

ARTIGO

NARRATIVAS AUTOMATIZADAS E A GERAÇÃO DE TEXTOS JORNALÍSTICOS:

a estrutura de organização do lead
traduzida em código

Copyright © 2016
SBPjor / Associação
Brasileira de Pesquisadores em Jornalismo

MÁRCIO CARNEIRO DOS SANTOS

Universidade Federal do Maranhão, Brasil

RESUMO - Descreve-se o experimento de construção de um software capaz de gerar leads e títulos jornalísticos de forma automatizada a partir de informações obtidas na internet. A possibilidade teórica já prevista por Lage no final do século passado baseia-se na estrutura simples e relativamente rígida desse tipo de construção narrativa, o que facilita a representação ou tradução da sua sintaxe em termos de instruções que os computadores podem executar. Discutem-se também as relações entre sociedade, técnica e tecnologia, fazendo um breve histórico sobre a introdução das soluções digitais nas redações jornalísticas e seus impactos. O desenvolvimento foi feito com a linguagem de programação Python e a biblioteca NLTK- *Natural Language Toolkit* – e usou os resultados do Campeonato Brasileiro de Futebol de 2013 publicados em portal da internet como fonte de dados.

Palavras-chave: Narrativas Automatizadas. Jornalismo online. Python. Inteligência artificial. NLTK.

NARRATIVAS AUTOMATIZADAS Y LA GENERACIÓN DEL TEXTO

PERIODÍSTICO: La estructura de la organización del lead traducida a código

RESUMEN - Se describe la experiencia de la construcción de un software capaz de generar leads potenciales y títulos de periódicos de forma automatizada de la información obtenida a través de Internet. La posibilidad teórica Lage ya previsto para el final del siglo pasado se basa en la estructura relativamente rígida y simple de este tipo de construcción de la historia, lo que facilita la representación o la traducción de su sintaxis en función de las instrucciones que la computadora puede ejecutar. El documento también analiza la relación entre la sociedad, la técnica y la tecnología, haciendo una breve historia de la introducción de soluciones digitales en las salas de redacción y sus impactos. El desarrollo se hizo con el lenguaje de programación Python y biblioteca NLTK y utiliza los resultados del Campeonato de Fútbol de Brasil 2013 publicados en el portal de Internet como fuente de datos.

Palabras clave: Narrativas automatizado. El periodismo en línea. Python. Inteligencia artificial. NLTK.

AUTOMATED NARRATIVES AND JOURNALISTIC TEXT GENERATION: The lead organization structure Translated into code

ABSTRACT - It describes the experiment of building a software capable of generating leads and newspaper titles in an automated fashion from information obtained from the Internet. The theoretical possibility Lage already provided by the end of last century is based on relatively rigid and simple structure of this type of story construction, which facilitates the representation or translation of its syntax in terms of instructions that the computer can execute. The paper also discusses the relationship between society, technique and technology, making a brief history of the introduction of digital solutions in newsrooms and their impacts. The development was done with the Python programming language and NLTK- Natural Language Toolkit library - and used the results of the Brazilian Soccer Championship 2013 published on an internet portal as a data source.

Keywords: Automated Narratives. Online journalism. Python. Artificial intelligence. NLTK.

1. Introdução - o medo e o fascínio das máquinas

Apesar de simplificada, a versão dualista das relações entre homens e tecnologia ainda hoje é utilizada. Sejam prometeicos ou fáusticos (RÜDIGER, 2007), apocalípticos ou integrados (ECO, 2006), ciberiluministas ou neoluditas¹, muito esforço tem sido dedicado por áreas como a Filosofia da Tecnologia e afins para discutir a questão, que se inicia com o conceito de técnica.

Se as origens da técnica repousam na antiguidade, o conceito de tecnologia veio bem depois. Ensina-nos Lemos (2002) que a tecnologia é a técnica moderna, muito distante do imaginário da antiguidade e liberta dos seus laços com o divino. É a técnica que, baseada na razão e no desenvolvimento científico, na física newtoniana, na matemática cartesiana e no empirismo, transforma a natureza em “objeto de livre conquista” (LEMOS, 2002, p. 45).

Para Rüdiger (2007, p. 175) “a técnica é, em essência, uma mediação do processo de formação da vida humana em condições sociais determinadas”. Já a tecnologia é

o conhecimento operacional que designamos pelo termo técnica enquanto se articula com a forma de saber que chamamos ciência, através da mediação da máquina e, potencialmente, em todas as áreas passíveis de automatização, conforme define o tempo que a criou, a Modernidade (RÜDIGER, 2007, p. 186).

Para Heidegger, a técnica é um modo de existência do homem no mundo, mas a partir da modernidade, esse existir tomará um rumo direto de agressão à natureza, agora sujeita ao conhecimento humano e à ideia de um progresso linear, constante e que não pode

ser parado. Para muitos, como Sennett (2009), abre-se aqui a caixa de Pandora, a deusa da invenção enviada por Zeus à terra e que, para os gregos, representava também a cultura das coisas produzidas pelo homem, através das quais este poderia causar danos a si mesmo.

Os grandes conflitos mundiais da primeira metade do século XX, o nazismo e o pesadelo da guerra fria e da ameaça nuclear materializaram os piores sonhos dos gregos num mundo que, em tese, deveria ser mais evoluído justamente pela existência da tecnologia.

As possibilidades de pensar as relações entre sociedade e tecnologia deram origem a novos campos como o que hoje conhecemos como *Science and Technology Studies* (STS). Nele pensadores como Castells (1999) e Feenberg (2002) têm se dedicado a formular um cenário compatível com os desafios de estudar uma relação obviamente multifacetada e complexa.

Em sua crítica a visões simplistas sobre o papel da tecnologia no mundo de hoje, Feenberg nos propõe inicialmente uma espécie de mapeamento das posições normalmente apresentadas e a partir delas tenta incorporar questões como democracia, poder e liberdade, como fatores também importantes a considerar nas discussões dos STS.

Na cartografia de Feenberg (2002) das sociedades modernas a tecnologia ocupa um lugar de destaque entre as fontes de poder que se articulam no meio social. Para ele, as decisões políticas que definem muitos dos aspectos da nossa vida cotidiana são direcionadas pela influência dos controladores dos sistemas técnicos, sejam eles das grandes corporações, militares ou de associações profissionais de grupos como físicos, engenheiros e mais recentemente, poderíamos sugerir também, desenvolvedores de software.

Ao fazer tal constatação o autor se remete ao pensamento de Marx que já no século XIX criticava a ideia de que a economia pudesse ser apenas regida por fatores extra- políticos, através de leis com a da oferta e da procura. Do mesmo modo, imaginar o papel da tecnologia sem avaliar as diversas relações que ela estabelece com a sociedade pode implicar numa visão reduzida do problema.

Num caminho semelhante à crítica marxista de uma economia regulada por uma ordem natural e inexorável, Feenberg (2010) relativiza a racionalidade da tecnologia a partir da constatação de que sua gênese e desenvolvimento acontecem no mundo dos homens e, por isso, também são influenciadas por ele.

Criação técnica envolve interação entre razão e experiência. Conhecimento da natureza é necessário para fazer um equipamento que funcione. Este é o elemento da atividade técnica que consideramos como racional. Mas o equipamento deve funcionar num mundo social e as lições da experiência nesse mundo influenciam o design (FEENBERG, 2010, p. 17)².

