

Instituto de Ciências Exatas Departamento de Ciência da Computação

Uso de banco de dados orientado a grafos na detecção de fraudes nas cotas para exercício da atividade parlamentar

Gabriel M. Araujo

Monografia apresentada como requisito parcial para conclusão do Bacharelado em Ciência da Computação

Orientadora Prof.a Dr.a Maristela Terto de Holanda

> Brasília 2018



Instituto de Ciências Exatas Departamento de Ciência da Computação

Uso de banco de dados orientado a grafos na detecção de fraudes nas cotas para exercício da atividade parlamentar

Gabriel M. Araujo

Monografia apresentada como requisito parcial para conclusão do Bacharelado em Ciência da Computação

Prof.a Dr.a Maristela Terto de Holanda (Orientadora) CIC/UnB

Prof. Dr. Donald Knuth Dr. Leslie Lamport Stanford University Microsoft Research

Prof. Dr. Rodrigo Bonifácio Coordenador do Bacharelado em Ciência da Computação

Brasília, 7 de setembro de 2018

Dedicatória

Eu dedico esse trabalho aos colegas de curso, que ajudaram a sustentar o sofrimento constante que esse curso proporciona. Dedico também para a galera da central e da CJR, onde aprendi muito mais do que assistindo aulas de cálculo 3.

Agradecimentos

Gostaria de agradecer a todos os meus colegas de curso que me ajudaram nesse período, a todos os professores que se empenharam para ensinar seus alunos e melhorar a cada dia, e principalmente ao Stack Overflow.

Resumo

Este trabalho propõe o uso da tecnologia de banco de dados orientado a grafos para a detecção de fraudes na Cota para o Exercício da Atividade Parlamentar – CEAP. Além do uso do banco de dados orientado a grafos, será feita uma plataforma web para expor informações importantes a população e um estudo sobre o impacto dessas informações na política brasileira. O uso dessas tecnologias facilita muito a manipulação de dados bastante relacionados entre si, tanto em questão de complexidade na consulta, quanto em relação a visualização da informação. A proposta em questão foi validada com um estudo de caso, utilizando os dados abertos da Cota para o Exercício da Atividade Parlamentar da câmara dos deputados. Foi desenvolvido um ETL para extrair os dados e popular o banco, em seguida as consultas foram realizadas para detectar as fraudes e obter informações a respeito dos dados, finalmente foi desenvolvido um sistema web que se comunica via REST com o banco de dados para expor as informações a população de forma mais clara e simples. Para trabalhos futuros, seria interessante o uso de aprendizagem de máquina para obter mais informações valiosas sobre a CEAP.

Palavras-chave: Banco de dados orientado a grafos, fraude, política

Abstract

This work proposes the use of graph oriented databases to detect frauds in Quota for the Exercise of Parliamentary Activity. In addition to the use of the graph oriented databases, a web platform will be developed to expose important information to the population and a study on the impact of this information on Brazilian politics. The use of these technologies, greatly facilitates the manipulation of closely related data, both in terms of query complexity, and information visualization. The proposal in question was validated with a case study, using the open data of the Quota for the Exercise of the Parliamentary Activity of the Chamber of Deputies. It was developed an ETL to extract the data and fill the database, then the queries were made to detect the fraud and to obtain information about the data, finally a web system was developed that communicates via REST with the database to expose the information to the population more clearly and simply. For future work, it would be interesting to use machine learning to obtain more valuable information about CEAP.

Keywords: Graph oriented databases, fraud, politics

Sumário

1	Introdução	1
	1.1 Objetivos	2
	1.2 Metodologia	2
	1.3 Estrutura do Trabalho	3
2	A Classe UnB-CIC	4
	2.1 Gerando o PDF	4
	2.2 Opções	4
	2.3 Informações do Trabalho	5
	2.4 Arquivos	6
	2.5 Documento	6
	2.5.1 Capítulos	6
	2.5.2 Figuras	8
	2.5.3 Equações	10
	2.5.4 Tabelas	11
	2.5.5 Abreviaturas e Siglas	13
3	Trabalho de Conclusão de Curso	15
	3.1 UnB	15
	3.1.1 Os tipos	15
	3.1.2 Teoria e prática	16
	3.1.3 Começo do projeto	16
	3.1.4 Estrutura e regras	17
	3.1.5 O projeto está pronto. E agora?	18
4	Apresentações	19
	4.1 Falando em Público	19
	4.2 BEAMER	21
\mathbf{R}	eferências	22

Apêndice	22
A Fichamento de Artigo Científico	23
Anexo	27
I Documentação Original UnB-CIC (parcial)	27

Lista de Figuras

2.1	Marca P/B	9
2.2	Marca colorida	9
2.3	Outra marca colorida	10

Lista de Tabelas

2.1	Exemplo de tabela	12
2.2	Matriz de Decisão de Eisenhower	13
2.3	Outro exemplo de tabela	13

Capítulo 1

Introdução

O Brasil é reconhecido mundialmente por seus aspectos culturais, por ser membro do grupo de países BRICS e por ser líder da américa latina. Porém, nosso país também é bastante conhecido por ser um país corrupto, e por possuir um cenário político conturbado. Um estudo feito por Abramo [1] compara as relações entre índices de percepção de corrupção e outros indicadores de alguns países latino americanos, e o Brasil se encontra na 49ª (quadragésima nona) posição em um ranking de corrupção dentre 90 países. Já o estudo feito por Filgueiras [2] mostra que de acordo com a percepção dos brasileiros, a Câmara dos vereadores e a Câmara dos deputados são as instituições com maior presença de corrupção.

Atualmente, operações como a Lava Jato, mostram que o país vive um período político muito sensível, que acabam causando vários problemas nos mais diversos setores do país e principalmente na economia. Dessa forma, o objetivo geral deste trabalho é utilizar a tecnologia de banco de dados orientado a grafos, para auxiliar na detecção de possíveis fraudes em um determinado conjunto de dados. Como foi mencionado acima, a câmara dos deputados está entre as instituições com maior presença de corrupção no país, portanto, o conjunto de dados utilizado neste trabalho é a Cota para o Exercício da Atividade Parlamentar – CEAP (antiga verba indenizatória), que é uma cota única mensal destinada a custear os gastos dos deputados exclusivamente vinculados ao exercício da atividade parlamentar.

