

**Mineração de dados não estruturados sobre eventos em
sistemas de energia elétrica**

Rodrigo Rossetto Gati

Dissertação de Mestrado do Programa de Mestrado Profissional em
Matemática, Estatística e Computação Aplicadas à Indústria (MECAI)

Rodrigo Rossetto Gati

Data Mining of non-structured data on events in electrical energy systems

Dissertation presented to the Institute of Mathematical and Computer Sciences - ICMC-USP, as part of the requirements to obtain the title of Master of Science - Professional Master in Mathematics, and Computing Applied to Industry. *EXAMINATION BOARD PRESENTATION COPY.*

Concentration Area: Mathematics, Statistics and Computing.

Advisor: Prof. Dr. Alexandre Cláudio Botazzo Delbem

USP - São Carlos

November 2018

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito: 29/11/18

Assinatura: 

Rodrigo Rossetto Gati

Mineração de dados não estruturados sobre eventos em sistemas de energia elétrica

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para obtenção do título de Mestre em Ciências – Mestrado Profissional em Matemática, Estatística e Computação Aplicadas à Indústria. *EXEMPLAR DE DEFESA*.

Área de Concentração: Matemática, Estatística e Computação.

Orientador: Prof. Dr. Alexandre Cláudio Botazzo Delbem

USP - São Carlos

Novembro de 2018

**Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)**

G261m Gatti, Rodrigo Rossetto
Mineração de dados não estruturados sobre eventos
em sistemas de energia elétrica / Rodrigo Rossetto
Gatti; orientador Alexandre Cláudio Botazzo Delbem. -
- São Carlos, 2018.
81 p.

Dissertação (Mestrado - Programa de Pós-Graduação
em Mestrado Profissional em Matemática, Estatística
e Computação Aplicadas à Indústria) -- Instituto de
Ciências Matemáticas e de Computação, Universidade
de São Paulo, 2018.

1. Mineração de dados. 2. Damicore. 3. Dados não
estruturados. 4. Complearn. 5. Sistemas de energia
elétrica. I. Delbem, Alexandre Cláudio Botazzo,
orient. II. Título.

Bibliotecários responsáveis pela estrutura de catalogação da publicação de acordo com a AACR2:
Gláucia Maria Saia Cristianini - CRB - 8/4938
Juliana de Souza Moraes - CRB - 8/6176

RODRIGO ROSSETTO GATI

MINERAÇÃO DE DADOS NÃO ESTRUTURADOS SOBRE EVENTOS EM SISTEMAS DE ENERGIA ELÉTRICA

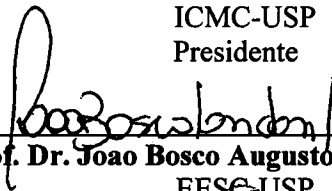
Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação - ICMC-USP, como parte dos requisitos para obtenção do título de Mestre.

Aprovado em 18 de março de 2019.

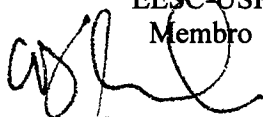
COMISSÃO JULGADORA,



Prof. Dr. Alexandre Cláudio Botazzo Delbem
ICMC-USP
Presidente



Prof. Dr. Joao Bosco Augusto London Junior
EESC-USP
Membro



Prof. Dr. Carlos Dias Maciel
EESC - USP
Membro



Prof. Dr. Anderson da Silva Soares
UFG
Membro

*Este trabalho é dedicado às crianças adultas que, quando pequenas,
sonhavam em se tornar cientistas.*

*Em especial aos pesquisadores do Instituto de Ciências Matemáticas e de
Computação (ICMC).*

AGRADECIMENTOS

Agradeço ao Mestre Jesus pela oportunidade de me conceber esta passagem por este planeta.

Agradeço ao meu orientador Alexandre C. B. Delbem, pela sua dedicação, ideias, críticas, paciência e oportunidades concedidas.

Aos amigos e companheiros que juntos conseguimos o nosso objetivo, sendo o término do mestrado profissional.

Agradeço ao meu pai Natalino e minha mãe Iara, meus irmãos, pela presença, apoio constante e paciência.

A todos os servidores públicos da escola Instituto Federal de São Paulo (IFSP) campus Catanduva, que buscam aprimoramento e conhecimento em sua atividade de trabalho, onde possam dar o retorno à sociedade.

Finalmente agradeço a minha esposa Sileni e filha Ana Luiza, por ser minha luz e fonte de coragem em todos os momentos de desafios.

“Há homens que lutam um dia e são bons. Há outros que lutam um ano e são melhores. Há aqueles que lutam muitos anos e são muitos bons. Mas há os que lutam por toda a vida, esses são os imprescindíveis.”

(Bertold Brecht)

RESUMO

GATI, R.R. **Mineração de dados não estruturados sobre eventos em sistemas de energia elétrica**. 2018. 81p. Dissertação (Mestrado em Ciências – Programa de Mestrado Profissional em Matemática, Estatística e Computação Aplicadas à Indústria) – Instituto de Ciências matemáticas e Computação, Universidade de São Paulo, São Carlos – SP, 2018.

Atualmente com avanço da tecnologia em todos os setores, temos uma geração e um volume muito grande de dados ou informações, que podem ser transformadas e utilizadas para o gerenciamento e tomada de decisão. Para a transformação desses dados verificamos também um grande desenvolvimento de máquinas e programas computacionais. Este trabalho propõe a aplicação da ferramenta computacional *Damicore* em dados não estruturados, fornecido por uma empresa de distribuição de energia, da qual estão relacionados com a interrupção no fornecimento de energia aos consumidores. Tendo como finalidade realizar a classificação e agrupamentos das informações que possam gerar relatórios para tomada de decisão gerencial. Por outro lado propor uma estruturação e padronização dessas informações, conforme a necessidade de cada empresa.

Palavras-chave: Mineração de dados, Dados não-estruturados, *Damicore*, *Complearn*, Sistemas de energia elétrica.

ABSTRACT

GATI, R.R. **Data mining of non-structured data on events in electrical energy systems**. 2018. 81p. Dissertação (Mestrado – Programa de Mestrado Profissional em Matemática, Estatística e Computação Aplicadas à Indústria) – Instituto de Ciências matemáticas e Computação, Universidade de São Paulo, São Carlos – SP, 2018.

Today, with advances in technology in all sectors, we have a generation and a very large volume of data or information that can be transformed and used for management and decision making. For the transformation of these data we also see a great development of machines and computer programs. This work proposes the application of the Damicore computational tool in unstructured data, provided by an energy distribution company, which are related to the interruption in energy supply to consumers. Its purpose is to carry out the classification and grouping of information that can generate reports for managerial decision making. On the other hand, propose a structuring and standardization of this information, according to the need of each company.

Keywords: Data mining, Non-structured data, Damicore, Complearn, Electrical energy systems.

LISTA DE ILUSTRAÇÕES

Figura 1 - Especificações da Máquina Virtual	40
Figura 2 - Especificações da Máquina Virtual	40
Figura 3 - Planilha do Relatório de Interrupções Descrito pela Subestação em Excel.ods	42
Figura 4 - Planilhas e Conversão de Arquivo	42
Figura 5 - Exclusão da linha sem Informação na Planilha Alimentadores.....	43
Figura 6 - Pasta de Arquivos de Texto.....	43
Figura 7 - Arquivos de Texto.....	44
Figura 8 - Pastas de Arquivos de Texto.....	45
Figura 9 - Terminal.....	46
Figura 10 - Visualização das Saídas (Gráficos).....	49
Figura 11 - Planilha Proposta de Dados Estruturados Padronizada.ods.....	55
Figura 12 - Gráfico <i>Trees Transform Branches Todosdados</i>	58
Figura 13 - Exemplos de Saídas do <i>Damicore de Dados Não Estruturados</i>	60
Figura 14 - Gráfico <i>Trees Transform Branches descreva</i>	64
Figura 15 - Gráfico <i>Trees Transform Branches Estruturado</i>	66
Figura 16 - Exemplos de Saídas do <i>Damicore de Dados Estruturados</i>	66
Figura 17 - Gráfico <i>Trees Transform Branches Todosdados x Descrição x Estruturado</i>	68
Figura 18 - Agrupamentos da Pasta de Texto Todosdados x Descrição.....	69

LISTA DE TABELAS

Tabela 1 - Valores de Agrupamentos x Não Agrupamentos Todos dados.....	63
Tabela 2 - Valores de Agrupamentos x Não Agrupamentos Descrição.....	65
Tabela 3 - Comparação do Aumento de Agrupamentos da Pasta Todos dados x Descrição.....	74

LISTA DE SÍMBOLOS

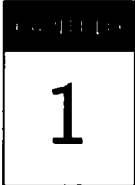
IEA	International Energy Agency
EIA	Energy Information Administration
TWh	Tera Watt Hora
ANEEL	Agência Nacional de Energia Elétrica
kW	Kilo Watt
<i>Smarts Grids</i>	Sistema de energia elétrica que se utiliza da tecnologia da informação
SGBD	Gerenciamento de Banco de Dados
SQL	Structured Query Language
XML	Extensible Markup Language
JSON	Java Script Object Notation
DAMICORE	Data Mining of Code Repositories (Ferramenta Computacional)
POSTGRESSQL	Sistema de Gerenciamento de Banco de Dados Relacional
EXPLAIN ANALYZE	Comando para visualizar e analisar o comportamento
BI	Business Intelligence é uma ferramenta computacional
DW	Data Warehouse é um banco de dados SQL Server
CAD	Software Comercial chamado de AutoCAD
DBF	Extensão de Arquivo para Banco de Dados
ECER	Interface gráfica denominada Modelo Entidade - Relacionamento
ER	Entidade – Relacionamento
PDM	Software Comercial para Gestão de dados do Produto
SKOLEM	Função para identificar objetos
XPath	Linguagem de consulta
GPP	General purpose processor (sistemas híbridos customizados)
FPGA	Field-programmable gate array (sistemas híbridos customizados)
KERNELS	Trechos de códigos para plataformas híbridas
NCD	Normalized Compression Distance (Algoritmo)
NJ	Neighbor Joining (Algoritmo)

FN	Fast Newman (Algoritmo)
SE	Subestação de Energia Elétrica
SID	Software Integrity Diagnosis para Detecção de Plágio

Sumário

1.	INTRODUÇÃO.....	25
2.	PESQUISA BIBLIOGRÁFICA	29
2.1	Dados Estruturados	29
2.2	Dados Semi-Estruturados	31
2.3	Dados Não Estruturados	33
2.4	Soluções Baseadas no Método <i>Compleam</i>	33
2.5	Soluções Baseadas nos Métodos <i>Damicore</i>	34
3.	MATERIAIS E MÉTODOS	37
3.1	Ferramenta Computacional <i>Damicore</i>	37
3.2	Bibliotecas / Ferramentas.....	38
3.3	Programa <i>Damicore.py</i>	39
3.4	Máquina Virtual	39
3.5	Dados de Eventos em Sistemas de Energia Elétrica.....	41
3.6	Estrutura dos Dados em Excel	42
3.7	Estrutura de Dados no Software <i>Damicore</i>	43
3.8	Criação das Pastas de Arquivos de Texto	44
3.9	Execução do Software <i>Damicore</i>	45
3.10	Geração de Saídas (Gráficos)	48
4.	PROPOSTA.....	53
4.1	Dados Estruturados	53
4.2	Estruturação de Dados.....	54
4.2.1	Termos Técnicos Padronizados	54
4.2.2	Planilha de Dados Estruturados e Padronizados.....	55
5.	RESULTADOS	57
5.1	Dados Não Estruturados	57
5.1.1	Pasta de Arquivo de Texto Completa (todosdados)	57
5.1.1.1	Agrupamentos Obtidos.....	58
5.1.1.2	Agrupamentos Relevantes	59
5.1.1.3	Exemplos de Saídas do <i>Damicore</i>	59
5.1.1.4	Saídas do <i>Damicore</i> da Pasta Todosdados	62

5.1.2	Pasta de Arquivo de Texto(descricao).....	63
5.1.2.1	Saídas do <i>Damicore</i> da Pasta Descricao	64
5.2	Dados Estruturados.....	65
5.2.1	Exemplos de Saídas do <i>Damicore</i>	66
5.3	Eficácia das Saídas de Agrupamento.....	67
5.3.1	Saída Gráfica do <i>Damicore</i>	67
5.3.2	Comparação de Agrupamentos.....	69
5.4	Discussão dos Resultados	72
6.	CONCLUSÃO	77
	REFERÊNCIAS	79



INTRODUÇÃO

1. INTRODUÇÃO

A energia elétrica é considerada fundamental para o crescimento, desenvolvimento econômico e social de qualquer país. Tem como função a prestação de serviços em diversos setores da economia como no setor industrial, residencial, comercial, rural e público. É classificada como investimento de infraestrutura nos países.

Segundo o *International Energy Agency* (IEA) a energia elétrica representa 17% de toda a demanda global, tendo como previsão um aumento de 23% até ano de 2050 (CORREA; TAFFAREL; MENON, 2016).

A *Energy Information Administration* (EIA) relata que o consumo mundial de energia de 2003 a 2030, sofrerá um aumento de consumo de 71%, onde essa demanda será maior na Ásia, América Latina, África e Oriente Médio. Porém, demanda por energia não indica redução e sim crescimento futuro (SILVA; PRIMO, 2013).

As fontes mais relevantes de energia elétrica estão alocadas nos combustíveis fósseis, como o petróleo, carvão mineral e gás natural. Outras fontes importantes são os combustíveis nuclear e hídrico (HIRANO, 2015).

No ano 2005, a geração de energia elétrica mundial foi de 17.450 TWh, sendo representada por 40% do carvão mineral, 20% do gás natural, 16% nuclear, 16% hídrica, 7% do petróleo e apenas 2% para energias renováveis (SILVA; PRIMO, 2013).

Algumas das principais fontes de geração de energia elétrica são obtidas através dos geradores hídrico, nuclear, térmico, eólico, solar, maré-motriz, motores à combustão e biomassa. Com a preocupação crescente dos países em relação ao aquecimento global e as possíveis mudanças climáticas, foi determinada em 12 de fevereiro

de 2005 pelo *Protocolo de Kyoto*, a redução de emissões de gases que provocam o efeito estufa (FALLEIRO; GASTALDINI; ANDRADE, 2014).

O autor Hirano (2015), menciona que segundo a *Agência Nacional de Energia Elétrica (ANEEL)*, o Brasil possui 79,5% da energia elétrica gerada de forma renovável.

No Brasil os valores percentuais das fontes utilizadas para a geração de energia elétrica são 67,07% hídrica - 19,04% fóssil - 9,58% biomassa - 2,85% eólica e 1,45% nuclear (HIRANO, 2015).

O autor Rosa (2007), relata que o Brasil é considerado como o primeiro país do mundo em recursos hídricos, mas classificado como o quarto na utilização deste recurso. O Brasil possui 25% desse recurso, enquanto nos outros países como Estados Unidos, Canadá e China são 80%, 67% e 65%.

