

POTENCIAIS PARA O JORNALISMO COMPUTACIONAL NA ESTRUTURAÇÃO DE
NARRATIVAS DE DIFUSÃO CIENTÍFICA BASEADAS EM WIKIDATA

POTENTIAL FOR COMPUTATIONAL JOURNALISM IN STRUCTURING SCIENCE
DISSEMINATION NARRATIVES BASED ON WIKIDATA

Érica Camillo Azzellini

Supervisor científico: Prof. Dr. Fernando J. da Paixão

Supervisor junto à empresa: Prof. Me. João Alexandre Peschanski

São Paulo
2018

Potenciais para o Jornalismo Computacional na estruturação de narrativas de difusão científica baseadas em WikiData

O projeto enquadra-se na linha de atuação do CEPID Neuromat no desenvolvimento de atividades de difusão científica. Propõe-se a experimentar as potencialidades de estruturação de narrativas jornalísticas a partir do banco de dados Wikidata, a fim de verificar as novas possibilidades de atuação para a prática jornalística científica no contexto das novas mídias digitais. Teoricamente, visa aprofundar a reflexão sobre os conceitos de Jornalismo Computacional e de narrativas estruturadas diante do ecossistema informacional vinculado ao Big Data e à expansão das TICs.

Palavras-chave: Jornalismo computacional, narrativa estruturada, Wikidata, Web semântica

Potential for Computational Journalism in structuring science dissemination narratives based on WikiData

This project is part of the CEPID Neuromat's activity in the development of scientific dissemination activities. It proposes to experiment journalistic structured narratives potential from Wikidata database, in order to verify new possibilities for a science-journalism practice in the context of new digital media. Theoretically, it aims at deepening the current reflection on the concepts of Computational Journalism and structured narratives in the information ecosystem based on Big Data and the expansion of ICTs.

Keywords: Computational Journalism, structured narratives, Wikidata, Web Semantics

1) Título e resumo do conjunto de projetos ou programas de pesquisa científica ou tecnológica que será objeto da pesquisa, identificando as fontes de financiamento desses projetos.

O objeto do projeto de pesquisa de jornalismo científico aqui proposto é o Centro de Pesquisa, Inovação e Difusão em Neuromatemática - CEPID Neuromat (2013/076990), financiado pela Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) e localizado na Universidade de São Paulo (USP).

Apesar de a FAPESP ser a principal fonte de financiamento, o NeuroMat também conta com as seguintes fontes de financiamento: NUMEC/MaCLinC (USP), FINEP PROINFRA, FAPERJ HOSPITAIS UNIVERSITÁRIOS, FAPERJ SEDIADAS 210.737/2014, Edital Universal MCT/CNPq/14/2012 480108/2012-9, CAPES NUFFIC, IBRO (International Brain Research Organization), Edital Universal MCT/CNPq 14/2014- 459335/2014-6, PROBAL CAPES/DAAD - Proc. 430/15, CAPES (PAEP 8621/2013-32), Edital Universal MCT/CNPq 14/2012 478537/2012-3, CNPq (Eventos Nacionais 456821/2013- 9), CNPq-BPP, CNPq, CAPES, USP, UNICAMP, UFRN, IMPA, UFRJ, UFABC, UFSCar, UFMG, UFPE e UFOP.

Em consonância com a missão dos CEPIDs da FAPESP, o NeuroMat propõe a criação de um centro de matemática que integre a modelagem matemática com pesquisa básica e aplicada na fronteira da neurociência para responder à crescente importância da matemática na neurociência teórica.

A criação do centro teve como motivação a grande quantidade de dados capazes de serem gerados atualmente nos laboratórios, cuja análise depende de novos modelos matemáticos, e o papel do desenvolvimento da linguagem e das estruturas matemáticas para a elaboração de teorias que expliquem fatores experimentais e sugiram previsões que possam ser testadas. Hoje, a neurociência vive uma situação de desequilíbrio, entre uma grande capacidade de produção de dados experimentais e uma insuficiente capacidade de compreensão teórica, em que a matemática é fundamental para fazer a ponte entre dados e explicação.

