

Modalidade: 1) Presentación oral de Ponencia en EBAM 2015

DA INDEXAÇÃO À RECUPERAÇÃO: a criação de um índice para auxílio no mecanismo de busca para a página eletrônica do curso de Arquivologia da Universidade Estadual da Paraíba (UEPB)

Manuela Eugenio Maia¹
Wellington da Silva Gomes
Jefferson Higino da Silva

1 INTRODUÇÃO

O uso das tecnologias na busca e no compartilhamento das informações produzidas e publicizadas na internet é, sem dúvida, o marco comunicacional desde os finais do século XX. Cada vez mais, sua evolução encontra-se no armazenamento multidimensional e nas relações de interatividade. Mais rápida e pontual, o envolvimento e a dependência entre indivíduos e rede ampliam o seu escopo e é algo inevitável no circuito humano nos diversos ambientes informacionais.

Assim, atualmente, interagir com as informações disponíveis em *websites* é tão comum e normal quanto rotineiro. Os investimentos e as ações na internet voltam-se para configurações de estruturas de rede simples e interativas, visando atrair e ampliar cada vez mais usuários. Contudo, embora se perceba por um lado esse fluxo, por outro, visualizamos a fragilidade da internet quanto à precisão dos sistemas de recuperação da informação. É comum encontrarmos problemas relativos ao controle terminológico, em principal, dentre as várias questões, os aspectos envolvendo a polissemia, fruto da riqueza vocabular da linguagem natural.

Essa riqueza gera os “inconvenientes” nos processos de recuperação da informação. Por exemplo, ao pesquisar em buscadores pelo termo “manga”, pode gerar as seguintes respostas: se referir à parte de uma roupa, ou à fruta, ou ao ato de rir (esse sentido, em especial, é peculiarmente usado no nordeste brasileiro). É na direção de controlar questões desse tipo, polissemia terminológica, e outras situações relativas à linguagem, que se constroem mecanismos vocabulares, representado por diferentes instrumentos de auxílio à pesquisa. Esse controle visa a situar o usuário no universo

¹ Doutoranda pelo Programa de Pós-Graduação em Ciência da Informação da Universidade Federal da Paraíba (UFPB). Mestre em Educação pela UFPB. Professora Assistente do curso de Arquivologia da Universidade Estadual da Paraíba (UEPB) e Editora Coordenadora da Revista Analisando em Ciência da Informação (RACIn).

linguístico referente a um campo específico onde a informação está circunscrita (GARCIA; SILVA, 2005).

Diante do número de acessos registrados pela empresa que disponibiliza o nosso *site* via internet, RedeHost, desde março de 2013 a abril de 2015, temos uma média acumulada de 106.522 visitas (MAIA; FERREIRA, 2015). Somamos também o volume de informação produzido pelo *site* do curso de Arquivologia da Universidade Estadual da Paraíba (UEPB), que possui mais de 100 (cem) matérias postadas. Em função desses dados, percebemos a necessidade de incorporar à página eletrônica acima citada ferramentas de auxílio de busca para atender demandas dos seus usuários. Para tal, a escolha de termos associados ao conteúdo de cada *link* do *site* deve ser crucial para o processo de recuperação da informação para o usuário. Assim, construímos entre 2014 e 2015 um índice com o intuito de realizar o controle da terminologia desse *site*, criando condições de inserir o usuário no contexto vocabular para a sua busca e recuperação informacional.

Dois princípios basilares sustentaram as nossas escolhas terminológicas: usar termos que permitam precisar quanto revocar as matérias postadas na nossa página. Nessa direção, o objetivo desse artigo é o de descrever como foi realizado a construção do índice para o *site* do curso de Arquivologia da UEPB.

Justificando a relevância desse estudo, estamos produzindo um debate bastante frutífero entre Biblioteconomia, área perita em controle terminológico e em tratamento da informação, Arquivologia, área de objeto de nosso estudo, e Ciência da Computação, espaço em que o conteúdo do *site* está disposto e fundamental para os processos de recuperação da informação. São três áreas que necessitam estreitar cada vez mais seus pontos de articulação no sentido de gerar produtos de informação que potencializem e garantam o seu pleno acesso e uso, finalidade da interlocução entre o tratamento e a recuperação da informação. Essa interseção, para a nossa proposta, reside na construção de um índice e o seu uso no processo de indexação na aplicação de motor de busca na página eletrônica do curso de Arquivologia da UEPB.