O Ato da Mesa nº 43 de 2009, detalha as regras para o uso da CEAP, entretanto um deputado pode realizar algumas transações que não são observadas facilmente pelos responsáveis em fiscalizar essas transações. Por exemplo, o artigo 4, parágrafo 13 do Ato da Mesa nº 43 de 2009, diz: Não se admitirá a utilização da Cota para ressarcimento de despesas relativas a bens fornecidos ou serviços prestados por empresa ou entidade da qual o proprietário ou detentor de qualquer participação seja o Deputado ou parente seu até terceiro grau. Dessa forma, o Deputado pode realizar transações que violam essa regra,

sendo inviável verificar as relações de parentesco de cada Deputado em cada transação, especialmente se utilizarem tecnologias inadequadas.

Portanto, justifica-se o uso de um banco de dados orientado a grafo para identificar os relacionamentos envolvendo cada transação de um Deputado. Um banco de dados relacional também consegue resolver esse problema, entretanto, com um custo e complexidade bem maior em relação a um banco de dados orientado a grafo. Isso se deve porque os relacionamentos são evidenciados na estrutura de um grafo de forma muito mais natural e simples, onde cada entidade é representada como um nó do grafo e se relaciona com outras entidades por meio de arestas. Devido a essas particularidades, os bancos de dados em grafo vem ganhando bastante popularidade ultimamente [3], registrando a maior taxa de mudança de popularidade de 2013 até 2017.

1.1 Objetivos

O objetivo geral deste trabalho é utilizar o SGBD OrientDB para evidenciar relacionamentos nas transações dos Deputados que violam o artigo 4, parágrafo 13 do Ato da Mesa nº 43 de 2009, que regula a CEAP. Para essa análise, será feito um sistema web que expõe essas informações a população de forma mais simples e clara. Além de estudar como essas tecnologias impactam a política brasileira.

Os objetivos específicos deste trabalho são:

- Implementar o banco de dados em grafo com os dados da CEAP.
- Executar consultas que evidenciem relacionamentos fraudulentos.
- Identificar vantagens e desvantagens no uso de banco de dados orientado a grafo para a detecção de fraudes em conjuntos de dados genéricos.
- Desenvolver um sistema web para expor informações a respeito dos dados.
- Estudar o impacto dessas tecnologias na política e sociedade brasileira.

1.2 Metodologia

Este trabalho foi dividido em duas partes, a primeira teórica e a segunda prática. Na parte teórica foi realizado um estudo baseado em livros, artigos e páginas da web sobre os assuntos relacionados a banco de dados, SGBD orientado a grafo, NoSQL, as leis que regem a CEAP e os impactos dessas iniciativas na política. Já na parte prática foi desenvolvido um ETL para ler os arquivos que contém os dados das transações e popular o banco, em seguida foram realizadas consultas que buscam evidenciar os relacionamentos

atrelados a cada transação, posteriormente foi desenvolvido um sistema web de forma que forneça as informações claramente a todos. Por fim, foi realizado uma análise dos resultados obtidos e conclusões finais.

1.3 Estrutura do Trabalho

Este trabalho está dividido nos seguintes capítulos:

- Capítulo 2: Introduzo os conceitos relacionados a grafos, necessários para a compreensão de um SGBD orientado a grafo. Forneço uma visão geral de SGBD NoSQL, apontando suas principais características e utilidades. Finalmente explico as características de um SGBD orientado a grafo e sua evolução.
- Capítulo 3: Nesse capítulo apresento os meios utilizados para resolver o problema, e uma explicação mais detalhada de todo o processo de desenvolvimento e como as tecnologias auxiliaram nessa resolução.
- Capítulo 4: Apresento a análise dos resultados obtidos ao realizar as consultas e do sistema desenvolvido para fornecer as informações. Essa análise abrange tanto os resultados das consultas, ou seja, se foi possível identificar uma transação ilícita, quanto o impacto da divulgação desse tipo de informação na política brasileira.
- Capítulo 5: Finalmente, apresento minhas conclusões do trabalho realizado e sugestões para trabalhos futuros relacionados com a área.

Capítulo 2

A Classe UnB-CIC

Este capítulo descreve a classe UnB-CIC, e demonstra os comandos disponíveis. A última versão foi atualizada pelo Prof. Ralha, em 2008 (vide Anexo I). A melhor forma de entender o funcionamento é observar o arquivo principal deste documento (monografia.tex).

2.1 Gerando o PDF

Para gerar corretamente as referências cruzadas, é necessário processar os arquivos mais de uma vez com a seguinte sequência de comandos (supondo que o arquivo principal seja monografia.tex).

pdflatex monografia bibtex monografia makeglossaries monografia pdflatex monografia

O primeiro comando processa os arquivos, indicando quais referências foram citadas no texto (bibliográficas ou cruzadas), o segundo comando processa o arquivo .bib que contém as informações bibliográficas, o terceiro gera o índice de siglas/abreviaturas, e o último comando junta todas estas informações, produzindo um texto com referências cruzadas funcionais.

2.2 Opções

O documento é gerado em função do curso dado como opção [obrigatória] a classe. Os cursos disponíveis são:

bacharelado Bacharelado em Ciência da Computação

licenciatura Licenciatura em Computação

engenharia Engenharia de Computação

mestrado, ppginf Mestrado em Informática

doutorado, ppginf Doutorado em Informática

mestrado, ppca Mestrado Profissional em Computação Aplicada

No caso dos cursos de pós-graduação, há o exame de qualificação do discente, a qual deverá constar a definição, pertinência do projeto, a sua abrangência, comprovação da eficiência e eficácia da metodologia proposta, uma revisão bibliográfica detalhada e o cronograma para conclusão do projeto [4]. Para gerar o documento referente a este exame, use a opção qualificação.

