



## O VOCABULÁRIO CONTROLADO COMO INSTRUMENTO DE ORGANIZAÇÃO E REPRESENTAÇÃO DA INFORMAÇÃO NA FINEP

*Tatiana Almeida Rosali Fernandes Souza*

**Resumo:** A Financiadora de Estudos e Projetos (FINEP), como empresa nacional de financiamento, possui potencial relevante de dados das propostas recebidas de pesquisa e desenvolvimento em ciência, tecnologia e inovação no país. Como tal, se caracteriza como sistema de recuperação da informação. O trabalho analisa o Vocabulário Controlado FINEP (VCF) enquanto instrumento de organização e representação da informação das propostas de financiamento. A análise do VCF é feita por abordagem histórico-metodológica da construção do vocabulário, ressaltando aspectos da concepção e das etapas de desenvolvimento do instrumento, destacando as principais mudanças no tempo. Investiga a viabilidade de aplicação do processo de categorização dos descritores em uso como contribuição para avaliação e reestruturação do vocabulário e conclui sobre a viabilidade de aplicação do método de categorização no VCF, destacando a importância fundamental da definição dos descritores nesse processo.

**Palavras-chave:** Sistema de Recuperação da Informação. Vocabulário Controlado. Categorização.

**Abstract:** *The Financial National Agency (FINEP) has relevant data originated from research proposals on fields in science, technology and innovation. As such FINEP may be characterized as an information retrieval system. This work analyses the FINEP Controlled Vocabulary (VCF) as an instrument of organization and representation of information originated from the research proposals received. The VCF analyses are based on a historical and methodological approach on the vocabulary construction, pointing out aspects of its conception and improvements over time. It investigates the feasibility of applying classification methods for descriptors in use, aiming to contribute to the present vocabulary. Results indicate advantages of applying categorization methods on the controlled vocabulary, and emphasize the fundamental role played by the descriptor's definition as an effective element to the whole process.*

**Palavras-chave:** Information Retrieval Systems. Controlled Vocabulary. Categorization.

### 1 INTRODUÇÃO

O volume informacional gerado nos dias atuais desafia os profissionais de informação no que diz respeito ao armazenamento e ao tratamento da informação. Como representar o conteúdo de forma satisfatória de modo a permitir que os usuários tenham acesso a informações relevantes? A organização do conhecimento assume cada vez mais um papel estratégico nas tomadas de decisão pelas Empresas.



Somente com a modernização dos instrumentos que auxiliam essa organização será possível adequar-se às rápidas mudanças no crescimento da importância do uso e da capacidade de recuperação da informação.

Além disso, um dos maiores problemas enfrentados pelas grandes empresas hoje em dia é a quantidade de informação produzida por seus departamentos e serviços. A gestão da informação pressupõe a atuação dos profissionais de informação como mediadores entre a informação localizada nos estoques informacionais (qualquer que seja o tipo de documento, formato ou suporte) e os seus usuários (entendidos como indivíduos que buscam informação para uma necessidade específica). O aumento da informação em circulação nas empresas traz diversos problemas relacionados ao tratamento e à recuperação de informação, sendo necessários estudos para aprimorar os instrumentos de busca. A conjuntura atual de produção de informação representa um desafio em relação ao acesso. Numa empresa como a Financiadora de Estudos e Projetos (FINEP), na qual circula uma quantidade significativa de informações relevantes sobre o desenvolvimento e o avanço da ciência, da tecnologia e da inovação no país, a realidade não é diferente.

É cada vez mais importante, no contexto de empresas como a FINEP, poder contar com ferramentas que aumentem a agilidade e a precisão na recuperação da informação e auxiliem na padronização da linguagem documentária. Para que as etapas de seleção e análise das propostas de financiamento sejam cumpridas com êxito, a empresa necessita de uma infraestrutura informacional que esteja de acordo com a política nacional de Ciência, Tecnologia e Inovação (C,T&I).

A FINEP atua no fomento aos projetos de pesquisa voltados à inovação tecnológica. Como órgão oficial do Ministério da Ciência e Tecnologia (MCT), atende às necessidades de institutos de pesquisa, empresas privadas e públicas, agências internacionais, investidores e organizações do Terceiro Setor. Um elemento importante nesta infraestrutura informacional é a organização das informações das propostas de financiamento, alocadas no Núcleo de Documentação, e do material de apoio à análise realizada pelos técnicos, que são as revistas especializadas e os livros, que constituem o acervo da Biblioteca da empresa. É primordial que haja um padrão na organização e representação da informação, visando eficácia na recuperação.

