

Felipe Francesco Ramthum Vegini

**PROTÓTIPO DE SISTEMA BASEADO EM
CONHECIMENTO PARA CONFIGURAÇÃO DE PRODUTO:
APLICAÇÃO NA INDÚSTRIA DE REVESTIMENTOS**

Dissertação submetida ao Programa de Pós-graduação em Engenharia Mecânica da Universidade Federal de Santa Catarina para a obtenção do Grau de Mestre em Engenharia Mecânica
Orientador: Prof. Jonny Carlos da Silva, Dr. Eng.

Florianópolis, 2018

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Vegini, Felipe Francesco Ramthum
Protótipo de Sistema Baseado em Conhecimento para
Configuração de Produto: Aplicação na Indústria de
Revestimentos / Felipe Francesco Ramthum Vegini ;
orientador, Jonny Carlos da Silva, 2018.
149 p.

Dissertação (mestrado) - Universidade Federal de
Santa Catarina, Centro Tecnológico, Programa de Pós
Graduação em Engenharia Mecânica, Florianópolis, 2018.

Inclui referências.

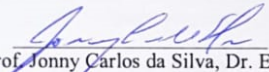
1. Engenharia Mecânica. 2. Inteligência
Artificial. 3. Árvore de decisões. 4. Projeto. 5.
Configuração de produto. I. da Silva, Jonny Carlos.
II. Universidade Federal de Santa Catarina.
Programa de Pós-Graduação em Engenharia Mecânica. III.
Título.

Felipe Francesco Ramthum Vegini


**PROTÓTIPO DE SISTEMA BASEADO EM
CONHECIMENTO PARA CONFIGURAÇÃO DE PRODUTO:
APLICAÇÃO NA INDÚSTRIA DE REVESTIMENTOS**

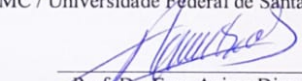
Esta Dissertação foi julgada adequada para obtenção do Título de “Mestre em Engenharia Mecânica”, e aprovada em sua forma final pelo Programa de Pós-Graduação em Engenharia Mecânica.

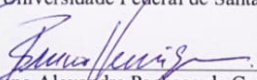
Florianópolis, 15 de junho de 2018.

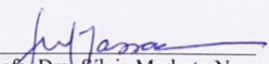

Prof. Jonny Carlos da Silva, Dr. Eng.
Coordenador do Curso

Banca Examinadora:


Prof. Jonny Carlos da Silva, Dr. Eng. - Orientador
EMC / Universidade Federal de Santa Catarina


Prof. Dr. Eng. Acires Dias
EMC / Universidade Federal de Santa Catarina


Prof. Dr. Bruno Alexandre Pacheco de Castro Henriques
EMC / Universidade Federal de Santa Catarina


Prof. Dra. Silvia Modesto Nassar
INE / Universidade Federal de Santa Catarina

Dedico este trabalho a meus
pais, Juarez e Marcia e minha noiva
Allana.

AGRADECIMENTOS

Gostaria de agradecer primeiramente a Deus, que me concedeu o dom da vida e do discernimento para fazer as escolhas certas para meu futuro.

A meus pais por terem me apoiado durante todo o mestrado, desde o momento que saí de casa para realizar as disciplinas até quando retornei para realizar o estágio e concluir a dissertação.

A minha noiva Allana Ferreira Boschen por ter entrado em meu destino durante a realização deste mestrado e por ter me aturado enquanto eu falava incessantemente sobre o projeto.

Ao meu orientador, o professor Jonny Carlos da Silva por ter proposto me ajudar e orientar em um momento em que eu estava prestes a desistir de todo o mestrado.

A toda a equipe do LASHIP por todo o apoio durante a etapa de disciplinas e ao Victor por ter me orientado durante meu primeiro ano em Florianópolis.

Ao departamento de Sistemas de Engenharia do grupo WEG pelo apoio financeiro e intelectual durante o desenvolvimento do projeto.

Ao Ramses Libera e Jean Kahl por terem aprovado minha sugestão para o desenvolvimento do projeto e darem todo o apoio para sua conclusão.

“Por vezes sentimos que aquilo que fazemos não é senão uma gota de água no mar. Mas o mar seria menor se lhe faltasse essa gota.”
(Madre Teresa de Calcuta)

RESUMO

Para que um vendedor possa iniciar efetivamente o processo de venda para um potencial cliente existem dois pontos fundamentais: extrair do cliente suas necessidades ou requisitos de projeto e conhecer as soluções que estão no portfólio da empresa que possam atender a estes requisitos. Empresas de grande porte necessitam de um elevado número de profissionais que consigam realizar habilmente estes dois pontos, mas existe certa dificuldade em encontrá-los em um número suficiente, o que demanda tempo e dinheiro da empresa para contratar e treinar os iniciantes. Este trabalho busca utilizar técnicas de sistemas especialistas e mineração de dados para sistematizar o processo de vendas por meio de um software configurador de produto, também denominado de venda guiada, de modo que sirva como plataforma de treino para os profissionais iniciantes e também como plataforma para clientes realizarem a compra diretamente. Para tal foi feito o mapeamento das principais ações realizadas pelos profissionais durante todas as etapas de uma venda, estas ações foram convertidas em perguntas simples e objetivas, onde as respostas destas são conectadas a bancos de dados especiais desenvolvidos para convergirem até uma solução. Para a seleção da ordem das perguntas foi utilizado um sistema híbrido entre Sistemas Especialistas e um classificador do tipo Árvore de Decisão, fazendo com que sua sequência seja selecionada dinamicamente de acordo com as respostas previamente escolhidas pelo usuário. O protótipo desenvolvido foi verificado e validado por membros da equipe de vendas técnicas da WEG Tintas, confirmando que o mesmo consegue simular com bastante precisão o procedimento normalmente adotado por eles para tratar com clientes e concluir vendas com sucesso.

Palavras-chave:

1. Inteligência Artificial 2. Árvore de decisões 3. Projeto 4. Configuração de produto

KNOWLEDGE-BASED SYSTEM PROTOTYPE FOR PRODUCT CONFIGURATION: APPLICATION IN THE COATINGS INDUSTRY

ABSTRACT

In order for a salesperson effectively start the sales process for a potential customer, there are two key points: extract from the customers their needs or project requirements and to know the solutions that are in the company's portfolio that can meet these requirements. Large companies need a large number of professionals who can skillfully perform these two points, but there is a certain difficulty in finding them in sufficient numbers, which cost to the company time and money hiring beginners. This work seeks to use techniques from expert systems and data mining to systematize the sales process through a product configurator software, also known as guided sales, so that it serves as a training platform for beginning professionals and also as a platform for customers to make the purchase directly. For this purpose, the mapping of the main actions carried out by the professionals during all the stages of a sale was done, these actions were converted into simple and objective questions, where the answers of these are connected to special databases developed to converge to a solution. For the selection of the order of the questions, a hybrid system between Expert Systems and a Decision Tree type classifier was used, making its sequence dynamically selected according to the answers previously chosen by the user. The developed prototype was verified and validated by members of the technical sales team of WEG Tintas, confirming that it can simulate quite accurately the procedure normally adopted by them to deal with customers and successfully complete sales.

Key-words:

1. Artificial Intelligence
2. Decision Trees
3. Project
4. Product Configurator

LISTA DE FIGURAS

Figura 2-1 – Ciclo dos Metais	29
Figura 2-2 – Pilha de Corrosão.....	32
Figura 2-3 – Corrosão do tipo uniforme.....	33
Figura 2-4 – Corrosão do tipo placas	34
Figura 2-5 – Corrosão do tipo alveolar.....	34
Figura 2-6 – Corrosão do tipo puntiforme.....	34
Figura 2-7 – Corrosão do tipo frestas	35
Figura 2-8 – Corrosão do tipo galvânica	35
Figura 2-9 – Corrosão do tipo intergranular	36
Figura 2-10 – Corrosão do tipo transgranular	36
Figura 2-11 – Corrosão do tipo filiforme	37
Figura 2-12 – Corrosão do tipo tensão e fadiga.....	37
Figura 2-13 – Placa sem biocida (esquerda) e Placa com biocida (direita)	41
Figura 2-14 - Organismos marinhos incrustantes.....	42
Figura 2-15 – Processo de fabricação de tintas líquidas.....	47
Figura 2-16 – Esquema simplificado de cura das tintas	48
Figura 2-17 – Composição básica de uma tinta em pó.....	52
Figura 2-18 – Esquema de fabricação de tinta em pó.....	52
Figura 2-19 – Leito fluidizado simples	54
Figura 2-20 – Leito fluidizado eletrostático	55
Figura 2-21 –Aplicação de tinta pó com pistola eletrostática.....	56
Figura 2-22 – Sistemas de pintura tradicionais	57
Figura 3.1 – Elementos básicos de um sistema especialista.....	61
Figura 3.2 – Modelo para desenvolvimento de um sistema especialista	64
Figura 3.3 – Rede semântica	68
Figura 3.4 – Exemplo de diagrama de classes.....	69
Figura 3.5 - Árvore de decisão	73
Figura 3.6 - Árvore com uma boa e uma má pergunta como nós.....	74
Figura 3.7 –Classificação de dados por uma Árvore de Decisão	75
Figura 3.8 – Criação de uma árvore de decisão por etapas	77
Figura 3.9 – Árvore de decisão com primeiro nó	85
Figura 3.10 – Representação do problema de superadaptação.	86
Figura 3.11 – Árvore de decisão com um atributo de muitos valores ...	87
Figura 4.1 – Etapas do projeto conceitual	92
Figura 4.2 – Diagrama de classes do conjunto de execução.....	97
Figura 4.3 – Diagrama de classes do conjunto de atualização.	98
Figura 4.4 - Estrutura geral de operação do protótipo.....	100

Figura 4.5 - Banco de dados aberto com Excel.....	102
Figura 4.6 – Representações do JSON: a) minimalista, b) normal, c) Árvore.....	104
Figura 4.7 – Formato de um objeto em JSON.....	104
Figura 4.8 – Formato de um array em JSON	105
Figura 4.9 – Formato de um valor em JSON	105
Figura 4.10 - Cálculos de entropia do protótipo durante sua execução	109
Figura 4.11 - Tela de pergunta do protótipo.....	111
Figura 4.12- Exemplo de pergunta com e sem alternativas ocultas	112
Figura 4.13 - Tela de respostas do protótipo.....	113
Figura 5.1 - Antecipação de erros de compilação pelo Eclipse.....	115
Figura 5.2 –Verificação do encadeamento de perguntas.....	118
Figura 5.3 - Exemplo de verificação das respostas para as perguntas.	118
Figura 5.4 – Log de depuração do protótipo	121
Figura 5.5 - Log de depuração do protótipo.....	122
Figura 5.6 - Log de depuração do protótipo.....	123
Figura 5.7 – Criação dinâmica da árvore de decisões.	124
Figura 5.8 - Árvore de decisão gerada dinamicamente	125
Figura 5.9 - Vista geral da tela do protótipo	129
Figura 5.10 – Tela de pergunta e explicação do protótipo	130
Figura 5.11 – Menu para remover ou repetir perguntas	130
Figura 5.12 – Fim da execução com perguntas puladas.....	131
Figura 5.13 - Tela de pergunta sobre o tamanho da aplicação.....	132
Figura 5.14 - Tela de resultados	133
Figura 5.15 - Redirecionamento para site de e-commerce WEG.....	134
Figura 5.16 - Relatório da venda guiada	134

LISTA DE QUADROS

Quadro 2.1 – Formas de corrosão	33
Quadro 2.2 – Características para o crescimento de microrganismos ...	39
Quadro 2.3 – Vantagens e desvantagens das tintas líquidas.....	43
Quadro 2.4 - Tipos de resinas e solventes apropriados	46
Quadro 2.5 - Vantagens e desvantagens das tintas em pó	51
Quadro 3.1 - Conjunto de treino para classificar se ambientes necessitam de primer epóxi ou não.	76
Quadro 3.2 - Conjunto de teste para classificar se ambientes necessitam de primer epóxi ou não.	83
Quadro 4.1 – Quadro para tratamento de especialidades	106
Quadro 4.2 – Conjunto de dados simplificados.....	107
Quadro 5.1 - Planilha de cálculos de entropia	116

ABREVIATURAS E SIGLAS

ABNT	Associação Brasileira de Normas Técnicas
CLIPS	<i>C Language Integrated Production System</i>
CSV	<i>Comma-Separated Value</i>
EASY	<i>Engineering And Sales System</i>
EC	Engenheiro de Conhecimento
EH	Especialista Humano
GUI	<i>Graphical User Interface</i>
IDE	<i>Integrated Development Environment</i>
IA	Inteligência Artificial
ISO	<i>International Organization for Standardization</i>
JSON	<i>JavaScript Object Notation</i>
POO	Programação Orientada a Objetos

ALFABETO LATINO

$gain(S,T)$	Ganho de Informação do conjunto S após divisão pelo atributo T	[bits]
$gain\ ratio(S,T)$	Razão de ganho da divisão do conjunto S pelo atributo T	[1]
$H(S)$	Entalpia do conjunto S	[bits]
$H(S,T)$	Entalpia do conjunto S após divisão pelo atributo T	[bits]
$info\ split_{(T)}S$	Divisão de Informação do atributo T do conjunto S	[bits]

SUMARIO

1.	INTRODUÇÃO	21
1.1	INTELIGÊNCIA ARTIFICIAL	21
1.2	CONTEXTO DA PESQUISA	22
1.3	JUSTIFICATIVA DO TRABALHO	23
1.4	OBJETIVOS DO TRABALHO	24
1.4.1	Objetivo geral	24
1.4.2	Objetivo específico	24
1.5	ESTRUTURA DO TRABALHO	25
2.	INDÚSTRIA DE REVESTIMENTOS	27
2.1	FUNÇÃO DOS REVESTIMENTOS	28
2.1.1	Corrosão	28
2.1.1.1	Importância do estudo da corrosão	30
2.1.1.2	Mecanismos dos processos corrosivos	31
2.1.1.3	Formas de Corrosão	32
2.1.1.4	Maneiras de Mitigar a Corrosão	38
2.1.1.5	Casos benéficos da corrosão	38
2.1.2	Microrganismos	39
2.1.2.1	Emersos	39
2.1.2.2	Submersos	40
2.2	TINTAS LÍQUIDAS	43
2.2.1	Matérias-primas	44
2.2.1.1	Resina	44
2.2.1.2	Pigmento	44
2.2.1.3	Solvente	45
2.2.1.4	Aditivos	46
2.2.2	Processo de fabricação	46
2.2.3	Formação da película de tinta	47
2.2.4	Tipos de tinta	48
2.2.4.1	Tintas com veículos não convertíveis	48
2.2.4.2	Tintas com veículos convertíveis	49
2.3	TINTA PÓ	51
2.3.1	Matérias-primas	51
2.3.2	Processo de fabricação	52
2.3.3	Aplicação de tinta em pó	53
2.3.3.1	Leito fluidizado simples	53
2.3.3.2	Leito fluidizado eletrostático	54
2.3.3.3	Pistola eletrostática	55
2.4	ESQUEMAS DE PINTURA	56

2.5	SELEÇÃO DOS ESQUEMAS	57
3.	INTELIGENCIA ARTIFICIAL E MINERAÇÃO DE DADOS.....	59
3.1	SISTEMAS ESPECIALISTAS.....	59
3.1.1	Surgimento dos sistemas especialistas.....	59
3.1.2	Estrutura de um Sistema Especialista.....	60
3.1.2.1	Base de Conhecimento.....	61
3.1.2.2	Memória de Trabalho.....	61
3.1.2.3	Máquina de Inferência.....	61
3.1.2.4	Interface	62
3.1.3	Desenvolvimento de Sistemas Especialistas	63
3.1.3.1	Análise de Viabilidade	64
3.1.3.2	Aquisição do Conhecimento	64
3.1.3.3	Representação do Conhecimento	65
3.1.3.4	Implementação.....	70
3.1.3.5	Verificação e Validação	70
3.2	MINERAÇÃO DE DADOS	71
3.2.1	Árvores de decisão.....	72
3.2.2	Algoritmos para árvores de decisão	78
3.2.2.1	Seleção de Atributos	79
3.2.2.2	Seleção dos atributos pelo ID3.....	80
3.2.2.3	Seleção de atributos pelo C4.5	87
4.	DESENVOLVIMENTO DO PROTÓTIPO	91
4.1	PROJETO CONCEITUAL.....	91
4.1.1	Identificação do Problema.....	92
4.1.2	Definição do Público Alvo	93
4.1.3	Definição da Ferramenta de Desenvolvimento	94
4.1.4	Definição do Método de Inferência.....	95
4.1.5	Definição do Modelo de Desenvolvimento.....	95
4.2	AQUISIÇÃO DE CONHECIMENTO.....	95
4.3	REPRESENTAÇÃO DO CONHECIMENTO	96
4.3.1	Conjunto de execução	96
4.3.2	Conjunto de atualização	97
4.3.3	Estrutura Geral.....	99
4.4	IMPLEMENTAÇÃO	101
4.4.1	Banco de Dados	101
4.4.1.1	JSON.....	103
4.4.1.2	Conversão.....	105
4.4.2	Módulo especialista para seleção de perguntas	106
4.4.3	Módulo estatístico para seleção de perguntas.....	107
4.4.3.1	Cálculos.....	107

4.4.3.2	Algoritmo	110
4.4.4	Interação com o Usuário.....	111
5.	VERIFICAÇÃO E VALIDAÇÃO	114
5.1	VERIFICAÇÃO	114
5.1.1	Teste do protótipo.....	120
5.2	VALIDAÇÃO	126
5.3	TELAS DO PROTÓTIPO	128
5.3.1	Interação com o usuário.....	129
5.3.2	Apresentação dos Resultados	132
6.	CONCLUSÕES	137
6.1	CONSIDERAÇÕES E COMENTÁRIOS FINAIS	137
6.2	CONTRIBUIÇÕES	138
6.3	APRENDIZADOS	138
6.4	RECOMENDAÇÕES PARA TRABALHOS FUTUROS	139
	BIBLIOGRAFIA.....	141

1. INTRODUÇÃO

1.1 INTELIGÊNCIA ARTIFICIAL

O termo Inteligência Artificial (IA) foi proposto pela primeira vez em 1956 por John McCarthy na Conferência o Dartmouth College (STAIR e REYNOLDS, 2002). Mas definir o que é a IA se torna uma questão bastante complexa, pelo fato de possuir aspectos biológicos, psicológicos e filosóficos, não existe uma resposta definitiva ou um consenso no meio acadêmico para sua definição.

De acordo com Luger (2004), “Inteligência artificial pode ser definida como o ramo da ciência da computação que se ocupa da automação do comportamento inteligente”, já Winston (1992) diz que a IA é o estudo dos conceitos que permitem aos computadores serem inteligentes e Coppin (2013) utiliza duas definições “Inteligência Artificial é o estudo dos sistemas que agem de um modo que a um observador qualquer pareceria ser inteligente” e “Inteligência Artificial envolve utilizar métodos baseados no comportamento inteligente de humanos e outros animais para solucionar problemas complexos.”

Todas estas quatro definições, apesar de diferentes, se apropriam das palavras inteligente ou inteligência, mas estas também trazem um conceito difícil de ser expressado e que gera discussões no meio acadêmico. Eysenck (1998) traz uma definição de 1994, publicada no Wall Street Journal e assinada por 50 pesquisadores da época, e diz que a inteligência é uma capacidade mental que envolve a habilidade de planejar, raciocinar, pensar de forma abstrata, resolver problemas, aprender rápido, compreender ideias complexas e aprender com a experiência.

O desenvolvimento da IA tem girado em torno de três objetivos: 1) o desenvolvimento de modelos formais para a inteligência humana (ciência cognitiva ou psicologia computacional), 2) a exploração de técnicas computacionais com potencial para simular comportamento inteligente e 3) o desenvolvimento de aplicações industriais ou educacionais através de técnicas de IA (BITENCOURT, 2006).

A IA é uma área que vem sendo muito atrativa para pesquisadores de diversas áreas ao longo de toda sua história, principalmente por conta de sua grande divulgação dentro de filmes de ficção científica, apresentando robôs com um comportamento inteligente e habilidades muito superiores à dos seres humanos. Apesar da ideia de IA parecer ser

relativamente simples, a criação de máquinas realmente inteligentes ainda é um grande desafio (ARTERO, 2009).

Existe um campo conhecido como ciência cognitiva, que trabalha com uma ideia onde, para se poder dizer que um programa realiza tarefas do mesmo modo que os seres humanos, é necessário primeiro determinar exatamente como estes pensam e agem. Existem três maneiras utilizadas para se avaliar e desenvolver teorias em cima deste conceito (RUSSEL e NORVIG, 2004):

- Introspecção: Tentar capturar e anotar os pensamentos conforme estes forem se desenvolvendo;
- Experimentos Psicológicos: Observar uma pessoa em ação sob determinadas condições;
- Imagens Cerebrais: Observar diretamente o cérebro enquanto se passa por condições.

Muitas aplicações de Inteligência Artificial são baseadas em construir um modelo que represente o conhecimento utilizado por especialistas humanos, normalmente a introspecção é suficiente para capturar o conhecimento e implantar dentro de um software. Em alguns casos a atividade do especialista pode ser considerada semelhante a um processo de classificação, utilizando um conjunto de parâmetros para dizer se determinada situação pode acontecer ou para classificar produtos como aplicáveis ou não ao cenário avaliado (RUSSEL e NORVIG, 2004), este último será o principal cenário abordado neste trabalho.

1.2 CONTEXTO DA PESQUISA

Este projeto iniciou de uma demanda industrial, onde o autor deste trabalho fez uma proposta para o desenvolvimento de um Sistema Especialista voltado para a divisão de tintas do grupo WEG. Esta proposta se enquadrou em um dos projetos que a empresa pretendia implantar.

O projeto foi realizado com ênfase na divisão de tintas do grupo WEG com possibilidade de expansão para as outras unidades de negócio da empresa. O setor de tintas do grupo possui duas unidades produtivas no Brasil, em Guaramirim-SC e Mauá-SP, além de uma unidade na Argentina. As unidades produzem tintas em pó, tintas líquidas e vernizes eletroisolantes, abrangendo linhas especiais para motores, transformadores, anticorrosivas, marítimas, automotivas, linha branca plásticos, vidros e espelhos. Mensalmente produzem cerca de 1,7 milhões

de litros de tintas líquidas e 2.300 toneladas de tintas em pó, divididos em milhares de tipos de tintas diferentes.

O grupo WEG foi fundado em 16 de setembro de 1961 com o nome de Eletromotores Jaraguá produzindo apenas motores elétricos, mas com o passar dos anos a empresa foi crescendo e explorando outros nichos de mercado. Atualmente possui 5 unidades de negócios distintas: Motores, Automação, Energia, Transmissão & Distribuição e Tintas, contando juntas com 22.600 colaboradores no Brasil e mais 8.300 espalhados no resto do globo. Apresenta filiais em 29 países e fábricas em outros 12. Contando com cerca de 2.700 engenheiros na equipe, a empresa demonstra grande interesse contratar mão de obra especializada. Possui um centro de treinamento para adolescentes com mais de 10 cursos para áreas estratégicas dentro da empresa.

1.3 JUSTIFICATIVA DO TRABALHO

Este trabalho representa uma solução para pelo menos três problemas enfrentados pelo grupo WEG: 1) a equipe de vendas tem problemas em controlar o extenso portfólio da empresa, 2) necessidade de treinamento técnico para novos vendedores (em especial para o segmento de tintas) e uma plataforma de e-commerce apresentando somente informações técnicas, gerando confusão para sua utilização por usuários leigos.

Quando uma empresa expande seus negócios, naturalmente seu portfólio de produtos passa a aumentar também, podendo chegar a um momento que a própria empresa não tem total controle/conhecimento de sua própria capacidade produtiva. Em relação as equipes de vendas, muitas vezes elas estão tão focadas em seus produtos mais vendidos, os "carros-chefes", que acabam deixando de ofertar outros produtos menos conhecidos, mas que poderiam também atender as necessidades dos clientes. Esse cenário tem se agravado em uma empresa como a WEG, pois em seu modelo de negócios atual ela tem feito diversas aquisições de novas empresas, cada uma com seu catálogo próprio de produtos, resultando em um número cada vez maior de produtos que os vendedores devem "gravar" para poder ofertar aos seus clientes. Atualmente a WEG conta com mais de 460 linhas de produtos, algumas contendo milhares de itens distintos. Neste cenário, a venda guiada auxiliaria para se ter um fácil acesso à informação dos produtos WEG para se ofertar aos clientes.

No Brasil não existe um curso oficial de capacitação para profissionais do segmento de revestimentos (tintas e vernizes), os cursos que mais se aproximam são os de Química e Engenharia Química, mas

mesmo estes muitas vezes não possuem uma disciplina que sequer mencione os revestimentos de maneira direta, ficando completamente a cargo das empresas treinar profissionais quase do zero para garantir um serviço de qualidade. O problema se agrava ainda mais conforme se observa cidades com poucos habitantes ou então afastadas dos grandes centros comerciais, pois os profissionais treinados tendem a recusar propostas que exijam residir nestes locais. Neste cenário, a venda guiada poderia ser utilizada por um vendedor experiente para que, mesmo sem o treinamento específico para revestimento, consiga realizar um bom trabalho.

A plataforma de *e-commerce* da empresa atualmente disponibiliza uma grande quantidade de informações sobre seus produtos, mas estas são em sua maioria informações técnicas e requerem um alto grau de conhecimento por parte do usuário. Mas a grande maioria dos usuários pode ser considerada leiga em se tratando de tintas industriais, logo, não possuem o conhecimento mínimo necessário para compreender o que tais informações significam e acabam por abandonar a plataforma antes de selecionar um produto e concretizar sua compra. Novamente destaca-se a demanda potencial para um sistema de venda guiada, desta vez com o intuito de deixar as informações mais claras para o usuário, incluindo explicações detalhadas do que cada uma das etapas da seleção estará fazendo e como cada característica pode impactar no produto final.

1.4 OBJETIVOS DO TRABALHO

Os objetivos do presente trabalho foram subdivididos em um objetivo geral e uma lista de objetivos específicos, conforme descrito a seguir.

1.4.1 Objetivo geral

Sistematizar e disponibilizar o conhecimento do setor de vendas da empresa WEG Tintas para melhorar o controle de seu portfólio e facilitar a rotina de trabalho de seus vendedores através de um protótipo de Sistema Especialista integrado a técnicas de árvore de decisão.

1.4.2 Objetivo específico

- Sistematizar o conhecimento do setor de vendas e o estruturar como um banco de dados "flat";

- Desenvolver um protótipo para realizar o gerenciamento do conhecimento;
- Criar uma estrutura lógica incrementável para que a empresa possa ampliar o protótipo;
- Testar e validar o protótipo, avaliando seu impacto na rotina da empresa.

1.5 ESTRUTURA DO TRABALHO

Neste primeiro capítulo buscou-se contextualizar a inteligência artificial dentro do escopo do projeto, apresentar brevemente a empresa que prestou auxílio tanto intelectual quanto financeiro para sua realização, mostrar os objetivos do trabalho e os resultados esperados quando o mesmo estiver concluído.

O capítulo 2 apresenta conceitos sobre o universo das tintas e revestimentos, suas peculiaridades e as principais diferenças físicas e químicas dentre as linhas de produtos da WEG. A necessidade de se utilizar tintas também foi abordada, quais os inconvenientes que elas podem evitar, com ênfase especial para a corrosão.

O capítulo 3 traz uma revisão bibliográfica mais aprofundada sobre inteligência artificial, sistemas especialistas e mineração de dados, explicando com maiores detalhes os fundamentos e conceitos que foram utilizados para a criação do protótipo.

O capítulo 4 é dedicado a criação do protótipo, apresentando a metodologia aplicada em sua concepção, e a forma com que o conteúdo dos capítulos anteriores foi utilizado para realizar as etapas de análise, aquisição, representação e implementação. Apresentando as classes desenvolvidas e como estas interagem para que o protótipo execute sua tarefa.

O capítulo 5 apresenta as etapas de verificação e validação do protótipo, onde tanto o engenheiro de conhecimento quanto os especialistas da empresa fazem suas avaliações para aprovar o protótipo, no fim do capítulo são apresentadas imagens e explicações detalhadas sobre a sua execução.

O capítulo 6 finaliza este trabalho com conclusões acerca dos resultados obtidos, as contribuições geradas pela pesquisa e sugestões para trabalhos futuros dentro desta área.

2. INDÚSTRIA DE REVESTIMENTOS

Este capítulo está dividido em quatro partes: a primeira explica a função e os problemas que a utilização ou não de revestimentos (tintas) pode causar a estruturas, a segunda aborda as propriedades das tintas líquidas, a terceira sobre as propriedades das tintas em pó e a quarta parte sobre o processo de seleção de uma tinta de acordo com suas propriedades e do ambiente a que será exposta.

Revestir superfícies através de técnicas de pintura é uma prática que vêm sendo utilizada há milhares de anos, sendo que inicialmente o propósito da pintura era puramente estético, conferindo uma característica distinta para um objeto visando criar um diferencial e possivelmente aumentar seu valor comercial ou beleza para fins decorativos (TIWARI, RAWLINS e HIHARA, 2015).

No período da revolução industrial, quando começou a se pensar em maneiras de aumentar a resistência das máquinas e equipamentos industriais os revestimentos começaram a ter um papel mais significativo iniciando efetivamente a ideia das indústrias de revestimentos que existem atualmente (MAKHLLOUF, 2014).

De acordo com Kränkel (2015), existem muitas características que podem influenciar no sucesso da aplicação de um revestimento ou na sua falha, o sucesso significa que o revestimento protegerá a superfície e manterá sua aparência durante o tempo previsto, e a falha que a proteção e/ou a aparência da tinta ficará comprometida antes do esperado. As principais causas para estas falhas são:

- Escolher uma tinta especificada para outro ambiente;
- Utilizar tintas incompatíveis entre uma camada e a posterior;
- Misturar catalisadores ou diluentes na proporção errada, ou ainda catalisadores incompatíveis;
- Aplicar uma camada superior à especificada;
- Não respeitar a distância ideal para aplicação (quando feita com pistola de ar);
- Não aguardar o tempo de cura (secagem completa) das tintas;

2.1 FUNÇÃO DOS REVESTIMENTOS

Antes de iniciar uma explicação aprofundada sobre os tipos de revestimentos, serão apresentados os principais fatores que acarretam na necessidade de sua utilização.

Equipamentos de todos os tipos e de todos os materiais estão sujeitos as mais diversas ações, sejam do meio-ambiente, impactos mecânicos com outros componentes, e até mesmo reações que podem ocorrer internamente. Muitas destas ações e reações são maléficas para o equipamento em questão, e os revestimentos servem para oferecer certo grau de proteção aumentando sua vida útil (KNUDSEN e FORSGREN, 2017).

As principais funções para os revestimentos são: servir como uma barreira entre uma superfície e o meio corrosivo a fim de evitar a corrosão da superfície, eliminar bactérias e microrganismos presentes no meio ou evitar sua proliferação, antiderrapantes para pisos, marcadores, demarcadores e sinalizadores para áreas específicas ou restritas, anti-chamas e intumescentes, resistência a altas temperaturas, dentre outras (KNUDSEN e FORSGREN, 2017).

A corrosão e os microrganismos são os principais problemas que necessitam de revestimentos e serão melhor apresentados nos tópicos a seguir.

2.1.1 Corrosão

A corrosão pode ser definida como a deterioração de materiais (em geral metálicos) causada por esforços mecânicos, reações físico-químicas ou eletroquímicas geralmente associadas ao meio ambiente que o material se encontra. Esta deterioração normalmente representa alterações indesejáveis nas propriedades físico-químicas dos materiais, como desgaste, modificações estruturais e variações químicas, podendo reduzir o tempo de vida útil e qualidade do produto ou até mesmo inutiliza-lo (Gentil, 2006).

De acordo com a norma DIN EN ISO 8044 (2015, tradução nossa), a corrosão é uma "Interação física entre um metal e seu ambiente que resulta em mudanças das propriedades do metal e que pode levar ao comprometimento de funções do metal, do ambiente ou do sistema técnico que estes façam parte".

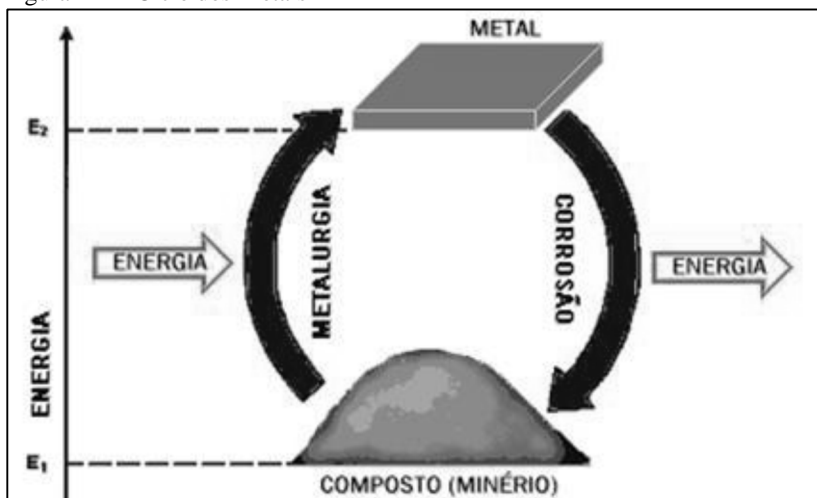
A norma ABNT NBR 15156 (2004) diz que a corrosão é a "Deterioração de um material geralmente metálico ou não, por ação química ou eletroquímica aliada ou não a esforços mecânicos"

Para os metais o conceito de corrosão é muito bem definido, porém existem controvérsias no meio científico quanto à caracterização da deterioração dos materiais conhecidos por não-metálicos ser ou não resultado de corrosão. Como por exemplo, a perda de elasticidade da borracha devida à oxidação através do ozônio presente no ar, a hidrólise da celulose da madeira quando exposta a ácidos, a deterioração do concreto por ação de sulfatos, alguns autores consideram todos como efeitos de corrosão, enquanto outros acabam por utilizar termos diferentes, mas que, fundamentalmente representam a mesma ideia (CICEK, 2017).

Por se tratar de um fenômeno normalmente espontâneo e se apresentar de maneira acelerada em materiais metálicos, ele assume um papel de grande importância nos dias atuais, onde grande parte dos componentes fabricados e utilizados possuem metais e ligas metálicas em suas composições.