2.3 Informações do Trabalho

O passo seguinte é definir as informações do trabalho, identificando os autores e os membros da banca (atenção a definição do gênero!). Por exemplo, para este documento foram utilizadas as seguintes definições:

\orientador{\prof \dr Guilherme Novaes Ramos}{CIC/UnB}% %\coorientador{\prof \dr José Ralha}{CIC/UnB} \coordenador[a] {\prof[a] \dr[a] Ada Lovelace}{Bibliothèque universelle de Genève}% \diamesano{24}{dezembro}{2014}%

\membrobanca{\prof \dr Donald Knuth}{Stanford University}%
\membrobanca{\dr Leslie Lamport}{Microsoft Research}%

\autor{Guilherme N.}{Ramos}%

Sobre o texto, definiu-se:

\titulo{UnB-CIC: Uma classe em LaTeX para textos do Departamento de Ciência da Computação}%

\palavraschave{LaTeX, metodologia científica}%
\keywords{LaTeX, scientific method}%

O título, apesar do tamanho reduzido, deveria apresentar uma ideia clara de todo o trabalho. As palavras-chave devem indicar os conceitos genéricos mais relevantes utilizados, e servem para indexação e busca de documentos que tratam os mesmos temas.

2.4 Arquivos

Os seguintes arquivos são exigidos:

tex/abstract.tex Contém o abstract do texto.

tex/agradecimentos.tex Contém os agradecimentos do autor.

bibliografia.
bib Contém as referências bibliográficas no formato ${\rm BibT_EX^1}.$

tex/dedicatoria.tex Contém a dedicatória do autor.

tex/siglas.tex Contém as definições de siglas/abreviaturas.

tex/resumo.tex Contém o resumo do texto.

Os alunos dos Programas de Pós-Graduação da Universidade de Brasília devem incluir a ficha catalográfica em seus documentos, gerada pela Biblioteca Central (BCE). Neste caso, o aluno deve substituir o arquivo PDF doc/BDM.pdf pelo fornecido pela BCE. Atenção, para que o arquivo seja incluido automaticamente pela classe, o nome deve ser obrigatoriamente BDM.pdf.

Demais arquivos não são inseridos automaticamente, mas a classe oferece comandos para inclusão, facilitando a organização destes.

2.5 Documento

Todo documento em IATEX é delimitado pelo ambiente document. O caso aqui não é diferente, mas a interação é simplificada. Basicamente, a classe UnB-CIC funciona "automagicamente" em função dos comandos e dos nomes dos arquivos.

2.5.1 Capítulos

O texto de cada capítulo deve estar em seu próprio arquivo, dentro do diretório correto tex. A inclusão do texto é feita pelo comando:

\capitulo{arquivo}{título}%

Os dois argumentos são:

arquivo argumento obrigatório que define o nome do arquivo que contém o texto do capítulo.

¹http://www.bibtex.org

título argumento obrigatório que define o título do capítulo.

Por exemplo, este texto está no arquivo 2_UnB-CIC.tex, e para criar os dois capítulos vistos até agora, o documento seria:

```
\begin{document}%
```

```
\capitulo{1_Introducao}{Introdução}% inclui o arquivo 1_Introducao.tex \capitulo{2_UnB-CIC}{A Classe \unbcic}% inclui o arquivo 2_UnB-CIC.tex \end{document}%
```

Para incluir um terceiro capítulo neste texto, cujo conteúdo trata de trabalhos conclusão de curso, basta criar o arquivo tex/3 TCC.tex e adicioná-lo com o comando descrito.

No caso de apêndices ou anexos necessários, o texto de cada um deve estar em seu próprio arquivo, também dentro do diretório tex/capitulos. Para facilitar as referências cruzadas, estes devem ser inclusos com os seguintes comandos (respectivamente):

```
\apendice{arquivo}{título}%
\anexo{arquivo}{título}%
```

\begin{document}%

Os dois argumentos funcionam exatamente como \capitulo. Desta forma, o exemplo de um documento "completo" seria:

```
\capitulo{1_Introducao}{Introdução}%
\capitulo{2_UnB-CIC}{A Classe \unbcic}%
\capitulo{3_TCC}{Trabalho de Conclusão de Curso}%

\apendice{Apendice_Fichamento}{Fichamento de Artigo Científico}%
\anexo{Anexo1}{Parte da Documentação Original}%
\end{document}%
```

Usando estes comandos, o rótulo de cada capítulo/apêndice/anexo é criado automaticamente a partir do nome do arquivo para posterior referência cruzada. Por exemplo, este capítulo pode ser referenciado com o comando \ref{2_UnB-CIC} (cujo resultado é: 2), mas a classe oferece opções mais interessantes. Os comandos para referenciar çapítulos são:

```
\refCap{referência}%
\refCaps{referência inicial}{referência final}%
```

Onde os argumentos são:

referência nome da referência do capítulo.

referência inicial nome da referência do capítulo inicial da sequência de capítulos.

referência final nome da referência do capítulo final da sequência de capítulos.

O Capítulo 1 é referenciado com o comando:

\refCap{1_Introducao}%

Considerando Capítulo 1 e também o Capítulo 2, é possível referenciar a sequência de Capítulos 1 a 2 com o comando:

\refCaps{1_Introducao}{2_UnB-CIC}%

Embora estes comandos não "simplifiquem" a inclusão de figuras, eles certamente facilitam a referência a elas com um padrão uniforme, e nada impede o uso dos comandos padrões.

2.5.2 Figuras

Para manter a organização dos arquivos de seu documento, as figuras devem ficar separadas no diretório img. As funções de inclusão de figuras permanecem as mesmas, mas a classe UnB-CIC oferece uma forma mais simples de inserir uma figura (e de referenciá-la). Basta executar o comando:

\figura[posição]{arquivo}{legenda}{referência}{tamanho}%

Os 5 argumentos são:

posição argumento [opcional] para posicionar a figura no texto².

arquivo nome do arquivo da imagem.

legenda da figura.

referência nome da referência da figura para referências cruzadas.

tamanho tamanho da imagem³.

Por exemplo, a Figura 2.1, inserida com o seguinte comando:

\figura[!h]{contorno preto}{Marca P/B}{unbPB}{width=0.5\textwidth}%

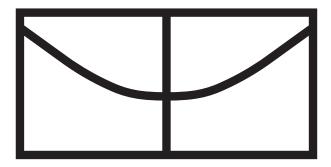


Figura 2.1: Marca P/B.

