



A importância da prática de indexação para a recuperação da informação: relato da BV-FAPESP¹

Fabiana Andrade Pereira
Rosaly Favero Krzyzanowski
Thais Fernandes de Moraes
Juliana Calherani

Resumo: Relata a experiência da Biblioteca Virtual da FAPESP (BV-FAPESP) no processo de indexação de assuntos, enfatizado a partir 2012, para o tratamento informacional dos seus registros referenciais. Relevantes para a organização e recuperação da informação nas bases de dados da BV-FAPESP, disponível na Web desde 2005, os assuntos são elementos essenciais da representação temática dos seus conteúdos e contribuem para o acesso rápido às informações, referentes aos auxílios à pesquisa e bolsas em ciência, tecnologia e inovação, apoiados pela Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP).

Palavras-chave: Bibliotecas virtuais; Indexação; Assuntos; Tratamento da informação; Recuperação da informação.

1 INTRODUÇÃO

A Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) é uma das principais agências de fomento à pesquisa do Brasil. Segundo Lafer (2014), a FAPESP tem autonomia garantida pelo Artigo 271 da Constituição Estadual e recebe 1% do total da receita tributária do Estado de São Paulo, cujo escopo essencial é apoiar a pesquisa em ciência e tecnologia (C&T), por meio de bolsas e auxílios. Por este motivo, faz parte de sua missão divulgar aos cidadãos paulistas suas ações na utilização dos recursos públicos. Para atingir a missão, a Fundação torna acessíveis as informações referenciais sobre auxílios e

¹ Relato de experiência apresentado oralmente no XXVI Congresso Brasileiro de Biblioteconomia e Documentação 2015: Biblioteconomia, Ciência e Profissão, 21-24 jul. 2015, São Paulo. Resumo expandido disponível em: <http://siscone.com.br/Uploads/CBBB15/Trab14400205020150331_000000.pdf>.



bolsas, por ela apoiados, em diferentes canais, entre eles, a Biblioteca Virtual da FAPESP (BV-FAPESP) <<http://www.bv.fapesp.br/>>.

Desde o seu lançamento, ocorrido em 2005, a BV-FAPESP utiliza de modernos recursos da tecnologia de comunicação e informação, e adota regras e padrões no tratamento e organização dos conteúdos em meio digital, disponibilizando informações de qualidade na internet para todos os interessados. Portanto, a BV-FAPESP, além de se configurar como uma importante ferramenta on-line também assume a responsabilidade de garantir o acesso equitativo, aos mais de 200 mil registros de projetos de pesquisa, à comunidade científica e à sociedade em geral, expressados através da sua significativa visibilidade alcançando mais de 4 milhões de acessos em 2014.

A BV-FAPESP apresenta as informações referenciais relativas aos auxílios e bolsas financiados, disponibilizando dados como títulos, resumos, assuntos e demais elementos sobre o fomento. Além disso, disponibiliza informações de valor agregado sobre os pesquisadores e os resultados oriundos das pesquisas, como as produções científicas e acadêmicas, em interfaces de acesso público, em português e inglês.

Os assuntos compõem uma parte relevante na organização e recuperação da informação, e contribuem para o acesso rápido às informações sobre os projetos. Ao pesquisar alguma palavra na BV-FAPESP utilizando o campo de busca da página inicial, pode-se obter sugestões dos assuntos existentes na BV e, caso essa palavra faça parte da Lista de Assuntos da BV-FAPESP, o usuário será direcionado para a “Página do Assunto” que, além de reunir projetos referentes ao tema, também apresenta outros termos relacionados. Portanto, dada a importância dos assuntos para a representação temática dos projetos de pesquisas, o objetivo primordial deste trabalho é demonstrar como ocorre o tratamento dos assuntos e como eles auxiliam na indexação, na organização e na recuperação de informação na BV-FAPESP. Do mesmo modo, será exposto como a experiência no tratamento dos assuntos tem colaborado para o fornecimento de informações relevantes e precisas para os usuários, além dessas soluções contribuir continuamente para o processo de indexação automática pelo sistema, com ênfase para a adoção de uma



pré-política de indexação, o controle de remissivas, o uso de qualificadores, o tratamento de assuntos genéricos, o uso de notas explicativas e a seleção de novos assuntos.

