

Aplicação para coleta e análise de dados de reputação de produtos em comércios eletrônicos

Eduardo Augusto Ferreira da Silva¹, Viviane Torres da Silva², Ricardo Choren³

¹ Universidade Federal Fluminense (UFF) – Niterói, RJ - Brasil

² IBM Research (on leave from UFF) – Rio de Janeiro, RJ – Brasil

³ Instituto Militar de Engenharia (IME) – Rio de Janeiro, RJ – Brasil

eferreira@ic.uff.br, vivianet@br.ibm.com, choren@ime.eb.br,

***Abstract.** Reputation systems applied in the e-commerce domain aims to collect and provide information on system objects, that may be products, users (customers and providers) and organizations. A common problem of these systems is that the reputation of an object is not integrated, i.e., each system provides an different reputation for the same object. As a solution, this paper presents an application responsible for extracting data from various information sources regarding a product chosen by the user. It is able to integrate this information and present the reputation of the product on a consolidated way.*

***Resumo.** Sistemas de Reputação aplicados no domínio de comércio eletrônico tem como objetivo coletar e disponibilizar informações sobre objetos do sistema, que podem vir a ser usuários do sistema (vendedores e compradores), produtos e organizações que comercializam os produtos. Um problema comum destes sistemas de reputação é a não integração das informações sobre um mesmo produto presente em diferentes bases de dados. Como solução, este trabalho apresenta um aplicativo responsável por coletar em diversas fontes de dados informações referentes a um produto escolhido pelo usuário, integrar estas informações e apresentar a reputação do produto de maneira consolidada.*

1. Introdução

A definição de comércio eletrônico (ou *e-commerce* em inglês) é dado pelo uso da internet para facilitar, executar e processar transações de negócios [Delone e Mclean 2004]. Estas transações de negócios envolvem um comprador (consumidor), um vendedor (fornecedor) e troca de produtos ou serviços por dinheiro. Toda transação, como, por exemplo, a venda de um determinado produto feita através do sistema, pode ser registrada. Portanto, é comum nos sistemas de comércios eletrônicos, como ocorre na Amazon, Buscapé e Ebay a existência de *Sistemas de Reputação* que possuem o objetivo de avaliar a reputação de usuários do sistema, dos produtos negociados e das organizações que negociam o produto.

Os sistemas de reputação têm a responsabilidade de coletar, distribuir e considerar o histórico de interações dos usuários [Mui, Mohtashemi e Halberstadt 2002]. Um problema comum nos sistemas de reputação, em especial, nos sistemas

associados às aplicações de comércio eletrônico, é a não integração das informações sobre um mesmo produto disponíveis em diferentes sistemas de reputação (i.e., diferentes fontes de dados). Em [Khatiban 2012] é feito um experimento deste tipo integrado, porém sem ser automatizado. Como solução, este trabalho apresenta um aplicativo responsável por coletar em diversas fontes de dados informações referentes a um produto escolhido pelo usuário, integrar estas informações e apresentar a reputação do produto de maneira consolidada.

Os resultados dessas informações de reputação consolidadas pela aplicação em diversos comércio eletrônico diferentes podem ser úteis nas seguintes problemáticas: (i) se obter reputação de um produto em que o sistema de reputação que ainda não tenha tido nenhuma experiência de interação com aquele produto; (ii) agir como uma fonte de informação complementar de reputação, considerando as interações referentes ao produto em outros comércio eletrônico;

2. Proposta de Aplicação

Em busca de solução para resolver o problema da não integração de dados entre os sistemas de reputação para comércio eletrônico, é proposta uma aplicação responsável por:

- Coletar em diversas fontes de dados, i.e., diferentes comércio eletrônico, informações referentes a um produto escolhido pelo usuário;
- Integrar estas informações extraídas;
- Apresentar a reputação do produto de maneira consolidada.

Nas próximas seções, serão apresentados os detalhes da proposta da aplicação: a arquitetura; processo de coleta de dados; informações de reputação; armazenamento; e consulta de resultados.

