

# Métodos Digitais: A Internet e as Redes como Instrumentos de Pesquisa

Digital methods: The Internet and networks as research tools

Márcio Carneiro dos Santos  
Universidade Federal do Maranhão (Brasil)

[mcszen@gmail.com](mailto:mcszen@gmail.com)

## Resumo

Discutem-se as possibilidades que a internet oferece para a pesquisa em Comunicação através do uso intensivo de recursos computacionais e técnicas diretamente ligadas às características dos objetos originalmente criados pelos processos digitais. Diferentes de modalidades anteriores que foram portadas ou adaptadas para o ambiente das redes, os métodos digitais tem foco na lógica da máquina e na relação que estabelece com a estruturação de sentido produzida pelo humano. A proposição de uma epistemologia especializada que incorpora, entre outros métodos, as soluções via desenvolvimento de código para coletar, analisar ou visualizar dados é explorada utilizando três exemplos: uma pesquisa a partir da conexão com a API do *Twitter*, um experimento de narrativas automatizadas e o acesso à memória digital do site *web.archive.org*, conhecido como *WaybackMachine*.

## Abstract

The paper discusses the possibilities that the Internet and networks offer for research in Communication through the intensive use of computer and technical resources directly linked to the characteristics of the original objects created by digital processes. Different from previous methods that have been ported or adapted to the network environment, digital methods focus on the logics of the machine and on the relationship with the structuring of meaning produced by human beings. The proposition of a specialized epistemology that includes, among other methods, social network analysis (SNA), agent-based modeling (ABM), as well as solutions via code development to collect, analyze or visualize data is explored using three examples: a survey from the access to the Twitter API, an experiment of automated narratives and the access to digital memory through the site *web.archive.org* known as *WaybackMachine*.

**Palavras-Chave:** epistemologia, ontologia sistêmica, métodos digitais.

**Key Words:** epistemology, ontology systemic, digital methods.

## 1. INTRODUÇÃO

O impacto dos processos de digitalização em grande parte da produção de sentido humana tem hoje seus efeitos estudados em diversas frentes. Entre elas, a) o surgimento de novos modelos de negócio e de uma economia onde há excesso de informação e a atenção das pessoas transforma-se em ativo extremamente valorizado; b) a série de mudanças comportamentais impulsionadas por novas formas de sociabilidade e interação; c) os desdobramentos em termos de relações de poder e participação cívica a partir de uma esfera pública expandida e mais complexa, povoada por um número bem maior de atores, com agendas e interesses diversos.

Diante das possibilidades de angulações e de um ainda pouco explorado sentido de aceleração no ritmo das mudanças em andamento é ainda tímido o processo de atualização do ferramental teórico e metodológico no campo da Comunicação para enfrentar os novos problemas que estamos nos propondo a estudar atualmente.

Tal fato, em parte, deve-se à lógica particular do desenvolvimento científico que precisa sempre de mais tempo de depuração para estabelecer suas bases, em tese, mais sólidas e provenientes da maturação das ideias, da validação da prova e da crítica entre pares.

Outro aspecto mais específico é a própria origem histórica do que se fez em Comunicação a partir do conhecimento e experiências de áreas como a Sociologia, a Psicologia e a Antropologia. Essa ligação, embora fundamental, de certa forma parece hoje também acrescentar um grau a mais de dificuldade na proposição de um pensamento gestado originalmente em questões ligadas às mídias digitais.

Seria improdutivo argumentar contra o fundamental papel que todas essas inter-relações e origens aportaram à Comunicação, bem como sua função estruturadora de grande parte do que é feito em pesquisa na área até hoje. Entretanto, é possível pensar nas possibilidades de reconfiguração e expansão do campo a partir de novas conexões, por exemplo, com a Ciência da Computação e a Neurociência, com os estudos da interação homem-máquina (IHM), ou ainda com a Filosofia da Tecnologia.

Além disso, o pensamento comunicacional do ambiente digital reforçou iniciativas internas como a *Media Ecology* oriunda da tradição dos estudos com ênfase nos meios, anteriores inclusive ao advento da popularização dos computadores.

A linhagem teórica estabelecida por McLuhan (2007), Bolter e Grusin (2000) e ainda Scolari (2008), para citar apenas alguns, teve muitos dos seus trabalhos revisitados diante das transformações contemporâneas.

Cada definição de Comunicação está fundada numa metáfora. A Comunicação já foi vista sucessivamente como canal, instrumento, flecha, projétil, conflito, contrato, orquestra, espiral e rede. [...] Neste texto faremos uma aposta muito clara pela metáfora do ecossistema, ou seja a Comunicação entendida como um conjunto de intercâmbios, hibridações e mediações dentro de um entorno onde confluem tecnologias, discursos e culturas (SCOLARI, 2008, p.26)<sup>1</sup>.

Falando sobre as pesquisas ligadas à internet, Rogers (2013) estabelece uma diferença fundamental entre objetos, conteúdos, equipamentos e ambientes nativamente digitais e aqueles que foram digitalizados, ou seja, que, com outras origens, foram portados ou migraram para o digital usando os termos de Vilches (2003).

