

# O mecanismo de busca do Google e a relevância na relação sistema-usuário

## *Google's search engine and the relevance in system-user interface*

Fátima Hassan Caldeira<sup>1</sup>

<sup>1</sup> Doutoranda em Ciências da Linguagem pela Universidade do Sul de Santa Catarina. Mestre em Ciências da Linguagem pela mesma instituição. Este artigo deriva de pesquisa para projeto de tese já qualificado. [fatima.caldeira@unisol.br](mailto:fatima.caldeira@unisol.br)

**RESUMO:** Este artigo apresenta pesquisa teórica que tem como objeto de análise o mecanismo de busca do Google. Parte-se da hipótese, inaugurada por Yus através da Ciberpragmática, de que a Teoria da Relevância é capaz de dar uma explicação científica às atividades humanas na Net, já que, para este teórico, as pesquisas sobre os procedimentos cognitivos da mente humana em cenários físicos implicam necessariamente um paralelismo em ambientes virtuais. Assim sendo, em decorrência da diluição da fronteira entre o virtual e o real, a relevância, além de orientar a comunicação na Internet, também orienta qualquer atividade humana *on-line*. Dentre as atividades humanas na Net, a pesquisa privilegia a *relevância procurada pelo sistema para o usuário* como lugar de análise, pois é nessa área que se insere o resultado de um mecanismo de busca tal como o Google. Do ponto de vista operacional, defende-se a hipótese de que as ferramentas fornecidas pela Teoria da Relevância (Sperber e Wilson), aliadas aos achados da Ciberpragmática (Yus), são capazes de analisar e descrever o mecanismo de busca do Google. O estudo presume que o mecanismo de busca 'metarrepresenta' a meta de um usuário que utiliza sua interface de busca, maximiza a relevância do *input* (a consulta, a busca) para disponibilizar *outputs* (resultados do mecanismo de busca) que atendam à relevância de cada indivíduo, aumentando os efeitos cognitivos e diminuindo o esforço de processamento do usuário para que este alcance a meta que o levou à consulta no buscador. Para corroborar tal presunção, apresentamos os avanços tecnológicos operados no mecanismo de busca do Google, com a introdução de seu novo algoritmo - o *Hummingbird* - e da Web Semântica, o reconhecimento do comportamento do usuário na internet e a personalização dos resultados das buscas de tal modo a atender, de modo relevante, às necessidades e intenções dos usuários, numa tentativa de contextualizar a experiência do usuário.

**PALAVRAS-CHAVE:** Mecanismo de busca; Google; Ciberpragmática; Teoria da Relevância; Websemântica.

**ABSTRACT:** This paper presents a theoretical research that has as its object of analysis Google's search engine. It starts with a hypothesis, inaugurated by Yus through Cyberpragmatic, explaining that the Relevance Theory is able to give a scientific explanation to human activities all over the Net, considering that according the mentioned theorist, the researches concerning to cognitive procedures of human mind in physical scenarios necessarily imply a parallelism over virtual environments. Therefore, due to the blurring of the distinction between virtual and real, the relevance, in addition to guiding communication mediated by Internet, also guides any online human activity. Among the human activities on Net, this research focuses the relevance sought by the system for an user as a place of analysis, in view of that the result of Google search engine fits exactly on that area. From an operational point of view, the defended hypothesis refers that tools provided by Relevance Theory (Sperber and Wilson), associated with the findings of Cyberpragmatic (Yus), have ability to analyze and describe the Google search engine. Moreover, this study assumes that the search engine 'metarepresents' a goal of an user who uses its search interface, maximizes the relevance of its input (inquiry, search) supplying outputs (search engine results) according to the relevance of each user, increasing the cognitive effects and decreasing the processing effort applied to reach the goal which motivated the search. In support of this assumption, it presents the technological advances concerning Google search engine, with the introduction of its new algorithm - Hummingbird - and Semantic Web, besides the recognition of user behavior on Internet and the personalization of search results in order to attend, in a relevant way, to the needs and intentions of users, in an attempt to contextualize the user experience.

**KEYWORDS:** Search engine; Google; Cyberpragmatic; Relevance Theory; Semantic Web.



## Introdução

Desde 2001, Francisco Yus propôs um projeto investigativo denominado Ciberpragmática, que preconiza a aplicação da Teoria da Relevância (TR) de Sperber e Wilson (1986, 1995) à comunicação mediada pela Internet. Mas, em trabalho mais recente, Yus apregoa que a TR pode fornecer uma explicação válida para toda a atividade humana na Net<sup>1</sup>, já que, segundo ele, a cognição humana não é orientada apenas para a relevância na comunicação, mas também para todos os aspectos de nossa vida diária. Assim, a relevância, além de orientar a comunicação na Internet, também orienta qualquer atividade humana *on-line*. Isso decorreria do fato de não podermos divisar a linha que separaria o *off-line* do *on-line* em nossas vidas, diante da tendência de realizarmos cada vez mais atividades na Rede, o que leva a uma diluição das fronteiras entre o virtual e o real (cf. YUS, 2013)<sup>2</sup>.

Para Gabriel (2013, p. 57), tal diluição de fronteiras leva ao chamado cibridismo que se caracteriza por permitir ao ser humano “transferir parte de si para o mundo digital”. De acordo com a autora, isso decorreu da hiperconexão e da proliferação de plataformas digitais móveis que foram popularizadas socialmente, os denominados *mobiles* ou dispositivos móveis (*tablets* e *smartphones*, por exemplo). Assim, podemos dizer que somos híbridos, pois “somos on e off-line ao mesmo tempo, simbioticamente, formando um ser maior que o nosso corpo/cérebro biológico, nos expandindo para todo tipo de dispositivo e abrangendo outras mentes e corpos” (GABRIEL, 2013, p. 58).

Diante desse contexto, Yus (2012) afirma que todas as pesquisas sobre os procedimentos cognitivos da mente humana em cenários físicos implicam

necessariamente um paralelismo em ambientes virtuais. Daí advém a possibilidade, apontada pelo autor, de a TR poder fornecer uma explicação válida para toda atividade humana na Net.

Dentre as áreas da mencionada atividade humana na Net classificadas por Yus (2012), nossa pesquisa encaixa-se *na relevância procurada pelo sistema para o usuário*, já que o objeto de nossa pesquisa é o mecanismo de busca<sup>3</sup> do Google.

Yus (2014) ressalta que, a princípio, a pragmática teria como escopo a análise da comunicação usuário-a-usuário, porém ela não pode deixar de analisar a comunicação sistema-usuário uma vez que esta desempenha um importante papel “na quantidade de conteúdo que é disponibilizado para processamento, no tipo de conteúdo que realmente será processado e como este será (ou não) proveitosamente combinado com o conhecimento de *background* do usuário para gerar conclusões relevantes”.

Segundo Yus (2012), a popularidade do Google resulta da utilização de algoritmos<sup>4</sup> que produzem os resultados mais relevantes, já que “o Google estuda o costume e o rendimento dos usuários e produz resultados personalizados” para cada um deles. Para o autor, o sistema é capaz de aprender com a atividade humana na Net e propor automaticamente fontes de informação por julgar serem potencialmente relevantes para cada usuário (2012).

Sugerimos, como hipótese de pesquisa, que o Google tem buscado, a cada melhoria de seu algoritmo, inserir mecanismos capazes de ‘metarrepresentar’ a meta de cada usuário quando estes utilizam a interface do buscador para uma consulta. Defendemos que a finalidade das constantes mudanças em

<sup>3</sup> Também conhecidos como *serch engines*, sites de busca, motores de busca ou buscadores.

<sup>4</sup> Em informática, um algoritmo é “uma sequência ordenada, definida e finita de ações que visam à solução de um determinado problema computacional. Em suma, o problema contém um conjunto de dados de entrada (*input*) e o algoritmo, na sequência das ações resolventes, produz os dados de saída (*output*)”. (ALGORITMO, 2014).

<sup>1</sup> Tomaremos como sinônimos os termos Internet, Net e Rede.

<sup>2</sup> É nossa a tradução das citações extraídas de obras em língua estrangeira.

seu algoritmo visa, portanto, a que os resultados de seu buscador alcancem a relevância para um indivíduo, aumentando os efeitos cognitivos<sup>5</sup> e reduzindo o esforço de processamento de seus usuários na tentativa de atender às metas de cada um deles.

