

A REPRESENTAÇÃO DA INFORMAÇÃO: CLASSIFICAÇÃO E INDEXAÇÃO AUTOMÁTICA DE DOCUMENTOS DE ARQUIVO

THE REPRESENTATION OF INFORMATION: AUTOMATIC CLASSIFICATION AND INDEXING OF ARCHIVES RECORDS

Renato Tarciso Barbosa de Sousa

Resumo: A classificação é uma função matricial do que-fazer arquivístico e um dos processos do programa de gestão de documentos. É, também, um elemento importante para a transparência e compartilhamento de informações, que são caminhos seguros para a tomada de decisão, para preservação da memória técnica e administrativa das instituições contemporâneas e para o pleno exercício da cidadania. A classificação de documentos de arquivo tem três objetivos: facilitar o acesso à informação; manter, preservar o vínculo arquivístico; fundamentar as funções arquivísticas de avaliação e de descrição. Essa importância, reconhecida pela literatura nacional e internacional, não é compatível com o nível de verticalização que se tem sobre esse tema. O pressuposto é que o arcabouço teórico-metodológico da área não consegue resolver as questões impostas pela grande massa de documentos e informações acumulada diariamente pelas organizações contemporâneas, pela disseminação das novas tecnologias e pelo desenvolvimento social. A pesquisa tem como objetivo a construção de um modelo para classificação e indexação automática de documentos de arquivo. Ela se utiliza dos métodos das Ciências Humanas e Sociais para abordar o tema a partir do material empírico privilegiado nesse trabalho, que é a literatura arquivística nacional e internacional e aquela que trata da Teoria da Classificação, da Teoria do Conceito, da Organização e Representação do Conhecimento, da Administração, da Tecnologia da Informação, da Ciência da Informação e da Diplomática Contemporânea.

Palavras-chave: Arquivística. Classificação de documentos arquivísticos. Organização de documentos de arquivo. Controle de vocabulário em arquivo. Indexação automática.

Abstract: The classification is a matrix function of the archival what-to-do one the processes of document management program. It is also an important element for transparency and information sharing, which are safety directions for making decisions, to preserve the memory of the technical and administrative contemporary institutions and the full exercise of citizenship. The classification of records has three objectives: facilitate access to information; maintain, preserve the archival bond; substantiating the archival functions of evaluation and description. This importance, recognized by national and international literature, is not compatible with the level of vertical integration we have about this topic. The assumption is that the theoretical and methodological framework of the area cannot resolve the issues imposed by the great mass of documents and information accumulated daily by contemporary organizations, the dissemination of new technologies, and social development. The research aims to build a model for automatic classification and indexing of records. It uses the methods of Humanities and Social Sciences to address the issue from the empirical field in this work, which is the national and international archival literature and the one that deals with the Classification Theory, the Theory of Concepts, Organization and Representation knowledge Management, Information Technology, Information Science and Contemporary Diplomatic.

Keywords: Archival. Classification of records. Organization of records. Vocabulary file control. Automatic indexing.

1 INTRODUÇÃO

A classificação de documentos de arquivo aparece nos estudos da área como uma preocupação há vários anos. A importância apontada pela literatura da área não teve, no mesmo nível, o estabelecimento de um instrumental teórico-metodológico consolidado. Fala-se e trabalha-se com classificação de documentos de arquivo sem agregar o desenvolvimento desse conceito proporcionado pela Filosofia, pela Teoria da Classificação e, mais recentemente, pela área de Representação e Organização da Informação. Costuma-se apresentar o caminho sem discutir os meios necessários para percorrê-lo. Importam-se, sem críticas, métodos desenvolvidos para objetos de características e natureza diferentes dos da Arquivística. E tudo isso tendo como pano de fundo a imprecisão terminológica, que parece uma marca comum da área.

