7
Conclusão e Trabalhos Futuros

7.1.
Conclusão

Foi introduzida nesta dissertação uma abordagem para um problema recorrente hoje: obtenção de informação relevante no ambiente da *Web*. Apoiada por Mineração de Textos (MT), a metodologia proposta dispensa a necessidade de configurações extensas e conjuntos de dados para treinamento da ferramenta de coleta e análise de dados textuais.

Os resultados obtidos por esta metodologia de coleta de dados mostram que a utilização de um documento exemplo como objetivo do processo de coleta de dados é uma abordagem interessante e que deve ser ainda mais investigada como meio de obtenção de informação.

Os benefícios adquiridos pela abordagem de textos como textos, e não como dados puramente estatísticos, mostram a centralidade da linguagem no processo de Mineração de Textos. Novas técnicas de Processamento de Linguagem Natural tendem a incrementar os resultados satisfatórios neste campo.

A utilização do dicionário *Thesaurus* permitiu que o processo de coleta não fosse desorientado por ocorrências de figuras de linguagem como sinonímia. Entretanto, há a necessidade de participação maior da comunidade científica para a criação cooperativa de uma base maior de termos *thesaurus* para a Língua Portuguesa.

A fundamentação teórica comprova que embora recente, a área de Mineração de Textos não é novidade e que já vem sendo empregada com sucesso no mundo acadêmico e corporativo. Bastante interdisciplinar, o estudo sobre MT remete a extensas pesquisas de outras áreas como Recuperação de Informação e Inteligência Artificial.

Seguindo uma metodologia para MT proposta em outros trabalhos encontrados na literatura e composta, nesta ordem, pelas etapas de coleta, pré-

processamento, indexação, mineração e análise, o processo de Mineração de Textos obteve resultados satisfatórios.

7.2. Trabalhos Futuros

Sugestões para futuros trabalhos envolvem as diversas tarefas de Mineração de Textos que podem ser empregadas no contexto do problema que é abordado por este estudo. Entretanto, uma destas tarefas seria especialmente útil a este problema: a Sumarização. Neste contexto, a Sumarização de Documentos irá permitir uma análise rápida de uma grande base de documentos por humanos.

Outros algoritmos para realizar a tarefa de Clusterização podem ser utilizados. A abordagem baseada em Lógica Nebulosa é uma das possibilidades, pois, permitiria maior flexibilidade na formação de *clusters* e até mesmo a multipertinência de documentos a mais de um *cluster*, o que ocorre em muitos casos do mundo real.

O emprego de técnicas desenvolvidas para o processo de Descoberta de Conhecimento na *Web* irá permitir a obtenção de informações adicionais e peculiares sobre um documento hipertexto (*Web Mining* de Contexto). Além disso, a análise do grafo de contexto (*Web Mining* de Estrutura) dos *web-sites* poderá indicar as autoridades sobre determinado assunto, o que permitirá obter com maior eficiência os documentos de maior importância sobre aquele tópico.

Assim como ocorreu com o processo de *KDD*, o processo de *KDT* também passa pelos seus estágios evolutivos. Atualmente, realizar uma tarefa de *KDT* exige grande interação humana, o que já não ocorre em um processo de *KDD*. Com o amadurecimento tecnológico desta área, será possível evoluir nesta direção: automatização de um processo de Descoberto de Conhecimento em Textos.

Esta dissertação teve como motivação a tese de doutorado de Christian Nunes Aranha (ARANHA C. N., 2007) e a dissertação de Mestrado de João R. Carrilho Jr. (CARRILHO, 2007), e terá continuidade com a dissertação de mestrado de Roberto Miranda Gomes, dentre outras.