

# Arquitetura de um Sistema de Recomendação Baseado em Ontologia para Anúncios de Carros

Fábio A. P. de Paiva<sup>1</sup>, José A. F. Costa<sup>2</sup>, Cláudio R. M. Silva<sup>3</sup>, Ricardo S. França<sup>4</sup>

<sup>1,2,4</sup> Departamento de Eng. Elétrica – Universidade Federal do RN (UFRN)  
Natal – RN – Brasil

<sup>3</sup> Departamento de Eng. de Comunicações – Universidade Federal do RN (UFRN)  
Natal – RN – Brasil

fabiopaiva@yahoo.com,  
{jafcosta, claudio.rmsilva, ricardoluizsf}@gmail.com

**Abstract.** *Recommender systems have emerged as one interesting approach to tackle the problem of information overload, however most of them have a problem. They fail when there are no identical keywords for an exact match of a search. In order to overcome this limitation, recently several proposals for systems have been presented. Some of them have proposed the integration of ontologies to improve the recommendation process. This paper presents an architecture for an ontology-based system and implements a prototype which demonstrates how it can be used to inside a portal to sell cars.*

**Resumo.** *Os sistemas de recomendação surgiram como uma abordagem interessante para resolver o problema da sobrecarga de informação. Entretanto a maioria deles falha quando não há palavras-chave idênticas para uma correspondência exata em uma pesquisa. Para minimizar essa limitação, recentemente várias novas propostas têm sido apresentadas. Algumas delas têm procurado explorar os benefícios das ontologias no processo de recomendação. Este trabalho apresenta uma arquitetura de um sistema baseado em ontologias e utiliza uma implementação de protótipo para demonstrar como ela pode ser usada em um portal de vendas de veículos.*

## 1. Introdução

Diariamente os usuários da Internet e suas aplicações criam cerca de 2,5 quintilhões de bytes de dados. Algumas estimativas calculam que 90% dos dados de hoje foram criados nos dois últimos anos [Zikopoulos *et al.* 2012]. Nesse contexto, um dos principais desafios de um usuário web é identificar informações que atendam às suas preferências e é por isso que os serviços personalizados de recomendação tornaram-se cada vez mais necessários [Kang e Choi 2011] e amplamente utilizados em várias áreas.

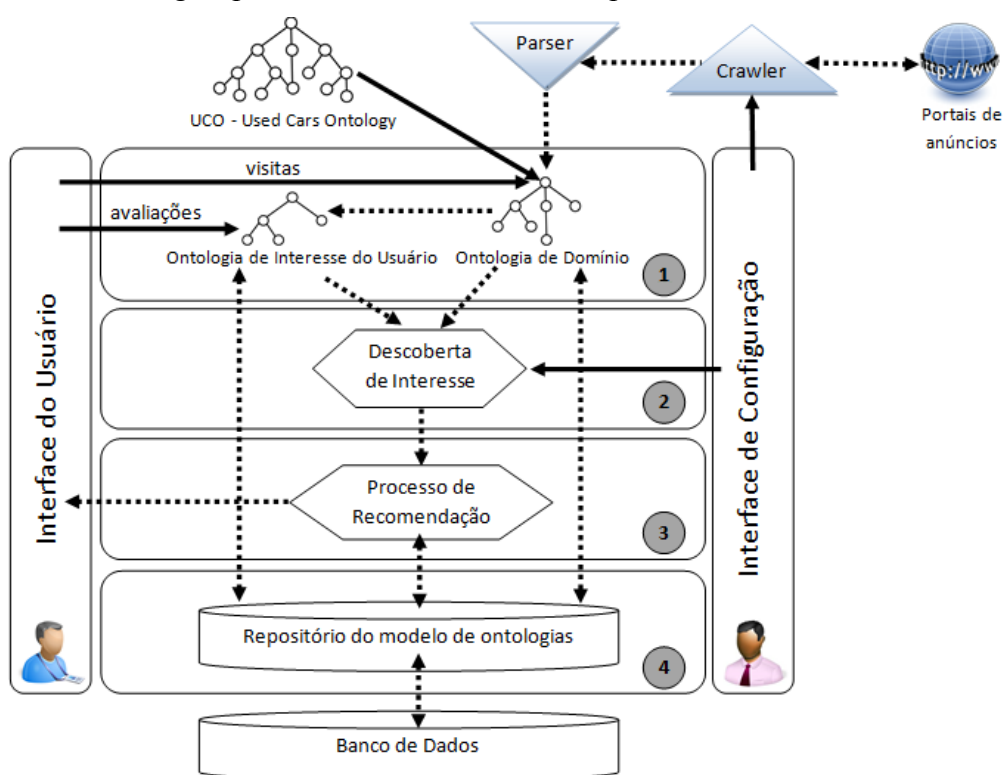
Os sistemas de recomendação surgiram como uma abordagem para resolver o problema da sobrecarga de informação. Eles são considerados aplicações especiais que fornecem sugestões personalizadas sobre produtos (ou serviços) que podem ser interessantes aos usuários. Os sistemas tradicionais sugerem itens (*e.g.*, uma música, um filme ou um livro) usando técnicas de mineração de texto [Gruber 1993]. No entanto, esses sistemas falham quando não há palavras-chave idênticas, mesmo existindo uma relação semântica entre elas [Kang e Choi 2011]. Para minimizar esse problema, nos últimos anos, vários estudos propuseram o uso de ontologias [Gao *et al.* 2008], [Zhen *et*