A organização dos documentos tem sido um dos grandes desafios da Arquivística contemporânea. Lidar com massas tão grandes de documentos acumulados diariamente pelas instituições exige uma maior sofisticação dos esquemas de classificação. O documento arquivístico é um artefato humano com pressupostos e características específicas. O ambiente e o conteúdo são delimitados e definidos pelo sujeito acumulador, que pode ser uma pessoa física ou jurídica (organização). Ao contrário daqueles encontrados em bibliotecas, por exemplo, os documentos arquivísticos não constituem um conjunto formado em vista de uma finalidade específica: eles representam o produto da atividade do sujeito acumulador. Entender o modo como as organizações estruturam-se e como executam suas funções e atividades é compreender como os documentos são acumulados. Ele é resultado de um ato desenvolvido e, na maioria dos casos, cotidianamente repetido. A gênese se dá quando a organização tem algo a cumprir, a provar, a determinar. Surge naturalmente como resultado das ações desenvolvidas pelo sujeito acumulador. Após o registro das informações em suportes (papel, mídia magnética, microfilme, películas fotográficas, películas cinematográficas etc.), é necessário mantê-las pelos valores administrativos, técnicos, legais, fiscais, probatórios, culturais e históricos que possam conter. À medida que os documentos vão sendo acumulados, estabelecem relações entre si. Eles estão unidos por um elo criado no momento em que são produzidos e recebidos, determinado pela razão de sua elaboração e que é necessário à própria existência e a capacidade de cumprir seu objetivo. Eles são um conjunto indivisível de relações intelectuais.

A explosão do volume de documentos arquivísticos ocorreu de forma paralela com o aumento geral da informação registrada pelas sociedades humanas. Talvez, em um ritmo mais lento, mas nem por isso menos significativo. Os fenômenos da industrialização e da

urbanização, o crescimento da presença dos estados nacionais e da complexidade de seus aparatos e, secundariamente, a banalização dos meios tecnológicos de reprodução de documentos explicam em parte a afirmação anterior. A partir da década de 1980, a consolidação da microinformática fez aparecer e tornar extremamente comum nas organizações um novo tipo de documento arquivístico: o eletrônico. Não é simplesmente um novo suporte para o registro de informações originadas das funções e atividades das instituições; ele tem características próprias, que impõem a definição de modos específicos de criá-lo, classificá-lo, avaliá-lo, descrevê-lo e de atestar sua autenticidade.

O conceito de classificação de documentos de arquivo não é estranho à literatura arquivística, muito pelo contrário. Os grandes manuais sempre dedicaram capítulos para tratar do tema. Entretanto, a presença não garantiu uma verticalização teórica sobre o mesmo. A classificação, que a Lógica define como a distribuição de indivíduos em grupos distintos, de acordo com caracteres comuns e caracteres diferenciadores, aparece na literatura arquivística cindida em dois outros conceitos: classificação e arranjo.

Essas constatações dão a dimensão da necessidade de se proceder a um estudo sobre a questão, tão crucial para o profissional do arquivo e que tem sido dominada pelo empirismo e pelo senso comum. Esta parece, em alguns momentos, ser uma marca indelével da área. Silva et al. (1999, p. 208) chegam a falar de um empirismo eclético direcionado à integração das novidades tecnológicas no “velho” quadro da prática funcional de classificação, ordenação, representação descritiva e difusão de informação escrita. Há, inclusive, uma tendência de “duvidar da possibilidade de se encontrarem bases científicas para o saber arquivístico. Os seus argumentos incidem sobre o fato de estarmos perante uma prática multissecular que dispensa fundamentação teórica especial. Segundo este ponto de vista, os modelos estão criados e será na experiência, na intuição e no bom senso que se encontrarão as soluções adequadas para enfrentar a realidade.” (SILVA et al., 1999, p. 203-204).

O trabalho parte do pressuposto formulado pelo arquivista canadense Jean-Pierre Wallot, ex-presidente do Conselho Internacional de Arquivos para quem sem a pesquisa, a Arquivística está condenada a permanecer como um corpo de práticas e de receitas sem base científica. O fato de ser uma disciplina com finalidades pragmáticas a coloca em uma fronteira perigosa com o empirismo. Busca-se, ainda, com o estudo proposto, uma fundamentação teórica para o tratamento da questão da classificação de documentos de arquivo.

2 O PROBLEMA

A informação, tanto quanto os recursos humanos, materiais e financeiros, é um elemento vital para a sobrevivência de qualquer organização. Ela deve ser gerenciada de forma eficaz e eficiente, o que implica no reconhecimento de sua importância e que lhe seja dado um tratamento no mesmo nível que aquele dispensado aos outros recursos.

A organização da informação na esfera pública é uma condição para o pleno acesso à informação. O acesso deve ser visto sob dois aspectos: nível legal (*corpus* de leis) e nível técnico-científico. No nível legal, a entrada em vigor da Lei 12.527/2011 alterou essa questão no Brasil. A Lei inverteu a lógica transformando a exceção em regra, isto é, existe agora uma Lei que garante o acesso à informação pública e a exceção é o sigilo. A restrição é aquilo que pode colocar em risco a segurança do Estado e da sociedade e a privacidade das pessoas.