2.1. Arquitetura

A arquitetura da aplicação é dividida em dois módulos. O módulo *batch* é responsável pela extração e processamento de dados e o módulo *front-end* é responsável pela interface com o usuário final da aplicação. Esse módulo provê a funcionalidade de consulta da reputação do produto através das informações processadas previamente pelo módulo *batch*. A figura 1 representa a arquitetura da aplicação:

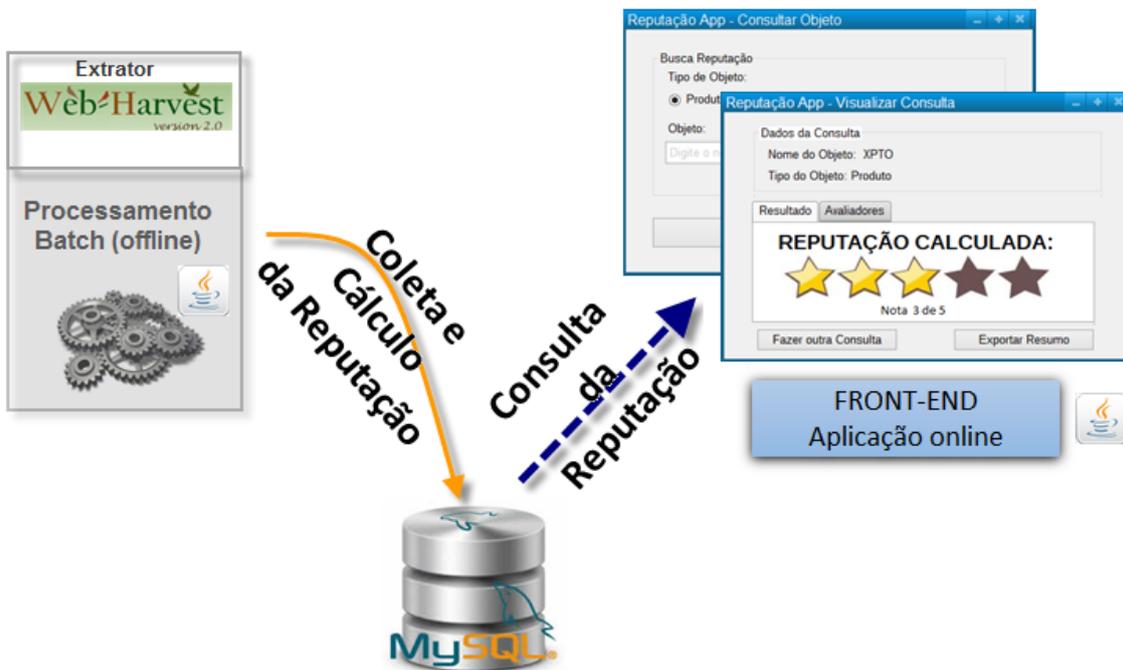


Figura 1. Arquitetura da Aplicação

Por se tratar de um trabalho acadêmico, foram adotados o uso de software livre. A solução prevê o uso da plataforma Java para ambos os módulos. Para o banco de dados relacional é utilizado o SGBD (Sistema de Gerenciador de Banco de Dados) MySQL versão 5.6.23.

2.2. Módulo Batch: Coleta de dados

A ideia da aplicação é coletar em diferentes fontes de dados informações sobre as reputações de produtos. A extração da informação é feita *offline* pois o processo de extração é bastante custoso. O processo pode envolver a análise de páginas web ou a utilização de scripts disponibilizados pelos sistemas de comércio eletrônico para a extração de dados de suas páginas web. Nesta versão da ferramenta utilizamos um *plugin* feito em Java de uma ferramenta de extração (*scrapping*): Web-haverst* 2.0.

O módulo *batch* é capaz de colocar informação não somente sobre as avaliações feitas por usuários sobre um produto, mas também informações qualificando estas avaliações e informações associadas aos avaliadores do produto. As informações capturadas relativas a avaliações sobre um produto são, por exemplo: nota da avaliação, data da avaliação e características prós e contras do produto. As informações relativas a qualificação das avaliações são, por exemplo: quantos usuários avaliaram a avaliação e percentagem de avaliações positivas e negativas. Já as informações capturadas sobre o avaliador são, por exemplo: quantidades de avaliações realizadas sobre produtos similares e percentagem de avaliações positivas e negativas. Toda informação é armazenada em banco de dados.