Atualmente, os vocabulários controlados tornaram-se instrumentos importantes para os sistemas informatizados. Esses vocabulários têm por objetivo principal o controle da polissemia que ocorre na linguagem natural, fazendo com que uma palavra que pode eventualmente assumir diversos significados passe a ter só um, determinado pelo contexto no qual a palavra está inserida. Este controle terminológico diminui a polissemia existente na linguagem natural, fazendo com que indexadores (na entrada de dados) e usuários (na busca da informação) consigam resultados mais eficazes.

Neste contexto, o trabalho de desenvolvimento do Vocabulário Controlado FINEP (VCF) surge, objetivando uma padronização e unificação da linguagem, em princípio da Biblioteca e do Núcleo de Documentação e, posteriormente, para uso de toda a FINEP, na recuperação de informação.

A elaboração do projeto VCF visou contribuir com a organização do conhecimento e,



consequentemente, com o trabalho da equipe de técnicos e analistas da FINEP, numa iniciativa conjunta da Biblioteca e do Núcleo de Documentação. A melhoria dos serviços de informação prestados encontra aliado nas novas tecnologias de informação e comunicação, considerando a importância da recuperação de informação relevante e seu valor na tomada de decisão da empresa.

O VCF, elaborado por uma equipe de profissionais da informação (Arquivistas, Bibliotecários e Cientistas da Informação), visa unificar a terminologia no tratamento da informação dos acervos bibliográficos e de projetos aprovados pela FINEP através do uso de uma linguagem de indexação padronizada. O VCF é a base para indexação das propostas de financiamento do Núcleo de Documentação e do material do acervo contido na Biblioteca.

Atualmente, o uso de instrumentos de controle terminológico para apoio à indexação e recuperação da informação torna-se imprescindível diante dos Sistemas de Recuperação da Informação, acesso online à base de dados na internet e organização de bibliotecas visando o incremento na qualidade da informação recuperada pelos usuários.

Com base nos elementos expostos, o objetivo principal do presente estudo é analisar o VCF, enquanto instrumento de organização e representação da informação, no contexto da FINEP, enfatizando a metodologia de construção deste vocabulário e, mais especificamente, seu processo de categorização. A categorização foi ressaltada neste estudo, por se tratar de um processo que se encontra em fase de reavaliação e de reestruturação para atender às necessidades de informação dos usuários.

## 2 CATEGORIZAÇÃO

O método de Categorização ou Facetação foi desenvolvido na década de 20, do século XX, por Shiyali Ramamrita Ranganathan para a construção de uma tabela de classificação bibliográfica. Ranganathan conduziu seu trabalho no sentido de discutir a natureza dos domínios de conhecimento a partir da fundamentação do que se entende por conhecimento e formação de conceitos para definir o modo pelo qual a Teoria da Classificação Facetada se propõe a dividir estes domínios. Para Campos e Gomes (2008):

A Categorização é um processo que requer pensar o domínio de forma dedutiva, ou seja, determinar as classes de maior abrangência dentro da temática escolhida. Na verdade, aplicar a categorização é analisar o domínio a partir de recortes conceituais que permitem determinar a identidade dos conceitos (categorias) que fazem parte deste domínio.

Esta divisão se concretiza a partir do método das categorias fundamentais. Categorias fundamentais são ideias que permitem recortar um “universo de assunto” em classes abrangentes. Atuam como primeiro corte classificatório e fornecem uma visão do conjunto dos agrupamentos que ocorrem na estrutura, possibilitando o entendimento global da área. Para Vickery (1960), categorias são conceitos de alta generalidade e ampla aplicação empregados na interpretação do mundo. As categorias também são utilizadas para determinar as relações entre as facetas. Para Ranganathan, facetas são manifestações



de categorias do universo de um conhecimento estudado, são as classes mais abrangentes dentro de um universo de ideias em que se formam os renques e as cadeias. Para o entendimento das Categorias Fundamentais, encontramos no trabalho de Campos; Gomes; Motta (2006) a seguinte síntese:

- **Personalidade** - Categoria fundamental de grande dificuldade de identificação. Ranganathan propõe o método do resíduo para identificar sua manifestação: não é Tempo, não é Espaço, não é Energia, ou Matéria, portanto é considerada uma manifestação da categoria fundamental Personalidade. Aqui ele aplica o princípio hindu “Não é isso, não é isso”

- **Matéria** - As manifestações da categoria Matéria são de duas espécies: Material e Propriedade.

