

A INDEXAÇÃO NA RECUPERAÇÃO DA INFORMAÇÃO EM ARQUIVOS: UMA ABORDAGEM INICIAL

Mariane Costa Pinto

Bibliotecária da Coordenação de Documentos Audiovisuais e
Cartográficos do Arquivo Nacional
E-mail: mariane@arquivonacional.gov.br

Resumo: A indexação é a representação do conteúdo dos documentos por meio da transcrição de uma linguagem natural em linguagem documentária. Para o desenvolvimento da atividade, deve-se elaborar procedimentos de indexação nos quais se defina quais critérios serão utilizados pelo corpo indexador, levando-se em consideração a infraestrutura da organização, seus recursos humanos e seus usuários. Os instrumentos resultantes do processo de indexação podem ser listas de cabeçalhos de assuntos, tesouros ou vocabulários controlados. Esse controle implica diretamente na organização de acervos, já que nomeia de forma consistente os pontos de acesso que levam à informação dos documentos. Apesar de normalmente realizada em documentos bibliográficos, a indexação adquiriu importância na arquivologia a partir do momento em que a documentação arquivística adquiriu diversidade de gênero e conteúdo. Para uma indexação consistente e aplicável aos documentos arquivísticos, alguns procedimentos precisam ser adotados, a fim de evitar repetição de informações e resultados ineficientes.

Palavras-chave: Recuperação da Informação. Política de Indexação. Controle de Vocabulário. Documentos Arquivísticos.



1 INTRODUÇÃO

Indexar é construir representações do conteúdo do documento de forma que sua inclusão possa ser recuperada em algum tipo de base de dados. As suas etapas geralmente compreendem o conhecimento prévio do documento, a determinação de seu tema principal, a identificação dos elementos do conteúdo que devem ser descritos e a extração dos termos correspondentes, a verificação da pertinência dos termos escolhidos, a tradução dos termos da linguagem natural para termos correspondentes da linguagem documentária, a verificação

da pertinência da descrição e a formalização da descrição, quando o sistema prevê regras especiais de apresentação ou de escrita (LANCASTER, 2004).

A indexação pode acontecer de duas maneiras: por derivação ou extração – palavras ou expressões que estão/ocorrem no documento são selecionadas para representar o seu conteúdo temático – e por atribuição, que envolve a atribuição de termos ao documento a partir de uma fonte que não é o próprio. A escolha da linguagem a ser utilizada é essencial para o desenvolvimento de padrões.

A linguagem de indexação afeta o desempenho de um sistema de recuperação de informação tanto na estratégia de busca (estabelece a precisão com que o técnico de busca pode descrever os interesses do usuário) quanto na indexação (estabelece a precisão com que o indexador pode descrever o assunto do documento). (FUJITA; RUBI, 2006, p. 52)

A partir daí, outras decisões são tomadas: se o sistema a ser escolhido será a pré-coordenação, no qual os termos já estão combinados no momento da indexação, como nos vocabulários controlados, nas listas de cabeçalhos de assunto e nas classificações bibliográficas, ou a pós-coordenação, no qual os termos são combinados no momento da busca; se o critério for a exaustividade, que implica em uma maior quantidade de termos para descrever o documento, ou da especificidade, em que se opta pelo uso de termos mais específicos para representar o conteúdo do documento.

Todas as escolhas têm como objetivo a recuperação da informação com vistas à sua revocação (quantidade de documentos recuperados) e à precisão (quantidade de documentos úteis encontrados).

A definição desses fatores leva à criação de uma política de indexação, responsável por normalizar os procedimentos de descrição de conteúdo de uma instituição.

De acordo com Carneiro (1985, p. 221), uma política de

indexação,

[...] deve servir como um guia para tomada de decisões, deve levar em conta os seguintes fatores: características e objetivos da organização, determinantes do tipo de serviço a ser oferecido; identificação dos usuários, para atendimento de suas necessidades de informação e recursos humanos, materiais e financeiros, que delimitam o funcionamento de um sistema de recuperação de informações.

A política ajuda a padronizar o processo de indexação, facilitando sua aplicação por vários indexadores de uma mesma instituição.

2 A RECUPERAÇÃO DA INFORMAÇÃO

A organização é peça chave para a recuperação das informações.