2 METODOLOGIA

O número 106.522 refere-se aos acessos a página eletrônica do curso de Arquivologia da UEPB desde a sua criação, março de 2013 até abril de 2015, mês anterior a elaboração desse artigo. Esse substancial número talvez ocorra pelo intenso

compromisso nas atualizações das matérias postadas, que desde sua origem até o presente, foram mais de 100 (cem) notícias, que tratam de diversas informações envolvendo a área. Considerando esses dois dados de natureza quantitativa, percebemos a necessária organização entre a produção da informação e quem dele procura e faz uso.

Tendo esse dados como ponto de partida, a nossa análise deu-se sobre o prisma qualitativo, pautando-se na tríade tratamento – informação – recuperação. Considerando o valor total de acessos, estabelecemos uma relação de mensal (26 meses) de 4.097, conduzindo-nos a pensar em estratégias que facilitassem a localização das matérias de forma eficiente para os seus usuários.

Assim, com base em pesquisa estritamente documental (MINAYO, 1996; RICHARDSON, 1999; SEVERINO, 2007), realizamos leituras na área de organização, tratamento e recuperação da informação. Percebemos que foi pertinente a construção de um instrumento de pesquisa que auxiliasse o indexador, controlando o vocabulário dos termos, e o usuário, orientando-o no processo de busca. Dos instrumentos estudados, optamos pelo índice, que lista os termos e os relaciona, estabelecendo os descritores representativos e significativos (LOPEZ, 2002; SMIT, 2003). Nesse aspecto, foram instituídas formas para identificar as matérias referentes à nossa página eletrônica em um processo de indexação, com o intuito de auxiliar na busca da informação desejada.

Após essa escolha, listamos as matérias por ordem cronológica crescente, de 2013 até o presente. De cada matéria lida, selecionamos os termos por extração e por atribuição, considerando a revocação e a especificidade do conteúdo de cada notícia. Assim, além de documental, essa pesquisa é descritiva, pois a análise de cada conteúdo é central no processo de construção do índice (RICHARDSON, 1999). A necessária criação de um instrumento de pesquisa visa a aperfeiçoar de forma controlada as informações contidas no nosso *site*. Nesse sentido, o índice é criado como instrumento que vem auxiliar essa indagação no intuito de garantir eficácia ao usuário em suas buscas.

Em síntese todos esses elementos já citado têm sua linearidade voltada para recuperação da informação, fazendo com que o usufrutuário venha executar suas atividades de interatividade e pesquisa no *website* com precisão.

3 DA PÁGINA ELETRÔNICA DO CURSO DE ARQUIVOLOGIA DA UEPB

A página eletrônica do curso de Arquivologia da UEPB caracteriza-se como

sendo um ambiente de internet, criada a partir de uma ação extensionista voltada para os alunos do referido curso. Nascida em março de 2013, é composta de dados sobre o curso, estrutura curricular e administrativa, docentes, informações sobre discentes, publicações em revista indexada, legislação e outros conteúdos.

Por meio dos dados pesquisados acerca do curso de Arquivologia da UEPB e da área como um todo, estabelecemos a classificação e nomeamos os rótulos, deixando-os arquitetados com termos familiares para os usuários, resultando em uma página com o designer dinâmico, simples e interativo, conforme a Figura 1:

Figura 1: Página principal do curso da Arquivologia da UEPB



Fonte: Dados da pesquisa, 2015 (MAIA; FERREIRA, 2015)

No rótulo "curso", temos todos os conteúdos referente a fundação, histórico, projeto político pedagógico, bem como os horários e componentes curriculares do curso de Arquivologia. Já em "estrutura administrativa", encontramos informações sobre coordenação, departamentos, direção e colegiado de curso em tela.

Os rótulos "docente" e "discente" apresentam elementos característicos desses dois seguimentos. Para o primeiro rótulo, temos informações como o currículo Lattes; quanto ao segundo, estão vinculados assuntos como a lista de Trabalho de Conclusão de Curso (TCC) defendidos e controle acadêmico (sistema de matrícula e notas das disciplinas).

Podemos citar também o rótulo "links" que trata de atalhos para *sites* de eventos, instituições e legislação arquivísticas e conteúdos afins. Atentamos para o rótulo "pesquisa acadêmica", em que sugerimos *sites* de revistas, bases de dados, bibliotecas digitais etc para os nossos usuários.

O nosso *site* é constituído também por notícias recentes, informando e atualizando os usuários. São postadas matérias semanais direcionadas à submissão de trabalhos científicos, inscrição em eventos acadêmicos, congressos nacionais e

internacionais, entre outros conteúdos.

