



UNIVERSIDADE DE BRASÍLIA
Faculdade de Ciência da Informação
Curso de Graduação em Biblioteconomia

METODOLOGIA DE *DOCUMENT REVIEW* COM BASES NA DOCUMENTAÇÃO
APLICÁVEL AO *ELETRONIC DISCOVERY*

Natália Bianca Mascarenhas Puricelli
Orientador: Prof. Dr. Marcílio de Brito

Brasília
2018

Natália Bianca Mascarenhas Puricelli

METODOLOGIA DE *DOCUMENT REVIEW* COM BASES NA DOCUMENTAÇÃO
APLICÁVEL AO *ELETRONIC DISCOVERY*

Monografia apresentada como parte das exigências para obtenção do título de Bacharel em Biblioteconomia pela Faculdade de Ciência da Informação da Universidade de Brasília

Orientador: Prof. Dr. Marcílio de Brito

Brasília

2018

P985m

PURICELLI, Natália Bianca Mascarenhas.

Metodologia de *document review* com bases na Documentação aplicável ao *eletronic discovery* / Natália Bianca Mascarenhas Puricelli. – Brasília, 2018.

46 f.

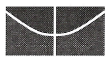
Orientação: Prof. Dr. Marcílio de Brito

Monografia (Bacharelado em Biblioteconomia) – Universidade de Brasília, Faculdade de Ciência da Informação, Curso de Biblioteconomia, 2018.

Inclui bibliografia

1. Documentação. 2. Recuperação da informação. 3. *Eletronic Discovery*. I. Título.

CDU025.4



Título: Metodologia de document Review com bases na Documentação aplicável ao eletrônico Discovery.

Aluna: Natália Bianca Mascarenhas Puricelli.

Monografia apresentada à Faculdade de Ciência da Informação da Universidade de Brasília, como parte dos requisitos para obtenção do grau de Bacharel em Biblioteconomia.

Brasília, 29 de novembro de 2018.

Marcílio de Brito - Orientador

Professor da Faculdade de Ciência da Informação (UnB)
Doutor em Ciências da Informação e da Documentação

Simone Bastos Vieira – Membro

Professora da Faculdade de Ciência da Informação (UnB)
Doutora em Ciência da Informação

Rita de Cássia do Vale Caribé – Membro

Professora da Faculdade de Ciência da Informação (UnB)
Doutora em Ciência da Informação

AGRADECIMENTOS

A minha família, que me deu apoio e incentivo nas horas difíceis;

Aos meus amigos e namorado, que não me deixaram ser vencida pelo cansaço, que me estimularam durante todo o processo e compreenderam minha ausência pelo tempo dedicado aos estudos;

Aos meus colegas de trabalho, pela paciência nos meus momentos de ansiedade em especial ao Marcelo, por me incentivar a pesquisar e me dedicar ao trabalho que faço buscando melhorias ao processo e crescimento profissional;

Aos colegas de sala e professores pelas trocas enriquecedoras de conhecimento, pelos momentos divertidos e difíceis nessa busca por uma nova formação;

Ao meu orientador, pela paciência e destreza em orientar por essa turbulenta fase de pesquisa e reflexão, pelo apoio e pelas excelentes conversas que despertaram visões adormecidas do conhecimento adquirido ao longo da trajetória acadêmica, o meu muitíssimo obrigada.

RESUMO

Electronic discovery é um processo jurídico para descoberta de evidências em um processo de litígio. Um dos estágios deste processo, o *document review* lida com a recuperação da informação utilizando estratégias de busca distintas dos métodos preconizados pela Ciência da Informação. Esta monografia utiliza-se de um estudo metodológico para identificar e descrever os processos de busca do estágio de *document review* sob a perspectiva das teorias da Documentação que regem a recuperação da informação. Foram contextualizadas na pesquisa, a análise facetada e a teoria do conceito, evidenciando suas correlações e contribuições para a realização do *document review*. Neste processo identificou-se que as teorias da Documentação auxiliam no processo de estruturação da pesquisa, desde a contextualização da pergunta e extração de conceitos até a escolha de táticas de busca. Isto contribui para a concretização de um método que possa suprir a lacuna do modelo de *e-Discovery* no qual a pesquisa é aplicada. Salientou-se também que o profissional da informação, por seus conhecimentos e formação possui competências e habilidades para destacar-se neste nicho de trabalho. Questiona-se também sobre a influência do conhecimento jurídico neste processo.

Palavras-chave: *electronic discovery*. *Document review*. Documentação. Recuperação da informação. Profissional da informação.

ABSTRACT

Electronic Discovery is a legal process for discovering evidence in a litigation process. One of the stages of this process, the document review deals with the retrieval of information using search strategies distinct from the ones advocated from Information Science. This monograph uses a methodology case to identify and describe the processes of searching the stage of document review from the perspective of the theories of documentation governing the retrieval of information. At this study, the faceted analysis and the theory of the concept were contextualized, evidencing their correlations and contributions to the realization of the document review. In this process, it was identified that the theories of the documentation help in the process of structuring the research, from the contextualization of the question and extraction of concepts to the choice of search tactics. This contributes to the realization of a method that can supply the gap of the e-Discovery model in which the study is applied. It was also emphasized that the information professional, for his knowledge and training has skills to stand out in this niche of work. It is also questioned about the influence of legal knowledge in this process.

Keywords: electronic discovery. Document review. Documentation. Information retrieval. Information professional.

LISTA DE ILUSTRAÇÕES

Figura 1 - Modelo referencial de <i>eletronic discovery</i>	21
Figura 2 - Guia de revisão	22
Figura 3 - Mapa de processos	33
Figura 4 - Identificação e preservação.....	34
Figura 5 – Coleção e processamento	35
Figura 6 - Revisão, análise, produção e apresentação	36

LISTA DE TABELAS

Tabela 1 – Táticas de monitoramento	26
Tabela 2 - Táticas de estrutura do arquivo	26
Tabela 3 - Táticas de formulação da pesquisa	27
Tabela 4 - Táticas de termos	27

LISTA DE ABREVIATURAS E SIGLAS

CAST	<i>Center of Applied Science and Technology</i>
CI	Ciência da Informação
EDRM	<i>Eletronic Discovery Reference Model</i>
ESI	<i>Eletronic Stored Information</i>
IR	<i>Information Retrieval</i>
NIST	National Institute of Standards and Technology
SRI	Sistema de Recuperação da Informação
TREC	<i>Text Retrieval Conference</i>

SUMÁRIO

AGRADECIMENTOS	5
RESUMO	6
ABSTRACT	7
LISTA DE ILUSTRAÇÕES	8
LISTA DE TABELAS	9
LISTA DE ABREVIATURAS E SIGLAS	10
SUMÁRIO.....	11
1 INTRODUÇÃO	12
2 CONSTRUINDO O OBJETO DE ESTUDO E O REFERENCIAL TEÓRICO	14
2.1 DEFINIÇÃO DO PROBLEMA E JUSTIFICATIVA.....	14
2.2 OBJETIVOS DA PESQUISA	14
2.2.1 Objetivo geral	14
2.2.2 Objetivos específicos.....	15
2.3 DELIMITAÇÃO DO ESTUDO	15
3 REVISÃO DE LITERATURA.....	16
3.1 <i>A Information Retrieval</i> e a Ciência da Informação.....	16
3.2 <i>Eletronic Discovery</i>	19
3.3 A recuperação da informação	24
4 PROCEDIMENTOS METODOLÓGICOS.....	31
5 APRESENTAÇÃO E ANÁLISE DOS DADOS.....	32
6 DISCUSSÃO.....	38
7 CONCLUSÃO	40
REFERÊNCIAS	42
APÊNDICE A – MAPA DE PROCESSOS	46

1 INTRODUÇÃO

O processo jurídico de busca de documentação em um processo de litígio, também conhecida como *Discovery*, passou a lidar com um aumento considerável de material a ser analisado. Na Ciência da Informação, desde final do século XIX, pesquisadores como Paul Otlet, Bradford e Saracevic, já se preocupavam com a recuperação da informação desse mar de informação. Os estudos sobre recuperação da informação (ou *information retrieval* – IR) sofreram mudanças significativas com a introdução dos sistemas computacionais.

Em um processo de litígios, a procura de evidências se opera sobre massas de informação cada vez maiores. Processos de litígios nos Estados Unidos, a partir de 2006, aplicaram regras para aceitar documentos eletronicamente armazenados (ou *Electronic Stored Information* - ESI), como explicam os princípios da Conferência Sedona (CONFERENCE, 2007). A partir de então, as pesquisas em torno dessa nova forma de *discovery* de um processo legal, o *electronic discovery*, passou a fazer parte das rotinas dos advogados.

Alguns pesquisadores como Saracevic, Bates, Oard, Grossman e Lewis se voltaram para a aplicação da recuperação da informação em o *e-Discovery*, a fim de aprimorar as técnicas de IR aplicadas a este novo processo jurídico.

O advento da Operação Lava-Jato da Polícia Federal apreendeu diversos documentos e aparelhos eletrônicos para processamento das investigações e obtenção de provas. O tratamento dessas informações compreende processos de *e-Discovery*. Este termo aparece no *Google Trends* desde 2004, numa curva crescente apresentando um pico máximo em 2006, o que coincide com o período em que as regras para utilização de ESI em processos jurídicos norte-americanos foram discutidas. No contexto mundial, pouquíssimas são as ocorrências relativas ao Brasil. Alguns artigos em jornais jurídicos, como os artigos de Stopanovski (2015) no *Consultor Jurídico* e de Reinaldo Filho (2006) no *Migalhas*¹, explicam o processo e informam que estas novas práticas constituem uma realidade internacional. Não obstante, essas novas práticas revelam-se pouco conhecidas no Brasil.

As buscas por *e-Discovery*, *electronic Discovery* e recuperação da informação em bases de dados de Ciência da Informação não retornaram resultados significativos sobre essas práticas.

¹ Stopanovski (2015) traz alguns dos conceitos de *e-Discovery* em sua coluna sobre Suporte a litígios no periódico *Consultor Jurídico*, enquanto o artigo do Reinaldo Filho (2006) aborda a ideia de ESI e os processos do *e-Discovery* no processo jurídico brasileiro.

Ao se ampliar as pesquisas para bases de dados jurídicas, as respostas de artigos e trabalhos possibilitaram um levantamento de mais de 40 artigos sobre o tema.

A partir de estudos das conferências e pesquisas desenvolvidas pelo *National Institute of Standards and Technology* (NIST), a TREC² Legal Track, encontrou-se uma série de estudos envolvendo a recuperação da informação e o Direito. Nas atividades de empresas que usam o *eletronic discovery* em seus processos, percebe-se um campo de atuação perfeitamente adaptado aos profissionais da informação, uma vez que se verificou semelhanças entre os processos realizados em estágios do *e-Discovery* e estágios de recuperação da informação, tal como explicados por Kent (1972).

Este trabalho apresenta uma metodologia de documentação aplicável ao estágio de *document review* em *e-Discovery*. A análise das metodologias de busca visa identificar influências e contribuições entre as teorias da Ciência da Informação em recuperação da informação e *e-Discovery*, no estágio de *document review*, adotando como referencial as estratégias de busca (*queries*) e o resultado nas realizações das pesquisas.

Após esta introdução ao tema de estudo, o escopo da pesquisa com suas delimitações e objetivos a serem alcançados é apresentado. O terceiro capítulo propõe uma revisão de literatura sobre os campos teóricos que fundamentam os processos e metodologias estudados. O capítulo seguinte expõe a metodologia de pesquisa e o capítulo cinco trata da coleta e análise dos dados. Por fim, um capítulo com as discussões acerca dos resultados da pesquisa e o último capítulo com considerações finais incluindo possibilidades para pesquisas futuras.

² *Text Retrieval Conference (TREC)* é uma conferência promovida pelo Instituto Nacional de Padrões e Tecnologias (NIST) dos Estados Unidos que, em 2006, desenvolveu uma linha de pesquisa apenas para questões legais sobre *Information retrieval* e os processos do *e-Discovery* denominada *TREC Legal Track*.