Entretanto, o maior entrave ao pleno acesso à informação não são as poucas restrições impostas pela lei, mas a condição de organização da informação pública. Diante disto, fica patente que os procedimentos que envolvem a produção, tratamento técnico e preservação dos acervos arquivísticos e informacionais devem estar dispostos de modo a possibilitar a sua recuperação para atender as demandas informacionais da sociedade. Aos usuários da informação devem ser franqueados todos os mecanismos que facilitem o seu acesso às informações públicas, a fim de garantir a transparência do estado.

Assim, o pleno exercício do direito à informação somente vai se realizar com a organização dos documentos públicos. Dessa forma, teremos que trazer para o cenário o conceito de gestão de documentos, que é reconhecido nacionalmente a partir da Lei 8.159/1991, a “Lei dos Arquivos no Brasil”, e internacionalmente a partir de várias normas ISO, principalmente a 15.489, a 30.300 e a 30.301.

A tecnologia da informação tem proporcionado grandes avanços nas áreas que trabalham com grandes volumes de documentos e informações. Entretanto, ela, em alguns momentos, é entendida como uma panacéia para todos os males. Em relação à classificação de documentos de arquivo, por exemplo, apesar de uma compreensão de que a tecnologia da informação, com suas várias aplicações, pode substituir completamente as rígidas estruturas de classificação por formas de acesso definidas pelos usuários.

Aos incautos devemos lembrar que o acesso, a recuperação, a busca são apenas um dos objetivos da classificação de documentos de arquivo. Autores, como Fiorella Foscarini (2010, p. 42), chegam a entender que a recuperação dos documentos é somente um benefício colateral da classificação, sendo seu propósito primário colocar os documentos individuais nos conjuntos aos quais pertencem, baseando-se no mandato e nas funções do criador, como

defende Luciana Duranti (2003, p. 43). Uma das características principais do documento de arquivo é exatamente a relação que os documentos mantêm entre si e que precisa ser preservada para dar sentido aos próprios documentos e as atividades ou ações que lhes deram origem. Inclusive, a manutenção desse contexto (representadas por essas relações entre os documentos) é um dos elementos mais destacados nas pesquisas que estudam a questão da autenticidade dos documentos digitais.

A classificação é uma função arquivística, tanto quanto a avaliação, a descrição, a criação, a preservação, a aquisição e a difusão. De acordo com Lopes (1996), o coração das práticas é formado pela classificação, avaliação e descrição. A partir desse entendimento, considero a classificação uma função matricial, pois é por meio dela que se realiza a avaliação. O instrumento arquivístico de gestão dos prazos de guarda e da destinação final (tabela de temporalidade), independente da metodologia aplicada e de sua configuração sempre tem como ponto de partida os conjuntos documentais definidos na classificação. Parte-se para a avaliação tendo como ponto inicial a classificação. No caso da descrição, a situação é semelhante. A norma internacional (ISAD-G) e a nacional (Norma brasileira de descrição arquivística, Arquivo Nacional, 2006) definem como primeiro elemento de descrição o conjunto documental, representado pelo fundo, série ou subsérie, isto é, a classificação.

Dessa forma, a classificação ganha uma envergadura, que sustenta toda a construção da prática arquivística. Ter qualidade na definição dos conjuntos documentais e de suas relações é fundamental para o que-fazer arquivístico.

A classificação é parte integrante de um Programa de Gestão de Documentos. A ISO 15.489/2001 (gestão de documentos) define sete processos de gestão de documentos, dentre eles o de organização lógica para a recuperação: classificação e indexação. A norma internacional e o E-Arq Brasil (Arquivo Nacional, 2011) indicam como instrumentos específicos e obrigatórios de gestão de documentos o plano de classificação de documentos.

Empiricamente, podemos afirmar que a excelência em situação arquivística é atestada pela existência das seguintes condições na instituição: política arquivística, existência de uma unidade político-administrativa responsável pela gestão dos documentos, de recursos humanos qualificados, de recursos materiais, de recursos financeiros e, por último, por instrumentos de gestão arquivística (plano de classificação, tabela de temporalidade etc.).

