

UNIVERSIDADE FEDERAL DE MINAS GERAIS
ESCOLA DE CIÊNCIA DA INFORMAÇÃO

ALESSANDRA LEVCOVITZ

**CONSTRUÇÃO DE UMA TAXONOMIA CORPORATIVA DE DADOS BANCÁRIOS
GERADOS EM CANAIS DE ATENDIMENTO ELETRÔNICO**

Belo Horizonte

2018

ALESSANDRA LEVCOVITZ

**CONSTRUÇÃO DE TAXONOMIA CORPORATIVA DE DADOS BANCÁRIOS
GERADOS EM CANAIS DE ATENDIMENTO ELETRÔNICO**

Monografia apresentada ao programa de Especialização do Núcleo de Informação Tecnológica e Gerencial – NITEG, no curso Gestão Estratégica da Informação da Escola de Ciência da Informação, da Universidade Federal de Minas Gerais, como requisito para a obtenção do certificado de Especialista em Gestão Estratégica da Informação.

Linha de Pesquisa: Gestão e Organização do Conhecimento

Orientadora: Elisângela Cristina Aganette

Belo Horizonte

2018



Universidade Federal de Minas Gerais
Escola de Ciência da Informação
Núcleo de Informação Tecnológica e Gerencial

Trabalho de Conclusão de Curso de Especialização em Gestão Estratégica da Informação, intitulado "*Construção de uma taxonomia corporativa de dados bancários gerados em canal de atendimento eletrônico*" autoria de **Alessandra Levcovitz** aprovada pela banca examinadora constituída pelas seguintes professoras:

Profa. Dra. Elisângela Cristina Aganetti
Curso Gestão Estratégica da Informação
Escola de Ciência da Informação – UFMG
Orientadora

Profa. Dra. Benildes Coura Moreira Maculan
Escola de Ciência da Informação – UFMG

Data da aprovação: Belo Horizonte, 18 de dezembro de 2018

DEDICATÓRIA

Aos queridos Beto, Henrique e Guilherme, homens da minha vida, que, cada um do seu jeito, me incentivam sempre na busca de novos conhecimentos.

AGRADECIMENTOS

À Elisângela, minha orientadora, que me ajudou a encontrar o caminho com suas orientações, sugestões preciosas e incentivo na elaboração deste trabalho.

Aos professores do curso de Gestão Estratégica da Informação do NITEG que abriram minha mente e meus horizontes para uma nova área de estudo e uma nova forma de pensar a respeito de dados, informações e conhecimento.

Aos meus pais pelo incentivo e apoio sempre em tudo.

Aos colegas do curso, em especial ao Chico, pelas trocas constantes de ideias, experiências e conhecimentos ao longo do nosso curso.

RESUMO

Na era da informação, a gestão e o uso eficiente dos ativos informacionais de uma empresa são essenciais para mantê-la competitiva. Em uma instituição bancária, por exemplo, é produzido, diariamente, um grande volume de dados relacionados às transações realizadas por clientes em seus canais de atendimento eletrônico. Esses dados são gerados, armazenados e acessados por meio de diferentes sistemas de informação, na forma de logs de transações não estruturados. Tais sistemas interpretam e usam os referidos dados com o objetivo exclusivamente operacional. Na era atual do Big Data, os bancos perceberam que esses logs, produzidos, armazenados e acessados por sistemas de informação, são excelentes fontes de informação interna e, desse modo, podem e devem ser usados estrategicamente, indo além do objetivo primário do sistema em que são gerados. A organização adequada destes dados para uso por diversas áreas de negócio, com interesses e necessidades diferentes, se tornou relevante para extração de conhecimento e tomada de decisões. O referido cenário evidencia e justifica a necessidade de se criar um repositório devidamente organizado de dados corporativos, que atenderá os usuários gestores da instituição. O armazenamento desses dados em um repositório de dados semiestruturado apresentou-se como uma alternativa adequada, considerando-se a diversidade de transações, o volume pré-existente de dados não estruturados armazenados em sistemas de informação com foco em canais eletrônicos, a estrutura irregular das transações e a evolução constante de produtos e serviços oferecidos pelos bancos a seus clientes. Neste cenário surge a necessidade de definição de um vocabulário controlado cujos termos serão usados como rótulos que serão agregados aos dados presentes em um log de uma transação bancária. Justifica-se a construção de uma taxonomia corporativa de produtos e serviços bancários oferecidos aos clientes em canais de atendimento como alternativa para padronizar termos e conceitos para organizar e armazenar o universo de logs de transações bancárias. O presente estudo tem como objetivo avaliar a aplicabilidade de uma taxonomia corporativa para organização de dados bancários. Para isso, utilizou-se o procedimento operacional de elaboração de taxonomia corporativa proposto por Aganette e Teixeira (2017). Tal procedimento foi avaliado e aplicado em uma instituição bancária. A aplicação do referido procedimento resultou na construção de uma taxonomia corporativa de dados e em recomendações de melhoria do procedimento operacional para elaboração de taxonomia corporativa sob a ótica da estruturação de dados.

Palavras-chave: Taxonomia corporativa. Taxonomia corporativa de dados. Dados semiestruturados. Vocabulário controlado. Big Data. Indústria bancária.

ABSTRACT

In the information era, management and efficient use of company's information assets are essential to keep its competitiveness. In a banking institution, for example, a large volume of data is produced daily related to transactions performed on electronic channels by customers. These data are generated, stored and accessed by different information systems, in the form of unstructured transaction logs. Such systems work with these data for operational purpose only. In today's Big Data era, banks realized that these logs, produced, stored and accessed by information systems, are excellent sources of internal information and, thus, should be used strategically, going beyond the primary purpose for what they were generated. The right organization of these data, to be used by different business areas with different needs, becomes relevant for knowledge acquisition and decision-making. The need to create a properly organized corporate data repository, which will support company's managers and decision makers, is justified in this scenario. Store these data in a semi-structured data repository seems to be an adequate alternative, considering the diversity of transactions, the large volume of unstructured data current stored by banking channels information systems, the irregular transaction structures and the continuous evolution of products and services offered by banks to their customers. In this scenario, the need to define a controlled vocabulary, which terms will be appended as labels to the existing banking transaction log's data, arises. The construction of a corporate taxonomy for banking products and services is desirable as a solution to standardize terms and concepts to store the banking transaction logs data. The present study aims to evaluate the applicability of a corporate taxonomy to organize bank data. For this purpose, the operational procedure for construction of corporate taxonomy proposed by Aganette and Teixeira (2017) was used. This procedure was evaluated and applied in a banking institution. The procedure application resulted in the construction of corporate taxonomy for data and in some recommendations for the operational procedure improvement to create corporate taxonomy in data organization perspective and data structuring point of view.

Keywords: Corporate Taxonomy. Corporate Data Taxonomy. Semi-structured Data. Controlled Vocabulary. Big Data. Banking Industry.

LISTA DE FIGURAS

FIGURA 1 – Notação JSON – chave/valor.....	24
FIGURA 3 - Notação JSON - valor.....	25
FIGURA 2 – Notação JSON - <i>array</i>	25
FIGURA 4 - Notação JSON - String	26
FIGURA 5 - Notação JSON – Número.....	26
FIGURA 6 - Exemplo de estrutura JSON	27
FIGURA 7 - Processo de extração de obtenção de inteligência a partir de dados	28
FIGURA 8 - <i>Big data</i> – estruturação de dados	29
FIGURA 9 - Passos para construção de conhecimento em uma base de dados	29
FIGURA 10 - Crescimento estrutural da complexidade entre os tipos de vocabulário controlado	31
FIGURA 11 - Mapa Conceitual das Práticas de Construção de Taxonomias Corporativas.....	36
FIGURA 12 – XML e sua representação em árvore	55
FIGURA 13 – Estrutura JSON de um log de pagamento.....	56

LISTA DE QUADROS

QUADRO 1 - Dado, informação e conhecimento	15
QUADRO 2 - Dados estruturados x Dados semiestruturados	23
QUADRO 3 - Termos para pesquisa bibliográfica	40
QUADRO 4 - Termo principal x Termo JSON	54

LISTA DE ABREVIATURAS

ATM	<i>Automated Teller Machine</i>
BACEN	Banco Central do Brasil
CIP	Câmara Interbancária de Pagamentos
ECMA	<i>European Computer Manufacturers Association</i>
FEBRABAN	Federação Brasileira de Bancos
IETF	<i>Internet Engineering Task Force</i>
JSON	JavaScript Object Notation
KDD	<i>Knowledge Discovery in Databases</i>
NSU	Número de Sequência Único
PCR	Plataforma de Cobrança de Recebíveis
POS	<i>Point of Sale</i>
TI	Tecnologia da Informação

SUMÁRIO

1 INTRODUÇÃO	14
1.1 PROBLEMA	17
1.2 OBJETIVOS	18
1.2.1 OBJETIVO GERAL.....	18
1.2.2 OBJETIVOS ESPECÍFICOS	18
1.3 JUSTIFICATIVA	19
1.4 ESTRUTURA DO TRABALHO DE CONCLUSÃO DE CURSO.....	20
2 CONCEITOS GERAIS E REVISÃO DA LITERATURA	21
2.1 LOGS DE TRANSAÇÕES BANCÁRIAS	21
2.2 BANCO DE DADOS ESTRUTURADO VERSUS BANCO DE DADOS SEMIESTRUTURADO	22
2.3 PADRÃO JSON.....	24
2.4 BIG DATA E DATA MINING.....	27
2.5 VOCABULÁRIO CONTROLADO	30
2.6 TAXONOMIA CORPORATIVA.....	33
2.7 PROCEDIMENTO PARA ELABORAÇÃO DE TAXONOMIAS CORPORATIVAS.....	35
3 METODOLOGIA	38
3.1 CARACTERIZAÇÃO	38
3.2 CONTEXTO, COLETA DE DADOS E APLICAÇÃO.....	39
3.2.1 Procedimentos Específicos - Fase 1	39
3.2.2 Procedimentos Específicos - Fase 2	41
3.2.3 Procedimentos Específicos - Fase 3	42
4 APRESENTAÇÃO E ANÁLISE DOS RESULTADOS	43
4.1 FASE 1 - REVISÃO DE LITERATURA QUANTO ÀS PRÁTICAS E TÉCNICAS EXISTENTES NA LITERATURA PARA CONSTRUÇÃO DE TAXONOMIAS CORPORATIVAS.....	43

4.2 FASE 2 - ANÁLISE DA LITERATURA ENCONTRADA, A FIM DE SE DEFINIR UM MÉTODO PARA CONSTRUÇÃO DE UMA TAXONOMIA CORPORATIVA NO DOMÍNIO BANCÁRIO	43
4.3 FASE 3 - APLICAÇÃO DE UM PROCEDIMENTO DE CONSTRUÇÃO DE TAXONOMIA CORPORATIVA NO CONTEXTO BANCÁRIO	45
5 CONSIDERAÇÕES FINAIS	57
REFERÊNCIAS	59
APÊNDICE A – LISTA DE CANAIS	63
APÊNDICE B – PAGAMENTOS ACEITOS PELO BANCO	65
APÊNDICE C – PAGAMENTOS ACEITOS EM CADA CANAL DE ATENDIMENTO	66
APÊNDICE D – DIAGRAMA BPM DE UM PROCESSO DE PAGAMENTO DE UM CONVÊNIO COM CÓDIGO DE BARRAS NO INTERNET BANKING	67
APÊNDICE E – MAPA CONCEITUAL	68
APÊNDICE F – DIAGRAMA HIERÁRQUICO DE TERMOS	69
APÊNDICE G – EXTRATO DE LISTA DE TERMOS E DESCRIÇÃO	70
APÊNDICE H – TAXONOMIA	78
APÊNDICE I – DIAGRAMA HIERÁRQUICO DO LOG	79
ANEXO A – PROCEDIMENTO OPERACIONAL PARA CONSTRUÇÃO DE TAXONOMIA CORPORATIVA	80

1 INTRODUÇÃO

Ter a informação necessária, correta, precisa e no momento desejado é um requisito cada dia mais almejado e necessário numa organização¹. A informação é um ativo cada vez mais importante para operação, avaliação de desempenho e tomada de decisões nas organizações, de forma a manter a empresa competitiva e atrativa no mercado em que atua e, até mesmo, promover mudanças na sua estratégia de atuação, como já mencionado por McGee e Prusak (1994).

Numa economia de informação, a concorrência entre as organizações baseia-se em sua capacidade de adquirir, organizar, interpretar e utilizar a informação de forma eficaz. As organizações que liderarem essa competição serão as grandes vencedoras do futuro, enquanto as que não o fizerem serão vencidas por seus concorrentes. (MCGEE; PRUSAK, 1994, p. 3)

Para garantir o acesso eficiente à informação, não basta ter sistemas e tecnologias capazes de armazenar o grande volume de dados produzidos pelos processos de negócio da organização. É necessário que os dados armazenados sejam organizados, compreendidos, recuperados e tratados de uma forma simples, eficiente, confiável, segura, coesa e no momento exigido pelo negócio, conforme as etapas de gestão da informação apresentadas por Davenport (1998). Segundo o autor, as etapas do processo de gestão da informação consistem de: identificação da necessidade; coleta das informações; armazenamento, tratamento e organização da informação; criação de produtos e serviços para disseminar a informação; análise e uso da informação para tomada de decisão.

O acesso eficiente à informação se torna cada vez mais um diferencial competitivo para uma organização. Informações precisas e atualizadas contribuem para a tomada de decisões de forma ágil em todos os níveis da empresa, sejam eles operacionais ou decisórios. Numa organização do setor bancário, um grande volume de dados e informações, relacionados às transações realizadas pelos clientes nos canais de atendimento eletrônico, é produzido diariamente (PRIYANKA; NAGARAJ, 2017; SUN, N. *et al.*, 2014). No contexto deste trabalho, são consideradas as definições de dado e informação apresentadas por Davenport (1998) conforme Quadro 1.

¹ A organização pode ser entendida como a estrutura empresarial construída com fins específicos de mercado, tendo em vista a necessidade de consolidar objetivos de lucro e geração de novas receitas (INNOVIA, 2018).

QUADRO 1 - Dado, informação e conhecimento

Dado	Informação	Conhecimento
Simple observações sobre o estado do mundo	Dados dotados de relevância e propósito	Informação valiosa da mente humana
Facilmente estruturado	Requer unidade de análise	Inclui reflexão, síntese e contexto
Facilmente obtido por máquinas	Exige consenso em relação ao significado	De difícil estruturação
Frequentemente quantificado	Exige necessariamente a mediação humana	De difícil captura em máquinas
Facilmente transferível		Frequentemente tácito
		De difícil transferência

Fonte: Adaptado de Davenport (1998).

Cada transação realizada pelo cliente no banco gera um log de transação contendo dados que permitem identificar e comprovar a realização daquela transação específica, pelo cliente, ao longo do tempo.

O conjunto de dados gerados por máquina, produzidos para registrar detalhes de cada transação operacional e função automatizada realizada nos sistemas de negócios ou de informação do banco, é o que é denominado log de transação (tradução nossa) (TURNER; SCHROECK; SHOCKLEY, 2013). As transações bancárias, tais como saques, consultas de extratos, pagamentos, transferências e empréstimos, são realizadas pelos clientes em canais de atendimento eletrônico. Canais de atendimento eletrônico são o meio pelo qual o cliente realiza os serviços bancários e tem acesso às informações de sua conta e produtos contratados. São canais de atendimento eletrônico: *internet banking*, aplicativos para dispositivos móveis, caixas de agência, *Automated Teller Machines (ATMs)* do próprio banco – também conhecido como caixa eletrônico, central de atendimento telefônica, ATMs de parceiros como o Banco 24 Horas, *point of sale (POS)* de redes parceiras (máquinas de pagamento com cartão) e sistema de atendimento personalizado a um cliente, disponível para gerentes em uma agência.

Desde a implantação dos canais eletrônicos bancários para atendimento a clientes, na década de 1980, os logs de transações bancárias são gerados e armazenados pelos sistemas informatizados, que apoiam os canais eletrônicos, com o objetivo exclusivamente operacional. Na era atual do *Big Data*, os bancos perceberam que os dados internos, produzidos e armazenados por esses sistemas de informação, são uma excelente fonte de informação interna (CHOO, 1994) e podem e devem ser usados muito além do

objetivo primário do sistema em que são gerados (TURNER; SCHROECK; SHOCKLEY, 2013).

Observou-se que, com a utilização dos logs, a análise de informações pode ser favorecida, uma vez que, por meio deles, é possível extrair informações relevantes para tomada de decisões em diversas áreas de uma instituição bancária. A partir da avaliação das informações contidas nos logs de transação, é possível analisar o comportamento e o perfil de um cliente, a movimentação das agências, a disponibilidade de um canal, os produtos e serviços necessários para um determinado cliente no momento em que ele utiliza um canal de atendimento bancário, o risco de autorização de uma determinada transação, dentre outras informações.

Para se manterem competitivos, os bancos passaram a usar os seus ativos informacionais internos, muito além do uso estritamente operacional e previamente definido. Percebeu-se que a análise dessas informações de forma dinâmica permite maior conhecimento do uso dos canais, do comportamento e necessidade dos clientes, dos relacionamentos e produtos, e orienta, ainda, o direcionamento das ações a executar, de maneira mais rápida e eficiente.

Apesar da possibilidade de usar os logs de transação como fontes de informação interna para extrair informações importantes para o negócio, a sua utilização hoje vem se mostrando um desafio para as áreas de negócio e de tecnologia da informação (TI) de um banco brasileiro. Os logs de transações são armazenados de forma não estruturada e são acessíveis e interpretados apenas pelo sistema gerador deles. Todas as consultas desejadas por um usuário neste conjunto de logs só podem ser feitas por meio deste sistema, exigindo o desenvolvimento pela equipe de TI de novas funcionalidades e de programas específicos para cada nova consulta solicitada pelo usuário. Isto é lento e não atende ao requisito do usuário de ter a informação que ele deseja, na hora que o negócio exige e poder minerar e trabalhar com os dados da forma mais aderente às suas necessidades.

O cenário acima evidencia a necessidade de se definir e estruturar um repositório de logs de transações, em que o conteúdo de cada log seja facilmente identificado e compreendido de maneira única por toda a organização. Assim, pessoas de diferentes áreas de negócio da empresa, desde que autorizadas, podem ter acesso a este repositório e extrair, minerar, consolidar as informações presentes nos logs, de forma simples e intuitiva, no momento exigido pelo negócio, e de acordo com as suas necessidades específicas.

No entanto, diante da heterogeneidade dos tipos de transações e das informações presentes nos logs de transações distintas (volume de milhares de tipos de

transações diferentes e que cresce a cada dia), a ideia inicial de armazenar os logs de forma totalmente estruturada em um banco de dados relacional não se mostrou viável e tampouco adequada.

Após exaustiva pesquisa sobre o assunto, percebeu-se que uma boa solução para armazenar os logs num repositório seria num formato semiestruturado conforme apresentado por Buneman (1997) e Hammer, McHugh e Garcia-Moline (1997). Assim, os dados presentes em um log poderiam ser representados e armazenados, de forma semiestruturada, como uma coleção de metadados (dados sobre dados, segundo definição de NISO (2005)), no formato JavaScript Object Notation - JSON (JSON, 2017; ECMA, 2017). O JSON possibilita organizar a informação em uma estrutura hierárquica de rótulos (que no JSON são tratados como chaves) e valores. O JSON organiza a informação como uma coleção de pares chave/valor. A *chave* (rótulo) classifica (identifica do que se trata, define o contexto) o dado que vem a seguir e o valor é o conteúdo do dado. Esse formato é legível desde que as chaves representem termos de um vocabulário da área de conhecimento do leitor. No entanto, definir as palavras-chave não se trata de uma tarefa fácil e aleatória. Faz-se necessária a criação de um vocabulário de termos controlados, compreensível por todas as pessoas da organização que terão acesso ao repositório.