Transpondo esse raciocínio para o trabalho de pesquisa, Rogers também separa os métodos eminentemente digitais dos que ele denomina de virtuais, ou seja, que tem 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 *email*, as entrevistas mediadas pelo computador e pelas redes são algumas dessas formas adaptadas, diferentes, por exemplo, da mineração e raspagem de dados (*data mining e scraping*), do acesso direto às APIs<sup>2</sup> das plataformas de mídia social, da utilização de métricas com o *Page Rank*<sup>3</sup> ou de ferramentas como Open Refine<sup>4</sup> para respectivamente coletar, classificar e organizar dados. A proposta pretende “reorientar o campo da pesquisa relacionada com a internet estudando e adaptando o que chamo de métodos do meio, ou talvez de forma simplificada, métodos inseridos nos objetos digitais (ROGERS, 2013)”<sup>5</sup>.

---

<sup>1</sup> Tradução do autor.

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

<sup>3</sup> PageRank é uma métrica ligada à Teoria de Redes que identifica centralidade ou importância a partir das conexões entre os elementos da rede, sendo uma das estratégias utilizadas pelo Google para ranquear resultados de busca.

<sup>4</sup> Mais detalhes sobre a solução Open Refine em [www.openrefine.org](http://www.openrefine.org) .

<sup>5</sup> Tradução do autor

## 2. POR UMA EPISTEMOLOGIA ESPECÍFICA DO DIGITAL

Vargas (1994), ao desenvolver seu pensamento sobre uma filosofia da tecnologia, propõe a ideia de que em diferentes períodos da humanidade estabeleceu-se uma conexão entre crenças, ciências e metafísica, esta última pensada na concepção de Ortega y Gasset. Escreve Vargas (1994, p. 27), “pois que a metafísica é entendida por Ortega como o tratado teórico sobre a raiz da realidade, sobre a qual os homens, em cada cultura e em cada época, edificam seu mundo”.

Assim, com os gregos, a partir do questionamento fundamental sobre a natureza das coisas e a crença de que por trás das aparências do mundo havia algo permanente, verdadeiro e eterno que eles denominavam de *physis*, surgem a metafísica de Aristóteles, a *physica* e a matemática gregas, com a aritmética e a geometria.

Quando os objetos matemáticos puderam se estabelecer como objetos racionais, revelou-se, através deles, a possibilidade de entender as ideias platônicas ou as substâncias aristotélicas como entidades eternas, imutáveis e permanentes, fontes de toda realidade lógica e inteligível. Assim o caminho da ciência matemática conduziu, na Grécia, à Metafísica (VARGAS, 1994, p. 30).

Na idade média a eternidade da *physis* natural é reconfigurada no ocidente a partir da disseminação do cristianismo e da crença na divindade única, criadora do mundo. Em Santo Agostinho a busca da verdade e a busca por Deus tornam-se a mesma coisa e na teologia de São Tomás de Aquino a razão e a fé andam juntas já que a primeira vai sustentar e confirmar os enunciados da segunda, sendo a lógica, agora, a ferramenta para essa tarefa.

Com o Renascimento e nos anos que se seguiram, mais uma vez uma grande mudança aconteceu. A confiança ilimitada apenas na razão humana e na lógica como formas de conhecimento do mundo e da verdade perderam força. A geografia, as grandes viagens e a observação mais atenta dos fatos naturais começaram a expor dúvidas e diferenças entre as antigas crenças e o que era encontrado através da experimentação, não ainda formatada no sentido atual, mas já considerando a natureza como uma instância capaz de fornecer aos interessados segredos e descobertas, mesmo que ainda, através do poder da Igreja, esses fossem um reflexo da criação divina.

Com Galileu muda-se o critério de validação do conhecimento, que deixa de ser feito pela fé e passa a ser resultado da comprovação experimental de algo que antes era apenas conjectura racional. Mesmo sendo condenado em 1633 pelo Santo Ofício, seu pensamento já

representava a visão de um mundo natural concebido como uma máquina em movimento. As ideias de força, trabalho e energia trouxeram a metafísica de Leibniz e a concepção de uma natureza destituída de alma, afastada de Deus e que, por ser inóspita e hostil, precisaria ser dominada; ideias que Newton irá consolidar com sua Física, dando início ao que entendemos hoje como ciência moderna.

Usando as palavras de Vargas (1994) poderíamos então pensar sobre a “raiz da realidade” que os homens contemporâneos usam para edificar seu mundo? Em que acreditam, como percebem o ambiente em que estão e que forças os movem nos dias de hoje? Diante da enorme lista de possibilidades, talvez seja viável indicar que um mundo hiperconectado, regido pelos velozes fluxos de informação que trafegam em redes digitais, impactando ciência, economia, política, cultura, crenças e comportamento humano, seria um dos itens dessa complexa metafísica do século XXI.

O conceito de sociedade informacional de Castells (2009) de certa forma corrobora essa visão. “Uma revolução tecnológica concentrada nas tecnologias da informação começou a remodelar a base material da sociedade em ritmo acelerado (CASTELLS, 2009, p. 39)”.

Um aspecto interessante sobre essa linha de pensamento traduz-se no fato de que a transformação material de que nos fala o autor está intimamente ligada ao advento dos processos de digitalização, pelos quais grande parte da produção de sentido humana, antes dependente dos suportes materiais para seu registro, torna-se agora uma enorme massa de informação numérica, traduzida em sequências de 0 (zero) e 1 (um), processadas de forma automatizada e, por muitas vezes, totalmente transparente ao entendimento comum, fluindo ao nosso redor, sem que saibamos direito o que realmente está acontecendo.

Seguindo a construção dessa metafísica do mundo digital contemporâneo é necessário um esforço adicional para compreensão dos seus elementos constituintes, dos entes que sustentam sua existência, bem como nas formas e métodos para que possamos estudá-los e entendê-los. Ao digital caberiam, portanto, ainda que de forma restrita ou especializada, uma ontologia e uma epistemologia, capazes de ajudar-nos na descrição do mundo que nos rodeia.