Se no campo da Filosofia é amplo o debate, o cinema ao longo de décadas tem traduzido esse imaginário de medo e fascínio em diversos filmes onde as soluções tecnológicas são representadas por robôs, autômatos, máquinas e até sofisticados programas de computador. Naves controladas por entidades automatizadas que se rebelam contra os humanos, como o computador HALL 9000 em “2001 – Odisseia no Espaço” de Kubrick (1968); que decretam sua extinção como em “Exterminador do Futuro” de James Cameron (1984) ou ainda que os escravizam, num mundo digitalmente criado, a “Matrix”, para utilização da humanidade como simples fonte de energia, dos irmãos Wachowski (1999), são apenas alguns dos inúmeros exemplos que poderíamos citar.

Na série de TV *Star Trek: The next generation*, que também ganhou os cinemas (*Star Trek - First Contact*, Jonathan Frakes, 1996), uma das piores ameaças alienígenas já enfrentadas foi a dos *Borgs*, raça de seres híbridos, biológicos e maquínicos, que rapidamente assumiam o controle das áreas que invadiam, a partir da conversão dos seres que encontravam em sua própria espécie, através da inserção de implantes que faziam as vítimas completamente integradas ao comando central, agindo como uma colônia de insetos, num exército cada vez maior.

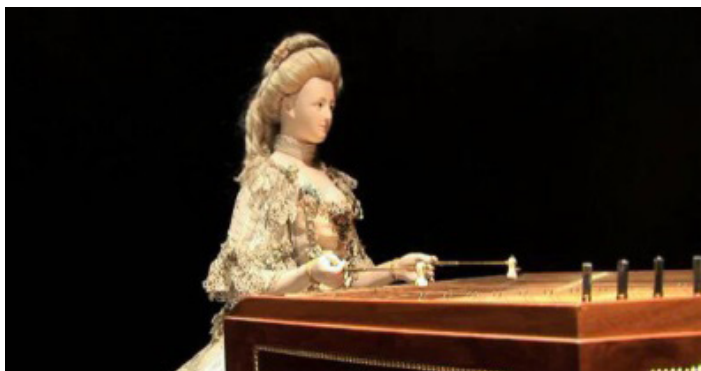
Por outro lado, o fascínio pelas máquinas é muito anterior à quase inevitável dependência contemporânea que estabelecemos com celulares, *smartphones*, *tablets* e tantos outros *gadgets* tecnológicos dos quais não queremos mais nos separar.

Na antiguidade e na idade média os relatos sobre autômatos eram restritos, sendo o século XVIII considerado sua época áurea. No trecho de Devaux (1964) é descrita a apresentação de uma dessas peças que ainda hoje podem ser vistas em Paris, a “Tocadora de Xilofone” de Roentgen, uma boneca musicista que, se supõe, tenha sido inspirada na figura de Maria Antonieta³.

Numa sala do Palácio de Versalhes, entre as saias de balão e os vestidos da corte, o exímio automatista Roentgen, apresenta a Luís XVI outra obra-prima. Aquela Tocadora de xilofone, de corpete decotado e vestido de seda bordada, provoca a curiosidade geral; fala-se do corpo da boneca divinamente

modelado debaixo do vestido, pasma-se com a precisão e a sua altiva graciosidade. Uma pequena ária muito viva saltita desperta pela saraivada dos martelos de marfim; todo um século se desprende dessa música elegante e seca; e quando a jovem, debaixo dos anéis do seu penteado alto, volta a cabeça para saudar, a semelhança levanta um sussurro em toda a sala... Mais tocante e mais perfeita no mecanismo que o Escrivão ou a Musicienne, dos Jaquet-Droz, ou o Pato, de Vaucanson, e os seus Flautistas, a Tocadora de xilofone evoca fielmente, quanto a nós, a Primeira Idade do Automatismo (DEVAUX, 1964, p. 7).

Figura 1- Tocadora de Xilofone de Roentgen, restaurada em 1864 por Robert Houdain.



Fonte: Lutice Créations ([200-]).

2. Jornalismo e Tecnologia

Desde os primórdios, o jornalismo esteve ligado a algum tipo de tecnologia, sendo o processo de impressão de Gutemberg e seu desenvolvimento um dos principais fatores que alavancou a expansão dessa atividade.

Muito tempo depois, já no final do século XX, a chegada das redes, da internet e dos computadores às redações iniciou um ciclo de profundas mudanças que até hoje está em andamento e que alguns como Soria (2014) descrevem simplesmente como um *tsunami*, traduzindo o impacto devastador que positiva e negativamente a digitalização de grande parte do processo de produção jornalística tem causado.

Machado (2003), ao descrever o início da mudança, nos ensina que duas posições se estabeleceram para compreender o que estava acontecendo. A primeira, que poderíamos chamar de instrumentalista, entendia que computadores eram apenas mais uma ferramenta a

disposição dos jornalistas, artefatos adicionais a serem utilizados na execução do seu trabalho, como antes haviam também sido as inovações do telégrafo, da máquina de escrever e do telex, entre outras.

Já na segunda forma de entender a transformação, a chegada do digital representava uma alteração muito mais extensa.

A falta de clareza sobre as consequências para o jornalismo da disseminação do suporte digital dificulta a compreensão plena das particularidades da prática jornalística nas redes, das mudanças no perfil do profissional, na estrutura organizacional das empresas jornalísticas e das funções que o usuário passa a ocupar no sistema de produção de conteúdos (MACHADO, 2003, p. 2).

Bradshaw e Rohumaa (2011), traçam um breve histórico sobre o início do jornalismo *online* no ocidente e indicam os britânicos *Today* de 1986 como o precursor na produção de conteúdo usando tecnologia digital e o *Daily Telegraph* como um dos primeiros jornais impressos a ter sua versão transposta⁴ na ainda pouco conhecida internet de 1994.

Hoje, na redação digital, as partículas de informação chegam em fluxo contínuo e, a partir delas, construímos nossas histórias, iniciando as vezes apenas com poucas palavras na área de “últimas notícias” e, quando merecem, chegando às grandes reportagens contadas nos portais jornalísticos, como a premiada *Snow Fall* (BRANCH, ([200-])) do *New York Times*.

As mudanças tecnológicas e seus impactos, entretanto, não podem ser avaliados de forma isolada já que fatores sociais e econômicos parecem constituir também vetores imbricados no complexo cenário da mídia contemporânea.

Haak, Parks e Castells (2012) apresentam um panorama de tendências pensando o futuro dentro de uma era digital interconectada. Para eles as novas possibilidades tecnológicas não geraram uma crise no jornalismo (que continua tendo seu papel social fundamental) e sim o reconhecimento de que os modelos de negócio sobre os quais se articulavam as grandes empresas de mídia precisavam ser revistos ou atualizados.

As funções essenciais do jornalismo resumidas nas etapas de observar fatos relevantes e fazer boas perguntas para as pessoas certas, tentar compreender as observações e respostas dentro de um determinado contexto e, por fim, explicar os resultados das etapas anteriores para os outros são, pelos autores, sintetizadas em termos de coleta de dados, interpretação e narração. Para eles esse núcleo

essencial do fazer jornalístico não mudou e sim foi reconfigurado e expandido pelas novas possibilidades tecnológicas.

Tentando detalhar essas premissas Haak, Parks e Castells (2012) listam novas ferramentas e práticas que seriam as principais tendências desse jornalismo do futuro que se propõem a descrever. São elas: jornalismo em rede (*networked journalism*); inteligência coletiva e conteúdo gerado por usuários (*crowdsourcing and user-generated content*); mineração de dados, análise de dados, visualização de dados e mapeamento (*data mining, data analysis, data visualization and mapping*); jornalismo visual (*visual journalism*); jornalismo de ponto de vista (*point of view journalism*); jornalismo automatizado (*automated journalism*) e jornalismo global (*global journalism*).