- **Energia** - A manifestação da categoria Energia se caracteriza pela ação de uma espécie ou outra. A ação pode ser entre e por todas as espécies de entidade, inanimada, animada, conceitual, intelectual e intuitiva.

- **Espaço** - A categoria Espaço é entendida como o local de pertencimento de um dado objeto, seja ele indivíduo, coisa, ideia, fenômeno, entre outras entidades. A superfície da terra, o espaço dentro dela e o espaço fora dela são manifestações desta categoria.

- **Tempo** - A categoria Tempo está de acordo com o que geralmente entendemos por esse termo. As ideias isoladas de tempo, como milênio, século, década, ano, e assim por diante.

A partir dessas categorias podemos então identificar e aplicar a categorização em conceitos utilizados por diferentes áreas do conhecimento que, no domínio interdisciplinar, se agregam nas pesquisas desenvolvidas em seu interior, preservando a sua ideia fundamental, o seu atributo. Segundo Campos; Gomes; Motta (2006)

[...] economia: teorias, política econômica (instrumentos/agentes); planejamento (operações); e assim por diante. Na construção civil: edificações, como casas, edifícios, teatros, praças (personalidade); técnicas de edificação (técnicas); cimento, pedra, cerâmica, vidro (materiais); os profissionais envolvidos (agentes); e assim por diante. Na biblioteconomia: instituições e organizações, bibliotecas, documentos, usuários, suporte documental (personalidade); tratamento documentário, recuperação de informação, aquisição (processos); linguagens documentárias (agentes dos processos); profissionais envolvidos (agentes dos processos); e assim por diante.

Posteriormente, o Classification Research Group (CRG) - fundado nos anos 50 do século XX com o objetivo de desenvolver estudos teóricos e práticos no âmbito da classificação - desdobrou estas categorias para a elaboração de classificações facetadas. Tais categorias, como bem colocado por Campos e Gomes (2006), são desdobramentos das categorias fundamentais de Ranganathan (PMEST) e se apresentam da seguinte forma:



|                                     |                                       |
|-------------------------------------|---------------------------------------|
| Coisas, substâncias, entidades      | Medidas                               |
| que ocorrem naturalmente            | processo, comportamento               |
| produtos                            | Objeto da ação (paciente)             |
| instrumentos                        | Relações entre coisas, interações     |
| constructos mentais                 | efeitos                               |
| Suas partes                         | reações                               |
| constituintes                       | Operações sobre coisas                |
| órgãos                              | experimentos, ensaios                 |
| Sistemas de coisas                  | operações mentais                     |
| Atributos de coisas                 | Propriedades de atributos, relações e |
| qualidades, propriedades, incluindo | operações                             |
| Estrutura                           | Lugar, condição                       |
|                                     | Tempo                                 |

O processo de categorização consiste em identificar as possíveis classes gerais (categorias) de conceitos que a área do conhecimento abrange, serve para orientar os profissionais no levantamento dos termos e auxilia na organização do domínio de conhecimento. Dahlberg (1978 apud CAMPOS; GOMES, 2006) enfatiza a importância fundamental da categoria na estruturação do conceito e do sistema de conceitos:

Podemos ver que as categorias têm uma capacidade de estrutura: não apenas estruturam, de fato, todos os nossos elementos de conhecimento e unidades do conhecimento; elas fornecem, ao mesmo tempo, por este meio, o esqueleto, os ossos e tendões para estruturar todo o nosso conhecimento. Com seu uso conscientioso, então, o corpo do nosso conhecimento pode se manter unido, pode se mover, pode se manter flexível - e pode crescer organicamente (DAHLBERG, 1978, p. 34 apud CAMPOS; GOMES, 2006).