A utilização de uma abordagem ontológica para enfrentar problemas que envolvem situações complexas é sugerida por Vieira (2008) quando diz que

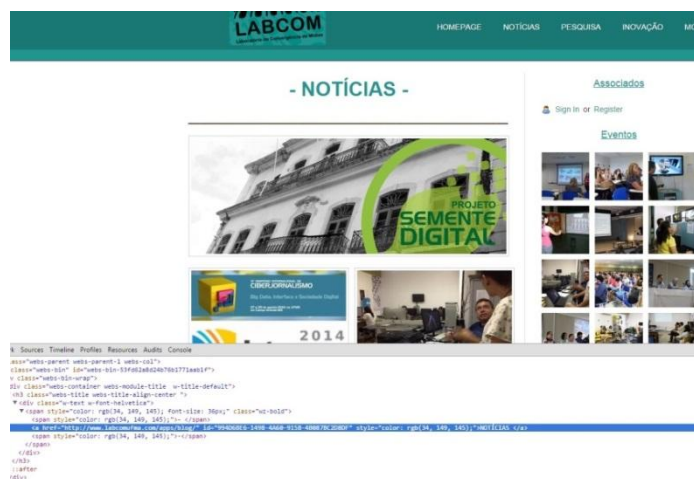
A complexidade exige que possamos entender e modelar a interação entre coisas e processos de natureza muitas vezes bem diversas, sob pena de não captação do que há de fundamental nesses sistemas. É a Ontologia que pode facilitar isso, com seu enfoque em busca do geral e do completo (VIEIRA, 2008, p. 25).

Assim, Vieira afirma que “uma das vantagens da prática ontológica é que, ao lidarmos com traços muito gerais das coisas, podemos utilizar os mesmos para fazer comparações e conexões inter e transdisciplinares” (VIEIRA, 2008, p. 26).

O trabalho de Manovich (2001) no intuito de descrever as características dos objetos digitais, dentro da discussão que trava sobre a dificuldade teórica em delimitar novas e velhas mídias, nos parece oferecer, ainda que o autor não use esses termos diretamente, uma proposta que nos aproximaria de uma ontologia dos entes digitais. Para Manovich (2001), os objetos digitais apresentam cinco traços ou características que podem ou não estar presentes simultaneamente em sua existência, a saber: descrição numérica, modularidade, automação, variabilidade e transcodificação.

Segundo essa última condição, cada objeto digital é constituído de duas camadas ou *layers*, uma utilizada para carregar o sentido a ser interpretado e processado pelos humanos, a camada da representação, que nos oferece o material para que possamos lidar com tal objeto. Entretanto, pela transcodificação, existe ainda uma segunda camada (FIG. 1), que também descreve ou traz informações sobre esse objeto só que para o processamento maquínico, automatizado, o *layer* dos dados estruturados que os computadores entendem e que é usado para fazer esse objeto trafegar pelas redes digitais.

FIGURA 1 – Imagem do site do LABCOM/UFMA (www.labcomufma.com) com seu respectivo código HTML aparente. Fonte: do autor



Que impactos tal proposta ontológica teria quando a pesquisa envolve temas e objetos que tenham vinculações diretas com o digital?

A resposta de Rogers (2013) indica que, mesmo portando métodos tradicionais para o emprego em pesquisas ligadas ao digital podemos, em algumas situações, estar utilizando um ferramental inadequado, por não considerar os aspectos específicos desses objetos, ou ainda, poderíamos acrescentar, estar em condição desconfortável para inferir ou avançar em conclusões mais sólidas, já que estamos processando apenas parte da informação que nos é disponibilizada. A ideia de métodos do meio, ou seja, métodos que exploram a lógica interna inerente aos objetos digitais, ou nos termos que estamos propondo, que consideram sua ontologia específica, permitem novas abordagens e formas mais eficientes de enfrentar dificuldades implícitas em algumas temáticas contemporâneas.

Por exemplo, varredura e extração de dados, inteligência coletiva e classificações baseadas em redes sociais, ainda que de diferentes gêneros e espécies, são todas técnicas baseadas na internet para coleta e organização de dados. *Page Rank* e algoritmos similares são meios de ordenação e classificação. Nuvens de palavras e outras formas comuns de visualização explicitam relevância e ressonância. Como poderíamos aprender com eles e outros métodos *online* para reaplicá-los? O propósito não seria tanto contribuir para o refinamento e construção de um motor de buscas melhor, uma tarefa que deve ser deixada para a Ciência da Computação e áreas afins. Ao invés disso o propósito seria utiliza-los e entender como eles tratam *hiperlinks*, *hits*, *likes*, *tags*, *datestamps* e outros objetos nativamente digitais. Pensando nesses mecanismos e nos objetos com os quais eles conseguem lidar, os métodos digitais, como uma prática de pesquisa, contribuem para o desenvolvimento de uma metodologia do próprio meio (ROGERS, 2013).<sup>6</sup>

A proposta de Rogers vai ao encontro do percurso que ora propomos partindo de uma visão do mundo contemporâneo onde o digital apresenta uma centralidade crescente, composto por entes com características específicas e, por isso, demandando também uma adequação ou extensão metodológica capaz de colaborar no esforço de pesquisas cujos objetos de alguma forma têm essa característica.

