

Estudo comparativo de softwares de construção de tesouros

Maria Luiza de Almeida Campos

Doutora em Ciência da Informação
Professora do Departamento de Ciência da Informação da UFF e do
Programa de Pós-Graduação em Ciência da Informação UFF/IBICT
Rua Tiradentes 210,602. Inglá Niterói—Brasil
E-mail: mlcampos@nitnet.com.br

Hagar Espanha Gomes

Livre Docente
Consultora
End.: Trav. Maria Elmira n.37, Santa Rosa, Niterói, RJ Brasil
E-mail: hagarespanha@terra.com.br

Maria Luiza Machado Campos

Doutora em Ciência da Computação
Professora do Departamento de Ciência da Informação da Computação e do
Programa de Pós-Graduação em Ciência da Computação NCE/UFRJ
Av. Brigadeiro Trompovsky s/n, Prédio do CCMN, C.P. 68530, CEP 21941-
590, Ilha do Fundão, Rio de Janeiro—Brasil
E-mail: mluiza@nce.ufrj.br

Alissandra Evangelista Martins

Mestranda em Sistemas de Informação
Av. Brigadeiro Trompovsky s/n, Prédio do CCMN, C.P. 68530, CEP 21941-
590, Ilha do Fundão, Rio de Janeiro—Brasil
E-mail: alissandra_martins@yahoo.com.br

Linair Maria Campos

Mestre em Sistemas de Informação
Av. Brigadeiro Trompovsky s/n, Prédio do CCMN, C.P. 68530, CEP 21941-
590, Ilha do Fundão, Rio de Janeiro—Brasil
E-mail: linair@nce.ufrj.br

Luana Farias Sales

Mestranda em Ciência da Informação UFF/IBICT, Bolsista CAPES
Av. Marica 984, Colubandê—São Gonçalo—Rio de Janeiro—Brasil
E-mail: luanabiblio@yahoo.com.br

Para auxiliar a construção de vocabulários o uso e, conseqüentemente, a escolha de um *software* de construção de tesouros é uma necessidade. Entretanto, esses *softwares* possuem diferentes características de funcionalidades, custo e apoio metodológico, cuja análise criteriosa deve ser considerada, visando orientar a sua escolha de acordo com a finalidade e complexidade do tesouro a ser construído. Com base nessa necessidade de avaliação, o objetivo do presente trabalho é apresentar um estudo comparativo desses *softwares*, levando em consideração as utilizações dos tesouros nos dias de hoje.

Palavras-chaves: *Softwares* de tesouros; Avaliação de *software*

Recebido em 25.10.2005

Aceito em 20.03.2006

Introdução

Com a difusão do uso dos computadores e das tecnologias de manipulação de documentos, surgiu a necessidade de organizar e recuperar a grande quantidade de informação a partir daí produzida, especialmente considerando o universo das grandes empresas, que muitas vezes disponibilizam diferentes tipos de dados através de uma rede corporativa, ou pela Internet. O objetivo é compartilhar os documentos de maneira rápida e fácil, deixando fluir o conhecimento do negócio, e, desta forma, a correta e eficiente recuperação da informação tem se tornado vantagem competitiva para as corporações, na medida em que aumenta a sua produtividade (MILSTEAD, 1998). Nesse contexto, o uso de tesouros têm se destacado como ponto de apoio para a organização e acesso multifacetado da informação, bem como para a recuperação de conceitos relacionados.

Para auxiliar a construção desses vocabulários o uso de *softwares* de construção de tesouros é uma necessidade, devido às suas facilidades de armazenamento, manipulação e apresentação dos termos definidos e suas relações. Desta forma, o objetivo dos tesouros e a escolha criteriosa de um *software* para sua implementação é importante caracterizar o que se entende por este instrumento e como pode se dar a sua elaboração.

Definir tesouro como uma linguagem documentária utilizada para recuperação de informação, de certa forma resume demais uma ferramenta que, entende-se e expande-se em aplicações, a partir da observação de seu histórico. Um século de história e aplicações explicam a importância de uma ferramenta que adequadamente construída com um suporte metodológico, torna-se um potente aliado na recuperação de informações.

A partir da década de 70 a *Unesco* apresenta definição para o instrumento tesouro que aponta dois aspectos de sua aplicação, ou seja, aquele onde o tesouro é considerado como uma estrutura de termos relacionados semanticamente em dado contexto de conhecimento; e aquele onde se enfatiza sua função, isto é, um dispositivo de controle terminológico que visa o tratamento e a recuperação de informações.

É um vocabulário controlado e dinâmico de termos relacionados semântica e genericamente cobrindo um domínio específico do conhecimento... É um dispositivo de controle terminológico usado na tradução da linguagem natural dos documentos, dos indexadores ou dos usuários numa linguagem do sistema (linguagem de documentação, linguagem de informação) mais restrita. (UNESCO, 1973, p.6)

As questões enfrentadas pelos profissionais de informação são as mais variáveis possíveis e historicamente o resultado disso pode ser observado em uma divisão bastante clara das vertentes dos tesouros que surgem na América do Norte e na Europa. Após observar o diagrama evolutivo proposto por Lancaster (FIG. 1), Campos (2001) observa: O que fica evidente no diagrama é a dicotomia na linha evolutiva dos tesouros. De um lado, a vertente de abordagem alfabética, de grande influência na América do Norte, e, de outro, a abordagem sistemática, que tem seus pressupostos estabelecidos pela classificação facetada. (CAMPOS, 2001)

O diagrama de Lancaster ilustra a evolução dos tesouros de vertente alfabética (linha americana) e dos tesouros que usam a abordagem sistemática (linha europeia), porém, devido à sua representação estar contextualizada em dado período histórico de sua evolução, ainda não evidencia o surgimento do tesouro de origem conceitual, para o qual a definição do conceito é requisito fundamental. Este tipo de tesouro surge no contexto de estudos derivados das teorias do conceito (DAHLBERG, 1978) e da teoria da classificação facetada (1967), sendo caracterizado com mais detalhes na seção 3.