Os comandos para referenciar figuras são:

\refFig{referência}%
\refFigs{referência inicial}{referência final}%

Onde os argumentos são:

referência nome da referência da figura.

referência inicial nome da referência da figura inicial da sequência de figuras.

referência final nome da referência da figura final da sequência de figuras.

A Figura 2.1 é referenciada com o comando:

\refFig{unbPB}%



Figura 2.2: Marca colorida.

Considerando a Figura 2.2 e também a Figura 2.3, é possível referenciar a sequência de Figuras 2.1 a 2.3 com o comando:

\refFigs{unbPB}{unb2}%

²Mais informações na documentação do ambiente *figure*, mas este é um bom começo: http://en.wikibooks.org/wiki/LaTeX/Floats,_Figures_and_Captions.

³Mais informações na documentação do comando \includegraphics.

Algumas vezes deseja-se usar a figura de uma das referências bibliográficas. Neste caso, utilize o comando:

\figuraBib[posição]{arquivo}{legenda}{bib}{referência}{tamanho}%

Os argumentos são os mesmos do comando \figura, acrescidos de:

bib nome da referência bibliográfica que originou a figura.

Por exemplo, a Figura ?? foi gerada com o comando:

\figuraBib{miktex}{\LaTeX\ vs MS Word} {pinteric latex 2004}{latexvsword}{width=.45\textwidth}%

Embora estes comandos não "simplifiquem" a inclusão de figuras, eles certamente facilitam a referência a elas com um padrão uniforme, e nada impede o uso dos comandos padrões.



Figura 2.3: Outra marca colorida.

2.5.3 Equações

As funções de inclusão de equações permanecem as mesmas, mas a classe UnB-CIC oferece uma forma mais simples de inserir uma equação (e de referenciá-la). Basta executar o comando:

\equacao{referência}{fórmula}%

Os 2 argumentos são:

referência nome da referência da equação para referências cruzadas.

fórmula a equação em si.

Por exemplo, a Equação 2.1, inserida com o seguinte comando:

 $\qquad \qquad \equacao{pitagoras}{a^2 + b^2 = c^2}\%$

$$a^2 + b^2 = c^2 (2.1)$$

Além disso, é possível quebrar em linhas, como na Equação 2.2, com o mesmo comando:

$$a = (x+y)^2$$

$$b = (x*y)^2$$
(2.2)

Os comandos para referenciar equações são:

\refEq{referência}%

\refEqs{referência inicial}{referência final}%

Onde os argumentos são:

referência nome da referência da equação.

referência inicial nome da referência da equação inicial da sequência de equações.

referência final nome da referência da equação final da sequência de equações.

Considerando a Equação 2.1 e também a Equação 2.3, é possível referenciar a sequência de Equações 2.1 a 2.3 com o comando:

\refEqs{pitagoras}{eq}%

Embora estes comandos não "simplifiquem" a inclusão de equações, eles certamente facilitam a referência a elas com um padrão uniforme e nada impede o uso dos comandos padrões.

$$d = c^3 - \frac{a}{h} \tag{2.3}$$

2.5.4 Tabelas

As funções de inclusão de tabelas permanecem as mesmas, mas a classe UnB-CIC oferece uma forma mais simples de inserir uma tabela (e de referenciá-la). Basta executar o comando:

\tabela{legenda}{referência}{especificações}{tabela}%

Os 4 argumentos são:

Tabela 2.1: Exemplo de tabela.

Item	Descrição
1	Descrição 1
2	Descrição 2
3	Descrição 3

legenda da tabela.

referência nome da referência da tabela para referências cruzadas.

especificações alinhamento de cada coluna da tabela.

tabela o conteúdo da tabela⁴.

Por exemplo, a Tabela 2.1, inserida com o seguinte comando:

 $\label{lem:lemplo} $$ \tilde{E}xemplo de tabela}{exemplo}{| c | c |}%$

{\hline

\textbf{Item} & \textbf{Descrição} \\\hline

- 1 & Descrição 1 \\hline
- 2 & Descrição 2 \\hline
- 3 & Descrição 3 \\hline}%

Os comandos para referenciar tabelas são:

\refTab{referência}%

\refTabs{referência inicial}{referência final}%

Onde os argumentos são:

referência nome da referência da tabela.

referência inicial nome da referência da tabela inicial da sequência de tabelas.

referência final nome da referência da tabela final da sequência de tabelas.

Considerando a Tabela 2.1 e também a Tabela 2.3, é possível referenciar a *sequência* de Tabelas 2.1 a 2.3 com o comando:

\refTabs{exemplo}{exemplo2}%

Algumas vezes deseja-se usar a tabela de uma das referências bibliográficas. Neste caso, utilize o comando:

⁴Mais informações na documentação do ambiente tabular.

Tabela 2.2: Matriz de Decisão de Eisenhower (Fonte: [5]).

	Urgente	Não Urgente
Importante	Crises	Planejamentos
Não importante	Interrupções	Distrações

Tabela 2.3: Outro exemplo de tabela.

#	A	В	Comentário
1	a_1	b_1	comentário 1
2	a_2	b_2	comentário 2
3	a_3	b_3	comentário 3

\tabelaBib{legenda}{bib}{referência}{especificações}{tabela}%

Os argumentos são os mesmos do comando \tabela, acrescidos de:

bib nome da referência bibliográfica que originou a tabela.

Por exemplo, a Tabela 2.2⁵ foi gerada com o comando:

Embora estes comandos não "simplifiquem" a inclusão de tabelas, eles certamente facilitam a referência a elas com um padrão uniforme, e nada impede o uso dos comandos padrões.

2.5.5 Abreviaturas e Siglas

Abreviaturas e siglas devem ser definidas no arquivo tex/siglas.tex, e a inserção feita com o comando:

\sigla{sigla}{descrição}%

Onde os argumentos são:

sigla a própria sigla/abreviatura.

 $^{^5 \}rm{Vale}$ a pena assistir o vídeo da palestra $\it{Time~Management}$ de Randy Pausch: http://www.cs.virginia.edu/~robins/Randy/

descrição definição completa do que representa a sigla/abreviatura.

