

Uma investigação no apoio da avaliação da usabilidade em Sistemas Sociais usando Processamento da Linguagem Natural

Marília S. Mendes¹, Elizabeth Furtado², Miguel Franklin¹

¹Universidade Federal do Ceará (UFC), Fortaleza, CE – Brasil

²Universidade de Fortaleza (Unifor), Fortaleza, CE - Brasil.

{mariliamendes, elizabethsfur, miguel.franklin}@gmail.com

Abstract. *In HCI (Human-Computer Interaction) there are several techniques that assess the usability of systems. However, such techniques do not consider the data collection from messages when users are interacting and expressing their feelings related to some difficulty in interaction. This paper proposes an investigation to support the evaluation of the usability of social systems through the analysis of texts posted by users during the interaction in the system. We apply a methodology to search for posts of users on Twitter, composed of standards that may be related to use of system. As a contribution we report some lessons learned in the evaluation of Social Systems using Natural Language.*

Resumo. *Em IHC existem várias técnicas que avaliam a usabilidade de sistemas. No entanto, tais técnicas não consideram a coleta de dados por mensagens no momento em que os usuários estão interagindo e expressando seus sentimentos relacionados à alguma dificuldade de interação. Este artigo apresenta uma investigação a fim de apoiar a avaliação da usabilidade de Sistemas Sociais. Foi aplicada uma metodologia de pesquisa por postagens de usuários no Twitter, composta de padrões relacionados ao uso do sistema. Como contribuição são apresentadas algumas lições aprendidas para avaliação de Sistemas Sociais usando Processamento da Linguagem Natural.*

1. Introdução

Nos últimos anos, os Sistemas Sociais (SS) têm crescido consideravelmente. Segundo Pereira e outros (2010), SS são sistemas que permitem às pessoas, em suas particularidades e diversidades, comunicarem-se (interagirem, colaborarem, compartilharem ideias e informações), mediando e facilitando qualquer forma de relacionamento social. YouTube, Wikipédia e redes sociais, como o Facebook e o Twitter são exemplos de SS. Com a utilização em massa desses sistemas, muitos trabalhos têm aplicado técnicas de avaliação da usabilidade para testar SS. Segundo Cybis (et al., 2007) e Preece (et al., 2005), as principais técnicas utilizadas para se coletar opiniões diretas dos usuários e suas percepções são questionários e entrevistas.

Os questionários têm como objetivo coletar rapidamente dados (principalmente quantitativos) de muitos usuários. No entanto, se as perguntas não forem bem formuladas ou forem tendenciosas, poderão induzir ou confundir o usuário. Além disso, ainda existe o risco do questionário ser muito longo e os usuários responderem sem ler cuidadosamente as perguntas. Entrevistas, tanto individuais como em grupo são de fácil

entendimento, qualitativas e de baixo custo. No entanto, existe o esforço de recrutamento de usuários. Também como nos questionários, as perguntas devem ser bem formuladas para se atingir o objetivo almejado e não confundir o usuário.

Além destes motivos, o uso das técnicas citadas não permite a espontaneidade do usuário. Em SS, os usuários usam o sistema continuamente e comentam seu uso de forma espontânea durante o mesmo. Segundo Zhao (et al., 2012), perguntar aos usuários o que eles pensam a respeito de um produto, como eles realizam o que querem; se o produto é esteticamente atraente; se eles enfrentaram problemas ao utilizá-lo – é uma maneira óbvia de obter feedback. No entanto, nem sempre a pergunta sobre o uso do sistema vem “na hora certa”. Imagine o seguinte cenário: de repente o usuário se depara com um problema no sistema e resolve comentar sobre ele a fim de tirar a dúvida com os demais usuários ou mesmo expressar sua insatisfação sobre o problema. Como o avaliador irá precisar em que momento o usuário fez isso?