Neste contexto, a construção de uma taxonomia corporativa, que atuará como instrumento de representação e organização de dados e informações, surge como uma solução para o cenário apresentado. Essa taxonomia corporativa de dados será proposta a partir das recomendações metodológicas de Aganette e Teixeira (2017) considerando-se que as autoras fizeram uma avaliação de vários métodos e práticas de construção de taxonomias presentes na literatura e descreveram um procedimento detalhado com as várias etapas e atividades que mapearam a partir dos métodos analisados.

1.1 Problema

No contexto atual de um banco brasileiro de médio porte, temos a seguinte situação:

- Cada transação realizada por um cliente em um canal de atendimento eletrônico gera um log de transação.
- O log de transação é armazenado em forma não estruturada, em arquivos texto. Cada registro contém um conjunto de caracteres alfanuméricos, interpretados e acessíveis apenas pelo sistema de informação que armazena os logs.
- Milhões de logs são gerados por dia.

- Cada transação realizada em um determinado canal gera um conjunto de informações específicas e relevantes para o tipo de transação e o canal onde foi realizada.
- Existem mais de 2.000 tipos de transações diferentes geradas pelos canais, cada qual com o seu conjunto específico de informações.
- Os logs são armazenados, de modo não estruturado, por um período mínimo de 20 anos para atendimento de necessidades de auditoria e processos jurídicos. Constituem uma excelente fonte de informação interna e podem ser usadas com diferentes focos, tais como: (i) avaliação de comportamento de clientes, (ii) segurança de transações, (iii) desempenho de canais e produtos, por áreas de negócio distintas do banco, com o objetivo de melhorar a competitividade e a eficiência do banco. No entanto, a forma como os logs de transações são armazenados atualmente impedem o seu uso rápido e eficiente pelos usuários de negócio. Apenas o sistema gerador é capaz de interpretar e utilizar dados dos logs armazenados.

1.2 Objetivos

A partir dos problemas apresentados, esta pesquisa se propõe a alcançar os objetivos descritos nos itens 1.2.1 e 1.2.2.

1.2.1 Objetivo geral

Propor uma taxonomia corporativa aplicada a produtos e serviços bancários de pagamentos e transferências, oferecidos em canais de atendimento eletrônico de um banco de médio porte de Minas Gerais, para organização de dados.

1.2.2 Objetivos específicos

- Analisar se as etapas propostas no procedimento de construção de taxonomia corporativa de (AGANETTE; TEIXEIRA, 2017) são adequadas e suficientes para a construção da taxonomia de dados de produtos e serviços bancários de pagamentos e transferências.
- Verificar se a taxonomia de produtos e serviços bancários de pagamentos e transferências elaborada é adequada à definição de rótulos (*tags*) a serem agregados aos dados dos logs de transação para armazenamento destes logs num repositório de dados.

1.3 Justificativa

Na era do *Big data* e mineração de dados, faz-se necessária a criação de um repositório de dados para armazenamento dos logs de transações bancárias para acesso, consultas e recuperação de informações de forma eficiente e precisa, por usuários de negócio com necessidades e interesses distintos, no momento em que o usuário precisa da informação.

Armazenar os logs em uma base de dados totalmente estruturada não mostrou ser uma solução adequada devido à diversidade de transações e informações presentes em cada log e a velocidade com que novos tipos de transação com informações diferentes são criadas. A criação de um repositório de dados semiestruturado, usando o padrão JSON é uma alternativa interessante para solucionar o problema.

Contudo, a recuperação de informações deste repositório por usuários de negócio só será eficiente se as chaves forem definidas usando termos de um vocabulário controlado de fácil compreensão e entendimento único por todos os usuários, mesmo que de áreas distintas. Isto evitará que usuários com perfis distintos, visões diferentes do processo de negócio, usem termos diferentes e não documentados para representar o mesmo objeto de negócio.

A proposta de criação da taxonomia corporativa de dados tem por objetivo a definição de um vocabulário controlado a ser usado para a definição das chaves que serão usadas para armazenamento de logs de transações de forma semiestruturada por meio de uma coleção de metadados no formato JSON. Para um banco brasileiro de médio porte, que precisa se manter competitivo no mercado em que atua, é muito importante ter informações precisas e rápidas que ajudam a identificar o comportamento de seus clientes, produtos e serviços. Os logs de transações bancárias de clientes nos canais de atendimento se mostraram nos últimos anos como uma fonte de informação interna muito rica para avaliação desse comportamento. A estruturação destes logs e o fácil acesso por usuários de áreas de negócio diferentes para extração e consolidação de informações para tomada de decisões se mostrou um desafio e uma necessidade.

A taxonomia corporativa tornará a terminologia compreensível e compartilhada por toda a organização e se apresenta como um requisito fundamental para o armazenamento dos logs em um repositório de dados semiestruturado, usando o padrão JSON (tag/valor). A utilização da taxonomia corporativa de dados como modo de definição de um vocabulário controlado e unificado, para uso como *tags* na estrutura JSON, mostrou-se uma alternativa possível, tendo em vista que as taxonomias permitem a organização e

classificação de termos e conceitos facilitando os processos de organização da informação das instituições (AGANETTE; ALVARENGA; SOUZA, 2010).

A definição de taxonomia corporativa para representação de produtos e serviços bancários oferecidos a clientes, com o objetivo de organizar dados em repositórios, é uma área de pesquisa pouco explorada. A aplicação prática do processo proposto por Aganette e Teixeira (2017) no segmento bancário, com a finalidade do desenvolvimento de uma taxonomia corporativa, mostra-se como uma possibilidade de se construir o referido instrumento de modo padronizado e, conseqüentemente, replicável, também para organização e estruturação de dados.

1.4 Estrutura do Trabalho de Conclusão de Curso

Esse trabalho está estruturado da seguinte maneira: Seção 1 introduz o tema, apresenta o problema e a justificativa da solução proposta; seção 2 apresenta o referencial teórico, relativo às áreas de conhecimento: setor bancário, Ciência da Computação e Ciência da Informação, usado para o desenvolvimento do trabalho; a seção 3 apresenta a caracterização e definição da metodologia utilizada na pesquisa; a seção 4 descreve a aplicação prática do procedimento operacional para criação de uma taxonomia corporativa no setor bancário e uma breve avaliação dos resultados alcançados; por fim, a seção 5 apresenta a conclusão do trabalho com uma avaliação dos principais resultados alcançados, contribuições verificadas, limitações encontradas e faz recomendações para trabalhos futuros .

2 CONCEITOS GERAIS E REVISÃO DA LITERATURA

A construção de uma taxonomia corporativa de dados, aplicada a produtos e serviços bancários oferecidos em canais de atendimento eletrônico, para uso em um repositório de dados de logs bancários semiestruturado, será realizada a partir do entendimento dos seguintes conceitos: Logs de transações bancárias, banco de dados semiestruturados, padrão JSON para intercâmbio de dados entre sistemas, *big data* e *data mining*, vocabulário controlado e taxonomia corporativa. Será estudado também o procedimento operacional de construção de taxonomia corporativa, proposto por Aganette e Teixeira (2017), e será avaliada a viabilidade de sua aplicação no setor bancário para organização de dados.

2.1 Logs de transações bancárias

Logs de transações bancárias são dados gerados por máquina, produzidos por sistemas informatizados de um banco, para registrar os detalhes de cada transação automatizada realizada dentro do processo de negócio disponível (TURNER; SCHROECK; SHOCKLEY, 2013). Logs são registros de atividades gerados por programas e sistemas de informação que podem ficar armazenados em arquivos, na memória do computador ou em bases de dados.

Pode-se dizer, ainda, que um log de transação bancária é um registro de realização de uma transação que é inserido no momento da execução da transação. O log não pode ser alterado em nenhum momento após a sua inclusão. Um log de transação bancária representa uma evidência que uma determinada transação foi realmente realizada no banco. O log contém dados suficientes para identificar a data e hora de realização da transação, o identificador único da transação (NSU) para o banco, o canal eletrônico em que a transação foi concluída, os dados de negócio específicos da transação, dentre outras informações. O log é um meio usado pelos bancos brasileiros para atender os artigos apresentados a seguir da resolução CMN/BACEN Nº 4.283, de 4 de novembro de 2013:

Art. 1º As instituições financeiras e demais instituições autorizadas a funcionar pelo Banco Central do Brasil, na contratação de operações e na prestação de serviços, devem assegurar:

II - a integridade, a confiabilidade, a segurança e o sigilo das transações realizadas, bem como a legitimidade das operações contratadas e dos serviços prestados;

IV - o fornecimento tempestivo ao cliente ou usuário de contratos, recibos, extratos, comprovantes e outros documentos relativos a operações e a serviços;

IX - a identificação dos usuários finais beneficiários de pagamento ou transferência em demonstrativos e faturas do pagador, inclusive nas situações em que o serviço de pagamento envolver instituições participantes

de diferentes arranjos de pagamento (BANCO CENTRAL DO BRASIL, 2013, *online*).

Os logs são usados para atestar a realização de transações para: órgãos reguladores de instituições financeiras brasileiras, clientes, processos judiciais e parceiros de negócio de banco. Os logs também são usados para demonstrar operações bancárias para auditorias internas e externas e comprovar a autenticidade de um comprovante de transação apresentado por qualquer entidade.

2.2 Banco de dados estruturado versus banco de dados semiestruturado

Banco de dados estruturado é aquele que possui uma estrutura de representação ou um esquema rígido, previamente definido. Um banco de dados relacional representa bem a definição de um banco de dados estruturado, pois é composto de um conjunto de tabelas que correspondem às entidades e cada tabela possui um conjunto de colunas que representam os atributos dessa entidade. Todas as entidades e atributos são previamente mapeados e com uma estrutura fixa. A estrutura do banco de dados é totalmente dissociada dos dados. Todos os registros de uma tabela devem se encaixar no esquema previamente definido para tal tabela. Num banco de dados relacional estruturado o esquema deve ser criado antes que os dados sejam armazenados.

Dados semiestruturados são normalmente modelados como uma coleção de objetos, sendo que cada objeto pode ter qualquer quantidade de atributos cujos valores são outros objetos ou dados atômicos. Modelos propostos para dados semiestruturados consistem de grafos com rótulos em que nodos do grafo correspondem a objetos ou valores e as arestas correspondem aos atributos (ABITEBOUL; BUNEMAN; SUCIU, 2000; BUNEMAN, 1997; HAMMER, MCHUGH; GARCIA-MOLINE, 1997).

Em modelo de dados semiestruturados, a informação que é normalmente associada a um esquema, está contida e é armazenada junto ao dado que é algumas vezes chamado de dado autodescritivo (BUNEMAN, 1997). Dados semiestruturados não possuem um esquema rígido e previamente definido, mas também não contêm apenas dados em estado natural (*raw data*). Permitem a definição de esquemas flexíveis que podem ser ajustados e adaptados na medida em que novos dados surgem e relações entre registros são estabelecidas. Usualmente, os dados são autodescritivos e misturados à informação do esquema.

Algumas características de dados semiestruturados, segundo Mello (2000) e Abiteboul, Buneman e Suciú (2000), são:

- Definição à posteriori: esquemas para dados semiestruturados são usualmente definidos após a existência dos dados, com base em uma investigação de suas

estruturas particulares e da análise de similaridades e diferenças. No entanto, isto não significa que sempre existe um esquema associado a um dado semiestruturado;

- Estrutura irregular: estruturas de dados semanticamente similares podem ser organizados de maneiras diferentes, podendo ter um conjunto diferente de atributos de uma estrutura para outra. A definição do tipo e tamanho de atributos não é rígida e pode variar de um registro para outro;
- Estrutura implícita: muitas vezes existe uma estrutura básica para os dados, porém, essa estrutura está implícita na forma como os dados são apresentados;
- Estrutura parcial: apenas parte dos dados disponíveis pode ter alguma estrutura, seja implícita ou explícita.

As principais diferenças entre dados estruturados e dados semiestruturados estão descritas no QUADRO 2.

QUADRO 2 - Dados estruturados x Dados semiestruturados

Dados estruturados	Dados semiestruturados
Esquema pré-definido	Nem sempre há um esquema pré-definido
Estrutura regular	Estrutura irregular
Estrutura independente dos dados	Estrutura embutida no dado (dado sobre dado)
Estrutura reduzida	Estrutura extensa
Estrutura fracamente evolutiva	Estrutura fortemente evolutiva
Estrutura prescritiva	Estrutura descritiva
Distinção entre estrutura e dado é clara	Definição entre estrutura e dado não é clara

Fonte: (MELLO *et al.*, 2000).

O campo de aplicação de dados semiestruturados vem crescendo rapidamente. A característica de ser independente de esquema facilita o armazenamento de novos registros com novos tipos de dados, que podem ser definidos de forma incremental, sem trazer impactos para os dados já armazenados. Repositório de dados com característica de estruturas evolucionárias, ou seja, em que a estrutura de dados se modifica frequentemente, assim como os dados armazenados, são bem representados por dados semiestruturados.

Diante das características de dados semiestruturados apresentadas acima e das características, descritas anteriormente, dos logs de transações bancárias, conclui-se que a utilização de dados semiestruturados para armazenar os logs é uma solução bastante interessante.

2.3 Padrão JSON

JavaScript Object Notation (JSON) é um formato leve para intercâmbio de dados. (JSON, 2017; ECMA, 2017).

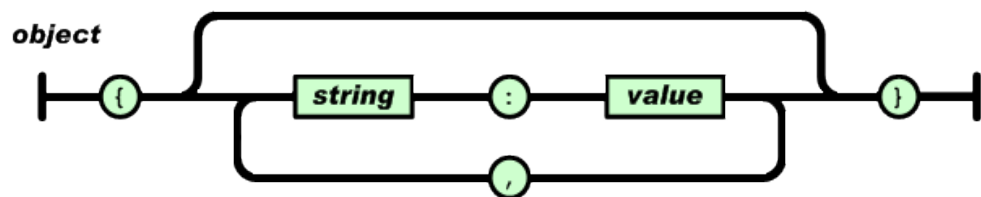
JSON foi apresentado pela primeira vez por meio do site JSON.org em 2001. A definição da sintaxe JSON foi publicada em julho de 2006 por *Internet Engineering Task Force* (IETF) RFC 4627 (CROCKFORD, 2006). ECMA-262, na quinta edição, de 2009, incluiu uma especificação normativa da especificação da gramática JSON (ECMA, 2009). A especificação ECMA-404 substituiu as especificações anteriores da sintaxe JSON. Concorrentemente, a IETF publicou a RFC 7158/7159 e, em 2017, a RFC 8259 como atualização da RFC 4627.

O formato JSON consiste de duas estruturas: i) uma coleção de pares chave/valor (*tag/value*); ii) uma lista ordenada de valores que em linguagens de programação correspondem a vetores (*arrays*) (JSON, 2017, online, tradução nossa)

Características da notação JSON:

- i Objeto: sequência sem ordem pré-estabelecida de pares (chave(*string*)/valor (*value*)). Um objeto inicia como um "{" (abre chave) e termina com um "}" (fecha chave). Os nomes são seguidos de ":" (dois pontos) e os pares chave/valor são separados por "," (vírgula) como na FIGURA 1.

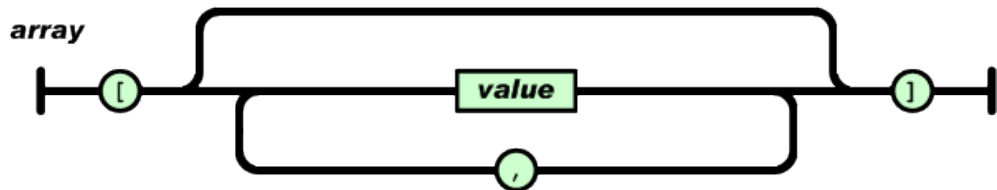
FIGURA 1 – Notação JSON – chave/valor



Fonte: ECMA (2017).

- ii Vetor (*array*): coleção ordenada de valores. Um vetor inicia com “[” (abre colchete) e termina com um “]” (fecha colchete). Os valores são separados por “,” (vírgula) como na FIGURA 2.

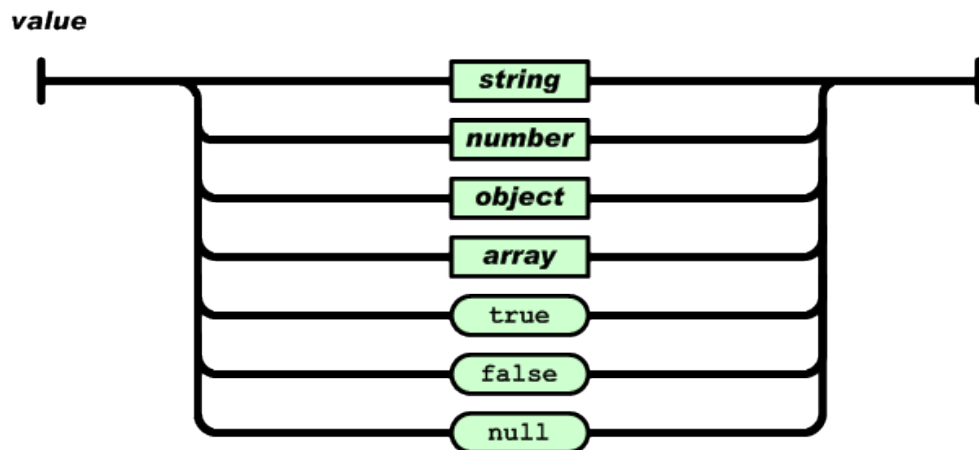
FIGURA 2 – Notação JSON - *array*



Fonte: ECMA (2017).

- iii Um valor pode ser um *string* entre aspas duplas (“”) ou um número ou um objeto ou um vetor (*array*) Estas estruturas podem ser encadeadas como na FIGURA 3.

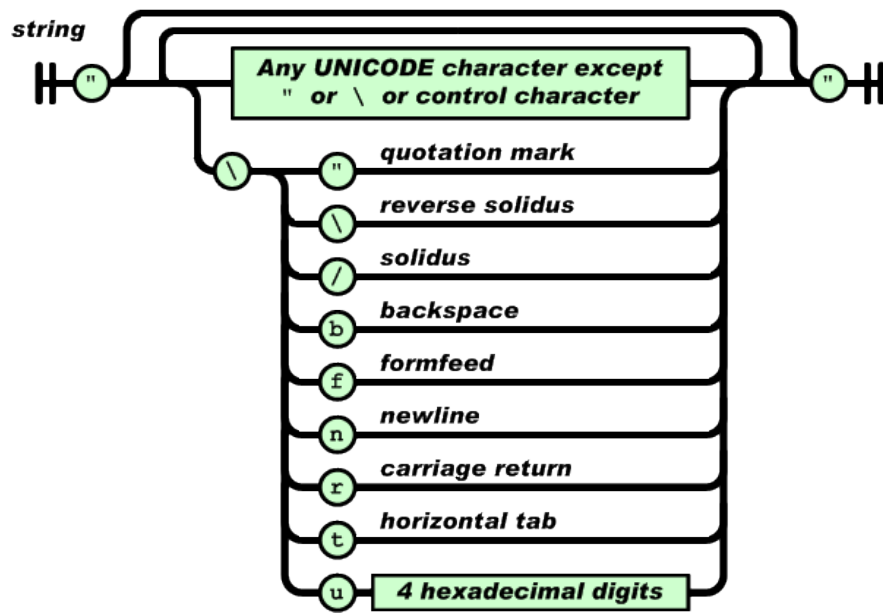
FIGURA 3 - Notação JSON - valor



Fonte: ECMA (2017).

- iv Um *string* é uma sequência de zero ou mais caracteres entre aspas duplas (“”) como na FIGURA 4.