Em alguns casos, admite-se a corrosão como o fenômeno inverso do metalúrgico, revertendo o metal para um formato semelhante ao encontrado na natureza, conforme apresentado na Figura 2-1. A composição mais abundante na natureza do óxido de ferro é a hematita, Fe_2O_3 , e a ferrugem que surge sobre metais ferrosos é o óxido de ferro hidratado, $\text{Fe}_2\text{O}_3 \cdot n\text{H}_2\text{O}$, ou seja, o metal tendendo a voltar a sua forma natural e estável (LAHIRI, 2017).

Figura 2-1 – Ciclo dos Metais



Fonte: Acervo WEG (2017)

2.1.1.1 Importância do estudo da corrosão

A corrosão gera frequentes problemas nas mais diversas atividades, como por exemplo: Indústrias petroquímicas, químicas, naval, construção civil, transporte aéreo, metroviário, ferroviário, marítimo, meios de comunicação, e até mesmo em atividades menores como restaurações odontológicas metálicas, próteses ortopédicas e obras de arte (POPOV, 2015).

De acordo com o National Physics Laboratory (NPL, 2017) a corrosão pode gerar prejuízos diretos e indiretos, tais como:

- Custo de manutenção em peças equipamentos, incluindo gastos com energia e mão-de-obra;
- Substituição de peças quando a corrosão atinge níveis críticos;
- Redução do rendimento, e aumento do consumo de energia (redução da transferência de calor, aumento de atrito);
- Perda de produtos e matérias-primas (vazamento de gás, óleo, líquidos em geral), ou então sua contaminação por produtos resultantes da corrosão;
- Equipamentos parados durante as manutenções não previstas;
- Acidentes (vazamentos de líquidos podem deixar a área em volta escorregadia, líquidos e gases podem causar explosões, estruturas como pontes ou prédios cedendo)

Todos estes problemas podem gerar não somente perdas financeiras, como de vidas humanas. Alguns casos históricos de problemas ocorridos devido à negligência com a corrosão podem ser citados: em 1952 três aviões Comet se desintegraram em pleno voo, em 1988 um Boeing 737-200 perdeu parte de sua fuselagem, a queda da ponte Silver Bridge em Ohio em 1967, na Suíça em 1987 a cobertura de uma piscina térmica caiu em cima de seus ocupantes, dentre centenas de outros casos (Gentil, 2006).

No Brasil um exemplo muito conhecido foi a queda de uma grade de alumínio no estádio do Maracanã em 1992 devido a parafusos corroídos, resultando em 3 fatalidades e diversos feridos.

Casos mais recentes incluem o vazamento de água radioativa dos tanques de armazenamento da usina de Fukushima em 2013, neste mesmo ano uma seção de 120 metros da ponte de Wisconsin, nos Estados Unidos

cedeu cerca de 60 cm devido à corrosão em sua base de sustentação (NACE, 2017)

2.1.1.2 Mecanismos dos processos corrosivos

De uma forma geral, os processos corrosivos podem ser classificados em dois grandes grupos, abrangendo todos os casos de deterioração por corrosão existente na natureza, a eletroquímica e a química (NPL, 2017).

2.1.1.2.1 Corrosão eletroquímica

É o processo de corrosão mais frequente na natureza, em geral ocorre na presença de água através do transporte de eletricidade através do contato dos diferentes elementos que compõe um metal com eletrólitos, formando pilhas de corrosão (CICEK, 2017) conforme apresentado na da Figura 2-2.

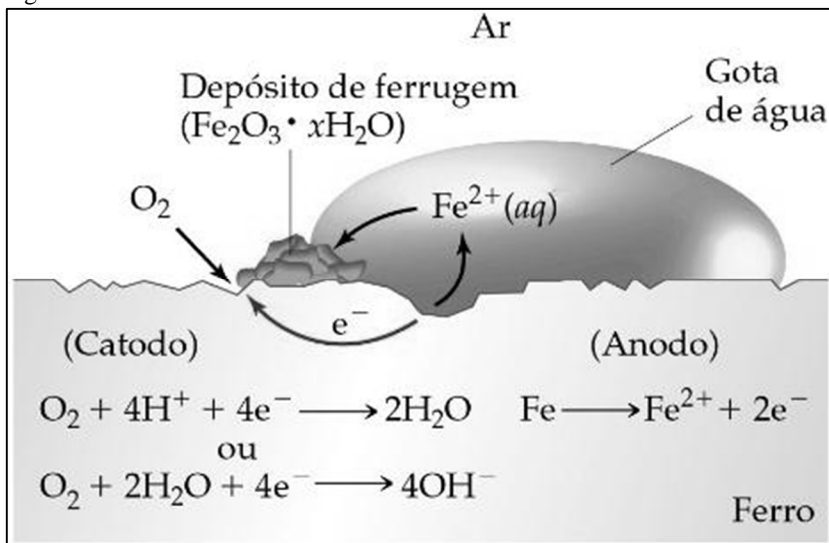
De acordo com Arruda (2009), ocorre principalmente por conta de:

- Corrosão em água ou em soluções aquosas;
- Corrosão atmosférica;
- Corrosão no solo;
- Corrosão em sais fundidos.

Quando em temperatura abaixo do ponto de orvalho, a umidade do ar se condensa na forma de gotículas de água de maneira natural sobre as superfícies que está em contato, estas conduzindo os eletrólitos para propiciar a ocorrência deste tipo de corrosão. Como normalmente a temperatura ambiente se encontra abaixo deste ponto, este tipo de corrosão pode ocorrer em praticamente todos os locais (ELAYAPERUMAL e RAJA, 2015).

O ambiente marítimo é considerado o mais corrosivo destes ambientes naturais, pois apresenta uma umidade relativamente alta aliada a uma grande concentração de sais minerais dissolvidos na água do mar, em especial o cloreto de sódio que se está presente em uma concentração de 0,5 mol.L, uma concentração que representa um pico em uma relação corrosividade/concentração (MCCAULEY, 2013).

Figura 2-2 – Pilha de Corrosão



Fonte: ARRUDA, 2009

2.1.1.2.2 Corrosão química

Também denominada de corrosão seca ou em meio não-aquoso, são menos frequentes na natureza envolvendo simplesmente o contato entre o material e o meio corrosivo, por vezes aliado a altas temperaturas.

As principais características deste tipo de corrosão são:

- Ocorrem necessariamente na ausência de água
- Não existe deslocamento de elétrons, apenas a interação direta entre o material e o meio.

As substâncias agressivas com maior índice de ocorrência são: enxofre, hidrogênio, amônia, carbono, sódio e vanádio (CICEK, 2013).

2.1.1.3 Formas de Corrosão

A corrosão pode ser analisada utilizando critérios morfológicos ou avaliando as causas, meios e mecanismos envolvidos em seu surgimento, existem ensaios específicos que são utilizados para determinar tanto como esta ocorreu como também monitorar sua evolução (MIRANDA, 2009). As principais maneiras de se classificar a corrosão estão descritas no Quadro 2.1 e nos tópicos a seguir.

Quadro 2.1 – Formas de corrosão

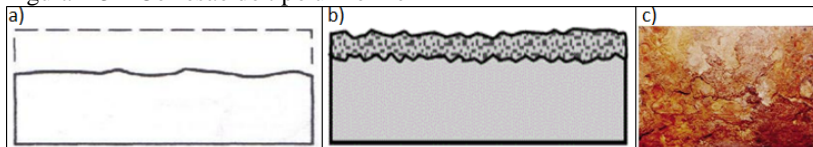
Classificação	Característica
Uniforme	Corrosão por toda a extensão da superfície
Placas	Corrosão em regiões separadas, mais largas que profundas;
Alveolar	Corrosão em regiões separadas, mais profundas que largas;
Puntiforme	Pontos de corrosão, de largura pequena e profunda com fundo em forma angulosa
Frestas	Corrosão na região de contato entre duas superfícies
Galvânica	Corrosão devido ao contato entre ligas com potenciais de corrosão distintos
Intergranular	Corrosão entre os grãos da rede cristalina dos metais
Transgranular	Corrosão que atravessa os grãos da rede cristalina dos metais
Filiforme	Corrosão em filamentos pouco profundos e compridos
Tensão/Fadiga	Corrosão devido a esforços físicos aliados a um meio corrosivo

Fonte: Elaborado pelo Autor.

2.1.1.3.1 Uniforme

A corrosão atinge toda a extensão da superfície metálica, sem uma área preferencial para seu crescimento, ocorre quando a superfície é relativamente uniforme e o meio corrosivo possui acesso à toda sua extensão (PEREZ, 2016).

Figura 2-3 – Corrosão do tipo uniforme

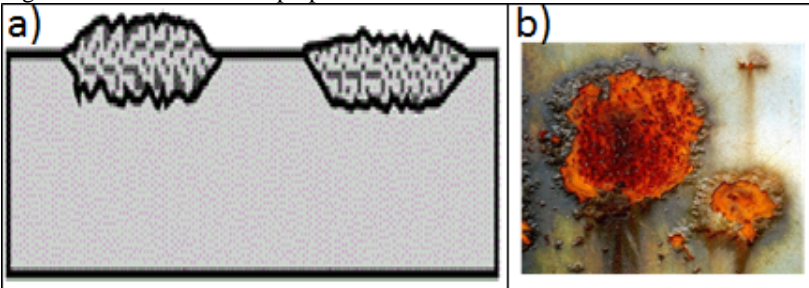


Fonte: a) Matos (2003) b) Nunes (2014) c) Gentil (2006)

2.1.1.3.2 Placas

A corrosão ocorre em algumas regiões da superfície, formando escavações irregulares mais largas que profundas (LAZZARI, 2017).

Figura 2-4 – Corrosão do tipo placas

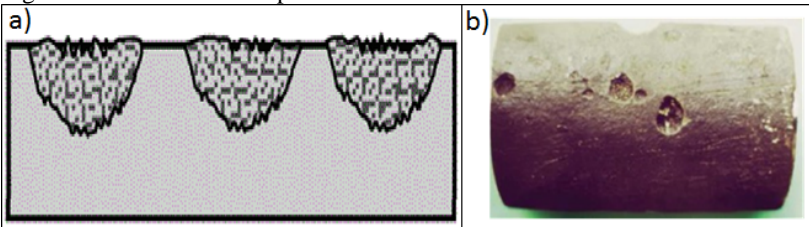


Fonte: a) Nunes (2014) b) Gentil (2006)

2.1.1.3.3 Alveolar

A corrosão produz escavações ou sulcos de fundo arredondado e geralmente com maior profundidade que diâmetro (MILELLA, 2013).

Figura 2-5 – Corrosão do tipo alveolar

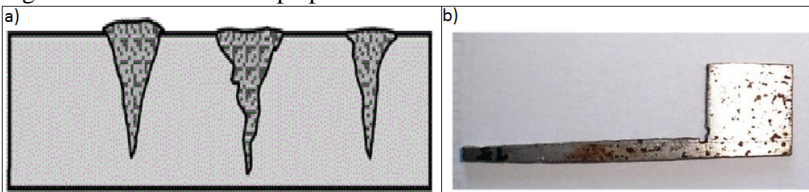


Fonte: a) Nunes (2014) b) Gentil (2006)

2.1.1.3.4 Puntiforme

Também conhecida apenas como pite, a corrosão ocorre em pequenas áreas da superfície metálica, são mais profundas que seu diâmetro e possuem um fundo em forma angulosa. Sua pequena dimensão dificulta sua localização durante inspeções visuais (PEREZ, 2016).

Figura 2-6 – Corrosão do tipo puntiforme



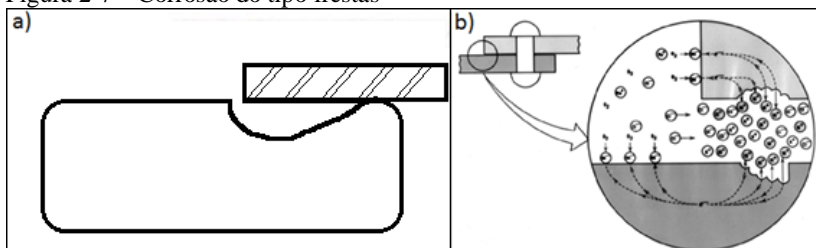
Fonte: a) Nunes (2014), b) (SILVA, et. Al. 2006)

2.1.1.3.5 Frestas

Ocorre de duas maneiras distintas:

- Regiões isoladas onde ocorra o contato entre dois materiais diferentes e um forneça uma condição que propicie a corrosão do outro.
- Regiões isoladas onde ocorra a junção de dois materiais iguais ou distintos, e a junção destes crie uma região que favoreça o acúmulo de água ou qualquer outra substância que cause a corrosão em um ou em ambos os materiais (MILELLA, 2013).

Figura 2-7 – Corrosão do tipo frestas

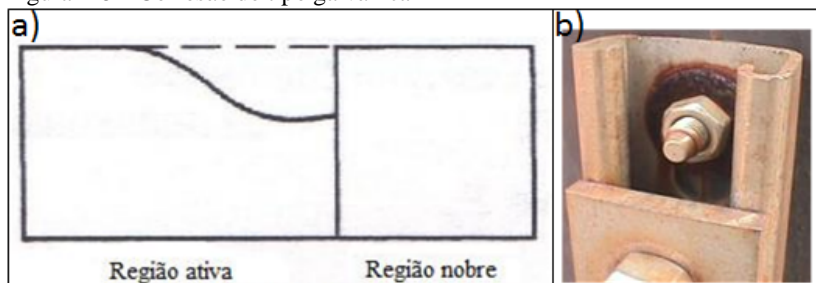


Fonte: Adaptado de Matos (2003)

2.1.1.3.6 Galvânica

Quando metais de ligas metálicas com diferentes valores de potencial de corrosão estão em contato tanto entre si quanto com um meio corrosivo, a liga mais nobre estará protegida, pois a corrosão tende a ocorrer somente na outra. Conforme se distancia do ponto de contato entre as duas ligas a corrosão tende a normalizar (LAZZARI, 2017).

Figura 2-8 – Corrosão do tipo galvânica



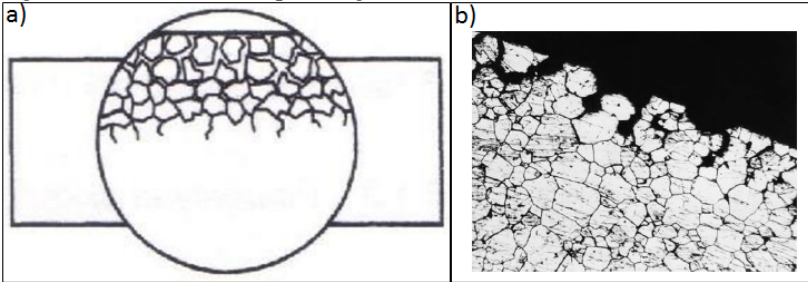
Fonte: Matos (2003), Acervo WEG

2.1.1.3.7 Intergranular

Corrosão que ocorre entre os grãos da rede cristalina do material metálico, ocorre quando impurezas reativas presentes na liga ou

elementos passivadores se precipitam para os interstícios dos grãos, tornando essa região mais suscetível à corrosão (CICEK, 2013).

Figura 2-9 – Corrosão do tipo intergranular

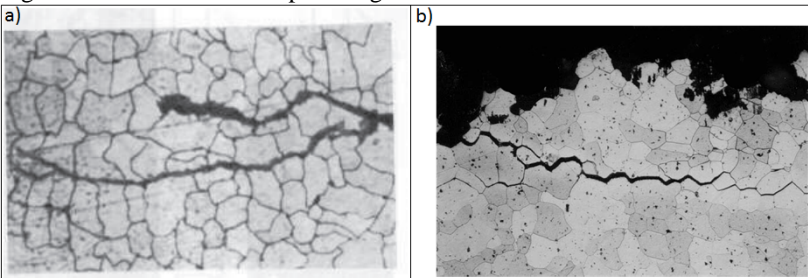


Fonte: a) Matos (2003), Acervo WEG

2.1.1.3.8 *Transgranular*

Ocorre quando a corrosão atravessa internamente os grãos da rede cristalina, podendo chegar a atravessar toda a superfície e deixando o material extremamente suscetível a fraturas por esforços mecânicos (POPOV, 2015).

Figura 2-10 – Corrosão do tipo transgranular

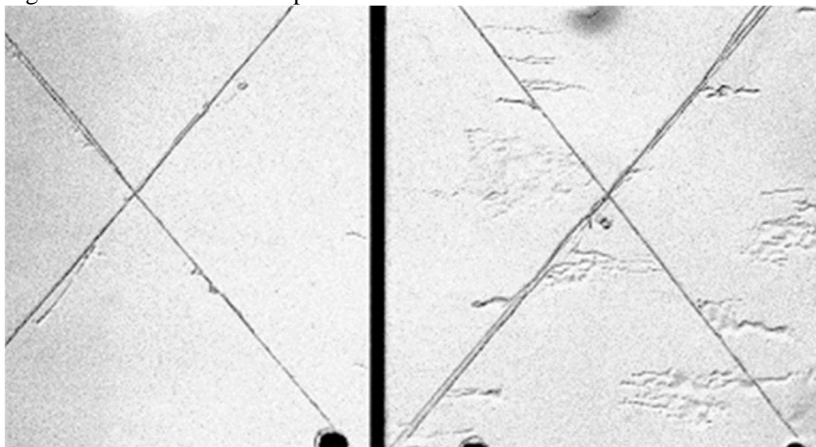


Fonte: a) Nunes (2014), Acervo WEG

2.1.1.3.9 *Filiforme*

Ocorre sob a forma de pequenos filamentos pouco profundos que se propagam em todas as direções, ocorrendo normalmente em materiais revestidos com tintas (ELAYAPERUMAL e RAJA, 2015).

Figura 2-11 – Corrosão do tipo filiforme



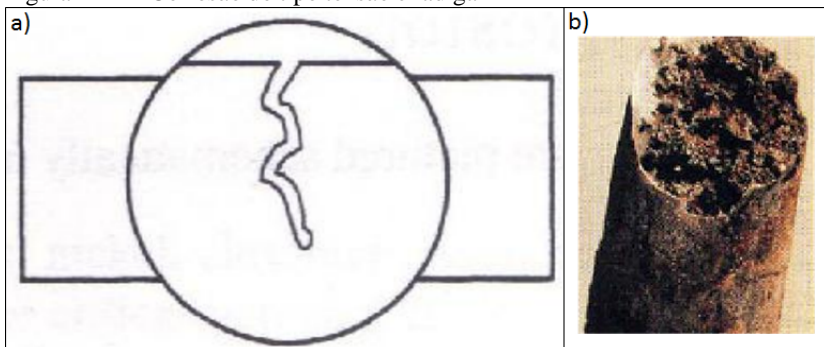
Fonte: Conpleq Consultoria (2016)

2.1.1.3.10 Tensão ou Fadiga

Ocorrem devido a esforços mecânicos cíclicos quando o material está em contato com um meio corrosivo. O meio corrosivo reduz a resistência à fadiga que o material originalmente possuía, causando uma ruptura em um tempo menor que o esperado.

Os casos mais comuns são: Cabos de aço submarinos, Eixos de hélices de barcos, Timões, Tubos de evaporadores, Caldeiras, Componentes de turbinas, motores e bombas, Tubulações transportadoras de líquidos corrosivos (LAPEC, [201-?]).

Figura 2-12 – Corrosão do tipo tensão e fadiga



Fonte: a) Matos (2003), b) Gentil (2006)

2.1.1.4 Maneiras de Mitigar a Corrosão

Os métodos mais práticos para reduzir a taxa de corrosão de materiais podem ser realizados modificando o processo, o meio corrosivo, os materiais ou através de revestimentos protetores (Gentil, 2006 apud Vernon, 1957).

- Modificação do Processo:
 - Projeto da estrutura;
 - Condições da superfície;
 - Proteção catódica.
- Modificação do Meio Corrosivo:
 - Desaeração da água ou solução neutra;
 - Purificação ou diminuição da umidade do ar;
 - Adição de inibidores de corrosão.
- Modificação dos Materiais:
 - Aumento da pureza;
 - Adição de novos elementos químicos;
 - Tratamentos térmicos.
- Aplicação de Revestimentos Protetores:
 - Tratamentos químicos ou eletroquímicos da superfície;
 - Revestimentos orgânicos, inorgânicos ou metálicos;
 - Protetores temporários.

Ao longo do trabalho é mostrada a utilização de revestimentos protetores como técnica para o controle da corrosão, pois geralmente é a que exige menos esforço, tempo e custos, além de apresentar manutenção relativamente simples para ser executada por terceiros.

2.1.1.5 Casos benéficos da corrosão

Apesar de vários processos de corrosão serem indesejáveis, existem aqueles que são benéficos e de grande importância industrial. Gentil (2006) cita alguns exemplos:

- Oxidação das ligas de ferro e cromo formando os aços inoxidáveis;
- Anodização do alumínio e suas ligas conferindo uma camada protetora e de aspecto decorativo;
- Fosfatização de superfícies metálicas para melhorar a aderência;

- Proteção catódica com ânodos de sacrifício ou galvânicos para proteção em instalações submersas ou enterradas;
- Corrosão superficial em bronze conferindo aspecto decorativo característico para monumentos e esculturas.

2.1.2 Microrganismos

Os microrganismos podem ser definidos de maneira resumida como o grupo das bactérias, arqueas, fungos, protozoários e vírus (PERUCHI, 2011). Por possuírem dimensões ínfimas, podem ser difíceis de serem identificados até que estejam em uma concentração grande o suficiente para gerar manchas visíveis ou causar doenças (TIMENETSKY, 2012).

De maneira geral, deseja-se eliminar ou controlar o crescimento de microrganismos, pois estes podem ser nocivos aos seres humanos ou ao meio ambiente ou aos equipamentos que tiverem contato. A Quadro 2.2 apresenta algumas características que propiciam o crescimento de microrganismos:

Quadro 2.2 – Características para o crescimento de microrganismos

<i>Características</i>	<i>Bactérias</i>	<i>Fungos</i>	<i>Algas</i>
Luz	Não necessita	Não necessita	Necessita
pH ideal	Levemente alcalino	Levemente ácido	Neutro
T ideal	25 – 40°C	20 – 35°C	15 – 30°C
Nutrientes	C, H, N	C, H, N	CO ₂
Oxigênio	Aeróbio ou anaeróbio	Aeróbio	Aeróbio
Água	Líquida ou vapor	Líquida ou vapor	Líquida ou vapor

Fonte: Bechtold, 2011 apud Matteucci, 2009

2.1.2.1 Emersos

Muitos ambientes possuem abundância destes elementos, e portanto são ambientes propícios ao surgimento de microrganismos: casas, hospitais, asilos, creches e indústrias alimentícias devem ter cuidados especiais (PERUCHI, 2011).

Além dos cuidados com a higiene e limpeza destes locais, normalmente se recomenda a utilização de revestimentos que contenham biocidas em sua composição para garantir que os microrganismos não se

desenvolvam para causar mal às pessoas que convivem nestes ambientes (McCAULEY, 2013).

2.1.2.2 Submersos

Ambientes submersos apresentam um grande problema que se inicia com o crescimento de microrganismos nas superfícies, e estes propiciam o crescimento de seres de maior complexidade e tamanho, conhecidos como bioincrustação.

A bioincrustação nada mais é que o acúmulo de microrganismos, animais e/ou algas em estruturas submersas. Este processo pode causar danos diretos e indiretos e tanto ambientais quanto financeiros (ABREU, 2016).

Como exemplo de danos ambientais, pode-se citar o transporte de espécies entre regiões (quando incrustadas em embarcações móveis) gerando desequilíbrios nos ecossistemas chegando até mesmo a eliminar populações de organismos aquáticos importantes como alimentos para outras espécies e/ou comercializados (ALMEIDA et al., 2007).

A bioincrustação acelera o processo de corrosão, causando grandes prejuízos em embarcações ou outras estruturas manufaturadas, aumentando a necessidade de manutenção e docagens, podendo provocar ainda acidentes, entupimento de tubulações e inutilização de equipamentos. Atua indiretamente aumentando o peso total da embarcação e aumentando a superfície de contato da embarcação com a água, o que leva a um aumento das forças de arraste ocasionando um aumento no consumo de combustível (YEBRA et al., 2004; KOTRIKLA, 2009; PERINA, 2009).

De acordo com Feio (2015), "Um navio sem um revestimento anti-incrustante pode conduzir a um aumento no seu consumo de combustível na ordem dos 40%, podendo chegar aos 70% em viagens de longo percurso. Com o aumento do consumo de combustível, aumentam as emissões dos gases de efeito de estufa." Na Figura 2-13, é possível ver a diferença entre uma chapa com e outra sem biocidas após certo tempo imersas.

Figura 2-13 – Placa sem biocida (esquerda) e Placa com biocida (direita)



Fonte: Endures, 2010

O processo de bioincrustação ocorre de modo progressivo, em um primeiro momento materiais orgânicos vão se depositando na superfície atraindo bactérias e formando uma espécie de biofilme, este serve de alimento para que algas efêmeras e perenes comecem um processo de assentamento, criando um ambiente atrativo para seres maiores como esponjas, cracas, mexilhões, dentre outros. O tipo de espécies incrustadas varia entre regiões, mas seu efeito nocivo é semelhante (PERINA, 2009). A Figura 2-14 apresenta uma ideia do tempo necessário para a colonização das incrustações.

Figura 2-14 - Organismos marinhos incrustantes



Fonte: Feio, 2015 apud Rosenhahn et. al, 2010

Para controlar o surgimento das bioincrustações existem dois grupos distintos de ações que podem ser tomadas: reativos e proativos.

Os reativos são os métodos aplicados para remover as incrustações já presentes na estrutura. Estes métodos normalmente são utilizados quando as incrustações não representam grandes problemas ou perigos para as instalações, e podem ser simplesmente removidas, são representadas por (MCCAULEY, 2013):

- Choque Térmico in loco;
- Limpeza Mecânica in loco;
- Retirada da água e limpeza mecânica;
- Retirada da água e congelamento.

Os proativos são os métodos aplicados de maneira preventiva, de modo a evitar o surgimento das incrustações e diminuir os riscos de danos precoces as estruturas (PEREZ, 2016).

- Proteção eletrolítica
- Métodos acústicos
- Tratamento UV
- Revestimentos químicos com biocidas

- Revestimentos químicos sem biocidas

Em relação dos revestimentos químicos, objeto de estudo deste trabalho temos as classes com e sem biocidas.

Os revestimentos com biocidas atuam através da liberação em pequenas doses de substâncias nocivas ou repelentes para manter afastados ou eliminar os microrganismos e assim evitar que estes se fixem na estrutura. Diversas leis regulamentam as matérias-primas que podem ser utilizadas nestes revestimentos para evitar agressão demasiada aos ambientes marinhos (MCCAULEY, 2013).

Os revestimentos sem biocidas utilizam com matérias-primas baseadas em silicone e criam uma superfície extremamente lisa, fazendo com que os microrganismos não consigam se fixar na estrutura.

2.2 TINTAS LÍQUIDAS

Tinta líquida pode ser considerada uma mistura química, geralmente viscosa e constituída por um ou mais pigmentos homogeneizados dentro de um aglomerante líquido que quando estendido em uma película fina sofre processo de cura e se torna um filme aderente ao substrato (DONADIO, 2011).

O Quadro 2.3 apresenta algumas vantagens e desvantagens das tintas líquidas sobre as tintas em pó.

Quadro 2.3 – Vantagens e desvantagens das tintas líquidas

<i>Vantagens</i>	<i>Desvantagens</i>
Tecnologia Conhecida	Alta dependência do petróleo
Fácil obtenção de camadas finas	Alto custo por m ² pintado
Trocas de cor rápidas	Necessidade de preparação (diluição)
Maior variedade de cores	Riscos de incêndio
Sistema tintométrico (tintas pré-fabricadas para rápida produção)	Processo altamente poluente, gerando efluentes industriais.
	Alto índice de rejeição
	Custo operacional maior

Fonte: Adaptado de (Kränkel, 2015)

2.2.1 Matérias-primas

De acordo com Donadio (2011), a composição básica de uma tinta líquida é uma mistura de resinas, pigmentos, solventes e aditivos, conforme apresentado nos itens a seguir:

2.2.1.1 Resina

"Componente de uma tinta que confere a esta as propriedades de coesão e adesão." (ABNT, 2004), também conhecida como ligante ou veículo a resina é uma das partes não voláteis da tinta e um dos constituintes que mais caracteriza suas propriedades finais, é a responsável pela formação do filme, aderência ao substrato, impermeabilidade, flexibilidade e auxilia na dispersão das partículas de pigmento. Normalmente as tintas são nomeadas e classificadas de acordo com a resina básica que as compõe, como acrílicas, alquídicas, epoxídicas, poliuretânicas, dentre outras.

Em tintas bicomponentes, a resina presente no componente B normalmente é um agente de cura, responsável por reagir com a resina do componente A para formar a película de tinta (TIWARI, RAWLINS e HIHARA, 2015).

2.2.1.2 Pigmento

"Componente sólido finamente dividido, presente na tinta, que tem como principais finalidades conferir cor, opacidade e propriedades anticorrosivas à película" (ABNT, 2004). Também são utilizados para encorpar a película de tinta. De acordo com Mannari e Patel (2015) podem ser classificados de acordo com sua:

- Natureza: orgânicos e inorgânicos
- Finalidade: tintoriais, cargas (inertes), anticorrosivos e especiais
- Ação: ativos e inertes.

2.2.1.2.1 Tintoriais

São utilizados principalmente para conferir cor e opacidade a tinta, normalmente inseridos apenas em tintas de acabamento. São encontrados com praticamente todos os tipos de cores: primárias, secundárias, terciárias e até mesmo as metálicas, mas o mais comum é se utilizar uma combinação de cores básicas para que a tinta final resulte na cor desejada. Os pigmentos mais comuns são: dióxido de titânio (Branco), amarelo de cromo, azul de ftalocianina, laranja de cromo, óxido de ferro (vermelho),

negro de fumo (preto) e alumínio (metálico) (KNUDSEN e FORSGREN, 2017).

2.2.1.2.2 *Cargas*

São pigmentos utilizados para encorpar e reforçar a tinta, tais elementos auxiliam a regular o brilho final e por serem em sua maioria translúcidos interferem pouco em sua tonalidade final. Servem como substituto de menor custo quando as tintas não necessitam de grande resistência ou quando é necessário aumentar o percentual de sólidos por volume. Os mais comuns são a calcita, dolomita, barita, sílica e caulim (MANNARI e PATEL, 2015).

2.2.1.2.3 *Anticorrosivos*

São caracterizados por conferir propriedades anticorrosivas as tintas, são mais encontrados em tintas de fundo. São divididos de acordo com o tipo de proteção que conferem: passivação anódica, proteção catódica ou proteção por barreira. Os mais utilizados são o cromato de zinco, fosfato de zinco e zinco metálico (MANNARI e PATEL, 2015).

2.2.1.2.4 *Especiais*

Os pigmentos especiais são os capazes de conferir propriedades diferentes das mencionadas anteriores. Os mais utilizados são para auxílio em caso de incêndios (anti-chamas e intumescentes), antibactericidas, antiincrustrantes, impermeabilizantes, fluorescentes. (DONADIO, 2011).

2.2.1.3 *Solvente*

"Constituinte da tinta com a finalidade de solubilizar o veículo e ajustar a viscosidade para aplicação" (ABNT, 2004). É um composto orgânico 100% volátil, ou seja, após a aplicação da tinta irá evaporar, é necessário pois sem ele as tintas apresentariam uma viscosidade muito elevada para serem aplicadas em finas camadas.

Sua utilização em tintas deve ser em uma proporção bem definida, pois possui muitos inconvenientes se utilizado em excesso: reduzir muito a película de tinta após sua evaporação, criar poros e crateras em pontos não tão diluídos, representa um custo adicional pois será evaporado e "perdido" (KNUDSEN e FORSGREN, 2017).

Por ser volátil, este elemento normalmente é fracionado em duas partes, a primeira fica dentro da própria tinta servindo para esta ter uma viscosidade suficiente para poder ser misturada e manipulada pelos pintores, mas a segunda parcela fica em uma embalagem separada para poder além de ajustar a viscosidade para a hora da aplicação, poder

compensar eventuais perdas por evaporação. Solventes possuem compatibilidade limitada com determinados tipos de resina, a Quadro 2.4 apresenta os solventes apropriados para as principais resinas (DONADIO, 2011).

Quadro 2.4 - Tipos de resinas e solventes apropriados

<i>Resina</i>	<i>Solvente</i>
Alquílica	Aguarrás, xileno ou mistura destes
Acrílica	Misturas de acetatos, xileno, cetonas e álcoois cíclicos
Epóxi	Misturas de MEK e MIBK com xileno e álcool butílico
Poliuretano	Misturas de acetato de etila, acetato de butila e MEK.
Etil silicato de Zinco	Xileno e Tolueno

Fonte: Adaptado de Kränkel (2015)

2.2.1.4 Aditivos

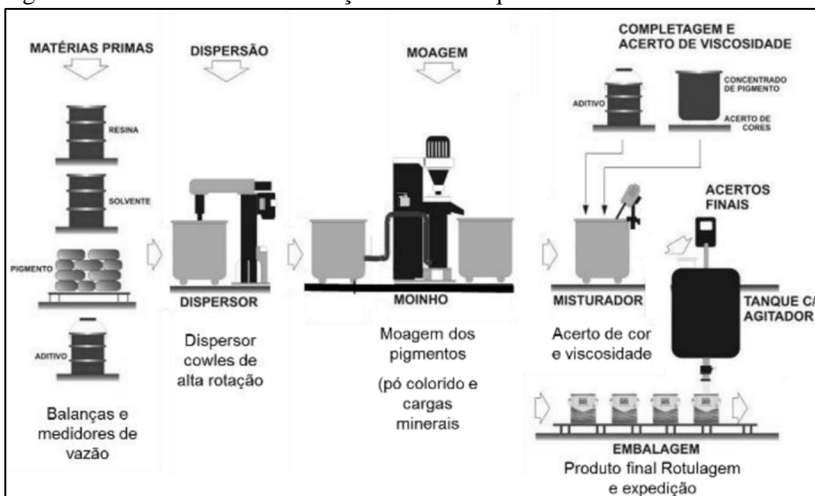
"Substâncias integrantes de composição que conferem à tinta líquida ou à película formada propriedades específicas e/ou desejadas." (ABNT, 2004), são utilizados em frações que normalmente não ultrapassam 1% do total por aditivo, mas que possuem influência significativa na manufatura, aplicação, aspecto e qualidade do filme de tinta (REIS, 2012). Knudsen e Forsgren (2017) divide os aditivos em quatro grandes grupos, cada um com sua função, existindo aditivos com mais de uma função:

- Cinética: Secantes, catalisadores e antipeles;
- Reologia: Espessantes e antiescorrimento;
- Processo: Surfactantes, umectantes, dispersantes, antiespumantes e nivelantes;
- Preservação: Biocidas, estabilizantes de ultravioleta.

