

A DEPENDÊNCIA DO RASTREAMENTO COMPORTAMENTAL ONLINE PARA A ECONOMIA GLOBALIZADA

RESUMO

A crescente quantidade de dados pessoais, cotidianamente, produzidos na Internet possibilitou a criação de uma infraestrutura invisível formada por empresas de análise de audiência, agência de marketing especializada em mídia eletrônica, publicidade online direcionada, entre outras atividades de análise de dados baseadas no comportamento online de usuários. São empresas que coletam, analisam e interpretam dados de acessos em várias fontes, sobretudo de sites "terceiros", e representam hoje uma grande economia mundial a partir de dados pessoais. Há uma grande variedade de tecnologias para o rastreamento do comportamento online dos usuários. Contudo, este trabalho investigou a tecnologia de rastreamento baseada em web cookies inseridos na navegação dos dez sites de notícias mais acessados por usuários de Internet brasileiros em abril de 2015 e maio de 2016. Este cenário específico é apresentado e consolidado por meio de gráficos e tabelas que apontam as empresas, suas origens e as características dos web cookies analisados.

Palavras chave: rastreamento de comportamento, cookies, vigilância, web tracking.

La dependencia del rastreamiento comportamental online para la economía globalizada

RESUMEN

La cantidad cada vez mayor de datos personales producidos todos los días en la internet nos hizo posible la creación de una infraestructura invisible hecha por empresas de análisis de audiencia, agencia de marketing especializada en mídia eletrônica, publicidad en línea direcionada, además de actividades de análisis de datos baseadas en el comportamiento en línea de usuarios. Son empresas que recogen, analizan e interpretan datos de accesos en varias fuentes, especialmente en sitios "terceros", y representan hoy una grande economía mundial a partir de datos personales. Hay una grande variedad de tecnologías para rastreo del comportamiento en línea de usuarios. Sin embargo, este trabajo investigo la tecnología de rastreo basada en web cookies insertados en la navegación de diez sitios de noticias con más visitados por usuarios brasileños en abril de 2015 y mayo de 2016. Este escenario específico es presentado e construido a través de gráficos y tabelas que señalan las empresas, sus fuentes y características de los web cookies analizados.

Palabras clave: rastreamiento comportamental, cookies, vigilancia, seguimiento web.

The dependence of online behavioral tracking for the global economy.

ABSTRACT

The increasing amount of personal data, daily, produced on the Internet has enabled the creation of an invisible infrastructure made up of audience analysis companies, marketing agency specialized in electronic media, targeted online advertising, among other data analysis activities based on online users' behavior. These are companies that collect, analyze and interpret access data from various sources, especially "third party websites", and today represent a major world economy from personal data. There is a great variety of technologies for tracking the online behavior of users. However, this work investigated the tracking technology based on web cookies inserted in the navigation of the ten most accessed news websites by Brazilian Internet users in April 2015 and May 2016. This specific scenario is

presented and built by means of graphs and tables that link companies, their origins and features of the web cookies analyzed.

key words: behavior tracking, web cookies, Surveillance, web tracking.

RODOLFO DA SILVA AVELINO*

SÉRGIO AMADEU DA SILVEIRA**

INTRODUÇÃO

Em sua primeira fase comercial, em meados de 1995, a Internet era constituída por sites estáticos e ferramentas pouco interativas. Nessa fase, as ferramentas de monitoramento e rastreamento disponíveis nos serviços da Internet eram restritas apenas a registros de metadados, como origem e destino dos acessos realizados. Segundo Castells (2003, p.139), “A privacidade era protegida pelo anonimato da comunicação na Internet e pela dificuldade de investigar as origens e identificar o conteúdo de mensagens transmitidas com o uso de protocolos da Internet”. Entretanto, as possibilidades de identificação de quem navegava na rede era muito grande. A falta de transparência técnica em uma simples navegação por um site dava ao usuário a sensação de segurança de acesso. Giddens, definiu esta sensação de confiança e segurança em sistemas e na tecnologia como a confiança em sistemas abstratos que seria “a condição do distanciamento tempo-espaco e das grandes áreas de segurança na vida cotidiana que as instituições modernas oferecem em comparação com o mundo tradicional (1991,p.102).”

