

## **WEB SEMÂNTICA: BOLHAS INDIVIDUAIS OU UM FACILITADOR DE BUSCAS?**

**GODOY, Valdir Alves de**

*Doutor em Administração Educativa,  
Gestor Educacional, Professor UNIESP*

**MACHADO, Marcos**

*Especialista em Tecnologia e Sistemas de Informação,  
Coordenador de curso, Professor UNIESP*

**LOPES, J. L. S.**

*Graduado em Sistemas de Informação*

**LOPES, Thiago J.**

*Graduado em Sistemas de Informação*

### **RESUMO**

O objetivo deste artigo consiste em um estudo sobre as mudanças que estão ocorrendo na internet, onde se evidencia uma das maiores ameaças a democratização da informação, o filtro-bolha. Nessa perspectiva, o trabalho visa proporcionar uma reflexão crítica ao leitor fazendo referência aos motores de pesquisa da chamada web semântica ou web 3.0 e suas possíveis consequências. Para balizar esse estudo, tomamos por base os trabalhos de Benkler (2006), Berners-Lee (2010), Breitman (2005), Garrigós, Gomes e Houben (2010), Goldman (2008), Introna e Nissenbaum (2000), Rawls (1971), Sunstein (2011) e Pariser (2011). A relevância do trabalho consiste no fato de se demonstrar que a web semântica, apesar de ser uma solução para os problemas relativos à sobrecarga de informações, agrega consigo, aspectos conflitantes no que tange aos direitos de liberdade e autonomia, além de uma possível transformação da web em meras bolhas individuais.

**Palavras-chave:** Filtro-Bolha, Algoritmos de Personalização, Web Semântica, Facilitador de Buscas.

### ***Abstract***

*The goal of this article consists of a study on the changes that are occurring on the internet, where it highlights one of the greatest threats to the democratization of information, the filter bubble. From this perspective, the work aims to provide to the reader a critical reflection by referring search engines from the so-called semantic web or web 3.0 and its possible consequences. To delimit this study, we mapped out on the basis of the work of Benkler (2006), Berners-Lee (2010), Breitman (2005), Garrigós, Gomes and Houben (2010), Goldman (2008), and Introna Nissenbaum (2000), Rawls (1971), Sunstein (2011) and Pariser (2011). The relevance of the work consists in the fact that it is shown that the semantic web, despite being a solution to the problems relating to information overload, aggregates with that, conflicting aspects with regard to the rights of freedom and autonomy, as well as a possible transformation of the web in mere individual bubbles.*

**Key-words:** Filter bubble, Personalization algorithms, Semantic Web, Search facilitator.

## INTRODUÇÃO

A Internet foi concebida inicialmente com o intuito de ser uma fonte de informações diversificada, plural, aberta e que fosse livre no aspecto de pontos de vistas diferentes e conflitantes.

Nesse sentido, a internet tem possibilitado a formação de novas formas de interação, organização e atividades sociais, demonstrando dessa forma, sua função socializadora, no sentido de aproximar pessoas das mais diferentes culturas, comandadas pelos mais diferentes tipos de governos, independente de sua localização física.

Assim, devido à facilidade na publicação de documentos na internet, uma enorme distribuição de informações está disponível na internet, trazendo ao seu utilizador enormes benefícios. Em contrapartida, uma imensa quantidade de dados descartáveis está disponibilizada na rede, ocasionando um verdadeiro caos de informações para quem procura algo na internet.

A web semântica<sup>1</sup> vem como uma resposta a esse problema da grande diversidade de informações disponíveis na web. Podemos dizer que, atualmente, temos uma web de documentos, ao invés de uma web de informações, de modo que, as ferramentas de busca enfrentam dificuldades para executar pesquisas que não estão diferenciadas em termos de assunto, qualidade e relevância.

Nessa perspectiva, a solução que surge é a web semântica, que agrega significado as páginas da web, oferecendo às pessoas a capacidade de criarem repositórios de dados na web, construir vocabulários e escreverem regras para interoperar com esses dados, ou seja, trata-se de uma web que possui toda sua informação organizada, permitindo uma melhor interação ao usuário.

A web semântica propõe ajudar os navegadores a "ler", "utilizar" e "entender" o conteúdo contido nas páginas navegadas. A ideia central é a de que metadados<sup>2</sup> adicionados às páginas web possam torná-las legíveis para computadores.

