

UNIVERSIDADE DE SÃO PAULO  
FACULDADE DE FILOSOFIA CIÊNCIAS E LETRAS DE RIBEIRÃO PRETO  
CURSO DE BIBLIOTECONOMIA E CIÊNCIA DA INFORMAÇÃO

**Leitura e vocabulário controlado do documento do Café**

Mediação e Recepção da Informação  
Profa. Dra. Silvia Maria do Espírito Santo

Aline Amorim – 10883300  
Arinara Ornaghi – 10731745  
Pietra Vitoreti - 10731807

Ribeirão Preto 2020

O seguinte resumo tem como referência bibliográfica o artigo científico denominado “Leitura e vocabulário controlado do documento do Café”, de autoria de Silvia Maria do Espírito Santo. Publicado em 2015, pelo periódico de ciências e tecnologia de informação e comunicação, “prisma.com”, vinculado ao Centro de Investigação Transdisciplinar Cultura, Espaço e Memória (CITCEM), com sede na Faculdade de Letras da Universidade do Porto. O escrito foi dividido em quatro partes, além da apresentação e considerações parciais, às quais pretendem tratar do paradoxo da busca, das etapas da política da indexação dos documentos históricos, dos propósitos para política de construção de Vocabulário Controlado, e dos conceitos coletados do passado; nessa dada ordem.

Segundo a autora, o objetivo do artigo seria o de

relatar uma experiência com leitura dos registros documentais, sobre a temática da História do Café, no ambiente rural brasileiro, e apresentar um esboço para elaboração de proposta para política de indexação do contexto documental da produção cafeeira no Brasil, no período final do século XIX e início do XX (ESPÍRITO SANTO, 2015, p. 4)

Para alcançá-lo, o estudo feito previamente partiu de dois fenômenos da organização da informação: controlar e represar informações. Embora, na área da ciência da informação, pudéssemos discorrer longamente acerca desses conceitos, iremos nos conter ao principal papel que tais fenômenos vieram a desempenhar na experiência relatada. *Controlar* as ambiguidades terminológicas é essencial a quem pretende desenvolver um vocabulário controlado. Enquanto isso, *represar* deve ser a conduta do indexador quando, propriamente embasado, compreender que determinada informação não dialoga com os interesses e motivações do vocabulário em construção.

Tendo em vista o atual contexto no qual nos deparamos, envoltos por constante evolução dos aparatos tecnológicos e das Tecnologias da Informação e Comunicação (TIC's), é natural que possamos nos sentir sufocados pelo intenso e constante fluxo de informações produzidas a cada minuto, em diferentes plataformas. Exatamente por isso, não pode-se esperar que simplesmente atirar o leitor em meio a esse fluxo incontrolável e desordenado, será o suficiente para que ele chegue ao conhecimento de que necessita. Faz-se necessária a organização da informação pautada em critérios científicos, o que implica na existência de profissionais qualificados.

Por essa razão, e também pela importância sócio-histórica que a produção e comercialização de café, nosso “ouro verde”, já representou na sociedade brasileira, é de extrema importância que continuem sendo realizadas pesquisas que têm como objeto de

estudo conjuntos documentais das fazendas de Café no Brasil. Uma das colaborações da autora referenciada resultou no artigo sobre o qual discorreremos agora, que virá a se aprofundar nos procedimentos e metodologias para o desenvolvimento de vocabulários controlados de registros documentais do Ciclo do Café, que estejam pautados na qualidade científica, e não nos moldes do acaso.

A segunda seção do artigo tem início expondo as etapas necessárias para o desenvolvimento de uma indexação pertinente de documentos, sendo elas 1) metodologia, 2) procedimentos técnicos e 3) capacidade de investigação e síntese. Falando um pouco mais sobre essa primeira, podemos refletir sobre sua importância. Por que é necessário que, em um projeto, exista uma metodologia afirmada, validada pela instituição na qual está inserida, e seguida por aqueles que desempenham suas funções? Simples, pois a metodologia é um dos fatores que atribui caráter científico ao projeto em andamento. Sem ela, os caminhos traçados para chegar a determinado fim, tornam-se nublados; e até mesmo os resultados alcançados com e sem sua aplicação, poderão variar?. Da mesma forma, os resultados finais de uma pesquisa ou trabalho só podem ser validados mediante a existência de uma metodologia compatível aos mesmos. Por fim, sem ela

a forma de leitura, análise, síntese e representação dos conteúdos informacionais seguirão parâmetros mercadológicos fugazes, talvez baseados em saberes inscritos e orientados por interesses de mercado, ou do extremo subjetivismo, pouco adequados ao atendimento das necessidades institucionais, dos públicos amplos e especializados nas demandas informacionais (ESPÍRITO SANTO, 2015, p. 6)

Portanto, uma metodologia de indexação deve visar também à padronização dos procedimentos a serem utilizados ao indexar os termos referentes a um determinado sistema de informação, os quais deverão ser seguidos por todos os pesquisadores envolvidos nesta prática. Em seguida, a autora do artigo trata de especificar quais seriam esses procedimentos; sendo, basicamente, três: 1) leitura documentária, 2) síntese documentária e 3) representação documentária. No entanto, sugere ainda um quarto, de grande relevância porém frequentemente negligenciado, que seria a revisão da literatura especializada do assunto que se representa.

