

# 1 Introdução

## 1.1. Motivação

Informações podem ser armazenadas das mais variadas maneiras. O modo mais clássico de armazenamento de informação é através da palavra escrita, impressa. O acesso à informação estocada desta forma é lento, difícil, e de baixo rendimento. Nestes casos, para todas as etapas da manipulação da informação é necessária a presença do ser humano, que com suas limitações na capacidade de aquisição de conhecimento e processamento de grande volume de informação, constitui o principal gargalo do processo (MANDEL, SIMON & DELYRA, 1997).

A modernização dos últimos anos tornou as tecnologias de informação uma realidade inerente às vidas de todos nós. Das grandes multinacionais às pequenas empresas, das instituições públicas ao ensino e na nossa própria casa, termos como informática, computador, Internet e multimídia, entre tantos outros, passaram a fazer parte das tarefas do dia-a-dia, transformando-se em instrumentos fundamentais do trabalho. Todo este avanço tecnológico proporcionou meios muito mais eficientes para o armazenamento e disseminação de informação, esta, no formato digital, uma condição necessária para o amplo uso dos computadores no seu processamento.

Contudo, o maciço crescimento da oferta de dados traz consigo grandes desafios em termos de obtenção de informação. Embora usados muitas vezes como sinônimos, os termos dados e informação possuem significados distintos:

- Dado: um fenômeno qualquer, desprovido de significado e contexto (LAUDON & LAUDON, 2002).
- Informação: resultado do processamento, manipulação e organização de dados que passam a ter significados e, portanto, podem ser contextualizados, interpretados e compreendidos (GOLDSCHMIDT & PASSOS, 2005).

Os dados estão por toda parte. A maioria das organizações não sofre falta de dados, mas, sim, de uma abundância de dados redundantes e inconsistentes (SINGH, 2001). A informação desejada encontra-se entre os bits e bytes armazenados, por exemplo, em um disco rígido. Esses dados, após uma série de processamentos que envolvem, por exemplo, operações lógicas, serão transformados em informação. Finalmente, quando essa informação é recuperada, interpretada e analisada, chega-se ao conhecimento. É o que está resumido no diagrama da Figura 1.



Figura 1 - Processo de obtenção de conhecimento

Métodos tradicionais de análise de dados, baseados principalmente no manuseio direto dos dados pelo homem, simplesmente não permitem a manipulação de conjuntos volumosos de dados (SINGH, 2001).

Técnicas de Mineração de Dados (MD) há algum tempo são utilizadas para lidar com o processo de obtenção de conhecimento em grandes bases de dados, essas em formato rigidamente estruturado. Porém, recentemente, comprovou-se que oitenta e cinco por cento de toda a informação do mundo está armazenada sob a forma de documentos textuais (GDS PUBLISHING, 2008) (IBM, 2008), ou seja, texto em formato livre, desprovidos de estruturas de dados, inviabilizando a utilização de técnicas de Mineração de Dados.

Técnicas de Recuperação de Informação sempre foram utilizadas para o armazenamento de documentos textuais e a recuperação de informação associada a eles (BAEZA-YATES & BERTIER, 1999) (MANNING, RAGHAVAN, & SCHÜTZE, 2007). Antes desta explosão de informação, as tarefas de recuperação de informação eram restritas a bibliotecas, nas quais com a ajuda de um bibliotecário, qualquer assunto poderia ser encontrado. Entretanto, a utilização

dessas técnicas na enorme massa de dados disponíveis, atualmente, não garante que a relevância das informações retornadas atenda às necessidades. Além disso, como todo processo de Recuperação de Informação precisa ser formalizado pelo usuário, por exemplo, através de uma consulta, estas técnicas oferecem suporte somente à obtenção da informação requisitada pelo mesmo, ignorando a possibilidade da existência de conhecimento, até então, desconhecido, nestes dados.

Em virtude do crescimento contínuo do volume de dados eletrônicos disponíveis, em formato textual, técnicas de extração de conhecimento automáticas tornam-se cada vez mais necessárias para manipular essa gigantesca massa de dados. **Mineração de Textos**<sup>1</sup> ou **Descoberta de Conhecimento em Textos**<sup>2</sup> surge, neste contexto, como uma abordagem à obtenção de informação útil a partir de bases de dados em formato textual. O principal objetivo das técnicas de Mineração de Textos é a manipulação de documentos em formato textual com o objetivo da obtenção do conhecimento implícito presente nestes (ARANHA & PASSOS, 2006).

