UNIVERSIDADE FEDERAL FLUMINENSE

NOME(S) DO(S) AUTOR(ES)

* ESTA PÁGINA É Capa Dura a ser impressa na gráfica.
* **Com hifenização**.

(remova as caixas amarelas)

TÍTULO DO TRABALHO

Niterói

ANO

NOME(S) DO(S) AUTOR(ES)

TÍTULO DO TRABALHO

Trabalho de Conclusão de Curso submetido ao Curso de Tecnologia em Sistemas de Computação da Universidade Federal Fluminense como requisito parcial para obtenção do título de Tecnólogo em Sistemas de Computação.

Orientador(a):

Nome

NITERÓI

ANO

Folha reservada para a ficha catalográfica

NOME(S) DO(S) AUTOR(ES)

TÍTULO DO TRABALHO

SUBTÍTULO

Trabalho de Conclusão de Curso submetido ao Curso de Tecnologia em Sistemas de Computação da Universidade Federal Fluminense como requisito parcial para obtenção do título de Tecnólogo em Sistemas de Computação.

Niterói, \_\_\_ de \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ de ANO.

Banca Examinadora (provisório):

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Prof. ou Profa. <NOME>, <Título>. – Orientador ou Avaliador

<Sigla da Universidade> - <Nome da Universidade>

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Prof. ou Profa. <NOME>, <Título>. – Orientador ou Avaliador

<Sigla da Universidade> - <Nome da Universidade>

A dedicatória é opcional, no qual o autor presta homenagem ou dedica seu trabalho a alguém.

Dedico este trabalho a(o) minha(meu) esposa(o) e aos meus estimados filhos.

AGRADECIMENTOS

(EXEMPLO) A Deus, que sempre iluminou a minha caminhada.

A meu Orientador Fulano de Tal pelo estímulo e atenção que me concedeu durante o curso.

Aos Colegas de curso pelo incentivo e troca de experiências.

A todos os meus familiares e amigos pelo apoio e colaboração.

Epígrafe é a inscrição colocada no início de um trabalho, de um capítulo ou partes principais. É opcional para a monografia.

Delete as caixas amarelas

“A Escola é uma arena onde grupos sociais lutam por legitimidade e poder”.

Dinair Leal da Hora

RESUMO

O resumo deve apresentar os pontos relevantes de um texto, no mesmo momento ele terá que dar uma visão rápida e clara dos conteúdos, metodologias e das conclusões do trabalho. As frases têm que ter uma sequência lógica e objetiva. A fonte do estilo é Arial e tamanho 12.

Palavras-chaves: palavra1, palavra2 e palavra3.

*O resumo e abstract* devem ser digitados sem parágrafo inicial e o espaçamento entre linhas é simples. Lembre-se que já existem estilos que formatam o texto adequadamente. Remova as caixas amarelas antes de imprimir.

ABSTRACT (opcional)

The summary must present the excellent points of a text, at the same moment it will have that to give a fast and clear vision of the contents, methodologies and of the conclusions of the work. The phrases have that to have a logical and objective sequence. The source of the style is so great Arial and 12.

Key words: word1, word2 and word3.

LISTA DE ILUSTRAÇÕES

[Figura 1: Exemplo de Figura. 19](#_Toc378694305)

LISTA DE TABELAS

[Tabela 1: Exemplo de Tabela. 19](#_Toc378694310)

LISTA DE GRÁFICOS

[Gráfico 1: Exemplo de um gráfico 18](#_Toc378694318)

LISTA DE ABREVIATURAS E SIGLAS

API – *Application Programming Interface*

ADR – *Adverse drug reaction*

FDA – *Food and Drug Administration (U.S. Federal agency)*

HPSG – *Head-driven phrase structure grammar*

MD5 – *Message-Digest algorithm 5*

ICMS – *Institute of Chinese Medical Sciences*

SUMÁRIO

[RESUMO 7](#_Toc378694359)

[ABSTRACT (opcional) 8](#_Toc378694360)

[LISTA DE ILUSTRAÇÕES 9](#_Toc378694361)