Trataremos agora, de forma breve, do que tange à leitura documentária. Esta, consiste na utilização de estratégias de leitura para a identificação, por parte do bibliotecário, das partes do texto nas quais se encontram as informações verdadeiramente relevantes para o propósito adotado em cada caso. Tais estratégias sofrem mudanças de acordo com alguns fatores, como por exemplo, o nível de especificidade do texto, e a

superestrutura textual (definida pelos constituintes básicos de um determinado tipo de texto). Cabe ao indexador averiguar quais as técnicas apropriadas para o material a ser trabalhado.

Dito isso, podemos começar a pensar o caso específico dos documentos históricos. Histórica do Café é a antiga denominação dada às regiões situadas ao Oeste da Corte Imperial do Rio de Janeiro. Atualmente, tal localização compreende a Região Administrativa de Ribeirão Preto e parte de Campinas. Foram selecionados documentos referentes a este contexto para o exercício relatado no artigo de Espírito Santo (2015). No acervo de um centro de documentação de uma antiga fazenda cafeeira, foram encontrados diversos tipos documentais, datados de 1890 a 1930, entre eles livros, cartas, mapas, atas, relatos, registros econômicos e diários; além dos documentos avulsos, como bilhetes, ofícios, notas e desenhos, sendo estes fontes de informações pessoais, mercantis e sócio-históricas.

É preciso ter em vista que a discussão elaborada no artigo do qual se trata esse resumo faz referência ao curso organizado pelo Projeto de Implantação do Centro de Documentação da Fazenda Santa Cecília, ministrado pela Profa. Dra. Maria Cristiane Barbosa Galvão, entre 2012 e 2013, com participação de alunos da Universidade de São Paulo. Circunscrito ao curso está a experiência de indexação de documentos históricos de registros da produção cafeeira da fazenda, liderada pela Profa. Dra. Silvia Maria do Espírito Santo. Sendo assim, podemos dialogar sobre a realização da leitura documentária em outras fazendas antigas de café.

Se for este o caso, a autora considera necessário salientar dois fatos com os quais a equipe responsável pela indexação e construção de um vocabulário poderá se deparar ao trabalhar com essa categoria de documentos. A primeira seria a necessidade de que um projeto desse tipo leve em conta a leitura integral dos textos selecionados do arquivo no qual se pretende trabalhar, já que somente ao avaliar de forma completa cada objeto material, poderão ser verificadas as ocorrências de assuntos presentes no corpus escolhido. Ademais,

Faz-se notar que em cada objeto material existem informações que foram registradas uma única vez e em uma só linha. Assim, toda informação encontrada em tais suportes materiais tem potencial relevância para a pesquisa histórica e científica (ESPÍRITO SANTO, 2015, p. 8)

O segundo fator é o de que, após a seleção e análise dos documentos, nem sempre será possível dominar com propriedade o “português antigo”, como chama a autora,

datado na linguagem da época em que o país se encontrava em meio a República Velha, primeira fase da república brasileira. São comuns também as complicações geradas pelos estilos caligráficos, de difícil compreensão, dos registros manuscritos. Nesse caso, é importante consultar um paleógrafo, o que ressalta o aspecto interdisciplinar da Ciência da Informação.

Após a devida leitura documentária, tem início o processo de escolha das palavras-chave, provenientes dos materiais analisados. Aqui, a metodologia aplicada priorizou o seguinte: 1) Substantivos, 2) Substantivos simples, 3) Substantivos compostos, quando imprescindível, 4) Substantivos no singular, exceto para grupos humanos. Também foram adotados substantivos próprios.

Em seguida, será realizada a representação documentária em um movimento de tradução, “etapa na qual os termos serão traduzidos para os termos de uma linguagem documental que compatibilizará tanto os termos indexados quanto os termos que identificam as necessidades dos usuários” (FUJITA, 2013, p. 43). Far-se-á necessária a habilidade de compreensão entre o conteúdo do texto identificado e a linguagem documentária adotada, visando evitar uma tradução generalista.

