



Textos gerados por Software – surge um novo gênero jornalístico?¹

*Texts generated by Software –
does a new journalistic genre begin?*

Márcio Carneiro de Santos

Doutor, Universidade Federal do Maranhão (UFMA), São Luís, MA, e-mail: mcszen@gmail.com

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.

¹ Trabalho apresentado no GP Gêneros Jornalísticos do XIV Encontro dos Grupos de Pesquisa em Comunicação, evento componente do XXXVII Congresso Brasileiro de Ciências da Comunicação.

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 possam 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: Gêneros. Narrativas. Inteligência Artificial. Software. Ciberjornalismo.

Abstract

Describes the experiment of building a software able to generate leads and newspaper titles in an automated manner from information obtained from the Internet. The theoretical possibility that Lage already provided by the end of last century, is based on relatively rigid and simple structure of this type of story building, which facilitates the representation or translation of its syntax in accordance with instructions that the computer can execute. Also discusses the relationships between society, technology and technique, 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 the Natural Language Toolkit library – NLTK – and used the results of the Brazilian Football Championship 2013 published in the Internet portal as a data source.

Keywords: Genres. Narratives. Artificial Intelligence. Software. Cyberjournalism.

Tecnologia – temor e fascínio, no mundo real e na ficção

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

Em sua revisão das visões simplistas sobre o tema, geralmente representadas por defensores e críticos do papel da tecnologia no mundo de hoje, Feenberg propõe inicialmente uma espécie de mapeamento dessas posições e, a partir dele, 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, pode-se sugerir também, desenvolvedores de software.

Ao fazer tal constatação, o autor remete-se ao pensamento de Marx que já no século XIX criticava a ideia de que a economia pudesse ser apenas regida por fatores extrapolíticos, por meio 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 a dualidade simplificadora entre uma tecnologia benéfica ou perigosa para o homem parece estar superada, o cinema ao longo de décadas tem traduzido esse imaginário de medo e fascínio em diversos filmes em que 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 alie-

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.

nígenas já enfrentadas foi a dos *Borgs*, raça de seres híbridos, biológicos e mecânicos, que rapidamente assumiam o controle das áreas que invadiam, a partir da conversão dos seres que encontravam em sua própria espécie, por meio 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 se estabelece com celulares, *smartphones*, *tablets* e tantos outros *gadgets* tecnológicos dos quais não se quer mais se separar.

Se na antiguidade e na idade média os relatos sobre autômatos eram restritos, o século XVIII é 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, [2000-].

³ A história da boneca pode ser conhecida no documentário “*L’Androïde de Marie-Antoinette*”, disponível em: <<http://www.youtube.com/watch?v=pSxWmJLAaEg>> .

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, ensina que duas posições se estabeleceram para compreender o que estava acontecendo. A primeira, que se poderia 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, capaz de impactar todas as etapas do processo de produção como também as habilidades necessárias para exercer a função de jornalista, os modelos de negócio dessa cadeia produtiva e os próprios papéis desempenhados tradicionalmente por emissores e receptores em relação aos veículos de massa.

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), no histórico que traçam sobre o início do jornalismo *online* no ocidente, 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. Era o *Electronic Telegraph*. Em 1993, o primeiro *browser*⁵, o *Mosaic*, havia sido lançado. *BBC Online* em 1997 e *Guardian Unlimited* em 1999

⁴ Mielnickzuk (2001) 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 Um *browser* é um software cliente de Internet que solicita ao servidor as páginas que o usuário assinala por meio do endereço que digita.

são destaques numa lista de iniciativas ligadas à ideia de levar a atividade jornalística para o ciberespaço.

Na época havia uma grande confiança entre as empresas de mídia de que a Internet seria apenas mais um veículo, mais um espaço a conquistar, uma nova fronteira, em que vendendo publicidade e utilizando os modelos de negócio tradicionais todos poderiam prosperar. Tal certeza atraiu muitos investimentos e novas empresas “*dot.com*” surgiram rapidamente, mas, em 2001, com a crise que ficou conhecida como o estouro da bolha da Internet, percebeu-se que não era tão simples assim lucrar com as iniciativas digitais e que, pelo contrário, o que estava começando era uma corrida pela sobrevivência, em que apenas os que se adaptassem ao novo cenário de forma mais eficiente poderiam continuar.