2.2.2 Processo de fabricação

A Figura 2-15 apresenta um desenho esquemático com a sequência de etapas para o processo de fabricação de tintas líquidas, e logo em seguida é apresentado cada uma de suas etapas.

Figura 2-15 – Processo de fabricação de tintas líquidas



Fonte: Kränkel, 2015

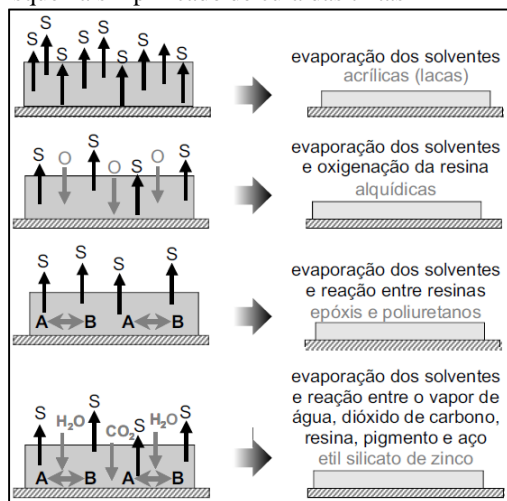
- Pesagem das Matérias-primas: As matérias-primas são pesadas e adicionadas a um recipiente (tacho) seguindo a proporção definida pelos formuladores;
- Dispersão: Esta é uma etapa de pré-mistura, as resinas, cargas e uma parcela dos solventes são misturadas e dispersadas em um misturador de alta rotação;
- Moagem: A mistura é encaminhada para um moinho (esferas ou de rolos) onde fica circulando até atingir o grau de fineza desejado em toda a mistura;
- Completagem: Aditivos e solventes são adicionados e homogeneizados em um misturador de baixa rotação;
- Acertos Finais: Controle de qualidade avalia se os parâmetros da tinta estão de acordo, ou se novas matérias-primas devem ser adicionadas;
- Embalagem e Expedição: Tintas são transferidas para embalagens menores e armazenadas para serem vendidas;

2.2.3 Formação da película de tinta

A formação da película seca, normalmente tratada por cura da tinta pode acontecer pela evaporação direta dos solventes, ou através de um conjunto de reações químicas, podendo estas reações ocorrerem pela oxigenação da tinta através do contato com o ar, pela reação forçada

através da mistura de resinas diferentes (tintas bicomponentes) ou pela combinação de qualquer um destes 3 fatores, conforme apresentado na Figura 2-16 (TIWARI, RAWLINS e HIHARA, 2015).

Figura 2-16 – Esquema simplificado de cura das tintas



Fonte: Kränkel, 2015

2.2.4 Tipos de tinta

As tintas podem ser agrupadas em três grandes grupos, de acordo com o tipo de veículo que as compõe: veículos não convertíveis, veículos convertíveis e veículos inorgânicos.

2.2.4.1 Tintas com veículos não convertíveis

São tintas cujas resinas formam uma película seca após a simples evaporação dos solventes de sua formulação, sem nenhum tipo de reação química entre seus componentes, usualmente denominadas de tintas do tipo laca (REIS, 2012). Os principais representantes desta classe de tintas são:

- **Acrílicas:** Possuem excelente retenção de brilho e resistência à decomposição por raios ultravioleta e óleos e graxas, não

amarelando quando expostas a intempéries (MAKHLOUF, 2014);

- Estirenoacrilato: Possui uma retenção de cor e brilho razoável, com pouca resistência aos raios ultravioleta e não são saponificáveis. Estão em desuso em função de sua alta reatividade e toxidez (REIS, 2012);
- Borracha clorada: Resistente à água, ácidos e álcalis, muito flexíveis e com ação bactericida, com o ponto negativo de degradar em temperaturas superiores a 65°C liberando ácido clorídrico (MAKHLOUF, 2014);
- Vinílicas: Também chamadas látex PVA possuem grande rendimento; durabilidade; ótimo desempenho nas repinturas; excelente acabamento. Apesar de não ter boa resistência a solventes, possui boa resistência a ácidos. Também possui alta resistência à água, a álcalis e à abrasão (Anghinetti, 2012);
- Nitrocelulose: Como aspectos positivos apresenta secagem rápida, alta dureza, lixamento rápido e resistência a aguarrás e gasolina, e como negativos: tendência ao amarelamento, película de baixa espessura, baixa resistência a grande maioria dos solventes podem necessitar de polimento para alcançar alto brilho (WEN e DUSEK, 2017);
- Betuminosas: Feitas a base de asfalto e piche, só podem ser fabricadas na cor preta. Possuem excelente resistência à umidade, mas baixa resistência química muito utilizadas em ambientes submersos (DONADIO, 2011);

2.2.4.2 Tintas com veículos convertíveis

São as tintas cuja película seca é formada devido a reações químicas, estas podem ocorrer em reações por mistura de componentes internos, ou reações com o meio ambiente que estejam expostas.

- Óleo: Possuem secagem lenta e são saponificáveis, sendo recomendadas para ambientes pouco agressivos e secos (REIS, 2012).
- Alquílicas: Tinta de baixo custo, mas com baixa resistência a umidade e produtos químicos (WEN e DUSEK, 2017).
- Fenólicas: Duras e quebradiças, apresentam pouca retenção de cor, mas com boa resistência aos raios ultravioleta. Possuem alta

aderência, sendo muito utilizadas como intermediárias em superfícies lisas (TIWARI, RAWLINS e HIHARA, 2015);

- Epóxi: Apresentam alta dureza, impermeabilidade e flexibilidade, com grande resistência a ácidos, álcalis, abrasão, solventes e altas temperaturas, porém com baixa resistência a intempéries (Anghinetti, 2012);
- Alcatrão de hulha: Tintas de altas espessuras, excelente resistência química, e a imersão em água, porém são agressivas ao meio ambiente (MAKHLOUF, 2014);
- Epóxis tar free: Tinta livre de metais pesados, sendo caracterizadas como tintas ecológicas e ainda assim mantendo suas propriedades semelhantes a dos epóxis tradicionais, porém de custo elevado (REIS, 2012);
- Epóxi shop primer: Alta aderência sobre metais não ferrosos, aplicados a baixas espessuras, não interfere nas soldas, ótima resistência mecânica e secagem rápida. Muito utilizado como uma pré-pintura para garantir a preservação de componentes que serão armazenados para uso futuro (KNUDSEN e FORSGREN, 2017);
- Epóxi rico em zinco: Possuem excelente resistência à umidade e fornecem proteção catódica ao aço carbono, evitando a progressão da corrosão em áreas com danos mecânicos. Recomendadas para pintura de comportas, boias de sinalização e outras estruturas imersas em água doce ou salgada (REIS, 2012);
- Acrilato de cobre: Tinta antiincrustante de auto-polimento, conforme a água do mar passa por este tipo de tinta ela vai arrastando aos poucos finas camadas da tinta. O óxido cuproso presente em sua composição repele os microrganismos de se fixarem na superfície (MANNARI e PATEL, 2015);
- Poliuretano: Excelente resistência a componentes químicos, umidade e intempéries, possuem alta flexibilidade e dureza, mas baixa resistência a corrosão. Normalmente utilizadas como acabamento, sendo aplicadas sobre outras tintas (ANGHINETTI, 2012);
- Silicone: Possuem baixa resistência química, porém conferem alta dureza e impedem a absorção de água permitindo ainda a realização de trocas gasosas, resistem a temperaturas de até 600°C e são altamente flexíveis (MAKHLOUF, 2014);
- Silicatos: Tintas com alto teor de zinco metálico são de base inorgânica e suportam temperaturas de até 500°C, possuem alta

resistência à umidade e conferem grande aderência as camadas posteriores (MANNARI e PATEL, 2015).

2.3 TINTA PÓ

De acordo com Camargo (2002), "As tintas em pó são definidas como partículas finamente divididas de polímero orgânico, termoplástico ou termorrígido, que geralmente contêm pigmentos, cargas e aditivos e que permanecem finamente divididas durante a armazenagem sob condições adequadas".

As tintas em pó são utilizadas em diversos segmentos, como: estruturas metálicas, eletrodomésticos, acessórios automotivos, máquinas, ferramentas, móveis metálicos, equipamentos elétricos, dentre outros.

A Quadro 2.5 apresenta as vantagens e desvantagens da tinta pó em relação a líquida:

Quadro 2.5 - Vantagens e desvantagens das tintas em pó

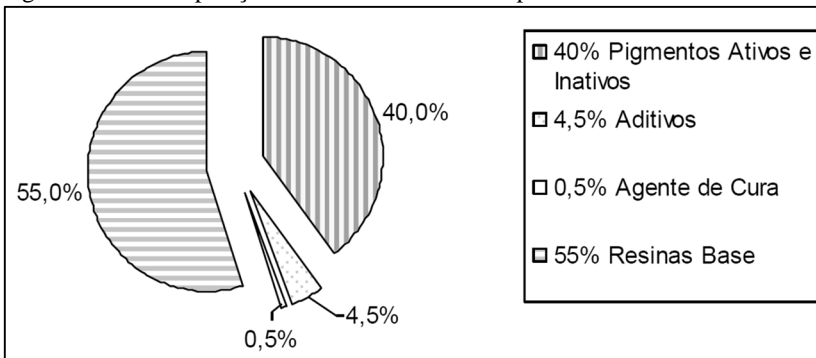
<i>Vantagens</i>	<i>Desvantagens</i>
Não utiliza solventes	Necessita de mais equipamentos
Baixos riscos de incêndio	Troca de cores demoradas
Redução nos prêmios de seguro	Tintas não podem ser misturadas
Tinta pronta para uso	Difícil aplicar em camadas abaixo de 30 µm
Baixos índices de rejeição	Difícil para pintar partes internas de objetos
Normalmente, aplicação em única demão	Difícil para pintar substratos não metálicos (madeira, plástico, etc)
Geração mínima de poluentes	Acabamento deficiente para certas finalidades (ex: setor automotivo)

Fonte: Adaptado de (Kränkel, 2015)

2.3.1 Matérias-primas

As matérias-primas utilizadas pelas tintas em pó são praticamente as mesmas das tintas líquidas, com exceção dos solventes. Uma simples mistura de resinas, aditivos, pigmentos e agentes de cura, todos no estado sólido, em uma proporção semelhante à apresentada na Figura 2-17 (MANNARI e PATEL, 2015).

Figura 2-17 – Composição básica de uma tinta em pó

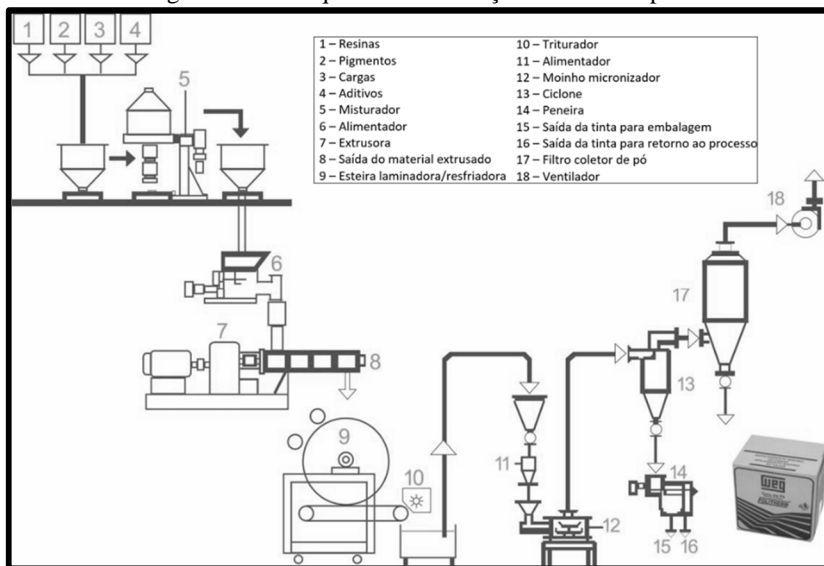


Fonte: Kränkel, 2015

2.3.2 Processo de fabricação

A produção de tintas em pó segue uma série de etapas distintas: pré-mistura, extrusão, resfriamento, micronização, classificação e embalagem (KRÄNKEL, 2015). A Figura 2-18 representa este esquema de fabricação.

Figura 2-18 – Esquema de fabricação de tinta em pó



Fonte: Adaptado de (Kränkel, 2015)

Pré-mistura: As matérias-primas são misturadas em um misturador até atingirem uma homogeneização adequada. Mistura puramente mecânica e à temperatura ambiente.

Extrusão: A mistura é aquecida e extrusada, de modo que todas as partículas sólidas se fundam formando uma massa uniforme e pastosa.

Resfriamento: O material fundido é resfriado passa por uma calandra de resfriamento o mais rápido possível assim que sair da extrusora, formando assim uma lâmina sólida. Esta lâmina é encaminhada por esteiras para um triturador, que quebra a lâmina em pequenos pedaços, conhecidos como chips.

Micronização: Os “chips” de tinta passam por uma operação de moagem ou micronização, até que suas partículas fiquem do tamanho desejado. Cada tinta possui uma especificação diferente para o tamanho de partículas.

Classificação e embalagem: Após micronizadas, as tintas passam por peneiras para garantir que nenhuma partícula com granulometria acima do permitido vá para o cliente final.

2.3.3 Aplicação de tinta em pó

A aplicação de tintas em pó requer certos equipamentos especiais para aspergir o pó sobre a superfície e para curar a tinta, o que inviabiliza sua utilização de maneira doméstica, geralmente sendo aplicada somente por empresas e indústrias (TIWARI, RAWLINS e HIHARA, 2015).

Para que a tinta aspergida seja atraída e fixada na a superfície que se deseja pintar, as partículas de pó devem ser carregadas com cargas elétricas. Existem duas abordagens comumente utilizadas para isso, os leitos fluidizados e as pistolas eletrostáticas (KRÄNKEL, 2015).

As tintas em pó não podem ser curadas a temperatura ambiente, elas necessitam de uma alta temperatura para realizar seu processo de cura, está devidamente informada no boletim técnico de cada tinta, mas geralmente superior a 140°C. As duas principais técnicas a cura das tintas em pó são a utilização de chapas pré-aquecidas ou estufas (KRÄNKEL, 2015).

2.3.3.1 Leito fluidizado simples

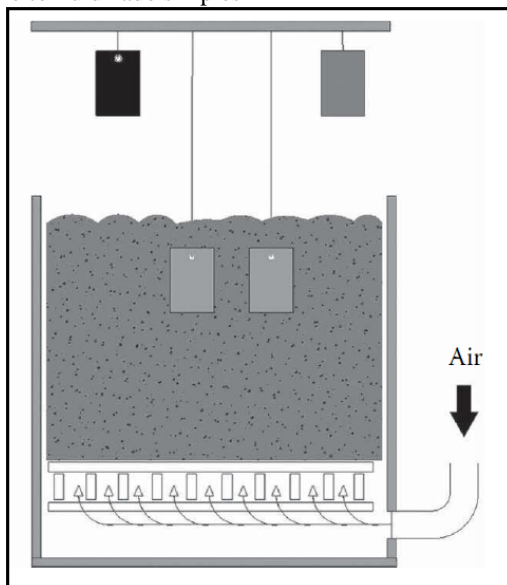
Leitos fluidizados simples são utilizados juntamente com objetos pré-aquecidas, podendo ser seguidos ou não de uma estufa.

Dentro de uma câmara a tinta em pó é insuflada em uma vazão de ar determinada, o pó é então suspenso formando uma densa nuvem com

comportamento semelhante a de um líquido (KNUDSEN e FORSGREN, 2017).

O objeto a ser pintado, aquecido a temperatura superior a de fusão do pó é inserido nesta câmara, entrando em contato com a tinta em pó, fazendo com que está se funda e fique aderida ao objeto. Caso seja necessário, o objeto pintado pode ser levado a uma estufa para completar o processo de cura da tinta. A Figura 2-19 representa este processo (KNUDSEN e FORSGREN, 2017).

Figura 2-19 – Leito fluidizado simples

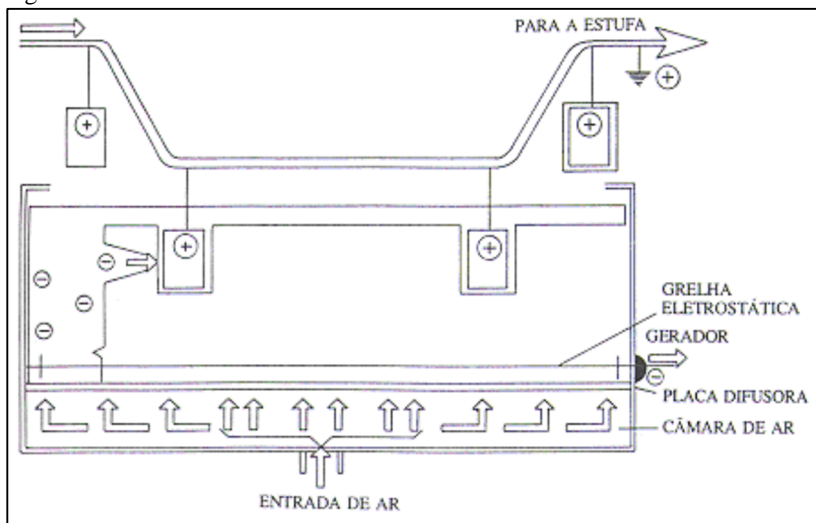


Fonte: Knudsen e Forsgren, 2017

2.3.3.2 Leito fluidizado eletrostático

Neste sistema, o objeto a ser pintado é aterrado e as partículas de tinta são atraídas para ele por forças eletrostáticas, não sendo necessário seu pré-aquecimento. Em seguida o objeto é conduzido para uma estufa para realizar o processo de cura da tinta. A Figura 2-20 representa este método.

Figura 2-20 – Leito fluidizado eletrostático



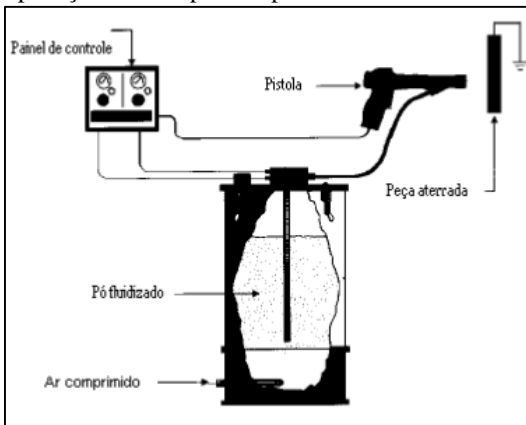
Fonte: Kränkel, 2015

Este sistema permite a pintura de peças com geometrias mais complexas e um controle mais preciso da camada final de tinta.

2.3.3.3 Pistola eletrostática

Neste sistema uma mistura de ar e tinta sai pelo bico da pistola onde está presente um eletrodo de alta tensão, responsável por carregar eletricamente as partículas de tinta. O objeto a ser pintado deve estar aterrado para atrair as partículas para sua superfície, conforme Figura 2-21.

Figura 2-21 –Aplicação de tinta pó com pistola eletrostática



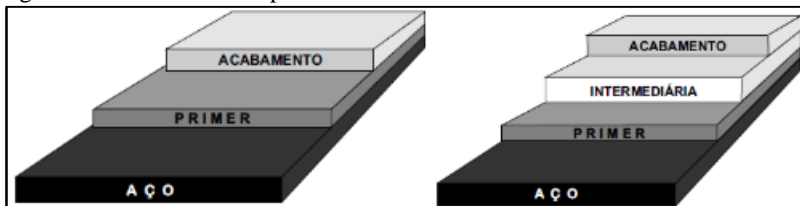
Fonte: Kränkel, 2015

2.4 ESQUEMAS DE PINTURA

Comercialmente as tintas são negociadas por meio de esquemas de pintura, que são conjuntos compostos por pelo menos uma tinta com espessuras de camada especificadas para garantir sua eficácia para cada condição, como por exemplo, "uma camada de N 2630 com $150\mu\text{m}$ + duas camadas de N 2677 com 80μ cada" representa um esquema de pintura composto por duas tintas, N 2630 e N 2677, onde primeiro deve ser aplicada uma camada de N 2630 com uma espessura de $150\mu\text{m}$ e após duas camadas de N 2677 com $80\mu\text{m}$ cada, respeitando-se o intervalo de repintura de cada tinta (KRÄNKEL, 2015).

A grande maioria dos esquemas de pintura é composta por até três tipos diferentes de tinta onde cada uma apresenta uma função diferente na proteção do substrato. Elas são normalmente separadas em tintas de fundo, intermediárias, acabamento (MANNARI e PATEL, 2015). A Figura 2-22 representa os sistemas de pintura mais comuns.

Figura 2-22 – Sistemas de pintura tradicionais



Fonte: Kränkel, 2015

Tintas de fundo são responsáveis principalmente pela aderência do esquema de pintura ao substrato, podem ainda conter pigmentos anticorrosivos e tintoriais, normalmente possuem péssima resistência a radiação solar e intempéries (MANNARI e PATEL, 2015).

Tintas intermediárias por vezes denominadas como Tie-Coat, oferecem espessura e auxiliam moderadamente na proteção, apresentam um custo bastante reduzido em relação às tintas de fundo e de acabamento (MANNARI e PATEL, 2015).

Tintas de acabamento são responsáveis principalmente por proteger o sistema contra intempéries, umidade, radiação UV e a manter o brilho e cor originais do esquema por maior tempo (KNUDSEN e FORSGREN, 2017).

Não existe um limite para a quantidade de tintas ou de camadas, por exemplo, a pintura de uma torre eólica possui apenas 3 camadas, mas suas pás necessitam de esquemas diferentes com até 7 camadas (TIWARI, RAWLINS e HIHARA, 2015).

2.5 SELEÇÃO DOS ESQUEMAS

Como a comercialização das tintas é feita na forma de esquemas de pintura, começa a surgir certo grau de complexidade para classificar e selecionar os melhores esquemas para determinadas situações. As características próprias de cada tinta não são suficientes para isso, e ainda não foi proposta uma equação as envolvendo e que possa ser utilizada para chegar a uma escala global de proteção da tinta (MANNARI e PATEL, 2015).

Existem diversas normas utilizadas para classificar as tintas, elas estabelecem padrões e níveis em diversos aspectos: Grau de aderência sobre diferentes substratos, resistência à névoa salina, umidade, intemperismo, ácidos, graxa, grau de escorrimento, resistência a impactos mecânicos, dentre outros.

Clientes mais experientes podem se utilizar destas normas na hora de comprar uma tinta, solicitando um esquema de pintura que resista a X horas em névoa salina, que perca no máximo certo percentual de brilho após 1000 horas em exposição a câmara UV-B, etc. Mas a grande maioria dos clientes não possui conhecimento para fazer uma solicitação tão precisa.

Por conta disso começa a se analisar a heurística do problema e a experiência dos vendedores/analistas se torna imprescindível na hora de questionar o cliente sobre sua aplicação e então fazer uma avaliação para recomendar um esquema de pintura ideal.

Algumas lógicas que os vendedores utilizam seguem estas linhas de pensamento: Ambientes úmidos precisam de um acabamento impermeável, já ambientes muito agressivos precisam de um primer anticorrosivo, ambientes expostos a luz solar necessitam de um acabamento com alta resistência UV, ambientes agressivos e expostos a luz solar necessitam tanto de primer anticorrosivo quanto acabamento, aço inoxidável necessita de um promotor de aderência ou primer especial com aderência ao aço, madeira necessita de um selador que deve ser compatível com as próximas camadas, exposição a produtos químicos necessitam de tintas de acabamento altamente específicas, e assim por diante.

Este capítulo é exposto de modo introdutório para apresentar ao leitor a real necessidade de se utilizar esquemas de pintura para a proteção de superfícies, junto com os problemas para selecionar e indicar um esquema ideal para as necessidades dos clientes.

Uma maneira de avaliar este problema é que existe uma lista de esquemas de pintura disponíveis e que devem ser classificados como “atende” ou “não atende” aos requisitos informados pelo cliente. Este tipo de classificação possui uma base predominantemente heurística, logo, a utilização da técnica de sistemas especialistas para analisar a heurística do problema e a técnica de árvore de decisão para realizar a classificação dos esquemas apresenta um grande potencial de sucesso.

O próximo capítulo aborda ambas as técnicas como preparação para a construção do protótipo no capítulo 4 que envolve os conceitos aqui mostrados.

3. INTELIGENCIA ARTIFICIAL E MINERAÇÃO DE DADOS

No início deste capítulo são apresentadas técnicas relacionadas à inteligência artificial, com uma maior ênfase em sistemas especialistas (SE) e árvores de decisão, que foram as principais técnicas utilizadas no desenvolvimento do projeto.

Na sequência, três tipos de algoritmos para auxiliar na geração de árvores de decisão serão apresentados com maiores detalhes sobre os cálculos utilizados na seleção de atributos.

3.1 SISTEMAS ESPECIALISTAS

São programas computacionais que apresentam um comportamento que se assemelha ao de um especialista humano em determinado domínio de conhecimento (BARRETO, 2001).

De acordo com Giarratano e Riley (1994), um sistema especialista (SE) é “...um programa computacional inteligente que utiliza conhecimento e procedimentos de inferência para resolver problemas com grau de dificuldade suficiente para requerer significativa especialidade humana em sua solução.”

Deste modo, pode-se definir um SE como um programa computacional cujo comportamento visa simular as ações de um especialista humano, buscando encontrar uma solução para determinado problema de acordo com seus conhecimentos prévios.

3.1.1 Surgimento dos sistemas especialistas

Até a metade da década de 60 a inteligência artificial (IA) era dominada pela ingênua crença de que uma sequência lógica genérica acoplada a poderosos computadores resultaria em um programa de performance sobre-humana e resolveria todos os problemas da humanidade. Conforme foi-se ganhando experiência na área, o poder limitado destas estratégias genéricas para tentar solucionar problemas de alta complexidade foi se tornando cada vez mais evidente (NEWELL, 1969).

Na busca por alternativas que pudessem solucionar estes problemas de naturezas mais complexas, Newell (1969) constatou que através de regras específicas do tipo “**SE** isso **ENTÃO** aquilo” era possível criar estruturas que se assemelhavam a aparente sequência

cognitiva utilizada por profissionais especialistas, sendo possível então obter uma solução rápida e relativamente precisa.

No início da década de 70, os SE começaram a se popularizar, programas de alta performance para domínios de conhecimento específicos, estimulando uma busca para encontrar as bases de conhecimento que fundamentam expertise humana e criando assim todo um conjunto de princípios, ferramentas e técnicas que formaram a base da engenharia de conhecimento (RAYES-ROTH, WATERMAN, e LENAT, 1983).

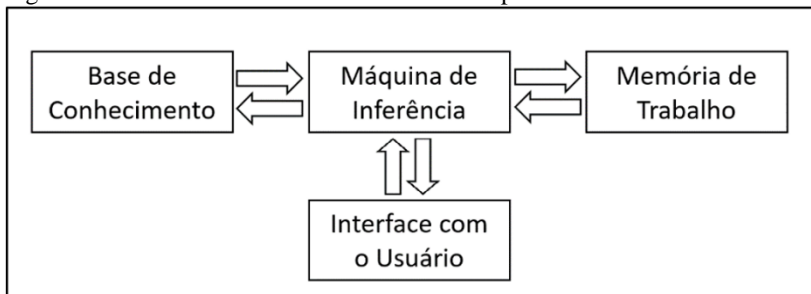
Diversos SE de sucesso foram desenvolvidos durante este período, ampliando a visibilidade e amadurecendo rapidamente seu campo de aplicação (RAYES-ROTH, WATERMAN, e LENAT, 1983), na sequência se encontram alguns exemplos desenvolvidos neste período:

- Dendral: Começou a ser desenvolvido em 1965 na Universidade de Standford, seu objetivo era encontrar estruturas moleculares orgânicas com base na espectrometria de massa das ligações químicas de uma molécula desconhecida (ZUBEN, 2011);
- R1: Desenvolvido pela Digital Equipment Corporation em 1978, para configurar sistemas computacionais do tipo VAX de acordo com especificações dos clientes (RAYES-ROTH, WATERMAN e LENAT, 1983);
- Prospector: Desenvolvido pela SRI International em 1977 e utilizado para auxiliar geólogos na busca por depósitos com recursos geológicos, conseguiu localizar uma jazida de molibdênio avaliada em U\$ 100,000,00 (BARRETO, 2001);
- Mycin: Criado no mesmo laboratório do Dendral em 1972, utilizado para o diagnóstico e tratamento de doenças infecciosas do sangue causadas por bactérias (ZUBEN, 2011).

3.1.2 Estrutura de um Sistema Especialista

Fundamentalmente um SE é formado por quatro módulos distintos: base de conhecimento, memória de trabalho, máquina de inferência e interface (DURKIN, 1994). A interligação entre estes módulos é apresentada na Figura 3.1.

Figura 3.1 – Elementos básicos de um sistema especialista



Fonte: Adaptado de Matelli (2008)

3.1.2.1 Base de Conhecimento

A base de conhecimento representa uma memória de longo prazo, expressando o conhecimento relevante para a solução do problema, normalmente é imutável dentro do sistema e sua representação mais comum é através de regras e objetos que são ativados de acordo com dados e fatos locais, a base é adquirida/construída através da experiência do especialista na área (COPPIN, 2013).

3.1.2.2 Memória de Trabalho

Representa a memória humana, local e de curto prazo, armazena informação, fatos e dados relevantes apenas à sessão atual do SE e é necessária para ativar as regras da base de conhecimento (Silva, 1998).

Normalmente a memória de trabalho é alterada conforme o usuário interage com o SE, estas informações podem ser adicionadas, removidas e modificadas, atuando tanto como um guia como quanto um histórico para a seção (MATELLI, 2008).

Quando uma sessão do SE é encerrada, todas as informações da memória de trabalho são prontamente eliminadas para que a próxima sessão do SE esteja livre de informações referentes ao problema/situação anterior (MATELLI, 2008).

3.1.2.3 Máquina de Inferência

De acordo com o dicionário online Dicio (2017), inferência pode ser definida como o "raciocínio através do qual uma proposição é considerada verdadeira pela sua ligação com outras já tidas como verdadeiras; a proposição que se assume como sendo verdadeira." A máquina de inferência de um SE pode ser configurada para operar com

diferentes tipos de inferência, sendo que as mais comuns são: encadeamento direto (*forward chaining*) e encadeamento reverso (*backward chaining*) (GIARRATANO e RILEY, 1994).

A máquina de inferência é o mecanismo intermediário que controla todo o SE, pode ser comparado com o raciocínio humano, decidindo quais das regras da base de conhecimento satisfeitas na memória de trabalho devem ser executadas e em qual ordem e se encarrega de atualizar as informações da memória de trabalho modificadas pelas regras ativas da base de conhecimento (MECABÔ, 2007).

A máquina de inferência possui uma agenda, que é uma lista com todas as regras ativas no momento em ordem de prioridade de ativação, esta agenda pode ser programada para funcionar de diferentes maneiras, entre elas pode-se citar: F.I.F.O. (First In First Out) onde as regras são disparadas (ativadas) por ordem de ativação e L.I.F.O. (Last In First Out) onde a regra ativa mais recente a ser ativada será a próxima a ser disparada (COPPIN, 2013).

3.1.2.3.1 Encadeamento Direto

No encadeamento direto o usuário utiliza uma interface para inserir dados dentro da memória de trabalho do SE, a máquina de inferência utiliza esses dados para ativar regras da base de conhecimento que por sua vez irão gerar novos dados na memória que ativarão novas regras da base, e assim por diante. As regras ativadas podem manipular dados existentes, criar novos, ou ainda, solicitar que usuário os insira (BITTENCOURT, 2006).

3.1.2.3.2 Encadeamento Reverso

No encadeamento reverso temos que:

“Um usuário informa um conjunto de fatos ao sistema e o sistema analisa se pode provar qualquer uma das possíveis hipóteses usando esses fatos. Em alguns casos, ele precisará de fatos adicionais e, nestes casos, o sistema especialista frequentemente fará perguntas ao usuário para verificar fatos que possam permitir a ativação de outras regras” Coppin (2013).

3.1.2.4 Interface

A interface é a parte de comunicação do SE com o usuário, é uma das partes mais importantes quando o SE é projetado para ser operado por pessoas leigas. Uma interface projetada ergonomicamente, com as informações facilmente localizadas e uma maneira simples e intuitiva de

inserir os dados pode ser a diferença entre o sucesso e o fracasso de um SE, pois de nada adianta criar uma poderosa ferramenta que ninguém consiga utilizar e compreender (BITTENCOURT, 2006).

Uma das principais características do SE em relação a sua interface é a capacidade de explicação, permitir que o usuário tenha acesso a informações para saber porque uma regra foi escolhida em detrimento de outra, auxiliando na disseminação do conhecimento aplicado ao SE para o usuário (COPPIN, 2013).

3.1.3 Desenvolvimento de Sistemas Especialistas

O desenvolvimento de um SE requer certo número de tarefas que podem variar devido principalmente a quantidade de áreas de conhecimento envolvidas e o número de especialistas necessários (SILVA, 1998).

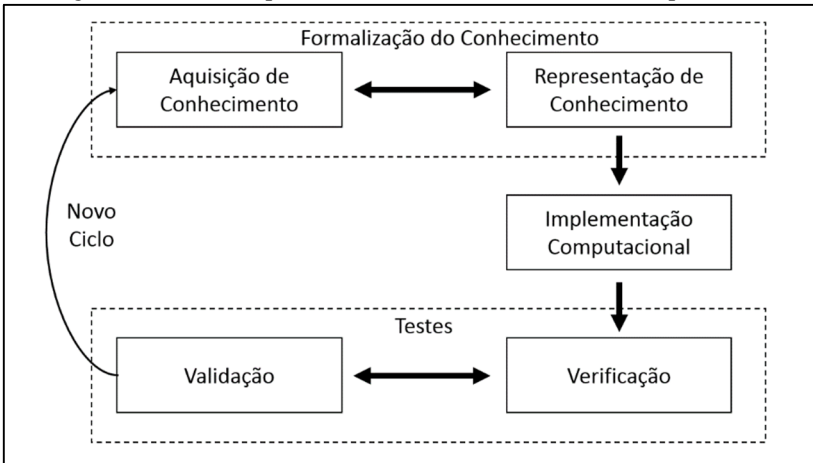
Normalmente, as pessoas com maior envolvimento com a parte de programação e desenvolvimento do SE são denominadas engenheiros de conhecimento (EC), que se envolvem diretamente com os especialistas no assunto e detentores do conhecimento, denominados especialistas humanos (EH) (COPPIN, 2013).