[LISTA DE TABELAS 10](#_Toc378694363)

[LISTA DE GRÁFICOS 11](#_Toc378694364)

[LISTA DE ABREVIATURAS E SIGLAS 12](#_Toc378694365)

[1 INTRODUÇÃO 15](#_Toc378694366)

[2 DESENVOLVIMENTO DO TRABALHO 16](#_Toc378694367)

[2.1 TEXTO DO TRABALHO 16](#_Toc378694368)

[2.1.1 REFERÊNCIAS BIBLIOGRÁFICAS 17](#_Toc378694369)

[2.1.2 CITAÇÕES 17](#_Toc378694370)

[2.1.3 IDIOMA ESTRANGEIRO 18](#_Toc378694371)

[2.1.4 FIGURAS, GRÁFICOS E TABELAS. 18](#_Toc378694372)

[2.1.5 NOTAS ENTRE O ORIENTADOR E O ALUNO 19](#_Toc378694373)

[CONCLUSÕES E TRABALHOS FUTUROS 20](#_Toc378694374)

[REFERÊNCIAS BIBLIOGRÁFICAS 21](#_Toc378694375)

[ANEXOS 22](#_Toc378694376)

# INTRODUÇÃO

Parte inicial do texto, onde devem constar a delimitação do assunto tratado, o problema ou oportunidade, objetivos da pesquisa, motivação para fazer a pesquisa, métodos usados e resultados encontrados, quando for o caso, e outros elementos necessários para situar o tema do trabalho, bem como a estrutura do documento.

A introdução pode ter a seguinte ordem:

* O que é o trabalho? (1 ou 2 parágrafos)
* Motivação para desenvolvê-lo? (1 parágrafo)
* Métodos usados, quando for o caso (linguagens, protocolos, metodologias, etc.) ( 1 ou 2 parágrafos)
* Organização do trabalho:
  + no Capítulo 2....;
  + o terceiro capítulo é dedicado ...;
  + ...; e
  + finalmente, no Capítulo “tal”, temos as conclusões e indicações para futuros trabalhos.

# TRABALHOS RELACIONADOS

Analisamos publicações acadêmicas preferenciando àquelas relacionadas ao nosso domínio de interesse e listamos neste tópico o conteúdo que consideramos mais relevante a nossa pesquisa (é importante observar que não encontramos muitos artigos no campo farmacêutico, sendo a maior parte do conteúdo disponível na internet voltado para a generalização da atividade de *web scraping* e suas metodologias ou para domínios próximos como a biomedicina).

Glez-Peña et al. [1] descrevem a dependência de tecnologias, ferramentas e dados *online* da pesquisa biomédica em função do crescimento do volume de informação e da necessidade de otimizar os benefícios resultantes de novas pesquisas através da extração de conteúdos de interesse em um tempo razoavelmente baixo e sem consumir muitos recursos, identificando os elementos necessários para a aplicação das técnicas de *web scraping*, comparando diferentes *softwares* e *frameworks* disponíveis e voltados para este fim, e propondo um escopo para aplicações no domínio biomédico tendo em vista o extenso uso de APIs para disponibilizar informações nesta área. Dentre as diversas tecnologias consideradas, os autores citam a biblioteca jsoup [2] para a linguagem de programação Java [3] e os *frameworks* Scrapy [4] e Beautiful Soup [5], ambos para a linguagem Python [6].

Feldman et al. [7] apresentam uma metodologia para extração de dados relacionados a reações adversas a medicamentos (ADRs) de fóruns médicos da internet e demonstram a habilidade do método em prever reações não listadas em testes clínicos antes que elas sejam reportadas a agência federal Estadunidense *Food and Drug Administration* (FDA), responsável por assuntos relacionados a saúde pública no país. A metodologia se baseia na combinação do conceito *head-driven phrase structure grammar* (HPSG) com padrões relacionais conhecidos do domínio de interesse aplicados a grandes quantidades de textos avulsos pós-processando a informação obtida com algoritmos automatizados para refinar o resultado obtido. A implementação da solução proposta utiliza o *framework* Scrapy [4], a biblioteca (nativa) MD5 da linguagem Python [6], a plataforma NLTK [8], também para a linguagem Python, e o *framework* URE, sobre o qual não encontramos informações a respeito, mas que é descrito como uma ferramenta de aproximação entre o HPSG e o processo *Automatic acquisition of lexicon*.