Passaram-se os anos e novas tecnologias foram continuamente sendo incorporadas ao fazer jornalístico. As bases de dados, a integração de múltiplas mídias para contar uma única história, a capacidade de customizar e segmentar o conteúdo em função dos interesses de usuários cada vez mais exigentes e difíceis de atrair. Surgiram os sistemas de CMS (*Content Management Systems*) que permitiram aos jornalistas publicar diretamente seu conteúdo sem a intermediação de um programador ou especialista em HTML, a linguagem que organiza os elementos de qualquer página na web e que os *browsers* utilizam para construir o que os leitores veem em seus computadores.

Chamar os tradicionais consumidores de notícias de leitores também não é mais tão preciso. A digitalização, o barateamento dos equipamentos para produzir imagens e som, a expansão da infraestrutura da Internet e a ubiquidade dos dispositivos móveis fez dos cidadãos produtores de conteúdo, dando a eles um espaço crescente no processo de produção jornalística e constituindo o que alguns chamam de *user generated content* (UGC), conteúdo gerado por usuários, e também de jornalismo participativo, termo que traduz uma série de iniciativas com escopo e dimensão diversos⁶, indo do jornalismo produzido por ou para pequenas comunidades até grandes iniciativas que, via Internet, ganham alcance internacional.

O impacto da tecnologia no jornalismo também obrigou a revisão de alguns conceitos clássicos como o da pirâmide invertida e do lead. A necessidade da atualização constante e a pressão do tempo criaram novas formas narrativas nas quais a notícia é construída em camadas, a partir das unidades de informação que vão se tornando disponíveis, sendo conectadas pelos *hiperlinks* e cuja estrutura pode ir de materiais praticamente brutos, sem qualquer edição, a pacotes completos do jornalismo tradicional incluindo análises, desdobramentos e contextualização. O conceito de resolução semântica de Fidalgo (2003) descreve o processo, fazendo uma analogia com as imagens digitais que, a partir do aumento do número de

⁶ Ver em Knight e Cook (2013) a distinção entre os dois conceitos.

*pixels*⁷ que as formam, permitem gradualmente melhor visualização e compreensão. Na redação digital, as partículas de informação chegam em fluxo contínuo e com elas se constroem histórias, iniciando às vezes apenas com poucas palavras na área de “últimas notícias” e, quando merecem, chegando às grandes reportagens contadas no ambiente digital como a premiada Snow Fall (BRANCH, [2000-]), do *New York Times*.

Mesmo com a necessidade de aprendizado em novas habilidades e devendo agora ser capaz de trafegar por tecnologias, equipamentos e conceitos como o de SEO⁸ e SMO, a figura do jornalista, reconfigurada pela tecnologia ainda permanece essencial como elemento do processo de produção; mas em 2010, os primeiros exemplos de algo mais radical começaram a aparecer.

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*⁹, 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, em que 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.

⁷ O conceito de pixel parte da ideia de que as imagens digitais são formadas por matrizes de pontos que definem a resolução da tela e traduz a menor unidade constituinte da representação das imagens quando são gerenciadas por computadores.

⁸ Sigla para *Search Engine Optimization*, conjunto de técnicas para melhorar o posicionamento da página de Internet em mecanismo de busca como Google. SMO (*Social Media Optimization*) seria o equivalente para as redes sociais.

⁹ Disponível em: <<http://narrativescience.com/>>.

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 (story packages were made public, but it was not the first disclosure) (Source: Wikimedia Commons)

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."

Figura 2 - Print de matéria sobre jornalismo automático

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.