A necessidade de iniciativas nessa linha pode ser justificada também por algumas condições verificáveis relacionadas à produção de informação a partir das redes: volume, variedade, velocidade. Não à toa esses termos estão associados a outro conceito contemporâneo, o de *big data*, que de forma simplificada poderia ser definido como o conjunto

---

<sup>6</sup> Tradução do autor

de métodos, ferramentas e processos destinados a lidar com a verdadeira enxurrada informacional com a qual nos deparamos hoje, tema que Gleick (2013) descreve numa perspectiva histórica e técnica.

Para o método científico um desafio adicional está ligado às estratégias de amostragem. À medida que temos fenômenos que acontecem com grande velocidade, variedade e volume, como os comentários no Twitter sobre determinado tema ou evento, por exemplo, até que ponto poderemos estabelecer amostras demasiado pequenas e ainda conseguir inferir algo com validade razoável?

É óbvio que cada pesquisa tem suas características particulares e em muitas delas o trabalho com amostras reduzidas é plenamente viável. Contudo à medida que o universo de estudo se expande, em termos relativos, tais amostras representam percentuais cada vez menores e, talvez, menos significativos.

São situações assim que exigem a incorporação de métodos que considerem as características inerentes aos objetos digitais, entre elas a transcodificação nos termos de Manovich. Como veremos a seguir, uma alternativa viável, para casos onde os dados são gerados dentro de uma plataforma de mídias sociais como o Twitter, é o contato direto com os servidores que a sustentam ou, em termos técnicos, a utilização da sua API (*Application Programming Interface*) para realizar consultas e extração de informação a partir do *layer* da máquina.

### **3. A CURVA EPISTÊMICA E SEU DESLOCAMENTO**

A inclusão dos métodos digitais no ferramental disponível para o pesquisador da Comunicação pode ajudar a enfrentar um problema que, com a situação de volume, variedade e velocidade, foi agravado e tem sua representação sugerida através do que denominamos aqui de curva epistêmica. O gráfico abaixo relaciona duas variáveis ligadas aos resultados advindos de pesquisas científicas na área de Ciências Sociais: profundidade e inferenciabilidade, ou seja, a capacidade de fazer inferências a partir dos dados e análises que se tem.

Ao operar com abordagens qualitativas, que dão ênfase ao *layer* do sentido e da significação humana, é possível conseguir um elevado grau de profundidade e especificidade já que o pesquisador tem contato direto com o nível dos indivíduos, entrevistando-os, observando em profundidade seu comportamento e coletando, a partir de suas opiniões e visão de mundo, os dados para análise. O contraponto dessa forma de pesquisa é justamente a

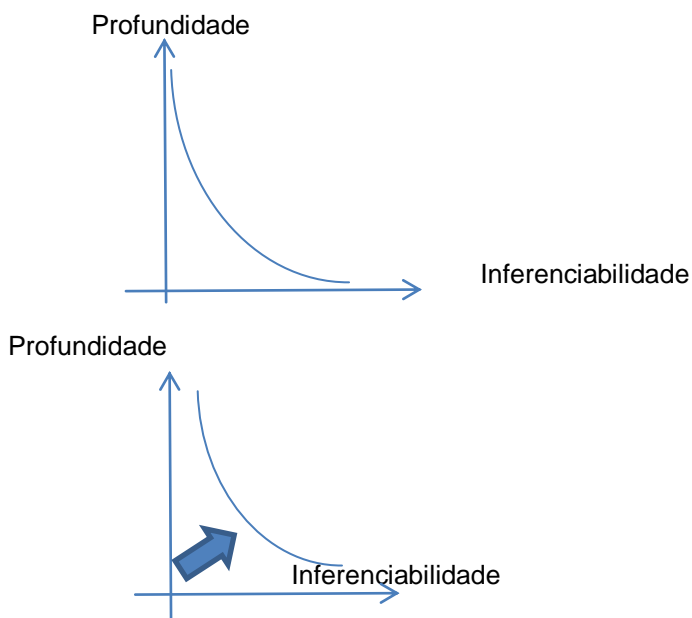


dificuldade de estabelecer padrões mais gerais. Em muitos casos as estratégias de amostragem não são probabilísticas e, por isso, oferecem potencial reduzido de gerar inferências. Longe estamos aqui de fazer críticas a essa forma de pesquisa que em muitas situações é amplamente justificável, e até, em alguns casos, a única alternativa viável. Mesmo assim a restrição existe e nos ambientes digitais pode ser agravada, gerando um distanciamento entre resultados e procedimentos científicos, levando às vezes ao caminho de um questionável ensaísmo como exemplifica Machado (2012).

As abordagens quantitativas que utilizam procedimentos estatísticos e estratégias de amostragem probabilísticas oferecem, ao contrário, maior potencial de extração de inferências e suporte mais sólido à identificação de padrões e tendências diante de universos maiores. Ao mesmo tempo, sofrem da falta de aprofundamento na compreensão do nível individual, sendo seus resultados criticados justamente por reduzirem a complexidade humana a modelos extremamente simplórios.