A seção a seguir tem como objetivo apresentar o histórico e a metodologia de construção do VCF. Esta apresentação é de fundamental importância para o desenvolvimento e o embasamento dos procedimentos de análise realizados, tendo em vista que a metodologia de elaboração do instrumento estudado (o VCF) tem características específicas. Essas especificidades se destacam por se tratar de um instrumento para a recuperação da informação de uma empresa com a temática voltada para a área de Ciência, Tecnologia e Inovação (C,T&I) e se preocupa com a organização, representação e recuperação da informação, visando atender seus usuários com maior eficiência.

### **3 HISTÓRICO E METODOLOGIA DE CONSTRUÇÃO DO VOCABULÁRIO CONTROLADO FINEP (VCF)**

A Biblioteca e o Núcleo de Documentação da FINEP perceberam, além da importância,



a necessidade do tratamento para a recuperação das informações contidas tanto nos projetos, que solicitam financiamento, quanto nos documentos, que compõem o acervo da Biblioteca. Neste sentido, essas unidades tiveram a iniciativa conjunta de desenvolver um instrumento que desse conta da representação temática de seus documentos, ou seja, por assunto. Até então, o que vinha sendo feito era a representação descritiva dos aspectos formais dos documentos<sup>1</sup> e sua representação temática por meio de palavras e sem padronização terminológica. Esta prática mostrou-se pouco eficaz no momento da busca pela informação desejada, pois dispersa documentos que tratam dos mesmos conceitos, expressos por palavras diferentes. Para suprir esta deficiência, foi criado um projeto para a elaboração do VCF.

Desde a sua concepção, o VCF vem passando por diversas modificações decorrentes de mudanças na sua coordenação, acarretando modificações metodológicas.

No ano de 2003, a elaboração de uma proposta metodológica para a construção de um novo instrumento, que tinha como pré-requisito atender às necessidades tecnológicas atuais da FINEP, teve como fundamento os princípios metodológicos que regem a Teoria da Classificação Facetada de Shiyali Ramamrita Ranganathan (1963), a Teoria da Terminologia de Eugene Wüester (1981) e a Teoria do Conceito de Ingrid Dahlberg (1978). Este novo instrumento foi definido, na época, como um vocabulário sistematizado, pois era regido por princípios terminológicos com uma estrutura sistematizada de conceitos, no qual um termo, que no interior de um domínio, juntamente com os outros conceitos, formavam um sistema de conceitos, permitindo uma interface mais simples e útil para o pesquisador final.

Esta primeira etapa objetivou determinar os domínios temáticos para a elaboração do Vocabulário Sistematizado. Definiu-se que os recortes temáticos seriam determinados pelos Fundos Setoriais<sup>2</sup>. Os Fundos foram recortados por domínios de conhecimento ou atividade, o que facilitou a reunião de conceitos dentro de uma mesma área, formando um todo coeso. As áreas escolhidas foram Biotecnologia e Energia, que somaram um total de 845 termos levantados a partir da lista já existente na biblioteca (proveniente da indexação do acervo bibliográfico), dentro de cada domínio e a partir da indexação dos projetos dos Fundos Setoriais destas áreas (CT-BIOTEC e CT-ENERG). Esta primeira fase envolveu as seguintes atividades: Definição das fontes para o levantamento dos termos; Análise dos conceitos; Organização da estrutura do vocabulário (estabelecimento das relações, elaboração das notas de definição, implantação do vocabulário no software e validação da forma do termo, da definição e dos relacionamentos pelos especialistas); e Apresentação do Vocabulário para os setores da FINEP.

O processo de categorização nesta fase foi apoiado na análise das definições e nas categorias estabelecidas tendo como base as categorias fundamentais de Ranganathan (PMEST). Por exemplo, na área de Biotecnologia, foram estabelecidas apenas três categorias: Domínios de atividade, que são

1 No caso da Biblioteca: registro do nome do autor, título, editor, ano de publicação, títulos de revistas, etc.; no caso do Núcleo de Documentação: nome da instituição, ano e Estado do projeto, nº do contrato, etc.

2 Os Fundos Setoriais, criados a partir de 1999, são instrumentos de financiamento de projetos de pesquisa, desenvolvimento e inovação no País. Há 16 Fundos Setoriais, sendo 14 relativos a setores específicos e dois transversais.



as áreas em que a Biotecnologia atua; Entidade, que é formada por conceitos que representam objetos concretos e abstratos; e as Técnicas Biotecnológicas, que representam os processos e ações dentro do domínio de Biotecnologia.