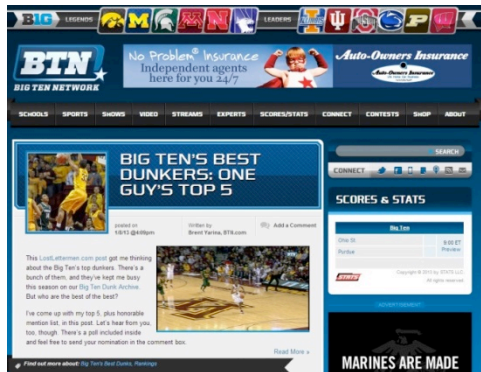


Figura 3 - Portal Big Ten Network (cliente)

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.

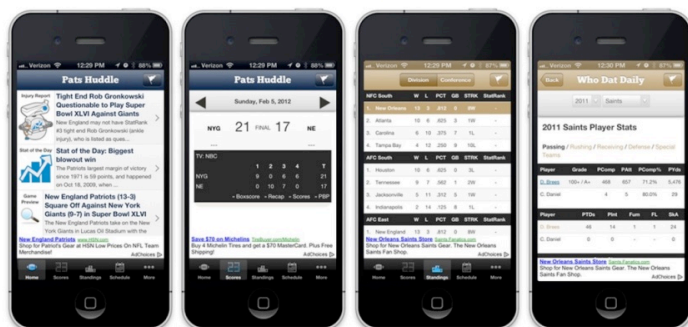


Figura 4 - Exemplos de conteúdo publicado por AI em plataformas móveis

Fonte: *Automated Insights*, 2013.

Arce (2009) já havia feito uma avaliação da possibilidade de automatização dos discursos incluindo aí 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”.

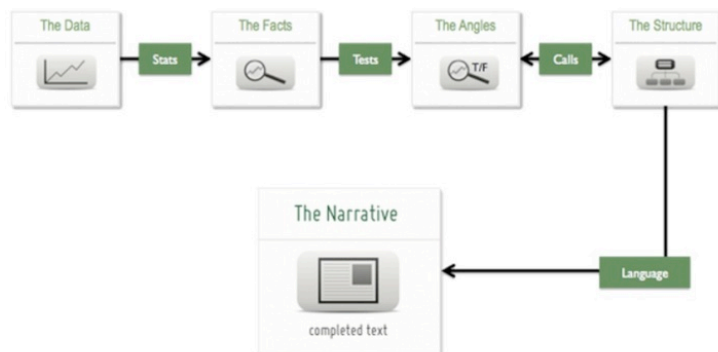


Figura 5 - Processo de transformação de dados brutos em narrativas, utilizado pela NS.

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 ser realizadas por uma máquina.

Ao se pensar no conceito de gênero como resultado de uma tentativa de organização ou estruturação de determinado conjunto de produção simbólica a partir do agrupamento em categorias com características comuns, pode-se, ainda que de forma inconclusiva, relacionar os fatos acima ao possível surgimento de um novo gênero ciberjornalístico com um traço bastante característico: não ser gerado por humanos. O experimento descrito a seguir tenta aprofundar as consequências dessa possibilidade, ainda que em caráter exploratório.

Modelagem de experimento com resultados do futebol

Para construir nosso experimento de narrativa automatizada, propõe-se a desenvolver um código de programação capaz de escrever pequenos textos sobre os resultados do campeonato brasileiro de futebol de 2013. Utiliza-se a linguagem de programação Python¹⁰ por considerá-la de mais fácil aprendizagem para não especialistas em programação como jornalistas e profissionais da comunicação¹¹.

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 NLTK¹² – *Natural Language Toolkit* – que é utilizada 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;

¹⁰ Disponível em: <www.python.org>.

¹¹ 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. Disponível em: <<http://gijn.org/>>.

¹² Disponível em: <www.nltk.org>.