Parte-se da idéia de que o problema da classificação apresenta quatro facetas. A primeira, a indefinição e imprecisão do objeto de estudo da Arquivística. A segunda, a necessidade de um tratamento interdisciplinar da questão. A terceira, o caráter positivista das

práticas e intervenções, verificado na trajetória da produção do conhecimento na área. E, por último, a ausência de procedimentos metodológicos para a coleta de dados necessários à construção de instrumentos de classificação.

O uso cada vez mais frequente da Tecnologia da Informação possibilitou, em muitas organizações públicas e privadas, a substituição do registro e controle manual da tramitação por sistemas automatizados. É possível encontrá-los disponibilizados em redes locais e remotas. Esses sistemas têm sido desenvolvidos por profissionais de Informática. A estrutura das bases de dados, criadas para este fim, reproduz os mesmos campos definidos nas fichas, formulários e livros de protocolo.

Um dos principais problemas encontrados nos sistemas informatizados é o preenchimento do campo intitulado *Assunto* ou *Descrição do Assunto*, que tem como objetivo permitir uma identificação do conteúdo informacional do documento. Nesse campo, deve ser descrita a razão de ser do registro documental, qual o motivo de sua criação.

A análise do conteúdo desse campo nos registros de sistemas de protocolo informatizados indicou uma total falta de padrão para o preenchimento do mesmo. É impossível imaginar que uma organização, por mais complexa que ela seja, tenha uma variedade tão expressiva de assuntos. Os ganhos de eficiência no acesso à informação perdem-se num emaranhado de termos vazios, fragmentados, sem conexão e sem uma vinculação muito clara com o conteúdo informacional dos documentos. É preciso lembrar que os recursos informáticos não permitem ainda uma interpretação que resolva a imprecisão do preenchimento do campo *Assunto*. A lógica é binária.

O problema identificado é o seguinte: os dois principais campos para o acesso ao documento e à informação contida no documento de arquivo são preenchidos de forma imprecisa, tornando pouco eficiente, em termos de busca, um instrumental extremamente sofisticado, que são os sistemas utilizados nas organizações contemporâneas.

3 OBJETIVOS

3.1 Objetivo geral

O objetivo geral desta pesquisa é o de construir um modelo para a classificação de documentos de arquivo, a partir da indexação automática desses documentos mantidos em repositórios digitais, relacionando com a gestão dos prazos de guarda.

Busca-se, portanto, a construção de um modelo para a classificação de documentos de arquivo, a partir da indexação automática desses documentos. O modelo operará com um instrumento de classificação e de avaliação consolidados: o Código de Classificação de

Documentos de Arquivo das Atividades-meio e a Tabela de Temporalidade de Documentos de Arquivo das Atividades-meio, elaborados para serem aplicados nos órgãos e entidades da Administração Pública Federal. Será necessária, ainda, a construção de um vocabulário controlado para ser vinculado àquela estrutura de classificação.

Procura-se uma tentativa de solução dos problemas impostos pela organização dos arquivos. Procuramos resolver, em parte, o desafio imposto com a elaboração de um vocabulário controlado vinculado à estrutura de classificação.

Com a presença cada vez mais significativa dos documentos digitais nas instituições, avançamos nossa pesquisa para o processamento automático, vinculando diretamente classificação e avaliação de documentos. A ideia é submeter um repositório de documentos digitais a um *software* de mineração de dados. Essa mineração seria feita a partir de vocabulário controlado. A identificação dos termos desse vocabulário estaria vinculada a uma estrutura de classificação (resultado da primeira fase da pesquisa) e esta à tabela de temporalidade. Ao final do processamento teríamos um documento indexado, classificado e avaliado, isto é, com todos os elementos para seu pleno acesso e capaz de ser gerenciado arquivisticamente. Essa é a ideia.

Serão introduzidos os procedimentos micro de controle de vocabulário, que tem por objetivo o controle dos termos ou expressões. O controle de vocabulário foi inicialmente desenvolvido pela documentação e introduziu-se na Arquivística, especialmente em razão da informatização crescente dos arquivos por meio da elaboração de listas de assuntos para protocolos informatizados. Ele é um recurso para organizar e recuperar documentos e informações com consistência, gerando, conseqüentemente, confiança no sistema. É, também, um meio para produzir confiança no sistema de organização e busca de informações arquivísticas.

3.2 Objetivos específicos

- Elaborar uma metodologia para construção de vocabulário controlado para documento de arquivo;
- Relacionar o vocabulário em uma estrutura de classificação de documentos de arquivo;
- Identificar *software* de mineração de textos, que permita a mineração a partir de termos previamente selecionados;
- Analisar experiências de indexação automática;
- Elaborar modelos de tipos documentais.

