

A classificação e a taxonomia como instrumentos efetivos para a recuperação da informação arquivística

Renato Tarciso Barbosa de Sousa

Professor da Universidade de Brasília (UnB). Doutor em História Social pela Universidade de São Paulo (USP).

Mestre em Biblioteconomia e Documentação pela Universidade de Brasília (UnB) – Brasília, DF - Brasil.

E-mail: renasou@unb.br

Rogério Henrique de Araújo Júnior

Professor da Universidade de Brasília (UnB). Doutor em Ciência da Informação pela Universidade de Brasília

(UnB). Mestre em Ciência da Informação pela Universidade de Brasília (UnB) - Brasília, DF - Brasil.

E-mail: araujojr@unb.br

Recebido em: 15/08/2014. Aprovado em: 27/08/2014. Publicado em: 25/3/2015.

Resumo

trata das possibilidades de uso da classificação de documentos de arquivo em seus objetivos básicos, manutenção do vínculo arquivístico, fundamentação da avaliação e da descrição, combinados às finalidades de uma linguagem documentária, ou seja, traduzir da linguagem natural dos documentos arquivísticos para uma linguagem documentária, que permite a condensação e síntese das informações em descritores (palavras-chave) que apoiem a recuperação da informação contida nos documentos de arquivo. Com o pressuposto de que as organizações necessitam de novas práticas de gestão da informação ancoradas em estratégias de recuperação da informação mais eficientes e eficazes, o objetivo da pesquisa é contextualizar e caracterizar o processo de classificação no âmbito da prática arquivística associando-o às funcionalidades de uma linguagem documentária que contribua para a melhoria do processo de busca e recuperação da informação. O resultado será a proposição de uma metodologia que associe e complemente a atividade de classificação, que abrange as áreas funcionais, aos temas dos documentos que deverão estar coligidos em descritores a serem selecionados e padronizados por meio de uma taxonomia. Desse modo, pretende-se firmar a concepção de classificação de documentos de arquivo, com o aporte da classificação dos temas dos mesmos documentos, trazendo novos pontos de acesso, benéficos à melhoria do processo de busca e recuperação da informação.

Palavras-chave: Classificação de documentos de arquivo. Taxonomia. Recuperação da informação.

Classification and taxonomy as effective instruments for archival information retrieval

Abstract

Deals with the possibilities of using the classification of records in its basic objectives, such as the maintenance of the archival bond, foundation evaluation and description, combined with the purposes of a documentary language, which means translating from natural language of the archival documents to a documentary language, which allows the condensation and synthesis of information in descriptors (keywords) that support retrieval of information contained in the records. Assuming that organizations need new information management practices, anchored in more efficient and effective information recovery strategies, this research contextualizes and characterizes the classification process within the archival practice, associating it with the features a documentary language that contributes on improving the research and retrieval process. The result will be the development of a methodology that joins and complements the classification activity, which covers the functional areas, the documents' subjects that should be collected in descriptors selected and standardized by means of a taxonomy. Thus, it is intended to establish the design of record classification, with the contribution of the subjects' classification of the same documents, bringing new access points, beneficial to the improvement of the search and information retrieval process.

Keywords: *Classification of records. Taxonomy. Information retrieval.*

Clasificación y taxonomía como instrumentos efectivos para recuperación de la información archivística

Resumen

Trata de las posibilidades de utilizar la clasificación de los documentos en sus objetivos básicos, el mantenimiento del vínculo archivístico, motivación de la evaluación y descripción, combinados a las finalidades de un lenguaje documental, o sea, traducir del lenguaje natural de los documentos archivísticos para un lenguaje documental, que permite la condensación y síntesis de la información en descriptores (palabras clave) que apoyen la recuperación de la información contenida en los documentos de archivo. Con la suposición de que las organizaciones necesitan nuevas formas de gestión de la información anclados en estrategias de recuperación de la información más eficientes y eficaces, el objetivo de la investigación es contextualizar y caracterizar el proceso de clasificación en el ámbito de la práctica archivística asociándolo con las funcionalidades de un lenguaje documental que contribuya para mejorar el proceso de búsqueda y recuperación de información. El resultado será la propuesta de una metodología que se una y complemente las actividades de clasificación, que abarca las áreas funcionales, los asuntos de documentos que deben recopilarse en descriptores seleccionados y estandarizados por medio de una taxonomía. Por lo tanto, la intención es de registrar el diseño de clasificación de documentos de archivo, con la contribución de la clasificación de los temas de los mismos documentos, trayendo nuevos puntos de acceso, benéficos para la mejora del proceso de búsqueda y recuperación de la información.

Palabras clave: Clasificación de registros. Taxonomía. La recuperación de información.