Comparando as fontes de energia nuclear, hídrica e térmica no Brasil, verifica-se que o investimento financeiro é maior na nuclear, a geração de kW tem o menor custo na hídrica e a taxa de retorno maior está na térmica (ROSA, 2007).

Atualmente existe um novo paradigma em sistemas elétricos, chamado de *Smarts Grids*, onde se aplicam técnicas/tecnologias baseadas na extração, armazenamento, classificação e conhecimento dos dados, com a finalidade de monitorar, corrigir, reconfigurar o sistema elétrico, propondo tomadas de decisões com ações antecipadas de forma automática na proporção que são identificadas situações anormais de operação, resultando na confiabilidade, eficiência e segurança aos consumidores de energia elétrica. Para tanto, é necessária a aplicação de ferramentas computacionais com sistemas integrados, combinando aquisição, processamento e análise dos dados (BRITO, 2002).

O conceito de banco de dados pode ter várias definições, como uma coleção de dados, um conjunto de dados, um armário de arquivamento de dados, na qual este banco deve conter essas informações (SILBERSCHATZ et al., 2012; DATE et al., 2004).

Um sistema de gerenciamento de banco de dados (SGBD) é um conjunto de programas, que tem acesso às informações contidas nos dados do banco, com a finalidade de modificar ou alterar quando necessário (SILBERSCHATZ et al., 2012; DATE et al., 2004).

Bancos de dados estruturados recebe a nomenclatura como *banco de dados relacionais*, pois são estruturados em registros de formato fixo com vários tipos, sendo

composto por uma coleção de tabelas, para representar tanto os dados como as relações entre eles, onde cada tabela contém registro de um tipo específico (SILBERSCHATZ et al., 2012; DATE et al., 2004).

Malacrida et al. (2014) define um banco de dados relacional está organizado em tabelas, chaves, atributos e relacionamentos, onde a tabela é um conjunto não ordenado de linhas, e a linha é composta por vários atributos. As chaves estabelecem relações entre as linhas das tabelas e os atributos são elementos dos dados de forma particular da tabela.

A finalidade do modelo de banco relacional é apresenta-lo de forma simples, flexível, independente e inter-relacionado representando-o por tabelas, fortalecendo as relações entre os dados e execução da estrutura, não sendo necessária a reconstrução de todo o banco de dados, descartando assim os conceitos anteriores. A linguagem padrão utilizada é a SQL (*Structured Query Language*), desenvolvida pela empresa IBM (SILBERSCHATZ et al., 2012; DATE et al., 2004).

O banco de dados denominado semi-estruturado, possui dados relacionais e dados não convencionais. As duas linguagens mais utilizadas para este tipo de banco de dados é a XML (*Extensible Markup Language*) e a JSON (*Java Script Object Notation*), que possui suas raízes no gerenciamento de documentos, sendo transformada numa linguagem para estruturar informações dos dados (SILBERSCHATZ et al., 2012; DATE et al., 2004).

No modelo de bancos não estruturados ou denominado *banco de dados não convencionais*, os dados não seguem um modelo de banco de dados. São dados do tipo temporais, espaciais e de multimídia (SILBERSCHATZ et al., 2012; DATE et al., 2004).

Neste trabalho será aplicada a ferramenta computacional *Damicore* (*Data Mining of Code Repositories*), apresentada por Sanches, Cardoso e Delbem (2011), em dados reais, não estruturados, fornecido por uma subestação de energia elétrica de distribuição de uma concessionária de energia. Estes dados contemplam informações sobre a causa ou motivo que resultou na interrupção do fornecimento de energia elétrica aos consumidores. Também será proposto a estruturação das informações dos dados para a aplicação do *Damicore*.

A finalidade da aplicação da ferramenta computacional *Damicore* é obter agrupamentos de informações relevantes, que possam ajudar na tomada de decisões gerenciais

PESQUISA BIBLIOGRÁFICA

2. PESQUISA BIBLIOGRÁFICA

2.1 Dados Estruturados

Em geral, dados estruturados apresentam-se organizados com base em projetos de banco de dados, os autores Teorey et al. (1986) apresentaram uma metodologia para construção de design de grandes bancos de dados relacionais, utilizando o modelo de relacionamento de entidade estendido. Resultando em representações precisas, Flexível, acomodar futuros requisitos de processamento, integridade dos dados e reduzindo o número de dependências de dados que deve ser analisadas.

O modelo entidade – relacionamento denominado ECER, é um modelo de interface gráfica que facilita aos usuários a realização de consultas e atualizações dos dados em bancos de dados relacionais existentes, através da manipulação de diagramas de esquema (CZEJDO et al., 1990).

Quando a informação possuir imprecisão e há incompletude em documentos textuais, aplicar-se a Teoria de Conjuntos Fuzzy e Lógica Fuzzy para a sua recuperação dos dados, obtendo resultados estimuladores ao aprofundamento e aperfeiçoamento do modelo proposto (MARQUES; BOSCARIOLI, 2002). Apesar do autor não ter validado o modelo, os resultados obtidos mostram-se satisfatórios com a tecnologia aplicada.

Através de ferramentas computacionais para a mineração de dados, é utilizada uma análise com o uso de técnicas de diagnóstico, para obtenção de informações da

vida útil dos cabos de distribuição energia elétrica na tomada de decisões, como apoio para a engenharia elétrica, utilizando banco de dados SQL (QUAK et al., 2002).

No trabalho desenvolvido Hassan et al. (2003) foi descrito um mecanismo para localizar informações em banco de dados relacionais para usuários com pouca experiência em realizar pesquisas por palavras-chaves na web, a abordagem utiliza a estimativa estatística de busca. O resultado é eficiente, pois o mecanismo é flexível, mantém a autonomia e a heterogeneidade das bases de dados.

A transformação de um banco de dados não convencional em um banco de dados relacional, utilizando a linguagem SQL em imagens obtidas por um sensor, formando um banco de dados geográfico para a aplicação de um planejamento territorial (BOLFE et al., 2009).

A partir de dados não estruturados de saúde bucal, coletados pela faculdade de Odontologia da Pontifícia Universidade Católica do Rio Grande do Sul, foi desenvolvido um banco de dados estruturado completo para armazenamento, organização, utilização das informações e recuperação de dados provenientes de prontuários odontológicos (BLOMBERG et al., 2009).

Para a migração das informações de uma base de dados já existente para uma outra base, mantendo a integralidade das informações, Malacrida et al. (2014) desenvolveram uma ferramenta de suporte ao usuário, que otimiza a referida migração, reduzindo o índice de erros e eliminando o processo manual.

Informações reais do cotidiano empresarial obtidas pelo método *Explain Analyze*, foram submetidas a uma análise de algoritmos de consultas do sistema de gerenciamento de dados PostgreSQL. Para consultas com doze ou mais tabelas, o algoritmo de programação dinâmica converge para resultados satisfatórios com maior eficiência. Com tabelas inferiores, é utilizado o algoritmo genético, excluindo as consultas com *inner join* (MAGALHÃES et al., 2015).

Aplicação de uma ferramenta computacional chamada *Business Intelligence (BI)*, em um banco de dados SQL Server chamado *Data Warehouse (DW)*, com a finalidade de análise dos dados para decisões estratégicas com foco em Shopping Center. Portanto os resultados foram satisfatórios (PEREIRA et al., 2016).

2.2 Dados Semi-Estruturados

Apresenta uma técnica para armazenar dados não estruturados em um banco de dados estruturados, através do formato de arquivo XML. Usa ontologias para estruturar a lógica do banco de dados. Utiliza mecanismo de identificação de instâncias de dados XML baseado em funções Skolem e XPath. E para atualização dos dados aplica uma técnica de manutenção incremental (SACCOL, 2001).

Os autores Carlos et al. (2002) aborda um estudo completo sobre a aplicação da linguagem XML, com a finalidade de realizar consultas e extração de dados pela Web baseado em Ontologias. Também desenvolveu uma ferramenta/aplicação sendo disponibilizada pela URCCAMP-BAGÉ/RS para uso na sua Home-Page.

Proposta de solução à integração de esquemas semi-estruturados com a linguagem XML através de um processo bottom-up, que adota um modelo conceitual para esquemas canônico e global. Por ser uma abordagem semi-automática, garante resultado desejado para o esquema global (MELLO, 2002).

Desenvolvimento de um banco de dados em XML, com a finalidade de disponibilizar informações via internet sobre a criação de Ovinos, permitindo aos associados uma maior interação junto a Associação Brasileira de Criadores de Ovinos (COSTA, 2002).

Para consultar fontes de dados semi-estruturados em XML aplica-se o modelo conceitual global XPath, porém, foi especificado um novo critério CXPath para validar consultas, servindo de referência para planejadores de consulta, usuários e implementadores da linguagem (FEIJÓ, 2005).

Para problemas relacionados com a consulta de fontes em XLM integradas, foi proposta uma abordagem que define operadores de fragmentação de um esquema global em vários esquemas locais e um algoritmo para decomposição de consultas escritas CXPath (SILVEIRA, 2006).

Desenvolvimento de cinco técnicas para realizar o casamento de forma semi-automática entre esquemas XML e relacionais, sendo uma técnica linguística e quatro estruturais. Os resultados demonstraram que o casamento linguístico é eficiente em todas as consultas, porém os casamento estruturais depende da nomenclatura usada nos esquemas (MERGEN, 2005).

Apresentação de uma plataforma em linguagem XML para desenvolvimento de sistemas colaborativos na web, vinculados a sistemas CAD e PDM existentes, para aplicações na fase de projeto conceitual do produto, garantindo portabilidade, armazenamento e conversão em 3D de esboços prescritivos (RAJESH; MAGESHBABU, 2008).

Desenvolvimento de uma abordagem em teoria de gráficos para problemas de aninhamento em esquemas XML, com a finalidade de realizar o mapeamento de esquemas ER (Entidade – Relacionamento), proporcionando uma otimização no desempenho de consulta e validação (FRANCESCHET; GUBIANI; MONTANARI, 2013).

Os autores Oliveira et al. (2014) apresentam uma nova abordagem em LBSN específico ao sistema Crowd4city, para aplicações de redes sociais baseadas em localização através da consulta de dados espaciais. Utilizando técnica Map/Reduce, banco de dados NoSQL e o desenvolvimento de extensões para indexação e consulta de Big Data espacial. Concluiu que a nova abordagem funciona muito melhor que a anterior em termos de tempo de resposta, porém se comparada ao ambiente LBSN sensível ao usuário, os tempos de resposta alcançados ainda devem ser melhorados.

Análise sobre o gerenciamento de dados, utilizando o SQL em dados estruturados e o JSON em dados semi-estruturados. Como resultado, oferece benefícios aos desenvolvedores de aplicativos, porque permite o gerenciamento de dados estruturados e semi-estruturados (LIU et al., 2014).

Desenvolvimento de uma ferramenta para extração de dados não estruturados via Web, para formação de Big Data de dados semiestruturados com a finalidade de ajudar no processo de tomada de decisão. Criação de um robô de busca dos dados por ontologia realizada pelo usuário. Os resultados foram satisfatórios na extração dos dados em relação à velocidade de busca e relevância das informações armazenadas (FURTADO et al., 2015).

2.3 Dados Não Estruturados

Dados não estruturados são informações que não contém um modelo predefinido. Sendo compostas de texto, datas, números, fatos que dificultam o entendimento destas informações através de programas tradicionais. Técnicas como mineração de dados, processamento de linguagem natural e análise de texto resultam em métodos para encontrar padrões e interpretar informações (LYNCH et al., 2010).

Temos que 80% dos dados referentes a negócios são dados não estruturados, da qual não contém uma estrutura interna identificável. A aplicação de mineração de dados como ferramenta computacional, utilizando métodos como finalidade a conversão ou transformação destas informações em uma forma compreensível para uso futuro (WAGH; GONDANE; DUKARE, 2016).

As informações contidas em Big Data são do tipo estruturadas e não estruturadas, porém quando analisadas são indispensáveis para as áreas de negócios. Aplicação de tecnologias BigInsights administração, desenvolvimento, visualização e para extração de *insights* se tornam intuitivas e fáceis. Resultando para organização vantagens competitivas (BORLONE, 2015).

Spangler (2018), relata um processo de três fases mais simplificado para aplicar a mineração de dados em dados não estruturados. As fases são explorar, entender e analisar, aproveitando os diferentes recursos que foram construídos uns sobre os outros. Também há necessidade da utilização de um analista como uma assistência algorítmica para interpretar e analisar os dados obtidos.

2.4 Soluções Baseadas no Método *Complearn*

Aplicando as técnicas Complexidade de Kolmogorov é possível realizar a análise de worms da internet, através de agrupamentos de várias espécies e tráfego de rede, através de agrupamentos de tipos de tráfegos. Para obtermos resultados satisfatórios de agrupamentos de worms devem ser compactados com UPX e a compres-

são foi embaralhada. Em relação ao tráfego de rede podemos utilizar como uma ferramenta eficaz para distinguir entre tráfego legal e ilegal na mesma porta, porém teremos problemas quando um invasor consegue alterar o tráfego na rede (WEHNER, 2018).

MANUEL (2015), em sua tese de mestrado para analisar os traçados de batimento cardíaco fetal por características biológicas, foi aplicado o algoritmo em dados pelo pH do cordão umbilical e da variabilidade curta dos traçados. Em relação aos resultados obtidos não houve relação entre o pH, porém para variabilidade curta foram separados com sucesso.

Neste trabalho foi proposto uma nova métrica para detecção de plágio, baseada na complexidade de Kolmogorov e utilizando um algoritmo de compressão heurística, da qual foi denominado Software Integrity Diagnosis(SID). Os resultados demonstraram vantagens quando comparados a outros detectores de plágio. O servidor do sistema SID está disponível online (CHEN et al., 2004).

Para reconhecer e classificar sequências de caracteres de qualquer tipo como sequências de tempo, linguagem, sequências genéticas (DNA, proteínas, etc), o método apresentado utiliza o conceito de entropia relativa, aplicando a compressão de dados e pelo cálculo das distancias relativas, fazendo o agrupamento em uma estrutura hierárquica. Sendo aplicado em problemas motivados por linguística, apresentando resultados altamente precisos para reconhecimento de linguagem, atribuição de autoria e classificação de idiomas (BENEDETTO; CAGLIOTI; LORETO, 2002).

2.5 Soluções Baseadas nos Métodos *Damicore*

A abordagem de mineração de dados denominada Damicore, quando aplicada para fusão de kernels em software, resultou em núcleos FPGA inferiores ao geral de hardware. Dos 11 códigos utilizados as reduções alcançaram 7,45% em LUTs e 35,39% em Registros (SANCHES; CARDOSO; DELBEM, 2011).

Segundo Moro (2014), foi aplicado à abordagem Damicore para realizar a análise dos dados de um sistema tutor educacional inteligente (MathTutor), com a intenção de identificar padrões de comportamento dos alunos. Apresentou resultados positivos, porém se torna necessário validar com um grupo maior de alunos.