Ambos EC e EH devem desenvolver um sistema visando o usuário final, pois ele é quem realmente irá utilizar e tirar proveito do SE, e se ele não ficar satisfeito com o projeto, acabará por evitar utilizá-lo sempre que possível (COPPIN, 2013).

Um dos modelos mais adotados para o desenvolvimento de um SE é o modelo Incremental, onde o sistema é desenvolvido em etapas incrementais de suas funcionalidades, como um ciclo contínuo de desenvolvimento de funções. Pode-se dizer que novas funções vão sendo incrementadas constantemente ao protótipo até que ele torne o sistema definitivo (SILVA, 1998).

Para organizar e ordenar as etapas dos ciclos incrementais existe diversos modelos já desenvolvidos, Silva (1998) apresenta uma sequência de etapas que pode ser traduzida em: 1) Análise de viabilidade, 2) Aquisição de Conhecimento, 3) Representação do Conhecimento, 4) Implementação, 5) Verificação e Validação, conforme apresentado na Figura 3.2.

Figura 3.2 – Modelo para desenvolvimento de um sistema especialista



Fonte: Adaptado de Alves (2001)

3.1.3.1 Análise de Viabilidade

Nesta etapa algumas perguntas devem ser analisadas para aumentar as chances de sucesso no final do projeto, de acordo com Gonzalez e Dankel (1993) e Bittencourt (2006) algumas destas perguntas são:

- O problema realmente existe?
- A técnica de SE é aplicável?
- O tipo de problema reproduz o conhecimento humano na busca de solução?
- O conhecimento ou especialização muda periodicamente ou permanece constante?
- O conhecimento envolvido é relativamente bem entendido e aceito?
- O problema pode ser resolvido por outros meios?
- Existe apoio gerencial para o projeto?
- Existe apoio da parte do EH?
- O EH é competente?

3.1.3.2 Aquisição do Conhecimento

Pode ser considerada a parte mais sensível durante o desenvolvimento de um SE, nesta etapa o conhecimento por parte dos EH

deve ser extraído e refinado para evitar que informações incoerente ou não relevantes sejam incluídas e gerem resultados não satisfatórios (ruídos) (BITTENCOURT, 2006).

Barreto (2001) menciona algumas técnicas que podem ser utilizadas para auxiliar na aquisição do conhecimento dos especialistas:

- Observação: o EH deve ser observado durante sua rotina de trabalho pelo EC, normalmente esta etapa serve para que o EC se familiarize com os tipos de problemas envolvidos no dia a dia e avalie como a técnica de SE pode gerar melhorias, pensar em funcionalidades além das planejadas que também podem ser úteis;
- Entrevista com o especialista: o EC realiza uma entrevista com o EH, normalmente é composta de perguntas simples e diretas e suas respostas podem ser tanto anotadas quanto gravadas em áudio para posterior avaliação, dúvidas preferencialmente devem ser tiradas durante a própria entrevista para um melhor aproveitamento do tempo;
- Análise de protocolo: Esta técnica consiste em criar uma situação que gere um problema que o EH deva solucionar, gerando assim um relatório das suas ações e do porquê de ter realizado cada uma;
- Ensino reverso: Os papéis são invertidos e o EC tenta “ensinar” o EH utilizando o conhecimento adquirido nas etapas anteriores, esta técnica auxilia com a sintonia do pensamento da equipe e deixa mais claro se o EC realmente está compreendendo o assunto.

A grande maioria do conhecimento adquirido para este trabalho foi através de entrevistas com especialistas e algumas observações pontuais quando a ligação de um cliente ocorria durante uma entrevista.

A análise de protocolo foi utilizada em dois cenários para confirmar algumas informações que estavam divergindo entre dois especialistas, mas com a utilização de um cenário bem definido os especialistas chegaram a um consenso sobre a resposta que melhor se enquadrava.

3.1.3.3 Representação do Conhecimento

O conhecimento adquirido na etapa anterior deve ser representado de modo que possa ser utilizado pelo SE, o método escolhido para

representar o conhecimento deve ser expressivo a ponto de permitir sua representação de modo completo e eficiente (WATERMAN, 1986).

3.1.3.3.1 Regras

A representação de conhecimento através de regras se baseia em utilizar a sintaxe "**SE** condição **ENTÃO** ação", deste modo as regras ficam dispostas de uma maneira natural e intuitiva, separando e auxiliando na visualização de suas condicionais e ações, além de facilitar a localização caso surja a necessidade de eventuais mudanças (COPPIN, 2013).

Segue um exemplo da utilização de regras para a seleção de quais tintas devem compor um esquema de pintura para determinadas situações:

SE (umidade é alta) **E** (ambiente é marítimo)
ENTÃO (o esquema de pintura necessita de uma demão de primer epóxi)
E (uma demão de um intermediário de alta espessura)
E (uma demão de um acabamento em poliuretano)

SE (umidade é alta) **E** (ambiente industrial)
ENTÃO (o esquema de pintura necessita de uma demão de primer epóxi)
E (uma demão de um acabamento em poliuretano)

SE (umidade é baixa) **E** (ambiente industrial)
ENTÃO (necessita de uma demão de epóxi dupla função)

Deve-se tomar cuidado com a maneira que as regras são criadas, pois uma regra mal formulada pode comprometer todo o funcionamento de um SE. Abaixo estão os principais erros que podem ser cometidos durante a criação de regras, Waterman (1986) menciona algumas:

- Encadeamento Infinito: trata-se de regras que não possuem fim, ou que são cíclicas.
- Regras com erros de compilação que não permitem o SE seja sequer executado.
- Regras nunca alcançadas/impossíveis: São regras que se deseja que sejam disparadas, mas alguma condição impossível as impedem. No exemplo anterior poderia se aplicar caso uma regra fosse ativada somente quando a umidade fosse tanto alta quanto baixa, condição esta que logicamente nunca ocorre.

- Regras nunca utilizadas/inúteis: Regras inúteis são regras que jamais serão disparadas, e que sua não ativação não deixa de causar algum efeito desejado no sistema. No exemplo anterior se aplicaria no caso de uma regra necessitar do registro de umidade, mas o SE não apresentar este valor em nenhum momento.
- Regras contraditórias: Regras com as mesmas condições, mas ações contraditórias. No exemplo anterior, se as condições para umidade fossem removidas, haveria duas regras que seriam ativadas quando o ambiente fosse industrial, mas recomendariam produtos diferentes.
- Regras sem saída: Regras que não geram modificação no sistema ou cujas modificações não interferem no resultado final. “SE (umidade é alta) ENTÃO (nada)”.
- Regras com condições sempre verdadeiras: Regras que sempre serão disparadas, ou com pelo menos uma condição sempre verdadeira. Como por exemplo, uma regra “SE $(1+1=2)$ ENTÃO (necessita de um primer epóxi)”, deste modo todos os esquemas de pintura passariam a utilizar primer epóxi.

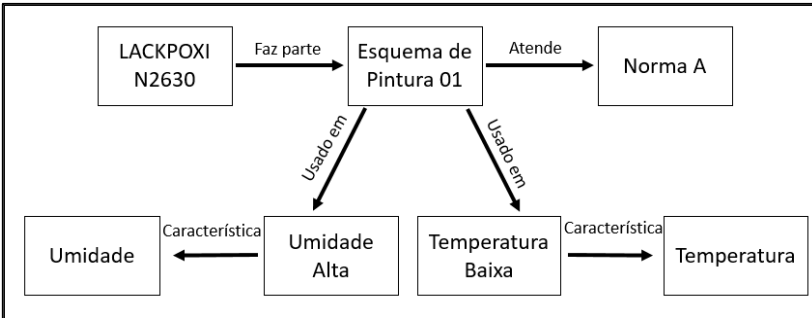
3.1.3.3.2 *Redes Semânticas*

Uma Rede semântica representa o conhecimento através de nós e arcos, onde os nós podem representar um tipo de conceito e os arcos definem qual a relação semântica entre os nós. Dependendo da área a ser aplicada a rede, um nó pode ter diferentes usos e interpretações podendo representar classes, predicados, até mesmo simples palavras (GIARRATANO e RILEY, 1994).

De acordo com Aragão (2002), conceitualmente a rede semântica foi proposta em 1913 por Selz como um modelo para explicar fenômenos psicológicos, mas foi em 1966 que Quillian implementou estas ideias propondo um modelo computacional para a memória humana e para o representar utilizou o relacionamento entre dois objetos.

As relações dentro das redes são de importância fundamental para garantir uma estrutura que além de representar e organizar o conhecimento, seja de fácil entendimento para quem a estiver lendo. Normalmente os arcos são definidos por termos como "faz parte de", "é um", "composto por", dentre outros (MCCUE, 2015). A Figura 3.3 apresenta um exemplo simples de rede semântica.

Figura 3.3 – Rede semântica



Fonte: Elaborado pelo Autor.

A rede diz que a tinta Lackpoxi N2630 faz parte do Esquema de Pintura 01, o esquema de pintura 01 atende a norma A além de ser usado em umidade alta e temperatura baixa, sendo que a umidade alta é uma das características da umidade e a temperatura baixa é uma característica de temperatura.

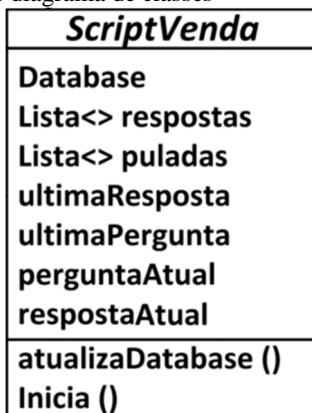
3.1.3.3.3 Orientação a Objetos

Objetos podem ser definidos como pacotes de informação e procedimentos lógicos que, quando combinados entre si criam uma entidade “auto-suficiente” que normalmente são criadas para representar o modelo de uma entidade que exista no meio físico ou computacional (BARRETO, 2001).

Objetos são representações de entidades e são caracterizados por um conjunto de propriedades ou atributos e um conjunto de métodos, as propriedades representam suas características e os métodos são as ações que o objeto pode realizar (BARRETO, 2001).

Utiliza também o conceito de classes, que são modelos estruturais que caracterizam os objetos, definindo quais propriedades e métodos eles deverão ter, deste modo é possível criar diversos objetos com as mesmas propriedades, mas com valores diferentes entre si, a Figura 3.4 representa o diagrama de classes da classe Dataset desenvolvida para o projeto.

Figura 3.4 – Exemplo de diagrama de classes



Fonte: Elaborado pelo Autor.

Os diagramas são divididos em três partes: a parte superior contém o nome da classe, a intermediária suas características e a inferior os métodos que ela pode executar. Deste modo pode-se “ler” a classe como “Uma classe chamada ScriptVenda que possui os atributos Database, ultimaResposta, ultimaPergunta, perguntaAtual, respostaAtual, uma lista de respostas e uma lista de ‘puladas’ (perguntas ignoradas) e executa os métodos atualizaDatabase e Inicia”.

O paradigma de orientação a objetos (POO) define a criação de diversos objetos diferentes representando as entidades que compõe o sistema, cada um com suas propriedades e funções (métodos) bem definidas (TAYLOR, 1990).

Deste modo o programa não mais é executado através da chamada de funções, mas sim, da interação entre os diferentes objetos desenvolvidos, deixando o código mais modular e facilitando a reutilização de partes específicas do código (BARRETO, 2001).

A POO possui quatro propriedades chave que auxiliam na criação das classes:

- **Abstração:** Consiste em ignorar aspectos ou propriedades que não são relevantes ao problema que será analisado, permitindo concentrar os esforços nos aspectos mais relevantes da aplicação.
- **Encapsulamento:** Oculta informações desnecessárias ou sigilosas dentro de métodos, apresentando para o usuário somente o desejado;

- Herança: Permite que propriedades e métodos sejam transferidos entre classes.
- Polimorfismo: Permite que o mesmo método realize ações distintas para diferentes objetos.

Neste trabalho a POO foi amplamente utilizada, sendo que as principais classes desenvolvidas foram: Database, ProcessadorDeDatabase, Perguntas, Explicações, Dados, Linhas, Colunas, Valores, ScriptVenda, App e Cria_GUI, estas serão explicadas com detalhes no próximo capítulo

3.1.3.4 Implementação

A implementação envolve o mapeamento do conhecimento coletado e formalizado nas etapas anteriores e sua conversão para uma estrutura lógica (linguagem de programação) que possa ser interpretada por um software previamente escolhido (RAYES-ROTH, WATERMAN e LENAT, 1983).

Tanto a estrutura lógica quanto os programas utilizados podem variar de acordo com o tipo de projeto e conhecimento prévio dos envolvidos (COPPIN, 2013). Para este projeto em questão duas linguagens foram escolhidas: java e javascript e estas executadas pelo software Eclipse.

Nesta etapa, o domínio de conhecimento, a estrutura das informações, regras e estratégias de controle são apresentadas de maneira explícita. É neste momento que inconsistências ficarão mais evidentes, a solução de conflitos é algo esperado, pois podem existir fatores globais sobre algumas regras que gerem influências inesperadas e necessitem ser reavaliadas, transformadas ou removidas (RAYES-ROTH, WATERMAN e LENAT, 1983).

A implementação é uma etapa extremamente importante na construção de um SE, pois é ela quem realmente dará “vida” ao sistema, as linhas de código devem ser muito bem pensadas e testadas para garantir que o conhecimento realmente esteja sendo expresso de maneira adequada (WATERMAN, 1986).

3.1.3.5 Verificação e Validação

Esta é a etapa de testes do SE, nela são realizados testes que normalmente são divididos em etapas com funções bastante distintas: Uma para verificar o conjunto de regras e a execução do sistema e a outra

para verificar o conhecimento envolvido e a maneira que foi implementado (RAYES-ROTH, WATERMAN e LENAT, 1983).

A verificação consiste em fazer uma análise apenas do ponto de vista de software, procurar bugs, quebra de regras, analisar se não existem regras ativando na hora errada ou pelo motivo errado, verificar erros de digitação ao longo do código fonte e se o programa está se comportando da maneira que o EC deseja.

A validação envolve a análise do conhecimento envolvido no SE, normalmente esta etapa é feita pelos EH que devem testar diferentes cenários no SE e avaliar se a resposta dele é realmente válida ou se possui erros conceituais.

3.2 MINERAÇÃO DE DADOS

O avanço da tecnologia tem favorecido a geração de uma grande quantidade de dados nas mais diversas áreas, por um lado estes dados representam um grande potencial para se descobrir informações que ainda não foram notadas, porém, os seres humanos são limitados pelas suas capacidades físicas e intelectuais para ler e processar manualmente todos estes dados (YE, 2003).

Esta limitação humana gerou a necessidade de desenvolver técnicas e ferramentas automáticas para analisar grandes volumes de dados e extrair deles informações e conhecimento úteis, a estas técnicas e ferramentas é utilizado o termo mineração de dados (*data mining*) (YE, 2003).

Existem diversas técnicas consolidadas para se trabalhar os conjuntos de dados de acordo com o tipo de resultado esperado, os mais comuns são: (TAN, STEINBACH e KUMAR, 2009)

- Árvores de Decisão: Utilizada para prever e classificar modelos encontrando padrões nos dados fornecidos;
- Regras de Associação: Utilizada para encontrar a frequência em que dados ocorrem simultaneamente (Ex: produtos de uma loja que normalmente são encontrados na mesma cesta de compras);
- Redes Neurais Artificiais: Previsão e classificação de dados com possibilidade de aprendizado constante;
- Processos Estatísticos: Utilizados para detectar anomalias em dados, normalmente utilizado com dois conjuntos de dados, um normal e um com erros;

- **Clusterização:** Utilizado para agrupar dados em conjuntos relativamente homogêneos para encontrar padrões de comportamento;

Estas técnicas representam apenas uma ideia superficial de seu funcionamento, existem diversos algoritmos para cada uma e muitos outros que utilizam elementos de mais de uma técnica (ARTERO, 2009).

Neste trabalho uma fração da técnica de árvores de decisão foi explorada e adaptada para criar um mecanismo capaz de selecionar as perguntas dinamicamente, de acordo tanto com as respostas anteriores quanto com o número de resultados do banco de dados, permitindo uma adição constante de novos dados à base sem a necessidade de manutenção nas regras.

Para realizar a classificação de dados, a conexão entre as diferentes classes e suas propriedades pode ser tão simples quanto um pequeno fluxograma ou tão complexa e desestruturada quanto um procedimento inteiramente manual.

Restringindo o estudo para apenas modelos executáveis, ou seja, aqueles que podem ser representados como programas em um computador, temos duas vertentes principais, os já mencionados sistemas especialistas, capturando o conhecimento através da interação com especialistas e o transferindo para um software (ARTERO, 2009), e os modelos indutivos que buscam analisar certa quantidade de dados para encontrar padrões de comportamento (LAKATOS & MARCONI, 2003).

Mccorduck, Feigenbaum e Nii (1988) apresentam como exemplo um sistema desenvolvido pela American Express para auxiliar na autorização de crédito, neste caso as propriedades são os detalhes da proposta e o histórico de créditos do cliente, que quando unidos são classificados em crédito aprovado ou negado.

Como a mineração de dados utiliza somente os dados que lhe são informados, existe a possibilidade de conduzir a informações incorretas, estes podem ser causados por informações erradas ou conflitantes dentro da base de dados. Estes modelos buscam encontrar padrões para classificar, mas existe sempre um percentual de opções que fogem dos padrões, e estes naturalmente serão classificados de maneira equivocada, de acordo com seu padrão teórico (COPPIN, 2013).

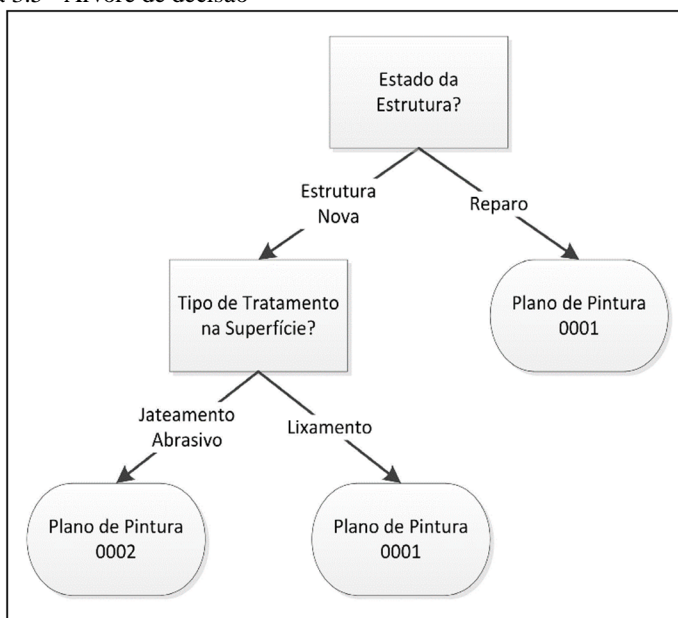
3.2.1 Árvores de decisão

Árvores de decisão são ferramentas poderosas e populares para classificação e predição de resultados. Sua atratividade se deve

principalmente pelo fato de seu resultado se apresentar em um formato semelhante ao de regras, que podem ser expressas em uma linguagem facilmente compreendida por pessoas (YE, 2003).

Árvores de decisão possuem três tipos de elementos: nós, ramos e folhas. Um nó representa uma pergunta, os ramos são suas respostas, e uma folha é um nó que não possui mais nenhuma pergunta para ser feita. Deste modo os nós são conectados pelos ramos até que todas as pontas da árvore sejam do tipo folha (YE, 2003), conforme demonstrado na Figura 3.5.

Figura 3.5 - Árvore de decisão



Fonte: Elaborado pelo Autor.

Na Figura 3.5 existem dois nós representados pelas perguntas “estado da estrutura” e “tipo de tratamento na superfície”, quatro ramos com as respostas “estrutura nova”, “reparo”, “jateamento abrasivo” e “lixamento” e três folhas, duas “plano de pintura 0001” e um “plano de pintura 0002”.

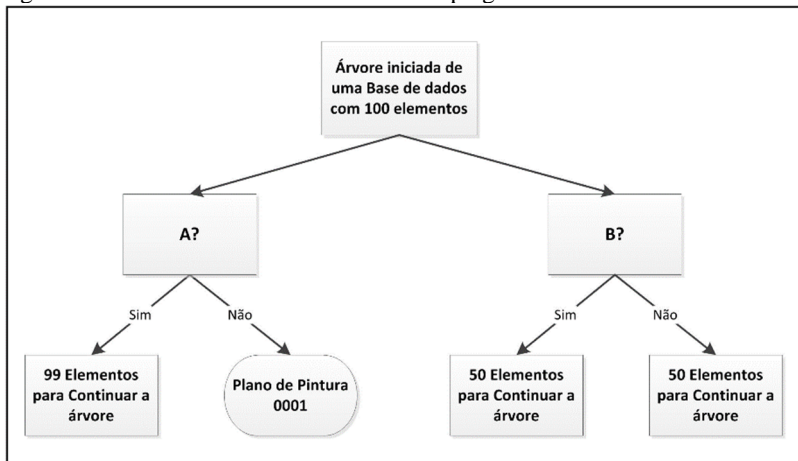
Deste modo pode-se “ler” a árvore como “Se a estrutura for nova e passar por um jateamento abrasivo ela precisa do plano de pintura 0002, já se for uma superfície nova e passar por lixamento, ou uma estrutura para reparos ela precisará do plano de pintura 0001”.

Não existe nenhuma norma ou padrão sobre a representação gráfica de uma árvore de decisão, mas o usual é que os nós e as folhas sejam representados com formas geométricas e/ou cores distintas.

Uma árvore deve possuir nós (perguntas) que garantam o particionamento (divisão) do conjunto de dados de modo expressivo, para que seja possível chegar a um resultado com um número reduzido de perguntas.

Supondo uma base de dados com 100 elementos que serão classificados de acordo com o plano de pintura ideal e diversos atributos que podem ser utilizados como nós (perguntas), analisando um atributo qualquer, denominado “A?” com respostas “sim” ou “não” e que possua 99 elementos “sim” e 1 “não” e outro atributo denominado “B?” com 50 elementos “sim” e 50 “não”, conforme apresentado na Figura 3.6.

Figura 3.6 - Árvore com uma boa e uma má pergunta como nós



Fonte: Elaborado pelo Autor.

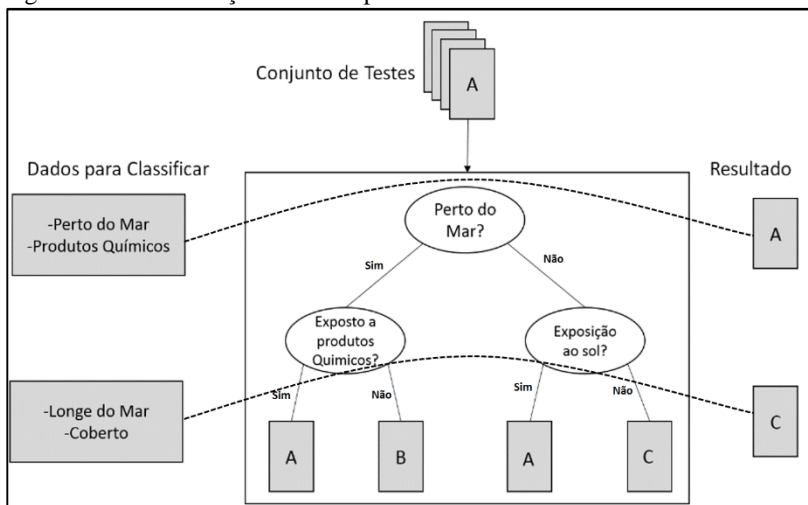
Após a divisão do conjunto de dados original pela pergunta A, ele é subdividido em dois conjuntos: um com 99 elementos que ainda necessitam ser classificados e um subconjunto com apenas 1 elemento já classificado. Após a divisão pela pergunta B, ele é subdividido em dois conjuntos, ambos com 50 elementos que ainda precisarão ser classificados.

O objetivo final de uma árvore de decisão é transformar todos os seus nós em folhas, ou seja, nós em que seu conjunto de elementos pertença a uma classe única, e como conjuntos de menor tamanho

possuem maior probabilidade de pertencerem a uma mesma classe, as perguntas que dividem mais homogêaneamente o conjunto de dados são as “melhores” perguntas, logo, na figura anterior pode-se dizer que a pergunta “B?” é uma pergunta melhor que a pergunta “A?”.

Para criar uma árvore de decisão é necessário um conjunto de testes, as informações deste conjunto servem como base para escolher as perguntas que seguirão para a árvore, estas informações devem ser confiáveis e representativas para que seja possível criar uma boa árvore. A Figura 3.7 apresenta uma pequena árvore de decisão para classificar tintas em classes A, B ou C, e mostra duas tintas sendo classificadas.

Figura 3.7 –Classificação de dados por uma Árvore de Decisão



Fonte: Elaborado pelo Autor.

Em um primeiro momento o conjunto de testes é utilizado para criar a árvore de decisão, e em um segundo momento os dados que se deseja classificar são comparados com a árvore de decisão, de acordo com a Figura 3.7 o primeiro conjunto “perto do mar e exposto a produtos químicos” pertence à classe A e o conjunto “longe do mar e coberto” pertence à classe C.

O conjunto de testes ou de treinamento normalmente se encontra no formato de tabelas, onde uma de suas colunas (geralmente a primeira ou a última) representa a característica que se deseja classificar enquanto as outras são as informações que a árvore utiliza para realizar a classificação.

O Quadro 3.1 apresenta um pequeno conjunto de dados para o treino de uma árvore de decisões que deve classificar se ambientes necessitam de primer epóxi ou não.

Quadro 3.1 - Conjunto de treino para classificar se ambientes necessitam de primer epóxi ou não.

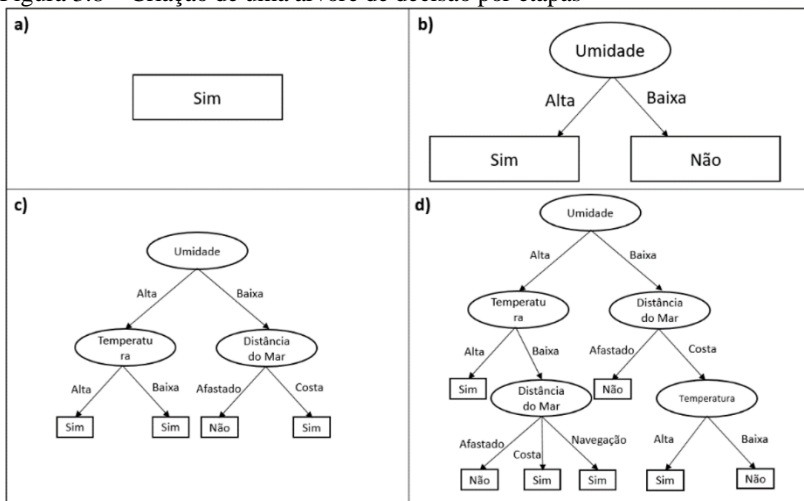
Ambiente	Umidade	Distância do Mar	Temperatura	Necessita de Primer Epoxi?
1	Alta	Navegação	Alta	Sim
2	Baixa	Afastado	Baixa	Não
3	Baixa	Costa	Alta	Sim
4	Alta	Costa	Baixa	Sim
5	Baixa	Afastado	Alta	Não
6	Alta	Afastado	Alta	Sim
7	Alta	Afastado	Baixa	Não
8	Baixa	Costa	Baixa	Não
9	Alta	Navegação	Baixa	Sim

Fonte: Elaborado pelo Autor.

Cada uma das linhas do quadro é um registro individual com sua “identidade” própria, todos os registros compartilham as mesmas características/propriedades, porém são entidades diferentes e, portanto possuem diferentes.

Para ilustrar a construção de uma árvore de decisão, os dados do Quadro 3.1 são utilizados e a Figura 3.8 apresenta esta criação, partindo de uma árvore com apenas um nível até uma com quatro níveis.

Figura 3.8 – Criação de uma árvore de decisão por etapas



Fonte: Elaborado pelo Autor.

O objetivo desta árvore de decisão é classificar os resultados em “sim”, onde necessitam de um primer epóxi ou “não”, quando não necessitam de um primer epóxi. Na Figura 3.8 destaca-se que:

- A árvore possui apenas um nó do tipo folha cuja resposta é “sim”, ou seja, para qualquer dado classificado por esta árvore é recomendado a utilização de um primer epóxi. Se comparada com o Quadro 3.1, temos que 4 dos 9 resultados seriam classificados de maneira errada.
- A árvore utiliza um novo nó com a pergunta sobre a umidade do ambiente, e este nó conduz para duas folhas, o número de casos classificados de maneira errada diminui para 2.
- A árvore cresceu mais um nível, e desta vez pode-se reparar que o nó do segundo nível é diferente para cada um dos ramos, o número de erros continua em 2.
- A árvore cresceu mais um nível, pode-se perceber que alguns nós do terceiro nível são do tipo folha, portanto não existem mais perguntas para se fazer para eles, mas os outros nós continuaram seu processo de crescimento, nesta última árvore todos os casos do conjunto de treinamento conseguem ser perfeitamente classificados, ou seja, nenhum erro.

Dependendo do tipo de aplicação, certo percentual de erros é considerado aceitável, portanto uma árvore de decisão não necessariamente deve ser criada até que todos os seus nós sejam do tipo folha, seu crescimento pode ser interrompido ou ela pode ser “podada” para remover dados que o seu criador deseja ocultar ou considere inúteis.

3.2.2 Algoritmos para árvores de decisão

Um dos primeiros pensamentos para se escolher uma árvore de decisão é o de gerar todas as árvores possíveis que classificam corretamente os dados e então selecionar a mais simples de todas (COPPIN, 2013). Porém, o número de árvores diferentes que podem ser criadas cresce exponencialmente com o número de atributos, mesmo um conjunto de dados relativamente pequeno como o do Quadro 3.1 possui centenas de alternativas, inviabilizando a seleção de uma árvore de decisão deste modo (TAN, STEINBACH e KUMAR, 2009).

Existem algoritmos bastante eficientes para auxiliar no processo de criação, gerando árvores relativamente precisas em uma quantidade razoável de tempo. Estes algoritmos normalmente utilizam uma estratégia que desenvolve a árvore de decisão através de uma série de decisões localmente ótimas sobre qual o melhor atributo para se particionar os dados (TAN, STEINBACH e KUMAR, 2009).

As métricas para selecionar o melhor atributo utilizadas na maioria dos algoritmos provem de equações extraídas das bases da teoria da informação desenvolvidas por Shannon e Weaver (1949) e apresentadas através de seu artigo “*A Mathematical Theory of Communication*”.

O uso da teoria da informação como uma estimativa para seleção de atributos em técnicas de classificação foi proposta por Hunt et al. (1966) e veio a se popularizar através dos algoritmos desenvolvidos por Quinlan: ID3 em 1979 e seu sucessor C4.5 em 1993 (SHMUELI et al., 2017).

Ambos ID3 e C4.5 podem gerar árvores de decisão de maneira automática, necessitando apenas de um conjunto de dados de entrada (conjunto de treinamento) e que se informe qual será o atributo que representará as classes, ou seja, o “resultado” da árvore.

A ideia geral de ambos os algoritmos é montar uma árvore de decisão de modo que cada nó da árvore represente um teste simples e objetivo (uma pergunta ou um tipo de filtro) sobre um atributo (qual a idade?, possui emprego?) podendo agrupar os resultados de cada teste em classes como “< 30, 30~40, > 40”, em valores unitários como “25,26,27”

ou “Branco,Preto,Marrom”, e em resultados binários “Sim/Não” (BRAMER, 2016).

Após a seleção de um nó, suas folhas são avaliadas individualmente e, se necessário, transformadas em novos nós para serem associadas a um novo atributo de teste até que todos os nós sejam do tipo folha.

3.2.2.1 Seleção de Atributos

Um dos principais aspectos para uma árvore de decisão ser representativa é a escolha dos atributos que serão selecionados para cada um dos nós da árvore (NAGABHUSHANA, 2006).

No algoritmo de Hunt, a árvore de decisão cresce de forma recursiva através do particionamento dos registros de treino em sucessivos subconjuntos cada vez mais puros, ou seja, com uma menor quantidade de classes diferentes (NOGUEIRA, 2015).

O algoritmo possui basicamente três passos:

- Se todos os registros de um nó pertencerem a mesma classe, então este nó é do tipo folha;
- Se um nó possuir registros pertencentes a mais de uma classe, uma condição de teste é selecionada para particionar os registros em subconjuntos e para cada resultado possível desta condição de teste um novo nó filho é criado e os registros do nó que os originou é distribuído baseado nos resultados.
- O algoritmo é então aplicado recursivamente para cada nó filho.

Existem muitas métricas que podem ser usadas para escolher os atributos e sua ordem para garantir que os registros sejam divididos da melhor forma, normalmente estas se baseiam na análise da distribuição de classes dos registros antes e depois da divisão.

O primeiro algoritmo desenvolvido por Quinlan (1979), o ID3 utiliza duas métricas para selecionar os atributos: entropia e ganho de informação. Seu segundo algoritmo, o C4.5, acrescentou duas novas métricas: informação de divisão e razão de ganho. Estas métricas serão definidas e sua fundamentação matemática apresentada nos tópicos a seguir.

3.2.2.2 Seleção dos atributos pelo ID3

Uma das métricas vindas da teoria da informação é denominada entropia, que apresenta um conceito onde:

“A informação transmitida por uma mensagem depende de sua probabilidade e pode ser medida em bits como menos o logaritmo em base 2 de sua probabilidade. Então, por exemplo, se existem oito mensagens de igual probabilidade de serem enviadas, a informação transmitida por cada uma delas é $-\log_2(1/8)$ ou 3 bits”
Quinlan (1993, tradução nossa).

Quando a entropia é aplicada para a seleção de atributos, pode ser definida como o nível de desordem da informação (NOGUEIRA, 2015) ou a impureza de um conjunto de dados (NAGABHUSHANA, 2006).

Uma maneira simples para interpretar estes conceitos é dizer que a entropia representa a chance de que, ao selecionar aleatoriamente sucessivos elementos de um conjunto, cada novo elemento selecionado seja diferente do anterior.

De acordo com Tan, Steinbach e Kumar (2009), a entropia apresenta duas propriedades principais:

- A entropia de um conjunto de dados é zero quando todos os seus elementos pertencem à mesma classe (nó do tipo folha);
- A entropia de um conjunto de dados é máxima quando todas as classes possuem o mesmo número de elementos.