al. 2010], [Kang e Choi 2011], [Ge *et al.* 2012] como uma maneira de aumentar o desempenho dos sistemas de recomendação.

Neste trabalho, é proposta uma arquitetura de um sistema de recomendação baseado em ontologia o qual fornece informações personalizadas por meio dos relacionamentos entre os interesses do usuário e os anúncios de carros disponíveis na web. Na seção 2, são apresentadas a arquitetura e a forma de cálculo do grau de interesse; na seção seguinte, a técnica e o mecanismo de recomendação utilizados são descritos e, na última seção, as considerações finais são apresentadas.

## 2. Arquitetura Proposta

A Figura 1 apresenta a arquitetura hierárquica proposta de um sistema de recomendação baseado em ontologias para auxiliar usuários na compra de carros usados.



**Figura 1. Arquitetura proposta do sistema**

A arquitetura é baseada em quatro camadas, conforme a descrição abaixo:

1. Camada de Contexto — formada pelas ontologias de domínio (subseção 2.1) e de interesse do usuário (subseção 2.2);
2. Camada de Descoberta de Interesse — a partir das avaliações do usuário, é calculado o grau de interesse pelos conceitos da ontologia;
3. Camada de Recomendação — responsável por: a) calcular a similaridade entre os usuários, b) prever itens que serão interessantes para um determinado usuário e c) *rankear* a lista de anúncios de acordo com o interesse do usuário;
4. Camada de Ontologias — responsável pelo armazenamento das instâncias que representam os artefatos dos modelos de ontologia utilizados na arquitetura.

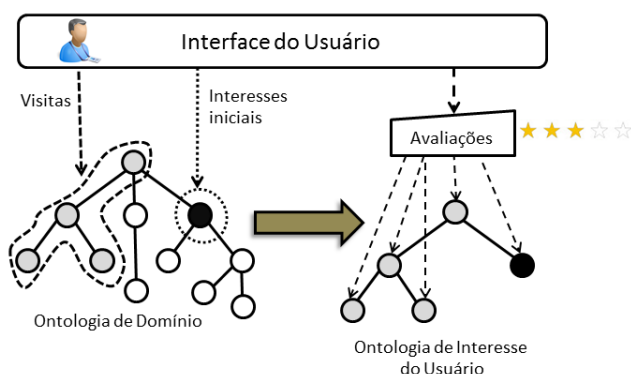
## 2.1. Ontologia de Domínio

Algumas vezes é interessante que apenas uma parte de uma ontologia seja reaproveitada. Este processo é chamado de modularização e consiste na extração de um subconjunto (também chamado de módulo) da ontologia original [Doran 2006]. Aqui, está sendo usado um módulo do modelo *Used Cars Ontology* (UCO) [MakoLab 2012].

