

Vamos falar sobre deficiência? Uma análise dos *Tweets* sobre este tema no Brasil

Fábio Manoel França Lobato¹, Marcelo da Silva¹, Krislen Coelho²,
Simone da Costa Silva², Fernando Pontes²

¹Instituto de Engenharia e Geociências - Universidade Federal do Oeste Pará
Santarém, Pará, Brasil

²Núcleo de Teoria e Pesquisa do Comportamento - Universidade Federal do Pará
Belém, Pará, Brasil

fabio.lobato@ufopa.edu.br

{marcelo.t.pain, krisllenmayra2010, symon.ufpa, farpl304}@gmail.com

Resumo. *As deficiências estão mais relacionadas a um contexto social do que com condições médicas. No entanto, a falta de atenção ao assunto e de suporte social impactam negativamente na vida das pessoas com deficiência e seus familiares. Em um estudo exploratório percebeu-se uma tendência à depreciação dos deficientes em detrimento da construção de uma rede de apoio. À luz de tais fatos, este trabalho analisou postagens relacionados às deficiências mentais, físicas e intelectuais, a fim de identificar os principais temas discutidos e as circunstâncias de utilização.*

Abstract. *Disabilities are more related to a social context than to medical conditions. However, the lack of attention to the subject and social support negatively impacts the lives of people with disabilities and their relatives as well. In an exploratory study, it was perceived a trend towards the depreciation regarding this theme, instead of the construction of a support network. In light of these facts, this paper analyzed postings related to mental, physical and intellectual disabilities to identify the main topics discussed and the circumstances of use.*

1. Introdução

Estima-se que mais de 10% da população mundial sofre de algum tipo de deficiência física, mental ou intelectual, sendo que 80% dessas pessoas vivem em países em desenvolvimento [Setareh Forouzan et al. 2013]. De acordo com uma pesquisa conduzida pelo *Pew Research Center*, nestes mesmos países, observa-se que a maior parte dos usuários adultos de internet são engajados nas Redes Sociais Online como Facebook e Twitter [Poushter 2016]. Neste contexto, diversos estudos mostram que pessoas com deficiência e seus familiares estão cada vez mais presentes nas mídias digitais, compartilhando experiências e procurando aconselhamento de outras pessoas em condições semelhantes [Naslund et al. 2016].

As conexões geradas por meio da interação entre os pares online podem impactar positivamente no bem-estar dessas pessoas, aflorando o sentimento de pertencimento a grupos sociais coesos e afins, possibilitando também a troca de conhecimentos e estratégias para lidar com desafios do dia a dia [Naslund et al. 2016]. Apesar dos benefícios

dessas plataformas, fenômenos relacionados à segregação e depreciação envolvendo o tema deficiência são observados com frequência. Diversos trabalhos dedicaram-se ao estudo do discurso de ódio nesta temática. [Mckay et al. 2015] analisaram expressões categorizadas como símbolos depreciativos e que podem ter implicações para pessoas e suas experiências de vida, por exemplo, como expressões comuns na linguagem moderna, podemos citar: "retardado", que evoluiu de um diagnóstico médico à insulto; e também expressões ligadas ao suicídio, usada para estigmatizar e banalizar pessoas que experimentaram ideações ou experiências suicidas.

