

**UNIVERSIDADE FUMEC  
FACULDADE DE CIÊNCIAS EMPRESARIAIS  
MESTRADO EM SISTEMAS DE INFORMAÇÃO E GESTÃO DO  
CONHECIMENTO**

**APLICAÇÃO DE ESTRATÉGIA DE SISTEMAS PERGUNTAS-RESPOSTAS  
PARA RECUPERAÇÃO DE INFORMAÇÕES: análise de informações em um  
sistema de registro de ocorrências de *bugs* e melhorias**

**VIVIÂNE DE ALMEIDA TÔRRES**

**Belo Horizonte – MG**

**2018**

**VIVIÂNE DE ALMEIDA TÔRRES**

**APLICAÇÃO DE ESTRATÉGIA DE SISTEMAS PERGUNTAS-RESPOSTAS  
PARA RECUPERAÇÃO DE INFORMAÇÕES: análise de informações em um  
sistema de registro de ocorrências de *bugs* e melhorias**

Projeto de pesquisa apresentado ao Curso de Mestrado em Sistemas de Informação e Gestão do Conhecimento, como parte dos requisitos para obtenção do título de Mestre em Sistemas de Informação e Gestão do Conhecimento.

Área de Concentração: Recuperação da Informação, Processamento de Linguagem Natural e Sistemas Especialistas (Inteligentes).

Linha de Pesquisa: Tecnologia e Sistemas de Informação.

Trilha: T3 - Organização, Recuperação e Representação da Informação.

Prof. Orientador: Luiz Cláudio Gomes Maia.

**Belo Horizonte – MG**

**2018**



UNIVERSIDADE  
FUMEC

Dissertação intitulada “**APLICAÇÃO DE ESTRATÉGIA DE SISTEMAS PERGUNTAS-RESPOSTAS PARA RECUPERAÇÃO DE INFORMAÇÕES: análise de informações em um sistema de registro de ocorrências de bug’s e melhorias**” de autoria de Viviane de Almeida Torres, aprovada pela banca examinadora constituída pelos seguintes professores:

---

Prof. Dr. Luiz Cláudio Gomes Maia – Universidade FUMEC  
(Orientador)

---

Profa. Dra. Cristiana Fernandes De Müylder – Universidade FUMEC  
(Examinador Interno)

---

Profa. Dra. Josmária Lima Ribeiro de Oliveira – PUC MINAS  
(Examinador Externo)

---

Lázaro Mariano de Souza Júnior, Esp. – Davinti Soluções  
(Consultor *Ad Hoc*)

---

Prof. Dr. Fernando Silva Parreiras  
Coordenador do Programa de Pós-Graduação em Sistemas de Informação e Gestão do  
Conhecimento da Universidade FUMEC

Belo Horizonte, 13 de agosto de 2018.

REITORIA

Av. Afonso Pena, 3880 - Cruzeiro  
30130-009 - Belo Horizonte, MG  
Tel. 0800 0300 200  
www.fumec.br

CAMPUS

Rua Cobre, 200 - Cruzeiro  
30310-190 - Belo Horizonte, MG  
Tel. (31) 3228-3000  
www.fumec.br

### **Dados Internacionais de Catalogação na Publicação (CIP)**

T693a Tôrres, Viviâne de Almeida, 1985 -  
Aplicação de estratégia de sistemas perguntas-respostas  
para recuperação de informações: análise de informações em um  
sistema de registro de ocorrências de bug's e melhorias / Viviâne  
de Almeida Tôrres. – Belo Horizonte, 2018.  
138 f : il. ; 29,7 cm

Orientador: Luiz Cláudio Gomes Maia  
Dissertação (Mestrado em Sistemas de Informação e  
Gestão do Conhecimento), Universidade FUMEC, Faculdade de  
Ciências Empresariais, Belo Horizonte, 2018.

1. Sistemas de consultas e respostas. 2. Recuperação da  
informação - Brasil. 3. Supermercados – Belo Horizonte. I.  
Título. II. Maia, Luiz Cláudio Gomes. III. Universidade  
FUMEC, Faculdade de Ciências Empresariais.

CDU: 65.011:681.3.6



Aos meus pais,  
a quem amo tanto  
e mesmo sem terem grandes oportunidades de escolaridade,  
me incentivaram a estudar,  
acreditaram em mim  
e contribuíram imensamente para me tornar o que sou hoje.

## **Agradecimentos**

Obrigada, Deus. Força maior que rege minha vida, sem Ti nada sou. Que o conhecimento adquirido seja aplicado para melhorar o dia-a-dia da vida das pessoas, porque do contrário não haveria sentido algum.

Ao orientador, Prof. Dr. Luiz Maia, por me orientar e por estar comigo nessa trajetória.

À Prof<sup>a</sup>. Dr<sup>a</sup>. Cristiana Muylder, pelo incentivo, pelas dicas, por mostrar novos caminhos e alternativas.

Ao Prof. Dr. Josmária Oliveira e ao Lázaro Mariano, por participarem deste momento.

Aos professores do Mestrado FUMEC pelo aprendizado e por contribuírem para o meu crescimento como pessoa, como estudante e como profissional.

À FUMEC, por me proporcionar essa oportunidade.

À Organização Verdemar Ltda., pelo apoio à presente pesquisa.

A Marcos Chaves, Superintendente de TI da Organização Verdemar Ltda., por seu apoio e pela liberação das informações necessárias para a composição do presente trabalho.

A meus pais, por me incentivarem a estudar e buscar conhecimento. Por me fazerem acreditar que sou capaz, que sou forte e que posso sempre mais.

Aos meus parentes e amigos, pelas orações e pelas vibrações de energia positivista.

À Tissiâne Tôrres Vieira, por estar comigo ao longo dessa caminhada e por ter compartilhado momentos tão ímpares e difíceis de descrever.

A Tiago Alves, pelo auxílio técnico, pelo suporte e pelo companheirismo.

*O conhecimento serve para encantar as pessoas, não para humilhá-las.*

Mário Sérgio Cortella.

## Resumo

Nos dias atuais, a tecnologia está presente nos diversos segmentos e sozinha não é suficiente para ser um diferencial nas empresas, principalmente quando não se detém o conhecimento. O conhecimento pode ser apresentado em diversos formatos, estando este intangível ou devidamente registrado. Quando há registro desse conhecimento, a recuperação dessas informações deve ser de fácil acesso para permitir agilidade e eficiência para a tomada de decisão. No entanto, recuperar informações com precisão não é tarefa simples. O presente trabalho trata sobre a geração de uma coletânea de documentos (*corpus* em formato adequado para *chatbots*) criada a partir de uma base histórica de atendimentos do sistema Mantis® (*software* livre que permite o registro de *bugs* e melhorias) de uma rede de supermercados. A base foi submetida a um serviço de perguntas e respostas que possui técnicas de aprendizado de máquina, *bot* e conversação natural, a fim de realizar análise dos retornos recuperados. O resultado do trabalho é a conversão de vários registros de soluções já revisadas anteriormente, em um direcionamento único e objetivo para a solução do problema questionado. Os atendimentos são considerados base de conhecimento para a recuperação de informações e resposta para o usuário. Com isso, cogita-se a possibilidade de aumento de produtividade, otimização do tempo de atendimento e o compartilhamento do conhecimento já existente, minimizando a perda de informações ou a dificuldade de localização das informações, a partir da utilização de um *chatbot* que consolida as possíveis soluções já utilizadas anteriormente.

Palavras-chave: Sistemas Perguntas e Respostas. *Chatbot*. Mantis®. Recuperação da Informação. Rede de Supermercados.

## Abstract

In the present day, technology is present in a wide range of segments and by itself is not enough to be a differential in companies, especially when it is not possible to halt knowledge. Knowledge can be presented in distinct formats, either being intangible or properly registered. When this knowledge is recorded, the retrieval of information should be easily accessible to allow agility and efficiency for decision making. However, retrieving information accurately is not a simple task. The present work copes with the generation of a collection of documents (*corpus* in suitable format for *chatbots*) created from a historical base of attendance of the Mantis® system (free software that allows the registration of *bugs* and improvements) of a supermarket chain. The database was submitted to a question and answer service that has machine learning techniques, either *bot* and natural conversation, in order to perform an analysis of the recovered feedbacks. The result of the work is the conversion of several records of previously reviewed solutions into a single, objective targeting solution for the problem under analysis. The attendance results are considered knowledge base for information retrieval and response to the user. Thus, it is to be considered the possibility of increase in productivity, optimization of the time of service and sharing of existing knowledge, minimizing loss of information or difficulty in locating information, through the use of a *chatbot* that consolidates the possible solutions that have been used previously.

Keywords: Questions and Answers Systems. Chatbot. Mantis®. Information Retrieval. Supermarket chain.

## Lista de Figuras

Figura 1 - Processo para recuperação da informação .....	26
Figura 2 - Arquitetura genérica de um sistema de perguntas e respostas. ....	34
Figura 3 – Processamento de documentos da coleção – sistema QA .....	37
Figura 4 – Processamento da pergunta – sistema QA .....	39
Figura 5 – Resultado varejo brasileiro 2016 - <i>ranking</i> ABRAS .....	53
Figura 6 - Exemplos ilustrativos: equipamentos utilizados em supermercados. ....	58
Figura 7 – <i>Ranking</i> supermercados no Brasil .....	59
Figura 8 – <i>Ranking</i> de supermercados em Minas Gerais. ....	60
Figura 9 - <i>Ranking</i> de loja igual a supermercado. ....	60
Figura 10 - Células tecnologia informação Verdemar. ....	63
Figura 11 - Fatores que atingem a competitividade. ....	66
Figura 12 – Capa da Revista Gôndola. ....	69
Figura 13 - Componentes da infraestrutura de TI.....	72
Figura 14 – Tela resumo do Mantis® .....	75
Figura 15 – Tela Abertura do Caso no Mantis® .....	76
Figura 16 – Tela pesquisa do Mantis®.....	77
Figura 17 – Processo de atendimento TI no Verdemar.....	79
Figura 18 – Etapas da metodologia adotada.....	83
Figura 19 - Exemplo de transformação. ....	89
Figura 20 - Esquema QNA <i>Maker</i> . ....	91
Figura 21 - Exemplo caso Mantis – visualização interna.....	96
Figura 22 – Exemplo caso Mantis – visualização externa.....	97
Figura 23 - <i>Step</i> remoção de caracteres especiais. ....	99
Figura 24 – Criação da base de conhecimento. ....	100
Figura 25 – Conexão do serviço QnA <i>Maker</i> .....	101
Figura 26 – Inclusão nome para base de conhecimento.....	101
Figura 27 – População da base de conhecimento. ....	102
Figura 28 - Conclusão da base de conhecimento. ....	102
Figura 29 – Visualização da base de conhecimento importada. ....	103
Figura 30 – Perguntas e respostas da base de conhecimento importada.....	103
Figura 31 – Realização de testes na base de conhecimento importada. ....	104

Figura 32 – Inspeção na base de conhecimento importada.....	104
Figura 33 – Recurso de publicação do QnA <i>Maker</i> .....	105
Figura 34 – Publicação da base conhecimento no QnA <i>Maker</i> .....	106
Figura 35 – Serviço de <i>Bot</i> .....	106
Figura 36 – Identificação serviço de <i>bot</i> .....	107
Figura 37 – Tipo de <i>bot</i> .....	107
Figura 38 – Configurações do QnA <i>Marker</i> .....	108
Figura 39 – Informações de acesso do QnA <i>Maker</i> – antes.....	108
Figura 40 – Pergunta 1 – <i>chatbot</i> Viviâne.....	114
Figura 41 – Pergunta 2 – <i>Chatbot</i> Viviâne, .....	115
Figura 42 – Pergunta 3 – <i>Chatbot</i> Viviâne.....	115
Figura 43 – Pergunta 4 – <i>Chatbot</i> Viviâne.....	116
Figura 44 – Pergunta 5 – <i>Chatbot</i> Viviâne.....	117
Figura 45 – Mensagem QnA <i>Maker</i> para salvar alterações.....	118
Figura 46 – Refinamento no QnA <i>Maker</i> – uma pergunta várias anotações.....	119
Figura 47 – Exemplo de pergunta para exclusão no QnA <i>Maker</i> .....	120
Figura 48 – Edição da pergunta no QnA <i>Maker</i> .....	121

## Lista de Tabelas

Tabela 1 - Processos e técnicas da extração da informação .....	28
Tabela 2 - Linhas de pesquisa para sistemas de perguntas e respostas .....	41
Tabela 3 - Causas que ocasionam problemas no estoque .....	62
Tabela 4 - Perguntas e possíveis respostas sistêmicas .....	80
Tabela 5 - Campos do Mantis®.....	94
Tabela 6 – Objetivos x resultados x fundamentação teórica .....	123



## Lista de Abreviaturas e Siglas

ABRAS	Associação Brasileira de Supermercados
AIIM	<i>Association for Information and Image Management</i>
AIML	<i>Artificial Intelligence Markup Language</i>
API	<i>Application Programming Interface</i>
BI	<i>Business Intelligence</i>
CLEF	<i>Cross Language Evaluation Forum</i>
CMC7	Caracteres Magnéticos Codificados em Sete Barras
CMMI	<i>Capability Maturity Model Integration</i>
CobIT	<i>Control Objectives for Information and related Technology</i>
CPU	Unidade Central de Processamento
DI	Declaração de Importação
DRU	Definição de Requisitos de Usuários
DUN	<i>Distribution Unit Number</i>
EAN	<i>European Article Numbering</i>
ECR	<i>Efficient Consumer Response</i>
EDI	<i>Electronic Data Interchange</i>
EI	Extração de Informação
ECF	Emissor de Cupom Fiscal
ERP	<i>Enterprise Resource Planning</i>
ETL	<i>Extract Transform Load</i>
EUA	Estados Unidos da América
FCA	<i>Formal Concept Analysis</i>
FUMEC	Fundação Mineira de Educação e Cultura
GAM	Gestão Administrativa de Mercadorias
GIONSENG	<i>Guided Input Natural Language Search Engine</i>
HD	<i>High disk</i>
IBM	<i>International Business Machines</i>
ICMS	Imposto sobre operações relativas à circulação de mercadorias e sobre prestações de serviços de transporte interestadual, intermunicipal e de comunicação
ISBN	<i>International Standard Book Number</i>

ISO	<i>International Standards Organization</i>
ISSN	<i>International Standard Serial Number</i>
ITIL	<i>IT Infrastructure Library</i>
LASER	<i>Light Amplification by Stimulated Emission of Radiation</i>
LSTM	<i>Long Short-Term Memory</i>
MB	<i>Megabyte</i>
MFD	Memória de Fita Detalhe
NFCe	Nota Fiscal de Consumidor Eletrônica
PANTO	<i>Portable nAtural laNguage inTerface to Ontologies</i>
PAPEL	<i>Palavras Associadas Porto Editora Linguateca</i>
PC	<i>Personal Computer</i>
PDA	<i>Personal Digital Assistant</i>
PDI	<i>Pentaho Data Integration</i>
PDV	Ponto de Venda
PHP	<i>Hypertext preprocessor</i>
PLN	Processamento de Língua Natural
PLNPLU	<i>Price Look-Up</i>
PMBOK	<i>Project Management Body of Knowledge</i>
PPGSIGC	Programa de Pós-Graduação em Sistemas de Informação e Gestão do Conhecimento
QA	<i>Question Answering</i>
QuestIO	<i>Questionbased Interface to Ontologies</i>
RAM	<i>Random Access Memory</i>
RDF	<i>Resource Description Framework</i>
RI	Recuperação de Informação
RNN	<i>Recurrent Neural Network</i>
ROM	<i>Read-Only Memory</i>
RSL	Revisão Sistemática de Literatura
SAT	Sistema Autenticador e Transmissor de Cupons Fiscais Eletrônicos
SEFAZ	Secretaria da Fazenda
SPR	Sistemas de Perguntas-Respostas
SQAS	<i>Shallow Question Answering System</i>
SQL	<i>Structured Query Language</i>

SRI	Sistemas de Recuperação de Informações
START	<i>SynTactic Analysis using Reversible Transformations</i>
SVM	<i>Support Vector Machines</i>
SVN	<i>Subversion</i>
TCP	Terminal de Consulta de Preço
TED	Transferência Eletrônica de Dados
TEF	Transferência Eletrônica de Fundos
TI	Tecnologia da Informação
TREC	<i>Text REtrieval Conference</i>
UCC	<i>Uniform Code Council</i>
URL	<i>Uniform Resource Locator</i>
XML	<i>Extensible Markup Language</i>

## Sumário<sup>1</sup>

<b>1</b>	<b>Introdução.....</b>	<b>16</b>
1.1	Motivação e justificativa.....	19
1.2	Lacuna a ser explorada.....	20
1.3	Objetivos.....	21
1.3.1	<i>Objetivo geral</i> .....	21
1.3.2	<i>Objetivos específicos</i> .....	21
1.4	Contribuições.....	21
1.5	Adequação do projeto de mestrado e a linha de pesquisa.....	22
1.6	Estrutura da pesquisa.....	23
<b>2</b>	<b>Referencial Teórico.....</b>	<b>24</b>
2.1	Recuperação da informação.....	24
2.2	Extração da informação.....	27
2.3	Sistemas de perguntas e respostas.....	30
2.3.1	<i>Processamento da pergunta</i> .....	35
2.3.2	<i>Processamento dos documentos</i> .....	36
2.3.3	<i>Processamento da resposta</i> .....	39
2.3.4	<i>Linhas de pesquisas</i> .....	40
2.3.5	<i>Recursos para auxílio dos sistemas de perguntas e respostas</i> .....	42
<b>3</b>	<b>Trabalhos Relacionados.....</b>	<b>44</b>
<b>4</b>	<b>Softwares de Perguntas/Respostas Relacionados.....</b>	<b>46</b>
<b>5</b>	<b>Contextualização do Ambiente.....</b>	<b>52</b>
5.1	Tecnologia da informação no varejo.....	52
5.1.1	<i>Sistemas para gerência de informações do frente de caixa</i> .....	54
5.1.2	<i>Sistemas para gerência de informações do retaguarda</i> .....	56
5.2	Tecnologia da informação na rede de supermercados.....	59

---

<sup>1</sup> Este trabalho foi revisado de acordo com as novas regras ortográficas aprovadas pelo Acordo Ortográfico assinado entre os países que integram a Comunidade de Países de Língua Portuguesa (CPLP), em vigor no Brasil desde 2009. E foi formatado de acordo com a Norma APA, 2017.

<b>5.2.1 Célula de suporte técnico</b> .....	<b>63</b>
<b>5.2.2 Célula de suporte a sistemas</b> .....	<b>65</b>
<b>5.2.3 Célula de desenvolvimento interno</b> .....	<b>68</b>
<b>5.2.4 Célula de infraestrutura</b> .....	<b>71</b>
<b>5.3 Mantis - Sistema de registro de bugs e melhorias</b> .....	<b>73</b>
<b>5.3.1 Cenário atual</b> .....	<b>77</b>
<b>5.3.2 Cenário desejado</b> .....	<b>79</b>
<b>6 Metodologia</b> .....	<b>81</b>
<b>6.1 Etapas e procedimentos</b> .....	<b>81</b>
<b>6.2 Mantis: registro de atendimento e exportação de dados/tabelas</b> .....	<b>84</b>
<b>6.3 Montagem do corpus</b> .....	<b>86</b>
<b>6.3.1 Projeto do corpus</b> .....	<b>87</b>
<b>6.3.2 Compilação e manipulação do corpus</b> .....	<b>87</b>
<b>6.3.3 Extração e conversão do banco</b> .....	<b>88</b>
<b>6.4 Elaboração das perguntas x respostas</b> .....	<b>89</b>
<b>6.4.1 Campos do Mantis® – utilização nas perguntas e respostas</b> .....	<b>92</b>
<b>6.5 Retirada de caracteres especiais ou inválidos</b> .....	<b>98</b>
<b>6.6 Importação do arquivo TXT/XLSX</b> .....	<b>100</b>
<b>6.7 Adequação e análise das respostas</b> .....	<b>102</b>
<b>6.8 Publicação da base</b> .....	<b>105</b>
<b>6.9 Funcionamento do QnA Maker</b> .....	<b>108</b>
<b>7 Testes e Resultados Obtidos</b> .....	<b>112</b>
<b>8 Considerações Finais</b> .....	<b>125</b>
<b>8.1 Limitações da pesquisa e recomendações para trabalhos futuros</b> .....	
<b>Referências</b> .....	<b>128</b>
<b>Apêndice</b> .....	<b>136</b>

## 1 Introdução

Segundo Buckland (1991), os três principais usos do termo informação podem ser classificados em: processo, conhecimento ou coisa:

- Como processo: o ato de informar ou comunicar o conhecimento, notícias sobre um fato ou ocorrência;
- como conhecimento: quando o que é percebido pela informação como processo e o conhecimento comunicado tendo como sua principal característica a intangibilidade;
- como coisa: tudo aquilo que pode ser identificado como informativo - objetos, documentos, textos, dados ou eventos, destacando-se por sua materialidade ou tangibilidade.

Nos três casos, dentro de uma organização empresarial, o que é relevante é a possibilidade de acesso, a disponibilidade e localização nos momentos oportunos para tomada de decisão. É ineficiente uma informação tardia mediante um questionamento com tempo de resposta determinado.

Lopes, Muylder e Judice (2011) conceituam a informação como um ativo organizacional em um cenário em que a competitividade, inovações tecnológicas e a diminuição da distância tornaram o processo de tomada de decisão e o tempo diferenciais diante da concorrência. As boas práticas da governança da tecnologia de informação classificam a informação como um ativo da tecnologia de informação que deve ser gerenciado de forma segura nas empresas para que possam agregar valor ao negócio (Puhl & Preuss, 2013).

O volume de informações e de dados criado nas organizações para utilização interna e externa por empresas de médio e grande porte é espantoso (Parreiras & Bax, 2003). Trata-se de um volume muito alto de documentos entre relatórios, *websites* e publicações. Com isso, é necessário avaliar o armazenamento dessas informações, a sua recuperação, uma gestão de conteúdo eficiente para centralizar os dados e liberar acesso aos indivíduos autorizados.

O estudo realizado pela *Association for Information and Image Management* (AIIM) e descrito por Puhl e Preuss (2013) apurou que até 80% das informações de uma empresa podem estar em dados não estruturados, espalhados pela empresa, com armazenamento em computadores pessoais ou via troca de *e-mails* ou em dispositivos portáteis, como *pen drive*, por exemplo. Em média, de maneira geral, o estudo aborda empresas que não possuem um gerenciador de conteúdo e podem vir a realizar 19 cópias de cada documento organizacional e perder um em cada 20 (1/20) documentos, além de gastar 400 horas por ano procurando e tentando recuperar documentos perdidos (Puhl & Preuss, 2013).

Para Machado (2014), a EMC (empresa de âmbito internacional de armazenamento de dados) realizou estudo sobre o universo digital e relatou que já existem disponíveis hoje no mundo quase 1 septilhão de *bits* de informação. O estudo prevê que até o ano de 2020 o número de dados armazenados em computadores, servidores, celulares, *smartphones* e *tablets* seja multiplicado no mínimo por seis, volume este referente restritamente aos dados virtuais armazenados. Com isso, estratégias de recuperação de informação podem ser consideradas imprescindíveis.

Para as redes de supermercados, o volume de dados também é ponto de atenção. O número de informações que são de caráter fiscal, ou seja, obrigações legislativas, também tende a crescer (Cico, 2017). O *site* do *E-CommerceBrasil* relata sobre duas novas siglas Nota Fiscal de Consumidor Eletrônica (NFCe) e Sistema Autenticador e Transmissor de Cupons Fiscais Eletrônicos (SAT-CFe) que deverão ser inclusas no cupom fiscal, documento que toda loja e empresa varejista é obrigada a emitir no momento da venda. O novo documento automatizará o processo de emissão do cupom, reduzindo custos de obrigações acessórias aos contribuintes, e facilitará o controle fiscal pelas administrações tributárias.

Além disso, o consumidor poderá conferir a validade do documento e escolher entre receber o papel ou somente o arquivo digital. Esse processo, a princípio, segundo Cico (2017), tem previsão de implantação até 2020, no entanto, também será mais uma informação a ser armazenada, assim como a obrigatoriedade do armazenamento do *Extensible Markup Language* (XML), formato digital utilizado para as notas fiscais eletrônicas) por cinco anos.

Assim como a *web*, o volume de informações nas empresas cresce exponencialmente, e administrar o volume de informações que trafegam nas redes internas e externas tornou-se uma preocupação. O excesso de dados ou sua redundância ou a dificuldade de localização prejudicam a produtividade das organizações (Núcleo, 2017).

Todas as organizações atualmente pertencem a esse cenário, cujos dados e informações necessários às suas atividades crescem exponencialmente e, com eles, a complexidade de manipulação e extração de informações alinhadas aos problemas e tarefas do dia a dia. Os sistemas de informação propõem gerenciar os dados históricos auxiliando em tomadas de decisão. No entanto, estão vinculados à necessidade de recuperação da informação e à resposta a questionamentos, normalmente em tempo real, com foco no próprio processo decisório. Nesse ambiente as bases de dados possuem crescente volume de informações que podem gerar dificuldade em identificar um dado específico ao problema a ser tratado.

De acordo com Ferreira, Campos e Macedo, (2006, p. 2) “a tecnologia de informação está redefinindo os fundamentos dos negócios” (p. 1). Desde o atendimento ao cliente até a distribuição dos produtos, existe uma dependência dos sistemas de informação. Sendo assim, a Tecnologia de Informação (TI) e seus custos passaram a fazer parte integrante da rotina das empresas. Uma das alternativas para obter ganhos de produtividade via redução de custos e a construção de alianças estratégicas com os fornecedores e distribuições são investimentos em tecnologia de informação, transformando o setor varejista e modificando sua maneira de competir, gerando ganhos de produtividade (Ferreira *et al.*, 2006).

As informações analíticas ou sintetizadas são essenciais e dependem da especificidade da solicitação. Seu valor é de difícil mensuração, mas podem possibilitar oportunidades mercadológicas, proporcionar redução de custos, aumento de produtividade e aumento de espaço no mercado competitivo, além de garantir diferencial para a organização.

Cesarino (1985) argumenta que os “sistemas de recuperação podem ser definidos como um conjunto de operações executadas consecutivamente em uma totalidade de



informações visando encontrar as informações realmente relevantes. E para esse fim, executam funções de seleção, análise, indexação e busca das informações” (Cesarino, 1985, p. 157).

Já os sistemas de perguntas e respostas são (Silva, 2011, p.55):

Tarefas provenientes da combinação das áreas recuperação de informação (RI) e de extração de informação (EI) que visam à obtenção de uma resposta exata e precisa com base em uma pergunta formulada em linguagem natural (Bilotti, 2004, p. 15).

A complexidade das pesquisas, a recuperação de regras de negócios e a localização das informações críticas e de dados relevantes de maneira eficiente no momento oportuno com alta disponibilidade e de fácil acesso/recuperação também constituem uma situação cada vez mais solicitada, assim como apresentar respostas a questionamentos diversos.

Diante desse cenário, busca-se compreender o seguinte problema de pesquisa:

**Como responder de maneira assertiva e rápida dúvidas e questionamentos de usuários a partir de uma base histórica em uma rede de supermercados frente à necessidade diária de atendimento de suporte a sistemas?**

### **1.1 Motivação e justificativa**

Ferreira *et al.* (2006) prelecionam que a inovação, o conhecimento e as estratégias são os principais fatores que definem a competitividade e o desenvolvimento de nações, regiões, setores, empresas e até indivíduos. As empresas varejistas em resposta ao processo de mercado financeiro mundial passaram a dar mais ênfase às estratégias competitivas baseadas no uso da tecnologia de informação.

Existe uma crescente competição internacional que trouxe a necessidade de introduzir eficientemente os avanços das tecnologias de informação e das políticas e estratégias mais modernas e eficientes (Ferreira *et al.*, 2006). Os sistemas de informação auxiliam no gerenciamento das informações das empresas e possuem extrema importância.

Quando esses sistemas disponibilizam a consulta rápida e a resposta adequada a determinados levantamentos, permite melhor tomada de decisão.

Com isso, a motivação deste trabalho é contribuir para o levantamento de técnicas e métodos de sistemas perguntas e respostas que auxiliem o suporte de sistemas, nas empresas, no esclarecimento de dúvidas relativas aos *softwares* e processos utilizados na organização, visto que sistemas gerenciais comuns podem não atender nesse quesito ou atender de forma parcial.

## 1.2 Lacuna a ser explorada

A recuperação da informação estuda os aspectos relativos à eficiência e à eficácia (precisão e revocação) das pesquisas realizadas. A informação correta no momento adequado permite atendimentos eficientes. O foco desta pesquisa é buscar informações de tecnologias de sistemas de perguntas e respostas que atendam à demanda de esclarecimento de dúvidas, pesquisando na base de dados histórica já existente no sistema de registro de ocorrências de *bugs* e melhorias, mesmo diante do crescimento dessa base.

Tendo isso em mente, a lacuna a ser explorada é a identificação de técnicas de sistemas de perguntas e respostas que possam auxiliar no aperfeiçoamento da recuperação da informação no sistema utilizado atualmente ou na substituição desse sistema. É necessária a análise de métodos e abordagens de sistemas de perguntas e respostas que possam proporcionar retornos de consultas textuais (perguntas) mais precisas. Pretende-se, portanto, a partir desse cenário científico e tecnológico, avançar nos estudos e apresentar direcionamentos que auxiliem nesse propósito.

O estudo permitirá a interação com conceitos interdisciplinares como a recuperação da informação, a organização e a representação da informação. Com isso, as áreas de Ciência da Informação, Ciência da Computação, Biblioteconomia, Aprendizagem de Máquina e Inteligência Artificial serão parte do objeto de estudo.

## 1.3 Objetivos

### 1.3.1 *Objetivo geral*

O objetivo deste trabalho é testar e/ou aplicar uma das técnicas de sistemas de perguntas e respostas em uma base histórica, a fim de realizar análises dos resultados recuperados, avaliando sua assertividade.

### 1.3.2 *Objetivos específicos*

Os objetivos específicos são:

- a) Elaborar uma coletânea de documentos (*corpus*) baseada no banco de dados existente.
- b) Gerar arquivo (*corpus*) em formato eletrônico para utilização do *chatbot* baseado nas etapas metodológicas de (McEnery & Wilson, 1996 citados por Alúcio & Almeida, 2006).
- c) Elaborar perguntas para realização de testes utilizando os atendimentos da base de conhecimento.
- d) Avaliar o retorno (respostas) do *chatbot* para as perguntas realizadas de acordo com a solução do atendimento.

## 1.4 Contribuições

Com esta pesquisa pretendeu-se contribuir para a discussão de novas tecnologias e novas possibilidades para diminuir o tempo de resposta aos usuários que possuem dúvidas sobre determinado domínio de assunto.

A busca por novas técnicas de sistemas perguntas e respostas pode auxiliar na recuperação das informações a partir do processamento de linguagem natural em uma base de dados específica. Com isso, visa-se, para o meio acadêmico, ao estudo das teorias existentes e ao apoio à pesquisa entre a universidade e as empresas, assim como para o meio empresarial, que poderá utilizar como base o levantamento realizado para proceder a adaptações ou reestabelecer novos sistemas.

Além disso, o estudo pode ser utilizado como suporte para outras pesquisas, gerando possibilidade de comparação com outros trabalhos e reflexões para empresas de ramos de atividades semelhantes ou com cenários similares.

De maneira geral, Amorim, Cury e Menezes (2014) acreditam que o grande desafio dos sistemas de perguntas e respostas é retornar à resposta que mais se aproxima da expectativa do solicitante para uma pergunta feita em linguagem natural. O processo completo requer muita atenção porque se trata de algo de alta complexidade e por esse motivo buscou-se aprofundar esta investigação.

### **1.5 Adequação do projeto de mestrado e a linha de pesquisa**

O Programa de Pós-Graduação em Sistemas de Informação e Gestão do Conhecimento (PPGSIGC) da Fundação Mineira de Educação e Cultura (FUMEC) possui como objetivo articular e aplicar conhecimento profissional no abrangente campo de sistemas de informação. O programa tem como missão a produção de atividades técnicas, de ensino e pesquisa, enfatizando a busca, o desenvolvimento e a integração do conhecimento e das atividades que dele se originam, com vistas à formação de docentes e pesquisadores.

O presente trabalho compreende os estudos de sistemas perguntas e respostas visando apresentar metodologias que auxiliem a recuperação das informações e o seu respectivo retorno de maneira eficiente. O projeto está classificado na linha de pesquisa tecnologia e sistemas de informação<sup>2</sup> na trilha T3 - organização, recuperação e representação da informação. Com isso, espera-se contribuir com pesquisas já existentes.

---

<sup>2</sup> Essa linha de pesquisa compreende estudos sobre os conceitos e processos de desenvolvimento de tecnologias e sistemas de informação integrados com banco de dados e dotados de recursos gráficos e usabilidade avançada, de acordo com os preceitos de gestão de projetos e qualidade de software. Trata também dos impactos dos sistemas baseados na Internet e das novas tecnologias no comportamento do consumidor e na gestão logística. Recuperado de: <http://www.fumec.br/pesquisa/propic/linhas-de-pesquisa-strictu-sensu/>.

Em alinhamento ao programa, o projeto contribui para a análise de técnicas, métodos e boas práticas utilizadas em sistemas de perguntas e respostas para auxiliar no processamento de linguagem natural. Foi feita revisão sistemática da literatura, permitindo a exploração de conceitos interdisciplinares com várias áreas do conhecimento. Trata-se de assunto que estabelece relação com ramos de conhecimentos correlatos, como a recuperação da informação, atingindo as áreas da Ciência da Informação, Ciência da Computação, Inteligência Artificial, entre outras.

## **1.6 Estrutura da pesquisa**

Além deste primeiro capítulo, que abordou o tema, cenário, problema, justificativas, objetivos, contribuições e a aderência ao Programa de Pós-graduação, o projeto da pesquisa contém outros sete capítulos.

O segundo capítulo aborda o referencial teórico dos construtos envolvidos no problema de pesquisa. Relata a recuperação e extração da informação, sistemas de perguntas e respostas com seus respectivos tipos de processamento (pergunta, documento e resposta), além de sua arquitetura.

O terceiro e quarto capítulo apresentam os trabalhos relacionados e *softwares* de sistemas de perguntas e respostas e *chatbots* já desenvolvidos e testados.

O quinto capítulo contempla informações relativas à contextualização do ambiente: a tecnologia no varejo e na empresa objeto de estudo, o sistema atual utilizado para busca de informações (sobre o sistema desenvolvido internamente) e o cenário atual vivido mediante a necessidade de respostas assertivas.

O sexto capítulo descreve a metodologia utilizada para o desenvolvimento do projeto, tendo o sétimo capítulo os resultados obtidos, seguido das considerações finais, as referências e os apêndices.

## 2 Referencial Teórico

Almansa e Macedo (2016) informam que o volume das publicações em bibliotecas digitais e em repositórios da web cresceram consideravelmente nos últimos anos e que junto a elas a complexidade e lentidão de localizar uma informação, um documento. Para minimizar esses problemas, a área da Ciência da Computação procura definir métodos eficientes para manipular grandes quantidades de informações em tarefas de coleta e de indexação de informações (Frakes e Baeza-Yates, 1992 como citado em Almansa e Macedo, 2016).

Ainda na linha de pensamento de Baeza-Yates e Ribeiro-Neto (1999, como citado em Almansa & Macedo, 2016) a recuperação de informação (RI) é um processo que se estende desde a representação da informação até o seu acesso, que pode ser uma alternativa, tal que a representação e a organização da informação facilitem a obtenção dos dados por parte do usuário.

### 2.1 Recuperação da informação

Os sistemas de recuperação da informação são interfaces entre uma coleção de recursos de informação, em meio impresso ou não, e uma população de usuários; e desempenham as seguintes tarefas: aquisição e armazenamento de documentos; organização e controle desses; e distribuição e disseminação aos usuários (Souza, 2006).

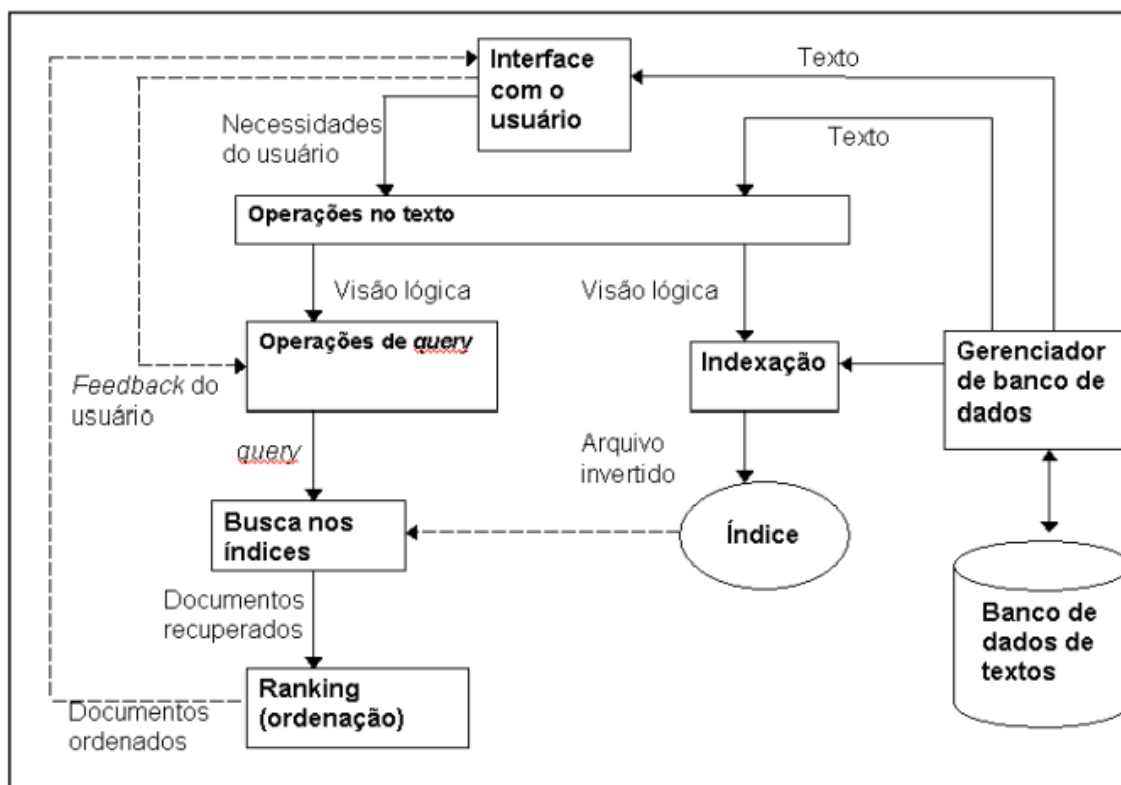
Para Souza (2006) o conceito de recuperação de informações é autoexplicatório, podem ser divididos em subsistemas de documentos, de usuários e de busca/recuperação. Eles servem de ponte entre o mundo dos criadores de informações e os usuários destas e, para isso, colecionam-nas e as organizam.

Monteiro, Fernandes, DeCarli e Trevisan (2017) opinam que “os sistemas de recuperação da informação devem representar, armazenar, organizar e localizar os itens de informação” (p. 163). Sua indexação é sua principal função, sendo que seus componentes devem incluir documentos, necessidade do usuário, realizar a consulta formulada e o processo de recuperação propriamente dito.

Recuperar informações é uma tarefa com dificuldades intrínsecas ao conceito de informação, como, por exemplo, a própria dificuldade do usuário em determinar sua real necessidade e o seu melhor atendimento aos documentos que fazem parte do acervo de informações do sistema (Souza, 2006).