2 CONSTRUINDO O OBJETO DE ESTUDO E O REFERENCIAL TEÓRICO

2.1 DEFINIÇÃO DO PROBLEMA E JUSTIFICATIVA

Processos de litígios nos Estados Unidos, a partir de 2006, aplicaram regras para aceitar documentos eletronicamente armazenados (ESI), como explicam os princípios da Conferência Sedona (2007). A partir de então, as pesquisas em torno dessa nova forma de *discovery* de um processo legal, o *eletronic discovery*, passou a fazer parte das rotinas de advogados. Alguns pesquisadores se voltaram para a aplicação da *Information Retrieval (IR)* para o *e-Discovery*, para aprimorar e aplicar as técnicas de IR a este novo processo jurídico. Tanto Mooers (1951 apud SARACEVIC, 2009) quanto Saracevic (2007, 2009) falam que IR se trata de pesquisar e recuperar a informação que é útil. Em um processo de litígio, isso é uma necessidade para a coleta e apresentação de evidências. Portanto, o envolvimento de profissionais da informação nessa área pode colaborar para melhorias nas metodologias aplicadas às práticas de *e-Discovery*.

Do ponto de vista jurídico, há um debate evidente sobre as práticas e metodologias empregadas ao longo de todo o processo de *eletronic discovery*. Contudo, do ponto de vista da recuperação da informação, há apenas discussões sobre as tecnologias de busca empregadas na recuperação da informação em bases de texto integral, mas as discussões aplicadas ao processo de busca não exploram todos os aspectos teóricos da ciência da informação. Estudar, portanto, a contribuição metodológica da atuação de um bibliotecário no processo de *eletronic Discovery* se torna interessante. Diante do acima exposto, esta pesquisa propõe se debruçar sobre o seguinte problema de pesquisa: demonstrar como as teorias e técnicas da ciência da informação podem ser aplicadas ao processo de *eletronic Discovery* com a finalidade de evidenciar um possível campo de atuação para os profissionais da informação, e contribuir na estruturação metodológica do ponto de vista informacional para o processo de revisão de documentos em *eletronic Discovery*.

2.2 OBJETIVOS DA PESQUISA

2.2.1 Objetivo geral

Identificar e descrever como as teorias da documentação auxiliam metodologicamente no processo de *document review* em *e-Discovery*.

2.2.2 Objetivos específicos

Demonstrar as contribuições possíveis da análise facetada e da teoria do conceito para o estágio de *document review* em *e-Discovery*;

Elencar as técnicas de recuperação da informação que podem ser aplicadas ao processo de *e-Discovery*.

Identificar uma metodologia de documentação aplicável ao estágio de *document review* em *e-Discovery* em uma empresa de suporte a litígios.

2.3 DELIMITAÇÃO DO ESTUDO

O escopo deste trabalho abarca um dos estágios do *eletronic discovery*, o *document review*, ou revisão de documentos. Este estágio será tratado do ponto de vista da Ciência da Informação. Apesar de ser um processo do campo jurídico, não serão discutidas questões legais sobre os documentos enquanto prova ou evidência.

O *eletronic discovery* é bastante discutido juridicamente. O modelo referencial de *eletronic discovery* (EDRM) é citado na maioria da literatura sobre o assunto, contudo, não há uma metodologia para a recuperação da informação. O modelo EDRM (2010) traz um guia sobre o estágio de *document review*, porém este guia não discute técnicas ou modelos de como buscar a informação necessária. Portanto, esta pesquisa será em torno do estágio de *document review*, o qual será analisado conjuntamente com técnicas de recuperação da informação. A teoria de análise facetada e a teoria do conceito também serão abordadas a fim de darem embasamento teórico para a construção da metodologia aplicável a este estágio do *e-Discovery*.

Com base no acima exposto, o estudo será realizado com um sistema baseado no modelo booleano no qual o *eletronic discovery* é empregado. Este processo legal tem como objetivo a obtenção e análise de evidências pelas partes envolvidas no processo jurídico. Em processos que envolvem grandes operações policiais, como por exemplo a Lava-Jato, a quantidade de informação a ser processada pelos advogados se apresenta bastante volumosa. Neste ponto, estruturar uma metodologia que colabore para a busca da informação necessária torna o processo mais eficaz.

3 REVISÃO DE LITERATURA

3.1 *A Information Retrieval* e a Ciência da Informação

A documentação surge, enquanto termo de uma ciência, nas primeiras décadas de 1900 com Paul Otlet, o qual apresenta uma fundamentação metodológica, em sua obra “*Traité de documentation*”, de 1934, para este novo campo de pesquisa. Neste tratado, Otlet (1934) propõe métodos e operações para facilitar a organização e acessibilidade às informações, que à época já se demonstravam em ascensão exponencial. Logo no início do tratado, alguns conceitos são elucidados e a ideia de documento é apresentada como termo que remeta a qualquer elemento que indique ou reproduza uma ideia em qualquer formato. Ainda, Otlet (1934) identifica o documento sobre dois aspectos: como resultado do trabalho intelectual do homem e como objeto criado pela civilização e suscetível de agir sobre ela.

A partir daí, identifica-se vários outros estudiosos que se debruçaram sobre essa nova ciência, como explanam Alvares e Araújo Júnior (2010), a qual se espalhou não só na Europa, mas nos Estados Unidos também. Buckland (1996) faz um apanhado histórico acerca do desenvolvimento da Documentação nos Estados Unidos até se tornar Ciência da Informação. Ele aborda as diferenças entre a Escola de Chicago e a Documentação como era tratada pelos europeus.

Alguns autores, expoentes da Documentação, apresentam ideias fundamentais para a área na época, as quais propiciaram especificações à futura Ciência da Informação. É o caso da organização e tratamento da informação que são apresentados pelos autores Paul Otlet (1934), Suzanne Briet (1951) e Samuel Bradford (1961).

Bradford (1961, p. 68) define documento, de forma mais conservadora, como sendo o registro da atividade intelectual. Em oposição, Briet (1951) que o precedeu, perpetua a ideia mais abstrata de documento, introduzida por Otlet, dizendo: “todo indício concreto ou simbólico conservado ou registrado a fim de representar, reconstituir ou provar um fenômeno físico ou intelectual.” (BRIET, 1951, p. 7).

Naquele momento histórico, portanto, Biblioteconomia e Documentação pareciam se preocupar com o mesmo objeto. Porém, Bradford (1961) define um objetivo específico para a Documentação que soa diferente do que se buscava na Biblioteconomia. Apesar de ambos tratarem informação, à época se enfatizava o livro e documentos da massa cinzenta, para o autor, a Documentação se ocupa de “tornar disponível a informação original” (BRADFORD,

1961, p. 69). Ele destaca, ainda, que ela se ocupa de entregar, de fato, a informação ao usuário, sem necessariamente se tratar de livros ou documentos acadêmicos.

Os objetivos da Documentação para Bradford (1961), assim como para Briet (1951) e Otlet (1934) são, em linhas gerais, selecionar, classificar, organizar e preservar para tornar a informação acessível. O que diferencia Bradford dos outros dois é sua intenção de mostrar que a Documentação, além desses objetivos, tem outro mais específico, o da pesquisa e entrega do documento.

O período histórico do pós-segunda guerra mundial, presenciou grande evolução tecnológica em muitos campos do conhecimento. Por se tratar de uma ciência interdisciplinar, como evidenciam alguns autores (BUCKLAND, 1996; LE COADIC, 2004; SARACEVIC, 2009), a Ciência da Informação surgiu e se transformou neste período com a colaboração de outras ciências. É possível identificar uma abundância de termos para áreas de estudo que hoje são englobadas pela Ciência da Informação. O termo *Information Retrieval*, por exemplo, cunhado pelo físico e matemático Calvin Mooers, em 1950, em seu estudo sobre a recuperação da informação, foi por certo tempo o termo utilizado para Ciência da Informação (ALVARES; ARAÚJO, 2010; SARACEVIC, 2009).

Bates (1999 apud SARACEVIC, 2009, p. 3) elabora três questões sobre os objetos de pesquisa dessa nova área, a Ciência da Informação, que podem ser divididas nas áreas física (referente a regras e leis), sociais (referente à relação homem-informação) e de design (referente ao acesso, rapidez e eficiência da informação). Em continuidade a esta ideia, o estudo dos autores White e McCain (1998 apud SARACEVIC, 2009, p. 4) divide a Ciência da Informação em dois grandes campos de estudos: a área de recuperação da informação e a área de comportamento informacional humano e estudo de métricas.

As três áreas de Bates (1999 apud SARACEVIC, 2009) englobam todos os campos e práticas que Saracevic (2009) e Le Coadic (2004) citam no que se refere ao campo de estudo da Ciência da Informação. Fica evidente, a contextualização da *Information Retrieval*, ou recuperação da informação, por esses autores como um campo de estudo dentro dessa nova ciência.

A especificidade abordada por Bradford (1961) para a Documentação é observada na definição de Mooers de recuperação da informação quando este diz:

[...] encontro ou processo de descoberta em relação à informação armazenada... útil para [um usuário]. Recuperação de informação abrange o aspecto intelectual da descrição da informação e sua especificação para pesquisa, e também quaisquer sistemas, técnicas ou máquinas que são empregadas para realizar a operação. (MOOERS, 1951 apud SARACEVIC, 2009, p. 5, tradução nossa).

Alguns autores acrescentam outra questão importante para o conceito de recuperação da informação, o qual apesar de não estar identificado nesta fala de Mooers, também é tratado em sua pesquisa. Trata-se da relevância da informação. Saracevic (2007, 2009, 2012) tem um amplo estudo sobre esta questão. Para o autor, recuperação da informação é a recuperação de objetos informacionais úteis e relevantes. Le Coadic (2004) também concorda que relevância é conceito fundamental para a recuperação da informação ao citar as experiências de Cranfield que definiu “[...] a relevância é uma relação entre um documento e uma questão. O êxito da recuperação é medido pela relação entre documentos relevantes e não-relevantes, recuperados ou não recuperados” (LE COADIC, 2004, p. 80).

A primeira exploração do conceito de relevância aconteceu em 1958, na "*International Conference for Scientific Information*" (ICSI), como relata Figueiredo (1977). Alguns dos pontos definidos sobre o conceito na época são:

- relevância não é, exclusivamente, uma propriedade de documentos;
- relevância não é uma propriedade dicotômica;
- existe uma "relevância para o usuário" que deve ser julgada. (FIGUEIREDO, 1977, p. 2).

Posteriormente, Goffman apresentou uma teoria sobre relevância em que esta é definida como uma medida da relação entre o documento encontrado e a pergunta que o originou (FIGUEIREDO, 1977; SARACEVIC, 2007).

O estudo de Saracevic (2007) sobre relevância revisa o conceito ao longo dos anos e coloca certos questionamentos pontuais sobre o termo e a sua importância no contexto da recuperação da informação.

A noção fundamental usada na descrição bibliográfica e em todos os tipos de classificações, ontologias ou categorizações, incluindo aquelas usadas em bancos de dados contemporâneos, é a do *aboutness*. A noção fundamental usada em IR é *relevância*. Não é sobre qualquer tipo de informação, mas sobre informação *relevante*. Fundamentalmente, a descrição bibliográfica e a classificação concentram-se na descrição e categorização de objetos de informação; IR também é sobre isso, *mas, e isso é um “mas” muito importante*, além disso, IR também trata de pesquisar, e pesquisar é sobre relevância. (SARACEVIC, 2007, p. 5, tradução nossa, grifo do autor).

Em estudo posterior, Saracevic (2009, p. 5) retoma a diferença entre base de dados de recuperação da informação, os quais são discutidos em termos de dados estruturados e não estruturados, sendo que o primeiro trata de semelhança e o segundo, de relevância.