INTRODUÇÃO

Desafios constantes à busca e recuperação da informação, a organização e a representação do conhecimento têm sido cada vez mais considerados como solução para a representação de quantidades crescentes de acervos arquivísticos nas organizações. Esses acervos, apesar de imprescindíveis à gestão estratégica, vêm se constituindo em um problema pela falta de espaço físico para guarda de grandes quantidades de documentos em suporte papel e o alto custo do armazenamento daqueles em formato digital. O valor de manter espaços nas dependências das organizações, bem como nos seus sistemas informatizados, vem desequilibrando a relação entre custo e benefício da manutenção dos acervos arquivísticos, sem contar as despesas envolvidas na manutenção de pessoal especializado para o tratamento desses acervos, que elevam sobremaneira o dispêndio de recursos financeiros nas corporações.

Em contrapartida, a tecnologia da informação tem possibilitado importantes avanços nas áreas que tratam de grandes volumes de documentos, sobretudo enfatizando o processo de busca e recuperação da informação arquivística. Entretanto, ela, em alguns momentos, é entendida como solução definitiva de todo o problema, mas que, na realidade, é mais um valioso ator no âmbito da gestão documental.

A classificação de documentos de arquivo tem

três objetivos: manter o vínculo arquivístico, fundamentar a avaliação e a descrição e possibilitar a recuperação da informação contida nos documentos de arquivo. Diversos autores, tais como Foscarini (2010), chegam a entender que a recuperação dos documentos armazenados é somente um benefício colateral da classificação, sendo seu propósito básico compor os documentos individuais nos conjuntos a que pertencem, baseando-se no mandato e nas funções do criador, como defendem Duranti, Eastwood e Macneil (2003). Uma das características principais do documento de arquivo é a relação que os documentos mantêm entre si e que precisa ser preservada para dar-lhes sentido, bem como as atividades ou ações que lhes deram origem. A compreensão e a manutenção desse contexto, representadas pelas 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 descrição, a aquisição e a difusão (disseminação), em que a tríade formada pela avaliação, descrição e classificação são as atividades capitais da prática arquivística. Podemos considerar a classificação como uma função matricial, pois

é por meio dela que se realiza a avaliação, que em uma etapa posterior, vai possibilitar a gestão dos prazos de guarda e da destinação final (tabela de temporalidade), independentemente da metodologia aplicada e de sua configuração, sempre tem como ponto de partida os conjuntos documentais definidos na classificação. A partir daí inicia-se o processo de avaliação, tendo como ponto partida a classificação. No caso da descrição, a norma do Conselho Internacional de Arquivos, ISAD(G) (2011) e a Norma Brasileira de Descrição Arquivística do Conselho Nacional de Arquivos, (Nobrade) (2006), definem como primeiro elemento a ser descrito o conjunto documental, representado pelo fundo, série ou subsérie, isto é, os agrupamentos documentais resultados do processo de classificação. Assim, a classificação ganha uma envergadura que sustenta toda a construção da prática arquivística. Assim, definir com precisão os conjuntos documentais e as suas relações é imprescindível para o que-fazer arquivístico (SOUSA, 2007).

A classificação, como operação matricial de todo o trabalho arquivístico, é uma etapa importante para a transparência e o compartilhamento das informações, que por sua vez, são caminhos seguros para o processo de tomada de decisão para a preservação da memória técnico-administrativa das organizações, além de permitir o pleno exercício da cidadania. Nesse instante a classificação se junta à atividade de indexação de documentos, por meio da construção de linguagens documentárias que apoiem o estabelecimento das relações entre os termos e os conteúdos de que tratam os documentos, espelhando suas funções/atividades e conduzindo os usuários dos sistemas de arquivo à recuperação da informação.

O acesso, consubstanciado pela recuperação da informação, deve ser considerado sob dois níveis: o legal (*corpus* de leis) e o técnico-científico. No nível legal, a entrada em vigor da Lei 12.527/2011 (BRASIL, 2011) alterou as condições e regras de acesso à informação 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 tudo aquilo que

pode colocar em risco a segurança do Estado e da sociedade, além da privacidade das pessoas.

Entretanto, o maior entrave ao pleno acesso à informação não é o conjunto de restrições impostas pela lei, mas as condições inadequadas de organização da informação pública, que não permitem o estabelecimento de um efetivo processo de busca e recuperação da informação. Atualmente, a coleta, o tratamento e a análise da grande quantidade de dados disponíveis nos arquivos públicos e privados são um obstáculo ao acesso à informação, que acaba por impactar na qualidade da decisão dos gestores. Tal situação é agravada por conta dos desafios tecnológicos que hoje se colocam, dentre eles o advento do documento digital, a profusão de bases de dados e a dispersão de informações.

Além dos desafios tecnológicos, deve-se ter em conta também o despreparo dos gestores públicos nos aspectos relacionados à cultura da organização da informação, tendo como consequência direta a análise superficial para subsidiar os processos de avaliação e de tomada de decisão.

Partimos do pressuposto que as organizações apresentam, cada vez mais, novas práticas informacionais, que requerem estratégias de recuperação da informação mais eficientes e eficazes. Portanto, é difícil ancorar a recuperação da informação contida nos documentos de arquivo apenas no instrumento de classificação. Precisamos incorporar nesse esforço novos instrumentos que possibilitem a recuperação da informação de modo compatível com as demandas informacionais atuais.