Apresentação de uma ferramenta computacional denominada de Damicore, que realiza tarefas de classificação e agrupamento em conjuntos de dados de forma livre de configuração. Da qual apresentou resultados satisfatórios (CESAR, 2016).

Sumoyama (2016), relata que a utilização da ferramenta computacional Damicore para classificação de kernels é uma opção bastante interessante em arquiteturas híbridas, da qual apresentou resultados satisfatórios.

De acordo Silva (2016), para otimizar desempenho e consumo de energia em arquiteturas multi-cores heterogêneas em FPGA, podemos aplicar o Damicore para realizar análise offline do código fonte.

Apresentação de uma nova técnica através da abordagem Damicore para geradores síncronos distribuídos, tendo como finalidade a formulação de um novo algoritmo de decisão para proteção anti-ilhamento. Constatou-se que a melhor solução para proteção do sistema contra o ilhamento, é a utilização de duas proteções operando com redundância, sendo data-mining de 5 ciclos e um relé de frequência ajustado (GOMES, 2016).

Desenvolvimento de uma técnica nova denominada Binary Code Clustering (BCC), que sintetiza procedimentos para análise de códigos binários de programas executáveis, utilizando a abordagem Damicore com o objetivo de traçar o perfil de consumo. Os resultados experimentais mostraram-se satisfatórios (PINTO, 2017).

Aplicação da ferramenta computacional Damicore em sistemas de distribuição de energia, como suporte à decisão referente a estimadores de distância com a finalidade de reduzir o efeito de múltiplas estimativas de algoritmos de localização de falhas. Os resultados confirmam o potencial da ferramenta em minerar dados complexos e reconhecer padrões, reduzindo 91,95% casos de múltiplas estimativas para falhas, com isso aumentando a confiabilidade energética (RECHE et al., 2018)

MATERIAIS E MÉTODOS

3. MATERIAIS E MÉTODOS

3.1 Ferramenta Computacional *Damicore*

Nesta seção vamos abordar de forma sucinta a teoria sobre a ferramenta computacional *Damicore*, da qual será aplicada neste trabalho.

A ferramenta computacional *Damicore* é composta pelo funcionamento de três algoritmos, o *Normalized Compression Distance (NCD)*, *Neighbor Joining (NJ)* e *Fast Newman (FN)*, sendo desenvolvida com a finalidade de realizar a classificação e agrupamento de informações ou dados. Trabalha com diferentes tipos de representações de código, a entrada de informações pode ser de qualquer tipo de dados (SANCHES; CARDOSO; DELBEM, 2011).

Conforme CESAR (2016), descreve de forma completa toda a teoria sobre a ferramenta computacional *Damicore*, inclusive a descrição do funcionamento e aplicações dos três algoritmos, sendo:

- 1º Passo: o cálculo da matriz de distância entre instâncias de um conjunto de dados, utilizando o algoritmo *Normalized Compression Distance (NCD)*.

- 2º Passo: após o cálculo da matriz de distância, será transformado em uma árvore sem raiz, tendo as instâncias como folhas e o comprimento das arestas demonstram a grau de similaridade, utilizando o algoritmo *Neighbor Joining(NJ)*.
- 3º Passo: pela técnica de detecção de comunidades, obtemos um agrupamento hierárquico das instâncias, utilizando o algoritmo *Fast Newman(FN)*.

3.2 Bibliotecas / Ferramentas

Para a execução da ferramenta computacional Damicore, tem que utilizar o sistema operacional *Linux*. Segundo (CESAR, 2016) as bibliotecas utilizadas possuem interfaces por linha de comando, sendo utilizadas como *scripts* de *shell* ou pelo interpretador de *Python*. As ferramentas utilizadas neste trabalho são:

- *ncd2.py*, realiza o cálculo *Normalized Compression Distance (NCD)*.
- *tree_simplification.py*, implementa o *Neighbor Joining(NJ)*.
- *clustering.py*, implementa *Fast Newman(FN)*.
- *partition.py*, calcula índices de validação para agrupamentos gerados.
- *dataset.py*, Executa pré-processamentos em conjunto de dados.
- *classification.py*, realiza a classificação.
- *damicore.py*, executa todas as funções como métrica, simplificação, agrupamento e validação.

3.3 Programa Damicore.py

Para executar o *download* da ferramenta computacional *Damicore.py*¹, que foi utilizada neste trabalho para realizar a classificação e agrupamento de dados.

3.4 Máquina Virtual

A ferramenta computacional *Damicore.py* é um aplicativo que deve ser executado no sistema operacional *Linux*.

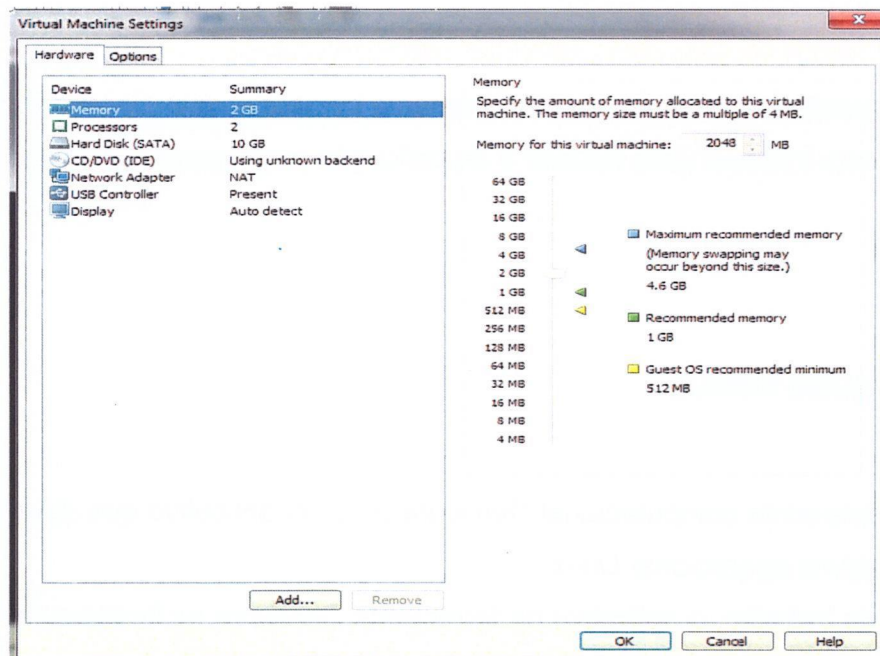
Neste trabalho, o aplicativo da ferramenta *Damicore.py* foi executado por uma máquina virtual *VMware Workstation Pro 14.1.3*² com as especificações descritas abaixo ou ilustradas na Figura 1 e na Figura 2.

- Modelo *VMware Workstation Pro 14.1.3*
- Sistema operacional *Linux*
- Versão *Ubuntu 64 bit*
- Memória 2GB
- *Processors 2*
- *Hard Disk (SATA) 10GB*

¹ O programa *Damicore.py* esta disponível em <https://gitlab.com/adaptsys/damicorepy.git> ou <https://gitlab.com/vhcandido/damicorepy> ou <https://github.com/sidgleyandrade/damicore-python> <https://github.com/brenocaetano/damicore>. Acesso em 09/08/2018.

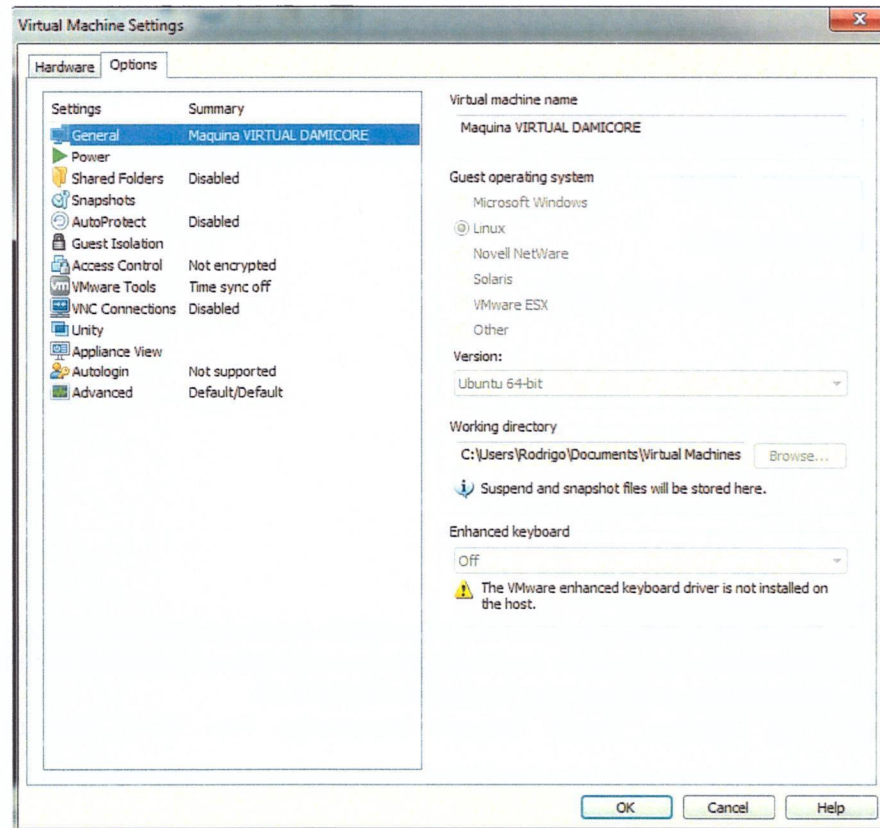
² O programa da Máquina Virtual *VMware Workstation Pro 14.1.3* está disponível em <https://www.vmware.com/products/workstation-pro/workstation-pro-evaluation.html>. Acesso em 16/08/2018.

Figura 1 - Especificações da Máquina Virtual



Fonte: Autor

Figura 2 - Especificações da Máquina Virtual



Fonte: Autor

3.5 Dados de Eventos em Sistemas de Energia Elétrica

Para este trabalho foi disponibilizado dados com informações reais, não estruturado. Através de relatórios descritos pelos operadores de uma subestação de energia elétrica de distribuição. O nome da subestação precisou ser anonimizado. Neste trabalho foi denominada SBE.

As informações dos dados se referem a “Causa” ou motivos que resultaram à interrupção do fornecimento de energia elétrica nas redes de distribuição de energia dos consumidores.

O dados foram disponibilizados numa planilha em formato Excel.ods conforme ilustra na Figura 3, contendo sete atributos: SE, Alimentador, Descrição, Observações, Data Inicio, Data fim e Operador. Cada atributo contém 174 informações.

- SE: nome da Subestação de energia elétrica
- Alimentador: ponto de entrega do fornecimento de energia elétrica ao consumidor
- Descrição: causa ou motivo da interrupção no fornecimento da energia elétrica no consumidor.
- Observações: são os relatos da equipe de manutenção no local onde ocorreu a interrupção do fornecimento de energia.
- Data Inicio: data e hora do inicio da interrupção no fornecimento de energia elétrica
- Data fim: data e hora do retorno do fornecimento de energia elétrica ao consumidor
- Operador : são os operadores da subestação de energia elétrica

Figura 3 - Planilha do Relatório de Interrupções Descrito pela Subestação em Excel.ods








	A	B	C	D	E	F	G
1	SE	Alimentador	Descricao	Observacoes	Data Inicio	Data fim	Operador
2	SBE1	ALM1	CAUSA: ARVORE NA REDE AT NA AREAS	Equipe A3110	01/11/2006 18:35	01/11/2006 19:21	OP1
3	SBE1	ALM1	Causa: Isolador danificado por descarga	Equipe T3208	1/13/2006 9:18	1/13/2006 9:36	OP2
4	SBE1	ALM3	Retorno a configuracao normal 10:44h,	Equipe T3208	1/13/2006 10:39	1/13/2006 10:44	OP1
5	SBE1	ALIM2	Causa do desligamento: Acidente com e	Equipe C3830	1/17/2006 10:24	1/17/2006 10:31	OP3
6	SBE1	ALIM3	Causado por galhos tocando a rede 22:0	Equipe T3208	1/21/2006 18:06	1/21/2006 22:00	OP4
7	SBE1	ALIM1	Causado por descarga atmosferica 20:35	Equipe T3208	1/21/2006 18:40	1/21/2006 20:35	OP2
8	SBE1	ALIM4	CAUSA: Teste em um trafo novo, tinha u	Equipe T3208	1/23/2006 16:55	1/23/2006 17:05	OP3
9	SBE1	ALIM2	Causa falta de ca disjuntor foi rearmado	Equipe EP692	1/28/2006 2:09	1/28/2006 3:15	OP4
10	SBE1	ALIM4	Causa do desligameto: cabos rompidos	Equipes T3708	02/06/2006 17:02	02/06/2006 17:28	OP5

Fonte: SEB1

3.6 Estrutura dos Dados em Excel

- Criar planilhas individuais para cada atributo e suas informações em arquivo Excel, conforme Figura 1.
- Converter as planilhas e suas informações em arquivo.csv, conforme Figura 4.
- Excluir da planilha, toda a linha sem informação, conforme ilustra o exemplo da linha nº 6 da Figura 5.

Figura 4 - Planilhas e Conversão de Arquivo

	Alimentadores.csv	06/10/2016 22:35	Arquivo de Valore...	4 KB
	Data Hora Fim.csv	06/10/2016 22:36	Arquivo de Valore...	3 KB
	Data Hora inicio.csv	06/10/2016 22:36	Arquivo de Valore...	3 KB
	Descrição.csv	06/10/2016 22:37	Arquivo de Valore...	37 KB
	Observações.csv	06/10/2016 22:37	Arquivo de Valore...	20 KB
	Operador.csv	06/10/2016 22:38	Arquivo de Valore...	5 KB
	tudo.csv	30/08/2017 18:19	Arquivo de Valore...	37 KB

Fonte: Autor

Figura 5 - Exclusão da linha sem Informação na Planilha Alimentadores

	A	B	C
1	Alimentador		
2	ALM1		
3	ALM1		
4	ALM3		
5	ALIM2		
6			
7	ALIM1		
8	ALIM4		
9	ALIM2		
10	ALIM4		

	A	B	C
1	Alimentador		
2	ALM1		
3	ALM1		
4	ALM3		
5	ALIM2		
6	ALIM1		
7	ALIM4		
8	ALIM2		
9	ALIM4		
10	ALM3		

Fonte: Autor

3.7 Estrutura de Dados no Software Damicore

Como neste trabalho foi utilizada uma máquina virtual, ligar e acessar as pastas de arquivo de texto: Pasta pessoal, *damicorepy*, *damicore*, conforme Figura 6.

Dentro da pasta *damicore*, colocar as planilhas com suas informações em formato.csv e converter cada linha da planilha em um arquivo de formato de texto, conforme Figura 7.