Obviamente que, quando falamos em Metarrepresentação na presente pesquisa, não estamos sugerindo que o mecanismo de busca do Google seja capaz de leitura da mente no sentido que é atribuído aos seres humanos. Como sugerimos adiante, os mecanismos utilizados para a personalização e a contextualização da experiência do usuário é que permitem que o buscador ‘metarrepresente’ a meta do usuário.

Defendemos que tal ‘metarrepresentação’ da meta é necessária para que o mecanismo de busca do Google heteroconcilie seus resultados às metas do usuário que utiliza sua interface, nos moldes da Teoria de Conciliação de Metas de Rauen (2014). Trata-se de uma atividade colaborativa entre o buscador do Google e seu usuário que pressupõe reconhecimento de intenções do usuário, para que os resultados da busca efetivamente colaborem para a meta deste e, ao mesmo tempo, para a meta do Google que está em prestar os resultados mais relevantes, com menor esforço de processamento e com maiores efeitos cognitivos para a satisfação plena de seu usuário.

Assim, baseando-nos no Princípio Cognitivo de Relevância, que afirma que o ser humano tem sua cognição biologicamente orientada à maximização da relevância, e na Relevância para um indivíduo, que afirma que uma suposição será relevante para um indivíduo se, na relação custo-benefício, os efeitos cognitivos alcançados quando ela é otimamente processada

<sup>5</sup> Dentre os efeitos cognitivos descritos pela TR temos: a) Implicação contextual: que ocorre quando suposições resultam da síntese de uma informação antiga com uma informação nova; b) Fortalecimento ou enfraquecimento de suposições: que ocorre quando a informação nova fornece evidências que fortalecem ou enfraquecem uma informação antiga; c) Eliminação de suposições contraditórias: que ocorrem quando a informação nova fornece evidências que levam ao abandono de suposições antigas (cf. SPERBER; WILSON, 2001; SILVEIRA; FELTES, 2002).

forem amplos ou, ainda, quando o esforço requerido para o processamento ótimo for pequeno (cf. SPERBER; WILSON, 2001), o objetivo de nossa pesquisa está em descrever e explicar o mecanismo de busca do Google e as constantes inovações em seus algoritmos (em especial a introdução da Web Semântica e da Personalização) de forma a confirmar nossa hipótese de pesquisa.

O presente artigo foi dividido em cinco seções: a primeira trata da chamada Era da Busca; a segunda apresenta as alterações no processamento de informações trazidas pela Era Digital; a terceira introduz os Mecanismos de Busca; a quarta aprofunda os achados especificamente no buscador do Google e a última apresenta alguns aspectos acerca da Web Semântica e da Personalização, já que essas interferem no sentido de ‘metarrepresentação’ adotado em nossa pesquisa.

## 1 A Era da Busca

A busca, no sentido atribuído em nossa pesquisa, é definida como o processo em que obtemos informações relevantes mediante “a aplicação de critérios e filtros de seleção em grandes volumes de informação” (GABRIEL, 2012, p. 24). Com a Web 2.0<sup>6</sup>, abandonou-se o modelo de mero consumo de conteúdos em uma internet estática que só possibilitava a leitura (Web 1.0), para dar lugar à Web participativa, dinâmica e colaborativa, que pode ser usada como base para todo tipo de interações (GABRIEL, 2013, p. 22) e que permite, potencialmente, a qualquer pessoa criar, publicar ou compartilhar conteúdos na internet.

Isso tem levado ao que Gabriel (2013, p. 26) denomina de “info-obesidade”, devido à multiplicação de conteúdos na Web em velocidade

<sup>6</sup> Foi Tim O’Reilly que cunhou o termo Web 2.0. Trata-se de uma plataforma de participação que permite não só o consumo, mas também a colocação de conteúdos. (cf. GABRIEL, 2012).

vertiginosa e sem qualquer controle de qualidade. Para Yus (2012), essa maior quantidade de informação disponível pode ter alterado a boa vontade do usuário e a capacidade de filtrá-la e de processá-la de forma eficiente. Pesquisas apontam que o ser humano não consegue operar em um ambiente de grande volume informacional sem que utilize um sistema que seja capaz de filtrar o que realmente interessa (cf. GABRIEL, 2013). Ou seja, tornou-se muito difícil, diante da explosão informacional, a obtenção de informações relevantes. Assim, por causa da “info-obesidade” na Web<sup>7</sup>, passou a existir a necessidade de uma “mediação tecnológica para filtrar e validar a informação” e, também, para aliviar a angústia gerada no ser humano pela dificuldade de encontrar aquilo de que precisa (GABRIEL, 2013, p. 33). Essa mediação dá-se através dos mecanismos de busca.

A mencionada angústia, segundo Gabriel (2012), derivaria do denominado “paradoxo da escolha”: quanto mais nossas opções de escolha aumentam, maior nossa angústia diante das inúmeras opções equivalentes disponíveis. Assim, vivemos a dúvida angustiante de saber se a nossa escolha realmente foi a melhor.

Além disso, conforme Davenport e Beck (2002 apud GABRIEL, 2013, p. 32), “a atenção é um recurso finito e, conforme a quantidade de informação aumenta, conseguimos prestar menos atenção em tudo”. Assim, para os autores, “quanto maior a riqueza da informação, maior a pobreza da atenção”. Trata-se do fenômeno, por eles denominado, de “Economia da atenção”. Já para Clay Shirky ([s.d.] apud GABRIEL, 2013, p. 33), o problema não estaria no excesso de informações, mas na falta de filtros.

<sup>7</sup> Web e Internet são coisas distintas. “A Internet é [...] uma rede mundial de computadores ou terminais ligados entre si, [...] de uma forma que os usuários conectados possam usufruir de serviços de informação e comunicação de alcance mundial [...]” (MORAIS; LIMA; FRANCO, 2012, p. 42). A *World Wide Web* ou Web é parte integrante da Internet e “pode ser vista como um imenso banco de dados, contendo informações providas por centenas de milhares de autores (pessoas e instituições)” (MORAIS; LIMA; FRANCO, 2012, p. 58).

A chamada computação ubíqua<sup>8</sup> também contribuiu para a ampliação da busca *on-line*. Com as melhorias na conectividade, o acesso à banda larga e os dispositivos móveis, podemos acessar qualquer informação independente de tempo e de lugar, potencializando a conveniência da busca (cf. GABRIEL, 2012).

Assim, o conjunto de todos esses fenômenos mencionados contribuiu para a criação de um ambiente adequado para o surgimento dos chamados mecanismos de busca: filtros que facilitaríamos nossas escolhas e reduziriam a nossa angústia diante da explosão informacional na Web.

## 2 Alterações no processamento de informações

As alterações no processamento de informações provocadas pela era digital também contribuíram para que nos tornássemos cada vez mais dependentes dos buscadores e para que estes, constantemente, estejam melhorando a relevância de seus resultados.

Diante da mencionada info-obesidade, há um bombardeio de estímulos sem precedentes que tem provocado “adaptações” cognitivas no ser humano, em especial nos denominados nativos digitais<sup>9</sup>.

Segundo Yus (2011), diante da tela de um computador, várias fontes de informação textual, visual e multimodal competem por nossa atenção. Dada essa competição, o leitor passa a ser incapaz de concentrar-se em apenas uma tarefa cognitiva, levando o cérebro a tornar-se inquieto e ávido por satisfação imediata com o mínimo de esforços.

<sup>8</sup> “A computação ubíqua surge [...] da necessidade de se integrar a mobilidade com a funcionalidade da computação pervasiva, ou seja, qualquer dispositivo computacional, enquanto em movimento conosco, pode construir, dinamicamente, modelos computacionais dos ambientes nos quais nos movemos e configurar seus serviços dependendo da necessidade” (ARAÚJO, 2003, p. 50).

<sup>9</sup> Indivíduos com menos de 20 anos que já nasceram imersos no acelerado ritmo dos avanços tecnológicos (cf. GABRIEL, 2013).