Analisar todas elas está além do escopo desse texto. Nosso foco está na penúltima, a do jornalismo automatizado - JA, que pode ser resumido pelo fato de que hoje parte do conteúdo jornalístico publicado já não é mais escrito por humanos e sim por máquinas, via software, através de ferramentas e soluções que envolvem desde simples listas de palavras (como o experimento que apresentamos a frente) até complexos modelos de inteligência artificial.

Independente da modalidade escolhida para realizar tal tarefa, o fato de desconectar texto e jornalista de forma tão radical nos parece representar a tendência tecnológica mais problemática e ao mesmo tempo interessante, justamente por ser aquela que vai operar com todo o imaginário já descrito na introdução.

É preciso observar também que tal temática é bastante recente dentro dos estudos do Jornalismo e, em nosso entendimento, não pode ser confundida com a do Jornalismo Digital em Base de Dados - paradigma JDBD (BARBOSA, 2007, 2008, 2009, 2011; FIDALGO, 2004, 2007; MACHADO, 2006; RAMOS, 2011a, 2011b) já bastante explorado por vários autores e que organiza suas conclusões tendo como eixo principal as funções das bases de dados “como definidoras da estrutura e da organização, bem como da composição e da apresentação de conteúdos de natureza jornalística” (BARBOSA, TORRES, 2013, p.154).

Mesmo dentro do amplo espectro de funcionalidades coberto pelo paradigma JDBD a parte dedicada à automação (nele denominada automatização) é descrita como abaixo.

Inerente ao uso de bases de dados nos processos de armazenamento, estruturação, organização e apresentação das informações. Permite agilidade nos processos de apuração,

formatação de conteúdos a partir do que está armazenado no arquivo e também as chamadas estatísticas dinâmicas ou Sistemas de Recomendação de Notícias (SRN), entre outros. Há três tipos de automatização: parcial, procedimental (nível intermediário) e total. (BARBOSA, TORRES, 2013, p.154-155).

A primeira diferença a observar é que bases de dados, ponto central sobre o qual se articula o JDBD, constituem um tipo de software específico, como também são as planilhas eletrônicas, os processadores de texto e os softwares de apresentação e autoria. Os algoritmos de inteligência artificial (IA), que suportam as narrativas automatizadas, apesar de operarem eventualmente acoplados ou acessando bases de dados, pertencem a uma categoria diferente e não devem ser confundidos com outras, principalmente pela lógica de procedimentos específica sob a qual operam.

De forma simplista, se a essência do trabalho das BDs é estabelecer e comutar relações entre dados, que podem ser reconfigurados de diversas formas para oferecer, como saída, múltiplas combinações entre eles; como veremos abaixo, os softwares da IA enfrentam basicamente o problema da representação de processos do mundo real dentro do ambiente computacional. Softwares de IA aprendem e realizam novas funções a partir do processamento que fazem sobre os conjuntos de dados que recebem como entrada e, por isso, são logicamente mais complexos e ricos em termos do que podem oferecer.

É essa complexidade que também se reflete no tipo de manipulação que BDs e soluções de IA proporcionam. As funcionalidades das BDs no processo de produção do jornalismo digital operam, poderíamos dizer, num nível macro, montando, por exemplo, uma página dedicada a todo o material sobre um determinado time de futebol, a partir dos conteúdos pré-existentes e dos metadados⁵ a eles acoplados. Já os algoritmos de IA, no jornalismo automatizado, vão operar num nível micro, o da constituição do próprio texto, quebrando por dizer uma barreira nova na inserção dos recursos computacionais dentro das atividades do jornalismo.

Definido, portanto, o recorte e as principais linhas de diferenciação entre JDBD e o jornalismo automatizado, pretendemos, nesse trabalho:

- a - compreender como isso está sendo feito e por quem;
- b- avaliar, mesmo que de forma exploratória, que tipos de impactos é possível perceber a partir da análise dos primeiros trabalhos acadêmicos já tratando desse objeto;

c- seguindo a linha experimental sob a qual operamos, aprofundar o conhecimento sobre essa tendência, a partir do desenvolvimento de uma prova de conceito que possa, ainda que de forma simplificada, replicá-la.

3. Narrativas Automatizadas - *Narrative Science e Automated Insights*

Morozov (2012) utilizando um sugestivo título: “Um robô roubou o meu Pulitzer!” relata os primeiros movimentos de empresas de inteligência artificial, entre elas a *Narrative Science*⁶, no negócio de gerar notícias. O produto da empresa: conteúdo jornalístico automatizado vendido como serviço para portais de notícias, principalmente da área de esportes e finanças, onde uma boa parte da informação utilizada advém de números e relações entre grandezas mensuráveis como a cotação do dólar ou o resultado de uma partida de futebol.

Figura 2 - Print de matéria sobre jornalismo automático

HOME / FUTURE TENSE: WHAT'S TO COME?

Article from **future tense**
ASU | NEW AMERICA | SLATE

A Robot Stole My Pulitzer!

How automated journalism and loss of reading privacy may hurt civil discourse.

By **Evgeny Morozov** | Posted Monday, March 19, 2012, at 7:11 AM ET



Automated journalism like that produced by Narrative Science could perhaps save media jobs, but it can also hurt civil discourse.
William Gottlieb/Library of Congress

Can technology be autonomous? Does it lead a life of its own and operate independently of human guidance? From the French theologian Jacques Ellul to the *Unabomber*, this used to be widely accepted. Today, however, most historians and sociologists of technology dismiss it as naive and inaccurate.

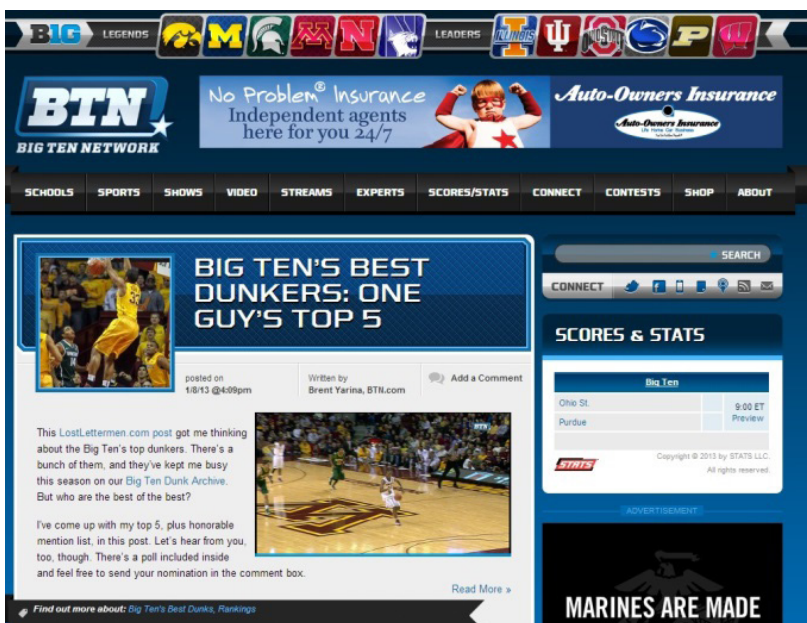
Yet the world of modern finance is increasingly dependent on automated trading, with sophisticated computer algorithms finding and exploiting pricing irregularities that are invisible to ordinary traders.