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, parte-se para uma solução que, a partir de um endereço específico na Internet em que esses dados fossem disponibilizados, conseguia ler todas as informações iniciais de forma automática e mais rápida. Para os testes, selecionou-se 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)¹³. 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, fez-se a parte do código que salvava esses elementos, associando-os a cada time numa estrutura que na linguagem Python é conhecida por dicionário, em que a cada elemento, chamado de chave, são associados valores diversos, cada um representando alguma informação gerada a partir dos resultados dos jogos.

```

CBPontos5.py - D:\Python\ParserTestes\CBPontos5.py
File Edit Format Run Options Windows Help
from BeautifulSoup import BeautifulSoup
from operator import itemgetter
import urllib, urllib2
input=urllib.url2opener("http://esportes.terra.com.br/futebol/brasileiro-serie-a")
data=str(input.read())
input.close()
times = {u"Atlético-MG": [0,0,0,0,0,0,0,0], u"Vasco": [0,0,0,0,0,0,0,0], u"Fluminens
u"Cruzeiro": [0,0,0,0,0,0,0,0], u"Internacional": [0,0,0,0,0,0,0,0], u"São
u"Ponte Preta": [0,0,0,0,0,0,0,0], u"Flamengo": [0,0,0,0,0,0,0,0], u"Coriti
u"Santos": [0,0,0,0,0,0,0,0], u"Bahia": [0,0,0,0,0,0,0,0], u"Vitória": [0,0,

```

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

Fonte: Autor, 2014.

¹³ Atualmente, o endereço da tabela é <<http://esportes.terra.com.br/futebol/brasileiro-serie-a/tabela>>.

Ao iniciar o código, o usuário é demandado apenas a escolher o número da rodada que deseja explorar. A partir daí, 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 acrescentadas 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.

A parte do texto também está associada a essa lógica. Foi criada uma lista de verbos a ser usada de acordo com o contexto do resultado. Assim “vence”, “ganha”, “bate” e outras construções semelhantes são escolhidas de forma aleatória pelo *software*. Quando a diferença de gols entre o vencedor ou o perdedor é maior que dois, indicando uma forte superioridade no placar, outra lista de verbos é acionada com opções como “arrasa”, “liquida” e até “humilha”, pensando num texto mais sensacionalista. Assim, o título é gerado concatenando as informações básicas que já existem, no caso, o nome dos times e o placar, como em “Flamengo vence Botafogo no Engenhão por 3 a 2”. Considerando que o *software* capturou da Internet os times envolvidos e o resultado, bem como o lugar onde aconteceu, basta comparar os gols do placar, definir o vencedor e daí escolher um dos verbos da lista disponível, concatenando todos esses elementos numa estrutura simples.

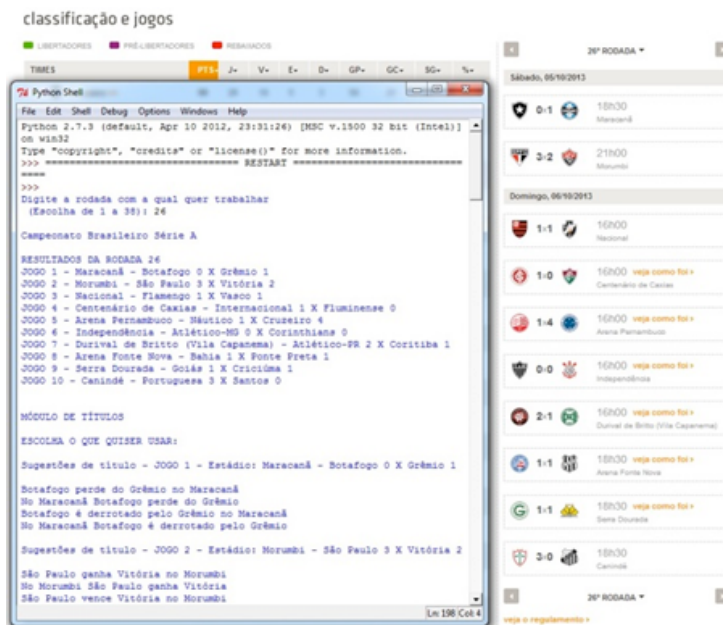


Figura 7 - Tela que compara a página do portal com os resultados e a tela gerada pelo programa em que podem ser vistos primeiro os dados registrados e depois as sugestões de título baseadas nos resultados

Fonte: Autor, 2014.

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.