Isso vem ao encontro da Teoria de Carr (2010 apud YUS, 2011) que, ao caracterizar a leitura de gêneros textuais digitais como superficial, sugere que temos a tendência de “engoli-los” sem reflexão profunda, habituando o cérebro a receber informações “de forma rápida e em pequenas porções”, privilegiando “a eficácia do instantâneo e a imediatez do estímulo acima do sossego e da reflexão”. Isso leva à diminuição da capacidade de concentração, reflexão e contemplação que se reflete, inclusive, no nível físico. Diante disso, estaríamos perdendo a nossa capacidade de sustentar uma linha de raciocínio durante um longo período de tempo.

Segundo Yus (2013), ao comentar o pensamento de Carr (2010) acerca dos efeitos nocivos da Internet, nossos cérebros estão viciados em microestímulos, precisando receber *inputs* o tempo todo, inquietos e impacientes em vista da necessidade de gratificação constante. Uma das razões

é o aumento da quantidade de micromensagens de relevância imediata para o usuário, mas que exigem pouco esforço de processamento em troca. [...] acabamos viciados em uma represa direta de micromensagens cujo equilíbrio de efeitos e esforço está transformando nossas mentes em mecanismos cognitivos preguiçosos na medida em que muitos usuários intensos da internet já não são capazes de dedicar o esforço necessário para processar um texto longo, como um romance. Estamos muito impacientes. Nossas mentes desejam relevância imediata, a satisfação imediata de flashes diretos de informação. (YUS, 2013)

As características de grande parte dos gêneros digitais, em especial dos que Yus (2011) classifica como *autóctonos*, isto é, que não possuem um correlato impresso, têm contribuído para essas mudanças. Os usuários da Web estão bastante acostumados às qualidades de hipertextualidade, não linearidade e multimodalidade dos gêneros digitais.

Tanto a navegação quanto a conexão na internet são processos não lineares.

depende sensivelmente de por onde entramos para começar a navegação e se desenvolve [...] de forma imprevisível. Nem a própria pessoa, quando entra na internet, pode prever os caminhos que seguirá na rede, e a cada acesso ou a cada clique, a navegação se desenvolve de forma diferente [...]. (GABRIEL, 2013, p. 115)

A não linearidade deriva da característica de hipertextualidade dos textos digitais. Para Lévy, um hipertexto caracteriza-se por um conjunto de nós (palavras, páginas, imagens, etc.) ligados por conexões (1993, p. 33). Esses nós são os chamados *links*: elos hipertextuais que conectam blocos de conteúdo que se caracterizam por sua multimodalidade (BRITO; SAMPAIO, 2013, p. 299-300). Trata-se, portanto, de um texto que foi criado a partir de uma estrutura de *hiperlink* e que, por isso, permite uma leitura não linear, multidirecional e imprevisível (cf. YUS, 2011). Por fim, os textos digitais são multimodais, pois neles convivem diferentes níveis semióticos: verbais, imagéticos e sonoros.

Outro fator importante reside no fato de que a evolução tecnológica gradativamente tem-nos libertado de “limitações geográficas e temporais, acelerando cada vez mais os processos comunicacionais” (GABRIEL, 2013, p. 16). A banda larga distribuiu “o poder entre os nós das redes (pessoas), transformando o cenário de criação, publicação e distribuição de informações e conteúdos no mundo” (GABRIEL, 2013, p. 17), como já ressaltamos ao falar da Web 2.0.

A multitarefa (*multitasking*), outro ponto destacado por Yus (2011), caracteriza a atual tendência de as pessoas executarem várias tarefas simultâneas. Aliás, os nativos digitais são também conhecidos como “geração Z” exatamente pela característica de “zapear”, ou seja, de realizar várias atividades simultaneamente ou mudar rápida e alternadamente de uma atividade para outra (cf. ALVES; CERQUEIRA, 2013, p. 7). Segundo Yus, isso é um desafio para nosso sistema cognitivo de processamento e reflete no



desempenho da mente e na inequação interesse-esforço (2011). Para o autor, todas essas atividades em paralelo acabam por exigir atenções parciais e recursos cognitivos adicionais para voltar à atividade inicial, pois a atenção do usuário está focada em outras atividades paralelas no computador.

Gabriel acresce, ainda, que a multitarefa tem afetado nosso cérebro, “diminuindo nossa capacidade de concentração e produtividade” (2013, p. 52). Esses efeitos também são destacados por Yus (2011) ao afirmar que a multitarefa tem como desvantagem tornar os que a ela se dedicam “menos produtivos em seus trabalhos do que aqueles que realizam uma atividade e logo passam a realizar a seguinte”, já que “o cérebro humano não está preparado para saltar de uma tarefa para outra, pois se dispersa com facilidade e perde a capacidade de estimar o equilíbrio entre interesse e esforço”. Apesar disso, para Gabriel (2013), a multitarefa pode ser classificada como um dos comportamentos digitais que tendem a ser mais viciantes.

Em relação à memória, como as informações estão amplamente disponíveis, não há mais porque armazenar informações em nossos cérebros, ainda mais com o auxílio de mecanismos de busca. Gabriel (2013), com base em diversos estudos em relação às transformações causadas pelo Google em nosso cérebro, destaca o fato de não lembrarmos mais das informações em si, mas sim de onde as encontramos.

Mas há uma nova necessidade que está emergindo: “Em um contexto sobrecarregado de informações, a principal habilidade necessária passa a ser como escolher a informação correta em cada situação, como validar, organizar, extrair significado, refletir e solucionar problemas” (GABRIEL, 2013, p. 104). É aqui que os mecanismos de busca têm importante papel ao atender as necessidades de cada usuário no momento da busca e ao validar as informações de seus resultados.

Por isso, como veremos adiante, os mecanismos de busca constantemente estão melhorando seu algoritmo, pois precisam atender, de modo veloz e

cada vez mais preciso, a essa necessidade de imediatez de informação, na procura de diminuir os esforços de processamento do usuário e aumentar os efeitos cognitivos advindos de seus resultados.

### 3 Mecanismos de busca (*search engines*)

Os mecanismos de busca são sistemas de recuperação de informações cuja finalidade está em “auxiliar na busca de informações armazenadas em ambientes computacionais” e cuja utilidade pode ser mensurada na relevância (qualidade) e na rapidez de seus resultados (velocidade) (cf. GABRIEL, 2012, p. 36).

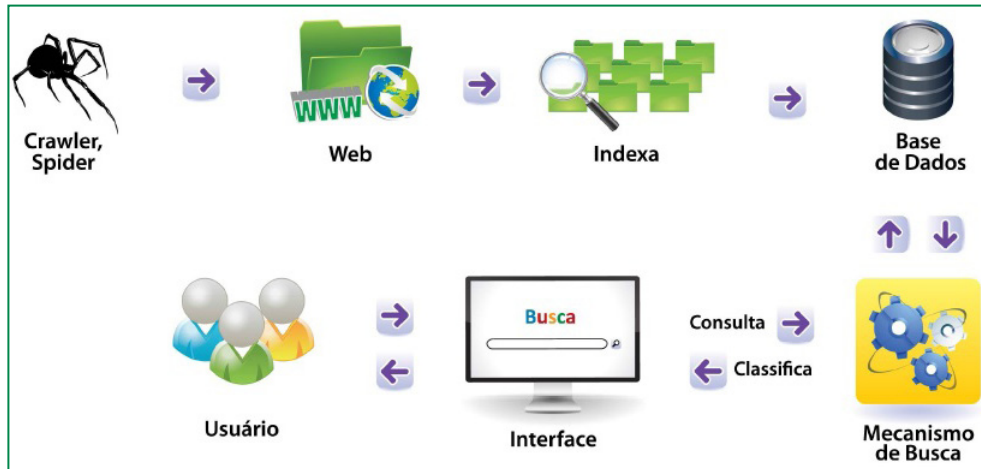
Segundo Fernandes et al. (2012), são três as funções essenciais para a qualidade e velocidade de tais mecanismos. Ordenando-as a partir de sua operacionalidade, temos:

- a) *crawling*: um programa denominado *crawler*<sup>10</sup> varre a Web coletando páginas novas e atualizadas para serem incluídas no índice;
- b) *indexing*: as informações recolhidas são armazenadas e indexadas na base de dados;
- c) *searching*: uma interface de busca é exibida para o usuário realizar a pesquisa, a interface e o software relacionado que conecta a busca do usuário com o índice executa um algoritmo para encontrar e exibir as páginas relevantes.