A associação entre o registro e o seu conteúdo informado pode ser vago, o que acarreta uma série de problemas nas respostas às questões específicas enviadas, ocasionando baixas taxas de revocação (trata-se da abrangência, ou seja, a razão entre o número de documentos recuperados sobre o total de documentos disponíveis na base de dados; com essa taxa é possível medir o sucesso do sistema de recuperação de acordo com os documentos pertinentes recuperados) e precisão (trata-se da razão entre os número de documentos atinentes recuperados sobre o total de documentos recuperados; com essa taxa é possível medir o sucesso do sistema de recuperação em não reaver documentos que não sejam relevantes de acordo com a necessidade do usuário) (Souza, 2006).

A boa relação entre essas taxas de revocação e precisão, como explicita Souza (2006), pode oferecer maior referência de documentos relevantes. O diagrama proposto por esse autor exemplifica o processo de um sistema de recuperação automatizado:



**Figura 1** - Processo para recuperação da informação

Fonte: Souza, R. R. (2006). Sistemas de recuperação de informações e mecanismos de busca na web: panorama atual e tendências. *Perspectivas em Ciência da Informação*, 11(2), 161–173. Recuperado de: [http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S1413-99362006000200002](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-99362006000200002). <https://doi.org/10.1590/S1413-99362006000200002>.

Atualmente, a relevância é uma propriedade essencial para o desenvolvimento dos sistemas de recuperação de informações (SRI), porque sua apresentação é feita de diferentes maneiras de acordo com a modelagem de recuperação das informações. Os modelos podem ser booleano, vetorial, probabilístico, *fuzzy*, redes neurais, entre outros (Scarinci & Oliveira, 2017).

Para modelo booleano, é utilizada a teoria dos conjuntos, restringindo-se apenas aos resultados que pertencem à lógica de busca, sem a possibilidade de ordenar/classificá-los do mais para o menos relevante. No modelo vetorial, as palavras e expressões possuem valores de representação e os resultados são ordenados do mais relevante para o menos relevante. Para o modelo probabilístico, é usada a “teoria matemática das probabilidades, para apresentar os resultados que o sistema considera relevante, sendo que a relevância é determinada de acordo com a interação do usuário com o sistema” (Monteiro *et al.*, 2017, p. 167; Silva, Santos & Ferneda, 2013).



No modelo *fuzzy* são atribuídos valores no intervalo 0-1 para destacar os mais relevantes. Na apresentação dos resultados e nas redes neurais artificiais é feita uma simulação computacionalmente. Os resultados relacionados a termos equivalentes podem ser recuperados como itens relevantes. Para Monteiro *et al.* (2017, p.168), é:

Pautado em aspectos humanos, em algoritmos genéticos, esse modelo consegue exibir resultados prévios para que o usuário possa aplicar mais filtros e torná-los mais relevantes, posteriormente registra essas interações para otimizar pesquisas futuras (Silva *et al.*, 2013)

No tocante às máquinas de busca atuais são empregados os sofisticados processos de coleta, indexação, análise e interface de busca, com linguagens e programas para representar e descrever semanticamente a informação; as ontologias de domínio; as bases de conhecimento (chamadas de bases secundárias com entidades descritas semanticamente), fornecendo dados estruturados para os agentes inteligentes; programas e tecnologias de visualização e apresentação das informações em contextos (*Knowledge Graph* ou mapa do conhecimento) nas interfaces de busca (Monteiro *et al.*, 2017).

Dessa forma, o motor de busca participa de um sistema maior que utiliza aplicativos (*crawler*, *spider* ou robô) para capturar palavras-chave e frases que identifiquem o conteúdo de uma página *web* para indexá-la e armazená-la em sua base de dados. Com isso, as máquinas de busca permitem a organização do conhecimento em um ciberespaço por meio da indexação que realizam (Monteiro, 2006).

## 2.2 Extração da informação

Silva (2011, p.56) relata que:

Segundo Moens (2006), EI é a identificação, classificação e estruturação das informações textuais específicas em classes semânticas tornando-as mais apropriadas para o processamento de informações. Dessa forma, a área de EI analisa as partes dos documentos que potencialmente contêm informação relevante segundo os critérios de extração definidos (Kowalski & Maybury, 2000, p. 336).

A extração da informação é diferente da recuperação da informação (Scarinci & Oliveira, 2001). As técnicas utilizadas na recuperação da informação localizam

documentos pertinentes em uma coleção, porém estão impossibilitadas de extrair informação de documentos pertinentes de acordo com critérios especificados. Os sistemas de extração de informação possuem a habilidade de extrair dados relevantes nos artigos e sua representação em estruturas de acordo com critérios específicos.

Segundo Álvarez (2007), a EI é uma tarefa de encontrar informações específicas a partir de grandes volumes de documentos. Esses documentos podem ser apresentados com algum nível de estruturação na apresentação das informações ou podem ser apresentados de forma livre, sem nenhuma organização.

Para gerar ou explicar o conhecimento, inúmeras técnicas de análise intradocumental “visam obter fatos (blocos de informação relevante sobre o contexto tratado) ou ainda integrá-los e relacioná-los a fim de derivar novos fatos que podem ser utilizados (Grishman, 1997 como citado em Silva, 2011, p. 56).

Silva (2011) reconhece que “dentre os processos e técnicas é possível destacar o reconhecimento de entidades, resolução de correferências e as técnicas de análises léxica, sintática e semântica sobre o conteúdo textual”(p. 56). Também é possível destacar:

**Quadro 1 - Processos e técnicas da extração da informação**

<b>Processos e técnicas</b>	<b>Significado</b>
<b>Tokenization e normalização</b>	Consiste em dividir uma ou mais sentenças desconsiderando caracteres irrelevantes do documento. As sentenças são chamadas de <i>tokens</i> (instância de caracteres que representam uma semântica útil para processamento). Cada unidade passa por um processo de normalização transformando em uma forma equivalente e única (Kowalski & Maybury, 2000; Manning; Raghavan & Schütze, 2008). A normalização é sensível ao idioma e também acompanha as alterações, evoluções e acordos ortográficos de cada língua (Silva, 2011, p. 58).
<b>Consideração de <i>stop-words</i></b>	Trata-se de preposições, artigos e pronomes, termos comuns e frequentes nos textos. De maneira geral, desconsiderados em boa parte dos métodos de recuperação da informação (Manning; Raghavan & Schütze, 2008), podendo ser utilizados para identificação de idiomas, para classificação de perguntas e respostas, para identificação de padrões e aplicação de heurísticas com base na língua (Silva, 2011, p. 58).
<b>Reconhecimento de entidades (<i>Named-Entity Recognition</i>)</b>	Trata-se de uma técnica de extração de informação que identifica ou classifica as entidades em classes predefinidas, com auxílio de dicionários, taxonomias e ontologias (Silva, 2011, p. 58).
<b>Resolução de correferências (<i>Coreference resolution</i>)</b>	Busca-se identificar expressões anafóricas no texto para reconhecimento de terminologias escritas que referenciam uma mesma entidade. Exemplo: na frase “o presidente dos [Estados Unidos da América] EUA viaja a Londres amanhã. A capital da Inglaterra receberá Barack Obama com festa”, os

	termos “Presidente dos EUA” e “Barack Obama” representam a mesma entidade no texto, assim como “Londres” e “capital da Inglaterra”. O objetivo principal da tarefa de resolução de correferências é identificar essa relação entre os termos (Silva, 2011, p. 58).
<b>Part-Of-Speech Tagging (POS-Tagging)</b>	Consiste em classificar as sentenças ( <i>tokens</i> ) segundo as classes gramaticais do idioma. Entre essas classificações citam-se: substantivo, artigo, pronome, verbo, advérbio, numeral, etc. (Silva, 2011, p. 59).
<b>Reformulação e expansão de perguntas (reformulation and question expansion)</b>	Propõe alternativas às consultas do usuário por meio de uso de sinônimos, hierarquia e relações semânticas entre conceitos (Silva, 2011, p. 59).
<b>Desambiguação (word-sense disambiguation)</b>	Identifica e resolve as ambiguidades presentes nas entidades textuais de acordo com o domínio (Moldovan; Tatu & Clark, 2009). Exemplo: “São Paulo” pode estar associada a um time de futebol, a uma cidade ou ainda um estado brasileiro. O uso de ontologias é aplicado para a classificação de termos com base nas classes de domínio e também para a resolução de ambiguidades (Silva, 2011, p. 59).
<b>Lemmatization e Stemming</b>	Visam à redução das flexões ou formas relativas de palavras para uma base comum ou para o seu radical (Manning; Raghavan & Schütze, 2008). Exemplo: estudante, estudando, estudavam poderia considerar apenas a forma estudar (Silva, 2011, p. 59).
<b>Classificação de perguntas/respostas</b>	Realiza tratamento diferenciado das respostas conforme o tipo da pergunta. Exemplo: a pergunta “onde Abraham Lincoln nasceu?” espera-se como resposta um local geográfico (cidade, estado ou país). Já na variante “quando Abraham Lincoln nasceu?” a resposta deveria ser uma entidade diferente que representa o tempo (Silva, 2011, p. 59-60).
<b>Análises sintática e semântica</b>	Identificam-se as árvores ou funções sintáticas dos termos nas orações como o sujeito, predicado, objeto direto, indireto e padrões linguísticos. Na análise semântica, os relacionamentos de sinonímia, hipônimos, hiperônimos, antônimos, merônimos, associativos, são estabelecidos. Redirecionando-se para o uso de ontologias (Silva, 2011, p. 60).

Fonte: Silva, D. C. D. (2011). Uma arquitetura de business intelligence para processamento analítico baseado em tecnologias semânticas e em linguagem natural. *Business*, 161, 58-60. Recuperado de: <http://btd.egc.ufsc.br/wp-content/uploads/2011/04/DhiogoCardosoDaSilva.pdf>.

Existem várias técnicas que podem ser utilizadas para a construção dos sistemas de extração de informação (Álvarez, 2007). No entanto, o tipo de texto influencia muito na escolha do método. Entre os tipos de textos podem-se ressaltar:

- a) **Estruturado:** quando apresentam regularidade no formato de exibição das informações. Essa regularidade pode ser identificada por um sistema de extração de informação, permitindo que cada elemento de interesse seja identificado com base em regras uniformes. As regras são consideradas marcadores textuais, assim como delimitadores e/ou ordem de apresentação de elementos, como, por exemplo, formulários preenchidos;
- b) **Semiestruturado:** quando apresentam certa regularidade na disposição dos dados. Alguns dados do texto adotam uma formatação, enquanto as demais informações são apresentadas de forma irregular.

- c) Não estruturado:** os textos livres ou não estruturados são aqueles que não exibem regularidade na apresentação dos dados. Os dados a serem extraídos não são facilmente detectados, a menos que exista um conhecimento linguístico sobre eles, como, por exemplo, uma página *web*.

As técnicas de processamento de linguagem natural são amplamente utilizadas no processo de extração de informações de documentos semiestruturados e livres (Álvarez, 2007).

Ainda sobre essas técnicas, Álvarez (2007) comenta que possuem como objetivo compreender os textos em um formato de língua natural para localizar informações relevantes para serem extraídas. De maneira geral, possuem diferentes domínios com etapas de processamento e módulos específicos de extração.

Com a *web* permitiu-se que as extrações não ficassem limitadas apenas a uso de bases estruturadas e domínios de conhecimento fechado. Por esse motivo houve uma expansão de universo de consultas sobre os mais variados conteúdos, estendendo-se a imagens, áudios, vídeos, entre outros tipos de arquivo. Com isso, os sistemas de perguntas e respostas evoluíram para diversos tipos de aplicação (Silva, 2011).

### **2.3 Sistemas de perguntas e respostas**

Machado Júnior, Foleiss e Souza (2009) ressaltam que a construção de sistemas de perguntas e respostas é uma tarefa da extração da informação (*information retrieving*), na área da Linguística Computacional. Esses sistemas possuem como objetivo “responder a perguntas formuladas em linguagem natural, formulando as respostas a partir de uma coleção de textos” (Machado Júnior *et al.*, 2009, p.1).

Os sistemas de pergunta-resposta, conforme Quaresma, Rodrigues, Prolo e Vieira (2006), estão sendo debatidos em conferências internacionais como *Text REtrieval Conference*<sup>3</sup> (TREC) e o *Cross Language Evaluation Forum*<sup>4</sup> (CLEF), por se tratar de uma área de processamento de linguagem natural e por estarem ligados à

---

<sup>3</sup> <http://trec.nist.gov>.

<sup>4</sup> <http://www.clef-campaign.org>.

recuperação da informação. Cada vez mais os sistemas de perguntas e respostas ganham espaço para discussão.

Os sistemas de perguntas e respostas podem ser considerados uma otimização dos sistemas de recuperação da informação, porque ao mesmo tempo em que os “sistemas de RI retornam documentos completos, muitas vezes sem a ‘resposta desejada’, os sistemas de perguntas e respostas retornam respostas curtas e diretas” (Almansa, 2017, p.16).

De acordo com Rabelo e Barros (2005), existe crescente interesse na área de pergunta e resposta pela linguagem natural. Diversos trabalhos encontrados apresentam sistemas conversacionais, de *front\_end* para consulta em banco de dados assim como perguntas livres em fontes textuais, similar a enciclopédias. De maneira geral, o objetivo dos sistemas de perguntas e respostas é responder à pergunta mediante pesquisa em uma base de documentos não estruturada, tendo ele uma coleção de documentos não estruturados e uma base de documentos pré-processados, ou seja, índices invertidos de busca.

Os sistemas de perguntas-respostas (SPR) recebem uma pergunta como entrada, em linguagem natural; e mediante pesquisa em alguma base de dados, retornam à resposta esperada (Rabelo & Barros, 2005). Esses sistemas tiveram grande avanço na inteligência artificial por representarem o conhecimento e o processamento natural.

Balduccini *et al.* (2008, como citados em Amorim, Cury & Menezes, 2012, p. 2) classificaram os sistemas de pergunta e resposta que incorporam o conhecimento e o raciocínio (ou inferência) em três abordagens: baseado em lógica formal, extração de informação e usando lógica formal na extração da informação. Nos três casos são utilizadas linguagens lógicas para extrair o conhecimento inferido. Os sistemas que fazem uso das ontologias para raciocinar e representar o conhecimento são conhecidos por utilizarem modelagem qualitativa ou raciocínio qualitativo.

Quando o usuário necessita realizar uma busca de informações sobre determinado assunto, ele elabora uma pergunta pretendendo que sua necessidade seja atendida. Cita-se como exemplo: o usuário possui uma dúvida sobre o motivo da rejeição de

uma nota fiscal eletrônica e sua respectiva solução para realizar na Secretaria da Fazenda (SEFAZ)<sup>5</sup> um cancelamento de nota. Atualmente, após as notas fiscais serem autorizadas pela organização, existe um prazo máximo de 24 horas para realizar a solicitação de cancelamento de forma automática. Depois desse prazo é necessário documentar uma denúncia espontânea sobre o erro no lançamento da nota Balduccini *et al.* (2008, como citados em Amorim, Cury & Menezes, 2012).

Nesse caso, a pergunta poderia ser: “quais são os motivos rejeição de uma nota fiscal?”. A resposta para o usuário irá depender da arquitetura do sistema de perguntas e respostas. De maneira geral, sua “arquitetura básica é composta de três módulos principais, são eles: a) processamento da questão; b) processamento do documento; c) processamento da resposta” (Almansa, 2017,p.16).

A primeira etapa de processamento da questão possibilita a transformação da pergunta realizada pelo usuário em linguagem natural em uma consulta de linguagem computacional. Logo em seguida é feito o entendimento (compreensão) da pergunta para extração das informações que auxiliem na seleção das possíveis respostas, ou seja, das respostas candidatas à pergunta inserida (Almansa, 2017).

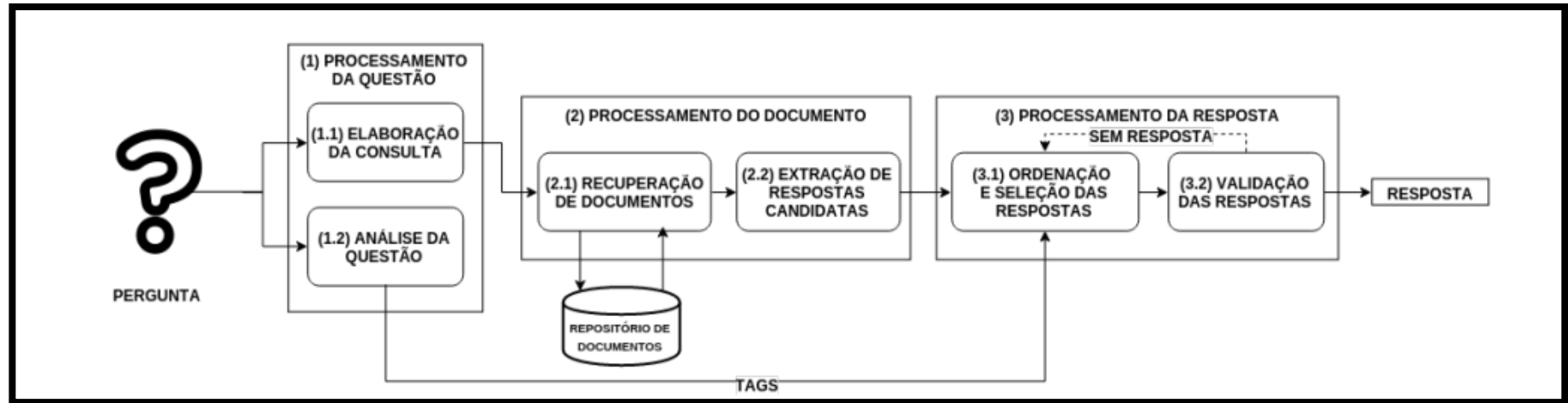
O processamento do documento consiste na segunda etapa, em que são feitas a recuperação dos documentos relevantes e a extração das respostas candidatas. Com isso, é executada a etapa final, processamento da resposta. “As respostas candidatas são ordenadas com base na similaridade entre a pergunta e a resposta candidata, apresentando o retorno ao usuário” (Almansa, 2017, p.16).

De maneira geral, os sistemas de perguntas e respostas são conhecidos por apresentarem somente uma única resposta para cada pergunta, valendo salientar que várias respostas podem ser ordenadas de acordo com a relevância para análise do usuário (Almansa, 2017).

---

<sup>5</sup> Secretaria de Estado da Fazenda (SEFAZ) é o órgão vinculado ao Ministério da Fazenda responsável pelo controle das receitas e das despesas de cada um dos estados e do Distrito Federal. Uma das principais obrigações cobradas por esse órgão para as empresas é o armazenamento das Notas Fiscais eletrônicas (NFes), atualmente pelo período de cinco anos.

A Figura 2 apresenta o esquema de uma arquitetura genérica para sistemas de perguntas e respostas com seus principais módulos:



**Figura 2** - Arquitetura genérica de um sistema de perguntas e respostas.

Fonte: Almansa, L. F. (2017). Uma arquitetura de question-answering instanciada no domínio de doenças crônicas. (p. 16). Recuperado de: <http://www.teses.usp.br/teses/disponiveis/95/95131/tde-10102016-121606/en.php>.



### **2.3.1 Processamento da pergunta**

Trata-se da primeira etapa e um dos seus objetivos é extrair e inferir informações extras para a pergunta fornecida. As informações extras são fundamentais para a seleção das respostas candidatas, como, por exemplo, o tema (nota fiscal eletrônica) ou o tipo da pergunta (uma definição ou múltiplas respostas). Outro objetivo que pode ser citado é a conversão da pergunta em linguagem natural para uma consulta de linguagem computacional, uma consulta para ser utilizada posteriormente na etapa de processamentos dos documentos (Almansa, 2017). A etapa de processamento da pergunta pode ser dividida em dois módulos: a) elaboração da consulta; b) análise da questão.

O módulo de análise da questão pode ser subdividido em tarefas, que são (Almansa, 2017): a) análise da questão; b) análise do foco da questão.

O módulo de análise da questão trata-se de um dos mais complexos e importantes no sistema de perguntas e respostas, porque uma pergunta possui informações explícitas e implícitas; entidades que compõem a sentença e análise de contexto e ambiguidade, respectivamente. Caso a pergunta seja classificada incorretamente, todas as etapas posteriores ficarão comprometidas e, conseqüentemente, a resposta final apresentada ao usuário (Almansa, 2017).

O foco da análise da pergunta realizada refere-se ao ponto central da questão, ao que ela se refere. Com isso, é possível eliminar ambigüidades que podem ocorrer na pergunta e ser mais assertivo na resposta. A classificação da pergunta é uma abordagem por correspondência de padrões baseado nas informações (Almansa, 2017).

Almansa (2017) exemplifica como ambigüidade: “qual a cor da manga do terno do presidente dos Estados Unidos durante seu discurso presencial no ano de 2008?”. A palavra “manga” pode fazer referência a uma fruta ou a uma parte da camisa.

Para evitar as ambigüidades e elaborar a consulta, um conjunto de palavras-chave é extraído para formar uma base de consulta computacional. Logo após essa

elaboração, a consulta gerada é transmitida para o processamento dos documentos disponíveis, na etapa de recuperação de documentos. Objetivando melhores resultados, é feita eliminação de *stopwords* (trata-se de palavras irrelevantes, por exemplo: as, e, os, de, para, com, sem, foi), reconhecimento das etiquetas nomeadas, etiquetador morfossintático de textos (*part-of-speech tagger*) e métodos complementares: “como, por exemplo, ontologias e os dicionários de sinônimos, podem ser utilizados na expansão do conjunto de palavras-chave” (Almansa, 2017, p.18).

### **2.3.2 Processamento dos documentos**

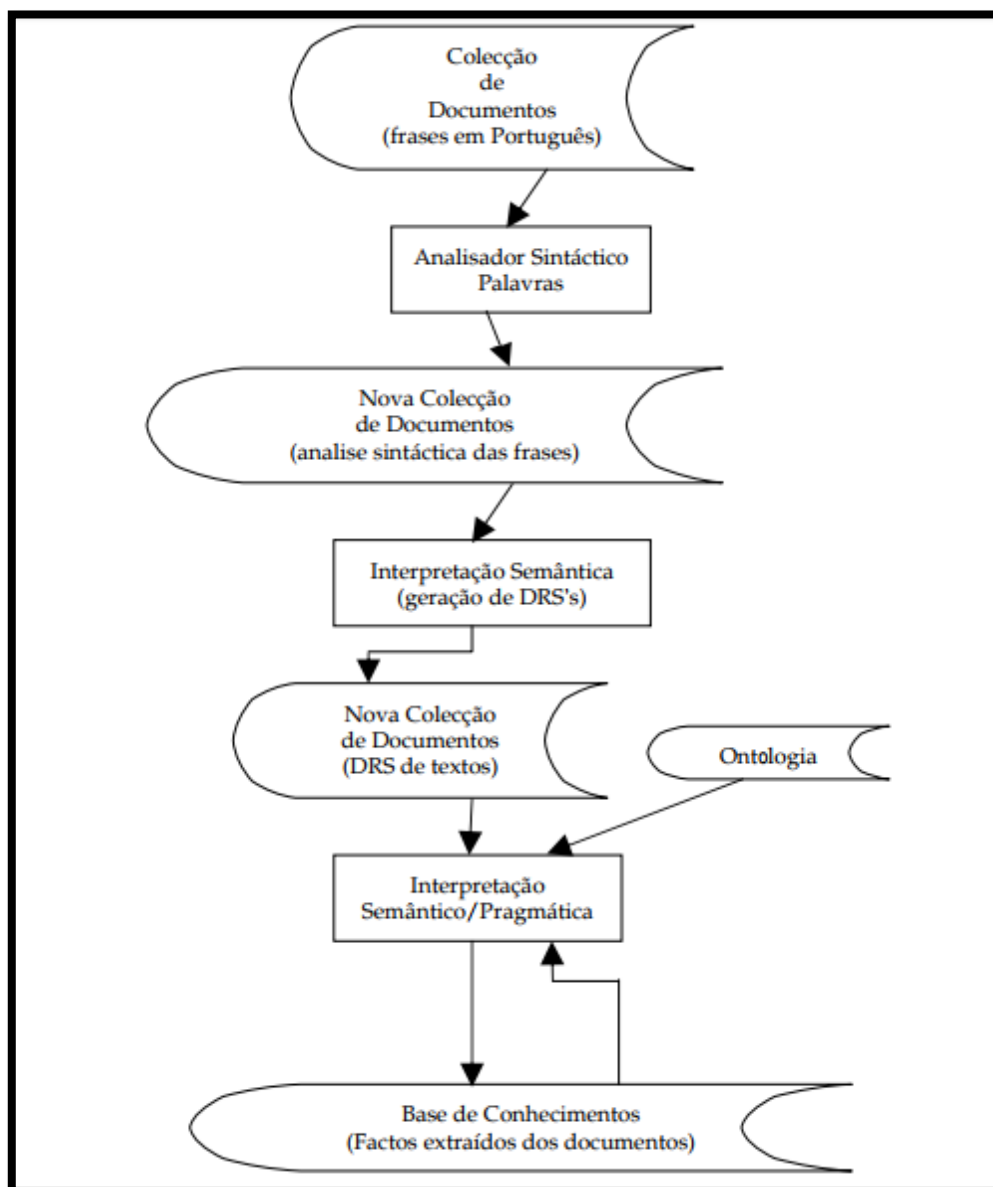
O processamento de documentos engloba desde a recuperação dos documentos candidatos até o tratamento e a extração das possíveis respostas. O usuário informa a pergunta, que é considerada como entrada de dados para a consulta criada. A recuperação de documentos possui como objetivo apresentar documentos precisos que contenham a resposta desejada (Almansa, 2017).

O rigor na recuperação dos documentos nos sistemas de perguntas e respostas é uma das diferenças entre os sistemas de recuperação da informação. Desejam-se, com os sistemas de perguntas e respostas, retornos objetivos, ao contrário dos “sistemas de RI que buscam por informações mais genéricas, apresentando como resultado documentos relevantes de forma ampla” (Almansa, 2017, p.19).

Localizar respostas precisas em documentos não é tarefas simples, porque as perguntas podem acrescentar níveis diferentes de complexidade e existem diferentes maneiras para analisar um documento em busca de respostas. Entre os métodos de pesquisa é possível citar: *tags* (encontradas na etapa de análise da questão e da resposta para direcionar quais artigos científicos serão recuperados e quais são as possíveis potenciais respostas associadas a esses artigos), árvores de análise sintática, grafos de dependência (Almansa, 2017, p. 20):

Proximidade linear (baseado na premissa de que as palavras-chave que representam a pergunta do usuário e a resposta candidata estão contidas em uma pequena quantidade N de parágrafos consecutivos). Caso a premissa se concretize na análise do documento, ele é retornado, caso contrário não.

A Figura 3 exemplifica o processamento de documentos de uma coleção, proposto por Quaresma *et al.* (2006):



**Figura 3** – Processamento de documentos da coleção – sistema QA

Fonte: Quaresma, P., Rodrigues, I., Prolo, C. A., & Vieira, R. (2006). Um sistema de pergunta-resposta para uma base de documentos. *Letras de Hoje*. Porto Alegre, 41(2), 47, jun. Recuperado de: <http://revistaseletronicas.pucrs.br/ojs/index.php/fale/article/view/594/425>.

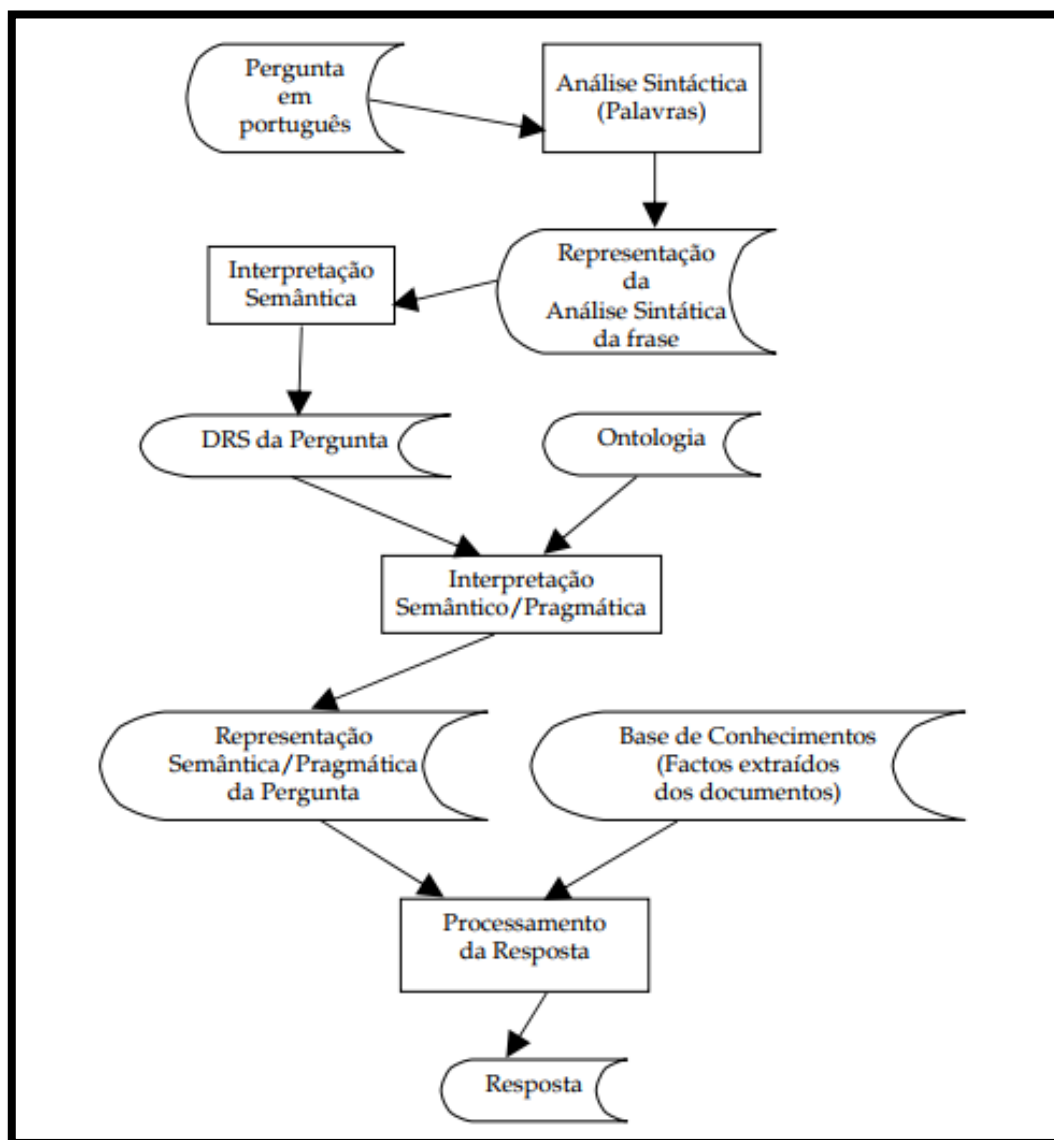
A coleção de documentos representa, para Quaresma *et al.* (2006), uma base de documentos não estruturados, como, por exemplo, a *web*. Os documentos pré-processados representam as bases de índices invertidos dos engenhos de busca. Segundo os autores, os sistemas de pergunta-resposta podem ser divididos em dois

módulos responsáveis pelo processamento do *corpus* e pelo processamento das perguntas: extração de informação e recuperação de informação.

O módulo de extração de informação (processa o *corpus* criando uma base de conhecimentos com a informação extraída dos documentos) e o módulo de recuperação de informação (processa a pergunta e retorna a resposta: um conjunto de palavras e a identificação do documento e da frase onde encontrou a resposta) seguem as seguintes etapas Quaresma *et al.* (2006):

- a) Análise Sintática: “as frases dos textos da coleção de documentos são processadas com analisador sintático de palavras” (Quaresma *et al.*, 2006). O processamento é feito documento a documento, criando-se um novo documento em que cada frase é substituída pela representação da sua análise sintática. Após essa fase tem-se uma nova coleção que tem o mesmo número de documentos da coleção inicial (Quaresma *et al.*, 2006);
- b) Análise Semântica: “a nova coleção de textos, em que cada frase foi substituída pela sua análise sintática, é reescrita” (Quaresma *et al.*, 2006) originando uma nova coleção de documentos na qual cada documento tem uma estrutura para representação do discurso (DRS), uma lista de referentes do discurso e um conjunto de condições (Quaresma *et al.*, 2006). “A partir da análise sintática constrói-se a estrutura do discurso, que representa a pergunta com os referentes da interrogativa marcados” (Quaresma *et al.*, 2006);
- c) Interpretação Semântica e Pragmática: nessa fase processa-se a coleção de documentos que se construiu na fase anterior. Algumas condições são reescritas e, tendo em conta a ontologia e o conhecimento já extraído, vai-se populando uma base de dados. Essa base de dados contém instâncias da ontologia e tabelas que representam predicados de aridade variável (Quaresma *et al.*, 2006);
- d) POrocessamento da Resposta: a representação final da pergunta é interpretada na base de conhecimentos com os fatos extraídos da coleção de documentos-alvo. Essa interpretação é feita a partir da unificação das entidades do discurso da pergunta com entidades do discurso dos documentos (Quaresma *et al.*, 2006).

Para o módulo de recuperação da informação Quaresma *et al.* (2006) apresentam o esquema da Figura 4:



**Figura 4** – Processamento da pergunta – sistema QA

Fonte: Quaresma, P., Rodrigues, I., Prolo, C. A., & Vieira, R. (2006). Um sistema de pergunta-resposta para uma base de documentos. *Letras de Hoje*. Porto Alegre, 41(2), 49, jun. Recuperado de: <http://revistaseletronicas.pucrs.br/ojs/index.php/fale/article/view/594/425>.

### 2.3.3 Processamento da resposta

Logo após o processamento da pergunta, é realizada a última fase, o processamento da resposta. As respostas candidatas são selecionadas e ordenadas de acordo com o seu grau de similaridade. Essa fase pode ser subdividida em: seleção e ordenação da resposta e validação da resposta. Não necessariamente apenas uma única

resposta é apresentada, podendo ocorrer várias respostas ordenadas, levando-se em consideração o grau de relevância (Almansa, 2017).

Para validar as respostas recuperadas podem ser utilizadas várias abordagens, por exemplo: ferramentas lexicais, por avaliação manual humana, de forma automática, por meio de *scripts* pré-programados, pela combinação e comparação dos resultados obtidos nas avaliações manuais e automáticas, entre outras. Alguns autores sugerem o uso combinado de avaliações efetuadas pelo sistema e por um colaborador e a comparação dos resultados. Uma das mais conhecidas é o *corpus* composto por duplas de perguntas e respostas (Almansa, 2017).

A *Conference and Labs of the Evaluation Forum* (CLEF), anteriormente chamada *Cross-Language Evaluation Forum*, por exemplo, é uma organização que oferece suporte à pesquisa no contexto multilíngue. Sua principal tarefa é elaborar repositórios de dados para o desenvolvimento de novas técnicas computacionais e propiciar aos pesquisadores participantes condições para a realização de experimentos de sistemas de recuperação de informações e sistemas de perguntas e respostas (Almansa, 2017).

As métricas utilizadas na avaliação dos sistemas de perguntas e respostas são: a precisão, a revocação e a *Fmeasure*<sup>6</sup>. Os valores dessas métricas são obtidos a partir dos dados encontrados pelo sistema e das análises realizadas pelo colaborador humano (Almansa, 2017).

### **2.3.4 Linhas de pesquisas**

O processamento da resposta é a última etapa do ciclo. Nela diversas técnicas podem ser aplicadas, desde metodologias superficiais, como o número de palavras em comum entre a pergunta e a frase de resposta e técnicas mais profundas, como algoritmos de aprendizado de máquina, que levam em conta aspectos sintáticos das frases de pergunta e resposta (Prestes, 2011). Burger, Cardie, Chaudhi, Gaizauskas,

---

<sup>6</sup> Média ponderada de precisão e revocação.

Harabagiu, Israel, Jacquemin... *et al.* (2001 como citado em Prestes, 2011), propõem 12 linhas de pesquisa para os sistemas de perguntas e respostas, que são:

**Quadro 2 - Linhas de pesquisa para sistemas de perguntas e respostas**

<b>Linhas de pesquisa</b>	<b>Significado</b>
<b>Classes de perguntas</b>	Compreendem os estudos: a) Taxonomia e da complexidade das perguntas; b) processamento de perguntas ( <i>a partir de bases de conhecimento e ontologias</i> ).
<b>Processamento de perguntas</b>	Estuda os modelos semânticos para o entendimento: Das perguntas, perguntas equivalentes, perguntas similares e perguntas ambíguas.
<b>QA e contexto</b>	As perguntas são feitas em um contexto, podendo servir para esclarecer a pergunta, resolver ambiguidades e manter a linha de raciocínio em uma série de perguntas.
<b>Fontes de dados</b>	Verifica as fontes de conhecimento para avaliar se são suficientes para responder as perguntas, podendo ser bases de dados, bibliotecas digitais, arquivos multimídia, entre outros.
<b>Extração da resposta</b>	Depende da complexidade da pergunta, do tipo de resposta (obtida no processamento da pergunta), da fonte de dados e foco da pergunta e do contexto. Compreende o estudo de métricas de avaliação, de completude, de correção e de justificativas para a resposta.
<b>Formulação da resposta</b>	A resposta deve ser formulada da maneira mais natural possível. Para isso, é preciso estudar como criar uma resposta com informações vindas de diferentes fontes, evitando informações sobrepostas e contraditórias.
<b>QA em tempo real</b>	Implementar sistemas que respondam em poucos segundos, independentemente da complexidade da pergunta, otimizando as etapas de busca dos documentos e a extração das respostas.
<b>QA multilíngue</b>	Desenvolver sistemas de perguntas e respostas que deem respostas em diversas línguas e busquem as respostas em textos escritos em línguas diferentes de qual pergunta foi elaborada.
<b>QA interativo</b>	Implementar modelos de diálogo que possuam detecção de intenções, objetivos e planos em comum com o usuário e resolução de correferências.
<b>Raciocínio avançado para QA</b>	Desenvolver sistemas que possam dar respostas fora do escopo de textos e bases de dados estruturadas, utilizando raciocínio e senso comum, além da capacidade de inferir fatos novos e montá-los para gerar uma resposta.
<b>Perfil de usuário</b>	Guardar informações do perfil de usuário a respeito do contexto, do domínio de interesse, esquemas de raciocínio normalmente utilizados, entre outros.
<b>QA colaborativo</b>	Estuda o desenvolvimento de modelos que possam detectar usuários no mesmo contexto e também modelos que buscam perguntas não relacionadas que têm respostas relacionadas, e vice-versa.

Fonte: Burger *et al.* (2001 como citado em Prestes, K. V. (2011). Avaliação de métodos de seleção da resposta em um sistema de perguntas e respostas. (p. 17-18). Recuperado de: <https://www.lume.ufrgs.br/bitstream/handle/10183/36881/000819142.pdf?sequence=1>.

