

UNIVERSIDADE FEDERAL DE OURO PRETO  
INSTITUTO DE CIÊNCIAS EXATAS E BIOLÓGICAS  
DEPARTAMENTO DE COMPUTAÇÃO

ANA LUIZA ALMEIDA SOARES  
Orientador: Prof. Dr. Rodrigo Cesar Pedrosa Silva  
Coorientador: Mestre Pedro Saint Clair Garcia

**TÍTULO DA MONOGRAFIA: AVALIAÇÃO DE DIVERSAS APIS DE  
GECODIFICAÇÃO  
SUBTÍTULO**

Ouro Preto, MG  
2023

UNIVERSIDADE FEDERAL DE OURO PRETO  
INSTITUTO DE CIÊNCIAS EXATAS E BIOLÓGICAS  
DEPARTAMENTO DE COMPUTAÇÃO

ANA LUIZA ALMEIDA SOARES

**TÍTULO DA MONOGRAFIA: AVALIAÇÃO DE DIVERSAS APIS DE  
GECODIFICAÇÃO  
SUBTÍTULO**

Monografia apresentada ao Curso de Ciência da Computação da Universidade Federal de Ouro Preto como parte dos requisitos necessários para a obtenção do grau de Bacharel em Ciência da Computação.

**Orientador:** Prof. Dr. Rodrigo Cesar Pedrosa Silva

**Coorientador:** Mestre Pedro Saint Clair Garcia

Ouro Preto, MG  
2023

# Errata

AUTOR, **Título da Monografia:**Avaliação de diversas APIs de gecodificação subtítulo. nº de páginas. Monografia- Departamento de Computação, Universidade Federal de Ouro Preto, Ouro Preto, 2023.

Página	Linha	Onde se lê	Leia-se
16	10		

Este elemento pré-textual é opcional e deve ser inserido no texto, geralmente após o trabalho já ter sido impresso, após a folha de rosto. Deve conter a referência do trabalho e o texto da errata com a indicação da página a linha ([ABNT, 2011](#)).

Ana Luiza Almeida Soares

**TÍTULO DA MONOGRAFIA: AVALIAÇÃO DE DIVERSAS APIS DE  
GECODIFICAÇÃO  
SUBTÍTULO**

Monografia apresentada ao Curso de Ciência da Computação da Universidade Federal de Ouro Preto como parte dos requisitos necessários para a obtenção do grau em Bacharel em Ciência da Computação.

Aprovada em Ouro Preto, XX de mês de Ano.

---

Prof. Dr. Rodrigo Cesar Pedrosa Silva  
Universidade Federal de Ouro Preto  
Orientador

---

Mestre Pedro Saint Clair Garcia  
Universidade Federal de Ouro Preto  
Coorientador

---

Prof. Dr. Membro da Banca 1  
Universidade Federal de ... - UFXX  
Examinador

---

Prof. Dr. Membro da Banca 2  
Universidade Federal de ... - UFXX  
Examinador

*Espaço para prestar uma homenagem ou dedicar o trabalho a alguém.*

# Agradecimentos

Espaço para agradecer às pessoas e/ou instituições que contribuíram de forma relevante à elaboração do trabalho.

Alguma citação importante para a construção do trabalho. A fonte deve ser indicada e também deve constar na lista de referências bibliográficas.

# Resumo

Síntese do trabalho contendo um único parágrafo. O resumo deve ser feito de forma clara, concisa e seletiva de todo o texto, ressaltando o objetivo, o método, os resultados e a conclusão (ABNT, 2003). A norma da ABNT ainda recomenda que a primeira frase seja uma explicação do tema principal, seguindo da informação da natureza do trabalho (pesquisa experimental, pesquisa bibliográfica, estudo de caso, pesquisa de campo, etc.). Apresente os objetivos (geral e específicos); justificativa e a metodologia desenvolvida. Também deve ser inserido as conclusões finais, apresentando uma síntese dos principais resultados alcançados e o valor da pesquisa no contexto acadêmico. Sugere-se entre 150 a 500 palavras.