Dentre as muitas tarefas desempenhadas em Mineração de Textos, a Categorização de Textos (CT) é a que tem recebido maior interesse da comunidade científica (LINDEN, 2008). Ao ser categorizado, um documento passa a pertencer a um grupo previamente definido que contém outros documentos que são semelhantes entre si. Uma vez identificados os documentos afins, torna-se mais fácil distinguir a informação relevante solicitada por um usuário, em meio a um conjunto de documentos irrelevantes, e repassá-la às pessoas que podem utilizá-la e a transformar em ação.

Separar a informação em categorias de conhecimentos que facilitam a sua manipulação e recuperação é o objetivo principal de Categorização de Textos. Além disso, classificar não é somente uma tendência natural do conhecimento; é igualmente uma necessidade da inteligência humana (PEIXOTO, BATISTA & CAPELO, 2003).

Contudo, por ser uma área ainda em fase de amadurecimento, as ferramentas de Mineração de Textos estão em processo de desenvolvimento.

---

<sup>1</sup> Do termo em inglês, *Text Mining*.

<sup>2</sup> Do termo em inglês, *Knowledge Discovery in Texts – KDT*.

Segundo (GOLDSCHMIDT & PASSOS, 2005), as ferramentas de Mineração de Dados podem ser classificadas quanto à usabilidade em:

- Primeira geração: ferramentas de análise dedicadas à realização de uma única tarefa de Mineração de Dados, como a construção de classificadores ou a descoberta de *clusters*, sem suporte às demais etapas do processo.
- Segunda geração: suítes de aplicativos capazes de realizar diversas operações de pré-processamento, como análises e transformação dos dados, além de tarefas de descoberta, como classificação, clusterização e visualização.
- Terceira geração: soluções de Mineração de Dados embarcadas. Constituem aplicativos para a solução de um problema específico de uma organização, em que as técnicas de Mineração de Dados são empregadas sem a necessidade de intervenção ou até mesmo conhecimento do usuário final. Alguns exemplos dessa geração são os *softwares* para análise de concessão de empréstimos.
- Quarta geração: aplicações que auxiliam o homem na condução do processo de descoberta do conhecimento.

Adotando esse mesmo critério para as ferramentas de Mineração de Textos, a maior parte das ferramentas disponíveis está classificada entre a segunda e a terceira geração.

Outra peculiaridade relacionada ao processo de Mineração de Textos é que diferente de tabelas e relações estatísticas, documentos textuais possuem uma linguagem associada a eles. Apesar do processamento total da linguagem natural ainda estar fora de alcance com a tecnologia atual (NEVES, 2012), existem técnicas que são capazes de extrair valiosas informações da linguagem natural presente nos documentos.

Embora possa ser abordado de forma puramente estatística, quando abordado de forma semântica, isto é, empregando a linguagem natural presente nos documentos textuais para obter melhor representação da informação, o processo de Mineração de Textos tende a melhores resultados (ARANHA C. N., 2007). Segundo (MONTEIRO, GOMES & OLIVEIRA, 2006), a Mineração de

Textos ainda é uma área pouco explorada e precisa do desenvolvimento de projetos que tratem as particularidades relacionadas ao Português do Brasil.

## 1.2. Objetivos do Trabalho

Os objetivos principais deste trabalho são pesquisar, propor, implementar e avaliar uma metodologia para realizar a Categorização Automática de Textos em Português do Brasil. O processo de Mineração de Textos proposto por (ARANHA C. N., 2007) com algumas adaptações é empregado para apoiar a Categorização de Textos, permitindo a utilização de processamento linguístico para o Português do Brasil e maior automação das tarefas.

Para isso, os seguintes objetivos intermediários são desejados:

- Prover um *framework* para a execução do processo de Categorização de Textos em Português do Brasil.
- Empregar técnicas de Mineração de Textos para o fornecimento de ferramentas que auxiliem a execução da tarefa de Categorização de Textos.
- Fornecer tratamento linguístico específico à Língua Portuguesa do Brasil, isto é, que faça proveito das ricas informações semânticas presentes em qualquer linguagem natural.
- Automatizar as etapas necessárias para a realização de Categorização de Textos, otimizando as técnicas de pré-processamento textual e a escolha dos algoritmos de Aprendizado de Máquina e seus respectivos dos parâmetros.