Como exemplo, se um time disputou 10 jogos, ou seja, 30 pontos, valor que teria se vencesse todas as partidas e de fato tem apenas 3, uma vitória e um empate ou três empates, seu rendimento seria de apenas 10%, ou seja, bastante baixo.

Esses números permitem ao *software* escrever textos com mais informações.

classificação e jogos

LIBERTADORES	PRÉ-LIBERTADORES	RESERVADOS											
TIMES	Pts	J	V	E	D	GP	GC	SG	%				
1 Cruzeiro >>	59	27	18	5	4	50	23	35	72				
2 Grêmio >>	48	27	14	6	7	33	23	10	59				
3 Botafogo >>	46	27	13	7	7	40	31	9	56				
4 Atlético-PR >>	45	27	12	9	6	45	35	10	55				
5 Atlético-MG >>	39	27	10	9	8	32	28	4	48				
6 Vitória >>	37	27	10	7	10	38	40	-2	45				
7 Flamengo >>	37	27	9	10	8	33	32	1	45				
8 Internacional >>	37	27	9	10	8	39	39	0	45				
9 Goiás >>	4	37	27	9	10	8	30	33	-3	45			
10 Santos >>	3	36	27	9	9	32	29	3	44				
11 Bahia >>	3	36	27	9	9	29	32	-3	44				
12 Corinthians >>	3	36	27	8	12	7	22	17	5	44			
13 Portuguesa >>	3	34	27	9	7	11	41	40	1	41			
14 Fluminense >>	3	34	27	9	7	11	31	34	-3	41			
15 Coritiba >>	34	27	8	10	9	30	35	-5	41				
16 São Paulo >>	33	27	9	8	12	28	29	-3	40				
17 Vasco >>	32	27	8	8	11	36	42	-6	39				
18 Criciúma >>	29	27	8	5	14	34	45	-11	35				
19 Ponte Preta >>	26	27	7	5	15	28	40	-12	32				


```

Python Shell
File Edit Shell Debug Options Windows Help
PROCESSANDO RESULTADOS DA RODADA 22
PROCESSANDO RESULTADOS DA RODADA 23
PROCESSANDO RESULTADOS DA RODADA 24
PROCESSANDO RESULTADOS DA RODADA 25
PROCESSANDO RESULTADOS DA RODADA 26
[{"time": "Cruzeiro", "pts": 59, "j": 27, "v": 18, "e": 5, "d": 4, "gp": 50, "gc": 23, "sg": 35, "%": 72, "l": 1}, {"time": "Grêmio", "pts": 48, "j": 27, "v": 14, "e": 6, "d": 7, "gp": 33, "gc": 23, "sg": 10, "%": 59, "l": 2}, {"time": "Botafogo", "pts": 46, "j": 27, "v": 13, "e": 7, "d": 7, "gp": 40, "gc": 31, "sg": 9, "%": 56, "l": 3}, {"time": "Atlético-PR", "pts": 45, "j": 27, "v": 12, "e": 9, "d": 6, "gp": 45, "gc": 35, "sg": 10, "%": 55, "l": 4}, {"time": "Atlético-MG", "pts": 39, "j": 27, "v": 10, "e": 9, "d": 8, "gp": 32, "gc": 28, "sg": 4, "%": 48, "l": 5}, {"time": "Vitória", "pts": 37, "j": 27, "v": 10, "e": 7, "d": 10, "gp": 38, "gc": 