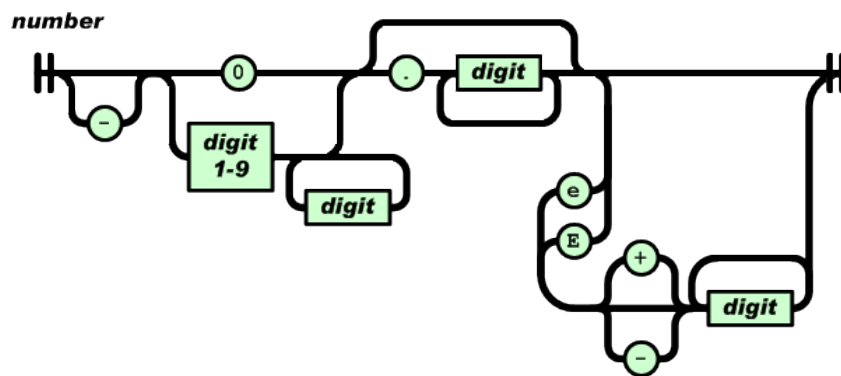
FIGURA 4 - Notação JSON - String



Fonte: ECMA (2017).

- v Um número é uma coleção de dígitos conforme notação apresentada na FIGURA 5.

FIGURA 5 - Notação JSON – Número



Fonte: ECMA (2017).

Na FIGURA 6, a seguir, pode-se ver um exemplo prático de uma estrutura JSON.

FIGURA 6 - Exemplo de estrutura JSON

```
{
  "nome": "Alessandra",
  "sobrenome": "Levcovitz",
  "idade": 51,
  "endereco": {
    "logradouro": "Rua XXX",
    "numero": 200,
    "cidade": "BeloHorizonte",
    "estado": "MG",
    "cep": 30000
  },
  "telefone": [
    {
      "tipo": "fixo",
      "numero": "31 3225-1234"
    },
    {
      "tipo": "celular",
      "numero": "31 9955-4567"
    }
  ]
}
```

Fonte: Elaborado pela autora.

A ideia utilizada pela notação JSON para representar informações é simples: para cada valor representado atribui-se um nome (ou rótulo ou *tag*) que descreve o seu significado. JSON pode ser considerado um esquema de metadados, assim como XML. Segundo a norma ANSI NISO z39 (NISO, 2005), esquemas de metadados estendem o conjunto de elementos de metadados adicionando regras de sintaxe – forma e estrutura – e em alguns casos regras codificadas. O uso de esquemas de metadados facilita a interoperabilidade entre sistemas. JSON é uma notação adequada para representar dados semiestruturados. É uma alternativa bastante interessante para armazenar os logs de transações bancárias, pois é leve (bem mais leve e mais curto que XML), facilmente legível e autodescritivo.

2.4 Big data e Data mining

Há três dimensões, comumente aceitas na comunidade, para definir e caracterizar *Big data*, conhecidas como três V's. São elas: volume, variedade e velocidade.

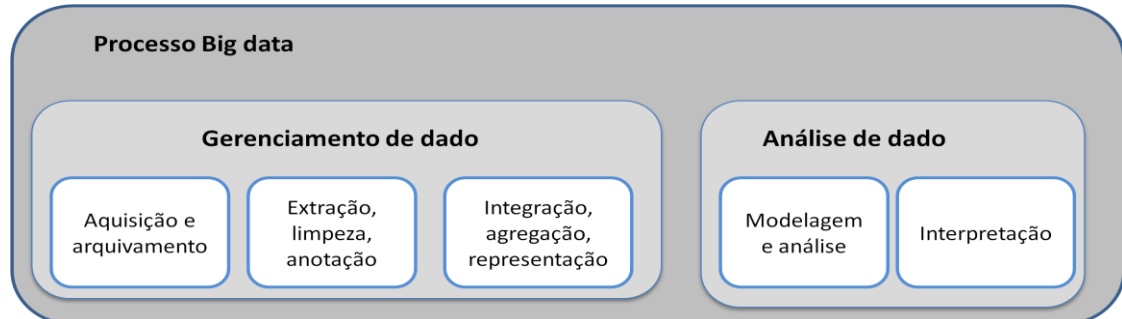
O volume refere-se à quantidade de dados produzidos medido em bytes. Variedade consiste na heterogeneidade estrutural de um conjunto de dados, ou seja, os dados podem ser encontrados, armazenados e usados nos seguintes modos: estruturado, semiestruturado, não estruturado. A velocidade refere-se à taxa na qual os dados são gerados *versus* a velocidade com que devem ser analisados. (GANDOMI; HAIDER, 2015).

Segundo a definição do *IT Glossary*:

Big data é um ativo de informação de alto volume, alta velocidade e grande variedade que exige formas inovadoras e econômicas de processamento da informação para melhorar a avaliação e a tomada de decisão. (GARTNER, 2018, *online*, tradução nossa)

O processo de extrair significado e informação do *big data* pode ser dividido em dois grandes processos: gerenciamento de dados e análise de dados. O gerenciamento de dados consiste nos processos e tecnologias que suportam a aquisição e armazenamento de dados, a preparação do dado para recuperação e análise. Análise de dados (*Analytics*) refere-se às técnicas de obtenção de inteligência a partir dos dados, ou seja, técnicas para modelagem, análise e interpretação dos dados.

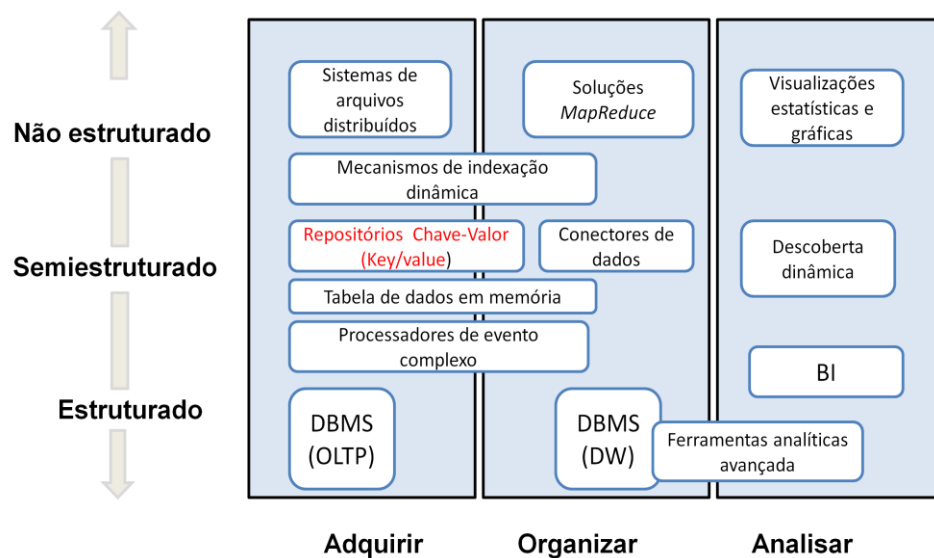
FIGURA 7 - Processo de extração de obtenção de inteligência a partir de dados



Fonte: adaptado de Gandomi e Haider (2015).

O processo de *big data* compreende as etapas de aquisição, organização e análise de dados que variam de acordo com o nível de estruturação dos dados, como apresentado na FIGURA 8.

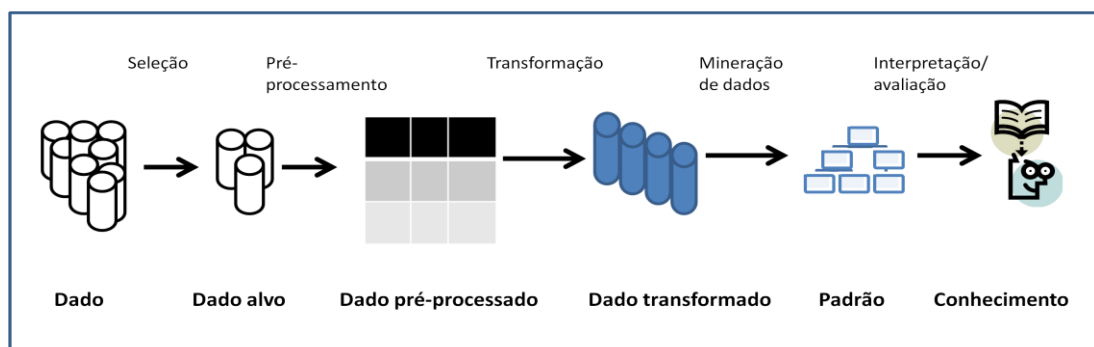
FIGURA 8 - Big data – estruturação de dados



Fonte: adaptado de Oracle Corporation (2012).

A utilização de *big data* permite aos gestores tomarem decisões baseadas em evidências e não apenas na intuição (MCAFEE; BRYNJOLFSSON, 2012). Segundo Fayyad, Piatetsky-Shapiro e Smyth (1996), a mineração de dados ou *data mining* é considerada o processo de descoberta de conhecimento útil a partir de dados - *Knowledge Discovery in Databases* (KDD). Ela consiste no processo de analisar grandes volumes de dados sob diferentes perspectivas, a fim de descobrir informações úteis que dificilmente são encontradas num processo puramente visual quando é tratado um grande volume de dados. Para o processo de descoberta de conhecimento em banco de dados de grandes volumes (*Big Data*), Fayyad, Piatetsky-Phapiro e Smyth (1996) propõem os passos descritos na FIGURA 9.

FIGURA 9 - Passos para construção de conhecimento em uma base de dados



Fonte: adaptado de Fayyad, Piatetsky-Phapiro e Smyth (1996).

A revolução do *big data* encontrou um espaço nas instituições bancárias tendo em vista o grande volume de dados valiosos que estas instituições produzem e armazenam por anos. A utilização desses dados revelam informações preciosas para tomada de decisões, pois permite analisar comportamento de clientes, segurança, desempenho de produtos, dentre outras diversas análises possíveis. As informações de qualidade podem ser extraídas rapidamente, de maneira mais fácil, e transformadas em informações relevantes e bastante significativas para tomada de decisão (SRIVASTAVA; GOPALKRISHNAN, 2015). A análise de transações de clientes realizadas em canais eletrônicos pode identificar o comportamento de um cliente e as necessidades dele no momento. Análise de transações realizadas em determinado canal podem demonstrar qual o tipo de serviço mais interessa os clientes que utilizam aquele canal e, com isso, direcionar os investimentos para evolução do canal.

A indústria financeira é uma das indústrias mais direcionadas a dados, diferentemente de uma indústria manufatureira. Na indústria financeira não há produtos a fabricar/manufaturar. O principal ativo produzido pelos bancos é o dado. Em um dia, milhões de transações são realizadas em canais eletrônicos de um banco de médio porte. Logo, trabalhar bem com os dados disponíveis e produzidos internamente por um banco é um diferencial competitivo. A geração e uso dos dados estruturados, por sistemas aplicativos do banco, já são bem consolidados e implementados, mas apenas recentemente os bancos descobriram o poder do uso de dados não estruturados produzidos. A união do uso dos dados estruturados com os dados não estruturados internamente pode produzir resultados valiosos para a tomada de decisão e aumento de competitividade.

Para competir em uma economia com poder de consumo, é cada vez mais claro que os bancos e as empresas do mercado financeiro precisam usar seus ativos de informação para obter uma compreensão de mercados, clientes, canais, produtos, regulamentos, concorrentes, fornecedores e funcionários. As instituições financeiras perceberão o valor gerenciando efetivamente e analisando o volume, velocidade e variedade de dados novos e existentes, e colocando as habilidades certas e ferramenta adequadas para entender melhor suas operações, os clientes e o mercado como um todo. (TURNER, SCHROECK; SHOCKLEY, 2013, *online*).

2.5 Vocabulário Controlado

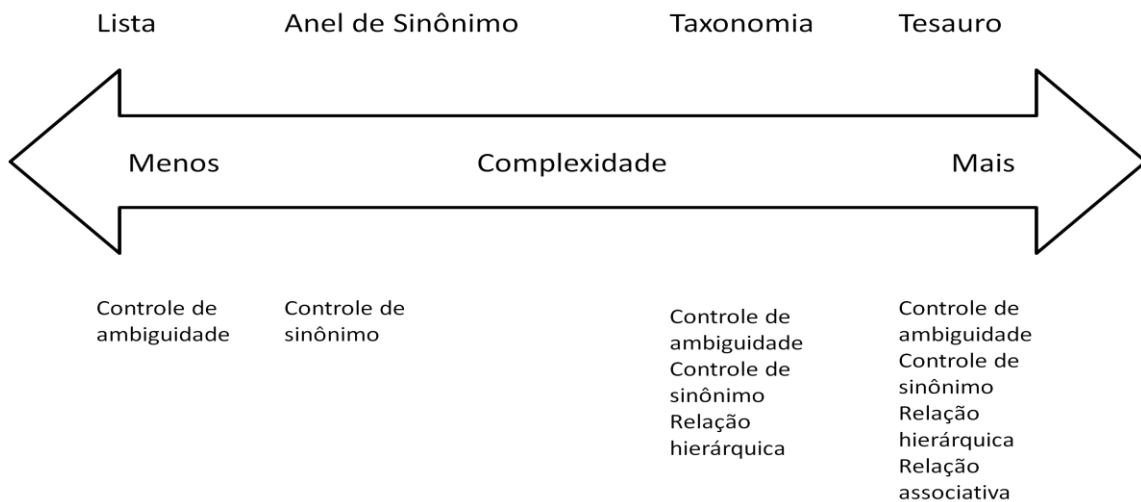
Segundo a norma ANSI/NISO Z39.19-2005 (NISO, 2005), o propósito primário de um controle de vocabulário é ter uma consistência na descrição do conteúdo e facilitar a recuperação de uma informação. O vocabulário controlado consiste de um conjunto de termos listados explicitamente. Essa lista é controlada por uma autoridade de registro. Todos os termos num vocabulário controlado devem ter uma definição não ambígua e não redundante. No mínimo, duas regras devem ser respeitadas: 1) Se o mesmo termo é usado

comumente com significados diferentes, então ele deve ser mais bem qualificado para resolver esta ambiguidade; 2) Se vários termos são usados para designar a mesma entidade, um dos termos deve ser identificado como o termo preferido no vocabulário controlado e os demais devem ser listados como sinônimos ou apelidos.

Um termo consiste de uma ou mais palavras que designam um conceito. Um sinônimo é uma palavra ou termo que possui exatamente o mesmo significado ou um significado bem próximo de outra palavra ou termo, Termos são selecionados de uma linguagem natural para inclusão em um vocabulário controlado. O objetivo principal de um vocabulário controlado é garantir que cada conceito distinto seja descrito por uma única forma linguística.

Na norma ANSI/NISO Z39.19-2005, são apresentados quatro tipos diferentes de vocabulário controlado. São eles: lista, anéis de sinônimos, taxonomia e tesouros. Os tipos variam dos menos complexos aos mais complexos como pode ser observado na FIGURA 10.

FIGURA 10 - Crescimento estrutural da complexidade entre os tipos de vocabulário controlado



Fonte: Aganette (2010) e NISO (2005).

Uma lista é um conjunto limitado de termos organizados em ordem alfabética ou de alguma outra maneira lógica. Listas são comumente usadas para descrever o conjunto de valores possíveis de uma entidade ou tipo de objeto.

Anéis de sinônimo consistem de um conjunto de termos que são considerados equivalentes para o propósito de recuperação de conteúdo. Não são usados no processo de indexação, apenas durante o processo de recuperação de conteúdo.

Taxonomia é um vocabulário controlado composto de termos preferenciais organizados em uma hierarquia ou poli-hierarquia.

Tesauros é um vocabulário controlado organizado em uma ordem conhecida e estruturado de tal forma que os vários relacionamentos entre os termos são exibidos claramente e identificados por indicadores de relacionamento padronizados.

No caso prático deste trabalho, os dados de um log serão representados no formato JSON como uma coleção de pares chave/valor. Para garantir a eficiência na recuperação de dados dos logs, entende-se que a criação de um vocabulário controlado, composto de um conjunto de termos que devem ser usados como chaves para representar os dados do log, é necessária. Dessa forma teremos uma única forma linguística para representar uma entidade de negócio e uma informação contida em um log.

Caso não seja definido um vocabulário controlado, haverá um risco do uso de termos diferentes para significar o mesmo tipo de dado. Como exemplo, pode-se citar o caso de ter o número de uma agência de uma conta presente em dois logs diferentes. Em um deles o termo usado para a “chave” do dado que contém a agência da conta foi “agência da conta”. No outro log, o termo usado foi “número da dependência”. Nesse caso, a recuperação de todos os logs cuja agência da conta seja igual a um número, usando o termo “agência da conta”, não será eficiente, pois o log que representou a agência da conta como “número de dependência” não será recuperado.

Outro ponto relevante é que usuários com perfis distintos e visões de negócio distintas podem definir termos diferentes para o mesmo objeto de negócio caso não haja um controle e a definição de um termo preferencial. Um exemplo interessante, relacionado a uma transação de pagamento de conta, é: na visão de um canal de atendimento, um valor é pago e na visão do convênio o pagamento é recebido. Logo, pode-se ter um termo “valor pago” e um termo “valor recebido” representando a mesma coisa, ou seja, o pagamento de uma conta efetuado por um cliente num canal de atendimento e repassado para a empresa conveniada.

A necessidade de um vocabulário controlado bem estabelecido e documentado é de grande importância para a qualidade do novo repositório de logs armazenado de forma semiestruturada usando conjunto de pares <chave>/<valor> no padrão JSON (JSON, 2017). A taxonomia é um tipo de vocabulário controlado adequado para o armazenamento dos dados de logs.

2.6 Taxonomia Corporativa

De acordo com a norma ANSI/NISO Z39.19-2005 (NISO, 2005), a taxonomia é um vocabulário controlado que consiste de termos preferidos, todos eles conectados em uma hierarquia ou poli-hierarquia. A taxonomia é uma coleção de termos de um vocabulário controlado organizados em uma estrutura hierárquica. Cada termo em uma taxonomia está em um relacionamento pai/filho, mais amplo ou mais restrito em relação a outros termos na taxonomia.

Conforme Terra *et al.* (2005), o uso da taxonomia na Ciência da Informação tem o seguinte significado:

A taxonomia é um sistema para classificar e facilitar o acesso à informação, e que tem como objetivo: representar conceitos através de termos; agilizar a comunicação entre especialistas e entre especialistas e outros públicos; encontrar consenso; propor formas de controle de diversidade de significados; e oferecer um mapa de área que servirá como guia em processo de conhecimento. É, portanto, um vocabulário controlado de uma determinada área do conhecimento e, acima de tudo, um instrumento ou elemento de estrutura que permite alocar, recuperar e comunicar informações dentro de um sistema, de maneira lógica. (TERRA *et al.*, 2005, *online*)

Segundo Martinez *et al.* (2004 apud VITAL; CAFÉ, 2007):

A taxonomia, em um sentido amplo, é a criação da estrutura (ordem) e dos rótulos (nomes) que ajudam a localizar a informação relevante. Em um sentido mais específico, é o ordenamento e rotulação de metadados, que permite organizar sistematicamente a informação primária. (MARTINEZ *et al.*, 2004, p. 106).

De acordo com Conway e Sligar (2002 apud AGANETTE; ALVARENGA; SOUZA, 2010), em um contexto geral, a taxonomia tem o seguinte objetivo:

A taxonomia é desenvolvida para prover a uma instituição ou grupo uma estrutura comum de conceitos e relações entre esses conceitos, para estruturar os elementos léxicos da linguagem, produzindo uma rede semântica comum. [...] permite a elaboração de um vocabulário controlado para recuperar a informação, criar metadados, além de fornecer esquemas que orientam estruturas e leiaute de página da web. (CONWAY; SLIGAR, 2002, *online*).

Para Terra *et al.* (2005), taxonomia não é um termo muito comum no mundo dos negócios, mas taxonomias são elementos estruturantes, estratégicos e centrais para negócios baseados em informação e conhecimento.

Segundo Leher (2015), a taxonomia pode ser usada para rotulagem de conteúdo corporativo.