Meanwhile, *Forbes*—one of financial journalism's most venerable institutions—now employs a company called Narrative Science to automatically generate online articles about what to expect from upcoming corporate earnings statements. Just feed it some statistics and, within seconds, the clever software produces highly readable stories. Or, as *Forbes* puts it, “Narrative Science, through its proprietary artificial intelligence platform, transforms data into stories and insights.”

Fonte: Morozov (2012).

Narrative Science (NS) nasceu de um projeto de pesquisa chamado “*Stats Monkey*” desenvolvido por alunos e professores de Ciência da Computação e Jornalismo da *Northwestern University* a partir do Infolab e que basicamente escrevia resumos sobre resultados de jogos do *baseball* americano. Em 2010 a empresa mudou de nome e logo em seguida patenteou uma plataforma de autoria baseada em inteligência artificial chamada *Quill*.

Figura 3 - Portal Big Ten especializado em notícias do esporte e cliente da NS



Fonte: Big Ten Network (2014).

Automated Insights (AI) é outra companhia que já fornece conteúdo jornalístico automatizado para diversos clientes. Nascida com o nome de *StatSheet*, em 2008, a empresa recebeu financiamento de uma entidade de apoio à inovação no estado da Carolina do Norte nos EUA e iniciou um percurso de desenvolvimento que em 2014 contabilizou, segundo seu site oficial (AUTOMATED INSIGHTS, 2013), mais de 300 milhões de textos escritos automaticamente, entre relatórios empresariais e notícias jornalísticas.

Figura 4 - Exemplos de conteúdo publicado por AI em plataformas móveis

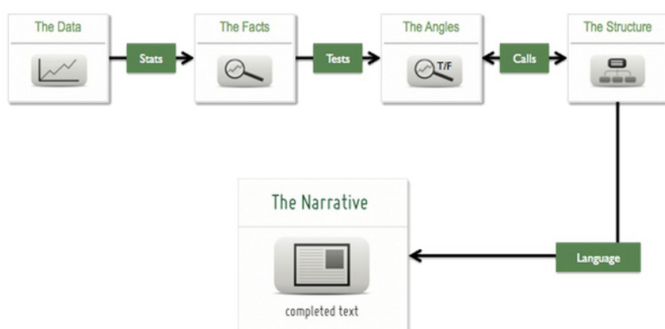


Fonte: Automated Insights (2013).

Arce (2009) já havia feito uma avaliação da possibilidade de automatização dos discursos incluindo nela as ideias de Lage (1997) sobre o tema, ambos, entretanto, em caráter teórico e não experimental.

Coppin (2010, p. 24) esclarece que, no campo da inteligência artificial, uma das principais questões está relacionada à representação da realidade que será utilizada pelo programa de computador, já que “para um computador poder solucionar um problema relacionado ao mundo real, ele primeiro precisa de um meio para representar o mundo real internamente. Ao lidar com aquela representação interna, o computador torna-se capaz de solucionar problemas”.

Figura 5 - Processo de transformação de dados brutos em narrativas da NS



Fonte: Narrative Science (2010).

Na questão específica do conteúdo jornalístico, as empresas citadas começaram a produzir *leads* basicamente por ser uma forma

que apresenta uma estrutura interna bastante definida e por isso traduzível de modo mais fácil para uma sequência de instruções a serem realizadas por uma máquina.

Carlson (2014) considera o jornalismo automatizado, entre as novas práticas jornalísticas centradas em dados, a com maior potencial disruptivo devido à limitada intervenção humana no processo, restrita basicamente às escolhas durante o desenvolvimento do código de programação. Devemos observar, inclusive, que tal participação humana mínima, pode estar acontecendo com pouca ou nenhuma intervenção de jornalistas profissionais, uma vez que as principais soluções já implementadas no mercado são oriundas de empresas privadas de inteligência artificial, que protegem seus processos de desenvolvimento e as equipes envolvidas, inclusive via registro de patentes, num modelo de propriedade industrial fechado.

Para o autor, que em sua pesquisa avalia as reações publicadas por jornalistas à utilização dos serviços da *Narrative Science* em redações, a emergência do JA traz sérias questões sobre o futuro da prática jornalística enquanto atividade laboral, sobre os padrões tradicionais de composição do conteúdo jornalístico e ainda sobre a própria identidade e autoridade do jornalismo como função social, bem como suas práticas de operação.

Já Clerwall (2014) aborda a questão do conteúdo gerado por software a partir de uma abordagem experimental. Apresentando textos jornalísticos, escritos por humanos e por máquinas, para um grupo de leitores, sem identificar sua origem, o autor avalia a percepção dos receptores sobre o material, principalmente em termos de qualidade, credibilidade e objetividade, observadas pelos participantes nessas narrativas.

Apesar de tratar-se de um estudo exploratório alguns resultados merecem destaque. Os textos gerados por algoritmos foram definidos como descritivos e entediantes, entretanto, destacaram-se no critério da objetividade. O estudo mostra também que nos resultados não ficou muito clara a distinção, por parte dos que tinham acesso ao material, entre o que era noticiário gerado por jornalistas e o que era produzido por computadores.

Dalen (2012) foca sua análise nas habilidades requeridas para a execução do trabalho jornalístico e no processo de comercialização de notícias. Partindo da percepção dos profissionais diante das narrativas automatizadas o autor reúne um quadro de pontos positivos e negativos levantados.

Os jornalistas ouvidos destacaram a capacidade de análise, a personalização e a criatividade como pontos importantes diante de características como factualidade, objetividade, simplificação e velocidade, essas últimas mais próximas do alcance do JA. Mesmo diante das ressalvas feitas pelos profissionais, foi apontada a possibilidade positiva de que a automatização de parte das narrativas propicia mais tempo para a investigação e o aprofundamento em matérias de maior relevância.

Não tão focado no jornalismo automatizado, o trabalho de Lewis e Usher (2014) se propõe a avaliar as possibilidades do encontro entre profissionais do jornalismo e desenvolvedores de software a partir do estudo de caso da iniciativa da rede global *Hacks/Hackers*. Utilizando o conceito de zonas de troca, os pesquisadores analisam as possibilidades de engajamento e cooperação entre os dois grupos, discutindo as implicações, desafios e oportunidades que podem advir dessa combinação.

Mais próximo de uma postura negativa diante da automatização, Latar (2014) descreve a nova lógica da extração de dados dos grandes repositórios digitais como uma tendência disposta a traduzir a complexidade dos sistemas sociais a partir das partículas de informação que geramos, através das interações e engajamento, via plataformas de mídias sociais e dispositivos móveis, tendências emergentes no cenário da comunicação digital.

A isso ele associa o nascimento de uma nova ciência que denomina de *social physics*, numa analogia com os átomos, partes microscópicas da matéria cujo estudo permite inferir conclusões sobre o todo do qual fazem parte. Dentro desse contexto, o autor avalia o surgimento do que chama de jornalismo robótico como um subproduto dessa tendência, baseado na extração automatizada de informação dos grandes silos de dados e na conversão desse conhecimento, via software, em narrativas para leitura sem o envolvimento humano na etapa da produção.

Considerando os custos envolvidos nos procedimentos tradicionais de geração de notícias, o autor alerta para um processo de obsolescência dos jornalistas em detrimento da consolidação de engenheiros de software e gestores de bases de dados como os empregados mais importantes dentro das empresas de mídia.

