

Pôster

OS SPAMS E A POLIFONIA INFORMACIONAL: PRÁTICAS DE INDEXAÇÃO SOCIAL E O COMPORTAMENTO LINGUAGEIRO DOS SPAMMERS EM SITES DE COMPARTILHAMENTO DE VÍDEOS

Priscila Oliveira da Mata – UFMG

Maria Aparecida Moura – UFMG

Resumo

Os sistemas que utilizam a linguagem natural como forma de gerenciamento da informação tornaram-se um alvo para os *spammers*, que utilizam-se deste novo cenário interativo e relativamente aberto para disseminarem *spams*, acoplados aos conteúdos de maior popularidade nos sites de compartilhamento de vídeos. Se por um lado comprometem a qualidade na recuperação da informação nestes sistemas, por outro, engajam práticas de indexação e comportamento linguageiro que efetivam a recuperação de *spams*. Neste sentido, tendo em vista este contexto informacional digital, a presente pesquisa, que se encontra em andamento, tem o propósito de identificar, sistematizar e analisar, pelo viés da semiótica peirceana o comportamento dos *spammers* em sites de compartilhamento de vídeos que utilizam a linguagem natural como forma de organização e recuperação da informação. Para tanto, tendo em vista a complexidade do objeto de estudo, aliada à teoria semiótica de Peirce, a triangulação de métodos apresenta-se como percurso metodológico mais indicado para o desenvolvimento da pesquisa. Desta forma, serão utilizadas algumas técnicas de pesquisa como a Netnografia, a Análise de Redes Sociais – ARS e a observação participante. Acredita-se que a compreensão do comportamento dos *spammers* em sites de compartilhamento de vídeos, como *You Tube*, *Vimeo* e *Myspace*, poderá auxiliar na criação de mecanismos eficazes para compreender as práticas de indexação social na *web*, assim como prover dispositivos dialógicos que contribuam para o mapeamento dessas práticas e reduzam os seus efeitos em processos de recuperação da informação na rede.

Palavras-chave: Linguagem natural. *Spammers*. Recuperação da informação. *Web* colaborativa. Semiótica Peirciana.

Abstract

The systems using natural language as a form of information management have become a target for spammers, who use up this new interactive scenario and relatively open to disseminate spam, coupled with higher contents of popularity on video sharing websites. On one hand committed to quality in information retrieval in these systems, on the other, engage indexing practices and *linguageiro* behavior that actualize retrieval of spam. Therefore, having in view this new digital scenery, the present research being developed has the purpose to analyze, understand and systematize, by Peirce's semiotics' approach, the behavior of spammers on video-sharing websites that make use of the natural language as a form of organization and retrieving of the information. For doing such, taking into account the complexity of the object of study, aligned to Peirce's semiotics theory, the triangulation of methods presents itself as a methodological pathway more indicated for the research development. In this sense, certain researching techniques such as Netnography, Analysis of Social Networks – ASN, the participant observation and the application of questionnaires will be utilized. It is believed that the understanding of spammers' behavior on web pages where

videos are uploaded, like You Tube, Vimeo and Myspace, may be useful in the creation of efficient mechanisms to understand the practices of social indexing the web, as well as providing dialogical devices that contribute to the mapping of these practices and reduce their effects on processes of information retrieval in the network.

Keywords: Natural Language. Spammers. Information Retrieval. Cooperative Web. Peircean Semiotic.

1 INTRODUÇÃO

Com a implementação dos processos colaborativos na *web*, em virtude do uso cada vez mais constante das tecnologias informacionais, o usuário passou a assumir o papel de *produser*, termo cunhado por Bruns (2008) que define este novo usuário da *web* que, ao mesmo tempo assume as tarefas de produção, publicação, seleção e mediação. Neste cenário de produção, compartilhamento e disseminação de informação digital, verifica-se fundamentalmente o uso da linguagem natural nos processos de descrição e recuperação dos conteúdos. Neste sentido, a etiquetagem social é usada como uma alternativa para o gerenciamento da informação realizada por distintos sujeitos informacionais. Trata-se de uma indexação livre em linguagem natural no qual não há o controle de vocabulários ostensivo e nem políticas de indexação pactuadas.