Os documentos não têm inimigos externos contra os quais devem se proteger, mas sua acumulação acarreta o surgimento de inimigos internos: a falta de organização ou a desorganização do arquivo e a conseqüente dificuldade ou impossibilidade de nele encontrar os documentos procurados. Para que um arquivo, qualquer um, preencha sua missão de ser “o arsenal da administração e o celeiro da história”¹, pressupõe-se a sua organização. Organizar, segundo o dicionário, significa “dispor de forma ordenada; dar forma regular às partes de um todo; prover de ou adquirir certa estrutura”², ou seja, organizar significa, forçosamente, categorizar, incluir, em função de algum critério, determinados elementos em uma classe e, nesse mesmo movimento, excluir outros elementos. (SMIT; KOBASHI, 2003, p. 11)

¹Expressão do arquivista francês Charles Braibant, transcrita por Heloísa Bellotto em *Arquivística: objetos, princípios e rumos*. São Paulo: Associação de Arquivistas de São Paulo, 2002. (Scripta, 1), p. 9.

²Dicionário Houaiss da língua portuguesa. Rio de Janeiro: Objetiva, 2001.

A indexação, como representação dos conceitos contidos no documento e transcrição dos conceitos em uma linguagem documentária, dá origem às classificações, tesauros, vocabulários controlados, listas de cabeçalho de assuntos (CHAUMIER, 1988).

O vocabulário controlado é, essencialmente, uma lista de termos autorizados. Sua importância se deve por: controlar sinônimos, optando por uma única forma padronizada com remissivas para as outras; diferenciar homógrafos e reunir ou ligar termos cujos significados apresentem uma relação mais estreita entre si. Podem ser de três tipos: classificações bibliográficas, listas de cabeçalhos de assuntos e tesauros.

As classificações bibliográficas são os agrupamentos realizados também considerando a organização e disposição física dos documentos (PEREIRA *et al.*, 2009).

As listas de cabeçalho de assunto representam os temas sob a forma de cabeçalhos estruturados (ARAÚJO, 2011), dessa forma são elaboradas listas que podem ser alfabéticas, por áreas, cronológicas, com os termos criados de acordo com a política de indexação adotada.

Tesauro é um conjunto controlado de termos nos quais se estabelecem relações hierárquicas e analógicas. Se necessário, também são feitas relações de sinonímia com palavras da linguagem natural e que se aplicam a um campo particular do conhecimento.

Podemos dizer que o vocabulário controlado permite ao usuário identificar pontos de acesso ao documento. Esses pontos podem estar localizados em diversas áreas: termos de indexação, tipo documental, órgão de origem ou data (SMIT; KOBASHI, 2003, p. 13).

O controle de vocabulário intervém na organização dos arquivos ao nomear, de forma consistente, os pontos de acesso aos documentos e à informação neles contida. O objetivo a ser alcançado pelos arquivos, por essa óptica, é sempre o da recuperação da informação: somente esse objetivo justifica os cuidados com o controle de vocabulário.

Todos esses sistemas de padronização e representação da informação têm em comum um único objetivo, representar a informação para propiciar a recuperação dos documentos indexados.

3 A INDEXAÇÃO APLICADA A ARQUIVOS

Nas instituições arquivísticas de acervo permanente, a indexação pode se tornar grande aliada na pesquisa de documentos. Enquanto a descrição arquivística contempla o contexto histórico-administrativo do fundo, a indexação pode relacionar: o fundo com outros fundos, o fundo com seus dossiês e dossiês de um fundo com dossiês de outros fundos. Essa relação produz diferentes caminhos de busca, o que enriquece as possibilidades de pesquisa do usuário.

Essa atividade não se restringe, simplesmente, à criação de termos de recuperação da informação dentro do dossiê ou item. Por meio da indexação, ou do instrumento resultante de seu processo, pode-se também elaborar e descrever um arranjo de série temático padronizado.

De acordo com Dodebei (2006), os documentos arquivísticos possuem duas ordens de leitura. A primeira contempla as características essenciais dos documentos isolados de seu conjunto orgânico. Nessa se faz uma análise informativa do conteúdo do dossiê ou item. A segunda considera o dossiê ou item dentro do todo orgânico do qual faz parte, da cadeia estrutural do fundo ao qual pertence.