Em paralelo à emergência do novo campo da física social, a narração, arte de contar histórias, também está se transformando num empreendimento científico através de algoritmos de inteligência artificial (IA) que se aproveitam do grande conhecimento desenvolvido pelos campos da linguística e do estudo da linguagem natural. Os algoritmos de IA são estruturados para converter fatos em histórias para leitura numa fração de segundos⁷. (LATAR, 2014, p.65)

4. Modelagem de experimento com resultados do futebol

Para construir nosso experimento de narrativa automatizada nos propomos a desenvolver um código de programação capaz de escrever pequenos textos sobre os resultados do Campeonato Brasileiro de Futebol de 2013. Utilizamos a linguagem de programação Python⁸ por considerá-la de mais fácil aprendizagem para não especialistas em programação como jornalistas e profissionais da comunicação⁹.

A linguagem Python permite a utilização de diversos módulos de programação já desenvolvidos previamente e com finalidades específicas, facilitando a construção das soluções a partir da combinação de funções cujo código já existe. A biblioteca NLTK¹⁰ – *Natural Language Toolkit* – que utilizamos nesse projeto é um desses exemplos e incorpora um grande número de recursos para o processamento de textos.

A modelagem do problema foi feita a partir da seguinte sequência: obter resultados dos jogos e informações complementares tais como local da partida e número da rodada; registrar essas informações em alguma estrutura simples de arquivo que pudesse posteriormente ser consultada para a construção do material; traduzir as próprias regras do torneio em termos de variáveis e relações para que a sintaxe do regulamento pudesse orientar a concatenação dos elementos do texto; gerar as frases a partir dos resultados das operações realizadas com os dados coletados nas partidas.

Assim partimos para uma solução que, a partir de um endereço específico na internet onde esses dados fossem disponibilizados, conseguia ler todas as informações iniciais de forma automática e mais rápida. Para os testes selecionamos a página do portal Terra dedicada à área de esportes que publicava a cada rodada os resultados e a tabela atualizada do campeonato (PORTAL TERRA, 2014)¹¹. A tabela foi usada como instrumento de validação dos cálculos do software já que ela também totalizava as métricas que as regras do torneio geravam tais como número de jogos, pontos ganhos, gols feitos, gols sofridos, saldo de gols e índice de aproveitamento.

Definida a estratégia de coleta dos dados fizemos a parte do código que salvava esses elementos associando-os a cada time, numa estrutura que na linguagem Python é conhecida por “dicionário”, onde a cada elemento, chamado de chave, são associados valores diversos, cada um representando alguma informação gerada a partir dos resultados dos jogos.

Figura 6 – Parte do código que mostra o endereço de extração dos dados e os times na estrutura de chaves do dicionário, inicialmente com todos os campos zerados.

```

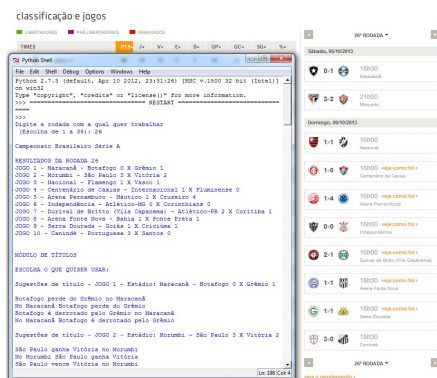
CBPontos3.py - D:\Python\ParserTestes\CBPontos3.py
File Edit Format Run Options Windows Help
from BeautifulSoup import BeautifulSoup
from operator import itemgetter
import urllib, urllib2, urllib
input=urllib.urlopen("http://esportes.terra.com.br/futebol/brazil-serie-a")
data=urllib2.urlopen(input).read()
input.close()
times = {
    u"Atlético-MG": [0,0,0,0,0,0,0,0,0,0], u"Vasco": [0,0,0,0,0,0,0,0,0,0], u"Fluminense":
    u"Cruzeiro": [0,0,0,0,0,0,0,0,0,0], u"Internacional": [0,0,0,0,0,0,0,0,0,0], u"São
    u"Flamengo": [0,0,0,0,0,0,0,0,0,0], u"Corinthians": [0,0,0,0,0,0,0,0,0,0],
    u"Santos": [0,0,0,0,0,0,0,0,0,0], u"Bahia": [0,0,0,0,0,0,0,0,0,0], u"Vitória": [0,0,

```

Fonte: Elaborado pelo autor.

Ao iniciar o código o usuário é demandado apenas a escolher o número da rodada que deseja explorar. O software coleta os resultados de todas as rodadas até chegar à selecionada e assim vai registrando os resultados e acumulando-os na estrutura do dicionário. É interessante observar que o que é retirado do portal da internet são apenas os resultados dos jogos. Com eles o software aplica as regras do torneio para calcular os outros valores associados ao time. Por exemplo, ao coletar o resultado de determinada partida, o software compara o número de gols dos dois times envolvidos, se um deles é maior do que o outro, o de valor maior ganhou a partida e, por isso, no registro referente a pontos ganhos são acrescidas três unidades. O perdedor não soma nada no registro e no caso de saldo de gols iguais, uma unidade é acrescentada a cada um dos times, indicando os pontos por um empate.

Figura 7 - Tela que compara a página do portal com os resultados e a tela gerada pelo programa onde podem ser vistos primeiro os dados registrados e depois as sugestões de título baseadas nos resultados

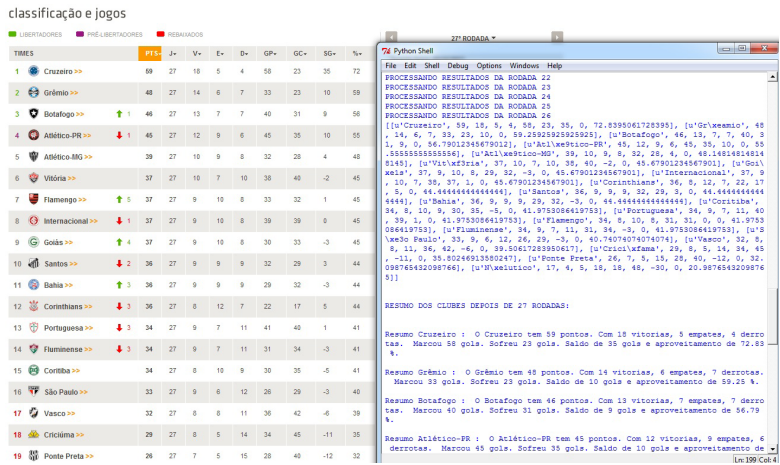


Fonte: Elaborado pelo autor.

À medida que o software registra os jogos das rodadas ele vai atualizando todos os parâmetros adicionais já listados, que são representações definidas pelas próprias regras do torneio, incluindo, na estrutura de dicionário que foi criada, um conjunto de dados que será utilizado para inferir várias outras informações como a própria posição do time na tabela, o número de pontos que cada um tem e seu aproveitamento (calculado dividindo o total de pontos conseguidos pelo total de pontos disputados).

Esses números permitem ao software escrever textos com mais informações.