**Palavras-chave:** Palavra-chave 1. Palavra-chave 2. Palavra-chave 3.

As palavras-chave devem estar separadas por ponto e finalizadas também por ponto. Devem ser escolhidos termos que descrevem o conteúdo do trabalho.



# Abstract

This is the english abstract.

**Keywords:** Keywords1, Keywords2, Keywords3.

# Lista de Ilustrações

Figura 3.1 – Mapa de clusters que mostra a quantidade de escolas em cada região. Ao aproximar o mapa, o usuário consegue ver a localização de cada uma das escolas presentes no banco de dados. . . . .	5
Figura 3.2 – Mapa que mostra a cidade de Belo Horizonte, desenvolvido pela Prodabel. Na barra de pesquisa, é possível pesquisar os endereços e marcá-los no mapa.	6
Figura 3.3 – Esquematização do processo de preparação e geocodificação dos dados . . .	7

# Lista de Tabelas

# **Lista de Algoritmos**

# Lista de Abreviaturas e Siglas

ABNT	Associação Brasileira de Normas Técnicas
DECOM	Departamento de Computação
UFOP	Universidade Federal de Ouro Preto

# Lista de Símbolos

$\Gamma$	Letra grega Gama
$\Lambda$	Lambda
$\zeta$	Letra grega minúscula zeta
$\in$	Pertence

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Justificativa	1
1.2	Objetivos	1
1.3	Organização do Trabalho	1
1.3.1	Estrutura da Monografia	1
<b>2</b>	<b>Revisão Bibliográfica</b>	<b>3</b>
2.1	Trabalhos Relacionados	3
2.2	Fundamentação Teórica	3
2.3	Instruções	3
<b>3</b>	<b>Desenvolvimento</b>	<b>5</b>
3.1	Bases de Dados	5
3.2	Erro e métricas	6
3.3	Preparação dos Dados e Geocodificação	7
<b>4</b>	<b>Resultados</b>	<b>9</b>
<b>5</b>	<b>Considerações Finais</b>	<b>10</b>
5.1	Conclusão	10
5.2	Trabalhos Futuros	10
5.3	Publicações Realizadas	10
	<b>Referências</b>	<b>11</b>
	<b>Apêndices</b>	<b>12</b>
	<b>APÊNDICE A Diferença entre Anexo e Apêndice</b>	<b>13</b>
	<b>APÊNDICE B Formatação</b>	<b>14</b>
	<b>Anexos</b>	<b>15</b>
	<b>ANEXO A Exemplo</b>	<b>16</b>

# 1 Introdução

Constantemente é necessário que representemos endereços físicos de forma computacional. Para que isso seja feito, é necessário traduzir o endereço, que geralmente é representado por uma sequência de palavras, para coordenadas de um plano. A forma mais comum dessas coordenadas é a representação por latitude e longitude. Esse processo de tradução do endereço se chama Geocodificação. E como todos os processos, ele possui erro. O presente trabalho pretende avaliar a qualidade da geocodificação de diferentes tecnologias para duas cidades grandes do Brasil, usando uma base de referência e analisando o erro, a discrepância e a acurácia de cada uma das tecnologias.

## 1.1 Justificativa

A representação dos endereços é utilizada em diversas aplicações, desde as mais óbvias, como softwares de roteamento e pesquisa de endereços. Até aplicações mais complexas como a indicação de produtos por meio de inteligência artificial, que por vezes utiliza de informações geográficas. Sendo assim, é de extrema importância que os softwares que realizam a tarefa de traduzir endereços façam isso com a melhor qualidade possível. É importante também

## 1.2 Objetivos

Descrever o objetivo principal e os objetivos específicos da pesquisa. Os objetivos constituem a finalidade de um trabalho científico. O objetivo principal é mais amplo e deve descrever de forma clara e sucinta qual a meta que se deseja atingir (uma proposta que solucione um problema; uma proposta de melhoria; uma análise de uma situação, entre outros exemplos).