Na visão jurídica, no *e-Discovery*, a relevância ou *relevancy* significa responsivo, aquilo que respondeu à pergunta, como definem Oard et al (2010, p. 15). Esses autores apresentam os diferentes conceitos para relevância, o qual para a IR pode significar utilidade ou a relação entre o documento e um tópico. Eles informam, ainda, que um estudo de Bales e Wang (2006 apud OARD et al, 2010, p. 15) identificou 14 critérios para a definição de relevância em *e-Discovery*.

3.2 *Eletronic Discovery*

Eletronic discovery ou *e-discovery*, como comumente é utilizado, se refere, em linhas gerais, ao processo de *Discovery* aplicado à informação eletronicamente armazenada (ESI). O glossário de revisão por tecnologia assistida de Grossman e Cormack (2013) define *eletronic Discovery* como sendo “o processo de identificar, preservar, coletar, processar, pesquisar, revisar e produzir informação eletronicamente armazenada que possa ser relevante em um caso civil, criminal ou regulatório” (GROSSMAN; CORMACK, 2013, p. 15).

Consoante a esta definição, a Associação para Gestão Inteligente da Informação (AIIM, 2018) acrescenta que o *e-Discovery* é um processo de litigação que percorre do momento do processo ao momento de apresentação da prova digital em um tribunal. Outros autores confirmam estas ideias, portanto, em suma, o processo de *eletronic Discovery* faz parte da litigação de um processo legal, o qual perpassa diversos estágios relacionados à coleta, tratamento e difusão de informações eletronicamente armazenadas que servirão como evidência e até mesmo prova legal (CHISHOLM, 2010; FLOOD, 2014; LAWTON; STACEY; DODD, 2014; OARD et al., 2010; SATHIYANARAYANAN; TURKAY, 2016; WANG; SOERGEL, 2010).

A informação eletronicamente armazenada (ESI) de que trata o *e-Discovery*, de acordo com os princípios da Conferência de Sedona de 2007, diz respeito a “[...] e-mails, páginas web, arquivos word, arquivos de áudio e vídeo, imagens, base de dados, e qualquer item virtualmente armazenado em um computador– incluindo mas não limitado a servidores, desktops, laptops, celulares, hard drives, flash drives, PDAs e MP3 players” (CONFERENCE, 2007). O glossário EDRM (2016) complementa ainda esta definição

dizendo que a ESI diz respeito à informação armazenada eletronicamente independente da mídia em que se encontra e independente do formato original na qual foi produzida.

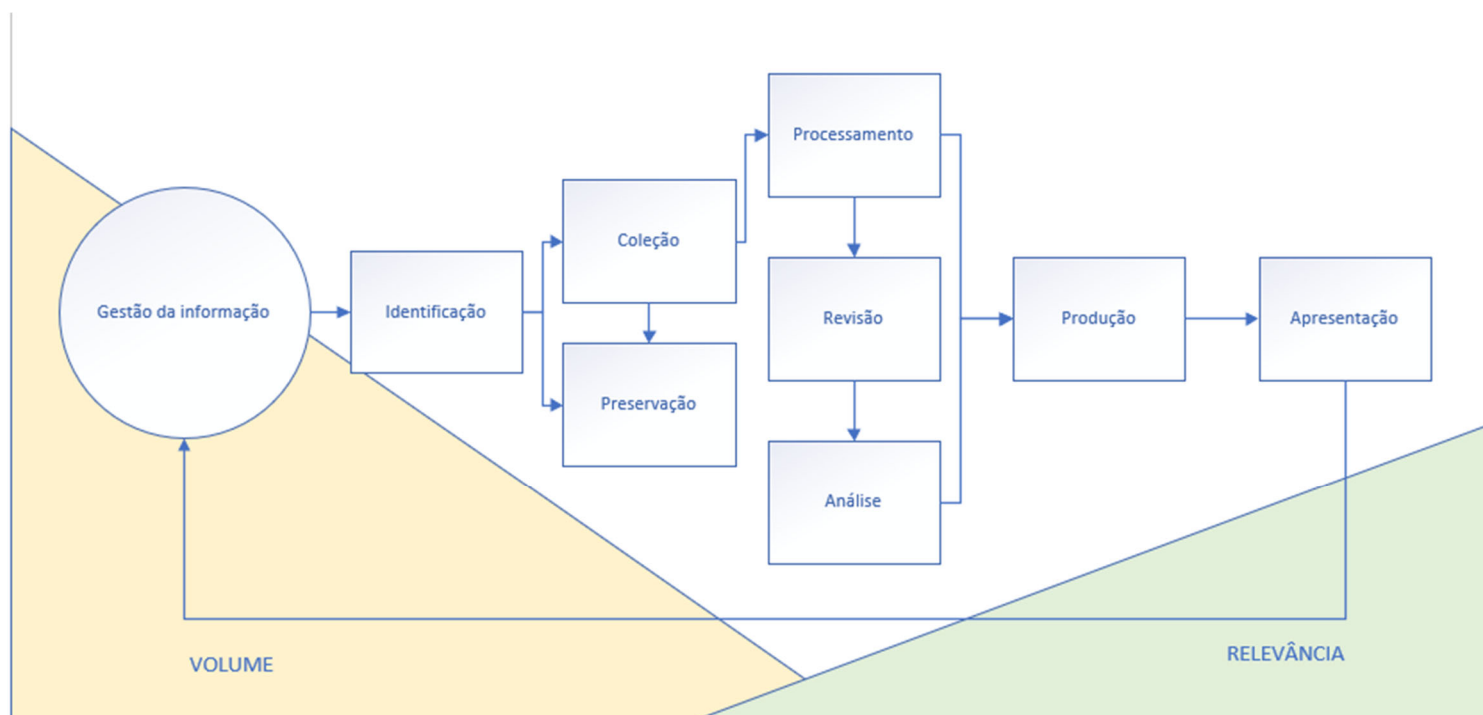
Há um destaque significativo de definições que é apontado por Roitblat (2013), doutor e consultor em soluções de *e-Discovery*. Logo no início de seu artigo, ele diferencia o entendimento de *e-Discovery* para a ciência da informação e para o Direito sendo que:

Do ponto de vista da ciência da informação, o *e-Discovery* trata da separação dos documentos responsivos dos não responsivos. Para os advogados, é claro, o *e-Discovery* envolve muito mais, mas essa perspectiva de informações é um precursor natural da análise jurídica. Essa separação, às vezes chamada de primeira revisão ou mesmo avaliação inicial de caso (mais uma vez focada na perspectiva da ciência da informação), tem sido tradicionalmente feita com a revisão de cada documento - também chamado de revisão linear. (ROITBLAT, 2013, p. 1, tradução nossa)

A definição de *e-Discovery* do Centro de Ciência Aplicada e Tecnologia (CAST) do Reino Unido está de acordo com as definições dos autores acima apresentados e ainda informa essa área como campo de pesquisa do Direito com interface a Ciência da Computação (LAWTON; STACEY; DODD, 2014). Observa-se que este processo engloba, assim, diferentes áreas competentes a questões específicas. Há ainda outro ponto sobre esta prática em que “Discovery é o processo no qual uma parte (a parte que produz) do caso legal disponibiliza para a outra (a parte requerente) os materiais em suas posses que são pertinentes para o caso” (OARD; WEBBER, 2013, p. 7, tradução nossa). Isto evidencia que o processo possui diversos estágios e passa da defesa à acusação e ao juiz em momentos distintos.

Em vista disso, a comunidade EDRM³ (*Electronic Discovery Reference Model*), em 2005, formalizou este processo em um diagrama de fluxo o qual apresenta sua eficiência nos trabalhos de autores da área (BARON, 2011; CHISHOLM, 2010; EDRM, 2018; FLOOD, 2014; LAWTON; STACEY; DODD, 2014; OARD et al., 2010; OARD; WEBBER, 2013). Este diagrama é composto por seis estágios: gestão da informação, identificação; preservação e coleção; processamento, revisão e análise; produção e apresentação (EDRM, 2018), que encontra-se representado na figura 1.

³ A comunidade EDRM, conforme apresentação no site deles EDRM (2018), é composta por profissionais legais e de *e-Discovery* que desenvolvem recursos e guias para melhorar o desenvolvimento e o uso das tecnologias para *Discovery*.

Figura 1 - Modelo referencial de *eletronic discovery*

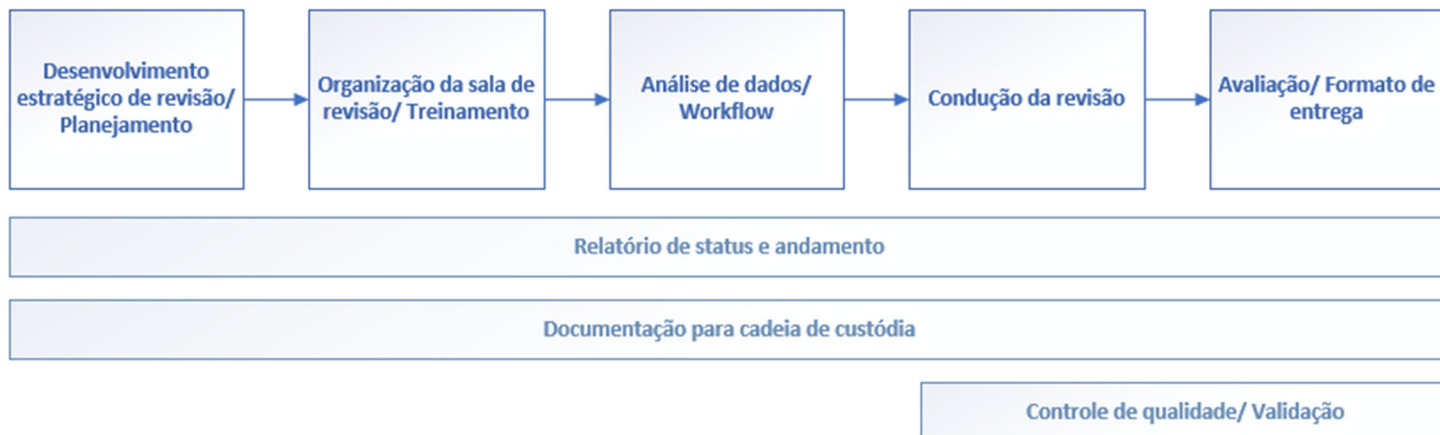
Fonte: adaptado de EDRM (2018, tradução nossa).

O primeiro estágio, gestão da informação, é composto por outro diagrama; o modelo de referência de governança da informação. Este modelo foi criado para demonstrar a necessidade de discutir melhor a gestão da informação, principalmente no que tange à colaboração dos *stakeholders*, contudo não será trabalhado nesta pesquisa.

O segundo estágio tem como objetivo desenvolver e executar um plano de identificação e validação de fontes potenciais de relevância de ESI incluindo pessoas e sistemas. Os próximos dois estágios, preservação e coleção, dizem respeito a coletar a informação que será utilizada e preservá-la de forma que ela possa ser utilizada legalmente.

Os três estágios posteriores são estágios de ação: processamento, revisão e análise. O processamento é a identificação do ESI coletado de forma que será normalizado e estruturado para que a revisão possa ser feita. A revisão é a identificação dos documentos que estão na coleção que poderão ser utilizados ou que devem ser retidos. Este estágio é composto por nove passos, os quais se repetem até que o resultado obtido seja o desejado. Os passos que o EDRM (2010) identifica são: desenvolvimento de um plano estratégico de revisão, configuração de uma sala para revisão, análise de dados, condução da revisão, avaliação, relatório de progresso, controle de qualidade/validação, recomendação e riscos. Estes passos são identificados no diagrama representado na figura 2 a seguir.

Figura 2 - Guia de revisão



Fonte: adaptado de EDRM (2010, tradução nossa).

A análise é um estágio concomitante a vários outros do modelo de *e-discovery*, e consiste em analisar as circunstâncias, fatos e potenciais evidências a serem utilizadas na coleção de documentos identificados. A produção é a preparação do documento encontrado em um formato que possa legalmente ser utilizado. E o estágio final de preservação diz respeito ao isolamento da informação potencialmente relevante de forma que seja eficiente, auditável e que mitigue riscos.