Matematicamente, a entropia é apresentada conforme equação 3-1 (QUINLAN, 1979).

$$H(S) = - \sum_{j=1}^k \frac{freq(C_j, S)}{|S|} * \log_2 \left(\frac{freq(C_j, S)}{|S|} \right) \quad (3-1)$$

onde:

$H \rightarrow$ Entropia [bits]

$S \rightarrow$ Conjunto de registros

$k \rightarrow$ Número de classes diferentes em S

$C_j \rightarrow$ j-esima classe de S

$freq(C_j, S) \rightarrow$ Número de registros C_j que pertencem a S

$|S| \rightarrow$ Total de registros em S

Para se entender ainda melhor este conceito, pode-se comparar a diferença na entropia do lançamento de um dado, uma moeda e uma moeda viciada que tende a cair $\frac{3}{4}$ das vezes para o lado “cara” (estes conjuntos possuem somente um atributo: “face” cujas alternativas são “1,2,3,4,5,6” para o dado ou “Cara,Coroa” para a moeda).

$$H(\text{Dado}) = - \sum_{j=1}^6 \frac{1}{6} * \log_2 \left(\frac{1}{6} \right) = 2,584 \text{ bits}$$

$$H(\text{Moeda}) = - \sum_{j=1}^2 \frac{1}{2} * \log_2 \left(\frac{1}{2} \right) = 1,000 \text{ bits}$$

$$H(\text{Viciada}) = \left(-\frac{1}{4} * \log_2 \left(\frac{1}{4} \right) \right) + \left(-\frac{3}{4} * \log_2 \left(\frac{3}{4} \right) \right) = 0,811 \text{ bits}$$

Com estes exemplos a definição proposta anteriormente que a entropia representa “a chance de que, ao selecionar aleatoriamente sucessivos elementos de um conjunto, cada novo elemento selecionado seja diferente do anterior”, fica mais clara.

Um dado apresenta um conjunto de 6 resultados possíveis {1, 2, 3, 4, 5, 6}, com todos os registros diferentes e de igual probabilidade de ocorrerem, deste modo, tem-se uma probabilidade de 1/6 de se tirar o mesmo resultado duas vezes, o que resulta em uma entropia maior que a dos outros exemplos.

Uma moeda apresenta somente 2 resultados possíveis {Cara, Coroa}, os dois diferentes entre si e também com a mesma probabilidade de ocorrer de $\frac{1}{2}$, deste modo sua entropia é menor que a do dado.

A moeda viciada apresenta $\frac{3}{4}$ de chances de se cair “cara”, convertendo esta probabilidade para um conjunto de casos tem-se {Cara, Cara, Cara, Coroa}, como o número de casos para cada classe possível é diferente, sua entropia é menor que a de uma moeda não viciada apesar de possuir o mesmo conjunto de respostas possíveis {Cara, Coroa}.

A entropia sozinha não é suficiente para selecionar um atributo, pois ela apenas representa o quão impuro o conjunto de dados está no momento “atual”. Deste modo, uma segunda métrica é utilizada em conjunto com a entropia, denominada ganho de informação (Quinlan, 1986).

Este ganho de informação representa a diferença de entropia antes e depois de se dividir o conjunto de dados por determinado atributo, portanto, um ganho de informação maior representa uma diminuição na

entropia do subconjunto resultante desta divisão e conseqüentemente uma menor impureza em seus registros. A equação 3-2 representa o ganho de informação (QUINLAN, 1979).

$$Gain(S, T) = H(S) - H(S, T) \quad (3-2)$$

onde:

$Gain(S, T) \rightarrow$ Ganho de informação do conjunto S após divisão pelo atributo T [bits]

$H(S) \rightarrow$ Entropia do conjunto S [bits]

$H(S, T) \rightarrow$ Entropia do conjunto S após divisão pelo atributo T [bits]

É preciso ter em mente que um conjunto de dados pode possuir vários atributos e por sua vez cada atributo pode possuir várias alternativas, e cada alternativa de cada atributo resultará em um subconjunto diferente.

Dividindo o grupo de dados já apresentado no Quadro 3.1 pelo atributo “umidade” resultará em um subconjunto para a resposta “alta” e outro para a resposta “baixa”, e caso o conjunto original seja dividido por outro atributo, serão outras respostas, e conseqüentemente outros subconjuntos resultantes.

Para contemplar este problema o ganho de informação deve ser calculado de maneira isolada para cada atributo, e as diferentes respostas de cada atributo serão combinadas através de uma média ponderada.

A equação 3-3 representa a entropia após a divisão de um conjunto S por um atributo T (QUINLAN, 1979).

$$H(S, T) = - \sum_{j=1}^n \frac{|T_j|}{|T|} * H(S_j) \quad (3-3)$$

onde:

$H(S, T) \rightarrow$ Entropia do conjunto S após divisão pelo atributo T [bits]

$S \rightarrow$ Conjunto S (antes da divisão)

$T \rightarrow$ Atributo para divisão

$n \rightarrow$ Número de alternativas para o atributo T

$S_j \rightarrow$ Subconjunto S após divisão pelo atributo T com resposta j

$H(S_j) \rightarrow$ Entropia do subconjunto S após divisão pelo atributo T com resposta j

$|T| \rightarrow$ Total de ocorrências do atributo T

$|T_j| \rightarrow$ Total de ocorrências do atributo T com resposta j

Para exemplificar melhor estes conceitos, o Quadro 3.2 apresenta uma modificação do Quadro 3.1, onde a coluna Temperatura foi substituída pela classificação através da ISO 12944 (2015), vale ressaltar que ambos os quadros foram extraídas da base de dados original do protótipo.

Quadro 3.2 - Conjunto de teste para classificar se ambientes necessitam de primer epóxi ou não.

Ambiente	Umidade	Distância do Mar	ISO 12944	Necessita de Primer Epoxi?
1	Alta	Navegação	C5-M	Sim
2	Baixa	Afastado	C3	Não
3	Baixa	Costa	C5-M	Sim
4	Alta	Costa	C5-M	Sim
5	Baixa	Afastado	C3	Não
6	Alta	Afastado	C3	Sim
7	Alta	Afastado	C3	Não
8	Baixa	Costa	C3	Não
9	Alta	Navegação	C5-M	Sim

Fonte: Elaborado pelo Autor.

Para iniciar a criação da árvore de decisão com os dados da tabela deve-se iniciar selecionando sobre qual atributo será realizada a classificação, neste caso o atributo selecionado será a última coluna, “necessita de primer epóxi?” com as respostas (classes) “sim” e “não”.

O próximo passo é selecionar um atributo para ser o primeiro nó da árvore, e este deve ser o atributo que representa o maior ganho de informação para o conjunto de dados.

O início do cálculo para se obter o ganho de informação de cada atributo é calcular a entropia inicial. O conjunto de dados possui 5 registros classificados como “sim” e 4 classificados como “não”, resultando em um total de 9 registros, utilizando estas informações na equação 3-1 temos:

$$H(S) = \left(-\frac{4}{9} * \log_2 \frac{4}{9}\right)_{N\tilde{A}O} + \left(-\frac{5}{9} * \log_2 \frac{5}{9}\right)_{SIM} = 0.991 \text{ bits}$$

Ou seja, o conjunto de dados do Quadro 3.2 possui uma entropia de 0.991 bits.

Em seguida deve-se calcular a entropia após a divisão para cada um dos atributos através da equação (3-3). A ordem de qual atributo é calculado primeiro é irrelevante, apenas é necessário que todos sejam calculados. Para o atributo “distância do mar” o cálculo é:

$$\begin{aligned}
 H(S, Dist) = & \left(\frac{2}{9} * \left(\left(-\frac{2}{2} * \log_2 \frac{2}{2} \right)_{SIM} + \left(-\frac{0}{2} * \log_2 \frac{0}{2} \right)_{NAO} \right) \right)_{Navegação} + \\
 & \left(\frac{4}{9} * \left(\left(-\frac{1}{4} * \log_2 \frac{1}{4} \right)_{SIM} + \left(-\frac{3}{4} * \log_2 \frac{3}{4} \right)_{NAO} \right) \right)_{Afastado} + \\
 & \left(\frac{3}{9} * \left(\left(-\frac{2}{3} * \log_2 \frac{2}{3} \right)_{SIM} + \left(-\frac{1}{3} * \log_2 \frac{1}{3} \right)_{NAO} \right) \right)_{Costa} \\
 H(S, Dist) = & \frac{2}{9} * 0.000 + \frac{4}{9} * 0.811 + \frac{3}{9} * 0.918 = 0.667 \text{ bits}
 \end{aligned}$$

Dos 9 registros originais existem: 2 registros com a resposta navegação, e destes, 2 são da classe “sim” e nenhum¹ da classe “não”, 4 registros com a resposta afastado, 1 da classe “sim” e 3 da classe “não” e 3 registros da classe costa, 2 da classe “sim” e 1 da classe “não. Deste modo, os cálculos sugerem que a entropia esperada após a divisão pelo atributo distância do mar é de 0.667 bits.

Com acesso às informações da entropia inicial do conjunto e da entropia depois da divisão pelo atributo clima é possível utilizar a equação 3-2 para calcular qual seria o ganho de informação ao se dividir o conjunto de dados pelo atributo distância do mar.

$$Gain(S, Dist) = 0.991 - 0.667 = 0.324 \text{ bits}$$

Resultando em 0.324 bits de ganho de informação ao se dividir o conjunto de dados pelo atributo “distância do mar”.

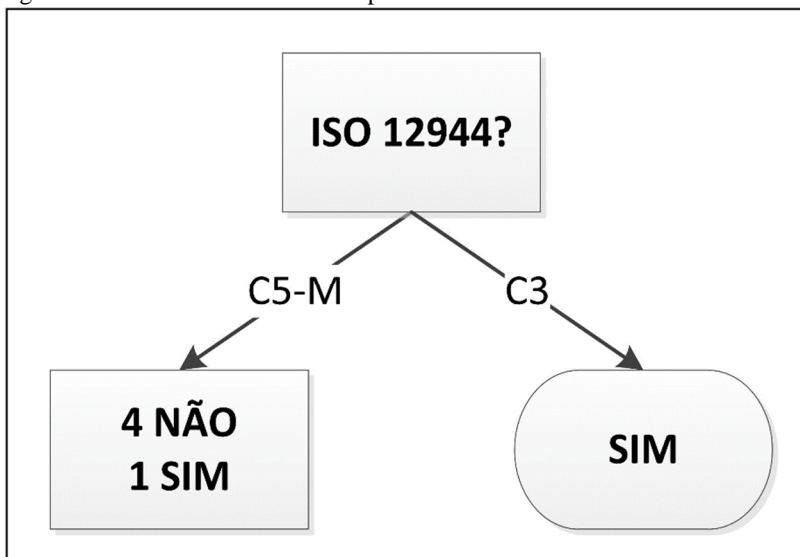
O próximo passo é repetir os cálculos para os outros atributos do conjunto, “umidade” e “ISO 12944” o que resulta em:

¹ Para fins práticos, o resultado do $\log_2(0)$ que apresenta um resultado indefinido é considerado simplesmente como 0, para evitar o trabalho de remover todas as classes nulas dos cálculos.

$$\begin{aligned} \text{Gain}(S, \text{Dist}) &= 0.247 \text{ bits} & \text{Gain}(S, \text{Umidade}) &= 0.229 \text{ bits} \\ \text{Gain}(S, \text{ISO}) &= 0.590 \text{ bits} \end{aligned}$$

Pode-se perceber que o atributo com o maior ganho é o “ISO 12944”, com 0.590 bits, o que resulta em um conjunto com uma impureza relativamente menor. A Figura 3.9 representa a árvore utilizando a ISO como o primeiro nó.

Figura 3.9 – Árvore de decisão com primeiro nó



Fonte: Elaborado pelo Autor.

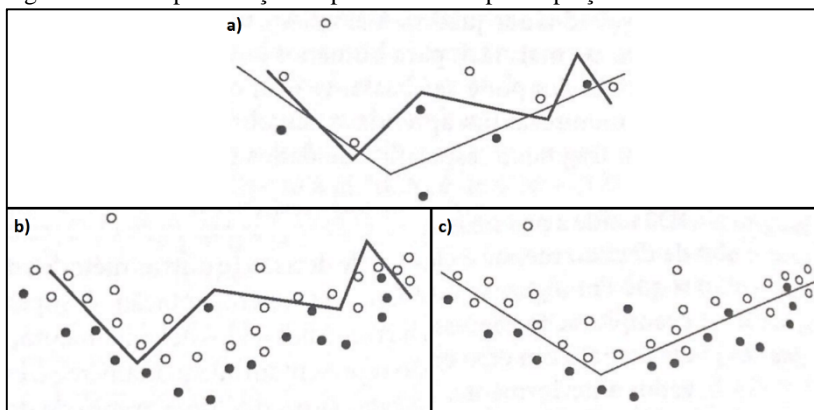
Selecionando o atributo ISO para ser o primeiro nó da árvore, suas respostas geram dois novos nós através dos ramos C5-M e C3. Como a resposta C3 gera apenas resultados da classe “sim”, este é um nó do tipo folha e não precisa mais ser expandido. Mas o C5-M possui resultados de classes diferentes e deve ser repetido o processo de cálculo para encontrar o próximo atributo para a divisão e completar a árvore.

A utilização do ganho de informação é um excelente parâmetro para selecionar o melhor nó, porém ele apresenta uma tendência a selecionar atributos com um grande número de alternativas, o que pode acarretar em dois tipos de problemas: Superadaptação e fragmentação (MITRA E ACHARYA, 2003).

- Superadaptação: A árvore fica dimensionada para funcionar somente com o conjunto de treinamento, sem representar adequadamente novos cenários e realizar sua função de classificação.
- Fragmentação: Os dados ficam altamente fragmentados, resultando em um grande número de pequenos conjuntos e uma árvore de baixa eficiência.

A Figura 3.10 representa o problema de superadaptação em uma notação diferente, mas que pode ser interpretada de modo semelhante para árvores de decisão.

Figura 3.10 – Representação do problema de superadaptação.



Fonte: Adaptado de Coppin (2013)

A Figura 3.10 apresenta um conjunto de treinamento que deve ser classificado com uma linha, de modo que os resultados que se encontram acima dela sejam brancos e os abaixo sejam pretos. Em a) dois modelos são gerados, um com linha fina que representa uma árvore de decisão “tradicional”, enquanto a linha grossa representa uma árvore de decisão superadaptada para contemplar todos os pontos e atributos do conjunto.

Na parte de baixo da Figura 3.10, um conjunto de testes é inserido para que o modelo classifique seus registros, porém percebe-se claramente que os erros de classificação da árvore superadaptada (b) são muito maiores que os da tradicional (c).

Este problema fica bastante evidente ao se realizar o cálculo para o ganho de informação levando em consideração o atributo Ambiente

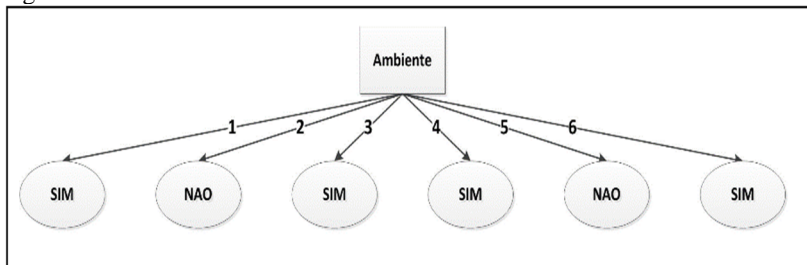
presente na primeira coluna do Quadro 3.2, onde cada um dos 9 ambientes possui apenas uma classe associada, o que resulta nos seguintes cálculos:

$$H(S, Ambiente) = - \sum_1^{14} \frac{1}{14} * \left(-\frac{1}{1} * \log_2 \frac{1}{1} \right) * \left(-\frac{0}{1} * \log_2 \frac{0}{1} \right) = 0$$

$$Gain(S, Ambiente) = 0.940 - 0 = 0.940 \text{ bits}$$

Como cada um dos 9 ambientes está relacionado a uma única classe, todos os seus resultados são nós do tipo folha, o ganho de informação é máximo, a entropia resultante é 0 e a árvore de decisão está finalizada com uma única pergunta. A Figura 3.11 mostra uma parte da árvore de decisão e o motivo claro do porque é uma árvore de eficiência baixa.

Figura 3.11 – Árvore de decisão com um atributo de muitos valores



Fonte: Elaborado pelo Autor.

Quanto menor o número de valores diferentes para um atributo, maiores as chances da pergunta ser de uma complexidade inferior e de dividir melhor o conjunto de dados, trazendo consigo maior quantidade de informação. Uma pergunta com respostas “sim” ou “não” é provavelmente menos complexa que uma pergunta que possua 10 alternativas diferentes.

Desenvolvida deste modo a árvore não possui praticamente nenhuma utilidade como classificadora, pois não utilizou nenhum dado significativo para sua criação.

Este problema foi parcialmente solucionado com a adição de novas equações no seu algoritmo, apresentadas no próximo tópico.

3.2.2.3 Seleção de atributos pelo C4.5

O algoritmo C4.5 possui diversas novas funcionalidades em relação ao seu antecessor ID3, novas técnicas para gerenciar as árvores

de decisão, auxílio em tomadas de decisão, predição de resultados incompletos e certo tratamento para dados incoerentes.

Em relação à seleção de atributos para selecionar os nós das árvores, duas novas equações foram inseridas para aumentar a qualidade das árvores: informação de divisão e taxa de ganho.

A informação de divisão representa quão dividida está a informação contida no atributo de teste, ou seja, quanto maior o número de classes relacionadas ao atributo mais dividido ele é considerado e menores são as chances do algoritmo o selecionar. A Equação 3-4 apresenta a informação de divisão (QUINLAN, 1993).

$$info\ split_{(T)}S = \sum_j^k \frac{|T_j|}{|T|} * \log_2 \frac{|T_j|}{|T|} \quad (3-4)$$

onde:

$info\ split_{(T)}S \rightarrow$ Divisão da informação do conjunto S pelo atributo T [bits];

$|T| \rightarrow$ Total de ocorrências do atributo T;

$k \rightarrow$ Número de valores diferentes do atributo T;

$|T_j| \rightarrow$ Total de ocorrências do atributo T com valor j;

Aplicando a equação no mesmo exemplo do Quadro 3.2 para os atributos ambiente e distância do mar, sendo que o ambiente apresenta 9 classes diferentes em seus 9 registros, cada classe com 1 ocorrência por registro e para a distância que apresentou 3 classes onde: 2 registros são da classe navegação, 4 registros da classe afastado e 3 registros da classe costa, tem-se:

$$info\ split_{(amb)}S = - \sum_i^9 \frac{1}{9} * \log_2 \frac{1}{9} = - \left(9 * \frac{1}{9} * \log_2 \frac{1}{9} \right) = 3.169\ bits$$

$$info\ split_{(dist)}S = - \sum_i^3 \frac{|T_j|}{9} * \log_2 \frac{|T_j|}{9} =$$

$$\left(-\frac{2}{9} * \log_2 \frac{2}{9} \right) + \left(-\frac{4}{9} * \log_2 \frac{4}{9} \right) + \left(-\frac{3}{9} * \log_2 \frac{3}{9} \right) = 1.530\ bits$$

Como todas as classes do atributo “distância do mar” possuem a mesma frequência, o somatório pode ser simplificado resultando em uma informação da divisão de 3.169 bits e para o atributo distância da costa a informação de divisão é de 1.530 bits. Esta diferença acontece porque a

informação em um atributo com 9 classes certamente está mais dividida que em um atributo com apenas 3 classes.

Repetindo os cálculos para os dois atributos restantes:

$$\text{info split}_{(umid)}S = - \sum_i^2 \frac{|T_j|}{9} * \log_2 \frac{|T_j|}{9} = 0.991 \text{ bits}$$

$$\text{info split}_{(ISO)}S = - \sum_i^2 \frac{|T_j|}{9} * \log_2 \frac{|T_j|}{9} = 0.991 \text{ bits}$$

A informação de divisão dos atributos ISO e Umidade é a mesma, pois ambos possuem apenas duas classes com 5 registros em uma e 4 registros na outra.

A próxima métrica combina o ganho de informação com a informação de divisão para uma característica denominada taxa de ganho (gain ratio), esta taxa de ganho é a característica final utilizada para determinar o atributo que será utilizado como nó, quanto maior a taxa de ganho, “melhor” é o atributo e menos impuro deve ser o resultado de sua divisão, a Equação 3-5 representa a taxa de ganho (QUINLAN, 1993):

$$\text{gain ratio}(S, T) = \frac{\text{gain}(S, T)}{\text{info}_{(T)}S} \quad (3-5)$$

Aplicando-a para os exemplos anteriores:

$$\text{gain ratio}(S, amb) = \frac{0,991 \text{ bits}}{3,169 \text{ bits}} = 0.311$$

$$\text{gain ratio}(S, dist) = \frac{0,211 \text{ bits}}{1,530 \text{ bits}} = 0.211$$

$$\text{gain ratio}(S, umid) = \frac{0,229 \text{ bits}}{0,911 \text{ bits}} = 0.231$$

$$\text{gain ratio}(S, ISO) = \frac{0,590 \text{ bits}}{0,911 \text{ bits}} = 0.595$$

A razão de ganho com o maior valor é o do atributo ISO, portanto este seria selecionado como o primeiro nó para a árvore.

É perceptível como a preferência por atributos com um grande número de classes foi reduzida, pois além do atributo ISO ter sido selecionado, a diferença entre as razões de ganho para um atributo com muitas classes e um com poucas classes ficou reduzida, quando utilizada

em uma grande base de dados estes valores se tornam cada vez mais evidentes.

Este capítulo também é utilizado de maneira introdutória para a construção do protótipo apresentada no capítulo 4.

Os pontos mais relevantes da criação de Sistemas Especialistas que foram utilizados na construção do protótipo estão aqui apresentados para preparar o leitor para a dinâmica acelerada do próximo capítulo.

A técnica de árvore de decisão é abordada e os algoritmos ID3 e C4.5 são apresentados com detalhes pois foram utilizados para controlar a dinâmica de todo o protótipo, para isso foram realizadas algumas modificações conceituais neles para os adaptar ao cenário proposto.

4. DESENVOLVIMENTO DO PROTÓTIPO

Este capítulo apresenta as diretrizes seguidas durante as etapas iniciais de desenvolvimento do protótipo, explica as motivações para sua realização e as dificuldades encontradas durante sua execução.

Apresenta a primeira parte da sequência de etapas já mencionada no capítulo anterior sobre o desenvolvimento de SE, abordando apenas as etapas de formalização de conhecimento e principal geração de conteúdo.

Inicia com o projeto conceitual (análise de viabilidade), aquisição de conhecimento, representação de conhecimento e finaliza com a implementação computacional.

O protótipo desenvolvido pode ser definido como um sistema híbrido entre sistemas especialistas com um classificador do tipo árvore de decisão, resultando em uma variação de um configurador de produtos que comercialmente é conhecido como venda guiada, onde perguntas são feitas para o usuário através de um processo interativo para coletar informações e as transformar em requisitos de projeto para poder lhe propor uma solução aparentemente adequada (aparentemente, pois por depender do conhecimento de um especialista, este pode não representar a totalidade do problema).

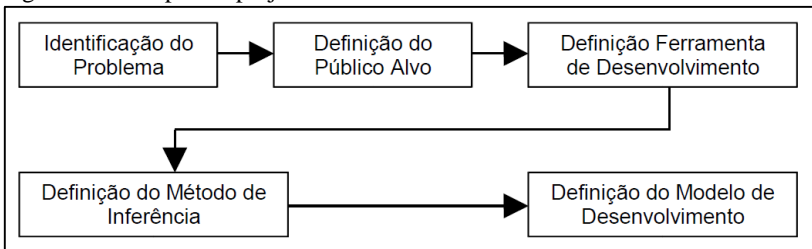
4.1 PROJETO CONCEITUAL

O projeto conceitual consiste em coletar informações sobre o projeto para realizar avaliações prévias antes de iniciar efetivamente a produção de conteúdo (ALVES, 2001), podendo ser considerado como uma análise de viabilidade do projeto, avaliando se o problema abordado está bem definido e qual o impacto que a solução proposta pode gerar sobre o mesmo.

Zimmermann (2003) afirma que “[...] decisões tomadas nas fases iniciais do desenvolvimento de produtos causam um grande impacto no ciclo de vida destes”. Ou seja, quanto antes os erros forem detectados, menor será seu impacto, o que acarreta em um menor custo com retrabalhos, entrega em um prazo menor, maior confiabilidade no sistema.

A Figura 4.1 apresenta as principais etapas de um projeto conceitual, estas serão devidamente descritas nos tópicos seguintes.

Figura 4.1 – Etapas do projeto conceitual



Fonte: Alves, 2001

4.1.1 Identificação do Problema

O projeto iniciou através de uma apresentação para a empresa WEG Tintas, daqui por diante mencionada como empresa-alvo, de um protótipo de SE desenvolvido durante a disciplina Sistemas Especialistas aplicados à Engenharia para a seleção de esquemas de pintura, que utilizava perguntas sobre requisitos de projeto do usuário para recomendar esquemas de pintura adequados.

A empresa relatou dois problemas que aparentemente poderiam ser solucionados pelo protótipo apresentado: 1) Revendedores autorizados necessitando de muito treinamento para conhecerem adequadamente o mercado de revestimentos e as soluções que a WEG tem a oferecer e 2) Dificuldades relatadas por clientes sobre a plataforma de e-commerce da empresa em relação ao grau de conhecimento necessário para sua utilização.

A proposta inicial consistia em utilizar uma modificação da base de dados e lista de perguntas do protótipo apresentado, originalmente desenvolvido com o software CLIPS, de modo que contemplasse as linhas de produto da empresa, uma tarefa que consistiria principalmente de duas etapas: 1) Aquisição e formalização do conhecimento da empresa através de análises a seus dados e entrevistas com seus vendedores e especialistas e 2) Criação de regras que utilizassem os novos dados coletados.

Na primeira reunião com os responsáveis, a empresa-alvo informou que desejava um software que fosse facilmente ampliável, de modo que fosse permitido a própria equipe de vendas inserir, modificar e remover perguntas e esquemas de pintura da base de dados. Este novo requisito passou a inviabilizar a utilização do CLIPS, pois sua principal força se encontra no encadeamento de regras pré-definidas, e para inserir ou remover novas perguntas, inevitavelmente as regras teriam de ser alteradas, o que exigiria a necessidade de um EC para a tarefa.

Durante a busca por uma alternativa viável, se deparou com o website "Akinator, o gênio da internet" (ELOKENCE, 2007), que aparentemente utilizava técnicas de árvores de decisão para identificar personagens através de perguntas com respostas "sim", "não", "não sei", "provavelmente sim" e "provavelmente não" (seu real algoritmo não é de domínio público).

Com as técnicas de árvore de decisão em mente, testes foram realizados utilizando modelos híbridos com SE e foi verificado que os novos requisitos seriam atendidos, porém sua implantação com o software CLIPS seria ineficiente, então tanto software quanto linguagem de programação foram alterados para opções que a empresa já utilizava: Eclipse como software e Java/Javascript como linguagem de programação.

4.1.2 Definição do Público Alvo

O protótipo deve ser utilizado por quatro tipos de pessoas: Vendedores experientes, vendedores inexperientes, clientes experientes e clientes inexperientes, descritos a seguir:

- Vendedores experientes: o utilizarão para facilitar seu trabalho, respondendo rapidamente a todas perguntas e provavelmente chegando ao resultado final sem grandes dificuldades, eles ainda podem o utilizar como ferramenta treino para desenvolver ainda mais seus conhecimentos sobre os produtos;
- Vendedores inexperientes: também o utilizarão para facilitar seu trabalho, provavelmente respondendo às perguntas de forma mais lenta e utilizando os menus de ajuda quando não compreenderem corretamente uma questão e ignorando as que mesmo com o a ajuda desconheçam a resposta, poderão também o utilizar como ferramenta de aprendizado para conhecer melhor os produtos e se especializar cada vez mais;
- Clientes Experientes: Conhecem muito bem os requisitos acerca de seu projeto e responderão as perguntas de forma rápida e utilizarão o menu de ajuda pontualmente;
- Clientes Inexperientes: Não possuem nenhum conhecimento do produto que necessitam e apresentam uma pequena noção dos requisitos de seu projeto, deverão utilizar o menu de ajuda em grande parte das questões.

Como os usuários podem ter níveis de conhecimento de informática variados, o protótipo é desenvolvido com uma interface simples e amigável. Apenas com as noções básicas para abrir um website e utilização de mouse/teclado deve ser possível o utilizar sem maiores problemas.

4.1.3 Definição da Ferramenta de Desenvolvimento

Duas ferramentas foram avaliadas para se desenvolver o protótipo:

- CLIPS: um ambiente de sistema especialista gratuito que utiliza uma linguagem semelhante a LISP, utiliza uma linguagem para se expressar com o uso de regras e orientação a objetos. Não apresenta muitas ferramentas de apoio, sendo comum a criação do código em um bloco de notas e posterior utilização do software para testar e depurar² (debuggar) o código.
- Eclipse: é um ambiente de desenvolvimento integrado (IDE), utilizado principalmente para o desenvolvimento de aplicações, com suporte às principais linguagens de programação utilizadas atualmente. Possui muitos recursos que auxiliam durante as etapas de geração de código para que se consiga focar ao máximo na tarefa de programar. Possui recursos para completar sentenças, nomes de variáveis, verifica a todo momento se o código possui erros de sintaxe, permite modificações em massa, apresenta um excelente depurador interno e diversas outras funcionalidades que facilitam e aceleram o trabalho de um programador (ECLIPSE, 2017).

O software selecionado foi o Eclipse devido ao amplo suporte que este oferece, e considerando que a empresa-alvo já o utiliza em diversas outras aplicações, esta já possuía um grande conhecimento prévio desta ferramenta.

² Processo para monitorar a execução de um programa, visando encontrar e reduzir defeitos, pode apontar linhas de código com problema, pausar uma execução, enviar mensagens informando o estado de variáveis, gerar relatórios, etc.

4.1.4 Definição do Método de Inferência

A venda guiada pode ser resumida como uma aplicação sucessiva de filtros sobre uma base de dados até que não existam novos filtros para se aplicar.

Este conceito segue o método de inferência direto, onde os dados inseridos pelo usuário são utilizados de maneira progressiva, sem a necessidade de se gerar hipóteses.

4.1.5 Definição do Modelo de Desenvolvimento

O protótipo foi construído de modo incremental, todas as funções planejadas foram separadas em etapas distintas, onde cada etapa consistia em sua programação, testes isolados e depois testes com funções desenvolvidas anteriormente.

4.2 AQUISIÇÃO DE CONHECIMENTO

Como a empresa-alvo tinha como requisito um software tão simples que fosse possível que a própria equipe de vendas realizasse a ampliação dos dados do sistema, podendo adicionar tanto novos esquemas de pintura quanto novas perguntas, foi decidido que a aquisição de conhecimento seria dividida em duas etapas, uma rápida feita em conjunto do EC com o EH para dar início ao desenvolvimento do protótipo, e uma segunda mais completa quando o protótipo já estivesse completamente desenvolvido, podendo ter a participação do EC apenas enquanto os EH não possuísem domínio da ferramenta.

Na primeira etapa, os dados utilizados serviram apenas como base para a criação de todas as regras do protótipo e apresentações da evolução do mesmo para as gerências.

Portanto foi coletada uma pequena quantidade de informações sobre apenas uma das linhas de atuação do segmento de tintas líquidas da empresa, e utilizada a ISO 12944 para escolher quais informações seriam suficientes e representativas para selecionar os esquemas.

Na segunda etapa, os dados utilizados na primeira foram removidos do sistema, e através de entrevistas e reuniões entre o EC e o EH uma base de dados mais complexa foi elaborada.

Uma adição contínua de novos dados está prevista para o banco de dados, e conforme a exploração de novos segmentos for progredindo, certas perguntas devem se tornar obsoletas, para isso foi definido um padrão de característica “não se aplica”, que o protótipo leva em

consideração na hora de selecionar a pergunta, deste modo irá evitar perguntas inadequadas para segmentos específicos.

4.3 REPRESENTAÇÃO DO CONHECIMENTO

Sendo utilizadas principalmente as linguagens Java e Javascript para a criação do protótipo, se optou por utilizar técnicas de programação orientada a objetos (POO) para representar o tipo de conhecimento analisado.

Foi desenvolvido um total de 12 classes com métodos e funções próprias, e estas por sua vez utilizam diversos métodos e instâncias herdados de classes nativas das bibliotecas Java. Estas classes são manipuladas de maneira direta por uma classe central denominada *ScriptVenda* seguindo uma rotina relativamente procedural mas sem abrir mão do dinamismo e descentralização da POO.

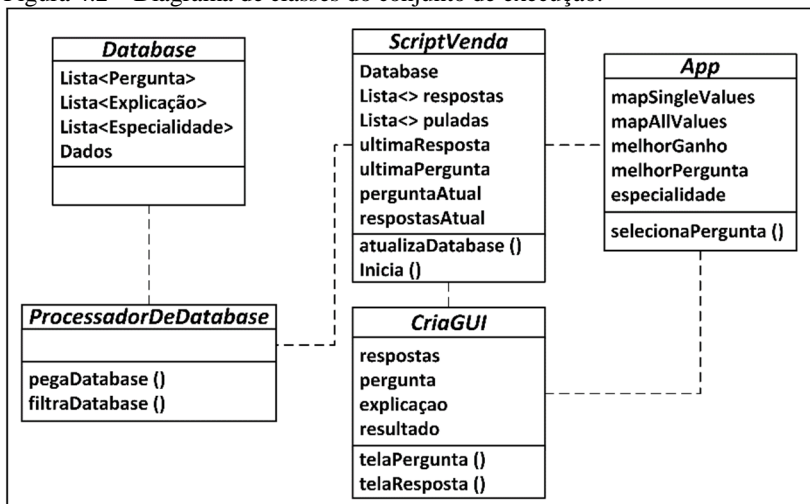
É possível separar as classes em dois conjuntos: execução e atualização. O conjunto de execução é composto por 5 classes, contemplando o script principal e todas as regras para iniciar e controlar o funcionamento do protótipo, já o de atualização é composto de 9 classes e realiza a atualização do banco de dados através de um arquivo externo. Duas destas classes estão presentes em ambos os conjuntos, mas são acionados métodos distintos em cada utilização.

4.3.1 Conjunto de execução

O conjunto de execução é o responsável pela grande maioria das operações do protótipo, responsável por carregar a base de dados do servidor, selecionar a sequência das perguntas, criar os elementos visuais para interagir com o usuário e permitir que se selecione respostas, que se tenha acesso aos menus de ajuda e ainda apresentar o resultado final no término de cada consulta.