Desta forma, este trabalho consistiu em investigar como avaliar a usabilidade desses tipos de sistemas por meio de seus textos escritos. Usabilidade é uma medida que assegura que os produtos sejam fáceis de usar, eficientes e agradáveis sob a perspectiva do usuário [Preece et al. 2005]. Segundo Preece (et al., 2005), a avaliação da usabilidade em sistemas envolve avaliar o efeito da interface junto ao usuário, verificando se ele está satisfeito ou não ao usar o sistema. Satisfação refere-se a percepções do usuário, sentimentos e opiniões, geralmente capturadas por meio de perguntas escritas ou orais [Jeffrey e Chisnell 2008].

Neste sentido, definimos uma metodologia para investigar as postagens dos usuários de um SS a fim de coletar informações para o desenvolvimento de uma ferramenta usando o Processamento da Linguagem Natural (PLN). O PLN consiste no desenvolvimento de modelos computacionais para a realização de tarefas que dependem de informações expressas em alguma língua natural, como tradução e interpretação de textos, busca de informações em documentos, etc. [Russell 1995]. O foco desta primeira análise foi o Twitter, que foi escolhido tanto por seu caráter interativo e social quanto por sua interação se constituir principalmente de textos escritos.

2. Metodologia de investigação

A metodologia de investigação teve os seguintes objetivos: (1) Propor padrões de busca para extração de postagens relacionadas ao uso do sistema e (2) Analisar as postagens que foram recuperadas, a fim de definir estratégias para apoiar na avaliação da usabilidade. A metodologia foi estruturada em 3 etapas, descritas a seguir:

2. 1. Pesquisa por postagens dos usuários

A etapa 1 teve como objetivo obter postagens dos usuários do Twitter relacionadas ao uso do sistema. Para tanto, decidiu-se aplicar a técnica chamada de Extração de Informações (EI) que tem como finalidade extrair informações específicas do texto em língua natural [Riloff e Jones 1999]. A informação extraída é determinada por um conjunto de padrões ou regras de extração específicas ao domínio de análise: padrões que podem ser escolhidos de maneira manual, por algum especialista, ou de forma automatizada [Gaizauskas e Wilks 1998]. Os padrões propostos para investigação foram definidos por especialistas em usabilidade, autores deste trabalho, levando em consideração palavras relacionadas ao uso do sistema, como:

- Substantivos genéricos relacionados ao domínio em avaliação de usabilidade: “usabilidade”, “interface”, “erro”, “problema” e “bug”;
- Interrogações, considerando dois motivos: (i) com o intuito de procurar perguntas sobre o uso do sistema, considerando que a maioria das questões perguntadas em SS terminam com interrogação [Paul e Hong 2011] e; (ii) a fim de obter postagens de cunho emocional, pois quando os usuários estão aborrecidos ou irritados, geralmente utilizam símbolos do teclado para enfatizar sua fúria ou frustração, como pontos de exclamação (!!!), letras maiúsculas e pontos de interrogação (???) [Preece et al. 2005];
- Adjetivos que qualificam o uso do sistema, como: “fácil”, “difícil”, “bom”, “ruim”, considerando que grande parte de pesquisas relacionadas à análise de sentimentos tem-se centrado sobre o uso de adjetivos [Taboada et al. 2011], [Benamara et al. 2007];
- Advérbios, a fim de caracterizar o comportamento dos usuários acerca do sistema: “bem”, “mal”, “muito”, “pouco”, “mais”, “menos”, também considerando pesquisas [Benamara et al. 2007] que utilizam advérbios na análise de sentimentos.

Associadas a estes padrões, as buscas foram combinadas com o nome do sistema e de suas funcionalidades. Com a seguinte explicação: se o usuário irá falar do sistema, ele usará o nome do sistema ou os nomes de suas funcionalidades. Os nomes das funcionalidades utilizadas foram: “tweets”, “followers” e “following”.