A rotulagem apropriada do conteúdo corporativo (definição adequada de rótulos/palavras-chave/termos/tagging) pode ajudar as empresas a obter benefícios, como acelerar processos ou aumentar as vendas e evitar riscos de negócios, como multas ou ações judiciais. Uma taxonomia é um conjunto de termos que representam conceitos importantes para um determinado

tópico. Uma taxonomia age como uma linguagem comum para uma organização usar ao aplicar metadados ao seu conteúdo corporativo. O objetivo essencial de uma taxonomia é organizar a informação. Qualquer empresa que leve a sério o aprimoramento do gerenciamento de informações deve investir em metadados e taxonomia. (LEHER, 2015, *online*, tradução nossa).

Conforme Centelles (2005 apud AGANETTE; ALVARENGA; SOUZA, 2010), a taxonomia é um tipo de vocabulário controlado com as seguintes características:

A taxonomia corporativa é um tipo de vocabulário controlado que reflete o contexto, a audiência e os conteúdos de uma determinada organização e pode ter diferentes usos, pois permite a representação de todos os seus objetos informativos para desenvolver diferentes funções dos sites corporativos: a organização de conteúdos, a busca, a navegação, a filtragem de informações, etc. (CENTELLES, 2005, *online*).

Segundo Aganette, Alvarenga e Souza (2010), as taxonomias atuam como um instrumento para tornar o processo de recuperação da informação mais eficiente e mais ágil. A taxonomia auxilia no controle conceitual de cada termo utilizado nas organizações, facilitando a identificação de termos relacionados e de sinônimos que podem ser usados em uma busca de informação num sistema de recuperação de informação (AGANETTE; ALVARENGA; SOUZA, 2010). Camargo (2016) apresenta a importância da taxonomia corporativa para organização de informação e conseqüentemente tornar as empresas mais competitivas em seus mercados de atuação.

Para as Organizações Corporativas, a necessidade de criação de instrumentos de organização dos fluxos informacionais torna-se essencial. Tal estruturação possibilita que as empresas tornem-se mais competitivas em seus mercados de atuação. Sendo assim, as Taxonomias Corporativas demonstram-se como importantes ferramentas para as empresas. (CAMARGO, 2016, *online*).

Alguns problemas enfrentados pelas organizações justificam, segundo Camargo (2016), a importância da construção de taxonomias para empresas. Dois desses problemas, apresentados a seguir, se encaixam bem no contexto avaliado nesta pesquisa.

Inabilidade informacional: A inabilidade informacional caracteriza-se pela dificuldade de expressão precisa das informações desejadas pelo usuário da informação. A inabilidade informacional caracteriza-se também pela ambigüidade da linguagem natural, gerando impacto negativo nos resultados obtidos nas buscas pela informação.

[...]

Ausência de padrões terminológicos organizacionais: A ausência desses padrões gera na empresa uma diversidade inumerável de formas de organização e estruturação de documentos e informações. Essa diversidade apresenta morosidade na recuperação da informação desejada e, em muitos casos, a não localização do dado desejado. (CAMARGO, 2016, *online*).

No contexto de logs de transações bancárias, entende-se que a elaboração de uma taxonomia corporativa de dados para os processos e serviços oferecidos a clientes nos

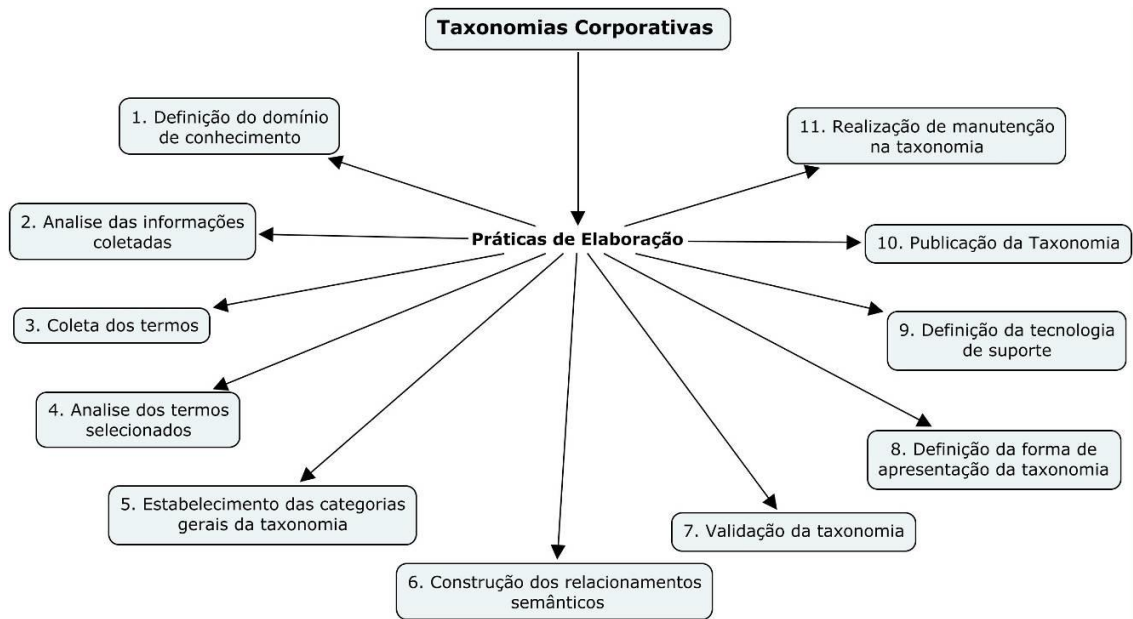
canais de atendimento é uma boa alternativa para o controle do vocabulário usado no repositório de dados. Ao analisar algumas transações e conceitos bancários, foi identificado que a estrutura hierárquica para a definição dos termos é bastante interessante no contexto das transações bancárias. Pode-se citar o seguinte exemplo: uma transação num ATM é realizada com um cartão; um cartão pode ser de diferentes tipos tais como: cartão exclusivamente de débito, cartão múltiplo, cartão exclusivamente de crédito; um cartão exclusivamente de débito pode ser de um dos tipos: cartão padrão Visa, cartão provisório, cartão com chip sem bandeira. Para recuperação de um conjunto de transações, o usuário pode desejar recuperar todas as transações realizadas com cartão ou somente as transações realizadas com cartão provisório. Através da relação hierárquica definida para “cartão” estabelecida na taxonomia, é possível que um sistema de recuperação de informação consiga recuperar, de forma simples, as transações para qualquer dos dois critérios apresentados anteriormente, desde que a hierarquia definida relacione o termo *cartão provisório* como filho do termo *cartão de débito* que por sua vez é filho do termo *cartão*. Um log de uma transação que só pode ser realizada com um cartão provisório, será armazenado com o número do cartão associado à chave (termo) *Cartão provisório*. Um log de uma transação que pode ser realizada com qualquer tipo de cartão de débito, será armazenado no repositório de logs no padrão JSON com o número do cartão presente no log associado à chave (termo) *Cartão de débito*.

Para a construção da taxonomia corporativa de dados será necessário aplicar uma metodologia que está apresentada na próxima seção.

2.7 Procedimento para elaboração de Taxonomias Corporativas

Aganette (2010) realizou um estudo em sua dissertação de mestrado, por meio de uma revisão de literatura, sobre métodos existentes para construção de taxonomias que resultou na proposição de uma metodologia para construção de taxonomias corporativas. Tal metodologia, de modo sucinto, apresentou onze etapas distintas, conforme Figura 11.

FIGURA 11 - Mapa Conceitual das Práticas de Construção de Taxonomias Corporativas



Fonte: (AGANETTE, 2010).

O referido estudo de Aganette (2010) reconheceu a importância do uso das Taxonomias Corporativas nas Organizações e apresentou as principais etapas que devem ser adotadas em sua construção. De acordo com a autora, o como executar essa estruturação taxonômica se apresenta como o grande desafio. Mais tarde, tal estudo resultou em um trabalho mais detalhado, de Aganette e Teixeira (2017). Este último trabalho apresentou uma proposta de um procedimento operacional, de modo mais exaustivo, com a exposição de atividades para a construção de taxonomias corporativas. Trata-se de um conjunto e atividades relacionadas em 8 etapas: (i) Levantamento terminológico, (ii) Seleção de termos, (iii) Definição de termos, (iv) Definição das categorias; (v) Ordem de citação das categorias; (vi) Normalização gramatical; (vii) Validação da taxonomia; (viii) Manutenção da taxonomia.

As autoras citam que o estudo se trata de uma proposta de procedimento passível de aplicação a sistemas de informação organizacionais. Apresentam, também, a necessidade de uma parceria entre teoria e prática para consolidação do procedimento de construção de taxonomias.

Como possíveis desdobramentos, pensa-se em se propor procedimentos específicos para cada etapa de construção da taxonomia, propor recomendações para o processo de construção de taxonomias em ambientes corporativos, sistematizar critérios que possam orientar a construção das taxonomias e analisar ferramentas de taxonomias existentes e, dentre essas, definir uma que auxilie até se chegar ao propósito do trabalho em questão. (AGANETTE; TEIXEIRA, 2017, *online*).

Concluem o trabalho citando a importância de conciliar o conhecimento científico às práticas de mercado e sugerem como desdobramentos, propor, conforme ANEXO A – Procedimento Operacional para Construção de Taxonomia Corporativa, procedimentos específicos para cada uma das 8 (oito) sub-etapas.

3 METODOLOGIA

Os próximos tópicos referem-se à caracterização da pesquisa e aos procedimentos metodológicos utilizados no processo de coleta e análise dos dados, etapas essas que se fazem necessárias para fornecer consistência, veracidade aos fatos e validade científica à pesquisa.

3.1 Caracterização

A metodologia adotada nesta pesquisa se pautará no método qualitativo e aplicado, pois realizará uma análise tanto da literatura da área quanto da aplicação de um procedimento de construção de taxonomia corporativa no contexto bancário, com intuito de propor uma taxonomia para um recorte do referido domínio. A utilização de diferentes formas de pesquisas se justifica, uma vez que o campo em que o objeto de estudo está inserido ainda não dispõe de um conjunto sistematizado e consolidado de conhecimentos e pesquisas. Assim, uma das maneiras encontradas de se minimizar a questão e ampliar a possibilidade de possíveis descobertas têm sido a exploração e análise das práticas e técnicas existentes na literatura, com o objetivo de realizar inferências, caracterizando, assim, o cunho qualitativo e exploratório deste estudo.

Segundo Silva e Menezes (2005), pesquisar significa procurar respostas para indagações propostas. Minayo² (1993 *apud* SILVA; MENEZES, 2005) considera que pesquisar é uma atividade de aproximação sucessiva da realidade que nunca se esgota, fazendo uma combinação particular entre teoria e dados. Considerando-se que a pesquisa aplicada é aquela em que o autor busca fazer um estudo científico voltado a solucionar algum problema específico, que já é conhecido e demonstrado no texto do trabalho, infere-se que tal pesquisa servirá tanto para gerar um novo conhecimento quanto para aplicá-lo na prática, em um caso real. Em relação à abordagem do problema, classifica-se como qualitativa, considerando que há uma relação dinâmica entre o mundo real e o sujeito que não pode ser quantificada e traduzida em números. O ambiente real é a fonte de coleta de dados e o pesquisador é o instrumento dessa coleta e avaliação.

A presente pesquisa percorreu as seguintes fases:

1. Revisão de literatura quanto às práticas e técnicas existentes para estruturação de dados, notação de metadados, mineração de dados e big data, construção de taxonomias corporativas;

² MINAYO, M. C. S. **O desafio do conhecimento**. São Paulo: Hucitec, 1993.

2. Análise da literatura encontrada, a fim de se definir um método para construção de uma taxonomia corporativa no domínio bancário para representação de dados de transações bancárias;
3. Aplicação de um procedimento de construção de taxonomia corporativa no contexto bancário.

3.2 Contexto, Coleta de Dados e Aplicação

3.2.1 Procedimentos Específicos - Fase 1

Para a Fase 1 - Revisão de literatura quanto às características, práticas e técnicas existentes na literatura para estruturação de dados, notação de metadados, mineração de dados e big data, construção de taxonomias corporativas, apresentada na seção 2 do presente estudo, abordou-se publicações e trabalhos científicos sobre os seguintes temas: logs de transações bancárias, repositório de dados estruturado e semiestruturado, notação para representação de dados de forma autodescritiva, *big data* e mineração de dados, vocabulário controlado e taxonomia corporativa.

Durante o levantamento preliminar de coleta de dados nas bases de dados do portal Capes, do *ResearchGate*, do *Google Acadêmico* e na base de relatórios e publicações do *Gartner Group*, utilizaram-se, nas estratégias de busca, os seguintes termos, agrupados por assunto, conforme Quadro 3 a seguir:

QUADRO 3 - Termos para pesquisa bibliográfica

Assunto	Termos
Repositório de dados estruturado e semiestruturado	<ul style="list-style-type: none"> • “banco de dados estruturado”, “banco de dados estruturados”, “bancos de dados estruturados”, “structured database”; • “banco de dados relacional”, “relational database”; • “banco de dados não estruturados”, “banco de dados não estruturado”; • “banco de dados semi-estruturado”, “banco de dados semiestruturado”, “semi-structured database”, “semi structured database”, “semistructured database”; • “dados estruturados”, “structured data”; • “dados não estruturados”, “unstructured data”; • “dados semi-estruturados”, “dados semiestruturados”, “dados semi estruturados”, “semi-structured data”, “semi structured data”.
Logs de transações	<ul style="list-style-type: none"> • “Transaction log”; • “Banking transaction log”; • “Banking log”; • “Log bancário”; • “Log de transação”, “log de transações”, “logs bancários”, “log bancário”, “log de transações bancárias”.
Big data	<ul style="list-style-type: none"> • “Big data”; • “big data analytics”; • “big data” and “data mining”; • “big data” and “bank”; • “big data” and “business intelligence”, “big data banking”.
Data mining	<ul style="list-style-type: none"> • “Data mining”, “mineração de dados”.
Canais de atendimento	<ul style="list-style-type: none"> • “Electronic banking channel”, “banking channel”, “canais eletrônicos”, “canal eletrônico”; • “canais de atendimento” e “bancos”; • “canal de atendimento” e “banco”.
JSON	<ul style="list-style-type: none"> • Json, “json metadata”.

Fonte: elaborado pela autora

Com relação ao assunto taxonomia corporativa e vocabulário controlado, a seleção de artigos e documentos foi realizada a partir dos artigos e trabalhos publicados pelos autores Aganette (2010), Vital e Café (2011) e das referências bibliográficas usadas por essas autoras nestes artigos.

Com a pesquisa realizada, conforme cada expressão de busca descrita no quadro 3, chegou-se a listas de resultados recuperados. Selecionaram-se artigos, livros, dissertações e relatórios que continham os termos apresentados no Quadro 3.

3.2.2 Procedimentos Específicos - Fase 2

Para a Fase 2 – Análise da literatura encontrada, a fim de se definir um método para construção de uma taxonomia corporativa no domínio bancário para representação de dados de transações bancárias – foram avaliados os textos selecionados com os seguintes focos:

- i. Na seção 2.1 *LOGs de transações bancárias*, foram analisadas algumas pesquisas e publicações de órgãos conceituados mundialmente para definição de soluções de TI, especificamente no setor bancário, com o objetivo de demonstrar a importância do uso dos logs de transações e da busca de uma solução eficiente para tratá-los.
- ii. A seção 2.2 *Banco de dados estruturado versus banco de dados semiestruturado* teve como objetivo avaliar, a partir dos textos selecionados, as características e aplicações mais adequadas dos tipos de banco de dados estruturado, não estruturado e semiestruturado e como representar os dados de acordo com o seu grau de estruturação.
- iii. Na seção 2.3 *Padrão JSON* foi selecionada e avaliada a documentação específica da notação JSON para validar a sua aplicabilidade para representação de dados de forma autodescritiva.
- iv. Na seção 2.4 *Big data e Data mining* foram avaliados trabalhos relacionados aos temas *big data* e mineração de dados na indústria bancária e a importância deles para a tomada de decisões estratégicas em um banco.
- v. Nas seções 2.5 *Vocabulário controlado* e 2.6 *Taxonomia Corporativa*, usando-se como base os artigos Aganette (2010) e Vital e Café (2011), foram selecionados outros trabalhos científicos publicados por essas autoras e as referências bibliográficas usadas por elas em seus trabalhos publicados. A partir do estudo e avaliação da bibliografia selecionada foram compreendidos o conceito de taxonomia corporativa e a sua aplicação como solução para definição de um vocabulário controlado para a organização de documentos e informações de uma empresa.
- vi. Na seção 2.7 *Procedimentos para elaboração de Taxonomias Corporativas*, foram avaliadas as técnicas propostas para construção de taxonomia corporativa e avaliado em detalhes o procedimento operacional proposto por Aganette e Teixeira (2017) para elaboração de uma taxonomia corporativa.

3.2.3 Procedimentos Específicos - Fase 3

Por se tratar de uma pesquisa aplicada, para construção de taxonomia corporativa no contexto bancário, a seção 4 descreve detalhadamente os passos usados, tendo como base o procedimento operacional proposto por Aganette e Teixeira (2017), para a criação da taxonomia, apresentando a aplicabilidade de cada passo no contexto do problema e os resultados obtidos. Como a diversidade de transações e tipos de logs gerados identificados no início do levantamento dos processos de negócio é muito grande, delimitou-se o universo de processos de negócio e transações avaliadas, como descrito na seção 4, para aplicação do procedimento e avaliação do resultado.

4 APRESENTAÇÃO E ANÁLISE DOS RESULTADOS

O estudo foi dividido em três fases como descrito anteriormente. A seguir serão apresentados e analisados os resultados de cada uma das fases.

4.1 Fase 1 - Revisão de literatura quanto às práticas e técnicas existentes na literatura para construção de taxonomias corporativas

A fase 1 - revisão da literatura quanto às práticas e técnicas existentes para estruturação de dados, notação de metadados, mineração de dados e big data, construção de taxonomias corporativas - foi descrita na seção 2.

4.2 Fase 2 - Análise da literatura encontrada, a fim de se definir um método para construção de uma taxonomia corporativa no domínio bancário

Na fase 2, buscou-se identificar métodos e procedimentos para a construção de taxonomia corporativa. Nesta fase foram usados como apoio os artigos, sites, livros e dissertações relacionados à taxonomia corporativa, recuperados na coleta de dados nas bases de dados como descrito na seção 3. Após análise da literatura selecionada, foi escolhido o procedimento operacional proposto por Aganette e Teixeira (2017) para aprofundar o estudo e executar a aplicação em um caso real. Essa escolha foi feita considerando-se que as autoras já haviam feito uma avaliação de vários métodos e práticas de construção de taxonomias presentes na literatura e descreveram um procedimento detalhado com as várias etapas e atividades que mapearam a partir dos métodos analisados.

O objetivo deste trabalho, como proposto por Aganette e Teixeira (2017), consiste na avaliação de uma parceria entre as constatações teóricas e a prática organizacional para a construção de uma taxonomia corporativa de dados bancários gerados em canais de atendimento eletrônicos. Partindo-se da avaliação do procedimento operacional proposto pelas autoras, que sistematiza critérios e detalha as etapas de construção de taxonomias em ambientes organizacionais, foram avaliadas em detalhes as atividades propostas no contexto das transações oferecidas a clientes em canais de atendimento de um banco.

Nesta pesquisa aplicada, a taxonomia foi desenvolvida e avaliada para o armazenamento de dados e recuperação dos conteúdos informacionais presentes nos logs de transações bancárias. Esses logs são gerados pelos sistemas de canais eletrônicos quando o cliente ou usuário solicita a realização de uma operação bancária no canal eletrônico e conclui a transação bancária. O armazenamento dos logs foi realizado em um

repositório de dados semiestruturado onde cada log é representado como um registro deste repositório usando-se a notação JSON (tag/valor) descrita anteriormente.

Segundo Belém, Almeida e Gonçalves (2017), *tags* (rótulos) são os suportes mais efetivos para recuperação de informações em textos. A notação JSON representa um conjunto de metadados em forma de um texto que contém um conjunto de pares <tag(chave)>/<valor>. Tendo-se como referencial teórico e definição de conceito de dados, informação e conhecimento apresentada por (DAVENPORT, 1998), pode-se considerar que o valor é o dado e a *tag* acrescenta significado ao dado, ou seja, informação. O uso de *tags* bem definidas se torna então um suporte eficiente para recuperação de dados com significado em textos.