A condução da revisão de fato, ou seja, a busca pelos documentos que respondem ao escopo da pesquisa identificada pela equipe legal não possui subpassos ou indicações de como devem ser realizadas. O EDRM (2010) indica apenas que é preciso fazer um teste com a estratégia de pesquisa escolhida para verificar se ela está de acordo com as questões a serem respondidas. Contudo, não há menção ou indicação de como formular tal estratégia de pesquisa dentro da base de dados. Nesse guia não há menção a uma metodologia de pesquisa para recuperação da informação buscada.

Outra instituição que promove a pesquisa e educação na área é a Conferência Sedona (*The Sedona conference*), fundada em 1997 por Richard G. Braham. Ela promove a educação e discussão sobre o tema a fim de provocar mudanças nas práticas legais, fato evidenciado pelo “*Civil Discovery Standards*” publicado pela Associação Americana de Advogados que se baseou nos princípios da Sedona Conference (2007). Esses princípios tratam de diversos assuntos dos estágios do processo de *e-Discovery*. Na declaração desses princípios, são enumeradas seis diferenças entre o documento físico e o ESI que evidenciam a necessidade de estudos para melhorar as práticas que envolvem esse tipo de documento.

Os desafios do *e-Discovery* levaram outras organizações a financiarem pesquisas. É o caso da *Text Retrieval Conference (TREC)* promovida pelo Instituto Nacional de Padrões e Tecnologias (NIST) dos Estados Unidos. Em 2006, foi desenvolvida uma linha de pesquisa apenas para questões legais sobre recuperação da informação e os processos do *e-Discovery* denominada *TREC Legal Track*.

O que motivou esta pesquisa foi exatamente um dos artigos publicados no *Legal Track* de 2010, sobre o julgamento de relevância no processo de *e-Discovery*. Este estudo de caso, liderado por Wang e Soergel (2010), pesquisou a diferença no julgamento de relevância no estágio de revisão para pesquisadores com formação em Ciência da Informação e em Direito. Este estudo suscitou a pergunta sobre a relação da classificação dos resultados de pesquisa como relevantes e a estruturação da consulta.

O estudo de Oard et al (2010) traz alguns conceitos importantes e continua a discussão sobre o julgamento de relevância do ponto de vista de quem tem o conhecimento do Direito e de quem tem o conhecimento da recuperação da informação. Neste artigo, os autores discutem diversos entendimentos sobre relevância, precisão, efetividade, recuperação da informação e outras ideias que estão diretamente relacionadas às práticas de recuperação da informação aplicadas ao *e-discovery*. Importante frisar que Oard et al (2010) abordam um ponto intrigante sobre a diferença do entendimento de palavra-chave, pesquisa por palavras-chave e pesquisa por conceito entre pesquisadores da área de recuperação da informação e advogados. Essas diferenças conceituais, ao longo do processo do *e-discovery*, podem gerar diferenças no julgamento de relevância dos resultados de uma pesquisa.

As bases de dados do *TREC Legal Track* possibilitaram aos pesquisadores de IR realizar pesquisas de avaliação de forma a compartilhar a mesma base de dados. Muitos estudos sobre a avaliação de relevância na recuperação da informação têm sido desenvolvidos e autores como Wang e Soergel (2010) e Oard et al (2010) têm questionado habilidades de pesquisa dos pesquisadores de recuperação da informação em oposição a advogados. Pois, como explicam Oard et al (2010, p. 5), quem de fato realiza as pesquisas são advogados e suas equipes de técnicos de informação, apesar de não necessariamente dominarem conceitos, ferramentas e modelos envolvidos no processo de recuperação da informação. A necessidade desse domínio conceitual para embasar as pesquisas é evidenciada ao declararem: “A necessidade de uma

compreensão mais precisa da utilidade de tecnologias específicas de IR em *e-Discovery* é clara, como é a necessidade de maior atenção ao processo global no qual elas são usadas” (OARD et al, 2010, p. 13, tradução nossa). Ao final deste artigo, Oard et al (2010, p. 33) informam que um certificado de padronização para implementação do *e-Discovery* seria uma ideia para uniformizar as práticas de IR para *e-Discovery*.

Consoante a esta ideia, Baron (2011, p. 28-29) apresenta em sua pesquisa alguns exemplos e declarações de juízes americanos que reforçam a necessidade de uma busca conceitual nas coleções de *e-Discovery*. As pesquisas apenas por palavras-chave podem ser falhas e esconder falsos negativos, aqueles documentos que são relevantes, mas não foram recuperados (OARD; WEBBER, 2013). Percebe-se, portanto, que há uma necessidade de se explorar métodos para a recuperação das bases de *e-Discovery*.

3.3 A recuperação da informação

Como explica Robredo (2005), a recuperação da informação é objetivo final em um ciclo documentário. Contudo, para conseguir uma boa recuperação, é preciso tratar a informação inicialmente. Assim, em uma unidade de informação, o documento que será disponibilizado passa pelo processamento técnico, ou seja, ele é descrito, analisado, indexado e classificado. A indexação aqui é assistida por uma linguagem documentária (LD), que se materializa sob a forma de um tesouro, um vocabulário controlado, ou qualquer outra ferramenta produzida para aquele fim. A análise do documento extrai conceitos que são representados por termos da linguagem controlada, assim a indexação tende a ser fidedigna ao conteúdo, como explicam Robredo (2005) e Lancaster (2004), porque o processo ainda sofre a influência do profissional indexador.

A recuperação do documento será possível por meio de busca no banco de dados, ou em um sistema de recuperação de informação (SRI), onde essas informações encontram-se armazenadas. Conforme Souza (2006 apud ARAUJO, 2012), as funções de um SRI são três: 1) representar as informações contidas nos documentos; 2) armazenar e gerir os documentos e suas representações; e 3) recuperar as informações e os próprios documentos contidos no sistema.

Chowdhury (2010) traz uma diferença entre os sistemas de recuperação de informação e os sistemas modernos de recuperação da informação. Para o autor, o SRI originalmente significa recuperação textual, por se tratar de documentos textuais; enquanto os sistemas modernos de

recuperação da informação lidam com informação multimídia (texto, áudio, imagem e vídeo), o que requer novas técnicas para recuperar a informação.

Kent (1972) define as operações unitárias como as etapas a serem empregadas em um sistema para a recuperação da informação. Estas operações são elencadas pelo autor em sete etapas: análise, controle de vocabulário e rubrica de assunto, registro dos resultados da análise em um instrumento, armazenagem de registros ou documentos-fonte, análise de questões e desenvolvimento de uma estratégia de pesquisa, condução da pesquisa e exposição dos resultados da pesquisa.

As estratégias de pesquisa elencadas por Kent (1972) somam oito, a saber: pesquisa de um único aspecto, estratégia de adição lógica, estratégia de produtos lógicos, estratégia dos produtos lógicos das adições lógicas, estratégia de subtração lógica, estratégia de sequência, estratégia de pesquisas entre barreiras, e estratégia maior que e menor que. Uma pessoa ao buscar informação no banco de dados precisa formular uma pergunta, passá-la para a linguagem utilizada naquele banco de dados e enfim realizar a busca com uma das estratégias de pesquisa (KENT, 1972; ROBREDO, 2005). Bates (1979) entende estratégia de pesquisa como algo um pouco mais abrangente, para a autora a estratégia é um plano de pesquisa. A autora explica que o termo “estratégia de pesquisa” era mais comumente utilizado no campo da computação, enquanto os bibliotecários se referem a este processo como “processo de referência” em que a estratégia de pesquisa é um dos passos a ser realizado.

Bates (1979) define ainda 29 táticas de pesquisa, que são movimentos a serem desempenhados no processo de pesquisa a fim de se chegar à resposta final. Nas tabelas a seguir é possível visualizar as táticas divididas em quatro grupos: táticas de monitoramento, táticas de estrutura do documento, táticas para formulação da pesquisa e táticas sobre os termos de pesquisa.

As táticas de monitoramento, apresentados na tabela 1, servem para auxiliar o pesquisador a manter a pesquisa no contexto e no foco da pergunta. São táticas que possibilitam a revisão e estruturação das estratégias de busca com o intuito de revisar os resultados da primeira consulta realizada.

Tabela 1 – Táticas de monitoramento

ID. NOME	DESCRIÇÃO
<i>M1. VERIFICAR</i>	Revisar o pedido original e compará-lo ao tópico de pesquisa atual para ver se é o mesmo.
<i>M2. PESAR</i>	Fazer uma avaliação de custo-benefício, em um ou mais pontos da pesquisa, de ações atuais ou previstas.
<i>M3. PADRONIZAR</i>	Tornar-se consciente de um padrão de pesquisa, examine-o e reproduza-o, se não for maximamente eficiente ou se estiver desatualizado.
<i>M4. CORRIGIR</i>	Observar e corrigir erros ortográficos e factuais no tópico de pesquisa.
<i>M5. RECORDAR</i>	Acompanhar as trilhas que se seguiram e as trilhas desejáveis não seguidas ou não concluídas.

Fonte: adaptado de BATES (1979, tradução nossa).

As táticas de estrutura do arquivo, listadas na tabela 2, devem ser empregadas para se pensar no arquivo de origem. São formas de estruturar as consultas pensando na estrutura do arquivo que originou aquela base de dados.

Tabela 2 - Táticas de estrutura do arquivo

ID. NOME	DESCRIÇÃO
<i>F1. BIBLIOGRAFIA</i>	Procurar uma bibliografia já preparada, antes de preparar uma; mas geralmente, verificar se o trabalho de pesquisa já foi feito por outra pessoa.
<i>F2. SELECIONAR</i>	Dividir consultas de pesquisa complexas em subproblemas e trabalhar em um problema de cada vez.
<i>F3. AVALIAR</i>	Revisar, em cada ponto de decisão da pesquisa, as opções disponíveis antes de selecionar.
<i>F4. CORTAR</i>	Ao selecionar entre várias maneiras de pesquisar uma determinada consulta, para escolher a opção que corta, elimina a maior parte do domínio de pesquisa de uma só vez.
<i>F5. ESTENDER</i>	Usar uma fonte diferente da sua finalidade.
<i>F6. AUXILIARES</i>	Projetar uma rota indireta auxiliar através dos arquivos de informações e recursos para alcançar as informações desejadas.
<i>F7. CLEAVE</i>	Empregar pesquisa binária na localização de um item em um arquivo ordenado.

Fonte: adaptado de BATES (1979, tradução nossa).

As táticas de formulação da pesquisa, elencadas na tabela 3, são formas de variar as estratégias de busca. Elas devem ser empregadas ao longo da pesquisa para auxiliar na exaustão das estratégias de busca.

Tabela 3 - Táticas de formulação da pesquisa

ID. NOME	DESCRIÇÃO
<i>S1. ESPECIFICAR</i>	Pesquisar em termos que são tão específicos quanto as informações desejadas.
<i>S2. EXAURIR</i>	Incluir a maioria ou todos os elementos da consulta na formulação de pesquisa inicial; adicionar um ou mais dos elementos de consulta a uma formulação de pesquisa já preparada.
<i>S3. REDUZIR</i>	Minimizar o número de elementos da consulta na formulação de pesquisa inicial; para subtrair um ou mais dos elementos de consulta de uma formulação de pesquisa já preparada.
<i>S4. PARALELOS</i>	Tornar a formulação de pesquisa ampla (ou mais ampla) incluindo sinônimos ou termos conceitualmente paralelos.
<i>S5. PRECISÃO</i>	Tornar a formulação de pesquisa precisa minimizando (ou reduzindo) o número de termos paralelos, mantendo os termos mais perfeitamente descritivos.
<i>S6. BLOQUEAR</i>	Rejeitar, na formulação de busca, itens contendo ou indexados por determinado termo, mesmo que isso signifique perder algumas seções de relevância do documento.

