

JORNALISMO E BIG DATA



Memorial

Protocolo de abordagem e suíte de ferramentas de coleta e análise de dados em grandes volumes para jornalistas e pesquisadores.

Resumo

Diante da situação de excesso de informação no ambiente digital, ao já difícil trabalho dos profissionais de jornalismo e pesquisadores da área, foi adicionado o desafio de lidar com aspectos de volume, variedade e velocidade na geração de dados, características associadas ao termo *big data*. Daí a necessidade de ferramentas capazes de automatizar a parte repetitiva dos processos de coleta e extração de material a partir de bases, sites, APIs e repositórios digitais de diversos formatos. Como resultado de um esforço de pesquisa aplicada, orientada à solução de problemas reais desta ordem, desenvolvemos um protocolo de abordagem e uma suíte de ferramentas para ajudar neste trabalho, entre elas, o LNEWS, para coleta de conteúdo em portais jornalísticos e o LTWEET para extração de conteúdo da plataforma Twitter.

1. APRESENTAÇÃO DO PRODUTO COM INTERFACES OU PROTÓTIPOS DO SISTEMA

A proposta de abordagem e ferramentas ora apresentadas procura facilitar o trabalho de jornalistas e pesquisadores que, cada vez mais, nas suas atividades diárias tem que lidar com massas de informações de diversas categorias e origens.



Figura 1 – Tela inicial do site labcomdata.com.br . Fonte: do autor.

Tal tipo de trabalho, sem o uso de recursos computacionais, é de acesso bem mais difícil.

O objetivo deste esforço de pesquisa aplicada é atender a uma demanda crescente de trabalho na área de jornalismo guiado por dados, entre profissionais, contemplando também diversas iniciativas de pesquisa no campo do Jornalismo, bem como da Comunicação em geral, mantendo o foco num conjunto de utilizadores mais distantes das práticas de programação e técnicas de raspagem de dados.

Em síntese, pesquisa aplicada é aquela orientada à solução de problemas reais, à inovação e ao desenvolvimento de artefatos, considerando um conjunto de beneficiários específico que vivencia tais problemas. No caso dos profissionais e pesquisadores do campo, apreender os movimentos do ecossistema de meios digitais, através de ferramentas que lhes permitam coletar e organizar esses dados de forma mais fácil e prática, economizando o seu tempo para as tarefas mais nobres de análise e interpretação, nos parece importante e fundamental diante do contexto contemporâneo.

Assim, nesta proposta para a candidatura se estabelecem aqui, o problema, os possíveis utilizadores, a tentativa de solução, os benefícios propostos e a contribuição para o desenvolvimento tanto de profissionais como pesquisadores, principalmente do jornalismo.

PROTOCOLO DE ABORDAGEM

Na imagem abaixo, apenas com finalidade didática, estabelecemos uma espécie de escala de utilização que mapeia os diversos níveis de ação para tratar a questão dos dados, conforme o nível de especialização e controle que se deseja em cada situação.

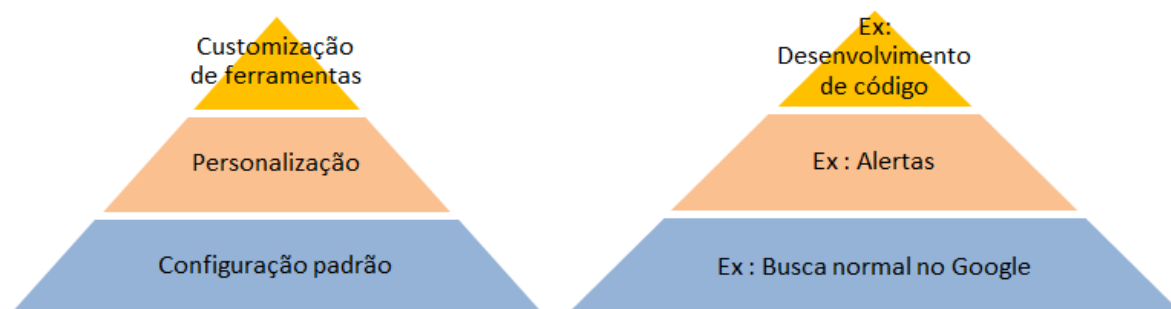


Figura 2: Representação da escala de utilização dos métodos digitais. Fonte: Do autor

Tal escala vai da utilização de ferramentas e técnicas já existentes em sua configuração padrão num nível inicial; com ajustes a fim de personalizá-las para atender nossas necessidades específicas, num nível médio; ou ainda, num nível mais alto, através da criação de soluções baseadas em programação e desenvolvimento de código.