Considerando-se as características e diferenças de um banco de dados estruturado e um banco de dados semiestruturado apresentadas anteriormente, avaliou-se que o banco de dados semiestruturado é o mais adequado para representação dos logs. Esta avaliação considerou as características de dados semiestruturados descritas por Mello *et al.* (2000), Buneman (1997) e Hammer, McHugh e Garcia-Moline (1997) e as características dos logs de transações bancárias em que: (i) os esquemas seriam definidos após a existência dos dados, (ii) os logs possuem uma estrutura irregular com um conjunto diferente de atributos de uma estrutura para outra, (iii) a definição de tipo e tamanho de atributos não é rígida, (iv) apenas parte dos dados disponíveis pode ter uma estrutura.

A taxonomia obtida será, então, usada para definição dos termos controlados, de compreensão única pelas áreas de negócio do banco, que serão usados como tags (chaves) de cada campo do log representado na notação JSON (chave/valor) armazenado no repositório .

A instituição em que a pesquisa foi realizada é um banco brasileiro de médio porte. A pesquisa teve foco nos serviços bancários oferecidos aos clientes nos canais de atendimento eletrônico de autosserviço e canais disponíveis para atender clientes nas agências do banco com a intermediação de um funcionário – atendente de caixa ou gerente. Os sistemas de canais eletrônicos de atendimento são responsáveis por interagir diretamente com os clientes ou com funcionários do banco que atendem clientes, oferecendo serviços bancários dentro das regras do Sistema Financeiro Nacional (FEBRABAN, 2018b). Devem garantir segurança e integridade das transações realizadas pelos clientes e gerar evidências – logs de transação - da efetivação da transação no canal com todas as informações relevantes associadas à transação realizada. Esses logs são usados tanto para garantir a conclusão e repasse das transações para parceiros de negócio, contabilidade e órgãos reguladores do Sistema Financeiro Nacional quanto para uso em

processos judiciais e órgãos governamentais, quando solicitados, para comprovação da realização de uma transação.

Como apresentado no item Big data e data mining, trabalhar com o conteúdo dos logs possibilita obter conhecimento de clientes e do negócio bancário para tomada de decisão. Nesta pesquisa aplicada, serão tratados os processos de gerenciamento de dados, ou seja, os processos e tecnologias que suportam a aquisição e armazenamento de dados e a preparação do dado para recuperação e análise posterior.

Considerando que a diversidade de transações disponíveis a um cliente em um canal de atendimento de um banco é grande, para o escopo deste trabalho foram consideradas apenas as transações de pagamentos e transferências que podem ser realizadas nos seguintes canais de atendimento: *Internet Banking*, *Mobile Banking* (APPs), ATMs do próprio banco (caixa eletrônico), caixas e plataforma de agência.

4.3 Fase 3 - Aplicação de um procedimento de construção de taxonomia corporativa no contexto bancário

A seguir apresentam-se os resultados da aplicação de cada atividade proposta, de acordo com o procedimento de Aganette e Teixeira (2017).

Etapa 1 - Levantamento terminológico - Atividade 1.1 Analisar processo de negócio

Esta etapa consistiu da análise do processo de negócio da área da organização em que a taxonomia foi elaborada. Para esta etapa, o procedimento operacional propõe que seja feita uma análise do processo de negócio e dos documentos gerados nos processos. Devem ser registrados assuntos, conceitos e documentos relacionados aos processos.

Como o objetivo proposto inicialmente foi a organização das informações dos logs de transações bancárias realizadas por clientes nos canais de atendimento de um banco, a primeira atividade realizada foi a identificação de todos os canais de atendimento que se caracterizam como canais de autosserviço ou canais de uso de gerentes ou caixas de agência para atendimento e realização de transações para o cliente. Foi criada uma lista desses canais, com uma breve descrição de cada um, em uma planilha, conforme Apêndice A. Além disso, foram identificados os gestores de negócio responsáveis por cada canal e os responsáveis pelos sistemas de TI desses canais.

O canal *Internet Banking* foi eleito como piloto para criar a primeira lista de transações de pagamentos disponíveis para os clientes correntistas do tipo pessoa física. Essa decisão considerou a facilidade de acesso a esse canal pela proponente e a sua visão, no papel de cliente correntista pessoa física do banco, das transações de pagamentos

disponíveis nesse canal. Como resultado desta primeira sub-atividade, foi gerado um documento contendo a lista de transações bancárias de pagamentos do canal Internet Banking para um cliente correntista pessoa física.

Para cada um dos demais canais identificados na lista de canais, foram identificadas as transações de pagamentos oferecidos neles para os clientes. Gerou-se uma lista de tipos de pagamentos oferecidos aos clientes pelo banco e uma associação do canal ao tipo de pagamento oferecido nele, conforme Apêndice B.

Para cada transação dessa lista, foi mapeado o processo de negócio envolvido na realização da transação, desde a escolha da transação no canal pelo cliente até a sua finalização, quando o log é gerado pelo sistema, o recibo é emitido para o cliente e a transação é concluída com êxito. As etapas de negócio que antecedem a oferta da transação no canal, tais como a criação dos produtos, as negociações com entidades externas parceiras, legislação envolvida, entre outras, não foram tratadas no levantamento realizado, assim como as etapas de negócio que sucedem a conclusão da transação pelo cliente no canal, tais como: repasse das transações para entidades externas e órgãos reguladores, integração financeira e contábil da transação.

A proponente, exercendo o papel de gestora da informação, com orientações e apoio dos gestores de canais (área de negócio específica do banco) e gestor de produtos e serviços de pagamentos, realizou buscas de documentos e informações, na base interna de normativos e procedimentos operacionais do banco, no portal institucional do banco e com os responsáveis na TI por cada sistema de canal definido no escopo, relacionadas aos tipos de canais, serviços e transações oferecidos nos canais, e pagamentos permitidos no banco. Além disso, solicitou o acesso aos sistemas aos responsáveis por cada sistema de canal, em ambiente de homologação dos sistemas de TI, com todos os meios possíveis de acesso em cada canal, para clientes pessoa física. Solicitou, também, ao gestor de pagamentos os dados específicos, para a realização do pagamento no canal, tais como: códigos de barra, formulários de pagamentos com dados preenchidos, entre outros.

As transações eleitas foram realizadas no canal de forma completa e todos os passos necessários para sua realização foram documentados através de diagramas BPM. Teve-se uma atenção aos termos exibidos aos clientes nas telas apresentadas pelo sistema e nos recibos emitidos. O processo mapeado foi avaliado pelo responsável pelo sistema de canal que ajudou a complementar os passos não visuais realizados pelo processo. Um dos passos não visuais de finalização da transação é exatamente a geração do log da transação, objeto do nosso estudo.

Nessa etapa também foi obtida a lista de todos os tipos diferentes de logs de transações de pagamentos gerados por cada canal com uma documentação de sistema dos layouts de cada um dos tipos de logs (campos e posição).

Como resultado desta etapa, foram gerados os seguintes produtos:

- Lista de canais e breve descrição de cada um (APÊNDICE A – Lista de canais)
- Tipos de pagamentos aceitos pelo banco (APÊNDICE B – Pagamentos aceitos pelo banco)
- Lista de pagamentos disponíveis em cada canal (APÊNDICE C – Pagamentos aceitos em cada canal de atendimento)
- Lista dos tipos de logs de transações de pagamento e documentação de sistema dos layouts de cada tipo de log.
- Documentação BPM das etapas para realização de uma transação de pagamento por um cliente em cada canal. Nessa documentação foram incluídas informações dos termos apresentados nas telas e recibos. (APÊNDICE D – Diagrama BPM de um processo de pagamento de um convênio com código de barras no Internet Banking)

Observa-se que a metodologia proposta por Aganette e Teixeira (2017) não provê orientações de como identificar os processos de negócio a serem analisados, mapeá-los e identificar os documentos gerados. Com isso, foi considerado relevante descrever em detalhes como foi feito o mapeamento de processos e o levantamento de documentos e informações gerados para auxiliar a aplicação da metodologia por outras pessoas.

O procedimento das autoras também não oferece exemplos de modelos de documentos a gerar. Para enriquecer e complementar o procedimento operacional proposto, foram citados todos os artefatos gerados nesta etapa e incluídos nos anexos os modelos de documento e exemplo para cada um deles.

Outro ponto bastante importante e que não foi explicitado no procedimento proposto é a identificação das pessoas chave para realização deste levantamento. Sugere-se que seja acrescentada ao procedimento uma atividade antecessora a esta de identificação das áreas de negócio e pessoas-chave para iniciar o levantamento.

Nessa etapa foi necessário envolver especialistas e responsáveis por áreas de canais, produtos e serviços e sistema de canais do banco.

Etapa 1 - Levantamento terminológico - Atividade 1.2 Analisar documentos e procedimentos gerados de forma minuciosa

Essa etapa teve por objetivo analisar os documentos e procedimentos gerados de forma minuciosa. O ponto de partida dessa atividade foram os documentos produzidos e o conhecimento adquirido na etapa anterior.

No procedimento operacional proposto para esta etapa não é prevista a geração de qualquer artefato. Na prática, observou-se a importância de gerar um artefato para auxiliar nesta revisão. Também se percebeu a necessidade de fazer ajustes nos documentos gerados na etapa anterior. Num primeiro momento, este resultado foi registrado na forma de texto, como um relatório de validação. Observou-se, no entanto, que uma representação gráfica, por meio de mapas conceituais, foi mais eficiente. Foi iniciado um mapeamento gráfico das informações obtidas e registradas nos documentos produzidos, usando-se mapas conceituais. Foram elaborados mapas conceituais como o apresentado em APÊNDICE E – Mapa Conceitual.

Como resultados desta etapa foram, então, gerados um mapa conceitual consolidando as várias informações registradas nos artefatos produzidos na etapa anterior e uma versão revisada dos artefatos da etapa 1.1.

Sugere-se acrescentar ao procedimento operacional o resultado ou produto gerado desta etapa. No caso da aplicação prática, os produtos obtidos foram:

- Versão revisada dos documentos elaborados na etapa anterior.
- Mapa conceitual

Etapa 1 - Levantamento terminológico - Atividade 1.3 Realizar entrevistas com especialistas e responsáveis pela área e seus respectivos processos e documentos conforme Roteiro de entrevistas

Esta etapa não foi realizada conforme proposta no procedimento operacional, pois as interações com especialistas e responsáveis aconteceram nas etapas 1.1 e 1.2. Na etapa 1.1 foram identificadas as pessoas-chave nas áreas de negócio e na área de TI envolvidas nos processos avaliados. O diálogo com essas pessoas foi iniciado desde a atividade 1.1, sem um roteiro prévio de entrevistas. Na etapa 1.2, dúvidas foram sanadas com essas pessoas-chave e com outras que foram sendo identificadas na medida em que o mapa conceitual foi elaborado e que dúvidas surgiram.

Por se tratar de elaboração de uma taxonomia para organização de dados brutos (raw data) e não para a organização de documentos, o roteiro de entrevista não se mostrou adequado para o escopo deste trabalho. Considera-se como diferença entre documentos e dados as apresentadas em Glushko (2013), item 1.4 “Documents {and, or, vs} Data”.

Foi, então, acrescentada uma nova etapa no procedimento descrita em (1.3x)

Etapa 1 - Levantamento terminológico - Atividade 1.3x - Realizar revisão do mapa conceitual com especialistas e responsáveis pelas áreas de negócio³

A etapa 1.3x foi usada para realizar a validação do mapa conceitual elaborado na etapa 1.2. A revisão foi feita com gestores de canais, gestores dos produtos identificados no mapa conceitual e os responsáveis pelos sistemas de TI.

Nesta etapa procurou-se entender, com os gestores de produtos e canais e com os responsáveis pelos sistemas (TI), o que eram os dados presentes nos logs e não presentes no mapa conceitual, qual o relacionamento dos dados dos logs com os termos do mapa conceitual. Em alguns casos foi necessário envolver outra área de negócio e outras áreas de TI que usam as informações presentes nos logs para processos de contabilização, de prestação de contas e repasse de informações e recursos financeiros para parceiros, para esclarecer as dúvidas. A “entrevista” com essas novas áreas envolvidas foi realizada de uma forma não estruturada, tendo como base o mapa conceitual elaborado, o diagrama BPM e os layouts dos logs (documento).

Etapa 1 - Levantamento terminológico - Atividade 1.4 - Listar os termos e conceitos identificados no levantamento

Usando-se os artefatos produzidos nas etapas anteriores, principalmente o mapa conceitual, foi elaborada uma lista de termos.

Segundo Vital e Café (2011), as classificações existentes na corporação devem ser usadas como recurso auxiliar e os termos precisam fazer sentido ao grupo de usuários que irá utilizá-los. Usando-se isso como base, procurou-se utilizar, sempre que possível, os termos exibidos ao cliente nos canais eletrônicos e os termos usados no portal corporativo do banco para descrição de produtos e serviços.

Cada termo foi descrito de acordo com o entendimento obtido pela proponente nas etapas executadas. Usou-se um glossário de termos existente no portal interno do banco para apoiar a elaboração da descrição. Além desse glossário, que foi considerado limitado e com poucas descrições específicas do banco e dos usuários de negócio, foram usados o glossário do Banco Central do Brasil (BANCO CENTRAL DO BRASIL, 2018), a cartilha de canais de atendimento da FEBRABAN (FEBRABAN, 2018a), os documentos da

³ Essa etapa não faz parte do procedimento operacional descrito por Aganette e Teixeira (2017).

CIP (CIP, 2018a) (CIP, 2018b) e FEBRABAN (FEBRABAN, 2018c) relacionados à Plataforma Centralizada de Recebíveis (PCR).

Nesta lista preliminar de termos e conceitos, não houve uma preocupação com a linguagem e a forma como os termos foram listados e descritos.

Etapa 2 - Seleção de termos

Nesta etapa, foram avaliadas similaridades nas descrições dos termos da lista elaborada na etapa 1.4. Procurou-se entender se dois termos distintos eram sinônimos ou o que os tornavam diferentes um do outro.

Para facilitar esta avaliação, foi criada uma coluna na planilha de termos para listar os termos equivalentes. Os termos considerados sinônimos foram associados ao termo considerado mais apropriado para uso pelo banco, usando-se essa nova coluna. As linhas contendo os demais termos considerados sinônimos foram excluídas da planilha de termos. Os termos que estavam com uma descrição que identificava dois ou mais significados diferentes foram redefinidos como dois ou mais termos diferentes.

Exemplo: *Terminal de autoatendimento, caixa eletrônico, ATM do banco* foram considerados sinônimos. Optou-se por usar o termo Caixa eletrônico como termo preferencial e os demais foram colocados na coluna de sinônimos. Ponto de atendimento e canal, apesar de apresentarem similaridade na descrição, foram considerados termos distintos e, portanto, foram descritos como dois termos na lista. Caixa, na lista de termos produzida na etapa anterior, foi descrito como um funcionário do banco com o perfil de atendente numa agência num balcão de caixa e como um canal eletrônico de realização de transação. Neste caso, o termo foi redefinido como dois termos distintos: terminal de caixa e atendente de caixa.

Etapa 3 - Definição de termos

As atividades 3.1 e 3.2 do procedimento operacional têm o objetivo de fazer uma descrição verbal do conteúdo de um termo, identificando claramente as diferenças entre dois ou mais termos similares, e identificar o contexto em que determinado termo deve ser usado.

Nesta etapa foi necessário definir previamente um padrão de descrição para os termos que contemplou: descrição resumida, classificação preliminar (canal, pagamento, forma de recebimento de recurso), tipo de transação em que foi identificado, local onde foi identificado (tela, recibo da transação, log da transação, etc.).

Esta etapa foi um refinamento da etapa anterior. A descrição dos termos foi feita aprimorando a descrição feita anteriormente, considerando-se o termo como sujeito da oração e usando-se oração transitiva. Abaixo estão exemplos de como foram realizadas as descrições:

- Canal de atendimento: Consiste no meio eletrônico pelo qual o cliente realiza operações bancárias e movimentações em suas contas. Podem ser canais de autosserviço (Internet Banking, Mobile Banking, caixa eletrônico, Central de atendimento telefônico, ATM de redes parceiras) ou canais de atendimento pessoal onde o atendimento é realizado por um gerente ou caixa de agência (terminal de caixa, Plataforma de atendimento).
- Caixa eletrônico: Consiste em um dispositivo eletrônico que permite que clientes realizem saque de dinheiro e realizem outras operações bancárias tais como: pagamentos, transferências, consulta de saldo, depósito, investimento, contratação de crédito, etc. em salas de atendimento do banco que podem funcionar em horário diferenciado do horário de funcionamento das agências. Também pode ser conhecido como terminal de autoatendimento ou ATM do banco. Corresponde a um canal de atendimento bancário.
- Agência: Consiste na dependência física do banco autorizada a funcionar pelo Banco Central do Brasil destinada à prática das atividades para as quais a instituição esteja regularmente habilitada
- Ponto de atendimento: Identifica unicamente uma instalação física de um sistema informatizado de canal eletrônico que gera logs de transações bancárias realizadas naquele canal e localidade. Diferencia-se de canal de atendimento, pois cada agência física possui uma instalação autônoma de sistema informatizado que gera logs de transações e atende o conjunto de terminais de caixa daquela agência. Para um canal do tipo terminal de caixa há vários pontos de atendimento, um para cada agência física. Para um canal de autosserviço, existe um ponto de atendimento para cada canal.
- Boleto bancário: Consiste em uma forma de cobrança emitida por meio de um banco. Essa cobrança pode ser definida como todas as contas a pagar que são emitidas por um banco, referentes a qualquer tipo de serviço que foi contratado e intermediado pela instituição bancária. É um documento emitido por alguém que presta serviços ou vende produtos e

que tem como função a cobrança dos mesmos. Também conhecido como boleto de pagamento, boleto de cobrança, título bancário, ficha de compensação.

- Pagamento de convênio: Consiste em um pagamento emitido por concessionárias de serviços como água, luz, telefone e gás ou por entidades da administração pública. Podem ser emitidos com um código de barras ou com informações preenchidas como campos de um formulário pré-impresso. Para que um banco possa receber o pagamento é necessário firmar um convênio com a empresa emissora do pagamento. São também conhecidos como contas.

Etapa 4 - Definição das categorias

Esta etapa consiste na identificação de categorias, buscando-se identificar classes e subclasses no contexto do negócio.

O procedimento operacional sugere os princípios teóricos importantes para a definição de categoria que, de forma resumida, são: relevância, diferenciação, permanência, homogeneidade, simplicidade, determinação.

Segundo Vital e Café (2011):

O estabelecimento das categorias, como primeira etapa de construção da taxonomia, deve seguir tanto uma abordagem dedutiva como indutiva. No método dedutivo, estabelecem-se as categorias gerais em um trabalho de equipe, entre especialistas, profissionais da informação e usuários, de acordo com as necessidades percebidas por esse grupo. A esse grupo cabe prever todas as possíveis categorias que representem o universo informacional da organização, constituindo, assim, a estrutura geral da taxonomia. (VITAL; CAFÉ, 2011, p.44).

Ainda considerando Vital e Café (2011), as categorias gerais precisam denotar características fundamentais do objeto/processo para o qual a taxonomia está sendo desenvolvida.

Usando os princípios apresentados no procedimento operacional e os princípios para definição de categorias gerais e específicas apresentados por Vital e Café (2011), foi feita uma representação gráfica das hierarquias identificadas, a partir dos termos descritos anteriormente, com o apoio dos gestores de negócio. A visualização da hierarquia em forma de árvore facilitou a avaliação crítica das categorias identificadas. O APÊNDICE F – Diagrama hierárquico de termos - apresenta um extrato das relações hierárquicas identificadas.