Como forma de avaliar os conceitos apresentados, foram realizados diversos experimentos de Categorização de Textos em Português do Brasil. Espera-se que esses estudos possam ser motivadores e um bom ponto de partida para outros pesquisadores.

### 1.3. Trabalhos relacionados

Esta seção tem por objetivo resumir as características dos principais trabalhos relacionados à área de Categorização de Textos com o emprego de Mineração de Textos e Processamento de Linguagem Natural.

O sistema Aíuri, desenvolvido por (SILVA A., 2007), é, segundo seu autor, um ambiente acadêmico cooperativo de alto desempenho, integrado a ambientes de *grids* computacionais, para a execução de algoritmos de Mineração de Textos. Faz uso dos *grids* do Intragrid NACAD, administrado pelo Núcleo de Computação de Alto Desempenho (NACAD) da COPPE, que agrupa máquinas heterogêneas do laboratório em um *grid* com finalidade didática, e do E-Infrastructure Shared Between Europe and Latin America (EELA), que é uma infraestrutura para o desenvolvimento e implantação de *grids* para uso científico, conectando a Europa e a América Latina.

Desenvolvido na linguagem de programação Java, o sistema Aíuri é provido de interface *web* e de toda uma infraestrutura para capacitá-lo a executar algoritmos em modo local ou em ambientes de *grids* computacionais. É composto por três módulos:

- O primeiro módulo realiza as atividades básicas para o funcionamento do sistema, como autenticação e carregamento dos arquivos de usuário.
- O segundo módulo é o responsável pelas tarefas de pré-processamento de dados textuais.
- O terceiro módulo contém as implementações dos algoritmos que são utilizados na etapa de Mineração do processo de descoberta de conhecimento: Classificadores Bayesiano e de Ranqueamento Linear. Além disso, contempla a fase de Pós-Processamento, com a disponibilização de diversas métricas de avaliação de resultados.

A etapa de pré-processamento empregada neste estudo é capaz de realizar abordagens (ver item 2.5) baseadas nas análises Estatística e Semântica. A primeira fase desta etapa é Tokenização (ver item 3.2.1). Em seguida, é realizada a eliminação de termos considerados irrelevantes ou *stopwords*. Por fim, utiliza-se

processamento linguístico para realizar a normalização morfológica dos termos (ver item 3.2.3.4) ou *stemming*.

Utiliza o Modelo de Espaço Vetorial para criar uma representação estruturada dos documentos (ver item 4.3.1). Os métodos utilizados para o cálculo de relevância (ver item 4.3.1.3) de cada termo são baseados na frequência desses no documento ou em toda a coleção: IDF e TF-IDF.

A base de documentos utilizada no estudo é um extrato do *corpus* CETENFolha (ver item 6.6). Experimentos nesta base foram realizados, ora com documentos distribuídos de forma balanceada entre as categorias, ora sem qualquer controle sobre a distribuição dos documentos e suas respectivas categorias.

A métrica utilizada para avaliar o desempenho dos classificadores foi a Medida F (ver item 3.5.3). A estratégia de treinamento empregada foi a de *hold-out*. Os estudos concluem que entre os dois classificadores utilizados, o classificador bayesiano, em média, foi mais rápido, e obteve modelos de classificação superiores aos obtidos pelo classificador de ranqueamento linear, sendo, desta maneira, uma boa opção de classificador para os textos em questão. Ambos os algoritmos tiveram desempenho superior nos conjuntos balanceados. Além disso, observou-se que para a maioria das classes quando houve a eliminação de *stopwords* e a utilização de *stemming*, ocorreram melhorias nos resultados obtidos, porém a utilização de *stemming* gerou grande aumento no tempo de processamento.

Em (MELO, 2007) é proposta uma abordagem ao reconhecimento de padrões textuais aplicada ao processo de classificação de documentos.

O modelo de representação de documentos utilizado é o Modelo de Espaço Vetorial (ver item 4.3.1.2). A seleção de atributos que compõem a representação vetorial dos documentos é baseada na escolha dos termos que possuem maior pontuação segundo as duas métricas de relevância utilizadas: escore de relevância e coeficiente de correlação.