Não obstante, uma análise exploratória mostrou que este fenômeno é inconteste em diversos termos ligados à deficiência. Neste sentido percebeu-se uma lacuna na literatura quanto à análise do discurso sobre deficiências nas redes sociais, sendo este o foco do presente trabalho. Dessa forma, considerou-se salutar distinguir os padrões de postagens contendo discurso de ódio dos que são de caráter informativo, sobretudo identificando conteúdo (expressões) e discurso (construção da postagem) que atinja o público-alvo (pessoas com deficiência). Esta tarefa pode ser vista como a realização de detecção de comunidades e segmentação de mercado [Lobato et al. 2017]. Uma tarefa que antecede a segmentação é a extração e seleção de atributos [Silva et al. 2017], sendo que uma etapa opcional é a anotação de dados, o que permitiria a classificação direta das instâncias anotadas. Neste âmbito, [Magalhães et al. 2017] demonstra a identificação de comentários ofensivos em um portal de notícias usando aprendizado supervisionado, para tal, a etapa anterior foi a de anotação do conjunto de dados de acordo com categorias pré-definidas. Ainda, [Cirqueira and Vinícius 2017] propõe uma ferramenta baseada em gamificação para anotação de dados a serem utilizados em análise de sentimentos. E em [Cirqueira et al. 2017] os autores avaliam o desempenho de métodos de análise de sentimento para português brasileiro.

O problema de pesquisa surgiu da necessidade de se estabelecer uma interlocução por intermédio de mídias digitais entre um consórcio de pesquisa e grupos de suporte parental relacionados à pessoas com deficiência. O consórcio é formado por pesquisadores de diferentes áreas, contemplando enfermagem, terapia ocupacional, psicologia, sistemas de informação, ciência da computação e comunicação. Já os grupos de suporte são diversos e abrangem membros/familiares de deficientes, dos mais variados espectros. No entanto, alguns problemas foram observados no planejamento de conteúdo direcionados a um público tão diverso e na construção de estratégias de marketing efetivas para recrutamento de deficientes e familiares para participação em projetos de pesquisa.

Visando contornar esta lacuna, o presente estudo analisou postagens relacionados às deficiências mentais, físicas e de aprendizado, a fim de identificar os principais temas discutidos e as circunstâncias de utilização. Os resultados obtidos dão subsídio para a construção de estratégias de comunicação mais efetivas para se estabelecer uma interlocução com os deficientes e seus familiares.

O restante deste artigo encontra-se organizado como segue. A metodologia e os resultados obtidos são descritos nas Seções 2 e 3, respectivamente. Os *insights* obtidos, as dificuldades encontradas e sugestões de trabalhos futuros são abordadas na Seção 4.

2. Metodologia

Inicialmente um estudo exploratório visando identificar padrões de comunicação na temática deficiência para melhorar a interlocução consórcio-grupo foi conduzido. Os achados auxiliaram na construção do desenho da pesquisa, o qual pode ser resumido em dois passos, delimitação do escopo e análise dos dados.

A delimitação do escopo abarca a definição e/ou validação dos termos relacionados às deficiências mais trabalhadas pelo consórcio de pesquisa. Os termos foram definidos em reunião entre um analista de rede social, um profissional da comunicação e um psicólogo. A validação dos termos deu-se por meio de consultas usando a interface de busca do Twitter. Os seguintes termos foram adotados Deficiência, Deficiência Mental, Cego(s), Cegueira, Surdo(s), Surdez, Autismo, Autista(s), Deficiência Auditiva, Deficiência Física, Deficiência Intelectual, Amputação, Síndrome de Down, Deficiência Visual, Pessoa com Deficiência, Paralisia Cerebral, Lesão Medular, Espinha Bífida, Mielomeningocele e Baixa Visão. Com os termos validados deu-se início ao segundo passo da pesquisa, que consistiu na aquisição, análise e anotação dos dados obtidos por meio dos termos de pesquisa. A aquisição de dados nesta fase foi feita utilizando a *Search API*.

Os dados obtidos foram analisados visando identificar os tópicos mais comentados e também classificando-os de acordo com o seu discurso. Para a identificação dos tópicos mais frequentes utilizou-se a modelagem de tópicos, a qual se mostrou potencialmente similar à análise baseada em teoria fundamentada [Baumer et al. 2017]. O método de modelagem de tópico avaliado foi o *Latent Semantic Allocation* (LSA). A classificação do discurso deu-se de forma manual, com três avaliadores para cada *Tweets* selecionado. Considerava-se de uma classe quando havia o consenso de pelo menos dois avaliadores, estratégia similar à adotada em [Magalhães et al. 2017]. Para tal, cinco categorias foram utilizadas para a classificação do discurso, a citar: “Informativo”, “Ofensivo e Pejorativo”, “Indignação e Denúncia”, “Relato de Experiências”, e “Outros”.