---

<sup>1</sup> Semântica, derivado de sema (sinal), refere-se ao estudo do significado em todos os sentidos do termo.

<sup>2</sup> Metadados ou Metainformação são dados sobre outros dados. Um item de um metadado pode dizer do que se trata aquele dado, geralmente uma informação inteligível por um computador. Os metadados facilitam o entendimento dos relacionamentos e a utilidade das informações dos dados.

Vale destacar que, essa ação não irá conceber inteligência artificial aos computadores, ou torná-los autoconscientes, mas dará às máquinas ferramentas para encontrar, trocar e, dentro de certo limite, interpretar as informações.

Dessa forma, não precisamos mais minerar a internet em busca daquilo que procuramos, sendo essa uma solução viável, no objetivo de filtrar esse enorme caos que a web se tornou.

Um bom exemplo de aplicação com conceitos de web semântica já são praticados pela empresa Google em seu site de buscas. Ao digitarmos a palavra “define” seguida da palavra desejada, ao invés de recebermos uma infinidade de páginas com resultados genéricos, são exibidas somente as páginas que definem o que o usuário escolheu.

A web semântica “é uma extensão da web atual” e não uma substituição da mesma, “na qual é dada a informação um significado bem definido, permitindo que computadores e pessoas trabalhem em cooperação” (BREITMAN, 2005).

O conceito é abstrato, mas, enquanto alguns sites já estão utilizando conceitos da web semântica, várias das ferramentas necessárias, ainda estão em desenvolvimento em relação a suas funções, bem como, na sua praticidade.

Vale ressaltar que, o implemento dessa tecnologia é recente e, segundo especialistas da área de TI, levará alguns anos ainda para ser implementada em sua totalidade.

Por outro lado, destacamos que essa solução, ou essa nova forma de se ver a web, não apresenta somente benefícios, pois esta tecnologia traz consigo *embutida* a possibilidade de personalização das páginas que são oferecidas aos usuários em suas consultas, surgindo assim, um novo paradigma no que tange a democratização da informação, o *Filtro-Bolha*<sup>3</sup>.

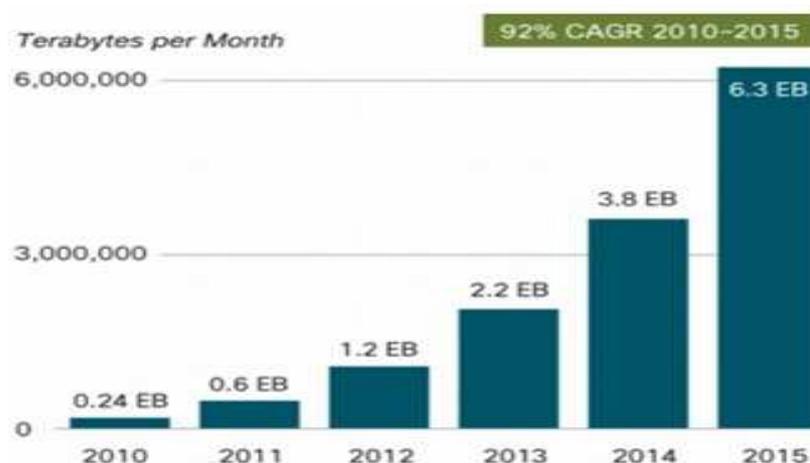
### **Algoritmos de Personalização: uma análise crítica dessa nova tecnologia**

Serviços de nuvem utilizados pela maioria dos usuários de Internet, como o Google e o Facebook, tão presentes em nosso dia-a-dia, lidam com uma imensa quantidade de informações geradas por seus usuários.

---

<sup>3</sup> Filtro-bolha: Termo utilizado por Eli Pariser, autor do livro “The Bubble Filtre: what the internet is hiding from you” (Filtro-Bolha: O que a internet está escondendo de você), alerta sobre o perigo existente dos serviços sob medida que selecionam e restringem as informações, criando bolhas de informações, cujo centro são os nossos supostos interesses.