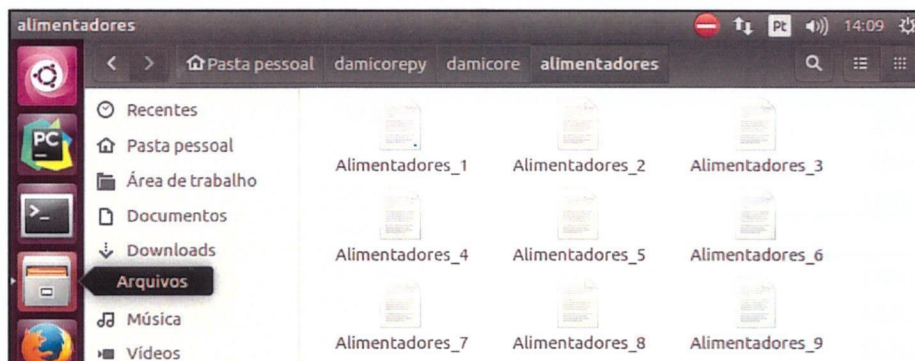
Figura 6 - Pasta de Arquivos de Texto



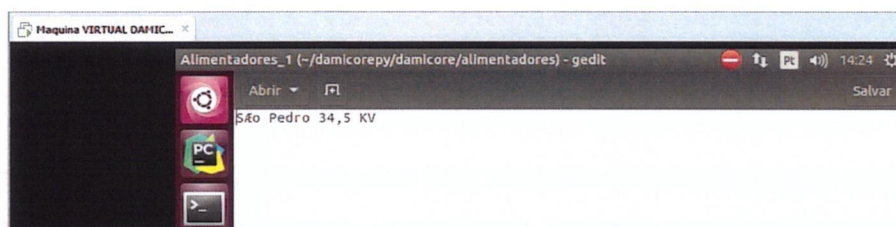
Fonte: Autor

Figura 7 - Arquivos de Texto

a) Conversão das Linhas da Planilha em Arquivos de Texto



b) Texto dentro do Arquivo Alimentadores_1



Fonte: Autor

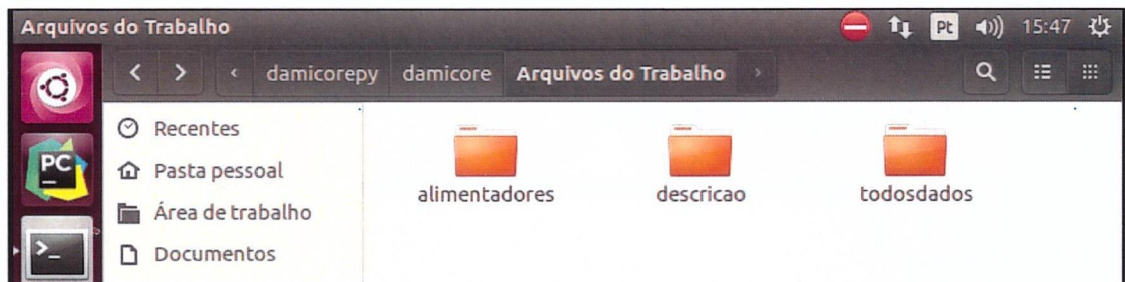
3.8 Criação das Pastas de Arquivos de Texto

Neste trabalho foram criadas as pastas de arquivos de texto para a execução no software *Damicore*, como demonstra a Figura 8.

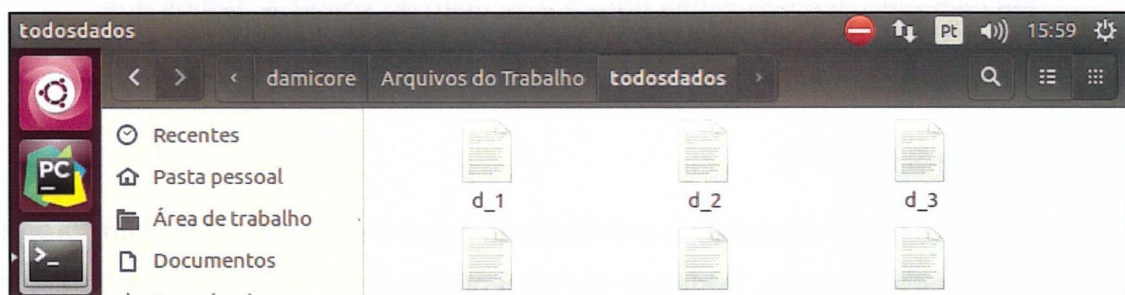
As pastas de arquivos de texto podem ser individuais com suas informações (alimentadores, datahorainicio, datahorafim, descrição, observações e operadores).

Também pode-se montar de qualquer maneira as pastas de arquivos, ou seja, colocar os arquivos importantes com suas informações.

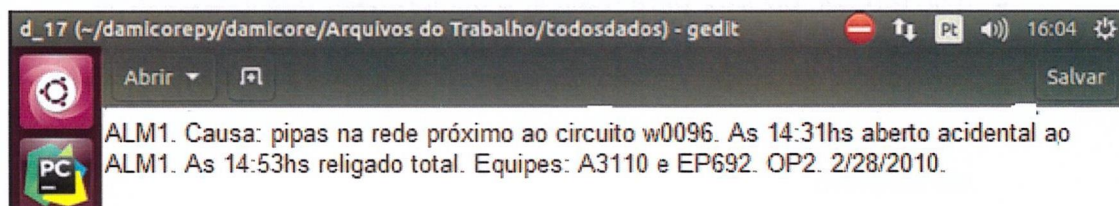
Figura 8 - Pastas de Arquivos de Texto

a) Pastas Utilizadas na Execução do *Damicore*

b) Arquivo de Texto da Pasta todosdados



c) Informações de Todas as Pastas de Arquivos de Texto



Fonte: Autor

3.9 Execução do Software *Damicore*

Neste trabalho para a execução do software *Damicore* foi utilizado o compressor *gzip* no comando, por ser rápido e amplamente indicado para realizar a compressão de arquivos de texto. Tipos e aplicações sobre compressores foram abordados de forma detalhada em (CESAR, 2016).

Para executar o software *Damicore*, seguir os passos descritos abaixo ou ilustrados na Figura 9:

- Dentro da pasta *Damicore*, clicar com botão esquerdo do mouse.
- Clicar, Abrir no terminal
- Digitar o comando: ponto ba

```
.\damicore.py descricao -c gzip -results-dir resultado15
```

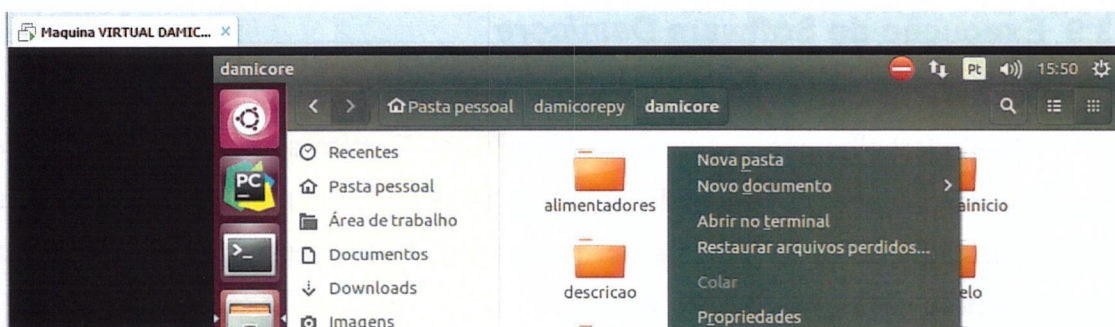
- Enter

No comando descrito acima para a execução do software, temos que:

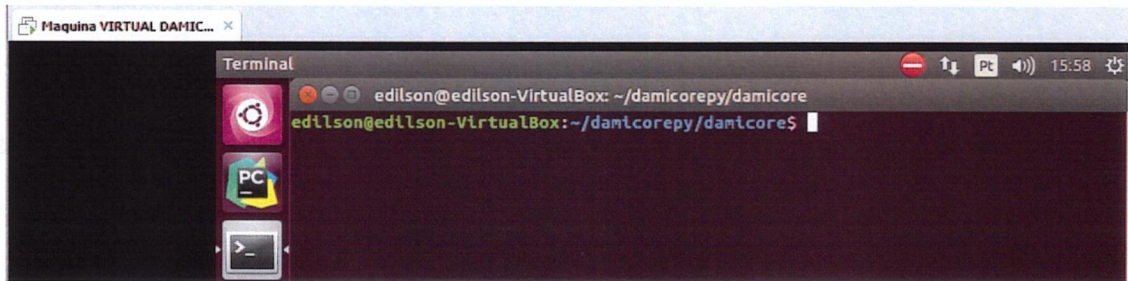
- Ponto
- Barra invertida
- No campo *descricao* é o arquivo de texto dentro da pasta *damicore* que deseja executar. Podendo ser *alimentadores*, *datahorafim*, *datahorainicio*, *observações* e *operador*, conforme Figura 6.
- *gzip* é o compressor para realizar a compressão das informações.
- No campo *resultado15*, foi dado este nome para a pasta que foi criada dentro da pasta *damicore*, onde serão colocados os resultados da execução do software.

Figura 9 - Terminal

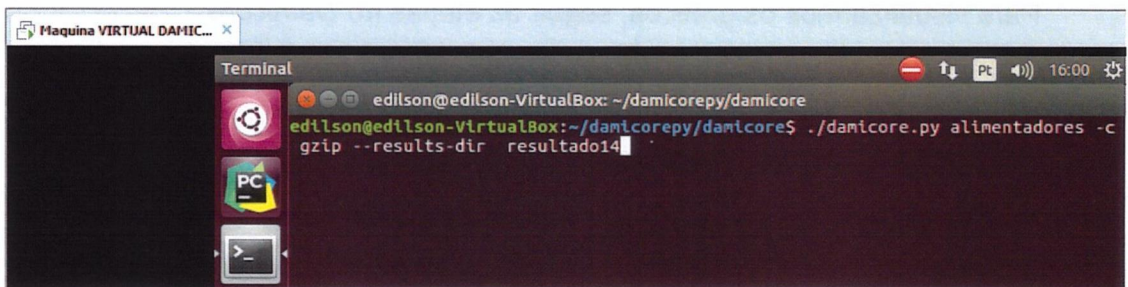
a) Pasta Damicore para Abrir Terminal



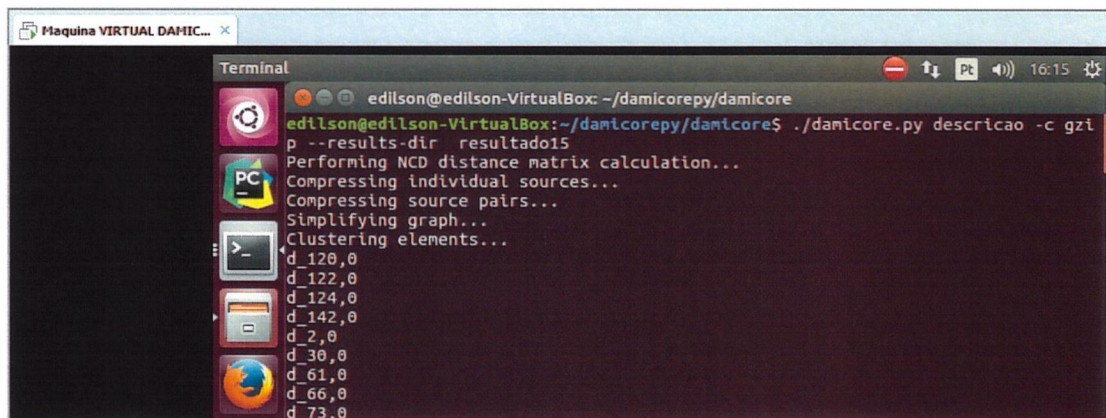
b) Terminal para Digitar Comando para Execução do Software Damicore



c) Descrição do Comando para Execução do Damicore



d) Inicio da Execução do Damicore



e) Fim da Execução do Damicore



Fonte: Autor

3.10 Geração de Saídas (Gráficos)

Após o processamento dos arquivos de textos contidos nas pastas, o *Damicore* resulta na classificação e agrupamento das informações como saídas em forma de gráficos, conforme descritos abaixo ou ilustrados na Figura 10.

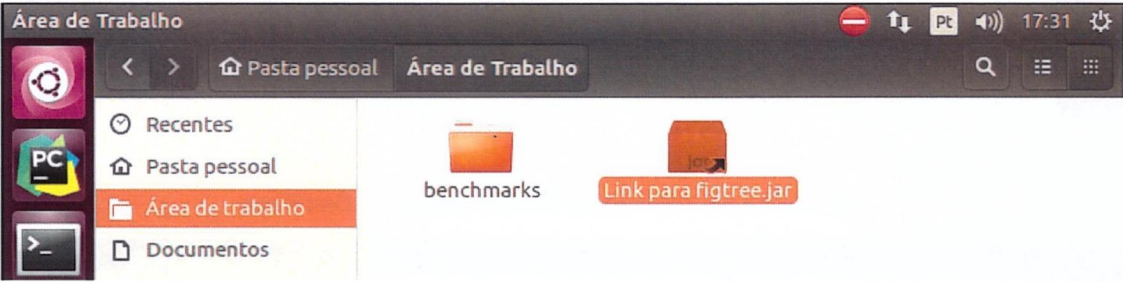
Neste trabalho, o gráfico da Figura 10 do item i é o *Trees Transform branches*, sendo utilizado para visualizar os agrupamentos da saída do *Damicore*. Para obter as informações dos agrupamentos, utilizar o gráfico da Figura 10 item h.