As classes foram propostas, discutidas e validadas consensualmente pelos autores. Na primeira classe, a “Informativo”, enquadraram-se os *Tweets* que possuíam como característica precípua a divulgação de informações sobre o tema. Os *Tweets* que utilizavam um dos termos sob escrutínio para denegrir, praguejar ou com conotação pejorativa foram incluídos na categoria “Ofensivo e Pejorativo”. A classe “Indignação e Denúncia” foi incorporada para definir os *Tweets* cujo conteúdo envolvia a denúncia de crimes relacionados ao tema deficiência ou reportando indignação frente à um evento/caso específico. A classe “Relato de Experiências” é auto-explicativa. Por fim, a categoria “Outras” representa *Tweets* que não se enquadram em nenhuma das classes propostas.

3. Resultados e Discussões

O intervalo de coleta foi compreendido de 6 à 13 de março, este período foi escolhido pois não incluía nenhuma data representativa qualquer Deficiência (e.g. Dia mundial do deficiente físico). Foram obtidos aproximadamente 200.000 *Tweets* sobre o tema. O número de amostras suficientes para o intervalo de confiança de 95% e taxa de erro de 5% para o total de dados coletados é de 377 postagens. Com o acréscimo dos 35% utilizados para aumento de confiabilidade, foram anotadas 509 amostras. O processo de anotação resultou em um conjunto de dados composto por 459 dos 509 *Tweets* totais, sendo que 5 haviam sido duplicados e em 36 não houve concordância entre os avaliadores.

A Figura 1 apresenta a distribuição das classes considerando os 459 *Tweets* com anotação consolidados.

Distribuição das classes

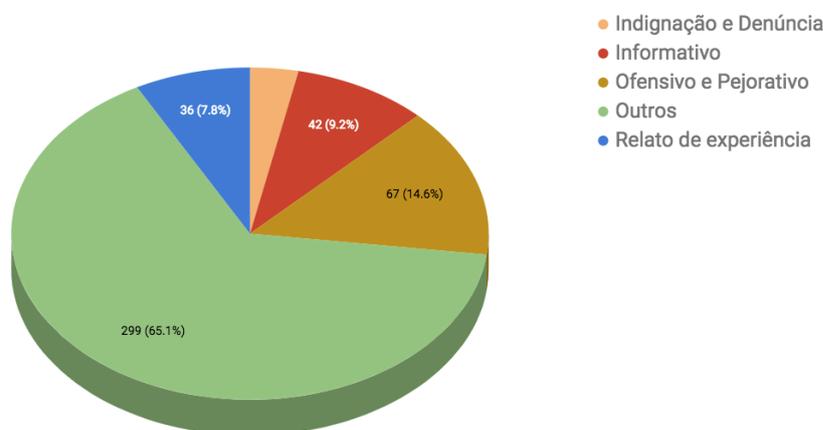


Figura 1. Distribuição das categorias para os *Tweets* classificados.

A maior parte dos *Tweets* foi classificada como “Outros” (299 *Tweets*), seguida de Ofensivo e Pejorativo (67), Informativo (42), Relato de experiência (36) e Indignação e Denúncia (15). Este fenômeno ocorreu justamente pela alta presença de ruído ligado à determinados termos utilizados como expressão popular, tais como: “Em terra de **cego**, quem tem olho é Rei”, “O pior **cego** é aquele que não quer ver”, “O amor é **cego**”, e “Mais perdido do que **cego** em meio de tiroteio”. Neste contexto, considerou-se que tais expressões não possuem caráter Ofensivo/Pejorativo aos deficientes visuais.