Durante muito tempo, a segunda leitura foi privilegiada e os níveis inferiores eram analisados circunstancial e subjetivamente em função do interesse do órgão mantenedor. Porém, com a informatização das bases de dados, uma nova ordem de leitura se fez necessária para que se atendesse a necessidade e as possibilidades de análise em vários níveis, e de forma simultânea, de um mesmo documento (DODEBEI, 2006).

Os documentos de arquivo possuem especificidades que

devem ser levadas em consideração na hora de indexar e, principalmente, distribuir as informações em seus campos específicos.

Durante muito tempo, acreditou-se que as técnicas biblioteconômicas não deveriam ser utilizadas em documentos de arquivo, pois os documentos de arquivo são únicos e não multicópias e por isso seriam mais complexos, do ponto de vista físico, que as publicações, ademais, o assunto não estaria explícito, por possuírem mais sentido coletivo do que unitário, como grupo do que como elementos singulares e serem mais significativos em relação à atividade do que ao assunto, além de terem origem heterogênea, podendo reportar a muitos assuntos diferentes (SCHELLENBERG, 1980).

Porém, com o passar dos anos, os documentos de arquivo ampliaram a variedade de gêneros e abrangências. Assim, passaram a ter a necessidade de ter “reconhecidas” suas características essenciais, dentro da individualidade do item ou dossiê, além de ter mantidas suas características orgânicas, que envolvem o conjunto documental em todos os níveis (fundo, seção, série, dossiê e item).

A indexação em documentos tão específicos e diferentes do documento bibliográfico requer alguma cautela e preparação de profissionais para o desenvolvimento da atividade, já que uma instituição arquivística não necessariamente contará na sua equipe com um bibliotecário experiente e/ou especializado.

4 O INDEXADOR NO CONTEXTO DA ORGANIZAÇÃO DA INFORMAÇÃO

De acordo com Lancaster (2004), diversos fatores influenciam na prática de um indexador, e na qualidade da indexação:

- 1) ligados ao indexador: conhecimento do assunto, experiência, concentração, capacidade de leitura e compreensão;
- 2) ligados ao documento: conteúdo temático, complexidade, língua e linguagem, extensão, apresentação e sumarização;
- 3) ligados a fatores ambientais: calor/refrigeração, iluminação, ruído.
- 4) ligados ao vocabulário: especificidade/sintaxe, ambiguidade/imprecisão, qualidade do vocabulário de entradas, qualidade da estrutura, disponibilidade de instrumentos auxiliares afins;
- 5) ligados ao “processo”: tipo de indexação, regras e instruções, produtividade exigida, exaustividade da indexação.

Ao indexador cabe definir, no processo de indexação, os níveis de exaustividade e de especificidade, e a escolha da linguagem, livre, controlada e pré-coordenada ou pós-coordenada.

Aos responsáveis pelo sistema de recuperação da informação e pelos instrumentos de pesquisa cabe analisar a capacidade de revocação e precisão do sistema, estratégias de busca, tempo de resposta do sistema, forma de saída.

À instituição detentora do acervo cabe verificar qual a preferência do usuário quanto à apresentação dos resultados e até que ponto o sistema satisfaz suas necessidades.

Para que a indexação seja realizada de maneira apropriada nos documentos de arquivo, o indexador deve analisar o documento como parte de fundo e como organismo independente. Dentre as informações que devem ser analisadas estão: história administrativa/biografia, história arquivística, título, âmbito e conteúdo, sistema de arranjo e notas.

Com a tomada de decisões, a instituição arquivística poderá padronizar procedimentos de análise conceitual, e mesmo descritiva, oferecendo ao usuário um ambiente de ampla possibilidade de acesso e pesquisa.

5 CONCLUSÃO

No mundo contemporâneo, onde é clara a necessidade de um processamento técnico mais eficaz e com respostas eficientes para a ampla gama de usuários possíveis, graças à informatização de dados das organizações, torna-se de suma importância o estabelecimento de padrões para descrição e indexação de documentos.

