

AS POTENCIALIDADES DE NARRATIVAS ESTRUTURADAS PARA O JORNALISMO COMPUTACIONAL: COMPETÊNCIAS JORNALÍSTICAS NA ELABORAÇÃO DE TEXTOS GERADOS COM BANCOS DE DADOS*

THE POTENTIAL OF STRUCTURED NARRATIVES FOR COMPUTATIONAL JOURNALISM: JOURNALISTIC SKILLS IN THE ELABORATION OF DATABASE GENERATED TEXTS

Érica Camillo Azzellini

Centro de Pesquisa, Inovação e Difusão em Neuromatemática (CEPID NeuroMat), Brasil
erica.azln@gmail.com

João Alexandre Peschanski

Centro Interdisciplinar de Pesquisa (CIP) da Faculdade Cásper Líbero (FCL), Brasil
japeschanski@casperlibero.edu.br

Fernando Jorge da Paixão

Instituto de Física Gleb Wataghin (IFGW) da Universidade Estadual de Campinas (UNICAMP), Brasil
paixao@ifi.unicamp.br

RESUMO: Neste artigo, explora-se as competências digitais desenvolvidas por jornalistas no contexto informacional do Big Data que levantam a viabilidade de intersecção entre Ciências da Computação e Jornalismo. Nesse sentido, autores propõem diferentes entendimentos sobre o Jornalismo Computacional, campo hipotético no qual a prática jornalística agrega um direcionamento técnico, o que expande o horizonte de entendimento da relação do jornalista com a construção narrativa no ambiente de abundância de dados. Observa-se nesse cenário a emergência de experimentações com narrativas estruturadas, entendidas como textos verbais automatizados a partir de moldes pré-determinados que processam dados de bancos de dados estruturados. Com isso, o artigo reflete sobre os softwares de *Natural Language Generation* (NLG) na composição de notícias e apresenta resultados do desenvolvimento da ferramenta Mbabel para geração de rascunhos estruturados para verbetes temáticos na Wikipédia a partir do banco de dados Wikidata.

PALAVRAS-CHAVE: Jornalismo Computacional; narrativa estruturada; Wikidata; jornalismo digital; Natural Language Generation; Mbabel.

ABSTRACT: In this article, we explore the digital skills developed by journalists in the informational context of Big Data that raise the feasibility of intersection between Computer Science and Journalism. In this sense, authors propose different understandings of the Computational Journalism, a hypothetical field in which the journalistic practice adds a technical direction, which expands the horizon of understanding of the relation of the journalist to the narrative construction in the environment of abundance of data. In this scenario we observe the emergence of experiments with structured narratives, understood as verbal texts automated from predetermined templates that process data from structured

* Esta pesquisa teve apoio da FAPESP, projeto 2013/07699-0.

databases. With this, the article reflects on Natural Language Generation (NLG) software in news composition and presents results of the development of the Mbabel tool for generating structured drafts for thematic entries in Wikipedia from the Wikidata database.

KEYWORDS: Computational Journalism; structured narrative, Wikidata; digital journalism; Natural Language Generation; Mbabel.

1 Introdução

O ecossistema informacional no qual o Big Data está inserido gera mudanças, desafios e oportunidades ao trabalho de jornalistas profissionais, cuja atuação envolve diretamente o tratamento da informação. Para além de alterações nos modelos de negócios do jornalismo e nas formas de interação com o público, notam-se transformações tanto nas práticas jornalísticas quanto nos formatos de suas narrativas ante as dinâmicas das mídias digitais (LIMA JÚNIOR, 2012). Tal percepção desperta uma reflexão sobre as variadas formas de elaboração narrativa nesse ambiente, principalmente no que diz respeito ao seu caráter multidimensional que viabiliza a integração de camadas de *software* com o texto legível por humanos (AZZELLINI, 2018).

Nesse sentido, narrativas originadas com base em bancos de dados estruturados permeiam potencialidades de produção de conteúdo e despertam debates sobre as competências digitais necessárias ao jornalismo profissional contemporâneo. A produção de notícias geradas por *software* se insere no contexto do Jornalismo Computacional que, embora possa ser considerado mais um campo hipotético do que um campo efetivamente estabelecido no jornalismo (STAVELIN, 2013, p. 36), traz perspectivas interessantes na intersecção entre Jornalismo e Ciências da Computação, propondo um direcionamento técnico às atividades jornalísticas.

Um indicativo da preocupação com a técnica no ambiente digital nessa área são as matrizes curriculares de instituições de ensino voltadas para o Jornalismo que não somente têm debatido sobre o impacto da tecnologia no *modus operandi* do jornalismo como também têm incluído matérias e oficinas práticas vinculadas ao uso de plataformas colaborativas de difusão de conhecimento (CARRERA et al., 2016), apontando a confluência entre habilidades jornalísticas de pesquisa e redação com a experimentação de novas plataformas voltadas não necessariamente para a produção de notícia.

No que diz respeito às narrativas dentro da perspectiva do Jornalismo Computacional, vale destacar a emergência de notícias geradas por *softwares* de Natural Language Generation (NLG), que tem adentrado redações e portais de conteúdo em diversos países e que modifica processos de trabalho e própria relação do jornalista com a narrativa (D'ANDRÉA; DALBEN, 2017). Em um sentido mais amplo, as construções de narrativas com suporte em bancos de dados vão além da prática jornalística e podem compor o que chamamos aqui de narrativas estruturadas - textos verbais automatizados a partir de moldes pré-determinados que processam dados de bancos de dados estruturados.

A fim de ilustrar o conceito, trazemos neste trabalho uma experimentação prática com narrativas estruturadas realizada com as plataformas Wikimedia. A partir delas, construiu-se Mbabel, uma ferramenta para geração de rascunhos de verbetes automáticos na Wikipédia a partir de informações armazenadas no Wikidata, o banco de

dados aberto e colaborativo de projetos Wikimedia. Desse processo, foi possível compreender empiricamente o arranjo de camadas diversas que compõem o texto no digital – especialmente em dinâmicas de elaboração do conhecimento de forma colaborativa, associando agentes humanos variados a fatores não-humanos.