Por exemplo:

\sigla{CIC}{Departamento de Ciência da Computação}%

A inserção de uma sigla/abreviatura no texto é simples, e pode ser feita de três formas diferentes:

\acrshort{CIC} CIC

\acrlong{CIC} Departamento de Ciência da Computação

\acrfull{CIC} Departamento de Ciência da Computação (CIC)

Capítulo 3

Trabalho de Conclusão de Curso

Este capítulo oferece sugestões para produção de um documento descrevendo um Trabalho de Conclusão de curso...

3.1 UnB

A Universidade de Brasília oferece diversas informações em seu sítio¹. O texto existente em 21/11/2014 é reproduzido a seguir:

Os cursos de graduação, especialização e pós-graduação têm como objetivo formar o aluno e prepará-lo para o exercício profissional. Como avaliação do aprendizado, a universidade exige um projeto que mobiliza os estudantes a colaborar com a pesquisa acadêmica. Desde a escolha do tema até a apresentação do trabalho final, o tempo do aluno é ocupado quase integralmente. Para facilitar a vida desses estudantes, o Portal UnB preparou uma série de dicas de professores especialistas no assunto.

3.1.1 Os tipos

A monografia, a dissertação e a tese são, respectivamente, os trabalhos de conclusão de curso de graduação ou especialização, mestrado e doutorado. A grande diferença é a profundidade exigida no projeto, aumentada de acordo com a importância do título de cada nível acadêmico. Mas, em todos os casos, a pesquisa deve abordar o tema selecionado com coerência, consistência e referencial teórico adequado.

Alguns cursos de graduação não exigem monografia, mas um relatório de estágios realizados, como acontece nas licenciaturas. A metodologia de pesquisar e apresentar resultados se mantém, como é exigido em todo projeto final.

¹http://www.unb.br/oportunidades/projeto_final_de_curso

Uma monografia é, genericamente, um relatório de pesquisa sobre o assunto estudado. É específico a um tema pré-definido dentro de uma área de conhecimento e aborda questões e análises de um problema, a construção de uma teoria ou o desenvolvimento de um produto.

Exigida no mestrado, a dissertação cobra do futuro mestre um conhecimento mais profundo. A pesquisa deve ser o resultado em relatório que representa o trabalho experimental ou exposição científica com um tema bem delimitado, e demonstrar o conhecimento de literatura existente sobre o assunto.

A mais densa entre todos os projetos finais, a tese de doutorado exige mais no que diz respeito a teoria e metodologia do tema pesquisado. Deve apresentar contribuições reais para o desenvolvimento específico da especialidade em questão. A base do estudo demanda uma investigação original.

3.1.2 Teoria e prática

Todo projeto de conclusão de curso exige um relatório escrito baseado em teorias, mesmo que o assunto estudado seja algo prático como uma campanha publicitária ou um projeto arquitetônico. Porém, o inverso não se aplica.

As divisões dos tipos de trabalho variam entre cada área de conhecimento. Em suma, o projeto pode ser teórico, prático ou uma união dos dois. Na primeira situação, o aluno pode fazer estudo de caso - pesquisar sobre um fato histórico ou evento importante - ou formular uma teoria - por meio de pesquisa ou reavaliação das semelhantes.

O projeto prático se dedica a criação e construção de um produto, que pode variar de um novo motor a uma composição musical. O curso de graduação costuma oferecer a opção de um trabalho prático aos alunos. No caso dos cursos de mestrado e doutorado, nem todos os departamentos da universidade dispõem de linhas de pesquisa que permitam um projeto que vá além da teoria acadêmica.

A união dos dois gêneros é comum quando o universitário relata a experiência de estágio ou na simulação de um projeto, como a construção de maquetes ou esquemas computacionais. As opções são vastas e o aluno deve explicar como e o que se deve fazer para que o projeto se torne possível.

3.1.3 Começo do projeto

Parece óbvio, mas muitos alunos esquecem a questão principal na hora de escolher o tema: o assunto deve interessar e estimular a pesquisa. Conviver meses com um tema que não agrada torna o trabalho mais complicado. Porém, escolher um bom tema não é abraçar

e desenvolver sobre tudo que ele \acute{e} e engloba. \acute{E} preciso delimitar o assunto de forma específica.

Um trabalho sobre a história do mundo, por exemplo, está fadado a se tornar superficial. Além de extremamente amplo, é grande o volume de informações a ser levantado e estudado. É importante ter foco para desenvolver um projeto coeso e com credibilidade.

Além disso, o estudante necessita desenvolver um problema e traçar uma hipótese. Em um exemplo bem simples: a Guerra no Iraque (tema) e o terrorismo mundial (problema) – o aumento dos ataques depois da invasão americana (hipótese); ou seja, o que o aluno quer tratar e onde ele espera chegar na pesquisa. A não comprovação da hipótese não inviabiliza o trabalho, desde que o desenvolvimento da análise enriqueça os conhecimentos sobre o tema tratado.

A prática essencial para o desenvolvimento de qualquer projeto é a pesquisa bibliográfica. As consultas às bibliotecas respaldam a parte teórica do estudo e podem elucidar diversas questões, sejam específicas do projeto ou sobre metodologias científicas. Nesse ponto, o papel do professor orientador é fundamental para a condução da pesquisa. Além da seleção dos livros, o docente analisa as melhores possibilidades de desenvolver o assunto, em todas as suas fases. Ele também pode indicar a aplicação de entrevistas e outros elementos de apoio ao conteúdo do projeto.

Atualmente, o meio mais difundido de pesquisa é a Internet. Além de facilitar o acesso a documentos, pela rede é possível saber quanto o tema escolhido já foi objeto de estudo de outros acadêmicos. Mas essa facilidade deve ser utilizada para indicar um caminho.