2 A IMPORTÂNCIA DA INDEXAÇÃO PARA AS BIBLIOTECAS PRESENTES NA WEB: REFERENCIAL TEÓRICO

Conforme aponta Fujita (2003, p. 61-62), a indexação está presente no processo de análise documentária, e trata-se de um método que busca combinar estrategicamente o tratamento do conteúdo dos documentos e a recuperação pelo usuário. A autora também diz que a indexação deve ser observada em dois aspectos: 1) descreve e identifica o documento por meio de conceitos que o representam; 2) auxilia na busca e no acesso à informação armazenada. Ainda, ao se tratar dos sistemas de informação, a autora reconhece que a indexação é a responsável por condicionar os resultados das estratégias de busca, ou seja, o bom desempenho da indexação refletirá sobre a recuperação da informação presente nas bases de dados. Os resultados recuperados mais pertinentes ao pesquisador são aqueles cujos conceitos correspondem precisamente às suas questões de busca.

Sobre esse aspecto Araújo Júnior (2007, p. 14) argumenta que a indexação serve para viabilizar a escolha dos termos que representarão o conteúdo e são imprescindíveis no aumento do índice de precisão de resposta na recuperação da informação. Para o autor, a indexação ocorre na conversão da linguagem do documento para a linguagem documentária, a fim de que o documento seja posteriormente recuperado.

Lancaster (2004, p.1), por sua vez, descreve que qualquer tarefa, que envolva as técnicas de indexar e resumir um documento, visa sua representação para “inclusão em algum tipo de base de dados”. O autor descreve que, ao fazer uma busca numa base de dados, almeja-se satisfazer uma necessidade de informação e, dessa forma, deve-se evitar a recuperação de itens inúteis. Com isso, a relevância e a pertinência são termos que



descreverão as informações consideradas úteis, denotando o poder de revocação menor e com um índice maior de precisão².

Atualmente, a Internet como fonte de pesquisa, possibilitou-se o acesso a uma grande quantidade de informação. Estar presente na Web e disponibilizar informações relevantes, como é o caso das bibliotecas virtuais e digitais, tornou-se um grande desafio meio às outras fontes existentes. Então, as técnicas de análise documental, incluindo a indexação, devem ser aplicadas, uma vez que aumentam as possibilidades de visibilidade e encontrabilidade da informação³. Ao considerar os parâmetros necessários à otimização dos dados para serem indexados pelos motores de busca - ou SEO (Search Engine Optimization) -, cujo maior propósito consiste em fazer com que esses dados apareçam entre as páginas preliminares dos buscadores e, assim, serem considerados como os resultados mais relevantes, os termos que os indexadores atribuem serão os pontos de acesso para encontrar determinada informação veiculada no meio de tantos outros sites existentes na Web.

Conforme ocorre na BV-FAPESP, além da preocupação em tornar a Biblioteca Virtual uma ferramenta, com revocação e precisão adequada para o público, também há a apreensão de torná-la pertinente nos resultados das buscas por informação de qualidade na Web. Não há dúvidas quanto ao papel expressivo que a gestão dos conteúdos informacionais exerce em ambientes complexos como é a Web, principalmente com o uso de procedimentos de padronização como a classificação, a indexação e a compatibilização de recursos, que são efetivos para a busca, recuperação e uso da informação, conforme destacam Campos et al. (2005).

O controle da linguagem que se dá através da indexação, no uso de vocabulários controlados, de ontologias, de tesouros, das taxonomias, entre outros padrões, futuramente serão o suporte essencial para a ocorrência da Web semântica. Campos et al. (2005 p. 71-

² RUBI, M. P. Os princípios da política de indexação na análise de assunto para catalogação: especificidade, exaustividade, revocação e precisão na perspectiva dos catalogadores e usuários. In: FUJITA, M.S.L., (org.) et al. *A indexação de livros: a percepção de catalogadores e usuários de bibliotecas universitárias*. Um estudo de observação do contexto sócio cognitivo com protocolos verbais. São Paulo: Editora UNESP; São Paulo: Cultura Acadêmica, 2009.