Neste trabalho, um web *crawler* é encarregado de localizar anúncios web referentes a carros usados. O conteúdo descoberto nos anúncios é usado como entrada para construir uma lista de itens disponíveis para recomendação. Depois que o conteúdo dos anúncios é devidamente extraído por um componente com a função de *parser*, os anúncios são usados para popular a ontologia de domínio. A configuração do web *crawler* é definida pelo administrador do sistema na interface de configuração.

## 2.2. Ontologia de Interesse do Usuário

A ontologia que representa os interesses do usuário é um subconjunto da ontologia de domínio e, para construí-la, é realizado um mapeamento entre os interesses do usuário e os conceitos da ontologia de domínio (Figura 2). A fim de tornar as recomendações mais personalizadas, o usuário avalia os itens de acordo com as suas preferências.



**Figura 2. Processo de construção da Ontologia de Interesse do Usuário**

É importante garantir que as características dos itens recomendados combinem com os interesses do usuário [Ge *et al.* 2012] a fim de garantir recomendações que atendam às suas necessidades. O interesse de cada usuário é representado por um modelo chamado de Modelo de Interesse do Usuário e formalmente pode ser definido como:

$$\Theta = (D, I, G, S), \quad \text{onde:}$$

- $D = \{\text{login, sexo, estadoCivil, anoNascimento, numeroFilhos, renda}\}$  representa os dados pessoais do usuário;
- $I = \{I_1, I_2, \dots, I_n\}$  é o conjunto de interesses (interesses iniciais + interesses descobertos através de interações) representados na ontologia do usuário;
- $G = \{G_1, G_2, \dots, G_n\}$  é o grau de interesse do usuário por cada um dos conceitos de  $I$  e;
- $S = [S_{ij}]_{n \times n}$  é a matriz que representa a similaridade entre todos os usuários.

### 2.3. Grau de Interesse do Usuário

Depois de apresentadas as ontologias de domínio e de interesses, é calculado o grau de interesse do usuário por cada um dos conceitos de sua ontologia. Através de um formulário, realiza-se a avaliação dos itens {ótimo, bom, razoável, ruim, péssimo} de acordo com a pontuação {5, 4, 3, 2, 1}, respectivamente. O *feedback* explícito (avaliações) de um usuário em relação a um conceito  $C$ , representado por  $\text{Exp}(C_i)$ , está no intervalo  $[-1, 1]$  e é calculado de acordo com a Equação 1:

$$\text{Exp}(C_i) = \frac{\text{pontuacao}(C_i) - 3}{2} \quad (1)$$

Em seguida,  $\text{Exp}(C_i)$  é normalizado para o intervalo  $[0, 1]$ . Já o *feedback* implícito (visitas), é baseado nas frequências de acesso e é calculado conforme Equação 2:

$$\text{Imp}(C_i) = \frac{F(C_i)}{\sum_{j=1}^n F(C_j)} \quad (2)$$

onde  $F(C_i)$  é a frequência de acesso aos conceitos da ontologia de interesse. Por fim, o grau de interesse em relação a um conceito  $C$ ,  $G(C_i)$ , é calculado pela Equação 3:

$$G(C_i) = \alpha * \text{Exp}(C_i) + \beta * \text{Imp}(C_i), \quad (3)$$

onde  $\alpha$  e  $\beta$  são pesos que influenciam diretamente no cálculo do grau de interesse do usuário. A soma desses pesos é igual a 1 ( $\alpha + \beta = 1$ ) e os seus valores são definidos pelo administrador do sistema.

## 3. Mecanismo de Recomendação

### 3.1. Filtragem Colaborativa

Após o cálculo do grau de interesse, os usuários são agrupados de acordo com suas similaridades e, para tal, são utilizados os Mapas Auto-Organizáveis e o algoritmo *K-Means*. O objetivo é que o sistema utilize a Filtragem Colaborativa para recomendar a um determinado usuário anúncios considerados interessantes baseado na opinião de outros usuários que apresentam perfis similares ao seu.