Na segunda fase do projeto de elaboração do VCF, os termos já existentes na base de dados da Biblioteca foram definidos, padronizados e estruturados através de relacionamentos hierárquicos e associativos. E logo em seguida, o foco voltou-se para a indexação das propostas de financiamento aprovadas que compõem o acervo corrente do Núcleo de Documentação (arquivo central) da empresa.

Nesta fase, foi decidido utilizar uma nova forma de categorização, com base na Tabela de Áreas do Conhecimento (TAC) do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Para esta função foram utilizados somente dois níveis das áreas do conhecimento estabelecidas pelo CNPq, as grandes áreas e as subáreas, os outros níveis existentes não se mostraram necessários para o agrupamento de descritores, pois são níveis muito específicos. Por exemplo: Matemática, Estatística e Ciência da Computação, entre outros, foram agrupados em Ciências Matemáticas e Naturais. Já as subáreas de Engenharia Civil, Engenharia Sanitária e Ambiental, Engenharia Elétrica e outros tipos de Engenharia, foram agrupados na grande área Engenharias. E assim por diante. Além disso foi necessário também, inserir notações (códigos alfanuméricos) para identificar as grandes áreas e as respectivas subáreas. Como, por exemplo, a grande área Ciências Matemáticas e Naturais que recebeu a notação “A” e as respectivas subáreas agrupadas neste grupo que receberam a notação, de acordo com a ordem em que apareciam na tabela do CNPq. A subárea Matemática, por exemplo, tinha notação “A1”, Estatística tinha notação “A2”. No caso dos descritores não se enquadram em nenhuma das áreas do CNPq, ora seriam classificados na área Ciência e Tecnologia, ora numa área criada chamada Multidisciplinar. (FINANCIADORA DE ESTUDOS E PROJETOS, 2002, 2003, 2005, 2007; CAMPOS, 2004).

Porém, com o desenvolvimento do trabalho, este tipo de categorização começou a apresentar alguns problemas conflitantes com a metodologia originalmente concebida - Ranganathan (1963), Dalhberg (1978) e Wüester (1981). Para Ranganathan, a poli-hierarquia na categorização dos termos não é admitida, cada descritor deve ser inserido em uma só categoria / faceta. Esta metodologia foi utilizada na primeira fase de desenvolvimento do VCF, conflitando-se na segunda fase com a Tabela de Áreas do Conhecimento (TAC) do CNPq, por ser esta uma tabela enumerativa, subdividida em áreas canônicas do conhecimento. Exemplificando: um descritor muito utilizado na indexação dos documentos da FINEP - “Inovação de processo” poderia se inserir em várias áreas do conhecimento; pode haver inovação de processo na área Engenharia de produto ou na área de Agronomia, ou até mesmo na área de Administração. Desta forma, a TAC foi considerada como não adequada para o uso na função para a qual foi designada no desenvolvimento do VCF.

Atualmente o Vocabulário está passando por um processo de revisão e avaliação, tendo como foco principal a qualidade do tratamento, da indexação e da recuperação da informação relevante para a FINEP. Diretamente vinculada a esta revisão, encontra-se a categorização dos termos. O presente



trabalho explora o estabelecimento de novas categorias para a organização dos descritores do VCF, cujas análises e resultados são apresentados a seguir.

### 4 ANALISE E RESULTADOS

A etapa de coleta de dados foi realizada no VCF, que se encontra disponível no Portal da Informação da Intranet da FINEP.

Na busca realizada em outubro de 2010 foram identificadas 559 propostas de financiamento indexadas pelos Tipos de Inovação, sendo 113 por “Inovação de processo”, 421 por “Inovação de produto” e 25 por “Inovação de serviço”. Neste trabalho, para efeito de análise, foram selecionados os projetos indexados por “Inovação de processo”, descritor que é definido pela FINEP como:

Adoção de métodos de produção novos ou significativamente melhorados, incluindo métodos de entrega dos produtos. Tais métodos podem envolver mudanças no equipamento ou na organização da produção, ou uma combinação dessas mudanças, e podem derivar do uso de novo conhecimento. Os métodos podem ter por objetivo produzir ou entregar produtos tecnologicamente novos ou aprimorados, que não possam ser produzidos ou entregues com os métodos convencionais de produção, ou pretender aumentar a produção ou eficiência na entrega de produtos existentes. (FINANCIADORA DE ESTUDOS E PROJETOS, 2010).