³ VECHIATO, F. L.; VIDOTTI, S. A. B. G. *Encontrabilidade da informação*. São Paulo. Cultura Acadêmica, 2014.



72) descrevem que os sistemas de bibliotecas digitais, ou mesmo as virtuais, estão alinhados com os objetivos da Web semântica, que depende do desenvolvimento de ferramentas que facilitem a “geração e a gerência do vasto conjunto de descritores”, visto que essas bibliotecas são “ilhas de informação de reconhecida qualidade”.

3 A PRÁTICA DE INDEXAÇÃO PARA REPRESENTAÇÃO TEMÁTICA DOS CONTEÚDOS INFORMACIONAIS DA BV-FAPESP: PROBLEMAS E SOLUÇÕES

A indexação dos registros referente às bolsas e auxílios à pesquisa sempre fez parte da rotina de trabalho dos bibliotecários da BV-FAPESP, isto é, todos são bibliotecários-indexadores. Sempre houve a preocupação com a representação temática dos conteúdos que comporiam o “acervo” da Biblioteca e, até hoje, os dados são extraídos dos títulos, resumos e palavras-chave fornecidas pelos pesquisadores e, posteriormente, traduzidos para termos de linguagem documentária.

Após alguns anos presente na Web e diversas modificações ocorridas nos softwares e interfaces do site da BV, além da inclusão de novas tipologias de informações, percebeu-se a necessidade de organizar a lista de assuntos existente até então na BV, pois surgiram diversas problemáticas que poderiam afetar diretamente o resultado de busca dos usuários, advindos sobretudo do uso de diferentes vocabulários controlados no ato de indexação (Vocabulário Controlado do SIBiUSP⁴ e o DeCs - Descritores em Ciências da Saúde⁵). Estes instrumentos de controle de termos têm na BV a função de padronizar e reunir todos os documentos sob uma única entrada, tornando maximizada a chance de recuperar todos os registros de um mesmo assunto.

A partir de 2012, surgiram algumas inquietações sobre a interferência dos assuntos para recuperação das informações da BV-FAPESP e, com isso, o trabalho de tratamento

⁴ Vocabulário Controlado do SIBiUSP. Disponível em: <<http://vocab.sibi.usp.br/pt-br/index.php>>.

⁵ DeCS - Descritores em Ciências da Saúde. Disponível em: <<http://decs.bvs.br/>>.



dos assuntos foi enfatizado. Para tanto, foi necessária a admissão de procedimentos e técnicas pelos bibliotecários, assim como o suporte fundamental das tecnologias.

3.1 Adoção de uma Pré-política de Indexação

A primeira ação assertiva visando uma solução, foi adotar uma “pré-política” de indexação. A adoção de regras, conforme destaca Fujita (2012, p. 17) “não deve ser vista como uma lista de procedimentos a serem seguidos, e sim um conjunto de decisões que esclareçam os interesses e objetivos de um sistema de informação e, particularmente, do sistema de recuperação da informação”.

Logo, a adoção de uma política de indexação tornou-se necessária, pois ela ditaria os elementos norteadores e os critérios que serviriam de guia para os bibliotecários-indexadores, na otimização do serviço e na racionalização dos processos. Entre os critérios, destacam-se os que se seguem: a) Adotar o termo mais utilizado pela comunidade de usuários/pesquisadores; b) Escolher sempre o termo mais específico para representar o assunto, evitando termos genéricos; c) Em caso de sinonímia o termo mais conhecido é adotado (autorizado quando relevante em buscas na literatura especializada e na Web, por exemplo), fazendo a remissiva para os termos equivalentes (não-autorizados); d) Os termos equivalentes estão na mesma ordem alfabética dos cabeçalhos autorizados, mas são identificáveis evitando seu uso pelos bibliotecários; e) As relações hierárquicas, o padrão de uniformização e o uso de plurais/singulares seguem as regras dos vocabulários controlados utilizados.