Assim, o presente trabalho visa contextualizar e caracterizar o processo de classificação no âmbito da prática arquivística, bem como discutir os requisitos básicos para a concepção de uma taxonomia, entendida como uma linguagem documentária, associada à classificação de documentos de arquivo com a finalidade de aperfeiçoar o processo de busca e recuperação da informação em arquivos.

A CLASSIFICAÇÃO NO ÂMBITO DA GESTÃO DE DOCUMENTOS

A classificação é parte integrante de um programa de gestão de documentos, e de acordo com Sousa e Araújo Júnior (2013), refere-se a três aspectos básicos:

- organização de documentos em um arquivo, a partir de um plano de classificação, um código de classificação ou de um quadro de arranjo;
- análise do conteúdo dos documentos visando identificar a categoria de assuntos (rótulos), por meio do processo de indexação e com a atribuição de códigos de classificação que servirão para sua posterior recuperação em um sistema arquivístico; e
- determinação de graus de sigilosidade de documentos ou do conteúdo de documentos, conforme a legislação vigente.

A Norma ISO 15.489/2001, que trata da gestão de documentos, define sete processos de gestão, dentre eles o de organização lógica para a recuperação, ou seja, classificação e indexação. A norma internacional e o Modelo de Requisitos para Sistemas Informatizados de Gestão Arquivística de Documentos do Conselho Nacional de Arquivos, E-Arq Brasil (2011), apontam como instrumento obrigatório de gestão de documentos o plano de classificação de documentos.

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

A classificação aparece 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 de trabalha-se

com classificação sem agregar o desenvolvimento desse conceito proporcionado pela Filosofia e pela Teoria da Classificação. Apresenta-se o caminho sem discutir os meios necessários para percorrê-lo. Adotam-se, sem qualquer avaliação crítica, 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 ser uma marca da área.

A organização dos documentos tem sido um dos maiores desafios da Arquivística contemporânea. Lidar com massas tão grandes de documentos acumulados diariamente pelas instituições exige 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 se estruturam e como executam suas funções e atividades é compreender como os documentos são acumulados, como devem ser classificados e indexados, visando sua posterior recuperação (SOUSA, 2007).

O documento 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 pela mesma finalidade com que são produzidos e recebidos, determinados

pela razão de sua elaboração e que é necessário à própria existência e a capacidade de cumprir seus objetivos, por isso formam um conjunto indivisível de relações intelectuais e orgânicas.

A explosão do volume de documentos arquivísticos ocorreu de maneira paralela com o aumento geral da informação registrada pelas sociedades. 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 (SOUSA, 2007).

A CLASSIFICAÇÃO E OS PRINCÍPIOS ARQUIVÍSTICOS

O conceito de classificação não é estranho à literatura arquivística, pelo contrário. Os grandes manuais sempre dedicaram capítulos para tratar do tema. Todavia, a presença não garantiu uma verticalização teórica sobre o assunto. 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 importante 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) 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 disseminação da informação escrita.

Duas questões surgem a partir dessa disposição. A primeira diz respeito à preocupação teórica, pois, como alertam diversos autores, tal como Kobashi (1996), a organização e a representação da informação, com fins documentários, é fato relativamente recente, isto é, os elementos estão em construção. A segunda refere-se ao caráter específico do documento arquivístico. São os aspectos limitadores, mas, sobretudo, desafiadores.

Parte-se da ideia 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.

Não se deve perder de vista 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 interpretação que possibilite a manutenção da ordem original, desde que ela represente a própria gênese documental e as relações orgânicas estabelecidas pelos documentos. Preservar a ordem original significa manter o documento e suas relações orgânicas no contexto das atividades que o geraram.

Tomemos como exemplo a aplicação do princípio de proveniência e de seus desdobramentos na realidade arquivística brasileira. Ela é dificultada, sobremaneira, pela desorganização dos arquivos na origem (arquivos dispostos 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. 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 dessa estruturação teríamos o desenho do organograma, que é a representação gráfica da estrutura organizacional e que indica a forma como o poder administrativo é dividido, delimitando os níveis de autoridade.

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 deste trabalho, 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.

Nos fundamentos teóricos da classificação, classificar significa distribuir indivíduos em grupos distintos, de acordo com características semelhantes e diferenciadoras. 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; já no segundo caso, o princípio será o natural.

A classificação baseia-se em princípios de divisão ou de classificação artificiais, não se prendendo à essência dos objetos distribuídos. 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. Contudo, a classificação que parte de princípios de divisão ou de classificação naturais encontra e exprime a ordem seguida pela natureza no meio

das diferenças quase infinitas dos seres. Ela parte da observação dos indivíduos e, em seguida, compara-os por meio de características comuns em certo número de indivíduos, eliminando as diferenças, conservando as analogias e com elas constituindo tipos de variáveis de extensão.