4 JUSTIFICATIVA

O sucesso dessa pesquisa e dessa proposta pode resolver a importante questão da subjetividade na classificação de documentos e os problemas de qualificação de nossos recursos humanos.

Entendemos a organização dos documentos de arquivo como parte da função social do arquivista. Organizamos para possibilitar o acesso rápido, seguro e eficiente à informação. Em um primeiro momento, para a tomada segura de decisão, para a comprovação, para a garantia de direitos e deveres da instituição ou de pessoas, para um estudo retrospectivo, para a manutenção de um estoque informacional que possa servir de ponto de partida para novas atividades e, depois, para preservação do capital informacional, que permitirá um uso além das fronteiras da criação do próprio documento. Nesse sentido, a classificação assume um lugar privilegiado e aparece como ferramenta importante para esse trabalho, ou melhor, como a primeira intervenção, garantindo qualidade e os fundamentos para outras incursões no universo arquivístico, tais como: a avaliação e a descrição.

Acreditamos na possibilidade do estabelecimento de um novo marco referencial para a organização de documentos de arquivo. Esse marco, entretanto, deve ser, na nossa compreensão, o resultado do conjunto de conhecimentos cujos limites extrapolam as fronteiras da própria Arquivística. É na Filosofia, na Teoria da Classificação, na História, na Administração, na Sociologia, na Diplomática Contemporânea, na Representação e Organização do Conhecimento, na Tecnologia da Informação e na Ciência da Informação, além da Arquivística, que encontramos os fundamentos teórico-metodológicos para a operação com a organização de documentos de arquivo. Não é a adição simples de saberes, mas uma articulação a partir dos limites dos objetivos propostos.

5 METODOLOGIA

Uma pesquisa de tal envergadura requer o aporte, tanto de aspectos teórico-conceituais da Ciência da Informação e da Arquivística, quanto de métodos, técnicas e instrumentos pertencentes a estas e às disciplinas que com elas interagem, como a Teoria da Administração, a Tecnologia da Informação, dentre outras.

No âmbito da Arquivística, defende-se aqui o princípio da proveniência em toda a sua extensão, que tem como desdobramento, como diz Belotto (2004), ou duas caras, como prefere De Luigi, ou dois graus como afirmam os canadenses, o respeito ao fundo e à ordem original. Não se perde de vista aqui que a noção de proveniência (origem dos documentos) está ligada àquela de acumulação (produção ou recebimento) e não a de recolhimento.

O respeito à ordem original, como desdobramento do princípio da proveniência, é entendido aqui não no seu sentido mais restrito. Precisamos ter outra leitura, que possibilite a manutenção da ordem original desde que ela represente, reflita a própria gênese documental e as relações orgânicas estabelecidas pelos documentos. Preservar a ordem original significa manter o documento e as suas relações orgânicas no contexto das atividades que o geraram. Pela descrição feita da realidade arquivística brasileira, percebe-se que é necessária, na maior parte dos casos, uma atuação ativa do arquivista, elaborando os instrumentos que devolvam os documentos a seu lugar de origem. Lugar intelectual e, não necessariamente, físico.

A aplicação do princípio de proveniência e de seus desdobramentos na realidade arquivística brasileira não é muito pacífica. Essa dificuldade é formada pela desorganização dos arquivos na origem (arquivos montados nos setores de trabalho e massas documentais acumuladas), pela instabilidade institucional, pela ausência de trabalhos em história administrativa e pela falta de critérios para a estruturação das organizações públicas, ou melhor, a utilização de critérios não científicos. As organizações deveriam ser estruturadas a partir de uma análise dos seus objetivos, da sua missão, e da compreensão das funções e atividades desenvolvidas. Como resultado desse estudo seria desenhado o organograma, que é a representação gráfica da estrutura organizacional e que indica a forma como o poder administrativo é dividido e delimita os níveis de autoridade, a linha de comando em seu sentido vertical. Alguns autores chegam a afirmar que, na verdade, o que existe nas organizações públicas brasileiras não é um organograma, mas sim um “personograma”. Fatores, então, de ordem política ou mesmo político-partidária e a tradição burocrática impedem que as definições ocorram daquela forma.