Assim, a anatomia de um mecanismo de busca pode assim ser representada:

<sup>10</sup> O *crawler* é o rastreador (cf. Fernandes et al., 2012). Também pode ser denominado *spider* (aranha) (cf. FIGUEIREDO, 2006, p. 31) ou simplesmente robô ou *bot*.

Figura 1 - Anatomia de um mecanismo de busca



Fonte: Fernandes et al. (2012).

Toda busca inicia pela Web, mas somente a Web visível<sup>11</sup> é rastreada pelos buscadores, porém ela não consegue ser totalmente indexada por eles. Estima-se que a Web profunda seja pelo menos cem vezes maior do que a Web visível. Ou seja, menos de 1% de toda a Web é rastreada (cf. GABRIEL, 2012).

Após o rastreamento, os mecanismos têm o poder de determinar quais páginas serão indexadas ou não em suas bases de dados. Os buscadores têm critérios para essa filtragem. O Google, por exemplo, tem o poder de apagar ou eliminar dados anteriormente indexados como punição no caso de suspeita de *spam*, e, ainda, o poder de controlar ou filtrar os

<sup>11</sup> “A “Web visível” ou “indexável” (surface Web) é a parte da Web que os buscadores conseguem acessar (“ver”) para poder indexar (acrescentar dados ao seu índice). A “Web profunda” (deep Web) ou “Web invisível” refere-se à parte da Web que não pode ser acessada pelos buscadores – essa parte consiste em páginas dinâmicas e conteúdos bloqueados por trás dos bancos de dados e outros sistemas que os mecanismos de busca não conseguem extrair” (GABRIEL, 2012, p. 47).

resultados de busca sob a pressão da censura de alguns Estados<sup>12</sup> (cf. GABRIEL, 2012).

Após a indexação, entra em cena a busca do usuário através da interface do buscador. Para Lévy (1993 apud GABRIEL, 2011, p. 51), “uma interface homem/máquina designa o conjunto de programas e aparelhos materiais que permitem a comunicação entre um sistema informático e seus usuários humanos”. Segundo Gabriel (2011, p. 51), “a interface é uma relação a três: o computador, o usuário, e o designer que a projeta. Dependendo do modo como é projetada, a interface intermediará a relação entre homem e computador de maneiras diferentes”. Para a autora, “as interfaces e tecnologias a elas associadas afetam e guiam nossa percepção e o modo como agimos no mundo” (2011, p. 51) e “transformam o modo como criamos e nos comunicamos” (JOHNSON, 2001 apud GABRIEL, 2011, p. 51).

Diante da tendência de os usuários privilegiarem a primeira página de resultados do buscador, surge o fenômeno denominado por Gabriel (2012) de “Ditadura dos Top 10”. Pesquisas revelam

[...] que 52% dos usuários clicam em um resultado de busca na primeira página [...] que apenas 10% clicam em algum resultado além da terceira página. [...] as três primeiras posições da primeira página de resultados recebem metade dos cliques totais da página, e a primeira posição recebe quase o dobro de cliques que a segunda recebe. (GABRIEL, 2012, p. 49)

Tal fenômeno tem levado a que cada vez mais pessoas e empresas utilizem técnicas para otimizar seus sites (SEO – *Search Engine Optimization*) para que eles possam ser posicionados na primeira página de resultados em uma busca e, de preferência, nas primeiras posições, o que tem afetado a produção textual na internet em diversas áreas.

<sup>12</sup> Na China, o Google não apresenta, em seus resultados, informações sobre a revolta da Praça de Tian’anmen e sobre a luta pela independência tibetana.

## 4 O Google e seu mecanismo de busca

Quando abordamos a Era da Busca e as alterações no processamento de informações, ficou clara a necessidade de mediação tecnológica para filtrar e validar as informações. Quanto à validação de informações, são esclarecedores os achados de Sperber et al. (2010) quando abordam a vigilância epistêmica em uma escala populacional e consideram o Google como um sistema de avaliação sem o qual seria incapacitada a utilização da Web no todo:

O Google não é apenas um motor de busca, mas também é usado como um mecanismo de avaliação epistêmica. Ele representa, implicitamente, sob a forma de uma lista ordenada, os valores epistêmicos relativos dos documentos da Web encontrados na pesquisa. Quanto maior a classificação de um documento, mais provável é conter as informações relevantes e confiáveis.

Dados da *Pew Internet* (2012) demonstram que a grande maioria dos usuários de mecanismos de busca confia nessas ferramentas, acreditando que são justas e imparciais, e também que os resultados que elas geram são precisos e confiáveis.

Segundo dados estatísticos de pesquisa realizada em 2013, o Google ocupa a posição de líder entre os buscadores mais utilizados no Brasil com 85,45% de participação dos internautas (cf. COM..., 2013). Já, o mercado global do Google, segundo pesquisa de 2012 realizada pela *StatCounter*, é de 91,78% (cf. FERNANDES et al., 2012).

Como todo mecanismo de busca, o Google segue a anatomia anteriormente descrita. Tudo começa com o rastreamento (*crawling*) e a indexação das páginas (*indexing*). O rastreador do Google é conhecido como *Googlebot* e tem a função de descobrir páginas novas e atualizadas para serem incluídas no índice do Google. Em realidade, o Google utiliza um grande número de computadores, pois ele foi desenvolvido para ser distribuído em várias máquinas a fim de melhorar seu desempenho ao rastrear as páginas da Web

visível. “O Googlebot usa um processo de algoritmos [...] que determinam quais sites devem ser rastreados, com que frequência e quantas páginas devem ser buscadas em cada site” (GOOGLE, 2014b).

O processo de rastreamento do *Googlebot* começa com uma lista de URLs<sup>13</sup> de páginas da Web. A lista é gerada a partir de processos anteriores de rastreamento, mas é aumentada com dados dos Mapas de Site (*Sitemaps*) fornecidos por *webmasters*<sup>14</sup>. Os Mapas de Sites são listas de todas as páginas de um site que visam guiar e facilitar a navegação do robô do Google por esse site. À medida que o *Googlebot* “visita cada um desses sites, ele detecta os links [...] de cada página e os inclui na lista de páginas a serem rastreadas. Novos sites, alterações em sites existentes e links inativos serão detectados e usados para atualizar o índice do Google”. (GOOGLE, 2014b). O *Googlebot* segue um link de uma página para outra transmitindo os dados dessas páginas aos servidores do Google (cf. GOOGLE, 2014a).

Após o rastreamento, o Google mantém o controle de tudo em um índice que está em constante expansão e que contém, atualmente, mais de 100 milhões de gigabytes. (cf. GOOGLE, 2014a; GOOGLE, 2014b). A indexação consiste em uma forma de reunir e organizar as informações rastreadas na Web, levando em conta diversos aspectos distintos para que seus usuários possam encontrar o que procuram: “data de sua publicação, seu conteúdo em termos de fotos e vídeos, etc.” (GOOGLE, 2014c).

Desde que o Google adotou a busca semântica (veremos adiante), o Mapa ou Gráfico do Conhecimento, uma base semântica de conhecimento, tem sido utilizado para organizar informações sobre pessoas reais, lugares e coisas.

<sup>13</sup>URL – “O Localizador Uniforme de Recurso **URL**, ou (*Uniform Resource Locator*), é um caso particular do Identificador Uniforme de Recursos **URI** ou (*Uniform Resource Identifier*), [...] que descreve o endereço de uma página na Web (o servidor que a hospeda e o nome do documento neste servidor) e o mecanismo (protocolo) utilizado para o acesso (HTTP, FTP etc.)”. (MORAIS; LIMA; FRANCO, 2012, p. 61, grifo nosso).

<sup>14</sup>“Pessoa responsável pela edição e/ou administração de um site [...]” (WEBMASTER, 2014).



A ideia do Mapa do Conhecimento parte da crença de que as palavras que os usuários digitam na busca referem-se a elementos do mundo real: pessoas, livros, filmes, etc. Ao compreender as relações entre as coisas e ao organizá-las nesse Mapa, o Google busca compreender melhor o que exatamente um usuário procura, conectando as informações da Web ao que é consultado pelo usuário (cf. GOOGLE, 2014d).