### 2.3.5 Recursos para auxílio dos sistemas de perguntas e respostas

Prestes (2011) enfatiza que os sistemas de perguntas e respostas de maneira geral são complexos e podem utilizar recursos externos para obter ou melhorar os resultados, como, por exemplo:

- a) **Ontologias:** é um recurso de grande valor e de ampla cobertura, principalmente para área de inteligência artificial e processamento de linguagem natural. Trata-se de um modelo de dados que descreve um conceito dentro de um domínio específico (Prestes, 2011);
- b) *Palavras Associadas Porto Editora Linguatca (PAPEL)*: é um conjunto de relações entre palavras da língua portuguesa, identificando relações entre relações composicionais, hierárquicas e de sinonímia (Prestes, 2011);
- c) **Parser:** realiza análise sintática de um texto para determinar a sua estrutura gramatical (Prestes, 2011);
- d) **WordNet:** é uma ontologia utilizada em sistemas de domínio aberto voltado para língua inglesa. Trata-se de uma base de dados léxicos que possui palavras organizadas em *synsets* (conjunto de sinônimos) com suas respectivas relações: hiperonímia, hiponímia, meronímia e holonímia<sup>7</sup> (Prestes, 2011);
- e) **Algoritmos de seleção da resposta:** trata-se de procedimentos que auxiliam na solução de problemas na localização das respostas, como, por exemplo, *bag-of-words*, *n-grams*, distância *keywords*, *Google score*, *Wikipedia Score*, aprendizado de máquina J48 e *Support Vector Machines (SVM)* (Prestes, 2011);
- f) **Similaridade *bag-of-words*:** algoritmo simples que atribui um escore (pontos que representam a similaridade entre a pergunta e a resposta candidata) para

---

<sup>7</sup>**Hiperonímia:** relação semântica entre uma palavra de significado mais geral ou abrangente e outra ou outras com significado mais específico em relação à primeira = superordenação.

**Hiponímia:** relação semântica entre uma ou mais palavras e um significado mais específico e outra com significado mais geral ou abrangente em relação à(s) primeira(s).

**Meronímia:** relação semântica entre uma palavra que tem o significado de uma parte e outra com significado de um todo em relação à primeira.

**Holonímia:** relação semântica entre uma palavra que tem o significado de um todo e outra com significado de uma parte em relação à primeira.



- uma resposta candidata de acordo com o número de palavras em comum com a pergunta (Prestes, 2011);
- g) **N-Grams**: é utilizada quando a simples correspondência de palavras não é o suficiente para que uma das respostas seja escolhida. Estabelece uma similaridade mais forte entre a pergunta e a resposta - a similaridade de *n-grams* estudada (Prestes, 2011);
- h) **Google Score**: utilizado para gerar um escore numérico. Uma consulta consistindo da resposta candidata e das palavras-chave da pergunta é enviada ao *Google*, e os 10 primeiros *snippets* retornados são analisados utilizando o algoritmo 1, que calcula a distância entre as palavras-chave da pergunta e a resposta candidata (Prestes, 2011);
- i) **Algoritmos de aprendizado de máquina**: trata-se de algoritmos que são treinados para classificar o tipo de pergunta a partir de um *corpus* anotado, por exemplo, WEKA, que é um *software* de código aberto que agrega diversos algoritmos de aprendizado de máquina utilizando inteligência artificial: árvores de decisão e algoritmos bayesianos (Prestes, 2011);
- j) **Agrupamento de respostas**: na extração das respostas da *web* ou de documentos encontram-se diversas respostas iguais ou equivalentes, que se referem à mesma entidade, como, por exemplo, a lista de respostas candidatas para a pergunta “quem foi a primeira mulher no espaço?”. A resposta inclui Valentina Vladimirovna Tereshkova e Valentina Tereshkova, que se referem à mesma pessoa e estão ambas corretas. Sendo assim, criou-se um método para agrupar essas respostas equivalentes e explorar essa característica da extração das respostas;
- k) Distância das *keywords*, TF-IDF (termo-inverso da frequência nos documentos), agrupamentos de respostas, entre outros.

### 3 Trabalhos Relacionados

Os estudos sobre sistemas de perguntas e respostas estão presentes em diversas áreas do conhecimento, figurando entre eles:

Para Almansa e Macedo (2016), no ambiente médico a busca por informações de alta importância é tarefa que exige tempo e experiência do profissional para localizar respostas rápidas. O objetivo do trabalho de Almansa e Macedo (2016) foi de apresentar uma arquitetura de um sistema de perguntas e respostas específico para doenças crônicas, utilizando módulos de processamento das perguntas, processamento dos documentos e o processamento da resposta, ainda em fase de avaliações.

A respeito do *chatbot* Tical, desenvolvido para divulgar o atlas linguístico do Brasil utilizando processamento de linguagem natural e a ferramenta *WhatsApp*, Moreno, Manfio, Barbosa, & Brancher (2015) relatam que ele responde às perguntas relacionadas ao atlas linguístico brasileiro a partir de ferramenta de comunicação instantânea.

A interação da comunicação de voz entre o ser humano e o computador, principalmente em motores de busca e seu progressivo aumento, é reportada por Abdul-Khader. & Woods (2015). Os *chatbots* são indicados como foco principal, com suas técnicas de *design*, interatividade e bate-papo. Concluiu-se no artigo que falta uma solução mais comum e que há a necessidade constante de pesquisas para a criação de uma abordagem usual, apesar dos desacordos comerciais.

Souza e Moraes (2006) apresentam a viabilidade da construção de uma base de conhecimento *Artificial Intelligence Markup Language* (AIML) para o *chatbot*, reduzindo esforço humano e descrevendo as dificuldades encontradas para essa modalidade. O artigo realça a intensificação dos *chatbots* nos últimos tempos em diversas aplicações: atendimento ao cliente, *help desk*, navegação de *sites*, suporte

técnico, educação, entre outros, e em contrapartida os possíveis desafios para sua elaboração mediante as variantes linguísticas da língua portuguesa.

Moraes & Machado (2016) propuseram uma arquitetura para um *chatbot* educacional baseado na metodologia *Formal Concept Analysis* (FCA), utilizada para análise de dados, representação do conhecimento e gerência de informações. O artigo descreve a dificuldade de localizar artigos e material didático sobre a técnica. A proposta é apresentada como promissora por explorar as características da FCA e visa realizar testes em estudantes de graduação e posteriormente estender o estudo em outras áreas de conhecimento.

O *chatbot* Prof<sup>a</sup> Elektra foi desenvolvido baseado no *chatbot* Alice e retrata a educação à distância e a implementação do *chatbot* educacional para o ensino das disciplinas de Física e rede de computadores (Leonhardt, Castro, Dutra & Tarouco, 2003).

A proposta de “assistentes pessoais” para auxiliar no esclarecimento de dúvidas utilizando processamento de linguagem natural, padrões morfossintáticos a alimentação de bases de conhecimentos é exposta por Bada e Menezes (2012). As respostas aos questionamentos de dúvidas serão conduzidas por meio de *chatbot* do tipo Alice com o apoio de ontologia e bases de conhecimentos. Com isso, o artigo visa auxiliar os professores e estudantes em ambientes virtuais de aprendizado, diminuindo o tempo de espera pelo esclarecimento da dúvida.

O estudo sobre o *chatbot* TicTad auxilia alunos no aprendizado da linguagem C#. Ele colabora na memorização de funções dessa linguagem de programação de maneira interativa, segundo Kasinathan, Vinothini, Mustapha, Siow & Hopman (2018).

Wailthare, Gaikwad, Khadse & Dubey (2008) sugerem diversos sistemas de *chatbot* em domínios variados, como educação, negócios, recuperação de informação, *e-commerce*, entre outros. Trata-se de um avanço tecnológico que tende a contribuir muito e ao longo do tempo interagir em conversas em geral.

#### 4 Softwares de Perguntas/Respostas Relacionados

Os trabalhos relacionados aos sistemas de perguntas e respostas colaboraram para a elaboração de *softwares* em diferentes áreas de atuação, entre eles podendo citar os que se seguem:

O programa **STUDENT** em LISP, desenvolvido em 1964 com o intuito de resolver problemas de álgebra que possuía reconhecimento de linguagem (Varga, 2014). O sistema **BASEBALL**, que lia e respondia questões escritas por meio de cartões perfurados sobre dados relacionados aos jogos de beisebol da liga americana. As informações foram coletadas durante o período de uma temporada. Para realizar a consulta nas informações coletadas, o sistema convertia a questão em uma consulta ao banco de dados (Almansa, 2017).

O programa **ELIZA**, desenvolvido em 1966 com base de conhecimento interno e que realizava reconhecimento de linguagem natural respondendo as perguntas dos usuários com respostas pré-programadas e repassando a impressão de que se estava falando com uma pessoa comum porque utilizava correspondência de padrão de palavras (Varga, 2014). O sistema **SHRDLU**, de inteligência artificial, capaz de fazer planos e manter conversas simples sobre um conjunto de objetos geométricos (Almansa, 2017).

O sistema **LUNAR**, que utilizava como fonte de informação dados sobre análises geológicas de amostras de pedras recolhidas durante a missão de Apollo até a lua. O sistema transformava as perguntas feitas pelo usuário em consultas de banco de dados e retornava a resposta (Almansa, 2017). O programa **QUERIX**, que adota os mesmos procedimentos do NLPReduce, ou seja, é baseado no formalismo da *web* semântica, no entanto, usa perguntas completas, técnicas de *POS-tagging* e incorpora um método interativo com o usuário para a resolução de ambiguidades (Silva, 2011).

O programa **POWERANSWER**, que utiliza sete camadas hierárquicas (*concepts; semantic relation; contexts; event structure; event relation; macroevent*) para a

representação de conhecimento a partir de textos. Sua arquitetura é formada por componentes semânticos organizados nos três módulos:

Módulo para resolução de ambiguidades; bases de conhecimento baseados em *WordNet*, *EventNet* e *Wikipedia*; módulo de identificação de relações (temporais ou não) entre conceitos e eventos com base no domínio definido pela ontologia; módulo de criação e manutenção de ontologias a partir de textos (chamado de Jaguar); mecanismo de inferência (denominado Cogex) para identificação de respostas e raciocínio (Moldovan, Tatu & Clark, 2009 como citados em Silva, 2011, p. 62).

O sistema **MYCIN**, que possui especialidade na área da Medicina e foi desenvolvido na década de 70 para auxiliar profissionais da área da saúde na identificação dos agentes causadores de infecções como, por exemplo, a meningite, e na recomendação de medicamentos antibióticos com doses ajustadas para o paciente em tratamento (Almansa, 2017).

O sistema **UNIX Consultant**, que é programado para auxiliar usuários inexperientes a adquirir conhecimentos sobre o sistema operacional UNIX. O sistema utilizava abordagens de processamento de linguagem natural, uma base de conhecimento abrangente e um módulo que buscava adaptar as respostas para que maior número de usuários conseguisse entender o conteúdo da resposta (Almansa, 2017).

O sistema **QuestIO** (do acrônimo de *Questionbased Interface to Ontologies*): atua sobre domínios fechados definidos por ontologias para converter a pergunta em linguagem natural em uma consulta SPARQL, que é executada sobre a base de conhecimento para que a resposta que envolva os conceitos do domínio seja retornada. As consultas são executadas diretamente sobre o conjunto de instâncias do modelo da ontologia; sendo assim, os dados precisam ser carregados das fontes para o modelo (Silva, 2011).

O programa **ALICE**, desenvolvido em 1995, que utiliza *Artificial Intelligence Markup Language* (AIML) e simula uma conversação com o usuário (Varga, 2014). O programa de perguntas e respostas *SynTactic Analysis using Reversible Transformations* (**START**): instalado na *web*, foi o pioneiro em 1993. Possui base de conhecimento com várias fontes da *web* e realiza tratamento da informação estruturada e não estruturada (Varga, 2014);

O programa *Portable nAtural laNguage inTerface to Ontologies* (**PANTO**): formaliza as perguntas em uma estrutura semelhante ao modelo de triplas de *Resource Description Framework* (RDF), que é comparada ao modelo da ontologia de domínio. “A ontologia é utilizada por componentes da arquitetura principalmente para reconhecimento de entidades; determina as projeções, filtros e relacionamentos entre conceitos para extrair a resposta a partir de consultas em SPARQL” (Silva, 2011, p. 64).

O sistema **ORAKEL**, que é baseado em ontologias para criar uma estrutura formal de representação da pergunta e para obter respostas factuais diretas ou ainda derivá-las por meio de processos de inferências a partir de bases de conhecimento. Utiliza consulta SPARQL e F-Logic para obter os fatos e conceitos a partir da interpretação da pergunta (Silva, 2011).

O serviço de perguntas e respostas **QSABE**, que cataloga de forma dinâmica o perfil de especialistas do conhecimento (colaboradores) e atua como um distribuidor inteligente de mensagens, utiliza metáforas de ambientes *groupwares*, teorias e técnicas de compreensão do texto (Andrade, Nardi, Pessoa & Menezes, 2003; Menezes, Tavares & Pessoa, 1998).

O sistema **FREyA**, que realiza a tradução da “pergunta em linguagem natural ou palavras-chave em uma consulta SPARQL2 e retorna a resposta para o usuário após executar uma consulta na ontologia”. Sua dinâmica, de maneira resumida, é identificar e verificar os conceitos da ontologia, gerar a consulta SPARQL e identificar o tipo da resposta e apresentar o resultado para o usuário (Amorim *et al.*, 2012, p.2).

O sistema **SEMANTIC CRYSTAL**, que permite ao usuário manipular e selecionar os conceitos e relacionamentos da ontologia graficamente. Possui interface em que “as consultas podem ser criadas de modo interativo a partir do modelo da ontologia, em vez de uma entrada em linguagem natural” (Silva, 2011, p. 64).

O sistema **POWERAQUA**, trata-se de uma:

Evolução de outro sistema chamado *Aqualog*, baseado em ontologia. Na arquitetura do *PowerAqua*, o componente análise da questão utiliza um componente linguístico para processar a consulta e sua saída é um conjunto de triplas linguísticas (sujeito, predicado, objeto) que é mapeado para a consulta do usuário (Amorim *et al.*, 2012, p.3).

O sistema **OMeD**, que se apoia na tomada de decisões médicas por meio das quais respostas são fornecidas em tempo real para perguntas médicas. Os componentes centrais do OMeD são: componente de interface (que recebe consultas em linguagem natural e realiza busca da informação); componente de representação do conhecimento (que agrega e traduz informações de um específico cenário para uma representação semântica adequada ao uso do OMeD); o componente de raciocínio semântico (que deriva as respostas para uma consulta do usuário por meio de um raciocínio sobre conhecimento médico relevante). Em um dos experimentos realizados utilizando cinco conjuntos de dados contendo 1.000 registros de pacientes e 20, 30, 40, 50 e 60 de medicamentos (metade foi utilizada como dados de treino), foram produzidos 100% de respostas corretas (Amorim *et al.*, 2012, p.3).

O sistema **Personalized QA Framework**, que se baseia em respostas factuais em estudo de caso no contexto de *Business Intelligence* (BI) e utiliza a ontologia como forma de anotação semântica sobre documentos, classificação e expansão de perguntas e recuperação de informação sobre índices do motor de busca Lucene (Silva, 2011).

O programa **NLPREDUCE**, que é baseado no formalismo (RDF, OWL, SPARQL) proposto por Kaufmann e Bernstein (2007 como citado em Silva, 2011), permite que usuários entrem com palavras-chave, fragmentos de sentença ou perguntas completas e emprega sinônimos e técnicas de *stemming* para refinar essa entrada. Possui triplas (sujeito, predicado, objeto) com as relações dos termos da pergunta, tal como no modelo RDF, e comparadas via consultas na base de conhecimento por meio de SPARQL (Silva, 2011).

O sistema **WATSON**<sup>8</sup>, de perguntas e respostas de autoria da *International Business Machines* (IBM), projetado para desenvolver a ciência do processamento de

---

<sup>8</sup> Em homenagem a Thomas J. Watson, fundador da IBM.

linguagens naturais a partir de avanços na tecnologia de perguntas e respostas. Ele foi otimizado para uma carga de trabalho baseada na arquitetura IBM DeepQA (desenvolvida usando a Apache UIMA), arquitetura probabilística paralela maciça baseada em evidências. Para que a arquitetura se tornasse robusta foram feitas adaptações para diferentes aplicações comerciais e outros problemas, inclusive nas áreas de Medicina, pesquisa corporativa e jogos (IBM, 2011).

O protótipo **PERGUNTE!**, de um sistema em português que utiliza a *web* como base de dados e utiliza padrões superficiais de texto para localizar respostas nos documentos retornados pelos engines de busca. Sua implementação, segundo os autores, apresentou alta taxa de acerto nos experimentos realizados (Rabelo & Barros, 2005).

O programa *Guided Input Natural Language Search Engine* (**GIONSENG**): aplica uma linguagem controlada (forma que impõe vocábulos específicos e limita o uso de linguagem natural) baseada em menus sobre bases de conhecimento em OWL. Trata-se de um sistema baseado no formalismo utilizando métodos de reformulação e expansão para sugerir complementações durante a entrada da pergunta do usuário, permitindo anotação semântica dinâmica para adição e sinônimos dos conceitos da ontologia. O sistema traduz as perguntas em consultas SPARQL, que são executadas diretamente sobre as bases de conhecimento para encontrar as respostas (Silva, 2011).

O sistema **MEANS**, que é apoiado por abordagens semânticas compostas por técnicas de processamento de linguagem natural e *Web Semântica*. Utiliza análise de dados em dois níveis de complexidade: primeiro, busca por entidades médicas como, por exemplo, drogas, sintomas e doenças, e logo após realiza em nível de relacionamento tais como tratamentos, prevenções e causas (Almansa, 2017).

O sistema *Shallow Question Answering System* (**SQAS**), que é concentrado nos domínios de textos jornalísticos do idioma português, com resultados precisos na busca pelos parágrafos candidatos à resposta e na filtragem de alguns modelos de perguntas (Almansa, 2017).



O sistema **COMUNICA**, que possui a capacidade de buscar informações em banco de dados estruturados e não estruturados. Composto por módulos de reconhecimento e sintetização da fala, um módulo de processamento de texto, um módulo administrador responsável por gerenciar os demais módulos e um módulo de acesso aos bancos de dados. De acordo com os autores, os módulos de apoio à fala ainda estão em construção. Utiliza técnicas de processamento natural de linguagem e ontologias nos módulos de processamento textual (Almansa, 2017).

O sistema **ESFINGE**, que pergunta em português, elabora as respostas automáticas e calcula seus resultados obtidos. O sistema permite o estudo na área da Língua Portuguesa e apresenta uma panorâmica de análises (Costa, 2005).

## 5 Contextualização do Ambiente

A contextualização do ambiente apresenta informações sobre a tecnologia da informação no varejo e na empresa objeto de estudo, visto que a análise surgiu baseada na necessidade de uma rede de supermercados de localizar respostas relativas a dúvidas de sistemas para os clientes internos. A empresa investe em tecnologia de várias formas (maquinário, sistemas, desenvolvimento interno, infraestrutura, entre outros), no entanto, em uma de suas vertentes por busca de informações deparou-se com a necessidade de melhorias.

### 5.1 Tecnologia da informação no varejo

Segundo Santos (2005 como citado em Silva Júnior, Souza e Manhães, 2014, p. 106), empresas do setor de varejo, “assim como fornecedores e prestadores de serviço, estão cada vez mais interessadas em analisar as variáveis relativo ao negócio, visando gerenciar cada vez melhor utilizando os recursos tecnológicos”.

Na década dos anos 90, como preceituam Silva Júnior *et al.* (2014), a automação no setor varejista no Brasil acelerou-se e apresentou várias evoluções. Angelo e Silveira (2001 como citados por Silva Júnior *et al.*, 2014, p. 106) afirmam “maior dinamicidade no atendimento ao cliente através dos *check outs*” (ponto de venda, - PDV -, terminal utilizado pelo vendedor/caixa, em que o cliente entrega suas compras para registrar e apurar os valores e efetuar o pagamento da compra), mais confiabilidade de disponibilidade das informações, permitindo, assim, tomadas de decisões rápidas em menos tempo devido a respostas e retornos ágeis.

No setor varejista, há cerca de 20 anos, Santos e Costa (1997) já relatavam sobre a reestruturação das empresas em segmentos diversos visando à adequação ao cenário de competitividade mais acirrada. A disputa pelo cliente e por sua fidelidade provocou mudanças no perfil varejista. No entanto, não há um formato ideal e sim uma busca por eficiência no negócio (Santos & Costa, 1997).

O varejo integra funcionalidades clássicas de operação comercial como, por exemplo, procura e seleção de produtos, aquisição, distribuição, comercialização e entrega,

tendo como principais tipos de lojas os departamentos: tradicionais, de descontos, de vestuário, varejo de alimentos, eletrodomésticos e eletrônicos. Em todas as funções, a tecnologia da informação pode auxiliar, melhorando o controle e automatizando tarefas (Santos & Costa, 1997). Os supermercados fazem parte do segmento varejista.

O volume de informações e o processamento de dados exigem a participação da tecnologia para auxiliar nos controles, nas análises e nos processos gerenciais. Trata-se de um ramo de atividade que está em constante ampliação.

A Associação Brasileira de Supermercados (ABRAS<sup>9</sup>), de acordo com o *ranking* ABRAS/SuperHiper, acentuou que as 500 maiores empresas no segmento supermercadista tiveram aumento em todos os seus subgrupos. O aumento, conseqüentemente, influenciou no faturamento total de 2016. O somatório das 500 empresas movimentou, no ano, R\$ 265,5 bilhões, expansão nominal de 9,8% e real de 1%, mesmo diante de problemas econômicos no país. Veja a síntese dos resultados referentes a 2016 para a estrutura varejista brasileira na Figura 5.

Síntese dos resultados - 2016					
	Fat. Nominal (R\$ bilhões)	Nº de lojas	Nº de check-outs	Área de Vendas m <sup>2</sup> (milhões)	Número de funcionários***
Setor autosserviço*	338,7	89.009	225.052	21,7	1.809.852
Segmento Supermercado*	314,7	37.685	172.408	16,5	1.667.077
Ranking (500 Empresas)**	265,5	7.829	69.890	9,2	683.614
300 Maiores Empresas**	259,1	7.396	66.864	8,8	662.040
20 Maiores**	180,1	5.110	39.102	5,6	403.578
280 Outras**	79,0	2.286	27.762	3,2	258.462

\*Estrutura do Varejo Brasileiro - Nielsen \*\*Tabela eficiência 500 \*\*\*Rais/Caged

**Figura 5** – Resultado varejo brasileiro 2016 - *ranking* ABRAS  
 Fonte: Associação Brasileira de Supermercados. ABRAS. (2017). *Síntese dos resultados 2016*. [Figura]. Recuperado de: <http://www.abras.com.br/economia-e-pesquisa/ranking-abras/as-500-maiores/>.

<sup>9</sup> Entidade criada em 11 de novembro de 1968; atua em rede, de forma participativa com as 27 Associações Estaduais de Supermercados afiliadas, com foco na evolução das lojas; no estímulo ao saudável intercâmbio com os fornecedores; além do esforço dirigido ao melhor atendimento aos consumidores e à evolução do mercado de consumo no país. Recuperado de: <http://www.abras.com.br/abras/>

A automatização comercial e o processamento automático de informações melhoram o desempenho da gestão e possibilitam de maneira geral redução de custos, de tempo de operação e de quantidade de mão de obra. O diferencial da informatização não é apenas a geração da informação e sim a capacidade de transformá-la em ações antecipadas em prol de melhores resultados (Santos & Costa, 1997).

Consta no *site* Nielsen<sup>10</sup> (2012) que o mercado vive o momento dos avanços tecnológicos e digitais e que a eficiência exige um “verdadeiro banco de dados” em que se considera que a informação tem papel de alta relevância nas empresas e que deve servir de apoio aos executivos para compreender o negócio, traçar estratégias, além do posicionamento no mercado, sortimento, preço, entre outros. Por esse motivo, o preço e o prazo não são informações únicas a serem monitoradas e os dados devem ser confiáveis para não comprometerem as tomadas de decisões (Nielsen, 2012).

Mediante esses dados, é necessário preparar-se para recuperar as informações e responder a questionamentos com assertividade.

### **5.1.1 Sistemas para gerência de informações do frente de caixa**

O termo “frente de caixa” ou “frente de loja” designa os processos e atividades ligadas de forma direta às vendas como, por exemplo, informações sobre o cliente, formas de pagamento, serviço de entrega e emissão de cupons fiscais e cheques (Carvalho & Galeale, 2006).

A automação da frente de caixa ou frente de loja busca eficiência nos processos de venda, atendimento ao cliente, além de outros serviços ligados ao *check out*, caixa e saída do cliente da loja (Carvalho & Galeale, 2006). Nela estão concentradas as

---

<sup>10</sup> A Nielsen é uma empresa germânico-americana com sede em Nova Iorque, nos Estados Unidos, que oferece diversas informações do mercado, utilizando metodologias próprias. A empresa realiza estudos sobre os consumidores em mais de 100 países visando oferecer uma visão mais detalhada e completa das tendências e dos hábitos dos consumidores ao redor do mundo. A Nielsen realiza medições e pesquisas sobre o varejo, consumo e o mercado editorial. No Brasil, atua há mais de 40 anos.

tecnologias mais específicas ao negócio supermercadista e ao varejo, podendo-se destacar:

- **PDV:** é a sigla correspondente a ponto de venda, que compreende toda TI relacionada diretamente ao processo de efetivação da venda, localizada no *check out* do supermercado, o caixa, como microcomputador, impressora fiscal, *scanner*, *pin-pad*, TEF, *software* de frente de caixa e balança eletrônica quando acoplada no caixa, entre outros (Carvalho & Galegale, 2006, p.4).
- **Emissor de Cupom Fiscal (ECF):** é uma impressora desenvolvida para a emissão de cupons fiscais, de acordo com a legislação, e sua principal diferença de uma impressora comum é que possui uma memória interna (EPROM) lacrada, onde são acumulados e armazenados os valores totais de vendas e outras informações de tributações, para fins de fiscalização (Carvalho & Galegale, 2006, p.4).
- **Memória de fita detalhe (MFD):** tecnologia acoplada aos novos modelos de ECF, que permite o armazenamento eletrônico de todas as informações referentes aos cupons fiscais, não apenas das totalizações, como no ECF tradicional. Desta forma, elimina a necessidade de uma segunda via de bobina que continha os dados dos cupons e tinha que ser guardada para efeito de fiscalização por cinco anos e auxilia o fisco, que já obtém as informações eletronicamente (Carvalho & Galegale, 2006, p.4).
- **Transferência Eletrônica de Fundos (TEF):** compreende o conjunto de componentes de *hardware*, *software* e comunicação necessários para a integração do PDV (através de seu sistema de frente de caixa) com sistema bancário, sendo capaz de realizar transações com cartões de débito e crédito (Carvalho & Galegale, 2006, p.4).
- **Pin-Pad:** teclado compacto e com leitor de cartões, que permite a leitura de cartões de débito ou crédito, enviando os dados destes e suas respectivas senhas, através de um sistema de criptografia (sistema de segurança). É indispensável para a adoção de TEF. Os modelos mais atuais também fazem a leitura de *Smart Cards*, cartões inteligentes, XIII SIMPEP – Bauru, SP, Brasil, 06 a 08 de novembro de 2006. Com processador e capacidade de armazenar dados de identificação ou de perfil de consumo, por exemplo (Carvalho & Galegale, 2006, p. 4).
- **Balança eletrônica:** equipamento eletrônico utilizado para a pesagem de produtos, que pode ser interligada diretamente ao PDV, que captura o peso ou transmite para uma impressora de termotransferência, que irá gerar uma etiqueta com um código de barras capaz de ser lido pelo *scanner* na frente de caixa (Carvalho & Galegale, 2006, p. 5).
- **Terminal de consulta de preço (TCP):** dispositivo eletrônico dotado de visor e leitor de código de barras, em alguns casos com memória própria, que possibilita ao cliente, através da passagem do código de barras no leitor, a visualização do preço do produto no visor (Carvalho & Galegale, 2006, p.5).
- **Software de frente de caixa:** sistema de informação voltado ao nível operacional, responsável pelo registro e processamento de transações relacionadas ao processo de vendas, como a emissão do cupom fiscal, forma de pagamento, consulta de preços e de crédito do cliente, emissão de cheques, dentre outros, normalmente interligado ao sistema de retaguarda e TEF (Carvalho & Galegale, 2006, p.5).
- **Monitor tipo touch-screen:** dispositivo onde a tela do monitor de vídeo é sensível ao toque, conjugando em um mesmo equipamento funções de entrada e saída de dados (Carvalho & Galegale, 2006, p. 5).
- **Identificador biométrico:** dispositivo de reconhecimento de imagens, semelhante ao leitor de código de barras, capaz de converter uma impressão digital de uma pessoa, por exemplo, em formato que possa ser interpretado pelo computador (Carvalho & Galegale, 2006, p. 5).

### 5.1.2 Sistemas para gerência de informações do retaguarda

O sistema de frente de caixa registra os dados, no entanto, sua gerência é feita por parte dos sistemas de retaguarda. O termo “retaguarda” no supermercado compreende de maneira geral vários setores: financeiro, contabilidade, estoque, reposição, gestão de custos, gestão de preço de venda, recursos humanos, entre outros departamentos, no caso de empresas de maior porte. Sua automação trata de processos e atividades relativos aos setores citados, visando informações gerenciais (Carvalho & Galegale, 2006).

Silva Júnior *et al.* (2014, p. 106-107) afirmam que dois instrumentos têm colaborado para mais eficiência do serviço prestado ao consumidor, tendo o *Enterprise Resource Planning* (ERP) (Sistemas Integrados de Gestão) como a plataforma para seu funcionamento, são eles:

- a) **EDI** (*Electronic Data Interchange* – Transferência eletrônica de informações): segundo Porto *et al.* (2000), o EDI surgiu como “um tipo de tecnologia de informação com a capacidade de estreitar o relacionamento entre empresas. Pode ser definido como um fluxo eletrônico e padronizado de dados entre empresas, permitindo melhores desempenhos, tanto em termos operacionais quanto estratégicos”.
- b) **ECR** (*Efficient Consumer Response* – resposta eficiente ao consumidor): consiste em uma estratégia utilizada principalmente na indústria varejista, na qual distribuidores e fornecedores trabalham em conjunto para proporcionar maior valor ao consumidor final (Lavratti *et al.*, 2002, p.1). “O ECR é considerado como uma cadeia de abastecimento de produtos de mercadorias no início do século XXI, o que o *Just in Time* significou para a indústria automobilística japonesa na crise petrolífera dos anos 70 e o *Quick Response* para as indústrias de vestuário e têxtil (Ghisi, 2001, p. 5) (Silva Júnior *et al.*, 2014, p. 106-107)”.

Além do EDI e do ECR, existem outros fatores que contribuíram para os avanços tecnológicos que auxiliaram o setor varejista durante suas mudanças de adaptação no mercado, entre eles a consolidação do ERP com inúmeros módulos integrados (com funcionalidades operacionais de finanças, custos, recursos humanos, vendas, logística, distribuição, produção, entre outros incorporados) e tecnologias utilizadas em supermercados, destacando-se:

- **Microcomputador**: também conhecido como PC – *personal computer* (computador pessoal), *desktop* (computador de mesa). É um equipamento eletrônico dotado de uma unidade central de processamento (processador ou [Unidade Central de Processamento] CPU), memória de armazenamento primário (RAM/ROM [*Random Access Memory/ Read-Only Memory*]),

memória de armazenamento secundário (discos magnéticos – [*higt disk*] HD, discos ópticos, fitas magnéticas), periféricos de entrada (teclado, *mouse*, *scanner*), periféricos de saída (impressora, monitor de vídeo, áudio), dispositivos de comunicação (placas de rede, modem), com capacidade de processar dados, realizar instruções e operações (Carvalho & Galegale, 2006, p. 3).

- **Notebook:** microcomputador portátil que traz como principal característica a integração e miniaturização da maior parte dos componentes, tornando-o leve e de pequenas dimensões. Conhecido também como *laptop* e *desknote*. Muitos *notebooks*, hoje, têm capacidades de processamento similares aos microcomputadores *desktops* (Carvalho & Galegale, 2006, p. 3).

- **HandHeld:** microcomputador de mão, também conhecido como [*Personal Digital Assistan*] PDA, *Pocket-PC* ou *Palm-Top*. Equipamento portátil desenvolvido para servir como dispositivo de acesso, apesar de alguns modelos possuírem uma grande capacidade de memória e de processamento (Carvalho & Galegale, 2006, p.3).

- **Scanner ou leitor de códigos de barras:** dispositivo de reconhecimento de caracteres, capaz de converter o código de barras em formato digital, que pode ser interpretado pelo computador, em geral funciona com leitura óptica ou [*light amplification by stimulated emission of radiation*] laser (Carvalho & Galegale, 2006, p.3).

- **Coletor de dados:** microcomputador de mão, similar a um *handheld*, com dimensões e capacidade de memória reduzidas. Permite a captura dos dados pelo seu próprio teclado ou por um leitor de código de barras acoplado a ele. Utilizado no varejo para a realização de inventário, conferência de preços, recebimento de mercadorias, dentre outros processos (Carvalho & Galegale, 2006, p.3).

- **Software de retaguarda:** é um sistema de informação gerencial capaz de processar dados e produzir informações acerca das principais áreas e processos de um supermercado, como compras, gestão custos e preços de venda, controle de estoque, financeiro e contábil, dentre outros (Carvalho & Galegale, 2006, p.3).

- **Transferência eletrônica de dados (TED):** também conhecido com EDI – *Electronic Data Interchange*, é uma tecnologia que permite a troca de dados entre empresas, através de um padrão comum de transações, como por exemplo, pedidos, faturas, dados bancários, dentre outros (Carvalho & Galegale, 2006, p.3).

- **EAN:** [*European Article Numbering*]. Entidade internacional que administra o sistema-padrão de codificação dos produtos através dos códigos de barras, assim como a linguagem EANCOM para o XIII SIMPEP – Bauru, SP, Brasil, 06 a 08 de novembro de 2006. EDI. No país, sua representante é a EAN BRASIL (Associação Brasileira de Automação Comercial, que recentemente adotou a marca GS1). Os padrões EAN mais usados pelo varejo são o EAN-13 e EAN-8, além do EAN/ISBN [*International Standard Book Number*] e [*International Standard Serial Number*] EAN/ISSN para livros, o [*Distribution Unit Number*] EAN-DUN-14 para distribuição e também o [*Uniform Code Council*] EAN/UCC padrão norte-americano e canadense (Carvalho & Galegale, 2006, p.3).

- **Price Look-Up (PLU):** são códigos internos, que adotam o padrão EAN, utilizados pelos supermercados para identificar produtos que não possuem código de barras, como frutas, legumes e carnes, por exemplo. Segundo a EAN BRASIL (2005), existem cerca de 1.300 PLUs globais atribuídos a produtos de hortifrúti, criados para padronizar a codificação desta categoria e auxiliar na integração logística (Carvalho & Galegale, 2006, p. 4).

- **Impressora de termo-transferência:** tem como finalidade a impressão de códigos de barras em etiquetas para serem utilizadas nos produtos ou nas gôndolas. Utiliza um filme especial, chamado Ribbon, à base de cera, resina ou a mistura de ambas, que através da elevação de temperatura transfere a impressão para o papel; convém lembrar que existem diversos tipos de filmes e de papéis que são utilizados (Carvalho & Galegale, 2006, p. 4).

- **Leitora de cheques e boletos bancários:** semelhante ao leitor de código de barras, trata-se de um dispositivo de reconhecimento de caracteres e imagens, capaz de converter o código de barras contido em um boleto bancário ou o código padrão [Caracteres Magnéticos Codificados em Sete Barras] CMC-7 contido na parte inferior dos cheques, em formato digital, para que possa ser interpretado pelo computador (Carvalho & Galegale, 2006, p. 4).



**Figura 6** - Exemplos ilustrativos: equipamentos utilizados em supermercados.

Fonte: elaborado pela autora baseado em: Acil (2018). *Computador* [figura]. Recuperado de: <http://www.acillimeira.com.br/noticias:qual-a-melhor-hora-para-trocar-de-computador-na-empresa>; Canal Automação Comercial (2018). *Balança computador com impressora integrada* [figura]. Recuperado de: <https://www.canalautomacao.com.br/produto/521-balanca-toledo-etiquetadora-prix-5-plus-preta-30kg-x-2g-5g-10g>; Cut The Knot (2018). *Código de barras* [figura]. Recuperado de: <https://www.cut-the-knot.org/Curriculum/Arithmetic/EAN13.shtml>; Datalogic (2018). *Scanner Magellan* [figura]. Recuperado de: <http://www.datalogic.com/eng/products/retail/fixed-retail-scanners/magellan-9300i-pd-673.html>; Nonus (2018). *Leitor de código de barras* [figura]. Recuperado de: <http://www.nonus.com.br/leitor-codigo-barras.php>; Soluções Industriais (2018). *Coletor de dados* [figura]. Recuperado de: [http://www.solucoesindustriais.com.br/empresa/identificacao\\_etiquetagem\\_e\\_radio\\_frequencia/genecamp-servicos-em-automacao-comercial-ltda-/produtos/identificacao/coletor-de-dados-9](http://www.solucoesindustriais.com.br/empresa/identificacao_etiquetagem_e_radio_frequencia/genecamp-servicos-em-automacao-comercial-ltda-/produtos/identificacao/coletor-de-dados-9); Zoom (2018). *Notebook* [figura]. Recuperado de: <https://www.zoom.com.br/notebook/notebook-dell-i15-5566-intel-core-i7-7500u-15-6-8gb-hd-1-tb-linux>.

O segmento supermercadista possui equipamentos de automação para facilitar as atividades do dia a dia e utiliza os sistemas para gerenciar os dados. Esses dados podem ser avaliados para gerar informações para melhoria dos processos. A tecnologia de informação é fundamental em todos os segmentos e na rede de supermercados tornou-se também um diferencial.



## 5.2 Tecnologia da informação na rede de supermercados

A rede de supermercados objeto deste estudo é a Organização Verdemar Ltda., que atende a diversas regiões da região metropolitana de Belo Horizonte e município de Nova Lima, com foco nos clientes de classes A e B e com atuação desde 1993. Trata-se de uma rede com mais de 22.000 produtos, entre nacionais e importados, além dos produtos de fabricação própria, doravante denominada apenas por Verdemar.

O Verdemar é portador de vários prêmios e, segundo o *site* Supermercado Moderno (2017), no ano de 2016 ocupou a 44ª posição no país, oitava no estado e 31ª em nível de tipo de loja igual a supermercados (Figuras 7, 8 e 9).

44 <sup>o</sup>		<b>ORGANIZAÇÃO VERDEMAR LTDA</b>		R\$ 664.885.367,58			
		Minas Gerais					
		Nome Fantasia: VERDEMAR SUPERM. E PADARIA					
ANO	FATURAMENTO	FATURAMENTO/M²	FATURAMENTO/FUNC.	N° LOJAS	M²	CHECKOUT	N° FUNC.
2016	R\$ 664 MI	R\$ 56.265,16	R\$ 201.358,38	8	11.817	121	3.302
2015	R\$ 552 MI	R\$ 54.813,11	R\$ 178.182,35	7	10.087	106	3.103
2014	R\$ 501 MI	R\$ 49.757,76	R\$ 153.334,81	7	10.080	104	3.271

VARIÇÃO:		2016	2015				
FATURAMENTO (%)	FATURAMENTO M² (%)	FATURAMENTO/FUNC (%)	N° LOJAS (%)	M² (%)	CHECKOUT (%)	N° FUNC. (%)	
10,52	-5,65	3,87	14,29	17,15	14,15	6,41	

**Figura 7** – Ranking supermercados no Brasil .

Fonte: Supermercado Moderno (2017). *Ranking de supermercados*. Recuperado de: <http://www.sm.com.br/ranking-de-supermercados?tipo=BuscaBrasil&busca=VERDEMAR%20SUPERM.%20E%20PADARIA>.