No processo de elaboração de *software* de tesouros há que se levar em conta os aspectos metodológicos, os quais variam de acordo com a linha adotada para a sua construção, e também outros aspectos operacionais independentes da metodologia adotada, como por exemplo, mecanismos para a criação e controle de atualização de termos, possibilidade de reutilização de outros tesouros e elaboração de relatórios.

Considerando essas características, podemos afirmar que a construção de um tesouro não é uma tarefa trivial e que para o seu projeto é importante o uso de uma ferramenta de *software* que apóie não só as questões operacionais, mas também a metodologia usada para a sua elaboração. Este trabalho pretende apontar as funcionalidades que tal ferramenta deve ter para que possa atender aos requisitos identificados, bem como avaliar as ferramentas já existentes, à luz desses requisitos. Além disso, os atributos são agrupados em categorias, o que permite uma avaliação objetiva.

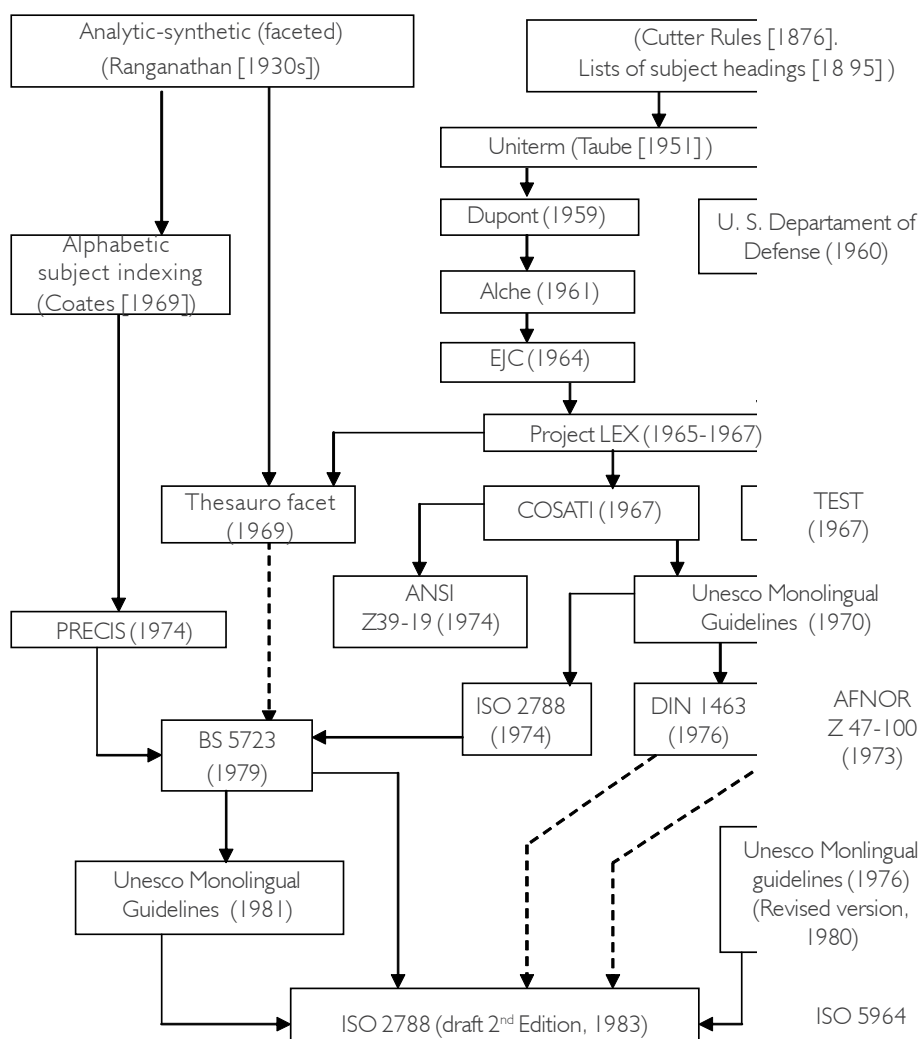


FIG. 1: Diagrama evolutivo dos tesouros (LANCASTER, 1986, p.30)

Utilização de tesouro

Tesauros têm sido utilizados para indexar e recuperar informação em diversos domínios, tais como: ciência da informação, mecatrônica, folclore, jurídico. Sua estrutura fornece um vocabulário uniforme para indexar a informação e permite aos seus usuários uma forma intuitiva e organizada de pesquisar assuntos de seu interesse, muitas vezes em mais de uma língua, como por exemplo, no *General European Multilingual Environment Thesaurus* (GEMET, 2005).

Além de seu uso para indexação de assuntos, o tesouro pode ainda oferecer outros recursos, através da exploração das relações entre seus termos, através de notas de escopo, ou outras informações, tal como a origem do termo.