Podemos então, 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. O primeiro princípio de divisão a ser utilizado é o da proveniência, que permite 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.

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. 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 deve ser o procedimento padrão a ser adotados na gestão documental.

O PAPEL DA TECNOLOGIA NA CLASSIFICAÇÃO DE DOCUMENTOS

O uso cada vez mais frequente da tecnologia da informação possibilita, 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 temática do registro documental, de qual assunto trata o documento.

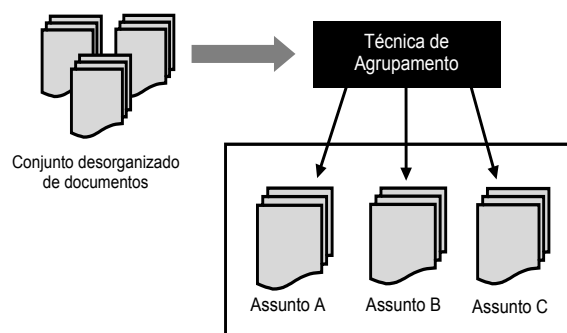
A análise do conteúdo desse campo nos registros de sistemas de protocolo informatizados indicou total falta de padrão para o preenchimento do mesmo. É impossível imaginar que uma organização, por mais complexa que seja, tenha uma variedade tão expressiva de assuntos, fato que se verifica na prática e pode ser atribuído à total falta de critérios de indexação, agravado pela inexistência de qualquer linguagem documentária para apoiar a padronização dos descritores selecionados. Os ganhos de eficiência na recuperação da informação, que poderiam junto à classificação constituir-se em pontos de acesso à informação demandada pelos usuários, perdem-se em um emaranhado de termos vazios, fragmentados, desconexos, 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 de assunto.

O relacionamento de uma linguagem documentária, bem como das suas funcionalidades com um instrumento de classificação, enseja uma metodologia para a construção dessa linguagem vinculada às unidades de classificação encontradas na estrutura do instrumento de classificação.

A proposta consiste na montagem de um protótipo, a partir de um *software* de mineração de textos - que pode ser conceituado, conforme propõe Araújo Júnior (2007), como uma ferramenta capaz de sumarizar um conjunto de documentos em agrupamentos - a ser aplicado à lista de assuntos de um sistema de registro e controle do trâmite documental.

A figura 1 apresenta, de modo esquemático, o objetivo do agrupamento de informações textuais.

Figura 1 – Objetivo do agrupamento de informações textuais



Fonte: Wives, 1999 apud Araújo Júnior, 2007.

A finalidade do protótipo é gerar uma lista com os termos que mais ocorrem. Serão introduzidos os procedimentos micro de controle de vocabulário, que tem por objetivo o controle dos termos ou expressões. O controle de vocabulário, inicialmente desenvolvido pela documentação, foi introduzido na arquivística especialmente em razão da informatização crescente dos arquivos por meio da elaboração de listas de assuntos para protocolos informatizados. Trata-se de um recurso para organizar e possibilitar a recuperação dos documentos e informações com consistência, gerando, conseqüentemente, confiança no sistema e no usuário.

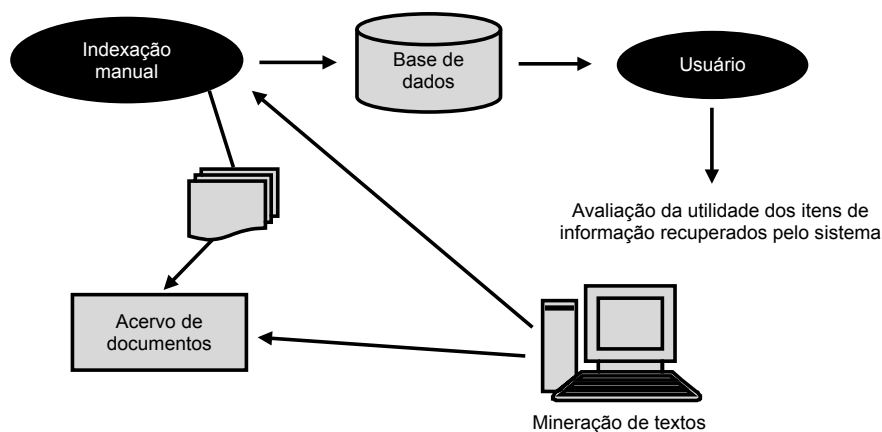
É, também, um mecanismo para aumentar o grau de precisão na recuperação das informações arquivísticas.

A precisão é um conceito importante para a avaliação da qualidade da recuperação da informação e pode ser entendida como a extensão com a qual itens de recuperados em uma base de dados são considerados úteis por quem demandou a pesquisa

(LANCASTER, 1998). A medição do grau de precisão foi proposta por Cleverdon (1962).

A figura 2 apresenta um exemplo da sistematização dos termos gerados a partir de mineração de textos para complementar o processo de indexação, com vistas a aumentar a precisão de resposta no processo de busca e recuperação da informação.