É composto por 5 classes representadas na Figura 4.2.

Figura 4.2 – Diagrama de classes do conjunto de execução.



Fonte: Elaborado pelo Autor.

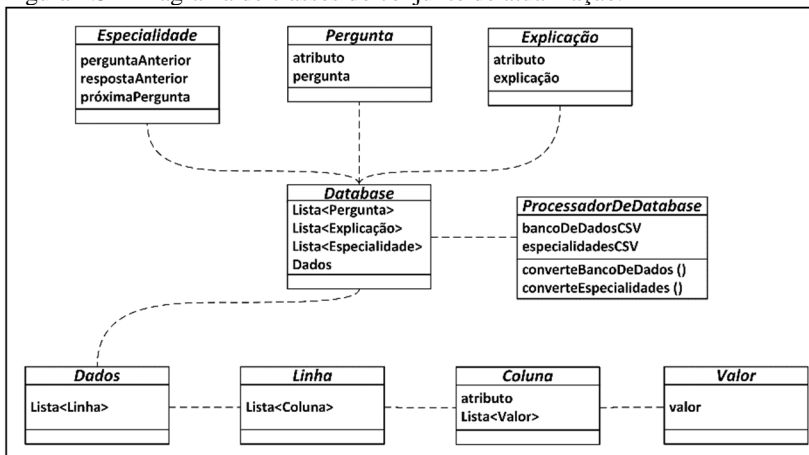
- *ScriptVenda*: A classe central responsável por gerenciar todas as outras classes do conjunto, controla a execução de todas as regras e métodos, é quem envia e recebe todas as informações para as organizar e apresentar ao usuário;
- *Database*: É a base de dados que se encontra armazenada no servidor;
- *ProcessadorDeDatabase*: Responsável por manipular a base de dados de acordo com as solicitações do *ScriptVenda*;
- *App*: Contém as regras e métodos para selecionar qual pergunta deve ser feita ao usuário;
- *CriaGUI*: Cria todas as informações visuais para interagir com o usuário, criando os ícones, menus de ajuda, perguntas e resultados.

4.3.2 Conjunto de atualização

O conjunto de atualização é responsável por realizar a leitura de um arquivo externo, o converter para a estrutura de classes desenvolvida e atualizar o servidor com o novo banco de dados. O arquivo externo deve estar em um formato especial de linhas e colunas, abordado na próxima seção.

As classes desenvolvidas para conjunto estão apresentadas na Figura 4.3.

Figura 4.3 – Diagrama de classes do conjunto de atualização.



Fonte: Elaborado pelo Autor.

- *ProcessadorDeDatabase*: Responsável por fazer o trabalho de leitura e conversão do arquivo externo criando todas as instâncias necessárias para preencher o banco de dados;
- *Database*: É a classe central da base de dados, ela armazena todas as instâncias das outras classes do conjunto;
- *Especialidade*: Contém os elementos que caracterizam uma especialidade;
- *Pergunta*: Contém o nome de um atributo e o texto da pergunta associada a ele;
- *Explicação*: Contém o nome de um atributo e o texto da explicação associada a ele;
- *Dados*: Contém uma lista com todas as instâncias da classe *Linha*;
- *Linha*: Contém uma lista com todas as instâncias da classe *Coluna*;
- *Coluna*: Contém o nome do atributo que está associada e uma lista de classes *Valor*;
- *Valor*: Possui um valor referente ao seu par linha/coluna;

De maneira resumida, as linhas representam os produtos e as colunas os atributos, sendo que cada coluna está relacionada a um atributo específico.

Deste modo, para representar um valor qualquer da base de dados deve existir uma Instância *Database*, que contém uma instância *Dados*, que contém uma instância *Linha*, que contém uma instância *Coluna*, que contém a instância *Valor*.

4.3.3 Estrutura Geral

A estrutura geral de operação do protótipo está apresentada na Figura 4.4.

O protótipo inicia com uma pergunta se o usuário quer atualizar o banco de dados (somente para os usuários-chave), caso a resposta seja sim ele acessa as classes do conjunto de atualização e converte um arquivo externo para o formato do banco de dados e o atualiza no servidor.

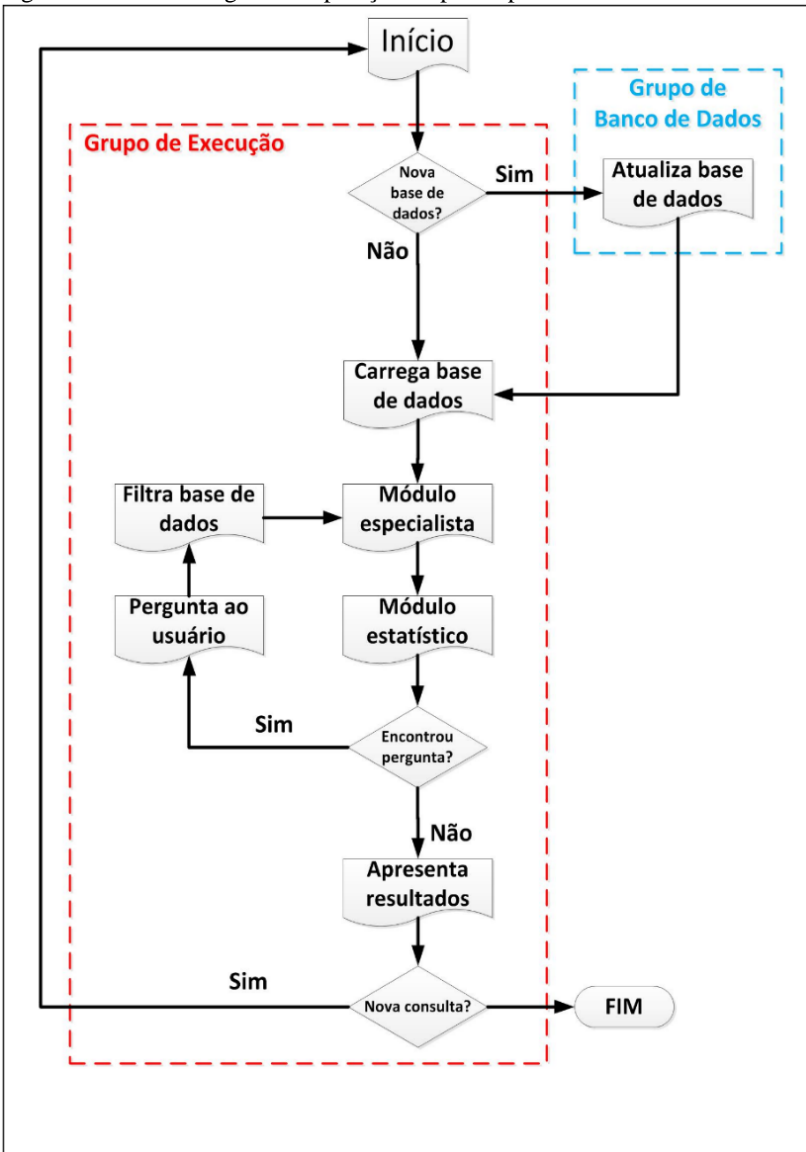
Na sequência, o banco de dados que se encontra no servidor é carregado e o protótipo executa os módulos especialista e estatístico para selecionar uma pergunta, ambos os módulos são compostos por um conjunto de funções específicas, onde:

- Módulo especialista: carrega a lista de especialidades que existe dentro do banco de dados e verifica se existe alguma condição válida para direcionar o protótipo para uma pergunta pré-definida;
- Módulo estatístico: Utiliza os cálculos de entropia para selecionar dinamicamente a próxima pergunta caso nenhuma especialidade tenha sido encontrada.

Caso uma pergunta tenha sido selecionada, o protótipo a apresenta em tela junto com suas alternativas, após a pergunta ser respondida a base de dados é filtrada de acordo com todas as respostas já informadas pelo usuário e reinicia o ciclo para encontrar a próxima pergunta.

Quando não existem mais perguntas a serem feitas, o resultado encontrado é apresentado na tela para o usuário e ele então pode optar por realizar uma nova consulta ou encerrar o protótipo e utilizar o resultado como desejar, sendo possível se dirigir diretamente ao site de e-commerce da empresa-alvo e realizar a compra.

Figura 4.4 - Estrutura geral de operação do protótipo



Fonte: Elaborado pelo Autor.

4.4 IMPLEMENTAÇÃO

A construção do código-fonte para o protótipo foi realizada através de uma sequência de etapas definida logo no início do projeto, estas etapas foram divididas de acordo com as funcionalidades previstas para a sua execução.

A utilização da ferramenta Eclipse facilitou e acelerou muito esta etapa devido a seu excelente depurador, que permite pausar, avançar e retroceder uma execução, além de permitir modificações no código-fonte e nas variáveis locais ainda em tempo de execução, sendo possível inserir manualmente valores antes mesmo de criar as funções que os gerariam, como por exemplo, um valor *perguntaSelecionada* antes de desenvolver as lógicas para selecionar uma pergunta.

A sequência de funcionalidades planejadas foi:

- I. Criação da estrutura de objetos e classes com dados inseridos manualmente;
- II. Leitura de um arquivo externo e conversão automática para as classes produzidas;
- III. Módulo estatístico utilizando o ganho de informação para selecionar as perguntas;
- IV. Criação da interface gráfica para o usuário responder as perguntas e apresentar as respostas;
- V. Módulo especialista através de uma lista de especialidades;
- VI. Adição da opção de ignorar perguntas;

A partir da finalização da etapa III o protótipo, apesar de limitado, já estava funcional, e começou a ser testado e avaliado por alguns dos especialistas da empresa-alvo, que puderam sugerir melhorias ainda durante a etapa de desenvolvimento.

4.4.1 Banco de Dados

Os dados analisados devem estar formatados para um modelo horizontal (flat file database), que consiste em uma estrutura onde cada linha representa uma nova instância de um objeto, e suas características estão divididas dentre diversas colunas (Ye, 2003).

As colunas não precisam apresentar uma estrutura fixa, as células são separadas entre si por algum caractere especial que sirva como indicador de fim de uma coluna e início da próxima, caractere este que não pode ser utilizado dentro do banco de dados para garantir que o

protótipo faça a leitura dos dados da maneira correta. Outro design possível é com um comprimento fixo para cada coluna, mas este pode gerar problemas quando se quer armazenar informações cujo número de caracteres apresente alta variação.

O protótipo desenvolvido utiliza arquivos no formato CSV (Comma-separated Value) que trabalham com o caractere especial ponto e vírgula “;” para separar uma coluna da outra, e foi optado pelo utilização de uma barra vertical “|” para separar múltiplos valores dentro de uma mesma célula. Além disso, as primeiras 4 linhas do banco de dados são para instâncias especiais, conforme descrito a seguir:

- Primeira Linha: Cada célula representa uma característica/atributo;
- Segunda Linha: Cada célula representa a pergunta que será mostrada ao usuário caso a característica seja selecionada;
- Terceira Linha: Cada célula deve conter a explicação sobre cada característica caso o usuário solicite;
- Quarta Linha: Cada célula deve informar se a coluna deve ser considerada ou não na hora dos cálculos (input/output);
- Outras Linhas: A partir da quinta, cada linha representará uma nova instância, onde cada célula é uma característica que o define, e a primeira coluna de cada uma destas linhas representa o esquema de pintura que tentará ser classificado.

A Figura 4.5 mostra o banco de dados quando aberto com o programa Excel.

Figura 4.5 - Banco de dados aberto com Excel.

Tintas	Linha de Produto	Ambiente	Espaco	Umidade	Substrato	Tipo de Tinta	Tipo de Ag	Velocidade	Cor
Insira Perg	Insira Pergunta sobre	Insira Pergun	Insira Pergu	Insira Pergunta sobre	Umidad	Insira Perg	Insira Perg	Insira Perg	Insira Perg
Insira Expli	Insira Explicação sob	Insira Explica	Insira Explic	Insira Explicação sobre	Umida	Insira Expli	Insira Expli	Insira Expli	Insira Expli
output	input	input	input	input	input	input	input	input	input
a1	Industrial	C3	Parcial	Baixa Climatizada	CIVIL 350	Galvanizac	Nao_Aplic	Nao_Aplic	Nao_Aplic
a10	Industrial	C1	Fechada	Baixa Climatizada	CIVIL 350	Galvanizac	Nao_Aplic	Nao_Aplic	Nao_Aplic
a100	Industrial	C3	Aberta	Baixa Climatizada	BOB, GAL	Liq	Nao_Aplic	Nao_Aplic	Nao_Aplic
a101	Industrial	C3	Parcial	Alta	CIVIL 350	Galvanizac	Nao_Aplic	Nao_Aplic	Nao_Aplic
a102	Industrial	C3	Fechada	Baixa Climatizada	CIVIL 350	Po	Nao_Aplic	Nao_Aplic	Nao_Aplic
a103	Industrial	C4	Fechada	Alta Baixa Climatizada	CIVIL 350	Liq	Nao_Aplic	Nao_Aplic	Nao_Aplic
a104	Industrial	C5-M-2	Parcial	Baixa	CIVIL 350	Liq	Nao_Aplic	Nao_Aplic	Nao_Aplic
a105	Industrial	C5-M-2	Parcial	Alta	CIVIL 350	Galvanizac	Nao_Aplic	Nao_Aplic	Nao_Aplic
a106	Industrial	C4	Fechada	Baixa Climatizada	CIVIL 350	Galvanizac	Nao_Aplic	Nao_Aplic	Nao_Aplic
a107	Industrial	C4	Aberta	Alta Baixa	BOB, GAL	Po	Nao_Aplic	Nao_Aplic	Nao_Aplic

Fonte: Elaborado pelo Autor.

O formato CSV é muito útil para escrita, leitura e compreensão do usuário, porém não é um formato que apresenta um grande pacote de

funções para manipular seu conteúdo e nem realizar uma conversão para o modelo de classes proposto para a orientação a objetos.

Portanto, o arquivo CSV contendo o banco de dados é convertido para o formato JSON, e este ficará salvo no servidor para ser acessado pelo conjunto de execução.

4.4.1.1 JSON

O formato JSON (JavaScript Object Notation) é muito utilizado por ser leve e rápido para a troca de dados, pode ser escrito em um simples formato de texto, logo, não depende de uma linguagem de programação específica para sua utilização, permitindo assim uma portabilidade de fácil acesso para a grande maioria das linguagens computacionais. Possui ainda uma grande vantagem por poder ser lido e escrito facilmente por seres humanos, além de ser de simples interpretação e geração para máquinas (SMITH, 2015).

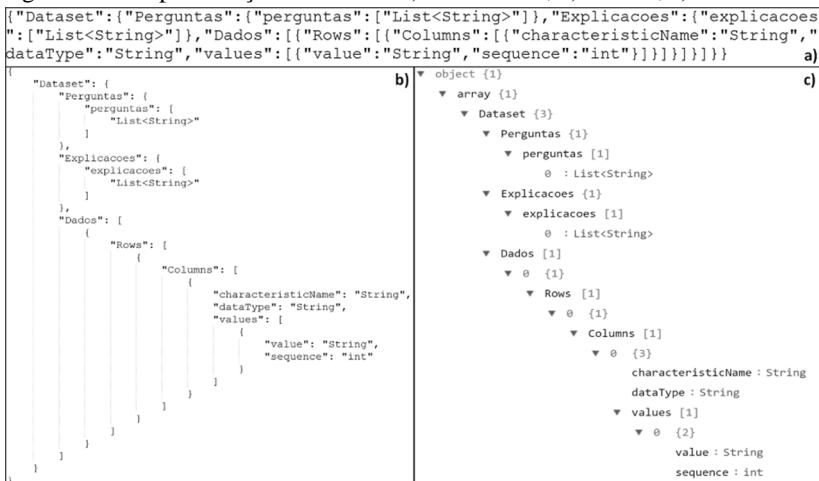
JSON é constituído basicamente de duas estruturas, que se repetem para formar o banco de dados:

- Uma coleção de pares nome/valor.
- Uma lista ordenada de valores.

A Figura 4.6 apresenta o modelo JSON em três representações distintas:

- **Minimalista:** apresenta o mínimo de caracteres possível, evitando espaços e quebras de linha, é extremamente rápido para a leitura de máquinas, mas muito ruim para um ser humano o compreender, normalmente gerada por compiladores;
- **Tradicional:** possui espaçamento razoável para facilitar a compreensão, normalmente os usuários utilizam esta representação para testes antes de converter para o modo minimalista;
- **Árvore:** normalmente é utilizada por programas para apresentar um banco de dados já carregado para o usuário.

Figura 4.6 – Representações do JSON: a) minimalista, b) normal, c) Árvore.

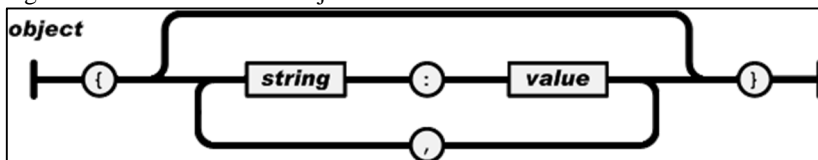


Fonte: Elaborado pelo Autor.

4.4.1.1.1 Objeto

“Um objeto é um conjunto desordenado de pares nome/valor. Um objeto começa com { (chave de abertura) e termina com } (chave de fechamento). Cada nome é seguido por : (dois pontos) e os pares nome/valor são seguidos por , (vírgula).” (JSON, 1999).

Figura 4.7 – Formato de um objeto em JSON

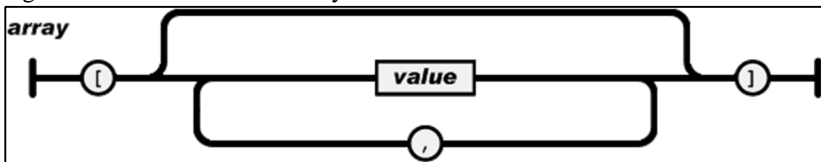


Fonte: ECMA-404, 2013

4.4.1.1.2 Array

“Uma array é uma coleção de valores ordenados. O array começa com “[(colchete de abertura) e termina com “]” (colchete de fechamento). Os valores são separados por “,” (vírgula).” (JSON, 1999).

Figura 4.8 – Formato de um array em JSON

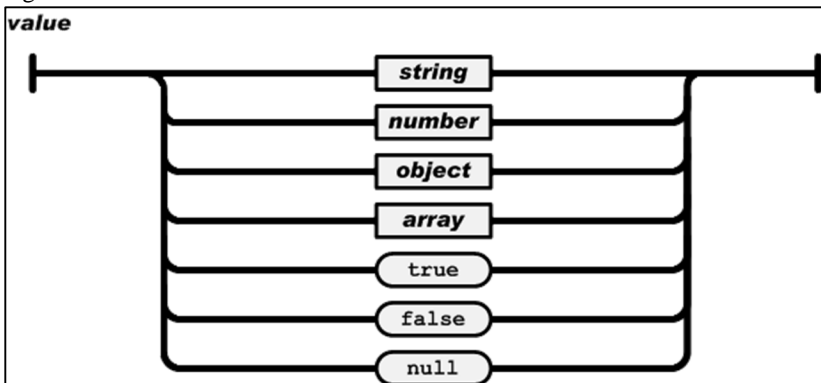


Fonte: ECMA-404, 2013

4.4.1.1.3 Valor

“Um valor [...] pode ser uma cadeia de caracteres (string), ou um número, ou true ou false, ou null, ou um objeto ou uma array. Estas estruturas podem estar aninhadas.” (JSON, 1999).

Figura 4.9 – Formato de um valor em JSON



Fonte: ECMA-404, 2013

4.4.1.2 Conversão

A conversão do banco de dados em CSV para o JSON é feita através de funções dentro da classe *ProcessadorDeDatabase*, de modo que os nomes de todos os elementos do JSON fiquem com a mesma notação de suas respectivas classes do diagrama de orientação a objetos. É possível ver isso na já apresentada Figura 4.6.

Como o JSON e as classes do grupo de atualização possuem a mesma notação, quando o JSON é carregado pelo protótipo, ele gera todas as instâncias relacionadas à classe *Database*.

4.4.2 Módulo especialista para seleção de perguntas

O módulo especialista possui uma implementação bastante simples através da classe *Especialidades*.

Da mesma forma que o banco de dados, esta classe utiliza um arquivo externo, que é convertido e carregado junto com o banco de dados e suas instâncias são inseridas dentro da classe *Especialidades* que por sua vez está alocada dentro da classe *Dataset*.

O arquivo possui apenas três colunas: *perguntaAnterior*, *respostaAnterior* e *proximaPergunta*. Sendo que a primeira linha do arquivo é o cabeçalho e as próximas linhas são instancias de especialidades, conforme apresentado no Quadro 4.1.

Quadro 4.1 – Quadro para tratamento de especialidades

Pergunta Anterior	Resposta Anterior	Próxima Pergunta
Nenhuma	Todas	Linha de Produto
Linha de Produto	Industrial	Substrato
Substrato	Civil 350 Bob. Galv	Umidade
Substrato	Não se aplica	Ambiente

Fonte: Elaborado pelo Autor.

O módulo utiliza o par pergunta/resposta da última pergunta realizada pelo protótipo e compara se este par existe em alguma das instâncias das *Especialidades*. Em caso positivo a próxima pergunta a ser selecionada será a *Próxima Pergunta* da instancia compatível e segue para o módulo estatístico, em caso negativo simplesmente segue pra o módulo estatístico.

As perguntas devem estar cadastradas com o mesmo nome que possuem no banco de dados e para permitir múltiplas respostas na mesma pergunta pode-se separar as respostas com uma barra vertical "|", pode-se utilizar a palavra "Todas" quando qualquer resposta deva direcionar para a mesma pergunta.

Este módulo é o que mais se assemelha com um tradicional SE, utilizando diretamente o conhecimento dos especialistas da empresa-alvo sobre as melhores técnicas de abordagem sobre os clientes. O mais provável é que ele seja utilizado para selecionar as perguntas iniciais da consulta, para direcionar o protótipo para a área de aplicação do usuário,

deixando então o módulo estatístico selecionar as próximas perguntas para encontrar uma solução.

4.4.3 Módulo estatístico para seleção de perguntas

Este módulo é representado pelos métodos encarregados de realizar os cálculos de entropia para selecionar a “melhor” pergunta, conforme já mencionado na seção 3.2.2.2.

4.4.3.1 Cálculos

Para realizar os cálculos de entropia o protótipo utiliza todas as instâncias presentes no banco de dados (após a aplicação dos filtros de perguntas já respondidas).

Como cada linha do banco de dados representa uma instância diferente não é possível agrupar os resultados como nos outros exemplos apresentados neste trabalho. Deste modo, ao invés de existirem “X instâncias com o plano a1”, “Y com a2” e “Z com a3” existe apenas uma instância para o plano “a1”, uma “a2”, uma a3 e assim por diante.

Sendo assim, a maneira que os cálculos de entropia são realizados pelo protótipo tem uma aparência um pouco diferente, mas seus princípios e conceitos permanecem os mesmos, de dividir os dados até que o conjunto resultante possua somente uma classe, no caso, como as classes são únicas, somente um elemento também, ou até que não existam mais perguntas para se fazer.

Para demonstrar como os cálculos são realizados, será utilizado o conjunto de dados simplificados apresentado no Quadro 4.2.

Quadro 4.2 – Conjunto de dados simplificados

Esquema	Umidade	Superfície	Tratamento
a1	Alta	Aço	Jateamento
a2	Alta	Aço	Jateamento
a3	Media	Ferro e Madeira	Lixa
a4	Media	Madeira	Lixa
a5	Baixa	Madeira	Jateamento
a6	Baixa	Madeira	Lixa

Fonte: Elaborado pelo Autor.

É possível agrupar cada um dos atributos pelas suas alternativas:

- Umidade { Alta=2, Media=2, Baixa=2};
- Superfície { Aço=2, Ferro=1, Madeira=4};

- Tratamento {Jateamento=3, Lixa=3}.

Tendo em vista que cada uma das linhas do Quadro 4.2 representa uma instância diferente, não existem conjuntos de respostas que podem ser agrupados, portanto em todos os cálculos a frequência para as classes será 1.

Para se iniciar os cálculos deve-se encontrar a entropia inicial do conjunto de dados, conforme Equação 3-1.

$$H(S) = - \sum_{k=1}^6 \frac{1}{6} * \log_2 \left(\frac{1}{6} \right) = -6 * \left(\frac{1}{6} * \log_2 \left(\frac{1}{6} \right) \right) = 2.58 \text{ bits}$$

Como todas as frequências são iguais, o somatório pode ser substituído por uma simples multiplicação pelo número de classes diferentes.

Nota-se que o valor da entropia (2.58 bits) é muito maior que o valor encontrado nos exemplos do capítulo anterior, isso se dá ao fato de não ser possível agrupar os esquemas de pintura, resultando no somatório de frequências 1.

O cálculo para a entropia após a divisão merece uma atenção especial, uma das respostas do atributo superfície possui 2 alternativas, o que resulta em um conjunto de 7 características associadas a 6 respostas, alterando o número total da fração ponderada, realizando o cálculo com a equação (3-3, tem-se:

$$\begin{aligned} H(S, sup) &= \left(-\frac{|T_{aco}|}{T} * H(T_{aco}) \right) + \left(-\frac{|T_{ferro}|}{|T|} * H(T_{ferro}) \right) + \left(-\frac{|T_{mad}|}{|T|} * H(T_{mad}) \right) \\ H(S, sup) &= \left(-\frac{2}{7} * \sum_{j=1}^2 \left(\frac{1}{6} * \log_2 \frac{1}{6} \right) \right) + \left(-\frac{1}{7} * \sum_{j=1}^1 \left(\frac{1}{6} * \log_2 \frac{1}{6} \right) \right) + \left(-\frac{4}{7} * \sum_{j=1}^4 \left(\frac{1}{6} * \log_2 \frac{1}{6} \right) \right) \\ H(S, sup) &= \left(-\frac{2}{7} * 2 * \left(\frac{1}{6} * \log_2 \frac{1}{6} \right) \right) + \left(-\frac{1}{7} * 1 * \left(\frac{1}{6} * \log_2 \frac{1}{6} \right) \right) + \left(-\frac{4}{7} * 4 * \left(\frac{1}{6} * \log_2 \frac{1}{6} \right) \right) \\ H(S, sup) &= 1,292 \text{ bits} \end{aligned}$$

Para os outros atributos o número total para a fração ponderada se mantém de acordo com seu conjunto.

$$\begin{aligned} H(S, trat) &= \left(-\frac{3}{6} * 3 * \left(\frac{1}{6} * \log_2 \frac{1}{6} \right) \right) + \left(-\frac{3}{6} * 3 * \left(\frac{1}{6} * \log_2 \frac{1}{6} \right) \right) = 1,292 \text{ bits} \\ H(S, umi) &= \left(-\frac{2}{6} * 2 * \left(\frac{1}{6} * \log_2 \frac{1}{6} \right) \right) + \left(-\frac{2}{6} * 2 * \left(\frac{1}{6} * \log_2 \frac{1}{6} \right) \right) + \left(-\frac{2}{6} * 2 * \left(\frac{1}{6} * \log_2 \frac{1}{6} \right) \right) \\ &= 0,861 \text{ bits} \end{aligned}$$

O cálculo informação de divisão também envolve a fração ponderada do número de atributos conforme equação (3-4):

$$info\ split_{(sup)}S = \left(-\frac{2}{7} * \log_2 \frac{2}{7}\right) + \left(-\frac{1}{7} * \log_2 \frac{1}{7}\right) + \left(-\frac{4}{7} * \log_2 \frac{4}{7}\right) = -1.387\ bits$$

$$info\ split_{(trat)}S = -\left(\frac{3}{6} * \log_2 \frac{3}{6}\right) * \left(\frac{3}{6} * \log_2 \frac{3}{6}\right) = -1.000\ bits$$

$$info\ split_{(umid)}S = \left(-\frac{2}{6} * \log_2 \frac{2}{6}\right) + \left(-\frac{2}{6} * \log_2 \frac{2}{6}\right) + \left(-\frac{2}{6} * \log_2 \frac{2}{6}\right) = -1.548\ bits$$

O cálculo para a taxa de ganho não necessita de nenhuma modificação, seguindo conforme equação (3-5):

$$Gain\ ratio(S, sup) = \frac{2.58 - 1.292}{1.387} = 0.94\ bits$$

$$Gain\ ratio(S, trat) = \frac{2.58 - 1.292}{1.000} = 1.292\ bits$$

$$Gain\ ratio(S, umid) = \frac{2.58 - 0.86}{1.584} = 1.09\ bits$$

O ganho de informação para a superfície e o tipo de tratamento é igual, porém com a taxa de ganho se percebe que o melhor atributo é o tipo de tratamento, pois este apresenta uma ramificação menor para os resultados, deste modo o primeiro atributo que o módulo especialista selecionaria para o Quadro 4.2 é o tipo de tratamento.

O protótipo realiza esses cálculos para todos os atributos cada vez que busca uma nova pergunta, a Figura 4.10 foi extraída diretamente do depurador do protótipo através de um *print screen* e apresenta os resultados destes cálculos durante uma execução.

Figura 4.10 - Cálculos de entropia do protótipo durante sua execução

UTILIZACAO	{3, 206, 210} - 419
UTILIZACAO	Entropia: 4.294 Ganho de info: 4.417 Informação de Divisão: 1.054
UTILIZACAO	Razão de Ganho: 4.19
IMERSAO	{416, 3} - 419
IMERSAO	Entropia: 8.587 Ganho de info: 0.124 Informação de Divisão: 0.061
IMERSAO	Razão de Ganho: 2.02
INTEMPERISMO	{208, 208, 3} - 419
INTEMPERISMO	Entropia: 4.294 Ganho de info: 4.417 Informação de Divisão: 1.054
INTEMPERISMO	Razão de Ganho: 4.19
SISTEMA	{301, 118} - 419
SISTEMA	Entropia: 5.186 Ganho de info: 3.525 Informação de Divisão: 0.858
SISTEMA	Razão de Ganho: 4.11

Fonte: Elaborado pelo Autor.

A Figura 4.10 apresenta o resultado dos cálculos de entropia para 4 atributos: Utilização, Imersão, Intemperismo e Sistema de Pintura. O número 419 representa o total de elementos que ainda existem na base, e entre chaves o número de ocorrências de cada uma das alternativas do atributo e apresenta o valor da entropia, ganho de informação, informação da divisão e razão de ganho para cada atributo.

Neste caso, o atributo selecionado seria utilização ou intemperismo com uma razão de ganho de 4.19, o protótipo trabalha com variáveis do tipo *double* que podem possuir até 16 dígitos, mas que são arredondados quando impressos na tela para evitar poluição visual.

4.4.3.2 Algoritmo

Para realizar os cálculos do ganho de informação, o protótipo faz uso de duas estruturas conhecidas como *HashMap*, que são listas relacionais desordenadas compostas por conjuntos pares-valor.

As duas estruturas armazenam os pares <Atributo, Respostas>, porém uma delas repete as respostas iguais para se ter acesso ao número total de ocorrências, enquanto a outra não repete para se ter acesso ao número de respostas diferentes, sendo preenchidas desta forma:

mapAllValues = Superfície: [Aço, Aço, Ferro, Aço, Ferro, Aço],
Sistema: [Pó, Liq, Pó, Liq, Liq, Pó, Pó, Liq]

mapSingleValues = Superfície: [Aço, Ferro],
Sistema: [Pó, Liq]

Tendo acesso a estas duas estruturas o protótipo consegue extrair todas as informações sobre o número de ocorrência de cada alternativa para cada atributo e realizar os cálculos para decidir qual será a próxima pergunta.

O código completo para realizar os cálculos de entropia encontra-se no apêndice A, mas um resumo no formato de pseudocódigo está descrito a seguir:

1. Cria as variáveis *mapAllValues* e *mapSingleValues*, e as preenche com as informações das classes de dados;
2. Seleciona um novo atributo do *mapSingleValues* que possua pelo menos dois valores associados e segue para 3, se não tiver mais nenhum segue para 4;

3. Realiza todos os cálculos de entropia com as informações do atributo selecionado e volta para 2;
4. Seleciona o atributo com a maior razão de ganho e o transforma na próxima pergunta.

Seguindo estes passos, o módulo estatístico consegue de maneira simples realizar seus cálculos e selecionar a melhor pergunta para fazer para o usuário.

Existem passos intermediários para validar se o atributo já foi ignorado ou respondido para evitar o evitar o gasto desnecessário de processamento (atributos já respondidos possuem razão de ganho 0 e atributos ignorados serão descartados no fim de qualquer modo).

4.4.4 Interação com o Usuário

A etapa de interação com o usuário é relativamente simples, a classe *CriaGUI* recebe o nome do atributo e as respostas que devem ser apresentadas em tela, o banco de dados é utilizado para coletar as outras informações do atributo como o texto para a pergunta e as explicações.

A tela é criada com um código HTML pré-definido onde as informações da pergunta atual são inseridas e substituem as informações antigas, gerando uma tela conforme apresentado na Figura 4.11.

Figura 4.11 - Tela de pergunta do protótipo



Fonte: Elaborado pelo Autor.

Em relação às alternativas apresentadas ao usuário, a classe *CriaGUI* utiliza somente a informação de quais alternativas permanecem no banco de dados após a aplicação do filtro com as respostas já

fornecidas pelo usuário. Deste modo, garante-se que o protótipo nunca gere uma situação do tipo “Nenhum resultado encontrado”, pois as alternativas que conduziriam a esta resposta nunca serão apresentadas ao usuário, conforme representado na Figura 4.12.

Figura 4.12- Exemplo de pergunta com e sem alternativas ocultas

The figure consists of two panels, a) and b), illustrating a question interface. Both panels ask: "Qual o substrato a ser pintado?" (Which substrate to be painted?).

Panel a) shows six radio button options:

- Aço Carbono
- Galvanizado Eletrolítico
- Galvanizado por Imersão a Quente
- Metais não Ferrosos
- Tinta envelhecida firmemente aderida
- Tinta envelhecida não aderida

Panel b) shows only two radio button options:

- Aço Carbono
- Galvanizado por Imersão a Quente

Both panels include a button labeled "Ignorar Pergunta" (Ignore Question) at the bottom.

Fonte: Elaborado pelo Autor.

A Figura 4.12 apresenta a mesma pergunta realizada em dois cenários distintos, sendo que em a) existem seis alternativas para o usuário selecionar e em b) existem apenas duas, garantindo assim que o usuário nunca chegue em resultados finais “nulos”.