Nesse contexto, a eficiência do uso de tesouros tem sido comprovada em estudos que apontam para ganhos de precisão nas consultas da ordem de 30% (SILVEIRA, 2003) e seu uso tem se difundido, passando da indexação de acervos de bibliotecas por meio de fichas catalográficas em papel, até a indexação de acervos multimídia digitais. É importante ressaltar que a forma de elaboração do tesouro vai influir na sua eficiência. Por exemplo, um tesouro pode ter poucas relações associativas¹ entre os termos, dificultando o resgate de termos correlatos. Questões dessa natureza remetem a aspectos metodológicos, que são abordados na seção seguinte.

Metodologia para construção de tesouro

Os tesouros americanos que seguem a vertente alfabética foram fruto do desenvolvimento que ocorreu a partir do cabeçalho de assuntos para o unitermo (CAMPOS, 2001). Essa linha de evolução privilegia a criação de termos pelo seu sentido lingüístico, não utilizando bases classificatórias ou terminológicas para a determinação do termo e suas relações. Conseqüentemente, neste tipo de tesouro não existe a preocupação de se criar categorias para agrupar os termos que possuam atributos comuns, sendo a lista alfabética a única forma de recuperação de informação.

Os tesouros que seguem a vertente européia utilizam categorias para organizar os termos. Esta vertente foi influenciada pela teoria da classificação facetada de Ranganathan (1967). O uso de facetas facilita evidenciar as relações entre os termos e a navegação pelo tesouro. Dessa forma, passa a ser possível buscar um assunto percorrendo as diversas facetas criadas, de forma sistemática, o que é uma vantagem quando não se sabe *a priori* o termo desejado, especialmente em tesouros com muitos termos.

Apesar desta vertente avançar quanto aos aspectos relacionados a uma apresentação sistemática do tesouro, seus autores apresentam um comportamento semelhante ao dos autores americanos no que se refere ao estabelecimento do termo propriamente dito, ou seja, ainda se privilegia o sentido lingüístico na sua criação.

De acordo com Campos (2001), verifica-se atualmente, devido à corrente que liga a teoria do conceito à teoria da classificação, uma tendência para um novo tipo de tesouro, o tesouro conceitual. Este é um tesouro com base em conceitos, onde na sua construção são necessários princípios para o

¹ Relação Associativa é uma relação não-hierárquica entre conceitos e classes (WERSIG, NEVELING, 1976)

estabelecimento do termo/conceito e das relações entre eles. Neste tipo de tesouro o sentido lingüístico não é mais determinante na criação e organização dos termos e sim o seu conteúdo conceitual.

É importante observar que as linhas de elaboração de tesouros possuem aspectos divergentes. Entretanto, apesar de suas particularidades, elas partilham de alguns requisitos comuns. Em todas as três tem-se a necessidade de se fazer um levantamento do escopo do domínio a ser representado, as relações entre os termos e suas sinónímias.

A linha européia e a linha do tesouro conceitual consideram os aspectos de categorização dos termos, ou seja, sua organização e apresentação sistemática (GOMES, CAMPOS, MOTTA, 2004). Esta organização pode influenciar no estabelecimento das relações entre os termos, por exemplo, pode limitar que relações do tipo *gênero/espécie* se dêem entre termos de categorias distintas. Porém, apenas no tesouro conceitual existe a preocupação com a definição do conceito que, por sua vez, vai influenciar a maneira da organização dos termos em categorias. A TAB. I ilustra esses aspectos.

TABELA I – Aspectos Metodológicos das Linhas de Construção de Tesouro

<i>Linhas/Aspectos</i>	Levantamento do Escopo	Relações entre termos	Categorização	Definição do conceito
Americana	Sim	Sim	Não	Não
Européia	Sim	Sim	Sim	Não
Conceitual	Sim	Sim	Sim	Sim

Além dos aspectos comuns e divergentes envolvidos nas metodologias, há que se considerar que a construção de um tesouro é um projeto que na maior parte das vezes envolve uma equipe multidisciplinar, composta por profissionais com o conhecimento do domínio a ser representado e profissionais com o conhecimento sobre como modelar um domínio, estes últimos geralmente oriundos da área de ciência da informação. Pode-se ainda acrescentar que a construção de um tesouro está sujeita a fatores semelhantes aos que ocorrem no desenvolvimento de um projeto de *software*, tais como: levantamento de requisitos, modelagem do problema, controle de versões e manutenção. Tendo em mente esses aspectos, na seção seguinte discutimos os requisitos propostos para a construção de tesouros.

Trabalhos relacionados

Existem algumas iniciativas que apontam características desejáveis em um tesouro. Dentre elas temos as normas internacionais, ilustradas na FIG. 1, e ainda outros estudos, como os de Gomes, Campos, Motta, (2004) e o de Soergel, (2002), que apontam características tais como: padrão para denominação de termos, tipos de relação entre termos, regras para atualização do tesouro, dentre outras. Como se espera que os *softwares* de tesouro contemplem essas características, elas podem ser consideradas para o

levantamento de alguns dos requisitos desses *softwares*. Entretanto, alguns aspectos não dizem respeito às características do tesouro em si, e sim à implementação do *software*, e não são considerados por estas iniciativas. Por exemplo: se o *software* é livre e gratuito, se roda em vários tipos de sistemas operacionais, se possui interface *Web*, dentre outras. Além disso, não é do escopo das iniciativas supracitadas a comparação de *softwares* de tesouro.

Neste contexto, o trabalho de Ganzmann (1990) é o mais detalhado que encontramos, apresentando uma lista de características bastante abrangente, cobrindo aspectos que vão desde a ergonomia do *software* até construção de relações multilingües entre os termos. Entretanto, alguns dos aspectos levantados por Ganzmann já se encontram obsoletos devido à data de sua publicação.

