

Linguagens de indexação em contextos cinematográficos: a experiência de elaboração do tesouro eletrônico do cinema brasileiro

Maria Aparecida Moura

Professora Adjunta da Escola de Ciência da Informação da UFMG

Luiz Nazário

Professor da Escola de Belas Artes da UFMG.

Daniela Cristina da Silva Rodrigues; Elisabete Quatrini Vieira; Iris da Silva; Lúcia Mara Barbosa de Oliveira Paoliello.

Bibliotecárias formadas pela Escola de Ciência da Informação da UFMG.

Relata a experiência de elaboração do Tesouro Eletrônico do Cinema Brasileiro, estruturado para auxiliar a organização e a recuperação de informações produzidas em contextos cinematográficos. Sua realização partiu da constatação da ausência de instrumentos especializados no tratamento temático dos acervos cinematográficos brasileiros e das crescentes demandas por linguagem especializada realizadas por centros de memória e referência audiovisual consolidados no país. Objetivou desenvolver um instrumento de indexação capaz de representar tematicamente os acervos cinematográficos específicos bem como o conhecimento produzido sobre o cinema brasileiro. Foram estabelecidas ações para mapear tematicamente a área com o objetivo de identificar os conceitos principais vinculados ao domínio e sua rede de relações e realizados estudos fundamentados na literatura da área, seguido de cotejamento conceitual baseado no acervo fílmico e nas publicações mantidas pela Ophicina Digital do Departamento de Fotografia, Teatro e Cinema da Escola de Belas Artes - UFMG. Propõe a estrutura facetada do instrumento com base na literatura e no conhecimento do especialista em cinema brasileiro integrante da equipe. Tais medidas visaram assegurar que os princípios fundamentais de garantia literária e de uso norteassem a elaboração do tesouro. Foram realizados pré-testes de indexação para assegurar a especificidade do instrumento e avaliar sua performance em contextos específicos. O tesouro resultante é composto por 1040 termos organizados em 34 facetas específicas.

Palavras-chave: *Tesouro eletrônico; Linguagem de indexação-elaboração; Cinema brasileiro*

Recebido em 24.09.2004

Aceito em 28.04.2005

Introdução

A ampliação dos domínios e a intensa disponibilização de conteúdos diversificados e integrais na *WEB* aumentaram significativamente a demanda por instrumentos de organização e tratamento da informação. Em função dessa oferta, os motores de busca, associados ao contexto digital, tiveram reconhecido destaque exatamente por proverem o acesso aos conteúdos informacionais disponíveis na rede, através da organização de estratégias de estruturação e tradução das necessidades de informação dos usuários em linguagem verbal. Contudo, a insuficiência das soluções dadas por esses instrumentos, em termos do fornecimento de respostas precisas às demandas informacionais enunciadas pelos usuários em interação com a rede, tornou crescente o interesse de pesquisa relativo aos processos de elaboração de linguagens de indexação.

Paralelamente, o campo audiovisual também vivenciou nos últimos anos uma considerável alteração de seus processos de criação e disseminação de produtos. Nesse sentido, ampliaram-se as demandas por instrumentos de organização e tratamento da informação que pudessem apoiar, em termos de acesso à informação, as diversas etapas dos processos da produção cinematográfica.

Este artigo objetiva descrever a experiência de criação do Tesouro Eletrônico do Cinema Brasileiro¹, elaborado para auxiliar a organização e a busca de informações produzidas em contextos cinematográficos. A atividade foi uma realização conjunta do Departamento de Organização e Tratamento da Informação da Escola de Ciência da Informação - ECI/UFMG e da *Ophicina* Digital do Departamento de Fotografia, Teatro e Cinema da Escola de Belas Artes – FTC/EBA/UFMG.

Linguagens de Indexação: aspectos históricos e perspectivas

A linguagem tornou-se, ao longo dos últimos anos, um tema de interesse central no âmbito da ciência da informação. Essa intensificação do interesse passou a incorporar a necessidade da efetiva articulação de uma perspectiva mais ampla desse conceito, exigindo a aproximação da CI com as demais áreas de conhecimento que se dedicam ao estudo do fenômeno da linguagem na atualidade. Destaca-se, nesse contexto, a filosofia da linguagem, a análise do discurso, a teoria da terminologia, a teoria do conceito, a semântica, dentre outras.

A linguagem é compreendida, no âmbito desse artigo, como um conjunto organizado e compartilhado de signos que permitem a expressão do pensamento. Nesse sentido, o acesso ao signo sempre exigirá a mediação e a interposição de múltiplas linguagens.

Em Platão encontram-se os argumentos para compreender a linguagem como instrumento através das teses da naturalidade e da convencionalidade. Para o filósofo, apenas o uso e, por conseguinte, a convenção pode prover de significado as palavras. Entretanto, o uso é a escolha repetida ou convalidada que levou a forjar determinado instrumento lingüístico e, assim como todos os outros instrumentos, os lingüísticos também podem resultar mais ou menos perfeitos e adequados à finalidade. (ABBAGNANO, 2000, p. 622).

¹ <http://www.eci.ufmg.br/tesaurodecinema>

A explosão informacional, experimentada nos últimos anos devido às inovações tecnológicas, tornou evidente a utilidade das linguagens de indexação em contextos mais amplos, embora o controle de vocabulário seja uma preocupação e uma prática consolidada na área da ciência da informação há décadas.

Nesse sentido, a área lidera os estudos teóricos e os experimentos relativos à análise dos processos de produção da informação com vistas à criação de instrumentos de organização e recuperação da informação cada vez mais adequados às realidades dos grupos que deles fazem uso.

Parece consensual que um dos aspectos que mais carece de estudos específicos refere-se à elaboração de linguagens de indexação mais adequadas ao provimento e tradução das necessidades dos usuários em interação com sistemas de informação tradicionais ou eletrônicos.

No processo de geração automática de tesouros Kimoto & Iwadera (1990) salientam que o processamento automático da linguagem natural apresenta dificuldades em relação à análise sintática e semântica.

Gomes e Campos (2004, p. 1-2) afirmam que os métodos automáticos, ao se basearem na frequência de ocorrência de termos, tornam os instrumentos de indexação ineficazes quanto à natureza semântica dos termos. Nesse sentido, os estudos relacionados aos fundamentos teóricos das linguagens de indexação (CAMPOS, 2001; CAMPOS & GOMES, 2003, 2004; RIVIER, 1992; SVENONIUS, 1986, 2000) realizados no Brasil e no exterior têm contribuído para dar maior consistência conceitual aos experimentos computacionais relativos à criação de tesouros. Em virtude desse fato, buscou-se assegurar que os princípios fundamentais de garantia literária e de uso pudessem nortear a elaboração do tesauro. Cabe destacar ainda que a base terminológica utilizada na elaboração desse instrumento originou-se de um profundo estudo realizado no domínio.

A garantia literária é um conceito introduzido por *Wyndam Hulme* em 1911. Segundo esse princípio os termos que compõem uma linguagem de indexação devem ser empiricamente derivados da literatura da área que se pretende descrever tematicamente. A garantia de uso visa privilegiar a abordagem temática adotada pelos usuários reais em situação de recuperação da informação.