3.1.4 Estrutura e regras

Antes do próprio trabalho escrito, o estudante deve fazer um projeto ou plano de pesquisa. O documento identifica o que deve ser feito, o porquê, como e onde será realizado o levantamento. Não há um modelo rígido para a apresentação do projeto de pesquisa, mas os seguintes elementos devem ser respondidos no texto:

- 1. Definição do objeto de estudo (tema/problema da pesquisa)
- 2. Justificativa
- 3. Hipóteses de trabalho
- 4. Discussão teórica
- 5. Metodologia
- 6. Pesquisa Bibliográfica

Seja monografia, dissertação ou tese, a parte escrita possui uma estrutura semelhante, embora cada uma tenha características próprias referentes à profundidade do tema estudado

De acordo com a Associação Brasileira de Normas Técnicas (ABNT), um trabalho acadêmico deve englobar os elementos pré-textuais (como resumo e índice), pós-textuais (bibliografia, anexos, entre outros) e textuais. Esses últimos compõem a parte central do trabalho - introdução, desenvolvimento e conclusão.

A introdução é a parte inicial do texto e deve constar o objeto de pesquisa, os objetivos, a justificativa da escolha do tema e outras informações que sejam necessárias para esclarecer o assunto.

A parte principal do trabalho está concentrada no desenvolvimento. É uma exposição sistematizada e ordenada de toda o estudo desenvolvido, apresentando análise e interpretação das informações e dados obtidos. A conclusão é a etapa final do texto. Nela, são apresentados os resultados tendo como referência os objetivos e hipóteses da pesquisa.

Em todo o trabalho a linguagem utilizada deve ser interessante, sem apelar para a linguagem coloquial. O trabalho deve estar de acordo com as normas da ABNT. Procure livros sobre estrutura e regras do tipo de projeto final específico de seu interesse.

3.1.5 O projeto está pronto. E agora?

Após a finalização do projeto, chega o momento de preparar a apresentação. Em geral, a banca examinadora é formada por três docentes, sendo um deles o professor orientador do projeto. Também é comum aos alunos o direito de escolha dos avaliadores, desde que seja pertinente ao assunto e ao objetivo do estudo.

Esses professores recomendam uma apresentação resumida do projeto, pontuando as características essenciais e como se chegou às conclusões. É sempre bom explicar o cronograma de todo o trabalho. É preciso, também, ficar atento ao tempo. Não é necessário explicar os conceitos já citados no projeto e pode influenciar a nota final. Lembre-se que as explicações são voltadas para os avaliadores, que já leram o seu trabalho.

Durante as considerações da banca examinadora não se deve interromper a avaliação dos professores, exceto quando eles dirigirem diretamente uma pergunta ao aluno. Educação e conhecimento dos procedimentos acadêmicos são essenciais para uma boa apresentação. Após a avaliação, os professores pedem para os presentes se retirarem da sala. É feita uma reunião onde será decidida a nota do projeto.

Cada departamento possui regras e orientações para a apresentação dos trabalhos de conclusão. Cabe ao aluno perguntar à coordenação do curso e ao orientador todas as etapas do processo de elaboração do projeto final.

Capítulo 4

Apresentações

Este capítulo oferece sugestões de como fazer a apresentação do trabalho. Uma apresen-

tação é necessária ao final do curso, é nela que se mostra os resultados obtidos de forma

resumida e, preferencialmente, simplificada. Embora o "verdadeiro" resultado seja o texto técnico, que de fato representa a contribuição científica obtida, a apresentação serve para

divulgar seus resultados e incentivar outros a se interessarem por seu trabalho.

Falando em Público 4.1

A ideia de uma apresentação não é mostrar todos os detalhes técnicos ou tentar impres-

sionar o público com seu conhecimento. O objetivo é apresentar suas principais ideias de

forma intuitiva, de modo que os presentes entendam o que foi feito e se interessem em

conhecer as minúcias lendo o texto técnico.

Vale lembrar que embora você veja seu trabalho como extremamente interessante,

geralmente seu público [ainda] não acha, e provavelmente têm coisas melhores para fazer...

É preciso atrair e manter a atenção deles, bem como garantir que eles se lembrem do que

foi dito (pelo menos da ideia principal).

Algumas noções importantes:

Motivação: Qual o problema e por que ele merece atenção?

Ideia Principal: Clara e explicitamente especificada.

Exemplos: A melhor forma de passar informações (ilustram motivação, funcionamento,

casos extremos, limitações, etc.).

19

Slides são uma excelente ferramenta **de apoio** ao apresentador, mas muitas vezes tomam vida própria e se tornam o elemento principal. É essencial, embora um pouco difícil, evitar a "morte por Powerpoint".

Existem muitas sugestões para fazer uma boa apresentação², por exemplo, imitar um bom apresentador³, boas práticas na elaboração de slides⁴, como organizar o conteúdo de um slide⁵ (ou mesmo "vida após a morte"⁶). Entretanto, as duas noções mais importantes são: você nunca se prepara demais para fazer uma apresentação, e a única regra de uma apresentação é a de atenção⁷.

Olivia Mitchell sugere as seguintes formas de manter a atenção da platéia⁸:

- 1. Fale sobre algo que interesse a platéia.
- 2. Diga porque deveriam prestar atenção.
- 3. Não apresente algo muito fácil ou muito difícil.
- 4. "Mudanças" prendem a atenção.
- 5. Conte estórias.
- 6. Faça pausas.
- 7. Seja breve.

Demonstrações ao vivo são impressionantes, desde que funcionem corretamente e não evidenciem as limitações do seu trabalho. Lembre-se que eventos importantes são, em sua maioria, regidos pela *Lei de Murphy*.

Por fim, lembre-se que é normal ficar nervoso perante uma platéia, e não há uma cura genérica para este problema. Há muitas sugestões de como lidar com isso⁹, inclusive uma que diz que o problema é você¹⁰. Tente descobrir o que funciona melhor para si (boa sorte!).

 $^{^{1}} http://www.smallbusinesscomputing.com/biztools/article.php/684871/\\ Death-By-Powerpoint.htm$

²https://hbr.org/2014/11/how-to-give-a-stellar-presentation

³https://www.youtube.com/watch?v=2-ntLGOyHw4

⁴https://www.youtube.com/watch?v=Iwpi1Lm6dFo

⁵https://hbr.org/2012/10/do-your-slides-pass-the-glance-test

⁶https://www.youtube.com/watch?v=lpvgfmEU2Ck

⁷http://finiteattentionspan.wordpress.com/2009/11/02/the-only-rule-about-giving-presentations-tha

 $^{^8}$ http://www.speakingaboutpresenting.com/content/7-ways-audience-attention-presentation

⁹http://www.wikihow.com/Overcome-Stage-Fright

¹⁰ http://seriouspony.com/blog/2013/10/4/presentation-skills-considered-harmful

4.2 BEAMER

A classe BEAMER, disponível no CTAN¹¹, é a recomendada para criar apresentações. Não só possibilita um resultado visualmente interessante, como também aproveita parte do texto escrito em LATEX. O manual¹² oferece instruções sobre o uso da classe e, principalmente, diretrizes para criar apresentações (especialmente as Seções 4 e 5 do Capítulo I).