Por outro lado, mesmo propondo uma lista detalhada, alguns aspectos não foram abordados, como por exemplo, questões relacionadas à metodologia de construção do tesouro, notadamente nas fases que estão relacionadas a análise do domínio e seu planejamento.

Nesse sentido, até onde temos notícia, não há trabalho publicado que leva em consideração essa visão integrada (características ligadas ao *tesouro e ao software*), e ainda, que a utiliza para fazer uma comparação dos *softwares* de tesouro. Na verdade, não encontramos qualquer iniciativa de comparação de *softwares* de tesouro publicada. O trabalho que mais se aproxima neste sentido é o de Will (2005). Nesse trabalho o autor faz apenas referências a princípios para elaboração de tesouros e, independente destes, faz um quadro onde analisa vários *softwares* de tesouro. Além disso, a análise dos *softwares* não é feita de forma sistemática, ou seja, não segue um quadro de referência onde cada aspecto é avaliado para todos os *softwares* considerados. A análise é feita em forma de texto resumido, onde para cada *software* há informações básicas sobre os recursos que ele oferece, juntamente com informações de contato e, quando pertinente, da página *Web* do *software* ou da instituição que o fornece. Por outro lado, não há na iniciativa de Will uma forma objetiva de se obter uma medida de avaliação de *softwares*. Neste sentido, a contribuição principal de nosso trabalho é a apresentação de um levantamento ponderado dos requisitos que um *software* de tesouro deve possuir, levando em consideração os aspectos metodológicos que influem na sua elaboração, bem como as suas características operacionais e de implementação.

Requisitos dos *softwares* de construção de tesouro

O levantamento de requisitos partiu de uma revisão dos trabalhos de Ganzmann (1990) e Gomes, Campos e Motta (2004), bem como da norma ISO 5964 (1985), onde os critérios propostos pelos autores foram analisados sob a perspectiva da tecnologia atual, sumarizados para permitir uma comparação mais objetiva, e estendidos para incluir aspectos metodológicos. Para facilitar esse repensar do levantamento de requisitos, vamos denominar doravante de *critérios* os requisitos originalmente propostos pelos autores e de *requisitos* os que foram reformulados no contexto deste trabalho.

Desta forma, critérios muito detalhados como: *não permitir a entrada duplicada de termos, controle de relações ilógicas entre níveis*

hierárquicos, não permitir relações duplicadas entre dois termos, foram resumidos para o requisito *consistência*. Cabe ressaltar que o fato de terem sido sumarizados, não implica em que seu detalhamento tenha sido ignorado ao avaliar o *software*. Ou seja, se um desses critérios específicos não é atendido, considera-se que o *software* não atende ao requisito em questão, porém na tabela comparativa dos *softwares* (Seção 6), é exibido apenas o requisito resumido.

Critérios já ultrapassados foram desconsiderados, como por exemplo, o uso do *mouse* como fator de ergonomia, pois hoje em dia o uso do *mouse* já está consolidado nos *softwares* de um modo geral. Critérios difíceis de avaliar de forma objetiva foram desconsiderados para facilitar a avaliação, como, por exemplo, os diferentes tipos de mensagem (erro, alerta, confirmação, dentre outras) contribuindo para a ergonomia do *software*. Neste caso, a ergonomia foi avaliada sob outros enfoques, tais como: *facilidade de uso*, ou seja, se o usuário é capaz de criar um pequeno conjunto de termos e suas relações sem auxílio de documentação. Embora esse requisito também possua algum grau de subjetividade, sua avaliação restringe-se a *sim* ou *não*, ao contrário das mensagens, onde não fica claro em que situação o *software* é considerado ergonômico ou não.

O objetivo é prover um conjunto mínimo de requisitos que se espera sejam atendidos por um *software* de tesouro, e classificá-los de acordo com seu grau de importância. Estes graus se dividem em *essenciais*, *importantes* e *inovadores*. *Essenciais* são os requisitos que o *software* deve atender para ser considerado adequado, e, portanto, seu peso deve ser pelo menos maior que a soma dos outros (em nosso trabalho adotamos o valor 50). *Importantes* são os requisitos que acrescentam funcionalidades além das essenciais ao *software*, tornando-o mais fácil de usar, e possuem peso 1. *Inovadores* são requisitos que destacam o *software* dentre os demais, fornecendo funcionalidades que contribuem de forma original para a elaboração do tesouro, e possuem peso 3. É importante ressaltar que os pesos podem variar, de acordo com a perspectiva de cada avaliador. Por exemplo, o fator custo gratuito pode ser considerado de peso 50 para uma empresa que não esteja disposta a investir na compra de um produto, mas pode ter peso 1 para uma outra que deseje priorizar as funcionalidades do *software*. Dessa forma, os pesos aqui sugeridos podem ser adaptados a situações específicas, de acordo com a necessidade. O importante é que a sua atribuição permite obter uma avaliação objetiva do *software* ao final. O peso de cada requisito é indicado entre parêntesis, à direita de cada critério na TAB. 2.