A política de indexação da BV-FAPESP atualmente está em construção com base nos princípios anteriormente descritos, também contemplando os documentos confeccionados durante a formação e melhoria das ferramentas e processos. Atualmente, a BV-FAPESP possui uma “Lista de Cabeçalho de Assuntos”, organizada alfabeticamente,



contendo mais de 25 mil termos e é um instrumento auxiliar para os bibliotecários na indexação dos registros informacionais sobre aos projetos financiados pela FAPESP.

Durante a indexação, os bibliotecários fazem a verificação dos resumos e títulos das bolsas e auxílios, e analisam as palavras-chave fornecidas pelos pesquisadores que apresentam uma série de divergências:

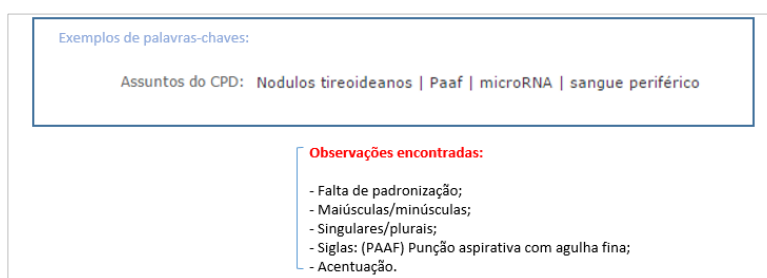


Figura 1 - Análise as palavras-chave fornecidas pelos pesquisadores: exemplo.

Fonte: Base administrativa da BV-FAPESP e autores (2015).

Os termos são traduzidos da linguagem natural (palavras-chave) para a linguagem controlada (descritores).

Ainda, entre as questões iminentes, outro enfoque é com relação à falta de união dos termos, inibindo um bom índice de precisão na recuperação da informação, mesmo na BV possuindo uma indexação controlada e contando com uma extensa lista de termos autorizados. Além disso, o uso de diferentes vocabulários trouxe também problemas com relação à sinonímia, plural/singular, termos genéricos, homografia, homonímia, siglas/descrição completa, etc., conforme pode-se verificar nos exemplos a seguir:

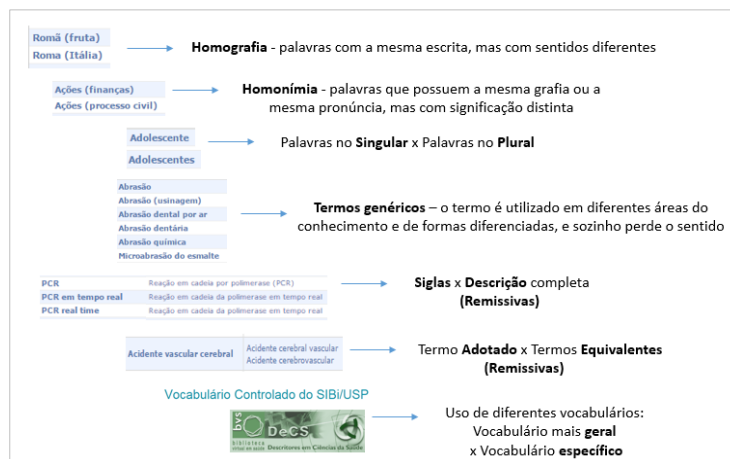


Figura 2 – Problemas identificados na Lista de Assuntos da BV-FAPESP: exemplos.

Fonte: Os autores (2015).

3.2 Controle de Remissivas

Objetivando solucionar a maioria das questões apontadas na Figura 2, foi estudada e desenvolvida, em conjunto com analistas de sistemas e bibliotecários, uma ferramenta que viabilizou a criação de remissivas, que possibilitou analisar as relações de equivalência entre os termos, permitindo eleger os preferenciais, entre as várias entradas distintas. Dessa maneira, cria-se várias entradas de assuntos conforme surgem nos registros da BV, mas somente um termo será adotado e “responsável” por reunir todos os registros referentes a uma determinada temática. A solução implantada na ferramenta de remissivas faz com que, independente do termo selecionado pelos bibliotecários no ato de indexação, o sistema faz a troca automática do termo equivalente para o adotado.