Figura 2 – Posição da mineração de textos no contexto da indexação manual



Fonte: Araújo Júnior, 2007.

A situação ideal que deve ser desenvolvida é remeter o controle de vocabulário ao plano de classificação e este à tabela de temporalidade. Essa solução deve contar, ainda, com a interpretação de arquivistas para funcionar em toda a sua plenitude.

Com a presença cada vez mais significativa dos documentos digitais nas instituições, avançamos nossa pesquisa para o processamento automático, que seja capaz de vincular diretamente classificação e avaliação de documentos. A ideia é submeter um repositório de documentos digitais a um *software* de mineração de textos. Essa mineração deve ser realizada com o apoio de uma linguagem documentária, por meio de um vocabulário controlado. A identificação dos termos desse vocabulário estará vinculada à 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 passível de ser gerenciado arquivisticamente.

A base de dados onde a representação do conteúdo dos documentos estará armazenada, o vocabulário do sistema, ou seja, o vocabulário controlado associado à estrutura de classificação, influenciará as estratégias doravante utilizadas na busca e recuperação dos documentos de arquivo.

UMA LINGUAGEM DOCUMENTÁRIA PARA A CLASSIFICAÇÃO

As linguagens documentárias estão no âmbito da representação do conteúdo dos documentos e destinam-se a expressar ou traduzir o conteúdo semântico dos documentos em uma linguagem artificial/especializada, normalmente sintética e/

ou analítica, a fim de recuperar os documentos pertinentes em resposta às consultas que tratam deste conteúdo, que por sua vez é operacionalizada por meio da análise documentária. A linguagem documentária é o contraponto da linguagem natural que é dada pelo discurso comum ou a linguagem usada habitualmente na escrita e na fala.

De modo geral, as linguagens naturais dificultam o tratamento da informação, causando problemas diversos relativos à imprecisão de termos, polissemia, considerações ambíguas sobre conceitos que impõem obstáculos à armazenagem e recuperação da informação. Logo, seu uso depende uma tradução documentária, ou seja, a utilização de linguagens documentárias para mediar seu armazenamento e recuperação. Por isso, a escolha aleatória de termos não produz qualquer efeito de descrição dos conteúdos dos documentos.

As linguagens documentárias são normalmente utilizadas em sistemas de informação ou unidades de informação para descrever o conteúdo dos documentos, a fim de prepará-los para o armazenamento e a posterior recuperação das informações que contêm.

Segundo Guinchat e Menou (1994), os elementos que compõem as linguagens documentárias são as palavras que descrevem as informações ou descritores. São extraídas da linguagem natural e reduzidas a uma configuração gramatical única e invariável (geralmente substantivo singular), palavras da linguagem natural relacionadas ao descritor correspondente por meio de uma remissiva e relações entre os descritores, tais como relações hierárquicas, de equivalência ou vizinhança.

A análise documentária se dá pela atividade da descrição de conteúdo, que é o conjunto de operações que descreve os assuntos de um documento e os produtos destas operações. Entre os seus objetivos, podemos destacar:

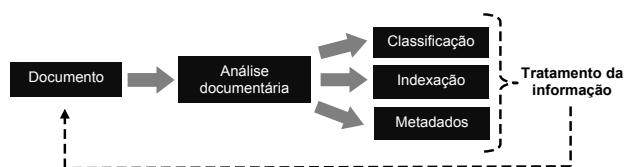
- conhecer o conteúdo dos documentos para informar os usuários;

- operar escolhas que levem a eliminar ou a conservar um documento;
- determinar sua forma e nível de tratamento; e
- estabelecer categorias para o armazenamento dos documentos (GUINCHAT e MENO, 1994).

Esta análise possibilita a extração de palavras-chave ou indexadores que passam a ser os rótulos de informação de cada documento e que podem, conforme a proposta deste trabalho, complementar a classificação de documentos de arquivo, a fim de tornar mais efetiva a busca e recuperação dos mesmos.

A figura 3, a seguir, apresenta esquematicamente o papel da análise documentária no tratamento da informação.

Figura 3 – A análise documentária no tratamento da informação



Fonte: baseado em Rowley, 1988 apud Araújo Júnior, 2007.

Segundo Araújo Júnior (2007), a representação do conteúdo dos documentos está na base do conceito de indexação, bem como de conceitos relacionados, tais como descritor, linguagem de indexação e termo de indexação. No caso da indexação, a identificação, seleção e análise dos documentos são partes precípuas de todo o processo.

Desse modo, podemos considerar os conceitos clássicos de indexação propostos por Cintra (1983), tradução de um documento em termos documentários, ou seja, em descritores que têm por função expressar o conteúdo dos documentos. E, também, a proposição de Robredo (2005), processo pelo qual se identificam os conceitos de que trata o documento, expressando-os na terminologia usada pelo autor (linguagem natural) ou com o apoio de códigos (linguagem documentária, descritores, sistemas de classificação, etc.).