8 °		ORGANIZAÇÃO VERDEMAR LTDA			R\$ 664.885.367,58		
44 ° NO BRASIL		Minas Gerais			Nome Fantasia: VERDEMAR SUPERM. E PADARIA		
ANO	FATURAMENTO	FATURAMENTO/M²	FATURAMENTO/FUNC.	N° LOJAS	M²	CHECKOUT	N° FUNC.
2016	R\$ 664 MI	R\$ 56.265,16	R\$ 201.358,38	8	11.817	121	3.302
2015	R\$ 552 MI	R\$ 54.813,11	R\$ 178.182,35	7	10.087	106	3.103
2014	R\$ 501 MI	R\$ 49.757,76	R\$ 153.334,81	7	10.080	104	3.271

VARIÇÃO: 2016 ▾ 2015 ▾						
FATURAMENTO (%)	FATURAMENTO M² (%)	FATURAMENTO/FUNC (%)	N° LOJAS (%)	M² (%)	CHECKOUT (%)	N° FUNC. (%)
10,52	-5,65	3,87	14,29	17,15	14,15	6,41

**Figura 8** – Ranking de supermercados em Minas Gerais.

Fonte: Supermercado Moderno (2017). *Ranking de supermercados*. Recuperado de: <http://www.sm.com.br/ranking-de-supermercados?tipo=BuscaBrasil&busca=VERDEMAR%20SUPERM.%20E%20PADARIA>.

31 °		ORGANIZAÇÃO VERDEMAR LTDA			R\$ 664.885.367,58		
44 ° NO BRASIL		Minas Gerais			Nome Fantasia: VERDEMAR SUPERM. E PADARIA		
ANO	FATURAMENTO	FATURAMENTO/M²	FATURAMENTO/FUNC.	N° LOJAS	M²	CHECKOUT	N° FUNC.
2016	R\$ 664 MI	R\$ 56.265,16	R\$ 201.358,38	8	11.817	121	3.302
2015	R\$ 552 MI	R\$ 54.813,11	R\$ 178.182,35	7	10.087	106	3.103
2014	R\$ 501 MI	R\$ 49.757,76	R\$ 153.334,81	7	10.080	104	3.271

VARIÇÃO: 2016 ▾ 2015 ▾						
FATURAMENTO (%)	FATURAMENTO M² (%)	FATURAMENTO/FUNC (%)	N° LOJAS (%)	M² (%)	CHECKOUT (%)	N° FUNC. (%)
10,52	-5,65	3,87	14,29	17,15	14,15	6,41

**Figura 9** - Ranking de loja igual a supermercado.

Fonte: Supermercado Moderno (2017). *Ranking de supermercados*. Recuperado de: <http://www.sm.com.br/ranking-de-supermercados?tipo=BuscaBrasil&busca=VERDEMAR%20SUPERM.%20E%20PADARIA>.

O Verdemar possui o setor de Tecnologia de Informação (TI) com 31 colaboradores. Os sistemas que auxiliam na gestão empresarial são de desenvolvimento próprio, existindo uma equipe de desenvolvimento e testes de *software* que analisam possíveis falhas e customizações inerentes à implementação dos sistemas de informações que atendem aos diversos setores da empresa (Verdemar, 2017).

Apesar de a empresa investir em tecnologia da informação e adotar o sistema descrito como ferramenta para gerenciar os processos internos do setor de TI, entende-se que o volume de dados é crescente. A própria organização, ao longo de 24 anos, registrou, em sua expansão, de duas para 12 lojas atualmente, com planejamento futuro para novas unidades, o que requer esforço da atualização da tecnologia. Os *softwares*, de maneira geral, requerem contínua avaliação e desenvolvimento frente às mudanças de mercado.

O Verdemar possui como missão tornar agradável o momento da compra do cliente e por esse motivo a diretoria da empresa solicita constantemente revisões dos procedimentos internos para melhorias. Para concretizar esse compromisso, a empresa conta com a tecnologia da informação, que é utilizada a favor de informatização de serviços e aceleração dos processos, visando melhorar e agilizar os serviços prestados aos clientes. Os diretores participam ativamente de decisões de tecnologia para que ela esteja a serviço do negócio principal da empresa.

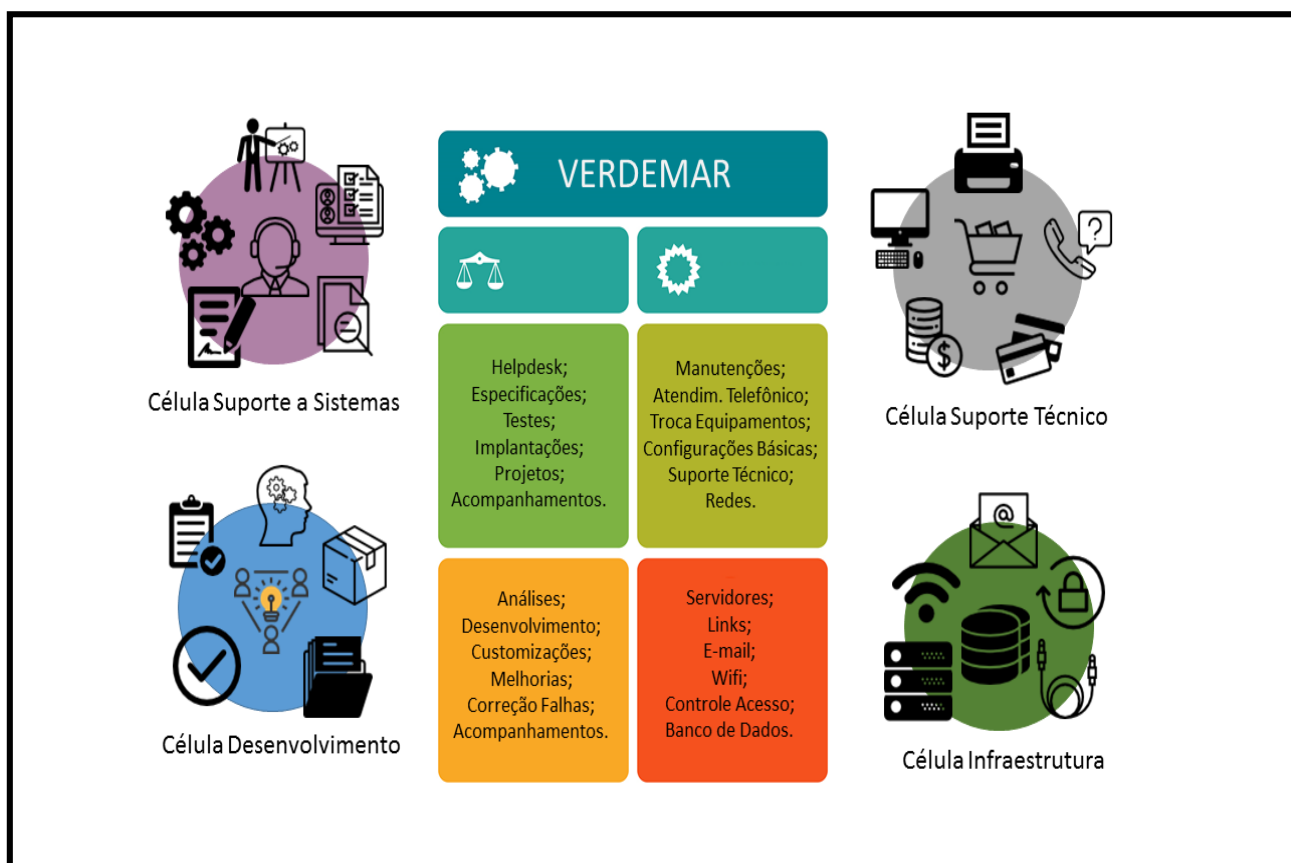
Badaró (2016) afirma que o coração do varejo é o estoque, principalmente em supermercados, e que existem diversas causas que podem acarretar problemas na sua validação. Entre as maiores causas Badaró destaca:

**Quadro 3 - Causas que ocasionam problemas no estoque**

<b>Causas que ocasionam problemas no estoque segundo Badaró (2016)</b>	<b>Ação realizada na Organização Verdemar com o intuito de minimizar os problemas</b>
Compra de estoque maior que o necessário	O Verdemar possui controle sistêmico da quantidade sugerida para compra baseado no giro dos produtos, associado ao <i>feeling</i> dos compradores frente aos fornecedores, visando efetuar compras cada vez <i>just in time</i>
Falta de inventários periódicos	Existe um departamento de inventário responsável pelo cronograma do ano inteiro de inventários rotativos, estratégicos e gerais nas lojas e no Centro de Distribuição. Os inventários são feitos via coletor de dados, com reconferências se necessário.
Falha na segurança	Somente colaboradores autorizados e devidamente identificados podem ter acesso ao depósito, produção, setores administrativos e áreas restritas. Além de contar com monitoramento de câmeras.
Mau armazenamento	A maioria dos produtos é armazenada no centro de Distribuição na região de Nova Lima, devidamente climatizado e com controle de temperatura para garantir a boa conservação dos produtos.
Não registrar todas as movimentações	Existe um setor de Gestão Administrativa de Mercadorias (GAM) que fiscaliza e audita os setores para avaliar se todas as movimentações de entrada e saída estão sendo executadas corretamente.
Utilização de Cadastro Genérico	Visando garantia de histórico de informações, todo produto que entra para a empresa é cadastrado e passa por um fluxo de etapas em diversos setores para preenchimento de dados e validação das informações. Todo produto é cadastrado somente mediante a autorização prévia da superintendência comercial e da diretoria. Desde informações fiscais até características e foto dos produtos são registradas.
Fator de conversão incorreto	O levantamento de dados do produto é feito durante seu recebimento no Centro de Distribuição. No ato de sua primeira entrega são coletados os dados logísticos do produto - altura da caixa, largura da caixa, quantidade de unidades em cada caixa, entre outras informações.
Cadastrar produtos em detalhes de sua descrição	O detalhamento do cadastro é feito devido à obrigatoriedade de processo de fluxo de etapas, que exige o preenchimento dos dados básicos e complementares do produto.
Não realizar conferências na entrada	No ato da entrega da mercadoria os produtos enviados na nota são conferidos fisicamente via coletor por um conferente de loja, visando garantir as quantidades e o produto correto. Vale ressaltar que antes da entrega da mercadoria o fornecedor envia o XML, que é importado automaticamente para o sistema que realiza as primeiras validações, repassando posteriormente para o setor especializado para vínculo de pedido de compras e análise prévia, para garantir mais eficiência na verificação física.
Vinculações incorretas	O sistema possui um simulador de receitas em que os produtos de fabricação própria passam por um processo rigoroso de validação e são submetidos à aprovação de vários setores.

Fonte: Badaró, A. (2016). *10 causas de problemas no seu estoque*. Infovarejo [Site]. 15 de agosto. Recuperado de: <https://www.infovarejo.com.br/10-causas-de-problemas-no-seu-estoque/>; Verdemar. (2017). *Documento recursos humanos: Manual Treinamento de Integração. Treinamento e Desenvolvimento*. Recuperado de: <http://trabalhenoverdemar.blogspot.com.br>.

A área da tecnologia da informação está dividida em quatro células para atender de maneira eficaz às demandas da empresa, são elas: suporte técnico, suporte a sistemas, infraestrutura e desenvolvimento interno. As áreas possuem interação entre elas e sua subdivisão permite a especialização dos atendimentos do dia a dia e nos projetos de melhorias.



**Figura 10** - Células tecnologia informação Verdemar.

Fonte: elaborado pela autora a partir de documentos institucionais/ rotinas do Verdemar. (2017). *Documento recursos humanos: Manual Treinamento de Integração. Treinamento e Desenvolvimento*. Recuperado de: <http://trabalhenoverdemar.blogspot.com.br>.

### 5.2.1 Célula de suporte técnico

A célula de suporte técnico é responsável por instalar, manter, controlar e acompanhar equipamentos, máquinas e acessórios que dão subsídio à execução de tarefas administrativas e operacionais das lojas. Varia desde a formatação de uma máquina até a configuração de um coletor de dados que executa operações como inventários e validação de preços. A célula possui como principal objetivo manter a disponibilidade

do ponto de venda para que os clientes efetuem suas compras. A equipe é dividida entre as lojas para garantir que os equipamentos estejam em bom funcionamento.

Os pontos de venda (PDVs) fazem parte da arquitetura principal e básica das lojas. Eles são equipados com emissores de cupom fiscal (ECF, equipamento de controle fiscal obrigatório decorrente da Lei Federal 9.532/97), teclado personalizado, *pin pad* (leitor de cartões de crédito, débito, tíquetes eletrônicos, entre outros), monitor, *scanner* (leitor com leitura ótica) para leitura dos códigos de barra e balança acoplada para pesagem de produtos, identificador biométrico, entre outros acessórios.

Os identificadores biométricos também fazem parte do parque de itens da responsabilidade da célula. Eles auxiliam nas operações de frente de caixa, proporcionando mais confiabilidade, e na segurança. Romano (2011) esclarece que a biometria é utilizada para verificação da identificação do usuário e é muito interessante, porque ao contrário de chaves ou senhas, a biometria não pode ser perdida ou roubada e está sempre com seu proprietário, promovendo o aumento da segurança da informação. Encarregados da frente de caixa e por gerentes de lojas, possuem sua identificação cadastrada para operacionalizar rotinas pertinentes ao PDV, como, por exemplo, cancelamento de itens e de cupons e emissão de relatórios.

As lojas são equipadas de forma que os clientes internos possam atender os clientes externos, prestando bom atendimento e levando a informação de maneira eficaz, além dos produtos de fabricação própria, os produtos nacionais e os importados. A linha de produtos oferecida é extensa. Para isso, manutenções preventivas são feitas regularmente e manutenções corretivas são feitas quando necessário. As lojas possuem vários equipamentos informatizados e eletrônicos para realização dos atendimentos, como, por exemplo:

- a) Terminais de consulta de preço: todos os produtos possuem código de barras para identificação. A partir dos terminais o cliente consegue visibilizar o preço atual e a descrição do produto;
- b) Coletor de dados: utilizado para execução de atividades operacionais (alteração de preços, controle de validade, avaliação de giro das mercadorias,

- inclusão/remoção de placas promocionais entre outros) e inventários parciais ou gerais de produtos;
- c) Celulares: utilizado para validação dos preços da concorrência, trata-se da pesquisa de preços. Rotina muito comum entre os varejistas. Anteriormente, a rotina era realizada por meio de anotações em papel e a tecnologia permitiu a informatização do processo manual;
  - d) Balanças: são usadas para pesagem de produtos a granel. Possuem ligação direta com o cadastro de produtos, o que permite a impressão de etiquetas com preço, descrição das receitas, informações nutricionais e muito mais;
  - e) Leitores de código de barras de boletos: as notas fiscais que possuem duplicatas anexas, devido ao grande volume e visualizando melhorias nos processos internos, passam por um processo de leitura para vinculação entre o boleto bancário e a nota fiscal, evitando digitação manual;
  - f) *Notebooks*: são utilizados para tarefas diversas nas lojas, tanto administrativas quanto operacionais. Cada técnico de suporte possui um *notebook* como uma de suas ferramentas de trabalho, entre muitos outros equipamentos.

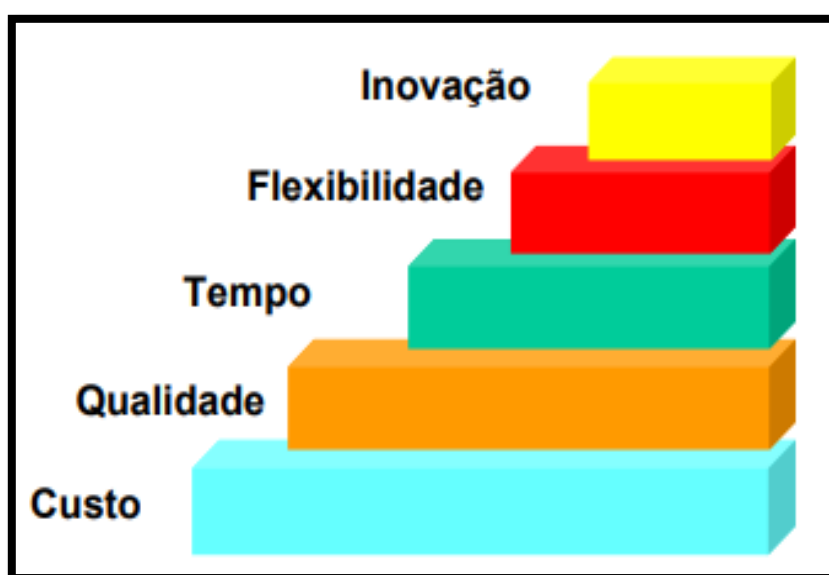
### **5.2.2 Célula de suporte a sistemas**

A tecnologia da informação sistêmica possui grande importância no cenário competitivo das empresas, principalmente quando torna os processos mais ágeis, transparentes e eficientes. Cada vez mais os resultados das empresas estão sendo analisados minuciosamente, visando encontrar falhas para correção, redução de falhas já identificadas e maximização dos lucros. Para isso, é cada vez mais utilizada a gerência por meio de sistemas.

A célula de suporte a sistema oferece à empresa todo o apoio necessário para informatização de processos manuais, para implantação de novos projetos e para manutenção de processos já existentes. Os sistemas atuais dividem-se em uma parte *desktop* e outra *web*, para disponibilizar mais mobilidade. Solicitações de compra, entrada de mercadorias, transferência de produtos, vendas, importações, agendamentos de entrega, entre outros processos, são controlados via sistemas.

Os sistemas permitem análise dos gestores e tomadas de decisões. Essas decisões são baseadas no cenário que o sistema apresenta e na interpretação da gestão. Nenhum sistema substitui o *feeling* e a experiência de um ser humano, mas a visão que o sistema possibilita permite ao responsável direto inferências mais assertivas.

Anteriormente, as organizações competiam levando em consideração somente os custos (Romano, 2011). Com o passar dos anos, essa competição tornou-se mais ampla, com várias dimensões e focos, conforme demonstra a Figura 11.



**Figura 11** - Fatores que atingem a competitividade.

Fonte: Bolwijn e Kumpe (1990 como citados em Romano, R. R. (2011). *Os impactos do uso de tecnologia da informação e da identificação e captura automática de dados nos processos operacionais do varejo*. Tese (Doutorado em Administração) - Escola de Administração de Empresas de São Paulo da Fundação Getúlio Vargas, São Paulo. Recuperado de: <http://bibliotecadigital.fgv.br/dspace/bitstream/handle/10438/8895/Tese%20Regiane%20Relva%20Romano%20-%20Versao%20Final%20Entregue.pdf?sequence=3&isAllowed=y>).

**Custo:** é capacidade da organização em atingir os objetivos, com a melhor relação custo-benefício possível, com o uso dos recursos disponíveis.

**Qualidade:** é a capacidade da organização em oferecer um produto ou serviço, de acordo com as especificações e/ou necessidades exigidas pelo cliente.

**Tempo:** é a capacidade da organização em responder, prontamente, aos novos requisitos do mercado.

**Flexibilidade:** é a capacidade da organização em oferecer uma grande variedade de produtos e serviços, num nível de customização que, no limite, atenda às necessidades individualizadas de cada cliente, ou seja, customização em massa.



**Inovação:** é a capacidade da organização de conceber ideias, práticas, produtos e/ou serviços que sejam percebidos como novos pelo mercado (Romano, 2011, p. 26).

Faz-se necessário acompanhar o mercado competitivo para se manter vivo nele, por esse motivo tomar decisões de forma assertiva tornou-se essencial. Várias rotinas foram informatizadas utilizando coletores de dados e celulares corporativos como meio de agilidade, rodando sob uma plataforma Java.

Rotinas que necessitam de respostas rápidas foram implementadas com a mobilidade de acesso, assim como alteração de contratos, autorizações de promoções, acompanhamentos de indicadores de venda, manutenções internas e liberação de notas. Os sistemas *mobiles* permitem que o acompanhamento da empresa seja feito mesmo sem estar *in loco* em suas dependências. Os processos continuam sendo acompanhados e feitos no tempo necessário de maneira segura.

Os sistemas utilizados também permitem interface com terceiros para que o controle de troca de documentos, informações e análise de dados seja realizado em forma de fluxo. Obrigações contábeis e agendamentos de entrega no Centro de Distribuição são exemplos de processo rotineiros, porém essenciais que funcionam de maneira transparente. A contabilidade envia solicitações de dados e a mesma é feita e registrada no sistema, assim como anexos, datas, horários e retornos.

Os agendamentos de entrega também são controlados via sistema, que informa o período de entrega que atende a empresa e o fornecedor agenda o melhor dia para descarregar a mercadoria de acordo com sua rota de entregas. Dessa forma ambos realizam o processo de satisfatoriamente.

Assim, os sistemas gerenciais crescem na empresa e são cada vez mais utilizados além do *hardware*. Os sistemas passaram a ser parte da rotina e do controle diário das operações. Os sistemas utilizados como destaques no Verdemar são de desenvolvimento próprio, e o suporte sistêmico realiza tarefas de entendimento do problema ou da solicitação de melhoria, cria especificações para implementação, realiza testes, implanta e acompanha sua utilização. Todas essas fases correspondem a uma fábrica de *software*, que hoje é considerada pela empresa como um diferencial.

A tecnologia trabalha para que o negócio esteja em primeiro lugar, assim como o cliente que mantém a empresa no mercado.

Segundo Silva & Moraes Filho (2015, p. 49):

“é razoável admitir que, atualmente, nenhuma organização deve conceber sua existência sem o uso de algum tipo de tecnologia de informação.”

“... é cada vez menos provável que alguma organização alcance sucesso sem o uso de modernas tecnologias da informação e comunicação (TICs).”

A esse respeito, Albertin e Albertin (2005, p. XVIII como citados em Silva & Moraes Filho 2015) mencionam:

“A oferta de TI e seu aproveitamento amplo e intenso pelas organizações têm sido considerados como uma realidade nos vários setores da economia e condição básica para as empresas sobreviverem e competirem.”

### **5.2.3 Célula de desenvolvimento interno**

O Verdemar possui desenvolvimento interno e próprio para atender às demandas da empresa de maneira ágil e em conformidade com as inovações que são consideradas como diferenciais para o negócio. Atendimento a legislação e processos contábeis e burocráticos é implementado, assim como rotinas que visam informatizar, melhorar o controle, criar histórico para tomada de decisões e contribuir para o constante aperfeiçoamento dos processos da empresa.

A ruptura de produtos, por exemplo, foi foco de desenvolvimento interno para geração de atividades diárias de análise e verificação dos produtos na área de venda. A ruptura de produtos é uma situação em que um produto regularmente comercializado ocupa determinado espaço na gôndola, e esse produto, por algum motivo desconhecido, não está disponível na área de vendas ao consumidor no momento da compra. A ruptura de produtos é conhecida como um problema crítico no mercado (Aguiar, Sampaio & Hilsdorf, 2010). Varejistas e fabricantes têm interesse direto na sua identificação e mapeamento de causas de ruptura e de suas respectivas cadeias de suprimento. “A solução desse problema poderia aumentar em até 5% de suas receitas” (Aguiar *et al.*, 2010).

O Verdemar, visando minimizar esse cenário, atua, exceto aos domingos e feriados, com atividades diárias de verificação de itens de entrada e giro. Trata-se de atividades executadas em um coletor de dados onde os produtos são registrados a partir da leitura do código de barras, informando seu preço de venda e garantindo, assim, que o produto a que foi dada entrada no depósito da loja esteja disponível na área de venda e que um produto com giro baixo ou alto também esteja ao alcance dos clientes.

A Revista *Gôndola* destacou, em sua capa do mês de novembro/2017, a seguinte questão: onde está o produto que deveria estar aqui? E nomeia a ruptura em sua matéria como a inimiga de todos (AMIS, 2017) (Figura 12).



**Figura 12** – Capa da Revista *Gôndola*.

Fonte: Associação Mineira de Supermercados. AMIS. (2017). Onde está o produto que deveria estar aqui? *Gôndola*, ano 23, n. 258, novembro. Recuperado de: <http://online.anyflip.com/glyq/coji/mobile/index.html>.

O desenvolvimento interno trabalha sempre buscando inovações e implementações que possam fazer diferença no dia a dia dos clientes internos e externos. Constantemente é feito acompanhamento das tarefas para minimizar, a partir da tecnologia, a possibilidade de erros e aumentar a produtividade.

O processo de importação de produtos, por exemplo, é um fluxo burocrático e detalhado. Ele atende à análise dos produtos e de suas respectivas quantidades até a declaração de importação (DI) na Receita Federal. A análise dos produtos e sua alocação em *containers* devem ser controladas com atenção específica, porque erros podem acarretar prejuízos financeiros. Para isso, o sistema auxilia na regularização das etapas e nas solicitações de compra de acordo com *lide time* dos produtos (os prazos de entrega, tempo entre o pedido e sua entrega efetiva; geralmente mais extensos, de três a seis meses de entrega, dependendo do produto e quantidade).

O abastecimento automático de produtos também é um exemplo de desenvolvimento interno. A maioria dos produtos é entregue no Centro de Distribuição, tendo nas lojas apenas o recebimento de produtos mais delicados e com data de validade mais rotativa e menor. Para transferir as mercadorias do Centro de Distribuição para as lojas, existe um mecanismo sistêmico que informa que os produtos foram vendidos e que precisam ser reabastecidos para evitar a ruptura dos produtos (item citado anteriormente).

Controlar o estoque assiduamente é um grande desafio para os varejistas em geral, principalmente para os varejistas que possuem fabricação própria. Os itens fabricados possuem itens relativos ao preparo e à embalagem para venda. Contar o estoque utilizado e coletar as receitas de cada item sem a informatização é algo extremamente moroso. Para evitar esse retrabalho, outro exemplo que pode ser citado é a informatização dos inventários semanais por intermédio de coletores de dados.

Os coletores de dados geram atividades semanais, de acordo com agendamento prévio, para execução de atividades de contagens. Essas contagens são inclusas automaticamente no sistema para que o comprador responsável analise o que já foi

gasto, para efetuar a compra de reposição. Assim, o desenvolvimento interno segue implementando funcionalidades que facilitem as tarefas do dia a dia da operação.

#### **5.2.4 Célula de infraestrutura**

A célula de infraestrutura é responsável por todo o arcabouço tecnológico da empresa, proporcionando um conjunto de elementos que dão apoio às áreas de suporte técnico, a sistemas e ao desenvolvimento, assim como à empresa como um todo. Realiza monitoramento, gerência e configuração da rede dados, o controle de equipamentos de telecomunicação e tráfego de transações. É responsável por desenvolver soluções lógicas a partir de recursos de *hardware/software* para rede.

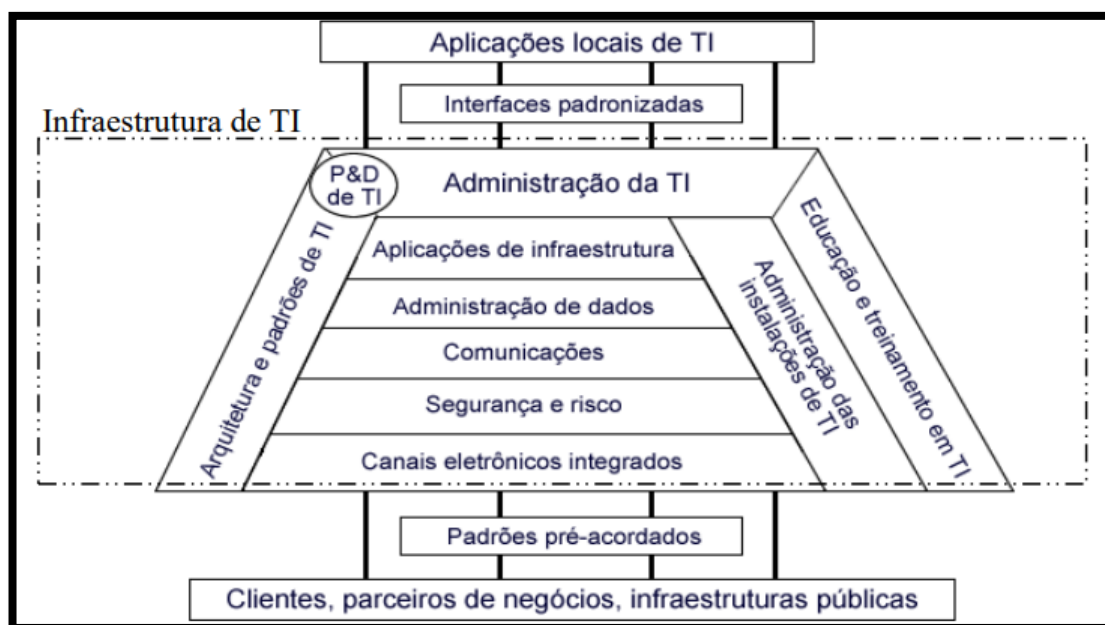
Também realiza gerenciamento da rede sem fio, é responsável pela criação e suporte aos *e-mails* da empresa, pela instalação, monitoramento e atualização dos servidores; pela segurança das informações, pela administração dos recursos *hardware* e dos sistemas operacionais. Realiza manutenções para garantir alta disponibilidade dos *links* de dados e dos sistemas, além de cuidar de *racks*, *data centers* e estudo de novas tecnologias.

No dicionário de Ferreira (2018), infraestrutura é aquilo que garante a existência de determinado grupo, instituição ou organização. Gurgel e Cezar Filho (2010) acrescentam que existem *frameworks* que regulamentam a infraestrutura da tecnologia da informação das organizações, semelhante às normas *International Standards Organization* (ISO).

Trata-se de *frameworks* como *Control Objectives for Information and related Technology* (CobIT), *IT Infrastructure Library* (ITIL), *Project Management Body of Knowledge* (PMBOK) e o *Capability Maturity Model Integration* (CMMI), que orientam como os processos devem ocorrer para que tenham bons resultados (Gurgel & Cezar Filho, 2010). Isso porque a infraestrutura é essencial para que as empresas exerçam suas atividades.

Gurgel e Cezar Filho (2010) complementam que os componentes da infraestrutura de TI podem ser representados (Figura 13) de acordo com o suporte de demandas

internas da organização (aplicações locais de tecnologia da informação) e com a interface com outras organizações externas (clientes, parceiros de negócios, infraestruturas públicas).



**Figura 13** - Componentes da infraestrutura de TI.

Fonte: Weill e Ross (2006, p. 48), como citados em Gurgel, G. P. F., & Cezar Filho, P. (2010). *Avaliação da infraestrutura de tecnologia da informação de empresas em Mossoró-RN*. 5, 45-55. Recuperado de: <http://www2.ifrn.edu.br/ojs/index.php/HOLOS/article/view/554>.)

Na linha de pensamento de Gurgel e Cezar Filho, (2010), os serviços da infraestrutura de tecnologia de informação totalizam 70 serviços identificados. Nesse sentido, cada um dos componentes representa um subconjunto desses serviços (Gurgel & Cezar Filho, 2010).

Os componentes foram agrupados em duas camadas: física e gestão. A camada física compreende os seguintes componentes: canais eletrônicos integrados, segurança e risco; comunicações; administração de dados; aplicações de infraestrutura; e administração das instalações de TI. A camada de gestão compreende componentes que tratam da gestão de toda a infraestrutura de TI: administração de TI; arquitetura e padrões de TI; educação e treinamento em TI e pesquisa e desenvolvimento de TI (Gurgel & Cezar Filho, 2010).

Logo, para que todas as demais áreas da TI funcionem (suporte técnico, suporte a sistemas e desenvolvimento interno), é essencial que a infraestrutura esteja bem arquitetada e bem consolidada.

### **5.3 Mantis - Sistema de registro de *bugs* e melhorias**

Os sistemas que auxiliam na gerência dos processos e atividades da empresa são desenvolvidos internamente. Por esse motivo, fez-se necessário gerenciar as possíveis falhas e customizações de melhorias para manutenção desses sistemas e acompanhar as novas demandas. O controle atual é gerenciado pelo *software* Mantis®.

O Mantis® é uma ferramenta *web* gratuita de código aberto que permite gerenciar bugs e melhorias de softwares (Mantis, 2017). Ele atende ao departamento de TI do Verdemar frente às necessidades de organização e gestão das solicitações feitas pelos clientes internos (gerenciando projetos de desenvolvimento através da inserção de registros, chamado de casos, pela ferramenta). Trata-se de um sistema livre que permite gerenciar projetos de *software*, tendo código aberto para modificações. Ele possui notificações de *e-mail*, controle de acesso, abertura de “casos” (relatos para verificação com a possibilidade de criar fluxos de processo), pesquisas de casos ativos e encerrados, gerenciamento de itens na agenda de análise e controle de registros e acessos, entre outras funcionalidades.

O Mantis® foi implantado na versão 1.2.9 desde meados de 2010, utilizando a plataforma de linguagem *Hypertext Preprocessor* (PHP) com implementação no banco de dados MySQL versão 5.0. A partir dessa implantação, não houve atualizações posteriores, devido à sua utilização intensa.

O foco desta pesquisa é o sistema Mantis®, considerando que sua utilização proporciona ao setor de TI a possibilidade de pesquisar informações sobre a implantação de projetos anteriores, assim como falhas identificadas de *software*, análise de documentações, verificação de definição de requisitos de usuários (DRU) e informações diversas sobre os sistemas da empresa. Quaisquer informações relativas a sistemas estão documentadas em um caso (registro) desse sistema e as

informações são centralizadas permitindo consultas, visualizando o histórico dos dados, auxiliando na resolução de problemas atuais já vividos anteriormente ou na análise de problemas parecidos ou na análise de projetos.

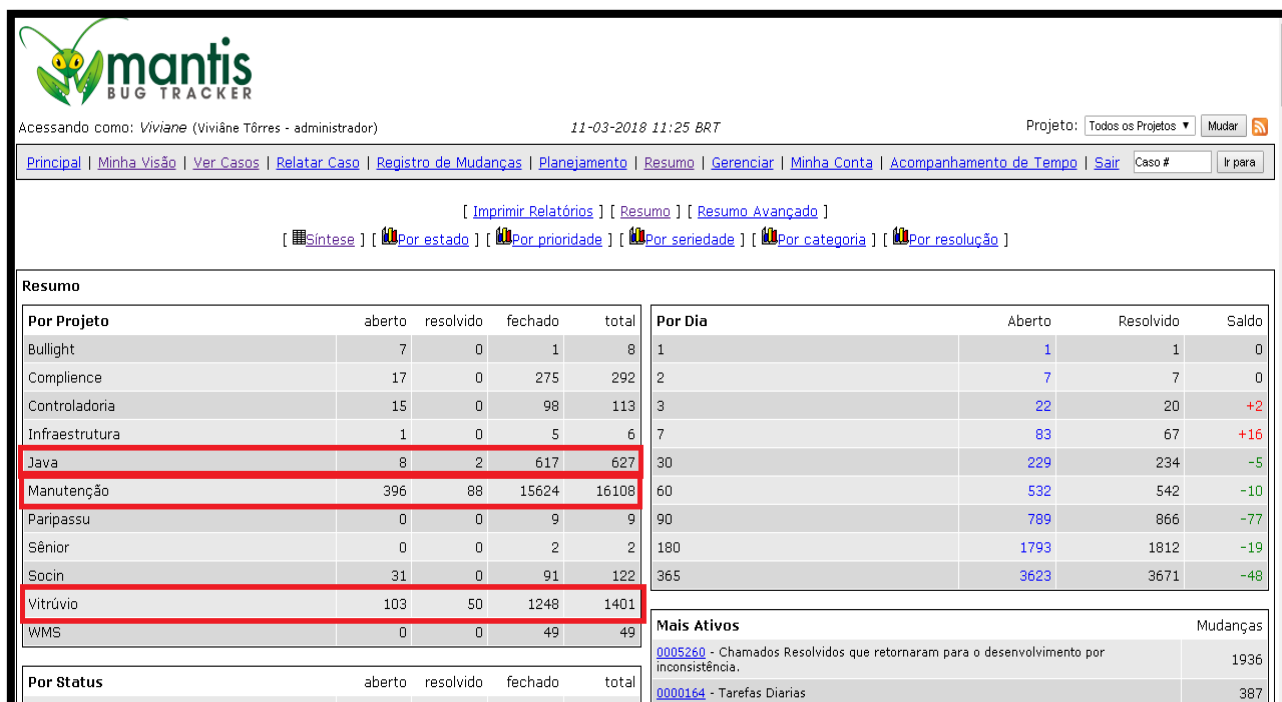
No entanto, o volume de informações, dados e documentos inseridos no Mantis® aumenta consideravelmente todos os anos, e essas informações registradas servem de auxílio para a realização dos atendimentos diários feitos pelos clientes internos relatando dúvidas e possíveis falhas. Ao tentar pesquisar assuntos já registrados, a quantidade de itens retornados é grande, dificultando o atendimento rápido.

Atualmente, cada sistema possui um projeto separado no Mantis®, no qual os registros referentes a ele são feitos. Existem casos com situação igual aberta, resolvida ou até mesmo fechada. Todos os registros servem de auxílio para realizar os atendimentos, principalmente os registros com situação igual a fechada, que possuem a informação do que foi feito para finalizar o atendimento, que no Mantis® é chamado de caso (similar a chamado, solicitação) (Figura 14).

Os principais projetos utilizados atualmente somam mais de 19.000 registros (casos) em um período de seis anos. Para cada caso aberto, os dados do atendimento são inseridos de acordo com as informações repassadas pelos clientes internos e posteriormente são utilizados como histórico para solucionar outros problemas similares (Figura 15).

A Figura 14 mostra a tela de resumo do Mantis® com os diversos projetos. E a Figura 15 ilustra a tela abertura de casos no Mantis®.





**Figura 14** – Tela resumo do Mantis®

Fonte: Verdemar. (2017). *Documento recursos humanos: Manual Treinamento de Integração. Treinamento e Desenvolvimento*. Recuperado de: <http://trabalhenoverdemar.blogspot.com.br>.

**mantis**  
BUG TRACKER

Acessando como: Viviane (Viviane Tárres - administrador) 11-03-2018 11:44 BRT Projeto: Manutenção Mudar

Principal | Minha Visão | Ver Casos | Relatar Caso | Registro de Mudanças | Planejamento | Resumo | Gerenciar | Minha Conta | Acompanhamento de Tempo | Sair Caso # Ir para

Recentemente Visitado: 0016538, 0016348, 0016090, 0016091, 0018974

**Digite os Detalhes do Relatório**

\*Categoria (selecione)

Frequência não se tentou

Gravidade pequeno

Prioridade normal

Selecionar Perfil

OU Preencha

Plataforma

SO

Versão SO

Versão do Produto

Atribuir a

Previsto para a Versão

\*Resumo

\*Descrição

Passos para Reproduzir

Informações Adicionais

Desenvolvedor  Agnaldo  Caio  Danilo  Lazaro  Eduardo  Lorrان  Renato  Tiago  Não Definido  Daniel  Rodrigo  Talita

Responsável\_DRU  agnaldo  grace.santos  mario.avila  marcus.oliveira  matheus.silva  rafael.marques  vanilda.reis  viviane  nao\_definido  nao\_necessario  dru\_ja\_concluida\_antes  marcilene.silva

Estimativa

LojaAbertura  01  02  03  04  05  06  07  08  09  10  11  12  13  14  15  CD

\*CanalAbertura  Email  Presencial  Telefone  Mantis

EsforcoTI  1  2  3  4  5

ImpactoCliente  1  2  3  4  5

\*Modulo

Solução

Carregar Arquivo (Tamanho máximo: 2,097k) Escolher arquivo Nenhum arquivo selecionado

Visibilidade  público  privado

Continuar Relatando  selecione para relatar mais casos

\* requerido Enviar Relatório

Copyright © 2000 - 2012 MantisBT Group  
mantis@riververdemar.com.br

**Figura 15 – Tela Abertura do Caso no Mantis®**

Verdemar. (2017). *Documento recursos humanos: Manual Treinamento de Integração. Treinamento e Desenvolvimento*. Recuperado de: <http://trabalhenoverdemar.blogspot.com.br>.