A análise da distribuição das classes, permite-nos concluir que, pelo menos no que tange ao Twitter, os usuários pouco postam denunciando ou demonstrando indignação. Se os exemplos ruidosos do *dataset* fossem excluídos (instâncias anotadas como “Outros”), menos de 10% dos *Tweets* representavam denúncia e indignação. Em contraste, cerca de 40% foram considerados Ofensivo e Pejorativo, e os 50% restantes quase que igualmente divididos entre Relato de Experiência e Informativo. Considerando a maior representatividade e coesão dos *Tweets* classificados como Outros, Ofensivo e Pejorativo, Informativo e Relato de experiência, aplicou-se neles a modelagem de tópicos para identificar os temas mais recorrentes. A Tabela 1 apresenta os resultados da modelagem de tópicos por categoria.

Por meio da análise da Tabela 1 é possível confirmar os achados na fase de anotação manual, já que a classe “**Outros**” é prevalente. Já sobre as impressões dos avaliadores acerca da classe “**Ofensivo e Pejorativo**”, por serem menos frequentes, os avaliadores não foram assertivos sobre os temas mais frequentes, justificando a utilização da modelagem de tópicos. Os *Tweets* anotados como “**Relatos de Experiência**” também contém ruídos, como a exposição de situações onde a pessoa se descreve em uma situação de perda de visão, necessitando de uma consulta em um oftalmologista (Tópico 2). Um achado interessante diz respeito ao Youtube como canal de comunicação bastante utilizado no nicho em estudo (Tópico 3), sendo que diversos canais são mantidos por deficientes e seus familiares com o intuito de compartilharem experiências, seguindo o estilo

Tabela 1. Tópicos mais frequentes por categoria.

Categoria	Tópico	Termos		
Outros	1	cego	tiroteio	amor
	2	surdo	ficar	mudo
	3	tiroteio	perdida	perdido
	4	amor	tiroteio	cupido
Ofensivo	1	cego	filho	melhor
	2	síndrome	down	nome
	3	autista	surdo	acho
	4	surdo	gritando	bocejar
Relato	1	síndrome	down	peessoas
	2	cego	ficando	oftalmo
	3	youtube	vídeo	gostei
	4	deficiência	aluno	auditiva
Informativo	1	deficiência	vagas	oferece
	2	down	síndrome	dia
	3	autismo	crianças	brasil
	4	pesquisa	sintomas	maconha

semelhante ao de um diário. No que tange aos *Tweets* “**Informativos**”, os temas mais discutidos sobre deficiência foram i) inserção no mercado de trabalho (Tópico 1); ii) divulgação de datas importantes (Tópicos 2 e 3); iii) e pesquisas relacionadas ao tema (Tópico 4). Tais achados permitem a construção de um *guideline* para se melhorar a eficácia da interlocução entre o consórcio de pesquisa e os deficientes e seus familiares.

4. Considerações Finais

É notório o crescimento de iniciativas que visam melhorar o bem-estar e qualidade de vida de pessoas com deficiência, neste ensejo, o suporte social por meio de plataformas computacionais tem ganhado destaque. Apesar dos esforços para uso consciencioso das redes sociais, é comum a presença de discurso de ódio ou de cunho pejorativo relacionado à deficiências, impactando negativamente na auto-estima de deficientes e seus familiares. Visando aprofundar as investigações nesta temática, este trabalho apresentou uma análise de *Tweets* sobre o tema em questão, identificando os principais tópicos discutidos e as circunstâncias de utilização.

Acerca das análises, a classificação do discurso deu-se de forma manual considerando-se cinco categorias, a citar: “Informativo”, “Ofensivo e Pejorativo”, “Indignação e Denúncia”, “Relato de Experiências”, e “Outros”. Para a identificação dos tópicos mais frequentes utilizou-se o algoritmo *Latent Semantic Allocation*. Por meio da análise dos resultados alcançados, destacam-se as seguintes contribuições científicas tangíveis: i) Obtenção de um *dataset* anotado considerando termos relacionados à deficiências; ii) A identificação de tópicos mais frequentes, considerando as categorias propostas; iii) A disponibilização de informações para abalizar a construção de *guidelines* para comunicação com o deficientes e familiares por intermédio de mídias sociais.