Tais serviços de nuvem iniciaram a utilização de algoritmos de personalização, tendo como objetivo lidar com o crescente aumento da quantidade de informações disponíveis *online* e a chamada sobrecarga de dados, demonstrada graficamente em estudos recentes realizados pela CISCO<sup>4</sup> (uma projeção), onde apontam que os dados gerados na Internet em 2015, serão quatro vezes maiores se comparados com os dados gerados em 2010, conforme figura abaixo:



**Figura 1:** Projeção dos dados gerados na internet entre 2010 e 2015

**Fonte:** CISCO

Como subterfúgio de tratar tal volume de dados, nesta técnica de personalização, ou como é tratada atualmente, filtragem dos dados, todas as ações dos usuários são gravadas e identificadas como uma identidade única. Assim, as informações são apresentadas levando-se em conta esta identidade.

O problema se evidencia na operação de tais filtros que ignoram o contexto e a neutralidade das informações e funcionam sem o controle e conhecimento dos usuários. Como resultado de tal processo, existe um tema pouco abordado pelos meios massivos de comunicação e pela mídia, que são os filtros-bolhas, onde Eli Pariser, ativista e coordenador da organização não governamental americana de cunho político MoveOn.org, define como “Um universo único de informações para cada um de nós” (PARISER, 2011).

Segundo Garrigós, Gomes e Houben (2010, p. 991), “a personalização da Internet se define como o processo de mudança do conteúdo e estrutura de uma aplicação web, com

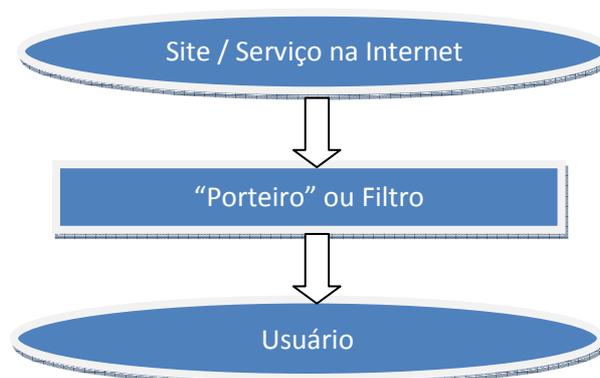
<sup>4</sup> CISCO - uma das maiores corporações da área de telecomunicações do mundo e uma das maiores em atuação nos elementos que fazem a Internet operar.

o objetivo de se adaptar a uma necessidade específica, objetivos, interesses e preferências de cada usuário”, onde nesses sistemas são identificados como uma identidade única.

A corporação Google, por exemplo, utiliza desde 2009, um mecanismo chamado de "57 sinais". Trata-se de um conjunto de características que envolvem desde o local que o usuário está conectado, identificação da placa de rede, tipo de navegador utilizado, no intuito de filtrar os resultados de busca do usuário.

Nesse sentido, podemos afirmar que não existe mais um Google padrão. Enquanto a maioria dos usuários acredita que suas buscas são ranqueadas pelo algoritmo do Google chamado de *Page Rank*, que deveria mostrar em ordem crescente as buscas realizadas pelos termos mais utilizados, na realidade, são mostrados apenas o que o mecanismo citado anteriormente sugere para um usuário em particular.

Baseado na afirmação de Goldman (2008) em sua publicação “*The Social Epistemology of Blogging*” (A Epistemologia Social de Blogging), os algoritmos de filtragem fazem o papel do que seria um “porteiro”, reduzindo o volume de informações que atingem os usuários. Dependendo de certos critérios, as informações são personalizadas utilizando esse mecanismo como sendo uma identidade única, conforme apresentado na figura abaixo:



**Figura 2:** Processo de filtragem

**Fonte:** dos autores

Assim, as informações são filtradas antes mesmo de se chegar ao usuário, sem que ele perceba o que está acontecendo por detrás do que está acessando. Uma vez que as informações são filtradas pelo sistema, assuntos de maior relevância para esse usuário

estão personalizados de acordo com seu perfil de interesse, ou seja, o usuário passa a ter sua própria *Bolha*.

Pariser (2011) observa o seguinte, fazendo menção aos filtros:

“Tome notícias sobre a guerra no Afeganistão. Quando você conversa com pessoas que dirigem sites de notícias, eles vão dizer que matérias sobre a guerra no Afeganistão não funcionam muito bem. Elas não recebem muitos cliques. As pessoas não se aglomeram nelas. E ainda, este é sem dúvida uma das questões mais importantes que o país enfrenta. Mas, nunca vai passar pelos filtros”.