No contexto desta pesquisa, em que foram avaliadas as transações disponíveis para clientes de um banco nos canais eletrônicos de atendimento, podemos citar o seguinte

exemplo de classe e subclasses identificadas nos termos avaliados. Pagamento foi considerado um termo geral e uma categoria de transação oferecida ao cliente num canal. Pagamento com código de barras foi considerado uma especialização do tipo de transação pagamento. Pagamento de boleto bancário (ficha de compensação) foi mapeado como um tipo de pagamento com código e barras, logo foi considerado como uma especialização de pagamento com código de barra.

As classificações foram feitas considerando as seguintes relações (BELÉM; ALMEIDA; GONÇALVES, 2017):

- Relação gênero/espécie: exprime os graus de superordenação e subordinação entre os conceitos. O termo superordenado representa o termo genérico (TG), enquanto o termo subordinado representa o termo específico (TE). O termo específico representa um conceito subordinado, que possui as mesmas características do conceito geral e mais alguma que o diferencia de outro conceito específico, subordinado ao mesmo conceito geral. Exemplo: Pagamento (TG) e Pagamento com código de barras (TE).
- Relação todo/parte: diferentemente da relação hierárquica gênero-espécie, na partitiva, o termo superordenado (um objeto ou conceito) é considerado um todo, do qual o termo subordinado é uma parte. Exemplo: O código de barras é uma parte do pagamento com código de barras, assim como a data de vencimento do pagamento e o valor do pagamento também são partes do pagamento.

Essas classificações foram realizadas refinando-se os diagramas hierárquicos elaborados anteriormente sendo que as relações entre os termos deveriam se encaixar em dois tipos: *é do tipo*, *é parte de*.

Etapa 5 - Ordem de citação das categorias

Na etapa 4 – Definição das categorias, utilizou-se a representação gráfica, em forma de árvore, para identificação das categorias. A ordem das categorias foi definida a partir do processo de negócio mapeado. Com essa forma de representação adotada, a ordem das categorias foi obtida naturalmente na etapa anterior.

Etapa 6 - Normalização Gramatical

Esta etapa foi relevante. Nas etapas anteriores não houve preocupação e cuidado com a normalização. Além de tratar a forma substantiva, masculina e singular, também foi normalizado o uso de letras maiúsculas e minúsculas nos termos.

Todas as siglas foram definidas com todas as letras maiúsculas. As demais palavras foram escritas com a primeira letra maiúscula e as demais minúsculas.

Foram definidos termos equivalentes com características especiais para serem usados no armazenamento de dados no repositório de logs como tags JSON. Para esses termos foram excluídos acentos e caracteres especiais dos termos principais para facilitar o armazenamento e recuperação em repositório de dados. Isso foi feito considerando que o tratamento e representação de caracteres especiais diferem de um sistema informatizado para outro de acordo com o sistema operacional e configuração de *software* utilizado. Também foram excluídas as contrações de preposição mais artigo, como “do” = “de” + “o”, e as preposições, sem perder a legibilidade do termo.

O Quadro 4 apresenta alguns exemplos de termos obtidos com o respectivos termos JSON.

QUADRO 4 - Termo principal x Termo JSON

Termo principal	Termo JSON
Pagamento Multa DETRAN MG com Código de Barra	Pagamento Multa DETRAN MG Codigo Barra
Data da NSU	Data NSU
Pagamento em Espécie	Pagamento Especie

Fonte: elaborado pela autora

Etapa 7 - Validação da Taxonomia

O resultado final foi apresentado aos gestores de canais e aos gestores de negócio de pagamentos, além das áreas de TI envolvidas. Alguns ajustes foram realizados como, por exemplo, a melhoria em algumas descrições e a escolha do termo mais adequado. Um exemplo de escolha de termo mais adequado foi a escolha do termo “Caixa Eletrônico”. Foram identificados quatro termos com a mesma descrição: Caixa eletrônico, ATM do banco, Terminal de autoatendimento, Caixa automático. A escolha do termo “Caixa Eletrônico” foi feita usando-se o termo usado pela Febraban para identificação do canal.

Etapa 8 - Manutenção da Taxonomia

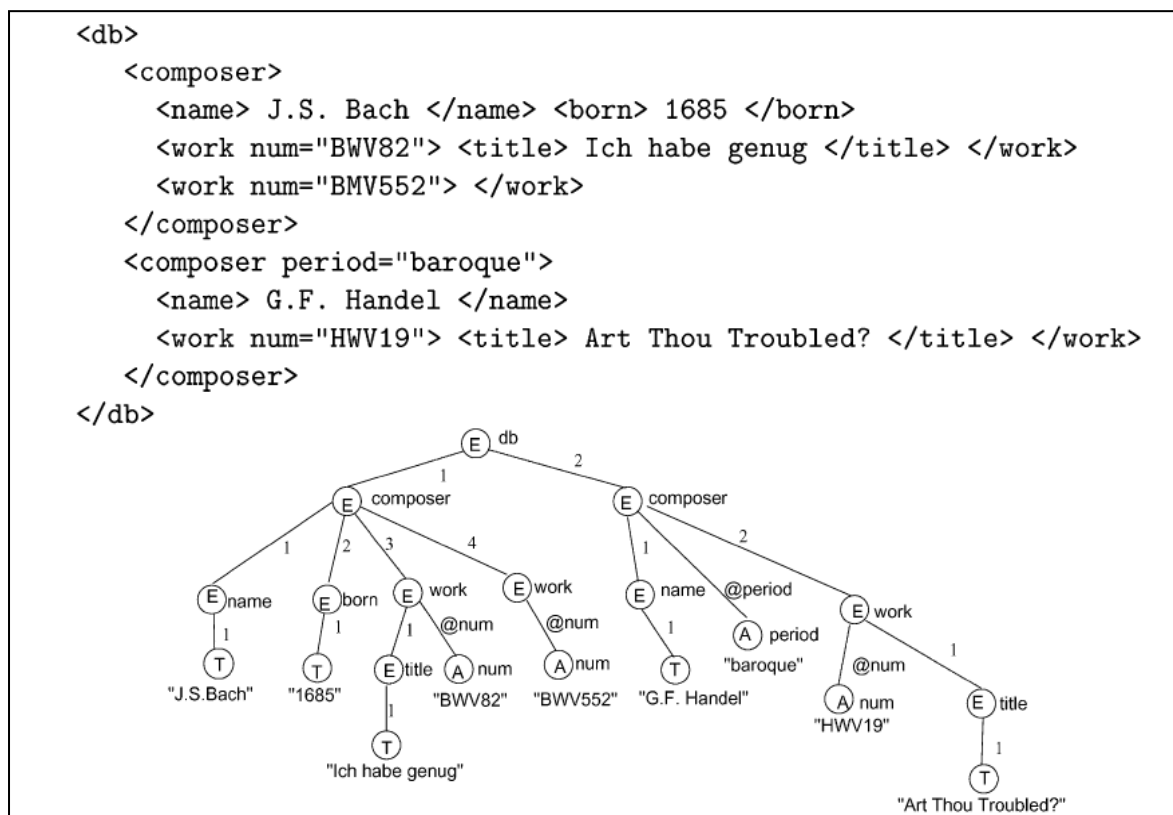
Essa etapa não foi realizada no contexto dessa pesquisa, pois a pesquisa concentrou-se na elaboração inicial de uma taxonomia de produtos e serviços bancários em canais de atendimento. No entanto, foi identificado que a inclusão de novos serviços e produtos oferecidos nos canais exigirá uma revisão da taxonomia elaborada.

Etapa 9 - Aplicação da taxonomia para representação de logs bancários⁴

Considerando-se que o objetivo da aplicação da taxonomia foi a definição e utilização do vocabulário controlado para representação dos logs de transações bancárias no padrão JSON, esta etapa consistiu na seleção de alguns tipos de logs de pagamento com código de barras gerados nos canais Internet Banking e Caixa Eletrônico e na representação desses logs na notação JSON usando os termos identificados na planilha para representação das chaves.

Para esta etapa foi usada a proposta apresentada por Buneman *et al.* (2002) que apresenta uma estrutura XML em forma de uma árvore hierárquica.

FIGURA 12 – XML e sua representação em árvore



Fonte: (BUNEMAN *et al.*, 2002).

⁴ Essa etapa não faz parte do procedimento operacional proposto por Aganette e Teixeira (2017).

Usando-se a representação gráfica apresentada no APÊNDICE I – Diagrama hierárquico de um LOG de transação em forma de uma árvore hierárquica e a taxonomia obtida nas etapas anteriores, um LOG de transação no padrão JSON foi representado como a FIGURA 13.

FIGURA 13 – Estrutura JSON de um log de pagamento

```
{
  "LOG":{
    "Identificacao LOG":{
      "Ponto atendimento":991,
      "Data NSU":"15-10-2018",
      "NSU canal":123765,
      "Data transação":"12-10-2018",
      "Hora transação":"10:43:54",
      "Data contabil transacao":"15-10-2018",
      "NSU estorno":0
    },
    "Canal":{
      "Codigo Canal":3,
      "Caixa Eletronico":{
        "Numero terminal":7611,
        "Sala atendimento":2001,
        "Agencia vinculo":296,
        "Agencia cofre":1
      }
    },
    "Pagamento":{
      "Meio pagamento":{
        "Debito conta":{
          "Numero agencia conta":574,
          "Numero conta debitada":562255,
          "Tipo conta debitada":"CC"
        }
      },
      "Pagamento convenio codigo barra":{
        "Pagamento codigo barra":{
          "Codigo barra":
"846500000059901090110035671768300114706491",
          "Forma captura":1,
          "Data movimento":"15-10-2018",
          "Codigo convenio":14534,
          "Codigo arrecadacao":51,
          "Valor pagamento":134.25,
          "Agencia recebedora":213,
          "Forma repasse":1,
          "Pagamento conta consumo":{
            "Data vencimento":"17-10-2018"
          }
        }
      }
    }
  }
}
```

Fonte: elaborado pela autora.

5 CONSIDERAÇÕES FINAIS

O objetivo geral do estudo, que foi a construção de uma taxonomia corporativa de dados para produtos e serviços bancários oferecidos em canais de atendimento eletrônico, por meio da aplicação do procedimento operacional de construção de taxonomia corporativa proposto por Aganette e Teixeira (2017), foi alcançado.

O vocabulário controlado obtido e a hierarquia definida na taxonomia foram aplicados com sucesso na representação de logs de transações bancárias no padrão JSON.

O objetivo de definição de uma taxonomia corporativa de dados para prover uma visão unificada do dado e para introdução de uma terminologia e semântica comuns a toda a organização foi alcançado.

Ao longo do trabalho, a importância de elaboração de uma taxonomia corporativa para um processo de negócio de um banco foi validada, reconhecida e valorizada pelas pessoas chave envolvidas na pesquisa aplicada. A taxonomia corporativa mostrou-se eficiente para a unificação de conhecimento e definição adequada e compartilhada de termos de negócio pelas diversas áreas da empresa. Estudos teóricos, realizados ao longo da pesquisa, relativos a características e importância de criação de uma taxonomia corporativa para organização de informações e geração de conhecimento foram comprovados na prática da pesquisa realizada.

As etapas propostas no procedimento operacional apresentado por Aganette e Teixeira (2017) foram uma boa orientação e muito relevantes para a elaboração da taxonomia corporativa de dados de produtos e serviços bancários em canais de atendimento. No entanto, foram necessárias adaptações e acréscimo de novas etapas no procedimento, principalmente considerando-se que o objetivo da elaboração da taxonomia corporativa nesta pesquisa foi a organização de dados e não de documentos. Como complementos, foram propostas orientações de como executar as etapas e quais os artefatos gerados em cada etapa. Também foi apresentada uma proposta de como mapear e documentar o processo de negócio. A partir da aplicação prática, foram avaliadas e sugeridas formas que se apresentaram mais eficientes para registrar e apresentar os resultados intermediários das etapas e para dialogar com os usuários chave para revisão e levantamento de informações necessárias. A atividade 1.3 sugerida no procedimento operacional não foi adequada para a construção de taxonomia de dados. Foi, então, proposta e descrita uma nova atividade em substituição a esta para realizar a revisão e complementação do levantamento terminológico necessário com os especialistas e responsáveis pelas áreas. A etapa de normalização gramatical foi adaptada para a utilização dos termos como chaves (tags) numa estrutura JSON que será armazenada num repositório de dados. Foi acrescentada uma etapa final

ao procedimento que apresenta como usar a taxonomia elaborada para organização dos dados em uma estrutura JSON. Como resultado, foi gerada uma versão mais detalhada do procedimento operacional, com novos insumos, adequada para a elaboração de uma taxonomia corporativa de dados.

O trabalho conseguiu estabelecer uma parceria entre teoria e prática para consolidação do procedimento de construção de taxonomia corporativa de dados. Foi possível fazer um elo entre um conteúdo teórico advindo da academia com uma aplicação prática deste conteúdo no setor bancário, demonstrando a importância de conciliar o conhecimento científico às práticas de mercado.

O alinhamento de disciplinas da Ciência da Computação com disciplinas da Ciência da Informação, tendo como objetivo um armazenamento e utilização eficiente dos dados gerados por sistemas de informação de uma empresa, para obtenção de informações para tomada de decisões estratégicas, também é considerado uma contribuição desta pesquisa.

Como limitação, pode-se citar que a taxonomia não abrangeu todos os serviços e operações bancárias disponíveis nos canais de atendimento. Foi necessário delimitar o escopo a apenas transações de pagamentos e transferências. Essa delimitação foi devida ao tempo disponível pela proponente e o prazo regulamentar para a defesa da monografia, à quantidade de pessoas da organização necessárias para levantamento de todos os processos e à disponibilidade delas durante o período da pesquisa. Contudo, fica aqui como apontamento para trabalhos futuros.

Como proposto inicialmente, o trabalho teve foco na representação e armazenamento dos logs de transações. Não era objetivo desta pesquisa avaliar a recuperação e o uso dos logs por usuários de negócio para extrair informações e tomar decisões. Como continuação deste trabalho, planeja-se avaliar a recuperação e o uso de informações dos logs, no padrão JSON, armazenados no repositório semiestruturado de logs de transações bancárias. Outra avaliação futura importante e relevante é a da evolução e manutenção da taxonomia elaborada ao longo do tempo. Pode-se também avaliar a aplicação da taxonomia corporativa para organização e armazenamento de dados gerados por outras indústrias.

REFERÊNCIAS

- ABITEBOUL, S.; BUNEMAN, P.; SUCIU, D. *Data on the Web: from relations to semistructured data and XML*. San Francisco: Morgan Kaufman, 2000.
- AGANETTE, E. N. C. **Taxonomias Corporativas**: Um estudo sobre definições e etapas de construção fundamentado na literatura publicada. 2010. Dissertação (Mestrado em Ciência da Informação) - Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2010. Disponível em: <<http://www.bibliotecadigital.ufmg.br/dspace/handle/1843/ECID-87BJSR>>. Acesso em: 12 mai 2018.
- AGANETTE, E. N. C.; ALVARENGA, L. D.; SOUZA, R. R. Elementos Constitutivos do Conceito de Taxonomia. **Informação & Sociedade**: Estudos, p. v.20, n.3, p.77-93, 2010. Disponível em: <<http://www.brapci.inf.br/index.php/article/download/14352>>. Acesso em: 17 jun 2018.
- AGANETTE, E. C.; TEIXEIRA, L. M. D. **Taxonomias Corporativas**: uma proposta de procedimento operacional para construção baseada na teoria e na prática. XVIII ENANCIB, Marília, SP, 2017. Disponível em: <<http://enancib.marilia.unesp.br/index.php/xviiienancib/ENANCIB/paper/view/34>>. Acesso em: 07 jul. 2018.
- BANCO CENTRAL DO BRASIL. **Resolução nº 4.283, de 2013**. 2013. Disponível em: <<https://www.bcb.gov.br/pre/normativos/busca/normativo.asp?tipo=Resolu%C3%A7%C3%A3o&data=2013&numero=4283>>. Acesso em: 11 nov. 2018.
- BANCO CENTRAL DO BRASIL. **Glossário completo**. 2018. Disponível em: <<https://www.bcb.gov.br/pt-br/#!/c/GLOSSARIO/89>>. Acesso em: 10 nov. 2018.
- BELÉM, F. M.; ALMEIDA, J. M.; GONÇALVES, M. A. A Survey on Tag Recommendation Methods. **Journal on the Association for Information Science and Technology**, New York, p. 830-844, 2017.
- BUNEMAN, P. Semistructured data. **Proceedings of the sixteenth ACM SIGACT-SIGMOD-SIGART symposium on Principles of database systems**, Tucson, 11-15 May 1997. p.117-121. Disponível em: <<https://dl.acm.org/citation.cfm?id=263675>>. Acesso em: 16 nov. 2018.
- BUNEMAN, P. *et al.* Keys for XML. **Computer Networks**, p. 473-487, 2002. Disponível em: <<https://dl.acm.org/citation.cfm?id=371984>>. Acesso em: 21 set. 2018.
- CAMARGO, M. F. M. A Construção de Taxonomia para Estruturação e Recuperação de Informações Corporativas. **Perspectivas em Ciência da Informação**, Belo Horizonte, v.21, n.4, 2016. Disponível em: <<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/3433>>. Acesso em: 12 set. 2018.
- CENTELLES, M. Taxonomies for categorization and organization in web sites. **Hipertexto.net**, Barcelona, n. 3, 2005. Disponível em: <<https://www.upf.edu/hipertextnet/en/numero-3/taxonomias.html>>. Acesso em: 09 set. 2018.
- CHOO, C. W. Perception and use of information sources by chief executives in environmental scanning. **Library & Information Science Research**, v.16, n. 1, p. 23-40, 1994.

CHOO, C. W. **A Organização do conhecimento**: como as organizações usam a informação para criar significado, construir conhecimento e tomar decisões. São Paulo: SENAC, 2006.

CIP. **Plataforma centralizada de recebíveis**. 2018a. Disponível em: <<https://www.cip-bancos.org.br/Paginas/PCR.aspx>>. Acesso em: 11 nov. 2018.

CIP. **Manual de leiautes de arquivos da nova Plataforma de Cobrança NPC**. 2018b. Disponível em: <<https://www.cip-bancos.org.br/Compliance/MAPX-OP016-2009%20-%20Manual%20de%20Leiautes%20de%20Arquivos%20do%20DDA.pdf>>. Acesso em: 11 nov. 2018.

CONWAY, S.; SLIGAR, C. Building taxonomies. In.: __. **Unlocking knowledge assets**. Redmont: Microsoft Press, 2002. Cap. 6. Disponível em: <http://www.microsoft.com/mspress/books/sampchap/5516a.asp#100>. Acesso em: 20 set. 2018.

CROCKFORD, D. **The application/json Media Type for JavaScript Object Notation (JSON)**. 2006. Disponível em: <<https://tools.ietf.org/html/rfc4627>>. Acesso em 15 nov. 2018.

DAVENPORT, T. **Ecologia da Informação**: por que só a tecnologia não basta para o sucesso na era da informação. São Paulo: Futura, 1998.

ECMA. **ECMA-262**: ECMAScript Language Specification. 2009. Disponível em: <<https://www.ecma-international.org/publications/files/ECMA-ST-ARCH/ECMA-262%205th%20edition%20December%202009.pdf>>. Acesso em: 28 set. 2018.

ECMA. **Standard ECMA-404**: The JSON Data Interchange Syntax. 2017. Disponível em: <<https://www.ecma-international.org/publications/standards/Ecma-404.htm>>. Acesso em: 01 nov. 2018.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. The KDD Process for Extracting Useful Knowledge from Volumes of Data. **Communications of the ACM**, New York, p.27-34, 1996.

FEBRABAN. **Cartilha de Canais de Atendimento**. 2018a. Disponível em: <https://cmsportal.febraban.org.br/Arquivos/documentos/PDF/Cartilha%20Canais%20de%20Atendimento_31_10_17.pdf>. Acesso em: 2 nov. 2018.

FEBRABAN. **Canais de Atendimento**. 2018b. Disponível em: <<https://portal.febraban.org.br/pagina/3055/30/pt-br/canais-de-atendimento>>. Acesso em: 11 nov. 2018.