Assuntos:

Vulnerabilidade Saúde bucal Cárie dentária Fatores de risco Adolescente
Adolescente
Adolescentes

Figura 3 – Ferramenta de remissiva no ato de indexação: exemplo.
 Fonte: Base administrativa da BV-FAPESP e autores (2015).

Com as implementações no sistema para o Controle de Remissivas, por exemplo, foi possível exercer a gestão da lista de assuntos de maneira mais precisa, com a possibilidade de remeter os termos sinônimos e incorretos para termos adotados de forma automatizada. Através de filtros específicos contidos na planilha administrativa, pode-se ter acesso às particularidades dos assuntos, saber quais são os termos genéricos, ou, quais termos não foram certificados. Outra implementação importante foi a possibilidade de visualizar todos os registros vinculados a um determinado assunto.

Com essa ferramenta, no menu “Controle de Remissivas” é possível visualizar a lista de todos os assuntos e suas remissivas (atualmente contendo mais de 1500).

Controle de remissivas	
Termo adotado	Termo equivalente
8-OH-DPAT	8-Hidroxi-2-(di-n-propilamino)tetralina
Abacaxi	Ananás
Acampamentos	Acampamento
Ação capilar	Capilaridade
Acelerador de partículas	Aceleradores de partículas
Zeolitas	Zeólitos
Zircônia tetragonal policristalina estabilizada com ítrio (Y-TZP)	Y-TZP
Zircônio	Óxido de zircônio Zircônia
Zoologia (classificação)	Taxonomia animal
Zoonoses	Zoonose
Total de remissivas	1588

Figura 4 – Controle de remissivas.
 Fonte: Base administrativa da BV-FAPESP e autores (2015).



3.3 Uso de Qualificadores

O qualificador (ou qualificador parentético) é usado para esclarecer o significado, definir a área do assunto ou para os casos de homografia.



Figura 5 – Exemplos de uso de qualificadores.

Fonte: Base administrativa da BV-FAPESP e autores (2015).

Nos exemplos anteriores, pode-se verificar que o caso 1) busca solucionar uma situação de homografia, pois o sistema ainda não é capaz de identificar se a palavra Roma, por mais que esteja acentuada ou não, trata-se de uma fruta (romã) ou de uma cidade italiana (Roma). Já o caso 2) trata-se de uma ocorrência de homônimos (ações em finanças e ações em processo civil) - de termos que tem significados diferentes em áreas distintas, mas possuem a mesma grafia e pronúncia.

Quando casos assim são identificados, o “Termo chave” (no exemplo de Roma e Ações) é colocado como “Termo genérico” e, eventualmente, os bibliotecários-indexadores buscam especificar o sentido do termo em questão.

3.4 Tratamento de Assuntos Genéricos

Para a BV-FAPESP os termos genéricos equivalem aos assuntos que sozinhos não fazem sentido diante da temática exposta no documento, uma vez que não especificam o assunto, são vagos, podendo abranger várias ideias e significados.



Levando em conta que o intuito é formar uma hierarquização dos assuntos existentes e disponibilizá-los para a comunidade científica e tecnológica, é preciso tratar os registros informacionais, que contém termos genéricos para que os assuntos representem a especificidade das pesquisas científicas. Logo, no ato da indexação dos registros, os assuntos genéricos não ficam disponíveis, de imediato, para os bibliotecários, porém o sistema automatizado faz o agrupamento dos projetos que possuem os termos genéricos. Posteriormente, os bibliotecários-indexadores analisam os registros dos projetos e definem os assuntos mais característicos à temática da pesquisa.

3.5 Uso de Notas Explicativas

As notas explicativas são usadas para definir os termos, principalmente quanto ao seu escopo ou uso.