Nas pirâmides acima exemplificamos a escala numa situação de coleta de dados que utiliza a busca do Google, inicialmente com sua interface normal, depois a partir de uma solução com maior poder de personalização como os alertas¹ e por fim através de um código específico para coletar e armazenar esses dados.

Procedimentos Básicos de Utilização

Em termos gerais, a abordagem que propomos resume-se às seguintes etapas:

Etapa 1 - Identificar a estrutura que contém os dados que precisamos. Algumas possibilidades apresentam-se com mais frequência:

¹ <https://www.google.com/alerts>

- a) Bases de Dados que permitem consultas amigáveis via preenchimento de formulários ou procedimentos simples. Exemplo: portais de transparência governamentais onde é possível requisitar dados sobre determinado tema e período.
- b) APIs² que exigem requisições estruturadas no formato que estabelecem, ou seja, respeitando sua sintaxe própria. Exemplo: APIs do *Twitter* e do *Facebook* que precisam ou de uma aplicação específica para solicitar conteúdo, como os aplicativos que as acessam em nossos celulares, ou de um código customizado que consiga estabelecer tal diálogo e coletar as informações que a API entrega a partir de cada tipo de requisição.
- c) Conteúdo disponível em páginas de internet que podem ser extraídos diretamente via técnicas de *scraping* (raspagem de dados)³. Como textos de matérias em portais jornalísticos ou tabelas e informações gerais publicadas, tais como previsão do tempo, cotação do dólar e resultados de competições esportivas.
- d) Informações protegidas em ambientes fechados, acessadas apenas por usuários cadastrados e que contam com mecanismos de proteção como encriptação de dados e outros. Tais ambientes eventualmente podem ser acessados por técnicas de *hacking* que estão além do escopo deste texto.

Etapa 2 – Formatar a consulta ou requisição de dados alinhada ao tipo de repositório onde eles se encontram de acordo com as opções acima.

Etapa 3 – Analisar os dados coletados a partir do processamento possível partindo do que foi efetivamente conseguido.

Podemos combinar então a escala de utilização da figura 2 com as quatro formas de estruturação de dados mais comuns numa tabela. A partir deste cruzamento, listamos algumas possibilidades de aplicação em seus três níveis.

² Uma API – *Application Programming Interface* (Interface de Programação de Aplicações) é o conjunto de rotinas, padrões e instruções de programação que permite a comunicação entre máquinas possibilitando que os desenvolvedores criem aplicações que possam acessar e interagir com determinado serviço na internet, inclusive extraindo dados dele.

³ Para ler mais sobre *scraping* para jornalistas recomendo ver Bradshaw, 2014

Estrutura dos Dados	Nível Inicial: Ferramentas Padrão	Nível Médio: Ferramentas com personalização	Nível Alto: Desenvolvimento de Código
a) Bases de Dados e Repositórios Acessíveis	Solicitação de dados através da própria interface da base de dados, recebendo o resultado no formato padrão de entrega. Ex: Acesso à base SIDRA do IBGE e download do arquivo no formato do Excel ou em CSV.	Utilização de filtros e recursos de análise e visualização oferecidos pela plataforma, alterando a forma de entrega do resultado de acordo com as opções oferecidas. Ex: Uso das funções avançadas da SIDRA e geração de gráficos.	Código para automatizar o acesso ao banco de dados fazendo requisições sucessivas, customizadas, coletando e salvando os registros em outro tipo de estrutura ou formato de dados. Ex: Python com módulos Splinter ou Selenium
b) Servidores com acesso via API específica	Acesso via aplicação oficial da plataforma ou através de sua página web padrão. Ex: Uso do app do Facebook no celular ou acesso à página www.facebook.com .	Acesso através de aplicativos de terceiros que também acessam o servidor da plataforma mas oferecem funcionalidades adicionais. Ex: Node XL.	Código para acessar diretamente a API da plataforma coletando todas as informações disponibilizadas por ela e também fazendo requisições sucessivas capazes de coletar volumes maiores de dados.
c) Conteúdo em páginas web com padrão HTML ⁴	Busca do Google, acesso manual e eventual coleta via CTRL+C e CTRL+V	Utilização de ferramentas específicas para scrapping. Ex: Portia	Desenvolvimento de código para coleta e análise. Ex: Python com módulo BeautifulSoup
d) Dados protegidos mediante acesso logado	Acesso via solicitação de cadastro e <i>log-in</i> normal.	Ferramentas de hacking geral tipo brute force ou engenharia social.	Desenvolvimento de códigos de invasão tipo worm ou trojan.