Figura 8 - Tela do software que demonstra a estrutura do dicionário atualizada e um pequeno texto de resumo da situação do time no campeonato a partir dos elementos registrados

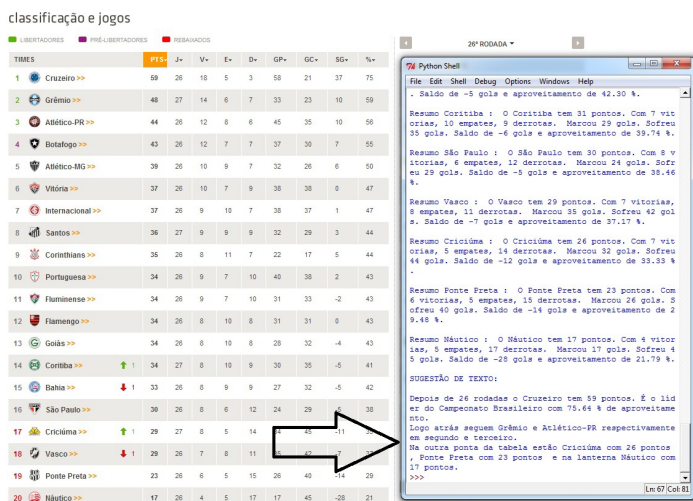


Fonte: Elaborado pelo autor.

Num nível com um pouco mais de complexidade é escrito então um lead com informações gerais sobre a situação do campeonato naquela rodada. Basicamente a partir do resumo que é inferido com a situação individual de cada time, o software faz o texto indicando os líderes com seus números e os lanterνας do torneio, aspectos que normalmente são enfatizados em noticiário desse tipo. A construção desse conteúdo, apesar de um pouco mais complicada, também parte da ideia de concatenar unidades menores de informação a partir de listas de palavras e expressões comuns nesse tipo de texto.

A título de ilustração seria como definir uma estrutura prévia onde alguns elementos, no caso o nome dos times e suas métricas, podem ser imaginados como lacunas a ser preenchidas por quem estiver naquelas posições em uma determinada rodada. A ideia de arquivos dinâmicos, ou seja, que se alteram ao longo do tempo, pode ser utilizada aqui.

Figura 9 - Tela do software com o que seria o lead construído a partir das informações lidas sobre o campeonato em determinada rodada



Fonte: Elaborado pelo autor.

Conclusões

Apesar de ter sido conduzido apenas em caráter exploratório, o experimento indica a possibilidade real e não apenas teórica de produzir alguns tipos de estruturas jornalísticas de forma automatizada.

Ficou claro que conteúdos baseados em informações numéricas e relações que podem ser traduzidas mais facilmente em expressões matemáticas, numa sintaxe mais restrita como a que pode ser extraída das regras de um torneio esportivo, por exemplo, são mais fáceis de reproduzir.

Da mesma forma que fizemos com os resultados do Campeonato Brasileiro, não seria difícil construir algo semelhante para gerar pequenos textos informando as variações do câmbio ou

de ações das bolsas de valores, a previsão do tempo para cidades ou regiões e outros conteúdos que, se observarmos, são construídos com uma estrutura que se repete com pequenas variações.

A capacidade de coletar e processar informações em grande quantidade e variedade parece indicar o potencial de uso desse tipo de solução, principalmente no jornalismo online e nos grandes portais da internet que precisam atualizar seus conteúdos com mais agilidade.

Em termos teóricos é importante ressaltar que Jornalismo em Base de Dados (JBD) e Jornalismo Automatizado (JA) não devem ser confundidos e operam com lógicas diferentes, apesar de estarem inseridos no processo maior de evolução das rotinas de produção jornalística a partir da utilização de recursos tecnológicos.

O fato de sua implementação dar-se através da utilização de algoritmos de categorias diferentes é uma das razões para isso. Enquanto o JBD utiliza-se das bases de dados e de sua capacidade combinatória, resultando saídas que se estabelecem a partir das relações previamente definidas na sua construção; o JA tem como principal suporte os algoritmos de inteligência artificial que são capazes de inferir relações novas e literalmente aprender à medida que são utilizados e processam os dados que lhes vão sendo disponibilizados.

Outra diferença a observar é sobre o nível de granularidade de sua atuação. Enquanto o JBD vai atuar num nível macro, concatenando notícias inteiras a partir dos metadados que foram associadas a elas para criar, por exemplo, uma página personalizada para torcedores de determinado time de futebol; o JA opera num nível micro, interno, na esfera das palavras e da construção mais básica de sentido mediante seu ordenamento e de suas relações sintáticas e semânticas.

Os desdobramentos desse tipo de tecnologia no mercado ainda não podem ser avaliados. É importante ressaltar também que mesmo as soluções mais complexas de inteligência artificial ainda estão distantes de replicar as sutilezas e complexidades de um bom texto jornalístico, principalmente numa língua como a portuguesa, que até hoje apresenta dificuldades para outras categorias de software como os de reconhecimento de voz e tradução, para conseguirem níveis altos de acerto.

Um dos mais recentes estudos sobre o tema avalia a situação atual do JA e afirma que

Nos últimos anos o uso de algoritmos para gerar notícias automaticamente a partir de dados estruturados abalou a indústria do jornalismo – especialmente desde que a Associated Press, uma das maiores e mais conhecidas companhias jornalísticas do mundo, começou a automatizar a produção

dos seus relatórios trimestrais de resultados. Uma vez desenvolvidos, os algoritmos podem não apenas criar milhares de novas histórias sobre determinado tópico mas também fazer isso mais rápido, mais barato e potencialmente com menos erros do que qualquer jornalista humano. Obviamente esse avanço tem aumentado o medo dos jornalistas de que o conteúdo automatizado vá eliminar empregos nas redações apesar de pesquisadores e profissionais observarem o potencial dessa tecnologia para melhorar a qualidade das notícias¹². (GRAEFE, 2016, p. 4)

Por outro lado, a precarização do trabalho e a replicação indiscriminada de releases e conteúdos gerados por fontes, justificados de forma simplista pela pressão do tempo e pela necessidade de atualização constante, são um risco para os profissionais já que, como foi demonstrado, as operações simples e baseadas em estruturas comuns têm muito mais chances de serem replicadas automaticamente.

O aprofundamento no trabalho de apuração, o jornalismo investigativo, a extração de relações complexas a partir de dados inter-relacionados e a criação de infográficos e formas alternativas de visualização de informações nos parecem bons exemplos de como a atividade humana pode continuar sendo essencial no que se considera um jornalismo de qualidade. A melhoria dos currículos e dos programas de formação na área também terá papel importante nos impactos dessas novas tecnologias.

Essas premissas com as quais temos trabalhado nos últimos três anos de forma geral foram confirmadas pelo estudo do *Tow Center* (GRAEFE, 2016) que alinha algumas das principais constatações e consequências ligadas ao crescimento do jornalismo automatizado.

O trabalho, por exemplo, indica a adoção do JA por grandes companhias de notícias, basicamente guiada pela crescente disponibilidade de dados estruturados, condição importante para o trabalho dos algoritmos, e pelo objetivo das empresas jornalísticas de reduzir custos e aumentar a quantidade de conteúdo disponível. Como já dissemos anteriormente, o estudo constata o potencial de crescimento do JA em rotinas de produção ligadas a tópicos repetitivos que, via software podem ser executadas de forma mais rápida, em larga escala e “potencialmente com menos erros do com jornalistas humanos” (GRAEFE, 2016, p. 5).

Outro aspecto interessante abordado é o potencial do JA para criar notícias sob demanda, a partir das questões dos usuários sobre determinados temas permitindo um maior nível de personalização,

bem como gerar, a partir do mesmo conjunto de dados, conteúdo em diferentes línguas e angulações.