Fonte: adaptado de BATES (1979, tradução nossa).

Por fim, as táticas de termos, apresentadas na tabela 4, devem ser empregadas também com a finalidade de variar as consultas. Elas são táticas a serem usadas nos termos definidos para a construção das consultas e podem ser combinadas com as outras táticas.

Tabela 4 - Táticas de termos

ID. NOME	DESCRIÇÃO
<i>T1. SUPERORDENADO</i>	Subir hierarquicamente para um termo mais amplo (superordenado).
<i>T2. SUBORDENADO</i>	Descer hierarquicamente para um termo mais específico (subordinado).
<i>T3. RELACIONADO</i>	Mover lateralmente na hierarquia para um termo coordenado.
<i>T4. VIZINHO</i>	Procurar termos de pesquisa adicionais, observando os termos vizinhos, seja em ordem alfabética imediata, por similaridade de assunto ou de outra forma.
<i>T5. TRAÇOS</i>	Examinar informações já encontradas na pesquisa, a fim de encontrar termos adicionais a serem usados no aprofundamento da pesquisa.

<i>ID. NOME</i>	<i>DESCRIÇÃO</i>
<i>T6. VARIAR</i>	Alterar ou substituir seus termos de pesquisa de várias maneiras.
<i>T7. MORFEMAS</i>	Tentar os afixos alternativos, sejam prefixos, sufixos ou infixos.
<i>T8. REARRANJAR</i>	Inverter ou reorganizar as palavras em termos de pesquisa em qualquer ou todos os pedidos razoáveis
<i>T9. CONTRÁRIOS</i>	Procurar o termo logicamente oposto àquele que descreve a informação desejada.
<i>T10. GRAFIA</i>	Pesquisar uma grafia diferente.
<i>T11. ESPAÇAMENTO</i>	Tentar variantes de espaçamento.

Fonte: adaptado de BATES (1979, tradução nossa).

Um sistema de recuperação de informação pode ser configurado com base em diferentes modelos. Le Coadic (2004) identifica quatro modelos diferentes: o modelo booleano, o vetorial, o probabilístico e o linguístico. O modelo booleano, como informa Le Coadic (2004) é o modelo que primeiro teve aplicação industrial nos bancos de informação. O autor explica também que o modelo booleano identifica dois tipos de relação de dependência com os operadores booleanos, no qual o E (produto lógico) une os componentes de uma expressão e o OU (adição lógica) une termos (LE COADIC, 2004). As estratégias de pesquisa de Kent (1974) também abordam essas relações de dependências dos operadores booleanos.

Rasmussen (2011, p. 97, tradução nossa) afirma que “cada modelo oferece uma forma de pensar sobre a recuperação da informação”. Ainda, o autor explica que o modelo booleano não é sabiamente utilizado. É preciso entender a lógica booleana para que as consultas sejam bem estruturadas e os operadores sejam corretamente empregados considerando a lógica de relacionamento dos termos. Portanto, em um sistema que emprega o modelo booleano é preciso pensar na lógica booleana para construir as consultas e também utilizar a melhor estratégia de pesquisa.

Além de pensar no modelo para estruturar uma consulta, é necessário pensar na forma como aquele conteúdo é indexado. Bases bibliográficas e arquivistas utilizam a indexação por aquisição, como explica Lancaster (2004), que significa a indexação por um agente humano. Outra forma de indexar, que começou a crescer a partir de 1960, foi a indexação por extração, que é a indexação automática por computador na qual a busca é realizada em texto completo. Dessa forma, permite construir a consulta com base no texto integral.

Lancaster (2004) discorre sobre alguns estudos que foram realizados à época para saber a eficácia desse tipo de indexação. Neste ponto, as TRECs tiveram um papel de desenvolvimento na área, pois como explica Lancaster (2004), as conferências permitiram a diversos grupos de pesquisa realizar testes em seus sistemas de recuperação em condições controladas, pois se utilizavam as mesmas bases de dados, consultas e avaliações.

O campo jurídico aproveitou a TREC, como explicado no capítulo anterior, para testar a recuperação de documentos no que tange a suas necessidades. Os sistemas de recuperação jurídicas nos Estado Unidos, o LEXIS e o WESTLAW, foram bases para diversas pesquisas. Blair e Maron fizeram, ainda, uma outra comparação com esses sistemas de recuperação jurídica, sobre um sistema de recuperação para litígios, o STAIRS/TLS (DABNEY, 1986; LANCASTER, 2004). Os autores explicam que a dificuldade de realizar queries de pesquisas em sistemas como os citados anteriormente se concentra em três quesitos: sinonímia, ambiguidade e expressões complexas. Guinchat e Menou (1994) concordam quanto a essas dificuldades de se pesquisar em uma base de linguagem natural e citam, ainda, a dificuldade de identificar os termos ou expressões utilizados pelos autores.

Dabney (1986) discorre sobre as pesquisas de Blair e Maron e sobre as dificuldades de se pesquisar em um sistema de recuperação em suporte a litígios, por conter documentos que, além de estarem em linguagem livre, possuem formalidades distintas na escrita. Em resumo, portanto:

[...] se a linguagem natural for usada como chave para recuperação da informação, o processo de análise e de geração de conteúdos deve se pautar em método consistente. Para se obter o sucesso de um método para captar os conteúdos e representá-los em linguagem natural torna-se necessário que se recorra a procedimentos diferenciados (COSTA, 2010, p. 171).⁴

Em um sistema com texto integral e indexação automática, a estratégia de pesquisa possibilita uma variedade de busca a partir da raiz das palavras como afirma Foskett (1973). Para tanto, é interessante pensar nos componentes semânticos dos conceitos como explica Soergel (2003), pois, um conceito pode ser expresso pela combinação de seus componentes semânticos. O autor afirma que o conceito é formado pela raiz semântica acrescido de um modificador. Ele

⁴ Faz-se necessário distinguir linguagem natural de linguagem livre. A autora Costa (2010) utiliza o termo linguagem natural como sinônimo de discurso comum. Esta pesquisa trabalha com texto integral ou linguagem livre, pois entende linguagem natural como uma linguagem composta por uma gramática, sintaxe e semântica. O processamento de linguagem natural é uma área da inteligência artificial na qual a máquina compreende a estrutura da linguagem. No caso deste trabalho, tem-se, portanto, apenas a leitura dos termos pela máquina, mas não a compreensão do que está escrito.

assegura, ainda, que “A análise facetada identifica os componentes semânticos de um conceito e organiza os elementos resultantes em facetas” (SOERGEL, 2003, p. 1), o que facilita a pesquisa por permitir a busca com conceitos amplos ou combinações específicas.

Lancaster (2004) também versa sobre a estruturação da query com base em fragmentos de palavras, o que o autor chama de truncamento. Ele fala sobre a formação de classes de palavras e seus fragmentos estruturados em uma estratégia de busca como um processo de “pós-controle”. Contrariamente a esta ideia, o processo de análise do documento para a indexação leva em conta duas ideias fundamentais para a compreensão da informação: a teoria do conceito e a análise facetada. Para Dahlberg (1978), um conceito é composto de enunciados verdadeiros que se articulam a respeito de determinado objeto. Estes enunciados são os atributos daquele conceito. Sendo assim, quando se tem uma base indexada com texto integral, a busca pelo conceito pode ser feita por meio de seus enunciados verdadeiros também.

A análise facetada de Ranganathan (1962), também versa sobre aspectos ou características de um objeto, que o autor chama de faceta. A análise facetada, como explica Ranganathan (1962), possibilita dividir um assunto em grupos de mesmas características. Para o autor, existem cinco categorias fundamentais a qualquer assunto que são aplicáveis dentro da classificação: personalidade, matéria, energia, espaço e tempo. Ranganathan (1962) explica que não necessariamente um objeto precisa ter as cinco categorias, mas ele terá sua faceta básica. Pode-se explorar apenas algumas dessas facetas.

Quando se divide um objeto em suas facetas, obtém-se o que o autor chama de isolado ou foco. A faceta apresenta a manifestação geral de uma categoria fundamental, enquanto o foco apresenta a manifestação particular. Esta análise permite também que os conceitos tenham dois tipos de relacionamentos, dentro de uma mesma faceta e entre facetas, como explica Broughton (2006), o relacionamento semântico e sintático respectivamente.

4 PROCEDIMENTOS METODOLÓGICOS

Esta pesquisa possui uma abordagem qualitativa pois pretende entender e descrever um processo do ponto de vista da Ciência da Informação. Como explicam Prodanov e Freitas (2013), a abordagem qualitativa é empregada em pesquisas com foco no processo e não na discussão estatística dos dados.

De acordo com Gil (2012), pesquisas que tem por objetivo descrever e analisar um fenômeno são pesquisas descritivas. O objetivo deste estudo é analisar um fenômeno e identificar teorias que possam estruturar um dos passos realizados em *eletronic discovery*, portanto, em vista da definição de Gil (2012) tem-se um nível de pesquisa descritivo.

Para analisar o fenômeno, foi necessário desenvolver, inicialmente, uma pesquisa bibliográfica afim de se entender as pesquisas já realizadas acerca do assunto e de se estabelecer pontos de análises que contribuam para o desenvolvimento do mesmo do ponto de vista da Ciência da Informação, em paralelo com o campo jurídico.

Para se atingir os objetivos propostos pela pesquisa, optou-se pelo procedimento experimental que torna a verificação de variáveis uma necessidade para se averiguar a contribuição dessa metodologia ao processo escolhido (GIL, 2012; PRODANOV; FREITAS, 2013).

As variáveis independentes são as teorias a serem aplicadas no processo de *eletronic Discovery* e que foram descritas na revisão de literatura. As variáveis dependentes, portanto, são as contribuições da aplicação dessas variáveis para a realização da pesquisa e contribuição metodológica ao processo de *e-Discovery*.