No processo de elaboração do tesauro, além dos princípios de garantia literária e uso, foram consideradas algumas experiências nacionais e internacionais de implementação de tesouros eletrônicos. Dentre eles, destacam-se os projetos nacionais: o Tesauro Brasileiro de Educação - BRASED² mantido pelo INEP, o Tesauro de Folclore e Cultura Popular Brasileira³ mantido pelo Centro Nacional de Folclore e Cultura Popular – CNFCP- do IPHAN e o Tesauro Eletrônico do Mundo do Trabalho mantido pela Rede Unitrabalho. Dentre as experiências internacionais destacam-se: o Tesauro da Unesco⁴ e o tesauro *multílingue Eurovoc*⁵ mantido pela União Européia em 16 idiomas.

As linguagens de indexação são instrumentos elaborados intencionalmente para proceder à mediação entre os processos de organização e recuperação da informação e resultam do trabalho coletivo desenvolvido entre os profissionais da informação e os especialistas das áreas a que se destinam. Para fazer isso, as linguagens de indexação adotam o postulado da monoreferencialidade a partir do qual restringe-se a significação de um dado termo. Assim, o termo deve representar um conceito relativo à determinada área de conhecimento e com ele guardar uma relação unívoca.

² <http://www.inep.gov.br/pesquisa/thesaurus/>

³ <http://www.museudofolclore.com.br/tesauro/Index.htm>

⁴ <http://www.ulcc.ac.uk/unesco/>

⁵ <http://europa.eu.int/celex/eurovoc/>

De acordo com Svenonius (2000) existem atualmente inúmeras linguagens de indexação que exibem um variável grau de refinamento. Essas linguagens quando utilizadas nos sistemas de recuperação da informação podem provê-lo de qualidade adicional de modo que por seu intermédio possamos transformar informação em conhecimento. Segundo a autora, quando isso acontece a linguagem de indexação torna-se análoga ao próprio conhecimento. Nesse sentido, as linguagens de indexação como representação do conhecimento são fundamentais não apenas na recuperação da informação, mas também em outros contextos que referenciam o conhecimento estruturado tais como, o processamento automático da linguagem e a engenharia de conhecimento.

Dessa forma, as linguagens de indexação em geral e os tesouros em particular mapeiam conceitualmente os termos e buscam efetivar e tornar evidentes as relações existentes entre eles. A partir de um diagrama de relacionamento entre termos, revelam-se os aspectos conceituais envolvidos. Dentre essas relações destacam-se as relações de hierarquia (todo/parte, gênero/espécie), de oposição, de funcionalidade, de associação e de equivalência.

As linguagens de indexação atuam nos sistemas de informação para orientar o indexador sobre quais os melhores termos para a representação do assunto de um documento e para guiar os pesquisadores sobre o modo de escolher os termos indexados que representam no sistema o assunto procurado. Possuem vocabulário e sintaxe próprios. Têm ainda como função recuperar documentos com conteúdo semelhante, recuperar documentos relevantes sobre um assunto específico, recuperar documentos por grandes áreas de assunto, possibilitar a conversão dos termos de indexação entre diferentes linguagens, auxiliar na escolha do termo adequado para a estratégia de busca, representar o assunto de maneira consistente permitindo a compatibilidade e o diálogo entre a linguagem do autor, do indexador e a do pesquisador.

Hoje, tem sido cada vez mais complexo o desenvolvimento de tais linguagens, visto que elas objetivam, em última instância, proceder à mediação entre o usuário e sua necessidade de informação. Dada a dinamicidade dos processos de produção de conhecimento, torna-se cada vez mais difícil prover os sistemas de representações que de fato reflitam esse conhecimento.

Ao analisar os aspectos comuns existentes entre a teoria do conceito, a terminologia e a classificação facetada, Campos (2001, p.117) enfatiza a importância dessas para a construção de linguagens e o movimento contraditório que esse esforço de construção exige. Segundo a autora, as abordagens onomasiológica e semasiológica, presentes no contexto de elaboração de linguagens, atuam no sentido de orientar a concepção da linguagem de indexação no que concerne à dinamicidade própria da linguagem, tendo em vista conferir maior representatividade ao termo nela incorporado.

Nos tesouros, há a preocupação com a precisão dos termos, dando ênfase ao termo preferencial. Por estarem em um ambiente de recuperação de informação, os termos devem ser submetidos a controles terminológicos rígidos. (CAMPOS, 2001, p. 129).

No estudo desenvolvido, objetivou-se elaborar um tipo de linguagem de indexação - o tesouro. O tesouro é um instrumento de indexação pós-coordenado no qual os conceitos podem ser identificados através da máxima "*diga-me com quem tu andas e eu te direi quem és*". Nesse sentido, os tesouros

permitted to comprehend conceptually a given field of knowledge through a relational structure of terms arranged schematically.

Trata-se de uma metalinguagem composta por um sistema de símbolos que utiliza a linguagem natural como um esquema de codificação da informação. Derivam desse esquema parte da complexidade do gerenciamento dos sistemas de recuperação da informação com o uso de linguagem verbal.

Na medida em que a língua é um organismo vivo, torna-se complexo elaborar um instrumento de representação da informação com base na linguagem natural. Tal complexidade pode ser atribuída à permanente mutação sócio-histórica que tem na linguagem uma de suas principais manifestações.

A linguagem cinematográfica

O mundo do cinema possui uma linguagem própria com um vocabulário altamente especializado, que se formou durante os últimos cem anos, assimilando as mais diversas conquistas tecnológicas, desde os primitivos equipamentos de captação e projeção de imagens até os modernos *softwares* de produção gráfica para efeitos especiais. Essa língua, falada e escrita pelos habitantes do *planeta cinema*, tem diversas dicções, diferentes dialetos - os quais provêm, contudo, de dois idiomas básicos: o francês e o inglês. Os cinéfilos de todo o mundo assimilaram o idioma francês porque foram os irmãos *Lumière* os inventores do cinematógrafo, o primeiro aparelho de projeção de imagens para uma platéia, substituindo os aparelhos de projeção individual - invento devidamente patenteado em 1895; e também porque foram franceses os primeiros teóricos a sistematizar o vocabulário da nova arte.

Os cinéfilos assimilaram igualmente o idioma inglês porque, a partir da Primeira Guerra Mundial (1914-1918), que arrasou a Europa, os USA tomaram a dianteira na produção cinematográfica, importando talentos, desenvolvendo tecnologias, fazendo do cinema uma grande indústria que não cessará de renovar a linguagem da câmara acrescentando-lhe, a cada ano, novos termos técnicos especializados. À diferença da literatura e da pintura, o cinema não é feito de palavras escritas ou de imagens pintadas por um único instrumento - caneta, máquina de escrever ou computador; pincel, *spray* ou aerógrafo - e sim de imagens silenciosas ou sonoras compostas por uma multiplicidade de equipamentos convergentes em constante evolução e refinamento: câmara, grua, *stead cam*; moviola, ilha de edição, estação gráfica; mono, *stereo*, *surround*; 16mm, 35mm, cinerama; preto e branco, colorido, *tecnicolor*; trucagem, efeito especial, efeito digital.