Quando um usuário inicia uma busca no Google, o algoritmo começa a trabalhar procurando pistas para compreender melhor o que o usuário quer dizer e extrair os documentos relevantes do índice. Dentre essas pistas temos (cf. GOOGLE, 2014a): o método de pesquisa (por texto, por imagem, por voz; etc.); o Google *Instant* (aperfeiçoamento de pesquisa que apresenta os resultados enquanto o usuário digita); a ortografia (identifica e corrige possíveis erros de ortografia, oferecendo alternativas); o *Autocomplete* (ferramenta que apresenta previsões de pesquisa que podem ser semelhantes aos termos de busca que o usuário está digitando, inclusive termos com mais de um significado); os Sinônimos (reconhece palavras com significados semelhantes); Entendimento da Consulta (*Query Understanding* – que procura obter os significados mais profundos das palavras que o usuário digita).

Em seguida, os resultados são classificados com base em mais de 200 fatores, cuja maioria, por razões estratégicas, não é revelada pela empresa. Dentre eles (cf. GOOGLE, 2014a): a confiabilidade, a reputação e a autoridade da página ou site (*PageRank*<sup>15</sup>); a novidade e atualidade; o contexto do usuário<sup>16</sup>; a adaptação dos resultados com base no idioma e no país

<sup>15</sup> Algoritmo “que atribui uma pontuação (um PageRank) a páginas Web, de acordo com a quantidade e a qualidade das ligações (externas ou internas) que apontem para ela [...] quanto mais ligações existirem apontando para uma página, maior é seu grau de importância no Google. Como consequência, essa página tem maior probabilidade de obter um bom posicionamento nas buscas [...]” (GOOGLE..., 2012).

<sup>16</sup> “O Google tem sido criticado por colocar por longos prazos cookies nas máquinas dos usuários para armazenar suas preferências, uma tática que lhes permite acompanhar termos de pesquisa dos usuários [...]” (GOOGLE..., 2012). “A maioria dos usuários do Google tem um cookie de preferências chamado “PREF” em seus navegadores. [...] O cookie PREF pode armazenar suas preferências e outras informações, [...]” (GOOGLE, 2014e).

do usuário e a busca universal (que combina, em uma única página de resultados, conteúdos relevantes, tais como imagens, notícias, vídeos, etc. para que o usuário possa localizar mais facilmente as informações que julgar mais importantes).

Os algoritmos do Google constantemente estão sendo atualizados a partir da análise dos seus resultados. Os resultados das buscas<sup>17</sup> podem tomar diversas formas (cf. GOOGLE, 2014a): Mapa de Conhecimento; pequenas visualizações da informação; notícias; respostas imediatas e informações acerca de condições meteorológicas, resultados esportivo, etc.; vídeos; miniaturas de imagens; repostas orais se a busca deu-se desse modo; etc.

**O algoritmo *Hummingbird* e a busca semântica.** Em 2013, nas comemorações do seu 15º aniversário, o Google anunciou uma nova atualização em seu algoritmo do mecanismo de busca: o *Hummingbird*. A proposta é proporcionar maior agilidade e precisão em seus resultados de busca. O *Hummingbird*, “na prática, é uma melhora na compreensão da intenção do usuário ao realizar uma pesquisa” (MACEDO, 2013).

A inovação surge diante de duas mudanças nas buscas, conforme Macedo (2013). A primeira refere-se às pesquisas por conversação, já que, cada vez mais, intensifica-se a busca móvel e através de pesquisa por voz<sup>18</sup>. A segunda é o aumento de buscas baseadas em perguntas, tais como: “‘como fazer...?’, ‘onde está o mais próximo...?’, ‘onde eu posso conseguir...?’, etc.” Para interpretar essas frases mais longas, o Google precisa fornecer resultados diferentes para cada uma delas baseando-se em contexto. Assim, a ferramenta terá que se concentrar não só no significado das palavras, mas também no contexto para compreender a intenção de uma busca.

<sup>17</sup> Denomina-se SERP (*Search Engine Result Page*) a página de resultados de uma busca. (GABRIEL, 2012).

<sup>18</sup> Vale lembrar que, quando o usuário utiliza um dispositivo móvel, há diferença nos resultados, pois o Google leva em consideração a localização.

Segundo Singhal<sup>19</sup>, “o Hummingbird está focado mais em classificar as informações com base numa compreensão mais inteligente das solicitações de busca, [...]” (cf. OPELBAUM, 2014). Para ele, o buscador do Google pretende funcionar “da mesma forma que os humanos entendem o mundo” (cf. BUSCADOR..., 2012). Singhal afirma, ainda, em entrevista ao *Wall Street Journal*, que o Google “perseguirá um modelo de busca que torne as respostas do computador similares às que um humano daria caso ouvisse uma pergunta qualquer” (cf. AGUIARI, 2012).

O Google pretende tornar os resultados de busca mais pessoais. O que leva à chamada busca personalizada. Segundo Ivo (2013), “o Google quer dar respostas mais individuais ao que os usuários querem saber”. O foco estará na intenção do usuário e não nas palavras exatas, para que o sistema de busca fique cada vez mais adequado ao uso humano (cf. FONSECA, 2013). E, além disso, o novo algoritmo introduz a chamada “busca semântica”, além das atuais buscas por palavras-chave. Assim, o Google quer unir a tecnologia da Web Semântica com o seu atual sistema de *PageRank*. O objetivo de tais mudanças está no desenvolvimento de um sistema capaz de encontrar resultados mais relevantes e, também, melhorar o desempenho para interpretar buscas que, como sabemos, são realizadas em linguagem natural.

Na definição de Singhal, Web semântica é uma tecnologia que interliga significados de diferentes palavras e consegue atribuir um sentido aos conteúdos publicados na internet. [...] O recurso permitirá que o computador reconheça as características de cada usuário e, de certa forma, “adivinha” quais resultados aquela pessoa quer obter ao buscar determinadas palavras na Web. O cálculo é feito a partir da análise do comportamento daquele usuário na internet. [...] Singhal definiu a busca semântica como uma forma “mais simples e amigável” dos usuários fazerem buscas na Web. “Será como conversar com uma pessoa” (AGUIARI, 2012).

<sup>19</sup>Amit Singhal é vice-presidente sênior, engenheiro de *software* e chefe da equipe do Núcleo de classificação (*ranking*) do Google.

Segundo Opelbaum (2014), o *Hummingbird* otimiza o Mapa do Conhecimento do Google, que, atualmente, contém informações acerca de 570 milhões de conceitos. “Ele utiliza uma equação elaborada por Landsman: Palavras + Contexto + Gráfico do Conhecimento”.

Para que isto funcione da maneira esperada, o Google terá a necessidade de detectar a intenção da busca e, conseqüentemente, como sugerimos, ‘metarrepresentar’ a meta do usuário. Segundo Macedo (2013), há três tipos de busca de acordo com essas intenções: Navegacionais (que se caracterizam pela intenção de encontrar um site específico. Ex.: encontrar o site da Unisul); Informativas (que se caracterizam pela intenção de encontrar algo específico ou tirar uma dúvida. Ex.: encontrar dicas de emagrecimento) e Transacionais (que ocorrem quando o usuário tem a intenção de realizar alguma atividade. Ex.: encontrar um restaurante; comprar um objeto qualquer; alugar uma casa; etc.).

Diante desse quadro, na próxima seção, apresentamos aspectos acerca da Web Semântica e da personalização com a finalidade de compreensão dos motivos que levaram o Google a introduzi-las.

## 5 Web semântica e personalização

Segundo Lima e Carvalho (2004), a maior parte do conteúdo de informações disponibilizadas na Web está direcionada ao entendimento humano. Isso leva a uma grande dificuldade na execução dos serviços de recuperação de informação processados por máquinas, como é o caso do mecanismo de busca do Google. Diante desse contexto, várias iniciativas buscam “tornar as informações processáveis pelas máquinas” (LIMA; CARVALHO, 2004, p. 2).

“A Web semântica é uma extensão da Web atual, que permitirá aos computadores e humanos trabalharem em cooperação” (JUSTINO, 2014).