Para cumprir o objetivo principal muitas vezes é necessário delimitá-lo em partes menores, ou seja, em objetivos específicos. Os objetivos específicos devem ser partes do trabalho, que após cumpridos pelo pesquisador, atinge-se o objetivo principal do trabalho. Devem vir listados, apresentados por tópicos e separados por ponto-e-vírgula.

## 1.3 Organização do Trabalho

Um parágrafo fazendo uma descrição dos capítulos restantes do documento.

### 1.3.1 Estrutura da Monografia

Segue uma **sugestão** para a estrutura da monografia:



**Capítulo 1:** Introdução.

**Capítulo 2:** Revisão Bibliográfica/ Embasamento Teórico (com o referencial teórico e trabalhos relacionados).

**Capítulo 3:** Metodologia ou Desenvolvimento (material e métodos).

**Capítulo 4:** Resultados e Discussões.

**Capítulo 5.1:** Conclusão (e trabalhos futuros).

## 2 Revisão Bibliográfica

Este capítulo deve apresentar uma contextualização da sua pesquisa com um resumo das discussões já feitas por outros autores sobre o assunto abordado e os conceitos principais relativos ao tema. O nome deste capítulo de **Revisão Bibliográfica** ou **Embasamento Teórico** deve ser acordado com seu orientador.

A revisão bibliográfica é a base que sustenta qualquer pesquisa científica e é indispensável para a delimitação do problema em um projeto de pesquisa, para obter uma ideia precisa sobre o estado atual dos conhecimentos sobre um tema, sobre suas lacunas e sobre a contribuição da investigação para o desenvolvimento do conhecimento (MARCONI; LAKATOS, 2003).

Para a escrita deste capítulo, as citações e referências devem estar de acordo com a norma (ABNT, 2002), que destina-se a orientar a preparação e compilação das referências bibliográficas de todo o documento.

### 2.1 Trabalhos Relacionados

Descreva os principais trabalhos realizados por outros autores sobre a temática escolhida para ser desenvolvida, apresentando os conceitos mais importantes, justificativas e características sobre o tema, do ponto de vista da análise feita pelos autores.

É importante destacar, no contexto da pesquisa, quais os resultados já alcançados e os respectivos responsáveis e se possível uma análise dos trabalhos consultados. Finalize a seção comparando a sua proposta de pesquisa com os trabalhos citados, destacando as semelhanças (caso existam) e a sua contribuição (o que pretende desenvolver).

### 2.2 Fundamentação Teórica

Nesta seção, devem ser fornecidas uma fundamentação teórica e explicações prévias acerca dos conceitos discutidos ao longo do texto. É importante situar o leitor (considerando que este seja leigo no assunto) sobre os termos usados com o intuito de prepará-lo para a leitura do capítulo de metodologia.

### 2.3 Instruções

Este documento está em conformidade com normas vigentes da ABNT. Foi usado a classe abntex2, cuja documentação que pode ser obtida no *link* <<http://www.abntex.net.br/>> ou <<https://www.ctan.org/pkg/abntex2>> . Para melhor orientação do funcionamento da classe e

ambientes, bem como possíveis alterações, recomenda-se a consulta dos manuais do abntex2 disponíveis para *downloads* nos *links* em destaque.

## 3 Desenvolvimento

Para avaliar a qualidade das APIs de geocodificação utilizadas no TerraLAB duas bases de dados padrão ouro foram usadas como referência. Chamaremos essas bases de Bases Gold. Com as bases, foi obtida a medida de erro e realizadas métricas diversas utilizando essa medida.

### 3.1 Bases de Dados

Foram coletadas duas bases de dados distintas para o presente trabalho.

A primeira base coletada foi a base do [Centro de Estudos da Metrópole \(CEM\)](#). A base consiste 12.500 endereços de escolas públicas e particulares do ensino básico da região metropolitana de São Paulo. Essa base foi coletada de forma manual pelo CEM utilizando o GPS para a coleta das coordenadas. Além de informações sobre o endereço, a base também conta com informações diversas sobre as escolas, permitindo com que se façam avaliações diversas em relação a esses dados. O CEM também disponibilizou um [mapa de cluster](#), com todas as escolas, permitindo uma melhor visualização da localização de cada uma delas e da densidade das escolas em São Paulo e região.