Mesmo o mais primitivo dos instrumentos de produção, como a claquete, feita com dois tabletes de madeira, hoje se sofisticou a ponto de incorporar um contador digital - a claquete eletrônica. O cinema cria e renova sua linguagem através da criação e da renovação do conjunto dos seus instrumentos de produção. Cada novo equipamento cria novas possibilidades expressivas que produzem novas expressões cinematográficas carentes de nomenclatura, tarefa assumida pelos técnicos que fazem uso desses instrumentos: é da natureza da linguagem cinematográfica incorporar termos esdrúxulos que nascem mais da prática que da teoria cinematográfica.

Exemplo disso é o verbo *renderizar*, do substantivo *renderização* – palavras derivadas do inglês *render*. Ou o curioso termo bilíngüe *contra-plongée*, derivado do francês *contre-plongée*, palavra composta que manteve o segundo termo no original e traduziu apenas o primeiro para o português. Da mesma forma, a expressão *decupagem* deriva de *découpage*. A palavra câmara, usada de preferência à palavra portuguesa câmara, é o aportuguesamento da palavra inglesa *camera*, provando o gosto dos brasileiros pelos americanismos (basta dizer que em Portugal o *mouse* do computador é chamado de rato). Portanto, o vocabulário cinematográfico é todo feito de estrangeirismos e neologismos, um código técnico em constante mutação.

Cada cinematografia possui as suas idiossincrasias e termos próprios para descrevê-las. Não existe uma tradução consagrada em português para o termo *heritage film*, que não se refere a patrimônio cinematográfico (*film heritage*), mas a um gênero de filme tipicamente inglês. O filme de patrimônio - termo que nos soa tão estranho - refere-se a filmes ingleses que celebram o passado, o patrimônio, a tradição, filmes de costumes com ênfase na identidade nacional. Da mesma forma não há um termo em português que traduza adequadamente a expressão alemã *Heimatfilm*, que designa filmes exclusivamente produzidos na Alemanha (eventualmente na Áustria) associados às tradições locais. Uma tradução aproximada do termo resultaria no esdrúxulo *filme de torrão natal*. Nesses casos, é preferível, a uma tradução literal, manter o termo original com uma nota de rodapé explicando o sentido do termo e o tipo de filme que o encarna. O mesmo ocorre com o *Kulturfilm* alemão, que se poderia traduzir como filme cultural, mas que implicou, durante o *Terceiro Reich*, em propaganda política do regime nazista.

Os nativos do *planeta cinema* falam em *plongée* e *raccord*, *close-up* e *flashback* com a *nonchalance* de *experts* que tiveram o privilégio de nascerem bilíngües. Comunicam-se através de um idioma de base franco-americano enriquecido com termos próprios da língua nacional. Os cinéfilos nascidos no Brasil, que falam oficialmente o português, quando encontram um concidadão de sua pátria de adoção entende-se com ele através de frases onde predominam termos como *fade in*, *make-up*, *travelling*, *glamour*, *star system*, *noir*, *happy end*, *making of* e *fade out*. Cinéfilos do mundo falam um esperanto técnico-cinematográfico que contém alguns termos já popularizados pelos programas de TV sobre cinema e outros que ainda guardam sua aura de mistério para os não iniciados.

O teórico e o historiador de cinema sempre se basearam em livros datados e, sobretudo, na própria memória (freqüentemente traidora e afetada por emoções e circunstâncias do momento) para escrever sobre os filmes. Mas os bancos de dados eletrônicos estão revolucionando a pesquisa no campo cinematográfico. Hoje, o *site* <http://www.imdb.com/> substitui as memórias individuais por uma gigantesca memória coletiva do cinema, realimentada diariamente por milhões de usuários em todo o mundo, numa revisão e atualização constante dos dados. O resultado é que uma elite de críticos está sendo substituída por uma massa de cinéfilos.

A facilitação da pesquisa e seu aprimoramento através do instrumento eletrônico têm sua contrapartida: a banalização da linguagem cinematográfica. O que era antes um sacerdócio, um culto mais ou menos secreto, foi profanado pelo amadorismo. Muito do prazer proporcionado pelo mistério cinematográfico foi diluído pelos meios técnicos de difusão. A sacralidade do cinema foi substituída pela desmistificação em massa. Assim, no centenário de Alfred Hitchcock

o mercado livreiro lançou uma infinidade de títulos sobre o cineasta – biografias, análises, filmografias, diários, roteiros, até notas de produção. Apesar de tudo isso, o melhor livro sobre Hitchcock continuou sendo *Hitchcock/Truffaut – Entrevistas* (1983), ou seja, nem sempre a quantidade de informação em que se mergulha hoje representa ganho qualitativo de conhecimento.

Alguns termos da linguagem cinematográfica de significado aparentemente semelhante, como diretor e cineasta, possuem uma distinção sutil: o diretor é um profissional; o cineasta é um artista. Essa distinção associa-se à teoria do cinema de autor, que os críticos da revista francesa *Cahiers du Cinéma* lançaram nos anos 1960. Nessa época, as discussões sobre a linguagem do cinema eram tão intensas que, mesmo no interior desse grupo, surgiram duas correntes opostas que receberam curiosas designações: os *Hitchcocko-hawksiense* e os *Mac Mahoniens*. Os primeiros, que predominavam, defendiam a causa de Alfred Hitchcock e Howard Hawks, enquanto os segundos, que freqüentavam a sala de arte Mac Mahon, privilegiavam Fritz Lang, Joseph Losey, Otto Preminger e Raul Walsh (Journot, 2002). Hoje, essas discussões parecem bizantinas, uma vez que esses cineastas foram igualmente consagrados pela história do cinema. Da mesma forma, os debates teóricos do cinema russo dos anos 1920 e 1930, com Serguei Eisenstein opondo o seu *cine-punho ao cine-olho* de Dziga Vertov, foram superados pela reconciliação das duas tendências na perspectiva histórica que os valoriza na mesma medida.

Às vezes, o objeto cinematográfico ganha diferentes termos conforme a época: a obra é chamada de fita, película e filme segundo as gerações. Contudo, na atualidade ocorre uma confusão generalizada entre os termos filme e vídeo, porque o público não consegue mais distinguir, pela textura das imagens, o que é uma projeção de película e uma exibição em vídeo. Muitas instituições culturais anunciam mostras de filmes e apresentam vídeos; há cineclubes que só programam filmes em DVD; e mesmo os cinemas iniciam suas sessões com propagandas em vídeo, de modo que o público passa a falar de vídeo como se tratasse de filme. “*Vi um filme ontem na TV*”: na verdade, se entendermos a imagem como um conjunto de informações envolvendo luz, textura, foco, pontos, definição de cor, formato de tela, ritmo, som etc, o que se vê no monitor é apenas uma transmissão empobrecida das informações contidas no filme.