O projeto visa, ainda, construir um centro de pesquisa avançada em neurociência teórica, reunindo uma equipe de ponta composta de matemáticos, cientistas da computação, neurocientistas e clínicos especialistas em reabilitação. O NeuroMat conta com um plano de

transferência de tecnologia e de inovação, que inclui o desenvolvimento de produtos para a saúde pública em neuroreabilitação, e um plano de difusão de conhecimento científico, que inclui a criação de cursos e oficinas destinados a estudantes, professores e jornalistas.

Portanto, com base no conceito de jornalismo computacional, o projeto de pesquisa de jornalismo científico proposto para o Programa José Reis de Incentivo ao Jornalismo Científico (Mídia Ciência) tem o objetivo de contribuir para o plano de difusão científica do NeuroMat, com especial interesse nas potencialidades de estruturação narrativa baseada na plataforma Wikidata.

A proposta apresentada está relacionada ao contexto de transformações na prática jornalística vinculadas ao novo ecossistema informativo introduzido pela disseminação das Tecnologias da Informação e Comunicação (WT LIMA JÚNIOR, 2011; RIBEIRO, 2013; OLIVEIRA, 2013). O Jornalismo Computacional, entendido como a capacidade de o jornalista integrar às suas práticas conhecimentos tradicionalmente restritos às Ciências da Computação, é uma forma de os profissionais da área se apropriarem das potencialidades desse conjunto massivo de dados gerados pelas TICs, o Big Data, de forma a explorar diferentes maneiras de interação com fontes de informação e experimentar novos recursos narrativos (CODDINGTON, 2015). Dentre os possíveis objetos do jornalista no contexto do Big Data, estão os acervos digitais, bancos de dados abertos e projetos para o conhecimento colaborativo, como é o caso da plataforma Wikidata.

O CEPID Neuromat orienta-se na difusão científica também na aproximação entre ciência de ponta e projetos colaborativos como os Wikimedia. A plataforma Wikidata, inaugurada em 2012, integra as iniciativas para o conhecimento livre e colaborativo da Fundação Wikimedia, atuando como um banco de dados estruturados que servem de suporte tanto para projetos Wikimedia quanto para aplicações externas, sob domínio público pelas diretrizes da licença Creative Commons. Segundo Vrandečić e Krötzsch (2014), Wikidata é um dos projetos Wikimedia de maior atividade, com uma abordagem de aquisição de dados via *crowd-source*, ou seja, viabilizada por esforços de colaboradores de todo o mundo que editam os dados de forma aberta, realizando a curadoria do conteúdo de acordo com o código de conduta estabelecido pela comunidade Wikimedia.

A bolsa aqui proposta insere-se num grupo em construção sobre formas experimentais de difusão científica, aliando novas linguagens e tecnologias, do qual o CEPID NeuroMat, com apoio da FAPESP, tem papel de liderança. O grupo conta com três bolsistas Mídia Ciência, apoiados pela FAPESP, e com quatro bolsistas de graduação, com apoio PUB/USP.

2) Justificativa de escolha do projeto, conjunto de projetos ou programas, em termos de relevância científica, tecnológica, cultural, econômica ou social

Com a proposta aqui apresentada, o CEPID NeuroMat pretende dar continuidade aos projetos em andamento de difusão por novas mídias, em especial as colaborativas vinculadas às iniciativas Wikimedia. Em trabalhos anteriores, membros do CEPID NeuroMat demonstraram o papel de tecnologias colaborativas como a Wikipédia especificamente no potencial para a difusão do conhecimento científico (PAIXÃO et al., 2016; PESCHANSKI et al., 2016), apontando dimensões positivas tanto para cientistas e a comunidade científica quanto para “wikimedistas”, membros da comunidade de editores dos projetos Wikimedia, e o público geral, sendo que uma ampla bibliografia contempla a perspectiva educacional e pedagógica da plataforma.