A escolha de análise por este tipo de Inovação justifica-se pelo fato desse descritor ter sido usado mais enfaticamente pela FINEP nos últimos anos, indicando um interesse da instituição pelo financiamento de projetos desta temática. Além disso, as inovações de processo são sabidamente importantes como diferencial no desenvolvimento socioeconômico dos países. A segunda etapa da coleta de dados consistiu da extração da amostra para análise. Tendo como base o Portal de Informação da FINEP foi construído um relatório de todos os projetos indexados pelo descritor “Inovação de processo”.

Este relatório possibilitou a identificação de cada projeto por três itens de informação: o número de referência, o título do projeto e os descritores atribuídos.

A partir da leitura do relatório foi possível extrair os descritores que indexaram os projetos, nos quais necessariamente constava o descritor “Inovação de processo”. Foi contabilizado o total de 291 descritores únicos, após descarte dos descritores duplicados. Cumpre ressaltar que tais descritores foram atribuídos pela equipe de elaboração do VCF a partir da indexação dos projetos na íntegra, tendo como base o título, o objetivo e o plano de trabalho, uma vez que estas três partes são consideradas pela equipe do VCF como as que contem as principais informações da temática do projeto.

Importante ressaltar que somente a partir de 2007 os projetos aprovados pela FINEP passaram a ser indexados e que cada projeto recebe uma média de 5 a 7 descritores temáticos. Este total foi estabelecido pela equipe do VCF como satisfatório, em termos de exaustividade para os objetivos da FINEP em relação à recuperação da informação dos projetos sob financiamento. No entanto, em busca de maior especificidade de indexação por assunto das propostas foi criado no Portal de Informação o campo “busca livre”, através do qual o usuário tem a possibilidade de inserir qualquer palavra que



desejar e o programa, por sua vez, recupera tal palavra em qualquer parte do projeto. Isto torna o VCF um sistema híbrido, segundo Lancaster (1993, p. 223) “o termo híbrido é empregado para designar qualquer sistema que funcione com uma combinação de termos controlados e linguagem natural”. Lancaster ainda aponta que muitos autores chegaram à conclusão de que o sistema de recuperação ideal inclui uma parte de termos controlados e uma parte de texto livre.

A etapa seguinte do presente trabalho constituiu-se da definição dos descritores extraídos do relatório que formam a amostra analisada. Cumpre ressaltar que a definição do descritor para a elaboração de um instrumento de organização e recuperação da informação é etapa de extrema importância, porque este é o elemento de análise que vai permitir não apenas estabelecer a estrutura classificatória como identificar os termos equivalentes que constituem os parâmetros necessários para o processo de categorização dos descritores. Ou seja, somente com os descritores devidamente conceituados é possível identificar a categoria a qual pertence cada descritor.

No presente trabalho (como no processo habitual de elaboração do VCF) as fontes utilizadas para a definição dos descritores foram de natureza diversa. Estão mencionadas, a seguir, segundo a ordem de relevância adotada na busca das definições: Glossários e Dicionários Especializados em diversas áreas do conhecimento, Dicionários de Termos Técnicos, Tesauros, Legislações, Manuais Técnicos e Dicionários de Língua Portuguesa. Quando não foi possível encontrar uma definição para os descritores nas fontes impressas disponíveis no acervo da FINEP e nos Glossários Especializados existentes na internet, a busca foi realizada em sites institucionais, como de universidades e empresas públicas e privadas renomadas e também na enciclopédia livre Wikipédia. O uso da Wikipédia se justifica no caso dos termos deste vocabulário pelas seguintes razões. Primeiro por servir de fonte de referência inicial para o entendimento de termos muito complexos. E, segundo, por ser o VCF um vocabulário que lida com termos da área de inovação e, por esse motivo, muitos desses termos ainda não se encontram dicionarizados, principalmente no caso de equipamentos ou produtos novos, que ainda não foram inseridos no mercado.