Por isso, vale pontuar que o termo “narrativa” é utilizado neste trabalho tanto no sentido de produção textual de cunho jornalístico quanto de teor enciclopédico, uma vez que propomos um olhar para as práticas de narrativas estruturadas com *softwares* de Natural Language Generation (NLG) no jornalismo e para as práticas de automatização de rascunhos de verbetes na Wikipédia via ferramenta Mbabel.

Metodologicamente, recorreremos a uma investigação conceitual sobre narrativas estruturadas ante o contexto de desenvolvimento do Jornalismo Computacional, de forma que também foi realizada uma revisão de literatura sobre esse conceito como base. Por fim, houve uma exploração prática com narrativas estruturadas a partir da construção da ferramenta Mbabel para geração de rascunhos automáticos de verbetes na Wikipédia a partir da integração com o banco de dados aberto Wikidata. A experimentação prática com a ferramenta Mbabel ocorreu no contexto de bolsa do Programa José Reis de Incentivo ao Jornalismo Científico (Mídia Ciência) da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), no Centro de Pesquisa, Inovação e Difusão em Neuromatemática (CEPID NeuroMat).

2 Competências jornalísticas ante o cenário do Jornalismo Computacional

Em um mundo no qual a computação atinge um status de ubiquidade, autores atentam-se para as intersecções possíveis entre o Jornalismo e as Ciências da Computação, refletindo sobre a viabilidade do Jornalismo Computacional (JC) como um campo possível (DIAKOPOULOS, 2012; COHEN et al., 2011; FLEW et al., 2012; HAMILTON; TURNER, 2009). Para além da utilização de recursos computacionais apenas como uma ferramenta acessória de atividades jornalísticas - tais como apuração, elaboração narrativa e publicação - a computação é incorporada no centro de práticas que possuem propósitos jornalísticos:

Jornalismo Computacional é a aplicação de computação e do pensamento computacional às atividades do jornalismo, incluindo coleta de informações, organização e produção de sentido, comunicação e apresentação, difusão e resposta pública às informações das notícias, tudo isso mantendo os valores centrais do jornalismo, como a precisão e a verificabilidade. (DIAKOPOULOS, 2011, p. 1).

Com isso, autores propõem o Jornalismo Computacional como um campo interdisciplinar no qual há a centralidade nos potenciais de práticas e ferramentas computacionais para o jornalismo:

Eu defino o jornalismo computacional aqui como uma vertente do jornalismo tecnologicamente orientado centrado na aplicação da computação e pensamento computacional às práticas de coleta de informações, criação de sentido e apresentação de informações, em vez do uso jornalístico de dados ou métodos

sociais de maneira mais geral (CODDINGTON, 2015, p. 335).

Cabe dentro de propostas de Jornalismo Computacional (AZZELLINI, 2018) a busca pelo desenvolvimento de ferramentas quantitativas próprias que auxiliem na manipulação de dados estruturados e não estruturados. Vale pontuar que a estruturação de dados se dá pelo grau de organização que lhe é conferida a partir de padrões de arranjo e sistematização de valores que lhes são atribuídos. Por outro lado, dados não estruturados não apresentam formas de arranjo mais rígidas ou planejadas, dificultando sua análise e manipulação. Com isso, o jornalista na abundância de dados e no contexto do Jornalismo Computacional deve não somente atender-se para a aptidões de alfabetização midiática – entendida como o conjunto de habilidades necessárias para uma boa competência digital (GUTIERREZ; TYNER, 2012) – como também adotar um posicionamento *technology savvy*, no sentido de que deve ser capaz de se apropriar de variadas tecnologias com o propósito de solucionar problemas cognitivos mais sofisticados que um uso apenas instrumental oferece, ou seja, adquirindo um envolvimento que permite a análise de potencialidades de sua manipulação em um sentido criativo também (WANG et al., 2014).

Além do posicionamento *technology savvy*, o cenário desenhado pelo Jornalismo Computacional inclui a necessidade de pensamento computacional, entendido como uma competência essencial na compreensão do funcionamento de tecnologias e em sua consequente apropriação para fins jornalísticos:

No Jornalismo, por exemplo, a forma de pensar computacionalmente auxilia no entendimento de como funciona uma determinada tecnologia e quais as possíveis apropriações profissionais poder se fazer delas. Além disso, ajuda na compreensão dos processos de inovação disparados a partir da sua estrutura. Pode ser justamente a aquisição do novo conhecimento que se estabelece a ponte entre as funções do Jornalismo, na atual sociedade, e a capacidade de obter informações de relevância social por intermédio de sistemas computacionais (LIMA JÚNIOR, 2012, p. 213).

Com isso, o Jornalismo Computacional abre margem para que haja também o engajamento direto de jornalistas na análise de Arquitetura da Informação, em experimentações com Inteligência Artificial, na manipulação e estruturação de banco de dados, no uso e construção de APIs em aplicações, além de participação em processos de outras tecnologias de convergência midiática. Por isso, o jornalista assume uma posição na qual é habilitado a explorar camadas mais profundas na geração de conteúdo multimídia, de forma que passa a pensar o próprio jornalismo em termos de plataforma e de modelos replicáveis (STAVELIN, 2013), e não apenas como conteúdo final em si.