Já em sua segunda evolução pós era comercial, diretamente relacionada com o avanço das linguagens de programação, da capacidade de processamento e armazenamento dos computadores pessoais, aliados ao aumento das taxas de transmissão de dados nas conexões de dados, permitiram uma mudança na experiência de navegação dos usuários. Simultaneamente a este avanço, a digitalização de bens culturais, sobretudo a forma de distribuição e compartilhamento destes, possibilitou o surgimento de uma nova cultura de consumo, com ênfase na colaboração em rede e na troca de conteúdos

entre os internautas.

Esta evolução também possibilitou o avanço das ferramentas de rastreamento e vigilância, que antes eram limitadas a informações de metadados como endereço IP de origem e destino, tipo de sistema operacional e navegador web, bem como as páginas acessadas. Essa base de dados era pouco eficiente na personalização dos acessos e experiência do usuário. As linguagens de programação para web, o surgimento de blogs e plataformas de redes sociais abriram caminho para a evolução das ferramentas de rastreamento e vigilância, para a consolidação de mecanismos de personalização dos dados coletados, sobretudo, para buscar entender a experiência de navegação e os interesses de acesso do usuário, possibilitando assim o aprimoramento das técnicas de controle de navegação.

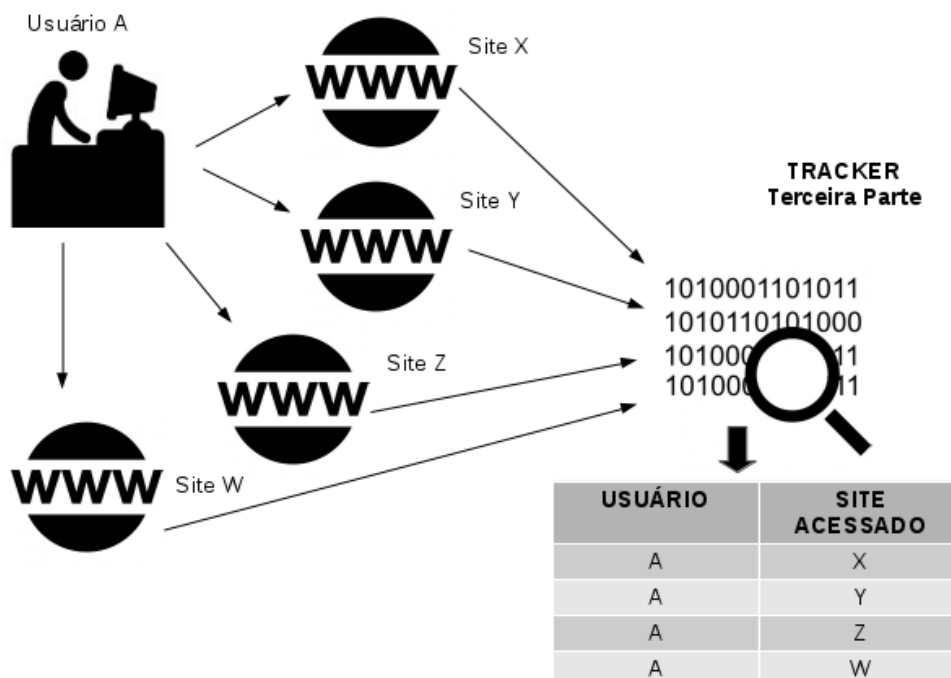
Este cenário corroborou para que a Internet se tornasse um campo fértil para a exploração do capital. Sendo assim, nela consolidou-se um ambiente de negócios cada vez mais competitivo com agressivas estratégias de marketing cujo principal objetivo era aperfeiçoar a compreensão corporativa das necessidades dos clientes. Colocar o cliente como um alvo permanentemente observado impulsionou o desenvolvimento pelas corporações de ferramentas sofisticadas para saber cada vez mais sobre os consumidores. Neste ambiente de controle, para Bauman e Lyon, “A vigilância suaviza-se especialmente no reino do consumo. Velhas amarras se afrouxam à medida que fragmentos de dados pessoais obtidos para um objetivo são facilmente usados para outro fim” (2013,p.10).