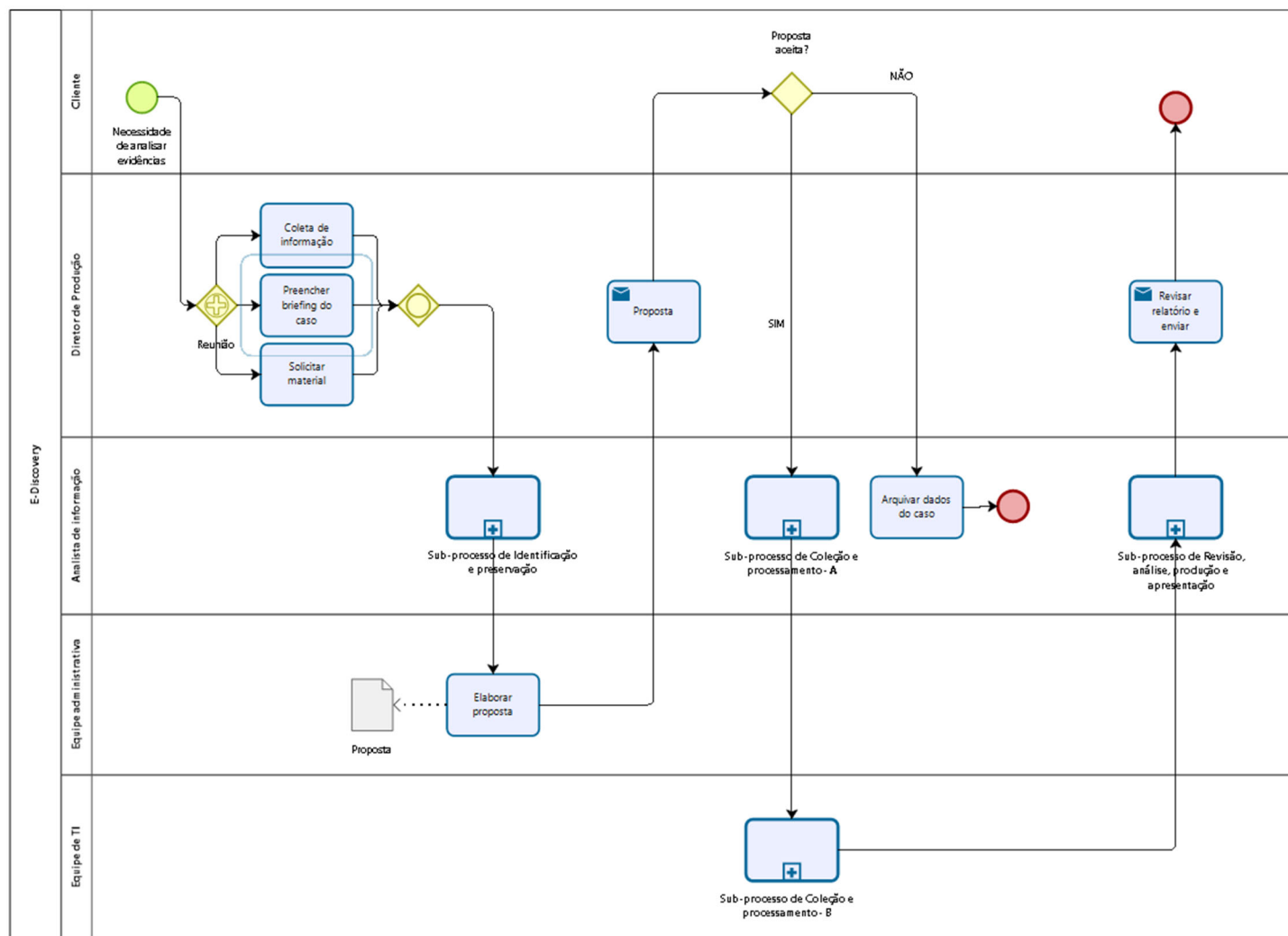
5 APRESENTAÇÃO E ANÁLISE DOS DADOS

O processo de *eletronic Discovery* lida com um volume de informações por caso na ordem de Terabytes (10^{12} bytes). As partes envolvidas no processo lidam com a necessidade de se organizar tal informação e de torná-la processável de acordo com a sua necessidade. Para um melhor entendimento do processo, vide a figura 3 para entender o procedimento no todo do ponto de vista da defesa.

A visão geral é apenas para contextualizar o processo⁵. Os pontos objetos do estudo metodológico são os subprocessos e serão divididos em três fases, contudo o foco será na terceira fase de processos.

⁵ Para melhor visualização do conteúdo, veja a Figura 3 replicada no Apêndice A.

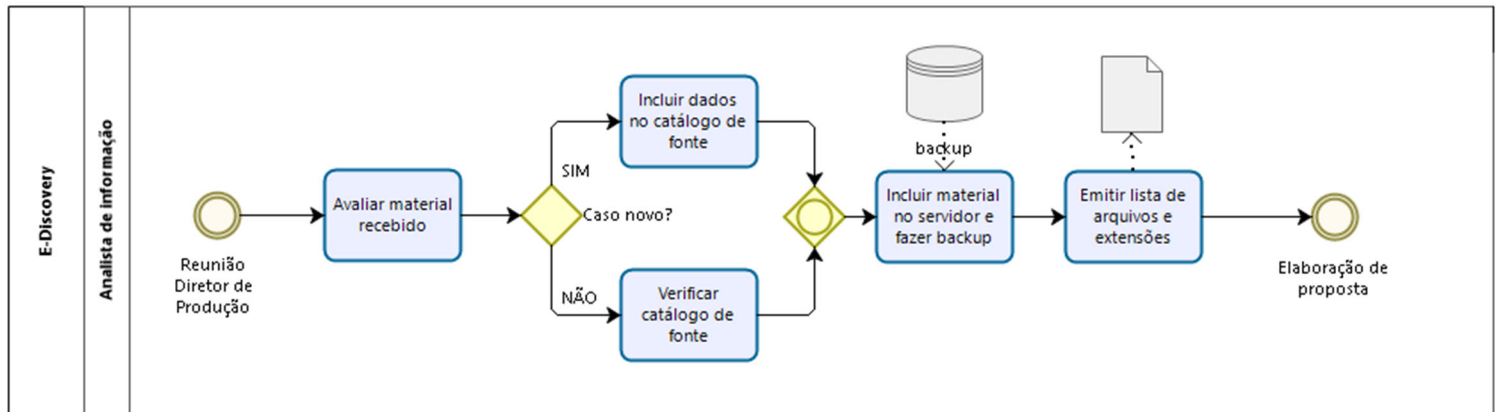
Figura 3 - Mapa de processos



Fonte: Elaboração própria.

A fase 1, como melhor apresentado na Figura 4, compreende a recepção do material para avaliação inicial e armazenamento. Em paralelo ao modelo EDRM, esta fase está inserida nos estágios de identificação e preservação.

Figura 4 - Identificação e preservação



Fonte: Elaboração própria.

Nesta primeira fase, algumas das operações unitárias de Kent (1972) podem ser aplicadas a fim de criar uma memória do caso e permitir explicar ao cliente ou ao juízo o caminho percorrido para se chegar àqueles resultados. Na Figura 4 é possível verificar que o primeiro passo é ‘avaliar material recebido’. Esta primeira análise do conteúdo, assim como a análise inicial de Kent (1972) serve para compreender o material recebido e escolher a melhor forma de estruturar aquela informação a fim de responder às questões levantadas pelo cliente. Quais os tipos de arquivos, tamanho, tempo de processamento estimado, por exemplo, são exemplos de algumas das considerações feitas inicialmente.

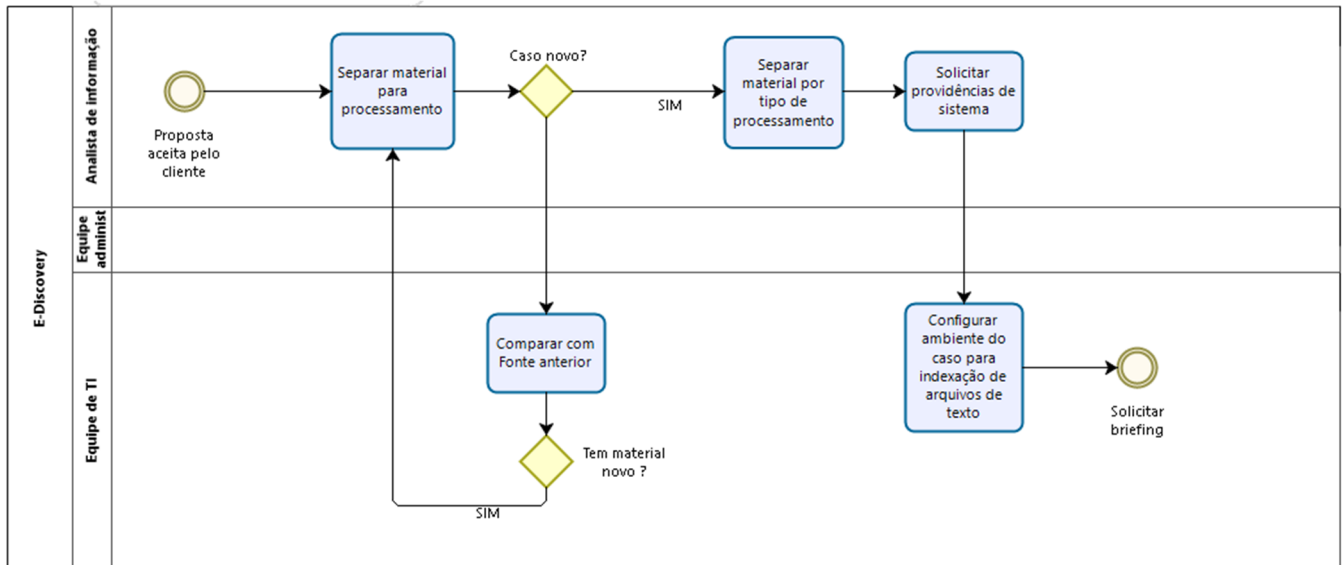
A construção de um catálogo de fontes permite identificar onde os documentos-fonte estão armazenados. Nesta etapa é preciso elaborar também uma ficha de identificação do material recebido, de forma que ela descreva as extensões e quantidades de arquivos para que, futuramente, isso possa ser revisto caso necessário. Esta etapa corresponde ao registro dos resultados das análises de Kent (1972).

Por fim, a inclusão do material no servidor e a realização de *backup* para garantir a integridade daquele material recebido. Desta forma, o processamento não afetará o material original. Este passo, pode ser relacionado com a armazenagem de registros ou documentos-

fonte descrito por Kent (1972) e corresponde ao estágio três do modelo EDRM, a preservação.

A fase 2, como melhor apresentado na Figura 5, corresponde à parte do processamento. Estágio subsequente aos anteriormente citados no modelo EDRM.

Figura 5 – Coleção e processamento



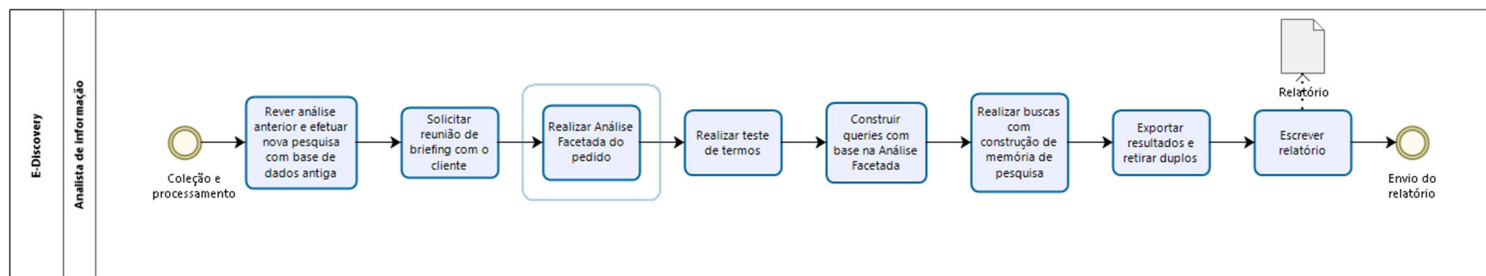
Fonte: Elaboração própria.

A separação do material para processamento corresponde à construção da coleção que será pesquisada, estágio quatro do modelo EDRM. Após a avaliação inicial e a construção da ficha de identificação da fonte recebida, a separação do material para processamento se torna mais simples, uma vez que já se sabe o conteúdo da massa de informação recebida.

Assim, o estágio posterior, o processamento, pode diante da identificação da massa recebida, normalizar e estruturar os documentos para a revisão. Separa-se por tipo de arquivo sendo os textuais enviados para indexação e os outros para procedimentos de carga que não são objetos deste trabalho.

A última fase, melhor detalhada na figura 6, apresenta o foco da análise e corresponde aos estágios subsequentes do modelo EDRM, mas principalmente ao estágio de Revisão.

Figura 6 - Revisão, análise, produção e apresentação



Fonte: Elaboração própria.

Esta fase é foco da pesquisa, pois é nela que serão validadas as variáveis independentes. A começar pela análise facetada. Neste estudo, quando o cliente solicita a realização do trabalho, é realizada uma reunião para entender as necessidades e perguntas a serem respondidas com o material disponibilizado. A aplicação da análise facetada à demanda do cliente (*briefing*) deve identificar o assunto a ser pesquisado e dele extrair conceitos e termos que serão distribuídos nas categorias fundamentais. Desta forma, contextualiza-se e estrutura-se a pergunta de forma que possibilite a construção das consultas. Por exemplo: tem-se um grupo de pessoas envolvidas em um trabalho, o qual foi realizado em um período determinado e este trabalho consiste em elaborar certas atividades. O cliente pede que sejam encontradas evidências da prestação deste serviço.