Nesse ponto, vale destacar que para os autores estudados, o Jornalismo Computacional difere de outros modelos de jornalismo digital - tais como Jornalismo de Dado, Jornalismo Guiado por Dados e Jornalismo de Base de Dados, por exemplo - principalmente no que diz respeito à sua ênfase na criação, adaptação ou uso de ferramentas computacionais aliadas a métodos jornalísticos como jornalismo em si. Requer, portanto, habilidades de programação e de pensamento computacional. Podem participar do ecossistema de desenvolvimento e elaboração de ferramentas usadas com propósitos jornalísticos ou semelhantes, assumindo uma atitude mais participativa e menos instrumental com a computação (STAVELIN, 2013, p. 54).

Além disso, devido à sua abrangência de abordagens técnicas e narrativas possíveis que abrigam agentes variados tais como programadores, empreendedores, jornalistas e pesquisadores, há espaço no JC para distinções entre áreas específicas. A aplicação de robótica no jornalismo ou experimentações com Inteligência Artificial e *machine learning*, por exemplo, trazem possibilidade de investigações específicas dentro do conceito de Jornalismo Computacional aqui apresentado. A prática de elaboração de conteúdos jornalísticos baseados em *software* de *Natural Language Generation* (NLG) chamada de Jornalismo Algorítmico (DÖRR, 2016) ou de Jornalismo Automatizado (GRAEFFE, 2016) pode ser entendida como um dos desdobramentos do Jornalismo Computacional especialmente pela sua repercussão em ambientes editoriais (D'ANDRÉA; DALBEN, 2017), trazendo oportunidades de exploração do que chamamos, a seguir, de “narrativas estruturadas”.

3 Narrativas estruturadas no JC: textos gerados com bancos de dados

O trabalho jornalístico envolve essencialmente o tratamento da informação, de forma que no ambiente do Big Data é necessário cada vez mais que os jornalistas desenvolvam competências de manipulação de dados em diferentes graus de estruturação (LINDEN, 2018). O levantamento e seleção de informações de interesse público em bases de dados, além da semantificação desses dados - ou seja, a atividade de organização padronizada de atribuição de sentido a determinados dados, de forma que sejam compreendidos por *softwares* e algoritmos - acabam, conseqüentemente, por integrar as práticas jornalísticas. O entendimento de narrativa para jornalistas é ampliado na medida em que interagem com interfaces digitais mais profundas que a camada narrativa final acessada pelo público, em uma espécie de “*backend* jornalístico” (BERTOCCHI, 2013). Tal proposta envolve o agenciamento entre fatores humanos e não-humanos na elaboração de narrativas, confluindo em experimentações de teor jornalístico no meio digital.

Vale ressaltar aqui a aproximação entre as habilidades jornalísticas mencionadas na seção anterior e o funcionamento de bancos de dados estruturados para a geração de textos no que chamamos de “narrativas estruturadas”. As narrativas estruturadas podem ser entendidas como o desenvolvimento de textos verbais, compreensíveis por humanos, automatizados a partir de moldes pré-determinados que processam dados oriundos de bancos de dados estruturados (AZZELLINI, 2018).

O grau de estruturação de bancos de dados, ou seja, de organização semântica da informação, viabiliza sua apropriação na elaboração de textos automáticos compreensíveis por humanos, proporcionando de certa forma um diálogo entre a camada humana e as camadas de *software* e algoritmos. O Wikidata, banco de dados utilizado por projetos como Wikipédia e Wikimedia Commons, é um exemplo de organização semântica de dados que favorece a extensão de seu uso, permitindo que, além dos projetos Wikimedia, aplicações de terceiros também possam se aproveitar deles (VRANDEČIĆ; KRÖTZSCH, 2014), conforme será discutido em mais detalhes na próxima seção.

A lógica de funcionamento de ferramentas e *softwares* de narrativas estruturadas se dá basicamente em cinco etapas: 1. *Coleta de dados*: o *software* realiza uma coleta de

dados a partir de fontes pré-definidas. Por isso a importância de bancos de dados estruturados, que colocam padrões para essa extração; 2. *Identificação de eventos de interesse*: métodos estatísticos identificam nesse conjunto de dados as informações que constituem um evento de interesse; 3. *Priorização*: com regras pré-determinadas, o software é capaz de priorizar detalhes do evento de interesse, seguindo critérios de noticiabilidade, por exemplo; 4. *Geração de narrativa*: a partir de especificações de enredo embutidas no *software*, os dados extraídos e filtrados são usados para gerar a narrativa estruturada; e 5. *Publicação*: o sistema publicador do jornal ou do portal de conteúdo recebe a narrativa estruturada, podendo ser publicada automaticamente ou depois da aprovação editorial, e é então acessada pelo público (GRAEFE, 2016).

Os *softwares* de Natural Language Generation (NLG) podem ser enquadrados no conceito de narrativas estruturadas, uma vez que geram textos de forma automática em linguagem bastante aproximada à humana a partir de informações oriundas de bancos de dados estruturados. Ambientes de produção de notícia, como Associated Press, Forbes, Le Monde, Los Angeles Times, ProPublica, The Washington Post e Yahoo, já usam os *softwares* de NLG Narrative Science, Syllabs, AX Semantics, Aexea e Quakebot para a geração de conteúdos (D'ANDRÉA; DALBEN, 2017), em especial para aqueles que possuem uma estrutura narrativa muito semelhante. É o caso de notícias sobre balanços financeiros, clima, esportes e trânsito. É possível observar que, ao mesmo tempo em que aumentam a eficiência de redações e a satisfação profissional pela automação de tarefas monótonas e propensas a erros, também suscitam desconfiança de profissionais que temem a redução do número de empregos no setor (LINDEN, 2018, p.7), embora também seja um fator que adiciona novas possibilidades de atuação:

Neste complexo processo de produção de notícias automatizadas, é importante salientar que, junto aos actantes não-humanos, há uma extensa rede de actantes humanos como programadores, jornalistas e cientistas de dados que desenvolvem os algoritmos, criam as templates das notícias, alimentam as tabelas com dados e informações, gerenciam e analisam os textos gerados de forma automática. E, a partir destas especificidades, novas atividades jornalísticas vão emergindo (D'ANDREA; DALBEN, 2017, p. 134).