### 5.3.1 Cenário atual

O sistema de registro de *bugs* e melhorias no Mantis® auxilia no controle dos atendimentos e casos. Sua pesquisa limita-se a filtros predeterminados ou a consultas de texto livre, o que permite que vários casos sejam retornados e que seja necessário analisar manualmente qual registro se refere ao atendimento que está sendo feito no momento da solicitação do cliente. Com isso, o atendimento se torna moroso e lento. Atualmente, uma pesquisa é feita com palavras-chave ou filtros- padrão do sistema. Ex: nota fiscal apresenta problemas relativos ao imposto sobre operações relativas à circulação de mercadorias e sobre prestações de serviços de transporte interestadual, intermunicipal e de comunicação (ICMS) (Figura 16).

The figure displays two screenshots of the Mantis Bug Tracker search interface. Both screenshots show the search results for the query "nota fiscal icms".

**Top Screenshot:** The search results show 392 cases. The search criteria are "Procurar: nota fiscal icms" and "Filtro: [ Filtros Avançados ]". The "Categoria" filter is set to "qualquer".

**Bottom Screenshot:** The search results show 121 cases. The search criteria are "Procurar: nota fiscal icms" and "Filtro: [ Filtros Avançados ]". The "Categoria" filter is set to "Falha".

The interface includes a navigation menu at the top with options like "Principal", "Minha Visão", "Ver Casos", "Relatar Caso", "Registro de Mudanças", "Planejamento", "Resumo", "Gerenciar", "Minha Conta", "Acompanhamento de Tempo", "Sair", and "Caso #". The search bar is located at the bottom of the results table, and the "Visualizando Casos" section shows the current count of results.

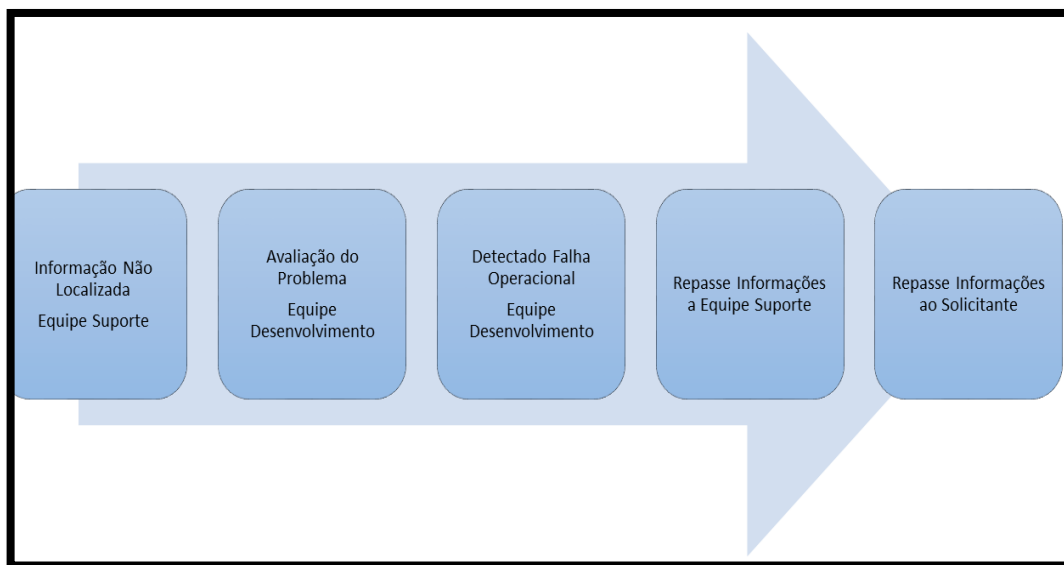
**Figura 16** – Tela pesquisa do Mantis®

Fonte: Verdemar. (2017). *Documento recursos humanos: Manual Treinamento de Integração, Treinamento e Desenvolvimento*. Recuperado de: <http://trabalhenoverdemar.blogspot.com.br>.

O exemplo simula um cenário real em que o cliente interno entra em contato informando que a nota fiscal que está sendo recebida nos depósitos das lojas apresenta problemas relativos ao ICMS. Trata-se de um tributo pago ao governo e que precisa ser apurado pelo sistema. Para orientar o cliente sobre qual é o procedimento correto a ser feito ou até mesmo para verificar se o atendimento se trata de uma possível falha no sistema ocorrida outras vezes, o analista realiza pesquisas na base histórica Mantis®.

São feitas duas consultas ao sistema Mantis® pela palavra-chave “nota fiscal ICMS” (Figura 16), sendo que o primeiro filtro está sem o campo categoria selecionado e o segundo especifica que são somente registros classificados como falha. Na primeira consulta, 392 registros são retornados para verificação do analista e na segunda 121 registros. Em ambos os casos o resultado é excessivo diante da pergunta inserida. Mesmo inserindo mais informações na consulta, o analista apresenta dificuldades em localizar uma resposta que o auxilie no atendimento, que em sua maioria são feitos pelo telefone.

O tempo de espera da resolução do problema de uma nota fiscal acarreta atrasos no descarregamento da mercadoria, e com isso a disponibilidade do produto na área de venda, ou seja, uma possível ruptura que foi destacada como um problema comum nos supermercados. Para evitar esse problema, outro pode ser gerado quando a equipe de suporte a sistemas não realiza a análise e aciona a equipe de desenvolvimento para avaliar diretamente no código-fonte o problema informado pelo cliente. Ao final da verificação, pode-se concluir por uma falha operacional de digitação. Trata-se de um problema em cadeia.



**Figura 17** – Processo de atendimento TI no Verdemar.

Fonte: elaborada pela autora.

### **5.3.2 Cenário desejado**

O registro dos atendimentos no sistema Mantis® é fundamental para a criação da base histórica de suporte já realizado para controle das demandas atendidas e para permitir a consulta posterior de soluções utilizadas na resolução dos problemas identificados. No entanto, espera-se que essa consulta seja breve, de fácil acesso e de boa usabilidade. Perguntas completas podem permitir mais agilidade nos atendimentos telefônicos, remotos e presenciais, visto que os filtros-padrão disponibilizados pelo sistema restringem as pesquisas.

A possibilidade de digitar uma pergunta e o sistema já apresentar todas as possíveis soluções para o problema informado é diminuir o tempo de análise e o tempo por parte da equipe de suporte. É diminuir o tempo de espera da solução por parte do cliente, otimizar o atendimento e proporcionar mais rapidez e interatividade nos atendimentos. Espera-se mais flexibilidade do sistema, porque até mesmo os analistas de suporte possuem dúvidas nos atendimentos ou estão mais capacitados em outros tipos de atendimentos e assuntos específicos. Veja-se a seguir possíveis perguntas e respostas que o sistema poderia retornar:

**Quadro 4** - Perguntas e possíveis respostas sistêmicas

<b>Perguntas</b>	<b>Possíveis Respostas</b>
Quais são os motivos de rejeição de uma nota fiscal?	A inscrição estadual do emitente ou destinatário é inválida.
	A classificação da natureza da operação é inválida.
	A numeração e/ou número de série estão incorretos.
Por que a atividade de alteração de preço (ruptura) não é exibida no coletor?	O coletor não possui sinal de antena.
	O coletor não está logado no usuário correto.
	A atividade ultrapassou o tempo determinado e está expirada.
Por que a nota fiscal não é apresentada no coletor de recebimento para conferência?	A nota não foi enviada para conferência no coletor.
	Houve problemas técnicos no coletor.
	A nota apresentou inconsistências na validação.

Fonte: Elaborada pela autora.

Há interesse em que, em vez de realizar a leitura de vários atendimentos registrados no sistema Mantis®, seja possível apenas incluir a dúvida e o sistema apresentar todas as possíveis soluções para execução e testes, a partir de respostas calcadas na base de conhecimento já registrada anteriormente, ou seja, os próprios atendimentos antigos.

## 6 Metodologia

A pesquisa desenvolvida no presente trabalho é de natureza aplicada, abordagem qualitativa com procedimentos técnicos de pesquisa bibliográfica (constituindo-se principalmente de teses, dissertações, livros e artigos científicos) e estudo de caso, utilizando métodos dedutivos. Ou seja, a dedução “parte do entendimento das leis e teorias que abrangem determinado fenômeno e a partir da definição de premissas e análise da relação entre elas se constrói o conhecimento” (Dresch, 2013, p.45).

A abordagem quanto aos objetivos é descritiva e a revisão sistemática de literatura (RSL) é a base para listar as principais técnicas e métodos dos sistemas de perguntas e respostas que podem ser utilizadas para aplicação no sistema de registro de *bugs* e melhorias para realizar adaptações ou pode-se avaliar a sua substituição.

### 6.1 Etapas e procedimentos

Este estudo pode ser classificado como experimental e qualitativo em sua abordagem metodológica, além de se tratar de um estudo de caso. Quanto aos fins, pode ser definido como descritivo, porque apresenta características dos sistemas perguntas e respostas, podendo estabelecer correlações entre as variáveis e definir a sua natureza (Vergara, 2006, p. 47).

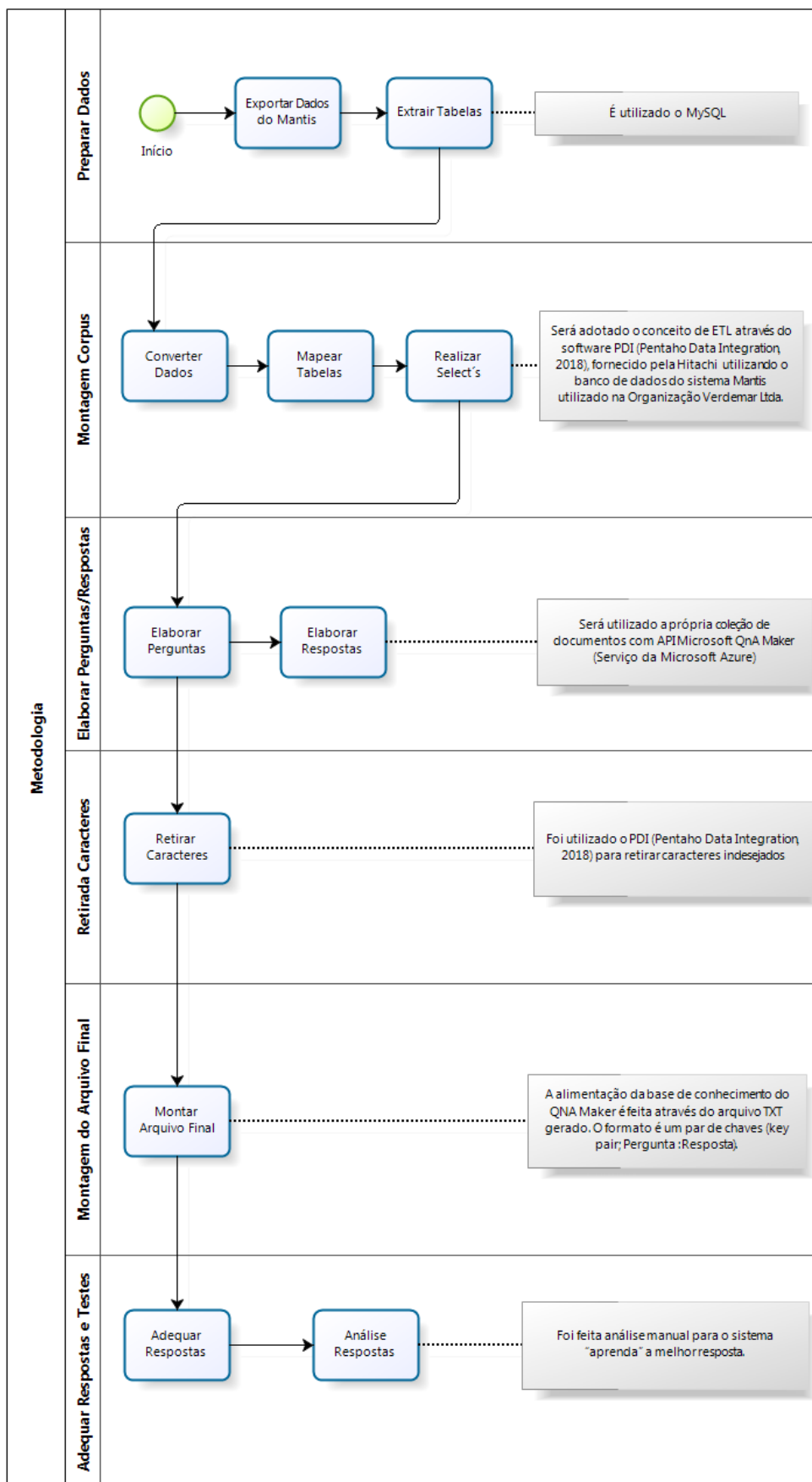
O trabalho foi dividido em sete etapas de acordo com os objetivos do trabalho, com o intuito de atingir o objetivo geral de testar e/ou aplicar uma das técnicas de sistemas de perguntas e respostas em uma base histórica, a fim de realizar análises dos resultados recuperados, avaliando-se sua assertividade. São elas:

- a) **Exportação dos dados do Mantis e extrair tabelas:** os dados que estão registrados no Mantis® foram exportados em tabelas que foram úteis para a realização do trabalho. Foi feito procedimento de verificação de quais dados e tabelas eram relevantes para o trabalho, utilizando o banco de dados MySQL;
- b) **Montagem do corpus:** fez-se a conversão do banco de dados do sistema Mantis® (que possui os registros de atendimentos) para o formato .TXT. Adotou-se o processo de conversão dos dados, mapeamento das tabelas e

- selects* (declaração SQL que retorna um conjunto de resultados de registros de uma ou mais tabelas) das possíveis respostas, a partir do conceito de *Extract Transform Load* (ETL - Extração Transformação Carregamento) do software PDI (*Pentaho Data Integration*, 2018), fornecido pela Hitachi;
- c) **Elaboração das perguntas x repostas:** foi feita a elaboração dos questionamentos para serem inseridos no sistema, tendo as respostas como referência para avaliação dos testes. utilizou-se o serviço da *Microsoft Azure* (2018) por meio da API *Microsoft QnA Maker* (serviço de perguntas e respostas com interface gráfica) para elaborar as perguntas a partir do conteúdo da base já existentes (Mantis®). Com isso, foi feito cruzamento entre as perguntas (casos do Mantis®, campo Resumo) e as possíveis respostas (campo Anotações) em uma tabela temporária;
  - d) **Retirada de caracteres especiais ou caracteres inválidos:** fez-se análise de caracteres especiais ou inválidos que atrapalham a verificação do contexto, utilizando uma função do *software Pentaho Data Integration* (PDI, 2018) para esse procedimento;
  - e) **Montagem do arquivo final:** foi elaborado um arquivo a ser importado pelo *QNA Maker*. A alimentação da base de conhecimento do *QNA Maker* é feita com base no arquivo .TXT gerado. O formato é um par de chaves (*key pair*, pergunta : resposta);
  - f) **Adequação das respostas (treinamento):** foi feita análise das relações entre as perguntas e a relevância das respostas, para que o *ranking* das possíveis respostas estivesse mais coerente com o esperado. E fez-se análise manual para o sistema “aprenda” a melhor resposta;
  - g) **Análise das respostas:** foram verificadas as respostas retornadas para cada pergunta. Os resultados retornados foram analisados manualmente.

O processo adotado encontra-se na Figura 18.





**Figura 18** – Etapas da metodologia adotada.  
 Fonte: Elaborada pela autora.

## 6.2 Mantis: registro de atendimento e exportação de dados/tabelas

Atualmente, o sistema de registro de bugs e melhorias (Mantis®) possui mais de 19.000 casos relatados no sistema oficialmente. Os registros são referentes a problemas de desenvolvimento (falhas na implementação), solicitação de *selects* (relatórios rápidos), customizações de melhoria, dúvidas, implantações de novos projetos, entre outros tipos de classificação. Para localizar informações utilizando apenas os filtros disponíveis pelo próprio sistema ou pesquisando por meio de palavras-chave, o processo é lento e de análise manual, além de apresentar problemas de propriedades linguísticas que podem influenciar no resultado, desde particularidades da língua portuguesa até os caracteres considerados inválidos.

Segundo Maia e Souza (2010), itens como a polissemia (a palavra tem vários significados), a sinonímia (duas palavras podem designar o mesmo significado) e duas e mais palavras podem combinar-se em ordem diferente designando ideias completamente diversas implicam o resultado da recuperação da informação e a taxa de ruído. Para a sinonímia, sua ocorrência pode impactar diretamente na taxa de silêncio. Com isso, a precisão e a revocação do sistema ficam comprometidas.

Exemplos:

a) A palavra pode ter vários significados (polissemia).

Chave (solução de um problema);

Chave (ferramenta para abertura de portas);

Chave (ferramenta para apertar parafusos).

b) Duas palavras podem designar um mesmo significado (sinonímia).

Abóbora – Jerimum.

c) Duas ou mais palavras podem combinar-se em ordem diferente, designando ideias completamente diversas.

Crimes, juvenis, vítimas (vítimas de crimes juvenis);

Crimes, juvenis, vítimas (vítimas juvenis de crimes).

Para o Mantis® pode-se citar como exemplo a palavra “nota fiscal”, que pode ser encontrada facilmente no sistema como: nota fiscal eletrônica, nfe, nf-e, nota, NF, etc. Todos os termos remetem ao mesmo significado dentro do contexto do sistema. No momento da consulta podem interferir negativamente na precisão dos resultados durante os procedimentos de recuperação da informação.

O sistema Mantis® possui MySQL (um dos componentes centrais da maioria das aplicações públicas da Internet) como gerenciador do seu banco de dados. A base utilizada possui nele foram os atendimentos registrados até abril/2018, cerca de 19.200 atendimentos. Esse histórico dos atendimentos resultou aproximadamente 7 GB de tamanho. O sistema Mantis® utiliza a linguagem de *script open source* PHP (um acrônimo recursivo para PHP: Hypertext Preprocessor) para aplicação do Mantis®.

Os dados inseridos no Mantis® até abril/2018 foram copiados para a criação de *dump* (registro da estrutura da tabela ou dados de um banco de dados) para armazenar uma cópia de segurança e realizar os experimentos a partir dos dados registrados.

O procedimento foi feito com comandos de “*backup*” fornecidos pelo próprio gerenciador de banco de dados MySQL. Logo após foi feito “*restore*” desse *backup* em uma outra máquina, com um novo banco de dados MySQL (o mesmo utilizado anteriormente). Procedeu-se à análise de quais tabelas seriam úteis e relevantes para o trabalho. Somente as tabelas que possuem as informações dos atendimentos foram selecionadas, sendo elas:

- a) *mantis\_bug\_table*;
- b) *mantis\_project\_table*;
- c) *mantis\_category\_table*;
- d) *mantis\_bug\_text\_table*;
- e) *mantis\_bugnote\_table*;
- f) *mantis\_bugnote\_text\_table*.

Os dados foram extraídos dessas tabelas por meio de sentenças *Structured Query Language* (SQL), o que possibilitou a geração bruta da informação utilizada posteriormente.

### 6.3 Montagem do *corpus*

Segundo Galisson e Coste (1983 como citados em Aluísio & Almeida, 2006, p. 157):

*Corpus* é um conjunto finito de enunciados tomados como objeto de análise. [...] Trata-se de uma coleção de documentos quer orais (gravados ou transcritos) quer escritos, quer orais e escritos, de acordo com o tipo de investigação pretendido. [...]

Já Trask (2004 como citado em Aluísio & Almeida, 2006, p. 157) opina: “*corpus* é um conjunto de textos escritos ou falados numa língua, disponível para análise”.

Para Ducrot e Todorov (2001 como citados em Aluísio & Almeida, 2006, p. 157): “é um conjunto, tão variado quanto possível, de enunciados efetivamente emitidos por usuários da referida linguagem determinada época”.

E Sinclair (2005 como citado em Aluísio & Almeida, 2006, p.156) defende: “*A corpus is a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research*”.

A moderna noção de *corpus* possui pelo menos quatro características fundamentais: a) amostragem e representatividade; b) tamanho finito; c) formato eletrônico; d) referência-padrão (McEnery & Wilson, 1996 como citados em Aluísio & Almeida, 2006). E para sua a compilação há três estágios principais a seguir: a) projeto do *corpus*; b) a compilação (ou captura), manipulação, nomeação dos arquivos de textos e pedidos de permissão de uso; c) anotação (Aluísio & Almeida, 2006). Para o presente trabalho foram utilizados os seguintes estágios:

### 6.3.1 Projeto do corpus

Foi utilizado o banco de dados do sistema Mantis adotado na Organização Verdemar Ltda.

### 6.3.2 Compilação e manipulação do corpus

A compilação e manipulação do *corpus* foram feitas de maneira automática pelo conceito de *Extract Transform Load* (ETL - Extração, Transformação e Carregamento) no *software Pentaho Data Integration* (PDI, 2018), fornecido pela Hitachi. Foi feita a limpeza do *corpus*, expurgando-se imagens, gráficos, tabelas e qualquer tipo de anexo. Para a extração e conversão foi realizado processo de transformação em um arquivo final de texto (.TXT). Já no formato necessário para importação, na ferramenta *QNA Maker*. Por padrão o *QNA Maker* espera receber um arquivo com três colunas: pergunta, resposta, *meta-tag* (sendo este opcional).

As anotações, segundo Aluísio e Almeida (2006), podem ser classificadas como estrutural (a marcação de dados externos e internos dos textos) e linguística (níveis morfossintático, sintático, semântico, discursivo, entre outros):

- a) Estrutural: os dados externos são utilizados para documentação do *corpus* na forma de um cabeçalho que inclui os metadados textuais, dados bibliográficos comuns, dados de catalogação, como tamanho do arquivo, tipo da autoria, a tipologia textual e informação sobre a distribuição do *corpus*. E os dados internos são as anotações de segmentação do texto em si: i) marcação da estrutura geral (capítulos, parágrafos, títulos e subtítulos, notas de rodapé e elementos gráficos como tabelas e figura); ii) marcação da estrutura de subparágrafos (sentenças, citações, palavras, abreviações, nomes, referências, datas e ênfases tipográficas do tipo negrito, itálico, sublinhado, entre outros);
- b) Linguística: nos níveis morfossintático, sintático, semântico, discursivo, entre outros. Possui três formas: manualmente (por linguistas), automaticamente (por ferramentas de processamento de língua natural – PLN) ou semiautomaticamente (correção manual da saída de outras ferramentas).

Para o presente trabalho foi utilizada a anotação linguística com formato semiautomático.

### **6.3.3 Extração e conversão do banco**

Para realizar a conversão dos dados, foi utilizado o conceito de ETL. Trata-se de ferramentas de *software* cuja função é a extração de dados de diversos sistemas, transformação desses dados conforme regras de negócios e, por fim, o carregamento dos dados.

Nesse caso, o *software* PDI (2018) foi escolhido. A ferramenta disponibiliza elementos que representam diversos objetos, como tabelas, colunas, conexões com bancos de dados, abertura e leitura de arquivos de diversos tipos (XML, TXT, CSV, XLS, DOC), integrações com *webservices*, etc. Também possui elementos pré-configurados, como funções *Java Script* e *Java* para execução e tratamento das informações, além de instruções SQLs genéricas e específicas para cada tipo de banco de dados.

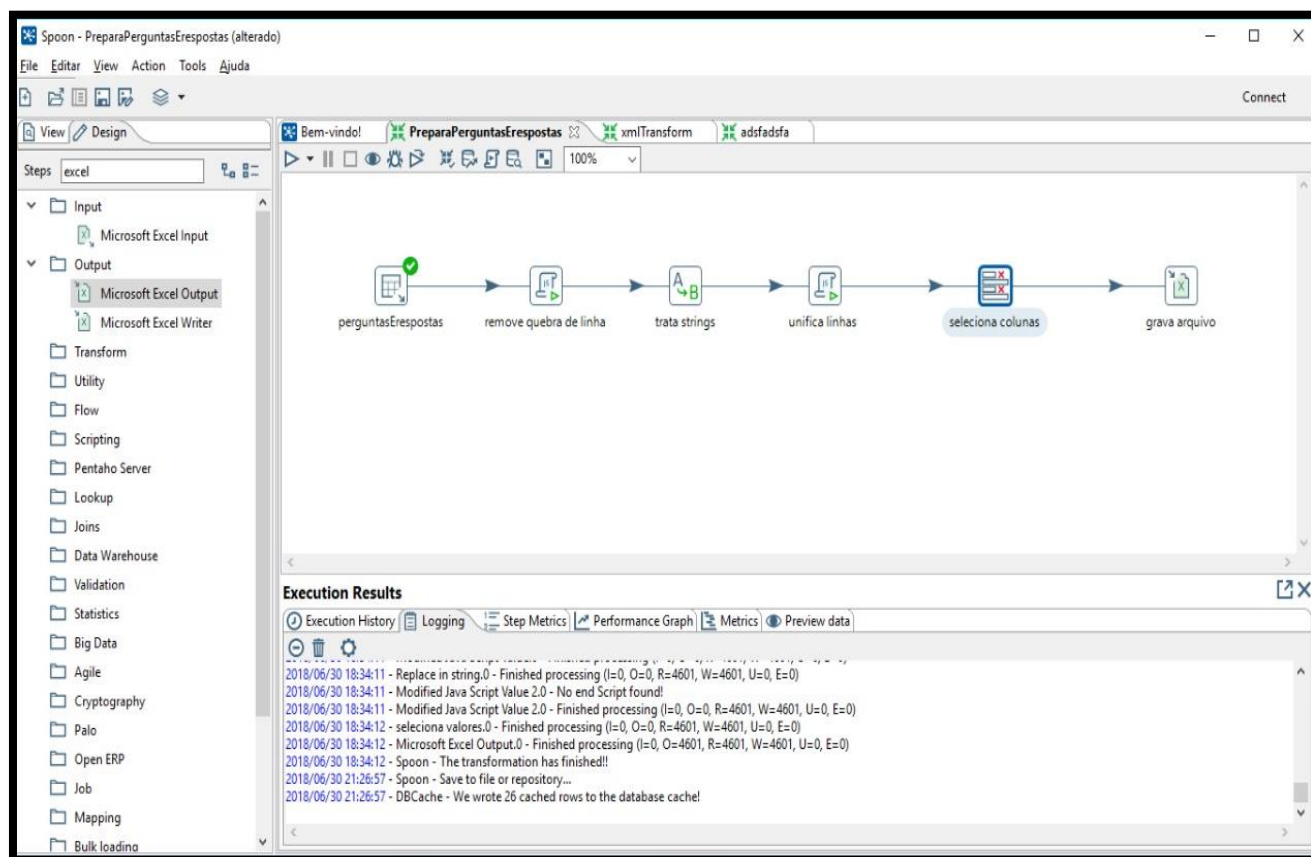
Com esse componente pronto na biblioteca-padrão, foi inserido o desenho de um diagrama, arrastando e soltando os componentes na página, ligando as tabelas umas às outras e definindo um fluxo e a ordem de execução de cada comando. Dessa forma, foi possível pré-visualizar o resultado do processo e ter uma visão concreta de como será o resultado da extração e transformação dos dados.

O diagrama representa a seleção dos dados em uma ou mais tabelas, o filtro de algum tipo de dado, como, por exemplo, a normalização das letras em maiúsculo ou minúsculo ou o relacionamento com alguma outra tabela e a saída para um arquivo ou até mesmo outra tabela no banco de dados. Cada diagrama é conhecido como uma transformação.

Em sequência, foi feita a conexão do banco com a transformação. As transformações podem ser executadas manual ou automaticamente, a partir de um agendamento fornecido pela ferramenta, que pode ser executada tanto em ambiente servidor, como em clientes *desktop*, uma vez que roda sob o *Java*. É necessário ter somente uma *Java Runtime Execution (JRE)* instalado. Neste trabalho foi utilizada a versão para

*desktop*. Logo após as validações de conexão e diagramação, gerou-se, como saída, um arquivo de texto (.TXT) pré-formatado, conforme necessidade do QNA Maker.

Tela de exemplo de “Transformação” utilizada:



**Figura 19** - Exemplo de transformação.

**Observação:** seleciona colunas = selecionar as colunas que possuem informações relevantes e que irão participar da transformação para gerar o arquivo final.

Fonte: Pentaho Data Integration (2018). Hitachi Vantara/Pentaho [site]. Recuperado de: <http://www.pentaho.com/product/data-integration#data-integration->

## 6.4 Elaboração das perguntas x respostas

Segundo Berber Sardinha (2000 como citado em Aluísio & Almeida, 2006), os textos devem ser autênticos. E por autenticidade, compreende-se que os “textos devem ter sido escritos em linguagem natural, não podendo ser textos “produzidos com o propósito de serem alvo de pesquisa linguística” e com representatividade, isto é, ser representativo da língua ou de uma variedade de língua que se deseja pesquisar. De maneira ideal, um “*corpus* deve ser elaborado de forma que represente determinadas

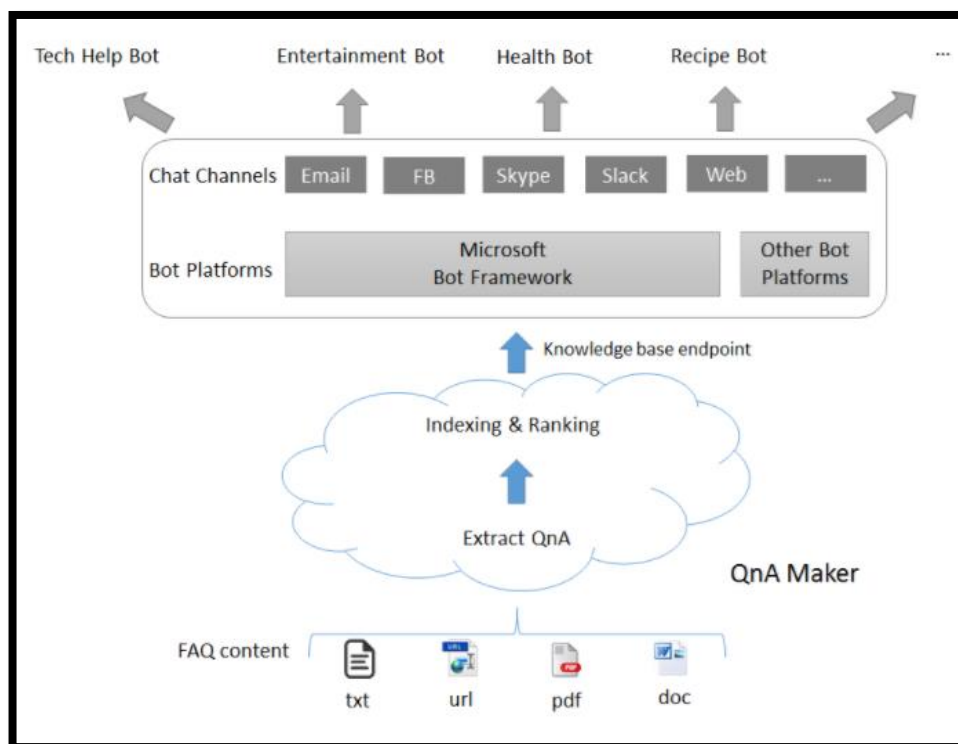
características linguísticas da comunidade cuja língua está sob análise” (Sinclair, 2005 como citado em Aluísio & Almeida, 2006 p.158-159).

Dessa forma, buscando o histórico verídico e autêntico de atendimento já resolvido anteriormente, foi feita a elaboração dos questionamentos para serem inseridos no sistema, tendo as respostas como referência para avaliação dos testes, pelo cruzamento entre as perguntas (casos do Mantis®, campo Resumo) e as possíveis respostas (campo Anotações). Essa é a forma como o Mantis® trabalha com a informação.

Nele é possível registrar várias anotações para a mesma pergunta. Isso aumenta o universo para análise. Outros campos como “data de inclusão”, “categoria”, “projeto” e “ID do caso” foram adicionados ao *corpus*, para aumentar as possibilidades de contextualização, utilizados pela ferramenta no momento de análise para formulação das perguntas e possíveis respostas.

Para analisar o desempenho das respostas, foi utilizado o serviço da Microsoft Azure (2018) a partir da *Application Programming Interface* (API - Interface de Programação de Aplicações ou Interface de Programação de Aplicação). Trata-se de instruções e padrões de programação para acesso a um aplicativo ou *software* do *QnA Maker*, utilizando-se o próprio *corpus* do Mantis®. Veja-se o esquema de processamento do *QnA Maker* na Figura 20.





**Figura 20** - Esquema QnA Maker.

Fonte: Microsoft Azure (2016). *QnA Maker Overview*. Publicado em 04 de dezembro. Microsoft Azure. [site] Recuperado de: <https://docs.microsoft.com/pt-br/azure/cognitive-services/qnamaker/home>.

Após a importação do arquivo na ferramenta QnA Maker da Microsoft, foi efetuada análise que retorna um formulário com sugestões de perguntas e as melhores respostas possíveis, de forma automática. Entretanto, é possível ainda fazer correções e as adequações necessárias para melhor caracterização tanto da pergunta, quanto da resposta. Trata-se de uma API de transferência de estado representacional, com pacotes gratuitos até um determinado limite, baseada na Web, que consiste em princípios/regras/restrições que permitem a criação de um projeto com interfaces bem definidas. É compatível com plataformas de desenvolvimento, serviços de hospedagem e canais.

As correções e adequações podem ser feitas inserindo-se outras perguntas e utilizando o sistema em nível de conhecimento, que ficará disponível gratuitamente no Microsoft Azure, com acesso controlado a partir de um link restrito, para realizar testes diversos e avaliar de maneira geral o sistema. Essa avaliação foi contemplada como observações.

Também é possível treinar a ferramenta para identificar melhor as perguntas e expressões informadas pelos usuários, como, por exemplo, categorias ou uma determinada palavra-chave que remeta a alguma sessão específica da base de conhecimento. Este processo é feito manualmente, avaliando cada pergunta e resposta e realizando marcações *de meta-tags* para identificação e agrupamento. A partir disso é possível obter mais assertividade entre as questões solicitadas pelos usuários.

#### **6.4.1 Campos do Mantis® – utilização nas perguntas e respostas**

O Mantis® possui vários campos predeterminados relativos a seu núcleo nativo do *software*. No entanto, permite que campos personalizados sejam criados para customizar sua utilização e aumentar as possibilidades de uso. Para elaboração das perguntas e respostas do sistema a ser desenvolvido, os principais campos utilizados serão os predeterminados: ID do caso (número de registro no sistema, autoincremental), data de inclusão, categoria (trata-se da classificação do caso: melhoria, falha, *select*, banco de dados, entre outros), projeto (trata-se da descrição do projeto, uma subdivisão, pode haver vários projetos registrados), resumo (trata-se de uma exposição sucinta da solicitação) e anotações (andamento da solicitação até sua resolução final).

Os campos que se referem a frequência, gravidade, prioridade, solicitante, versão, desenvolvedor responsável, loja, canal de abertura (*E-mail*, Mantis, telefone, entre outros), esforço em horas para desenvolvimento, impacto para o cliente, entre outros campos predeterminados ou customizados que serão descartados porque não irão agregar aos objetivos do trabalho.

Para exemplificar o motivo do descarte dos campos, veja-se o exemplo de um dos registros do sistema Mantis®, caso 18358, que relata a reclamação de um solicitante que informa que haver problemas nas atividades de conferência de preço da loja 3, localizada na Avenida Mário Werneck, número 1.500, no bairro Buritis. As atividades não foram geradas no coletor de dados para execução. A equipe de suporte sistêmico realiza a primeira análise e posteriormente repassa para a equipe de desenvolvedores para análise final. O desenvolvedor verifica e retorna o caso, informando que houve

um problema no banco de dados MySQL responsável pela carga de dados das atividades e que após sua estabilização as atividades voltaram a funcionar normalmente.

Para a elaboração das perguntas e análise das respostas, somente os campos: "ID" do caso, "resumo", "categoria" e "anotações" foram levados em consideração porque possuem os principais dados entendimento do problema e o procedimento utilizado como solução do problema.

Os demais campos foram desconsiderados: projeto, visibilidade, data de envio, última atualização, relator, prioridade, gravidade, frequência, estado, plataforma, SO, versão SO, versão do produto, previsto para a versão, corrigido na versão, passos para reproduzir, desenvolvedor, efetivação, loja abertura, canal abertura, interrupção, previsão fechado, revisão, classificação *update*, módulo e solução (campo criado recentemente com a grande maioria dos dados vazios ou preenchidos incorretamente, por exemplo, "OK").

O Quadro 5 mostra que os campos que serão descartados não apresentam conteúdo representativo para localização de uma solução ao problema relatado. Por esse motivo, foram analisados quais campos realmente possuem valor agregado para a verificação do sistema de perguntas e respostas.

Quadro 5 - Campos do Mantis® - continua

Descrição Campo	Tipo Campo	Utilizado na Empresa?	Preenchimento Campo	Campo Descartado?
Número Caso	Padrão	Sim	18358	Não
Projeto	Padrão	Sim	Manutenção	Sim
Categoria	Padrão	Sim	Falha	Não
Visibilidade	Padrão	Não	Público	Sim
Data de Envio	Padrão	Sim	02/01/2018 08:24 hs	Sim
Última Atualização	Padrão	Sim	27/04/2018 10:37 hs	Sim
Relator	Padrão	Sim	matheus.silva	Sim
Atribuído a	Padrão	Sim	matheus.silva	Sim
Prioridade	Padrão	Não	Normal	Sim
Gravidade	Padrão	Não	Pequeno	Sim
Frequência	Padrão	Não	Não se tentou	Sim
Estado	Padrão	Sim	Retorno/Fechado	Sim
Resolução	Padrão	Não	Reaberto/Fechado	Sim
Plataforma	Padrão	Sim	Dep. TI	Sim
SO	Padrão	Sim	Viviâne	Sim
Versão SO	Padrão	Não	-	Sim
Versão do produto	Padrão	Não	-	Sim
Previsto para a versão	Customizado	Sim	-	Sim
Corrigido na versão	Customizado	Sim	18.01	Sim
Resumo	Padrão	Sim	Sistema não gerou atividades de conferência de preço para a loja 03	Não
Descrição*	Padrão	Sim	Veja detalhamento abaixo	Sim
Passos para reproduzir	Padrão	Sim	-	Sim
Marcadores	Padrão	Sim	Semana de análise	Sim
Informações adicionais	Padrão	Sim	-	Sim
Desenvolvedor	Customizado	Sim	Renato	Sim
Responsável DRU	Customizado	Sim	-	Sim
Efetivação	Customizado	Sim	100	Sim
LojaAbertura	Customizado	Sim	3	Sim
Canalbertura	Customizado	Sim	Presencial	Sim
Esforço	Customizado	Sim	1	Sim
Impacto cliente	Customizado	Sim	-	Sim
Interrupção	Customizado	Sim	Interrupção	Sim
Previsão DRU	Customizado	Sim	-	Sim
Previsão resolvido	Customizado	Sim	-	Sim
Previsão fechado	Customizado	Sim	-	Sim
DRU entregue	Customizado	Sim	-	Sim
Entrega DRU	Customizado	Sim	-	Sim

Quadro 6 - Campos do Mantis® - conclui

Descrição Campo	Tipo Campo	Utilizado na Empresa?	Preenchimento Campo	Campo Descartado?
Revisão	Customizado	Sim	-	Sim
Classificação <i>update</i>	Customizado	Sim	Nenhum	Sim
Base	Customizado	Sim	-	Sim
Requer treinamento	Customizado	Sim	-	Sim
Esforço DRU	Customizado	Sim	-	Sim
Módulo	Customizado	Sim	Ciclone	Sim
Arquivos anexados	Padrão	Sim	-	Sim
Anotações*	Padrão	Sim	Foram feitas três anotações. Veja detalhamento abaixo	Não
Solução	Customizado	Sim	-	Sim

\*Descrição – Detalhamento das informações descritas nesse campo:

#### SUPORTE DE SISTEMA - ERRO:

\*\*\*\*\*

Cenários Simulados - COM ERROS - Atividades de Conferência de Preço

Descrição do problema: sistema não gerou atividades de conferência para a loja 03, atividade "LJ03 - 1 CONF ALT PRECO". Verificamos que foram geradas atividades para todas as lojas, menos para a loja 03. Mesmo executando o agendamento sistema não gerou a atividade.