Conforme o banco de dados do protótipo for sendo ampliado e novos cenários explorados, esta propriedade indiretamente deixará de ser utilizada.

Quando terminam as perguntas, o resultado com o menor custo (verificado no site de e-commerce) que restou no banco de dados é selecionado e apresentado em tela para o usuário em um formato semelhante ao das perguntas e um botão para imprimir relatórios é disponibilizado. Os resultados contêm *hyperlinks* para que o usuário possa seguir direto ao website de e-commerce da empresa-alvo, conforme apresentado na Figura 4.13.

Figura 4.13 - Tela de respostas do protótipo

NOVA CONSULTA		IMPRIMIR
Histórico de Respostas		Plano de Pintura Proposto
Pergunta	Resposta	
Qual Sistema de Pintura?	Tinta Pó	<ul style="list-style-type: none"> 1ª demão → POLITHERM 54 HB - 1 camada de 120µm Quantidade Recomendada: 7.3kg de Tinta 2ª demão → POLITHERM 54 HB - 1 camada de 120µm Quantidade Recomendada: 7.3kg de Tinta
Há contato com Sol e Chuva?	Estruturas ficarão abrigadas	
Qual a segmentação/utilização das estruturas?	Industrial	
Qual tipo de indústria?	Alimentícia	
Qual a distância da orla marítima?	Acima de 30 km	
Qual o substrato a ser pintado?	Aço Carbono	
Qual o tratamento de superfície?	Jato abrasivo padrão Sa 2 1/2	
Area	40m²	

Fonte: Elaborado pelo Autor.

Na Figura 4.13, é possível ver que a primeira parte do texto de cada demão de tinta proposta possui um *hyperlink* representado através de uma cor diferente e da adição do sublinhado.

Este capítulo apresentou as principais etapas de criação do protótipo, demonstrando como as informações dos capítulos anteriores foram utilizadas para desenvolver um algoritmo capaz de executar a configuração do produto como um sistema de venda guiada e mantendo o conceito de que a base de dados permaneça isolada e possa ser facilmente modificada sem exigir a necessidade de se reescrever o código-fonte.

O próximo capítulo aborda as duas etapas finais do desenvolvimento: verificação e validação, onde serão apresentados os métodos com que o protótipo foi testado e aprovado tanto pelo EC quanto pelos representantes da empresa-alvo.

5. VERIFICAÇÃO E VALIDAÇÃO

Neste capítulo são descritas as ações realizadas de modo a testar o protótipo, confirmando que a maneira com que ele sistematiza o conhecimento extraído e formalizado dos especialistas e o transfere para o usuário estão condizentes com os requisitos definidos.

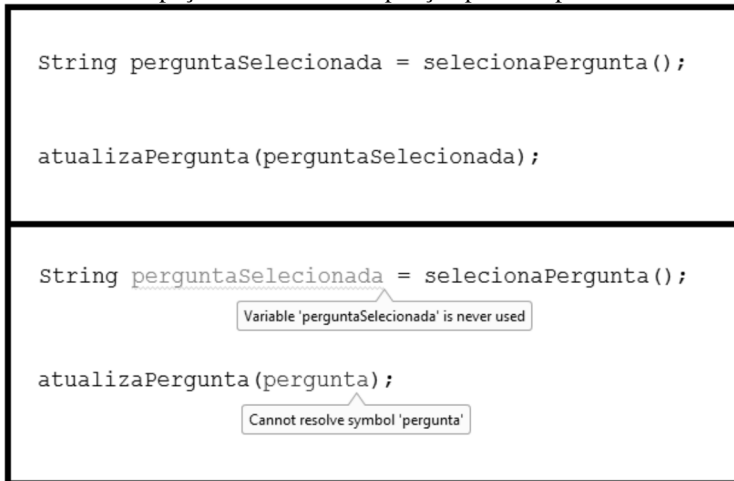
Também são avaliadas se as principais ações (perguntas) que os especialistas (vendedores experientes) realizam durante uma conversa com um potencial cliente são semelhantes às do protótipo e por fim se os produtos que o protótipo recomenda são os mesmos que eles recomendariam.

5.1 VERIFICAÇÃO

Por seguir o modelo incremental, a verificação do protótipo foi realizada constantemente durante sua criação pelo EC. A cada nova regra adicionada ao modelo, testes eram realizados com a ajuda do depurador interno do Eclipse.

O Eclipse realiza verificações em tempo real para garantir a integridade das regras e antecipar erros de compilação, logo, não houve problemas desta natureza durante o desenvolvimento. A Figura 5.1 demonstra como estes erros são antecipados enquanto o código ainda é escrito.

Figura 5.1 - Antecipação de erros de compilação pelo Eclipse



Fonte: Elaborado pelo Autor.

Na parte superior da Figura 5.1 é possível verificar que um atributo *perguntaSelecionada* é criado e utilizado corretamente em uma função, já na parte inferior, o mesmo atributo é criado, porém ocorre um erro de digitação tenta-se utilizar um atributo com nome semelhante: *pergunta*, mas que não existe dentro do código, resultando em uma mensagem instantânea de erro.

Foram encontrados alguns erros gerados pelo fato do EC não possuir domínio da linguagem de programação escolhida, estes envolveram principalmente o fato de que listas e variáveis são manipuladas de diferentes maneiras de uma linguagem para outra, gerando problemas como “utilizar n vezes o primeiro valor de uma lista ao invés de utilizar uma vez cada um dos n seus valores”, ou então realizar modificações em uma cópia de um atributo ao invés de modificar o atributo em si.

Estes erros foram mais frequentes durante a etapa de verificação das funções referentes aos cálculos de entropia. Para esta etapa foi criado um modelo através de uma planilha excel com funções internas para realizar os cálculos de entropia, o Quadro 5.1 representa esta planilha.

Quadro 5.1 - Planilha de cálculos de entropia

	Substrato		Orla Marítima	
Ocorrências	Total	620	Total	620
	Aço	300	Menor que 10 km	199
	Ferro Fundido	300	Maior que 30 km	242
	Não se Aplica	20	Entre 10 e 30 km	179
			Não se Aplica	0
	H(S)	9,276	H(S)	9,276
	H(S,T)	4,353	H(S,T)	3,142
	info(S,T)	4,923	info(S,T)	6,134
	info split _(T) S	1,173	info split _(T) S	1,573
	Gain ratio(S,T)	4,196	Gain ratio(S,T)	3,898

Fonte: Elaborado pelo Autor.

A planilha foi criada de modo que na parte superior se inseria o número de ocorrências das alternativas de um atributo e funções realizavam os cálculos e apresentava seus resultados na parte inferior.

A verificação dos cálculos normalmente envolvia criar um cenário, o executar dentro do sistema e comparar os resultados com o mesmo cenário feito na planilha, e no caso de divergências o depurador era então utilizado para identificar a fonte do erro.

Pelo fato do EC estar alocado no setor de Sistemas de Engenharia da empresa-alvo foi possível solicitar para dois programadores experientes realizarem uma verificação paralela com ênfase especial em performance, o que resultou em modificações na maneira como algumas variáveis são acessadas e carregadas, deixando o protótipo ainda mais ágil mas sem modificar sua lógica.

Para verificar a capacidade do protótipo processar a base de dados adequadamente, deve-se ter em mente que cada linha da base de dados representa um cenário distinto, de modo que cada uma de suas colunas pode ser considerada como as entradas que o definem, com exceção da primeira coluna que representa sua saída.

Como o protótipo foi criado baseado na necessidade da equipe adicionar novos cenários, com liberdade para criar perguntas e alternativas diferentes das já existentes na base de dados, a verificação dos cenários da maneira tradicional, avaliando se o sistema realiza todas as inferências programadas, não seria o suficiente.

Para não gerar uma dependência para a empresa-alvo de serviços futuros do EC, se fez necessário realizar uma verificação sólida de modo a garantir que ao adicionar dados à base, seja uma nova linha (cenário) que utilize as perguntas e alternativas já existentes ou adicionar/remover

uma pergunta ou modificar alguma alternativa no meio da base, em todos estes cenários de mudanças o protótipo deve ser capaz de processar os novos dados adequadamente.

Vale ressaltar que a inserção/remoção de uma pergunta (coluna) da base não é uma atividade trivial, pois esta ação gera uma modificação em todos os cenários da base. A remoção pode causar um problema menor, onde alguns cenários passem a se sobrepor, mas a adição de uma nova pergunta gera a necessidade desta ser avaliada e preenchida em todas as linhas (cenários) da base.

Deste modo, foi realizado um extenso trabalho de verificação dos módulos especialista e estatístico, responsáveis por gerenciar o encadeamento instantâneo das perguntas e respostas, garantindo que todos os dados de entradas e saídas existentes na base de dados sejam contemplados.

Esta verificação foi feita alterando o tamanho da base de dados, incluindo e removendo perguntas, repetindo cenários e inserindo entradas que ocorriam em um cenário e depois eram replicadas em outros. Em todos estes testes o protótipo realizou o adequado gerenciamento.

Uma das técnicas utilizadas para verificar se o protótipo conseguiria carregar e processar todas as entradas foi a de executar realizando uma análise de todas as perguntas e alternativas por ele indicadas, mas não respondendo nenhuma, através do botão de “ignorar perguntas”, que faz com que o protótipo adicione a pergunta para uma lista especial de perguntas ignoradas e execute novamente seus módulos para selecionar uma próxima pergunta (que não esteja na lista das ignoradas).

A Figura 5.2 e Figura 5.3 apresenta *print screens* de uma das execuções desta verificação das perguntas e respostas.

Figura 5.2 –Verificação do encadeamento de perguntas.

Histórico de Respostas	
Pergunta	Resposta
Qual Sistema de Pintura?	IGNORADA
Há contato com Sol e Chuva?	IGNORADA
Qual a segmentação/utilização das estruturas?	IGNORADA
Qual o Estado da obra?	IGNORADA
Qual a distancia da orla marítima?	IGNORADA
Qual o substrato a ser pintado?	IGNORADA
Qual o tratamento de superfície?	IGNORADA
Possui classificação de ambiente (ISO 12944) definida?	IGNORADA
Qual tipo de indústria?	IGNORADA
As estruturas ficarão imersas?	IGNORADA

Fonte: Elaborado pelo Autor.

Figura 5.3 - Exemplo de verificação das respostas para as perguntas

<p>Pergunta Corrente</p> <p>Possui classificação de ambiente (ISO 12944) definida?</p> <p><input type="radio"/> C2</p> <p><input type="radio"/> C3</p> <p><input type="radio"/> C4</p> <p><input type="radio"/> C5-I</p> <p><input type="radio"/> C5-M</p> <p><input type="radio"/> Não se aplica</p> <p><input type="button" value="Ignorar Pergunta"/></p>	<p>Pergunta Corrente</p> <p>Qual a distancia da orla marítima?</p> <p><input type="radio"/> Acima de 30 km</p> <p><input type="radio"/> Entre 10 e 30 km</p> <p><input type="radio"/> Menor que 10 km</p> <p><input type="radio"/> Não se aplica</p> <p><input type="button" value="Ignorar Pergunta"/></p>	<p>Pergunta Corrente</p> <p>Qual a segmentação/utilização das estruturas?</p> <p><input type="radio"/> Industrial</p> <p><input type="radio"/> Não se aplica</p> <p><input type="radio"/> outros (não é indústria de produção)</p> <p><input type="button" value="Ignorar Pergunta"/></p>
--	---	---

Fonte: Elaborado pelo Autor.

A Figura 5.2 apresenta o histórico de execução, mostrando as perguntas sendo encadeadas, enquanto a Figura 5.3 apresenta três exemplos das alternativas exibidas durante uma execução.

Este tipo de análise serviu como base para garantir que o protótipo realmente utiliza todos os dados da base e consegue encadear qualquer cenário nela presente.

Já para avaliar se todas as saídas poderiam ser alcançadas, foi utilizada uma função especial do depurador para antecipar a execução do método para “apresentar os resultados” de uma execução antes mesmo de responder a qualquer pergunta, forçando assim o protótipo a enviar como resposta todas as saídas possíveis, que em todos os testes coincidiu com as bases utilizadas.

Como etapa final para a verificação deste encadeamento automático, a base de dados foi completamente removida e substituída por uma nova, fornecida pela divisão de motores da empresa-alvo, esta base de dados continha 60 colunas (perguntas), 30000 linhas (cenários de entrada) e 400 saídas distintas. O elevado número de perguntas se deu pelo fato de muitas possuírem respostas do tipo “sim ou não”.

Dezenas de cenários foram testados nesta base de motores e o protótipo foi capaz de realizar o encadeamento de todos corretamente, confirmando assim a robustez do gerenciamento de entradas e saídas da base de dados independentemente de seu conteúdo.

Por fim, a base de dados desenvolvida e validada pelos especialistas da empresa foi inserida no protótipo. Ela contemplava cerca de 500 cenários distintos e 10 perguntas. O conjunto de perguntas e respostas desta base é apresentado na sequência:

- **Estado da Obra:** obra nova, reparo, não se aplica;
- **Tipo de Substrato:** aço carbono, galvanizado eletrolítico, galvanizado por imersão a quente, metais não ferrosos, tinta envelhecida não aderida, tinta envelhecida firmemente aderida, não se aplica;
- **Contato com Ambiente:** enterradas ou submersas, abrigadas, desabrigadas, não se aplica;
- **Distância Orla:** acima de 30 km, entre 10 e 30 km, menor que 10 km, navegação, não se aplica;
- **Tratamento de Superfície:** jateamento abrasivo comercial (brush off), jato abrasivo padrão Sa 2½, jateamento abrasivo padrão Sa 3, limpeza e lixamento, manual mecânico St 3, não se aplica;
- **Segmentação:** industrial, não se aplica;
- **Tipo de Imersão:** LM1, LM2, LM3, não se aplica;
- **Classificação ISO 12944:** C2, C3, C4, C5-M, C5-I, não se aplica;
- **Tipo de Indústria:** alimentícia, fertilizantes, metalúrgica, papel e celulose, química, não se aplica;

- **Sistema de Pintura:** tinta líquida, tinta pó, não se aplica.

Esta base foi utilizada para formalizar a conclusão do projeto e entregar o protótipo para a empresa, encerrando assim o papel do EC em seu desenvolvimento, então foi necessário confirmar que todas as verificações realizadas anteriormente continuaram válidas.

Para esta verificação final, cerca de uma centena de cenários foram realizados e inseridos em uma apresentação para o departamento de vendas, solicitante do projeto. Todos estes cenários foram executados corretamente e o projeto foi então aprovado.

5.1.1 Teste do protótipo

As imagens deste tópico apresentam o código de depuração durante algumas etapas de uma execução completa do protótipo partindo de uma base de dados com 10 perguntas e 419 registros.

As três figuras: Figura 5.4, Figura 5.5 e Figura 5.6 seguem o mesmo padrão: O lado esquerdo apresentando a pergunta selecionada, o lado direito apresenta a tabela do depurador e abaixo o resultado do depurador junto com sua ação. A tabela está organizada em blocos, onde cada bloco representa uma pergunta e suas informações estão divididas em três linhas: a primeira informa a frequência de cada uma das respostas que a pergunta possui, enquanto a segunda e terceira linhas apresentam o resultado dos cálculos de entropia.

Figura 5.4 – Log de depuração do protótipo

Pergunta Corrente

Qual Sistema de Pintura?

Tinta Líquida

Tinta Pó

UTILIZACAO	{3, 206, 210}
UTILIZACAO	Entropia: 4.294 Ganho de Info: 4.417 Informação de Divisão: 1.054
UTILIZACAO	Razão de Ganho: 4.19
IMERSAO	{416, 3}
IMERSAO	Entropia: 8.587 Ganho de Info: 0.124 Informação de Divisão: 0.061
IMERSAO	Razão de Ganho: 2.02
INTEMPERISMO	{208, 208, 3}
INTEMPERISMO	Entropia: 4.294 Ganho de Info: 4.417 Informação de Divisão: 1.054
INTEMPERISMO	Razão de Ganho: 4.19
SISTEMA_PINTURA	{301, 118}
SISTEMA_PINTURA	Entropia: 5.186 Ganho de Info: 3.525 Informação de Divisão: 0.858
SISTEMA_PINTURA	Razão de Ganho: 4.11
ESTADO_OBRA	{347, 72}
ESTADO_OBRA	Entropia: 6.232 Ganho de Info: 2.479 Informação de Divisão: 0.662
ESTADO_OBRÁ	Razão de Ganho: 3.746
SUBSTRATO	{168, 72, 149, 6, 12, 12}
SUBSTRATO	Entropia: 2.775 Ganho de Info: 5.936 Informação de Divisão: 1.877
SUBSTRATO	Razão de Ganho: 3.162
TIPO_INDUSTRIA	{209, 60, 60, 30, 30, 30}
TIPO_INDUSTRIA	Entropia: 2.659 Ganho de Info: 6.052 Informação de Divisão: 2.121
TIPO_INDUSTRIA	Razão de Ganho: 2.854
TRATAMENTO_SUPERFICIE	{97, 69, 69, 172, 12}
TRATAMENTO_SUPERFICIE	Entropia: 2.414 Ganho de Info: 6.296 Informação de Divisão: 2.02
TRATAMENTO_SUPERFICIE	Razão de Ganho: 3.117
CLASSIFICACAO_AMBIENTE	{81, 38, 38, 32, 210, 20}
CLASSIFICACAO_AMBIENTE	Entropia: 2.728 Ganho de Info: 5.983 Informação de Divisão: 2.079
CLASSIFICACAO_AMBIENTE	Razão de Ganho: 2.878
DISTANCIA_ORLA	{3, 164, 136, 116}
DISTANCIA_ORLA	Entropia: 2.92 Ganho de Info: 5.79 Informação de Divisão: 1.621
DISTANCIA_ORLA	Razão de Ganho: 3.573

Pergunta selecionada: INTEMPERISMO com razão de ganho igual a 4.1901
Exeção Localizada, modificando pergunta para: SISTEMA_PINTURA

Fonte: Elaborado pelo Autor.

O protótipo realiza os cálculos de entropia para os 10 atributos e seleciona “intemperismo”, pois este apresenta a maior razão de ganho do conjunto. Mas como o protótipo encontra uma especialidade, a pergunta é então alterada para “sistema pintura”.

Em seguida, a pergunta é respondida com a alternativa “tinta pó”, reiniciando o processo para localizar a próxima pergunta, conforme apresentado na Figura 5.5.

Figura 5.5 - Log de depuração do protótipo

	UTILIZACAO	[58, 60]
	UTILIZACAO	Entropia: 3.442 Ganho de Info: 3.44 Informação de Divisão: 1
	UTILIZACAO	Razão de Ganho: 3.441
	IMERSAO	[118]
	IMERSAO	JA RESPONDIDA OU FILTRADA
	INTEMPERISMO	[59, 59]
	INTEMPERISMO	Entropia: 3.441 Ganho de Info: 3.441 Informação de Divisão: 1
	INTEMPERISMO	Razão de Ganho: 3.441
Pergunta Corrente	SISTEMA_PINTURA	[118]
Há contato com Sol e Chuva?	SISTEMA_PINTURA	JA RESPONDIDA OU FILTRADA
<input type="radio"/> Estruturas ficarão abrigadas	ESTADO_OBRA	[118]
<input type="radio"/> Estruturas ficarão desabrigadas (com incidência de raios UV)	ESTADO_OBRA	JA RESPONDIDA OU FILTRADA
<input type="button" value="Ignorar Pergunta"/>	SUBSTRATO	[48, 22, 48]
	SUBSTRATO	Entropia: 2.517 Ganho de Info: 4.366 Informação de Divisão: 1.508
	SUBSTRATO	Razão de Ganho: 2.896
	TIPO_INDUSTRIA	[58, 30, 30]
	TIPO_INDUSTRIA	Entropia: 2.553 Ganho de Info: 4.33 Informação de Divisão: 1.508
	TIPO_INDUSTRIA	Razão de Ganho: 2.871
	TRATAMENTO_SUPERFICIE	[30, 22, 22, 44]
	TRATAMENTO_SUPERFICIE	Entropia: 1.88 Ganho de Info: 5.002 Informação de Divisão: 1.937
	TRATAMENTO_SUPERFICIE	Razão de Ganho: 2.583
	CLASSIFICACAO_AMBIENTE	[18, 18, 12, 60, 10]
	CLASSIFICACAO_AMBIENTE	Entropia: 2.22 Ganho de Info: 4.662 Informação de Divisão: 1.961
	CLASSIFICACAO_AMBIENTE	Razão de Ganho: 2.378
	DISTANCIA_ORLA	[50, 38, 30]
	DISTANCIA_ORLA	Entropia: 2.394 Ganho de Info: 4.488 Informação de Divisão: 1.554
	DISTANCIA_ORLA	Razão de Ganho: 2.889

Pergunta selecionada: INTEMPERISMO com razão de ganho igual a 3.441

Fonte: Elaborado pelo Autor.

O protótipo aplica um filtro na base de dados com as perguntas já respondidas anteriormente, para realizar os novos cálculos. É possível perceber pela imagem que mesmo só tendo respondido a pergunta “Sistema Pintura”, as perguntas “Imersão” e “Estado Obra” são filtradas no processo e, portanto, foram excluídas dos cálculos.

As perguntas foram respondidas até o fim da execução do protótipo, a Figura 5.6 apresenta log de depuração após a resposta da última pergunta.

Figura 5.6 - Log de depuração do protótipo

	UTILIZACAO	[1]
	UTILIZACAO	JA RESPONDIDA OU FILTRADA
	IMERSAO	[1]
	IMERSAO	JA RESPONDIDA OU FILTRADA
	INTEMPERISMO	[1]
	INTEMPERISMO	JA RESPONDIDA OU FILTRADA
Plano de Pintura Proposto	SISTEMA_PINTURA	[1]
	SISTEMA_PINTURA	JA RESPONDIDA OU FILTRADA
• 1ª demão -- POLITHERM 54 HB - 1 camada de 120µm Quantidade Recomendada: 9.2kg de Tinta	ESTADO_OBRA	[1]
	ESTADO_OBRA	JA RESPONDIDA OU FILTRADA
• 2ª demão -- POLITHERM 54 HB - 1 camada de 120µm Quantidade Recomendada: 9.2kg de Tinta	SUBSTRATO	[1]
	SUBSTRATO	JA RESPONDIDA OU FILTRADA
	TIPO_INDUSTRIA	[1]
	TIPO_INDUSTRIA	JA RESPONDIDA OU FILTRADA
	TRATAMENTO_SUPERFICIE	[1]
	TRATAMENTO_SUPERFICIE	JA RESPONDIDA OU FILTRADA
	CLASSIFICACAO_AMBIENTE	[1]
	CLASSIFICACAO_AMBIENTE	JA RESPONDIDA OU FILTRADA
	DISTANCIA_ORLA	[1]
	DISTANCIA_ORLA	JA RESPONDIDA OU FILTRADA
Pergunta Seleccionada: Fim das Perguntas		
Plano de Pintura Localizado: 1007738556		

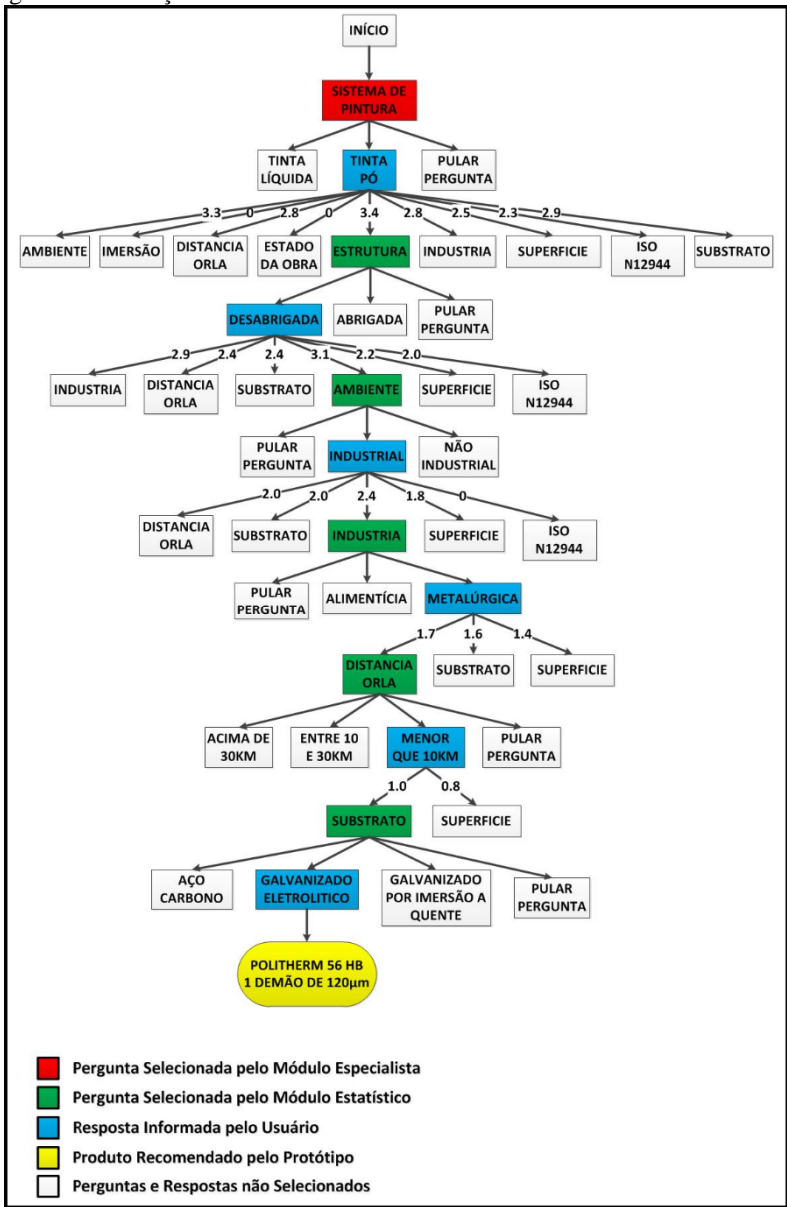
Fonte: Elaborado pelo Autor.

Conforme a Figura 5.6, todas as perguntas possuem somente um resultado, o que caracteriza o fim da execução da etapa de interação com o usuário, o protótipo então utiliza todas as respostas obtidas para filtrar a base de dados e propor uma solução na forma de um esquema de pintura.

Neste caso, o esquema de pintura proposto foi o: 1 demão de POLITHERM 54HB 120µm + 1 demão de POLITHERM 54HB 120µm”. Dependendo do conjunto de respostas é possível que mais de um esquema de pintura seja localizado, o protótipo então avalia qual o mais barato para indicar.

Pode-se dizer que a cada nova execução do protótipo é gerada de maneira indireta uma árvore de decisão de um ramo só, a Figura 5.7 apresenta o processo de criação da árvore.

Figura 5.7 – Criação dinâmica da árvore de decisões.

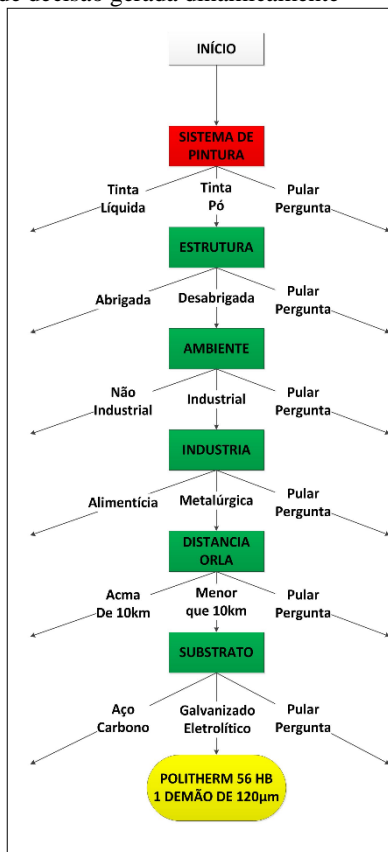


Fonte: Elaborado pelo Autor.

Todos blocos em vermelho ou verde são as perguntas seleccionadas pelo protótipo, em azul são respostas do usuário e por fim em amarelo o produto proposto, os blocos em branco são as perguntas ou respostas não seleccionadas. A resposta de uma pergunta é ligada às próximas possíveis perguntas com setas indicando a razão de ganho para cada uma, e as perguntas são ligadas as suas respostas por setas simples.

Para demonstrar a árvore de acordo com a notação usual de perguntas como nós e as respostas como ramos, a Figura 5.8 apresenta como seria esta árvore já concluída.

Figura 5.8 - Árvore de decisão gerada dinamicamente



Fonte: Elaborado pelo Autor.

Como o protótipo utiliza somente as informações fornecidas pelo usuário para gerar uma resposta instantânea, a árvore de decisão resulta em apenas um ramo preenchido e os outros permanecem incompletos. Em uma próxima execução, todas as informações são removidas e uma nova árvore será gerada.

5.2 VALIDAÇÃO

A etapa de validação consiste basicamente em confirmar se as informações que o protótipo retorna para o usuário estão condizentes com a realidade.

Esta etapa é de suma importância para o sucesso do projeto e para garantir que não ocorram erros capazes de gerar danos à imagem da empresa, pois como o protótipo será utilizado tanto para treinamento de colaboradores quanto para a recomendação direta de produtos para clientes, erros em um único cenário pode gerar problemas de natureza mais branda como a recomendação de um produto mais resistente (e possivelmente mais caro) que o necessário podendo acarretar na perda de uma venda, ou, em casos mais graves, ofertar um produto inadequado ou ineficiente para as necessidades do cliente podendo evoluir até um eventual processo.

De modo simplificado, pode-se dizer que o protótipo localiza um resultado utilizando todas as respostas informadas pelo usuário durante a execução, na forma de filtros no banco de dados, selecionando então o resultado mais barato dentre os que restaram (atendem aos requisitos).

O banco de dados inicial foi construído inteiramente por um EH da empresa-alvo que atua no segmento de “pintura de estruturas metálicas” com o auxílio do EC que através de entrevistas e reuniões demonstrou como o banco deveria ser estruturado e o que poderia ou não poderia ser inserido nele. Deste modo, os únicos erros esperados desta etapa seriam principalmente devido a lacunas no conhecimento do EH.

Uma das estratégias utilizadas para a criação de alguns dos cenários foi “selecionar” um esquema de pintura e então pensar em quais cenários o esquema poderia ser aplicado, deste modo, vários cenários podem direcionar para o mesmo esquema.

Deste modo, quando o protótipo estava com quase 500 cenários cadastrados foi considerado que o banco de dados estava com tamanho suficiente para uma avaliação “externa”, para isso, mais dois especialistas da empresa-alvo foram selecionados para a validação, esta gerou algumas discussões sobre cenários específicos onde cada especialista defendia a

utilização de um esquema de pintura diferente, de acordo com os esquemas que estava mais acostumado a ofertar aos seus clientes.

Estas discussões serviram para reforçar uma das vantagens de se utilizar SE que é a formalização do conhecimento, pois para o mesmo cenário é possível existir diversos esquemas de pintura distintos e que consigam atender aos “requisitos mínimos”, mas sempre irá existir um que será mais adequado, e uma arguição entre especialistas é uma excelente maneira para fazer esta avaliação e chegar a um consenso de qual realmente aparenta ser o melhor em cada cenário.

Depois do banco de dados estar completamente validado, os especialistas testaram a base junto com o protótipo para garantir que este estava utilizando as informações da base de modo preciso.

Em seguida o protótipo foi adicionado no servidor e o link para acessar o protótipo foi inserido no website da empresa para poder ser acessado por qualquer usuário com os acessos exigidos pela empresa-alvo.

Os três EHs que participaram do projeto encaminharam e-mails com o link do protótipo e instruções sobre a sua utilização com uma solicitação de avaliação técnica e prática do mesmo. Foram incluídas nestes e-mails 137 pessoas, destas, 35 eram especialistas da equipe de vendas técnicas, 50 representantes comerciais e 42 formuladores dos laboratórios de desenvolvimento de produtos.

Cerca de 15 pessoas fizeram questionamentos sobre esquemas diferentes em determinados cenários, mas somente um dos cenários questionados se julgou verdadeiro e foi atualizado no servidor. Esta atualização foi realizada por um dos EH encarregado, comprovando que a estrutura do banco atendeu aos requisitos da empresa-alvo de que os próprios especialistas pudessem realizar as atualizações.

Foram recebidas diversas sugestões para adicionar novas perguntas a base de dados, em sua maioria direcionadas para os segmentos de atuação de quem fez a sugestão. Muitas destas foram guardadas para o momento que os dados do protótipo forem expandidos e passem a englobar estes novos segmentos.

Houve também algumas sugestões para modificações no layout para dispor melhor as informações em tela, modificar o local de alguns ícones e algumas frases de ajuda que estavam incompletas, a maioria destas necessitava de alterações no código-fonte e foram realizadas pelo EC.

Um *script* foi utilizado para monitorar, contabilizar o número de acessos e gerar relatórios para apresentar aos gerentes da empresa-alvo mostrando se a ferramenta realmente estava sendo utilizada e justificar

custos de mão-de-obra dos especialistas para a ampliação do banco de dados.

O *script* fez registros por pouco mais de dois meses e verificou que o protótipo foi acessado quase 2000 vezes por cerca de 400 usuários diferentes, o que apontou que o link para o protótipo foi repassado para outros usuários verificarem. Alguns destes usuários “externos” se interessaram pela maneira que o protótipo operava e solicitaram avaliações para a criação de sistemas semelhantes para outros grupos de negócio da empresa-alvo.

A gerência da empresa-alvo aprovou o protótipo e seis especialistas foram designados para montar novos bancos de dados englobando diferentes segmentos e conforme cada um for concluído e devidamente validado, será incorporado aos dados já existentes no servidor..

5.3 TELAS DO PROTÓTIPO

O protótipo é inteiramente executado em apenas uma tela que pode ser acessada através de um link que está incorporado ao sistema da empresa-alvo EASY (Engineering And Sales sYstem). O usuário deve possuir um login de acesso com privilégios de colaborador, e futuramente será liberado também o acesso para clientes, a posição definitiva em que o link estará dentro do sistema ainda está em análise pelo departamento de marketing.

A tela do protótipo é dividida em cinco entidades: uma barra de funcionalidades na parte superior, outra na parte inferior e no centro existem as colunas: histórico de respostas, pergunta corrente e ajuda.