¹¹http://www.ctan.org/pkg/beamer

¹²http://www.ctan.org/tex-archive/macros/latex/contrib/beamer/doc/beameruserguide.pdf

Referências

- [1] Abramo, Cláudio Weber: Relações entre índices de percepção de corrupção e outros indicadores em onze países da américa latina. SPECK, Bruno W. et al. Os custos da corrupção. Cadernos Adenauer, (10):47–62, 2000. 1
- [2] Filgueiras, Fernando: A tolerância à corrupção no brasil: uma antinomia entre normas morais e prática social. Opinião Pública, 15(2):386–421, 2009. 1
- [3] engines db: Dbms popularity broken down by database model. https://db-engines.com/en/ranking_categories. Acessado em setembro de 2017. 2
- [4] PPGI/UnB: Regimento do programa de pós-graduação em informática da universidade de brasília. http://ppgi.unb.br/normatives, 2013. 5
- [5] Covey, Stephen R, A Roger Merrill e Rebecca R Merrill: First things first. Simon and Schuster, 1995. 13

Apêndice A Fichamento de Artigo Científico



Fichamento de Artigo Científico

Prof. Guilherme N. Ramos

Um fichamento reúne elementos relevantes do conteúdo, apresentando a estrutura do texto, e deve seguir a seqüência do pensamento do autor, destacando suas ideias, argumentos, justificativas, exemplos, fatos, etc.

1 Artigo Científico

Geralmente, um artigo científico é escrito com a seguinte estrutura (buscando responder algumas questões):

I. Introdução

- Qual o contexto do problema? (O que? Onde? Quando?)
- Qual a principal questão ou problema colocado? (Por quê? Como? Qual?)
- Qual o objetivo visado? O que se pretende constatar ou demonstrar? (investigar, analisar, refletir, contribuir,...)

II. Referencial Teórico

- Quais são os autores/teorias/conceitos que já estudaram os principais assuntos abordados e que sustentam ao texto?
- Quais os resultados mais recentes relacionados a eles?

III. Metodologia/Desenvolvimento

- Quais os procedimentos metodológicos adotados? (natureza do trabalho: empírico, teórico, histórico) (coleta de dados: questionário, entrevista, levantamento bibliográfico).
- Como a pesquisa foi desenvolvida? Quais as principais relações entre teoria e prática?
- Havendo artefato proposto, ele está disponível para utilização e/ou modificação?

IV. Resultados

- Houve validação (por meio de experimentação)? Como foi feita?
- Os resultados obtidos são corretos/válidos?

V. Conclusões

- Qual o problema atacado?
- Quais os resultados obtidos para os objetivos propostos?
- Quais conclusões podem ser tiradas destes resultados?
- Quais as limitações da metodologia utilizada?
- Quais as possibilidades de trabalhos futuros para o problema?

2 Fichamento

Neste contexto, um fichamento deve conter a seguinte estrutura:

- 1. *Identificação do aluno:* indicação precisa de quem é o autor do fichamento.
- 2. *Identificação do texto:* indicação precisa de quem são os autores do texto analisado e dos detalhes do documento, de modo que se possa buscá-lo para uma leitura completa.
- 3. **Pontos-chave:** noções mais relevantes do texto analisado. *Proposta* (o que é apresentado?), *mérito* (por que é relevante?), *validação* (como verificar a utilidade?), e *perspectivas* (o que pode ser melhorado?).
- 4. *Palavras-chave:* expressões que identificam o assunto abordado.
- 5. **Sinopse do texto:** resumo com suas palavras. Deve ser mais detalhado que um abstract, geralmente apresentando pelo menos um parágrafo por seção do texto original. No caso de inclusão de trechos, o texto deve ser identificado entre "aspas" e concatenado através de suas próprias palavras.
- 6. **Análise crítica:** posicionar-se em relação as seguintes questões: pertinência do assunto; forma como foi abordado; comparação com outras abordagens do mesmo assunto (caso conheça). Junto ao *resumo*, é a parte mais interessante para o leitor, pois apresenta uma avaliação do conteúdo apresentado.

2.1 Exemplo

- 1. *Identificação do aluno:* Alan Mathison Turing, 00/000000
- 2. *Identificação do texto:*Guilherme N. Ramos, Yutaka Hatakeyama, Fangyan Dong, and Kaoru Hirota, Hyperbox clustering with Ant Colony Optimization (HACO) method and its application to medical risk profile recognition, Applied Soft Computing, Vol. 9, Issue 2, pp 632-640, 2009. (doi:10.1016/j.asoc.2008.09.004)
- 3. Pontos-chave:

Proposta: HACO - método para aglomeração de dados utilizando hipercaixas com posicionamento otimizado via algoritmo de colônia de formigas.

Mérito: apresenta uma nova forma de fazer agrupamentos considerando a topologia do espaço de dados e fornecendo resultados intuitivos e facilmente utilizáveis.

Validação: comparação com algoritmos conhecidos em testes com dados padrões e com dados de infecção viral para diagnóstico auxiliado por computador.

Perspectivas: adequação das dimensões das hipercaixas, diminuição de parâmetros.

4. *Palavras-chave*:colônia de formigas, hipercaixa, otimização, reconhecimento de padrões.

5. **Sinopse do texto:** A Colônica de Formigas (ACO) é um método de otimização que pode ser utilizado para agrupar dados. Hyperbox clustering with Ant Colony Optimization (HACO) é um método de agrupamento que utiliza ACO tentar posicionar hipercaixas no espaço de forma a agrupar a maior quantidade de dados possível, e ainda gera uma forma simples de classificar novos dados.

