

A representação do conhecimento por meio de Redes Semânticas: um estudo sobre a contribuição da Mineração de Dados

Cleônidas Tavares de Souza Junior (Mestrando - MCTI), cleonidas@gmail.com;

Renelson Ribeiro Sampaio (Orientador - MCTI), renelson.sampa@gmail.com;

Hernane Borges de Barros Pereira (Coorientador - MCTI), hernanebbpereira@gmail.com;

Faculdade SENAI CIMATEC

Palavras Chave: Teoria de Grafos, Redes Semânticas, Mineração de Dados, Conhecimento.

Introdução

Em sua tese de doutorado, ROSA (2016) registra como diferentes autores entendem e explicam o que são Redes Semânticas. Para o autor, “Redes semânticas são sistemas de representação do conhecimento baseados em grafos cujos vértices são palavras e as arestas, os relacionamentos entre as palavras estabelecidos por alguma regra” (ROSA, 2016).

O estudo dos grafos, na matemática, foi iniciado pelo matemático suíço Leonhard Euler (1707-1783) ao analisar e propor uma solução para o problema das pontes de Königsberg. A partir de então, a Teoria dos Grafos (TG) tem sido usada para resolução de problemas relacionados a diferentes áreas como Ciências Sociais, Estatística, Logística, Redes Semânticas, entre outras.

Para os estudos referentes a Redes Semânticas (RS), a Teoria dos Grafos é uma das ferramentas que apoiam a análise dos resultados; além dessa teoria, a Mineração de Dados (MD) também exerce influência no processo de construção da Rede Semântica. Grosso modo, os estudos da RS envolvem: (a) definição do problema; (b) coleta e tratamento de um corpus (MD); (c) construção da Rede (TG); e (d) análise dos resultados.

A coleta e tratamento dos dados podem ser feitos de diferentes formas e, conseqüentemente, gerar diferentes resultados. Caldeira (2005), ao analisar diferentes textos, propôs um trabalho de tratamento dos dados baseado na exclusão de

artigos e preposições; na concatenação de termos que remetem a um significado; etc.

O estudo feito neste trabalho, baseado em um texto qualquer, apontou que a omissão de um artigo, por exemplo, é capaz de alterar o significado de uma sentença. Cabe investigar o que uma série de alterações de significados é capaz de produzir em um processo de representação do conhecimento.

Neste trabalho pretende-se analisar como diferentes autores elaboraram os tratamentos de dados em seus trabalhos e como essas modificações podem contribuir para a criação de uma representação do conhecimento.

Métodos e Resultados parciais

Para auxiliar nas análises deste trabalho, foi construído um programa na linguagem C++ que segue uma lista de regras. Inicialmente é proposta uma regra x que reconhece as palavras de um texto e gera uma lista; o resultado desse processo inicial são dois arquivos (nodes.csv e edges.csv). Os arquivos gerados servem de subsídios para a construção de uma rede no programa Gephi.

Para gerar a lista, a regra x seguiu os seguintes critérios: (a) excluir os pontos, as vírgulas, as interrogações e as exclamações do texto sob análise; e (b) criar relações (arestas) entre as palavras (vértices) um-a-um (i.e. a primeira palavra

Seminário Anual de Pesquisa – 2017

Faculdade SENAI CIMATEC

Programa de Pós-Graduação em Modelagem Computacional e Tecnologia Industrial

Programa de Pós-Graduação em Gestão e Tecnologia Industrial

III Workshop de Gestão, Tecnologia Industrial e Modelagem Computacional.

ISSN online 2447-9640

Workshop de Gestão, Tecnologia Industrial e Modelagem Computacional

com a segunda, a segunda com a terceira e assim sucessivamente).

Abaixo segue a rede gerada a partir da aplicação do Gephi nos arquivos criados pela regra x:

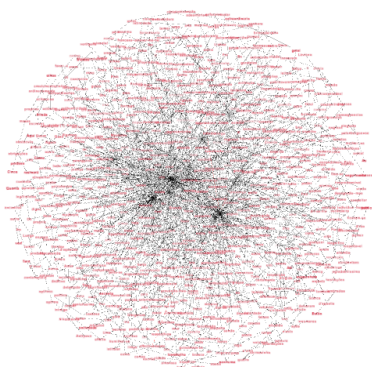


Figura 1: Rede elaborada no Gephi.

Os pontos mais escuros na rede ilustrada na Figura 1 são os vértices que apresentam os maiores graus (número de ligações entre vértices). Esses vértices representam as palavras “o”, “a”, “de”, “que”.

As palavras “o” e “a” podem ser artigo ou pronomes, dependendo de como elas são usadas nas sentenças. Numa estratégia de analisar e tratar os vértices de maior grau e, como próximos passos, este trabalho irá propor a análise da regra $x+1$ a ser aplicada em um texto e (a) fará a exclusão dos artigos; ou (b) concatenar os artigos aos seus referentes.

Sabe-se que, quando reduzimos um sintagma como “o dentista” para “dentista”, há uma mudança no seu significado. A palavra “dentista”, sem a presença do artigo, perde a informação de gênero masculino e de determinação (a informação de que o dentista em questão é conhecido e o único indivíduo no contexto).

Por outro lado, se, ao invés de excluir os artigos, eles fossem concatenados aos seus referentes, ter-se-ia, por exemplo, “odentista” e/ou “adentista”. Nesse caso, na rede gerada pela regra $x+1$, esses

artigos deixariam de ser vértices de maior grau, manteriam seu conteúdo semântico e mudariam o formato apresentado na Figura 1.

Conclusões

Nos estudos de RS, a etapa de MD é muito importante para a formulação das redes. Ela é capaz de restringir ou ampliar o significado das palavras contribuindo, dessa forma, para uma representação do conhecimento com maior (ou menor) poder de generalização.

Falta, para este estudo, analisar como diferentes autores trabalham com as questões de tratamento de dados e como o processo de mineração de dados pode contribuir para os estudos sobre Redes Semânticas e para a construção de uma representação do conhecimento.

Referências

- BRAGA, Patrícia Freitas. **Um Modelo Computacional para Extração Textual e Construção de Redes Sociais e Complexas**. Dissertação (Mestrado). Senai CIMATEC, Salvador, 2010.
- CALDEIRA, Silvia Maria Gomes. **Caracterização da rede de signos lingüísticos: um modelo baseado no aparelho psíquico de Freud**. Dissertação (Mestrado). Fundação Visconde de Cairu, Salvador, 2005.
- GOLDBARG, Marco César; GOLDBARG, Elizabeth. **Grafos: conceito, algoritmos e aplicações**. Rio de Janeiro: Editora Elsevier, 2012.
- ROSA, Marcos Grilo. **Modelo empírico para analisar a robustez de redes semânticas**. Tese (doutorado) Faculdade de Educação, Universidade Federal da Bahia, Salvador, 2016.

Workshop de Gestão, Tecnologia Industrial e Modelagem Computacional.

Seminário Anual de Pesquisa – 2017

Faculdade SENAI CIMATEC

Programa de Pós-Graduação em Modelagem Computacional e Tecnologia Industrial

Programa de Pós-Graduação em Gestão e Tecnologia Industrial

III Workshop de Gestão, Tecnologia Industrial e Modelagem Computacional.

ISSN online 2447-9640