A segunda base coletada foi a base de dados da [Prodabel](#), empresa de informática e informação da prefeitura de Belo Horizonte. A base de dados foi descoberta por meio da referência 1. É uma base de dados mantida e atualizada mensalmente por 27 empresas públicas e privadas

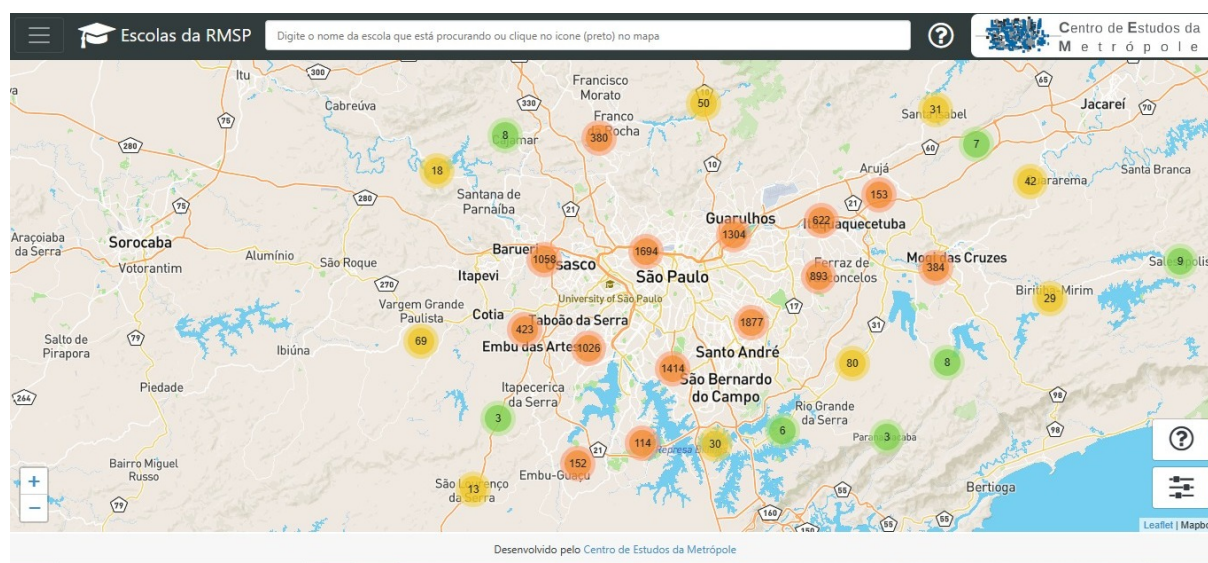


Figura 3.1 – Mapa de clusters que mostra a quantidade de escolas em cada região. Ao aproximar o mapa, o usuário consegue ver a localização de cada uma das escolas presentes no banco de dados.