⁴ *HiperText Markup Language*

Tabela 1 – Matriz de possibilidades de coleta via métodos digitais em função da forma e local dos dados e dos níveis de aplicação. Fonte: do autor.

DESENVOLVIMENTO DE FERRAMENTAS

Links diretos de acesso:

LNEWS: https://colab.research.google.com/drive/1e9_FdP-7g3KXAjsqglqPz-D1ziQ8NpKMI

LTWEET:

<https://colab.research.google.com/drive/1wGMvNWqIq6jWCPnSOq8iMNFLpXrNGnpx>

Página do Site com tutoriais e links de acesso: labcomdata.com.br/teste

Vídeo Tutorial LTWEET: <https://youtu.be/z1BUTcxUmJ0>

Vídeo Tutorial LNEWS: <https://youtu.be/aTUzx-GVTR4>



Figura 3 – Página do site com acesso às ferramentas, tutoriais e demais informações. Fonte: do autor.

1.1 – LNEWS – Para coleta de conteúdo em sites jornalísticos

No seu atual estágio a ferramenta LNEWS permite ao pesquisador:

- a) Selecionar um site ou portal jornalístico, com 3 opções pré-selecionadas UOL, ESTADÃO e GLOBO.COM e uma quarta opção onde o usuário pode entrar com um endereço específico para, por exemplo, explorar algum veículo da sua cidade ou estado, com relevância

específica para sua pesquisa;

- b) A partir da escolha feita no item anterior o software coleta uma lista de links disponíveis para acesso no momento da consulta onde é possível ver a estrutura de organização das editorias/categorias que são utilizadas pelos emissores na publicação do seu conteúdo;
- c) Com a lista o usuário pode escolher qualquer uma das matérias recuperadas e vê-la na íntegra com alguns dos seus meta-dados, quando acessíveis, tais como seus autores. Ainda nesse momento é criada uma lista das palavras mais frequentes e feito um sumário (resumo) do texto. Essa última função ainda apresenta pouca eficiência no caso da língua portuguesa.
- d) Após essa etapa é oferecido ao pesquisador salvar sua coleta criando dois arquivos, um com a lista de links para as matérias e o segundo com os textos das mesmas, podendo realizar essa última ação adicionando filtros por editoria ou até pela seleção de matérias que apresentem uma palavra específica. Por exemplo, do total de matérias coletadas é possível salvar apenas as da editoria Política em que a palavra “poder” esteja presente.