Nesse cenário surgem empresas para desenvolver uma gama de serviços e produtos específicos para o trabalho de mineração de dados e Marketing Digital, para um mercado constituído por firmas de todos portes. Nele, a matéria prima é a privacidade dos usuários. Por meio de complexos algoritmos de rastreamento e vigilância, as experiências produzidas nas redes permitem as companhias construir grandes bancos de dados de informações e preferências de usuários. O processamento e análise desses dados servem para modelar e controlar a experiência de navegação do usuário, sem o seu consentimento prévio, sobretudo, para o direcionamento e personalização de propagandas. Por ser mais discreto e por se favorecer pela falta de legislação, este modo de atividade é mais invasivo e pouco se sabe como ele funciona.

Estes rastros digitais gerados pelos usuários permitem a sua modulação e controle por parte das grandes corporações e estados. Para Assange e Andy (2013,p.75), “Eles não chamam as pessoas de “assinante”, “usuários” ou qualquer outro termo do gênero; eles as chamam de “alvos”, e aí você pode dizer: “Tudo bem, trata-se de um jargão de marketing””.

Todos os rastros gerados pelos usuários fazem parte do padrão de funcionamento da Internet, onde, geralmente, se faz um uso intenso de protocolos de rede. Kurose (2010,p.6), define um protocolo como, “ (...) o formato e a ordem das mensagens trocadas entre duas ou mais entidades comunicantes, bem como as ações realizadas na transmissão e/ou no recebimento de uma mensagem ou outro evento.” Desde sua concepção a Internet foi criada para rastrear e monitorar as interações realizadas por estas entidades comunicantes. Para Galloway (2004), esta arquitetura de rede favorece o controle, não a liberdade, e esta característica reside nos protocolos técnicos que fazem destas conexões.

Figura 1. Possibilidades de rastreamento por empresas terceiras (terceira parte) dos sites acessados por um usuário.



A EVOLUÇÃO NA CRIAÇÃO DOS BANCOS DE DADOS COMPORTAMENTAIS

A Internet é considerada uma das maiores plataformas de empreendimentos e inovação que impulsionam a economia global. Empresas como Google, Apple e Facebook estão entre as empresas mais valiosas do mundo. O modelo de negócio dessas grandes companhias está centrado na coleta, processamento e análise de dados pessoais como a matéria prima para a formação de uma rede de comercialização de publicidade e serviços.

A crescente quantidade de dados pessoais, cotidianamente, produzidos na Internet, possibilitou a criação de uma infraestrutura invisível formada por empresas de análise de audiência, agência de marketing especializada em mídia eletrônica, publicidade online direcionada, entre outras atividades de análise de dados. São empresas que coletam, analisam e interpretam dados de acessos em várias fontes, sobretudo de sites “terceiros”, e representam hoje uma grande economia mundial a partir de dados pessoais.

Estas grandes empresas criam ambientes controlados e personalizados para o usuário, geralmente, atuando em diversas plataformas (sistemas para telefonia móvel, para computadores pessoais, TV, entre outras) e serviços. Pariser descreveu estes ambientes como mecanismos que criam e refinam a experiência do usuário prevendo suas ações e desejos, o que ele passou a chamar de bolha dos filtros (PARISER,2012,p.14).

As empresas agora são capazes de atingir avaliações de mercado significativas, empregando modelos de negócios que se baseiam na utilização de dados pessoais. Segundo a Microsoft¹, dados pessoais estão se tornando rapidamente uma moeda fundamental para o relacionamento entre a marca e o consumidor. Nesta relação, muitas vezes não transparente para o consumidor, as marcas dependem cada vez mais de dados pessoais para fornecer produtos e serviços personalizados e competir de forma eficaz.