Partindo do exposto, os requisitos considerados são divididos em oito grupos, a saber:

a) Características gerais

Em características gerais temos a avaliação de requisitos genéricos que são importantes para a construção dos termos de tesouro. Nesta categoria se enquadram as limitações genéricas (restrição de tamanho, número, de termos e possibilidade de uso de termos compostos, a existência da informação da fonte ou origem do termo, a possibilidade de criar classes de assuntos, e se os termos podem ser expressos em mais de um idioma.)

b) Tratamento de relações

Em *tratamento de relações*, estão os requisitos que indicam de que forma o *software* permite conceber as relações entre os termos, ou seja, se o *software* permite relações definidas pelo usuário, relações padrão e relações adotadas no padrão ISO 5964 (1985).

c) Tratamento de dados

Em *tratamento de dados*, estão os requisitos que indicam de que forma se dá o armazenamento e a aquisição dos dados, ou seja, se o *software* permite guardar o tesouro em banco de dados, se permite importação de termos de e para outras fontes, se existe crítica em relação à consistência de dados, se é feito algum controle de restrição de acesso para atualização do tesouro e se permite o gerenciamento da situação de termos, como por exemplo, termos candidatos ou termos obsoletos.

d) Interface / Manipulação dos dados

Em *interface / manipulação dos dados*, os requisitos estão voltados para avaliar a maneira que o *software* permite interagir com os seus usuários, tanto no projeto do tesouro quanto na pesquisa, ou seja, se permite busca flexível por termos, se permite reorganizar a hierarquia de termos, se possui interface *Web*, e se a criação do tesouro é um processo intuitivo.

e) Relatórios

Em *relatórios*, estão os requisitos voltados para avaliar a apresentação da estrutura do tesouro ao usuário final, ou seja, que tipos de relatório possuem, dentre os seguintes: alfabético, sistemático, KWIC, hierárquico e definido pelo usuário.

f) Características de implementação

Em *características de implementação*, os requisitos têm como objetivo avaliar os pré-requisitos que o *software* possui para ser utilizado, ou seja, se depende de *software* não livre, se roda em *Unix*, se roda em *Windows* e se é gratuito.

g) Apoio metodológico

Em *apoio metodológico*, os requisitos buscam avaliar se o *software* foi projetado com recursos que forneçam apoio ao uso de alguma metodologia de construção de tesouros, ou seja, recursos que apóiem os aspectos ilustrados na TAB.1. Esses requisitos são: representação gráfica do escopo ajuda estabelecer o relacionamento entre termos, categorização com base no conceito e uso de características de divisão - que permitem agrupar termos de uma mesma hierarquia de acordo com algum atributo comum (GOMES, CAMPOS, MOTTA, 2004). Cabe ressaltar que este

último requisito, diz respeito à etapa de categorização de termos do tesouro.

8. Apoio ao uso

Em *apoio ao uso*, os requisitos estão voltados para avaliar o tipo de apoio que o fabricante do *software* disponibiliza aos seus clientes na utilização do produto, ou seja, se o *software* possui: suporte, grupo de usuários ativo, ajuda *online* e documentação.

Comparação de *softwares* de construção de tesouros

Os *softwares* selecionados foram obtidos a partir de pesquisa na Internet. São eles: *TermChoir* (2005), *Multites* (2005), *TheSmain* (2005), *TermTree* (2005), *Synaptica* (2005) e *TemaTres* (2005).

Alguns dos *softwares* possuem cópias de demonstração, que foram instaladas e utilizadas para a construção de um pequeno tesouro, a título de teste. Dentre esses, alguns apresentaram problemas na instalação, os quais não foram solucionados. Outros não possuem cópia de demonstração. Sendo assim, com exceção do primeiro grupo, os outros *softwares* foram avaliados com base em material disponível na página *web* do produto. Os *softwares* avaliados através de documentação são indicados por um asterisco ao lado do seu nome.

Foram desconsiderados *softwares* que não possuem versão para *windows* 2000 ou superior ou então algum tipo de *linux* ou *unix* e ainda os que apresentaram problemas na instalação. A escolha dos seis *softwares* listados se deve ao resultado de uma análise preliminar, onde foram selecionados aqueles cujo conjunto de funcionalidades foi considerado suficiente para esse estudo comparativo. A exceção é o *software* *TemaTres*, que, apesar de apresentar poucos recursos, foi selecionado por ser o mais completo dentre os *softwares* livres avaliados.

A avaliação da TAB. 2 permite concluir que, de uma maneira geral, os *softwares* de tesouro atendem de maneira satisfatória a maioria dos requisitos que estes devem ter, exceto pelos aspectos metodológicos. Sendo assim, têm-se as seguintes observações.

a) Características gerais

Em relação às *características gerais*, podemos observar que os *softwares* não gratuitos as atendem de forma bastante abrangente. De todos os avaliados, apenas o *TermTree* possui uma restrição: não é multilíngue, o que pode ser uma limitação séria, dependendo do propósito do tesouro. Já o *TemaTres*, gratuito, não possui capacidade para referenciar a fonte do termo, embora seja multilíngue.

b) Tratamento de relações

Em relação ao *tratamento de relações*, observamos que *TheSmain*, *TermTree* e *TemaTres* não permitem a criação de relações definidas pelo usuário. Esta limitação pode ser um problema no caso de se querer fornecer alguma

semântica adicional ao tesouro, como por exemplo, a caracterização de diferentes tipos de relações partitivas. Além disso, neste item, temos que a maioria dos *softwares* segue padrões mais antigos como o Z39-19 (1974), enquanto que o padrão ISO 5964 (1985) é utilizado apenas pelo *ThesMain*.

c) Tratamento de dados

Em relação ao requisito *tratamento de dados*, podemos verificar que é atendido de maneira integral pelos *softwares* não gratuitos. O TemaTres, por outro lado, não permite importação nem controle de atualização dos dados. Estes aspectos podem ter um impacto profundo na construção de tesouros de maior porte, onde não só a reutilização de termos de outras fontes, como também o controle de sua atualização é de grande importância, fazendo com que a solução gratuita não seja adequada.