ACO é baseado no comportamento de formigas reais, que otimizam o caminho percorrido entre o alimento e o formigueiro. Hipercaixas definem de forma muito simples uma região em um espaço n-dimensional, combinadas para definir regiões de topologia complexa, e utilizadas como um classificador de forma trivial.

HACO busca encontrar uma partição de dados, efetivamente definindo grupos. Primeiro, aplica ACO para tentar posicionar hipercaixas de forma que estas contenham a maior quantidade possível de dados. A seguir, se não há conhecimento prévio da quantidade de classes, considerase que as hipercaixas que se sobrepõem representam uma mesma classe de dados, e [grupos de] hipercaixas distintas representam classes diferentes. Caso o número de classes seja conhecido, HACO aplica o algoritmo Nearest-neighbor (NN) para definir a quantidade correta de grupos. Uma consequência de se usar hipercaixas é que o resultado do agrupamento define também um classificador: se um novo dado está dentro de uma hipercaixa, sua classe será a mesma da definida por esta hipercaixa.

Os resultados experimentais de HACO fora, comparados a três algoritmos que têm o mesmo fim: testado em NN, Fuzzy C-Means (FCM), e o próprio ACO (com uma abordagem diferente para agrupamento). O primeiro teste foi em conjuntos de dados sintéticos, e serviu como prova de conceito, oferecendo diversas informações sobre o comportamento do método em função de certas configurações. Um segundo experimento foi realizado com dados reais de pacientes para agrupá-los em "saudáveis" e "não saudáveis", e HACO obteve o melhor resultado dentre os algoritmos testados. A análise da estrutura do classificador gerado possibilita descobrir informações relativas às características das classes, indicando um "perfil de risco" para os pacientes.

Foi apresentado o método HACO para agrupar dados, utilizando a meta-heurística ACO e hipercaixas, que possibilita a extração de informações inerentes a estrutura dos dados. HACO foi validado com experimentos, e demonstrou grande potencial. Os resultados são muito influenciados pela configuração dos parâmetros, que será investigada.

6. Análise crítica: Este é o melhor artigo de todos os tempos. O artigo apresenta uma forma inovadora de agrupar dados, de forma não-supervisionada (embora possa aproveitar informações se houver). O resultado pode ainda ser utilizado como classificador de novos dados, e - o mais interessante - analisado para descobrir informações sobre as classes. Além disso, explora as vantagens de cada elemento que compõe o método, obtendo melhores resultados e diminuindo o custo computacional. A aplicação em um caso real, cujos resultados podem ser utilizados para auxiliar o diagnóstico de pacientes, dá mais destaque ao trabalho.

O problema de agrupamento de dados é muito pertinente e, em tempos de excesso de dados, a possibilidade de análise intuitiva da estrutura e descoberta de conhecimento é bastante interessante. Além disso, a solução proposta é de uso geral, oferecendo mais possibilidades de uso.

Os experimentos realizados foram coerentes e suficientes para demonstrar o que foi afirmado. Entretanto, o método só foi comparado a outros algoritmos simples, seria interessante uma comparação com algoritmos mais avançados, bem como específicos para aplicação. A comparação também foi em uma única aplicação específica, seria melhor que houvesse mais testes com outros dados para conclusões melhor embasadas. Além disso, é preciso uma análise mais profunda quanto às configurações de HACO, que influenciam muito o resultado.

Anexo I

Documentação Original UnB-CIC (parcial)

```
% -*- mode: LaTeX; coding: utf-8; -*-
: unb-cic.cls (LaTeX2e class file)
%% Authors : Flávio Maico Vaz da Costa
                                             (based on previous versions by José Carlos L. Ralha)
%% Version : 0.96
\hfill 
%%
                               : 0.75 [04/04/2005] - Fixed font problems, UnB logo
%%
                                                                                            resolution, keywords and palavras-chave
%%
                                                                                            hyphenation and generation problems,
%%
                                                                                            and a few other problems.
                               : 0.8 [08/01/2006] - Corrigido o problema causado por
%%
%%
                                                                                            bancas com quatro membros. O quarto
%%
                                                                                            membro agora é OPCIONAL.
%%
                                                                                            Foi criado um novo comando chamado
%%
                                                                                            bibliografia. Esse comando tem dois
%%
                                                                                            argumentos onde o primeiro especifica
%%
                                                                                            o nome do arquivo de referencias
%%
                                                                                            bibliograficas e o segundo argumento
%%
                                                                                            especifica o formato. Como efeito
%%
                                                                                            colateral, as referências aparecem no
%%
                                                                                            sumário.
%%
                               : 0.9 [02/03/2008]
                                                                                      - Reformulação total, com nova estrutura
%%
                                                                                            de opções, comandos e ambientes, adequação
%%
                                                                                            do logo da UnB às normas da universidade,
%%
                                                                                            inúmeras melhorias tipográficas,
```

%%	aprimoramento da integração com hyperref,
%%	melhor tratamento de erros nos comandos,
%%	documentação e limpeza do código da classe.
%%	: 0.91 [10/05/2008] - Suporte ao XeLaTeX, aprimorado suporte para
%%	glossaries.sty, novos comandos \capa, \CDU
%%	e \subtitle, ajustes de margem para opções
%%	hyperref/impressao.
%%	: 0.92 [26/05/2008] - Melhora do ambiente {definition}, suporte
%%	a hypcap, novos comandos \fontelogo e
%%	\slashedzero, suporte [10pt, 11pt, 12pt].
%%	Corrigido bug de seções de apêndice quando
%%	usando \hypersetup{bookmarksnumbered=true}.
%%	: 0.93 [09/06/2008] - Correção na contagem de páginas, valores
%%	load e config para opção hyperref, comandos
%%	\ifhyperref e \SetTableFigures, melhor
%%	formatação do quadrado CIP.
%%	: 0.94 [17/04/2014] - Inclusão da opção mpca.
%%	: 0.95 [06/06/2014] - Remoção da opção "mpca", inclusão das opções
%%	"doutorado", "ppginf", e "ppca" para identificar
%%	o programa de pós-graduação. Troca do teste
%%	@mestrado por @posgraduacao.
%%	: 0.96 [24/06/2014] - Ajuste do nome do curso/nome do programa.
%%	