Liu, Liang e Wishart [9] oferecem uma segunda versão para um sistema *online* chamado PolySearch [10] que é apresentado como uma ferramenta de *text-mining* significantemente melhorada para a descoberta de associações entre doenças humanas, genes, medicamentos, toxinas e muitas outras entidades biomédicas através de consultas (*queries*) generalizadas do tipo “dada uma variável X, encontre todas as ocorrências associadas da variável Y” onde X e Y podem assumir qualquer uma das entidades do domínio suportadas pelo sistema, que foi desenvolvido utilizando Python e uma tecnologia de motor de busca chamada ElasticSearch [11]. A partir dos parâmetros informados, uma ampla busca na internet é realizada em múltiplas e variadas fontes como artigos acadêmicos e bancos de dados amplamente utilizados pelas comunidades biomédica e farmacêutica como DrugBank [12], UniProt [13] e Human Metabolome Database [14].

Amalia, Afifa e Herriyance [15] abordam o hábito dos Indonésios de recorrer, como uma medida de primeiros socorros, a motores tradicionais de busca para encontrar informações sobre doenças e possíveis medicamentos ao passo que estes são por vezes incapazes ou ineficientes ao lidar com sinônimos populares para nomes científicos, além de retornar artigos não relacionados com o assunto desejado e propõem, como solução, a construção de um motor de busca baseado em *semantic web* [16] capaz de extrair vocabulário de *websites* populares que sabidamente ou potencialmente tenham o conteúdo desejado, criando um relacionamento entre termos científicos e populares, incluindo sinônimos para as doenças mais pesquisadas (o artigo foca em doenças tropicais, embora o contexto possa ser expandido). A solução utiliza técnicas de *web scraping* para obter dados de fontes como o banco de dados Drugs.com [17], o portal AloDokter [18] e o *website* da agência nacional de controle de alimentos e medicamentos da Indonésia [19] gerando, como resultado, serializações do modelo RDF [20] fazendo uso de *frameworks* do tipo CSV2RDF [21] para a linguagem Python.

Gräßer et al. [22] buscam conhecer a preferência e a experiência de usuários do domínio farmacêutico através da análise automatizada por *web scraping* de fóruns e *websites* que contenham opiniões e *reviews* destes usuários: primeiro uma análise de sentimentos [22, p.121] é feita e uma previsão é gerada quanto a satisfação geral, efeitos colaterais e efetividade de um determinado medicamento. Em seguida, é investigada a compatibilidade dos modelos de classificação obtidos através da sua transferência entre diferentes domínios e fontes de dados, mostrando que esta abordagem pode ser utilizada para apontar, explorar e comparar similaridades e análises de sentimentos entre esses domínios. Os dados são obtidos de duas fontes independentes distintas e amplamente utilizadas no campo farmacêutico por profissionais e consumidores – Drugs.com [17] e DrugLib.com [23] – através de um web *crawler* implementado na linguagem Python utilizando o *framework* Beautiful Soup [5].

## DISCUSSÃO

Han et al. [24] trazem um estudo sobre a predição da desintegração da fórmula de medicamentos comprimidos com administração via oral utilizando técnicas de redes neurais, destacando a importância de dois fatores para o sucesso das predições: os dados avaliados e os algoritmos de processamento utilizados, percebendo a extração de dados como uma etapa crítica do processo e citando o uso da tradicional extração manual de dados em pesquisas atuais por cientistas farmacêuticos com experiência na área:

“*In order to ensure a satisfied prediction accuracy, two key factors are to be considered: data and algorithm. The first issue is the reliable data in pharmaceutical research. Deep-learning attempts to learn these characteristics to make better representations and create models from reliable data. Thus, data extraction is a critical step. In current research, reliable formulation datasets were manually extracted and labeled from the research articles of Web of Science by experienced pharmaceutical scientists*.” [24, p. 339]

Entendemos, com isso, que há espaço para soluções mais viáveis de *data-mining* no domínio farmacêutico através uma metodologia simples, fundamentada na determinação de fontes confiáveis para os dados de interesse associada a aplicação das técnicas de *web scraping* para extrair, com velocidade e confiabilidade, as informações desejadas dos dados obtidos, liberando assim os recursos humanos para outras atividades da pesquisa.

# DESENVOLVIMENTO DO TRABALHO

Neste capítulo colocamos algumas informações relevantes para a elaboração do TCC.

O processo de desenvolvimento do TCC é interativo, e ocorre através de sucessivos *e-mails*. O aluno envia versões intermediárias para seu orientador, que o direciona para a próxima versão. Desta forma, o trabalho vai sendo elaborado e quando o orientador decide que o trabalho está pronto para ser avaliado pela Banca, este o encaminha para avaliação. A avaliação da banca pode resultar em um conjunto de sugestões para o aprimoramento do trabalho, cabendo ao orientador encaminhar ao seu orientando as modificações a serem efetuadas. Ao final desse processo o trabalho estará pronto para ser entregue em sua versão em capa dura, que é um requisito obrigatório para a aprovação no TCC.

## TEXTO DO TRABALHO

O presente documento foi digitado no estilo e formatos necessários para a elaboração do TCC, portanto guardem uma cópia, para salvar as informações nele contidas, e o utilize como base para desenvolver o TCC. Para digitar, selecione o texto de modelo e troque pela sua redação. O parágrafo está justificado com a primeira linha recuada em 02 centímetros, entrelinha em 1,5.

Na sequência temos uma lista das principais recomendações durante a escrita.

### REFERÊNCIAS BIBLIOGRÁFICAS

Na seção de “**REFERÊNCIAS BIBLIOGRÁFICAS**” temos dois exemplos para o formato das referências. Elas devem ser utilizadas no texto com o formato [1] para a primeira referência da lista e assim por diante, em ordem. Se utilizar um *site* como referência use o modelo em [2]. Quando se referir a uma página específica, como no caso de uma citação use [1, p.32]. Lembre-se de que definições, conceitos, estatísticas, gráficos, ou seja, qualquer afirmação ou dado contido no texto que não for seu deve exibir uma referência. Outra questão é que devemos ter cuidado na seleção das referências, por exemplo, o Wiki não é uma referência confiável, pois existem muitos erros. Uma boa dica é usar a referência automática que já coloca as citações na ordem correta e faz a referência (no final do documento) de forma automática. Com isso, caso alguma referência seja acrescentada depois, você não precisa rever toda a lista citada. Para isso vá no menu referência e clique em inserir citação [1].

### CITAÇÕES

Citação: É a menção do texto de informação extraída de outra fonte para esclarecer, ilustrar ou sustentar o assunto apresentado. Podemos classificá-las em Curta e Longa. “Curta: É transcrita entre aspas, com o mesmo tipo e tamanho da letra utilizados no parágrafo do texto no qual será inserido. O uso das aspas delimita a citação direta”. [1, p.154].

(Exemplo de Citação Longa) É transcrita em parágrafo distinto. Inicia na margem de parágrafo, sem deslocamento na primeira linha e termina na margem direita. Longa: É transcrita em parágrafo distinto. Inicia na margem de parágrafo, sem deslocamento na primeira linha e termina na margem direita. Longa: É transcrita em parágrafo distinto. Inicia na margem de parágrafo, sem deslocamento na primeira linha e termina na margem direita [1, p. 155].

### IDIOMA ESTRANGEIRO

Todos os termos que não pertençam à língua portuguesa devem ser destacados em *itálico*. Os termos não usuais devem ser definidos ou no texto ou em notas de rodapé.