Todavia, a escolha das palavras-chave ou descritores durante o processo de indexação deve estar apoiada no uso de linguagens documentárias, tais como tesouros, listas de palavras-chave, listas de descritores livres ou quaisquer outros vocabulários controlados, bem como no conhecimento que o profissional da informação tem acerca do acervo documental a ser indexado. Esta dinâmica permite a padronização dos descritores a serem usados na representação do conteúdo dos documentos, já que a indexação preconiza três etapas na tarefa de indexar, conforme propõe Rowley (1988):

Figura 4 – Etapas do processo de indexação



Fonte: baseado em Rowley, 1988 apud Araújo Júnior, 2007.

Nesse momento, a associação entre a atividade de classificação de documentos de arquivo e a possibilidade de sua indexação esbarra na falta de uma linguagem documentária adequada para expressar os termos típicos presentes em documentos arquivísticos, que possuem como característica a representação das atividades administrativas e funcionais de uma organização.

Tanto a estrutura hierárquica, os níveis de detalhamento, quanto as características dos termos em um tesouro estão mais voltados para aplicações em acervos bibliográficos; o mesmo ocorre com vocabulários controlados em geral, pois não espelham as estruturas funcionais de uma organização, tema central de toda a documentação de arquivo.

Assim, a linguagem documentária que mais se aproxima de uma classificação factível e representativa dos documentos aponta na direção das taxonomias, justamente pelo seu caráter hierárquico mais simples e flexível do que os tradicionais tesouros.

A taxonomia pode ser considerada como uma estrutura de termos organizados em camadas, por meio de uma hierarquia simples que pode ser detalhada conforme as necessidades de classificação e indexação dos

documentos de arquivo. Ela possibilita, de modo abrangente, classificar objetos, seres vivos, coleções de documentos em grupos ordenados hierarquicamente, permitindo sua identificação, localização e acesso. Na seção seguinte apresentamos uma metodologia de construção de taxonomias associadas à estrutura de classificação de documentos de arquivo, a fim de estabelecer novos pontos de acesso que beneficiem a busca e a recuperação da informação arquivística.

TAXONOMIA COMO INSTRUMENTO PARA CLASSIFICAÇÃO DE DOCUMENTOS DE ARQUIVO

A taxonomia pode ser definida como sistema de classificação que apoia o acesso à informação, permitindo classificar, alocar, recuperar e comunicar informações em um sistema de maneira lógica.

A taxonomia, vinculada à classificação de documentos, tem como finalidade estruturar e sistematizar o conjunto de assuntos correspondentes às atividades desempenhadas pelas áreas de trabalho de uma dada organização, a fim de garantir aos tomadores de decisão e usuários em geral, o acesso mais rápido aos documentos e às informações necessárias ao processo decisório da instituição.

Nesse sentido, a construção da taxonomia envolve o processo de representação do conhecimento criado e compartilhado nas unidades de uma organização que serve de panorama geral para a determinação dos conceitos e temas tratados na instituição. O conhecimento básico dos processos, da cultura organizacional, bem como da linguagem técnica empregada no cotidiano empresarial, são insumos valiosos, pois possibilitam a adaptação e validação das versões da taxonomia a ser construída.

Esse processo visa organizar a lista estruturada de termos, de modo a apoiar a organização e recuperação da informação, além de apresentar um mapa conceitual dos tópicos usados no registro do conhecimento na organização tendo como base o plano de classificação de documentos, que é um insumo importante para construção da linguagem documentária.

Outro aspecto que deve ficar claro na estruturação da taxonomia são as várias versões e atualizações do instrumento que devem ser viabilizadas por uma comissão que aprove a inclusão ou exclusão de termos. Além disso, ela deve complementar a finalidade desse tipo de linguagem documentária, ou seja, aperfeiçoar o compartilhamento, a recuperação e a disseminação da informação, por meio da compreensão e do registro da terminologia utilizada na organização.

A taxonomia, no contexto da organização e gestão da informação, deve ser acoplada ao plano de classificação para complementar a classificação sob duas dimensões:

- classificação do documento na perspectiva das áreas funcionais da organização; e
- classificação na perspectiva dos conteúdos de que tratam os documentos utilizando o protocolo de indexação a ser definido em uma política específica.

A taxonomia se destina a inúmeras finalidades, todas elas objetivando classificar e organizar. Quando este sistema de classificação é associado à organização da informação na organização, suas aplicações podem ser enumeradas nas seguintes ações:

- descrição dos conteúdos informacionais;
- definição de padrões para tratamento técnico dos conteúdos informacionais;
- melhoria dos sistemas de recuperação da informação;
- melhoria da interface de navegação dos sistemas de informação; e
- localização e reuso da informação disponível na instituição.

De acordo com Bräscher e Carlan (2010), as taxonomias adquirem importância no contexto das organizações como instrumentos auxiliares à gestão do conhecimento e à organização e recuperação da informação. Elas garantem que todas as aplicações da instituição utilizem a mesma linguagem para organizar,

armazenar e apresentar a informação, facilitando e permitindo o acesso e a recuperação da informação.