Foi utilizada uma ferramenta a fim de contar e coletar as postagens referentes às buscas efetuadas no Twitter usando os padrões propostos. A ferramenta foi desenvolvida usando a linguagem PHP e a API *Search* do Twitter para resgatar as postagens. As postagens obtidas foram àquelas definidas com perfil público. O perfil público do Twitter permite o acesso de qualquer pessoa à suas informações e postagens. Cada busca retornava uma quantidade de 0 a 1500 postagens e estas eram armazenadas em um arquivo Excel. As coletas ocorreram duas vezes por mês, durante três meses. No início e no meio de cada mês, nas seguintes datas de 2012: 02/10, 17/10, 02/11, 17/11, 02/12 e 17/12, seguindo a seguinte ordem: busca individual, somente o nome do sistema e de suas funcionalidades e, em seguida, associando-os aos 16 padrões definidos acima: 5 substantivos, 1 interrogação, 4 adjetivos e 6 advérbios, resultando em 17 arquivos para o nome do sistema e 17 para cada funcionalidade, totalizando 68 arquivos por coleta.

2.2. Análise das postagens

A análise efetuada consiste em contabilizar o resultado das postagens relacionadas ao uso do sistema a fim de verificar a eficácia do padrão utilizado. Por eficácia de um padrão, entende-se aquele cujo resultado da pesquisa traz postagens relacionadas ao uso. Por exemplo, o padrão “nome do sistema + erro” apresentou a seguinte postagem relacionada ao uso: “*O **Twitter** fica dando **erro** quando eu quero dar uma resposta*” mas também apresentou a seguinte postagem que não é relacionada ao uso: “*O maior **erro** da minha vida foi ter apresentado o **Twitter** pra minha mãe!*”. As postagens selecionadas sobre o uso podem ser perguntas, reclamações ou mesmo elogios sobre o Twitter. Também foram consideradas mensagens retuitadas (RT), considerando opinião similar (interesse ou sentimento) sobre o assunto. A análise foi realizada de forma manual pelos especialistas em usabilidade.

2.3. Resultados

Foram contabilizados, para cada arquivo de busca, o total de postagens obtidas e o total de postagens que se referiam ao uso do sistema. Em seguida, foi realizada a média das 6

coletas. Neste trabalho são apresentados os resultados referentes aos padrões individuais e associados a substantivos (Tabelas 1 e 2), descrevendo a média das postagens coletadas (n); a porcentagem (%) média de postagens relacionadas ao uso do sistema e o desvio padrão (σ) das porcentagens obtidas.

Tabela 1 - Média das Buscas individuais

Padrão	n	%	σ
Twitter	1437	2	0,7
Tweets	1189	6	7,4
Following	1292	2	0,6
Followers	1474	0	0

Tabela 2 - Média da busca associada a substantivos

Padrão de Busca		n	%	σ
Twitter	Usabilidade	22	78	14,1
	Interface	609	61	30,7
	Erro	1148	82	13,0
	Problema	1041	50	34,0
	Bug	978	85	16,8
Tweets	Usabilidade	5	25	14,1
	Interface	106	68	29,1
	Erro	512	65	17,6
	Problema	984	59	28,4
	Bug	711	81	26,8
Padrão de Busca		n	%	σ
Following	Usabilidade	4	0	0
	Interface	22	5	6,4
	Erro	185	8	17,1
	Problema	513	2	1,8
	Bug	333	5	5,6
Followers	Usabilidade	1	0	0
	Interface	10	2	4,4
	Erro	182	36	12,1
	Problema	251	2	1,5
	Bug	1049	12	2,9

Os resultados foram analisados de forma quantitativa, por meio dos dados obtidos, e de forma qualitativa, pela observação dos analistas durante a classificação das postagens. Observou-se que somente o nome do sistema ou de suas funcionalidades (Tabela 1) é insuficiente para recuperação de postagens relacionadas ao uso, pois, em geral, o usuário utiliza o nome do sistema para interagir com os demais, não somente para relatar problemas no uso. Quando os padrões eram associados a substantivos (Tabela 2), foram obtidas porcentagens mais altas relacionadas ao uso do sistema. Para o padrão "Twitter + erro", por exemplo, foi obtida uma média de $n=1148$ postagens e uma porcentagem média de 82% de postagens relacionadas ao uso. As porcentagens obtidas por este padrão nas 6 coletas foram: 85%, 59%, 91%, 84%, 96% e 76%. O desvio padrão foi calculado em cima das 6 porcentagens obtidas e deu um valor de $\sigma = 13$, significando que 66,6% dos valores de uma distribuição normal encontram-se dentro da faixa do desvio padrão, tanto para mais quanto para menos em relação à média. Apenas as porcentagens 59% e 96%, referentes a 2ª e 5ª coleta ficaram fora desta média.