- A barra superior possui os botões “novo”, “salvar” e “carregar”, estes servem para iniciar uma nova consulta, salvar e carregar uma consulta salva anteriormente.
- A barra inferior apresenta a opção para troca de idioma da aplicação para português do Brasil, inglês e espanhol.
- Histórico de respostas: Mostra as perguntas já apresentadas para o usuário, tanto as respondidas quanto as ignoradas, e cada pergunta possui botões para repetir ou anular;
- Pergunta corrente: Apresenta a pergunta atual, suas respostas e um botão para o usuário ignorar a pergunta;
- Ajuda: A parte de explicações do sistema, contém informações adicionais sobre cada uma das etapas do protótipo para melhorar a compreensão do usuário sobre o sistema;

A Figura 5.9 mostra uma vista geral da tela.

Figura 5.9 - Vista geral da tela do protótipo

NOVO SALVAR CARREGAR MAESTRO ID: 0

Histórico de Respostas ID: 0

Pergunta	Resposta

Pergunta Corrente

Qual o tipo de substrato que a tinta será aplicada?

Aço Carbono
 Galvanizado por imersão a quente

Ignorar Pergunta

Ajuda

Ajuda Tipo

Um dos principais fatores para o sucesso de uma tinta é que esta consiga aderir firmemente a superfície em que for aplicada.

Superfícies de aço carbono possuem uma boa aderência, mas devem ser livradas de sua carepa de laminação antes da aplicação das tintas.

Superfícies de aço galvanizado não possuem tantos problemas com corrosão, mas a aderência de tintas em sua superfície é mais complexa, necessitando de tintas especiais.

WTI_VENDA_GUIADA->WTI_BOTAO

Idioma

Fonte: Elaborado pelo Autor.

5.3.1 Interação com o usuário

O protótipo sempre inicia encadeando a primeira pergunta, não sendo necessária nenhuma ação além de se entrar no link para que este se inicie.

Todos cenários testados com o protótipo apresentaram tempo de resposta inferior a 1 segundo, levando mais tempo para a internet do usuário atualizar a página do que para a aplicação executar todos cálculos e verificações de especialidades, deste modo optou-se pelo protótipo realizar automaticamente todas as suas funções sem o uso de nenhum comando do usuário além das respostas para as perguntas.

O protótipo utiliza as classes Perguntas e Explicações para preencher o texto da pergunta e o texto das explicações, e os apresenta junto com as alternativas de maneira direta na tela, conforme pode ser visto na Figura 5.10.














Figura 5.10 – Tela de pergunta e explicação do protótipo

Pergunta Corrente	Ajuda
<p>Qual o tipo de substrato que a tinta será aplicada?</p> <p><input type="radio"/> Aço Carbono</p> <p><input type="radio"/> Galvanizado por imersão a quente</p> <p></p>	<p>Ajuda Tipo</p> <p>Um dos principais fatores para o sucesso de uma tinta é que esta consiga aderir firmemente a superfície em que for aplicada.</p> <p>Superfícies de aço carbono possuem uma boa aderência, mas devem ser livradas de sua carepa de laminação antes da aplicação das tintas.</p> <p>Superfícies de aço galvanizado não possuem tantos problemas com corrosão, mas a aderência de tintas em sua superfície é mais complexa, necessitando de tintas especiais.</p>

Fonte: Elaborado pelo Autor.

As perguntas já respondidas pelo usuário ficam no menu do lado esquerdo, nele as perguntas anteriores podem ser conferidas, repetidas ou removidas, conforme Figura 5.11.

Figura 5.11 – Menu para remover ou repetir perguntas

NOVO SALVAR CARREGAR MAESTRO ID: 0													
Histórico de Respostas ID: 0	Pergunta Corrente												
<table border="1"> <thead> <tr> <th>Pergunta</th> <th>Resposta</th> <th></th> </tr> </thead> <tbody> <tr> <td>Qual o tipo de substrato que a tinta será aplicada?</td> <td>Aço Carbono</td> <td> </td> </tr> <tr> <td>Qual a durabilidade exigida para o esquema de pintura?</td> <td>Alta (acima de 15 anos)</td> <td> </td> </tr> <tr> <td>Qual o tipo de ambiente ao redor</td> <td>IGNORADA</td> <td></td> </tr> </tbody> </table>	Pergunta	Resposta		Qual o tipo de substrato que a tinta será aplicada?	Aço Carbono	 	Qual a durabilidade exigida para o esquema de pintura?	Alta (acima de 15 anos)	 	Qual o tipo de ambiente ao redor	IGNORADA		<p>Qual a classificação (norma ISO 12944)</p> <p><input type="radio"/> C1</p> <p><input type="radio"/> C3</p> <p><input type="radio"/> C4</p> <p><input type="radio"/> C5-I</p> <p><input type="radio"/> C5-M</p> <p><input type="radio"/> Im2</p> <p></p>
Pergunta	Resposta												
Qual o tipo de substrato que a tinta será aplicada?	Aço Carbono	 											
Qual a durabilidade exigida para o esquema de pintura?	Alta (acima de 15 anos)	 											
Qual o tipo de ambiente ao redor	IGNORADA												

Fonte: Elaborado pelo Autor.

Quando o usuário solicita para responder uma pergunta novamente, o protótipo utiliza os filtros restantes no banco de dados e apresenta apenas as alternativas que restaram, podendo até mesmo apresentar somente uma alternativa (a mesma que ele já havia selecionado anteriormente) caso muitas perguntas já tenham sido respondidas, este é

o único meio de fazer o sistema apresentar uma pergunta com somente uma opção.

Caso o usuário solicite a remoção de uma pergunta, ela é meramente retirada da lista de perguntas respondidas e o protótipo executa novamente a verificação dos módulos especialista e estatístico, podendo eventualmente repetir a pergunta removida.

Ao fim da execução, caso o usuário tenha ignorado perguntas, o protótipo pede se ele deseja tentar as responder novamente ou se quer ver os resultados com a consulta incompleta, conforme Figura 5.12.

Figura 5.12 – Fim da execução com perguntas puladas

Histórico de Respostas ID: 0		Pergunta Corrente
Pergunta	Resposta	Você pulou perguntas importantes para uma recomendação mais precisa do melhor esquema de pintura para suas necessidades, o que deseja fazer? <input type="button" value="Ver resultados"/> <input type="button" value="Responder"/>
Qual o tipo de substrato que a tinta será aplicada?	Aço Carbono	
Qual a durabilidade exigida para o esquema de pintura?	Alta (acima de 15 anos)	
Qual o tipo de ambiente ao redor	IGNORADA	
Qual a classificação (norma ISO 12944)	IGNORADA	

Fonte: Elaborado pelo Autor.

Caso o usuário deseje tentar responder as perguntas ignoradas, o protótipo as remove da lista de perguntas já respondidas e executa novamente os módulos especialista e estatístico, sendo possível que as perguntas ignoradas não sejam mais relevantes por conta das outras já respondidas.

Seria possível fazer o protótipo avaliar se as perguntas ignoradas ainda são relevantes ou não antes de apresentar a opção para o usuário as responder novamente, porém, para fins de registro é importante ter o histórico completo de ações do usuário.

Por fim, quando não existem mais atributos para serem perguntados, o protótipo questiona o usuário sobre o tamanho ou peso da obra que ele deseja pintar, conforme apresentado na Figura 5.13.

Figura 5.13 - Tela de pergunta sobre o tamanho da aplicação

Histórico de Respostas ID: 0		Pergunta Corrente	Ajudá
Pergunta	Resposta	Encontramos o esquema de pintura ideal para suas necessidades.	Você pulou questões durante a execução da venda guiada.
Qual o tipo de substrato que a tinta será aplicada?	Aço Carbono	Caso conheça a área estimada de aplicação, informe no campo abaixo!	Caso realmente deseje ignorar estas perguntas, selecione a primeira opção para obter um esquema não tão preciso para necessidade.
Qual a durabilidade exigida para o esquema de pintura?	Alta (acima de 15 anos)	Área total a ser pintada(m²): 0	Caso queira que a venda guiada apresente novamente as questões puladas e ainda relevantes a suas necessidades, selecione a segunda opção.
Qual a classificação (norma ISO 12544)	C4	<input type="button" value="Confirmar"/> <input type="button" value="X"/>	
Qual o tipo de ambiente ao redor	IGNORADA	Caso conheça o tipo e peso de sua obra, informe no campo abaixo!	
		<input type="radio"/> Obra Pesada <input type="radio"/> Obra Leve	
		Peso da obra (kg): 0	
		<input type="button" value="Confirmar"/> <input type="button" value="X"/>	


Fonte: Elaborado pelo Autor.

5.3.2 Apresentação dos Resultados

Por fim, quando todas informações são respondidas pelo usuário, o esquema de pintura mais adequado ao cenário especificado pelo usuário é apresentado no painel central, no mesmo local onde as perguntas são feitas.

Os esquemas de pintura sempre são apresentados com a mesma grafia: Um texto “Nome da tinta + Número de Camadas + Espessura da camada + Quantidade recomendada”, conforme apresentado na Figura 5.14.

Figura 5.14 - Tela de resultados

NOVO SALVAR CARREGAR MAESTRO ID: 0	
Histórico de Respostas ID: 0	
Pergunta	Resposta
Qual o tipo de substrato que a tinta será aplicada?	Aço Carbono
Qual a durabilidade exigida para o esquema de pintura?	Alta (acima de 15 anos)
Qual a classificação (norma ISO 12944)	C4
Qual o tipo de ambiente ao redor	IGNORADA
<p>Area - 500m²</p> <ul style="list-style-type: none"> • WEGPOXI ZSP 31 - 1 camada de 60µm - Quantidade Recomendada: 73.9L de Tinta. • WEGGLACK CVD 12 - 1 camada de 180µm - Quantidade Recomendada: 257.1L de Tinta. 	
	

Pergunta Corrente

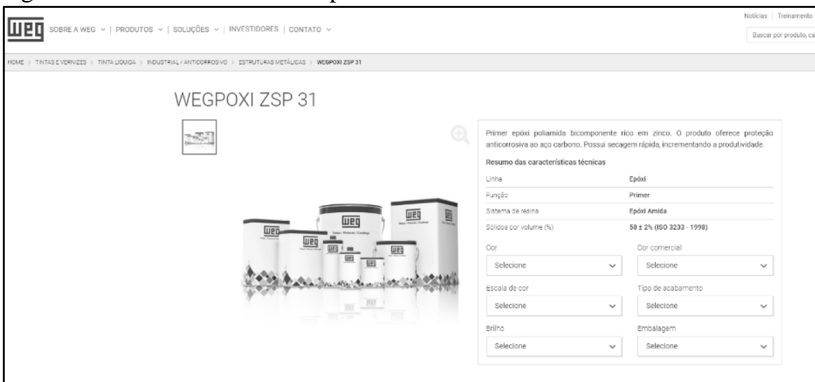
- [WEGPOXI ZSP 31 - 1 camada de 60µm - Quantidade Recomendada: 73.9L de Tinta.](#)
- [WEGGLACK CVD 12 - 1 camada de 180µm - Quantidade Recomendada: 257.1L de Tinta.](#)

Fonte: Elaborado pelo Autor.

O esquema apresentado deve ser composto de pelo menos uma camada de uma tinta, não existindo um limite nem para o número de tintas nem o de camadas.

O nome de cada tinta está ligado através de um *hyperlink* que abrirá uma nova aba no navegador direcionando para a página de cadastro da tinta no EASY, para que o usuário possa realizar o pedido de compra se assim o desejar, a página de cadastro de uma tinta é apresentada na Figura 5.15.

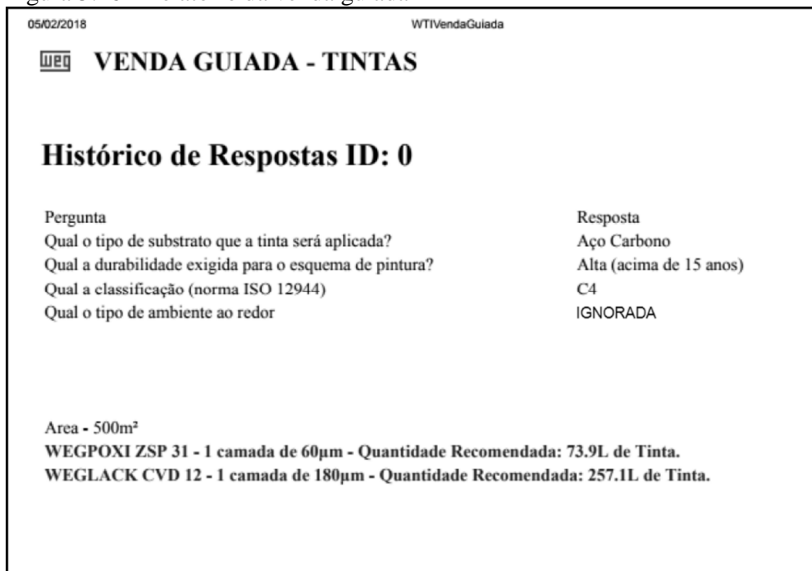
Figura 5.15 - Redirecionamento para site de e-commerce WEG



Fonte: Elaborado pelo Autor.

No menu esquerdo, existe um botão “imprimir” que gera um relatório das perguntas e respostas utilizadas para que tanto o usuário quanto a empresa possam ter um registro da consulta e de qual esquema foi recomendado, o modelo deste relatório está apresentado na Figura 5.16.

Figura 5.16 - Relatório da venda guiada



Fonte: Elaborado pelo Autor.

Este capítulo apresentou as últimas etapas do desenvolvimento deste protótipo, evidenciando os testes realizados pelo engenheiro de conhecimento e os pelos especialistas da empresa para confirmar que as informações contidas nele são verdadeiras.

Foram apresentados alguns números sobre a abrangência atual e a aprovação do protótipo pelas gerências, o que garantiu um novo projeto entre os especialistas para expandir o banco de dados explorando novos segmentos.

Imagens do protótipo em funcionamento foram inseridas para apresentar ao leitor a aparência final desenvolvida e apresentar todas as ações que os usuários do protótipo podem realizar durante uma consulta.

6. CONCLUSÕES

Este capítulo apresenta as principais conclusões acerca deste projeto. A primeira seção inclui comentários sobre o desenvolvimento do projeto e sobre o protótipo em si, em seguida algumas contribuições do trabalho são apresentadas e sugestões para trabalhos futuros visando continuidade a este projeto.

6.1 CONSIDERAÇÕES E COMENTÁRIOS FINAIS

O protótipo foi desenvolvido em parceria com a empresa WEG através de um estágio realizado junto ao departamento de sistemas de engenharia.

Devido ao fato de ser uma empresa de grande porte e atuar em muitos segmentos, o seu objetivo foi criar um SE dinâmico de modo a ser possível aplicar para outras unidades de negócio sem grandes dificuldades, o que gerou certo trabalho e tempo de pesquisa para encontrar uma alternativa viável em relação aos SE tradicionais.

O encadeamento de perguntas de forma dinâmica aliado à possibilidade de se utilizar especialidades quando necessário foi o principal fator para atender a demanda da empresa-alvo, possibilitando que o protótipo seja executado com “qualquer” base de dados, desde que formatada adequadamente.

Durante a etapa de verificação, uma base contendo mais de 30000 cenários para configuração de motores foi testada e apresentada para a empresa-alvo de modo a provar o dinamismo do sistema em qualquer tipo de base de dados.

A técnica de árvore de decisão é muito mais ampla do que a apresentada neste trabalho, possui diversos algoritmos e modos para previsão de resultados, porém, como o protótipo está ligado ao site de e-commerce da empresa, é necessário garantir que todos os dados estejam validados, bloqueando todas as possibilidades do protótipo realizar suposições e preencher automaticamente dados incompletos.

Foram realizados dois treinamentos com membros diferentes da equipe de vendas da empresa-alvo voltados à manutenção e ampliação do banco de dados e das regras internas do protótipo, com a criação de uma apostila e um manual sobre sua utilização. E uma apresentação com a equipe inteira sobre a utilização do protótipo e comentários sobre seu impacto na rotina de trabalho.

Vale ressaltar algumas propriedades do SE criado:

- Sistema de Venda Guiada independente da base de dados;
- Ordem das perguntas dinâmicas, de acordo com as respostas do usuário;
- Perguntas podem ser ignoradas;
- Somente alternativas válidas aparecem para o usuário;
- Perguntas cujas alternativas já foram filtradas pelas outras respostas não são apresentadas;
- Consulta pode ser salva dentro do sistema e gerado um ID ou então impressas em formulário próprio;
- Suporte a línguas estrangeiras: Inglês e Espanhol (sendo ainda possível inserir outros idiomas se necessário);
- Menu de explicações para todas as perguntas;
- Recomendação de quantidade a ser adquirida de cada componente;
- Ligação direta com o site de e-commerce da WEG.

6.2 CONTRIBUIÇÕES

Uma das maiores contribuições deste trabalho é o fato de se iniciar na empresa uma cultura para a organização sistemática de todos os seus produtos, disseminando e armazenando o conhecimento dos vendedores e evitando maiores perdas no caso de funcionários chave se desligarem dela.

Conforme novos produtos forem lançados, estes deverão ser avaliados para a criação formal de seus possíveis esquemas de pintura e adicionados ao sistema, evitando que somente os produtos mais conhecidos sejam ofertados aos clientes.

Todos os produtos estarão padronizados e avaliados conforme suas especificações, diminuindo as chances de erro por parte dos vendedores ofertando produtos incompatíveis com as necessidades do cliente e evitando que estes erros causem prejuízos para a imagem da empresa.

6.3 APRENDIZADOS

As principais dificuldades encontradas durante a execução deste trabalho ocorreram principalmente devido às mudanças de escopo e algumas divergências de opinião entre os colaboradores envolvidos diretamente no projeto.

A escolha de duas linguagens de programação e de um software que o autor desconhecia gerou certo desconforto no início do projeto, mas

conforme este foi se desenvolvendo e com o suporte da equipe do departamento, a escrita do código foi se tornando cada vez mais natural.

A escolha do tema e o início do estágio ocorreram próximos ao final do prazo regular para a defesa da dissertação, o que acarretou em um pensamento negativo de que mesmo com eventuais prorrogações, o prazo não seria o suficiente. No início, este foi um fator de preocupação, mas conforme o protótipo se desenvolvia com velocidade superior à esperada, este se tornou um fator motivador.

6.4 RECOMENDAÇÕES PARA TRABALHOS FUTUROS

Nenhum trabalho é tão bom que não possa ser melhorado, eis algumas sugestões para dar continuidade a evolução deste projeto:

- Desenvolver um novo módulo especialista para perguntas ou respostas especiais e que precisem sofrer algum tipo de tratamento para serem empregadas no banco de dados, como por exemplo características cuja resposta combine com outras para gerar algum número, ou então respostas numéricas que devam ser convertidas de acordo com algum parâmetro que dependa de outras respostas;
- Desenvolver um algoritmo para registrar a ocorrência do número de vezes que cada pergunta foi ignorada e utilizar esta informação para manipular o módulo estatístico e reduzir a ocorrência de perguntas normalmente ignoradas pelos usuários;
- Testar o algoritmo de seleção de perguntas em bases de dados mais complexas;

BIBLIOGRAFIA

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. ABNT NBR 15156: **Pintura industrial - Terminologia**. Rio de Janeiro, 2004.

ABREU, F. E. L. **Biocidas Anti-Incrustantes de Terceira Geração em Sedimentos Superficiais da Costa Brasileira**. Dissertação (Mestrado) em Oceanografia Física, Química e Geológica da Universidade Federal do Rio Grande, Rio Grande, 2016.

ALMEIDA, E. et. Al. **Marine Paints**: The particular case of antifouling paints. *Progress in Organic Coatings* vol. 59, 2–20, 2007, 19p.

ALVES, Guilherme Dionízio. **Sistema Especialista Protótipo para Diagnóstico de Falhas em um Sistema Hidráulico Naval**. Dissertação (Mestrado) em Engenharia Mecânica na Universidade Federal de Santa Catarina, Florianópolis, 2001.

ANGHINETTI, I. C. B. **Tintas, Suas Propriedades e Aplicações Imobiliárias**. Monografia (Especialização) em Construção Civil na Universidade Federal de Minas Gerais, Belo Horizonte, 2012.

Aragão, J. F. **Sistemas Especialistas Como Ferramenta Auxiliar Para o Ensino da Disciplina Bases da Técnica Cirúrgica**. Dissertação (Mestrado) em Informática da Universidade Federal de Campina Grande, Campina Grande, 2002.

ARRUDA, E. A. **Estudo Comparativo do Processo Corrosivo do Aço Patinável e do Aço Carbono Comum**. TCC (Graduação) em Engenharia Mecânica do Instituto de Tecnologia da Universidade Federal do Pará, Belém, 2009

ARTERO, A. O. **Inteligência Artificial Teórica e Prática**. 1. ed. São Paulo: Livraria da Física, 2009.

BARRETO, J. M. **Inteligência Artificial no limiar do Século XXI**. 3. ed. Florianópolis: Editora do Autor, 2001.

BECHTOLD, Micheline. **Síntese e Caracterização de Nanopartículas de Prata e Aplicação Como Agente Biocida em Tinta Poliuretânica Hidrossolúvel**. Dissertação (Mestrado) em Engenharia Química na Universidade Federal de Santa Catarina, Florianópolis, 2011.

BEN, Smith. **Beginning JSON**. Berkeley: Apress, 2015. ISBN 978-1-4842-0202-9

BERRY, M. J. A.; LINOFF, G. **Data Mining Techniques**: for marketing, sales and customer support. USA: Wiley Computer Publishing, 1997.

Bittencourt, G. **Inteligência Artificial**: Ferramentas e Teorias. 3. ed. Florianópolis: UFSC, 2006.

BRAMER, Max. **Principles of Data Mining**. London: Springer, 2016. ISBN 978-1-4471-7307-6.

CAMARGO, M. **Resinas Poliésteres Carboxifuncionais para Tinta em Pó: Caracterização e Estudo Cinético da Reação de Cura**. Tese (Doutorado) em Engenharia de Minas, Metalúrgica e de Materiais da Universidade Federal do Rio Grande do Sul, Porto Alegre, 2002.

CICEK, Volkan. **Corrosion Engineering and Cathodic Protection Handbook**: With an Extensive Question and Answer Section. John Wiley & Sons, 2017.

Haleck, Leonardo. Corrosão Filiforme. CONPLEQ Consultoria, 2016. Disponível em: <<http://www.conpleq.com.br/conpleq-consultoria-corrosao-filiforme>>. Acesso em 20 de Outubro de 2017.

COPPIN, B. **Inteligência Artificial**. Tradução Valério, Jorge Duarte Pires. Rio de Janeiro: LTC, 2013.

Dicio - Dicionário Online de Portugues. Disponível em <<https://www.dicio.com.br/>>. Acesso em 23 de agosto de 2017.

International Organization for Standardization. ISO 8044: **Corrosion of Metals and Alloys** - Basic terms and definitions. 2015

DONADIO, P. A. **Manual Básico sobre tintas**. Águia Química. Ponta Grossa, 2011. 15 p.

DURKIN, J. **Expers Systems Design and Development**. 1. ed. New York: Macmillan, 1994.

ECMA-404. **The JSON Data Interchange Format**. ECMA International, Geneva, 2013, Disponível em <<http://www.ecma-international.org/publications/files/ECMA-ST/ECMA-404.pdf>>. Acesso em 03 de Agosto de 2017.

ELAYAPERUMAL, K.; RAJA, V. S. **Corrosion Failures: Theory, Case Studies, and Solutions**. John Wiley & Sons, 2015.

Elokence. **Akinator, o Gênio da Internet**. 2007, disponível em <<http://pt.akinator.com>>. Acesso em 10 de de Agosto de de 2017

ENDURES - A TNO Initiative. **Antifouling And Protective Coating**. Disponível em <<https://www.endures.co.uk/antifouling-and-protective-coatings/>>. Acesso em 20 de Outubro de 2017.

EYSENCK, H. J. **Intelligence: a new look**. 1. ed. New Jersey: Transaction Publishers, 1998.

FEIO, R. F. G. B. **Microcápsulas com função biocida para o controlo da bioincrustação**. Dissertação (Mestrado) em Engenharia Química no Instituto Superior Técnico Lisboa, Lisboa, 2015.

GENTIL, Vicente. **Corrosão**. 6. ed. Rio de Janeiro: LTC, 2011.

GIARRATANO, J.; RILEY, G. **Expert Systems: Principles and Programming**. 2. ed. Boston: PWS Publishing Company, 1994.

GONZALEZ, A. J.; DANKEL, D. D. **The Engineering of KnowledgeBased Systems - Theory and Practice**. Prentice-Hall, Inc., 1993.

Hunt, Earl B. et. Al. **Experiments in induction**. Michigan: Academic Press, 1966.

International Organization for Standardization. ISO 12944: **Paints and varnishes** - Corrosion protection of steel structures by protective paint systems. 2015.

ECLIPSE Foundation. **IDE and tools**. Disponível em <<https://www.eclipse.org/ide/>>. Acesso em 10 de Outubro de 2017.

JSON. **Introdução ao JSON**. 1999, disponível em <<http://www.json.org/index.html>>. Acesso em 03 de de Agosto de de 2017.

KNUDSEN, O. O.; FORSGREN, A. **Corrosion Control Through Organic Coatings**. 2. ed. Boca Raton: CRC Press, 2017. ISBN 978-1-4987-6072-0

KOTRIKLA, A. **Enviromental management aspects for TBT anti-fouling wastes from the shipyards**. Journal of Environmental Management vol. 90 77-85, 2009, 8p.

KRÄNKEL, F. **Pintura Industrial com Tintas Líquidas**. Manual de Treinamento, WEG Tintas, Guaramirim, 2015.

LAHIRI, Amiya Kumar. **Applied Metallurgy and Corrosion Control**. Singapore: Springer, 2017.

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

LAPEC - Laboratório de Pesquisa em Corrosão. **Corrosão e Fadiga**. Disponível em <http://www.ufrgs.br/lapec/wa_files/corros_c3_a3o-fadiga.pdf>. Acesso em 20 de Outubro de 2017.

Lazzari, Luciano. **Engineering Tools for Corrosion: Design and Diagnosis**. United Kingdom: Woodhead Publishing, 2017.

LUGER, G. F. **Inteligência Artificial: Estruturas e Estratégias para a Solução de Problemas Complexos**. Tradução Engel, Paulo Martins. 4. ed. Porto Alegre: Bookman, 2004.

MAKHLOUF, Abdel Salam Hamdy. **Handbook of Smart Coatings for Materials Protection**. vol 64. Elsevier, 2014.

MANNARI, V.; PATEL, C. J. **Understanding Coatings Raw Materials**. Hanover: Vincentz Network, 2015. ISBN 978-3-86630-603-5

MATELLI, José Alexandre. **Sistemas Baseados em Conhecimento para Projeto de Plantas de Cogeração a Gás Natural**. Tese (doutorado) em Engenharia Mecânica na Universidade Federal de Santa Catarina, Florianópolis, 2008.

MATOS, L. A. C. **Estudo da Corrosão de Ligas Ferrosas em Diesel e Biodiesel**. Dissertação (Mestrado) na Universidade Estadual do Centro-Oeste, Guarapuava, 2013.

MATTEUCCI, A. J. **How to choose a correct biocide system for paints**. 11° Congress International of Paints - ABRAFATI, 2009.

McCauley, Ronald A. **Corrosion of Ceramic Materials**. 3. ed. Boca Raton: Crc Press, 2013.

MCCORDUCK, P.; FEIGENBAUM, E. A.; NII, P. H. **The Rise of the Expert Company: How Visionary Businesses are Using Intelligent Computers**. 1. ed. New York: Times Book, 1988.

McCUE, Colleen. **Data Mining and Predictive Analysis: Intelligence Gathering and Crime Analysis**. Butterworth-Heinemann, 2015. ISBN 978-0-12-800229-2.

MECABÔ, L. **Desenvolvimento de um protótipo de sistema especialista para apoio à manutenção de turbocompressores centrífugos de gás natural**. Dissertação (Mestrado) em Engenharia Mecânica, Universidade Federal de Santa Catarina, Florianópolis, 2007.

Milella, Pietro Paolo. **Fatigue and corrosion in metals**. Milan: Springer, 2013. ISBN 978-88-470-2336-9.

MIRANDA, E. F. **Comportamento quanto à corrosão de aços inoxidáveis na produção do biodiesel**. Dissertação (Mestrado) em Engenharia em Processos Químicos e Bioquímicos na Escola de Engenharia de Mauá, São Caetano do Sul, SC, 2009.

MITRA, S.; ACHARYA, T. **Data Mining: Multimedia, Soft Computing and Bioinformatics**. NJ: Wiley, 2003, 401p. ISBN 0-471-46054-0

NACE International. Corrosion Failure. Disponível em <<https://www.nace.org/corrosion-failure.aspx>>. Acesso em 10 de Outubro de 2017.

Nagabhushana, S. **Data Warehousing Olap and Data Mining**. New Age International, 2006.

NEWELL, A. **Heuristic programming: ill-structured problems**. Progress in Operations Research Vol. 3. New York: John Wiley and Sons, 1969.

NOGUEIRA, Rildo de Mendonça. **Análise dos Impactos Harmônicos em uma Indústria de Manufatura de Eletroeletrônicos Utilizando Árvores de Decisão**. Dissertação (Mestrado) na Universidade Federal do Pará, Belém, 2015.

NPL - National Physical Laboratory. **A Short Introduction to Corrosion and its Control**. Disponível em : <http://www.npl.co.uk/upload/pdf/basics_of_corrosion_control.pdf>. Acesso em 24 de Outubro de 2017

NUNES, G. F. M. **Formas de Corrosão**. Apresentação da disciplina Corrosão I da CEFET-MG. Disponível em: <http://www.tecquimica.cefetmg.br/galerias/arquivos_download/Formas_de_Corrosao_-_Gisele_-_Cefet_2014.pdf>. Acesso em 24 de Outubro de 2017.

PEREZ, Nestor. **Electrochemistry and corrosion science**. 2. ed. Vol. 412. Switzerland: Springer, 2016.

PERINA, F. C. **Avaliação da toxicidade de biocidas utilizados em tintas antiincrustantes**. Dissertação (Mestrado) em Oceanografia Física, Química e Geológica na Universidade Federal do Rio Grande, Rio Grande, 2009

PERUCHI, S. **Monitoramento de Microrganismos no Centro Cirúrgico do Hospital São Judas Tadeu, Meleiro, Sc, Brasil**. TCC (Graduação) em Farmácia da Universidade do Extremo Sul Catarinense, Criciúma, 2011.

POPOV, Branko N. **Corrosion engineering: principles and solved problems**. Elsevier, 2015. ISBN 978-0-444-62722-3.

Quinlan, J. Ross. **C4.5 Programs for Machine Learning**. San Mateo: Morgan Kaufmann series in Machine Learning, 1993.

Quinlan, J. Ross. **Induction of decision trees** - Machine learning. vol. 1.1. 81-106, 1986.

RAYES-ROTH, F.; WATERMAN, D.; LENAT, D. **Building Expert Systems**. 1. ed. Reading, Massachusetts: Addison-Wesley, 1983. 443p.

REIS, Ricardo de Freitas. **A Importância da Tinta Líquida Industrial**. Monografia (Especialização) na Universidade Candido Mendes, Rio de Janeiro, 2012.

ROSENHAHN A., et al. **The role of “inert” surface chemistry in marine biofouling prevention**. Physical Chemistry Chemical Physics, Vol. 12 (2010) 4275-4286.

RUSSEL, S. J.; NORVIG, P. **Inteligência Artificial**. Tradução Consultoria PubliCare. 2. ed. Rio de Janeiro: Elsevier, 2004.

SHANNON, C. E.; WEAVER, W. **A Mathematical Model of Communication Urbana**. University of Illinois Press, 1949.

SHMUELI, G., et Al. **Data Mining for Business Analytics: Concepts, Techniques, and Applications** in R. John Wiley & Sons, 2017.

SILVA, Jonny Carlos da. **Expert System Prototype for Hydraulic System Design Focusing on Concurrent Engineering Aspects**. Tese (Doutorado) em Engenharia Mecânica na Universidade Federal de Santa Catarina, Florianópolis, 1998.

SILVA, R. K. da S. et. Al. **Determinação da Corrosividade de Amostras de Petróleo do Espírito Santo**. XLVI Congresso Brasileiro de Química, 2006. Disponível em <<http://www.abq.org.br/cbq/2006/trabalhos2006/3/413-589-3-T1.htm>>. Acesso em 20 de Outubro de 2017.

STAIR, R. M.; REYNOLDS, G. W. **Princípios de Sistemas de Informação**. Rio de Janeiro: LTC, 2002.

TAN, P. N.; STEINBACH, M.; KUMAR, V. **Introdução ao DATA-MINING: Mineração de Dados**. 1. ed. Rio de Janeiro: Ciência Moderna, 2009.

TAYLOR, David A. **Object-Oriented Technology: A Manager's Guide**. Massachusetts: Addison-Wesley Publishing Company, 1990.

TIMENETSKY, Jorge. **Controle de Microrganismos: Conceitos básicos e termos**. São Paulo, ICB-USP, 2012.

TIWARI, A.; RAWLINS, J.; HIHARA, L. H. **Intelligent Coatings for Corrosion Control**. Kidlington: Butterworth-Heinemann, 2015. 728p. ISBN 978-0-12-411467-8

WATERMAN, D. A., **A Guide to Expert Systems**. Addison-Wesley Publ. Company, 1986.

WEN, M.; DUŠEK K. **Protective Coatings: Film Formation and Properties**. Springer, 2017.

WINSTON, P. H. **Artificial Intelligence**. 3. ed. Massachusetts: Addison-Wesley, 1992.

WITTEN, I.;FRANK, E. **Data Mining: Practical Machine Learning Tools and Techniques**. 2. ed. Morgan Kaufmann series in data management systems. San Francisco: Elsevier, 2005. ISBN: 0-12-088407-0

YE, Nong. **The Handbook of Data Mining**. Vol. 24. Mahwah, NJ/London: Lawrence Erlbaum Associates, Publishers, 2003.

YEBRA, D. M. et al. **Antifouling Technology** - Past, Present and Future Steps Towards Efficient and Environmentally Friendly Antifouling Coatings. Progress in Organic Coatings, vol. 50, 75-104, 2004, 29.

ZIMMERMANN, Marco Aurélio. **Sistema Especialista Protótipo para Auxílio na Seleção de Bombas Hidrostáticas**. Dissertação (Mestrado) em Engenharia Mecânica na Universidade Federal de Santa Catarina, Florianópolis, 2003.

ZUBEN, Fernando J. Von. **Sistemas Baseados em Regras e Árvores de Decisão**. Apostila, 2011. Disponível em <ftp://ftp.dca.fee.unicamp.br/pub/docs/vonzuben/ea072_2s11/topico6_EA072_2s11.pdf>. Acessado em 10 de Outubro de 2017