FEBRABAN. **Nova Plataforma de Boletos de Pagamento - Cobrança Registrada**. 2018c. Disponível em: <<https://portal.febraban.org.br/pagina/3150/1094/pt-br/servicos-novo-plataforma-boletos>>. Acesso em: 11 nov. 2018.

GANDOMI, A.; HAIDER, M. Beyond the hype: Big data concepts, methods, and analytics. **International Journal of Information Management**, Guildford, p.137-144, 2015.

GARTNER. **IT Glossary**. 2018. Disponível em: <<https://www.gartner.com/it-glossary/big-data>>. Acesso em: 20 nov. 2018.

GILLILAND, A. J. Setting the stage. In: BACA, M. **Introduction to Metadata**. Los Angeles: Getty Research Institute, 2016. Disponível em: <<http://www.getty.edu/publications/intrometadata/setting-the-stage/>>. Acesso em: 18 set. 2018.

GLUSHKO, R. J. Foundations for Information Organization Retrieval and User. In: _____. **The Discipline of Organizing**. [S.l.]: MIT Press, 2013. Cap.1.

HAMMER, J.; MCHUGH, J.; GARCIA-MOLINE, H. Semistructured data: the TSIMMIS experience. **Proceedings of the East European Conference on Advances in Databases and Information Systems (ADBIS)**, 1997. Disponível em: <<https://dl.acm.org/citation.cfm?id=2227685>>. Acesso em: 04 nov. 2018.

INNOVIA. **Administração de empresas**: entenda a diferença de organização e instituição. 2018. Disponível em: <<https://www.innovia.com.br/blog/business/administracao-de-empresas/>>. Acesso em: 01 nov. 2018.

JSON. **Introducing JSON**. 2017. Disponível em: <<http://json.org/>>. Acesso em: 1 out. 2018.

LEHER, M. **How to Efficiently Create a Corporate Taxonomy**. Denver. 2015. Disponível em: <https://www.wandinc.com/Data/Sites/1/pdf_documents/white-papers/wand-white-paper--how-to-efficiently-create-a-corporate-taxonomy.pdf>. Acesso em: 27 out. 2018.

MARTÍNEZ, A. *et al.* Las categorías o facetas fundamentales: una metodología para el diseño de taxonomías corporativas de sitios web argentinos. **Ciência da Informação**, Brasília, v. 33, n. 2, p. 106-111, 2004.

MCAFEE, A.; BRYNJOLFSSON, E. Big Data: The Management Revolution Exploiting vast new flows of information can radically improve your company's performance. **Spotlight on Big Data - Harvard Business Reviews**, p. 1-9, 2012.

MCGEE, J.; PRUSAK, L. **Gerenciamento estratégico da informação**: aumente a competitividade e a eficiência de sua empresa utilizando a informação como uma ferramenta estratégica. Rio de Janeiro: Campus, 1994.

MELLO, R. S. *et al.* Dados Semi-Estruturados. **XV Simpósio Brasileiro de Banco de Dados**. João Pessoa, PB, 2000. p. 475-513.

NISO. **Guidelines for the construction, format and management of monolingual controlled vocabularies - ANSI/NISO Z39.19**. 2005. Disponível em: <<https://www.niso.org/publications/ansiniso-z3919-2005-r2010>>. Acesso em: 1 nov. 2018.

ORACLE CORPORATION. **Initial Steps on the Journey through Big Data for Financial Services Institutions**. 2012. Disponível em: <<http://www.oracle.com/us/industries/financial-services/ofsa-big-data-wp-1728241.pdf>>. Acesso em: 29 out. 2018.

POKORNY, J. NoSQL databases: a step to database scalability in web environment. **International Journal of Web Information Systems**, v.9, n.1, 69-82, 2013.

PRIYANKA S. P.; NAGARAJ, V. D. Analysis of banking data using machine learning. **I-SMAC (IoT in Social Mobile Analytics and Cloud)**, p. 876-881, 2017. Disponível em: <<https://ieeexplore.ieee.org/document/8058305>>. Acesso em: 03 nov. 2018.

SILVA, E. L.; MENEZES, E. M. **Metodologia da pesquisa e elaboração de dissertação**. Florianópolis: Universidade Federal de Santa Catarina, 2005. Disponível em: <https://projetos.inf.ufsc.br/arquivos/Metodologia_de_pesquisa_e_elaboracao_de_teses_e_dissertacoes_4ed.pdf>. Acesso em: 11 nov. 2018.

SRIVASTAVA, U.; GOPALKRISHNAN, S. Impact of Big Data Analytics on Banking Sector: Learning for Indian Banks. **Procedia Computer Science**, New York, p. 643-652, 2015.

SUN, N. *et al.* iCARE: A framework for big data-based banking customer analytics. **IBM Journal of Research and Development**, Armonk, NY, 2014. Disponível em: <<https://ieeexplore.ieee.org/document/6964895>>. Acesso em: 20 out. 2018.

TERRA, J. C. *et al.* **Taxonomia**: elemento fundamental para a Gestão do Conhecimento. 2005. Disponível em: <<https://pt.slideshare.net/jcterra/taxonomia-elemento-fundamental-para-a-gestao-do-conhecimento>>. Acesso em: 05 out. 2018.

TURNER, D.; SCHROECK, M.; SHOCKLEY, R. **Analytics**: The real-world use of big data in financial services. How innovative banking and financial markets organizations extract value from uncertain data. Somers, NY. 2013. Disponível em: <https://www-935.ibm.com/services/multimedia/Analytics_The_real_world_use_of_big_data_in_Financial_services_Mai_2013.pdf>. Acesso em: 06 nov. 2018.

VITAL, L. P. **Recomendações para Construção de Taxonomia em Portais Corporativos**. Florianópolis: UFSC, 2007.

VITAL, L. P.; CAFÉ, L. M. A. Práticas de elaboração de taxonomias: Análise e recomendações. ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 8, Salvador. **Anais eletrônicos**, 2007. Disponível em: <<http://enancib.ibict.br/index.php/enancib/viiienancib/paper/viewFile/2833/1961>>. Acesso em: 12 out. 2018.

VITAL, L. P.; CAFÉ, L. Proposta para o desenvolvimento de taxonomias em portais corporativos. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 16, n. 4, p. 42-54, out/dez, 2011.

APÊNDICE A – Lista de canais

Canais de atendimento eletrônico	Definição	Sigla
Internet Banking	Ambiente bancário na internet. Site do banco onde os clientes conseguem realizar diversas transações bancárias (operações) como pagamentos de faturas, contas e tributos, consulta a saldo e extrato, aplicação e resgate de investimento, entre outros, sem ter que ir a uma agência bancária.	IBK
Mobile Banking ou APP do banco	Banco móvel. Aplicativos para <i>smartphones</i> e <i>tablets</i> que disponibilizam serviços bancários aos clientes. Por meios desses aplicativos o cliente pode realizar operações como pagamentos de faturas, contas e tributos, consulta a saldo e extrato, aplicação e resgate de investimento, entre outros.	APP /MBK
ATM do banco ou Caixa eletrônico ou Terminal de autoatendimento	Caixas eletrônicas ou caixas automáticas. <i>Automatic Teller Machine.</i> Dispositivo eletrônico que permite que clientes realizem saque de dinheiro e realizem outras transações bancárias tais como: pagamentos, transferências, consulta de saldo, depósito, investimento, contratação de crédito, etc. Terminais de autoatendimento de propriedade do banco, instalados em salas de atendimento de agência e locais autorizados pelo banco.	ATM GAA
Caixa de agência	Balcão de atendimento a cliente do banco por um funcionário de agência com a função de "caixa bancário". Funcionário de agência que realiza transações solicitadas por um cliente em um terminal de caixa. Transações tais como: saque, depósito, pagamentos, transferências; Termo pode ser usado para representar uma pessoa ou um ponto de atendimento a um cliente que contém um balcão, um computador, dispositivo (impressora) para autenticação e impressão de recibo e dispositivo para autenticação de cliente (cartão, senha) Sistema de atendimento utilizado por caixas de agência	ATA
Plataforma de atendimento	Sistema de atendimento usado por gerentes do banco para acompanhar as contas e produtos de um cliente do banco e realizar transações tais como investimento, empréstimos, abertura de conta, transferências, entre outras, solicitadas por um cliente.	PLA

Call Center ou Central de atendimento ou Gente Fone	Sistema de atendimento telefônico composto por uma URA (unidade de resposta audível) e atendentes que interagem com o cliente e realizam transações solicitadas pelo cliente tais como consulta de saldos, bloqueio e desbloqueio de cartão, solicitação de talões de cheque, dentre outras, usando um sistema de TI.	CCE
Banco 24 Horas (TecBan)	ATMs (caixas eletrônicos) da rede TecBan instalados em todo o país, em lugares conveniados com a Rede 24 Horas. Permitem aos clientes realizar saques e consultar extratos com cartões de débito ou cartões de crédito do banco. Rede parceira do banco que oferece milhares de terminais de atendimento espalhados por todo o território brasileiro.	B24H
Máquinas de cartão, Rede Visa ou POS de adquirentes Visa	Máquinas de cartão de débito ou crédito que permitem que o cliente pague uma despesa (conta) usando seu cartão de débito ou crédito com uma autenticação. O pagamento é intermediado pela rede Visa que garante o débito na conta do cliente no banco e o crédito na conta do lojista (adquirente).	VISA
Rede Saque e Pague	Rede parceira do banco. Possui caixas eletrônicos (ATMs) que permitem que clientes façam saques, depósitos e consulta de saldos.	RSP
Office Banking – Pagamentos Eletrônicos ou SPE - Sistemas de pagamentos eletrônicos	Canal que permite que clientes realizem transações de pagamentos em lote, usando arquivos CNAB240 para envio de pagamentos, usando o ambiente bancário na internet. Difere-se do Internet Banking pois permite ao cliente enviar pagamentos em arquivos e receber retornos de pagamentos também em arquivos. Os arquivos podem gerados e processados por sistemas de TI do cliente.	SPE
SPB	Sistema de Pagamento Brasileiro. Canal que permite o envio de TED para outro banco e o recebimento de TED na conta do cliente.	

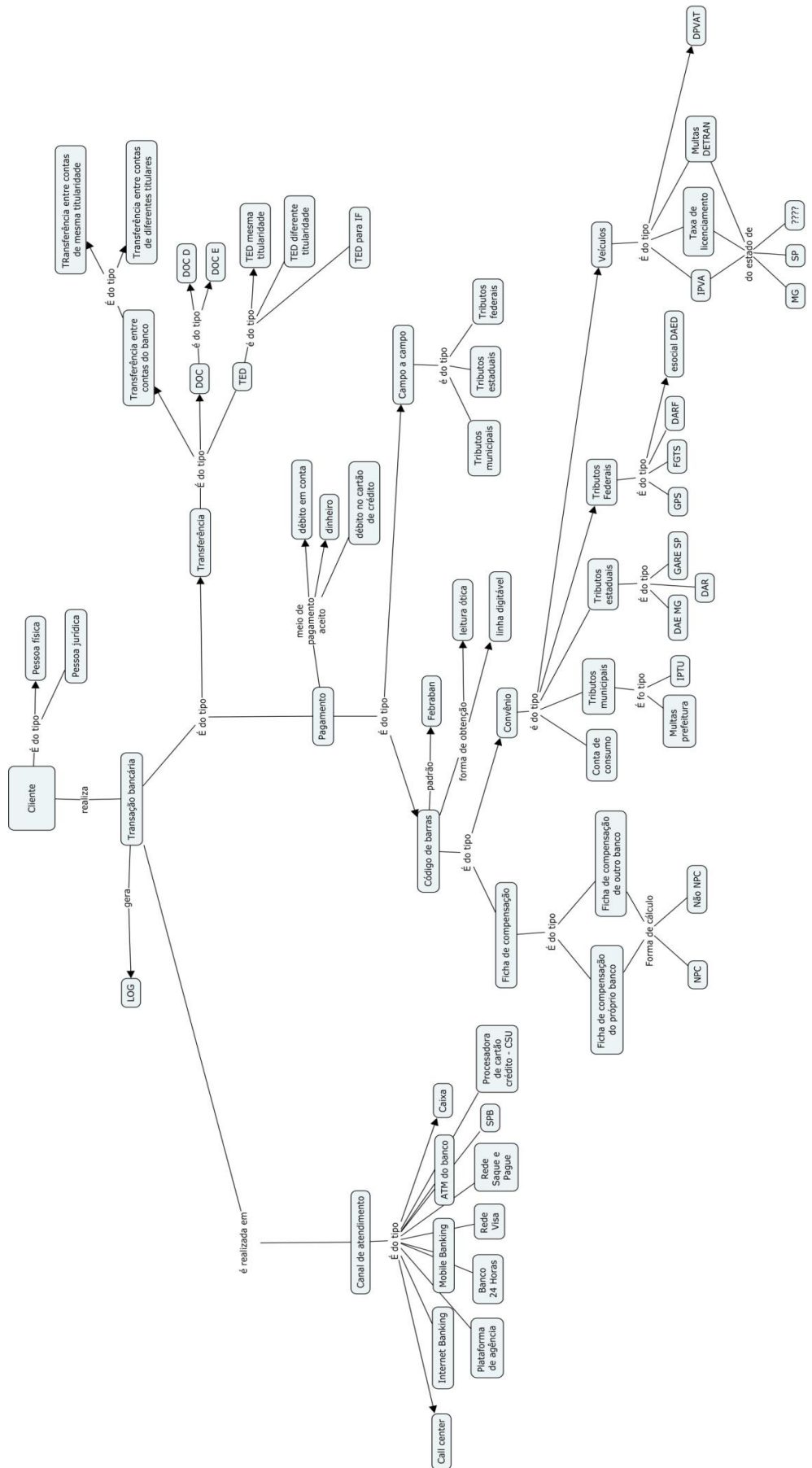
APÊNDICE B – Pagamentos aceitos pelo banco

Pagamento com código de barras	Pagamento de Ficha de compensação (boleto bancário) com código de barras	Pagamento de ficha de compensação do próprio banco
		Pagamento de ficha de compensação de outros bancos
	Pagamento de contas de Convênios com código de barras	Pagamento de Conta consumo
		Pagamento de Tributos municipais
		Pagamento de Tributos estaduais
		Pagamento de Tributos Federais
Pagamento de tributos de Veículos		
Pagamento campo a campo	Pagamento campo a campo de tributos municipais	
	Pagamento campo a campo de tributos estaduais	
	Pagamento campo a campo de tributos Federais	
	Pagamento campo a campo de tributos de veículos	
Transferência entre contas do banco	Transferência entre contas de mesma titularidade	
	Transferência entre contas de titulares diferentes	
	transferência para conta corretora	
DOC	DOC D	
	DOC E	
TED	TED de mesma titularidade	
	TED de titularidade diferente	
	TED para IF	
Depósito	Depósito em espécie	
	Depósito em cheque	

APÊNDICE C – Pagamentos aceitos em cada canal de atendimento

Canal	Tipo de Pagamento
Internet Banking	Pagamento com código de barras
	Pagamento campo a campo
	Transferência entre contas do banco
	DOC
	TED
Mobile Banking ou APP do banco	Pagamento com código de barras
	Transferência entre contas do banco
	DOC
	TED
ATM do banco ou Caixa eletrônico ou Terminal de autoatendimento	Pagamento com código de barras
	Pagamento campo a campo
	Transferência entre contas do banco
	DOC
	TED
	Depósito
Caixa de agência	Pagamento com código de barras
	Pagamento campo a campo
	Transferência entre contas do banco
	Depósito
Plataforma de atendimento	Transferência entre contas do banco
	TED
Call Center ou Central de atendimento ou Gente Fone	Nenhum. Foram descontinuados em julho/2018
Banco 24 Horas (TecBan)	Nenhum
Máquinas de cartão, Rede Visa ou POS de adquirentes Visa	Compras
Rede Saque e Pague	Depósito
SPE - Sistemas de pagamentos eletrônicos	Pagamento com código de barras
	Transferência entre contas do banco
	DOC
	TED

APÊNDICE E – Mapa Conceitual



APÊNDICE G – Extrato de lista de termos e descrição

id	Termo	Classe	Parte de	Definição	Sigla	Equivalente
1	Canal de atendimento			Consiste no meio eletrônico pelo qual o cliente realiza operações bancárias e movimentações em suas contas. Podem ser canais de autosserviço (Internet Banking, Mobile Banking, caixa eletrônico, Central de atendimento telefônico, ATM de redes parceiras) ou canais de atendimento pessoal onde o atendimento é realizado por um gerente ou caixa de agência (terminal de caixa, Plataforma de atendimento).		
2	Internet Banking	Canal		Ambiente bancário na internet (Site do banco) onde os clientes conseguem realizar diversas transações/operações bancárias sem ter que ir a uma agência bancária.	IBK	
3	Mobile Banking	Canal		Banco móvel. Aplicativos para Android ou iPhone que disponibilizam serviços bancários através de dispositivos móveis tais como celular e tablets.	MBK	APP
4	Caixa eletrônico	Canal		Dispositivo eletrônico que permite que clientes realizem saque de dinheiro e realizem outras transações bancárias tais como: pagamentos, transferências, consulta de saldo, depósito, investimento, contratação de crédito, etc. Terminias de autoatendimento de propriedade do banco, instalados em salas de atendimento de agência e locais autorizados	GAA	ATM do banco, Terminal de autoatendimento, Caixa automático

5	Terminais de Caixa	Canal		Consiste de um ponto de atendimento de um cliente que contém um balcão, um computador, dispositivo (impressora) para autenticação e impressão de recibo e dispositivo para autenticação de cliente (cartão, senha) em que o cliente pode fazer depósitos, saques e pagamentos. É operado por um funcionário de agência (atendente de caixa).	ATA	Terminal de caixa
6	Plataforma de atendimento	Canal		Sistema de atendimento usado por gerentes do banco para acompanhar as contas e produtos de um cliente do banco e realizar transações solicitadas pelo cliente.	PLA	
7	Call Center	Canal		Sistema de atendimento telefônico composto por uma URA (unidade de resposta audível) e atendentes que interagem com o cliente e realizam transações solicitadas pelo cliente usando um sistema de TI.	CCE	Central de atendimento telefônico, Gente Fone
8	Banco 24 Horas	Canal		ATMs da rede TecBan instalados em locais externos ao banco que permitem aos clientes realizar saques e consultar extratos. Rede parceira do banco que oferece milhares de terminais de atendimento espalhados por todo o território brasileiro.	B24H	TecBan, ATM da TecBan

9	Rede Visa	Canal		Conjunto de POS ou máquinas de cartão de débito ou crédito interligadas à Rede Visa que permitem que o cliente pague uma compra usando seu cartão de débito ou cartão de crédito. O pagamento é intermediado pela rede Visa que garante o débito na conta do cliente no banco associada a um cartão de bandeira Visa (emissor) e o crédito na conta do lojista (adquirente)	VISA	POS, VisaElectron
10	Rede Saque e Pague	Canal		Rede parceira do banco composta de caixas eletrônicos de sua propriedade (ATMs) espalhados por diversas localidades do Brasil que permitem que clientes façam saques, depósitos e consultem saldo.	S&P	
11	Sistema de Pagamento eletrônico	Canal		Canal que permite que clientes, principalmente PJ, realizem transações de pagamentos, usando o ambiente bancário na internet. Difere-se do Internet Banking pois permite ao cliente enviar lotes de pagamentos a efetuar e receber retornos de pagamentos em arquivos. Os arquivos podem gerados e processados por sistemas de TI do cliente.	SPE	SPE, Pagamentos eletrônicos, Office Banking

12	Pagamento	Serviço bancário		<p>Consiste da quitação de uma prestação de serviço ou aquisição de um produto por um cliente a um fornecedor do serviço ou produto.</p> <p>O pagamento pode ser realizado por meio de boleto bancário emitido pelo prestador de serviço, depósito bancário ou transferência a crédito na conta do prestador de serviço</p>		
13	Pagamento com código de barra	Pagamento		<p>Consiste em uma forma de realizar um pagamento de um produto ou serviço por meio de um documento de pagamento contendo um código de barras no padrão FEBRABAN</p>		
14	Pagamento campo a campo	Pagamento		<p>Consiste em uma forma de realizar um pagamento de um serviço ou tributo por meio de um documento de pagamento contendo campos pré-definidos que permitem a identificação do pagamento realizado que devem ser preenchidos pelo cliente com as orientações passadas pela entidade credora do pagamento.</p>		
15	Transferência	Pagamento		<p>Consiste em uma forma de realizar um pagamento a um terceiro por meio de transferência de valores de uma conta bancária para outra conta bancária.</p>		

16	Pagamento de ficha de compensação	Pagamento com código de barra, Pagamento		Consiste em uma forma de cobrança emitida por meio de um banco. Essa cobrança pode ser definida como todas as contas a pagar que são emitidas por um banco, referentes a qualquer tipo de serviço que foi contratado e intermediado pela instituição bancária. É um documento emitido por alguém que presta serviços ou vende produtos e que tem como função a cobrança dos mesmos.. Também conhecido como boleto de pagamento, boleto de cobrança, título bancário, ficha de compensação.		Pagamento de boleto bancário. Pagamento de título. Pagamento de boleto de pagamento
17	Pagamento de convênio com código de barra	Pagamento com código de barra, Pagamento, Convênio		Consiste em um pagamento emitido por concessionárias de serviços como água, luz, telefone e gás ou por entidades da administração pública. Contém um código de barra que identifica o pagamento. Para que um banco possa receber o pagamento é necessário firmar um convênio com a empresa emissora do pagamento. São também conhecidos como contas.		
18	Pagamento de ficha de compensação do próprio banco	Pagamento de ficha de compensação com código de barra, Pagamento com código de barra, Pagamento		Consiste no pagamento de uma ficha de compensação ou boleto bancário em que o banco emissor do boleto é o mesmo banco recebedor do pagamento.		