Para visualizarmos os gráficos, seguir as etapas no *Damicore*:

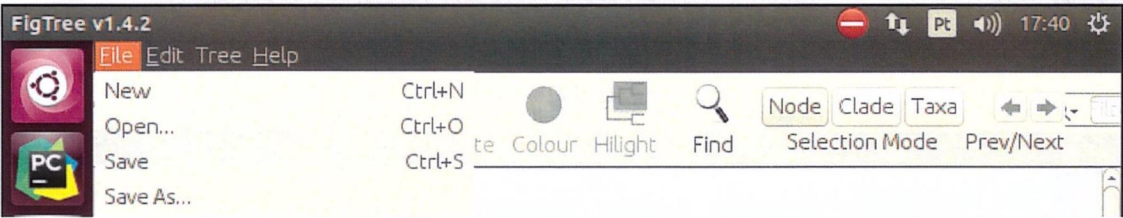
- Acessar Área de Trabalho
- Clicar em *link para figtree.jar*
- *File*
- *Open*
- *damicorepy*
- *damicore*
- pasta que nomeou no final do comando para execução, como exemplo a pasta resultado15 conforme ilustra na Figura 9 d.
- nome da pasta que executou o comando, como exemplo a pasta descricao conforme ilustra na Figura 9 d.
- *001-tree newick*
- *Trees*
- *Transform branches*
- Selecionar o formato do gráfico entre três modelos

Figura 10 - Visualização das Saídas (Gráficos)

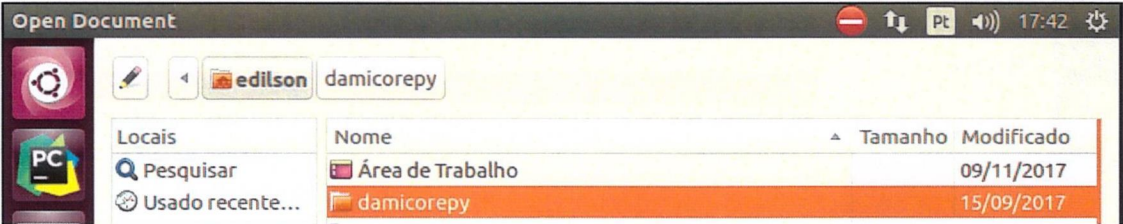
a) Área de trabalho e Link para figtree.jar



b) Pasta File



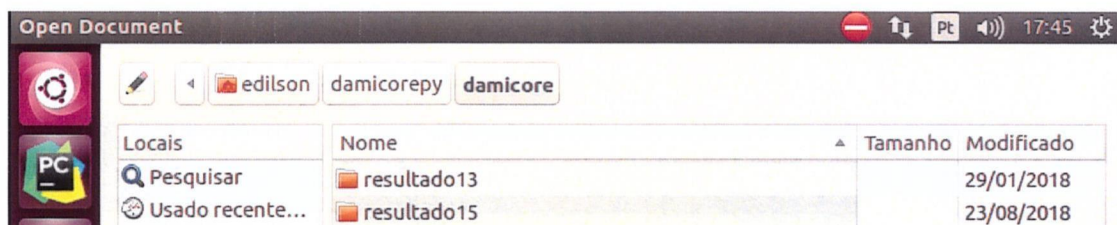
c) Pasta Damicorepy



d) Pasta Damicore



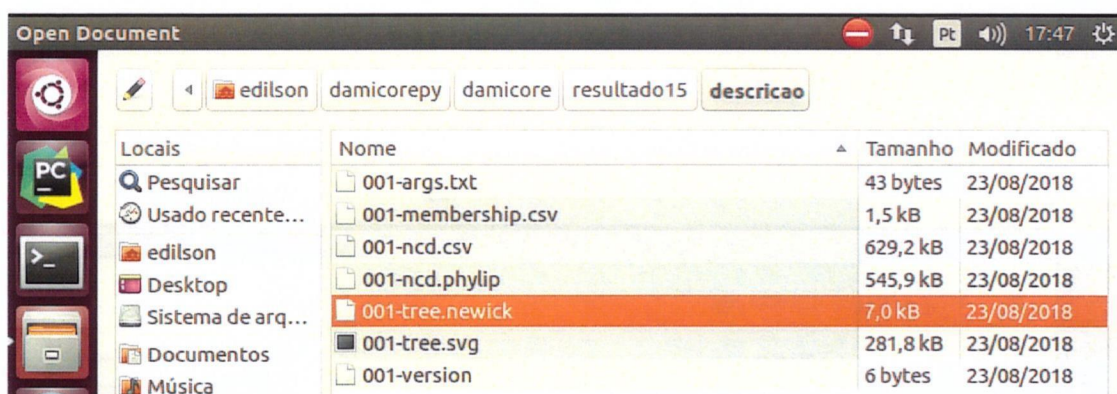
e) Pasta Onde Gerado as Saídas (Gráfico)



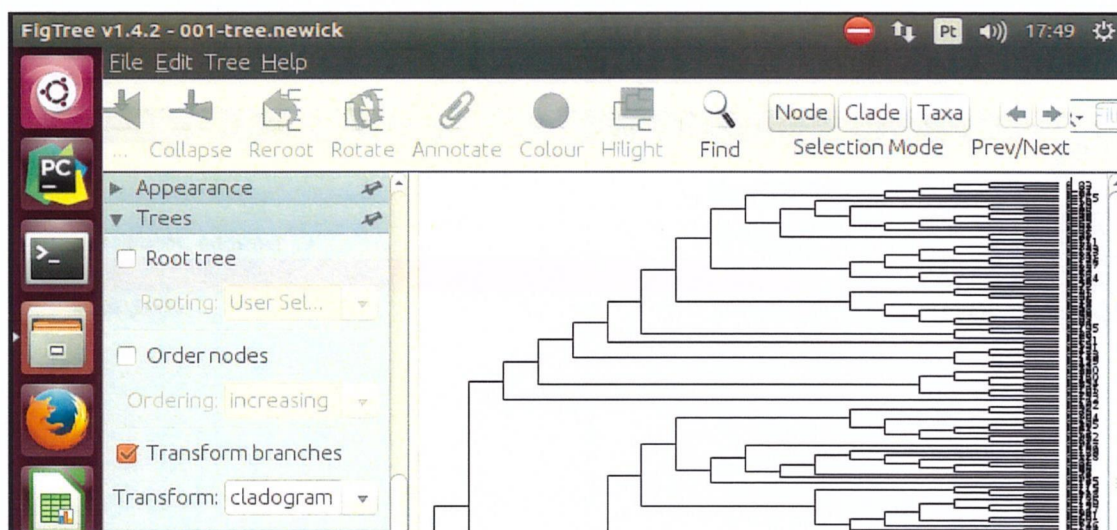
f) Pasta de Arquivos de Texto Usada pelo Damicore

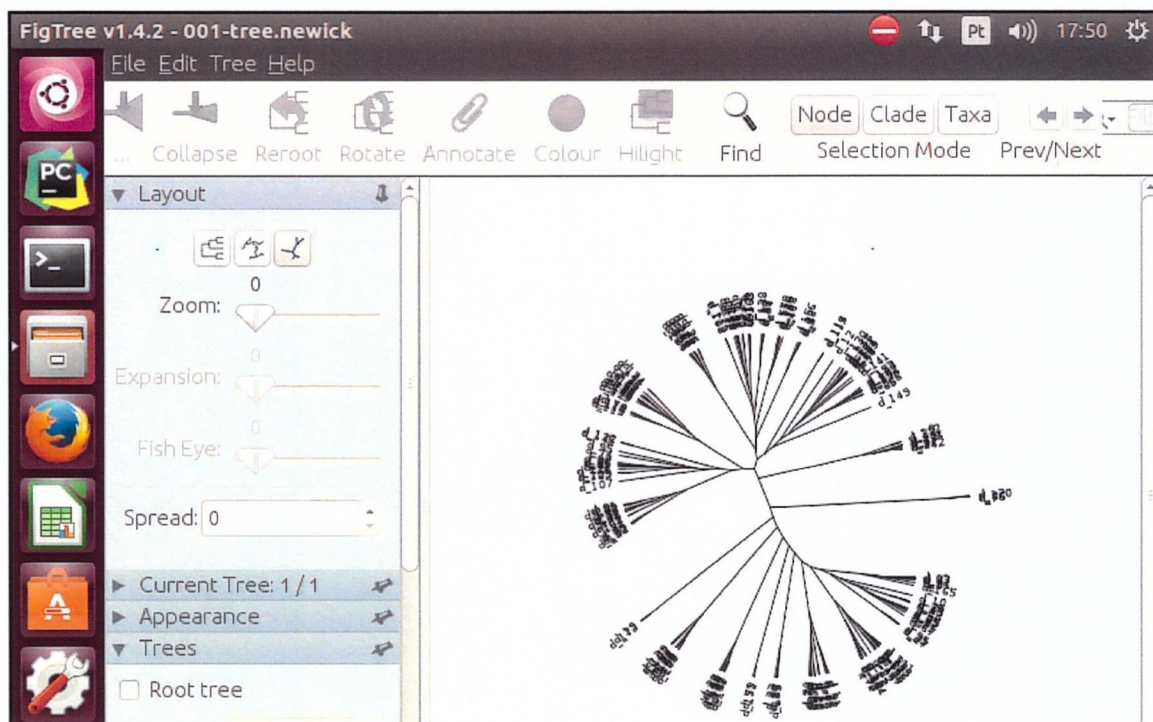


g) Pasta 0010-Tree.newick



h) Tipo do Gráfico Trees Transform Branches



i) Gráfico *Trees Transform Branches*

Fonte: Autor

PROPOSTA

4. PROPOSTA

4.1 Dados Estruturados

Como já foi citado, os dados fornecidos pela subestação de energia da concessionária, são descritos em forma de relatórios pelos operadores da subestação ou pela equipe de plantão, utilizando sua própria linguagem e termos técnicos. Com isso os dados estão na forma não estruturada, não padronizada, possuindo muitos termos e palavras repetidas.

A proposta de estruturar os a dados é criar um programa contendo opções de termos, frases ou palavras padronizadas sobre as subestações(SE), Alimentadores, Descrições, Data, hora início, hora fim, ano, operadores e observações. Onde os operadores e as equipes de manutenção escolhem as opções dos termos técnicos padronizados definidos.

Com isso teremos dados padronizados ou seja iguais, quando aplicados pelo *Damicore* para gerar agrupamentos, resultará em excelentes eficiências de agrupamentos, gerando informações para elaboração de relatórios gerenciais para tomada de decisão.

4.2 Estruturação de Dados

Para exemplificar uma estruturação e padronização dos dados, vamos utilizar os mesmos termos que foram utilizados nos dados não estruturados, fornecidos pela subestação.

4.2.1 Termos Técnicos Padronizados

Segue abaixo alguns exemplos de termos técnicos padronizados da descrição, que foram utilizados na aplicação do *Damicore*, considerando os mesmos conceitos dos dados reais fornecidos.

- Cabo AT Rompido por Curto-Circuito
- Cabo AT Rompido por Arvore ou Galhos
- Cabo AT Rompido por Queda de Poste
- Desligamento por Eucalipto ou Galhos Tocando na Rede
- Desligamento por Gambiarra na Rede
- Feito Inspeção e Nada foi Encontrado
- Desligamento por Descarga Atmosférica
- Desligamento por Isolador Danificado
- Desligamento por Pipa na Rede
- Desligamento por Vendaval
- Desligamento por Programado
- Desligamento Acidental Cliente Derrubou Poste
- Desligamento Acidental ao Fechar a Chave

4.2.2 Planilha de Dados Estruturados e Padronizados

Após a definição dos termos técnicos padronizados, foram adicionados em uma planilha de Excel.ods conforme Figura 11. Portanto foi criado uma pasta de texto com o nome estruturado. Para aplicação desta pasta no *Damicore* fazer todas as etapas de 3.6 a 3.10.

Figura 11 – Planilha Proposta de Dados Estruturados Padronizada.ods

	A	B	C	D	E	F	G	H
1	SE	Alimentador	Descricao	Observacoes	Inicio	Fim	Operador	Ano
2	SB1	ALM1	Cabo At Rompido por Curto-Circuito	Equipe A3110	18:35	19:36	OP1	2016
3	SB1	ALM1	Desligamento por Descarga Atmosferica	Equipe T3208	09:18	09:36	OP1	2016
4	SB1	ALM3	Feito Inspecao e Nada foi Encontrado	Equipe T3208 e A3110	10:39	10:49	OP4	2016
5	SB1	ALM2	Desligamento por Gambiarra na Rede	Equipe PLSQC	10:24	10:31	OP2	2016
6	SB1	ALM2	Desligamento por Descarga Atmosferica	Equipe EP692	18:06	22:00	OP2	2016
7	SB1	ALM1	Desligamento por pipa na Rede	Equipe T3208	18:40	20:35	OP3	2016
8	SB1	ALM4	Cabo AT Rompido por Arvore ou Galhos	Equipe T3207	16:55	17:05	OP1	2016
9	SB1	ALM2	Desligamento por Vendaval	Equipe A3110	02:09	03:15	OP4	2016
10	SB1	ALM4	Cabo AT Rompido por Arvore ou Galhos	Equipe EP692	01:34	03:35	OP3	2016

Fonte: Autor

RESULTADOS

5. RESULTADOS

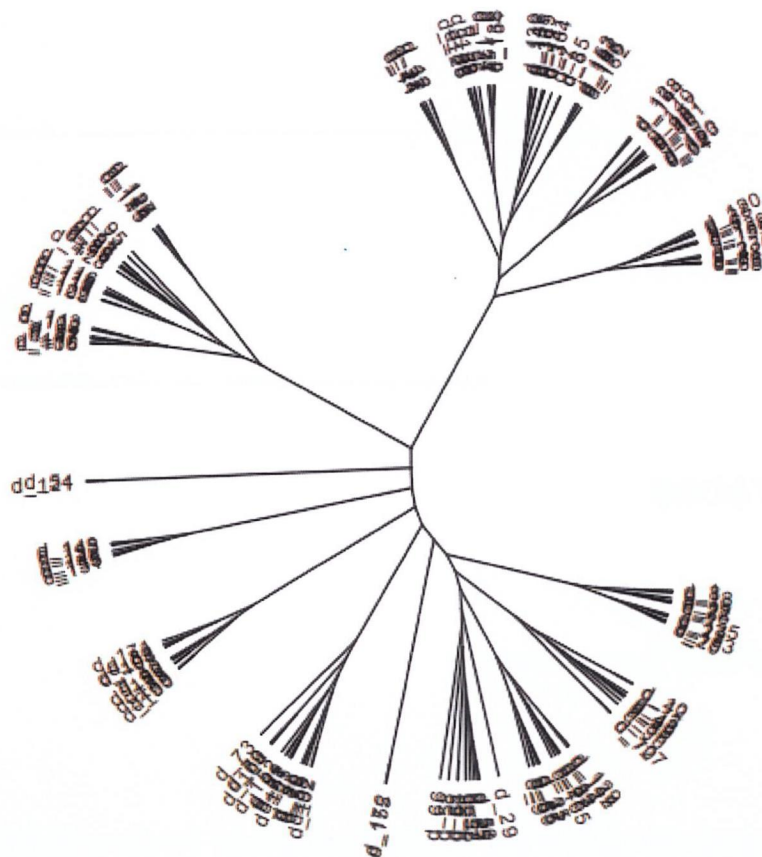
5.1 Dados Não Estruturados

5.1.1 Pasta de Arquivo de Texto Completa (todosdados)

Para executarmos o *Damicore* foi criado uma única pasta de arquivo de texto com o nome todosdados, onde foi adicionado todas as informações contidas nas pastas alimentadores, descrição, data hora inicio, data hora fim, operador e observações conforme ilustra a Figura 8 Figura 8 - Pastas de Arquivos de Texto, item a. Não foi adicionado à pasta SE(subestação), pois todas as informações de texto utilizadas são desta subestação de energia.

Após a execução do *Damicore*, resultou uma saída em forma de gráfico, ilustrando o agrupamento e classificação conforme Figura 12.

Realizando a análise destas informações, verificamos agrupamentos importantes que podem ser utilizados para tomada de decisão gerencial em relação a “causa” que provocou a interrupção no fornecimento de energia elétrica ao consumidor.