A automação no jornalismo envolve a “substituição de atividades que dependem de reconhecimento de padrões, coleta e extração de dados e algoritmos” (LINDEN, 2018, p.11), de forma que tarefas mais analíticas e criativas ainda precisarão do respaldo do fator humano. O software, portanto, atua como fator suplementar e complementar no Jornalismo Computacional (CASWELL; DÖRR, 2018, p. 6), abrindo caminhos para a criatividade de agentes humanos em termos narrativos.

Nesse sentido, é possível também observar a experimentação com o conceito de narrativas estruturadas na Wikipédia que, embora não seja um veículo jornalístico nem tenha esse objetivo, é uma plataforma relevante na lógica de consumo de informação contemporânea, utilizada para obtenção de conteúdos de cunho enciclopédico sobre variadas temáticas. Por conta da estrutura aberta e colaborativa dos projetos Wikimedia, é possível fazer uso de seus recursos e ilustrar o funcionamento de narrativas estruturadas a partir da ferramenta Mbabel, desenvolvida no Centro de Pesquisa, Inovação e Difusão em Neuromatemática (CEPID NeuroMat), no contexto da bolsa José Reis de Incentivo ao Jornalismo Científico (Mídia Ciência) da Fundação de Amparo à Pesquisa do Estado de

São Paulo (FAPESP), conforme será tratado a seguir.

4 Mbabel: narrativas estruturadas na Wikipédia com Wikidata

Conforme mencionado acima, é possível entender narrativas estruturadas como organizações textuais automatizadas, compreensíveis por humanos, geradas a partir de arranjos previamente determinados que fazem uso de informações provenientes de bancos de dados estruturados. Dessa forma, durante projeto de pesquisa¹ vinculado ao Programa José Reis de Incentivo ao Jornalismo Científico (Mídia Ciência) da FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo) no CEPID NeuroMat (Centro de Pesquisa, Inovação e Difusão em Neuromatemática), utilizou-se o Wikidata como banco de dados estruturado para o desenvolvimento da ferramenta Mbabel² para geração do que foi chamado de "rascunhos estruturados" na Wikipédia, seguindo a lógica do conceito de narrativas estruturadas.

O Wikidata é uma base de conhecimento colaborativa, livre, multilíngue e secundária, lançada em 2012 pela Fundação Wikimedia (VRANDEČIĆ, 2013). Isso significa que se trata de um banco de dados aberto que pode ser editado tanto por humanos quanto por robôs, operando sob a licença Creative Commons, com disponibilização imediata de informações em variados idiomas e armazenamento de conexões com outros bancos de dados. Além disso, o Wikidata é integrado aos outros projetos Wikimedia, tais como a Wikipédia e o Wikimedia Commons, e por projetos terceiros via API³. Além da sua capacidade de armazenar uma quantidade massiva de dados, o Wikidata mostra-se relevante pelo potencial de servir como base para o desenvolvimento de serviços e aplicações que, por seu aspecto aberto e colaborativo, podem ser criados e utilizados por qualquer pessoa ou instituição no mundo, além de servir como suporte para pesquisa acadêmica.

Os dados no Wikidata são estruturados de forma multilíngue em entidades de *itens* e *propriedades*, sendo que cada entidade possui sua própria página. Os itens são marcados com a letra Q, seguida por uma sequência numérica única (Qid), e as propriedades pela letra P, também seguida por sua própria sequência de números (Pid). Seguindo um modelo semântico, uma *declaração* no Wikidata descreve um determinado item, consistindo em uma propriedade P com um valor Q. Assim, a organização se dá no seguinte modelo: tendo Terra (Q2) como item e ponto mais alto (P610) como propriedade, obtém-se Monte Everest (Q513) como valor, gerando uma declaração que descreve o ponto mais alto do planeta Terra, em uma estrutura semântica de armazenamento de dados. Além disso, uma declaração pode conter *qualificadores*, que fornecem mais contexto sobre a declaração, seguindo a estrutura de propriedade P associada a um valor Q. Um item no Wikidata pode conter, ainda, *identificadores*, que são propriedades que o ligam a uma base de dados externa, potencializando a capacidade de convergência do Wikidata.

1 Processo 2018/07286-0.

2 Ferramenta Mbabel, disponível em: <<https://pt.wikipedia.org/wiki/Wikipédia:Mbabel>>. Acesso em 17 dez. 2018.

3 Wikidata: Introdução. Disponível em: <<https://www.wikidata.org/wiki/Wikidata:Introduction/pt-br>>. Acesso em 17 dez. 2018.

Terra (Q2)

terceiro planeta mais próximo do Sol no Sistema Solar

 editar

Mundo | Planeta Azul | Globo Terrestre | Planeta Terra | A Terra | Gaia |  | Globo | Tellus | Sol III | Sol d |  |  |  | 

Declarações

ponto mais alto P610	 Monte Everest Q513  editar
P2044 elevação acima do nível do mar	8 848±20 metro
P459 método de determinação	nível médio do mar Q125465
▼ 0 referência	
+ adicionar referência	

Figura 1: Exemplo de estruturação semântica no Wikidata com o item Terra (Q2).
Fonte: Captura de tela editada pelos autores.

A partir dessa organização semântica no Wikidata, foi possível implementar uma predefinição em Lua no domínio da Wikipédia para geração de textos automáticos, conforme a noção de narrativa estruturada apresentada neste trabalho. A ferramenta Mbabel desenvolvida no CEPID NeuroMat é uma adaptação e ampliação da aplicação homônima de *template* para rascunhos de verbetes sobre obras de arte elaborada pelo usuário Pharos⁴, envolvido no WikiProject Metropolitan Museum of Art⁵.