METODOLOGIA DE CONSTRUÇÃO E USO DA TAXONOMIA

A elaboração da taxonomia deve ser realizada em três etapas metodológicas:

Etapa 1: consiste no levantamento das áreas do conhecimento e das relações hierárquicas de termos que integraram a primeira versão da taxonomia. Desta etapa, dois elementos devem ser disponibilizados, a lista de termos que representam a versão do primeiro nível de termos que integrará a versão preliminar do instrumento e o quadro com as relações hierárquicas de termos;

Etapa 2: consiste na elaboração da versão preliminar da taxonomia. Após a definição do primeiro nível de termos, de caráter geral e abrangente, passa-se para a identificação da necessidade de especificar alguns dos termos que oferecem mais de uma opção de aplicabilidade no campo semântico. Assim, alguns deles apresentam subtermos classificados em níveis de acordo com sua especificidade.

Essa configuração facilita a construção da primeira versão da taxonomia, pois apresenta uma versão compatível com a hierarquização de termos. Assim, os termos do primeiro nível podem ser considerados como os termos gerais.

Etapa 3: consiste na validação, com os profissionais envolvidos na administração da instituição, da versão preliminar da taxonomia. Para tanto, devem ser realizadas reuniões de validação dos termos e hierarquia da versão definitiva da taxonomia. Nas reuniões a validação deve realizada a partir do conhecimento acumulado pelos especialistas sobre as áreas de trabalho da instituição, bem como em material bibliográfico e no código de classificação de documentos. O processo de validação deve contar com o conhecimento técnico e a experiência dos especialistas sobre as diversas áreas de trabalho da organização, por meio do preenchimento de um instrumento de coleta de dados.

A elaboração da taxonomia se dará por meio de um número previamente determinado de termos em que a relação hierárquica será definida do primeiro nível em relação aos demais. Para a seleção dos termos devem ser usados alguns documentos importantes:

- nomenclatura das áreas funcionais da organização, bem como o seu organograma;
- regimentos internos de cada área funcional descritos no organograma da instituição;
- normas, portarias e regulamentação das atividades do órgão; e
- plano de classificação de documentos de arquivo da instituição.

A criação dos níveis hierárquicos de termos deve também refletir a preocupação em detalhar os grandes temas e temas específicos do âmbito de atuação da organização, as áreas funcionais com os respectivos macroprocessos, processos de apoio e processos finalísticos, assim como a natureza de classificação dos seus documentos de arquivo. Logo, a criação dos níveis da taxonomia deve ser ponderada por meio das seguintes diretrizes:

- a) flexibilidade de uso dos termos, com a oferta adequada de pontos de acesso, refletindo as áreas funcionais, temas relacionados às atividades de apoio e finalísticas, bem como a nomenclatura utilizada na classificação dos documentos da instituição;
- b) normalização e padronização dos descritores a serem empregados na indexação dos documentos;
- c) uso da taxonomia em conjunto com o código de classificação de documentos, objetivando criar mais pontos de acesso aos documentos, providência que facilitar a busca e recuperação dos documentos; e
- d) instrumento complementar às funções e aplicações do código de classificação de documentos da instituição.

Finalmente, a taxonomia deverá ser usada como instrumento complementar à atividade de

classificação de documentos, devendo ser pautada pelos seguintes passos:

1. classificar o documento conforme o plano de classificação de documentos de arquivo;
2. identificar no primeiro ou no segundo nível de termos da taxonomia a área utilizada na classificação dos documentos;
3. utilizar os termos dos diversos níveis da taxonomia como descritores (termos de indexação) em complemento à classificação dos documentos; e
4. verificar a coerência da classificação combinada com a indexação nas dimensões funcionais e temáticas, respectivamente com os documentos.

Como regra geral para a classificação e a descrição dos documentos, a seleção dos descritores da taxonomia deve ser combinada com os diferentes níveis de termos, desde que seja preservada a unidade temática combinada com o código de classificação atribuído a cada um dos documentos processados.

Em um modelo de aplicação da taxonomia combinada à classificação dos documentos de arquivo, o instrumento deve trazer um aporte de termos de indexação para complementar a classificação dos documentos, a fim de criar pontos de acesso adicionais para a recuperação dos conjuntos documentais.

CONSIDERAÇÕES FINAIS

Entendemos a organização dos documentos como parte da função social do arquivista. A organização se dá para possibilitar o acesso rápido, seguro e eficiente à informação. No primeiro momento, para a correta tomada 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, na sequência, para preservação do capital informacional que permitirá o uso além das fronteiras da criação do próprio documento. Nesse sentido, a classificação assume posição privilegiada e aparece como ferramenta importante para a consecução do trabalho, garantindo qualidade e

os fundamentos para outras incursões no universo arquivístico, tais como a avaliação e a descrição.