Duas bases de documentos são utilizadas nos experimentos:

- A primeira é constituída por títulos e resumos de dissertações de mestrado e teses de doutorado do Departamento de Engenharia

Elétrica da COPPE/UFRJ. Possui cerca de quinhentos documentos organizados em cinco categorias.

- A segunda base é oriunda do corpus CETENFolha (ver item 6.6) e contém uma seleção de novecentos documentos também organizados em cinco categorias.

O processamento linguístico é empregado para identificar as classes gramaticais (ver item 3.2.3.2) dos termos que serão selecionados para o processo de classificação. Diversas combinações de classes gramaticais de termos são utilizadas, como por exemplo, somente o uso de termos substantivos, a combinação de substantivos e adjetivos, o conjunto de substantivos, nomes próprios e verbos, dentre outros.

Dois técnicas de Aprendizado de Máquina são utilizadas: Redes Neurais e Classificadores Bayesianos. Os melhores resultados foram obtidos com Classificadores Bayesianos, nas duas bases de documentos, utilizando cento e cinquenta termos selecionados entre as classes gramaticais substantivos, nomes próprios e adjetivos.

Em (CAMARGO, 2007) é apresentada uma abordagem linguística na classificação de textos em Português.

O modelo de representação de documentos utilizado também é o Modelo de Espaço Vetorial (ver item 4.3.1.2). Para a seleção de atributos representativos dos documentos são utilizadas duas métricas de relevância: escore de relevância e ganho de informação.

Semelhante ao trabalho de (MELO, 2007), duas bases de textos são utilizadas, sendo a primeira formada por um subconjunto CETENFolha de 855 textos classificados em cinco categorias: esporte, imóveis, informática, política e turismo, sendo 171 arquivos por classe. A outra base de textos é formada por um conjunto de textos gerados pela junção do título e resumo das teses de pós-graduação (mestrado e doutorado) da área de Engenharia Elétrica da Universidade Federal do Rio de Janeiro - COPPE-UFRJ. São 475 textos classificados nas categorias controle, microeletrônica, processamento de sinais, redes de computadores e sistemas de potência, sendo 95 arquivos por classe.

O processamento linguístico é utilizado para selecionar os termos pelas suas funções sintáticas e desta forma, tornar possível selecionar dentre as classes



gramaticais os termos que se pretende utilizar para representação dos textos. Para isso, o analisador sintático utilizado neste trabalho para extração de informações linguísticas dos documentos é o PALAVRAS, desenvolvido por (BICK, 2000) para a língua portuguesa. Ele realiza tarefas de processamento léxico-morfológico, análise sintática e faz parte de um grupo de analisadores sintáticos do projeto Visual Interactive Syntax Learning<sup>3</sup> (VISL), do Institute of Language and Communication da University of Southern Denmark.

Os classificadores utilizados são o Bayesiano, a Máquina de Vetor Suporte e um classificador baseado em regras de decisão. Dos resultados obtidos pode-se concluir que o classificador bayesiano é o que apresenta melhores resultados tanto para textos de jornal, com uma taxa de erro de 7,49%, quanto para textos científicos, com uma taxa de erro de 15,98%. Uma ressalva deve ser feita em relação ao classificador *Support Vector Machines* (SVM): embora apresente bons resultados em trabalhos anteriores (ver capítulo 5), não repetiu o desempenho para os experimentos realizados neste estudo.

O autor (LINDEN, 2008) utiliza a combinação de classificadores *K-Nearest Neighbors* (*k-NN*) e *Support Vector Machines* na categorização hierárquica de textos. O motivo pela categorização hierárquica de documentos, segundo o autor, é de que a distribuição de uma coleção de documentos em categorias auxilia na organização, busca e recuperação de informações, mas à medida que o número de documentos e o número de categorias aumentam, essa tarefa torna-se mais complexa para o ser humano; uma forma de ajudar a organizar informações no auxílio à compreensão humana é utilizar a organização hierárquica.

O modelo de representação de documentos utilizado é o Modelo de Espaço Vetorial (ver item 4.3.1.2). A seleção de atributos é realizada com a remoção de *stopwords* (ver item 3.2.2) e de termos que possuem frequência menor do que três em cada documento. Também é aplicada e a execução de um processo de lematização (ver item 3.2.3.4).