Com os descritores organizados e definidos, a etapa da categorização foi iniciada. Como já mencionado, o elenco de categorias escolhido para análise foi a do Classification Research Group (CRG). Justifica-se esta escolha pelo fato de a classificação do CRG ter um maior número de categorias que a PMEST e, por este motivo, possibilitar uma hospedagem mais flexível dos termos em categorias. Além disso, por ser um desenvolvimento das Categorias Fundamentais de Ranganathan (PMEST), a classificação do CRG não apresenta conflito com a metodologia original, anteriormente estabelecida para a elaboração do VCF.

No processo de desenvolvimento da categorização da amostra dos descritores pertencentes ao VCF fez-se necessário o entendimento de cada uma das categorias que foram utilizadas neste processo. Neste ponto cumpre destacar que a teoria do CRG não está centrada em fonte única, sendo apresentada por um número considerável de trabalhos escritos por diferentes membros deste grupo. Além disso, o CRG não apresenta sua teoria de forma explícita em uma lista organizada de princípios, como fez



Ranganathan. Sendo assim, não foi encontrado na literatura nenhum estudo descritivo sobre cada uma das categorias do CRG.

Para estabelecer o que foi entendido por cada uma das categorias utilizadas nesta amostra, a natureza exploratória deste estudo permitiu que se considerasse como base de análise: os conceitos das categorias fundamentais de Ranganathan, os dicionários de língua portuguesa, além do próprio agrupamento dos descriptores, que de certa forma descrevem as categorias às quais pertencem.

A seguir, serão apresentadas as categorias identificadas a partir da definição dos descriptores. Também será explicitado o significado de cada uma destas categorias, ou seja, o que aqui foi entendido como o conceito de cada categoria, bem como apontados os casos de equivalência ou semelhança com as categorias fundamentais de Ranganathan: **Efeito** – o resultado produzido por uma ação ou um agente; **Entidade** – equivalente à categoria Personalidade de Ranganathan, portanto utilizou-se o método do resíduo para identificar os descriptores que aqui se inseriram, ou seja, o que não se caracterizava como alguma das outras categorias foi inserida nesta; **Instrumento** – aparelhos, objetos ou utensílios que servem para executar uma obra ou levar a efeito uma operação mecânica em qualquer arte, ciência ou ofício; **Lugar** – equivalente à categoria Espaço de Ranganathan, o lugar diz respeito ao local de pertencimento de um dado objeto, seja ele indivíduo, coisa, ideia, fenômeno, entre outras entidades; **Operação** – semelhante à categoria Energia de Ranganathan, a operação foi aqui entendida como uma ação ocorrida entre coisas; **Operações mentais** – compreende os tipos de estudos, pesquisas ou análises; **Parte** – parte de entidade; **Processo** – categoria também semelhante à categoria Energia de Ranganathan, porém, se caracteriza por um tipo de ação que ocorre naturalmente, como por exemplo, as reações químicas ou as doenças; **Produto** – resultado de uma produção; **Propriedade** – qualidade ou atributo inerente a alguma entidade; **Sistema** – conjunto ou combinação de coisas ou partes de modo a formarem um todo complexo ou unitário; **Substância** - aquilo que subsiste por si, sem dependência de quaisquer outros elementos accidentais, a matéria pura.

No desenvolvimento do processo de categorização verificamos que algumas categorias/facetas - como, por exemplo: aparelho, técnica e material - foram também identificadas. Porém, como o número de descriptores encontrados para tais categorias não justificava a inclusão das mesmas, todos os descriptores dessas categorias puderam ser adequadamente agrupados em categorias afins, como indicado a seguir: As Ciências, aqui entendidas como domínios do conhecimento, foram inseridas na categoria Entidade; Os Equipamentos e os Aparelhos foram agrupados na categoria Instrumento; As Técnicas e os Métodos foram categorizados como Operação; Os Materiais foram agrupados na categoria Produto; Os Setores e os Ramos de um domínio do conhecimento foram categorizados como Parte de coisa; Os atributos foram agrupados juntamente com as propriedades.

Como síntese do processo de coleta, tratamento dos dados do presente trabalho, os descriptores foram organizados alfabeticamente numa planilha. Para cada desritor foi acrescentada a respectiva definição e a categoria em que foi classificado. No processo de categorização da amostra de descriptores do VCF, os principais resultados foram: 1 desritor na categoria **Efeito**, 32 descriptores na categoria



**Entidade**, 21 descritores na categoria **Instrumento**, 15 descritores na categoria **Lugar**, 67 descritores na categoria **Operação**, 3 descritores na categoria **Operações mentais**, 19 descritores na categoria **Parte**, 14 descritores na categoria **Processo**, 41 descritores na categoria **Produto**, 7 descritores na categoria **Propriedade**, 42 descritores na categoria **Sistema** e 29 descritores na categoria **Substância**.