Lancaster (2004) já chamava atenção para a necessidade de uma indexação mais democrática, na qual os usuários acrescentariam termos da sua própria escolha para facilitar a organização e a recuperação da informação de certos registros como imagens ou obras de ficção.

Em *sites* de compartilhamento de vídeos, o uso da linguagem natural é uma realidade na qual se acrescenta a complexidade do objeto e a dificuldade de representar o seu conteúdo. No entanto, a relativa ausência de controle de vocabulário permitiu aos *spammers* engajarem práticas de indexação e comportamentos languageiros que oportunizam a disseminação eficiente de *spams* causando “ruídos” que comprometem a qualidade da recuperação da informação.

Tendo em vista a grande complexidade de se entender os processos de trocas informacionais, interação e criação existentes hoje nos ambientes digitais colaborativos, esta pesquisa, que se encontra em andamento, tem o propósito de identificar, sistematizar e analisar pelo viés da semiótica peirceana, o comportamento dos *spammers* em *sites* de compartilhamento de vídeos que utilizam a linguagem natural como forma de organização e recuperação da informação.

1.1 OBJETIVOS

Com base nos conceitos peirceanos de semiose, objeto, signo e interpretante a pesquisa tem como objetivo geral analisar, no contexto da *web* colaborativa, o comportamento linguageiro dos *spammers* em linguagem natural, assim como suas práticas de indexação social em *sites* de compartilhamento de vídeos. Os objetivos específicos são: analisar as características da linguagem natural utilizada para a representação da informação em *sites* de compartilhamento de vídeos; compreender os processos de significação empreendidos pelos usuários que utilizam a linguagem natural para representar o conteúdo dos vídeos hospedados nos *sites* analisados; compreender as interações estabelecidas pelos *spammers* em redes de compartilhamento de vídeo; identificar, analisar e sistematizar a constituição das práticas de indexação e as motivações que levam os *spammers* a disseminarem deliberadamente conteúdo informacional polifônico em redes de compartilhamento de informação multimídia; investigar sob uma perspectiva semiótica, as repercussões da prática *spammer* nos processos de representação e recuperação da informação em sistemas de compartilhamento de vídeos.

2 REFERENCIAL TEÓRICO

A fundamentação teórica se baseia nos seguintes temas e autores, embora não se limite aos mesmos: (i) **Web colaborativa**, tendo em vista a revolução digital ocasionada pelo uso cada vez maior das tecnologias informacionais que trouxe mudanças consideráveis no processo de circulação da informação na *web*. (BAEZA-YATES; RIBEIRO NETO, 1999; GILL *et al.*, 2007; ZILLER; MOURA, 2011) (ii) **Semiótica**, em virtude do caráter signico dos processos de trocas informacionais na *web* atual. (PEIRCE, C. S, 1995; PINTO, 1995; SANTAELLA, 1985) (iii) **Etiquetagem social e linguagem natural**, uma vez que a etiquetagem surge como um novo conceito e como uma nova alternativa para a organização do conhecimento na *web*. (MATHES, 2004; MOURA, 2009; VOSS, 2007) (iv) **Spammers**. (BENEVENUTO *et al.*, 2008; HEYMANN *et al.*, 2007; KRAUSE *et al.*, 2008)

3 METODOLOGIA

Tendo em vista a complexidade da análise dos processos de significação no universo empírico proposto, optou-se pela utilização da metodologia qualitativa no desenvolvimento da pesquisa, abordagem em que a subjetividade do investigador e dos sujeitos estudados está intrinsecamente ligada ao processo de investigação (MINAYO, 2005).

Como base teórica, será utilizada a teoria semiótica de Charles S. Peirce, com o intuito de auxiliar as análises no que tange principalmente ao comportamento dos usuários de *sites* de compartilhamento de vídeos. Desta forma, serão estudados os conceitos peirceanos de semiose, objeto, signo e interpretante, buscando relacioná-los com os conceitos de representação da informação e indexação social adotados na Ciência da Informação. Acredita-se que a teoria semiótica de Peirce mostra-se adequada para auxiliar o percurso metodológico da pesquisa, uma vez que, permite explorar as manifestações simbólicas que ocorrem nas interações entre os usuários, nos processos de significação em sistemas colaborativos.