### FIGURAS, GRÁFICOS E TABELAS.

As figuras, gráficos e tabelas devem ser referenciadas no texto, por exemplo: no Gráfico 1 apresentamos um exemplo. Os índices já foram criados neste texto, para colocar novos objetos pressione o botão direito do *mouse* sobre o objeto, selecione “legenda” e digite sua descrição para o objeto. Após este procedimento basta atualizar o índice que ele será incluído.



Gráfico 1: Exemplo de um gráfico



Figura 1: Exemplo de Figura.

Tabela 1: Exemplo de Tabela.

|  |  |
| --- | --- |
| Tabela exemplo | Descrição |
| 1 | Campo 1 |
| 2 | Campo 2 |

### NOTAS ENTRE O ORIENTADOR E O ALUNO

Nossa interação é feita através de e-mails, mas essa forma de comunicação pode ser prejudicial se escrevermos um texto muito longo, portanto devemos ser objetivos. As observações diretamente no texto do TCC têm se demonstrado mais produtivas, a forma e estratégia são combinadas entre o tutor orientador e o orientando.

CONCLUSÕES E TRABALHOS FUTUROS

Neste capítulo devem ser colocadas as conclusões que o aluno obteve durante a elaboração do trabalho, bem como o que pretende após sua conclusão (especialização, mestrado, aplicar os conhecimentos em alguma área...).