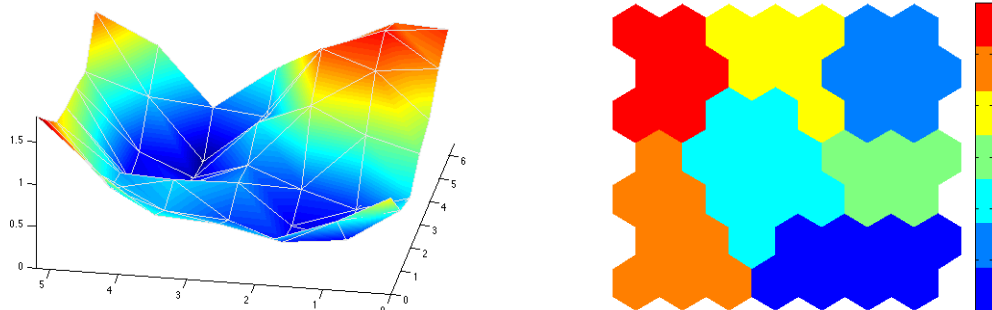
Os Mapas Auto-Organizáveis (*Self-Organizing Maps* ou simplesmente SOM) são algoritmos de redes neurais artificiais que se baseiam no aprendizado competitivo não-supervisionado, o que significa que o treinamento é inteiramente orientado pelos dados e que leva em consideração apenas os padrões de entrada [Kohonen 1997]. Cada neurônio  $i$  do mapa é representado por um vetor de peso  $p$ -dimensional  $m_i = [m_{i1}, m_{i2}, \dots, m_{ip}]^T$ , onde  $p$  é igual à dimensão do vetor de entrada [Costa e Netto 2001].

A Matriz-U é um método de visualização usada, normalmente, com os mapas SOM para análise de agrupamentos. Ela se baseia na distância do espaço de entrada entre um vetor de peso e os seus vizinhos no mapa [Yamaguchi e Ichimura 2011]. Já a Matriz-U\* [Utsch 2003] leva em consideração a densidade da informação a fim de aprimorar os resultados da Matriz-U convencional.

### 3.2. Resultados Preliminares

O vetor de entrada do algoritmo SOM é formado pelo conjunto de grau de interesses do usuário,  $G(C_i) = \{C_1, C_2, \dots, C_n\}$ , e por alguns de seus dados pessoais, representados numericamente. Os dados utilizados nos experimentos foram obtidos de forma

simulada. No entanto, a fim de aproximá-los da realidade, eles foram gerados a partir de uma distribuição normal e, em seguida, normalizados na faixa de valores [0, 1].



**Figura 3. Matriz- $U^*$  construída a partir dos dados de entrada do SOM** **Figura 4. Mapa SOM representando 7 grupos de usuários similares**

O conjunto de dados foi analisado por um mapa SOM, empregado na construção da Matriz- $U^*$  (Figura 3). O algoritmo *K-Means* foi utilizado para segmentar a  $U^*$ . E, para determinar o número adequado de agrupamentos, usou-se o *CDbw* (*Composed Density between and within clusters*) [Halkidi e Vazirgiannis 2008], um índice que avalia a compacidade e a separação de grupos definidos por um algoritmo de agrupamento. Ao fim da execução, são obtidos sete grupos que representam os perfis de usuários similares (Figura 4) e a matriz de similaridade correspondente,  $[S_{ij}]_{n \times n}$ .

### 3.3. Processo de Recomendação

A listagem de recomendação é gerada a partir de regras definidas pelo administrador. A performance do sistema é determinada por essas regras que atuam como parâmetros de configuração. Alguns exemplos das regras usadas são a) o *threshold* que determina o interesse (ou não) do usuário por um anúncio, b) a atribuição de valores aos pesos  $\alpha$  e  $\beta$  (ver seção 2.3), c) os parâmetros de inicialização do algoritmo SOM e outros.