Seguindo a investigação em torno dos Wikimedia como plataformas para o conhecimento livre e colaborativo, cabe investigar os potenciais do projeto Wikidata, banco de dados aberto. O funcionamento de Wikidata baseia-se na pluralidade de dados, uma vez que a plataforma permite a coexistência de dados conflitantes; nos dados secundários, ao armazenar tanto os fatos quanto as suas referências, aumentando a diversidade e a verificabilidade da informação; e dados multilíngue, bastando uma única entrada de dados em um idioma para que fique disponível em todos os outros idiomas disponíveis na Wikimedia. A partir de tais recursos, a importância da Wikidata se dá não somente por sua capacidade de armazenar uma quantidade massiva de dados, mas pelo potencial de servir como base para novos serviços e aplicações que, por conta do aspecto colaborativo, aberto e global, podem ser desenvolvidos por qualquer pessoa com acesso à Internet:

Wikipédia é um dos websites mais importantes hoje em dia, legado que Wikidata deve dar continuidade. Em apenas dois anos, Wikidata já é uma plataforma importante

para integrar informação de muitas fontes. Além disso, também agrega grandes quantidades de metadados incidentais sobre sua própria evolução e contribuição para a Wikipédia. O Wikidata, portanto, tem o potencial de ser um importante recurso tanto para pesquisa quanto desenvolvimento de aplicações novas e melhores. Wikidata, a base de conhecimento livre que qualquer pessoa pode editar, pode, portanto, nos aproximar de um mundo que compartilha livremente a soma de todos os conhecimento (VRANDEČIĆ, KRÖTZSCH, 2014, p. 85).¹

Esse banco de dados estruturados (*linked data*) atua como suporte tanto para os outros projetos Wikimedia quanto para iniciativas externas via APIs abertas, de forma a possibilitar o envolvimento de usuários em novas práticas jornalísticas (OLIVEIRA, 2013). Levando em consideração que o jornalismo profissional passa por mudanças em seu modelo de negócios e em suas atividades por conta dos novo ecossistema informacional derivado da expansão das TICs e da evolução do Big Data (WT LIMA JÚNIOR, 2012), a plataforma Wikidata representa a oportunidade de novas práticas midiáticas para jornalistas - que devem produzir conteúdos de relevância social a partir da extração de dados não triviais, em sintonia com a demanda informativa da sociedade conectada, segundo Maria Eugênia Ribeiro (2013, p. 7).

O bom uso de API - especialmente no caso do *The Guardian*, que contabiliza mais de 3 mil usuários cadastrados em sua plataforma - mostrou o poder de aplicativos que podem ser criados a partir da combinação de diversos conteúdos. Assim, o cruzamento entre comunicação social e a ciência da computação, que reflete em posturas abertas (tomando como exemplo a disponibilização de acervos digitais seguindo padrões de *linked data*), oferece uma oportunidade de estudo para a compreensão de processos e rotinas, bem como sua influência em mudanças na cultura jornalística (OLIVEIRA, 2013, p. 8)

Nesse contexto, a experimentação com narrativas estruturadas representa novos caminhos para a produção e para o consumo de conteúdo jornalístico. O protótipo do projeto *Structured Stories*, fundado por David Caswell e construído entre 2013 e 2014 com uma base de dados acessível por APIs, foi desenvolvido a partir do conceito de narrativa estruturada, de forma a poder ser manipulado tanto por jornalistas quanto por cidadãos comuns:

¹ Tradução livre do original: Wikipedia is by all accounts one of the most important websites today, a legacy Wikidata must live up to. In only two years, Wikidata is already an important platform for integrating information from many sources. In addition, it also aggregates large amounts of incidental metadata about its own evolution and contribution to Wikipedia. Wikidata thus has the potential to be a major resource for both research and development of new and improved applications. Wikidata, the free knowledgebase anyone can edit, may thus bring us all one step closer to a world that freely shares in the sum of all knowledge (VRANDEČIĆ, KRÖTZSCH, 2014, p. 85)