Como forma de garantir a comunicação, sugestões e o *feedback* com os usuários, temos o rótulo "fale conosco". Além disso, neste campo, é explicitado os responsáveis pela gestão da página.

O ápice do nosso *site* é o gerenciamento da Revista Analisando em Ciência da Informação (RACIn), cujo o Número de Padronização Internacional para Publicações Seriadas (ISSN) tem o registro 2317-9708 junto ao Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT). Nascida em 2013, vívida há 3 (três) anos e de periodicidade semestral, a Revista já se encontra indexada na base do Latindex, um sistema de informação cooperativo e de divulgação sobre revistas científicas pertencentes aos países da América Latina, Caribe, Espanha e Portugal (LATINDEX, 2014).

Todo esse trabalho é realizado por uma equipe de colaboradores formada por dois docentes efetivos e dois discente da UEPB.

4 PROCESSO DE INDEXAÇÃO E RECUPERAÇÃO DA INFORMAÇÃO: construindo um índice

Criar pontos de acesso em espaços que promovem e disseminam informações disponíveis na internet requer aguçada sensibilidade considerando a pluralidade de usuários potenciais que dela faz uso e sua agilidade comunicativa. A extração ou atribuição de descritores em documentos, sejam livros ou outros registros de informação, é realizado por meio de um processo analítico permeado por etapas, que perpassam desde o controle de vocabulário até a construção de mecanismos de busca como instrumentos de recuperação da informação. É relevante pontuar que entendemos vocabulário controlado como conjunto limitado de termos autorizados para a indexação e busca de documentos em uma determinada base de dados (LOPEZ, 2002, p. 42).

No sentido de estabelecer artifícios para facilitar e assegurar a localização de informações de maneira ágil e bem sucedida na *web*, especificamente na página eletrônica do curso de Arquivologia da UEPB, formulamos um índice. Este instrumento de acesso visa a auxiliar o usuário na busca pela informação desejada. De acordo com Lopez (2002, p. 33),

os índices têm como objetivo permitir uma rápida localização das unidades documentais que atendam a critérios específicos [...]. Na confecção de

índices, deve-se tomar muito cuidado com a escolha dos termos a serem utilizados. Em tais tarefas a utilização de vocabulários controlados e tesauros é imperativa.

Nesse contexto, para a construção do nosso índice, foi realizado um trabalho de análise das matérias do *site*, o qual gerenciamos, extraindo ou atribuindo um conjunto de termos de cada uma das notícias. Por conseguinte, ordenamos alfabeticamente a lista de descritores em uma tabela, estabelecendo as remissivas para cada termo significativo (SMIT; KOBASHI, 2003).

Em parte dos *sites* e das bases de dados da internet, são estabelecidos índices de forma automática, ou seja, instalam-se *softwares* nesses ambientes que promovem a captura de termos e os listam sem a intervenção humana durante esse processo. No nosso caso, optamos em construir manualmente o índice, lendo todos os conteúdos postados de março de 2013 a abril de 2015 e determinando a lista de descritores. Alia-se a esse trabalho o fato de conhecer e de participar da mesma comunidade de usuários para qual a página eletrônica é direcionada.

A vantagem da construção manual de índices se dá pela capacidade humana em julgar possibilidades da representação da informação, [refletindo] no processo de qualidade, como consequência uma precisão informacional maior concomitantemente com o controle de vocabulário (FERNEDA, 2003, p. 96).

De forma ampla, indexamos para representar o conteúdo temático de uma informação servindo de alicerce para facilitar com precisão os processos de sua busca. Por indexação, entendemos como um processo mental que atribui a um conjunto finito de termos controlados como pista na identificação de um conteúdo informativo maior (LANCASTER, 2004). Assim, quando elegemos determinados descritores, que são utilizados como pontos de acesso, objetivamos facilitar a recuperação e a busca para uma determinada informação desejada. Por exemplo, imaginemos um documento que trate da realização de "Congresso Nacional de Arquivologia" e, no processo de indexação, foi utilizado como uma das pontes de acesso o termo "evento".

Em função de sua abrangência, o resultado poderá tornar o processo de busca muito amplo, influenciando dessa forma na qualidade da recuperação, pois, além do "Congresso Nacional de Arquivologia", existem vários outros "encontros acadêmicos" como "conferência", "mostra", "seminário", "simpósio" etc. Essa amplitude facilita no caráter revocatório, contudo, dificulta na especificidade. Ou seja, o uso de um descritor amplo pode permitir respostas exaustiva no que se refere aos os números de termos atribuídos a matéria.