A análise facetada permite uma visão geral e estruturada do assunto. Uma vez que a pergunta está dividida em facetadas, é possível entender a relação entre os termos e, a partir daí, começar a planejar a pesquisa. Escolher qual das estratégias melhor se aplica para a construção da consulta, subdividir a pesquisa em partes quando se tratar de um assunto mais complexo e assim variar as consultas usando as táticas de Bates (M1, M5, T5, T6, T10 e T11). Após a primeira pesquisa e resultados, recombinar os termos a fim de aprimorar a pesquisa. Em alguns momentos, o retorno da pesquisa é nulo, o que pode significar que aquilo de fato não existe naquela base ou que há alguma especificação na consulta que está escondendo um possível resultado. Neste caso, é interessante revisar a consulta e reestruturá-la.

Aliado à análise facetada, a teoria do conceito também pode ser empregada para estruturar os termos das consultas. Esta pode ser empregada com base nas ideias de Dahlberg (1978), sobre os enunciados verdadeiros do conceito. Por exemplo: se um dos conceitos extraídos na análise facetada for relatório. Por se tratar de uma base de texto integral, é possível que os termos

utilizados não sejam exatamente aqueles expressados pelo cliente. Utiliza-se, então, os enunciados verdadeiros a fim de se encontrar a terminologia presente na base. Outra forma de expandir as possibilidades de pesquisa é reduzir a palavra ao conceito semântico, conforme explicitado por Soergel (2003), e dessa maneira realizar a pesquisa com a raiz desta. Isso pode acarretar uma alta taxa de falsos positivos, pois no exemplo acima citado esse emprego poderia retornar ‘relato’ e relatar’ que não são satisfatórios para a busca desejada. Se assemelha à truncagem, explicada por Lancaster (2004).

A teoria do conceito permite extrair variações do conceito para as consultas. Isto possibilita uma variação nas estratégias de pesquisa e a ampliação das possibilidades de encontrar respostas à pergunta de pesquisa com os termos buscados. Ao se trabalhar com uma base de texto integral, deve-se ter em mente que os resultados são a existência ou não daqueles termos especificados na consulta, portanto quanto mais restrita for a consulta menor a possibilidade de encontrar resultados que não respondam à pergunta inicial. Contudo, há de se avaliar os resultados obtidos a fim de manter a pesquisa condizente com seu objetivo. Portanto, as táticas de monitoramento de Bates (1979) devem ser aplicadas para este fim.

No guia de revisão do modelo EDRM (2010), fala-se em planejamento, mas no tocante à escolha da forma que será realizada a pesquisa, no escopo da pesquisa, dos formatos dos arquivos. A menção a estratégia de pesquisa se restringe apenas à utilização de operadores booleanos e definição de palavras-chave de buscas.

As estratégias de Kent (1972), porém, são necessárias para se definir a relação entre os conceitos ou termos definidos. Se será necessário relacionar duas ou mais pessoas com determinada atividade, se essas pessoas têm mais de um e-mail, por exemplo, ou se não interessa um domínio específico. Os operadores booleanos serão utilizados para construir essa chave, tendo como foco a relação entre os conceitos e termos. As táticas de Bates (1979) são necessárias para manter as pesquisas no foco da análise feita. As táticas de monitoramento são formas de verificar se as consultas representam a demanda, enquanto as táticas de design da formulação da pesquisa são formas de exaurir o assunto. Por fim, as táticas de seleção e revisão de termos específicos são maneiras para variar as formulações das consultas com base em uma primeira pesquisa. Essas táticas podem ser usadas com os resultados da análise facetada ou com os conceitos extraídos.

6 DISCUSSÃO

A revisão de literatura trouxe à luz, a percepção de que não há estruturação nas estratégias de busca para recuperação da informação nos estágios de *eletronic discovery* no tocante ao tratamento da demanda do cliente. O modelo EDRM (2018) apresenta estágios para estruturar e padronizar o processo da descoberta de evidências, incluindo um guia de revisão, em que operacionaliza a recuperação de um documento dentro da coleção. Este guia, contudo, não apresenta qualquer designação de estruturas ou estratégias de busca para se recuperar a informação desejada. Salienta-se que a busca por um documento pode ser realizada por diversos critérios, todavia, em se tratando de documentos eletronicamente armazenados, devem-se considerar as bases de texto integral, e os mecanismos de inteligência artificial baseados em analisadores linguísticos, para as buscas em linguagem natural, a fim de estruturar a busca pela evidência.

Ademais, identificam-se na literatura diferenças conceituais quanto à recuperação da informação e quanto à noção de relevância dos resultados. Tendo-se, portanto, tantas definições e conceitos sobre relevância, optou-se por não avaliar este quesito no estudo em questão, pois como explicado em capítulos anteriores a relevância no contexto jurídico e no contexto da ciência da informação podem divergir. Assim, para não acarretar resultados dúbios, optou-se por não tratar de relevância dos resultados na aplicação das teorias da CI, mas nas contribuições estruturais dessas teorias com o intuito de apresentar uma forma de suprir a lacuna identificada na literatura no estágio de *document review*.

Embora presente no campo jurídico, esperava-se que esta lacuna fosse preenchida com teorias e técnicas da Ciência da informação. A constatação recai, portanto, sobre a possibilidade de atuação deste profissional da informação. Devido ao seu nível aprofundado de compreensão da teoria subjacente à forma apresentada, a execução desta é esperada ser mais eficaz. A confirmação desta suposição constitui um aspecto importante para trabalhos futuros.

A análise facetada foi aplicada ao estágio de revisão do modelo EDRM (2018), fase 3, a qual permitiu compreender a pergunta e seus relacionamentos de forma mais simples, quando estruturada em facetas. Consequentemente, a identificação das relações terminológicas, semântica e sintática, ficou mais clara, e a escolha da estratégia de pesquisa, como descritas por Kent (1972) para a estruturação da consulta, também foi facilitada.

Percebeu-se que as operações unitárias de Kent (1972), se aplicadas às fases anteriores à análise facetada, permitem que o caso seja recuperado posteriormente. Colaboram igualmente para a construção da cadeia de custódia, ponto destacado no guia de revisão EDRM (2010), por ser necessário para o entendimento do caminho percorrido para se chegar àquele documento.

A teoria do conceito foi aplicada na variação das consultas, com foco nos conceitos extraídos da análise facetada. Em casos que a análise facetada é prejudicada por falta de conteúdo, a teoria do conceito também é prejudicada, mas a variação dos termos levado à exaustão permite uma chance maior de se obter resultados alinhados à pergunta inicial. Quando associada às estratégias de termos de Bates (1979), a teoria do conceito possibilita maior ou menor especificação de uma consulta.

Os resultados das primeiras pesquisas devem ser avaliados segundo um critério básico: se eles estão no contexto da pergunta. Esta avaliação é necessária para excluir falsos positivos, aqueles que possuem aqueles termos, mas estão inseridos em contextos diferentes do que se procura. Por se tratar de uma base de texto integral, a pesquisa com as consultas apenas indica que tal documento possui aqueles termos nas especificações ajustadas, mas não significa que eles correspondem ao contexto da pergunta. Aqui se aplicam as estratégias de monitoramento de Bates (1979) aliadas à análise facetada para conferir se os resultados podem seguir para o próximo estágio.

Assim, verifica-se que as teorias elencadas e as técnicas de *information retrieval* permitem estruturar formas de realizar o *document review*. Elas permitem explorar as opções de revisão, os relacionamentos entre termos e as possibilidades de resultados. Entretanto, a avaliação de relevância dos resultados carece de pesquisas aprofundadas, principalmente no que se refere à relevância jurídica do resultado encontrado.

7 CONCLUSÃO

O *e-Discovery* é um processo de recuperação da informação jurídica dentro do campo do litígio, em que se demandam informações de uma parte à outra para busca de evidências em um processo civil ou criminal. Nesse sentido, esta pesquisa buscou demonstrar o surgimento e a evolução do conceito de *Information Retrieval* e *eletronic Discovery* em Ciência da Informação, onde caracterizaram-se conceitos básicos sobre *IR* e *e-Discovery*, assim como conceitos mais complexos como o de relevância. Para a *IR*, este conceito significa ser útil, para o Direito, significa servir de prova ao processo. Há ainda o conceito de *relevancy* que se aproxima à ideia de relevância da Ciência da Informação. Existe ainda a definição de relevância de Cleveland que se destaca dessas, mas aplica-se ao *e-Discovery*. As contribuições teóricas herdadas da Documentação, e mais recentemente da Ciência da Informação, são perceptíveis, uma vez que o processo de *e-Discovery* aplica-se a *Information Retrieval* em um dos seus estágios. Há, todavia, outros conceitos e definições, como a Teoria do Conceito ou a análise facetada de Ranganathan, que estão relacionados ao processo de pesquisa.

A inserção do profissional de ciência da informação neste campo de atuação se mostra, internacionalmente, existente. Uma vez que pesquisadores de referência da área participam de conferências amplamente divulgadas sobre *e-Discovery*, como é o caso da TREC Legal Track e as pesquisas de Saracevic sobre avaliação da relevância. É, portanto, intrigante que no Brasil, país em que a lei de anticorrupção trouxe diversas demandas à escritórios de advocacia, as publicações e comunicações de pesquisas na internet, sobre *e-Discovery*, sejam reduzidas. Porém, alguns artigos em jornais de referência da área jurídica demonstram que este campo está crescendo no país e provavelmente isto seja uma indicação de tendência para se estudar e ampliar este nicho profissional.

Os principais países onde aparecem pesquisas realizadas sobre o tema são os Estados Unidos e Inglaterra. O mercado interno e a realidade jurídica brasileira dão sinais de crescentes demandas em *IR* e *e-Discovery*, nesse sentido, é fundamental investir na expansão do setor. Alguns estudos da TREC Legal Track questionam sobre as habilidades dos advogados e dos pesquisadores de *IR* para realizar as pesquisas e avaliar a relevância dos resultados. Este questionamento pode ser aprofundado ao questionar como as teorias e os conceitos fundamentais nos processos do ciclo informacional, como a Teoria do Conceito, a análise

facetada, as técnicas ou modelos de busca, podem colaborar para uma avaliação de relevância mais precisa dentro do processo de *e-Discovery*.

A realização do estudo metodológico permitiu enxergar que a aplicação da análise facetada no briefing de reunião com o cliente. Possibilitou também a estruturação da pergunta de pesquisa em facetadas, o que trouxe identificação da relação entre os conceitos identificados e permitiu o planejamento da pesquisa. A teoria do conceito aliada à análise facetada permitiu a variação dos conceitos e a definição dos termos relacionados para a exatidão nas queries de busca. Por fim, as estratégias de Kent (1972) colaboraram para a tradução das relações entre os conceitos para a linguagem de busca, os quais aliados às táticas de Bates (1979) permitiram a alterações nas queries de busca e o monitoramento dos resultados em relação à pergunta.

No estudo metodológico identificou-se e descreveu-se como as teorias da Documentação contribuem para o desenvolvimento do estágio de *document review* em *e-Discovery*. Neste sentido, entende-se que a análise facetada e a teoria do conceito, ausentes na metodologia de *e-Discovery*, acrescentaram passos estruturais aplicados na revisão de documentos. Elencaram-se também estratégias de busca e recuperação da informação, amplamente conhecidas em CI, mais eficazes também no processo de *e-Discovery*. Demonstrou-se também como aplicar tais estratégias para o alinhamento e foco da pesquisa com seus objetivos.

Os métodos aqui estudados para estruturar o estágio de revisão do modelo de *e-Discovery* podem ser assimilados por outros profissionais, contudo, devido às habilidades específicas da formação em Ciência da Informação, esses profissionais apresentam claramente maior facilidade em aplicar esses princípios em *document review*. Sobre este particular, verificou-se a necessidade de pesquisas aprofundadas para se avaliar a influência do conhecimento jurídico na definição de *queries*.