Entendemos como válida a utilização do princípio da proveniência e o seu desdobramento como princípios de divisão ou de classificação. Eles são compreendidos, no âmbito dessa pesquisa, como as bases que fundamentam toda a atividade de classificação. Na verdade, esses princípios são para a Teoria da Classificação a característica ou princípio da classificação ou da divisão, isto é, a qualidade ou o atributo escolhido para servir de base ao processo classificatório. Assim, existem tantas classificações quantas forem as características possíveis de serem empregadas como base da divisão.

De acordo com os fundamentos teóricos da classificação, classificar significa distribuir indivíduos em grupos distintos, de acordo com caracteres comuns e caracteres diferenciadores. Pode-se fazer essa distribuição observando-se características superficiais e mutáveis ou levando em consideração caracteres essenciais e permanentes. No primeiro caso,

a classificação é elaborada a partir de um princípio de divisão ou classificação artificial. No segundo caso, o princípio será natural.

A classificação que utiliza princípios de divisão ou de classificação artificiais não se prende à essência dos objetos distribuídos e podem ser modificados, suprimidos mesmo, sem a alteração dessa essência. É o caso, por exemplo, de distribuir os homens pela cor do cabelo, arranjar os livros em uma estante de acordo com a encadernação ou organizar os documentos arquivísticos a partir do seu suporte ou formato. Entretanto, a classificação que parte de princípios de divisão ou de classificação naturais encontra e exprime a ordem seguida pela natureza no meio das dissemelhanças quase infinitas dos seres. Ela parte da observação dos indivíduos e, em seguida, compara-os. Observando caracteres comuns em certo número de indivíduos dissemelhantes, elimina as diferenças, conserva as analogias e com estas constitui tipos de variável extensão.

Podemos, a partir dessas definições, entender a proveniência e o seu desdobramento como princípios de divisão ou de classificação naturais, pois são atributos essenciais e permanentes ao conjunto (arquivo) a ser dividido. A origem desse conjunto de documentos é uma marca indelével e inseparável do objeto e é o que lhe dá inteligibilidade, identidade e sentido. As características desse conjunto documental são delimitadas pelo sujeito acumulador (pessoa física ou jurídica) e pelas relações orgânicas que se estabelecem no interior desse conjunto.

De forma diferente ocorreu com os métodos de agrupamento dos documentos na fase anterior às determinações de Natalis de Wailly. Naquela época, os registros documentais eram reunidos a partir de características artificiais, isto é, de princípios de divisão ou de classificação baseados em características superficiais e fáceis de observar, mas que não representavam relações verdadeiras e, por esta razão, a classificação tornava-se menos perene e menos fiel às características dos objetos classificados.

A partir desse entendimento, podemos afirmar que o primeiro princípio de divisão a ser utilizado é o da proveniência, que permitirá a separação e a delimitação, dentro do patrimônio arquivístico de uma nação ou de um país, dos arquivos pertencentes a uma pessoa física ou jurídica, pública ou privada.

Enquanto o respeito aos fundos é o próprio princípio da divisão do patrimônio arquivístico de uma nação ou de um país, internamente, o princípio de respeito à ordem original indica os elementos que vão servir de base para as divisões do fundo. Esses elementos podem ser encontrados no contexto de gestação do registro documental. No sentido mais amplo, é a missão da organização que vai permitir o aparecimento dos documentos de

arquivo, mas é o desenvolvimento dessa finalidade maior que vai determinar a existência de funções, das atividades e das tarefas administrativas, que têm, nesse sentido mais restrito, como produto e subproduto o documento arquivístico. É esse processo que dá ao documento sua natureza, suas características e as suas especificidades (conforme discutido no capítulo três) e, também, seu sentido e sua própria existência. Portanto, a denominação do respeito à ordem original dada pelos canadenses, segundo grau do princípio da proveniência, parece mais adequada à compreensão.

Do mesmo modo que a identificação do fundo nem sempre é uma tarefa pacífica, os procedimentos administrativos internos que permitem o aparecimento do documento também são, em muitos casos, de difícil compreensão, pois estão, em geral, envolvidos em processos organizacionais informais, improvisados e sem registro. Defendemos aqui o estabelecimento de um ferramental teórico-metodológico para buscar a gênese documental e retirar dela as características ou atributos para fundamentar a divisão dos níveis de classificação internos do fundo de arquivo.