Entre as limitações Graefe (2016) ressalta o fato do software se basear em dados e inferências que podem estar sujeitos a distorções e erros, comprometendo o conteúdo e trazendo ainda questões adicionais como a da responsabilidade sobre o mesmo, que continua caindo sobre o jornalista ou editor, e a da transparência do processo, ou seja, a informação para o usuário de como o algoritmo opera. O autor também lembra que as soluções automatizadas não conseguem realizar algumas tarefas essenciais para o trabalho como explicar um fenômeno novo (sobre o qual não há dados anteriores), fazer perguntas e estabelecer causalidade.

Em relação aos jornalistas o estudo aposta num fortalecimento da relação homem-máquina dentro das redações e sugere justamente que estes devem se focar em tarefas que os algoritmos não realizam com facilidade, como a análise em profundidade, as entrevistas e as reportagens investigativas. Para a sociedade em geral o estudo aponta que o excesso de conteúdo noticioso gerado pelo JA vai dificultar o trabalho das pessoas para encontrar conteúdo que seja mais relevante para elas.

Se “resistir é inútil”¹³ parece ser uma afirmação intimamente ligada às relações entre homens e técnica na história das sociedades, no campo do jornalismo, um texto criativo e bem elaborado poderá nos garantir a convivência pacífica com as soluções automatizadas que tem seu valor em processos repetitivos e de baixo nível de execução.

As informações disponíveis ainda são inconclusivas também para que se estabeleça um novo gênero jornalístico, apesar de termos como “jornalismo automatizado” ou “robotizado” estarem já aparecendo com mais frequência na literatura do campo. Evidencia-se, entretanto, um tipo específico de modo narrativo, baseado na concatenação de dados estruturados, na possibilidade de geração de inferências e relações semânticas a partir do uso intensivo sobre grandes quantidades de informação e na ausência da ação humana no processo, resultando, segundo alguns como Clerwall (2014) e Gaefer (2016), em textos mais objetivos e confiáveis.

É interessante observar que o software que desenvolvemos se alimenta das informações intrínsecas ao evento ou contexto para o qual é direcionado, bem como das relações internas que ali se estabelecem; hoje lidando apenas com questões simples

mas, no futuro, talvez, sendo apto a identificar situações mais complicadas a partir da evolução de tecnologias como a geração de linguagem natural, a exemplo da solução NLTK que usamos em nosso experimento.

Muito mais nociva do que a geração de textos jornalísticos via software parece ser a automatização dos jornalistas que deixam de exercer a ação humana e complexa ligada à sua atividade, no exercício das práticas da profissão. Esse parece ser o grande problema que teremos que enfrentar, sejamos nós céticos, temerosos ou fascinados por tecnologia.

REFERÊNCIAS

ARCE, Tacyana. O lead automatizado: uma possibilidade de tratamento da informação para o jornalismo impresso diário. **Revista Exacta**, Belo Horizonte, v. 2, n. 3, 2009.

AUTOMATED INSIGHTS. 2013. **Site Internet**. Disponível em: <www.automatedinsights.com>. Acesso em: 10 jan. 2013.

BARBOSA, Suzana; TORRES, Victor. O paradigma 'Jornalismo Digital em Base de Dados': modos de narrar, formatos e visualização para conteúdos. *Galaxia* (São Paulo, Online), n. 25, p. 152-164, jun. 2013.

BARBOSA, Suzana. Jornalismo em ambientes dinâmicos: perspectivas, tendências e desafios para a criação de conteúdos em tempos de convergência. In: *Actas III Congreso Internacional de Ciberperiodismo y Web 2.0*. Bilbao, Espanha: Universidad del País Vasco, 2011.

_____. Modelo JDBD e o ciberjornalismo de quarta geração. In: FLORES VIVAR, J. M.; RAMÍREZ, F. E. (Ed.). *Periodismo Web 2.0*. Madrid: Editorial Fragua, 2009. p. 271-283.

_____. Modelo Jornalismo Digital em Base de Dados (JDBD) em Interação com a Convergência Jornalística. In: *Textual & Visual Media*. Revista de la Sociedad Española de Periodística. vol. 1, Madrid, 2008, p. 87-106.

_____. *Jornalismo Digital em Base de Dados (JDBD) – um paradigma para produtos jornalísticos digitais dinâmicos*. (Tese de Doutorado). PósCOM/UFBA, 2007. Disponível em: <<http://migre.me/aTuYN>>. Acesso em: 4 fev. 2012.

BIG TEN NETWORK. 2014. **Portal Internet**. Disponível em: <www.btn.com>. Acesso em: 12 abr. 2014.

BRADSHAW, Paul; ROHUMAA, Liisa. **The online journalism handbook:**

skills to survive and thrive in the digital age. Essex: Pearson Education, 2011.

BRANCH, John. Snow Fall: the avalanche at Tunnel Creeak. **The New York Times**, New York, (200-]). Disponível em: <<http://www.nytimes.com/projects/2012/snow-fall/?forcedirect=yes#/?part=tunnel-creek>>. Acesso em: 2 jun. 2014.

CARLSON, Matt. The Robotic Reporter: automated journalism and the redefinition of labor, compositional forms and journalistic authority. In: LEWIS, Seth C. (org.). **Digital Journalism**. Vol. 3, nº 3. New York: Taylor&Francis Online, 2014.

CASTELLS, Manuel. **A sociedade em rede**. São Paulo: Paz e Terra, 1999.

CLERWALL, Christer. Enter the Robot Journalist: User's perception of automated content. In: **Journalism Practice**. Special Issue – Future of Journalism in an age of digital media and economic uncertainty. Volume 8, Issue 5. New York: Taylor&Francis Online, 2014.

COPPIN, Ben. **Inteligência artificial**. Rio de Janeiro: LTC, 2010.

DALEN, Arjen. The Algorithms Behind the Headlines: How machine-written news redefines the core skills of human journalists. In: **Journalism Practice**. Volume 6, Issue 5-6. New York: Routledge, 2012.

DEVAUX, Pierre. **Autômatos, automatismo e automatização**. Tradução Luis Borges Coelho. Lisboa: Editorial Gleba, 1964. (Coleção Horizonte, n.3).

ECO, Humberto. **Apocalípticos e integrados**. São Paulo: Perspectiva, 2006.

ELLUL, Jacques. **A técnica e o desafio do século**. Rio de Janeiro: Paz e Terra, 1968.

FEENBERG, Andrew. E-book. **Transforming technology**: a critical theory revisited. New York: Oxford University Press, 2002.

_____. E-book. **Between reason and experience**. Essays in technology and modernity. Cambridge, MA: Mit Press, 2010.

FIDALGO, Antônio. A resolução semântica no jornalismo online. In: BARBOSA, S. (Org.). **Jornalismo digital de terceira geração**. Covilhã, PT: LivrosLabCOM, 2007. p. 93-102.

_____. Do poliedro à esfera: os campos de classificação. A resolução semântica no jornalismo online. In: **Anais II Encontro Nacional da SBPjor**. Salvador-BA/Brasil, 2004.

GRAEFE, Andreas. **Guide to Automated Journalism**. Tow Center for Digital Journalism. Janeiro, 2016. Disponível em <https://www.gitbook>.

com/book/towcenter/guide-to-automated-journalism/details . Acessado em 20/01/2016.

HAACK, Bregtje; PARKS, Michael; CASTELLS, Manuel. The Future of Journalism: Networked Journalism. In: **Internacional Journal of Communication**. Vol. 6. 2012.