Simulado em: base exemplo? Base produção? Base de produção

Passo a passo para simulação: ciclone: <http://10.0.20.16:8080/CicloneV2GUI/> [^]

Mensagem de erro: em anexo. Usuário utilizado para testes: Matheus.Silva

Auxílio necessário para análise? Não. Qual dificuldade? Nenhuma.

Versão do sistema. (Informe os dados da versão onde o cenário foi simulado).

Versão do sistema: Tirar *print* Ajuda Sobre

\*\*\*\*\*

\*Anotações – detalhamento das anotações:

#### **Anotação 1 - (0118349) - tempo gasto: 01:30**

Em análise do processo foi verificado que a atividade não gerou devido à carga da loja 03 não ter sido enviada.

Realizada análise e verificada a tabela do semáforo TB\_BDCOMUM\_AGEND\_CARGA, verificado que o *status* da carga estava igual a 3.

Verificado o *select* 1405 e não foi identificado nada parado.

Liguei para o Caio confirmar a situação.

Em análise verificado que a situação está como se o *E-Conect* tivesse importado a carga.

Passado para o Jair verificar.

#### **Anotação 2 - (0118372) - Tempo gasto: 00:05**

Verificado com o Jair e ele me disse que deu *Time out* no MySQL na importação da carga.

Com isso não é uma falha.

Segue caso para priorizar!

#### **Anotação 3 - (0118753) – Tempo gasto: 00:05**

Após a estabilização do MySQL a atividade foi gerada corretamente.

Não houve problemas no *Superus*.

Caso finalizado.

Observação: a resposta desejada, que o sistema de perguntas e respostas deve localizar, está na anotação 1, que relata que houve um problema na carga de dados e que a tabela que foi verificada é a TB\_BDCOMUM\_AGEND\_CARGA. Logo, o atendimento seria mais direcionado e essa primeira hipótese de solução já seria investigada

Fonte: elaborado pela autora.

**mantis**  
BUG TRACKER

Acessando como: viviane (Viviane Almeida - desenvolvedor) 27-04-2018 10:38 BRT Projeto: Todos os Projetos [Mais]

Principal | Minha Visão | Ver Casos | Relatar Caso | Registro de Mudanças | Planejamento | Minha Conta | Sair

Recentemente Visitado: 0018335, 0018279, 0018244, 0018206, 0018202

[ Voltar ao Caso ]

Núm	Projeto	Categoria	Visibilidade	Data de Envio	Última Atualização
0018338	Manutenção	[Todos os Projetos] Falha	publico	02-01-2018 08:24	27-04-2018 10:37

Relator: [matheus.silva](#) [Alterar]

Atribuído a: [matheus.silva](#)

Prioridade: normal Gravidade: pequeno Freqüência: não se tentou

Estado: retorno Resolução: reaberto

Plataforma: Dep. TI SO Viviane Versão SO:

Versão do Produto: Corrigido na Versão: 18.01

Resumo: Sistema não gerou atividades de conferência de preço para a loja 03

Descrição:
 

```

  SUPORTE DE SISTEMA
  ERRO:
  .....
  <b> Cenários Simulados - COM ERROS </b>
  atividades de Conferência de Preço
  <b> Descrição do problema</b>
  Sistema não gerou atividades de conferência para a loja 03, atividade "1303 - 1 CONF ALT PRECO".
  Verificamos que foram geradas atividades para todas as lojas, menos para a loja 03.
  Mesmo executando o agendamento o sistema não gerou a atividade.
  <b> Simulado em: Base Exemplo? Base Produção? </b>
  Base de Produção
  <b> Passo a Passo para simulação </b>
  Cíclones: http://39.46.26.18:8080/ciclonev2001/
  <b> Mensagem de Erro </b>
  Em anexo.
  <b> Usuário utilizado para testes</b>
  matheus.silva
  <b> Auxílio necessário para análise? </b>
  Não.
  <b> Qual dificuldade? </b>
  Nenhuma.
  .....
  <b> Versão do Sistema. </b>
  (Informe os dados da versão onde o cenário foi simulado).
  Versão do Sistema: Tirar print Ajuda Sobre/<b>
  .....
  
```

Passos para Reproduzir:

Informações Adicionais:

Desenvolvedor:  Agnaldo  Caio  Danilo  Lazaro  Eduardo  Lorrain  Renato  Tiago  Não Definido  Rodrigo  Talita  Mauricio

Efetivação: 100

Loja Abertura:  01  02  03  04  05  06  07  08  09  10  11  12  13  14  15  CD

Canal Abertura:  Email  Presencial  Telefone  Mantis

Interrupção:  Não  Interrupção  Retrabalho  Programado

Previsão Fechado:  Sim  Não

Revisao:  Sim  Não

Classificação Update:  Falha  Erro Operacional  Nenhum

Modulo: Ciclone

Adicionar Anotação:

Atualizar Informação

**Anotações**

[0118333] [viviane](#) (desenvolvedor) 08-01-2018 15:32  
Após a estabilização do Mysql a atividade foi gerada corretamente. Não houve problemas no Superus. Caso finalizado.

[0118372] [renato.almeida](#) (administrador) 02-01-2018 17:00 alterado em: 02-01-2018 17:00  
**Tempo gasto: 00:05**  
Verificado com o Jair e ele me disse que deu Time out no mysql na importação da carga. Com isso não é uma falha. Segue caso para priorizar!

[0118349] [renato.almeida](#) (administrador) 02-01-2018 10:45 alterado em: 02-01-2018 17:00  
**Tempo gasto: 01:30**  
Em análise do processo verificado que a atividade não gerou devido a carga da loja 03 não ter sido enviada. Realizada análise e verificado a tabela do semáforo TB\_BDCOMUM\_AGEND\_CARGA, verificado que o status da carga estava igual a 3. Verificado o select 1405 e não foi identificado nada parado. Liguei para o Caio confirmar a situação. Em análise verificado que a situação esta como se o e-Connect tivesse importado a carga. Passado para o Jair verificar.

Tempo total para o caso = 01:35

Copyright © 2000 - 2012 MantisBT Group  
www.mantisbt.org

Figura 21 - Exemplo caso Mantis – visualização interna.

Fonte: Verdemar. (2017). *Documento recursos humanos: Manual Treinamento de Integração. Treinamento e Desenvolvimento*. Recuperado de: <http://trabalhenoverdemar.blogspot.com.br>

The screenshot displays the Mantis Bug Tracker interface for a specific bug report. At the top, the Mantis logo and 'BUG TRACKER' are visible. The user is logged in as 'viviane (Viviane Almeida - Desenvolvedor)' on '27-04-2018 10:30 BRT'. The project is set to 'Todos os Projetos'. Navigation links include 'Principal', 'Minha Visão', 'Ver Casos', 'Relatar Caso', 'Registro de Mudanças', 'Planejamento', 'Minha Conta', and 'Sair'. A search bar and a 'Recientemente Visitado' list are also present.

The main section is titled 'Ver Detalhes do Caso' and includes a table with the following data:

Núm	Projeto	Categoria	Visibilidade	Data de Envio	Última Atualização
0018358	Manutenção	[Todos os Projetos] Falha	público	02-01-2018 08:24	08-01-2018 15:33

Below the table, the bug details are listed:

- Relator:** matheus.silva
- Atribuído a:** matheus.silva
- Prioridade:** normal
- Estado:** fechado
- Plataforma:** Dep. T.1
- Versão do Produto:** SD
- Previsão para a Versão:** Corrigido na Versão 18.01
- Gravidade:** pequeno
- Resolução:** corrigido
- Frequência:** não se tentou
- Versão SO:** Viviane

The **Resumo** section contains the following text:

0018358: Sistema não gerou atividades de conferência de preço para a loja 03

**Descrição**

SUPORTE DE SISTEMA

ERRO:

-----

**Cenários Simulados - COM ERROS**  
Atividades de Conferência de Preço

**Descrição do problema**  
Sistema não gerou atividades de conferência para a loja 03, atividade "L003 - 1 CONF ALT PRECO".

Verificamos que foram geradas atividades para todas as lojas, menos para a loja 03. Mesmo executando o agendamento o sistema não gerou a atividade.

**Simulado em: Base Exemplo? Base Produção?**  
Base de Produção

**Passo-a-Passo para simulação**  
Ciclone: <http://10.0.20.16:8080/CicloneV2GUI/>

**Mensagem de Erro**  
Em anexo.

**Usuário utilizado para testes**  
Matheus-Silva

**Auxílio Necessário para análise?**  
Não.

**Qual dificuldade?**  
Nenhuma

-----

**Versão do Sistema.**  
(Informe os dados da versão onde o cenário foi simulado).

**Versão do Sistema: Tirar print Ajuda Sobre**

-----

**Marcadores**  
A DES 01\_31 A 06/01/2018 X

**Aplicar Marcadores** (Separar por '/')  
[Aplicar] Marcadores atuais

**Desenvolvedor**  
Renato

**Responsável\_DRU**  
100

**Efetivação**  
03

**LojaAbertura**  
Presencial

**CanalAbertura**  
1

**EsforçoTI**  
5

**ImpactoCliente**  
Interrupção

**Previsão DRU**  
Previsão Resolvido

**Previsão Fechado**  
DRUEntregue Não Necessária com Documentação

**Entrega\_DRU**  
Revisão

**Classificação Update**  
Nenhum

**Base**  
Requer Treinamento

**Esforço DRU**  
Modulo Ciclone

**Arquivos Anexados**  
[Monitorar] [Mover]

**Relações**

**Usuários monitorando este caso**  
Lista de Usuários Não há usuários monitorando este caso.

Nome de usuário [Adicionar]

**Anotações**

[0118358] viviane (desenvolvedor) 08-01-2018 15:32 Após a estabilização do MysqI a atividade foi gerada corretamente. Não houve problemas no Superus. Caso finalizado.

[0118372] renato.almeida (administrador) 02-01-2018 17:00 alterado em: 02-01-2018 17:00 **Tempo gasto: 00:05**  
Verificado com o Jair e ele me disse que deu Time out no mysql na importação da carga. Com isso não é uma falha. Segue caso para priorizar!

[0118349] renato.almeida (administrador) 02-01-2018 10:45 alterado em: 02-01-2018 17:00 **Tempo gasto: 01:30**  
Em análise do processo verificado que a atividade não gerou devido a carga da loja 03 não ter sido enviada. Realizada análise e verificado a tabela do semáforo TB\_BDCOMUM\_AGEND\_CARGA, verificado que o status da carga estava igual a 3. Verificado o select 1405 e não foi identificado nada parado. Liguei para o Caio confirmar a situação. Em análise verificado que a situação esta como se o e-Connect tivesse importado a carga. Passado para o Jair verificar.

Tempo total para o caso = 01:35

**Acompanhamento de tempo**  
Data de Início: 2 Janeiro 2018  
Data de Término: 27 Abril 2018  
[Obter informações de acompanhamento de tempo]

Figura 22 – Exemplo caso Mantis – visualização externa.

Fonte: Verdemar. (2017). *Documento recursos humanos: Manual Treinamento de Integração. Treinamento e Desenvolvimento*. Recuperado de: <http://trabalhenoverdemar.blogspot.com.br>

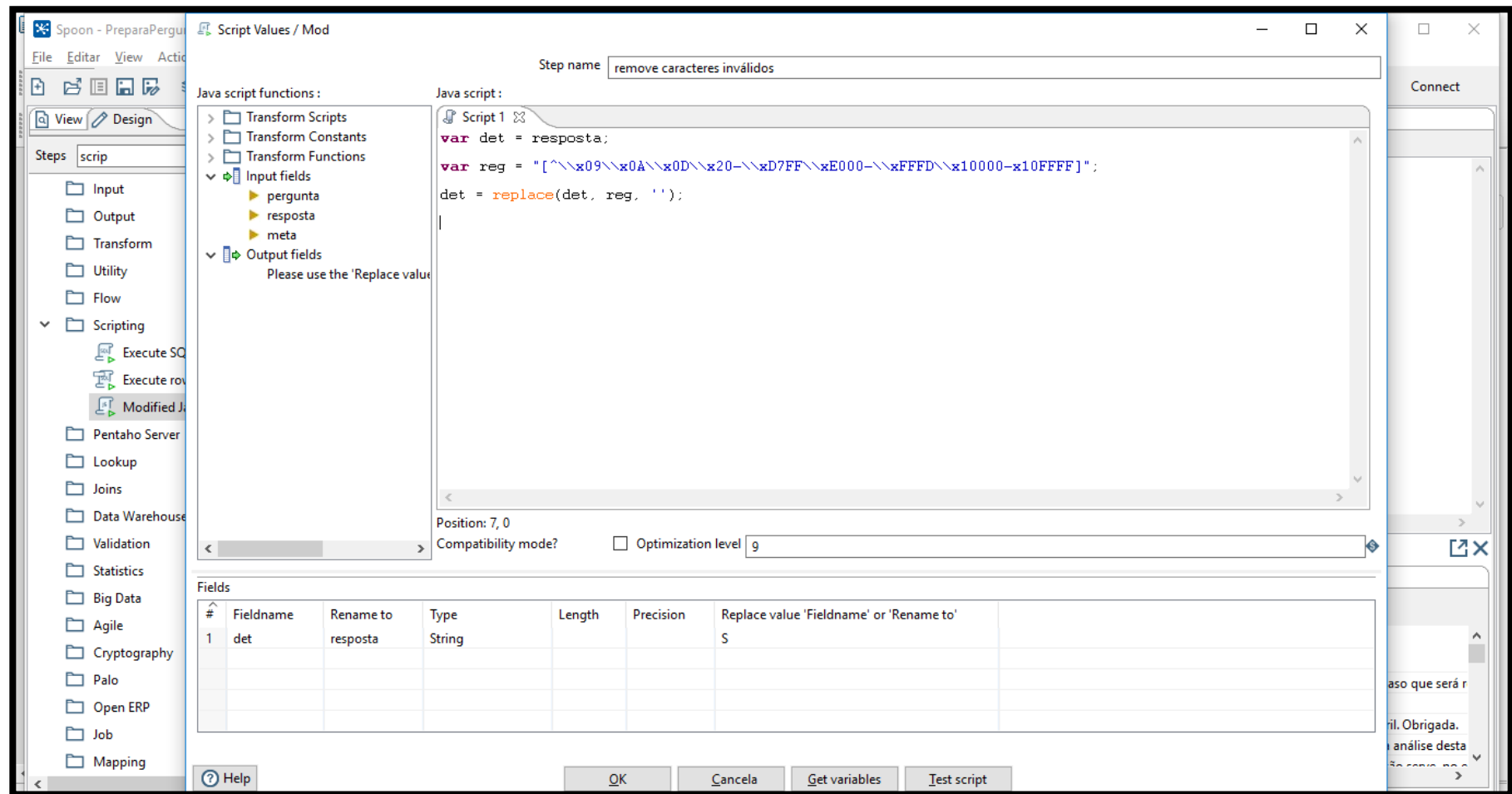
## 6.5 Retirada de caracteres especiais ou inválidos

O vocabulário da linguagem de marcação consiste em palavras, espaços e caracteres especiais, entre eles o asterisco, que funcionam como frase ou palavras e que são considerados “curingas”. A função dos curingas é a substituição de partes dos *strings*, propondo assim padrões de entrada mais abrangentes, para que as respostas não sejam repassadas sempre da mesma forma (Souza & Moraes, 2006). A presença dos caracteres especiais na base de conhecimento é considerada um ruído ou “sujeira”, atrapalhando a verificação do texto e diminuindo sua precisão e desempenho no envio de respostas.

Por esse motivo, a base de dados utilizada foi analisada e os caracteres especiais foram eliminados. O *Pentaho Data Integration* (PDI) possui uma funcionalidade que permite remover acentos e caracteres especiais. Foi utilizado o *step* (passo; etapa; componente da transformação ou fluxo do PDI) “*replace in string*”. Essa etapa permite capturar de forma programática todos os campos e dados enviados da etapa anterior. Por padrão, a ferramenta também oferece várias funções para tratamento de *string*, como, por exemplo, a função “*replace()*”, utilizada neste trabalho para localizar e substituir valores. Por aceitar uma programação simples em *JavaScript*, as funções conhecidas dessa linguagem também podem ser utilizadas para identificação e substituição dos caracteres indesejados, como quebras de linha (“\n”), acentos e tabulações (“\t”), bem como qualquer outro tratamento pertinente.

Veja-se a montagem do *script* de remoção desses caracteres na Figura 23.





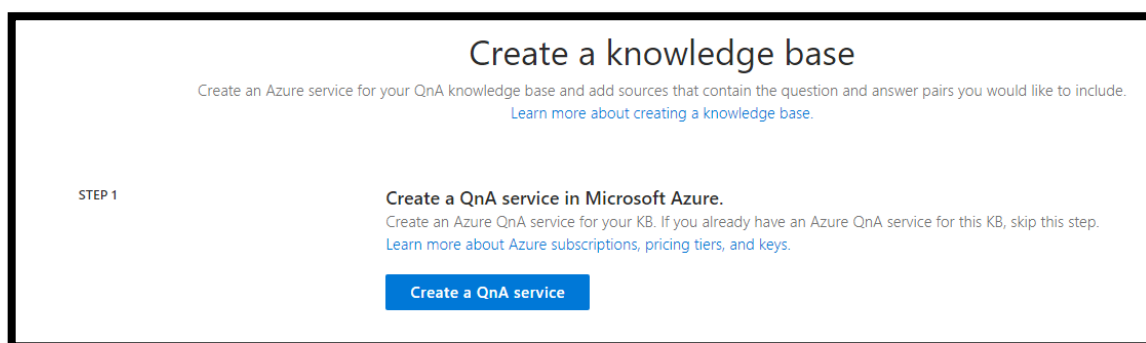
**Figura 23** - Step remoção de caracteres especiais.

Fonte: Pentaho Data Integration (2018). Hitachi Vantara/Pentaho [site]. Recuperado de: <http://www.pentaho.com/product/data-integration#data-integration->

## 6.6 Importação do arquivo TXT/XLSX

Após a geração do arquivo .TXT e sua limpeza de caracteres especiais, foi feita a importação do arquivo no QnA *Maker*, a partir do painel de controle e criação de bases de conhecimentos (“*knowledge-base*”). Sua utilização é restrita, necessita de cadastro prévio. É possível criar várias bases de conhecimento para o mesmo sistema. Nesse momento, o assistente do QnA *Maker*, conduz a importação com base em cinco etapas, sendo elas:

- a) A criação de um novo serviço de QnA na plataforma *Azure*, caso não haja.



**Figura 24** – Criação da base de conhecimento.

Fonte: Microsoft Azure (2018). *API do QnA Maker*. Microsoft Azure. [site] Recuperado de: <https://azure.microsoft.com/pt-br/services/cognitive-services/qna-maker/>.

- b) A conexão do serviço QnA *Maker* com a base de conhecimento foi criada com informações dos campos<sup>11</sup> *Microsoft Azure Directory ID*, *Azure Subscription Name* e *Azure QnA Service*.

<sup>11</sup> *Microsoft Azure Directory ID*: onde é informado o diretório padrão da conta do *Portal Microsoft Azure*. *Azure Subscription Name*: referente à forma de tarifação do serviço. *Azure QnA Service*: onde é informado o nome propriamente do serviço do *Portal Microsoft Azure*.

STEP 2

**Connect your QnA service to your KB.**  
After you create an Azure QnA service, [refresh this page](#) and then select your Azure service using the options below.

\* Microsoft Azure Directory ID  
Diretório Padrão

\* Azure subscription name  
Free Trial

\* Azure QnA service  
QnaVivi

**Figura 25** – Conexão do serviço QnA Maker.

Fonte: Microsoft Azure (2018). *API do QnA Maker*. Microsoft Azure. [site] Recuperado de: <https://azure.microsoft.com/pt-br/services/cognitive-services/qna-maker/>

c) Foi incluso o nome da base de conhecimento (*name your KB*).

STEP 3

**Name your KB.**  
The knowledge base name is for your reference and you can change it at anytime.

\* Name  
qna\_v

**Figura 26** – Inclusão nome para base de conhecimento.

Fonte: Microsoft Azure (2018). *API do QnA Maker*. Microsoft Azure. [site] Recuperado de: <https://azure.microsoft.com/pt-br/services/cognitive-services/qna-maker/>

- d) Foi feita a população da base de conhecimento. Nessa etapa, a importação dos dados também foi realizada. É possível importar dados a partir de uma *Uniform Resource Locator (URL) online* ou de algum arquivo pré-formatado.
- e) Para o presente trabalho, o arquivo pré-formatado foi utilizado. Ao clicar em “*Add File*”, um novo arquivo é enviado para a plataforma e fica disponível para processamento. Trata-se da próxima etapa.

STEP 4

**Populate your KB.**  
Extract question-and-answer pairs from an online FAQ, product manuals, or other files. Supported formats are .tsv, .pdf, .doc, .docx, .xlsx, containing questions and answers in sequence. [Learn more about knowledge base sources](#). Skip this step to add questions and answers manually after creation. The number of sources and file size you can add depends on the QnA service SKU you choose. [Learn more about QnA Maker SKUs](#).

URL

[+ Add URL](#)

File name

[+ Add file](#)

**Figura 27** – População da base de conhecimento.

Fonte: Microsoft Azure (2018). *API do QnA Maker*. Microsoft Azure. [site] Recuperado de: <https://azure.microsoft.com/pt-br/services/cognitive-services/qna-maker/>

- f) Foi concluída a criação da base (*create your KB*): por fim, nesta etapa, após clicar no botão “*create your KB*”, os arquivos importados na etapa 4 foram validados e processados.

STEP 5

**Create your KB**  
The tool will look through your documents and create a knowledge base for your service. If you are not using an existing document, the tool will create an empty knowledge base table which you can edit.

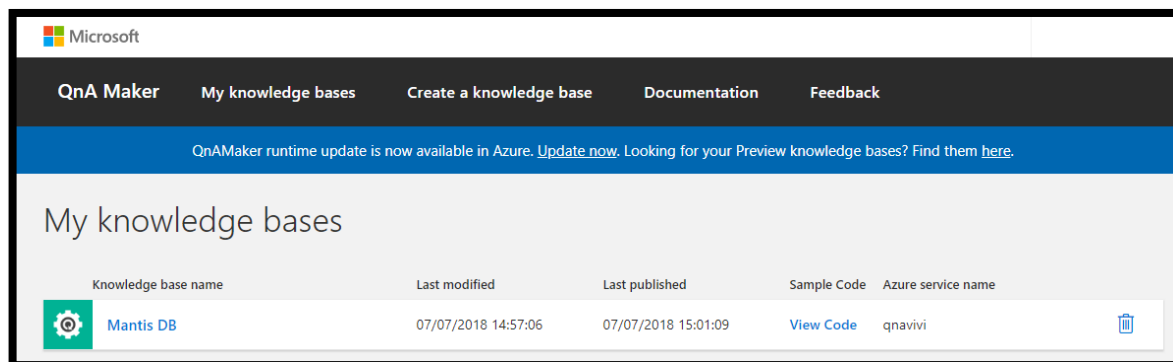
[Create your KB](#)

**Figura 28** - Conclusão da base de conhecimento.

Fonte: Microsoft Azure (2018). *API do QnA Maker*. Microsoft Azure. [site] Recuperado de: <https://azure.microsoft.com/pt-br/services/cognitive-services/qna-maker/>

## 6.7 Adequação e análise das respostas

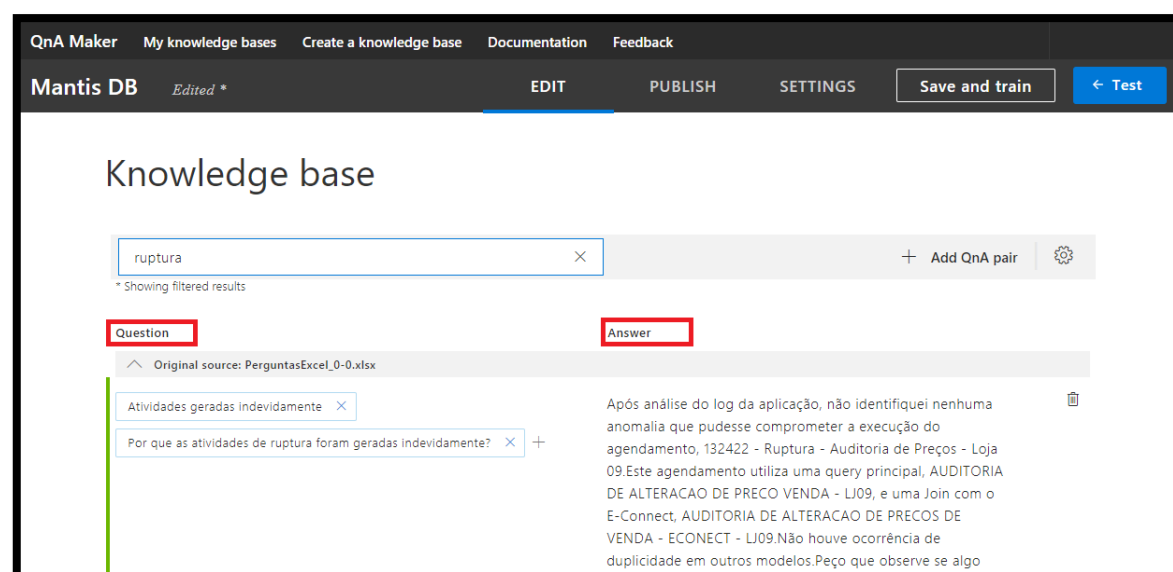
Após a importação do arquivo, o *QnA Maker* gera várias perguntas e suas respectivas respostas. Nesse momento, é possível conferir a relevância (*confidence score*) e editá-las, para melhorar a assertividade da ferramenta.



**Figura 29** – Visualização da base de conhecimento importada.

Fonte: Microsoft Azure (2018). *API do QnA Maker*. Microsoft Azure. [site] Recuperado de: <https://azure.microsoft.com/pt-br/services/cognitive-services/qna-maker/>

Cada novo registro é composto de uma pergunta (*question*) e uma resposta (*answer*). Opcionalmente, *metas-tags* podem ser adicionadas para potencializar a classificação de cada pergunta/resposta. Uma pergunta pode ter vários outros termos associados (*alternative phrase*), para manter o contexto e garantir que a mesma resposta seja alcançada.

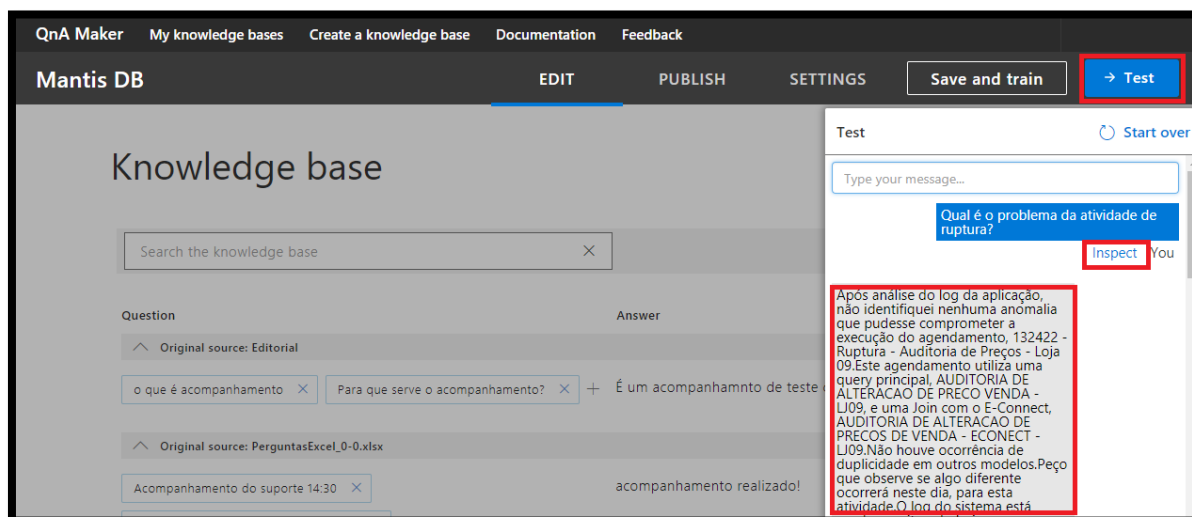


**Figura 30** – Perguntas e respostas da base de conhecimento importada.

Fonte: Microsoft Azure (2018). *API do QnA Maker*. Microsoft Azure. [site] Recuperado de: <https://azure.microsoft.com/pt-br/services/cognitive-services/qna-maker/>

O processo de ter vários termos associados ou várias perguntas ou várias respostas é um procedimento que se repete para todas as linhas da base de conhecimento. As linhas podem ser acrescentadas e editadas para refinamento. Para avaliar se o

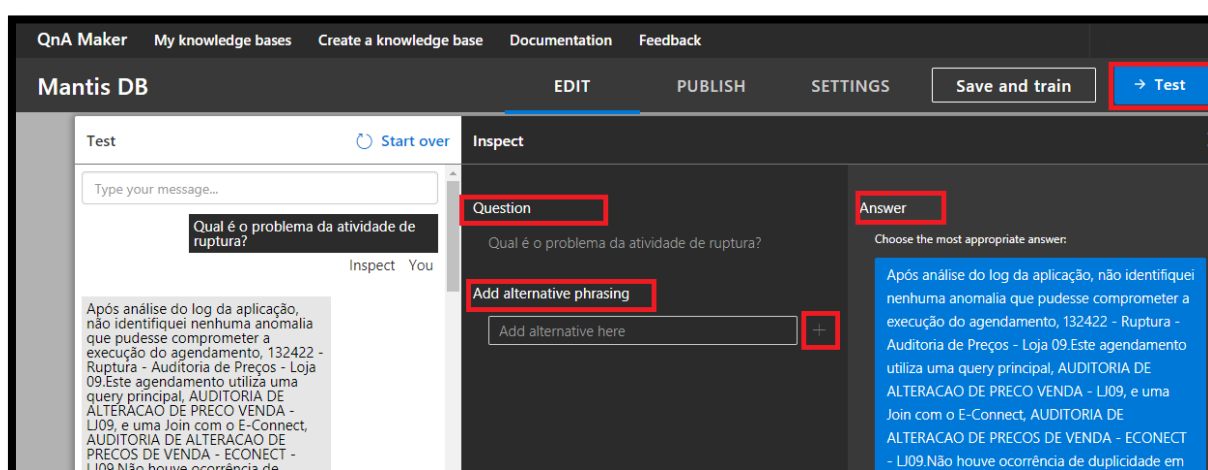
refinamento está sendo feito corretamente, é possível realizar simulações e testes diretamente na ferramenta antes da publicação do *Chatbot*.



**Figura 31** – Realização de testes na base de conhecimento importada.

Fonte: Microsoft Azure (2018). *API do QnA Maker*. Microsoft Azure. [site] Recuperado de: <https://azure.microsoft.com/pt-br/services/cognitive-services/qna-maker/>

O *QnA Maker* ainda permite realizar inspeções nas perguntas e respostas para que os ajustes sejam realizados. Novas perguntas podem ser acrescentadas como perguntas alternativas, ou seja, questionamentos diferentes, porém com a mesma resposta. Esse procedimento deve ser feito na base completa.



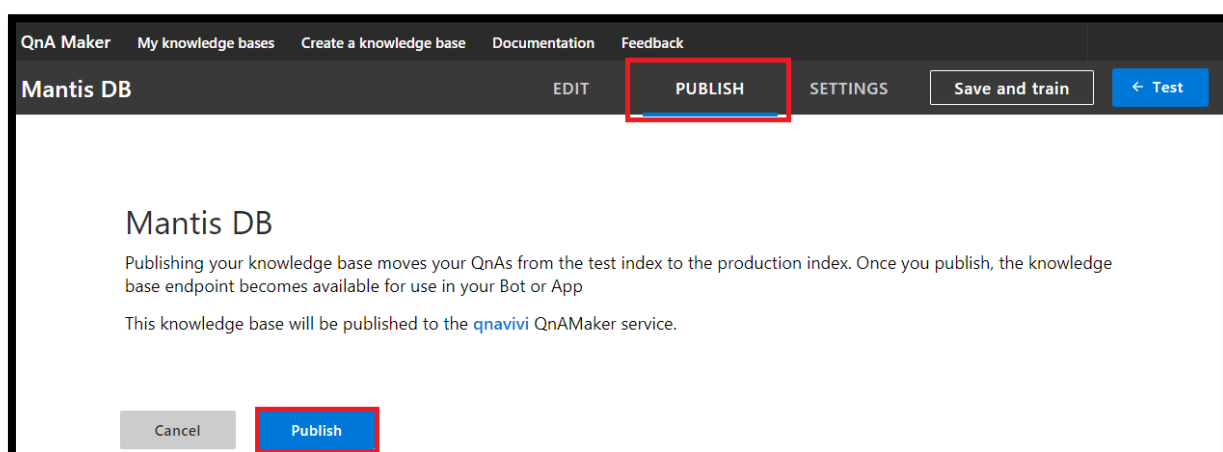
**Figura 32** – Inspeção na base de conhecimento importada.

Fonte: Microsoft Azure (2018). *API do QnA Maker*. Microsoft Azure. [site] Recuperado de: <https://azure.microsoft.com/pt-br/services/cognitive-services/qna-maker/>

Após a análise e os refinamentos da base de conhecimento, é feita sua publicação, para que seja gerado o *chatbot* (é gerado um *link*) para que seja inserida a pergunta. E o sistema apresenta a melhor resposta, de acordo com os refinamentos que foram realizados.

## 6.8 Publicação da base

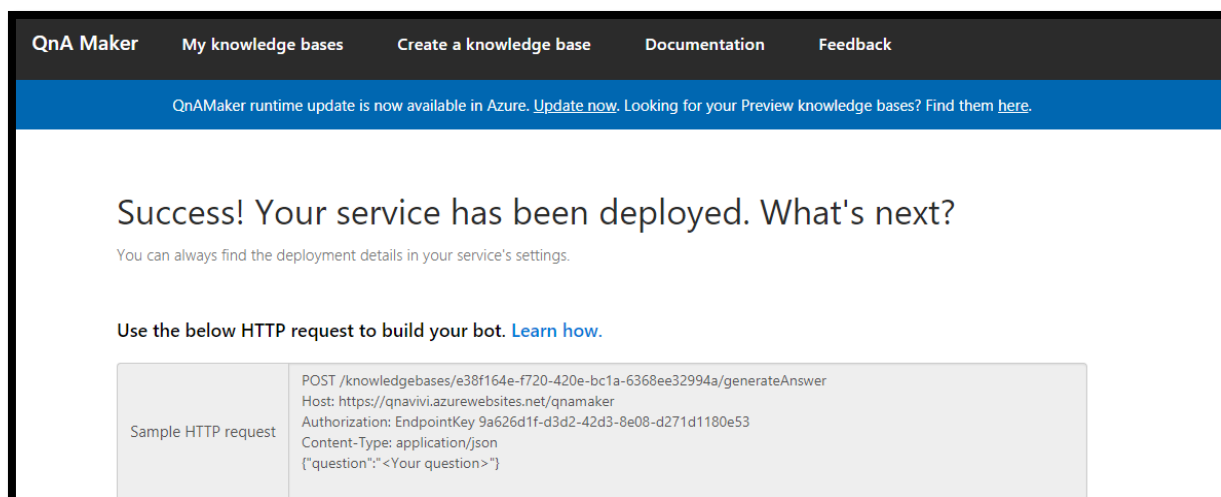
Depois dos refinamentos, é necessário publicar a base de conhecimento. Para isso, foi utilizado o recurso de “*Publish*” do *QnA Maker*. Ele possibilita colocar o *chatbot* para real utilização sem ser no ambiente de testes e inspeções.



**Figura 33** – Recurso de publicação do *QnA Maker*.

Fonte: Microsoft Azure (2018). *API do QnA Maker*. Microsoft Azure. [site] Recuperado de: <https://azure.microsoft.com/pt-br/services/cognitive-services/qna-maker/>

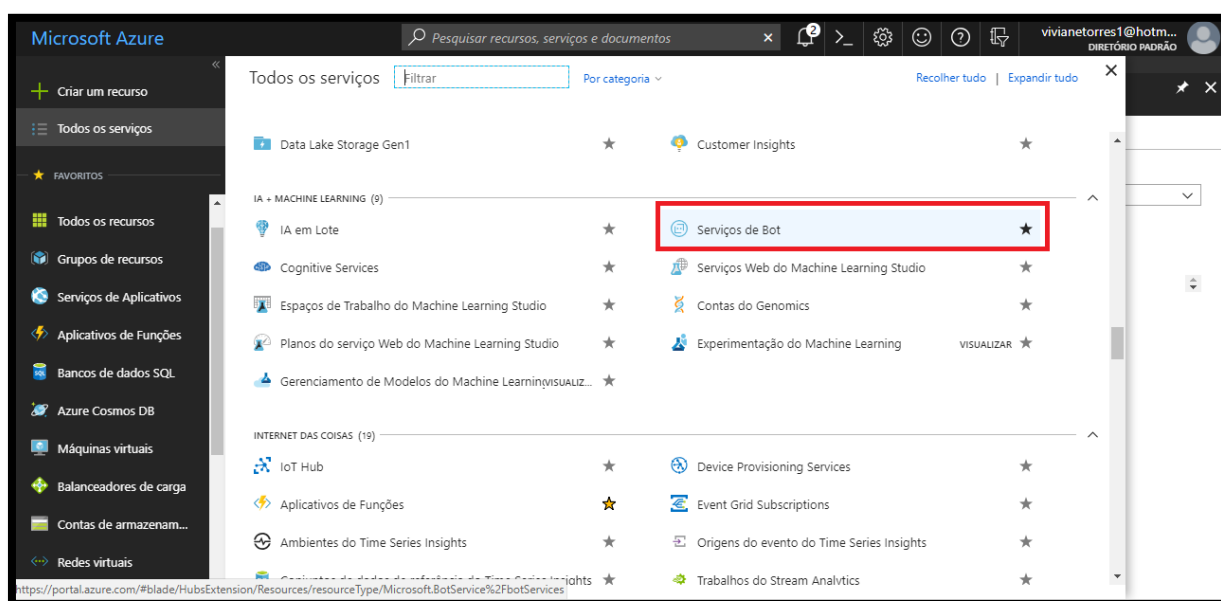
Logo após a publicação foram feitas a configuração das permissões e a liberação do *link*.



**Figura 34** – Publicação da base conhecimento no QnA Maker.

Fonte: Microsoft Azure (2018). *API do QnA Maker*. Microsoft Azure. [site] Recuperado de: <https://azure.microsoft.com/pt-br/services/cognitive-services/qna-maker/>

O *link* apresentado na publicação é reservado para ser incluso nas configurações do Portal *Microsoft Azure*. Por intermédio do portal é possível controlar os serviços e as aplicações da base de conhecimento que foi publicada e configurar o serviço de *bot* (robô).