De acordo com a empresa de marketing e fidelidade AIMA, em seu relatório anual Global Loyalty Lens report 2015², 80% dos consumidores entrevistados estão dispostos a compartilhar informações básicas como nome, e-mail e nacionalidade com as empresas. Já mais de 70% compartilham informações como data de nascimento,

ocupação, hobbies e interesses. O relatório ainda aponta a alta disposição dos brasileiros em compartilhar mais seus dados pessoais, fortalecendo assim a ação de empresas que atuam neste mercado a manterem um relacionamento mais próximo aos seus clientes.

Além das empresas de Marketing Digital, no ecossistema da Internet existem outros agentes que diante de seus interesses firmam acordos de cooperação ou desenvolvem tecnologias de rastreamento de comportamento e vigilância na Internet. Dentre estes agentes podemos assumir três grandes grupos:

- **Trackers:** Empresas de análise de audiência na Internet, agência de marketing especializada em mídia eletrônica, publicidade online direcionada, entre outras atividades de análise de dados. São empresas que coletam, analisam e interpretam dados de acessos em várias fontes, sobretudo de sites “terceiros”, ajudando estes a atingir suas campanhas publicitárias, gerar *leads*, ou seja, contatos capazes de se transformar em clientes, entre outras ações de marketing.
- **Governos:** Agências de Inteligência, defesa e vigilância mantidas por governos.
- **Plataformas de controle:** São grandes empresas que criam ambientes controlados e personalizados para o usuário, geralmente atuando em diversas plataformas e serviços. Google, Apple e Facebook, são grandes exemplos de plataformas de controle. Estas empresas desenvolvem múltiplos “sensores” em forma de ambientes digitais para “aprisionar” seus usuários privando-os de explorarem outras plataformas. São exemplos às plataformas de TV sob demanda, os ambientes limitados para download e instalação de aplicativos, a limitação de integração de produtos (celular, computador pessoal, tablet e dispositivos inteligentes), que as empresas Apple e Google criaram.

TECNOLOGIAS DE REGISTRO E RASTREAMENTO DE COMPORTAMENTO ONLINE

Há uma grande variedade de tecnologias para o rastreamento do comportamento online dos usuários. Contudo, este trabalho irá investigar a tecnologia de rastreamento

baseada em web cookies inseridos na navegação dos dez sites de notícias mais acessados por usuários de Internet brasileiros em abril de 2015 e maio de 2016.

Enquanto você navega na Internet por meio de um *browser* (navegador), constantemente, as páginas acessadas depositam um pequeno arquivo de texto, muitas vezes incompreensível para os usuários. Este arquivo conhecido como web cookie permite que os sites acessados obtenham informações específicas sobre quem está acessando suas páginas. Cada vez que você acessa um site, o seu navegador envia de volta o web cookie (ou cookies) correspondente à página acessada (servidor web) para que ele possa manter o controle de sua atividade, ou seja, sua experiência neste site.

O principal propósito do web cookies é identificar usuários e, possivelmente, preparar páginas personalizadas ou para salvar as informações de sessão de um site. Portanto, ao retornar a um site, ao invés de ser exibida uma página de boas-vindas genérica, você terá uma página de boas-vindas já com seu nome, além disso eles são usados para personalizar sua experiência de navegação, ou para exibir anúncios segmentados para você.

CLASSIFICAÇÃO DOS WEB COOKIES

Existe uma variedade de tecnologias de web cookies. A seguir iremos apresentar uma breve descrição das mais utilizadas.