A terceira seção do artigo propõe-se a tratar dos propósitos para a política de construção de um vocabulário controlado. A autora cita alguns, tais como a harmonização da linguagem de um sistema de informação especializado; o fornecimento de consistência ao processo de indexação; a explicitação das relações semânticas entre termos descritores; e o fornecimento de chaves para a recuperação da informação mais compatíveis com a linguagem dos usuários, assim como dos conteúdos indexados. Em seguida, disserta brevemente sobre a estrutura e o funcionamento de um vocabulário controlado, que

possui uma quantidade restrita de unidades constituintes (termos descritores) e delimitação do significado destas unidades (conceitos), arrançadas segundo uma ordem conhecida e estruturada segundo relações lingüísticas, relações lógicas, relações ontológicas e relações associativas. Dada a sua complexidade, o vocabulário controlado tende a contemplar contextos informacionais e comunicacionais específicos, também chamados de domínios e subdomínios (ESPÍRITO SANTO, 2015, p. 11)

Dessa forma, é necessário que os termos descritores da área da qual pretende tratar o vocabulário controlado sejam retirados de alguma fonte segura. Nesse caso, o indexador possui duas opções: seguir a *garantia literária* (selecionar os termos e/ou seus conceitos a partir da literatura do domínio) ou a *garantia de uso* (selecionar os termos e/ou seus conceitos a partir das demandas dos usuários daquele sistema de informação). Um dos

desafios da construção de vocabulários referentes a Histórica do Café reside aqui, já que os usuários são contemporâneos, mas a documentação não.

Sendo assim, ao adotar a garantia literária, percebe-se que, muitas vezes, os termos descritores extraídos dos documentos não são mais utilizados atualmente, e não se encontra seus conceitos na literatura dos dias de hoje; por isso, tal garantia deve ser buscada nas fontes bibliográficas referentes à época. Olhando a outra face da moeda, ao adotar a garantia de uso, a fonte oral torna-se uma das mais valiosas para a extração dos termos e seus conceitos, e deverão ser consultados agentes sociais diretamente ligados à produção cafeeira.

A autora demonstra, então, que os termos e conceitos a serem preferidos na construção desse gênero de vocabulário controlado, são os do passado. Assim, entre “Phosphoro” (passado) e “Fósforo” (contemporâneo), o primeiro será o termo preferido e o segundo, o termo não preferido. Já no caso do termo “Cravas”, cujo conceito do ciclo do café não é conhecido, mas o conceito empregado na contemporaneidade é; tal termo permanecerá sem conceito, até que se encontre algum validado pela garantia literária ou de uso. O termo “Canga” tem seus conceitos conhecidos, mas esses diferem entre si; nesse caso, o conceito a ser associado ao termo será aquele utilizado durante o Ciclo do Café.

Para a coleta de termos, foram utilizadas fichas terminológicas, por meio das quais será feita a análise e sistematização dos termos descritores. O conteúdo a ser inserido na ficha deve se basear no conceito coletado (por meio da garantia literária ou de uso), e não pelo conhecimento tácito do indexador. Tal fato pretende padronizar, na medida do possível, a metodologia empregada no preenchimento das fichas, visando diminuir a influência de saberes vinculados a outros contextos, assim como as ambiguidades e discrepâncias.

Nem todo conceito fornecerá, de imediato, todas as informações necessárias ao completo preenchimento de sua ficha terminológica. Nesse caso, será preciso desenvolver várias fichas, até que se alcance um mapeamento conceitual e relacional confiável. A partir desse volume de fichas, deve-se produzir uma ficha síntese, avaliando como se apresentam os conceitos coletados. Após a finalização do trabalho com as fichas terminológicas e as fichas síntese, o próximo passo seria sua inserção em um software específico para a criação de vocabulários controlados, a ser escolhido pela equipe envolvida no projeto.

Por fim, a autora conclui seu texto com algumas reflexões quanto a Ciência da Informação, quanto à informação como semente da qual poderá vir a brotar o

conhecimento, e da importância de sua organização por parte de um profissional qualificado para que esta possa cumprir de fato seu papel. Dessa forma, unindo em um nó tudo que foi abordado no artigo até agora, Espírito Santo (2015) escreve:

Esse processo do ato de pesquisar e estratificar a informação, comumente é aceito como processo de Tratamento da Informação, daquilo que, uma vez avaliado foi considerado o que é “essencial” para o conhecimento, tal processo dependa, obviamente, de normas, regras e dos clássicos critérios das classificações universalizadas e modais presentes na tecnologia da informação. O Café é um bom exercício (ESPÍRITO SANTO, 2015, p. 13-14)

---

## BIBLIOGRAFIA

ESPÍRITO SANTO, S. M. Leitura e vocabulário controlado do documento do Café. **prisma.com**, [s.l.], [s.v.] n. 27, p. 3-17, 2015. Disponível em: <<http://ojs.letras.up.pt/index.php/prisma.com/article/view/1851/1684>>. Acesso em: 18 maio 2020.

FUJITA, M. S. L. A representação documentária no processo de indexação com o modelo de leitura documentária para textos científicos e livros: uma abordagem cognitiva com protocolo verbal. **PontodeAcesso**, Salvador, v.7, n.1, p.42-66, abr. 2013. Disponível em: <<https://portalseer.ufba.br/index.php/revistaici/article/view/8135/5807>>. Acesso em: 20 maio 2020.