Seu intento está em atribuir um sentido aos conteúdos publicados na Net de modo a ser perceptível tanto ao humano quanto ao computador. Para alcançar essa finalidade, é necessário desenvolver tecnologias e linguagens que tornem a informação legível também para as máquinas. É necessário que os computadores

compreendam e vinculem dados colocados na rede com semântica, isto é, os computadores, através de ontologias e regras de inferências, encontrem respostas mais precisas para as consultas, descartando o que é irrelevante para o usuário.

O objetivo da introdução de semântica na Web é tornar a informação “compreensível” para o computador. Ela representa uma revolução no processamento da informação e, por consequência, uma revolução na maneira de se obter e organizar os conhecimentos. Pretende-se que os recursos disponibilizados sejam expressivos o bastante para que máquinas e/ou agentes de software sejam capazes de processar e “entender” o real significado dos dados (LIMA; CARVALHO, 2004, p. 3).

A internet atual é definida como a Web Sintática, na qual os computadores meramente apresentam as informações, mas cujo processo de interpretação ainda fica sob a responsabilidade de seres humanos. A Internet desenvolveu-se rapidamente como um meio de troca de documentos entre as pessoas, mas o conteúdo semântico das páginas é codificado de uma maneira somente acessível a seres humanos (cf. BREITMAN, 2005).

Para Grimaldo (2004), para se adicionar semântica à forma como os computadores veem uma página, é necessário:

1. Uma linguagem que estruture os objetos digitais sintaticamente, denominada XML (*eXtensible Markup Language*);
2. Um formato que estruture o significado da informação que possuem os objetos digitais (em conjunto com os metadados associados a ele) denominado RDF (*Resource Description Framework*);
3. Um programa de computador que recupere a informação existente, baseado em Inteligência Artificial, os denominados Agentes Inteligentes;

4. Um conjunto de regras que permitam aos Agentes Inteligentes moverem-se dentro da Web com liberdade e de acordo com o perfil informacional do usuário que o use, denominado Ontologia.

Ao afirmar que o Google tem expandido sua tecnologia, baseando-a em “localização, personalização, contexto e otimização semântica”, Simoni (2013) assim conceitua a otimização semântica:

é uma busca, uma pergunta ou uma ação que gera resultados relevantes mesmo quando os itens do resultado apresentado não contenham exatamente o termo da pesquisa. Isso significa um resultado que considera valor do conteúdo semântico, termos que tenham relação semântica (SIMONI, 2013).

Aqui, exemplificamos com uma simulação intuitiva da busca de um usuário.

- (1) Consulta: qual a idade do Guga?

Figura 2 – Resultado da consulta no Google utilizando o *Internet Explorer*

Fonte: resposta à consulta simulada na interface do Google em 25 de maio de 2015.

Verifica-se um resultado que já agrega a busca semântica. O resultado do Google expõe o recurso da busca universal, pois, por reunir diversas informações em uma só página de resultados, permite ao usuário localizar mais facilmente as informações que julgar mais relevantes. Assim, não apenas apresenta uma resposta direta à pergunta acerca da idade de Guga (perceba-se que os resultados apresentam-no como Gustavo Kuerten – termo que não fazia parte da consulta), mas também uma caixa contendo vários outros dados curiosos acerca dele, a biografia do atleta na Wikipédia (encabeçando o Top 10) e, ainda, uma seleção de outros jogadores de tênis, pois o Google “infere” que o usuário pode ter interesse pelo esporte, adiantando-se a possíveis outras buscas do usuário.

**Personalização.** Sugerimos, no decorrer do texto, que o Google está buscando personalizar os resultados de seu mecanismo de busca de tal modo a atender, de modo relevante, às necessidades e intenções de seus usuários. A personalização dos resultados visa a contextualizar a experiência do usuário.

Para Marchi (2010, p. 69), a pesquisa em

personalização de conteúdo para motores de busca tem como principal objetivo desenvolver mecanismos que forneçam coleções de páginas individualizadas, baseadas, de alguma forma, em modelos que representam as necessidades dos usuários e seu contexto de trabalho.

Segundo Pariser (2012, p. 36), o Google “monitora todo e qualquer sinal que consiga obter de nós. [...] Além disso, naturalmente, os termos que pesquisamos trazem grandes revelações sobre nossos interesses”. O Google instala, ainda, *cookies* de rastreamento em nossos computadores capazes de colher nossas informações pessoais. Em realidade, fala-se em cerca de 57 sinais, mas, estrategicamente, o Google revela poucos deles. Dentre os sinais já revelados, encontramos: histórico de pesquisa; região do usuário; sinais

sociais (que englobam o curtir, o compartilhar e o seguir em redes sociais – dentre as quais o Google+, e, ainda, interesses comuns e recomendações de amigos); o endereço de IP; o tipo de computador usado; o tipo de navegador; o idioma; etc. Ou seja, há uma coleta implícita de características e preferências de cada usuário.

Gabriel (2010) destaca que, como o Google tem comprado inúmeras plataformas, ele consegue ter acesso a muita informação. Disto têm surgido alguns sinais para a personalização. Dois nos chamaram a atenção na explanação de Gabriel: o Orkut está dando direito ao Google para usar as informações que seus usuários postam e o Gmail tem uma cláusula em seu contrato (que em geral aceitamos sem ler) que diz que ele pode ler todas as mensagens de seus usuários para oferecer anúncios adequados a eles (*links* patrocinados).

Para adaptar seus resultados (e, por consequência ‘metarrepresentar’ a meta do usuário), o Google considera a interação do usuário com o sistema, para determinar seu contexto e para conseguir seu perfil, quer explicitamente quando ele preenche formulários para um dos serviços do Google, por exemplo, ou até implicitamente quando ele passa a acompanhar o histórico de pesquisa do usuário e a própria navegação deste pela web. “Acompanhando a navegação do usuário por meio das ações do mouse, páginas visitadas, tempo de permanência nos links clicados, entre outras, é possível identificar seu interesse naquele momento e contexto” (CINI, 2002; BRUSILOVSKY, 2001; CLAYPOOL, 2000; PALAZZO, 2000 apud MARCHI, 2010, p. 71).

É fácil comprovar a personalização de resultados, basta solicitar a amigos que, utilizando seus computadores ou *mobiles*, digitem os mesmos termos na interface do buscador. Muitos se surpreenderão ao perceberem os diferentes resultados. Além disso, quem de nós não teve a desagradável experiência de, em buscas posteriores ou nas redes sociais, ser perseguido

pela publicidade de produtos que pesquisamos no Google. Tudo isso é efeito da personalização.

Diante do exposto, defendemos que é a relevância para um indivíduo que justifica o constante monitoramento das ações do usuário na Rede, a fim de personalizar os resultados do buscador, já que o Google procura conhecer o contexto e as intenções do usuário ao digitar na interface do buscador.

Sugerimos que contexto, aqui, deve ser compreendido no sentido da TR. Sperber e Wilson (2001) caracterizam sua noção de contexto, compreendendo-o como um construto psicológico que é representado por um subconjunto de suposições mentalmente representadas que são utilizadas para o processamento de uma dada informação e que ressaltam a história de vida de cada indivíduo, gerando diferenças nas informações memorizadas. Assim, representações diferentemente construídas resultam em diferentes suposições sobre o mundo, diferentes crenças, portanto. A cada nova experiência de um indivíduo, algo se acresce ao conjunto de contextos potenciais, o que pode justificar, uma vez mais, o monitoramento constante do usuário.

Quanto à relação dos resultados do buscador com os efeitos cognitivos descritos pela TR, em especial quando a intenção que move o usuário à busca é informativa, é importante destacar uma denúncia de Pariser (2011, 2012). O autor afirma que a filtragem de dados, através do constante monitoramento do comportamento do usuário na Net, está criando uma bolha que está isolando o usuário do acesso total à informação disponibilizada na Rede, a partir da desculpa de os buscadores retornarem resultados mais relevantes às consultas. Trata-se do que Pariser denomina de *Filtro Bolha*.