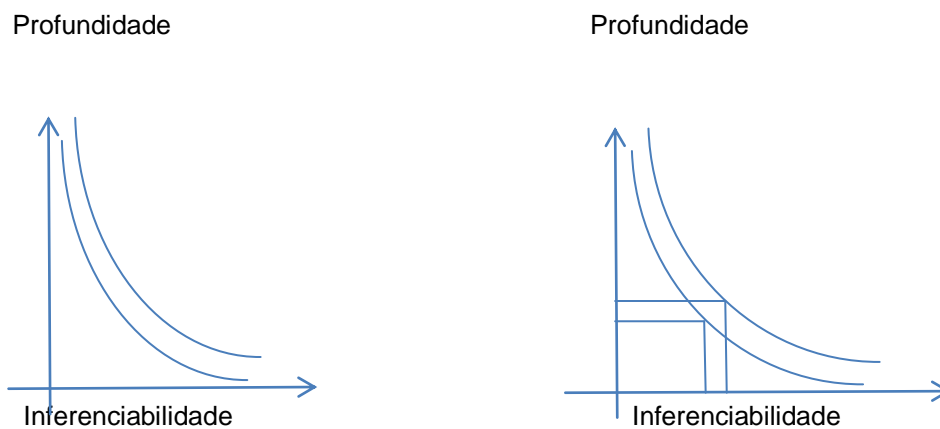
Considerando o ambiente digital e suas características, nossa proposta epistemológica baseada numa ontologia específica dos objetos estudados pretende minimizar as consequências das duas abordagens expostas à medida que, com os métodos digitais, pretende olhar para as duas camadas de que são constituídos, extraindo dados de ambas, dentro do possível, e ainda considerando as inter-relações que estabelecem entre si.

FIGURA 2 – Deslocamento da curva epistêmica considerando os eixos da profundidade na apreensão dos objetos de estudo em relação ao potencial de fazer inferências embasadas sobre eles. Fonte: do autor.



O deslocamento da curva epistêmica (FIG. 2) com a inclusão dos métodos digitais, que permitem a coleta de dados do *layer* da máquina, oferece ao pesquisador a possibilidade de operar com amostras maiores e ainda ter um potencial de aprofundamento considerável. Um exemplo seria coletar grandes massas de *tweets* e ainda poder analisar o conteúdo das mensagens através de ferramentas de específicas como o NLTK – *Natural Language Toolkit*<sup>7</sup>.

FIGURA 3 – Potencial de inferenciabilidade numa relação mais efetiva, minimizando sua proporcionalidade inversa. Fonte: do autor.



Se considerarmos na figura acima da direita (FIG. 3) que o ponto médio de cada curva é o ponto de equilíbrio entre nosso potencial de fazer inferências com sustentação científica mantendo ainda um razoável nível de profundidade na percepção dos objetos estudados, o deslocamento da curva epistêmica nos mostra que o potencial de geração de conhecimento com a adoção dos métodos digitais faz crescer, em tese, essa possibilidade.

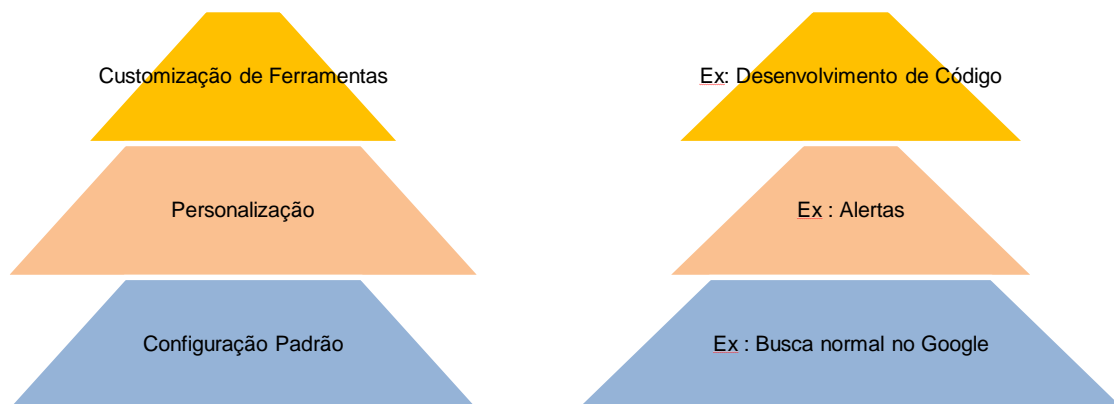
#### **4. A NOVA CAIXA DE FERRAMENTAS DO PESQUISADOR EM COMUNICAÇÃO**

<sup>7</sup> <http://www.nltk.org/>

A diversidade de novas ferramentas hoje disponíveis para o pesquisador de Ciências Sociais não poderia ser totalmente coberta nesse texto. Por isso, decidimos destacar algumas delas considerando sua abrangência e aplicabilidade. Como já foi dito anteriormente, uma série de técnicas como a mineração de dados, a extração direta de informações a partir das APIs das plataformas de mídias sociais, o desenvolvimento de códigos customizados para coleta e análise de material são apenas algumas possibilidades.

Tais soluções oferecem uma espécie de escala de utilização como representada abaixo (FIG. 4):

FIGURA 4 - 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 personaliza-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.

Na pirâmide ao lado 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 alertas<sup>8</sup> e por fim através de um código específico para coletar e armazenar esses dados.

<sup>8</sup> <https://www.google.com/alerts>

Desse modo, definimos métodos digitais como o conjunto de ferramentas , processos e abordagens de pesquisa que consideram a ontologia dos objetos digitais e as estruturas de redes por onde circulam, utilizando-se de recursos computacionais intensivos para coleta e análise de dados.

É importante dizer que a proposta dos métodos digitais não impede de forma alguma sua adoção em parceria com outros procedimentos ou abordagens tradicionais, sendo traduzida na maioria das vezes como um conjunto adicional de recursos à disposição do pesquisador.