Vale destacar, ainda segundo o autor, que a maneira como a informação é transmitida no Facebook é com o botão “like” (gosto), onde seu funcionamento se dá de uma forma muito particular: “É fácil clicar 'like' em 'Eu corri uma maratona' ou 'eu fiz um bolo realmente incrível'. É muito difícil de clicar 'like' em 'Guerra no Afeganistão entra no seu 10º ano’”.

No livro “*A Theory of Justice*” (Uma Teoria da Justiça), escrito por Rawls (1971), é introduzido o conceito chamado de bens primários (*primary goods*), onde todo homem racional presume o que deseja. O autor assume que os mais importantes bens primários na sociedade são: direito, oportunidade, liberdade, renda e riqueza.

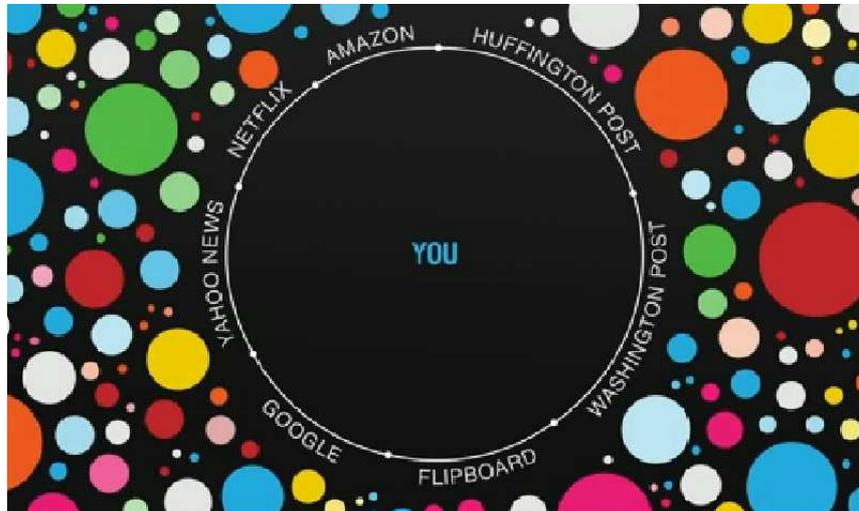
Atualmente todas as pessoas pesquisam informações sobre um assunto para tomar decisões importantes em suas vidas. As informações disponibilizadas na internet são utilizadas para que as pessoas possam planejar suas vidas e participar adequadamente em sociedade. Nessa perspectiva, além do acesso as informações pela internet, os filtros afetam o valor da liberdade que temos de acessar informações e transformá-las em conhecimento.

Desse modo, concordar com os filtros poderia afetar o valor moral da informação e as facilidades de um indivíduo obtê-la, ao contrário da internet de alguns anos atrás, em que o problema de acesso era devido ao alto custo de um equipamento de acesso.

Vale ressaltar, que sistemas tão complexos quanto o mecanismo de pesquisa da corporação Google, chegaram a um nível tal de complexidade que nem mesmo seus desenvolvedores explicam completamente qualquer que seja o resultado de uma busca efetuada por tal mecanismo (PARISER, 2011).

Torna-se *mister* destacar que, ao se alterar um algoritmo de forma invisível (uma vez que, não somos informados sobre tais mudanças), esses algoritmos personalizados criam para seus usuários o que Pariser (2011) chama de “bolha” em seu universo de

informações pessoal, tornando-se um perfil único, isolando-os da rede, conforme ilustrado na figura 3:



**Figura 3:** Ilustração da Bolha na Internet  
**Fonte:** TED Ideas worth spreading

Bernes-Lee (2010), em um artigo recente alerta que “corporações como Facebook, LinkedIn e Frinedster ameaçam a universalidade da web” e argumenta que, tais sites, criam “silos fechados de informações”.

Em artigo escrito para a revista *Scientific American*, o autor argumenta que a liberdade na internet está em risco e deve ser defendida. Caso isso não aconteça “perderemos a liberdade de se conectar com qualquer que seja o site que desejamos” (BERNERS-LEE, 2011).

O filtro bolha é um fenômeno que está próximo de se tornar o que Sunstein (2007) definiu como *Echo Chamber* (câmara vazia), em sua publicação chamada *Republic.com 2.0*. O autor alerta sobre os cidadãos que se utilizam de ferramentas tecnológicas para customizar as fontes de informação, terminando em algo que ele chama de *echo chamber* ou *information cocoons* - casulo de informação. Mas, existe uma grande diferença, o filtro bolha ocorre sem a autonomia do usuário.