Visando coerência interna da pesquisa em relação ao marco teórico adotado, será utilizado a Netnografia, que apoiará o monitoramento, a observação, a identificação, os registros de atividades e comportamentos dos usuários dos sites de compartilhamento de vídeos: *You Tube*, *Vimeo*, e *Myspace*. Nesses sites, serão caracterizados suas formas de representação das informações, bem como o comportamento dos usuários mantenedores de canais de vídeos. Buscando-se uma maior robustez nas análises, utilizaremos ainda a Análise de Redes Sociais (ARS) no intuito de compreendermos melhor as relações existentes entre os atores destes sistemas, a identificação dos vídeos populares e a polifonia informacional engajada pelos *spammers* por intermédio das práticas de indexação social e do comportamento linguageiro.

A utilização de diversos métodos de pesquisa tornou-se necessária, em virtude da complexidade em se estudar comportamentos de usuários de ambientes colaborativos digitais carregados de subjetividade e significação e que utilizam a linguagem natural como forma de representação da informação. Neste sentido, a triangulação de métodos foi fundamental uma vez que “leva em conta o comportamento dos atores no que tange à compreensão, inteligibilidade dos fenômenos sociais e o significado e intencionalidade que lhe atribuem os atores.”(MINAYO, 2005, p. 82). A triangulação é uma nova perspectiva no campo metodológico e ela surge da necessidade de se combinar abordagens qualitativas e quantitativas, tendo em vista as especificidades dos universos a serem estudados.

Finalmente, para completar o percurso metodológico da pesquisa, para a coleta de dados, serão utilizados a observação participante e o questionário.

4 RESULTADOS PRELIMINARES

A pesquisa ainda se encontra em fase inicial, na qual está sendo realizada a revisão de literatura e a identificação e caracterização dos vídeos populares no *You Tube* bem como seu desdobramento em *spams*.

Nesta etapa, o que se observa com as leituras é que as recentes discussões sobre as necessidades e desafios dos estudos voltados para o conteúdo multimídia no contexto da *web* colaborativa, apontam para o fato de que a compreensão do comportamento dos usuários é de grande relevância para identificar os *spammers* e entender como eles atuam.

A análise preliminar dos vídeos populares e dos “vídeos *Spams*” revelou um padrão de comportamento linguageiro e de indexação social na atividade de compartilhamento que enseja a recuperação simultânea de documentos pertinentes à busca e *spams* em um grau de relevância acentuado a polifonia informacional. Essa polifonia informacional pode ser compreendida, tanto como um novo conteúdo cultural que tende a fortalecer a importância do conteúdo original, quanto engajar oportunismos e ruídos que dificultam a recuperação da recuperação pertinente.

Tendo em vista a complexidade que permeia a questão da significação em interfaces digitais multimodais¹, somada à crescente utilização da linguagem natural para representar as informações nos *sites* de compartilhamento de vídeos e ainda, a necessidade compreender a ação dos *spammers*, a natureza da indexação social realizada e a proliferação de *spams* nestes *sites*, torna-se oportuno um estudo mais aprofundado que busque analisar o comportamento e as interações entre os usuários, visando encontrar formas para coibir esses modelos de disseminação de *spams*, tendo compreender os aspectos semióticos que tais práticas ensejam e melhorar a qualidade dos sistemas de informação que adotam a indexação social.

5 CONSIDERAÇÕES

Acredita-se que a compreensão destes comportamentos e práticas de indexação poderá auxiliar na criação de soluções integradas, contribuindo assim, para um sistema de indexação dinâmico, diminuindo os ruídos que dificultam a atenção em processos de recuperação da informação na rede, o que configura hoje, um dos grandes desafios para a área da Ciência da Informação.