#### **4.1 Experimentos de Customização de Código**

A seguir descreveremos de forma reduzida três trabalhos de pesquisa por nós desenvolvidos onde a ideia de métodos digitais foi utilizada em sua escala de aplicação mais intensa para operacionalizar ou complementar as etapas de coleta e análise de dados.

##### *4.1.1 Avaliando uma métrica de intensidade*

.Tópico: Plataformas de Redes Sociais e conexão com suas APIs para coleta de dados em grande volume.

A hipótese era ter a plataforma *Twitter* como um sistema de produção de notícias, onde usuários se comportariam como produtores de conteúdo e audiência, constituindo uma rede onde informação e atenção trafegam simultaneamente, a partir das relações de interesse e filiação criadas pelas categorias de seguidos e seguidores. Para isso decidiu-se buscar no sistema algo que fosse semelhante aos clássicos conceitos de valores-notícia ou critérios de noticiabilidade. O experimento modelado fez medições da métrica *tweets* por minuto (TPMs), considerada como uma variável capaz de representar o interesse dos emissores por determinados fatos ou temáticas e daí motivá-los a escrever sobre eles, como jornalistas que consideram tais critérios também de forma intuitiva para decidir o que vão publicar. Foi desenvolvido um código (FIG. 6) em linguagem Python, denominado *Social Tracker* (ST) para coletar e analisar *posts* durante a transmissão do desfile das escolas de samba do Carnaval do Rio de Janeiro de 2013 e verificar a variação dos TPMs de acordo com o que estava acontecendo na transmissão. Mais detalhes em Santos (2013).

FIGURA 6 - Imagem da resposta da API do Twitter recuperada por ST. Fonte: Santos (2013)

```
{
  "completed_in": 0.069,
  "max_id": 303248101462908928,
  "max_id_str": "303248101462908928",
  "next_page": "?page=2&max_id=303248101462908928&q=Social%20Tracker&rpp=100",
  "page": 1,
  "query": "Social+Tracker",
  "refresh_url": "?since_id=303248101462908928&q=Social%20Tracker",
  "results": [
    {
      "created_at": "Sun, 17 Feb 2013 21:02:36 +0000",
      "from_user": "mcezen",
      "from_user_id": 149853725,
      "from_user_id_str": "149853725",
      "from_user_name": "MCS",
      "geo": null,
      "id": 303248101462908928,
      "id_str": "303248101462908928",
      "iso_language_code": "pt",
      "metadata": {
        "result_type": "recent"
      },
      "profile_image_url": "http://a0.twimg.com/profile_images/1460824435/LabComReleituraFinal2_normal.JPG",
      "profile_image_url_https": "https://si0.twimg.com/profile_images/1460824435/LabComReleituraFinal2_normal.JPG",
      "source": "<a href='\"http://twitter.com/\"'>web</a>",
      "text": "Teste para a ferramenta Social Tracker desenvolvida pelo LABCOM para an\u00e9lise de conte\u00fado do Twitter.",
      "to_user": null,
      "to_user_id": 0,
      "to_user_id_str": "0",
      "to_user_name": null
    }
  ],
}
```

#### 4.1.2 Software pode escrever textos jornalísticos?

Tópico: Inteligência artificial e simulação

O experimento propôs-se a testar de forma prática a possibilidade já levantada por Lages (1997) e expandida por Arce (2009), sobre a produção de textos jornalísticos através de processos automatizados, ambos, entretanto, trabalhos eminentemente teóricos. Tal procedimento que já é praticado por pelo menos duas empresas americanas de inteligência artificial (*Narrative Science* e *Automated Insights*) serviu de base para o desenvolvimento de um protótipo simplificado de código capaz de escrever *leads* sobre os resultados do Campeonato Brasileiro de Futebol da Série A de 2013 a partir da extração automatizada dos dados sobre os resultados dos jogos publicados no portal Terra. A ferramenta (FIG. 7) extraia apenas os gols das partidas e utilizava as regras do próprio campeonato para inferir outras informações sobre a situação dos times e, a partir da concatenação de listas de palavras, escrever pequenos textos para publicação como demonstrado em Santos (2014).

FIGURA 7 - Tela do software com o lead construído a partir das informações lidas sobre o campeonato em determinada rodada. Fonte: Santos (2014)

classificação e jogos

TIME	Pts	J	V	E	D	GP	GC	SG	%
1 Cruzeiro	69	26	18	5	3	58	21	37	75
2 Grêmio	48	27	14	6	7	33	23	10	59
3 Atlético-PR	44	26	12	8	6	45	35	10	56
4 Botafogo	43	26	12	7	7	37	30	7	55
5 Atlético-MG	39	26	10	9	7	32	26	6	50
6 Vitória	37	26	10	7	9	38	38	0	47
7 Internacional	37	26	9	10	7	38	37	1	47
8 Santos	36	27	9	9	9	32	29	3	44
9 Corinthians	35	26	8	11	7	22	17	5	44
10 Portuguesa	34	26	9	7	10	40	38	2	43
11 Fluminense	34	26	9	7	10	31	33	-2	43
12 Flamengo	34	26	8	10	8	31	31	0	43
13 Goiás	34	26	8	10	8	28	32	-4	43
14 Coritiba	34	27	8	10	9	30	35	-5	41
15 Bahia	33	26	8	9	9	27	32	-5	42
16 São Paulo	38	26	8	6	12	24	29	-5	38
17 Criciúma	29	27	8	5	14	21	21	0	31
18 Vasco	29	26	7	8	11	21	21	0	31
19 Ponte Preta	23	26	6	5	15	26	40	-14	29
20 Náutico	17	26	4	5	17	17	45	-28	21