A possibilidade do estabelecimento de um novo marco referencial para a organização de documentos de arquivo aponta para um processo de inovação constante, no qual os fundamentos da organização de documentos de arquivo estão diretamente associados com a organização e representação da informação e do conhecimento, perfazendo não adição simples de saberes, mas uma articulação a partir dos limites dos objetivos propostos.

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 e, do mesmo modo, garantir pontos de acesso aos acervos de modo a aprimorar continuamente o processo de busca e recuperação da informação, visando beneficiar os gestores nas tomadas de decisão.

REFERÊNCIAS

- ARAÚJO JÚNIOR, R.H. de. *Precisão no processo de busca e recuperação da informação*. Brasília: Thesaurus, 2007.
- BRÄSCHER, M.; CARLAN, E. Sistemas de organização do conhecimento: antigas e novas linguagens. In: ROBREDO, J.; BRÄSCHER, M. (Org.). *Passeios no bosque da informação: estudos sobre representação e organização da informação e do conhecimento*. Brasília: Ibict, 2010. cap. 8, p.147-176. Disponível em: <www2.senado.leg.br/bdsf/item/id/189812>. Acesso em: 06 ago. 2014.
- BRASIL. CONSELHO NACIONAL DE ARQUIVOS. *e-ARQ Brasil: modelo de requisitos para sistemas informatizados de gestão arquivística de documentos/câmara técnica de documentos eletrônicos*. Rio de Janeiro: Arquivo Nacional, 2011.
- _____. *NOBRAD E: Norma Brasileira de Descrição Arquivística*. Rio de Janeiro: Arquivo Nacional, 2006.
- BRASIL. Lei n. 12.527, de 18 de novembro de 2011. Regula o acesso a informações previsto no inciso XXXIII do art. 5º, no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal... *Diário Oficial da União*, Poder Executivo, Brasília, DF, 18 nov. 2011. Seção 1, Edição Extra, p. 1.
- CINTRA, A.M.M. Elementos de linguística para estudos de indexação. *Ciência da Informação*, v. 19, n.1, p. 131-136, jan./abril, 1983.
- CLEVERDON, C.W. *Report on testing and analysis of investigation into comparative efficiency of indexing systems*. Cranfield: Aslib, 1962.
- CONSELHO INTERNACIONAL DE ARQUIVOS. *ISAD(G): Norma geral internacional de descrição arquivística*. 2. ed. adotada pelo Comitê de Normas de Descrição, Estocolmo, Suécia, 19-22 de setembro de 1999, versão final aprovada pelo CIA. Rio de Janeiro: Arquivo Nacional, 2001.
- DURANTI, L.; EASTWOOD, T.; MACNEIL, H. *Preservation of the integrity of electronic records*. Dordrecht: Kluwer Academic Publishers, 2003.
- FOSCARINI, F. La clasificación de documentos basada en funciones: comparación de la teoría y la práctica. *Tabula*, Salamanca, n. 13, p. 41-58, 2010.
- GUINCHAT, C.; MENO, M. *As linguagens documentária: introdução geral às ciências e técnicas da informação e documentação*. 2. ed. rev. e ampl. Brasília: Ibict, 1994. p.133-169.
- ISO. International Organization for Standardization. *ISO 15489-1: 2001; Information and documentation – records management. Part 1: General*. [S. l., 2001]. Disponível em: <<http://www.iso.org>>. Acesso em: 30 abr. 2014.
- KOBASHI, N. Y. Análise documentária e representação da informação. *Informare*, Rio de Janeiro, v. 2, n. 2, p. 5-27, jul./dez. 1996.
- LANCASTER, F.W. *Indexing and abstracting in theory and practice*. 2nd ed. London: Library Association, 1998.
- ROBREDO, J. *Documentação de hoje e de amanhã: uma abordagem revisitada e contemporânea da ciência da informação e de suas aplicações biblioteconômicas, documentárias, arquivísticas e museológicas*. Brasília: Edição do autor, 2005.
- ROWLEY, J. *Abstracting and indexing*. 2nd ed. London: Clive Bingley, 1988.
- SILVA, A.M. et al. *Arquivística: teoria e prática de uma ciência da informação*. Porto: Afrontamento, 1999.
- SOUSA, R.T.B. de S.; ARAÚJO JÚNIOR, R. H. de. (Org.). *A produção científica da arquivologia em classificação, descrição e recuperação: o estado da arte*. Rio de Janeiro: Associação dos Arquivistas Brasileiros, 2013.
- SOUSA, R.T.B. de. A classificação como função matricial do que-fazer arquivístico. In: SANTOS, V.; INNARELI, H.; SOUSA, R. *Arquivística: temas contemporâneos*. Brasília: SENAC, 2007. p.79-162.
- WIVES, L.K. *Estudo sobre agrupamento de documentos textuais em processamento de informação não estruturadas usando a técnica de clustering*. 1999. Dissertação (Mestrado em Ciência da Computação) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 1999.