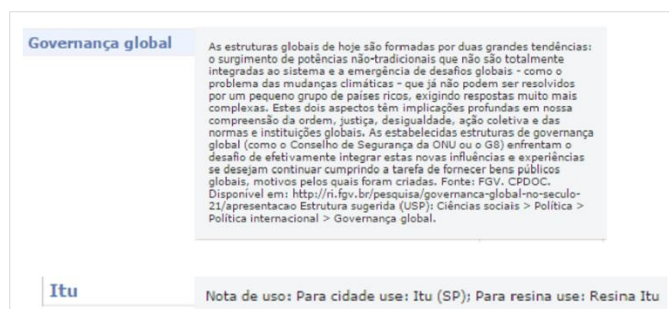


Figura 6 – Exemplos de notas explicativas para definição e uso.
Fonte: Base administrativa da BV-FAPESP e autores (2015).

3.6 Seleção de Novos Assuntos

A FAPESP é praticamente o berço das pesquisas, que entram no sistema da BV, assim que o Termo de Outorga da FAPESP é assinado. Logo, a ocorrência de assuntos



novos é constante na base da BV e, é imprescindível selecionar e adicionar novos assuntos diariamente. Para a adoção de novos assuntos é feito um trabalho minucioso de análise documentária e de sua relevância (internamente, dentro da própria BV; e externamente, em demais trabalhos científicos). Por isso, é indispensável localizar na literatura especializada, em artigos científicos, por exemplo, as grafias e descrições corretas dos termos, além de “pré-determinar” uma estrutura hierárquica.

Fica evidente a necessidade de estabelecer uma cooperação técnica com os responsáveis pelos vocabulários controlados utilizados para fornecer os novos assuntos.

4 RESULTADOS PRELIMINARES

O tratamento dos assuntos é essencial para a obtenção de resultados efetivos, por exemplo, a Página Inicial do site da BV-FAPESP apresenta sugestões de assuntos para auxiliar os usuários na busca de informações nas bases de dados:



Figura 7 – Auxilio dos assuntos na página inicial da BV-FAPESP.
Fonte: Base administrativa da BV-FAPESP e autores (2015).



Caso a palavra buscada na BV faça parte da Lista de assuntos, o usuário será direcionado para uma página específica que, além de reunir os auxílios e bolsas relacionados à temática, apresenta ainda, termos relacionados, mapa e gráficos do fomento.

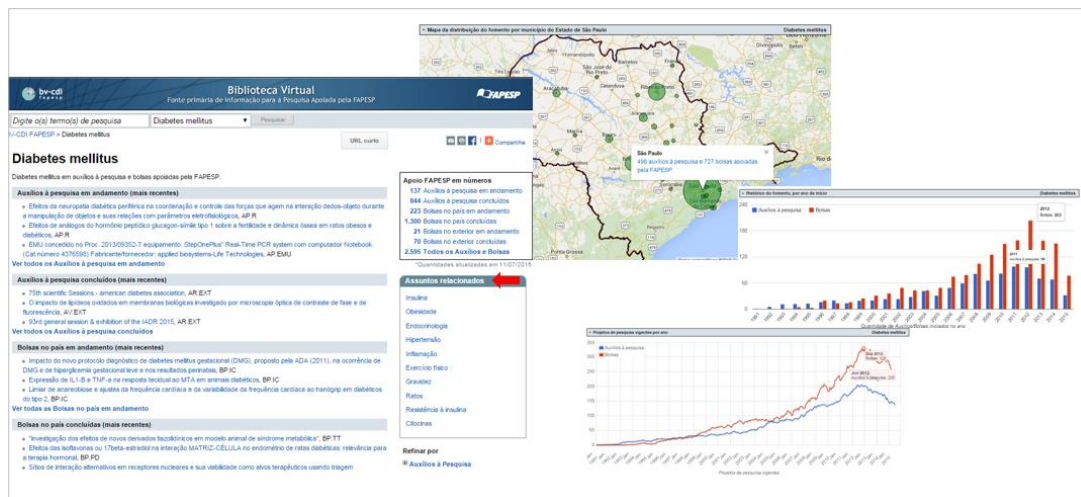


Figura 8 – Página de Assuntos, mapa e gráficos do apoio.
Fonte: Base administrativa da BV-FAPESP e autores (2015).

Destaque para a Página de Pesquisadores que, com base nos assuntos apresentados nos registros das pesquisas científicas, permite ao pesquisador e público em geral a visualização em forma de nuvem de *tags* das “palavras-chave” mais utilizadas nos projetos, fato que auxilia na informação sobre as expertises de cada pesquisador FAPESP.