A figura 2 ilustra a extração de dados nos comércios eletrônicos:

* Web-haverst: <http://web-harvest.sourceforge.net>

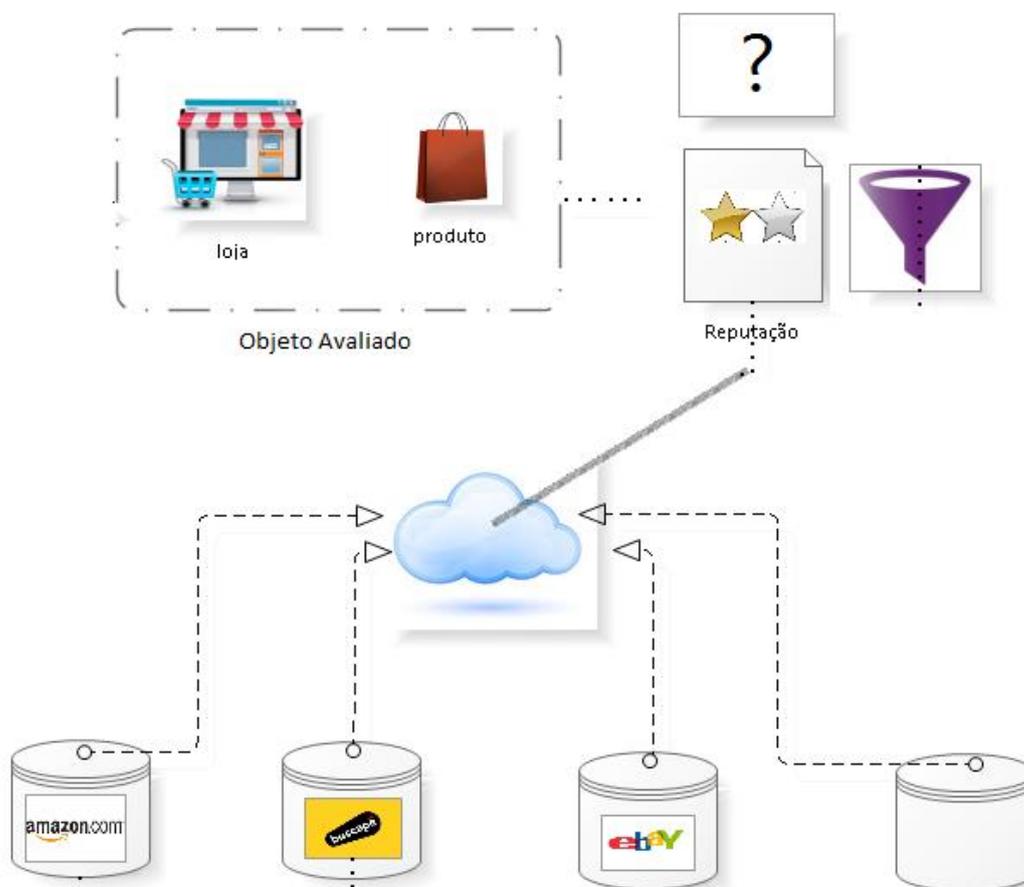


Figura 2. Coleta de dados

2.3. Informações de Reputação

Não é escopo deste trabalho detalhar o mecanismo utilizado para a consolidação das reputações de um mesmo produto. Este mecanismo ainda está em desenvolvimento. Contudo é possível afirmar que o mecanismo agrupa informações sobre as reputações dos produtos fornecida por diferentes usuários em diferentes sistemas de comércio eletrônico juntamente com informações relativas as avaliações sobre estas avaliações e informações relativas aos usuários que avaliaram o produto.

2.4. Consulta de Resultados

A consulta sobre a reputação de um determinado produto é realizada com base no nome deste produto. A figura 3 apresenta protótipos das interfaces gráficas utilizadas para consultar e para informar o resultado da consulta. O resultado da consulta informa não somente a reputação calculada mas também informações sobre os avaliadores que forneceram tais avaliações e sobre a qualificação de suas avaliações. Segue a figura 3:

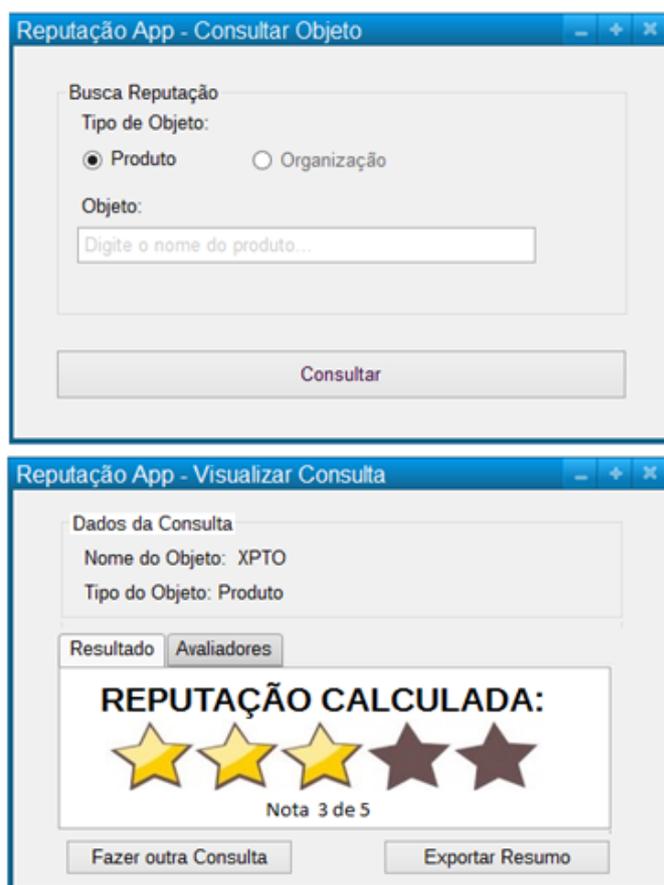


Figura 3. Interface gráfica da aplicação. Tela de Consulta e Tela de Visualização

3. Trabalhos Futuros

O módulo batch atualmente só utiliza um determinado script para a extração das informações. É nosso objetivo estender este módulo para que outros scripts sejam utilizados e outros sistemas de comércio eletrônico sejam analisados. Atualmente, conseguimos analisar os comércios eletrônicos Amazon, Buscape e Ebay.

Como comentado na seção 2.3, o mecanismo para consolidação das reputações ainda está em andamento. Atualmente estamos analisando as diferentes possibilidades de considerar as informações que capturamos com o objetivo de dar mais informação para o usuário sobre a reputação consolidada e sobre os avaliadores do produto. A medida que o mecanismo de consolidação for sendo definido acreditamos ser necessário adaptar a interface de apresentação dos dados.

4. Considerações

O objetivo deste trabalho é propor uma aplicação responsável por coletar em diversas fontes de dados informações referentes a um produto escolhido pelo usuário, integrar estas informações e apresentar a reputação do produto de maneira consolidada.

Este trabalho é inovador pois os sistemas de reputação disponíveis na literatura, como [Jacovi et al. 2014] [Sabater e Sierra 2002] [Mui, Mohtashemi e Halberstadt 2002] [Melnik e Alm 2002] [Resnick et al. 2006] [Valera et al 2001] calculam a

reputação dos produtos com base somente nas informações contidas no próprio sistema de reputação. Embora os sistemas de reputação digam que possuem uma abordagem para a Web, estes sistemas estão acoplados a somente um sistema de comércio eletrônico e utilizam somente as informações geradas por esse sistema para o cálculo da reputação dos produtos.

5. Referências

- Delone, W. H., e Mclean, E. R. (2004). Measuring e-commerce success: Applying the DeLone & McLean information systems success model. *International Journal of Electronic Commerce*, 9(1), 31-47.
- Jacovi, M. et al. (2014). The perception of others: inferring reputation from social media in the enterprise. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing* (pp. 756-766). ACM.
- Khatiban, S. (2012). Building reputation and trust using federated search and opinion mining. In *Proceedings of the 21st international conference companion on World Wide Web* (pp. 151-154). ACM.
- Melnik, M. I., e Alm, J. (2002). Does a seller's ecommerce reputation matter? Evidence from eBay auctions. *The journal of industrial economics*, 50(3), 337-349.
- Mui, L., Mohtashemi, M. e Halberstadt, A. (2002) "Notions of reputation in multi-agents systems: a review". In: *Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 1*. ACM, p. 280-287.
- Resnick, P. et al. (2006). The value of reputation on eBay: A controlled experiment. *Experimental Economics*, 9(2), 79-101.
- Sabater, J., e Sierra, C. (2002). Reputation and social network analysis in multi-agent systems. In *Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 1* (pp. 475-482). ACM.
- Valera, F. et al. (2001). Communication management experiences in e-commerce: using a multiagent system to provide intermediation service in an e-commerce environment. *Communications of the ACM*, 44(4), 63-69.