40, "sg": -2, "%": 45, "l": 6}, {"time": "Flamengo", "pts": 37, "j": 27, "v": 9, "e": 10, "d": 8, "gp": 33, "gc": 32, "sg": 1, "%": 45, "l": 7}, {"time": "Internacional", "pts": 37, "j": 27, "v": 9, "e": 10, "d": 8, "gp": 39, "gc": 39, "sg": 0, "%": 45, "l": 8}, {"time": "Goiás", "pts": 4, "j": 37, "v": 27, "e": 9, "d": 10, "gp": 8, "gc": 30, "sg": 33, "%": -3, "l": 9}, {"time": "Santos", "pts": 3, "j": 36, "v": 27, "e": 9, "d": 9, "gp": 32, "gc": 29, "sg": 3, "%": 44, "l": 10}, {"time": "Bahia", "pts": 3, "j": 36, "v": 27, "e": 9, "d": 9, "gp": 29, "gc": 32, "sg": -3, "%": 44, "l": 11}, {"time": "Corinthians", "pts": 3, "j": 36, "v": 27, "e": 8, "d": 12, "gp": 7, "gc": 22, "sg": 17, "%": 5, "l": 12}, {"time": "Portuguesa", "pts": 3, "j": 34, "v": 27, "e": 9, "d": 7, "gp": 11, "gc": 41, "sg": 40, "%": 1, "l": 13}, {"time": "Fluminense", "pts": 3, "j": 34, "v": 27, "e": 9, "d": 7, "gp": 11, "gc": 31, "sg": 34, "%": -3, "l": 14}, {"time": "Coritiba", "pts": 34, "j": 27, "e": 8, "d": 10, "gp": 30, "gc": 35, "sg": -5, "%": 41, "l": 15}, {"time": "São Paulo", "pts": 33, "j": 27, "e": 9, "d": 8, "gp": 12, "gc": 28, "sg": 29, "%": -3, "l": 16}, {"time": "Vasco", "pts": 32, "j": 27, "e": 8, "d": 11, "gp": 36, "gc": 42, "sg": -6, "%": 39, "l": 17}, {"time": "Criciúma", "pts": 29, "j": 27, "e": 8, "d": 5, "gp": 14, "gc": 34, "sg": 45, "%": -11, "l": 18}, {"time": "Ponte Preta", "pts": 26, "j": 27, "e": 7, "d": 5, "gp": 15, "gc": 28, "sg": 40, "%": -12, "l": 19}]]
RESUMO DOS CLUBES DEPOIS DE 27 RODADAS:
Resumo Cruzeiro : O Cruzeiro tem 59 pontos. Com 18 vitórias, 5 empates, 4 derrotas. Marcou 50 gols. Sofreu 23 gols. Saldo de 35 gols e aproveitamento de 72.83 %.
Resumo Grêmio : O Grêmio tem 48 pontos. Com 14 vitórias, 6 empates, 7 derrotas. Marcou 33 gols. Sofreu 23 gols. Saldo de 10 gols e aproveitamento de 59.25 %.
Resumo Botafogo : O Botafogo tem 46 pontos. Com 13 vitórias, 7 empates, 7 derrotas. Marcou 40 gols. Sofreu 31 gols. Saldo de 9 gols e aproveitamento de 56.79 %.
Resumo Atlético-PR : O Atlético-PR tem 45 pontos. Com 12 vitórias, 9 empates, 6 derrotas. Marcou 45 gols. Sofreu 35 gols. Saldo de 10 gols e aproveitamento de 55.