d) Interface e manipulação de dados

Em relação ao requisito *interface e manipulação de dados*, podemos afirmar que ele é atendido de maneira adequada em relação à facilidade de uso e busca flexível, que, pode-se considerar como os recursos mais importantes. Entretanto, a reorganização de hierarquias, que pode ser importante para tesouros muito grandes ou que mudam com muita frequência, só é atendida pelos *softwares* *TermChoir* e *TremTree*. O recurso *interface Web*, também não é atendido pelos *softwares* *Multites* e *ThesMain*, o que pode ser um problema no caso de se querer disponibilizar o acesso remoto a pessoas autorizadas a manipular o tesouro, ou no caso de a política de instalação de *softwares* da empresa estar voltada para o uso de *softwares* através da *Web*.

e) Emissão de relatórios

Em relação à *emissão de relatórios*, ela é, de um modo geral, bem explorada pelos *softwares* não gratuitos, inclusive em relação a relatórios na *Web*, onde apenas o *ThesMain* não oferece. Já o TemaTres, gratuito, oferece grandes limitações nesse aspecto, permitindo apenas o relatório alfabético.

f) Características de implementação

Em relação às *características de implementação* observamos sérias restrições em relação a *softwares* não proprietários. Apesar de apenas o *Multites* e o *ThesMain* não rodarem em ambientes baseados em *Unix*, todos os *softwares* não gratuitos avaliados dependem de algum *software* não livre, como por exemplo banco de dados da Oracle (2005) ou da Microsoft (2005). Em contrapartida, o *software* livre gratuito é bem inferior em relação aos demais. Esta é uma limitação das mais sérias, pois pode ir contra a política de *software* da empresa, especialmente se a empresa é pública, pois existe a atualmente a orientação crescente de se adotar *softwares* livres nestas instituições.

g) Aspectos metodológicos

Em relação aos *aspectos metodológicos*, nenhum dos *softwares* disponíveis possui apoio abrangente, ou seja, nenhum deles relaciona as etapas de utilização do *software* à uma base teórica-metodológica para construção de tesouros, ainda que possuam ferramentais que auxiliem na adoção de uma metodologia. Um exemplo de ferramental que pode servir como auxiliador é a ajuda na definição dos conceitos através da busca *Web*, por parte dos *softwares Synptica e TermChoir*.

h) Apoio ao uso

Em relação ao requisito de *apoio ao uso*, pode-se dizer que ele também é atendido de maneira geral pelos *softwares* não livres, exceto no aspecto de *grupos de usuários ativos*, o qual nenhum dos *softwares* possui, o que caracteriza possivelmente a pouca maturidade de uso dessas ferramentas.

Podemos observar também que poucos *softwares* possuem um conjunto relevante de características inovadoras. Dentre as nove possíveis, *TermChoir e Synptica* possuem três, *Multites e TermTree* possuem duas, *ThesMain e TemaTres* possuem apenas uma.

TABELA. 2 – Comparativo de *softwares* de construção de tesouro

	Critério	TemChoir	Multites	TheSmain (Superthes)	TermTree 2000	Synptica*	TemaTres*
1	1.1 Limitações genéricas (50)	não	não	não	não	não	não
	1.2 Fonte do termo (1)	sim	sim	sim	sim	sim	não
	1.3 Classes de assunto (1)	sim	sim	sim	sim	sim	não
	1.4 Multilíngüe (1)	sim	sim	sim	não	sim	sim
2	2.1 Relações definidas pelo usuário (50)	sim	sim	não	não	sim	não
	2.2 Relações padrão (1)	sim	sim	sim	sim	sim	sim
	2.3 Padrão ISO 5964 (1)	não	não	sim	não	não	não
3	3.1 Persistência em banco de dados (50)	sim	sim	sim	sim	sim	sim
	3.2 Importação de dados (1)	sim	sim	sim	sim	sim	não
	3.3 Exportação de dados (1)	sim	sim	sim	sim	sim	sim
	3.4 Consistência (50)	sim	sim	sim	sim	sim	sim
	3.5 Controle de Acesso (1)	sim	sim	sim	sim	sim	sim
	3.6 Controle de Atualização de Dados (1)	sim	sim	sim	sim	sim	não
4	4.1 Busca Flexível (1)	sim	sim	sim	sim	sim	sim
	4.2 Recursos para reorganizar hierarquias (1)	sim	não	não	sim	não	não
	4.3 Facilidade de uso (1)	sim	sim	sim	sim	sim	sim
	4.4 Interface Web (1)	sim	não	não	sim	sim	sim
5	5.1 Relatório Alfabético (1)	sim	sim	sim	sim	sim	sim
	5.2 Relatório Sistemático (1)	sim	sim	sim	sim	sim	não
	5.3 Relatório KVIC (1)	sim	sim	sim	sim	não	não
	5.4 Relatório Hierárquico (1)	sim	sim	sim	sim	sim	não
	5.5 Relatórios na web (1)	sim	sim	não	sim	sim	não
	5.6 Relatório definido pelo usuário (3)	sim	sim	sim	sim	sim	não
6	6.1 Depende de Software não Livre (3)	sim	sim	sim	sim	sim	não
	6.2 Roda em Uniz (1)	sim	não	não	sim	sim	sim
	6.3 Roda em Windows (1)	sim	sim	sim	sim	sim	sim
	6.4 Gratuito (3)	não	não	não	não	não	sim
7	7.1 Possui suporte (3)	sim	sim	não	sim	sim	não
	7.2 Possui grupo de usuários ativo (3)	não	não	não	não	não	não
	7.3 Possui ajuda online (1)	sim	sim	sim	sim	sim	não
	7.4 Possui documentação (1)	sim	sim	sim	sim	sim	não
8	8.1 Representação gráfica do escopo (3)	não	não	não	não	não	não
	8.2 Ajudar a relacionar termos (3)	não	não	não	não	não	não
	8.3 Categorização com base no conceito (3)	não	não	não	não	não	não
	8.4 Ajudar a definir os conceitos (3)	sim	não	não	não	sim	não
	8.5 Características de Divisão (3)	não	não	não	não	não	não