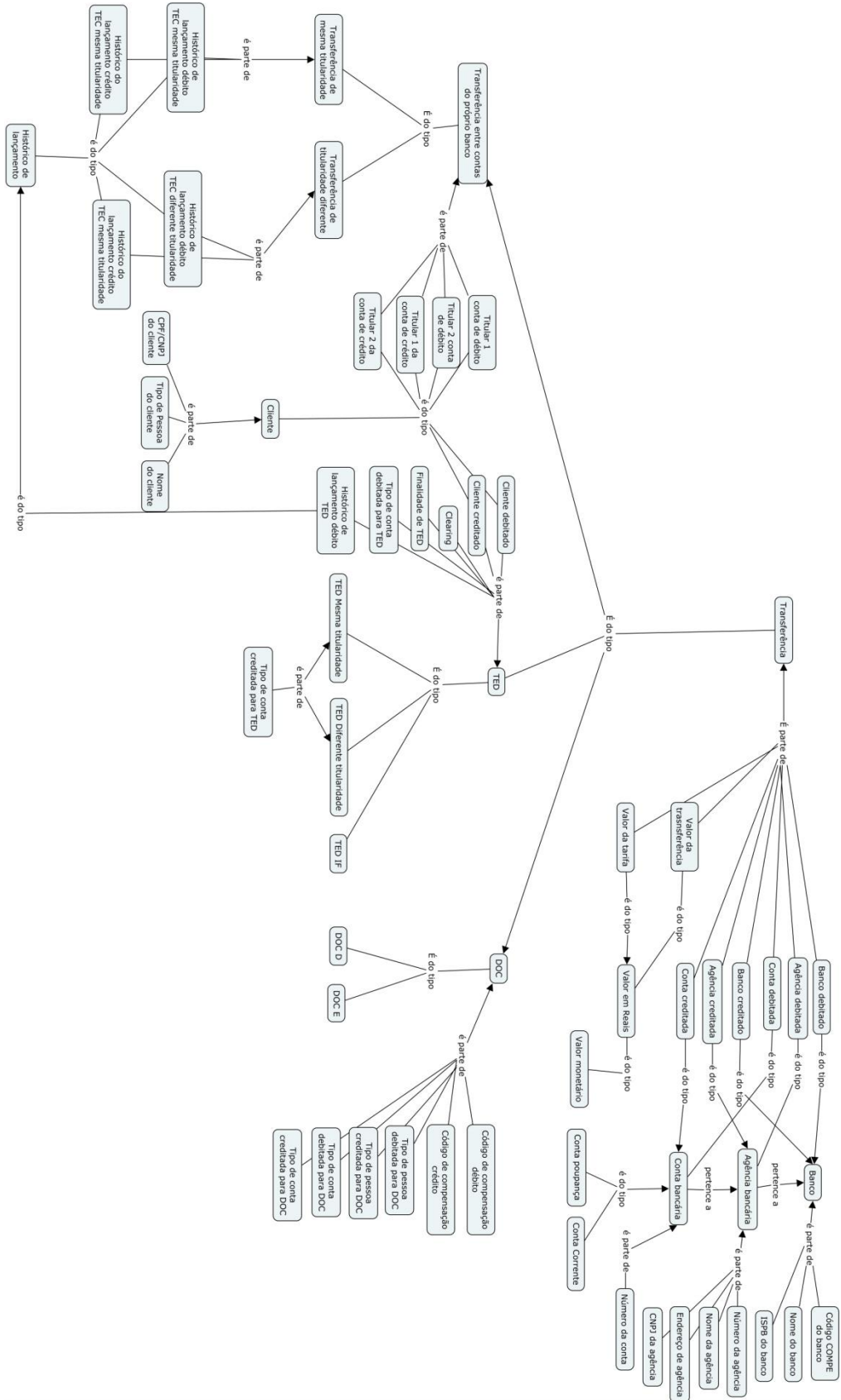
19	Pagamento de ficha de compensação de outro banco	Pagamento de ficha de compensação com código de barra, Pagamento com código de barra, Pagamento		Consiste no pagamento de uma ficha de compensação ou boleto bancário em que o banco emissor do boleto é diferente do banco recebedor do pagamento.		
20	Pagamento Conta de consumo código barra	Pagamento de convênio com código de barra, Pagamento com código de barra, Pagamento, Convênio		Consiste no pagamento de um convênio com código de barra, do tipo conta de consumo (água, luz, telefone, gás), emitido pela concessionária do serviço com um código de barras.		
21	Pagamento tributo municipal código barra	Pagamento de convênio com código de barra, Pagamento com código de barra, Pagamento, Convênio		Consiste no pagamento de um convênio com código de barra, do tipo tributo municipal, emitido pelo órgão municipal credor do tributo.		
22	Pagamento tributo estadual código barra	Pagamento de convênio com código de barra, Pagamento com código de barra, Pagamento, Convênio		Consiste no pagamento de um convênio com código de barra, do tipo tributo estadual, emitido pelo órgão estadual credor do tributo.		
23	Pagamento tributo federal código barra	Pagamento de convênio com código de barra, Pagamento com código de barra, Pagamento, Convênio		Consiste no pagamento de um convênio com código de barra, do tipo tributo federal, emitido pelo órgão federal credor do tributo.		

24	Pagamento prefeitura código de barras	Pagamento tributos municipais código de barras, Pagamento de convênio com código de barra, Pagamento com código de barra, Pagamento, Convênio		Consiste num pagamento de tributo municipal com código de barra emitido por uma prefeitura.		
25	Pagamento multa municipal código de barras	Pagamento tributos municipais código de barras, Pagamento de convênio com código de barra, Pagamento com código de barra, Pagamento, Convênio		Consiste num pagamento de tributo municipal com código de barra emitido pelo órgão municipal fiscalizador.		
26	Pagamento DAR SEFAZ-SP código de barra	Pagamento tributos estaduais código de barras, Pagamento de convênio com código de barra, Pagamento com código de barra, Pagamento, Convênio		Consiste num pagamento de tributo estadual com código de barra do tipo Documento de Arrecadação do estado de São Paulo emitido pela secretaria da fazenda do estado de São Paulo.		

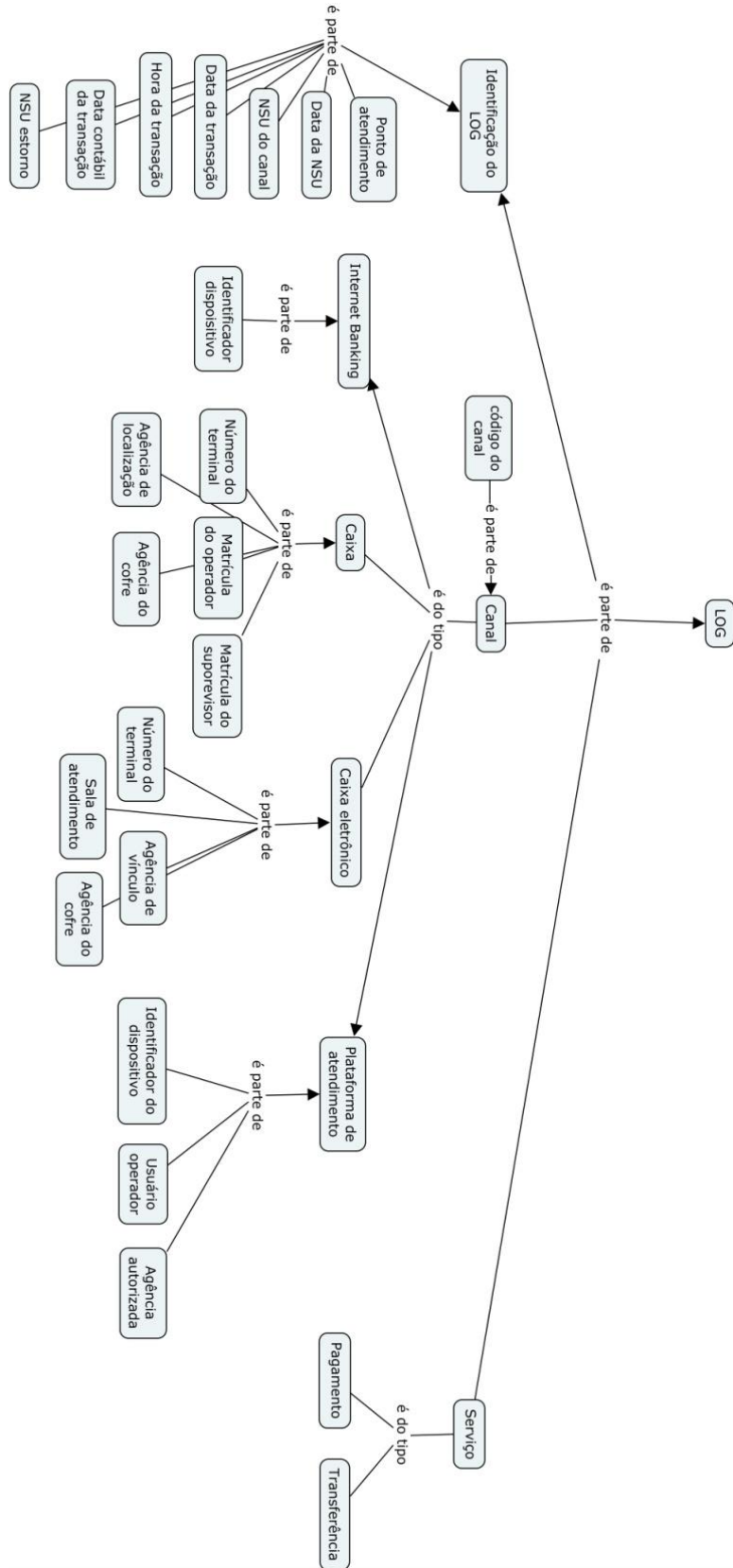
65	Meio de pagamento		Pagamento	Consiste no meio em que o banco obtém o recurso financeiro do cliente para efetuar o pagamento. O cliente pode realizar o pagamento por meio de autorização de débito na sua conta, fornecimento de numerário (dinheiro em espécie) suficiente para quitar o valor do pagamento ou por meio de autorização de débito em seu cartão e crédito.		
	Débito em conta	Meio de pagamento		Consiste no meio de pagamento em que o cliente autoriza o débito do valor diretamente em sua conta. Para isso é necessário que o cliente assine eletronicamente a transação com o uso de seu cartão bancário na função débito e senha ou com a entrega de um documento assinado pelo cliente.		
	Pagamento em Espécie	Meio de pagamento		Consiste no meio de pagamento em que o cliente fornece um valor em dinheiro (espécie) suficiente para quitar o valor total do pagamento. Este meio pode ser usado em terminais de caixa ou caixas eletrônicos com recurso para receber pagamento em espécie e fornecer troco.		
	Débito cartão de crédito	Meio de pagamento		Consiste no meio de pagamento em que o cliente autoriza o débito do valor diretamente do seu limite disponível de cartão de crédito. Para isso é necessário que o cliente assine eletronicamente a transação com o uso de seu cartão de crédito e senha.		

APÊNDICE H – Taxonomia

Extrato da taxonomia elaborada para operações de transferência.



APÊNDICE I – Diagrama hierárquico do LOG



ANEXO A – Procedimento Operacional para Construção de Taxonomia Corporativa

A seguir está o procedimento operacional proposto por Aganette e Teixeira (2017, *online*), usado como base na elaboração deste estudo.

Objetivo: Este procedimento visa definir as atividades necessárias para elaboração e manutenção da taxonomia em contextos corporativos, com intuito de garantir a qualidade do processo de organização, estruturação e recuperação das informações.

Abrangência: Aplicável a todas as áreas, setores e processos de uma instituição.

Responsabilidade: É de responsabilidade do Gestor da Informação as atividades descritas neste procedimento.

Descrição das Atividades:

1. Levantamento Terminológico

Seq.	Atividade	Responsabilidade
1.1	1. Analisar processo de negócio da área e suas respectivas atividades e documentos. Essa etapa do levantamento terminológico da área, deve-se registrar assuntos, conceitos e documentos, por meio da análise de: a) Processos de negócio; b) Documentos gerados nos processos de negócio.	Gestor da Informação
1.2	2. Analisar documentos e procedimentos gerados de forma minuciosa.	Gestor da Informação
1.3	3. Realizar entrevistas com especialistas e responsáveis pela área e seus respectivos processos e documentos, conforme Anexo 1- Roteiro de Entrevistas.	Gestor da Informação
1.4	4. Listar os termos e conceitos identificados no levantamento, considerando-se: a) Informações necessárias para recuperação dos documentos; b) Produtos e serviços desenvolvidos pela área; c) Padrões e documentos de referência utilizados pela área; d) Relações de equivalência; e) Siglas e acrônimos.	Gestor da Informação

2. Seleção de termos

Seq.	Atividade	Responsabilidade
2.1	1. Selecionar os termos a serem utilizados na taxonomia, observando os seguintes critérios: a) Equivalência: listar os sinônimos em português, siglas ou acrônimos e, se for o caso, os termos equivalentes em outras línguas em um tesouro ou junto às definições. b) Contexto: a definição de escopo de um termo (ou categoria) é fornecida pelo contexto e baseado em sua posição no sistema de classificação. c) Uso corrente: a terminologia usada num sistema de classificação deve refletir uso corrente de um campo de conhecimento.	Gestor da Informação

3. Definição dos termos

Seq.	Atividade	Responsabilidade
------	-----------	------------------

3.1	1.Elaborar a descrição verbal do conteúdo de um termo, visando distingui-lo de outro, caso necessário. A descrição só é necessária se o mesmo possuir margem para dúvidas.	Gestor da Informação
3.2	2.Determinar o contexto em que determinado termo deve ser utilizado.	Gestor da Informação
4. Definição das Categorias		
<i>Seq.</i>	<i>Atividade</i>	<i>Responsabilidade</i>
4.1	1.Identificar o assunto básico, suas classes e subclasses de forma lógica, em uma estrutura que reflita o conhecimento em seu contexto de negócio, a fim de definir as categorias a serem utilizadas na taxonomia.	Gestor da Informação
4.2	2.Seguir os seguintes princípios teóricos no momento da criação e definição das categorias: a) Relevância: refletir o propósito, o escopo do assunto, o processo na taxonomia. b) Categorias fundamentais: defini-las a partir da natureza do assunto a ser classificado. c) Diferenciação: usar uma característica de divisão para decompor a categoria em partes. Essa deve distinguir uma categoria das demais. d) Permanência: representar qualidades permanentes do item que está sendo dividido. e) Homogeneidade: o conteúdo de uma categoria não pode se sobrepor à outra. f) Simplicidade / Economia: estender a classificação até o nível de uso prático, considerando a quantidade de conteúdo que será categorizado. g) Determinação: escolher categorias que possibilitem a identificação das subcategorias subordinadas.	Gestor da Informação

5. Ordem de citação das categorias		
<i>Seq.</i>	<i>Atividade</i>	<i>Responsabilidade</i>
5.1	1.Ordenar as categorias por um critério relevante para a natureza, assunto e escopo da taxonomia (Sequência relevante). Nota (1): Sugere-se que a elaboração da taxonomia seja realizada via análise dos processos de negócio da empresa e pela terminologia adotada na concepção das informações pertinentes aos mesmos, de acordo com os processos mapeados. Nota (2): Uma vez estabelecida a ordenação das categorias, sequência consistente, deve evitar alterações constantes. Caso seja necessário, esta alteração deve ser registrada.	Gestor da Informação
6. Normalização Gramatical		
<i>Seq.</i>	<i>Atividade</i>	<i>Responsabilidade</i>

6.1	1.Descrever os termos na forma substantiva, masculina e singular. Exemplos: Premiação; Política. Nota: Exceções podem ser mencionadas. Exemplos: Recursos Humanos ----- Conceitos diferentes.	Gestor da Informação
7. Validação da Taxonomia		
Seq.	Atividade	Responsabilidade
7.1	1.Submeter a taxonomia à apreciação dos usuários e especialistas da área, que devem avaliar os seguintes pontos: a) Coerência Terminológica; b) Aplicabilidade dos conceitos definidos às atividades da área; c) Usabilidade dos termos eleitos em detrimento de seus sinônimos; d) Pertinência das alterações propostas no mesmo. Nota: A etapa de verificação pressupõe a análise da taxonomia por especialistas, para cada categoria modificada, devendo estes questionar a todo instante, a pertinência das alterações propostas na estrutura taxonômica. Essas etapas devem prever: a) A reclassificação dos documentos, com base na nova estrutura taxonômica; b) A verificação da classificação pelos especialistas; c) Os ajustes para homologação das alterações.	Gestor da Informação

8. Manutenção da Taxonomia		
Seq.	Atividade	Responsabilidade
8.1	1.Aplicar o Procedimento de Elaboração e Manutenção de Taxonomia.	Gestor da Informação
8.2	2.Acrescentar categorias na taxonomia: a) O acréscimo poderá ser sugerido por qualquer usuário da área. b) O gestor da informação será responsável por avaliar a pertinência da criação de um novo assunto, obedecendo aos princípios e conceitos estabelecidos neste POB. Outras orientações consideradas pertinentes: *A extensão do nome dos diretórios e seus respectivos documentos eletrônicos devem somar, no máximo, o número de 256 caracteres, a contar da raiz do diretório. Estas extensões favorecem o risco de perda da informação armazenada; *Deve-se evitar repetir o nome do diretório ao nomear arquivos subordinados ao mesmo (redundância); *O uso de siglas e abreviaturas é recomendável desde que o seu significado esteja claro, permitindo o seu reconhecimento por todos os usuários autorizados a acessar a informação; *Uma taxonomia deve direcionar de forma intuitiva a classificação dos documentos e informações de interesse comum de um grupo de usuários, favorecendo o conhecimento e o encadeamento lógico das atividades sob responsabilidade de uma equipe; *Para que esses benefícios possam efetivamente ser alcançados, é importante o esforço e o comprometimento de cada membro da equipe na implantação, utilização e manutenção da Taxonomia.	Gestor da Informação

Fonte: (AGANETTE; TEIXEIRA, 2017, *online*).