- **Web cookies de sessão:** São criados e armazenados temporariamente durante uma sessão de navegação em um site e são excluídos do dispositivo do usuário quando o navegador é fechado. Fundamentais para o funcionamento de sistemas onde o usuário tem que navegar de uma página para outra, como em uma compra, aulas online, webmail, de modo que os detalhes da conta e outras preferências não sejam perdidos durante a navegação.
- **Web cookies Persistente:** Diferentes do cookie de sessão, este tipo não é excluído após o navegador ser fechado, e sim depois de um período de tempo específico definido pelo seu servidor. Criados para rastrear o comportamento de

usuários, um web cookie persistente, ou também conhecido como cookie de rastreamento, é constantemente usado por anunciantes para registrar informações sobre os hábitos de navegação durante um período determinado de tempo. Além disso, estes web cookies são utilizados para personalizar a experiência do usuário, armazenando suas preferências (como por exemplo idioma, temas últimas pesquisas) de modo que elas possam ser lembradas em uma próxima visita.

- **Web cookies de terceira parte:** São web cookies criados a partir de um site que não seja o que o usuário acessou. A maior parte destes web cookies são gerenciados por trackers, corporações especializadas em marketing digital ou grandes bancos de dados privados de usuários. São utilizados pelos sites acessados como fonte para propagandas ou buscas direcionadas. Um único site pode depositar em computador dezenas de cookies de terceira parte. Estes web cookies são considerados a grande ameaça a privacidade do usuário. Ao longo do tempo, essas empresas conseguem desenvolver um histórico detalhado dos tipos de sites que você frequentou.

Com o surgimento de ferramentas para bloqueio das técnicas de rastreamento web e a limitação técnica dos cookies, novas formas de rastreamento vem sendo desenvolvidas. Não necessariamente criada pelos motivos destes bloqueios e sim por motivos de limitações técnicas, o armazenamento de dados local, disponibilizado pelo HTML5³ e suportado por todos os navegadores de Internet, vem sendo cada vez mais adotado por desenvolvedores web. O conceito é similar aos cookies, mas com possibilidade de injeção de objetos com tamanhos superiores, permitindo assim o desenvolvimento de rotinas persistentes e mais sofisticadas.

Outra técnica muito utilizada são os Flash cookies, que também possuem um comportamento similar aos web cookies, porém não são armazenados no mesmo diretório no navegador do usuário. Esta técnica é chamada de compartilhamento local de objetos, ou Local Shared Objects (LSOs), e não aparece na lista de cookies do navegador, sendo assim difíceis de serem detectados e excluídos.

APRESENTAÇÃO DOS RESULTADOS

Para o desenvolvimento desta pesquisa foi utilizado a ferramenta de análise de Web Cookies do site de auditoria livre webcookies.org. Ainda, para definir o escopo da pesquisa foram escolhidos os dez sites de notícias brasileiros mais acessados segundo rank do site Alexa⁴, em abril de 2015 e maio de 2016. A atualização diária de conteúdo em língua Portuguesa foi o critério para definir um site de notícias brasileiro. Os dez sites analisados no período são apresentados na Tabela 1.

Tabela 1: Os dez sites de notícias mais acessados no Brasil (2015 e 2016)

Nome do Site	Endereço (url)
UOL	uol.com.br
Terra	terra.com.br
Globo.com	globo.com
R7	r7.com
Folha de São Paulo	folha.uol.com.br
IG	ig.com.br
Estadão	estadao.com.br
MSN	msn.com
Tecmundo	tecmundo.com.br
BOL	bol.com.br

Os web cookies aferidos referem-se ao acesso na página inicial (homepage) de cada site. A quantidade de web cookies injetados por site nos dois períodos de coletas é apresentada na Tabela 2.

Tabela 2: Quantidade de web cookies inseridos por site.

Nome do Site	Abril de 2015	Mai de 2016
Uol	60	161
Terra	120	44
Globo.com	46	1
R7	25	25
Folha de São Paulo	70	35
IG	70	38
Estadão	34	34
MSN	15	27
Tecmundo	23	33
BOL	10	10
TOTAL	473	408

Alusivo a classificação dos web cookies analisados, foi possível observar a presença de

web cookies de sessão de terceira parte em quase todos os sites, com exceção as homepages da globo.com e MSN, conforme é apresentado na Tabela 3. Foi observado um sensível desvio dos dados coletados nas amostras dos dois períodos analisados nestes sites. Contudo, estes desvios não foram possíveis de serem detectados pela ferramenta webcookies.org, apontando assim a necessidade de uma investigação futura mais aprofundada.