Cabe ainda ressaltar que os *softwares* avaliados apresentam algumas características positivas, que embora não tenham sido consideradas isoladamente como requisitos nesse trabalho, merecem destaque. Dentre elas, selecionamos as que julgamos mais relevantes, as quais exemplificamos com algum dos *softwares* analisados que possuísse a característica em questão: (i) recursos de busca sofisticados incluindo linguagens de busca baseadas em SQL² que permitem recuperações poderosas, talhadas para o mecanismo de banco de dados que armazena o tesouro (*Thesmain*); (ii) adaptação do *software* sob medida e sem custo para atender pequenas demandas do usuário (*Synaptica*); (iii) suporte a múltiplos tipos de dados e a inclusão de imagens (*Termchoir*); (iv) facilidades de uso tais como recursos de clicar e arrastar e *menu* adicional sensível ao botão direito do *mouse* (*Termchoir*); (v) A possibilidade de escolha entre duas versões de normas padrão, sendo elas a ISO 2788 para tesouros de assunto e a AS4390/ISO 15489 para taxionomias (*TermTree*); (vi) A possibilidade de salvar os dados de configuração de relatórios, como por exemplo a ordem e tipos de relações a serem impressas (*Multites*).

Por outro lado, também existem alguns aspectos negativos que pudemos observar. Por exemplo: (i) o *Multites* não permite gerar um relatório sistemático único com hierarquia de gênero-espécie e parte-todo; (ii) não se conseguiu desfazer o relacionamento de USE/UP no *Termtree*; (iii) o *TemaTree* não possui um programa instalador, podendo ser difícil de instalar para usuários com pouca familiaridade com informática.

Funcionalidades ou limitações como às exemplificadas acima existem em vários dos *softwares* avaliados, sendo que o *Termchoir* é o que mais recursos oferece e o *TemaTree* o mais limitado. Em contrapartida, quanto mais recursos são adicionados em um *software*, geralmente mais complexo fica o seu uso. Os requisitos que apontamos nesse trabalho, têm como objetivo fornecer um guia para ajudar na avaliação de *softwares* de tesouro, mas eles são sintéticos, e, desta forma, não vão por si só avaliar o quanto um *software* é mais poderoso que outro em relação a um mesmo requisito, e qual a complexidade introduzida por conta disso. Por exemplo, a maioria dos *softwares* avaliados fornece recursos para importação de dados, porém alguns fornecem opções mais flexíveis e poderosas, e também mais complexas, do que outros, incluindo até a possibilidade de elaboração de *scripts*³ para configurar essa tarefa. Não é o objetivo desse trabalho avaliar qual o melhor *software*, mas sim apontar quais os requisitos mínimos para que o *software* seja considerado adequado. Dessa forma, a análise mais detalhada sobre como cada *software* explora cada requisito não foi abordada em profundidade. Até porque, para a avaliação do melhor *software* depende do objetivo que se pretende atingir. Por exemplo, se o que se deseja é construir um tesouro simples, talvez a escolha mais adequada seja uma ferramenta com os requisitos mínimos e sem muitas funcionalidades adicionais que dificultem a sua operação. Entretanto, cabe observar que os requisitos aqui apontados podem ser estendidos para uma análise mais aprofundada, ou ainda, notas podem ser atribuídas no lugar de *sim* ou *não*, no julgamento de cada requisito. Neste caso, deve-se ter o cuidado de estabelecer parâmetros para a atribuição de uma nota, de forma a diminuir a subjetividade na avaliação.

² SQL é uma linguagem padrão para acesso e manipulação de bancos de dados.

³ Um *script* aqui é considerado como um conjunto de instruções, em formato texto, utilizado para realizar alguma tarefa configurável, ou seja, que possa alterar algum aspecto de uma funcionalidade já pronta dentro do *software*, através de uma interação com este.

Conclusão

Atualmente os tesouros vêm ocupando posição de destaque como elemento central para organizar e divulgar informações, em um mundo onde a quantidade e variedade de dados têm criado dificuldades para a sua manipulação e resgate. Estes tesouros, muitas vezes extensos e com uma ampla rede de relações entre seus termos, requerem o uso de ferramentas e metodologias que venham apoiar a sua elaboração, tarefa que está longe de ser trivial.