Figura 9 – Nuvem de tags na Página do Pesquisador.
 Fonte: Base administrativa da BV-FAPESP e autores (2015).

5 CONSIDERAÇÕES FINAIS

Perante a análise dos assuntos da BV-FAPESP vem sendo possível observar que, devido ao avanço da pesquisa científica do Estado de São Paulo, constantemente surgirão termos novos, que não estão inseridos nos vocabulários até então utilizados na indexação, e que deverão ser adicionados à lista de assuntos da BV-FAPESP. Dessa forma, para a adoção de novos termos se faz necessária a verificação das grafias e descrições corretas dos termos que serão adicionados à lista de assuntos. Também é essencial estabelecer um trabalho de colaboração com os desenvolvedores dos vocabulários controlados utilizados,



como contribuição para que novos termos sejam apreciados. Um estudo sobre a maneira de hierarquizar os termos atualmente disponibilizados em lista na BV-FAPESP está sendo desenvolvido, assim como a análise da melhor forma de tornar a lista de assuntos pública, para auxiliar os pesquisadores na escolha das palavras-chave que descrevem suas pesquisas. Ainda, as ferramentas existentes na BV-FAPESP são interessantes para auxiliar a gestão de assuntos e oportunizar o compartilhamento de soluções simplistas, porém funcionais e qualitativas, para que sirvam de base às demais bibliotecas virtuais e digitais, no desafio de tornar a recuperação da informação em suas bases de dados mais consistentes.

The importance of the practice of indexing for information retrieval: BV-FAPESP experience

Abstract: This article reports the experience of the FAPESP Virtual Library (BV-FAPESP) regarding the indexing process, emphasized in 2012 to the informational processing of reference records. Relevant to the organization and retrieval of information in the database of BV-FAPESP, available on the Web since 2005, the issues are essential elements of thematic representation of contents and contribute to the quick access to information, concerning research grants and fellowships in science, technology and innovation, supported by the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP).

Keywords: Virtual libraries; indexing; Matters; Information processing; Information retrieval.

REFERÊNCIAS

ARAÚJO JÚNIOR, R. H. **Precisão no processo de busca e recuperação da informação.** Brasília: Thesaurus, 2007.



CAMPOS, M. L. M. et al. Web semântica e a gestão de conteúdos informacionais. In: Marcondes et al. (coord.). **Bibliotecas digitais: saberes e práticas**. Salvador: EDUFBA; Brasília: IBICT, 2005.

FUJITA, M. S. L. A identificação de conceitos no processo de análise de assunto para indexação. **Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, v. 1, n. 1, p. 60-90, jul.-dez. 2003. Disponível em: < <http://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/2089> >. Acesso em: jun. 2015.

FUJITA, M. S. L. A política de indexação para representação e recuperação da informação. In: LEIVA, I.G.; FUJITA, M. S. L. (ed.). **Política de indexação**. São Paulo: Cultura Acadêmica; Marília: Oficina Universitária, 2012. p. 17-28.

LAFER, C. A peculiaridade da autonomia da FAPESP. **O Estado de S. Paulo**, Opinião. São Paulo, 17 ago. 2014. Disponível em: < <http://opinioao.estadao.com.br/noticias/geral,a-peculiaridade-da-autonomia-da-fapesp-imp-,1545142> >. Acesso em: 17 ago. 2014.

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

Informações dos Autores

Fabiana Andrade Pereira

Bibliotecária da FAPESP

E-mail: fpereira@fapesp.br

Rosaly Favero Krzyzanowski

Coordenadora do Projeto Biblioteca Virtual da FAPESP

E-mail: rosalyfk@fapesp.br

Thais Fernandes de Moraes

Supervisora do Centro de Documentação e Informação da FAPESP

E-mail: tmorais@fapesp.br



Juliana Calherani

Bibliotecária da FAPESP

E-mail: jcalherani@fapesp.br

Agradecimento: Agência financiadora Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) (processo nº 2014/50982-7).