Tabela 3: Classificação dos web cookies analisados

Site de notícias	Nº de web cookies de terceira parte	Nº de web cookies persistentes	Nº de web cookies de sessão e de terceira parte
Uol	36	141	10
Terra	16	38	1
Globo.com	1	1	0
R7	12	20	1
Folha de São Paulo	21	27	3
IG	15	24	2
Estadão	15	23	3
MSN	10	22	0
Tecmundo	13	28	3
BOL	6	7	3

Por padrão a injeção e o acesso dos web cookies devem ser feita apenas pelo servidor da página web acessada pelo usuário. Contudo, é possível observar neste trabalho que a origem dos cookies injetados são de regiões diferentes do mantenedor dos sites acessados, como apresentados na Figura 2 e na Tabela 4.

Figura 2. Cookies por país.

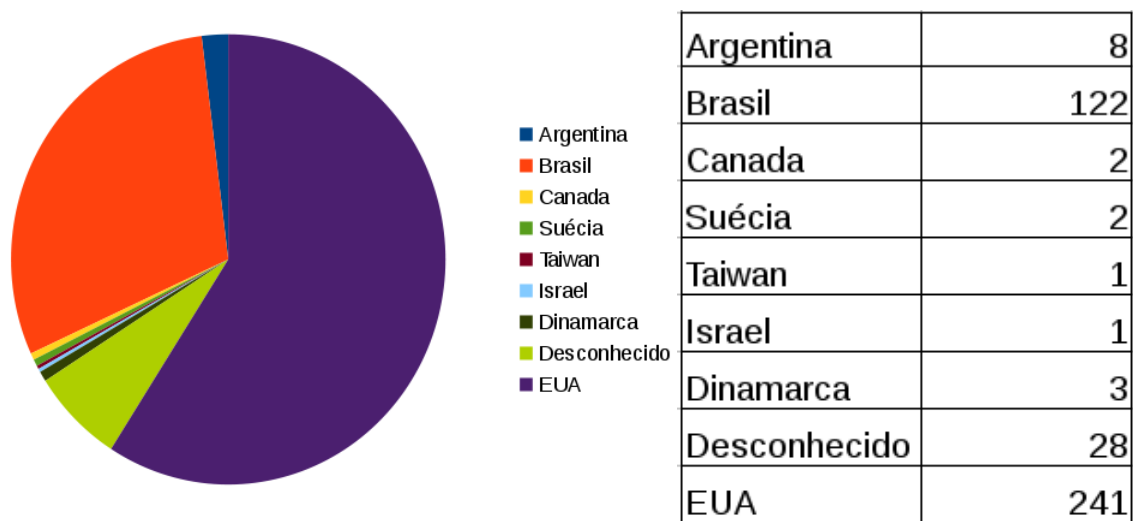


Figura 4. Relação de web cookies por empresa e país

Empresa	País	Qtde de web cookies	Empresa	País	Qtde de web cookies
Acuity	Canadá	1	Navegg	Brasil	5
Adform ApS	Dinamarca	3	NetSeer	EUA	2
Admotion	Argentina	8	OpenX Technologies	EUA	1
AdNexusMedia LLC	EUA	12	Pingdom	Suécia	1
Adobe	EUA	11	Predicta	Brasil	6
AlmondNet	EUA	8	Quantcast	EUA	3
App Nexus	EUA	5	Radium ONE	EUA	1
Appier	Taiwan	1	Reamp	Brasil	3
CHLEBA	Brasil	2	Rede Record	Brasil	10
ComScore Inc	EUA	18	revenueantra	EUA	1
Connexity	EUA	1	Rubicon Projects	EUA	74
Conversant	EUA	9	Run	EUA	2
DATA XU INC	EUA	5	Sekindo	Israel	1
Drawbridge	EUA	2	Tapad Inc.	EUA	2
Dstillery	EUA	3	Tecmundo	Brasil	6
DYNad	EUA	3	Terra	Brasil	8
eyereturn	Canadá	1	The Trade Desk	EUA	2
Facebook	EUA	3	Turn	EUA	9
Globo	Brasil	1	Twitter	EUA	5
Google	EUA	14	UOL	Brasil	38
Grupo Estado	Brasil	20	videology	EUA	1
IG	Brasil	21	Voice Five	EUA	5

