Pontifícia Universidade Católica
do Rio de Janeiro

Dárlinton Barbosa Feres Carvalho

Combining a process and tools to support the analysis of online communities applied to healthcare

TESE DE DOUTORADO

Thesis presented to the Programa de Pós-Graduação em Informática of the Departamento de Informática, PUC-Rio as partial fulfillment of the requirements for the degree of Doutor em Informática.

Advisor: Prof. Carlos José Pereira de Lucena

Rio de Janeiro
March 2013

Dárlinton Barbosa Feres Carvalho

Combining a process and tools to support the analysis of online communities applied to healthcare

Thesis presented to the Programa de Pós-Graduação em Informática, of the Departamento de Informática do Centro Técnico Científico da PUC-Rio, as partial fulfillment of the requirements for the degree of Doutor.

Prof. Carlos José Pereira de Lucena
Advisor
Departamento de Informática — PUC–Rio

Prof. Hugo Fuks
Departamento de Informática — PUC-Rio

Prof. Simone Diniz Junqueira Barbosa
Departamento de Informática — PUC-Rio

Prof. Sean Wolfgand Matsui Siqueira
Departamento de Informática Aplicada — UNIRIO

Prof. Mariano Pimentel
Departamento de Informática Aplicada — UNIRIO

Prof. José Eugenio Leal
Coordinator of the Centro Técnico Científico of PUC–Rio

Rio de Janeiro March 22nd, 2013

Dárlinton Barbosa Feres Carvalho

Dárlinton has a Master's Degree in Computer Science from the Pontifical Catholic University of Rio de Janeiro (PUC-Rio) and a Bachelor's Degree in Computer Science from the Federal University of Ouro Preto (UFOP). He is experienced in various business fields. He has worked at Siemens Corporate Research (Princeton, USA), Automatos (Rio de Janeiro) and the Brazilian Ministry of Education (Brasilia). He is currently a researcher of the Software Engineering Laboratory at PUC-Rio and R&D Coordinator at RNP.

Bibliographic data

Carvalho, Dárlinton Barbosa Feres

Combining a process and tools to support the analysis of online communities applied to healthcare / Dárlinton Barbosa Feres Carvalho; Advisor: Carlos José Pereira de Lucena. — 2013.

82 f. : il. (color.) ; 30 cm

Tese (doutorado) — Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Informática, 2013.

Inclui bibliografia

1. Informática – Teses. 2. Processo para estudar mídias sociais. 3. Ferramentas computacionais personalizadas. 4. Mapa de Associação de Comunidades. 5. Problema de Seleção de Conteúdo. 6. Análise de fóruns online. 7. Análise de comunidades online. 8. Análise de mídias sociais. 9. Estudo de questões de saúde. 10. Ciência social computacional. 11. Ciência da Web. I. Lucena, Carlos José Pereira. II. Pontifícia Universidade Católica do Rio de Janeiro. Departmento de Informática. III. Título.

CDD: 004

Acknowledgments

First of all, I would like to thank all those who have helped me in my personal development and made this work possible. Regarding this thesis, I am grateful to have worked with Rodrigo Pazzini, who came up with the research theme and further developed it with our team. I am also grateful to other colleagues, friends, and mentors who have contributed greatly to this thesis research, namely: Wilma Madeira, Paulo Vasconcellos, Ricardo Marcacini, Hugo Fuks, Solange Rezende, and Sérgio Rosa. Special thanks go to the research colleagues from the Software Engineering Lab/PUC–Rio, under the command of Prof. Carlos Lucena, where I found support to discuss ideas about research during my studies. Finally, thanks to the CNPq and PUC–Rio for the financial support. Thank you all!

Abstract

Carvalho, Dárlinton Barbosa Feres; Lucena, Carlos José Pereira. **Combining a process and tools to support the analysis of online communities applied to healthcare**. Rio de Janeiro, 2013. 82p. DSc Thesis — Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

This research thesis is aiming to exploit valuable social media, especially those available in online communities of social network sites, in order to perform social studies about healthcare issues. Based on a practical approach, a process was defined to conduct such studies. This process relied on tailored computational tools to provide support for specific tasks such as content retrieval, selection, and analysis. Two tools that stand out are presented because of their utility and the complexity of the process in which their development was based on. The first tool, for the benefit of online community analysis, is the Community Association Map, a process developed to support experts in understanding users' interests based on their associations within their communities. Our second tool (TorchSR) aims to aid analysts in the selection of discussions from online forums to be manually analyzed by (qualitative) research techniques (*e.g.* content and discourse analysis). This task, which was defined as solving the content selection problem, was tackled with a tool based on unsupervised machine learning techniques, such as hierarchical clustering. An exploratory study case shows that TorchSR helps analysts in dealing with the problem. The proposed process was employed in two studies about relevant healthcare issues (*i.e.* hepatitis C and drug abuse) which resulted in interesting findings in the field of public health. In conclusion, this thesis presents a practical application of computational social science to the field of health, through development of a process and tools used to support analysts and improve its application.

Keywords

Process to study social media. Tailored computational tools. Community Association Map. Content selection problem. Online forum analysis. Online community analysis. Social media analysis. Study of healthcare issues. Computational social science. Web science.

Resumo

Carvalho, Dárlinton Barbosa Feres; Lucena, Carlos José Pereira. **Combinando um processo e ferramentas para apoiar a análise de comunidade online aplicados à área de saúde**. Rio de Janeiro, 2013. 82p. Tese de Doutorado — Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Esta pesquisa de tese teve como objetivo explorar a análise de mídias sociais, especialmente as disponíveis em comunidades online de sites de redes sociais, a fim de realizar estudos sociais sobre questões de saúde. Com base em uma abordagem prática foi definido um processo para realizar esses estudos. Este processo contou com ferramentas computacionais adaptados para fornecer apoio em tarefas específicas, tais como recuperação de conteúdo, seleção e análise. Duas ferramentas que se destacam são apresentadas por causa de sua utilidade e a complexidade do processo em que a sua construção se baseou. Para o benefício da análise de comunidades online, o Mapa de Associação de Comunidades é um processo desenvolvido para apoiar especialistas em compreender os interesses dos usuários com base em suas associações dentro de suas comunidades. A outra ferramenta visa auxiliar analistas a selecionar discussões de fóruns online a serem analisados manualmente com técnicas de pesquisa qualitativa, por exemplo, análise de conteúdo e do discurso. Esta ferramenta, TorchSR, foi criada baseada em aprendizado de máquina não supervisionado, usando agrupamento hierárquico, para dar suporte na resolução do problema de seleção de conteúdo. Um estudo de caso exploratório mostra que esta ferramenta ajuda na resolução do problema. O processo proposto foi utilizado em dois estudos sobre questões relevantes de saúde (hepatite C e o abuso de drogas), que resultou em descobertas relevantes sobre saúde pública. Em conclusão, este trabalho apresenta a aplicação prática de ciência social computacional no campo da saúde, através do desenvolvimento de um processo e ferramentas utilizadas para apoiar os analistas e melhorar a sua aplicação.

Palavras–chave

Processo para estudar mídias sociais. Ferramentas computacionais personalizadas. Mapa de Associação de Comunidades. Problema de Seleção de Conteúdo. Análise de fóruns online. Análise de comunidades online. Análise de mídias sociais. Estudo de questões de saúde. Ciência social computacional. Ciência da Web.

Contents

List of Figures