A proposta da ferramenta Mbabel é auxiliar o usuário da Wikipédia nos primeiros passos de edição de um verbete temático, disponibilizando uma estrutura textual básica e sugestões de organização do texto dentro de um determinado assunto. Verbetes sobre filmes, por exemplo, seguem uma estrutura narrativa semelhante, uma vez que precisam apresentar informações essenciais como quando e onde o filme foi lançado, por quem foi dirigido e produzido, quais atores fizeram parte do elenco, por exemplo. De forma semelhante, no caso de obras de arte, os verbetes costumam trazer informações como criador da obra, data de criação, suas dimensões, número de inventário, a qual coleção pertence etc. Tais informações são cadastradas e armazenadas no Wikidata como propriedades e seus correspondente valores atuam como variáveis em um texto genérico de *template*, gerando automaticamente um texto no estilo rascunho de verbete, pronto para ser editado pelo usuário.

4 Página do usuário Pharos na Wikipédia, disponível em: <<https://en.wikipedia.org/wiki/User:Pharos>>. Acesso em: 17 dez. 2018.

5 Sobre WikiProject Metropolitan museum of Art. Disponível em: <https://en.wikipedia.org/wiki/Wikipedia:GLAM/Metropolitan_Museum_of_Art>. Acesso em: 17 dez. 2018.

The screenshot shows the Wikipedia page for the Mbabel tool. The page title is 'Wikipédia:Mbabel'. The main content area is titled 'Página inicial' and contains a welcome message: 'Bem-vindo(a) à página da ferramenta Mbabel'. Below this, there is a section titled 'Objetivos' with three numbered points: 1. 'Auxiliar o usuário nos primeiros passos para a criação de um artigo na Wikipédia por meio da geração de um rascunho estruturado necessariamente em sua própria página de testes.' 2. 'A criação de rascunhos estruturados eficientes depende da criação de um template prévio, capaz de lidar com as informações provenientes de uma determinada variedade de itens no Wikidata, tais como obras de arte, jornais, museus, sismos etc.' 3. 'Incentivar a integração entre os projetos Wikipédia e Wikidata. Em especial, a ferramenta pretende incentivar melhorias na estruturação de dados no Wikidata.'

Figura 2: Página inicial da ferramenta Mbabel.
Fonte: Captura de tela realizada pelos autores.

Com isso, nota-se que dentro de uma temática específica para verbetes, há possíveis propriedades em comum, cujo valor pode ser coordenado em um texto de rascunho de verbete quando associado com uma estrutura textual mais elaborada. Assim, os rascunhos estruturados são genéricos, devido à sua capacidade de adaptação a diferentes Qids dentro de um mesmo tema; pré-estabelecidos, uma vez que a ferramenta funciona a partir de um texto previamente inserido no *template* e editáveis, já que é fundamental na lógica das dinâmicas da Wikipédia que todos os verbetes apresentem um patamar mínimo de qualidade⁶ acordado pela comunidade antes de serem publicados, de forma que a ferramenta nesse caso não substitui a edição humana.

Além da geração de frases com base em dados cadastrados no Wikidata, o rascunho estruturado do Mbabel pode conter sugestões de seções para os editores, referências elaboradas a

partir de identificadores, infocaixa e listas de conteúdos, de tal forma a facilitar o processo de elaboração de verbete para o usuário e integrar diferentes projetos Wikimedia.

Vale pontuar também que, conseqüentemente, quanto mais propriedades sobre um determinado item estiverem cadastradas no Wikidata, mais completo será o rascunho estruturado gerado pelo Mbabel e menos trabalho manual o editor terá. As frases incluídas no verbete só são invocadas por completo quando há a existência do valor da propriedade cadastrado no item correspondente do Wikidata, uma vez que as frases disponíveis no rascunho do Mbabel são planejadas de acordo com a estrutura semântica do Wikidata.

Até o momento, foram criados *templates* para verbetes dos temas: museus, obras de arte, filmes, livros, jornais e sismos. Está em andamento o desenvolvimento de uma ampliação da ferramenta para criação de verbetes sobre eleições municipais e eleições estaduais brasileiras, envolvendo a manipulação de mais de um Qid por chamada e um

6 Guia de edição de verbetes na Wikipédia: <https://pt.wikipedia.org/wiki/Ajuda:Guia_de_edição/Como_começar_uma_página>. Último acesso em: 17/12/2018.

grau mais elevado de estruturação semântica no Wikidata. A coordenação dos itens de eleições no Wikidata irão viabilizar um rascunho estruturado mais completo e com frases mais complexas, com texto introdutório indicando a data e local da eleição; os candidatos e seus respectivos vices, partidos e coligações; a ocorrência ou não de um segundo turno e seus desdobramentos; a porcentagem e o número total de votos válidos, nulos e brancos, além da quantidade de eleitores aptos e de abstenções em cada turno. Também são geradas automaticamente tabelas com os resultados das eleições e uma infocaixa. A referência utilizada nesse caso é o Tribunal Superior Eleitoral, instância jurídica vinculada à administração eleitoral estadual e municipal no Brasil.

Para gerar o rascunho estruturado, o usuário pode acessar dentro da Wikipédia em português a página do Mbabel, na qual encontra informações sobre o propósito da ferramenta e sua forma de utilização, passo a passo. Por se tratar de uma plataforma colaborativa com uma comunidade ativa e engajada, também estão disponibilizadas instruções para o usuário que queira elaborar um rascunho estruturado para outra temática que ainda não foi desenvolvida. Para tal, recomenda-se o estudo dos Wikidata Labs, em especial a primeira⁷ e a nona⁸ edição, treinamentos técnicos de Wikidata realizados no CEPID NeuroMat entre 2017 e 2018 com o objetivo de promover competências na integração entre Wikidata e outros projetos Wikimedia, em especial a Wikipédia.