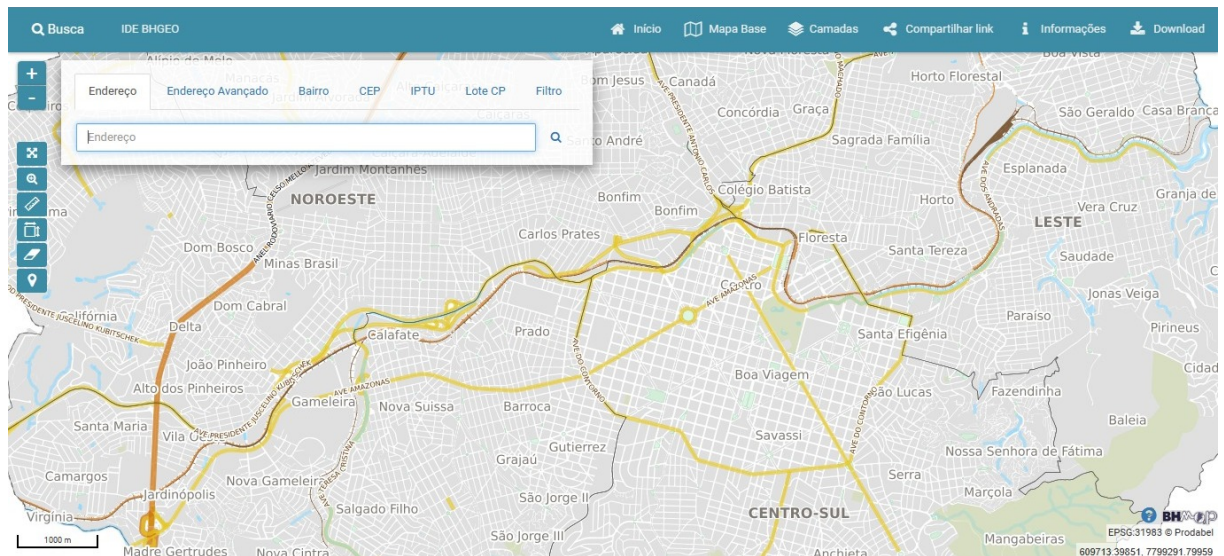


Figura 3.2 – Mapa que mostra a cidade de Belo Horizonte, desenvolvido pela Prodebel. Na barra de pesquisa, é possível pesquisar os endereços e marcá-los no mapa.

de Belo Horizonte. As empresas têm a responsabilidade de reportar qualquer inconsistência que encontrarem, bem como fornecer novos dados a medida que são adquiridos por ela. É uma base considerada confiável pois é constantemente atualizada e é utilizada por diversos serviços da prefeitura. Um exemplo de serviço que utiliza a base de dados é a distribuição dos alunos da rede pública por meio de georeferenciamento. A base conta com 740.000 endereços na data de coleta. A prefeitura também disponibiliza [site com um mapa](#) para visualização do endereços registrados. O endereço está posicionado em cima do edifício representado. Isso pode gerar erro de alguns metros devido a maioria das APIs colocar o endereço na frente do edifício representado.

## 3.2 Erro e métricas

A principal métrica utilizada para avaliar a qualidade da geocodificação é o erro do endereço. Esse erro é calculado como a distância entre o ponto de referência e o ponto geocodificado pela GeoAPI. Com base nesse erro, calcularemos medidas estatísticas, como a média, a mediana, o desvio padrão e a média aparada em 5%, para analisar a precisão das GeoAPIs.

Outra métrica utilizada é a taxa de resposta por API. Para alguns endereços da base de dados, as GeoAPIs podem retornar um erro, não fornecendo uma geocodificação válida. Nesse caso, nada é inserido no banco de dados. A taxa de resposta é calculada como a quantidade de endereços geocodificados dividida pela quantidade de endereços originais na base de dados. Esse valor, normalmente entre 0 e 1, é convertido em uma porcentagem para facilitar a compreensão dos resultados.