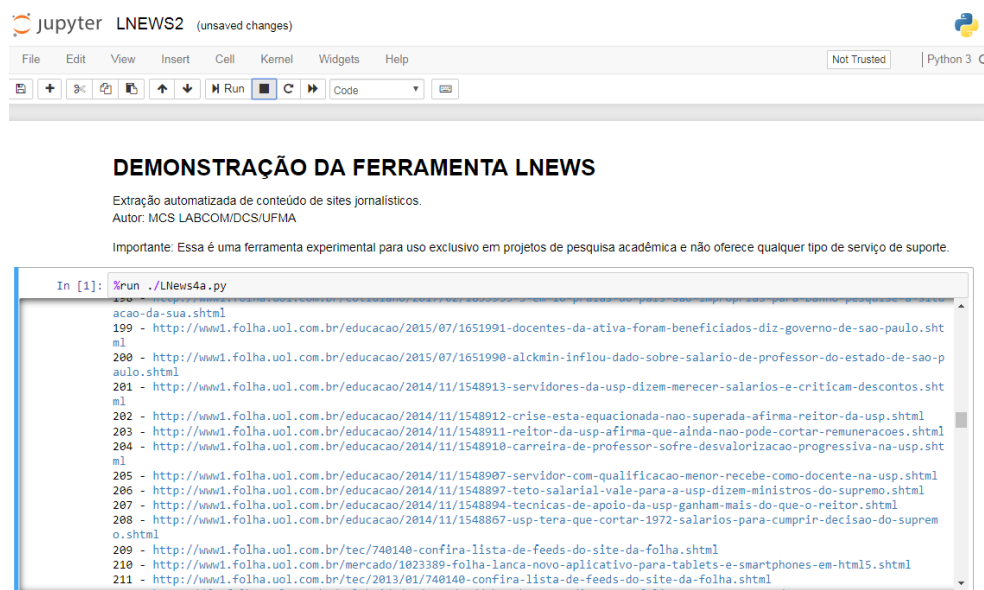


Figura 4 – Tela com lista de links extraídos. Fonte: do autor.

Outro módulo da solução, denominado LNEWS Análise, já trabalha não mais com os sites diretamente e sim com os arquivos de textos extraídos. A partir deles estão sendo desenvolvidas funções para calcular e visualizar métricas tais como a diversidade léxica, que traduz a riqueza na utilização das palavras do texto e as palavras frequentes que podem ser também listadas ou apresentadas num gráfico.

O trabalho com o LTWEET também já gerou frutos adicionais que temos colocado na página do LABCOM DATASETS, ou seja, na página disponibilizamos conjuntos de tweets já extraídos para consulta ou análise incluindo, entre outros, o do Governador do Maranhão, Flávio Dino e o do presidente Bolsonaro.



Figura 7 – Tela da página do site LABCOM DATASETS que oferece conjunto de dados já extraídos utilizando a ferramenta LTWEET. Fonte : do autor .

Hoje, apenas para controle, o acesso à página das ferramentas LNEWS e LTWEET é feito mediante cadastro em formulário que nos permite avaliar o grau de utilização e a diversidade dos usuários interessados nas mesmas. Atualmente quase 100 pesquisadores, de todas as regiões do país e representando diversos grupos de pesquisa e interesses diversos, já se cadastraram para utilizar.

Para alunos e profissionais de jornalismo também temos disponibilizado oportunidades de contato com as ferramentas através dos canais de comunicação do LABCOM bem como em eventos como Intercom Nordeste, Intercom Nacional e Encontro Nacional da ABCiber .



Figura 8 – Apresentação das ferramentas para alunos do Mestrado Acadêmico da UFMA em Imperatriz. Fonte: do autor.

O interesse pelas ferramentas entre pesquisadores e jornalistas também pode ser identificado através dos registros do Google Analytics do site do LABCOM – www.labcomdata.com.br

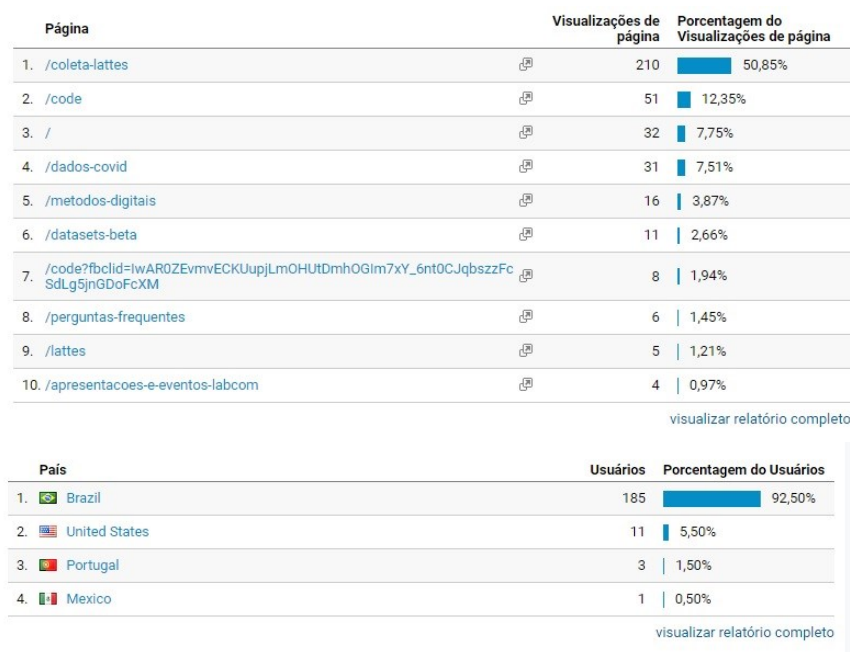


Figura 9 – Telas com métricas do Google Analytics a partir do site do LABCOM, indicando o crescimento do tráfego no site após o lançamento das ferramentas e os países de origem dos acessos. Fonte : do autor .