Figura 12 - Gráfico *Trees Transform Branches Todosdados*

Fonte: Autor

5.1.1.1 Agrupamentos Obtidos

- Cabo AT Rompido
- Eucalipto ou Galhos Tocando na Rede
- Gamba(Gambiarra) na Rede
- Feito Inspeção e Nada foi Encontrado
- Descarga Atmosférica
- Isolador Danificado
- Pipa na Rede
- Vendaval
- Para-raio Estourado
- Desligamento Voluntário

- Religamento Bloqueado
- Nada Encontrado
- Desconhecida
- Não Identificada
- Poste Danificado
- Fechamento da Chave
- Curto Banco de Capacitor
- Atuação ERAC
- Acidente com a Equipe Copel
- Jumper Rompido na Saída RA
- Amarração Rompeu
- Cliente Derrubou Eucalipto na Rede AT
- Cruzeta Danificada
- Teste Transformador Novo

5.1.1.2 Agrupamentos Relevantes

- Cabo AT Rompido
- Eucalipto ou Galhos Tocando na Rede
- Gamba(Gambiarra) na Rede
- Feito Inspeção e Nada foi Encontrado
- Descarga Atmosférica
- Isolador Danificado
- Pipa na Rede
- Vendaval

5.1.1.3 Exemplos de Saídas do *Damicore*

Segue abaixo exemplos de saídas do *Damicore* de agrupamentos e classificação das informações dos arquivos textos, demonstradas pelo gráfico da Figura 12.

Verificamos que os agrupamentos e classificação são realizados por similaridades ou igualdades das palavras contidas nos textos, conforme destacadas em vermelho na Figura 13.

Figura 13 – Exemplos de Saídas do *Damicore de Dados Não Estruturados*

a) Agrupamento Similar: Gamba

Causa: gamba na rede na entrada da CF 94673 - 21h32 - abertura accidental al. ALM1 - 22h22 - fechado al. ate a chave 3573 - 22h25 - aberto al. ALM1 a pedido do Equipe pl/wbz que avistou fogo no trecho anterior a chave 3573 - 22h34 - fechado al. ate a chave 3573 segundo pl/wbz nada encontrado no local. - 22h55 - fechado al. ate a chave 2343 - 23h11-aberto o al ALM1 p/fechamento da chave 2343.equipe identificou o defeito -23h 27-fechada a chave 2343 religando total

Causa: Encontrado gamba na estrutura. 03:00h Desligado ALM1 04:02h religado apos a SU 3573 por Calogeras, o defeito no primeiro trecho. Feito varios testes sem sucesso, inclusive com a cf 90763 aberta e nao aceitou. Equipe pl/wbz vai colocar manutencao no trecho ao amanhecer. As 11:05 desligado o RA SEM09 para fechar a chave SU 3573 As 11:06 religado o Al. ALM1.

b) Agrupamento Diferente:

Feito Inspeção e Nada Encontrado x Galhos Tocando a Rede AT

Causa : feito inspeção e nada encontrado As 15:02hs aberto ra 52-25 desligado AL. ALM3 As 15:31hs feito inspecao na urbana de WBZ e SBV feito teste sem sucesso. As 15:41hs aberta a SU 2620 feito teste sem sucesso As 15:55hs fechado o AL/BRM com a SU 3516 aberta. (59 Uc's religadas) As 16:13hs fechada a o trecho SU 1992 energizando urbano de SBV pelo -AL/BVT ate a SU 2620 Aberta (300 Uc's religadas) As 16:23hs fechado a su 3516 energizando ate a su 2619 aberta.(162 UC'S religadas) As 17:04hs fechado a su 2619 energizando ate a su 3053 aberta.(61uc's religadas) As 17:26hs fechado a su 3056 energizando ate a cf 5201 aberta. As 17:34hs fechado a cf 5201 energizando total. Retornar alimentado a configuracao normal.

Causa: No Al. ALM2, existem galhos de arvores tocando na rede de AT alguns poste antes da SU 3573. Estes galhos de arvores esta sobre a rede mas nao estavam tocando, porem com a chuva as folhas pesaram e caíram sobre a rede de AT. Desligou o Al. ALM2 e ao mesmo tempo o SQ015 (WBZ). 09:00 Hs:- Desligado o Al. SBE1 09:18 Hs:- Religado a SE/WBZ pelo Al. SIT ficando desligado apenas o Al. ALM2. 09:34 Hs:- Religado o Al. WBZ ate a SU 2344 10:17 Hs:- Religado o Al. ALM2 por Calogeras 10:35 Hs:- Desligado o RA 52-09 na SE Calogeras 11:26 Hs:- Religado o trecho da SU 2344 11:50 Hs:- Retornado a SE/WBZ em configuracao normal 12:30 Hs:- Religado o Al. ALM2 energizando total o trecho.

c) Agrupamentos Similares: Nada Encontrado

Causa nada encontrado. As 15:33 abriu. As 16:08 fechado.
Causa nada encontrado. As 16:16 abriu. As 16:35 fechado.
Causa nada encontrado. As 05:26 abriu. as 0551 fechado.
Causa nada encontrado. AS 12:45 abriu. As 13:05 fechado.
Causa nada encontrado. As 14:40 abriu. As 15:09 fechou.
Causa nada encontrado. Abriu as 20:30. Fechado as 20:42.

d) Agrupamento Similares: Descarga Atmosférica

Causa do por descarga atmosfericas 20:35h, religado total..
Causa do desligamento : Descarga atmosferica. As 22:25 h religado total.

e) Agrupamento Diferente:

(Cabo Rompido x Isolador Danificado x Descarga Atmosférica)

<p>ALM1. Causa: Cabo rompido lado fonte e lado carga e lado fonte na tronco proximo a CF 70016 16:13 Hs:- Desligado o Al. ALM1 16:38 Hs:- Religado o Al. ALM1 com a LA 2103 aberta (29 consumidores) 18:04 Hs:- Fechado a LA 2103 energizando o trecho total. Equipe 3208 sendo que motivo da demora: Equipes estava na inspecao do Alimentador ALM3 que estava desligado. Equipes que participaram da manobra: A3111 e T3207. Leitura final do Ra de Trecho: 1252, nao houve nenhuma tentativa de religamento sem sucesso. OP5.</p>
<p>ALM3. Causa: Isolador danificado devido a descarga atmosferica dois postes antes da LA 3668 16:02 Hs:- Desligado o Al. Ceu Azul 16:53 Hs:- Religado o Al. Ceu Azul com a LA 2702 aberta (863 consumidores) 17:32 Hs:- Fechado a LA 2702 energizando o trecho total. Equipe 3208. Foi o motivo da demora: Equipes A3110 e EP692 estavam no ALM5 que estava desligado. Equipes que participaram da manobra: A3111 e T3207. Leitura final do Ra de Trecho: 2464, houve tentativa de religamento sem sucesso OP3.</p>

f) Agrupamentos Similares: Feito Inspeção e nada Encontrado

Causa do desligamento acidental: feito inspecao e nao encontrado. CAUSA: 92-82. 17:28Hr - Desligado acidentalmente; 18:43Hr - Religado total.
Causa do desligamento acidental: feito inspecao e nada encontrado. 13:17Hr - Desligado acidental 13:29Hr - Religado total e na sua configuracao normal de fonte.
Causa do desligamento acidental: feito inspecao e nada foi encontrado. 18:17Hr - Desligado acidentalmente; 18:26Hr - Religado total e na sua configuracao normal.
Causa do desligamento acidental: feito inspecao e nada foi encontrado. 18:34Hr - Desligado acidentalmente; 19:11Hr - Religado total e na sua configuracao normal de fonte.
Causa do desligamento acidental: feito inspecao e nada foi encontrado. 11:56Hr - Desligado acidentalmente; 13:37Hr - Religado total e na sua configuracao normal de fonte;

g) Agrupamentos Similares: Vendaval

ALM2. Causa: vendaval as: 21:00h aberto as: 21:39h fechado. OP3. Equipe T3208 - A3110 – EP692.
ALM2. Causa : vendaval / Condutor AT as: 22:34h aberto as: 23:05h fechado. A3110 – C3830.

Fonte: Autor

5.1.1.4 Saídas do *Damicore* da Pasta Todosdados

Segue abaixo a Tabela 1 com os valores numéricos de agrupamentos e classificação da qual foram obtidos pela saída do software *Damicore* conforme a Figura 12.

Foram selecionados para análise os agrupamentos relevantes (item 4.1.1.1), da qual possuem informações importantes que provocaram a interrupção no fornecimento de energia elétrica e para tomada de decisão gerencial.

Tabela 1 – Valores de Agrupamentos x Não Agrupamentos Todos dados

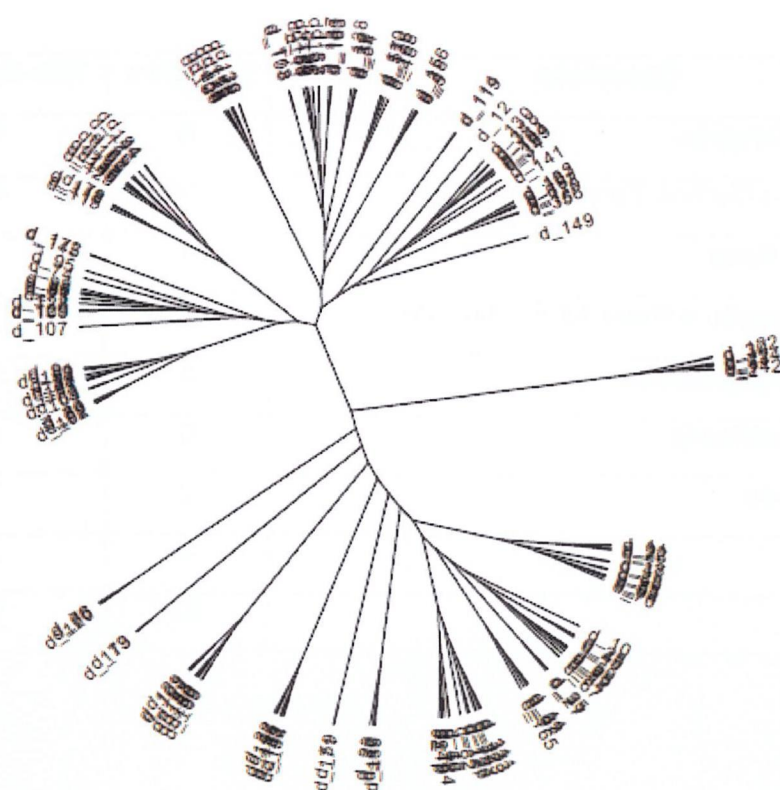
Descrição	Agrupou	Não Agrupou
Cabo AT Rompido	8	10
Eucalipto ou Galhos Tocando na Rede	14	16
Gamba na Rede	1	2
Feito a inspeção e Nada foi Encontrado	11	5
Descarga Atmosférica	5	10
Isolador Danificado	8	8
Pipa na Rede	2	5
Vendaval	4	2
Total	53	58

Fonte: Autor

5.1.2 Pasta de Arquivo de Texto(descricao)

Para executarmos o *Damicore* foi criado uma única pasta de arquivo de texto com o nome *descricao*, onde foi adicionado todas as informações contida na pasta *descrição* conforme ilustra a Figura 8 Figura 8 - Pastas de Arquivos de Texto, item a.

Após a execução do *Damicore*, resultou uma saída em forma de gráfico, ilustrando o agrupamento e classificação conforme Figura 14.

Figura 14 - Gráfico *Trees Transform Branches* descrica

Fonte: Auto

5.1.2.1 Saídas do *Damicore* da Pasta Descricao

Segue abaixo a Tabela 2 com os valores numéricos de agrupamentos e classificação da qual foram obtidos pela saída do software *Damicore* conforme a Figura 14.

Foram selecionados para análise os agrupamentos relevantes (item 4.1.1.1), da qual possuem informações importantes que provocaram “causa” na interrupção do fornecimento de energia elétrica e para tomada de decisão gerencial.

Tabela 2 - Valores de Agrupamentos x Não Agrupamentos Descricao

Descrição	Agrupou	Não Agrupou
Cabo AT Rompido	11	7
Eucalipto ou Galhos Tocando na Rede	18	12
Gamba na Rede	2	1
Feito a inspeção e Nada foi Encontrado	12	4
Descarga Atmosférica	6	9
Isolador Danificado	9	7
Pipa na Rede	3	4
Vendaval	5	1
Total	66	45

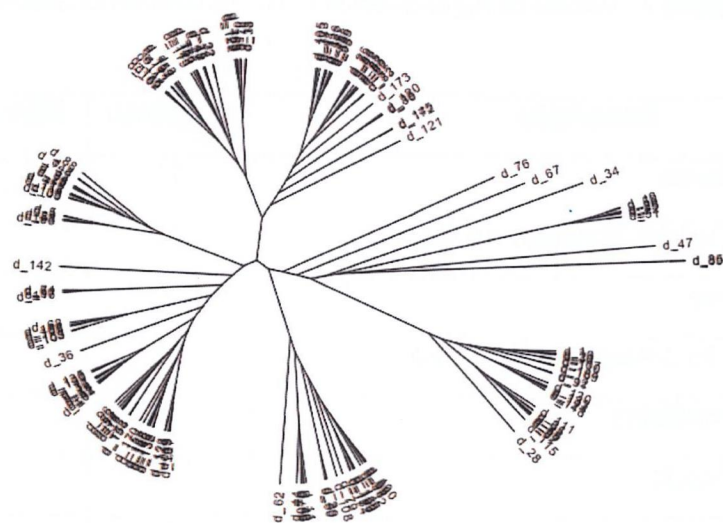
Fonte: Autor

5.2 Dados Estruturados

Para executarmos o *Damicore* foi criado uma única pasta de arquivo de texto com o nome estruturado, onde foi adicionado todas as informações dos alimentado- res, descrição, hora início, hora fim, ano, operador e observações conforme ilustra a Figura 11. Não foi adicionado à pasta SE(subestação), pois todas as informações uti- lizadas são desta subestação de energia e foi incluído o ano.

Após a execução do *Damicore*, resultou uma saída em forma de gráfico, ilus- trando o agrupamento e classificação conforme Figura 15. Realizando a análise des- tas informações, verificamos faixas de agrupamentos bem definidos, devido a estru- turação e padronização das informações dos dados.

Figura 15 - Gráfico *Trees Transform Branches Estruturado*



Fonte: Autor

5.2.1 Exemplos de Saídas do *Damicore*

Segue abaixo exemplos de saídas do *Damicore* de agrupamentos e classificação das informações dos arquivos textos, demonstradas pelo gráfico da Figura 16.

Pela figura abaixo podemos verificar o resultado de agrupamentos bem definidos, devido a estruturação e padronização dos dados.