Ao dissociar a estrutura da história a partir da forma com que ela é contada, Structred Stories pode permitir a comunicação de mesmas notícias em qualquer língua, em qualquer dispositivo, usando qualquer forma de mídia e em qualquer grau de detalhe. [...] E ao fazer notícias “computáveis”, Structred Stories pode liberar novos conhecimentos de notícias gerais usando queries, raciocínio e machine learning. Histórias estruturadas não substituem artigos de texto, mas os complementa significativamente ao permitir novas funcionalidades que não são possíveis com notícias baseadas em linguagem (*Why are Structred Stories useful?*, Structred Stories FAQ²).³

O caso de Structred Stories sustenta as potencialidades para a prática jornalística da manipulação de bancos de dados via API, servindo, também, como exemplo de uma prática de um Jornalismo Computacional, entendido como atividades e serviços construídos em torno de ferramentas computacionais com fins jornalísticos:

[...] Jornalismo computacional como uma vertente de jornalismo tecnologicamente orientado, centrado na aplicação de computação e no pensamento computacional às práticas de coleta de informação, elaboração de sentidos e apresentação de informação, ao invés de um uso jornalístico de dados ou de métodos de ciências sociais em geral (CODDINGTON, 2015, p. 135)⁴.

Com isso, nota-se a tendência de um jornalismo baseado no pensamento computacional e que se apropria de conhecimentos tradicionalmente restritos às Ciências da Computação, que ao unir-se a iniciativas de bancos de dados estruturados, como é o caso de WikiData, podem gerar novos serviços e formas inovadoras de consumo de conteúdo nas mídias digitais. Portanto, vale também investigar, teoricamente, as implicações do Jornalismo Computacional para a prática

² Structred Stories FAQ, disponível em: <http://www.structuredstories.com/#/intro>. Acesso em: 13 mar. 2018.

³ Tradução livre do original: By decoupling a story's structure from the way in which that story is told, Structred Stories can enable communication of the same news in any language, on any device, using any form of media or at any degree of detail. [...] And by making news 'computable' Structred Stories can unlock new knowledge from general news using queries, reasoning and machine learning. Structred stories do not replace text articles, but they significantly complement them by enabling new functionality that just is not possible with news based on language (*Why are Structred Stories useful?*, Structred Stories FAQ).

⁴ Tradução livre do original: [...] computational journalism as a strand of technologically oriented journalism centered on the application of computing and computational thinking to the practices of information gathering, sense-making, and information presentation, rather than the journalistic use of data or social science methods more generally (CODDINGTON, 2015, p. 135).

jornalística em transformação, contribuindo para a bibliografia sobre o assunto, ainda relativamente limitada no país.

3) Identificação dos pesquisadores envolvidos nas pesquisas científicas ou tecnológicas que serão objeto da pesquisa jornalística