Além das duas ferramentas detalhadas aqui, a suíte de soluções do LABCOM também inclui a LCLattes que gera relatórios de produção para a plataforma Sucupira e a LQUALIS que apresenta as avaliações Qualis das revistas acadêmicas da CAPES.

Dados adicionais sobre a suíte de ferramentas desenvolvida pela LABCOM.

LNEWS — ferramenta de coleta de conteúdo em sites jornalísticos. Extrai textos, links e títulos das matérias filtradas por editoria ou palavra-chave.

LTWEET — ferramenta de coleta de tweets que podem ser filtrados por usuário ou palavra-chave.

Link de acesso para ambas: <https://www.labcomdata.com.br/teste> . Observação: é necessário se cadastrar em formulário na home do site e depois solicitar acesso de membro já que o acesso é feito logado no site.

LTEXTO — Faz análise de arquivos de texto gerando visualização de palavras frequentes e dispersão léxica. Ainda em desenvolvimento. Utilização mediante solicitação de link.

LQUALIS — Verifica o QUALIS Periódicos de publicações da área de Comunicação e Informação oferecendo a pontuação nas tabelas antiga e nova (ainda não oficial) e consulta em lote.

Link de acesso: <https://www.labcomdata.com.br/code>

LCLATTES — Gera relatórios detalhados sobre a produção dos pesquisadores com base nos dados do CNPq. Inclui hoje a produção acadêmica em revistas, trabalhos em eventos, livros e capítulos.

Link de acesso: <https://www.labcomdata.com.br/coleta-lattes>

Vídeos Tutoriais para as ferramentas são encontrados no canal do LABCOM no YOUTUBE em <https://www.youtube.com/channel/UCz9xhatSPQ9MZFPf1e9kDUQ/videos>

LABCOM - Teste para Novo QUALIS

Com e Info - Faixa A - B2

Importante : Trata-se de um teste apenas.

MARQUE A CONSULTA QUE DESEJA EXECUTAR :

Na pesquisa por NOME, selecione as revistas nas quais publicou.

Assim posso calcular se seu Lattes melhorou.

Mostrar Tabela Qualis

Pesquisar pelo ISSN:

Pesquisar pelo Nome da Revista:

ESCOLHA A REVISTA:

ESCOLHA:

▼ []
0 : "INTERCOM: REVISTA BRASILEIRA DE CIENCIAS DA COMUNICACAO"

Figura 10 – Tela da solução LQUALIS que também pertence à suíte de ferramentas do LABCOM.

Fonte: do autor.

Este conjunto de ferramentas também consolida o trabalho do autor e do LABCOM no campo da pesquisa aplicada, com orientação à solução de problemas reais de jornalistas e pesquisadores da

área, articulando teoria, conhecimento empírico e desenvolvimento de artefatos via software, num esforço também de inovação, considerando a área de Ciências Sociais Aplicadas e, mais especificamente, Comunicação e Informação.



Figura 11 – Exemplo de Registro de Software do LTWEET. O LNEWS e todos os outros também passaram pelo mesmo procedimento.

Fonte: INPI

2. REFLEXÃO TEÓRICA SOBRE O JORNALISMO

A condição contemporânea do ecossistema comunicacional, baseada em redes digitais, massivo crescimento de emissores e explosão informacional, caracterizada por volume, velocidade e variedade, tem sido devidamente registrada nos últimos vinte anos a partir dos trabalhos de Castells (1999), Chwe (2000), Feenberg (2002), Lemos (2002), Santaella (2003), Vilches (2003) entre outros.