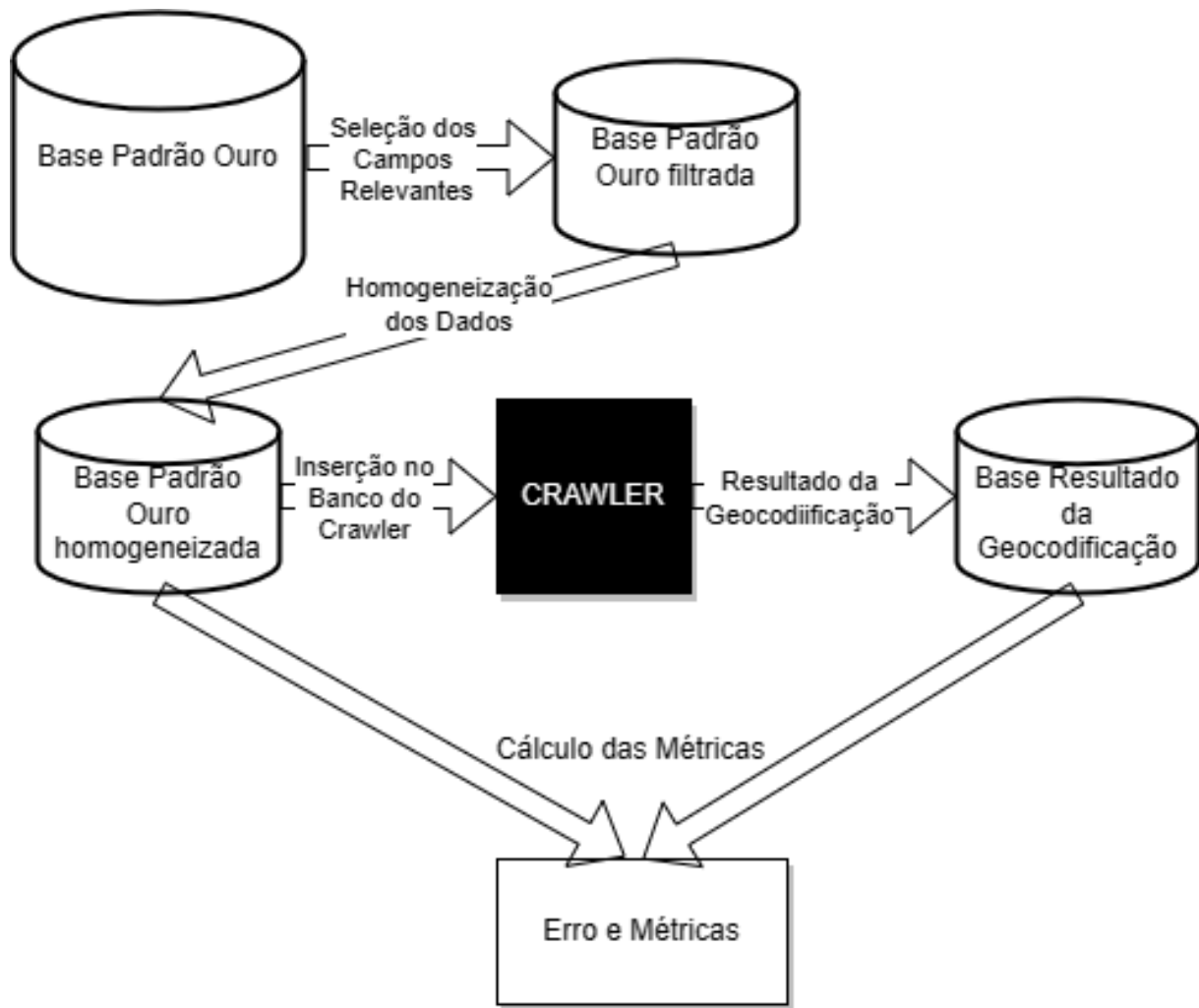


Figura 3.3 – Esquematização do processo de preparação e geocodificação dos dados

### 3.3 Preparação dos Dados e Geocodificação

A preparação de dados e geocodificação desempenham um papel crucial em muitos estudos e projetos que envolvem informações geográficas. Nesta pesquisa, esses processos desempenham um papel fundamental na obtenção de dados consistentes e na atribuição de coordenadas geográficas aos endereços. A etapa de preparação de dados envolve a seleção dos campos relevantes da base de dados, como o nome da rua, número, bairro, CEP e cidade. Além disso, é realizada uma homogeneização dos dados, onde abreviações comumente utilizadas são substituídas por suas formas completas correspondentes. Essa etapa é essencial para garantir resultados mais precisos na geocodificação. A geocodificação, por sua vez, consiste em atribuir coordenadas geográficas (latitude e longitude) a cada endereço presente na base de dados. Utilizando ferramentas adequadas, o processo de geocodificação é realizado, possibilitando a localização precisa de cada endereço no espaço geográfico. Para realizar a geocodificação, os endereços previamente preparados são inseridos no banco de dados do Crawler, onde as ferramentas de geocodificação estão disponíveis. Essas ferramentas utilizam algoritmos e informações geográficas para identificar