Um filme só pode ser visto numa sala de cinema por meio de um projetor de películas. DVD, vídeo transmitem apenas gravações da informação contida na película obtida por processos conhecidos como telecinagem (vídeo) ou autoração (DVD). O primeiro é a captação da imagem impressa na película em um suporte magnético; o segundo é a gravação dos arquivos digitais num disco versátil digital. Em ambos os processos há uma inevitável perda de informação. Os processos inversos – transferência de dados magnéticos e de dados digitais para a película – foram batizados com nomes ainda mais estranhos: respectivamente, kinescopagem e datacinagem. Enfim, diante da complexidade dos significados e significantes da linguagem cinematográfica, a criação do Tesauro eletrônico do cinema brasileiro pretende dar um primeiro passo na direção da sistematização de termos usados muitas vezes de forma incorreta, imprópria e até contraditória. A implementação desse Tesauro nas bibliotecas e instituições culturais será de grande utilidade para cineastas, atores e técnicos; jornalistas e agentes culturais; estudantes e professores de cinema.

O Tesouro eletrônico do cinema brasileiro

A construção de um vocabulário controlado surgiu a partir de um tópico da disciplina *Elaboração de Linguagens de Indexação*⁶, do Curso de Biblioteconomia da UFMG. A área de cinema foi escolhida e recortada em cinema brasileiro. Como produto dessa disciplina foi elaborado um tesouro experimental monolíngüe com 344 termos (283 descritores e 61 não-descritores). Em etapa posterior, foi elaborado um projeto de ampliação do instrumento construído na fase anterior. Nesse projeto, conforme já salientado, evidenciou-se o público-alvo, os objetivos do instrumento e os princípios metodológicos adotados.

Em virtude dessas ações o instrumento proposto, denominado *Tesouro do cinema brasileiro*, tornou-se mais exaustivo e contou com a co-orientação de um profissional da área pertinente⁷ para possibilitar a garantia de uso. Esse tesouro possui 1.040 termos (983 descritores e 57 não-descritores).

A fase inicial de elaboração desse vocabulário controlado foi o levantamento da literatura relevante na área. Com a ajuda de um historiador especialista em cinema, várias obras foram selecionadas para estudo. A leitura dessas obras possibilitou o contato com importantes aspectos do cinema brasileiro, o que proporcionou uma visão panorâmica de elementos técnicos, históricos e estéticos. Com o aumento do volume de leituras ficou explícita a necessidade de experiência e de conhecimento técnico na área. O grau de complexidade do assunto confirmou a necessidade do trabalho interdisciplinar.

A construção dessa linguagem documentária envolveu também pesquisas em outros instrumentos de indexação da mesma natureza. Foram observados dados como áreas temáticas cobertas, objetivos, fontes de informação pesquisadas, renovação/atualização dos instrumentos, idiomas disponíveis, simbologia adotada, e principais facetas. Em etapa posterior, a equipe de trabalho escolheu conjuntamente os termos que melhor definiam os conceitos veiculados na literatura, conforme a importância desses para seus usuários. Os critérios para escolha dos termos foram definidos em função de sua relevância na área, assim como sua especificidade.

A metodologia adotada, a partir de então, fundamentou-se na concepção do relacionamento entre conceitos através da análise de assunto "*processo por meio do qual o classificador, indexador ou catalogador identifica e determina de que assuntos trata um documento e quais desses assuntos devem ser representados [...]*" (NAVES, 2001, p. 206). Os relacionamentos ocorrem no campo das idéias (conceitualmente), no qual os conceitos são comparados e representados por relações hierárquicas e associativas. No campo semântico, (comparação entre palavras) são representados pela relação de equivalência.

Estudos da área (NAVES, 2001; FUGITA, 1999) ressaltam que a análise de assunto é uma das etapas mais complexas num sistema de recuperação da informação, pois o indexador sofre influência de fatores como a subjetividade, o conhecimento anterior, a preparação profissional e as vivências que influenciam na qualidade de crítica, de associação e de interpretação.

A etapa seguinte consistiu em organizar os termos que comporiam as facetas. O agrupamento desses elementos possibilitou a constituição do conceito, definindo-o como a compilação de enunciados verdadeiros a respeito de determinado objeto, fixada por símbolo lingüístico. (DAHLBERG, 1978).

⁶ Ministrada pela professora Maria Aparecida Moura, Departamento de Organização e Tratamento da Informação da Escola de Ciência da Informação da UFMG.

⁷ Professor Luiz Nazário, coordenador da Oficina Digital do Departamento de Fotografia, Teatro e Cinema da Escola de Belas-Artes da UFMG.

A utilização do glossário foi fundamental para o estabelecimento das similaridades de cada termo, o que possibilitou a inserção do mesmo na faceta mais apropriada. As facetas são em número de 34 e convencionou-se que seriam denominadas da seguinte maneira: Gêneros/Tipos de filmes; ciclos/movimentos do cinema; cinema de animação; antropologia; semiótica/linguagem do cinema; movimentos artísticos/culturais/filosóficos; documentação do filme; legislação do cinema; organizações profissionais/culturais relacionadas ao cinema; fontes de informação em cinema; preservação de filmes; atividades cinematográficas; profissões relacionadas ao cinema; eventos de cinema; ensino; produtoras de filmes; companhias cinematográficas; estúdios de cinema; salas de exibição; produção cinematográfica; local de filmagem; processos de edição; sonorização; tecnologia do cinema; sistema de cor; movimentos de câmera; movimentos de cena; enquadramentos de cena; cinematografia; origem/evolução do cinema; equipamentos e suportes utilizados em filmagem; tempo de projeção; processos de produção e exibição cinematográfica.

A escolha das facetas levou em consideração a compreensão que os especialistas possuem da área.

Para garantir a longevidade do tesauro, considerou-se fundamental o desmembramento dos nomes das instituições e dos eventos em tabelas de autoridade, permitindo assim, em qualquer momento, a atualização do tesauro através da inclusão de novos descritores sem afetar a estrutura do mesmo.

As facetas desmembradas são as seguintes: organizações profissionais / culturais relacionadas ao cinema (agências reguladoras de cinema, associações relacionadas ao cinema, órgãos governamentais relacionados ao cinema e sindicatos relacionados ao cinema), eventos de cinema, produtoras de filmes, escolas/instituições de ensino em cinema, legislação referente ao cinema brasileiro, cineastas/diretores brasileiros e atores/atrizes brasileiros. Após essa etapa foi realizado um pré-teste, através de alguns documentos, para mensurar a efetiva utilização desse instrumento no tratamento da informação para a área.

Cabe destacar que, na realização do projeto do Tesauro eletrônico do cinema brasileiro, considerou-se fundamental a realização das etapas apresentadas na FIG.1, visto que elas não ocorrem de forma isolada, mas interagem mutuamente. Além disso, salienta-se que as atividades de monitoramento e avaliação são etapas fundamentais para garantir a consistência, representatividade e longevidade da linguagem elaborada.