Nesse sentido, notamos que o valor da autonomia entra potencialmente em conflito com o elemento definidor da computação em nuvem, onde o controle das informações é retirado dos usuários e passados para um terceiro, que seriam os prestadores de serviços de computação em nuvem. Nessa situação, serviços como a personalização, podem ser adicionados sem a notificação devida ao usuário.

De acordo com Intronna e Nissenbaum (2000), “o usuário tem o direito de demandar a informação completa e verdadeira sendo bem divulgadas as regras subjacentes a algoritmos de indexação, de procura e priorização de informações, estabelecidas num sentido significativo para os usuários”.

Pariser (2011) afirma que, “para que os usuários tenham controle dos serviços que estão usando, deve-se saber qual informação está sendo utilizada na personalização, bem como, o modo com que os bancos de dados estão sendo usados”.

Ter muita informação sem poder separar o que nos interessa é o que Benkler (2006), em seu livro “*Wealth Of Networks*” chama de “*Babel Objection*” (Objeção de Babel): “Indivíduos têm de ter acesso a algum mecanismo que possa peneirar o universo da informação, conhecimento, e movimentos culturais a fim detalhá-los em âmbito administrável e utilizável<sup>5</sup>”.

É uma questão que renasce quando serviços atualmente ativos da internet não são controlados completamente pelo usuário. Na verdade, nem conhecimentos sobre tais filtros ou como funcionam esses filtros, são informados aos usuários, não existindo dessa forma, a possibilidade de liberdade de escolha e nem decisão sobre o que queremos ou não que seja apresentado.

Não existe uma divulgação completa de como as linhas de códigos do algoritmo funcionam, que nos leve a uma total transparência para que tenhamos uma melhor experiência quando pesquisamos algo. Não se trata apenas da possibilidade de abusos como *spams* e segredos corporativos, como também se tornam muito difícil mesmo para um usuário avançado compreender o algoritmo.

Com essas implicações, se for demonstrado o algoritmo para o usuário, o ideal seria que quando os filtros personalizados realmente fossem a nosso favor, o usuário pudesse ser notificado das atividades do filtro e ainda como funciona basicamente o sistema de filtragem. Dessa forma, ele saberia que poderia perder algumas informações que pudessem ser relevantes nesse tipo de procura personalizada.

Na personalização, através do rastreamento *online* associado ao perfil de um usuário, é criada uma representação dos traços da identidade do usuário. A personalização é baseada na interpretação da identidade.

---

<sup>5</sup> “Individuals must have access to some mechanism that sifts through the universe of information, knowledge, and cultural moves in order to whittle them down into manageable and usable scope (BENKLER, 2006)”.

Segundo Nissenbaum (2004), a “identidade se refere às pessoas entenderem quem elas são através dos tempos, abrangendo a continuidade e a descontinuidade<sup>6</sup>”. Em seu relato sobre privacidade dentro de um contexto de integridade, o autor argumenta que esse tipo de necessidade de privacidade depende de um contexto particular pessoal no sentido do fluxo das informações. Cada contexto individual possui diferentes expectativas de qual tipo de informação é apropriada e inapropriada, e como a informação deve ser distribuída. Quando as informações normativas são violadas, a privacidade de um indivíduo é infringida. A “privacidade envolve a habilidade de uma pessoa de controlar o fluxo de seus dados pessoais a serem distribuídos para cada contexto em particular”.

A ideia de que cada indivíduo tem diferentes expectativas de qual contexto quer compartilhar, pode justificar a necessidade de filtros para compartilhar apenas informações que deseja. O indivíduo tem a expectativa sobre qual informação que detém é apropriada a receber num determinado contexto. Analogamente, num contexto social, estar cheio de amigos e sustentar um relacionamento pode ser mais importante que a realização de ambições profissionais.

Quando são levadas em conta expectativas contextuais, a autonomia não é apenas depender da filtragem, mas ser mais específico na filtragem quando se aborda requerimentos num determinado contexto. Esses requerimentos estão relacionados aos traços da identidade de um indivíduo materializada num perfil usando algoritmos personalizados para filtragem.

No momento, filtros personalizados estão sendo usados na maior parte dos serviços na nuvem e, frequentemente, não tomam o contexto de perfil de um indivíduo. No Facebook, por exemplo, se não tivermos interesse no álbum de fotos de um contato, o sistema irá assumir que não temos interesse naquele contato. Porém, poderíamos estar apenas interessado nas atualizações de status sobre links de trabalho.