Os pesquisadores envolvidos nas pesquisas científicas e tecnológicas, que serão objeto da pesquisa jornalística aqui proposta incluem o diretor do projeto, Antonio Galves (NUMEC - IME/USP), os pesquisadores principais Antonio Carlos Roque da Silva Filho (FFCLRP/USP), Claudia D. Vargas (IBCCF/ UFRJ), Ernst Hamburger (IF/USP), Jorge Stolfi (IC/UNICAMP), Pablo Augusto Ferrari (NUMEC - IME/USP) e Yoshiharu Kohayakawa (NUMEC - IME/USP), os investigadores associados seniores Angela Sirigu (CNRS), Marzio Cassandro (Sapienza Università di Roma), Pierre Collet (CNRS - Ecole Polytechnique - NUMEC), Ricardo Fraiman (Universidad de la Republica, Uruguay), Roberto Fernandez (Universiteit Utrecht - NUMEC) e Sidarta T.G. Ribeiro (UFRN), os investigadores associados de São Paulo Alexsandro Giacomo Grimbart Gallo (UFSCar), Anatoli Iambartsev (IME/USP), André Frazão Helene (USP), Claudio Possani (IME/USP), Fabio Kon (IME/USP), Florencia Graciela Leonardi (NUMEC - IME/USP), Kelly Rosa Braghetto (IME-USP), Maria Elisa Pimentel Piemonte (USP), Miguel Abadi (IME/USP), Nancy Lopes Garcia (UNICAMP) e Osame Kinouchi Filho (FFCLRP/USP) e os investigadores associados fora de São Paulo Aldana González Montoro (Universidad Nacional de Córdoba, Argentina), Amanda Sávio Nascimento (DECOM/UFOP), Carlos Hoppen (UFRGS), Christophe Pouzat (Université Paris Descartes), Daniel Fraiman Borrazás (Universidad de San Andrés, Buenos Aires), Daniel Yasumasa Takahashi (Princeton University), Eva Loecherbach (Université de Cergy Pontoise), Errico Presutti (Istituto Nazionale di Fisica Nucleare/INFN), Guillermo Cecchi (IBM Thomas J. Watson Research Center), Jerome Paul Armand Laurent Baron (UFMG), Marcello Magnasco (Rockefeller University), Mauro Copelli (UFPE), Paul Balister (University of Memphis), Roberto Imbuzeiro Oliveira (IMPA), Sergio Neuenschwander (UFRN) e Valeria Della Maggiore (Universidad de Buenos Aires) (disponível em <http://neuromat.numec.prp.usp.br/pt-br/equipe>, acessado em 26/11/2016). O coordenador da área de difusão do CEPID Neuromat é o Prof. Fernando da Paixão (UNICAMP), que aliás atuará

como supervisor científico da pesquisa aqui proposta, com a participação do jornalista responsável da área de difusão e professor João Alexandre Peschanski (FCL).

4) Descrição e cronograma das atividades previstas para sua realização

O projeto de pesquisa de jornalismo científico propõe a realização de atividades práticas e teóricas, em especial na confluência entre jornalismo científico e museologia, na documentação de acervos e facilitação do acesso a bens culturais e o registro do projeto de exposição do CEPID NeuroMat, em andamento. Na consideração teórica, pretende-se entender, explorar o papel das novas mídias na museologia. Os objetivos práticos das atividades, que serão realizados sob a coordenação do supervisor científico e do jornalista responsável, a partir de reuniões semanais, são:

- I. Desenvolver pelo menos 12 produtos associados a narrativas estruturadas no contexto da iniciativa de difusão científica wiki do CEPID NeuroMat;
- II. Produzir um material para treinamento no uso do Wikidata, em sua integração a outros projetos Wikimedia;
- III. Desenvolver uma página wiki para a centralização de recursos Wikidata, para a comunidade em português;
- IV. Publicar um artigo em um meio de comunicação de público amplo sobre a importância da plataforma Wiki ou novas mídias de modo mais geral para a difusão de conhecimento e digitalização de acervos de galerias, arquivos e museus;
- V. Manter com pelo menos uma postagem por semana o blog “Traço de Ciência”, sobre difusão científica do NeuroMat, em língua portuguesa.

Já os objetivos das atividades teóricas, que também serão realizadas sob a coordenação do supervisor científico, são:

- I. Produção de um artigo de pesquisa sobre Jornalismo Computacional e narrativas estruturadas, em vista de uma publicação acadêmica a ser submetida no contexto da bolsa;
- II. Elaboração de um artigo de pesquisa sobre narrativas estruturadas em WikiData em vista de uma publicação acadêmica a ser submetida no contexto da bolsa;
- III. Composição de um relato de experiência da atividade da bolsa, a ser eventualmente elaborado e submetido com a renovação da bolsa por mais seis meses;
- IV. Acompanhamento de formação em difusão científica em especial com a realização de Curso de Introdução ao Jornalismo Científico, disponibilizado pela WSFJ.