Na aba de *templates*, o usuário encontra detalhes sobre rascunho estruturado disponível para uso, sendo que dentro de cada temática há um texto explicativo citando algumas propriedades disponíveis no *template*, indicando que tipo de informação as frases a serem geradas pela ferramenta poderão conter, quais seções são indicadas na formatação do verbete como sugestão e recursos adicionais, tal qual no caso de museus apresentado abaixo:

A narrativa estruturada de **museus** deve ser utilizada como suporte para a elaboração de verbetes sobre museus em sua página de testes. Ela gera automaticamente um rascunho de verbete que deverá conter informações como local, tamanho do acervo, número de visitantes anuais, atual curador, estilo arquitetônico, prêmios recebidos, situação patrimonial etc., de acordo com dados disponíveis em seu respectivo item no Wikidata. Além disso, esse rascunho estruturado contém um Listeria com até 100 peças de acervo expostas no museu. As seções sugeridas para esse tipo de verbete são descrição, história, acervo, arquitetura e administração⁹.

Listeria¹⁰ é uma ferramenta criada por Magnus Manske, cientista sênior no

7 Material sobre o Wikidata Lab I: Como automatizar infocaixas. Disponível em: <https://pt.wikipedia.org/wiki/Wikip%C3%A9dia:Edit-a-thon/Atividades_em_portugu%C3%AAs/Wikidata_Lab_I>. Acesso em: 13 dez. 2018.

8 Material sobre o Wikidata Lab IX: Narrativas estruturadas na Wikipédia. Disponível em: <https://pt.wikipedia.org/wiki/Wikip%C3%A9dia:Edit-a-thon/Atividades_em_portugu%C3%AAs/Wikidata_Lab_IX>. Acesso em: 13 dez. 2018.

9 Texto explicativo sobre o rascunho estruturado para museus na ferramenta Mbabel. Disponível em: <<https://pt.wikipedia.org/wiki/Wikip%C3%A9dia:Mbabel/Templates>>. Acesso em: 13 dez. 2018.

10 Sobre Listeria, ferramenta de geração de listas automáticas na Wikipédia via Wikidata queries: <<http://magnusmanske.de/wordpress/?p=301>>. Acesso em: 1 fev. 2019.

Wellcome Trust Sanger Institute e responsável pelo desenvolvimento de uma das primeiras versões do *software* MediaWiki, no qual a Wikipédia é executada. O Listeria gera listas automáticas de conteúdos na Wikipédia a partir de *queries* no Wikidata, de forma que o conteúdo na enciclopédia torna-se dinâmico na medida em que se atualiza de acordo com as entradas no Wikidata. No caso do *template* de museus no Mbabel, há a inserção da ferramenta para a geração de uma lista com até 100 itens de seu acervo, contendo imagens carregadas no Wikimedia Commons e informações como título da obra, criador, data de criação, tipo de obra, material utilizado e a coleção a qual pertence, tornando o rascunho de verbete mais completo em termos de conteúdo e incrementando as formas de visualização da informação possíveis na Wikipédia.

Para o usuário fazer uso do Mbabel, basta inserir o Qid de um item que ainda não possua um verbete na Wikipédia em português na caixa de texto do botão do Mbabel e o nome do *template* desejado ao lado, separado por uma vírgula, como no exemplo abaixo para geração de um rascunho estruturado de museu:

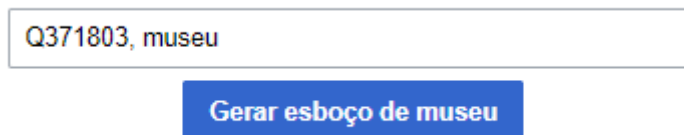


Figura 3: Botão da ferramenta Mbabel para geração de rascunho estruturado sobre museu.
Fonte: Captura de tela realizada pelos autores.

Após o usuário clicar para gerar o esboço, há a indicação para que o usuário clique em “mostrar previsão” para que possa visualizar o texto em forma de linguagem natural em sua página de testes. O código presente na tela intermediária não deve ser alterado pelo usuário, uma vez que se trata de uma predefinição que chama o *template* escolhido de acordo com a seleção realizada pelo botão.