A partir dessa denúncia de Pariser, é possível presumir que a personalização e a contextualização dos resultados, implantadas no mecanismo de busca do Google, podem estar privilegiando, dentre os efeitos cognitivos descritos pela TR, o fortalecimento de nossas suposições, o que pode levar à

saturação dos estímulos e tornar difícil que modifiquemos nossas crenças e o modo como compreendemos o mundo, já que, através do filtro, podemos ter menores chances de acesso a opiniões que divergem das nossas. Segundo Pariser (2012), estamos sendo conduzidos a um determinismo informativo por temos bolhas informacionais sendo criadas ao nosso redor, adaptadas aos nossos interesses e preferências. Como nossa pesquisa ainda está no início, terá que ser aprofundada em relação a esse aspecto.

## Conclusão

Cada vez mais os mecanismos de busca fazem parte do nosso cotidiano. Os dispositivos móveis permitem que a busca seja realizada por qualquer pessoa em qualquer tempo e lugar. Mas, muitas vezes, não estamos cientes de que os resultados de uma busca estão sendo personalizados para cada usuário com o objetivo de alcançar a relevância na relação entre ele e o sistema (buscador).

Os buscadores procuram desenvolver seu sistema operacional tendo em vista alcançar a relevância de seus resultados de acordo com as metas de seus usuários. Assim, defendemos que eles procuram ‘metarrepresentar’ tais metas através da introdução da personalização e da busca semântica em seus algoritmos com a finalidade de interpretar a real intenção do usuário quando este formula uma consulta (*Hummingbird*).

As buscas são realizadas em linguagem natural e os dispositivos móveis têm possibilitado “buscas conversacionais” que vêm substituindo a digitação de palavras-chave soltas por perguntas estruturadas, tais como: Onde se pode comer Sushi em Tubarão? Ou Como chegar ao endereço X? etc. A busca semântica, representada pelo *Hummingbird*, persegue a capacidade de lidar com tais demandas conversacionais e entregar ao usuário exatamente aquilo de que precisa (cf. GASPARINI, 2013).



Uma vez que esses processos promovem cada vez mais uma interação que simula àquelas próprias dos seres humanos, cremos que todo esse esforço integrado de atualizações que o Google tem introduzido em seus algoritmos pode ser explicado pela TR, principalmente a partir do Princípio Cognitivo de Relevância e da Relevância para um Indivíduo.

Obviamente, estamos no início de nossas pesquisas e muito mais terá que ser aprofundado para que nossa hipótese inicial seja corroborada, mas cremos que essa pesquisa se mostra interessante, pois diariamente utilizamos ou temos a possibilidade de utilizar mecanismos de busca para alcançar informações e metas. Assim, diante de uma atividade humana na Net tão universalizada, é essencial compreender como e porque alcançamos resultados tão diversos em uma busca, apesar de digitarmos as mesmas palavras-chave na interface do buscador. Cremos que a TR é capaz de dar uma explicação válida a esse fenômeno.

## Referências

AGUIARI, Vinicius. Google implementará recursos de busca semântica este ano. *INFO on-line*. 15 mar. 2012. Disponível em: <<http://info.abril.com.br/noticias/internet/google-prepara-nova-reforma-para-as-buscas-15032012-1.shl>>. Acesso em: 03 jun. 2014.

ALGORITMO. *Infopédia*. Enciclopédia e Dicionários Porto Editora. Disponível em: <[http://www.infopedia.pt/\\$algoritmo\(informatica\);jsessionid=rxoplzQY8HQZTNVWuXwJNw\\_](http://www.infopedia.pt/$algoritmo(informatica);jsessionid=rxoplzQY8HQZTNVWuXwJNw_)>. Acesso em: 03 jun. 2014.

ALVES, Anderson Fernandes; CERQUEIRA, Otto Silva. *O design como incentivo à leitura na geração digital*. 2013. 62 f. Monografia (Graduação). Curso de Desenho Industrial, Universidade São Judas Tadeu. São Paulo, 2013. Disponível em: <[http://www.usjt.br/cursos/lacce/desenho\\_industrial/arquivos/tcc/2013/design-incentivo-leitura-geracao-digital.pdf](http://www.usjt.br/cursos/lacce/desenho_industrial/arquivos/tcc/2013/design-incentivo-leitura-geracao-digital.pdf)>. Acesso em: 10 maio 2014.

ARAÚJO, Regina Borges de. *Computação ubíqua: princípios, tecnologias e desafios*. XXI Simpósio Brasileiro de Redes de Computadores. Natal, 19 a 23 de maio de 2003. p. 45-115. Disponível em: <[http://www.professordiovani.com.br/rw/monografia\\_araujo.pdf](http://www.professordiovani.com.br/rw/monografia_araujo.pdf)>. Acesso em: 04 jun. 2014.

BREITMAN, Karin Koogan. *Web semântica: a internet do futuro*. Rio de Janeiro: LTC, 2005.

BUSCADOR do Google passará pela maior reformulação da história, diz jornal. *Olhar Digital*. 15 mar. 2012. Disponível em: <<http://olhardigital.uol.com.br/noticia/buscador-do-google-passara-pela-maior-reformulacao-da-historia,-diz-jornal/24847>>. Acesso em: 3 jun. 2014.

COM 85% de participação, o Google é o buscador mais usado no Brasil. *G1. Tecnologias e Games*. São Paulo. 16 jul. 2013. Disponível em: <<http://g1.globo.com/tecnologia/noticia/2013/07/com-85-de-participacao-google-e-o-buscador-mais-usado-no-brasil.html>>. Acesso em: 03 jun. 2014.

FERNANDES, Rogério Paulo Muller et al. *Panorama atual do uso dos mecanismos de busca na Web*. XIII Encontro Nacional de Pesquisa em Ciência da Informação – XIII ENANCIB 2012. GT 8 Informação e Tecnologia. Comunicação oral. Disponível em: <[http://www.uel.br/grupo-pesquisa/ciberespaco/doc/panor\\_atual\\_uso\\_busca\\_enancib2012.pdf](http://www.uel.br/grupo-pesquisa/ciberespaco/doc/panor_atual_uso_busca_enancib2012.pdf)>. Acesso em: 03 jun. 2014.

FIGUEIREDO, Dayana Ester Andrade. *Recuperação da informação: uma análise sobre os sistemas de busca da Web*. 2006. 64 f. Monografia (Graduação em Biblioteconomia) – Universidade de Brasília - Faculdade de Economia, Administração, Contabilidade e Ciência da Informação. Departamento de Ciência da Informação e Documentação. Brasília, 2006.

FONSECA, João da. *O que é o Hummingbird Update do Google e como revoluciona o SEO*. Seomaster. 8 out. 2013. Disponível em: <<http://www.seomaster.com.br/blog/o-que-e-o-hummingbird-update-do-google-e-como-revoluciona-o-seo>>. Acesso em: 03 jun. 2014.

GABRIEL, Martha. *Sem e Seo: dominando o marketing de busca*. 2. ed. São Paulo: Novatec Editora, 2012.

GABRIEL, Martha. *Conversando com computadores: interfaces da voz na web*. Pinhais: Editora Melo, 2011.

GABRIEL, Martha. *Oráculo digital*. Café Filosófico. Instituto CPFL/ Cultura. 23 ago. 2010. Disponível em: <<http://www.cpfcultura.com.br/wp/2010/08/23/a-era-da-busca-oraculos-digitais-%E2%80%93-martha-gabriel-com-marcelo-tas/>>. Acesso em: 19 jun. 2014.

GABRIEL, Martha. *Educ@r: a (r)evolução digital na educação*. São Paulo: Saraiva, 2013.

GASPARINI, Cláudia. Como será o mecanismo de busca do futuro? *Exame.Com*. 13 dez. 2013. Disponível em: <<http://exame.abril.com.br/rede-de-blogs/digitos/2013/12/13/como-sera-o-mecanismo-de-busca-do-futuro/>>. Acesso em: 10 jun. 2014.

GOOGLE. *Como funciona a busca*: dos algoritmos às respostas. Disponível em: <<http://www.google.com/insidesearch/howsearchworks/thestory/index.html>>. Acesso em: 03 jun. 2014a.

GOOGLE. Googlebot. In: *Ferramentas do Google para webmaster*. Disponível em: <<https://support.google.com/Webmasters/answer/182072?hl=pt-BR>>. Acesso em: 13 jun. 2014b.

GOOGLE. Rastreamento e indexação. In: *Por dentro da pesquisa*. Disponível em: <<http://www.google.com/intl/pt-BR/insidesearch/howsearchworks/crawling-indexing.html>>. Acesso em: 12 jun. 2014c.