As principais dificuldades encontradas estão relacionadas à alta presença de ruído nos dados analisados. No que tange às ameaças a validade do estudo, a não utilização de uma medida estatística para avaliar a concordância entre os avaliadores e também a quantidade de dados anotados são as mais notórias. A consideração de tais fatos impeliu-

nos a estabelecer como trabalhos futuros a anotação de um *dataset* sem incluir os termos de busca ruidosos, e a adoção do coeficiente kappa de Cohen para analisar a proporção de concordância observada. Também vislumbrou-se o estudo da estrutura do *Tweet* quanto à utilização de *hashtags* e *links* a fim de saber quais as características das postagens com maior engajamento

Agradecimentos

Este trabalho foi parcialmente financiado pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). Agradecemos também aos voluntários, por seus esforços na anotação dos *datasets*, e aos revisores pelos comentários enriquecedores.

Referências

- Baumer, E. P. S., Mimno, D., Guha, S., Quan, E., and Gay, G. K. (2017). Comparing grounded theory and topic modeling: Extreme divergence or unlikely convergence? *Journal of the Association for Information Science and Technology*, 68(6):1397–1410.
- Cirqueira, D., Jacob, A., Lobato, F., de Santana, A. L., and Pinheiro, M. (2017). Performance Evaluation of Sentiment Analysis Methods for Brazilian Portuguese. In Abramowicz, W., Alt, R., and Franczyk, B., editors, *Business Information Systems Workshops: BIS 2016 International Workshops, Leipzig, Germany, July 6-8, 2016, Revised Papers*, pages 245–251. Springer International Publishing, Cham.
- Cirqueira, D. and Vinícius, L. (2017). Opinion Label : A Gamified Crowdsourcing System for Sentiment Analysis Annotation. In *XVI Workshop de Ferramentas e Aplicações*.
- Lobato, F., Pinheiro, M., Jacob, A., Reinhold, O., and Santana, Á. (2017). Social CRM: Biggest Challenges to Make it Work in the Real World. In Abramowicz, W., Alt, R., and Franczyk, B., editors, *Business Information Systems Workshops: BIS 2016 International Workshops, Leipzig, Germany, July 6-8, 2016, Revised Papers*, volume 263, pages 221–232. Springer International Publishing, Cham.
- Magalhães, G. G. M. S., Lima, F., Santos, E. F., Junior, P., and Rosa, L. (2017). Seleção de Técnicas de Mineração de Dados para Segmentação de Mercado. In *6th Brazilian Workshop on Social Network Analysis and Mining*.
- Mckay, K., Wark, S., and Mapedzahama, V. (2015). Sticks and stones : How words and language impact upon social inclusion. *Journal of Social Inclusion*, 6:146–162.
- Naslund, J. A., Aschbrenner, K. A., Marsch, L. A., and Bartels, S. J. (2016). The future of mental health care: peer-to-peer support and social media. *Epidemiology and Psychiatric Sciences*, 25(2):113–122.
- Poushter, J. (2016). Social networking very popular among adult internet users in emerging and developing nations. Technical report, Pew Research Center.
- Setareh Forouzan, A., Mahmoodi, A., Jorjoran Shushtari, Z., Salimi, Y., Sajjadi, H., and Mahmoodi, Z. (2013). Perceived Social Support Among People With Physical Disability. *Iranian Red Crescent Medical Journal*, 15(8):663–667.
- Silva, W., Santana, Á., Lobato, F., and Pinheiro, M. (2017). A Methodology for Community Detection in Twitter. In *Proceedings of the International Conference on Web Intelligence*, pages 1006–1009.