REFERÊNCIAS

- AIIM. **What is eDiscovery?** Disponível em: < <https://www.aiim.org/What-is-eDiscovery/> >. Acesso em: 1 maio. 2018.
- ALVARES, L.; ARAÚJO, R. H. DE. Marcos históricos da ciência da informação: Breve cronologia dos pioneiros, das obras clássicas e dos eventos fundamentais. **Transinformacao**, v. 22, n. 3, p. 195–205, 2010. Disponível em: < http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-37862010000300001&lng=pt&tlng=en > Acesso em: 03 set 2018.
- ARAUJO, V. M. A. P. Sistemas de recuperação da informação: uma discussão a partir de parâmetros enunciativos. **Transinformacao**, v. 24, n. 2, 2012. Disponível em: < <http://www.scielo.br/pdf/tinf/v24n2/a06v24n2.pdf> > . Acesso em: 23 set. 2018.
- BARON, J. R. Law in the Age of Exabytes: Some further Thoughts on “Information Inflation” and Current Issues in E- Discovery Search. **Richmond Journal of Law and Technology J.L. & Tech Richmond Journal of Law and Technology**, v. 17, n. 3, 2011. Disponível em: < scholarship.richmond.edu/jolt/vol17/iss3/3 >. Acesso em: 29 mar 2018.
- BATES, M. J. Information Search Tactics. **Journal of the American Society for Information Science**, v. 30, July 1979, p. 205-214. Disponível em: < <https://pages.gseis.ucla.edu/faculty/bates/articles/Information%20Search%20Tactics.html#Table1> >. Acesso em: 03 de set de 2018.
- BRADFORD, S. C. **Documentação**. 1. ed. Rio de Janeiro: Fundo de Cultura, 1961.
- BRIET, S. **Qu’est-ce que la documentation**. Paris: Édit, 1951.
- BROUGHTON, V. The need for a faceted classification as the basis of all methods of information retrieval. **Aslib Proceedings: New Information Perspectives**, v. 58, n. 1–2, p. 49–72, 2006. Disponível em: < <https://pdfs.semanticscholar.org/2075/a8e693425438536da6d51060274c15b95e50.pdf> >. Acesso em: 03 set 2018.
- BUCKLAND, M. Documentation, information science, and library science in the U.S.A. **Information Processing and Management**, v. 32, n. 1, p. 63–76, 1996. Disponível em: < people.ischool.berkeley.edu/~buckland/20THCENT.pdf >. Acesso em: 16 abr 2018.
- CHISHOLM, C. **Integrating forensic investigation methodology into eDiscovery**. Sans Institute, 2010. Disponível em: < <https://www.sans.org/reading-room/whitepapers/incident/paper/33448> >. Acesso em: 03 set. 2018.
- CHOWDHURY, G. G. **Introduction to modern information retrieval**. 3rd ed. London, England, UK: Facet Publishing; 2010. Disponível em: < https://books.google.com.br/books?id=cN4qDgAAQBAJ&pg=PA115&dq=Chowdhury,+G.+G.+Introduction+to+Modern+Information+Retrieval.+3rd+ed.+London,+England,+UK:+Facet+Publishing;+2010.&hl=pt-BR&sa=X&ved=0ahUKEwj1xYT54_LdAhXBI5AKHaBuACYQ6AEINjAC#v=onepage&q=Chowdhury%2C%20G.%20G.%20Introduction%20to%20Modern%20Information%20Retrieval.%203rd%20ed.%20London%2C%20England%2C%20UK%3A%20Facet%20Publishing%3B%202010.&f=false >. Acesso em: 23 set. 2018.

CONFERENCE, T. S. **The Sedona Principles addressing electronic document production**. 2. ed. Phoenix: The sedona conference, 2007. Disponível em: <<https://thesedonaconference.org/publication/the%20sedona%20principles>>. Acesso em: 08 mar 2018.

COSTA, L. S. F. Aproximações teórico-conceitual entre as categorias de Ranganathan, o discurso retórico e a narrativa literária. **Pesquisa Brasileira em Ciência da Informação e Biblioteconomia**, v. 3, n. 1, p. 169-184, 2010. Disponível em: <<http://inseer.ibict.br/ancib/index.php/tpbci/article/view/36/67>>. Acesso em: 03 set 2018.

DAHLBERG, I. Teoria do conceito. **Ciência da Informação**, v.7, n. 2, p. 101-107, 1978. Disponível em: <<http://revista.ibict.br/ciinf/article/view/115/115>>. Acesso em: 03 set. 2018.

DABNEY, D. P. The curse of Thamus: an analysis of full-text legal document retrieval. **Law Library Journal**, vol. 78, n. 5, p. 5-40, 1986. Disponível em: <<https://works.bepress.com/aallcallforpapers/57/>>. Acesso em: 28 ago 2018.

EDRM. **EDRM Model**. 2018. Disponível em: <<https://www.edrm.net/frameworks-and-standards/edrm-model/>>. Acesso em: 07 mar. 2018.

EDRM. **Glossary**. 2016. Disponível em: <<https://www.edrm.net/collections/edrmglossary/>>. Acesso em: 07 mar. 2018.

EDRM. **Review Guide**. 2010. Disponível em: <<https://www.edrm.net/frameworks-and-standards/edrm-model/review-guide/>>. Acesso em: 07 mar 2018.

FIGUEIREDO, L. M. de. O Conceito de Relevância e suas implicações. **Ciência da Informação**, v. 6, n. 2, p. 75-78, 1977. Disponível em: <<http://revista.ibict.br/ciinf/index.php/ciinf/article/view/1584>>. Acesso em 15 abr 2018.

FLOOD, J. **The Use of Technology in Irish eDiscovery with Reference to the EDRM Digital Investigation and Forensic Computing with the supervision of**. 2014, 70 p. Tese (mestrado em Investigação digital e computação forense). Dublin: University College Dublin. Disponível em: <http://www.academia.edu/8417850/The_Use_of_Technology_in_Irish_eDiscovery_with_Reference_to_the_EDRM>. Acesso em: 12 mar 2018.

FOSKETT, A. C. **A abordagem temática da informação**. São Paulo, SP: Polígono, 1973. 437 p.

GIL, A. C. **Métodos e técnicas de pesquisa social**. 6. ed. São Paulo: Atlas, 2012. 200 p.

GUINCHAT, C.; MENOU, M. J.; BLANQUET, M. **Introdução geral às ciências e técnicas da informação e documentação**. 2. ed. Brasília: Instituto Brasileiro de Informação em Ciência e Tecnologia, 1994. 540 p.

GROSSMAN, M. R.; CORMACK, G. V. The Grossman-Cormack glossary of technology-assisted review with foreword by John M. Facciola, U.S. Magistrate Judge. **Federal Courts Law Review**, v. 7, n. 1, p. 1-34, 2013. Disponível em: <www.fclr.org/fclr/articles/html/2010/grossman.pdf>. Acesso em: 14 maio 2018.

KENT, A. **Manual da recuperação mecânica da informação**. Brasília: Universidade de Brasília, 1972. 427 p.

LANCASTER, F. W.; **Indexação e resumos: teoria e prática**. 2. ed. Brasília: Briquet de Lemos, 2004. 452 p. Cap. 14 e 15.

LAWTON, D.; STACEY, R.; DODD, G. **eDiscovery in digital forensic investigations**. CAST, UK Home Office, n. 32, 2014. Disponível em: <www.gov.uk/government/uploads/system/uploads/attachment_data/file/394779/ediscovery-digital-forensic-investigations-3214.pdf>. Acesso em 23 abr 2018.

LE COADIC, Y.-F. **A Ciência da Informação**. 2. ed. Brasília: Briquet de Lemos, 2004.

OARD, D. W.; BARON, J. R.; HEDIN, B.; LEWIS, D. D.; TOMLINSON, S. Evaluation of information retrieval for E-discovery. **Artificial Intelligence and Law**, Texas, v. 18, n. 4, p. 347–386, 2010. Disponível em: <terpconnect.umd.edu/~oard/pdf/jail10.pdf>. Acesso em 16 maio 2018.

OARD, D. W.; WEBBER, W. Information Retrieval for E-Discovery. **Foundations and Trends R in Information Retrieval**, v. 7, n. 2–3, p. 99–237, 2013. Disponível em: <<https://terpconnect.umd.edu/~oard/pdf/fntir13.pdf>>. Acesso em: 9 abr 2018.

PRODANOV, C. C.; FREITAS, E. C. de. **Metodologia do trabalho** : métodos e técnicas da pesquisa e do trabalho acadêmico. 2. ed. Novo Hamburgo: Feevale, 2013. 277 p.

RASMUSSEN, E. Access models. In: RUTHVEN, Ian; KELLY, Diane (Ed). **Interactive information seeking, behaviour and retrieval**. London: Facet Publishing, 2011. p. 95-112

REINALDO FILHO, D. A exibição da prova eletrônica em juízo – necessidade de alteração das regras do processo civil? In: **Migalhas**. [s.l.], 2006. Disponível em: <<https://www.migalhas.com.br/dePeso/16,MI31031,21048-A+exibicao+da+prova+eletronica+em+juizo+necessidade+de+alteracao+das>>. Acesso em: 8 mar 2018.

ROBREDO, J. **Documentação de hoje e de amanhã: uma abordagem revisitada e contemporânea da Ciência da Informação e de suas aplicações biblioteconômicas, documentárias, arquivistas e museológicas**. 4. ed. Brasília: Ed. do Autor, 2005. 409 p.

ROITBLAT, H. **Measurement in eDiscovery** : a technical white paper measurement in eDiscovery. 2013. Orca Tec. Disponível em: <<https://theolp.wildapricot.org/Resources/Documents/Measurement%20in%20eDiscovery%20-%20Herb%20Roitblat.pdf>>. Acesso em 12 mar 2018.

SARACEVIC, T. Relevance: A review of the literature and a framework for thinking on the notion in information science. Part II: Nature and manifestations of relevance. **Journal of the American Society for Information Science and Technology**, v. 58, n. 13, p. 1915–1933, 2007. Disponível em: <tefkos.comminfo.rutgers.edu/articles.htm>. Acesso em: 15 abr 2018.

SARACEVIC, T. Information Science. In: **Encyclopedia of Library and Information Sciences**, 3. ed. New York: Taylor & Francis, 2009. p. 2570–2586. Disponível em: <<https://tefkos.comminfo.rutgers.edu/articles.htm>>. Acesso em: 12 mar 2018.

SARACEVIC, T. Research on relevance in Information Science: a historical perspective. Proceedings of the ASIST Annual Meeting. **Anais...2012**. Disponível em: <<https://tefkos.comminfo.rutgers.edu/articles.htm>>. Acesso em: 15 abr 2018.

SATHIYANARAYANAN, M.; TURKAY, C. **Determining and Visualising E-mail Subsets**

to Support E-discovery. 2016. [s.l: s.n.]. Disponível em: <[http://openaccess.city.ac.uk/16154/1/2016 - IEEE_VIS - Mithileysh - Determining and Visualising E-mail Subsets to Support E-discovery_Poster.pdf](http://openaccess.city.ac.uk/16154/1/2016-IEEE_VIS-Mithileysh-Determining-and-Visualising-E-mail-Subsets-to-Support-E-discovery-Poster.pdf)>. Acesso em: 12 abr 2018.

SOERGEL, D. **Beyond facets:** Semantic roots and modifiers as elements of a conceptual morphology. 2003. [s.l: s.n.]. Disponível em: <<http://www.dsoergel.com/cv/B36.pdf>>. Acesso em: 03 set 2018.

STOPANOVSKI, M. R. Softwares que os operadores jurídicos no Brasil ainda vão usar. In: **Consultor Jurídico.** Brasília, 2015. Disponível em: <<https://www.conjur.com.br/2015-fev-11/suporte-litigios-sofwares-operadores-juridicos-brasil-ainda-usar>>. Acesso em: 8 mar 2018.

WANG, J.; SOERGEL, D. A user study of relevance judgments for E-discovery. Proceedings of the ASIST Annual Meeting. **Anais...**2010 Disponível em: <<https://onlinelibrary.wiley.com/doi/abs/10.1002/meet.14504701157>>. Acesso em 15 mar 2018.

Apêndice A – Mapa de processos