Instituto Verificador de comunicação	Brasil	2	webspectator	EUA	1
MediaMath, Inc	EUA	10	Yahoo	EUA	1
Microsoft	EUA	21	You Tube	EUA	3

Por fim, neste trabalho foi possível observar que no período investigado, técnicas como flash cookies e armazenamento local (disponíveis no HTML5), vem sendo gradativamente adotada em substituição a técnica de injeção de web cookies, sobretudo por permitir o aumento da capacidade (tamanho) de injeção de dados e rotinas de rastreamento de navegação web. Ainda, que a maior parte dos web cookies observados são oriundos de plataformas de controle ou empresas de marketing digital americanas.

REFERÊNCIAS

- Assange, J. (2013). *Cypherpunks: liberdade e o futuro da internet*. Tradução Cristina Yamagami. São Paulo: Boitempo.
- Bauman, Z. (2013). *Vigilância líquida: diálogos com David Lyon*. Tradução: Carlos Alberto Medeiros. Rio de Janeiro: Zahar.
- Castells, M. (2003). *A Galáxia da Internet. Reflexões sobre a internet, os negócios e a sociedade*. Trad. Maria Luiza X. de A. Borges. Rio de Janeiro: Jorge Zahar Ed.
- Galloway, A. (2004). *Protocol: how control existis after decentralization*. Cambridge: MIT Press.
- Giddens, Anthony (1991). *As consequências da modernidade /Anthony Giddens*; tradução de Raul Fiker. - São Paulo: Editora UNESP.
- Kurose, J. F. e ross, K (2010). *Redes de Computadores e a Internet*. - 5ª Ed., Pearson.
- Pariser, E. (2012). *O filtro invisível: o que a internet está escondendo de você*. Tradução: Diego Alfaro. Rio de janeiro. Zahar.

* Doutorando no Programa acadêmico interdisciplinar de Ciências Humanas e Sociais da UFABC, onde concentra sua pesquisa na Linha Cultura, Comunicação e Dinâmica Social. E-mail para contato: rodolfo.avelino@ufabc.edu.br.

** Professor Doutor do Programa de Pós-Graduação em Ciências Humanas e Sociais da UFABC. E-mail para contato: sergio.amadeu@ufabc.edu.br.

¹ The Consumer Data Value Exchange: Building better terms of engagement for data (and why the future of brands depends on it).

² AIMA Global Loyalty Lens report 2015, disponível em: <http://aimia.com/content/dam/aimiawebsite/Aimia2015GlobalLoyaltyLensReport/Aimia%202015%20Global%20Loyalty%20Lens%20Report/AIMIA-Loyalty-Lens-Research-Report.pdf>. Acesso em 18/06/2016

³ Segundo o site DevMedia, com o surgimento da web, era necessário criar uma linguagem que fosse entendida por meios de acesso diferentes. Para tanto, Tim Berners-Lee desenvolveu o HTML, que é uma abreviação de Hypertext Markup Language, ou seja, Linguagem de Marcação de Hipertexto. Resumindo, o HTML é uma linguagem usada para a publicação de conteúdo (texto, imagens, vídeos, áudio etc.) na web. O HTML5 é a nova versão do HTML4 e um dos seus principais objetivos é facilitar a manipulação dos elementos, possibilitando o desenvolvedor modificar as características dos objetos de forma não intrusiva, fazendo com que isso fique transparente para o usuário final. Fonte: <http://www.devmedia.com.br/o-que-e-o-html5/25820>. Acesso em 02/11/2016.

⁴ <http://www.alexa.com/topsites/countries/BR>