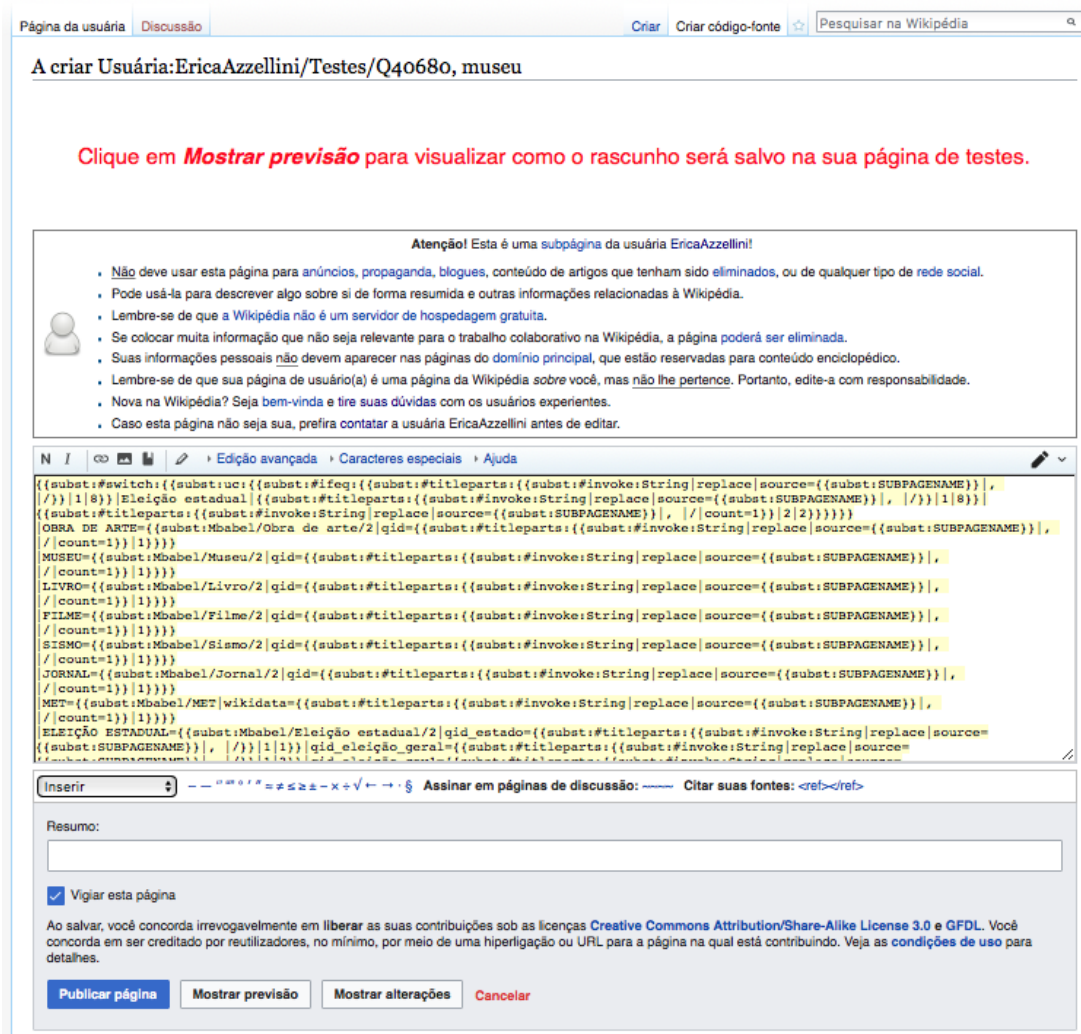


Figura 4: Tela intermediária na geração de rascunho automático na Wikipédia via Mbabel. O usuário deve clicar em “mostrar previsão” para poder visualizar o texto em linguagem natural.
Fonte: Captura de tela realizada pelos autores.

Em seguida, o usuário é direcionado para a sua página de testes, na qual pode visualizar uma previsão de como o verbete está estruturado, já com as informações do Wikidata inseridas no *template* escolhido em linguagem natural. Ao clicar em “publicar página”, o usuário pode começar a editar o verbete em sua página de testes pelo código-fonte ou por um editor visual da Wikipédia, sem precisar lidar necessariamente com códigos de programação. Depois que ampliar o rascunho estruturado com mais informações e contexto e realizar as devidas revisões no verbete, seguindo os padrões exigidos pela Wikipédia, o editor pode publicar o texto do verbete no domínio principal.

A partir do desenvolvimento da ferramenta Mbabel, foi possível disponibilizar para a comunidade Wikimedia uma forma dinâmica de concepção de rascunhos de verbetes na Wikipédia, integrando diferentes plataformas Wikimedia e explorando em especial as potencialidades dos recursos do Wikidata. Assim como outras aplicações de NLG voltadas para o jornalismo, a ferramenta Mbabel proporciona a redução de trabalho manual e repetitivo na elaboração textual, dinamizando processos e apresentando um entendimento renovado do texto na plataforma ao explorar sua multidimensionalidade.

A ampliação do uso de recursos vinculados ao Wikidata em outros projetos Wikimedia, em especial a Wikipédia, entra em um contexto de exploração de dinâmicas possíveis na *web semântica* em plataformas colaborativas. Os usuários podem, então, interagir com o texto em diferentes camadas, de forma que é possível ampliar a ideia do próprio texto em si para as camadas de *software* que interagem com o que é legível também por humanos. Ou seja, o desenvolvimento da ferramenta Mbabel ilustra a elasticidade das narrativas estruturadas em um ambiente de construção de conhecimento colaborativa - que agora agrega não somente a multiplicidade do fator humano, marcada pelas práticas de cooperação entre usuários na elaboração de um verbete na Wikipédia -, como também carrega em si uma interação com o texto originado em código de programação.

5 Considerações finais

Ao observar as práticas do jornalismo ante o contexto informacional do Big Data, nota-se que a viabilidade de intersecção entre Ciências da Computação e Jornalismo abre margem para o desenvolvimento de um jornalismo que, diferente de outras formas de jornalismo digital, não somente se volta para os métodos e as estruturas narrativas nesse ambiente no que diz respeito a formato, como também propõe um direcionamento técnico a seus objetivos jornalísticos. Nesse sentido, a relação do jornalista com a construção narrativa no ambiente de abundância de dados passa por alterações em seus formatos tradicionais, de forma que o jornalista pode ultrapassar a materialidade textual ao lidar com processos de organização de dados estruturados e com o desenvolvimento de plataformas e ferramentas pelo viés jornalístico.

Dentro do Jornalismo Computacional, é possível observar a emergência de explorações baseadas em narrativas estruturadas, principalmente para automatização de notícias e de outros conteúdos com base em bancos de dados estruturados. Nesse sentido, processos de automatização de textos com narrativas semelhantes, seja com *softwares* de *Natural Language Generation* (NLG), no caso de notícias, ou com a ferramenta Mbabel, no caso de rascunhos estruturados de verbetes da Wikipédia, sinaliza a abertura para experimentação com o conceito de narrativas estruturadas e suas consequentes possibilidades no contexto do Jornalismo Computacional, sugerindo ao jornalista a exploração de níveis de elaboração de conteúdos nas camadas de código e de organização semântica de dados, o que requer um posicionamento *technology savvy* do jornalista.

A experimentação com a ferramenta Mbabel serve como um exercício para exemplificar a aplicação do conceito de narrativa estruturada, entendida aqui como a elaboração de textos verbais automatizados a partir de moldes pré-determinados que processam dados de bancos de dados estruturados – no caso, do Wikidata. A partir de seu desenvolvimento, foi possível disponibilizar para a comunidade Wikimedia uma forma dinâmica de concepção de rascunhos de verbetes na Wikipédia, integrando diferentes plataformas Wikimedia e explorando em especial as potencialidades dos recursos do Wikidata.