¹ A multimodalidade pode ser entendida como sendo a linguagem além da linguagem. A comunicação multimodal é a utilização simultânea de vários meios de comunicação. Fonte: www.let.unb.br

REFERÊNCIAS

- BAEZA-YATES, Ricardo; RIBEIRO NETO, Berthier. **Modeling**. In: _____. Modern information retrieval. New York: Addison Wesley, 1999. Cap. 2. Disponível em: < ftp://mail.im.tku.edu.tw/seke/slide/baeza-yates/chap10_user_interfaces_and_visualization-modern_ir.pdf >. Acesso em: 20 maio 2013
- BENEVENUTO, F., *et al.* Detectando usuários maliciosos em interações via vídeos no YouTube. In: XIV Brazilian Symposium on Multimedia Systems and Web- WebMedia, Vila Velha, 2008. **Proceedings...** Vila Velha: WebMaedia, 2008. Disponível em: < <http://homepages.dcc.ufmg.br/~tiagorm/publications/papers/webmedia2008.pdf> > Acesso em 01 jun 2013.
- BRUNS, Axel. **Blogs, Wikipedia, second life, and beyond**: from production to produsage. NovaIorque: Peter LangPublishing, 2008.
- GILL, P. *et al.* YouTube traffic characterization: a view from the edge. In: VII ACM SIGCOMM conference on Internet measurement- ACM SIGCOMM, Sand Diego, 2007. **Proceedings...** San Diego: ACM SIGCOMM, 2007. Disponível em: < <http://www.hpl.hp.com/techreports/2007/HPL-2007-119.pdf> > Acesso em 01 mar 2013.
- HEYMANN, P.; KOUTRIKA, G.; GARCIA-MOLINA, H. Fighting spam on social Web sites: a survey of approaches and future challenges. **IEEE Internet Comput.**, vol. 11, no. 6, p. 36–45, Nov. 2007. Disponível em< <http://ilpubs.stanford.edu:8090/818/1/2007-34.pdf> > Acesso em 01 maio 2013.
- KRAUSE, B *et al.* The anti-social tagger: detecting spam in social bookmarking systems. In: IV International Workshop on Adversarial Information Retrieval on The Web- AIRWeb, 2008. **Proceedings...** ACM AIRWeb, 2008. Disponível em: < http://airweb.cse.lehigh.edu/2008/submissions/krause_2008_anti_social_tagger.pdf > Acesso em: 01 jan 2013.
- LANCASTER, F. W. **Indexação e resumos**: teoria e prática. 2. ed. Brasília: Briquet de Lemos, 2004.
- LIU, K; FANG, B.; ZHANG, Y. Detecting tag spam in social tagging systems with collaborative knowledge. In: VI International Conference on Fuzzy Systems and Knowledge Discovery- IEEE FSKD, 2009. **Proceedings...** FSKD , 2009 Disponível em:< <http://dl.acm.org/citation.cfm?id=1802229> > Acesso em 01 jan 2013.
- MATHES, Adam. Folksonomies - cooperative classification and communication through shared metadata. Computer Mediated Communication – LIS590CMC, Urbana: University of Illinois, 2004. Disponível em: < <http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html> >. Acesso em: 01 maio 2013.
- MINAYO, Maria Cecília de Souza; ASSIS, Simone Gonçalves de; SOUZA, Edinilsa Ramos de. (orgs.). **Avaliação por triangulação de métodos**: abordagem de programas sociais. Rio de Janeiro: Editora Fiocruz, 2005. 243 p.
- MOURA, M. A. Folksonomias, redes sociais e formação para o tagging literacy: desafios para a organização da informação em ambientes colaborativos virtuais. **Informação &**

Informação, Londrina, v. 14, n. especial, p. 25-45, 2009a. Disponível em: <<http://www.uel.br/revistas/uel/index.php/informacao/article/view/2196>>. Acesso em: 01 maio 2013.

PEIRCE, C. S. Collected papers. Org. John Deely. Versão eletrônica. Bloomington: Past Masters, 1995.

PINTO, Julio. **1,2,3 da semiótica**. Belo Horizonte: Editora UFMG, 1995.

SANTAELLA, Lucia. **O que é semiótica?** 3. ed. São Paulo: Ed. Brasiliense, 1985. (Primeiros Passos)

_____. **Semiótica aplicada**. São Paulo: Cengage Learning, 2002.

VOSS, Jakob. Tagging, folksonomy & Co – renaissance of manual indexing?. In: X International Symposium of Information Science. **Proceedings...** Disponível em: <<http://arxiv.org/abs/cs/0701072>>. Acesso em: 01 dez 2013.

ZILLER, Joana. Qualidade da informação e produsage: semiótica, informação e o usuário antropofágico. Orientadora: Maria Aparecida Moura. 2011. 311f. (Doutorado em Ciência da Informação) – Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de Minas Gerais, Belo Horizonte.