De um filtro para todos os princípios de interações podem ser omitidos quando a descontinuidade da identidade em diferentes contextos é levada em conta. Quando diferentes filtros personalizados forem desenvolvidos com diferentes configurações, o conflito específico desse requerimento no contexto, não estará mais atrapalhando.

---

<sup>6</sup> “Identity refers to people’s understanding of who they are over time, embracing both continuity and discontinuity (NISSENBAUM, 2004)”.

Os filtros entraram em nossas vidas do mesmo jeito que a internet: de uma forma explosiva se espalhando em todos os cantos. Porém, ao contrário da web, os filtros não foram introduzidos aos usuários, mas sim, impostos. Nos moldes atuais a internet está caminhando no sentido de nos mostrar o que ela "pensa" que queremos ver, mas não o que realmente necessitamos ou temos o direito de ver.

## CONSIDERAÇÕES FINAIS

Questões relacionadas ao filtro-bolha nos levam a um grande paradoxo: A web semântica surgiu da necessidade de se filtrar o "lixo" existente na web. Por outro lado, esses filtros da maneira que estão sendo personalizados, ferem os direitos de liberdade e autonomia dos usuários de internet, uma vez que, não decidimos o que entra e sai dessa bolha, e o mais importante: não conseguimos enxergar o que fica fora dela.

Nessa perspectiva, o grande desafio com esses filtros personalizados se deve ao fato desses filtros olharem principalmente o que clicamos primeiro podendo comprometer o equilíbrio da atual sociedade de informação e conhecimento, pois são baseados somente em relação ao seu grau de relevância.

A internet teve grande impacto na história recente da humanidade, pois possibilitou a todos uma fonte de informação plural e aberta, livre de "filtros humanos", entendendo-se aqui a expressão filtros humanos, representados pelos editores de grandes jornais e noticiários.

Surge aqui o grande paradigma. No passado, esses "filtros humanos" decidiam aquilo que poderíamos ou não ver ou ser informados, e aparentemente, com a web semântica, esses filtros foram substituídos por algoritmos de personalização, ou seja, os "porteiros" que determinam aquilo que eles julgam ser relevantes para nós.

Corroborando com Pariser (2011) enxergamos isso como um grande problema, uma vez que, esses filtros não possuem a ética arraigada neles, pois se tratam de robôs pré-programados, de modo que, precisamos "nos certificar de que esses algoritmos tenham codificado neles, um senso de vida pública, um senso de responsabilidade cívica" conforme destacou o autor no Congresso Anual da TEDTalk.

Para finalizar, destacamos que a web deve ser vista como uma fonte de inspiração e troca de conhecimento sob diferentes perspectivas e isso não será possível se nos deixar isolado em uma rede de um.

**REFERÊNCIAS BIBLIOGRÁFICAS**

BENKLER, Yochai. **The Wealth of Networks: How Social Production Transforms Markets and Freedom** (Yale Press), 2006.

BERNERS-LEE, Tim. **Information Management A Proposal**. Disponível em: <<http://www.w3.org/History/1989/proposal.html>>. Acesso em 15 mai. 2011.

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

CISCO Systems. **Cisco Visual Networking Index: Forecast and Methodology, 2010–2015**, whitepaper, 2011.

GARRIGÓS, Irene, GOMEZ, Jaime and HOUBEN, Geert-Jan. **Specification of personalization in web application design**. Information and Software Technology, 991-1010, 2010.

GOLDMAN, Alvin. **The Social Epistemology of Blogging**. Information Technology and Moral Philosophy. Cambridge University Press, 2008.

INTRONA, Lucas, D. and NISSENBAUM, Helen. **Shaping the Web: Why the Politics of Search Engines Matters**. The Information Society, 16, 169-185, 2000.

NISSENBAUM, Helen. **Privacy as Contextual Integrity**. Washington Law Review Vol. 79, No. 1, February 2004: 119-158.

RAWLS, John. **A Theory of Justice**, Cambridge Mass., Belknap Press of Harvard University Press, 1971.

SUNSTEIN, Cass. **Republic.com 2.0**, Princeton University Press, 2007.

PARISER, Eli. **The Filter Bubble: What the Internet Is Hiding from You**. Penguin Group USA, 2011.