado, a cada nova pesquisa feita, mais informações fornecemos aos grandes grupos comerciais. Seria essa uma troca justa?

Referências

AT&T. **1916**: Remembering Claude Shannon. Disponível em: <<http://www.corp.att.com/atlabs/reputation/timeline/16shannon.html>>. Acesso em 10 mai. 2015.

BERNERS-LEE, T. **Tim Berners-Lee**. Disponível em: <<http://www.w3.org/People/Berners-Lee/>>. Acesso em 4 jul. 2015

_____. **Weaving the web: the original design and ultimate destiny of the World Wide Web**. New York: HarperBusiness, 2001.

BLUM, A. **Tubes: A Journey to the Center of the Internet**. New York: Ecco, 2013.

BRADLEY, G. D. **The story of the Pony Express**. Nova York: Amazon, 2011.

BUSH, V. **As we may think**. Disponível em: <<http://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/>>. Acesso em: 1º. abr. 2015.

CARROLL, R. **Bell did not invent telephone, US rules**. Disponível em: <<http://www.theguardian.com/world/2002/jun/17/humanities.internationalaleducationnews>>. Acesso em: 2 mai. 2015.

CORNELL UNIVERSITY. **IN MEMORIAM**. Disponível em: <<http://www.cs.cornell.edu/Info/Department/Annual96/Beginning/salton.html>>. Acesso em: 14 jul. 2015.

DERTOUZOS, M. L. Communications, computers and networks. **Scientific American**, New York, n.256, p. 196, set. 1991.

GLEIK, J. **A informação**: Uma história, uma teoria, uma enxurrada. São Paulo: Schwarcz, 2013.

GOLDSTINE, H. H. **The computer from Pascal to von Neumann**. Princeton: Princeton University Press, 1980.

GROSVENOR, E. S.; WESSON, M. **Alexander Graham Bell**: The Life and Times of the Man Who Invented the Telephone. Nova York: Abrams Books, 1997.

INDIANA UNIVERSITY. **What is FTP, and how do i use it to transfer files?** Disponível em: <<https://kb.iu.edu/d/aerg>>. Acesso em: 1º. mai. 2015.

LEVY, P. **As tecnologias da inteligência**. São Paulo: Editora 34, 2010.

MCCAHILL, M. et al. **The Internet Gopher Protocol**. Disponível em: <<https://tools.ietf.org/rfc/rfc1436.txt>>. Acesso em: 1º. mai. 2015.

MCGILL UNIVERSITY. **Alan Emtage**. Disponível em: <<https://aoc.mcgill.ca/greatest-mcgillians/alan-emtage>>. Acesso em: 15 abr 2015.

MORSE, S. F. B. **Samuel F. B. Morse**: his letters and journals. 2.v. Nova York: FQ Books, 2010.

POLLAK, M. **F.Y.I.** Disponível em: <<http://www.nytimes.com/2004/04/18/nyregion/fyi-245399.html>>. Acesso em 1º. mai. 2015.

POSTEL, J.; REYNOLDS, J. **FILE TRANSFER PROTOCOL (FTP)**. Disponível em: <<https://www.ietf.org/rfc/rfc959.txt>>. Acesso em 4 maio 2015.

SALTON, G. **A Theory of Indexing (CBMS-NSF Regional Conference Series in Applied Mathematics)**. Philadelphia: Society for Industrial and Applied Mathematics, 1987.

SEIFE, C. **Decodificando o Universo**. Rio de Janeiro: Rocco, 2010.

STANDAGE, T. **The victorian internet**: The Remarkable Story of the Telegraph and the Nineteenth Century's On-line Pioneers. Nova York: Bloomsbury USA, 2014.

TANENBAUM, A. S. **Sistemas Operacionais Modernos**. 3.ed. São Paulo: Prentice Hall Brasil, 2010.

W3SCHOOLS. **SQL Create index Statement**. Disponível em: <http://www.w3schools.com/sql/sql_create_index.asp>. Acesso em 22 jun 2015.

WWW VIRTUAL LIBRARY. **About the virtual library**. Disponível em: <<http://vlib.org/admin/AboutVL>>. Acesso em 10 jul 2015.

São Paulo, v.3, n.1, ago 2015

ISSN: 2357-7126

INSS-L: 2357-7126

Brazilian Journal of Technology, Communication, and Cognitive Science é produzida pelo Grupo de Pesquisa Tecnologia, Comunicação e Ciência Cognitiva, credenciado pelo Programa de Pós-graduação em Comunicação Social da Universidade Metodista de São Paulo. A revista do TECCOG é uma publicação científica semestral em formato eletrônico e foi lançada em setembro de 2013.

A Comunicação Social, enquanto campo do conhecimento pertencente à área das Ciências Sociais Aplicadas, dispõe de contínuos esforços no sentido de estabelecer e compreender os fenômenos científicos comunicacionais sob uma perspectiva inter e transdisciplinar. Assim, o foco científico da publicação mira a complexidade das relações entre ciência e tecnologia e os seus impactos cognitivos no ser humano e na sociedade.

A proposta é acompanhar e compreender cientificamente os caminhos trilhados pela evolução tecnológica no campo da Comunicação Social, construindo ferramentas teórico-metodológicas nas pesquisas na área, se adaptando também aos instrumentos de verificação desenvolvidos em outras áreas do conhecimento – em especial, na Ciência Cognitiva.

É, portanto, um campo de profunda investigação científica, de ação e métodos transdisciplinares, para avançar na compreensão de como as informações são absorvidas, transmitidas e processadas pelo sistema sensorial e pelo conjunto mente/cérebro do ser humano

Editor

Walter Teixeira Lima Junior

Editora da edição v.3, n.1, ago 2015

Krishma Carreira

Comissão Editorial

Walter Teixeira Lima Junior (Universidade Metodista de São Paulo) * Lúcia Santaella (Pontifícia Universidade Católica de São Paulo) * Luis Martino (UNB) * João Eduardo Kogler (Universidade de São Paulo) * Ronaldo Prati (Universidade Federal do ABC) * Ricardo Gudwin (Universidade Estadual de Campinas) * João Ranhel (Universidade Federal de Pernambuco) * Eugenio de Menezes (Faculdade Cásper Líbero) * Reinaldo Silva (Universidade de São Paulo) * Marcio Lobo (Universidade de São Paulo) * Vinicius Romanini (Universidade de São Paulo)

Conselho Editorial

Walter Teixeira Lima Junior (Universidade Metodista de São Paulo) * Lúcia Santaella (Pontifícia Universidade Católica de São Paulo) * Luis Martino (UNB) * João Eduardo Kogler (Universidade de São Paulo) * Ronaldo Prati (Universidade Federal do ABC) * Ricardo Gudwin (Universidade Estadual de Campinas) * João Ranhel (Universidade Federal de Pernambuco) * Eugenio de Menezes (Faculdade Cásper Líbero) * Reinaldo Silva (Universidade de São Paulo) * Marcio Lobo (Universidade de São Paulo) * Vinicius Romanini (Universidade de São Paulo)

Assistente Editorial

Walter Teixeira Lima Junior
Krishma Carreira

Projeto Gráfico e Logotipo

Danilo Braga * Walter Teixeira Lima Junior * Leandro Tavares

Editoração eletrônica

Eduardo Uliana