- V. Participação em um programa de estudo de programação em Lua, linguagem do Wikidata, com no mínimo 14 horas.

Cronograma de atividades:

ATIVIDADES	Mês 1	Mês 2	Mês 3	Mês 4	Mês 5	Mês 6
Leituras sobre Jornalismo Computacional e narrativas estruturadas no contexto de divulgação científica	X	X	X	X		
Relato de experiências no blog	X	X	X	X	X	X
Curso de Jornalismo Científico da WSFJ	X	X	X	X		
Curso de programação em Lua	X	X				
Produtos de narrativas estruturadas associadas ao NeuroMat		X	X	X	X	X
Desenvolvimento de material de treinamento para Wikidata			X	X	X	
Produção de artigo acadêmico			X	X	X	X
Seminário de pesquisa						X

6) Bibliografia

CASWELL, David A. Structured Narratives as a Framework for Journalism: A Work in Progress. In: 6th Workshop on Computational Models of Narrative, 2015. Disponível em: <<http://drops.dagstuhl.de/opus/volltexte/2015/5279/>>. Acesso em: 3 mar. 2018.

CASWELL, David A. et al. Editorial Aspects of Reporting into Structured Narratives. In: Conference: Conference Computation + Journalism, Columbia University, New York, 2015. Disponível em:

<https://www.researchgate.net/publication/312940709_Editorial_Aspects_of_Reporting_into_Structured_Narratives>. Acesso em: 3 mar. 2018.

CODDINGTON, Mark. 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://www.tandfonline.com/doi/pdf/10.1080/21670811.2014.976400>>. Acesso em: 5 mar. 2018.

LIMA JÚNIOR, Walter Teixeira. Jornalismo computacional em função da "Era do Big Data". *Líbero*, São Paulo, v. 14, n. 28, p.45-52, dez. 2011.

MENDONÇA, Vinícius de Sousa. **Notícias geradas por software: O jornalismo sem repórter**. 2016. 39 f. TCC (Graduação) - Curso de Jornalismo, Centro Universitário de Brasília, Brasília, 2016.

OLIVEIRA, André Rosa de. Acervo digital e linked data: da memória jornalística às APIs abertas. *Brazilian Journal of Technology, Communication, and Cognitive Science*, n. 1, set. 2013.

PAIXÃO, Fernando da et al. Tecnologias colaborativas na difusão científica: Um relato dos usos e apropriações da Wikipédia no CEPID NeuroMat. In: Anais IX Simpósio Nacional ABCiber, São Paulo, 2016. Disponível em: <http://abciber2016.com/wp-content/uploads/2016/trabalhos/tecnologias_colaborativas_na_difusao_cientifica_um_relato_dos_usos_e_apropriacoes_da_wikipedia_no_cepid_neuromat_david_fernando_levon_alves.pdf>. Acesso em: 6 mar. 2018.

PESCHANSKI, João Alexandre et al. O uso da Wikipédia na difusão científica. In: Anais XXXIX Congresso Brasileiro de Ciências da Comunicação - 39, INTERCOM, São Paulo, 2016. Disponível em: <<http://portalintercom.org.br/anais/nacional2016/resumos/R11-1908-1.pdf>>. Acesso em: 6 mar. 2018.

RIBEIRO, Maria Eugênia. O jornalista na sociedade digital: possibilidades e desafios. In: Anais XXXVI Congresso Brasileiro de Ciências da Comunicação - 36, INTERCOM, Manaus, 2013. Disponível em: <http://www.intercom.org.br/papers/nacionais/2013/resumos/R8-1497-1.pdf> Acesso em: 4 mar. 2018.

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://cacm.acm.org/magazines/2014/10/178785-wikidata/fulltext>>. Acesso em: 11 mar. 2018.