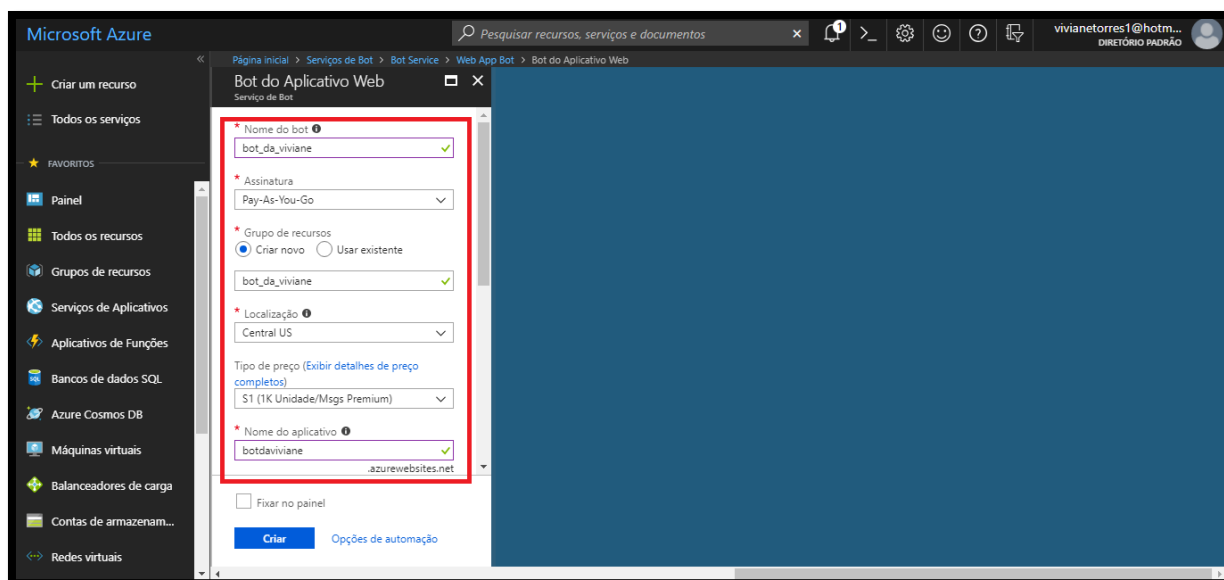


**Figura 35** – Serviço de *Bot*.

Fonte: Microsoft Azure (2018). *API do QnA Maker*. Microsoft Azure. [site] Recuperado de: <https://azure.microsoft.com/pt-br/services/cognitive-services/qna-maker/>

Os dados foram preenchidos, para identificação do *bot* e é a aderência do plano de tarifação por utilização do serviço.

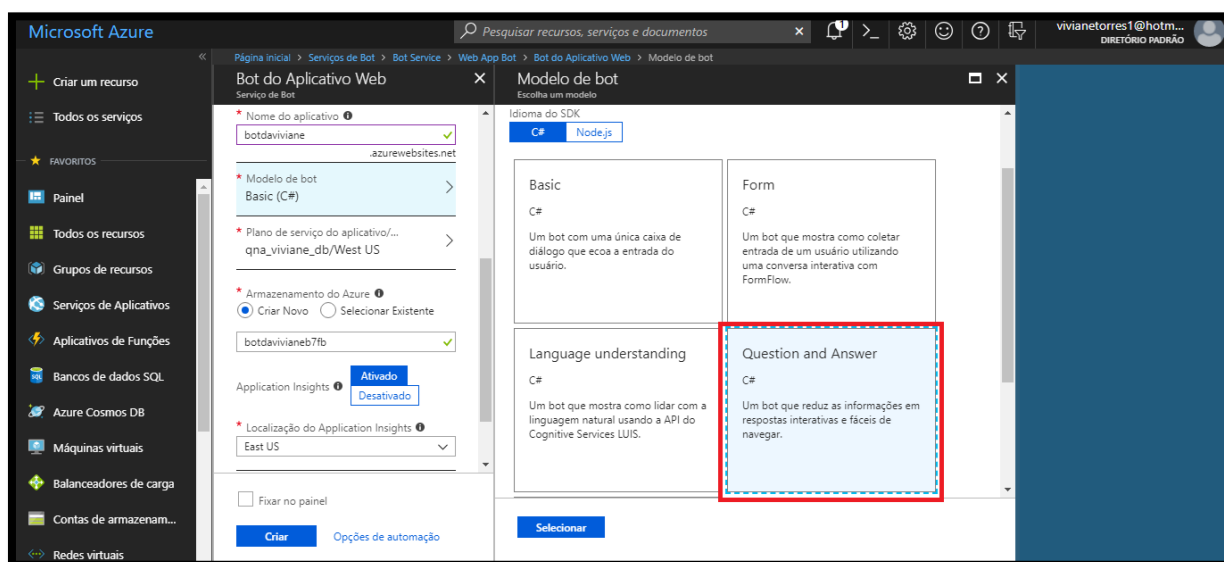




**Figura 36** – Identificação serviço de bot.

Fonte: Microsoft Azure (2018). *API do QnA Maker*. Microsoft Azure. [site] Recuperado de: <https://azure.microsoft.com/pt-br/services/cognitive-services/qna-maker/>

Ainda na criação, foi selecionado o tipo de *bot*: *question and answer* (perguntas e respostas).

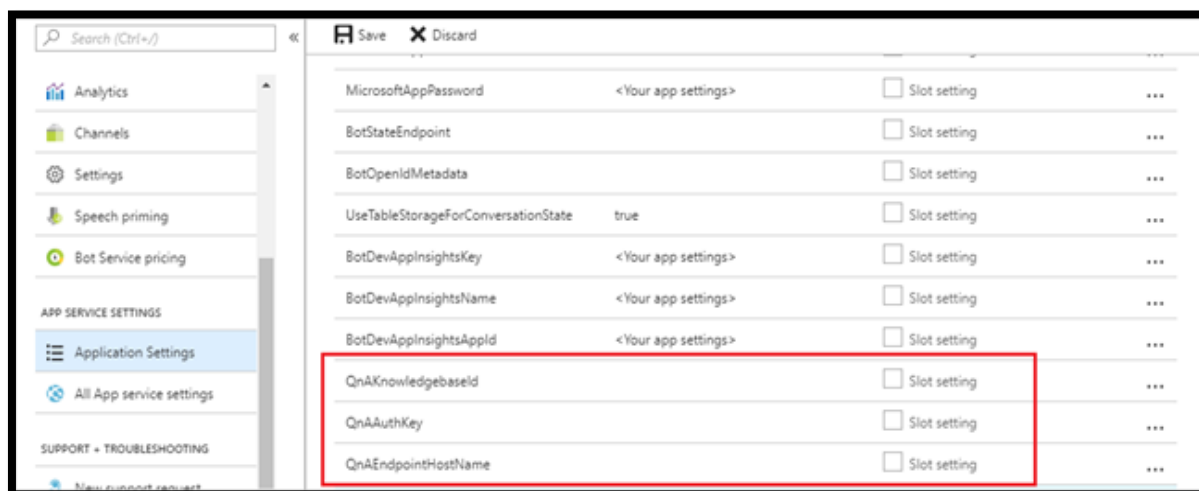


**Figura 37** – Tipo de bot.

Fonte: Microsoft Azure (2018). *API do QnA Maker*. Microsoft Azure. [site] Recuperado de: <https://azure.microsoft.com/pt-br/services/cognitive-services/qna-maker/>

A criação do *bot* foi finalizada e logo após foi feita sua integração com a base do *QnA Maker*, reservada anteriormente. O serviço do *QnA Maker* foi aberto para obter as chaves (*keys*) e então as configuração da aplicação foram feitas (*Application Settings*),

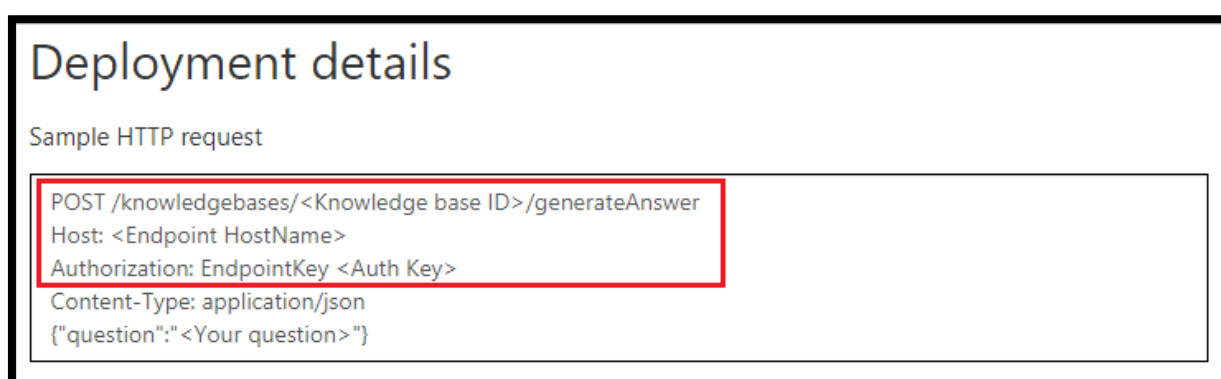
criando-se e editando as propriedades do *QnAKnowledgebaseId*, *QnAAuthKey* e *QnAEndpointHostName*.



**Figura 38** – Configurações do QnA Marker.

Fonte: Microsoft Azure (2018). *API do QnA Maker*. Microsoft Azure. [site] Recuperado de: <https://azure.microsoft.com/pt-br/services/cognitive-services/qna-maker/>

Para obter os valores, foi feito acesso QnA Maker e a partir das propriedades da base de conhecimento foram obtidas as informações de acesso. Veja-se o exemplo de detalhamento da base na Figura 39:



**Figura 39** – Informações de acesso do QnA Maker – antes.

Fonte: Microsoft Azure (2018). *API do QnA Maker*. Microsoft Azure. [site] Recuperado de: <https://azure.microsoft.com/pt-br/services/cognitive-services/qna-maker/>

## 6.9 Funcionamento do QnA Maker

Uma das técnicas que o QnA Maker utiliza é a Seq2Seq. Trata-se de um padrão para codificar e decodificar *corpus* utilizando redes neurais como base. As redes neurais

são estruturas computacionais que mapeiam uma entrada para uma saída com base em uma rede de elementos de processamento altamente conectados (neurônios) (Jones, 2017).

Jones (2017) salienta que uma rede neural recorrente (RNN) é uma classe de redes neurais e pode incluir conexões ponderadas dentro de uma camada. As RNNs também incluem *loops* (repetições) e podem armazenar informações ao processar novas entradas. Essa memória é que as torna ideais para tarefas de processamento cujas entradas anteriores devem ser consideradas. Por essa razão, as redes de *deep learning* atuais são baseadas em RNNs.

Pode-se explorar a aplicação de letras a uma rede neural. As redes neurais operam em valores numéricos, de modo que é necessária alguma representação para alimentar uma letra em uma rede. A codificação *one-hot* (grupo de *bits* com combinação de valores) converte uma letra em um vetor, sendo que um único elemento do vetor é definido. Essa codificação cria uma característica distinta, que pode ser usada matematicamente, por exemplo, cada letra representada obtém seu próprio peso aplicado na rede, enquanto nessa implementação representam-se letras por meio de *one-hot*. Da mesma forma também é possível fazer uma representação por palavras da mesma forma (Jones, 2017).

Essa técnica contém camadas em módulos que podem ser reutilizadas para a construção de um modelo próprio, assim como modelos internos já existentes que funcionam de forma imediata. Esses modelos podem ser compilados da maneira que foram construídos ou permitem incluir camadas para um modelo maior. Cada modelo possui pelos menos duas camadas primárias, o codificador e o decodificador.

Sutskever, Vinyals & Le (2014) alertam que é importante encontrar a codificação do problema que tenha o maior número de dependências de curto prazo, o que torna o problema de aprendizagem muito mais simples.

O codificador codifica a sequência de entrada para uma representação interna chamada “vetor de contexto”, que é usada pelo decodificador para gerar a sequência

de saída. Os comprimentos das sequências de entrada e saída podem ser diferentes, já que não há uma relação explícita entre as sequências de entrada e saída.

As sequências são partes do texto de forma estruturada para encontrar a resposta mais rapidamente. Trata-se da aprendizagem da máquina, porque é como se mantivesse uma memória, semelhante à memória humana.

Memória de Longo Prazo (LSTM - *Long Short-Term Memory*) foi introduzida por Hochreiter & Schmidhuber, em 1997, e é um tipo de redes neurais classificadas como *Recurrent Neural Network* (RNN) especificamente desenvolvidas para prever sequências. Elas funcionam em grande variedade de problemas e podem ser utilizadas amplamente. Já as sequências são estruturas de dados comuns em problemas como tradução de textos para geração de respostas para perguntas escritas (Moura, 2017). Por exemplo:

- a) Sequência de entrada:  $A = [I, love, beer]$ ,
- b) Sequência de saída  $B = [Eu, adoro, cerveja]$ .

É possível ver que são sequências de vetores simples. Chama-se "*time steps*" o tamanho da sequência. Por exemplo, no caso anterior têm-se sequências com *time steps* igual a três. Com isso, é possível classificar o problema de previsão de sequências (seq2seq) em algumas categorias, segundo Moura (2017), como:

- a) *One-to-One*: sequências de um *time step* para sequências de um *time step*;
- b) *Many-to-One*: sequências de  $n$  *time steps* para sequências de um *time step*;
- c) *Many-to-Many*: sequências de  $n$  *time steps* para sequências de  $n$  *time steps*.

O *Encoder-Decoder* é composto por dois modelos: um para ler a sequência de entrada e codificá-la em uma sequência de saída de tamanho fixo (*Encoder*). E outro para decodificar essa sequência de tamanho fixo que foi criada para gerar os resultados (Moura, 2017). A inovação desse modelo é a presença do vetor codificado de tamanho fixo, que é capaz de orquestrar a comunicação entre a sequência de entrada e saída. Por esse motivo, esse modelo pode também ser classificado como *sequence embedding*. Em geral, esse processo ocorre na seguinte sequência:

- a) Uma ou mais camadas LSTM para implementar o *Encoder*,
- b) uma camada auxiliar entre o *Encoder* e *Decoder* para conectar suas saídas e entradas (faz-se necessário porque o *Encoder* retorna uma estrutura de dados em 2D; no entanto, o *Decoder* espera uma estrutura em 3D);
- c) uma ou mais camadas LSTM para implementar o *Decoder*,
- d) uma camada *Dense* para implementar a saída do modelo (Moura, 2017).

## 7 Testes e Resultados Obtidos

Os testes realizados comprovam a autenticidade da ferramenta *QnA Maker* dentro dos propósitos a que se designa. Possui interação amigável com o usuário, apesar de não ter sido detectada possibilidade para alterar o idioma. Seus passos para a criação da base de conhecimento são sequenciais e intuitivos e as telas possuem *layouts* de fácil aprendizado. Suas funções são parametrizáveis, o que permite maior número de possibilidades para criação de avaliações e exames.

A ferramenta *Microsoft Azure* possui várias possibilidades de configuração e parâmetros, o que permite testes e ajustes para novas verificações, além de parametrizações de perfil para permissões, recursos para assinatura (compra de licença), foto do usuário, idioma preferido, entre outros. No entanto, durante a importação dos arquivos para composição da base de conhecimento, obtiveram-se dificuldades que trouxeram transtornos para verificação, como, por exemplo:

- a) Ao importar os arquivos em formato .TXT, estrutura sugerida pela própria ferramenta (pergunta \t resposta \t filtro1: valor1 | filtro2: valor2 | filtroN: valorN), por motivo desconhecido, não foi possível obter os dados de maneira correta. Em todas as tentativas, a ferramenta não identificou a tabulação e os filtros separados por símbolo pipe (“|”);
- b) A opção de formato *Excel* (XLSX, versão 2013) também apresentou erros em alguns arquivos, mesmo seguindo o padrão e as regras orientadas. Foi identificado que o problema ocorria somente com arquivos com mais de 1,3 *megabytes* (MB);
- c) Em ambos os casos, a ferramenta reportou nenhum log de erro para facilitar a verificação do problema e, ao acionar o suporte técnico da *Microsoft*, nenhuma respostas foi obtida em tempo hábil (até o presente momento).

A partir da importação de arquivos de com 1,3 MB ou menos, os testes foram realizados inserindo perguntas relativas a notas fiscais eletrônicas, impostos, atividades processuais do dia a dia de um supermercado (alterações de preço de venda, pedidos de compra, entregas de mercadorias, vendas, emissão de documentos entre outros) e itens específicos da empresa objeto de estudo, como

atividades de ruptura, atividades de importação de produtos, entre outros, e o sistema apresentou as respostas de acordo com o histórico de anotações importado e com o *ranking* criado pelo QnA *Maker.ai*.

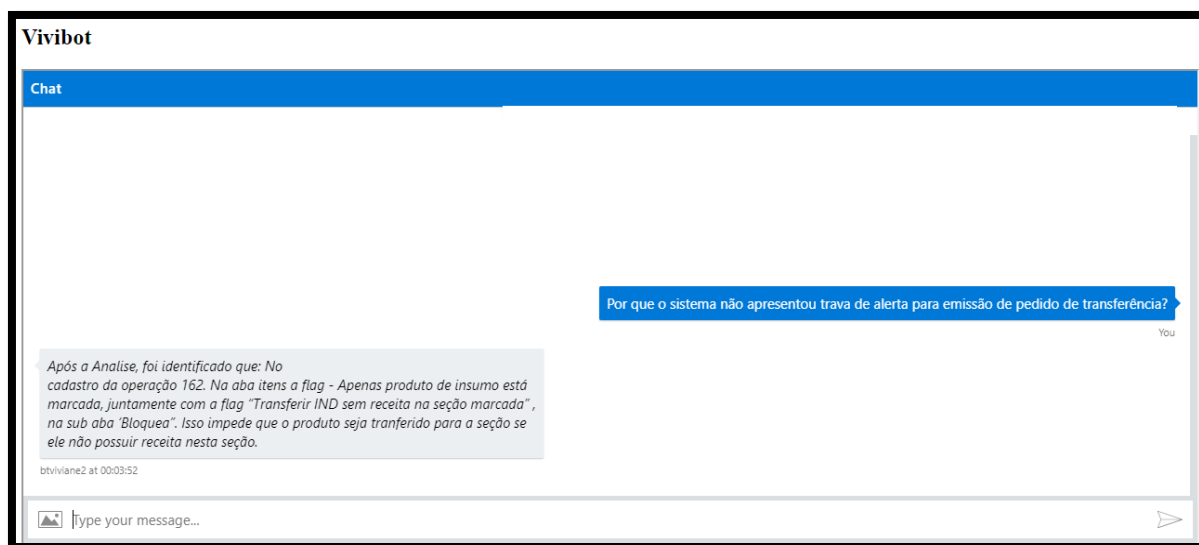
Os retornos mostrados pelo sistema foram satisfatórios, as respostas automáticas foram apresentadas com tempo de resposta muito rápido, em torno de um a dois segundos, revelando bastante coerência com a pergunta realizada. Isso pode auxiliar a equipe de suporte mediante os atendimentos do dia a dia de forma mais assertiva e rápida. No entanto, as respostas são apresentadas na íntegra de sua anotação, sendo necessário refinar da base para tornar as respostas mais amigáveis e naturais. Atualmente a resposta correta e mais próxima da solução definitiva é registrada de maneira direta sem gentileza ou qualquer cortesia. Citam-se como exemplos:

Pergunta 1:

**Usuário:** *por que o sistema não apresentou trava de alerta para emissão de pedido de transferência?*

**Sistema:** *após a análise, foi identificado que no cadastro da operação 162 - na aba itens a flag -, apenas produto de insumo está marcado, juntamente com a flag “transferir IND sem receita na seção marcada”, na subaba ‘Bloquea’. Isso impede que o produto seja transferido para a seção se ele não possuir receita nesta seção.*

Análise: tratava-se de uma configuração na operação.



**Figura 40** – Pergunta 1 – chatbot Viviâne.

Fonte: Microsoft Azure (2016). *QnA Maker Overview*. Publicado em 04 de dezembro. Microsoft Azure. [site] Recuperado de: <https://docs.microsoft.com/pt-br/azure/cognitive-services/qnamaker/home>.

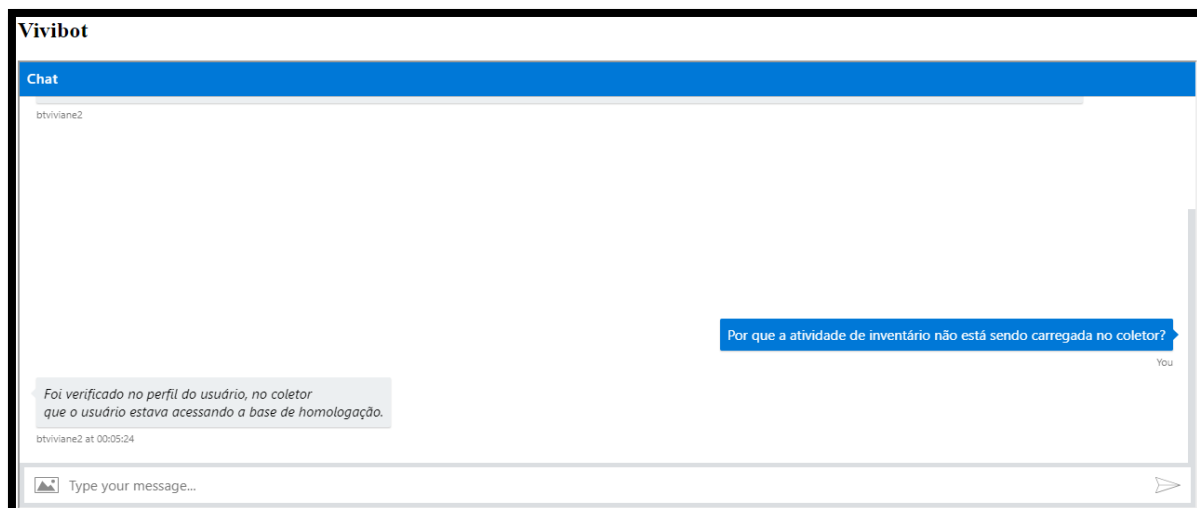
Pergunta 2:

**Usuário:** *por que a atividade de inventário não está sendo carregada no coletor?*

**Sistema:** *foi verificado no perfil do usuário, no coletor que o usuário estava acessando a base de homologação.*

Análise: o usuário estava utilizando a base testes em vez da base produção.





**Figura 41** – Pergunta 2 – Chatbot Viviâne,

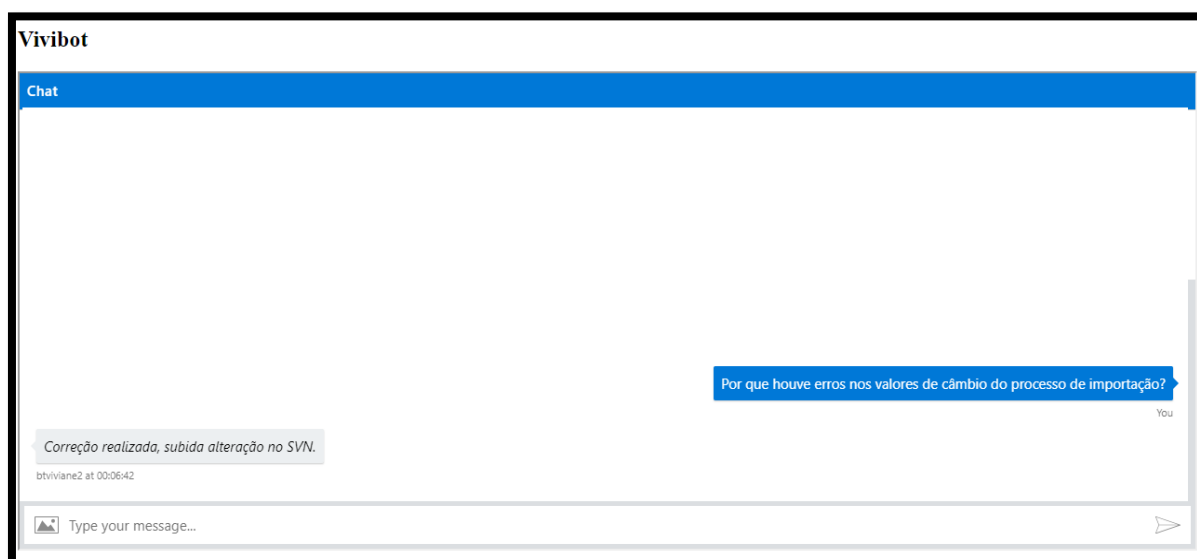
Fonte: Microsoft Azure (2016). *QnA Maker Overview*. Publicado em 04 de dezembro. Microsoft Azure. [site] Recuperado de: <https://docs.microsoft.com/pt-br/azure/cognitive-services/qnamaker/home>.

Pergunta 3:

**Usuário:** *Por que houve erros nos valores de câmbio do processo de importação?*

**Sistema:** *Correção realizada, subida alteração no SVN (Subversion).*

Análise: havia um erro de desenvolvimento e foi corrigido.



**Figura 42** – Pergunta 3 – Chatbot Viviâne.

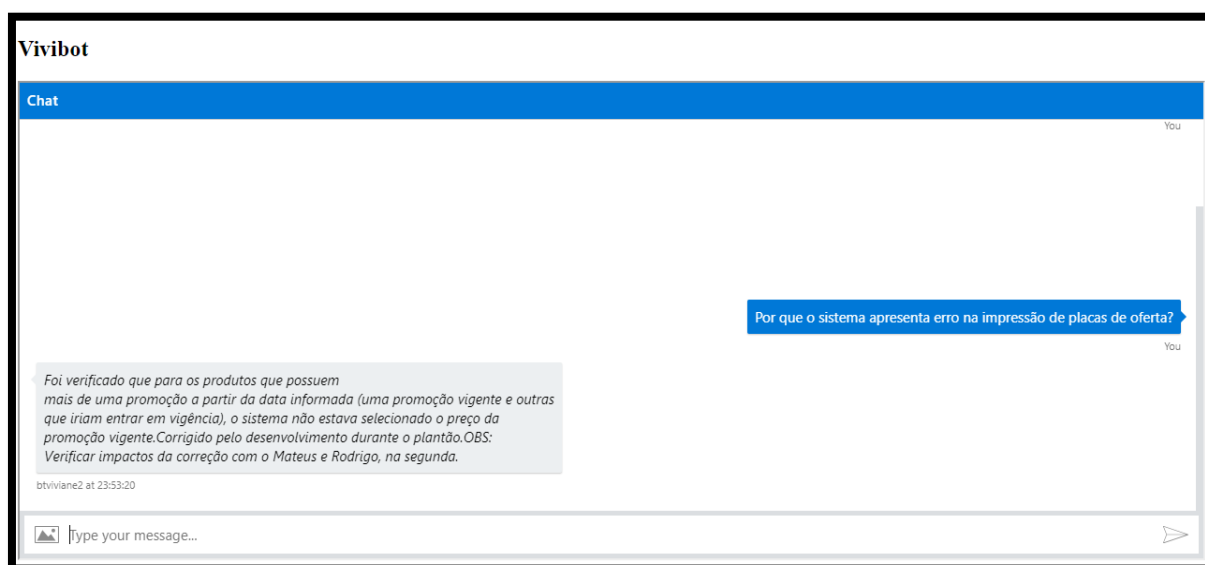
Fonte: Microsoft Azure (2016). *QnA Maker Overview*. Publicado em 04 de dezembro. Microsoft Azure. [site] Recuperado de: <https://docs.microsoft.com/pt-br/azure/cognitive-services/qnamaker/home>.

Pergunta 4:

**Usuário:** por que o sistema apresenta erro na impressão de placas de oferta?

**Sistema:** foi verificado que para os produtos que possuem mais de uma promoção a partir da data informada (uma promoção vigente e outras que iriam entrar em vigência), o sistema não estava selecionando o preço da promoção vigente. Corrigido pelo desenvolvimento durante o plantão. Obs: verificar impactos da correção com o Mateus e Rodrigo, na segunda.

Análise: o sistema não estava preparado para o cenário apresentado pelo usuário e alterado pela equipe de desenvolvimento.



**Figura 43** – Pergunta 4 – Chatbot Viviâne.

Fonte: Microsoft Azure (2016). *QnA Maker Overview*. Publicado em 04 de dezembro. Microsoft Azure. [site] Recuperado de: <https://docs.microsoft.com/pt-br/azure/cognitive-services/qnamaker/home>.

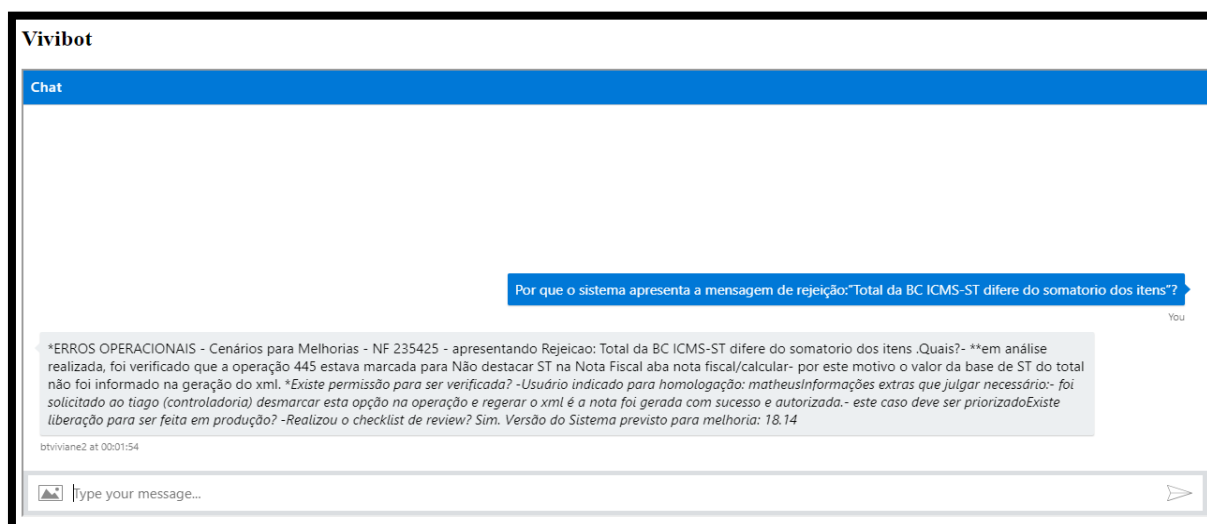
Pergunta 5:

**Usuário:** por que o sistema apresenta a mensagem de rejeição: “total da BC ICMS-ST difere do somatorio dos itens”?

**Sistema:** ERROS OPERACIONAIS - Cenários para melhorias - NF 235425 - apresentando Rejeição: total da BC ICMS-ST difere do somatorio dos itens. Quais? -

**em análise realizada, foi verificado que a operação 445 estava marcada para não destacar ST na nota fiscal aba nota fiscal/calcular - por este motivo o valor da base de ST do total não foi informado na geração do XML. Existe permissão para ser verificada? - Usuário indicado para homologação: Matheus Informações extras que julgar necessário: - foi solicitado ao Tiago (controladoria) desmarcar esta opção na operação e regerar o XML. E a nota foi gerada com sucesso e autorizada - este caso deve ser priorizado. Existe liberação para ser feita em produção? -Realizou o checklist de review? Sim. Versão do sistema previsto para melhoria: 18.14**

Análise: havia um erro operacional de parametrização, após ajuste o sistema funcionou corretamente.



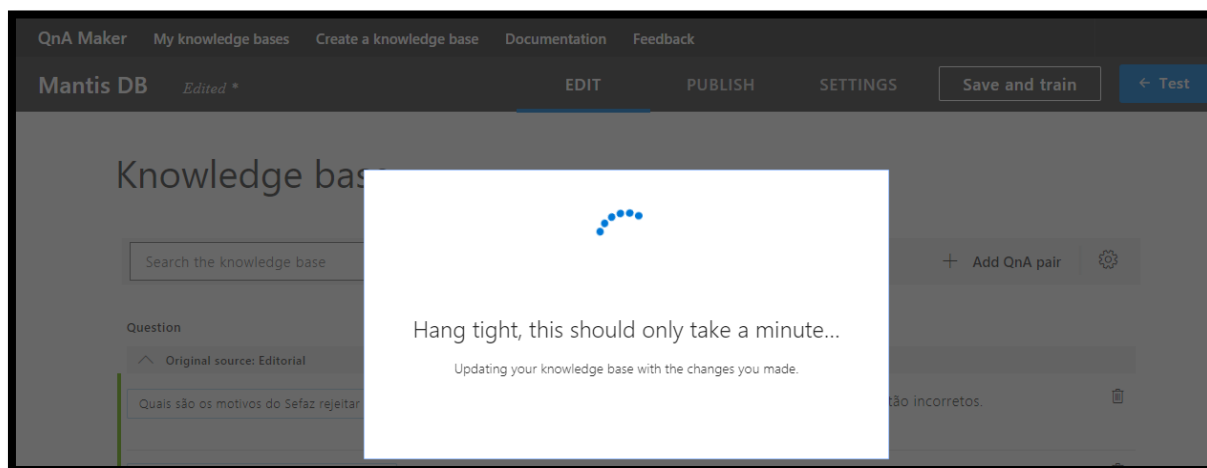
**Figura 44 – Pergunta 5 – Chatbot Viviâne.**

Fonte: Microsoft Azure (2016). *QnA Maker Overview*. Publicado em 04 de dezembro. Microsoft Azure. [site] Recuperado de: <https://docs.microsoft.com/pt-br/azure/cognitive-services/qnamaker/home>.

As respostas estão corretas e podem ser utilizadas, principalmente para os atendimentos telefônicos que requerem mais agilidade.

Notou-se lentidão para importação dos dados e para a publicação do *chatbot*. Com isso, para validação de parâmetros/configurações e desempenho da internet (trata-se de uma ferramenta *online*) para fins de testes, foi importada uma base com tamanho menor que a base completa para avaliação. E foi detectado que a lentidão manteve-se durante todas as etapas: importação dos dados, publicação, testes e inspeção. No

entanto, a lentidão não foi vista como dificultador, porque o tempo foi considerado aceitável para edições e gravação dos dados.

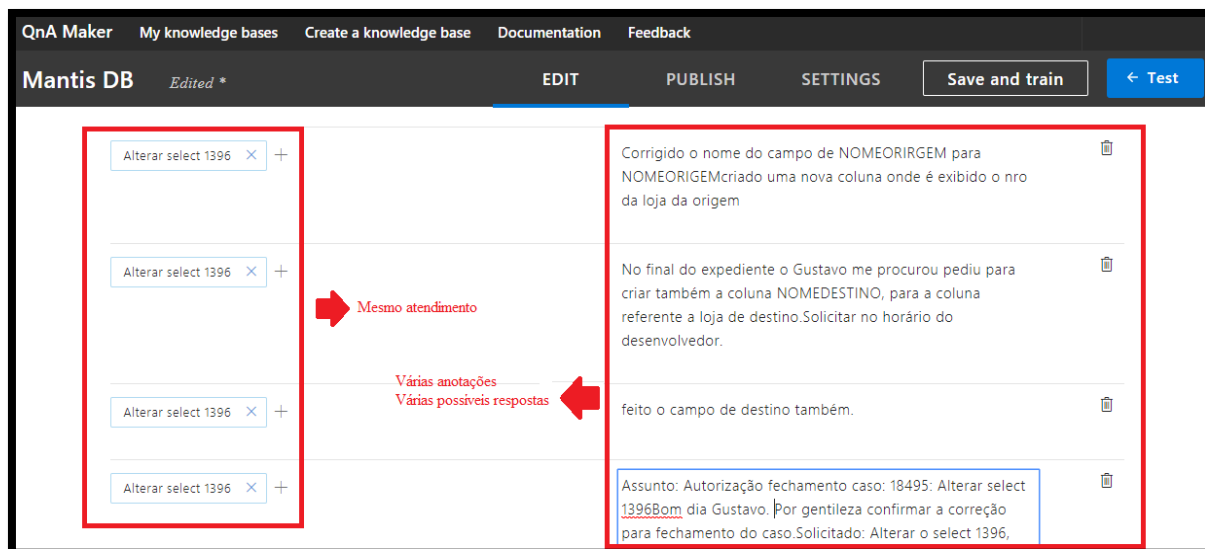


**Figura 45** – Mensagem QnA Maker para salvar alterações.

Fonte: Microsoft Azure (2016). *QnA Maker Overview*. Publicado em 04 de dezembro. Microsoft Azure. [site] Recuperado de: <https://docs.microsoft.com/pt-br/azure/cognitive-services/qnamaker/home>.

O refinamento das perguntas e das anotações é imprescindível para que o sistema apresente desempenho cada vez melhor. O Mantis® é utilizado como um sistema de interação entre as equipes de suporte e desenvolvimento. Logo, seu refinamento é algo moroso, porque um caso (atendimento registrado) pode ter uma para n anotações, ou seja, cada atendimento pode ter uma a várias anotações, não tendo número limite para as anotações. Essas anotações são irrelevantes para a solução definitiva do problema.

A Figura 46 ilustra um exemplo de um atendimento que foi registrado e que possui mais de quatro anotações. O QnA Maker repete a pergunta para cada anotação e assume que cada anotação é uma possível solução. Ao avaliar cada anotação observa-se que são registros importantes para a interação do atendimento, para documentação, mas que para a solução definitiva não são relevantes e foram desconsideradas e deletadas para não serem apresentadas para o usuário final.



**Figura 46** – Refinamento no QnA Maker – uma pergunta várias anotações.

Fonte: Microsoft Azure (2016). *QnA Maker Overview*. Publicado em 04 de dezembro. Microsoft Azure. [site] Recuperado de: <https://docs.microsoft.com/pt-br/azure/cognitive-services/qnamaker/home>.

Em todos os atendimentos foram retiradas as anotações referentes a: interação entre as equipes (suporte e desenvolvimento), solicitações de prioridade e urgência, solicitações de detalhamento e cenários de testes, informações de *selects*, alinhamentos internos e solicitações de autorização para encerrar o atendimento após solução final ser apresentada. Trata-se de informações relevantes para o andamento do atendimento e não para a sua solução.

Também foram identificados registros que poderiam ser expurgados de forma íntegra por não apresentarem importância na base de conhecimento, registros com classificação de “diário” para apontamento de horas e identificação de tarefas, registros de solicitação de alterações via banco de dados, registros para criação de relatórios avulsos para análise. Esses registros são referentes a formalidades de atendimentos e poderiam ter sido localizados de acordo com a “categoria” (campo padrão do Mantis® que permite classificar o atendimento) e serem desconsiderados no momento inicial da montagem do *corpus*.

Na Figura 47 encontra-se um exemplo de atendimento registrado apenas para acompanhamento da equipe de suporte.



**Figura 47** – Exemplo de pergunta para exclusão no QnA Maker.

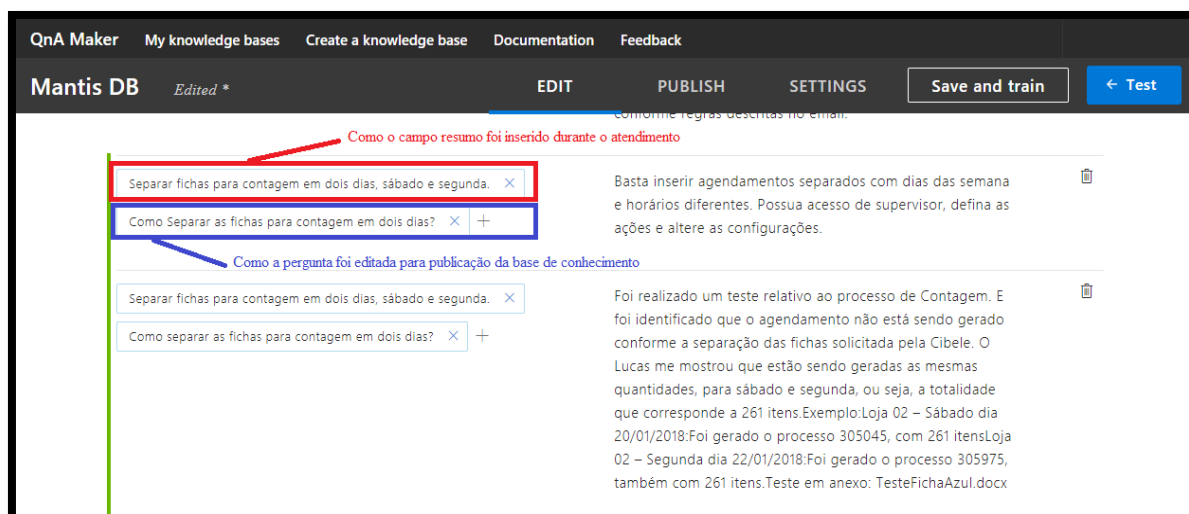
Fonte: Microsoft Azure (2016). *QnA Maker Overview*. Publicado em 04 de dezembro. Microsoft Azure. [site] Recuperado de: <https://docs.microsoft.com/pt-br/azure/cognitive-services/qnamaker/home>.