Python Shell

```

file Edit Shell Debug Options Windows Help
- Saldo de -3 gols e aproveitamento de 42.30 %.
Resumo Coritiba : O Coritiba tem 31 pontos. Com 7 vitórias, 10 empates, 9 derrotas. Marcou 29 gols. Sofreu 35 gols. Saldo de -6 gols e aproveitamento de 39.74 %.
Resumo São Paulo : O São Paulo tem 38 pontos. Com 8 vitórias, 6 empates, 12 derrotas. Marcou 24 gols. Sofreu 29 gols. Saldo de -5 gols e aproveitamento de 38.46 %.
Resumo Vasco : O Vasco tem 38 pontos. Com 7 vitórias, 8 empates, 11 derrotas. Marcou 35 gols. Sofreu 42 gols. Saldo de -7 gols e aproveitamento de 37.17 %.
Resumo Criciúma : O Criciúma tem 26 pontos. Com 7 vitórias, 5 empates, 14 derrotas. Marcou 32 gols. Sofreu 44 gols. Saldo de -12 gols e aproveitamento de 33.33 %.
Resumo Ponte Preta : O Ponte Preta tem 23 pontos. Com 6 vitórias, 5 empates, 15 derrotas. Marcou 26 gols. Sofreu 40 gols. Saldo de -14 gols e aproveitamento de 29.49 %.
Resumo Náutico : O Náutico tem 17 pontos. Com 4 vitórias, 5 empates, 17 derrotas. Marcou 17 gols. Sofreu 45 gols. Saldo de -28 gols e aproveitamento de 21.79 %.
SUGESTÃO DE TEXTO:
Depois de 26 rodadas o Cruzeiro tem 69 pontos. É o líder do Campeonato Brasileiro com 75.64 % de aproveitamento. Logo atrás seguem Grêmio e Atlético-PR respectivamente em segundo e terceiro. Na outra ponta da tabela estão Criciúma com 26 pontos, Ponte Preta com 23 pontos e na lanterna Náutico com 17 pontos.
>>>
Ln:67 Col:81

```

#### 4.1.3 Extrair dados da base de arquivos Internet Archive

Tópico: Sites como arquivo e raspagem de dados via código

A coleta automatizada de dados, também conhecida como raspagem (*scraping*) ou mineração é um recurso cada vez mais comum no jornalismo investigativo e pode, no caso do trabalho acadêmico, ser utilizada tanto para a execução de rotinas repetitivas, permitindo ao pesquisador mais tempo para as tarefas de maior complexidade, como para identificar padrões e tendências em grandes volumes de informação que, em algumas situações, podem passar despercebidos no processo exclusivamente manual. Nosso experimento inicial acessa o site conhecido como *WaybackMachine* (FIG. 8) ou *Internet Archive*<sup>9</sup> que constitui-se de uma biblioteca digital de sites de internet. Atualmente já é possível acessar as versões das páginas iniciais arquivadas (FIG. 9) no repositório e a partir delas analisar a evolução das versões, a intensidade e regularidade das alterações e também as palavras ou termos mais utilizados ao longo do tempo.

FIGURA 8 - Tela do Internet Archive com a marcação das versões arquivadas (399 entre 1997 e 2014) do site da UFMA em suas respectivas datas no ano de 2007 que podem ser extraídas via código. Fonte: Internet Archive (2014)

<sup>9</sup> <https://archive.org/>

**VIII Seminario Regional (Cono Sur) ALAI**  
“POLÍTICAS, ACTORES Y PRÁCTICAS DE LA COMUNICACIÓN:  
ENCRUJADAS DE LA INVESTIGACIÓN EN AMÉRICA LATINA”  
27 y 28 de agosto 2015 | Córdoba, Argentina

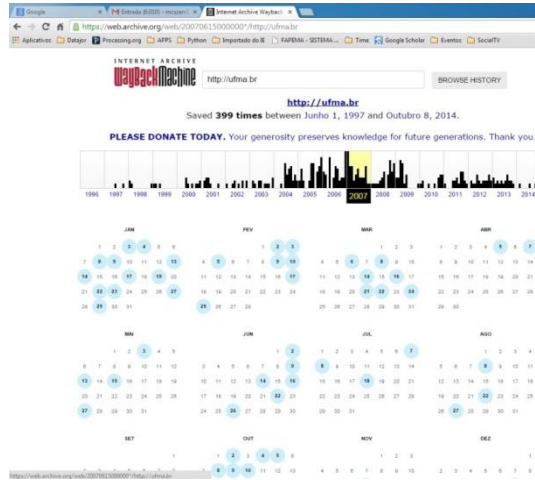
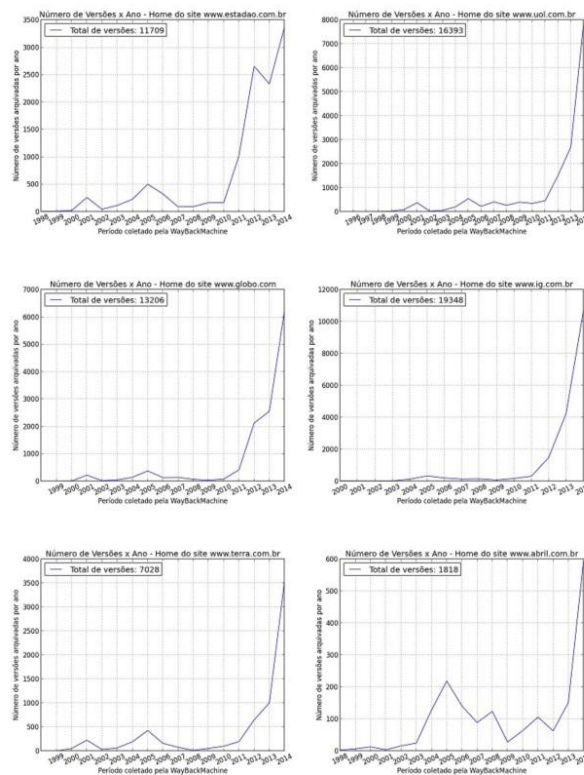