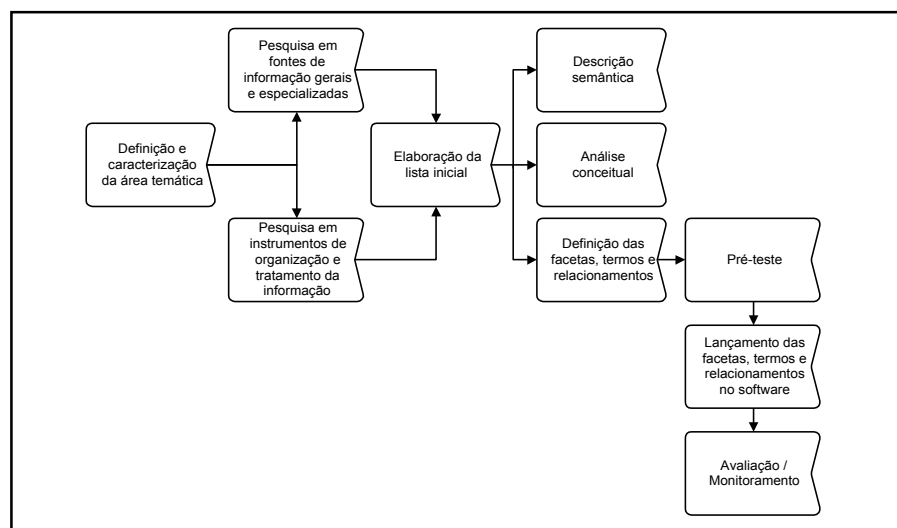


FIGURA 1 - Representação gráfica das etapas de elaboração do tesauro.

As atribuições da equipe na elaboração do instrumento

A elaboração do Tesouro eletrônico de cinema brasileiro permitiu a concretização de uma atividade de cooperação acadêmica da máxima relevância. Trata-se de uma contribuição fundamental à organização dos acervos cinematográficos brasileiros gerada a partir da conjugação de esforços e conhecimentos das artes e da ciência da informação. O processo de elaboração do instrumento contou com inúmeras ocasiões em que os conhecimentos das duas áreas envolvidas convergiram no sentido de consolidar um projeto de natureza interdisciplinar. Conforme já salientado anteriormente, esse trabalho tomou por base os princípios da garantia literária e de uso. Entretanto, esses princípios não são facilmente apreensíveis no domínio proposto. De modo semelhante, a dinâmica de operacionalização conceitual de um instrumento de indexação não se apresenta de maneira evidente entre os profissionais sem uma formação específica. Nesse sentido, foi necessária a realização de diversas reuniões com o intuito de promover e sedimentar o entendimento da equipe acerca das variáveis que estavam implicadas na efetivação do projeto.

O especialista em cinema brasileiro e sua equipe de pesquisadores foram fundamentais no sentido de promoverem discussões conceituais específicas em que as particularidades da linguagem cinematográfica foram evidenciadas. Tal dinâmica permitiu a ampliação do entendimento da equipe acerca do domínio de conhecimento perscrutado e a geração de uma proposta mais corrente com os usos e as particularidades da linguagem verificadas no campo.

Esforço semelhante foi realizado pela especialista em ciência da informação e sua equipe de pesquisadores no sentido de evidenciar o papel da linguagem de indexação nos processos de organização e recuperação da informação. Além disso, foram ressaltados os princípios teóricos e metodológicos adotados na efetivação de um instrumento dessa natureza.

As ações de seleção e descrição semântica dos termos bem como a proposição da estrutura facetada foram definidas com base nas ponderações apresentadas pelas duas equipes envolvidas de modo que fosse possível gerar um instrumento robusto o suficiente para representar as informações geradas em contextos cinematográficos.

Instruções de uso do tesouro

A finalidade do *Tesouro eletrônico do cinema brasileiro* é proporcionar tratamento formal para a área em questão, minimizando ambigüidades e facilitando a comunicação entre os profissionais envolvidos. Também atendem aos conhecedores dessa área e ao indexador que não tenha necessariamente uma formação específica. Foram adotados os significados das convenções e abreviaturas utilizados no *software Thesaurus Webchoir Educational - TCS-8*:

BT (*Broader Term*): Termo genérico. O termo introduzido após este símbolo representa um conceito de conotação mais ampla.

NT (*Narrower Term*): Termo específico. O termo introduzido por esse símbolo representa um conceito de conotação mais específica.

SN (*Scope note*): Nota de aplicação. Explicação concisa sobre o modo de emprego de um descritor. Também chamada de nota explicativa ou nota de escopo.

UF (*Used for*): Usado por. O termo introduzido após este símbolo é um não-descritores ou termo não preferido.

USE (*Use*): O termo introduzido após este símbolo é o descritor, ou termo preferido.

RT (*Related Term*): Termo relacionado. O termo introduzido por esse símbolo representa um conceito que possui algum tipo de relação com aquele sob o qual está submetido.

RTO (*Related Term Opposite*): Termo Relacionado Oposto. O termo introduzido por esse símbolo representa um conceito contrário, contraditório ou oposto.

A novidade, em termos de relacionamentos, foi a presença do termo relacionado oposto que permitiu a oposição entre termos.

O tesauro possui três classes gerais fundamentais de relacionamento: relacionamentos semânticos ou de equivalência, relacionamentos lógico-hierárquicos, relacionamentos associativos.

O *Tesauro* possui uma apresentação sistemática com um índice alfabético apresentado em duas partes. A primeira é das categorias ou hierarquias de termos ordenados de acordo com seus significados e relacionamentos lógicos. Contém o relacionamento genérico/específico indicado pela posição dos termos dentro de sua hierarquia e pelas diferentes margens. Inclui também notas de explicação e referências recíprocas a termos equivalentes e associativos. A segunda é o índice alfabético que leva o usuário ao ponto apropriado da seção sistemática. Contém também as notas de explicação e as referências recíprocas entre sinônimos e demais relacionamentos que não o genérico/específico. O número total de termos (descritores e não-descritores) constituintes do tesauro é de 1.040 sendo que são 983 descritores e 57 não-descritores. A disposição dos termos segue ordem alfabética palavra por palavra, como no exemplo:

ATIVIDADES CINEMATOGRAFICAS

Assistência de produção

Cenografia

Cinematografia (filme)*

Consultoria técnica

Crítica de cinema

Direção

Distribuição cinematográfica

Dublagem

Edição

A escolha da forma singular prevaleceu para os termos.

Utilizou-se o símbolo * (asterisco) para representar os termos com grafias iguais em mais de uma faceta, mas que correspondem a contextos diferentes.

Ex. Cinesdistri (estúdio)*

Cinesdistri (companhia cinematográfica)*

Aos conceitos diferentes, que foram expressos pela mesma forma lingüística, acrescentou-se um qualificador a cada forma.

Exemplo: Girafa (equipamento)

Como princípio geral, as siglas foram usadas como descritores. Uma das razões é que muitas siglas são mais reconhecidas na área do cinema do que a expressão completa. A forma por sigla figurou como descritor, com uma remissiva da grafia completa.

Software utilizado na elaboração do tesauro

Atualmente existem vários *softwares* com características de registro e de consulta dos termos, ou seja, de ferramenta de construção de tesouros. Para o gerenciamento do Tesauro eletrônico de cinema brasileiro foi utilizado o *software Thesaurus Webchoir Educational*, TCS-8, versão 1.16¹ que permite:

Gerenciar com facilidade a estrutura do Tesauro com criação, edição ou remoção de termos e relacionamentos (FIG. 2).