REFERÊNCIAS BIBLIOGRÁFICAS

1. Daniel Glez-Peña, Anália Lourenço, Hugo López-Fernández, Miguel Reboiro-Jato, Florentino Fdez-Riverola. **Web scraping technologies in an API world**, 2014. Briefings in Bioinformatics, Volume 15, Issue 5, September 2014, Pages 788–797.
2. HEDLEY, Jonathan. **Biblioteca Java para trabalhar com HTML.** <[https://jsoup.org](https://jsoup.org/)> Acesso em 06 set. 2019.
3. ORACLE. **Linguagem de programação**. <[https://java.com](https://www.java.com/)> Acesso em 06 set. 2019.
4. SCRAPINGHUB; et al. ***Framework* Python para extração de dados de *websites***. <[https://scrapy.org](https://scrapy.org/)> Acesso em 06 set. 2019.
5. RICHARDSON, Leonard. **Biblioteca Python para “*scraping* de tela”**. <[https://crummy.com/software/BeautifulSoup](https://www.crummy.com/software/BeautifulSoup/)> Acesso em 06 set. 2019.
6. PYTHON SOFTWARE FOUNDATION. **Linguagem de programação**. <[https://python.org](https://www.python.org/)> Acesso em 06 set. 2019.
7. Ronen Feldman , Oded Netzer , Aviv Peretz , Binyamin Rosenfeld, Utilizing Text Mining on Online Medical Forums to Predict Label Change due to Adverse Drug Reactions, Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 10-13, 2015, Sydney, NSW, Australia.
8. NLTK PROJECT. **Plataforma para criar programas Python para trabalhar com dados de linguagem humana**. <https://www.nltk.org> Acesso em 09 set. 2019.
9. Yifeng Liu, Yongjie Liang, David Wishart, PolySearch2: a significantly improved text-mining system for discovering associations between human diseases, genes, drugs, metabolites, toxins and more, Nucleic Acids Research, Volume 43, Issue W1, 1 July 2015, Pages W535–W542.
10. CANADIAN INSTITUTES OF HEALTH RESEARCH; et al. **Ferramenta *online* para *text-mining***. <http://polysearch.ca> Acesso em 09 set. 2019.
11. ELASTICSEARCH B. V.. **Motor de busca**. <[https://elastic.co](https://www.elastic.co/)> Acesso em 09 set. 2019.
12. CANADIAN INSTITUTES OF HEALTH RESEARCH; et al. **Banco de dados de medicamentos e substâncias bioquímicas**. <[https://drugbank.ca](https://www.drugbank.ca/)> Acesso em 09 set. 2019.
13. NATIONAL INSTITUTES OF HEALTH, U.S. Department of Health and Human Services; et al. **Banco de dados de informações funcionais e sequências de proteínas**. <[https://uniprot.org](https://uniprot.org/)> Acesso em 09 set. 2019.
14. CANADIAN INSTITUTES OF HEALTH RESEARCH; et al. **Banco de dados de informações sobre o metabolismo de pequenas moléculas encontradas no corpo humano**. <[http://hmdb.ca](http://www.hmdb.ca/)> Acesso em 09 set. 2019.
15. A. Amalia, R. M. Afifa and H. Herriyance, "Resource Description Framework Generation for Tropical Disease Using Web Scraping," 2018 IEEE International Conference on Communication, Networks and Satellite (Comnetsat), Medan, Indonesia, 2018, pp. 44-48.
16. WORLD WIDE WEB CONSORTIUM, W3C. **Padrões da “*web* semântica”**. <[https://w3.org/standards/semanticweb](https://www.w3.org/standards/semanticweb/)> Acesso em 09 set. 2019.
17. DRUGS.COM. **Banco de dados independente com informações sobre medicamentos**. <[https://drugs.com](https://www.drugs.com/)> Acesso em 09 set. 2019.
18. ALODOKTER. **Portal *web* de informações médicas e sobre saúde**. <https://alodokter.com> Acesso em 09 set. 2019.
19. AGÊNCIA NACIONAL DE CONTROLE DE ALIMENTOS E MEDICAMENTOS DA INDONÉSIA. ***Website* oficial da instituição**. <http://pom.go.id> Acesso em 09 set. 2019.
20. WORLD WIDE WEB CONSORTIUM, W3C. **Padrões do modelo *Resource Description Framework***. <[https://w3.org/RDF/](https://www.w3.org/RDF/)> Acesso em 09 set. 2019.
21. WORLD WIDE WEB CONSORTIUM, W3C. **Procedimentos e regras para conversão de dados tabulares em RDF**. <[https://w3.org/TR/csv2rdf/](https://www.w3.org/TR/csv2rdf/)> Acesso em 09 set. 2019.
22. Felix Gräßer, Surya Kallumadi, Hagen Malberg, and Sebastian Zaunseder. 2018. Aspect-Based Sentiment Analysis of Drug Reviews Applying Cross-Domain and Cross-Data Learning. In DH’18:2018 International Digital Health Conference, April 23–26, 2018, Lyon, France. ACM, New York, NY, USA, 5 pages.
23. DRUGLIB.COM. **Banco de dados sobre medicamentos e pesquisas farmacêuticas**. <http://druglib.com> Acesso em 09 set. 2019.
24. HAN, Run. **Predicting oral disintegrating tablet formulations by neural network techniques**, 2018. Artigo Científico (publicado no Asian Journal of Pharmaceutical Sciences 13) – Institute of Chinese Medical Sciences (ICMS), University of Macau, Macau, China.
25. […]
26. […] Os exemplos originais do *template* foram mantidos (abaixo) para referências até o término do trabalho:
27. CARDOSO, Alcionê Damásio. **Vantagens e Desvantagens na Forma de Escolha de Diretor de Escola na Rede Pública Estadual de Santa Catarina**, 2002. Dissertação (Mestrado em Gestão Institucional) – Curso de Pós Graduação em Educação, UnC-UNICAMP, Caçador, SC.
28. SOUZA, Celso de Oliveira. **Histórico da Fundação Educacional Barriga Verde.** <http://www.febave.org.br/historico.htm> Acesso em 25 abr. 2004.

ANEXOS

A função das duas definições, Anexo e Apêndice, é semelhante, mas com uma grande diferença entre elas: a autoria. O **ANEXO** de um trabalho acadêmico deve ser aquele texto ou documento que **não foi elaborado por você**, tendo como objetivo servir de legitimação. Já o **APÊNDICE** se configura como texto ou documento **elaborado por você**, tendo como objetivo complementar a sua argumentação.

ANEXO A – TÍTULO DO ANEXO A

ANEXO B – TÍTULO DO ANEXO B