Observou-se que houve a criação recente do campo personalizado “solução” no Mantis® e que em poucos casos registrados (atendimentos) ele está preenchido. Esse campo foi desconsiderado para a criação da base, por falta de documentação. Sua utilização, porém, aumentaria consideravelmente o desempenho do *chatbot*. Caso sejam revistos os atendimentos anteriores à sua criação, podem ser submetidos novamente e criada uma nova base de conhecimento.

O QnA Maker, a partir de processamento interno, retorna a resposta mais aproximada, mas seria de suma importância para o *ranking* de análise e para apresentar a solução imediata para novos atendimentos, sem necessidade de interação com uma outra equipe ou mesmo de consultas no Mantis®, que possui restrição de filtros predeterminados.

O campo “resumo” foi utilizado como pergunta e o campo “anotações” como possíveis respostas. Foi identificado que a informação inserida no campo resumo não tem formato de pergunta e sim como uma síntese do problema. Com isso foi necessário realizar ajustes para melhor apresentação.

Veja-se o exemplo ilustrativo na Figura 48.



**Figura 48** – Edição da pergunta no QnA Maker.

Fonte: Microsoft Azure (2016). *QnA Maker Overview*. Publicado em 04 de dezembro. Microsoft Azure. [site] Recuperado de: <https://docs.microsoft.com/pt-br/azure/cognitive-services/qnamaker/home>.

Contudo, os testes foram aprovados porque as respostas realizadas no *chatbot* foram respondidas com a solução correta para o questionamento. Para essa ação, foi feita comparação com a solução indicada no atendimento do sistema Mantis®. Fazem-se necessários refinamentos para que a aplicação interaja de maneira cortês com o usuário e de forma mais natural, simulando um agente virtual com características sociáveis.

Sobre o referencial teórico, notou-se que a informação pode ser considerada um bem imprescindível para as empresas. O seu registro no sistema Mantis® e posteriormente sua apresentação no *chatbot* demonstraram o valor efetivo no esclarecimento de dúvidas inseridas durante os testes, conforme afirmado por Lopes *et al.* (2011). Parreiras e Bax (2003) relatam sobre o grande volume de informações e de dados criado nas organizações que é confirmado pelo número de atendimentos registrados e seus constante crescimento. Também se verificou, durante os testes, que realizar o cruzamento das perguntas e das respostas dentro da base de conhecimento para organizar as informações e disponibilizar no *chatbot* (retirando os dados desnecessários e duplicados) não é tarefa simples (Souza, 2006). Para Souza (2006), a recuperação das informações possui dificuldades intrínsecas ao conceito de informação (Almansa, 2017, p. 20).

Além disso, é possível inferir que os resultados estão em conformidade com os objetivos e baseados no referencial teórico proposto:



Quadro 7 – Objetivos x resultados x fundamentação teórica

Objetivos	Resultado	Fundamentação Teórica
Elaboração da coletânea de documentos ( <i>corpus</i> ) baseado no banco de dados existente.	O <i>corpus</i> foi elaborado em formato eletrônico (.TXT), baseado nos dados da Organização Verdemar Ltda. registrados no sistema Mantis®.	<p>“A <i>corpus</i> is a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research” (Sinclair, 2005 como citado em Aluísio &amp; Almeida, 2006, p.156).</p> <p>O conjunto finito de enunciados tomados como objeto de análise. Trata-se de uma coleção de documentos quer orais, quer orais e escritos, de acordo com o tipo de investigação pretendido (Galisson &amp; Coste (1983 como citados em Aluísio &amp; Almeida, 2006, p.157).</p> <p>“Um conjunto de textos escritos ou falados numa língua, disponível para análise” (Trask, 2004 como citado em Aluísio &amp; Almeida, 2006, p. 157).</p>
Geração do arquivo ( <i>corpus</i> ) em formato eletrônico para utilização do <i>chatbot</i> baseado nas etapas metodológicas de McEnery & Wilson (1996 como citado em Aluísio & Almeida, 2006).	O <i>corpus</i> foi elaborado em formato eletrônico (.TXT), compilado e manipulado de maneira automática pelo conceito de ETL, no <i>software</i> PDI, fornecido pela Hitachi. Foi feita a limpeza expurgando imagens, gráficos, tabelas e qualquer outro tipo de anexo. Para a extração e conversão foi realizado processo de transformação em um arquivo final de texto (.TXT), para importação na ferramenta QnA <i>Maker</i> com três colunas: pergunta, resposta, <i>meta-tag</i> (sendo este opcional).	<p>A moderna noção de <i>corpus</i> possui pelo menos quatro características fundamentais: a) amostragem e representatividade; b) tamanho finito; c) formato eletrônico; d) referência padrão (McEnery &amp; Wilson, 1996 como citados em Aluísio &amp; Almeida, 2006).</p> <p>Para sua a compilação três estágios principais a seguir: a) projeto do <i>corpus</i>; b) a compilação (ou captura), manipulação, nomeação dos arquivos de textos, e pedidos de permissão de uso, c) anotação (Aluísio &amp; Almeida, 2006).</p>
Elaboração das perguntas para realização de testes utilizando os atendimentos da base de conhecimento.	As perguntas foram elaboradas a partir do campo “resumo” (trata-se de uma exposição sucinta da solicitação do usuário) dos atendimentos sistêmicos registrados no Mantis®.	<p>“Os textos devem ser autênticos. Por autenticidade, compreendem-se: a) os textos devem ter sido escritos em linguagem natural, não podendo ser textos ‘produzidos com o propósito de serem alvo de pesquisa linguística’” (Berber Sardinha, 2000 como citado em Aluísio &amp; Almeida, 2006).</p>
Avaliar o retorno (respostas) do <i>chatbot</i> para as perguntas realizadas de acordo com a solução do atendimento.	As respostas foram validadas a partir do campo “anotações” (andamento da solicitação até sua resolução final) dos atendimentos sistêmicos registrados no Mantis®.	<p>“O <i>corpus</i> deve ter representatividade, isto é, ser representativo da língua ou de uma variedade de língua que se deseja pesquisar. Idealmente, um <i>corpus</i> deve ser elaborado de forma a representar determinadas características linguísticas da comunidade cuja língua está sob análise” (Sinclair, 2005 como citado em Aluísio &amp; Almeida, 2006 p. 158-159).</p>

Fonte: Elaborado pela autora.

Espera-se que os estudos sobre sistemas de perguntas e respostas sejam cada vez mais ampliados, porque assim como abordado por Almansa e Macedo (2016), no ambiente médico, Moreno *et al.*, (2015) no ambiente de comunicação instantânea, Leonhardt *et al.* (2003) e Bada e Menezes (2012) no ambiente educacional, entre muitos outros autores, a busca por informações importantes e relevantes é uma tarefa que exige tempo e experiência do profissional para localizar respostas rápidas e que podem ser auxiliadas pela tecnologia de informação a partir desses sistemas.

## 8 Considerações Finais

Conforme Amorim *et al.* (2014), de maneira geral o grande desafio dos sistemas de perguntas e respostas é retornar a resposta que mais se aproxima da expectativa do solicitante para uma pergunta feita em linguagem natural. O processo completo requer muita atenção, porque se trata de algo de alta complexidade. Principalmente porque o volume de informações tende a crescer consideravelmente.

Foi realizado estudo pela *Association for Information and Image Management (AIIM)*, descrito por Puhl e Preuss (2013), que revelou que até 80% das informações de uma empresa podem estar em dados não estruturados, espalhados pela empresa, com armazenamento em computadores pessoais ou via troca de *E-mails* ou em dispositivos portáteis, como *pen drive*, por exemplo, entre outros tipos de armazenamento.

Em média, o estudo aborda empresas que não possuem um gerenciador de conteúdo e podem vir a realizar 19 cópias de cada documento organizacional e perder um em cada 20 (1/20) documentos, além de gastar 400 horas por ano procurando e tentando recuperar documentos perdidos (Puhl & Preuss, 2013). Com isso, a recuperação da informação é uma necessidade real em todas as áreas da ciência; sua eficiência e eficácia estão sendo cobradas cada vez mais, seja ela em um ambiente público ou privado. Recuperar informações não constitui uma tendência ou de um fator de modismo, mas uma constatação.

Os sistemas de perguntas e respostas surgiram como uma possibilidade de aumento de assertividade e interação com o usuário e os *chatbots*, por sua vez, como um desmembramento. São práticos e de fácil avaliação de desempenho, bastando inserir uma pergunta que esteja em acordo com o assunto de domínio da base de conhecimento e será possível analisar a resposta, sendo que a sua implementação não é algo trivial.

Os testes realizados nos serviços da *Microsoft Azure* e *QnA Maker* possibilitaram responder ao questionamento inicial deste trabalho, apresentando um serviço de perguntas e respostas frequentes, que permite interação com o usuário a partir da

importação de uma base histórica com conteúdo organizacional relativo a atendimentos de suporte. Esses atendimentos foram classificados pela própria empresa como dúvidas, falhas, *selects*, melhorias, entre outros, e que, devido ao volume, sua consulta tornou-se uma tarefa morosa. Essas ferramentas podem contribuir para o aumento da assertividade do atendimento de suporte a sistemas de forma ágil em uma rede de supermercados. Não apenas a supermercadistas com área de tecnologia da informação interna, mas também a outras empresas de ramos de atividades diferentes, desde que exista uma base de conhecimento sólida para importação dos dados. Os serviços são apenas exemplos de utilização, porque os sistemas de perguntas e respostas, independentemente da tecnologia utilizada, comprovam sua contribuição na localização de informações de forma eficiente.

Concluiu-se que, para responder questionamentos e dúvidas de usuários, não apenas em uma rede de supermercados, mas de qualquer área de conhecimento, os *chatbots* podem ser gerados a partir de base de conteúdo existentes. Os *chatbots* podem ser utilizados frente à necessidade de atender usuários na prestação de serviço de suporte e em outros tipos de frentes, como, por exemplo, na interação para captar clientes, conversas para ações de *marketing*, pesquisas de opinião, agendamentos, consultas de ensino e especialidades médicas, entre outros.

O fator de análise para desempenho de um sistema de perguntas e respostas será sempre o grau de assertividade e confiança que o sistema proporciona. O volume de informações nas empresas, o excesso de dados ou sua redundância, assim como a dificuldade de localização (Núcleo, 2017), são confirmações que *chatbots* refinados e preparados, principalmente por domínios (assuntos específicos), são uma alternativa promissora.

Os objetivos do presente trabalho foram concluídos completando-se a lacuna proposta de identificar técnicas de sistemas de informação que auxiliem no aperfeiçoamento da recuperação de informações. Não há somente técnicas, como serviços preestabelecidos com procedimentos nativos de aprendizado de máquina que favorecem a implementação de sistemas de perguntas e respostas.

Consideram-se o experimento e o estudo de caso satisfatórios, visto que os dados e informações crescem exponencialmente, e não há paralisações para a elaboração de conteúdo, crescendo também a complexidade de sua manipulação e extração de informações relevantes para tomadas de decisão. Os sistemas de perguntas e respostas, assim como os *chatbots*, são uma possibilidade de solução.

O estudo permitiu avanços nas análises e nos testes das ferramentas que possuem alta capacidade de aproveitamento, com possibilidades reais de agilidade e diminuição de atendimentos, item muito relevante em um setor de suporte. O sistema confirma sua utilidade, sendo possível reproduções em ambientes de áreas diversas, sendo necessário um especialista no assunto para realizar avaliações.

### **8.1 Limitações da pesquisa e recomendações para trabalhos futuros**

As limitações de pesquisa encontradas foram referentes à tarifação do serviço e ao tamanho da base, que dificultaram maiores refinamentos. Faz-se necessário o aprofundamento nos demais recursos que a ferramenta disponibiliza, entretanto, podem ser feitos posteriormente e ficando assim como sugestão futura.

Como estudo futuro sugere-se uma revisão para preenchimento do campo solução do sistema Mantis® para os registros antigos e para os novos atendimentos. Logo após essa ação, sugere-se uma nova submissão da base de conhecimento com avaliações das equipes de suporte e de desenvolvimento da organização. Por se tratar de áreas com funções e objetivos distintos, apesar da parceria e interação, a comparação de percepção entre as equipes e mensuração do número de assertividade do *chatbot* permitirá a criação de estatísticas para medir o nível de acertos e a eficiência das respostas. Além disso, sugere-se aprofundar nos recursos dos serviços *Microsoft Azure* e *QnA Maker* para testes e aplicações para a publicação do *chatbot*, além de realizar testes apenas implementando a técnica Seq-to-Seq, utilizada pela ferramenta.

## Referências

- Abdul-Kader, A., S., & Woods, J. (2015). Survey on chatbot design techniques in speech conversation systems. *International Journal of Advanced Computer Science and Applications*, 6(7), 72–80. Recuperado de: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.695.5415&rep=rep1&type=pdf>. <https://doi.org/10.14569/IJACSA.2015.060712>.
- Acil (2018). *Computador* [figura]. Recuperado de: <http://www.acillimeira.com.br/noticias:qual-a-melhor-hora-para-trocar-de-computador-na-empresa>
- Aguiar, F. H. O., Sampaio, M., & Hilsdorf, W. C. (2010). Gestão de ruptura no varejo de alimentos. *Anais do XXX Encontro Nacional de Engenharia de Produção. Maturidade e desafios da Engenharia de Produção: competitividade das empresas, condições de trabalho, meio ambiente*. São Carlos, SP, Brasil, 12 a 15 de outubro. Recuperado de: [http://www.abepro.org.br/biblioteca/enegep2010\\_tn\\_sto\\_113\\_741\\_15651.pdf](http://www.abepro.org.br/biblioteca/enegep2010_tn_sto_113_741_15651.pdf).
- Albertin, Alberto Luiz; Albertin, Rosa Maria de Moura (2008). *Tecnologia de Informação e desempenho empresarial no gerenciamento de seus projetos: um estudo de caso de uma indústria* (p. 599-629). Recuperado de: [http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S1415-6552008000300002](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1415-6552008000300002)
- Almansa, L. F., & Macedo, A. A. (2016). *Sistema para perguntas e respostas em sistema de informação doenças crônicas*. (p. 2587-2596). Recuperado de: <http://www.lbd.dcc.ufmg.br/colecoes/wim/2016/>.
- Almansa, L. F. (2017). *Uma arquitetura de question-answering instanciada no domínio de doenças crônicas*. Recuperado de: <http://www.teses.usp.br/teses/disponiveis/95/95131/tde-10102016-121606/en.php>.
- Aluísio, S. M., & Almeida, G. M. D. B. (2006). O que é e como se constrói um corpus? Lições aprendidas na compilação de vários corpora para pesquisa linguística. *Calidoscópico*, 4(3), 155-177. Recuperado de: <http://revistas.unisinos.br/index.php/calidoscopio/article/view/6002>.
- Álvarez, A. C. (2007). *Extração de informação de artigos científicos: uma abordagem baseada em indução de regras de etiquetagem*. (p. 131). Recuperado de: <http://www.teses.usp.br/teses/disponiveis/55/55134/tde-21062007-144352/pt-br.php>. <https://doi.org/10.11606/D.55.2007.tde-21062007-144352>.
- Amorim, M. T. C. F., Cury, D., & Menezes, C. S. (2012). Um helpdesk Inteligente baseado em Ontologias. *Anais do 23º Simpósio Brasileiro de Informática na Educação (SBIE 2012)*. ISSN 2316-6533, Rio de Janeiro, 26-30 de novembro. Recuperado de: <http://www.br-ie.org/pub/index.php/sbie/article/viewFile/1779/1540>.

- Amorim, M. T. C. F., Cury, D., & Menezes, C. (2014). Sobre a aplicação de ontologias para orientar agentes a responder perguntas. *In: Silva, D. C. D. Silva (2011). Uma arquitetura de business intelligence para processamento analítico baseado em tecnologias semânticas e em linguagem natural*. Business, 161. Recuperado de: <http://btd.egc.ufsc.br/wp-content/uploads/2011/04/DhiogoCardoso DaSilva.pdf>.
- Amorim, M. T. C. F., Cury, D., & Menezes, C. (2014). Sobre a aplicação de ontologias para orientar agentes a responder perguntas. *Revista Brasileira de Informática Na Educação*, 22(3), 1. Recuperado de: <http://www.lume.ufrgs.br/handle/10183/127992>.  
<https://doi.org/10.5753/rbie.2014.22.03.1>.
- Andrade, J. C., Nardi, J. C., Pessoa, J. M., & Menezes, C. S. (2003). Qsabe-um ambiente inteligente para endereçamento de perguntas em uma comunidade virtual de esclarecimento. *La-Web*, 7. Recuperado de: [http://www.cwr.cl/la-web/2003/stamped/19\\_andrade\\_j.pdf](http://www.cwr.cl/la-web/2003/stamped/19_andrade_j.pdf).
- Associação Brasileira de Supermercados. ABRAS. (2017). *Síntese dos resultados 2016*. [Figura]. Recuperado de: (<http://www.abras.com.br/economia-e-pesquisa/ranking-abras/as-500-maiores/>).
- Associação Mineira de Supermercados. AMIS. (2017 nov.). Onde está o produto que deveria estar aqui? *Gôndola*, ano 23 (258). Recuperado de: <http://online.anyflip.com/glyq/ coji/mobile/index.html>.
- Bada, E. M., Menezes, C. S. (2012). Uma proposta para extração de perguntas e respostas de textos. *Anais do XVII Congresso Internacional de Informática Educativa, TISE*, 44-49. Recuperado de: <http://www.tise.cl/volumen8/TISE2012/06.pdf>.
- Badaró, A. (2016). *10 causas de problemas no seu estoque*. Infovarejo [Site]. 15 de agosto. Recuperado de: <https://www.infovarejo.com.br/10-causas-de-problemas-no-seu-estoque/>.
- Baeza-Yates e Ribeiro-Neto (1999) Ricardo Baeza-Yates e Berthier Ribeiro-Neto. *Modern information retrieval*, volume 463. ACM Press New York, 1nd ed.
- Bilotti, M. W. (2004). *Query expansion techniques for question answering. department of electrical engineering and computer science*. Massachusetts Institute of Technology, Estado Unidos. Recuperado de: <https://www.cs.cmu.edu/~mbilotti/pubs/Bilotti04.pdf>.
- Buckland, M. K (1991). Information as thing. *Journal of the American Society for Information Science* (1986-1998), 42(5), 351-360. Recuperado de: <https://pdfs.semanticscholar.org/b3d4/d7980d6a628b503003ef4e7763a93544508e.pdf>;
- Burger, J., Cardie, C., Chaudhi, V., Gaizauskas, R., Harabagiu, S., Israel, D., Jacquemin, C., Lin, C.Y., Maiorano, S., Miller, G., Moldovan, D., Ogdem, B.,

- Pragert, J., Rillof, E., Singhal, A., Shrihari, R., Strzalkowski, T., Voorhees, E., & Weischedel, R. (2003). *Issues, Tasks and program structures to roadmap research in question & answering (Q&A)*. Recuperado de: [https://www.researchgate.net/publication/228770683\\_Issues\\_tasks\\_and\\_program\\_structures\\_to\\_roadmap\\_research\\_in\\_question\\_answering\\_QA](https://www.researchgate.net/publication/228770683_Issues_tasks_and_program_structures_to_roadmap_research_in_question_answering_QA). Página 40 e 41.
- Canal Automação Comercial (2018). *Balança computadora com impressora integrada* [figura]. Recuperado de: <https://www.canalautomacao.com.br/produto/521-balanca-toledo-etiquetadora-prix-5-plus-preta-30kg-x-2g-5g-10g>
- Carvalho, A., & Galegale, N. V. (2006). Tecnologia da informação no setor supermercadista: um estudo exploratório no interior de São Paulo. *Anais do XIII SIMPEP* – Bauru, SP, Brasil, 06 a 08 de novembro. Recuperado de: [http://www.simpeptestemigracao.feb.unesp.br/anais/anais\\_13/artigos/52.pdf](http://www.simpeptestemigracao.feb.unesp.br/anais/anais_13/artigos/52.pdf).
- Cesarino, M. A. N. (1985). Sistemas de recuperação da informação. *R. Esc. Bibliotecon*, Belo Horizonte: UFMG, 14(2), 157-168, set. Recuperado de: <http://basessibi.c3sl.ufpr.br/brapci/index.php/article/download/13794>.
- Cico, C. (2017). O novo cupom fiscal traz benefícios para varejistas e consumidores. *E-CommerceBrasil* [Site]. 08 de setembro. Recuperado de: <https://www.ecommercebrasil.com.br/artigos/o-novo-cupom-fiscal-traz-beneficios-para-varejistas-e-consumidores/>
- Costa, L. F. (2005). Esfinge - resposta a perguntas usando a rede. In: J. M. Gutiérrez, F. M., Santoro, P. Isaías, Pedro (Ed.) *Proceedings da conferência IADIS Ibero-Americana WWW/Internet 2005*, Lisboa, Portugal, 18-19 de outubro de 2005, IADIS Press, pp.616-619. Recuperado de: [https://comum.rcaap.pt/bitstream/10400.26/350/2/CostaCIAWI 2005.pdf](https://comum.rcaap.pt/bitstream/10400.26/350/2/CostaCIAWI%202005.pdf).
- Cut The Knot (2018). *Código de barras* [figura]. Recuperado de: <https://www.cut-the-knot.org/Curriculum/Arithmetic/EAN13.shtml>.
- Datalogic (2018). *Scanner Magellan* [figura]. Recuperado de: <http://www.datalogic.com/eng/products/retail/fixed-retail-scanners/magellan-9300i-pd-673.html>
- Dresch, A. *Design science e design science research como artefatos metodológicos para Engenharia de Produção*. 2013. 184 f. Dissertação (Mestrado em Engenharia de Produção e Sistemas) – Universidade do Vale do Rio dos Sinos – UNISINOS, São Leopoldo, 2013. Recuperado de: <http://www.repositorio.jesuita.org.br/handle/UNISINOS/4075>.
- Ferreira, A. B. H. *Dicionário Aurélio* (2018). Recuperado de <https://dicionariodoaurelio.com/>
- Ferreira, M. S., Campos, I. C., Macedo, M. A. S; (2006). Análise da importância do uso estratégico da informação no varejo supermercadista no Brasil. *Anais do XXVI*



*ENESEP* - Fortaleza, CE, Brasil, 9 a 11 de Outubro. Recuperado de: [http://www.abepro.org.br/biblioteca/enesep2006\\_tr530352\\_7451.pdf](http://www.abepro.org.br/biblioteca/enesep2006_tr530352_7451.pdf).

Frakes e Baeza-Yates (1992) William B. Frakes e Ricardo Baeza-Yates, editors. *Information Retrieval: Data Structures and Algorithms*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA. ISBN 0-13-463837-9

Gurgel, G. P. F., & Cezar Filho, P. (2010). *Avaliação da infraestrutura de tecnologia da informação de empresas em Mossoró-RN*. 5, 45-55. Recuperado de: <http://www2.ifrn.edu.br/ojs/index.php/HOLOS/article/view/554>.

IBM - Systems and Technology (2011). *Watson – um sistema projetado para respostas*. Recuperado de: [ftp://ftp.software.ibm.com/la/documents/br/systems/power/watson\\_pt.pdf](ftp://ftp.software.ibm.com/la/documents/br/systems/power/watson_pt.pdf).

Jones, M. T. (2017). *Recurrent neural networks deep dive*. IBM [site], published on August, 17. Recuperado de: <https://www.ibm.com/developerworks/library/cc-cognitive-recurrent-neural-networks/index.html>.

Kasinathan, Vinothini; Mustapha, Aida; Siow, Sonia; Hopman, Marshella (2018). TicTad: A Chatterbot for Learning Visual C# Programming based on Expert System; *Indonesian Journal of Electrical Engineering and Computer Science*. Vol. 11, No. 2, August 2018, pp. 740-746. Recuperado de: <http://iaescore.com/journals/index.php/IJEECS/article/view/12837>

Kowalski, G.J. & Maybury, M.T. *Information storage and retrieval systems: theory and implementation*. (2. ed., 336 p). Estados Unidos: Springer, 2000.

Leonhardt, M. D., Castro, D. D., Dutra, R. L. S., & Tarouco, L. M. R. (2003). ELEKTRA: Um chatterbot para uso em ambiente educacional Michelle. *RENTE: Novas Tecnologias na Educação*, 1(2), 1-11. Recuperado de: <http://seer.ufrgs.br/index.php/rente/article/view/14336/8251>. <https://doi.org/10.5753/CBIE.SBIE.2003.81-88>.

Lopes, B., Mulyder, C. F., Judice, V. M. M. (2011). Inteligência competitiva e o caso de um arranjo produtivo local de eletrônico brasileiro; *Revista Gestão & Planejamento*, 12(2), 213-231. Recuperado de: <http://www.spell.org.br/documentos/ver/912/inteligencia-competitiva-e-o-caso-de-um-arranjo-produtivo-local-de-eletronica-brasileiro/i/pt-br>.

Machado, A. (2014). Estudo da EMC prevê que volume de dados virtuais armazenados será seis vezes maior em 2020. *O Globo* [Site]. 10 de abril. Recuperado de: <https://oglobo.globo.com/sociedade/tecnologia/estudo-da-emc-preve-que-volume-de-dados-virtuais-armazenados-sera-seis-vezes-maior-em-2020-12147682>

Machado Júnior, D., Foleiss, J. H., & Souza, V. M. A. de. (2009). SQAS: um sistema automático de question-answering para textos jornalísticos. *Anais do 7th Brazilian Symposium in Information and Human Language Technology* (Stil 2009), 1-3.

Recuperado de: [http://www.nilc.icmc.usp.br/til/stil2009\\_English/Proceedings/olincom/Machado JrFoleissdeSouza.pdf](http://www.nilc.icmc.usp.br/til/stil2009_English/Proceedings/olincom/Machado JrFoleissdeSouza.pdf).

Maia, L. C. G., & Souza, R. R. (2010). *Uso de sintagmas nominais na classificação automática de documentos eletrônico*. 15(1), 154-172. Recuperado de: <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/875/717>

Mantis Bug Tracker (2017). Recuperado de: <https://www.mantisbt.org/>.

Menezes, C. S., Tavares, O. L., & Pessoa, J. M (1998). QSabe - Trocando experiências sobre informática educativa em uma rede de educadores. *Revista Brasileira de Informática na Educação*, (2). Recuperado de: <https://pdfs.semanticscholar.org/2217/6e38647842b189173366e0fdab84c6a7d4f3.pdf>.

Microsoft Azure (2018). *API do QnA Maker*. Microsoft Azure. [site] Recuperado de: <https://azure.microsoft.com/pt-br/services/cognitive-services/qna-maker/>.

Microsoft Azure (2016). *QnA Maker Overview*. Publicado em 04 de dezembro. Microsoft Azure. [site] Recuperado de: <https://docs.microsoft.com/pt-br/azure/cognitive-services/qnamaker/home>.

Monteiro, S. D., Fernandes, R. P. M., DeCarli, G. C., & Trevisan, G. L. (2017). Sistemas de recuperação da informação e o conceito de relevância nos mecanismos de busca: semântica e significação. *Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação*, 22(50), 161. <https://doi.org/10.5007/1518-2924.2017v22n50p161>

Monteiro, S. D. O ciberespaço e os mecanismos de busca: novas máquinas semióticas. *Ciência da Informação*, 35(1), 31-38, jan./abr. 2006. Recuperado de: [http://www.scielo.br/scielo.php?pid=S0100-19652006000100004&script=sci\\_abstract&tlng=pt](http://www.scielo.br/scielo.php?pid=S0100-19652006000100004&script=sci_abstract&tlng=pt)

Moraes, S., & Machado, R. (2016). Chatterbot for Education: a study based on formal concept analysis for instructional material recommendation. *Anais do Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação - SBIE)*, 27(1), 1347. Recuperado de: <http://br-ie.org/pub/index.php/sbie/article/view/6826>. <https://doi.org/10.5753/cbie.sbie.2016.1347>.

Moreno, F., Manfio, E., Barbosa, C. R., & Brancher, J. D. (2015). *Tical: Chatbot sobre o atlas linguístico do Brasil no WhatsApp*, (Sbie). 279. Recuperado de: <http://www.br-ie.org/pub/index.php/sbie/article/view/5170/3561>. <https://doi.org/10.5753/cbie.sbie.2015.279>.

Moura, W. (2017). LSTM Encoder-Decoder com attention para prever sequências. *Hacking Analytics* [site], 29 de outubro. Recuperado de: <https://hackinganalytics.com/2017/10/29/lstm-encoder-decoder-com-attention->

para-prever-sequencias/.

Na era da informação, a eficiência exige um verdadeiro banco de dados (2012). *Nielsen* [Site], 21 de novembro. FMCG. Recuperado de: <http://www.nielsen.com/br/pt/insights/news/2012/na-era-da-informacao-a-eficiencia-exige-um-verdadeiro-banco-de-dados.html>.

Nonus (2018). *Leitor de código de barras* [figura]. Recuperado de: <http://www.nonus.com.br/leitor-codigo-barras.php>.

Núcleo Health Care (2017). *Saiba como lidar com o excesso de informação nas empresas*, 10 de maio. Recuperado de: <http://nucleohealthcare.com.br/blog/2017/05/10/saiba-como-lidar-com-o-excesso-de-informacao-nas-empresas/>.

Pentaho Data Integration (2018). *Hitachi Vantara/Pentaho* [site]. Recuperado de: <http://www.pentaho.com/product/data-integration#data-integration->.

Parreiras, F. S., & Bax, M. (2003). A gestão de conteúdos no apoio à Engenharia de software. *Anais do KMBrasil*, 2003, São Paulo. São Paulo: SBGC - Sociedade Brasileira de Gestão Conhecimento. Recuperado de: [https://www.researchgate.net/publication/267790885\\_a\\_gestao\\_de\\_conteudos\\_no\\_apoio\\_a\\_engenharia\\_de\\_software](https://www.researchgate.net/publication/267790885_a_gestao_de_conteudos_no_apoio_a_engenharia_de_software).

Prestes, K. V. (2011). *Avaliação de métodos de seleção da resposta em um sistema de perguntas e respostas*. Recuperado de: <https://www.lume.ufrgs.br/bitstream/handle/10183/36881/000819142.pdf?sequence=1>.

Puhl, R. L., & Preuss, E. Gerenciamento de conteúdo corporativo (ECM) na Gestão de TI. *Anais do EATI - Encontro Anual de Tecnologia da Informação e Semana Acadêmica de Tecnologia da Informação*. Recuperado de: <http://www.eati.info/eati/2013/assets/anais/artigo61.pdf>.

Quaresma, P., Rodrigues, I., Prolo, C. A., & Vieira, R. (2006). Um sistema de pergunta-resposta para uma base de documentos. *Letras de Hoje*, Porto Alegre, 41(2), 43-63, jun. Recuperado de: <http://revistaseletronicas.pucrs.br/ojs/index.php/fale/article/view/594/425>.

Rabelo, J. C. B., & Barros, F. A (2005). Pergunte! Uma interface em português para pergunta-resposta na Web. *Anais do V Encontro Nacional de Inteligência Artificial/SBC*, São Leopoldo. *Anais do XXV Congresso da Sociedade Brasileira de Computação – SBC*. 2005. (v. 1, p.1114-1117), 2005. Recuperado de: <http://www.lbd.dcc.ufmg.br/colecoes/enia/2005/094.pdf>.

Romano, R. R. (2011). *Os impactos do uso de tecnologia da informação e da identificação e captura automática de dados nos processos operacionais do varejo*. Tese (Doutorado em Administração) - Escola de Administração de Empresas de São Paulo da Fundação Getúlio Vargas, São Paulo. Recuperado de:

<http://bibliotecadigital.fgv.br/dspace/bitstream/handle/10438/8895/Tese%20Regiane%20Relva%20Romano%20-%20Versao%20Final%20Entregue.pdf?sequence=3&isAllowed=y>.

Saiba como lidar com excesso de informação nas empresas (2017). *Health Care [Site]*. 10 de maio. Recuperado de: <http://nucleohealthcare.com.br/blog/2017/05/10/saiba-como-lidar-com-o-excesso-de-informacao-nas-empresas/>.

Santos, A. M. M. M., & Costa, Cláudia Soares (1997). Características gerais do varejo no Brasil. *BNDES Setorial [Biblioteca Digital BNDES]*, Rio de Janeiro, (5), 55-69, mar. Recuperado de: <https://web.bndes.gov.br/bib/jspui/handle/1408/7125>.

Scarinci, R. G., & Oliveira, J. P. M. (2001). Extração de informação como base para descoberta de conhecimento em dados não estruturados. *Cadernos de Informatica*, 1(1), 16-20. Recuperado de: <http://seer.ufrgs.br/index.php/cadernosdeinformatica/article/view/v1n1p15-20/8800>.

Scarinci, R. G., & Oliveira, J. P. M. de. (2017). Sistemas de recuperação da informação e o conceito de relevância nos mecanismos de busca: semântica e significação. *Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação*, 22(50), 161. <https://doi.org/10.5007/1518-2924.2017v22n50p161>.

Silva, D. C. D. (2011). Uma arquitetura de business intelligence para processamento analítico baseado em tecnologias semânticas e em linguagem natural. *Business*, 161. Recuperado de: <http://btd.egc.ufsc.br/wp-content/uploads/2011/04/DhiogoCardosoDaSilva.pdf>.

Silva, C. L., Morais Filho R. A. M (2015). Comportamento estratégico em matéria de gestão de TI: estudo de caso no setor atacadista de alimentos. *Produto & Produção*, 16(4), 48-69. Recuperado de: <http://seer.ufrgs.br/index.php/ProdutoProducao/article/viewFile/41819/35893>.

Silva Júnior, S. C., Souza, M. A. F., & Manhães, J. V. P. (2014). Ferramentas de TI que contribuem para as operações e a competitividade no setor varejista brasileiro. *Revista Eletrônica Sistemas & Gestão*, Universidade Federal Rural do Rio de Janeiro (UFRRJ), Seropédica, RJ, 9(1). Recuperado de: <http://www.revistasg.uff.br/index.php/sg/article/view/V9N1A9>. doi: <http://dx.doi.org/10.7177/sg.2014.V9.N1.A9>.

Silva, R. E., Santos, P. L. V. A. C., & Ferneda, E. (2013). Modelos e recuperação de informação e web semântica: a questão da relevância. *Inf Inf*, Londrina, 18,(3), 27-44, set./dez. Recuperado de: [http://www.uel.br/revistas/uel/index.php/informacao/article/viewFile/12822/pdf\\_3](http://www.uel.br/revistas/uel/index.php/informacao/article/viewFile/12822/pdf_3).

Soluções Industriais (2018). *Coletor de dados* [figura]. Recuperado de: [http://www.solucoesindustriais.com.br/empresa/identificacao\\_etiquetagem\\_e\\_radio\\_frequencia/genecamp-servicos-em-automacao-comercial-ltda-/produtos/identificacao/coletor-de-dados-9](http://www.solucoesindustriais.com.br/empresa/identificacao_etiquetagem_e_radio_frequencia/genecamp-servicos-em-automacao-comercial-ltda-/produtos/identificacao/coletor-de-dados-9)

- Souza, L. S. de, & Moraes, S. M. W. (2006). *Construção automática de uma base AIML para Chatbot: um estudo baseado na extração de informações a partir de FAQs*, 5. Recuperado de: <http://www.lbd.dcc.ufmg.br/colecoes/eniac/2015/021.pdf>.
- Souza, R. R. (2006). Sistemas de recuperação de informações e mecanismos de busca na web: panorama atual e tendências. *Perspectivas em Ciência da Informação*, 11(2), 161–173. Recuperado de: [http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S1413-99362006000200002](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-99362006000200002).  
<https://doi.org/10.1590/S1413-99362006000200002>.
- Supermercado Moderno (2017). *Ranking de supermercados*. Recuperado de: <http://www.sm.com.br/ranking-de-supermercados?tipo=BuscaBrasil&busca=VERDEMAR%20SUPERM.%20E%20PADARIA>
- Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural Networks, *ArXiv*, 14, dec., p.1-9. Recuperado de: <https://arxiv.org/pdf/1409.3215.pdf>. <https://doi.org/10.1007/s10107-014-0839-0>,
- Varga, S. (2014). Uma visão geral sobre o uso de sistemas de perguntas e respostas na ciência cognitiva. *Anais do 4º. Seminário Interno de Cognição Artificial - SICA 2014*. Unicamp, Campinas, SP. Recuperado de: <http://www.dca.fee.unicamp.br/~gudwin/courses/IA889/2014/IA889-17.pdf>.
- Vergara, S. C. (2006). *Projetos e relatórios de pesquisa em Administração*. São Paulo: Atlas.
- Verdemar. (2017). *Documento recursos humanos: Manual Treinamento de Integração. Treinamento e Desenvolvimento*. Manual interno fornecido para este estudo.
- Wailthare, S., Gaikwad, T., Khadse, K., & Dubey, P. (2008). Artificial Intelligence Based Chat-Bot. *International Research Journal of Engineering and Technology ISO Certified Journal*, 9001, 1060-1062. Recuperado de: <https://irjet.net/archives/V5/i3/IRJET-V5I3242.pdf>.
- Weiss, R. J., Chorowski, J., Jaitly, N., Wu, Y., & Chen, Z. (2017). Sequence-to-sequence models can directly translate foreign speech. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH, 2017–August*, 2625–2629. <https://doi.org/10.21437/Interspeech.2017-503>. Recuperado de: <https://arxiv.org/abs/1703.08581>
- Zoom (2018). *Notebook* [figura]. Recuperado de: <https://www.zoom.com.br/notebook/notebook-dell-i15-5566-intel-core-i7-7500u-15-6-8gb-hd-1-tb-linux>.

## Anexo

Termo de autorização da empresa: Organização Verdemar Ltda.



### TERMO DE AUTORIZAÇÃO – DISSERTAÇÃO/TESTE/ARTIGO CIENTÍFICO

A Superintendência de Tecnologia de Informação da empresa Organização Verdemar LTDA, CNPJ 65.124.307/0003-01, Inscrição Estadual 0627053960217, localizada no endereço Avenida Nossa Senhora do Carmo, nº 1900, Bairro Sion, CEP 30320-000, Belo Horizonte, Minas Gerais, Telefone: (031) 2105-0101, autoriza a produção intelectual baseado nas informações de atendimentos de suporte realizados no sistema Mantis para a dissertação da aluna Viviâne de Almeida Tôrres, a título de pesquisa científica podendo ser estendida a teses e artigos científicos. Serão utilizados somente os dados disponíveis no sistema de atendimentos (Mantis) da Célula de Sistemas e a informação que os dados são referentes a Organização Verdemar LTDA.

A aluna, de código de matrícula 5A226413651, cursa o Programa de Pós-Graduação: Mestrado em Sistemas de Informação e Gestão do Conhecimento na Universidade FUMEC, com a orientação do Prof. Dr. Luiz Cláudio Gomes Maia, localizada em Av. Afonso Pena, nº 3880, Cruzeiro, CEP 30.130-009, Belo Horizonte, Minas Gerais, Telefone: (031) 3280-5000 ou 0800 0300 200. As informações utilizadas serão utilizadas para estudo de melhorias em sistemas especialistas em recuperação de informação em um banco de dados.

Observação.:

Não é necessário assinatura dos diretores da empresa porque foi previamente autorizado pela Superintendência de T.I.

Superintendente Tecnologia Informação – Marcos Vinicius Chaves