Nesse trabalho, a coleção de textos utilizada para o desenvolvimento dos experimentos é a **Folha-RICol**<sup>4</sup>, derivada do corpus em língua portuguesa

---

<sup>3</sup> Projeto VISL disponível em <http://beta.visl.sdu.dk/visl/pt/>

<sup>4</sup> Disponível em <http://www.linguateca.pt/Repositorio/Folha-RICol/>

CETENFolha (Corpus do NILC/Folha de São Paulo). Essa coleção compreende um conjunto de documentos organizados em 28 categorias hierárquicas.

A avaliação do processo de categorização é realizada pelo método *hold-out* e com a utilização do software WEKA. Os experimentos utilizam dois métodos combinatórios de classificadores: a votação e a heurística  $k$ -NN+SVM. A heurística citada propõe que nós da árvore de categorias com mais de dois filhos sejam classificados pelo  $k$ -NN, e nós com um filho utilizam o classificador SVM. Os resultados obtidos por meio da votação de classificadores foram superiores (valor da medida F1 = 94,4%) aos resultados obtidos pela heurística proposta (valor da medida F1 = 84,9%).

Em (REIS, 2011) utiliza-se Mineração de Textos para a categorização de cadeias de caracteres formadas por sequências de aminoácidos. Segunda a autora, uma palavra pode ser entendida como uma cadeia específica de caracteres que possuem um valor semântico. Abordar as proteínas como sequências de aminoácidos que possuem valor biológico permitiu o emprego das técnicas de Mineração de Textos.

Baseando-se nisso, foi desenvolvida uma metodologia capaz de treinar um classificador de textos sobre sequências de proteínas. A tarefa do classificador é identificar a categoria a que determinada proteína pertence baseada na sequência de aminoácidos que possui em sua estrutura.

A ferramenta de mineração de textos utilizada nos estudos de casos é o sistema Aíuri (SILVA A., 2007), comentado anteriormente. Desta forma, o texto (ou a sequência de aminoácidos) é representado matematicamente por meio do Modelo de Espaço Vetorial. A métrica de atribuição de pesos (ver item 4.3.1.3) utilizada para definir a frequência dos termos é TF-IDF.

A eliminação de *stopwords* (ver item 3.2.2) ou utilização de Lematização (ver item 3.2.3.4) não faz sentido no contexto dessa aplicação. A seleção das características mais representativas dos documentos (ou proteínas) é baseada simplesmente no valor de frequência de cada termo (aminoácido).

Os resultados obtidos com os dois algoritmos de classificação disponíveis na ferramenta, Bayesiano e Ranqueamento Linear, são semelhantes, bons, de fácil interpretação e possuem tempos de execução mínimos quando comparados às

ferramentas desenvolvidas para esse propósito e que não utilizam técnicas de Mineração de Textos.

(NEVES, 2012) propõe uma metodologia baseada em Mineração de Textos para a extração de informação de textos não estruturados em Português a fim de popular um *Data Mart* para a gestão das informações adquiridas.

A etapa de pré-processamento inicia com o processo de Tokenização (ver item 3.2.1) e conta com o auxílio de um léxico que é construído manualmente durante a execução desta fase e que valida os termos obtidos no processo de tokenização, além de identificar as Entidades Nomeadas (ver item 3.2.3.5) definidas pelo usuário. Utiliza o Modelo de Espaço Vetorial como estrutura de representação dos documentos.

Em seguida, inicia-se o processamento linguístico:

- O primeiro passo visa condensar os termos validados pelo usuário na etapa anterior em uma versão resumida por meio de um processo de Sumarização dos documentos. Após a realização da Sumarização, é iniciada a análise morfológica (ver item 3.2.3.2), realizada por meio de consultas ao léxico que contém para cada termo a sua classificação morfológica.
- Logo em seguida, começa o processo de Análise Sintática (ver item 3.2.3.6) de cada termo do documento sumarizado. Os dados resultantes do analisador sintático são agregados aos dados morfológicos dos termos, identificando para cada termo, além de sua classe gramatical, a sua função sintática no período analisado. Durante o processo de Análise Sintática também ocorre a identificação dos períodos do documento sumarizado.

Após o processamento linguístico, são realizadas tarefas de remoção de *stopwords* (ver item 3.2.2) e *stemming* (ver item 3.2.3.4). O passo seguinte é a classificação dos termos que consiste em identificar, nos períodos, os termos que serão candidatos a popular as dimensões do *Data Mart*. O processo de classificação é baseado em heurísticas que utilizam as informações obtidas na durante o processamento linguístico.