Recuperar os termos, através do sistema de busca (FIG. 3).

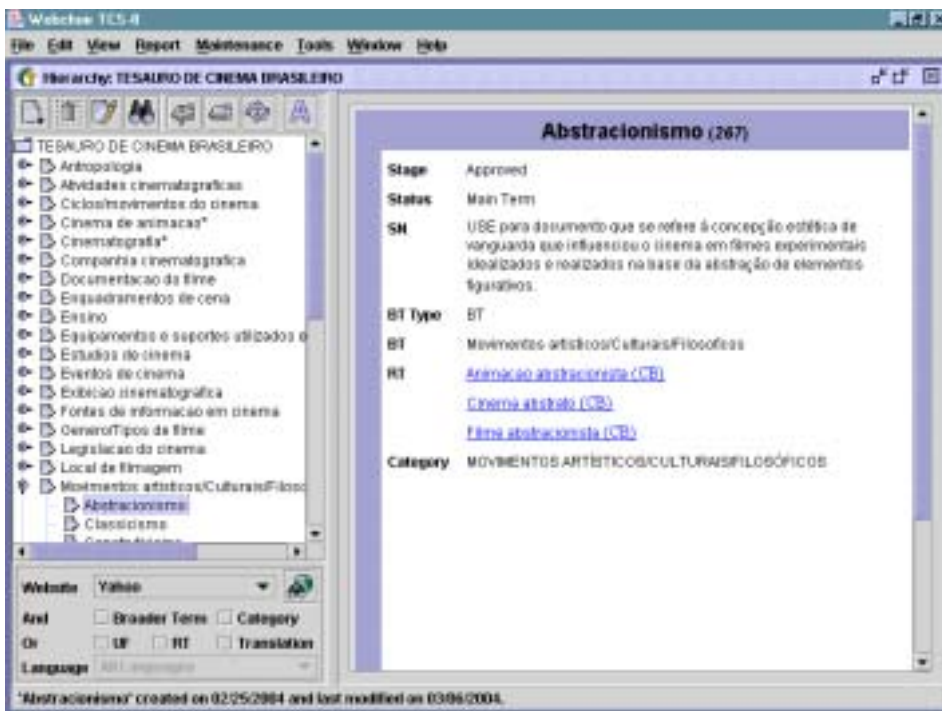


FIGURA 2 - Estrutura do Tesauro eletrônico do cinema brasileiro.

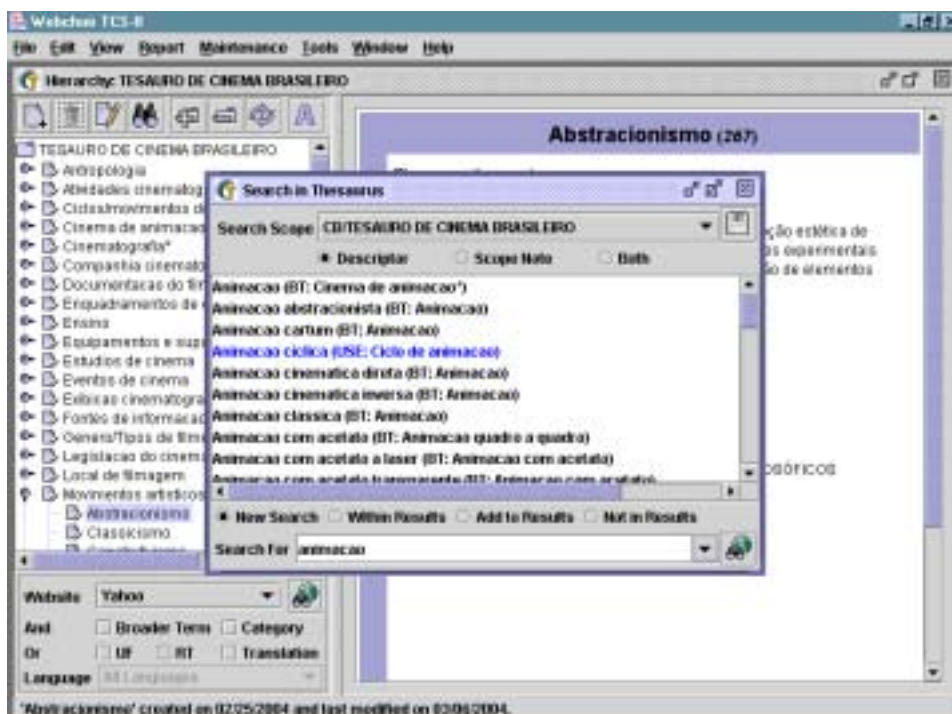


FIGURA 3 - Recuperação dos termos no Tesauro eletrônico do cinema brasileiro

¹ Disponível em: <<http://www.webchoir.com/TCS-8.html>>. Acesso em 30/08/2003.

Esse programa gera produtos, como *home page*, arquivos XML e possibilita a emissão de relatórios, dentre os principais:

a) *Relatório alfabético*: representa uma lista geral de todos os termos em ordem alfabética, incluindo os seus respectivos relacionamentos, notas de explicação, datas de criação e modificação, dentre outros. (FIG. 4).

b) *Relatório hierárquico*: representa uma lista de todas as categorias/facetas com os respectivos termos em ordem alfabética (FIG. 5).

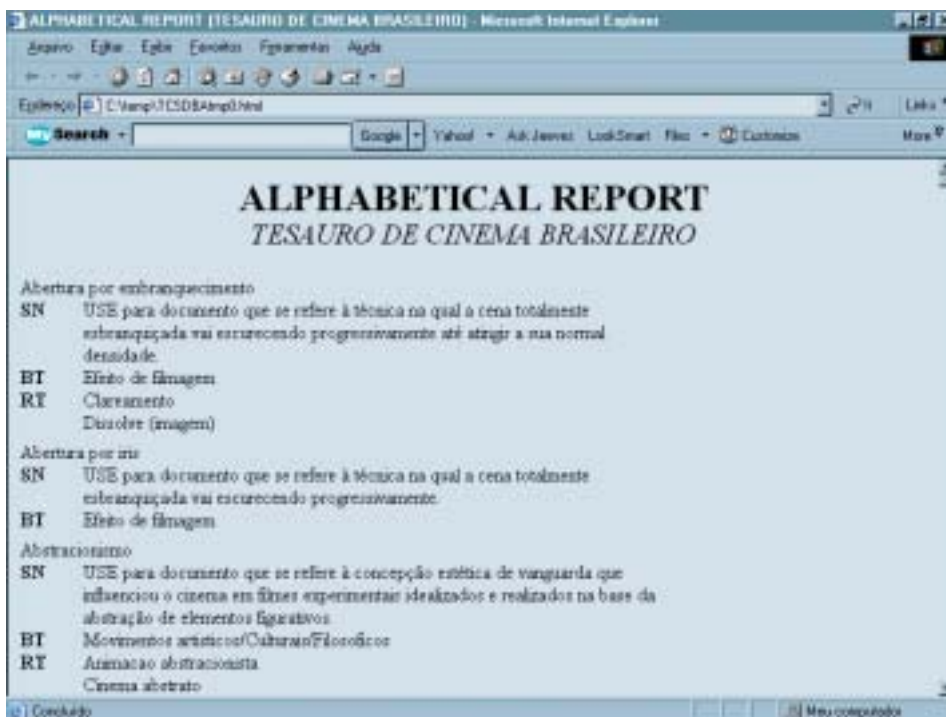


FIGURA 4 - Visualização do relatório alfabético.