LAGE, Nilson. O lead clássico como base para a automação do discurso informativo. In: CONGRESSO BRASILEIRO DE PESQUISADORES DA COMUNICAÇÃO INTERCOM, 20., 1997, Santos. **Anais...** Santos, SP. 1997.

LATAR, Noam. The Robot Journalism in the Age of Social Physics: The end of human journalism? In: **The New World of Transitioned Media**. Springer, 2015.

LEMONS, André. **Cibercultura**: tecnologia e vida social na cultura contemporânea. 4. ed. Porto Alegre: Sulina, 2002.

LEWIS, Seth; USHER, Nikki. Code, Collaboration and The Future of Journalism: A case study of the Hacks/Hackers global network. In: **Digital Journalism**. Routledge Online, 2014.

LUTICE CRÉATIONS. **Site Internet**, Paris, [2000-]. Disponível em: <<http://www.automates-boites-musique.com/>>. Acesso em: 7 abr. 2014.

KNIGHT, Megan; COOK, Clare. **Social media for journalists**: principles e practice. Londres: Sage, 2013.

MACHADO, Elias. **O ciberespaço como fonte para os jornalistas**. Salvador: Calandra, 2003.

_____. Elias. O Jornalismo Digital em Base de Dados. Florianópolis: Calandra, 2006.

MIELNICZUK, Luciana. **Características e implicações do jornalismo na web**. 2001. Disponível em: <http://200.18.45.42/professores/chmoraes/comunicacao-digital/13-2001_mielniczuk_caracteristicasimplicacoes.pdf>. Acesso em: 8 set. 2010.

MOROZOV, Evgeny. **A robot stole my Pulitzer!**: future tense. 2012. Disponível em: <http://www.slate.com/articles/technology/future_tense/2012/03/narrative_science_robot_journalists_customized_news_and_the_danger_to_civil_discourse_.html>. Acesso em: 11 abr. 2014.

NARRATIVE SCIENCE. 2010. **Site Internet**. Disponível em: <www.narrativescience.com>. Acesso em: 10 jan. 2013.

PORTAL TERRA. Esporte. 2014. **Portal Internet**. Disponível em: <<http://esportes.terra.com.br/futebol/brasileiro-serie-a>>. Acesso em: 31 maio 2014.

RAMOS, Daniela. Formato: condição para a escrita do Jornalismo Digital em Bases de Dados. Uma contribuição da semiótica da cultura. (Tese de

Doutorado). ECA/USP, 2011. Disponível em: <<http://migre.me/aTvzX>>. Acesso em: 30 abril 2015.

RÜDIGER, Francisco. **Introdução às teorias da cibercultura: tecnocracia, humanismo e crítica no pensamento contemporâneo**. 2. ed. Porto Alegre: Sulina, 2007.

SANTOS, M. Journalism and the Internet of Things: can raw data change everything, again? In: **International Conference on Integrated Journalism Education, Research and Innovation (Resumos)**. Integrated Journalism in Europe – IJE, 2015, Barcelona, pag. 59. Disponível em: http://eventia.estaticos.econgres.es/2015_IJE/BookProgramme.pdf. Acessado em 07/01/2015.

SENNETT, R. **O artífice**. Rio de Janeiro: Record, 2009.

SORIA, Carlos. **Convergência de mídias**. 2014. Palestra apresentada ao Seminário sobre Integração Multimídia, São Luís, 2014.

NOTAS

- 1 A ideia do ciberiluminismo está relacionada à visão extremamente positiva e às vezes até ingênua sobre a relação entre tecnologia e seres humanos, representada normalmente por suas características inovadoras, gerando transformações capazes de criar um mundo melhor e mais justo. Já o ludismo vem de Ned Ludd, supostamente, operário que liderou um movimento que pregava a destruição das máquinas nas tecelagens inglesas porque essas reduziam os postos de trabalho. Alguns autores citam Ludd como um personagem criado pelo movimento operário da época para facilitar a propagação da campanha contra a automatização do processo fabril no início da Revolução Industrial.
- 2 “Technical creation involves interaction between reason and experience. Knowledge of nature is required to make a working device. This is the element of technical activity we think of as rational. But the device must function in a social world, and the lessons of experience in that world influence design.” – Tradução nossa.
- 3 A história da boneca pode ser conhecida no documentário “*L’Androïde de Marie-Antoinette*”, disponível em: <http://www.youtube.com/watch?v=pSxWmJLAaEg>.

- 4 Mielnickzuk (2001) nos fala das fases de jornalismo digital, chamando a primeira de fase transpositiva, justamente porque o conteúdo do impresso era apenas copiado para a internet sem grandes alterações.
- 5 Metadados são dados sobre outros dados. Informações como o autor do texto, a data em que foi escrito ou os registros de todas as suas versões, bem como a classificação ou *tags* indicando a que editoria pertencem, são exemplos de metadados que normalmente são adicionados às matérias jornalísticas a partir dos softwares de gerenciamento e publicação de conteúdo hoje já comuns nas redações.
- 6 <http://narrativescience.com/>
- 7 “In parallel to the emergence of the new field of “social physics”, narration, the art of telling stories, is also becoming a scientific endeavor employing artificial intelligence algorithms taking advantage of the vast body of knowledge of the field of linguistics and the study of natural language. AI algorithms are being composed that can convert facts into readable stories in a fraction of a second.” Tradução do autor. A biblioteca NLTK (*Natural Language Toolkit*) que utilizamos no experimento descrito nesse artigo é um exemplo de software de linguagem natural.
- 8 www.python.org.
- 9 Projetos envolvendo programação e jornalismo têm sido desenvolvidos, com exemplos na área do Jornalismo Investigativo, no intuito de extrair e processar dados em grandes quantidades e utilizar essas informações para a construção de infográficos e narrativas no jornalismo digital. <http://gijn.org>.
- 10 www.nltk.org.
- 11 Atualmente o endereço da tabela é <http://esportes.terra.com.br/futebol/brasileiro-serie-a/tabela>.
- 12 In recent years, the use of algorithms to automatically generate news from structured data has shaken up the journalism industry—most especially since the Associated Press, one of the world’s largest and most well-established news organizations, has started to automate the production of its quarterly corporate earnings reports. Once developed, not only can algorithms create thousands of news stories for a particular topic, they also do it more quickly, cheaply, and po-

tentially with fewer errors than any human journalist. Unsurprisingly, then, this development has fueled journalists' fears that automated content production will eventually eliminate newsroom jobs, while at the same time scholars and practitioners see the technology's potential to improve news quality. Tradução do autor.

- 13 “*Resistance is futile*”, frase repetida pelos *Borgs* da série *Star Trek* para suas vítimas (tradução nossa).

Márcio Carneiro dos Santos é jornalista e doutor pelo programa de Tecnologias da Inteligência e Design Digital da PUC-SP, com pós-doutorado na UNB, dentro da linha de pesquisa em Teorias e Tecnologias da Comunicação. Mestre em Comunicação pela UAM-SP e MBA em Marketing pelo ISAN-FGV-Rio. Atualmente é professor adjunto do Departamento de Comunicação Social da UFMA na área de Jornalismo em Redes Digitais e coordena o Laboratório de Convergência de Mídias - LABCOM/UFMA. Tem trabalhos publicados nas áreas de TV Digital, Teoria de Redes, Análise de Redes Sociais, Filosofia da Tecnologia e Inovação no Jornalismo. Bolsista de Produtividade DT-II do CNPq em Desenvolvimento Tecnológico e Extensão Inovadora.

RECEBIDO EM: 12/12/2015 | ACEITO EM: 27/01/2015