Figura 16 - Exemplos de Saídas do *Damicore* de Dados Estruturados

a) Desligamento por vendaval

ALM1. Desligamento por vendaval. Equipe A3110 e C3830. 18:35 . 19:00. OP3. 2016.
ALM1. Desligamento por vendaval. Equipe A3110 e C3830. 16:16. 16:35. OP2. 2016.
ALM1. Desligamento por vendaval. Equipe A3110. 05:06. 05:51. OP2. 2016
ALM1. Desligamento por vendaval. Equipe T3208. 12:45. 13:05. OP5. 2016.
ALM1. Desligamento por vendaval. Equipe. EP692. 14:40. 15:09. OP5. 2016
ALM1. Desligamento por vendaval. Equipe T3207. 20:30. 20:42. OP1. 2016.

b) Desligamento por Descarga Atmosférica

ALM3. Desligamento por Descarga Atmosferica. Equipe T3207. 21:35. 22:00 . OP1. 2016.
ALM3. Desligamento por Descarga Atmosferica. Equipe T3207. 11:55. 13:00. OP1. 2016
ALM2. Desligamento por Descarga Atmosferica. Equipe T3207. 11:55 . 13:00 .OP2. 2016
ALM3. Desligamento por Descarga Atmosferica. Equipe C3830. 15:55 .16:20. OP2. 2016
ALM3. Desligamento por Descarga Atmosferica. Equipe C3830. 22:05 . 00:20. OP4. 2016.
ALM1. Desligamento por Descarga Atmosferica. Equipe PLSQC 17:02 . 18:20 . OP3. 2016.

c) Feito Inspeção e Nada foi Encontrado

ALM2. Feito Inspeção e Nada foi Encontrado. Equipe A3110. 18:35 . 19:00 . OP3. 2016.
ALM2. Feito Inspeção e Nada foi Encontrado. Equipe A3110. 14:35. 15:00 . OP3. 2016
ALM2. Feito Inspeção e Nada foi Encontrado. Equipe A3110. 23:35 . 00:10 . OP3. 2016
ALM2. Feito Inspeção e Nada foi Encontrado. Equipe A3110. 20:35 . 22:05 OP3. 2016
ALM5. Feito Inspeção e Nada foi Encontrado. Equipe C3830 18:10. 18:46. OP1. 2016
ALM5. Feito Inspeção e Nada foi Encontrado. Equipe C3830. 09:35 . 11:00. OP2. 2016
ALM5. Feito Inspeção e Nada foi Encontrado. Equipe T3207. 07:15. 8:50. OP4. 2016
ALM5. Feito Inspeção e Nada foi Encontrado. Equipe EP692. 11:15. 13:06. OP5. 2016.

Fonte: Autor

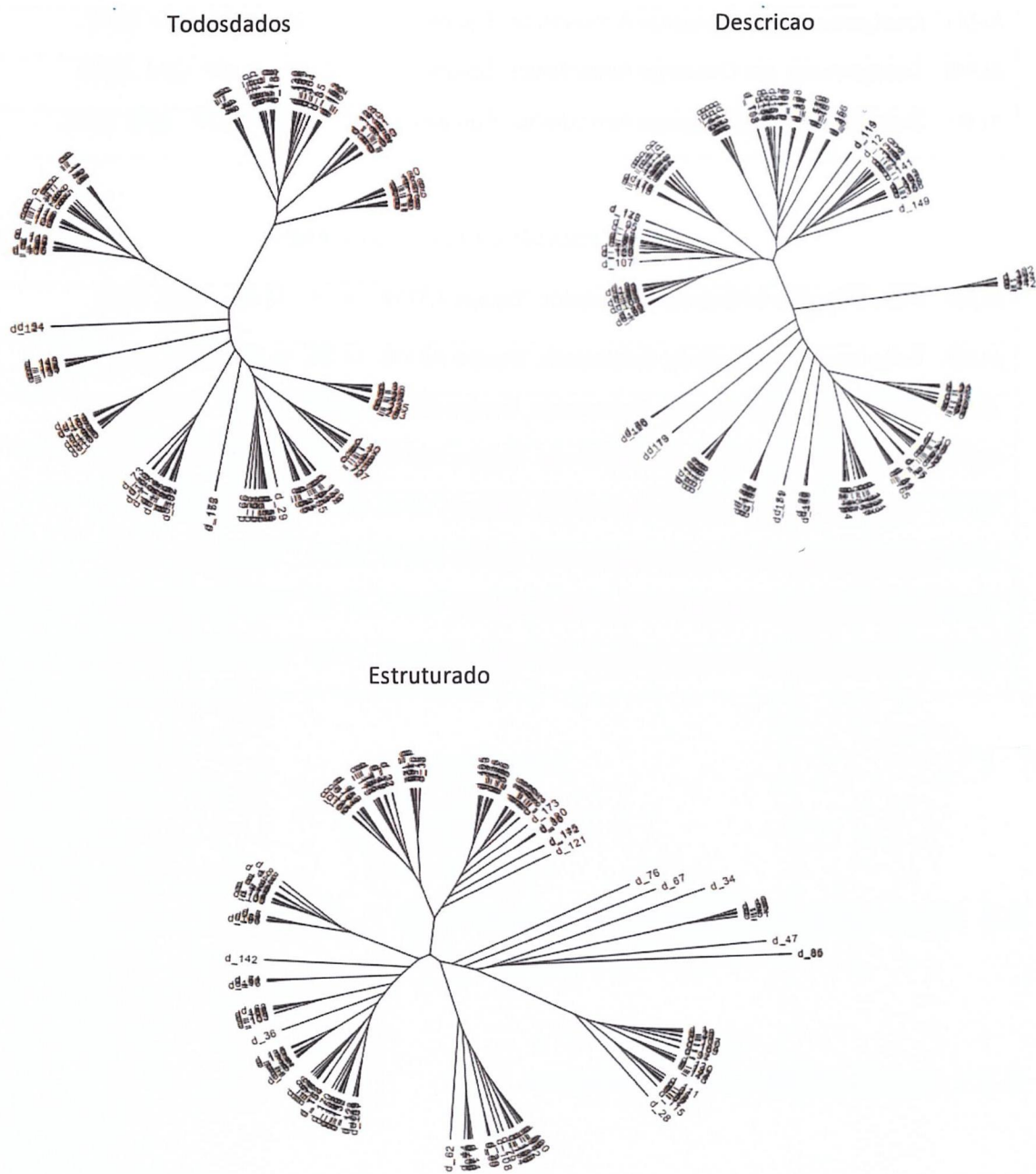
5.3 Eficácia das Saídas de Agrupamento

5.3.1 Saída Gráfica do *Damicore*

A Figura 17 demonstrada abaixo representa três figuras gráficas da saída do *Damicore*, sendo a primeira ou esquerda representada pela Figura 12, gerada pela aplicação da pasta de texto todosdados, a segunda ou a direita pela Figura 14 da

pasta de texto descrição e a terceira ou abaixo pela Figura 15 da pasta de texto estruturado.

Figura 17 - Gráfico *Trees Transform Branches* Todosdados x Descrição x Estruturado



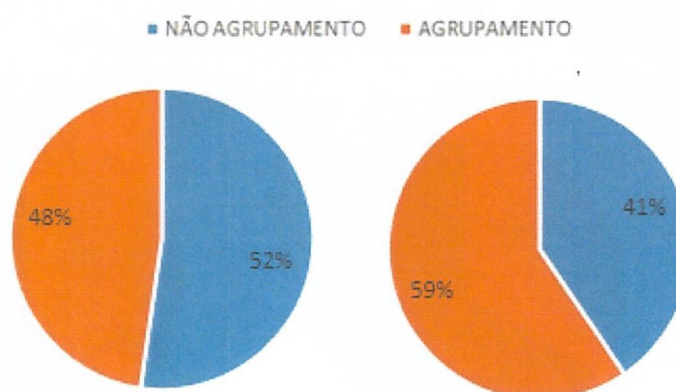
Fonte: Autor

5.3.2 Comparação de Agrupamentos

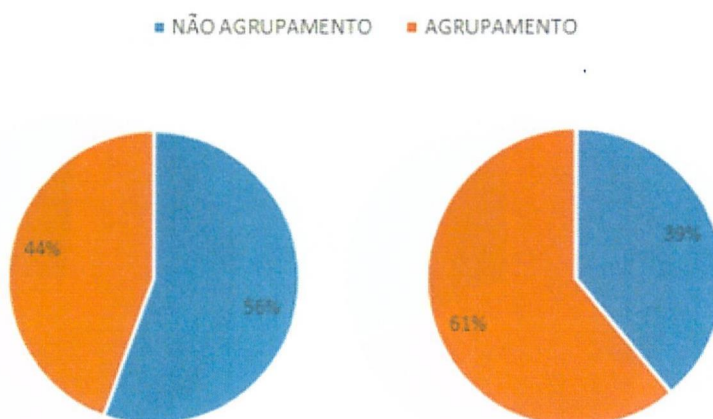
Os gráficos representados pela Figura 18 (a, b, c, d, e, f, g, h, i) abaixo foram obtidos pelos dados numéricos dos agrupamentos relevantes descritos no item 5.1.1.2, gerados pela saída do *Damicore* das pastas de texto todos dados e descrição, conforme Tabela 1 e Tabela 2.

Figura 18 – Agrupamentos da Pasta de Texto Todos dados x Descrição

a) Agrupamentos x Não Agrupamentos

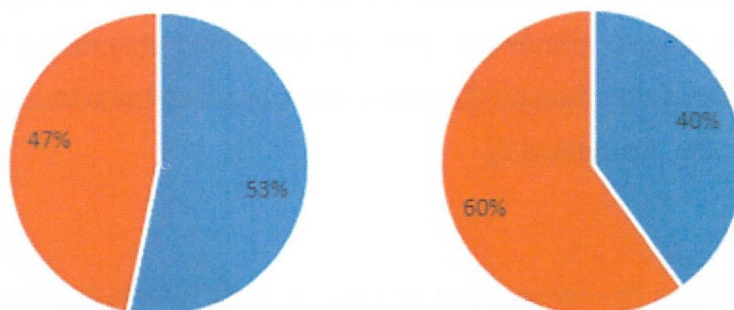


b) Cabo AT Rompido



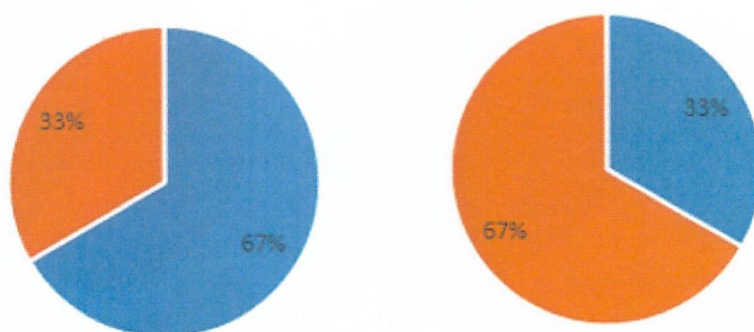
c) Eucalipto ou Árvore ou Galhos na Rede AT

■ NÃO AGRUPAMENTO ■ AGRUPAMENTO



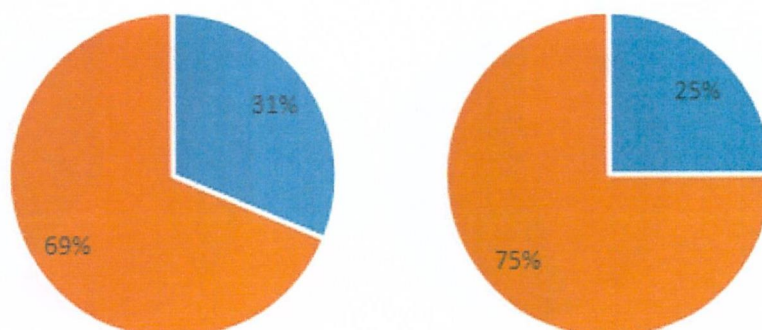
d) Gamba(Gambiarra) na Rede

■ NÃO AGRUPAMENTO ■ AGRUPAMENTO



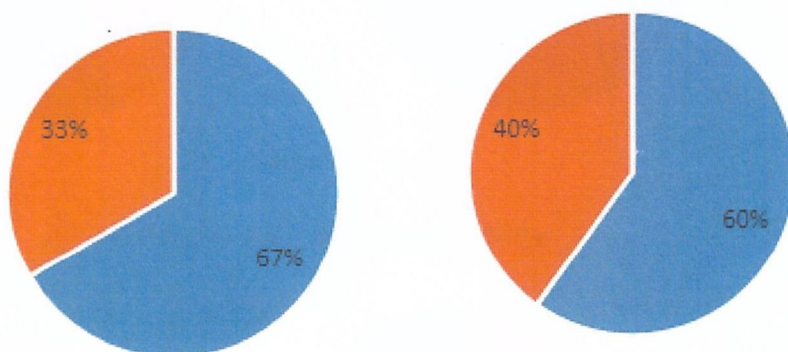
e) Feito Inspeção e Nada Foi Encontrado

■ NÃO AGRUPAMENTO ■ AGRUPAMENTO



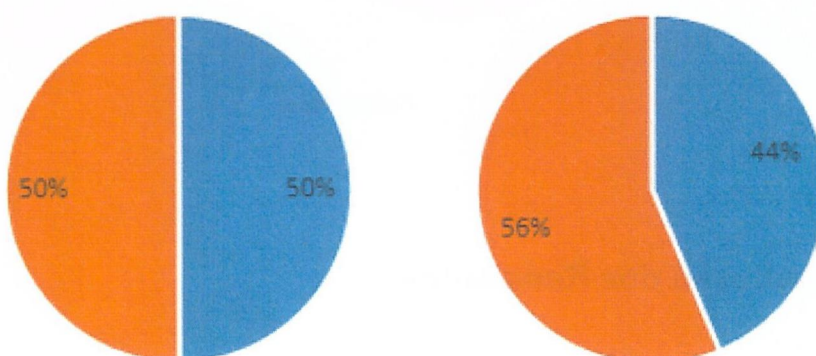
f) Descarga Atmosférica

■ NÃO AGRUPAMENTO ■ AGRUPAMENTO



g) Isolador Danificado

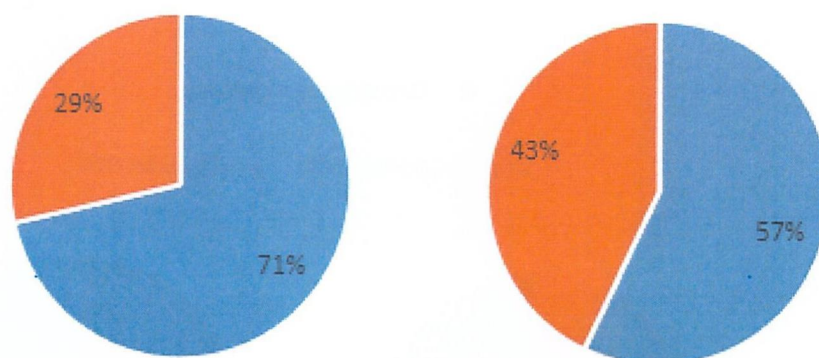
■ NÃO AGRUPAMENTO ■ AGRUPAMENTO



h) Pipa na Rede

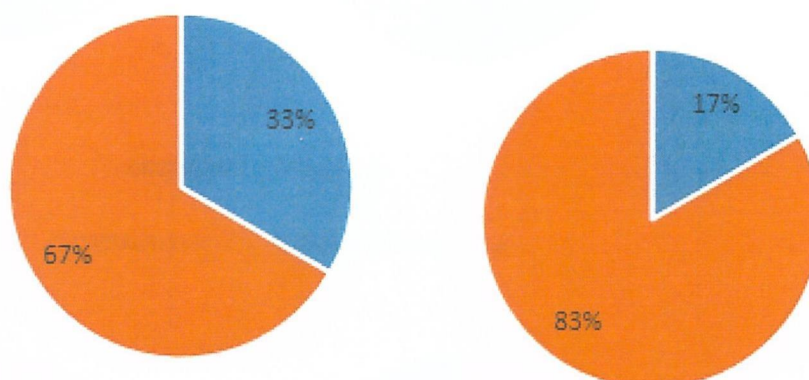
■ NÃO AGRUPAMENTO ■ AGRUPAMENTO





i) Vendaval

■ NÃO AGRUPAMENTO ■ AGRUPAMENTO



Fonte: Autor

5.4 Discussão dos Resultados

Os dados fornecidos pela concessionária de energia elétrica da qual foram submetidos ao *Damicore*, são informações do sistema elétrico de distribuição de energia com relação a interrupção no fornecimento de energia para os consumidores. Todas as informações contidas nestes dados são descritos em relatórios pelos operadores da subestação de energia ou relatados pela equipe de plantão, que utilizam sua linguagem própria resultando em textos não estruturados e não padronizados.