Suas características de interconexão (NEWMAN, 2010; EASLEY, KLEINBERG, 2010), excesso (GLEICK, 2013) e complexidade (SIMON, 1962; MORIN, 2005), aliadas a um conjunto próprio de especificidades, que constituem uma verdadeira ontologia dos entes binários, também já mapeada por Manovich (2001) e Santos (2016), indicam um quadro consolidado de percepção de profundas mudanças que, entretanto, parece ainda não ter impactado de forma proporcional os métodos e técnicas utilizados pelos pesquisadores da Comunicação para analisar objetos do ciberespaço.

A iniciativa denominada Métodos Digitais, que tem em Rogers (2013) um de seus principais interlocutores, explicita a necessidade, gerada pelas especificidades dos entes de origem binária, de estudá-los a partir métodos e técnicas originados em sua lógica interna.

Transpondo tal raciocínio ao trabalho de pesquisa, Rogers também separa os métodos eminentemente digitais dos que ele denomina de virtuais, ou seja, que têm sua origem em outros campos e têm sido adaptados para a internet e as redes sociais. A netnografia ou etnografia virtual, os questionários aplicados via e-mail, as entrevistas mediadas pelo computador e pelas redes são algumas das formas adaptadas, diferentes, por exemplo, da mineração e raspagem de dados (data

mining e *scraping*), do acesso direto às APIs das plataformas de mídia social, da utilização de métricas como o Page Rank ou de ferramentas como Open Refine para, respectivamente, coletar, classificar e organizar dados.

Os métodos digitais, dessa forma, não pretendem substituir os atuais, mas complementá-los em situações de pesquisa ou profissionais onde se está lidando com objetos cuja ontologia própria exige um tipo de abordagem mais efetiva.

Por mais estranho que possa parecer, foi a complexificação do comportamento humano que tornou necessária a inserção dos processos maquímicos nas áreas de conhecimento em que antes tais abordagens nunca foram comuns ou bem-vindas. As humidades digitais (LEMOS, 2002; MORETTI, 2007), a linguística computacional (Santos, 2014), o jornalismo automatizado ou guiado por dados (BRADSHAW, 2014; BRUNS, 2017; RODRIGUES, 2009), a matemática aplicada à sociologia (BONACHI; LU, 2012) e tantas novas formas híbridas de conhecimento desafiam os que se propõem a estudar os processos contemporâneos, hoje sustentados por redes e máquinas binárias, gerando objetos digitais com características e ontologia próprias (MANOVICH, 2001), incapazes de serem desvelados sem uma reorientação de métodos, ferramentas e técnicas de pesquisa (MORETTI, 2007; DOS SANTOS, 2015; SANTOS, 2015; VAN DIJCK, 2013).

Na Comunicação, foram os jornalistas investigativos e os profissionais no ambiente organizacional que primeiro tiveram que lidar com esse tipo de problema. Dados dos portais públicos de transparência e as histórias que podiam conter, escondidas atrás dos números; métricas dos indicadores relacionados à presença nas plataformas de mídias sociais; informações trazidas pelas ferramentas de monitoramento e a enxurrada de dados das soluções de *analytics*. Tudo isso trouxe ao campo não apenas um conjunto novo de problemas, mas uma espécie de reação em cadeia que começou a impactar também a necessidade de novas habilidades desses profissionais, diferentes formas de abordagem e (por que não também?) as revisões teórica e epistemológica de um saber que tinha sido construído num mundo analógico, bem diferente do atual.

O empoderamento público sustentado pelos meios digitais de comunicação, pela expansão material da infraestrutura de internet e pelo maior acesso a dispositivos facilitadores das conexões e da produção de conteúdo individualizada, criou uma explosão de emissores, reconfigurando as redes de difusão de informação do mundo analógico, antes concentradas em grandes *hubs* de atenção, como os veículos de mídia e as fontes oficiais.

Como numa cidade onde se abriam diversas novas vias, a circulação de dados ganhou velocidade e volume nunca antes vistos, engendrando processos só passíveis da compreensão a partir das teorias da complexidade, das redes e dos jogos, todas muito estranhas ao pesquisador tradicional das Humanidades e Ciências Sociais.

Em termos mais gerais a aproximação das Ciências Sociais com a temática do *big data* também tem sido foco de alguns trabalhos como em González-Bailón (2013), que argumenta que a partir desse tipo de abordagem novas questões podem ser feitas e antigas revisitadas; Mahrt e Scharkow (2012) que discutem as situações onde a análise de grandes quantidades de dados pode ser útil ou ainda Bruns (2013) que problematiza as dificuldades em conciliar procedimentos e terminologia oriundos das ciências ditas duras ao trabalho acadêmico das Humanidades e Ciências Sociais.