Este trabalho é, também, um esforço que vai além do tema proposto, pois, em todos os momentos deste estudo, parte-se da Arquivística enquanto disciplina do conhecimento humano, que, para alcançar um nível de desenvolvimento, não pode prescindir de trabalhos de pesquisa para aplicação e questionamento dos seus modelos de interpretação e de intervenção na realidade. Trabalha-se com os princípios e os conceitos de uma maneira crítica, procurando sempre tratá-los como pontos de partida para a reflexão, como o norte a ser seguido, mas de uma maneira a não limitar as possibilidades de solução aos problemas colocados. Busca-se, também, um diálogo com as áreas que possam contribuir para superar as bases intuitivas de tratamento da função classificação em Arquivística. O diálogo ocorrerá principalmente com a História, com a Filosofia, com a Teoria da Classificação, com a Administração, com a Diplomática Contemporânea, com a Tecnologia da Informação, com a Organização e Representação do Conhecimento e com a Ciência da Informação.

O percurso metodológico será apoiado em dois campos de investigação: teórico e empírico. Essa separação é apenas de fundo organizacional. A análise prescinde de uma interação e um diálogo constante entre os dois campos. Esclarecemos que o espaço empírico da pesquisa é a Administração Pública brasileira e seus instrumentos de gestão arquivística, principalmente aqueles elaborados pelo Conselho Nacional de Arquivos.

As seguintes etapas deverão ser cumpridas:

- realizar estudos para definir as bases teóricas para a elaboração de vocabulários controlados a serem aplicados em documentos de arquivo;

- analisar autores atuais da Arquivística e da Ciência da Informação que trabalham com a classificação de documentos de arquivo;
- realizar estudos a partir da Teoria da Classificação, visando estabelecer os requisitos necessários à construção de instrumentos de classificação de documentos de arquivo;
- identificar as propostas metodológicas para a elaboração de instrumentos de classificação de documentos de arquivo;
- identificar *software* de mineração de textos;
- testar os *softwares* existentes e definir quais são os requisitos necessários para o desenvolvimento da indexação a partir de uma lista selecionada de termos;
- elaborar uma metodologia para construção de vocabulário controlado em arquivo;
- definir um vocabulário controlado para as atividades-meio da Administração Pública Federal;
- relacionar o vocabulário controlado à estrutura do Código de Classificação da atividade-meio, elaborado pelo Conselho Nacional de Arquivos;
- relacionar a classificação com a avaliação de documentos de arquivo, a partir do Código de Classificação de Documentos de Arquivo e da Tabela de Temporalidade de Documentos de Arquivo, elaborados pelo Conselho Nacional de Arquivos;
- identificar um repositório digital de documentos de arquivo em um órgão da Administração Pública Federal;
- submeter os documentos digitais ao *software* minerador de textos;
- incluir os instrumentos (Código de Classificação, Tabela de Temporalidade e Vocabulário Controlado) em um sistema informatizado de gestão arquivística de documentos de arquivo.

6 CONCLUSÕES

A ideia foi a de submeter os documentos de arquivo, elaborados em meio digital, a um *software* de mineração de textos, que identificou os termos no documento a partir de um vocabulário controlado. Uma das etapas da pesquisa foi a construção desse vocabulário controlado vinculado à estrutura de classificação do Código de Classificação do Conarq. Dessa forma, quando o *software* encontrou os termos no documento, ele vinculou automaticamente a uma unidade de classificação. Com essa vinculação, será possível a determinação dos prazos de guarda desse documento. Assim, pretende-se ter, ao final do processamento, um documento de arquivo classificado, indexado e avaliado.

O protótipo foi testado em um repositório com quinhentos mil documentos (.doc, .pdf e HTML). As primeiras análises apontaram para resultados bastante satisfatórios. Os documentos foram indexados e receberam, automaticamente, um código de classificação. Entretanto, não foi possível resolver os problemas de sinonímia.

Entendemos que a grande contribuição do projeto de pesquisa em tela é a possibilidade de interrelacionar saberes de várias áreas do conhecimento, com o objetivo de um processamento técnico automatizado dos documentos de arquivo eficiente, eficaz e que gere segurança nos processos de organização, busca e avaliação desses registros documentais.

Além disso, o resultado pode criar as bases para uma metodologia para elaboração de vocabulário controlado para documentos de arquivo, o mapeamento da produção sobre classificação de documentos de arquivo e a definição de um modelo para vinculação automática entre classificação, avaliação e indexação de documentos de arquivo.

REFERÊNCIAS

ASSOCIAÇÃO DOS ARQUIVISTAS HOLANDESES. **Manual de arranjo e descrição de arquivos**. Rio de Janeiro: Arquivo Nacional, 1973.