GOOGLE. O mapa do conhecimento. In: *Por dentro da pesquisa*. Disponível em: <<http://www.google.com/intl/pt-BR/insidesearch/features/search/knowledge.html>>. Acesso em: 14 jun. 2014d.

GOOGLE. Tipos de *cookies* usados pelo Google. In: *Privacidade e termos*. Disponível em: <<http://www.google.com/intl/pt-BR/policies/technologies/types/>>. Acesso em: 17 jun. 2014e.

GOOGLE *search*. *Biografias e Coisas.com*. 18 set. 2012. Disponível em: <<http://desmanipulador.blogspot.com.br/2012/09/google-procura.html>>. Acesso em: 03 jun. 2014.

GRIMALDO, Wilmer Arturo Moyano. Sociedad de la información: metadatos y futuro de la Internet en la recuperación de información de calidad. *Bibliotecas & Tecnologías de La Información*. 2004. Disponível em: <[http://eprints.rclis.org/7010/1/Art%C3%ADculo\\_Sociedad\\_de\\_la\\_Informaci%C3%B3n\\_y\\_Metadatos.pdf](http://eprints.rclis.org/7010/1/Art%C3%ADculo_Sociedad_de_la_Informaci%C3%B3n_y_Metadatos.pdf)>. Acesso em: 19 jun. 2014.

HUMMINGBIRD. Google: o que mudou? *Internet Innovation*. Blog. 24 fev. 2014. Disponível em: <<http://www.internetinnovation.com.br/blog/hummingbird-google-o-que-mudou/>>. Acesso em: 03 jun. 2014.

IVO, Diego. *Codínome Hummingbird*: receba mais visitas com atualização do Google. 12 out. 2013. Disponível em: <<http://www.administradores.com.br/noticias/tecnologia/codínome-hummingbird-receba-mais-visitas-com-atualizacao-do-google/80859/>>. Acesso em: 03 jun. 2014.

JUSTINO, Carlos. *O Google está mudando conheça a nova busca semântica*. Pub Marketing. 18 fev. 2014. Disponível em: <<http://www.grupoipub.com.br/o-google-esta-mudando-conheca-nova-busca-semantica/>>. Acesso em: 3 jun. 2014.

LIMA, Júlio César de; CARVALHO, Cedric Luiz. *Uma visão da Web semântica*. Instituto de Informática. Universidade Federal de Goiás. 2004. Disponível em: <[http://www.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF\\_001-04.pdf](http://www.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_001-04.pdf)>. Acesso: 18 jun. 2014.

MACEDO, Gustavo. Google Hummingbird: o impacto do novo buscador para seu negócio. 01 nov. 2013. *Exame.com*. Blog Performance. Disponível em: <<http://exame.abril.com.br/rede-de-blogs/performance-digital/2013/11/01/google-hummingbird-o-impacto-do-novo-buscador-para-seu-negocio/>>. Acesso em: 03 jun. 2014.

MARCHI, Késsia Rita da Costa. *Uma abordagem para personalização de resultados de busca na web*. 2010. 141 f. Dissertação (Pós-Graduação em Ciência da Computação) - Universidade Estadual de Maringá, Maringá/PR, 2010.

MORAIS, Carlos Tadeu Queiroz de; LIMA, José Valdeni de; FRANCO, Sérgio Roberto K. *Conceitos sobre internet e web*. Porto Alegre: Editora da UFRGS, 2012. Disponível em: <<http://www.ufrgs.br/sead/publicacoes/documentos/livro-conceito-internet-e-web>>. Acesso em: 20 maio 2015.

OPELBAUM, Polina. *Os efeitos do Google Hummingbird nas buscas e no social*. Blog PR Newswire – Brasil. 19 mar. 2014. Disponível em: <<http://blogprnewswire.com/2014/03/19/>>. Acesso em: 3 jun. 2014.

PARISER, Eli. *O filtro invisível*: o que a internet está escondendo de você. Tradução Diego Alfaro. Rio de Janeiro: Jorge Zahar, 2012. Disponível em: <<http://books.google.com.br/books?id=R9zb5nAfw8C&printsec=frontcover&dq=O+filtro+invis%C3%ADvel&hl=ptBR&sa=X&ei=A36kU6xc57KxBLekfgK&ved=0CCYQ6AEwAA#v=onepage&q=O%20filtro%20invis%C3%ADvel&f=false>>. Acesso em: 19 jun. 2014.

PARISER, Eli. *Tenha cuidado com os “filtros-bolha” online*. TED. palestra. mar. 2011. Disponível em: <[http://www.ted.com/talks/eli\\_pariser\\_beware\\_online\\_filter\\_bubbles?language=pt-br](http://www.ted.com/talks/eli_pariser_beware_online_filter_bubbles?language=pt-br)>. Acesso em: 20 abr. 2015.

PEW INTERNET. *Search engine use 2012*. 9 mar. 2012. Disponível em: <[http://www.pewinternet.org/files/old-media/Files/Reports/2012/PIP\\_Search\\_Engine\\_Use\\_2012.pdf](http://www.pewinternet.org/files/old-media/Files/Reports/2012/PIP_Search_Engine_Use_2012.pdf)>. Acesso em: 03 jun. 2014.

RAUEN, Fábio José. For a goal conciliation theory: ante-factual abductive hypotheses and proactive modelling. *Linguagem em Discurso*, LemD, Tubarão, SC, v. 14, n. 3, p. 595-615, set./dez. 2014.

SIMONI, Raphael. *SEO e semântica web*: uma nova revolução com a busca de entidades. iMaster. 08 ago. 2013. Disponível em: <<http://imasters.com.br/search/seo/seo-e-semantica-Web-uma-nova-revolucao-com-a-busca-de-entidades/>>. Acesso em: 03 jun. 2014.

SPERBER, Dan et al. Epistemic vigilance. *Mind & Language*. 2010. p. 359–393. Disponível em: <<http://www.dan.sperber.fr/wp-content/uploads/Epistemic-Vigilance-published.pdf>>. Acesso em: 10 fev. 2014.

SPERBER, Dan; WILSON, Deirdre. *Relevância: comunicação e cognição*. Tradução de Helen Santos Alves. Lisboa: Fundação Calouste Gulbekian, 2001.

SPERBER, Dan; WILSON, Deirdre. *Relevance: communication & cognition*. 2. ed. Oxford: Blackwell, 1995. (1st ed. 1986).

WEBMASTER. *Infopédia*. Enciclopédia e Dicionários Porto Editora. Disponível em: <<http://www.infopedia.pt/lingua-portuguesa/Webmaster>>. Acesso em: 13 jun. 2014.

YUS, Francisco. *Putting relevance at centre stage in all research on human activity on the Internet*. Artigo entregue para Relevance: Discourse and Translation. Warsaw/Polônia, set. 2012.

YUS, Francisco. *Una aproximación discursiva a las identidades en línea*. In: Language and Identity in the Spanish-Speaking World (6th International Conference of Hispanic Linguistics and 4th Biennial Meeting of the International Association for the Study of Spanish in Society). 2013. Birkbeck: University of London. Disponível em: <<https://sites.google.com/site/franciscoyus/conferences>>. Acesso em: 20 mar. 2014.

YUS, Francisco. *Social media*. In: The Routledge Handbook of Pragmatics. Eds. A. Barron, G. Steen and G. Yueguo. London: Routledge, 2014. Disponível em: <<https://drive.google.com/g/?a=v&pid=sites&srcid=ZGVmYXVsdGRvbWFpbXmcmFuY2lzY295dXN8Z3g6NjgxNzE4NjIjNDhlYWY1OA>>. Acesso em: 15 set. 2014.

YUS, Francisco. *En qué difieren los géneros electrónicos de los analógicos*. En el Congreso Leer y Escribir en Español en la Red. Aprendizaje, Renovación y Propuestas. University of Comillas, Cantabria (Espanha), 2011. Disponível em: <<https://sites.google.com/site/franciscoyus/conferences/texto%20para%20actas%20Yus.doc?attredirects=0&d=1>>. Acesso em: 20 mar. 2014.

Recebido em 20 de dezembro de 2014.

Aceito em 20 de maio de 2015.