Em paralelo, a necessidade de traduzir enormes massas de dados em formas de mais fácil apreensão para os consumidores de informação também levou esses profissionais a utilizarem ferramentas de visualização e infografia (RODRIGUES, 2009; CORDEIRO, 2013), capazes de traduzir, em imagens mais simples, padrões, tendências e inferências, num tipo de lógica de síntese já comum em áreas como a Economia e a Estatística.

Se em termos de técnicas a extração de dados (*scraping*) e a visualização têm se transformado em novas habilidades fundamentais para o jornalista digital, mesmo conhecimentos mais distantes como o de programação tem recebido interesse crescente entre os profissionais.

3. MÉTODO

Para o desenvolvimento em pesquisa aplicada nossa opção metodológica é a Design Science (DS), “ciência que procura consolidar conhecimento sobre o projeto e desenvolvimento de soluções para melhorar sistemas existentes, resolver problemas e criar novos artefatos” (DRESCH; LACERDA; ANTUNES Jr., 2015, p. 59).

“O artefato, criado pelo homem, representa um intermediador entre um conjunto do conhecimento estabelecido em determinada área e as condições específicas que envolvem o problema que o artefato deverá resolver. (SANTOS, 2016, p.11)”.

A DS classifica os artefatos em um contínuo que vai de um extremo mais abstrato a outro mais tangível e inclui as categorias de constructos, modelos, métodos e instanciações.

O termo classe de problemas também faz parte dos conceitos importantes da DS.

Conjuntos de problemas práticos ou teóricos, que tem já estabelecido um conjunto de soluções ou artefatos a eles ligados, constituem-se numa classe de problemas.

Como exemplo da Comunicação e das Ciências Sociais, poderíamos citar a necessidade geral de coletar dados em repositórios na internet, que poderíamos nomear como coleta de dados digitais.

Seja para a produção de uma matéria jornalística, para um plano de gestão ou para a definição de uma política pública sobre determinado tema, com os processos de digitalização e o crescimento do uso de bases de dados, a necessidade de conseguir tais informações, acessando seus repositórios disponíveis na rede, tais como portais de transparência, por exemplo, caracteriza uma classe de problemas onde operam artefatos como os métodos de *scraping* (raspagem) e extração automatizada, bem como as instanciações disponíveis exemplificadas pelos algoritmos em determinada linguagem de programação, que operam para resolver tais problemas (SANTOS, 2016, p.13).

Deste modo a proposta de um protocolo de abordagem unido ao desenvolvimento das ferramentas do LABCOM se insere na categoria de problemas representada pela necessidade enfrentada pelos profissionais e pesquisadores de extrair conteúdo volumoso, seja a partir de sites jornalísticos ou da plataforma Twitter ou semelhante, de forma automatizada, garantindo uma amostragem mais significativa e ganho de tempo, evitando a coleta manual bem mais lenta e cansativa.

Os artefatos, utilizando a terminologia da DS, permitem também ao pesquisador da Comunicação ou ao jornalista a visualização da estrutura interna das editorias/categorias no caso do LNEWS, bem como buscas por perfil ou termo, no caso do LTWEET.

4. RESULTADOS ESPERADOS E OBTIDOS

De forma geral , o esforço de pesquisa aplicada acima descrito tem como linhas gerais os seguintes objetivos:

a) Ampliação epistêmica - incorporação das atividades ou tentativas de predição e prescrição nas iniciativas de pesquisa ligadas às classes de problemas do mundo real; incentivo à pesquisa aplicada e orientação às abordagens guiadas por dados;

b) Extensão metodológica - experimentação e teste de abordagens como a dos métodos digitais, baseadas na premissa de uma ontologia específica dos entes binários que impacta suas formas de apreensão, como, num exemplo mais específico, a recente adoção da metodologia de análise de redes sociais (ARS) pelo campo;

c) Interdisciplinaridade - abertura à conexão e operação de grupos de pesquisadores oriundos de outras tradições acadêmicas para o enfrentamento de problemas que também se constituem a partir da hibridização de temáticas de complexidade aprofundada;

d) Revisão teórica - adaptação e proposição de constructos, modelos e explicações com possibilidade de operação no atual ecossistema informacional contemporâneo;

e) Diversificação de habilidades - abertura à aproximação de abordagens quantitativas mais consistentes, que incluam análise e visualização de dados, estatística e até, em alguns casos, matemática e linguagens de programação, como Python.

f) Atualização nos programas de formação de profissionais e pesquisadores - o que implica revisão (ou renovação de propostas) de disciplinas, ementas, programas e, eventualmente, áreas de concentração e linhas de pesquisa, no caso dos programas de pós-graduação.