3. Considerações finais e trabalhos futuros

Este trabalho enfrentou e deverá enfrentar alguns desafios durante seu desenvolvimento. Um deles consistiu no grande volume de dados gerados a serem analisados. A finalidade de construir uma ferramenta para apoiar este tipo de trabalho consiste exatamente em efetivar o contexto de avaliação da usabilidade em sistemas de grande porte. Outro desafio consiste na discriminação de postagens que falem bem ou mal do sistema. No resultado quantitativo, a segunda maior porcentagem de postagens relacionadas com o uso do sistema concentram-se em padrões negativos, como: "Problema", "Erro" e "Bug", o que nós leva a supor uma maior quantidade de reclamações. Também foi observado que os padrões positivos foram mais utilizados a fim de fornecer um contexto

negativo ao sistema, como: "*não é fácil*", "*nada bom*", "*bem ruim*". Outro desafio consiste em identificar qual é o problema do usuário em relação ao sistema utilizado. Existirão situações em que é necessário determinar em que tarefa o usuário teve dificuldade. Por exemplo, frases como: "*Como eu faço para postar uma foto aqui?*" ou "*Este campo sempre dá erro!*". A resolução deste problema envolve conceitos como relacionamento proposicional, correferência ou expressões referenciais.

Embora esta investigação não tenha focado em estudos mais detalhados da sociolinguística [Tarallo 1990], existem vários desafios a serem considerados em um SS com grande diversidade de usuários que implicam em variações de classe social, sexo, localidade, etc.. Para localidade, um estudo interessante seria considerar variantes regionais presentes nas postagens dos usuários para se referirem ao sistema, como por exemplo, o termo "*Deu pau no sistema*" para se referir "*Problema no sistema*". Como trabalhos futuros, pretende-se aplicar as regras definidas em uma Ferramenta de apoio na análise de usabilidade a fim de filtrar as postagens relacionadas ao uso do sistema, classificando-as em positivas e negativas, identificar os problemas de usabilidade usando PLN e por fim, demonstrar a aplicabilidade da ferramenta proposta em uma situação real de avaliação.

4. Referências

- Benamara, F., Cesarano, C., Picariello, A. and Reforgiato, D. (2007). "Sentiment analysis: adjectives and adverbs are better than adjectives alone". In proceedings of international conference on weblogs and social media.
- Cybis, W.A., Betiol, A.H. e Faust, R. (2007). "Ergonomia e usabilidade: conhecimentos, métodos e aplicações". 1. ed. São Paulo: Novatec, 352 p.
- Gaizauskas, R. e Wilks, Y. (1998). "Information extraction: beyond document retrieval." In journal of documentation 54 (1), 70–105.
- Jeffrey, R. e Chisnell, D. (2008). "Handbook of usability testing: how to plan, design, and conduct effective tests". In wiley publishing.
- Paul, S. A., Hong, L., Chi, E. H.. (2011). "What is a question? crowdsourcing tweet categorization". In CHI 2011.
- Pereira, R., Baranauskas, M. C. C., Silva, S. R. P. (2010). "Softwares sociais: uma visão orientada a valores". In IHC '10, p. 149-158.
- Preece, J.; Rogers, Y.; Sharp, H. (2005). "Design de interação: além da interação homem computador". In bookman, 2005.
- Riloff, E. and Jones, R. (1999). "Learning dictionaries for information extraction by multi-level bootstrapping". In proceedings of aaai-99.
- Russell, S. and Norvig, P. (1995). Artificial Intelligence - a modern approach. In Prentice-Hall.
- Taboada, M., Brooke, J., Tofiloski, M., Voll, K. and Stede, M. (2011). "Lexicon-based methods for sentiment analysis". In Association for Computational Linguistics.
- Tarallo, F. (1990). "A pesquisa Sociolinguística". In Atica.
- Zhao, X., Sosik, V. S. and Cosley, D. (2012). "It's complicated: how romantic partners use facebook". In CHI'12.