Referências

- AZZELLINI, E. C. O conceito de Jornalismo Computacional ante as potencialidades das narrativas estruturadas. In: *41º Congresso Brasileiro de Ciências da Comunicação*, 2018, Joinville, Anais, INTERCOM. Disponível em: <http://portalintercom.org.br/anais/nacional2018/resumos/R13-1018-1.pdf>. Último acesso em: 15 fev. 2019.
- BERTOCCHI, D. *Dos dados aos formatos: um modelo teórico para o design do sistema narrativo no jornalismo digital*. (Tese de Doutorado). ECA/USP, 2013. Disponível em: <https://tinyurl.com/y95b8maz>. Acesso em: 8 mai. 2018.
- CARRERA, M. R.; DIELO, M. ; MORAES, R. ; PESCHANSKI, J. A. . A Wiki-Pedagogia no Jornalismo: O Caso do Projeto Wikipédia da Faculdade Cásper Líbero. *Rebej* (Brasília), v. 6, p. 75-100, 2016. Disponível em: <https://tinyurl.com/y6vx3uol>. Acesso em: 28 jun. 2018.
- CASWELL, D.; DÖRR, K. Automated Journalism 2.0: Event-Driven Narratives. In: *Journalism Practice*, v. 12, n. 4, p. 477-496, 2018. Disponível em: <https://tinyurl.com/y78utqqj>. Acesso em: 8 jun. 2018.
- CODDINGTON, M. Clarifying Journalism's Quantitative Turn: A typology for evaluating data journalism, and computer-assisted reporting. *Digital Journalism*, v. 3, n. 3, 331-348, 2015. Disponível em: <https://tinyurl.com/ycdevz6l>. Acesso em: 8 mai. 2018.
- COHEN, S.; LI, C.; YANG, J.; YU, C. Computational Journalism: A Call to Arms to Database Researchers. In: *5th Biennial Conference on Innovative Data Systems Research*, 2011. Disponível em <https://tinyurl.com/yabeywql>. Acesso em: 8 mai. 2018.
- D'ANDRÉA, C. F. B.; DALBEN, S. F. Redes Sociotécnicas e Controvérsias na Redação de Notícias por Robôs. In: *Comunicação e Cultura*, v. 15, n. 1, p. 118 - 140, 2017. Disponível em: <https://tinyurl.com/yar8mrrq>. Acesso em: 8 mai. 2018.
- DIAKOPOULOS, N. Computatinal Journalism and the Emergence of News Platforms. In: *The Routledge Companion to Digital Journalism Studies*, 2016. Disponível em: <https://tinyurl.com/ycml4um4>. Acesso em: 8 mai. 2018.
- DÖRR, K. Mapping the field of Algorithmic Journalism. In: *Digital Journalism*, v. 4, n. 6, p. 700-722, 2016. Disponível em: <https://tinyurl.com/yca2g3y7>. Acesso em: 8 mai. 2018.
- FLEW, T.; SPURGEON, C., DANIEL, A.; SWIFT, A. The promise of Computational Journalism. In: *Journalism Practice*, v. 6, n. 2, p. 157-171, 2012. Disponível em: <https://tinyurl.com/y98z3gq3>. Acesso em: 8 mai. 2018.
- GRAEFE, A. Guide to Automated Journalism. In: *Tow Center for Digital Journalism*, Columbia Journalism School, 2016. Disponível em: <https://tinyurl.com/ycgtguue>. Acesso em: 8 mai. 2018.
- HAMILTON, J.; TURNER, F. Accountability Through Algorithm: Developing the Field of

Computational Journalism. In: *Behavioral Sciences Summer Workshop*, Stanford, CA, 2009. Disponível em: <https://tinyurl.com/y7yy9c4h>. Acesso em: 8 mai. 2018.

GUTIÉRREZ, A.; TYNER, K. Media Education, Media Literacy and Digital Competence. In: *Comunicar*, v. 19, n. 38, p. 31-39, 2012. Disponível em: <https://tinyurl.com/ycwqj7xj>. Acesso em: 8 mai. 2018.

LIMA JÚNIOR, W.T. Big Data, Jornalismo Computacional e Data Journalism: estrutura, pensamento e prática profissional na Web de dados. In: *Estudos em Comunicação*, n. 12, p. 207-222, dez. 2012.

LINDEN, C. J. Algoritmos para jornalismo: o futuro da produção de notícias. In: *Líbero*, ano XX1, n. 41, p. 5-27, jan-jun 2018.

STAVELIN, E. *Computacional Journalism: When Journalism Meets Programming*. 2013. 107 f. Tese (Doutorado) - Curso de Information Science And Media Studies, University Of Bergen, Noruega, 2013.

VRANDEČIĆ, Denny. The Rise Of Wikidata. *IEEE Intelligent Systems Magazine*, v. 28, n. 4, p. 90-95, 2013.

VRANDEČIĆ, Denny; KRÖTZSCH, Markus. Wikidata: A Free Collaborative Knowledgebase. *Communications of the ACM*, New York, v. 57, n. 10, out. 2014, p. 78-85. Disponível em: <https://tinyurl.com/y8bf9q68>. Acesso em: 8 mai. 2018.

WANG et al. An investigation of middle school science teachers and students use of technology inside and outside of classrooms: considering whether digital natives are more technology savvy than their teachers. In: *Educational Technology Research and Development*, v. 62, n. 6, p. 637-662, 2014. Disponível em: <https://tinyurl.com/y7u4es5r>. Acesso em: 8 mai. 2018.

Recebido em dia 31 de dezembro de 2018.
Aprovado em dia 08 de março de 2019.