Os agrupamentos e classificação obtidos na saída do *Damicore* são por similaridade ou igualdade dos textos, podemos constatar através de exemplos citados na Figura 13. Os itens (c,d,f,g) nos mostra agrupamentos similares da "Causa" devido a

apresentar texto menores e idênticos. Por outro lado os itens (b,e) apresentam agrupamentos não similares da "Causa" devido á textos maiores, que apresentam termos, palavras idênticas e repetidas.

Pela análise gráfica da Figura 17 constatamos que a saída gráfica pelo *Dam-core* da pasta de texto descricao, obteve faixas de maior agrupamento do que a pasta todosdados, sendo validado pelo gráfico de dados numéricos da Figura 18 (a), onde temos 59% de agrupamento para 41% de não agrupamento.

No gráfico da Figura 18 (a) para pasta de texto todosdados, apresenta 52% de não agrupamentos para 42% de agrupamento, motivo pelo qual os dados desta pasta de arquivo contém todas as informações (alimentador, descrição, data, hora inicio, hora fim, operadores e observações) possuindo um tamanho maior e também como os textos são escritos por uma linguagem própria, não padronizada ou estruturada, apresentado termos, frases, palavras idênticas e repetidas.

Por outro lado na Figura 18 (a), a pasta de texto descricao, apresenta 59% de agrupamentos para 41% de não agrupamento, motivo pelo qual a pasta possui somente uma informação (descrição da causa da interrupção) com tamanho menor, que resulta num agrupamento melhor dos termos desejados.

Outro ponto interessante no gráfico da Figura 18 é que em todos os itens, verificamos um aumento no número de agrupamentos conforme Tabela 3 abaixo, quando reduzimos o tamanho das informações em forma de texto da pasta de todosdados para descrição.

a - Agrupamento x Não Agrupamento

b - Cabo AT Rompido

c - Eucalipto ou Árvore ou Galhos na Rede AT

d - Gamba(Gambiarra) na Rede

e - Feito Inspeção e Nada Foi Encontrado

f - Descarga Atmosférica

g - Isolador Danificado

h - Pipa na Rede

i – Vendaval

Tabela 3 – Comparação do Aumento de Agrupamentos da Pasta Todosdados x Descricao

Itens	Todosdados Agrupamento %	Descricao Agrupamento %	Aumento %
a	48	59	11
b	44	61	17
c	47	60	13
d	33	67	34
e	69	75	6
f	33	40	7
g	50	56	6
h	29	43	14
i	67	83	16

Fonte: Autor

Quando estruturamos as informações e padronizamos conforme pasta de texto estruturado, podemos constatar através de uma análise gráfica da saída do *Damicore*, faixas de agrupamentos bem definidas como demonstrada na Figura 17 (estruturado), como também através do agrupamento das informações contidas nos dados na Figura 16 (a,b,c).

Como as saídas do *Damicore* apresentam agrupamentos bem definidos das informações como demonstradas na Figura 16, podemos fazer diferentes análises e confrontar as informações para geração de relatórios gerenciais para tomada de decisão, como exemplos:

- Quantas interrupções de energia foram pela causa (descarga atmosférica ou vendaval ou galho ou pipa, etc).
- Quantas interrupções de energia ocorreu no Alimentador pela causa (Vendaval ou descarga atmosférica ou pipa ou galho , etc).
- Quantas interrupções de energia no Ano ocorreu pela causa (descarga atmosférica ou vendaval ou galho ou pipa, etc).
- Quanto tempo durou a interrupção pela causa (descarga atmosférica ou vendaval ou galho ou pipa, etc).
- Quanto tempo durou o restabelecimento da energia pela causa (descarga atmosférica ou vendaval ou galho ou pipa, etc).
- O Tempo total de interrupção pela causa no mês ou ano
- O tempo de cada equipe levou para restabelecer a energia pela causa

CONCLUSÃO

6. CONCLUSÃO

Atualmente neste contexto tecnológico de geração de grandes volumes de dados ou informações, as empresas necessitam de ferramentas computacionais que realizem a abstração das informações relevantes que possam ser utilizadas para tomada de decisões.

A ferramenta computacional *Damicore* apresenta uma grande vantagem na aplicação destas informações, pois pode ser aplicada a qualquer tipo de informações, não necessita de preparação dos dados. A sua função é realizar a classificação e agrupamento das informações, da qual são realizadas por similaridade ou igualdade.

Neste trabalho a aplicação da ferramenta computacional *Damicore* foi dividida em duas partes, sendo que na primeira foi aplicado em dados reais, não estruturados, não padronizados, descritos em forma de texto por relatórios, utilizando uma linguagem própria, que relatam as possíveis “causas” que geraram a interrupção no fornecimento de energia elétrica aos consumidores. Verificamos a classificação e agrupamento de informações relevantes da “causa”, que podem ser utilizados para geração de relatórios gerenciais para tomada de decisões. Quando utilizamos todas as informações, resultou em 52% de não agrupamentos a 48% de agrupamentos da “causa”, devido os textos serem grandes possui termos ou palavras repetidas, formando textos similares da qual foram agrupados. Por outro lado quando reduzimos o tamanho dos

textos, utilizando apenas a informação da “causa”, resultou em 59% de agrupamentos a 41% de não agrupamentos.

Na segunda parte, foi utilizado os mesmos conceitos de “causa”, porém as informações foram estruturadas e padronizadas. Resultando agrupamentos bem definidos, pois os textos ficam similares. Com base nestes agrupamentos, podemos gerar relatórios gerenciais para tomada de decisão.

REFERÊNCIAS

REFERÊNCIAS

- BENEDETTO, D.; CAGLIOTI, E.; LORETO, V. **Language Trees and Zipping**. p. 2–5, 2002.
- BLOMBERG, LUCIANO COSTA;MOTA, EDUARDO GONÇALVES;FIGUEIREDO, JOSÉ ANTÔNIO POLI;RUIZ, D. D. A. **Development of an oral health database for the management of clinical records**. Program, v. 24, n. 3, p. 249–253, 2009.
- BOLFE, É. L. et al. **Uso, ocupação das terras e banco de dados geográficos da metade sul do Rio Grande do Sul**. Ciência Rural, v. 39, n. 6, p. 1729–1737, 2009.
- BORLONE, D. R. **Análise comportamental por intermédio das tecnologias de big data**. 2015.
- BRITO, J. N. **Desenvolvimento de um Sistema Inteligente para Diagnóstico de Falhas em Motores de Indução Trifásicos**. X Simpósio Brasileiro de Automação Inteligente (SBAI), v. X, p. 474–479, 2002.
- CARLOS, G.; REGINA BASTOS HARO, B. **Cip – Catalogação Na Publicação**. 2002.
- CESAR, B. K. M. **Estudo e extensão da metodologia DAMICORE para tarefas de classificação**. p. 119, 2016.
- CHEN, X. et al. **Shared Information and Program Plagiarism Detection**. v. 50, n. 7, p. 1545–1551, 2004.
- CORREA, A.; TAFFAREL, M.; MENON, G. **Análise de eficiência : uma comparação das empresas estatais e privadas do setor de energia elétrica brasileiro**. Revista Catarinense da Ciência Contábil, v. 15, n. 46, p. 9–23, 2016.
- COSTA, J. A. M. **Desenvolvimento de um esquema XML para banco de dados sobre ovinos**. 2002.
- CZEJDO, B. et al. **A Graphical Data Manipulation Language for an Extended Entity-Relationship Model**. Computer, v. 23, n. 3, p. 26–36, 1990.
- FALLEIRO, A. D. M.; GASTALDINI, M. DO C. C.; ANDRADE, J. C. S. **Projetos de energia renovável no âmbito do MDL - perfil dos projetos localizados no RS**. Revista de Administração da UFSM, v. 7, n. 0, p. 100–112, 2014.
- FEIJÓ, D. DE V. **Consultando XML por meio de Modelos Conceituais: extensão e formalização de CXPath**. 2005.

FRANCESCHET, M.; GUBIANI, D.; MONTANARI, A. **A Graph-Theoretic Approach to Map Conceptual Designs to XML Schemas**. ACM Transactions on Database Systems (TODS), v. 38, n. 1, p. 6, 2013.

FURTADO, J. C. et al. **Ferramenta para extração de dados semi-estruturados para carga de um big data**. Revista Brasileira de Computação Aplicada, v. 7, n. 3, p. 43–52, 2015.

GOMES, E. A. P. **Detecção de Ilhamento de Geradores Síncronos Distribuídos por Correlações da Mineração Complexa de Dados**. 2016.

HASSAN, M. et al. **Database selection and keyword search of structured databases: Powerful search for naive users**. Proceedings of the 2003 IEEE International Conference on Information Reuse and Integration, IRI 2003, p. 175–182, 2003.

HIRANO, M. Y. **Geração de energia por microturbina alimentada à biogás m uma propriedade rural - estudo de caso**. p. 110, 2015.

LIU, Z. H.; HAMMERSCHMIDT, B.; MCMAHON, D. **JSON data management-Supporting schema-less development in RDBMS**. Proceedings of the ACM SIGMOD International Conference on Management of Data, p. 1247–1258, 2014.

LUCAS BIANCHI MAGALHÃES et al. **Análise Comparativa Dos Algoritmos De Otimização De Consultas Do Postgresql**. Colloquium Exactarum, v. 7, n. 1, p. 01–21, 2015.

LUIS FERNANDO DE SOUZA MORO. **Caracterização de Alunos em Ambientes de Ensino Online: Estendendo o Uso da DAMICORE para Minerar Dados Educacionais**. p. 631, 2014.

LYNCH, M. et al. **Unstructured data**. p. 6–9, 2010.

MALACRIDA, T. F.; ZAUPA, A. P.; PAZOTI, M. A. **Desenvolvimento De Uma Ferramenta Para Migração De Dados Entre Bancos De Dados Relacionais**. Colloquium Exactarum, v. 6, n. 1, p. 20–36, 2014.

MANUEL, J. **Análise de Sinal Fisiológico usando Métodos não Lineares**. 2015.

MARQUES, S.; BOSCARIOLI, C. **Sistemas gerenciadores de banco de dados relacionais Fuzzy: uma aplicação em recuperação de informação**. n. 1977, p. 1733–1743, 2002.

MELLO, R. DOS S. **Uma abordagem Bottom-UP para a integração semântica de esquemas XML**. v. 0, p. 145, 2002.

MERGEN, S. L. S. **Casamento de esquemas XML e esquemas relacionais**. 2005.

OLIVEIRA, M. G. DE; GABRIELLE, A.; FALCÃO, R. **SCALING SPATIAL BIG DATA IN A LOCATION-BASED SOCIAL NETWORK effective actions for the solution of the citizen ' s demands . The Crowd4City LBSN was conceived adopting RDBMS (Relational Database Management Systems) technologies enabled for spatial data h**. p. 141–155, 2014.

PEREIRA, O. et al. **Proposta Para Implantação De Business Intelligence Em Shopping Center**. Colloquium Exactarum, v. 8, n. 2, p. 13–26, 2016.

- PINTO, R. DE S. **Caracterização do perfil de utilização de recursos de programas a partir de arquivos executáveis utilizando mineração de dados**. 2017.
- QUAK, B. et al. **Database Support for Condition Assessment of Distribution Power Cables**. p. 1100–1104, 2002.
- RAJESH, P. K.; MAGESHBABU, B. **Web-based collaborative conceptual design: An XML approach**. International Journal of Advanced Manufacturing Technology, v. 38, n. 5–6, p. 433–440, 2008.
- RECHE, E. A. et al. **Data Mining-Based Method to Reduce Multiple Estimation for Fault Location in Radial Distribution Systems**. IEEE Transactions on Smart Grid, v. 3053, n. c, p. 1–1, 2018.
- ROSA, L. P. **Geração hidrelétrica, termelétrica e nuclear**. Estudos Avançados, v. 21, n. 59, p. 39–58, 2007.
- SACCOL, D. DE B. **Materialização de Visões XML**. 2001.
- SANCHES, A.; CARDOSO, J. M. P.; DELBEM, A. C. B. **Identifying merge-beneficial software kernels for hardware implementation**. Proceedings - 2011 International Conference on Reconfigurable Computing and FPGAs, ReConFig 2011, p. 74–79, 2011.
- SILBERSCHATZ, A.; KORTH, H. F.; SUDARSHAN, S. **Sistema de banco de dados**. 6º ed. Rio de Janeiro: Elsevier, 2012.
- SILVA, B. DE A. **Um método de otimização da relação desempenho/consumo de energia para arquiteturas multi-cores heterogêneas em FPGA**. 2016.
- SILVA, R. M. DA; PRIMO, M. A. M. **Framework para análise da sustentabilidade de fontes de energia**. Interciencia, v. 38, n. 11, p. 760–768, 2013.
- SILVEIRA, F. V. **Fragmentação e Decomposição de Consultas em XML**. 2006.
- SPANGLER, S. **Data mining : Three steps to mining unstructured data**. p. 2–8, 2018.
- SUMOYAMA, A. S. **Classificador de kernels para mapeamento em plataforma de computação híbrida composta por FPGA e GPP**. 2016.
- TEOREY, T. J.; YANG, D.; FRY, J. P. **A logical design methodology for relational databases using the extended entity-relationship model**. ACM Computing Surveys, v. 18, n. 2, p. 197–222, 1986.
- WAGH, J. J.; GONDANE, J. D.; DUKARE, A. T. **Unstructured data mining and its applications**. n. 3, p. 36–40, 2016.
- WEHNER, S. **Analyzing Worms and Network Traffic using Compression**. 2018.