BELLOTO, Heloísa Liberalli. **Arquivos permanentes: tratamento documental**. 2 ed. Rio de Janeiro: FGV, 2004.

BRASIL. ARQUIVO NACIONAL. **Classificação, temporalidade e destinação de documentos de arquivo relativos as atividades-meio da Administração Pública**. Rio de Janeiro: Arquivo Nacional, 2001.

BRASIL. e-ARQ Brasil: Modelo de Requisitos para Sistemas Informatizados de Gestão Arquivística de Documentos. Rio de Janeiro: Arquivo Nacional, 2011.

CAMPOS, Maria Luiza Almeida. **Em busca de princípios comuns na área de representação da informação: uma comparação entre o método de classificação facetada, o método de tesouro-baseado-em-conceito e a teoria geral da terminologia**. 1994. 207 f. Dissertação (Mestrado em Ciência da Informação). – Escola de Comunicação, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1994.

CONSELHO INTERNACIONAL DE ARQUIVOS. **Norma internacional de descrição arquivística**. Rio de Janeiro: Arquivo Nacional, 2001.

COUTURE, Carol. **Les fonctions de L'Archivistique contemporaine**. Quebec: Presses de l'Université du Québec, 1999.

DÍAZ RODRIGUEZ, Alfonso. La clasificación como proceso de gestión de documentos. **Tabula**, Salamanca, n. 13, p. 79-94, 2010.

DUCHEIN, Michel. O respeito aos fundos em Arquivística: princípios teóricos e problemas práticos. **Arquivo & Administração**, Rio de Janeiro, v. 10-14, n. 1, abr. 1986.

DURANTI, Luciana, EASTWOOD, Terry, MACNEIL, Heather. **Preservation of the integrity of electronic records**. Dordrecht: Kluwer Academic Publishers, 2003.

ESTEBAN NAVARRO, Miguel Angel. La representación y la organización del conocimiento en los archivos: los lenguajes documentales ante los procesos de clasificación, ordenación y descripción. In: MARCO, Francisco Javier García (ed.). **Organización del conocimiento em sistemas de información y documentación**. Zaragoza: Librería General, 1995. p. 65-90.

FOSCARINI, Fiorella. La clasificación de documentos basada en funciones: comparación de la teoría y la práctica. **Tabula**, Salamanca, n. 13, p. 41-58, 2010.

GAGNON-ARGUIN, Louise, BANNOURI, Rabii. La classification des documents institutionnels dans les ministères du gouvernement du Québec et le défi informatique. Montreal: Conseil du Trésor, 1998.

GIL, Fernando. Classificações. In: **Enciclopédia Einaudi**. Lisboa: Imprensa Nacional Casa da Moeda, 2000. p. 90-110.

HÉON, Gilles. Les plans de classification em archivistique. **Archives**, v. 27, n. 1, p. 73-90, 1995.

LANGRIDGE, Derek. **Classificação**. Abordagem para estudantes de Biblioteconomia. Rio de Janeiro: Interciência, 1977.

LIARD, L. **Lógica**. São Paulo: Companhia Editora Nacional, 1979.

LOPES, Luis Carlos. **A informação e os arquivos**. Teorias e práticas. Niterói: Eduff, 1996.

MARTÍN-POZUELO CAMPILLOS, Maria Paz. Dificultades para la concepción de una clasificación universal em Archivística. In: MARCO, Francisco Javier García (ed.). **Organización del conocimiento en sistemas de información y documentación**. Zaragoza: Librería General, 1995. p. 91-98.

PIEIDADE, M. A. Requião. **Introdução à Teoria da Classificação**. Rio de Janeiro: Interciência, 1983.

POMBO, Olga. Da classificação dos seres à classificação dos saberes. **Leituras**. Revista da Biblioteca Nacional de Lisboa, Lisboa, n. 2, p. 19-33, 1998.

ROBERGE, Michel. La classification universelle des documents administratifs. Québec: Documentor, 1985.

SILVA, Armando Malheiro *et al.* **Arquivística**. Teoria e prática de uma ciência da informação. Porto: Afrontamento, 1999.

SOUSA, Renato Tarciso Barbosa de. **Classificação Arquivística: trajetória e apropriação de um conceito**. 2005. Tese (Doutorado em História Social) – Universidade de São Paulo.

TESSITORE, Viviane. **Arranjo**: estrutura ou função? São Paulo: Arquivo do Estado, 1989.