Em relação ao que já foi feito poderíamos citar pelo menos uma dezena de publicações em revistas acadêmicas e nos principais eventos da área sobre as experiências e desenvolvimento nesta linha de trabalho. Além disso, toda a suíte de ferramentas já tem registro no INPI, sendo que uma sexta já está em desenvolvimento, denominada LWhats para análise de conversas em grupos do WhatsApp.

5. CARÁTER DE INOVAÇÃO

Ressalte-se que, de forma alguma, a inserção dessas mudanças deverá acontecer de forma acrítica ou contemplativa, a partir de uma visão simplista de que a aproximação com dados, algoritmos e ferramentas computacionais por si só nos salvará a todos.

Hoje, no Brasil, já existe, ainda que de forma embrionária, um conjunto de pesquisadores que têm explorado, por exemplo, o viés e os erros que a utilização de ferramentas de inteligência artificial (especificamente *machine learning*) têm gerado em processos que acabam descambando para invasão da privacidade, distorção dos fatos e até práticas como racismo e xenofobia, como demonstram Amadeu (2019) e Silva (2019).

Entretanto, é justamente para que a tradição crítica do campo da Comunicação possa continuar a ser exercida no ambiente digital, que se faz necessária uma aproximação direta de tais temas ou, pelo menos, de forma indireta, através do trabalho em equipes interdisciplinares, capazes de operar mais efetivamente com tais questões. Afastar-se dessas temáticas, isso sim, poderá nos

colocar em posição de fragilidade científica e até irrelevância, num movimento de distanciamento dos processos essenciais que guiam as transformações no ambiente digital.

Entendemos que a aplicação desse caminho, além de trazer resultados de difícil alcance por outros meios, ainda permite potencialmente enfrentar novas questões de investigação ou ainda submeter questões já abordadas a análises diferentes, caminho recomendável em toda investigação científica. A disposição de enfrentar tais desafios consiste, em nosso entendimento, na inovação da nossa proposta e do nosso trabalho.

6.0 REFERÊNCIAS

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

CHWE, Michael S. Communication and coordination in social networks. *Review of Economic Studies*, 67, p. 128-156, 2000.

DRESCH, Aline; LACERDA, Daniel Pacheco; ANTUNES JR, José Antonio Valle. Design Science Research: método de pesquisa para avanço da ciência e tecnologia. Porto Alegre: Bookman, 2015.

EASLEY, David; KLEINBERG, Jon. Networks, Crowds and Markets: reasoning about a highly connected world. Nova York: Cambridge University Press, 2010.

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

GLEICK, James. A informação: uma história, uma teoria, uma enxurrada. São Paulo, Companhia das Letras, 2013.

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

MANOVICH, L. The language of new media. Massachusetts: Mit Press. 2001.

MORIN, Edgar. Introdução ao pensamento complexo. 4.ed. Porto Alegre: Sulina, 2005.

NEWMAN, M. E. Networks: an introduction. Nova York: Oxford University Press, 2010.

ROGERS, Richard. Digital Methods. Cambridge: Mit Press. [E-book], 2013.

SANTAELLA, Lúcia. Culturas e artes do pós-humano: da cultura das mídias à cibercultura. São Paulo: Paulus, 2003.

SANTOS, Márcio. Comunicação digital e jornalismo de inserção: como big data, inteligência artificial, realidade aumentada e internet das coisas estão mudando a produção de conteúdo informativo. São Luís:

LABCOM Digital, 2016.

SIMON, Herbert A. The architecture of complexity. In: Proceedings of the American Philosophical Society. v. 106, n. 6. dez, 1962.

VILCHES, Lorenzo. A migração digital. São Paulo: Loyola, 2003.