e atribuir as coordenadas geográficas correspondentes a cada endereço. É importante ressaltar que o processo de geocodificação é realizado pela equipe de Back-end do TerraLAB, portanto, vemos esse processo como uma caixa preta. Uma vez concluída a geocodificação, os endereços geocodificados, juntamente com suas coordenadas geográficas, são armazenados no banco de dados. Esses dados geocodificados podem ser utilizados para análises espaciais, mapeamento e visualização de informações geográficas, contribuindo para a compreensão de padrões e tendências em determinada área de estudo. Portanto, a preparação de dados e geocodificação são etapas essenciais para garantir a qualidade e a utilidade das informações geográficas utilizadas neste estudo. Esses processos permitem a obtenção de dados consistentes e georreferenciados, facilitando a análise e interpretação dos resultados obtidos

## 4 Resultados

Neste capítulo são apresentados, interpretados e analisados todos os resultados alcançados no trabalho. A análise deve ser realizada de forma que fique claro que os objetivos específicos foram atendidos. Se possível, faça uma comparação com os resultados da literatura, destacando a importância da pesquisa realizada no contexto acadêmico.



## **5 Considerações Finais**

Neste capítulo deve ser explicitado se todos os objetivos descritos na introdução foram atingidos e ressaltar a contribuição do trabalho para o meio acadêmico.

São apresentados de forma sucinta os resultados obtidos e um fechamento de todo trabalho desenvolvido.

### **5.1 Conclusão**

Em resumo, nesta seção devem ser apresentadas as considerações finais do trabalho. Faça uma recapitulação a respeito de cada um dos objetivos específicos, sintetize os resultados obtidos e conclua se o objetivo principal do trabalho foi alcançado.

### **5.2 Trabalhos Futuros**

Apresente propostas de continuidade do seu trabalho.

### **5.3 Publicações Realizadas**

Caso o trabalho tenha originado publicações é válido acrescentar essa informação, visto que pode creditar ainda mais o estudo. Assim, elas devem ser apresentadas na forma de uma subseção do capítulo conclusão. Por exemplo:

Os trabalhos seguintes, que foram originados das metodologias propostas, foram aceitos para apresentação em conferências nacionais:

1. Autor. Título do Artigo. Cidade: Conferência, Ano.

# Referências

ABNT. *NBR 6023: Informação e documentação - referências - elaboração*. Rio de Janeiro, 2002. 24 p.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. *ABNT NBR 6028:2003: Informação e documentação – resumo – apresentação*. Rio de Janeiro, 2003. 2 p.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. *ABNT NBR 14724:2011: Informação e documentação – trabalhos acadêmicos – apresentação*. Rio de Janeiro, 2011. 15 p.

CARVALHO, M. de. *Construindo o saber: técnicas de metodologia científica*. [S.l.]: Papirus Editora, 1989. ISBN 9788530800710.

MARCONI, M. d. A.; LAKATOS, E. M. *Fundamentos de Metodologia Científica*. 5. ed.. ed. São Paulo: Atlas, 2003.

RAMPAZZO, L. *Metodologia científica*. [S.l.]: Edições Loyola, 2005. ISBN 9788515024988.

# **Apêndices**

# **APÊNDICE A – Diferença entre Anexo e Apêndice**

Os apêndices “São textos ou documentos elaborados pelo autor, a fim de complementarem sua argumentação, sem prejuízo da unidade nuclear do trabalho” (ABNT, 2011). Podem ser incluídos nos apêndices: os questionários da pesquisas, as tabulação de dados, ilustrações e outros documentos que necessariamente foram preparados pelo autor. Já os anexos, em conformidade com a norma (ABNT, 2011) “são textos ou documentos não elaborados pelo autor, que servem de fundamentação, comprovação ou ilustração à parte do trabalho”, como por exemplo leis, ilustrações, demonstrações de formulas, tabulações de dados de trabalhos referenciados, etc.

## **APÊNDICE B – Formatação**

Os apêndices devem ser identificados por letras maiúsculas consecutivas (APÊNDICE A, APÊNDICE B, etc), travessão e os respectivos títulos, devendo estar centralizados na folha.

# **Anexos**

# **ANEXO A – Exemplo**

Utiliza a mesma formatação dos apêndices.