```

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: Autor, 2014.

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 lanternas do torneio, aspectos que normalmente são enfatizados em noticiário desse tipo. É importante ressaltar que a tabela do site é aqui utilizada apenas para validar os resultados já que o *software* retira da Internet somente os jogos, seus resultados e o local onde aconteceram. O resto ele mesmo calcula, baseado nas regras do torneio que estão representadas para ele em termos de instruções de programação.

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 em que 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.

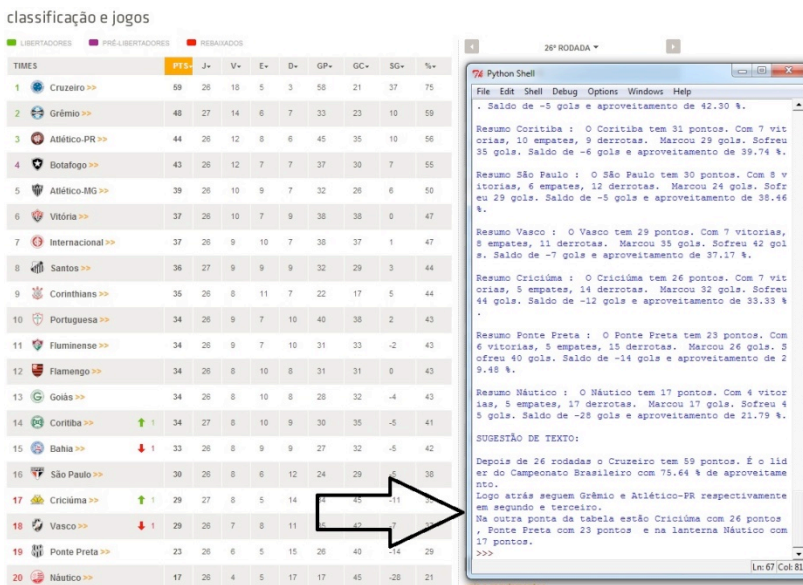


Figura 9 - Tela do software com o que seria o lead construído a partir das informações lidas sobre o campeonato em determinada rodada

Fonte: Autor, 2014.

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 reduzidas mais facilmente a expressões matemáticas, baseadas 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 se faz com os resultados do Campeonato Brasileiro, não seria difícil construir algo semelhante para gerar pequenos textos informando as variações do câmbio ou de ações em bolsas de valores, a previsão do tempo para cidades ou regiões e outros conteúdos que, se observados, são construídos com uma estrutura que se repete com pequenas variações.

A capacidade de coletar e processar informações com 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.

Formas mais complexas de programas já estão em uso comercial como nos exemplos da *Narrative Science* e da *Automated Insights*. Os desdobramentos desse tipo de tecnologia no mercado ainda não podem ser avaliados. É fato que muitas redações têm reduzido postos de trabalho e/ou se utilizado de mão de obra contratada de outras formas diferentes da relação formal registrada em carteira. Temporários, estagiários e colaboradores que são pagos como pessoa jurídica são situações comuns. Com o desenvolvimento de soluções mais sofisticadas de produção de conteúdo a partir de *software*, é possível considerar que novas alterações podem estar a caminho.

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.

Por outro lado, a precarização do trabalho e a replicação indiscriminada de re-leases e conteúdos gerados por fontes, justificados de forma simplista pela pressão do tempo e pela necessidade de atualização constante, são um risco para esses 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 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 terão papel importante nos impactos dessas novas tecnologias.

Se “resistir é inútil”¹⁴, 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 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 para que se estabeleça um novo gênero jornalístico. Entretanto, é interessante observar que o *software* alimenta-se 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.

¹⁴ “Resistance is futile”, frase repetida pelos *Borgs* da séria *Star Trek* para suas vítimas (tradução nossa).

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, simplesmente replicando conteúdos ou realizando de forma descuidada parte do seu trabalho. Esse parece ser o grande problema que se tem de enfrentar, sejam os homens céticos, temerosos ou fascinados por tecnologia.

Referências

ARCE, T. 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. Disponível em: <www.automatedinsights.com>. Acesso em: 10 jan. 2013.

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

BRADSHAW, P.; ROHUMAA, L. **The online journalism handbook: skills to survive and thrive in the digital age**. Essex: Pearson Education, 2011.

BRANCH, J. Snow Fall: the avalanche at Tunnel Creeak. **The New York Times**, New York, (2000-]. Disponível em: <<http://www.nytimes.com/projects/2012/snow-fall/?forcedredirect=yes#/?part=tunnel-creek>>. Acesso em: 2 jun. 2014.

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

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

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

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

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

FIDALGO, A. Sintaxe e semântica das notícias on-line. Para um jornalismo assente em base de dados. In: FIDALGO, A.; SERRA, P.(Org.). Informação e Comunicação Online. **Jornalismo Online**. v. 1. Covilhã: Universidade da Beira Interior/Portugal, 2003.

LAGE, N. 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.

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

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

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

MIELNICZUK, L. **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, E. **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. Disponível em: <www.narrativescience.com>. Acesso em: 10 jan. 2013.

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

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

Recebido: 20/10/2014

Received: 10/20/2014

Aprovado: 19/11/2014

Approved: 11/19/2014