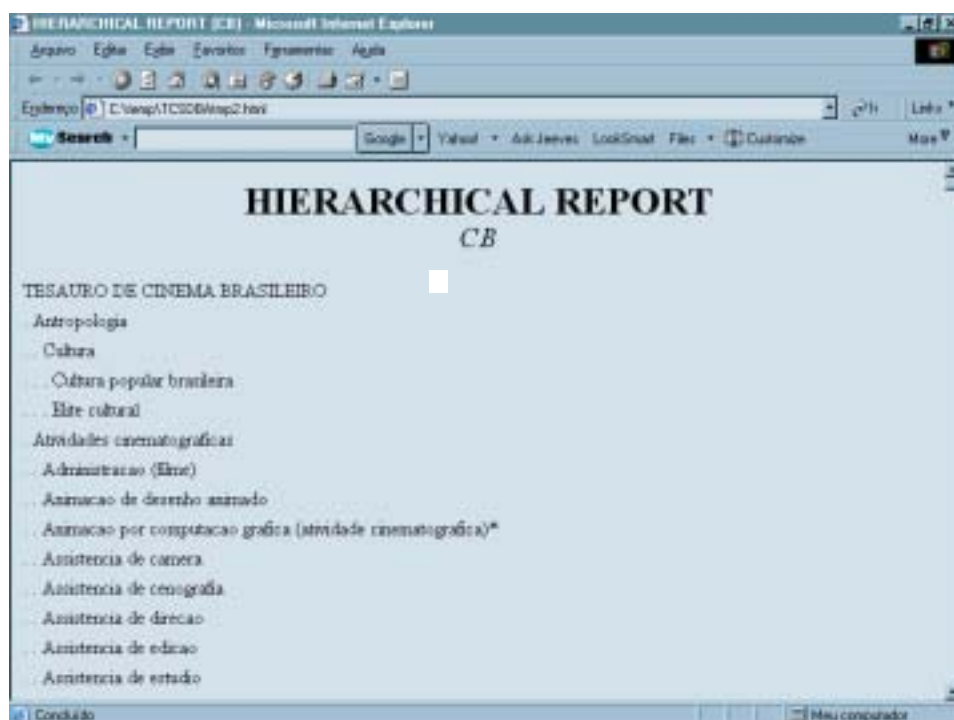


FIGURA 5 - Visualização do relatório hierárquico.

c) *Índice rotado*: representa uma rotação dos termos ordenados alfabeticamente, o que possibilita, através do conhecimento de uma palavra, a recuperação do termo (FIG. 6).

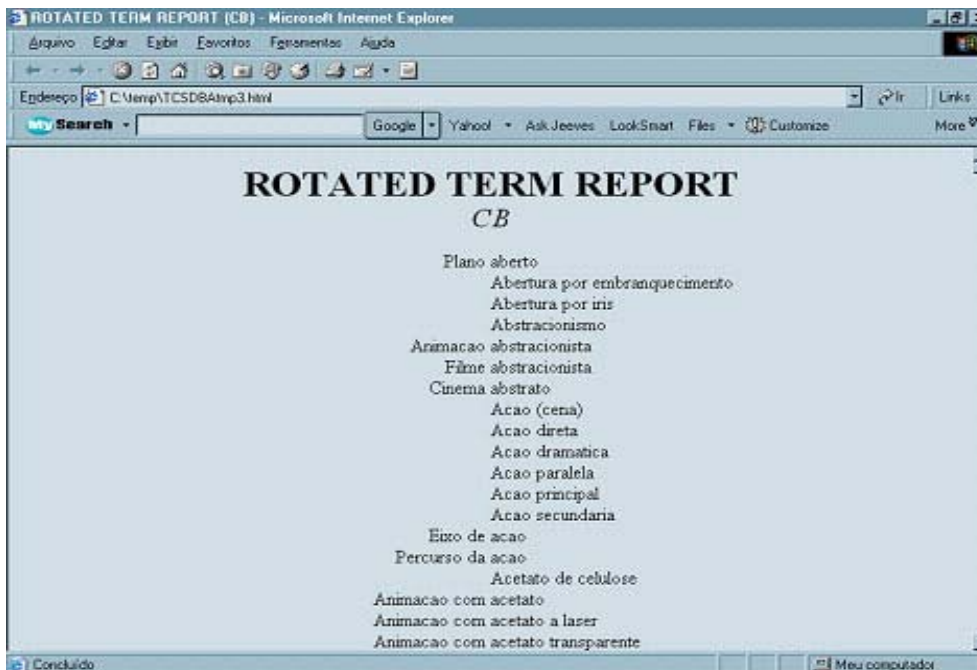


FIGURA 6 – Visualização do índice rotado

O tesouro eletrônico é uma ferramenta em permanente construção, o que garante sua longevidade. Ele reflete a área de conhecimento e representa sua dinâmica de produção. Dessa forma, pode-se afirmar que a originalidade do tesouro eletrônico não reside tanto no fato de ser eletrônico, mas em ser dinâmico, numa área do conhecimento em que a expansão e a consecutiva transformação da informação são relativamente freqüentes.

Considerações finais

A extrema especialização do conhecimento na atualidade tem exigido respostas ágeis em termos de organização e recuperação da informação. Os usuários interagem de forma mais direta com os sistemas de recuperação da informação, razão pela qual, a linguagem que provê o acesso à informação, deve possuir uma sintaxe própria, mas, ao mesmo tempo, refletir as demandas informacionais produzidas nesses ambientes.

O campo audiovisual é certamente um dos domínios que mais sofreu alterações de caráter tecnológico e conceitual ao longo dos últimos anos. Em virtude desse fato, realizar pesquisas ou organizar informações especializadas em cinema se reveste em tarefa extremamente delicada.

O cinema representa um domínio cuja tradução intersemiótica ocorre de maneira intensa, ou seja, os signos verbais são interpretados através dos signos não verbais e vice-versa. Em função disso, prover um instrumento verbal de representação da informação nesse domínio revestiu-se num grande desafio.

O principal desafio do projeto foi a criação das condições de possibilidade para o compartilhamento dos conhecimentos necessários ao desenvolvimento da proposta. Foi preciso construir uma base comum de conhecimentos entre

os integrantes do projeto que permitisse alguma equivalência em termos da estrutura conceitual concebida. Para tanto, a equipe envolvida no processo de elaboração do *Tesauro* teve como objetivo principal fornecer uma sistematização inicial da linguagem cinematográfica adotada no Brasil. Através do acompanhamento e sistematização dos termos usados no campo, foi possível estruturar um léxico articulado por uma rede de relacionamentos que, do ponto de vista da linguagem, pudesse prover um uso mais articulado e em consonância com o vocabulário adotado pelos teóricos do cinema, os profissionais e artistas.