Neste contexto, acompanhando a evolução da tecnologia de computação, os *softwares* voltados para a construção de tesouros surgem para suprir a demanda de diminuir a complexidade da tarefa de criação, armazenamento, atualização e divulgação dos vocabulários concebidos. Entretanto, ao que parece, não existe a preocupação no *software* de se explicitar o conhecimento tácito que o construtor do tesouro deve possuir. As ferramentas, até onde pudemos avaliar, não estão preparadas para representar ou guiar o processo de modelar o domínio de conhecimento abordado pelo tesouro. Ao contrário, seu foco é automatizar as tarefas operacionais, e, nesse caso, as ferramentas estão bem preparadas para fornecer recursos como: criar o termo, associar a este uma relação, armazenar em banco de dados, emitirem relatórios, e assim por diante. Entretanto, quando se aborda a questão da concepção do tesouro, do seu pensar, então a realidade é outra: nenhuma ferramenta provê recursos para expressar a modelagem do domínio, para planejar suas relações em mais alto nível, ou para guiar a categorização dos termos com base em seu conceito, o que parece indicar uma visão dissociada entre a tecnologia e as metodologias.

A concepção do tesouro pode ser considerada a parte mais complexa da sua elaboração, e hoje tem sido a fase que antecede o uso das ferramentas, talvez pela lacuna metodológica deixada por estas e apontada neste trabalho. É nessa fase que o escopo é traçado, os relacionamentos são pensados e o tesouro toma sua forma, provavelmente na memória tácita e em materiais de apoio usados pelo seu criador.

Nesse sentido, uma verdadeira inovação, que esperamos ser percebida pelos desenvolvedores de *software* de tesouro é uma visão integrada, onde aos aspectos operacionais venham se juntar os aspectos metodológicos criando uma ferramenta mais inteligente.

Comparative Study of Thesauri Softwares

To help the construction of vocabularies, the use and the consequent selection of a thesaurus construction software is a must. However, thesauri softwares have different characteristics of functionalities, costs, and methodological support, among other diversities, which must be carefully analysed according to the goals and complexity of the thesaurus to be built. Thus, this work aims at introducing a comparative study of thesauri software, taking into account present thesauri uses.

Key-words: Thesaurus software; Evaluation software

Referências

- ACTIVE CLASSIFICATION SOLUTIONS. *Termtree*. Disponível em: < <http://www.termtree.com.au> > Acesso em: 06 jun 2005.
- CAMPOS, M. L. A. *Linguagem documentária*. Niterói: EDUFF, 2001.
- DAHLBERG, Ingetraut. Teoria do conceito. *Ciência da Informação*, v. 7, n. 2, p. 101-07, 1978.
- FERREIRA, D. *Tematres*. Disponível em: < <http://www.r020.com.ar/tematres/index.html> > Acesso em 6 jun. 2005.
- GANZMANN, J. Criteria for the evaluation of thesaurus software. *International Classification*, v.17, n. 3/4, p. 148-157, 1990. Disponível em: < <http://www.willpower.demon.co.uk/ganzmann.htm> > Acesso em 6 jun. 2005.
- GEMET. General Environmental Multilingual Thesaurus, 2005. Disponível em: < <http://www.eionet.eu.int/gemet> > Acesso em 6 jun. 2005.
- GOMES, H.E., CAMPOS, M.L.A, MOTTA, D.F. *Elaboração do tesouro documentário*. tutorial, 2004. Disponível em: < <http://conexaorio.com/bit/tesouro> > Acesso em 6 jun. 2005
- ISO 5964. Guidelines for the establishment and development of multilingual thesauri, 1985. Disponível em: < <http://www.iso.ch/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=12159> > Acesso em 6 jun. 2005.
- LANCASTER, F.W. *Vocabulary Control for Information Retrieval*. 2.ed. U.S.A.: IRP, 1986.
- MICROSOFT Corporation 2005. Disponível em: < <http://microsoft.com> > Acesso em 6 jun. 2005.
- MILSTEAD, J. L. *Use of thesauri in the full-text environment*. Indian Head, MD, The Jelem Company, 1998. Disponível em: < <http://www.bayside-indexing.com/Milstead/useof.htm> > Acesso em 6 jun. 2005
- MULSYSTEMS. *Multites*. Disponível em: < <http://www.multites.com> > Acesso em 6 jun. 2005.
- ORACLE Corporation 2005. Disponível em: < <http://www.oracle.com> > Acesso em 6 jun. 2005.
- RANGANATHAN, S.R. *Prolegomena to library classification*. New Delhi: Ejnar Munksgaard, 1951.
- SILVEIRA, M. de L. *Recuperação vertical de informação*. um estudo de caso na area jurídica. Tese (Doutorado)— Universidade Federal de Minas Gerais, Belo Horizonte, 2003.
- SOERGEL, D. Thesauri and ontologies in digital libraries: tutorial. In: *European Conference on Digital Libraries (ECDL 2002)*, Rome, Italy, 2002. Disponível em: < http://www.dsoergel.com/cv/B63_rome.pdf > Acesso em 6 jun. 2005.
- SYNAPSE CORPORATION. *Synaptica*. Disponível em: < <http://www.synaptica.com> > Acesso em 6 jun. 2005.
- UMWELTBUNDESAMT. *Thesmañ*. Disponível em: < <http://www.umweltbundesamt.at/umwelt/produkte/superthes> > Acesso em 6 jun. 2005.
- UNESCO. *Guidelines for the establishment and development of monolingual thesauri*. Paris, 1973. 37p.
- WEBCHOIR. *Termchoir*. Disponível em: < <http://www.webchoir.com> > Acesso em 6 jun. 2005.
- WERSIG, G; NEVELING, U. *Terminology of Documentation*. Paris: Unesco, 1976. 274p.
- WILL, L. *Software for building and editing thesauri, 2005*. Disponível em: < <http://www.willpowerinfo.co.uk/thesoft.htm> > Acesso em 6 jun. 2005.