Quando um usuário  $n$  é similar a um usuário  $u$ , pode-se dizer que  $n$  é um vizinho de  $u$ . Depois de realizado o agrupamento de usuários, o próximo passo é prever os itens  $i$  que ainda não foram visualizados pelo usuário  $u$ , mas que já foram avaliados anteriormente pelos seus vizinhos  $n$ , conforme Equação 4 [Schafer *et al.* 2007]:

$$previsao(u, i) = \bar{r}_u + \frac{\sum_{n \in vizinhos(u)} simUsuario(u, n) * (r_{ni} - \bar{r}_n)}{\sum_{n \in vizinhos(u)} simUsuario(u, n)} \quad (4)$$

onde  $\bar{r}_u$  e  $\bar{r}_n$  são, respectivamente, as avaliações médias dos usuários  $u$  e  $n$ , enquanto  $r_{ni}$  é a avaliação do usuário  $n$  sobre o item  $i$ . Finalmente, os itens recomendados são ordenados baseado no valor de  $G(C_i)$ .

## 4. Conclusões

Neste trabalho, é apresentada uma arquitetura hierárquica de um sistema baseado em ontologias para recomendação de anúncios de carros usados. A arquitetura utiliza duas ontologias: uma de domínio e outra para representar os interesses do usuário. A fim de aplicar a técnica de Filtragem Colaborativa, um mapa SOM é empregado para agrupar usuários com características e interesses similares. Enquanto os sistemas de

recomendação tradicionais utilizam palavras-chave para representar os interesses do usuário, este trabalho utiliza conceitos de ontologia. Dessa forma, o modelo de interesses será mais adequado à realidade do usuário e conseqüentemente o mecanismo de recomendação sugerirá um maior número de anúncios que atendam às reais necessidades do usuário. O trabalho ainda está em fase de desenvolvimento e a continuidade na implementação de outras funcionalidades é o foco de trabalhos futuros.

## Referências

- Costa, J. A. F., Netto, M. L. A. (2001). Clustering of complex shaped data sets via Kohonen maps and mathematical morphology. In *Proceedings of the SPIE, Data Mining and Knowledge Discovery*. B. Dasarathy (Ed.), vol. 4384, pp. 16-27.
- Doran, P. (2006). Ontology reuse via ontology modularisation. In *Proceedings of Knowledge Web PhD Symposium*, pp. 1-6.
- Ge, J., Chen, Z., Peng, J. e Li, T. (2012). An ontology-based method for personalized recommendation. In *11th International Conference on Cognitive Informatics & Cognitive Computing (ICCI\*CC)*, pp.522-526.
- Gao, Q., Yan, J. e Liu, M. (2008). A Semantic Approach to Recommendation System based on User Ontology and Spreading Activation Model". In *International Conference on Network and Parallel Computing*, pp. 488-492.
- Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Journal Knowledge Acquisition*, pp. 199-220.
- Halkidi, M. e Vazirgiannis, M. (2008). A density-based cluster validity approach using multi representatives. *Pattern Recognition Letters*, vol. 29, pp. 773-786.
- Kang, J. e Choi, J. (2011). An ontology-based recommendation system using long-term and short-term preferences. In *International Conference on Information Science and Applications*, pp. 1-8.
- Kohonen, T. (1997). *Self-Organizing Maps*. 2nd. Ed., Berlim: Springer, Verlag.
- Makolab [site] (2012). Used Cars Ontology Metadata. Disponível em: <http://ontologies.makolab.com/uco/ns.html>. Acessado em: 15 jun. 2013.
- Schafer, J. Ben, Frankowski, D., Herlocker, J. e Sen, S. (2007). The adaptive web, P. Brusilovsky, A. Kobsa e W. Nejdl, Springer-Verlag, Alemanha, p. 291-324.
- Yamaguchi, T. e Ichimura, T. (2011). Visualization using multi-layered U-Matrix in growing Tree-Structured self-organizing feature map. *Systems, Man and Cybernetics (SMC), IEEE International Conference*, p. 3580-3585.
- Ultsch, Alfred. (2003). U\*-Matrix: a Tool to visualize Cluster in high dimensional Data. Technical Report No. 36, Dept. of Mathematics and Computer Science, University of Marburg, Germany.
- Zikopoulos, P., Eaton, C., Deutsch, T., Deroos, D. e Lapis, G. (2012). *Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data*. New York: McGraw-Hill, pp. 176.
- Zhen, L., Huang, G. Q. e Jiang, Z. (2010). An inner-enterprise knowledge recommender system. In *Expert Systems with Applications*, vol. 37, pp. 1703-1712.