Por fim, é executado um processo de **ETL**<sup>5</sup> que irá popular o *Data Mart* gestão das informações adquiridas.

A Tabela 1 apresentada abaixo resume para cada um dos trabalhos relacionados os corpus, o modelo de representação dos documentos e a métrica de atribuição de pesos utilizados.

Tabela 1 - Resumo comparativo dos trabalhos relacionados quanto ao corpus, modelo de representação dos documentos e atribuição de pesos utilizados

<b>Autor</b>	<b>Corpus</b>	<b>Representação dos documentos</b>	<b>Atribuição de pesos</b>
(SILVA A., 2007)	CETENFolha	Modelo de Espaço Vetorial	IDF
			TF-IDF
(MELO, 2007)	CETENFolha	Modelo de Espaço Vetorial	Escore de relevância
	Teses		Coefficiente de correlação
(CAMARGO, 2007)	CETENFolha	Modelo de Espaço Vetorial	Escore de relevância
	Teses		Ganho de Informação
(LINDEN, 2008)	Folha-RICol	Modelo de Espaço Vetorial	Ganho de Informação
			Coefficiente de correlação
(REIS, 2011)	Aminoácidos	Modelo de Espaço Vetorial	TF-IDF
(NEVES, 2012)	Websites	Modelo de Espaço Vetorial	TF-IDF

A Tabela 2 resume para cada um dos trabalhos relacionados as técnicas de Processamento de Linguagem Natural empregadas, as tarefas de MT realizadas e os modelos de classificadores utilizados.

---

<sup>5</sup> Do termo em inglês, Extract Transform Load (Extração Transformação Carga).

Tabela 2 - Resumo comparativo dos trabalhos relacionados quanto a técnica de PLN, tarefa de MT e modelos de classificadores utilizados

<b>Autor</b>	<b>PLN</b>	<b>Tarefa</b>	<b>Classificadores</b>
(SILVA A. A., 2007)	Stemming	Classificação	Ranqueamento Linear
			Classificador Bayesiano
(MELO, 2007)	Stemming	Classificação	Redes Neurais
	POS Tagging		Classificador Bayesiano
(CAMARGO, 2007)	Stemming	Classificação	Classificador Bayesiano
	POS Tagging		SVM
			Regras
(LINDEN, 2008)	Stemming	Classificação	KNN
	POS Tagging		SVM
			Comitê
(REIS, 2011)	N/D	Classificação	Ranqueamento Linear
			Classificador Bayesiano
(NEVES, 2012)	Stemming	Sumarização	N/D
	POS Tagging	Extração de Informação	

Os trabalhos aqui apresentados serviram de inspiração para a elaboração da abordagem utilizada nesse trabalho que visa lidar com o problema de automatizar a tarefa Categorização de Textos apoiada por técnicas de Mineração de Textos.

#### **1.4. Organização da Tese**

Este tese possui mais sete capítulos conforme organização descrita abaixo.

O capítulo 2 apresenta os principais fundamentos da área de Mineração de Textos. Conceitos como os principais elementos de um processo de Mineração de Textos, a estrutura de um documento em linguagem natural, as características representativas de um documento, as abordagens ao processo de Mineração de Textos e as áreas correlatas são explicados.

No capítulo 3 são descritas as etapas da metodologia de Mineração de Textos estudada, composta por: coleta, pré-processamento, indexação, mineração e análise de resultados. As principais métricas utilizadas na análise de resultados também são comentadas.

O Capítulo 4 apresenta os conceitos da área de Recuperação de Informação que são utilizados para o entendimento da representação de documentos textuais,

assim como os modelos de representação de documentos e as métricas de atribuição de pesos.

Um estudo sobre a área de Categorização de Textos é apresentado no capítulo 5. Também é apresentada uma pesquisa de softwares dedicados a essa tarefa.

No capítulo 6 é relatado o *framework* proposto nesta tese e a descrição dos elementos que o compõem.

No capítulo 7 é apresentado o estudo de caso realizado e os resultados obtidos.

O capítulo 7 descreve as conclusões e as principais contribuições proporcionadas pelo presente trabalho. Alternativas de trabalhos futuros também são sugeridas.