A normalização permite que se explore o cruzamento de dados nos documentos arquivísticos e se desenvolva técnicas para a análise de conteúdo e tradução de assuntos em linguagem documentária.

É preciso estar atento à formação de indexadores que conheçam o sistema de descrição multinível e os princípios da arquivologia, para que possam agregar as técnicas desenvolvidas pela biblioteconomia à recuperação da informação arquivística.

A relevância do desenvolvimento dessa atividade está nas inúmeras possibilidades de pesquisa que o ponto de acesso, ou seja, o “elemento de informação, termo ou código que, presente em unidade de descrição, serve à pesquisa, identificação ou localização de documentos” (VIANA, 2013, p. 9), criado com o termo de indexação é capaz de propiciar ao usuário/pesquisador, ao fornecer informações sobre vários documentos de um mesmo fundo e de diferentes fundos, porém, de mesmo tema.

Com o estabelecimento de critérios de análise e transcrição de assuntos, a instituição detentora de documentos arquivísticos será capaz de realizar uma indexação sem prejuízos na busca e recuperação da informação.

Deliberações nesse sentido implicam no desafogamento da sala de consulta por conta da agilização do tempo de pesquisa do usuário, na melhoria do grau de satisfação desse e na consequente melhoria da gestão da informação, sendo por isso uma atividade que pode ser melhor explorada na arquivologia.

REFERÊNCIAS

ARAÚJO, Helder *et al.* **Linguagens de indexação**: uso das linguagens presentes na prática da indexação. São Luís: Universidade Federal do Maranhão, 2011.

CARNEIRO, Marília V. Diretrizes para uma política de indexação. **Revista da Escola de Biblioteconomia da UFMG**, Belo Horizonte, v. 14, n. 2, p. 221-241, set. 1985.

CHAUMIER, J. Indexação: conceito, etapas, instrumentos. **Revista Brasileira de Biblioteconomia e Documentação**, São Paulo, v. 21, n. 1/2, p. 63-79, jan./jun., 1988.

DODEBEI, Vera L. D. **Curso de indexação de documentos para acervos arquivísticos**. Rio de Janeiro, 2006.

LANCASTER, F. W. **Indexação e resumos**: teoria e prática. Brasília: Briquet de Lemos, 2004.

PEREIRA, Edinete N. *et al.* **Classificação bibliográfica**: as diversas contribuições para o tratamento da informação. [Natal]: UFRN, 2009.

FUJITA, Mariângela S. L.; RUBI, Milena P. O ensino de procedimentos de política de indexação na perspectiva do conhecimento organizacional: uma proposta de programa para a educação à distância do bibliotecário. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 11, n. 1, p. 48-66, jan./abr. 2006.

SCHELLENBERG, T. R. **Documentos públicos e privados**: arranjo e descrição. 2. ed. Rio de Janeiro: Fundação Getúlio Vargas, 1980.

SMIT, Johanna W.; KOBASHI, Nair Y. **Como elaborar vocabulário controlado para aplicação em arquivos**. São Paulo: Arquivo do Estado; Imprensa Oficial do Estado, 2003.

VIANA, Alexandra Werneck *et al.* **Manual básico para indexação de documentos arquivísticos**: a experiência no Arquivo Nacional. Rio de Janeiro, 2013.

INDEXING TO INFORMATION RECOVERY IN ARCHIVES: AN INITIAL APPROACH

Abstract: *Indexing is the representation of documents' content through transcription of natural language to documentary language. Elaborating procedures of indexing in which criterias will be used by the indexers is necessary to the development of this activity. These procedures must consider the infrastructure of the organization or company, its human resources and its users. The final instruments resulted from the indexing process can be lists of subject headings, thesauri or controlled vocabularies. This control directly involves the archival description and organization, due its consistent appointment to the access points that lead to documents information. Although it is typically applied to bibliographic documents, the process of indexing gained importance in Archival Science meanwhile archival documents acquired diversity in gender and in content. To obtain a consistent indexation and application on archives, some procedures must be adopted in order to avoid information repetition and inefficient results.*

Keywords: *Information Retrieval. Indexing Policy. Vocabulary Control. Archival Documents. Documents.*

Originais recebidos em: 08/03/2016

Aceito para publicação em: 10/05/2016

Publicado em: 21/06/2016