É nessa lógica que optamos por indicar na indexação de cada matéria do *site* termos específicos quanto amplos, atendendo as duas expectativas, dependendo do que o usuário desejar. Voltemos ao nosso exemplo supra, o usuário que estiver interessado em uma matéria sobre o "Congresso Nacional de Arquivologia", e assim solicitar ao índice, este reemitirá as informações relacionadas apenas ao termo específico, ou seja, com um nível de revocação muito menor. Isto quer dizer que os resultados apresentados podem ser mais relacionados e pontuais acerca do tema no processo de busca.

Em uma pesquisa, a relação entre termos e o total de itens recuperados é denominada coeficiente de precisão, ou seja, são os itens que servem para sua pesquisa em relação a todos os que aparecem no resultado da mesma. No coeficiente de revocação, todos os itens procurados são encontrados, pois as pontes de acesso são descritores mais genéricos (FERNEDA, 2003; LANCASTER, 2004).

No índice, inserimos as remissivas, sinalizando o descritor representativo para aqueles conceitos que podem ser designados por vários termos. Ou seja, a remissiva refere-se a escolha de um único termo que é usado como padrão em relação a outros que possuem o mesmo significado. Um caso fácil de compreender é quando há siglas no índice. Como exemplo no nosso índice, temos os descritores "AN" e "Arquivo Nacional, ambos designam o mesmo objeto. Por isso, é necessário fazermos uma opção e, no nosso caso, escolhemos o descritor por extenso ao invés de sua sigla, deixando esclarecido no texto do índice com o uso da remissiva.

5 CONSIDERAÇÕES FINAIS

Com o auxílio de conhecer as características de quem procura o conteúdo na *web*, tornar-se prático estabelecer estratégia para o tratamento das informações estruturadas nas matérias da página do curso de Arquivologia da UEPB. Partindo dos princípios basilares que norteiam a indexação, a saber, especificidade, exaustividade, revocação e precisão, o seu uso é imprescindível na construção de mecanismos de controle de vocabulário e de instrumentos de busca para os usuários (GARCIA; SILVA, 2005).

Selecionamos descritores para construção do índice partindo das necessidades informacionais dos usuários, analisando o contexto dos significados de cada matéria postada no nosso *site*. Atentamos para os elementos como classificação, processo de busca e recuperação da informação, que são necessários para compreensão da página

eletrônica do curso de Arquivologia da UEPB no que se refere ao ensino, à pesquisa e à extensão.

REFERÊNCIAS

GARCIA, Rodrigo Moreira; SILVA, Helen de Castro. O comportamento do usuário final na recuperação temática da informação: um estudo com pós-graduandos da UNESP de Marília. **DataGramaZero - Revista de Ciência da Informação**, Rio de Janeiro, v. 6, n. 3, jun. 2005. Disponível em: <http://www.dgz.org.br/jun05/Art_02.htm>. Acesso em: 10 maio 2015.

FERNEDA, Edberto. **Recuperação de informação: análise sobre a contribuição da ciência da computação para a ciência da informação**. 2003. 146 f. Tese (Doutorado em Ciências da Comunicação)- Escola de Comunicações e Artes, Universidade de São Paulo, São Paulo, 2003.

LANCASTER, F. W. **Indexação e resumo: teoria e prática**. Brasília, DF: Brinquet Lemos, 2004.

LATINDEX. Sistema Regional de Información en Línea para Revistas Científicas de América Latina, el Caribe, España y Portugal. **Portal principal**. México: [s. n.], 2014. Disponível em: <<http://www.latindex.unam.mx/>>. Acesso em: 10 maio 2015.

LOPEZ, André Porto Ancona. **Como descrever documentos de arquivo: elaboração de instrumentos de pesquisa**. São Paulo: Arquivo do Estado de SP, 2002.

MAIA, Manuela Eugênio; FERREIRA, Danilo de Sousa. **Página principal do curso de Arquivologia da Universidade Estadual da Paraíba**. João Pessoa: RedeHost, 2015. Disponível em: <<http://www.arquivologiauepb.com.br/>>. Acesso em: 10 maio 2015.

MINAYO, Maria Cecília de Souza. **O desafio do conhecimento: pesquisa qualitativa em saúde**. São Paulo: HUSITEC, 1996.

RICHARDSON, Roberto Jarry. **Pesquisa social: métodos e técnicas**. São Paulo: Atlas, 1999.

SEVERINO, Antônio Joaquim. **Metodologia do trabalho científico**. São Paulo: Cortez, 2007.

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