Research Project Proposal for DCC-UFMG Class "*Sistemas de Recomendação*"

J.F.R. Fonseca*
BlueDog Technology Studio
Belo Horizonte, Brazil
jfra.fonseca@gmail.com

ABSTRACT

Proposal for a Research Project to be performed as part of the requisites for concluding the class "*Sistemas de Recomendação*" of the Department of Computer Science (DCC) of the Federal University of Minas Gerais (UFMG).

Keywords

Recommender Systems, Text Matching, Legal Compliance, Privacy, Privacy Policy

1. INTRODUCTION

The maintainers of most applications, services, web-pages, data or software packages, require users to agree and comply with a Privacy Policy Statement [1], a legal document that informs how the data collected from the user will be used or shared with third-parties. The act of agreeing with a Privacy Policy attests the user's compromise of compliance with the legal requirements as stated by the application maintainer, and concordance with the use given to the data collected. The user may be prompted to read the legal document during the installation of a software package or creation of a web-service account, and by proceeding with the operation at hand the user is considered to have given consent for the collect and use of data.

McDonald and Cranor (2008) estimate that a user would take 7 minutes to briefly skim through a medium-sized Privacy Policy Statement. They also estimated that the citizens of the United States of America collectively would spend 53.8 *billion* hours to properly read all the Privacy Policies prompted to them, every year [2]. In this paper, we propose a system capable of recommending sentences of interest from a document, to be carefully analyzed by the reader. The recommendation is based on the analysis of the document's contents and a previously elaborated knowledge and probabilistic model.

*Course-free class student.

We expect to minimize time cost of reading Privacy Policy Statements without compromising the user's knowledge of its inner proceedings.

2. METHODOLOGY

We used information retrieval techniques like fuzzy string matching [3] to detect sentences from documents that are more closely related to a set of predetermined legal sentences known to be a threat to user privacy [4]. Such sentences were elaborated after manual analysis of several Privacy Policies and consulting with a lawyer and different user privacy defender groups. Potentially harmful sentences identified in documents are presented for the user to review, in decreasing order of similarity with known harmful sentences.

2.1 Results Validation

Singh, Sumeeth and Miller (2011) evaluated how much users were able to comprehend of a Privacy Policy by applying a quiz test [5]. We propose a similar approach, prompting users to read the recommended sections of a Privacy Policy Statement, measuring the amount of time spent, and evaluating their scores on a quiz test about the document. We expect users to obtain similar or better scores than [5], spending, for each Privacy Policy analysed, less than the seven minutes estimated by [2] to skim through a medium-sized document.

3. REFERENCES

[1] C. Jensen and C. Potts. Privacy policies as decision-making tools: an evaluation of online privacy notices. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 471–478. ACM, 2004.

[2] A. M. McDonald and L. F. Cranor. Cost of reading privacy policies, the. *ISJLP*, 4:543, 2008.

[3] G. Navarro. A guided tour to approximate string matching. *ACM Comput. Surv.*, 33(1):31–88, Mar. 2001.

[4] A. Newitz. Dangerous terms: A user's guide to eulas. https://www.eff.org/wp/dangerous-terms-users-guide-eulas, 2005. Accessed: September 17, 2015.

[5] R. I. Singh, M. Sumeeth, and J. Miller. A user-centric evaluation of the readability of privacy policies in popular web sites. *Information Systems Frontiers*, 13(4):501–514, 2011.