Atenção especial foi dada também aos usuários de publicações e registros cinematográficos brasileiros. Dessa forma, o *Tesauro* foi gerado a partir de uma padronização da linguagem utilizada pelos produtores e usuários da informação. No caso, a garantia de uso foi assegurada por um especialista em cinema e sua equipe de pesquisadores. Através do seu ponto de vista, da experiência, da compreensão conceitual dos termos e das inter-relações existentes entre esses termos foi possível privilegiar uma estrutura mais conexa às lógicas de utilização da informação no domínio. Como resultado obteve-se um vocabulário controlado que representa a dinâmica informacional existente no campo. Avalia-se que, se no processo de elaboração do *Tesauro* não houvesse a garantia de uso, talvez o instrumento pudesse tornar-se ineficaz em sua finalidade de proporcionar o tratamento formal para a área. Verifica-se então que o instrumento concebido pode evitar as ambigüidades e tornar mais interativa a comunicação entre os profissionais que fazem uso das informações nesse campo.

Outro aspecto a ser destacado refere-se ao fato de que a construção de uma linguagem de indexação, independente da área, requer sempre um trabalho interdisciplinar. Estudos das várias obras selecionadas foram feitos por profissionais da informação, mas as peculiaridades da área foram esclarecidas por especialistas em cinema, uma vez que o instrumento tinha como pressuposto atender as demandas informacionais desse segmento. Nesse sentido, foi fundamental para o êxito do trabalho a existência de um projeto no qual estavam explícitos o público alvo, os objetivos e os princípios metodológicos. Devido a esse fato, cada etapa do projeto foi executada com transparência, método e rigor conceitual. Além disso, a escolha de um bom instrumento para o gerenciamento do tesauro foi imprescindível. Nesse aspecto, identificou-se um software que pudesse prover a atualização e o gerenciamento do tesauro e que, ao mesmo tempo, proporcionasse a organização e a recuperação especializada da informação. Cabe ressaltar, entretanto que o software não foi o elemento principal no desenvolvimento do tesauro, mas contribuiu de maneira significativa na gestão da base terminológica elaborada.

Finalmente, a experiência de elaboração do *Tesauro eletrônico do cinema brasileiro* revelou a necessidade de investir em iniciativas locais para o provimento de soluções informacionais que adotem, como princípio, as necessidades concretas de um dado campo de conhecimento e que a atividade de elaboração de linguagens de indexação seja considerada uma tarefa fundamentalmente interdisciplinar.

No caso específico do *Tesauro* a oportunidade de compor uma equipe interdisciplinar tornou evidente que é preciso estabelecer, cada vez mais, a aproximação de campos disciplinares diferentes para a solução de problemas específicos e que seja possível nesse movimento o compartilhamento de metodologias de trabalho (DOMINGUES, 2005, p. 24). Acredita-se, pois que são iniciativas como essas que tornam possível a geração novos conhecimentos.

Indexing languages in cinematographic contexts: the experience of creating an Electronic Thesaurus of Brazilian Cinema

The article describes the experience of creating an Electronic Thesaurus of Brazilian Cinema, which was developed to support the organization and the search of information generated in cinematographic contexts. The point of departure for this work is the realization of the fact that there are no specialized instruments available for the thematic organization of Brazilian cinematographic collections, as well as the increasing demand for specialized language on the part of renowned centers of memory and audiovisual reference in Brazil. The objective is to provide an indexation tool capable of representing not only the specific cinematographic collections thematically, but also the knowledge produced about Brazilian movies. For this, we determined procedures towards mapping the area thematically with the purpose of identifying the main concepts related to the area and its net of relationships. Studies were conducted based on the specific literature of the area, followed by a conceptual comparison based on the collection of films and publications housed by Ophicina Digital of the Department of Photography, Theater and Cinema of the School of Fine Arts - UFMG. With the help of available literature and the expertise of a Brazilian Cinema specialist who was part of the team of researchers, the article introduces the indexation tool. Fundamental principles of use and of literary warranty were used to verify the application of the tool to specific contexts. Pre-testing indexation testified to the specificity of the tool and evaluated its application in specific contexts. The result was the compilation of a thesaurus of 1040 terms organized in 34 specific facets.

Key-words: Electronic thesaurus; Indexing languages - elaboration; Brazilian Cinema

Referências

- ABBAGNANO, Nicola. *Dicionário de filosofia*. São Paulo: Martins Fontes, 2000.
- CAMPOS, M. L. de A. *Linguagem documentária: teorias que fundamentam sua elaboração*. 5. ed. Rio de Janeiro: Ed. UFF, 2001. 133 p.
- CAMPOS, M. L. de A., GOMES, H. E. Tesouro e normalização terminológica: o termo como base para o intercâmbio de informações. *Datagramazero*, v. 5, n. 6, dez. 2004. Disponível em <http://www.dgzero.org/>, dez.2004.
- _____. Organização de domínios de conhecimento e os princípios ranganathianos. *Perspectiva em Ciência da Informação*, Belo Horizonte, v. 8, n. 2, p. 150-163, jul./dez. 2003.
- DAHLBERG, J. Teoria do conceito. *Ciência da Informação*, Brasília, v. 7, n. 2, p. 101-107, 1978.
- DOMINGUES, I. Em busca do método. In: _____. (Org.). *Conhecimento e transdisciplinaridade II*: aspectos metodológicos. Belo Horizonte: Ed. UFMG, 2005. p.17-40.
- FUGITA, M. S. L. A leitura do indexador: estudo de observação. *Perspectivas em Ciência da Informação*, Belo Horizonte, v. 4, n. 1, p. 101-116, jan./jun. 1999.
- JOURNOT, M.-T. *Le Vocabulaire du Cinéma*. Paris: Nathan/VOUEF, 2002.
- NAVES, M. L. Estudo de fatores interferentes no processo de análise de assunto. *Perspectivas em Ciência da Informação*, Belo Horizonte, v. 6, n. 2, p. 189-203, jul./dez. 2001.
- KIMOTO, H.; IWADERA, T. *Construction of dynamic thesaurus and its use for associated information retrieval*. [s.l]: ACM, 1990. p.227-240.
- LARA, M. L. G. de. *A representação documentária: em jogo a significação*. São Paulo: ECA/USP, 1993. 133p.
- _____. Algumas contribuições da semiologia e da semiótica para a análise das linguagens documentárias. *Ciência da Informação*, Brasília, v. 22, n. 3, p. 223-226, set./dez. 1993.
- _____. Linguagens documentárias, instrumentos de mediação e comunicação. *Rev. Bras. Bibliotecon. e Doc.*, São Paulo, v. 26, n.1/2, p. 72-80, jan./jun. 1993.
- RIVER, A. Construção de linguagens de indexação: aspectos teóricos. *Revista da Escola de Biblioteconomia da UFMG*, Belo Horizonte, v. 21, n. 1, p. 56-99, jan./jun. 1992.
- SVENONIOUS, E. Unanswered questions in the design of controlled vocabularies. *Journal of American Society for information Science*, v. 37, n. 5, p.331-340, 1986.
- _____. *The intellectual foundation of information organization*. Cambridge: MIT Press, 2000.