FIGURA 9 – Prints dos gráficos com o número de versões por ano dos principais portais jornalísticos brasileiros gerados através dos links extraídos automaticamente para as páginas arquivadas. Fonte: do autor.



## 5. CONSIDERAÇÕES FINAIS

A proposta dos métodos digitais e suas diversas modalidades de aplicação, longe de estabelecer um conflito com as formas tradicionais utilizadas no desenho de pesquisas da Comunicação, em nosso entendimento, vêm acrescentar ao conjunto de recursos disponíveis novas ferramentas capazes de traduzir as especificidades dos objetos que se constituem a partir da lógica binária e que em profusão povoam as redes telemáticas contemporâneas.

A disposição de estudá-los considerando suas características específicas, principalmente no que se refere ao seu potencial de carregar não só o sentido para a interpretação humana, mas também as instruções e a sintaxe entendidas pelas máquinas, pode abrir oportunidades para o desenvolvimento da pesquisa em nossa área, integrando novas angulações e abordagens às já existentes.

O deslocamento da curva epistêmica que propomos no texto representa uma direção geral aplicável dentro do possível e talvez capaz de oferecer um eixo mais sólido a muitas discussões que hoje ainda carecem da objetividade e rigor que se esperam do trabalho denominado científico.

A escala de aplicação de tais ferramentas, que implica em um amplo gradiente de possibilidades de utilização, não obriga nenhum pesquisador a aprender a programar, mas aponta para um caminho onde a formação de equipes multidisciplinares e a compreensão técnica das características dos meios de comunicação, principalmente a internet, pode trazer fundamental diferença nos horizontes a serem vislumbrados.

A presença dos objetos digitais no mundo pode estar constituindo a anomalia de que nos fala Kuhn (2009), potencialmente capaz de gerar uma ampla revisão dos métodos e pressupostos teóricos em que nos baseamos hoje para executar o nosso trabalho. Não necessariamente destruindo o que se tem, mas permitindo que novas espécies epistêmicas, híbridas ou totalmente novas, também possam surgir e prosperar.

## Bibliografia

Bolter, J. ;Grusin, R. (2000).*Remediation*.Understanding New Media.Cambridge: The MIT Press.

Castells, M. (1999). *A sociedade em rede*. São Paulo: Paz e Terra.



- Gleick, J. (2013). *A informação. Uma história, uma teoria, uma enxurrada*. São Paulo, Companhia das Letras.
- Kuhn, T. (2009). *A estrutura das revoluções científicas*. 9ª Ed. São Paulo: Perspectiva.
- Machado, E. (2012). As limitações metodológicas nas pesquisas em Jornalismo. Um estudo dos trabalhos apresentados no GT de Jornalismo da Associação Nacional de Pós Graduação em Comunicação (COMPÓS, 2000-2005). *Anais do 10º Encontro Nacional de Pesquisadores em Jornalismo*. Curitiba. Disponível em <http://soac.unb.br/index.php/ENPJor/XENPJOR/paper/view/2146>>. Acesso em 26 jan. 2015.
- Macluhan, M. (2007). *Os meios de comunicação como extensões do homem*. Trad. Décio Pignatari. São Paulo: Cultrix.
- Manovich, L. (2001). *The language of new media*. Cambridge, MA. Mit Press.
- Moretti, F. (2007). *Graphs, maps, trees*. Abstract models for literary history. New York, Verso.
- Rogers, R. (2013). *Digital Methods*. Cambridge: Mit Press. E-book.
- Scolari, C. (2008). *Hipermediaciones - Elementos para una Teoría da Comunicación Digital Interactiva*. Barcelona: Gedisa Editorial.
- Vargas, M. (1994). *Para uma filosofia da tecnologia*. São Paulo: Alfa e Omega.
- Vieira, J. (2008). *Ontologia*. Formas de Conhecimento: Arte e Ciência. Uma visão a partir da complexidade. Fortaleza : Expressão Gráfica e Editora.
- Santos, M. (2013). Conversando com uma API: um estudo exploratório sobre TV social a partir da relação entre o twitter e a programação da televisão. *Revista Geminis*, ano 4 n. 1, p. 89-107, São Carlos. Disponível em: [www.revistageminis.ufscar.br/index.php/geminis/article/view/129/101](http://www.revistageminis.ufscar.br/index.php/geminis/article/view/129/101)>. Acesso em: 20 abr. 2013.
- Santos, M. (2014). Textos gerados por software. Surge um novo gênero jornalístico. *Anais XXXVII Congresso Brasileiro de Ciências da Comunicação*. Foz do Iguaçu. Disponível em: <http://www.labcomufma.com/biblioteca-digital>>. Acesso em 26 jan. 2014.
- Vilches, L. (2003). *A migração digital*. São Paulo: Loyola.