

8 Conclusões e Trabalhos Futuros

No presente capítulo são apresentadas as conclusões obtidas e sugestões para trabalhos futuros.

8.1. Conclusões

Para o desenvolvimento deste trabalho foi buscada uma fundamentação teórica sobre a MT e CT enfatizando os procedimentos e características comuns aos classificadores e métodos de categorização utilizados nesta Tese. Foi apresentada uma visão geral sobre a Categorização de Textos, destacando conceitos e características de Aprendizagem de Máquina. Na abordagem adotada, os classificadores fazem uso de uma coleção de textos previamente categorizados para construir um modelo estatístico de predição capaz de categorizar novos documentos.

A combinação de classificadores e o emprego de tratamento linguístico comprovam que esta abordagem obtém ótimos resultados.

Os trabalhos correlatos também contribuíram para a execução dessa dissertação. A escolha dos trabalhos teve um impacto na aplicação de ideias e conceitos.

Na elaboração da metodologia, foram apresentadas a coleção, as categorias, a representação dos documentos, a heurística proposta e a avaliação; enfim, todo o planejamento da metodologia que foi utilizada nos experimentos. Essa metodologia é crucial para a realização, descrição e análise dos experimentos, permitindo identificar e discutir os problemas encontrados.

Entre as principais contribuições inovadoras proporcionadas, direta ou indiretamente, por esta tese podem ser destacadas:

- Utilização de técnicas de processamento linguístico em favor da obtenção de resultados de maior qualidade.
- Elaboração de heurísticas que auxiliam os processos de Mineração de Textos, como não utilizar os algoritmos de lematização em verbos.
- Assistência ao processo de descoberta de conhecimento permitindo que usuários com pouco conhecimento nas áreas de Mineração de Textos ou Categorização de Textos possam utilizar a ferramenta com êxito.
- A preparação para utilização de um Léxico verbal com termos do Português Brasileiro contendo informações sobre verbos, suas flexões e formas canônicas.
- Comprovação de que a criação de léxicos auxilia a obtenção de termos com maior valor semântico.
- A construção de um Léxico de sinônimos e antônimos com termos do Português Brasileiro.
- Um ambiente de software para a realização de Categorização Automática de Textos em Português do Brasil baseada em Mineração de Textos.

Conclui-se que o trabalho atingiu o fim a que se propôs e apresentou resultados satisfatórios quando consideradas as obras existentes na literatura.

As ferramentas utilizadas para a construção do *framework* assim como a abordagem empregada na realização dos estudos de casos foram descritas e fundamentadas, de maneira a permitir que outros pesquisadores possam vir a agregar novos algoritmos ao sistema, com um mínimo de esforço.

8.2. Trabalhos Futuros

Como trabalhos futuros, este trabalho sugere:

- A inclusão de outros algoritmos de classificação ao *framework*.

- Maior estudo sobre a combinação de classificadores também é desejável.
- Expansão do léxico criado seria de grande utilidade para muitos trabalhos futuros.
- Utilização de processamento distribuído ou paralelo, principalmente para aplicações de tempo real.

Outro ponto interessante é avaliar o *framework* em um domínio diferente, exigindo a necessidade de etiquetar uma coleção de documentos para esse fim.