Como etapa complementar de tratamento de dados foi realizada a análise das definições atribuídas aos descritores contidos na amostra selecionada para este estudo. O objetivo desta análise foi identificar palavras e expressões, presentes nas definições dos descritores, que tiveram maior influência no processo de classificação no elenco definido de categorias anteriormente descrito. O resultado desta análise será descrito a seguir:

- Efeito - com apenas 1 descritor, a categoria foi caracterizada com seu próprio nome, já que o descritor, aqui categorizado, foi Efeito especial e foi definido como Efeito mecânico;
- Entidade - esta categoria apresentou poucas características comuns entre as definições de seus descritores, uma vez que esta é uma categoria bastante diversificada. Nesta categoria foram encontradas manifestações como Coisas, Plantas, Frutas, Indústrias, Disciplinas, Empresas, entre outras;
- Instrumento - as palavras mais significativas encontradas nas definições dos descritores foram: Dispositivo, Máquina, Equipamento, Aparelho e Ferramenta;
- Lugar - não há considerações acerca desta categoria, pois a maioria de seus descritores são Identificadores e no VCF este tipo de descritor não é definido;
- Operação - as palavras mais significativas encontradas nas definições dos descritores foram: Ação, Atividade, Implementação, Investigação, Método, Processo, Procedimento, e Utilização;
- Operações mentais - as palavras mais significativas encontradas nas definições dos descritores foram: Estudo e Pesquisa;
- Parte - as palavras mais significativas encontradas nas definições dos descritores foram: Parte, Ramo e Setor;
- Processo - as palavras mais significativas encontradas nas definições dos descritores foram: Alteração biológica, Doença e Processo;
- Produto - as palavras mais significativas encontradas nas definições dos descritores da categoria foram: Agente, Material, Mistura e Produto;
- Propriedade - as palavras mais significativas encontradas nas definições dos descritores foram: Capacidade, Equilíbrio, Propriedade;
- Sistema - as palavras mais significativas encontradas nas definições dos descritores foram: Conjunto, Sistema e Variedade;
- Substância - as palavras mais significativas encontradas nas definições dos descritores foram: Substância, Material e Composto químico.

Importante destacar que os resultados desta análise complementar corroboraram os resultados das análises anteriores do processo de definição das categorias, anteriormente descrito. Em outras



palavras, os projetos com pedidos de financiamento aprovados pela FINEP, indexados por “Inovação de processo”, puderam ser categorizados satisfatoriamente com respaldo da definição dos descritores.

Como principal resultado do processo de categorização, destacamos que as categorias Operação e Processo são entendidas como ação entre coisas, condição para a implantação de um Sistema, cujo resultado é um Produto, definido e/ou pesquisado por uma Entidade e/ou Parte. A Substância é a própria matéria, que pode ser objeto da pesquisa científica. A categoria Instrumento nos indica como operacionalizar a pesquisa, localizando-a em Lugar. A categoria Propriedade indica os benefícios, ou não, advindos do objeto pesquisado através das Operações Mentais e seus Efeitos.

Outro aspecto a destacar da análise é a ordem das categorias que após o estudo estatístico apresentou a seguinte sequencia de ocorrência: Operação, Sistema, Produto, Entidade, Substância, Instrumento, Parte, Lugar, Processo, Propriedade, Operações Mentais e Efeito.

Esses resultados não só respaldam satisfatoriamente o encaminhamento teórico-metodológico adotado no presente trabalho, como também indicam uma possível nova forma de organizar as categorias na indexação de propostas de financiamento encaminhadas e aprovadas à FINEP para recuperação da informação.

## 4 CONSIDERAÇÕES FINAIS

O presente trabalho teve como objetivo analisar o Vocabulário Controlado FINEP enquanto instrumento de organização e representação da informação e investigar a aplicação do método de categorização neste instrumento.

A partir das análises realizadas foi possível observar que o trabalho de elaboração de um instrumento de controle terminológico é bastante minucioso e requer uma série de estudos, pesquisas, decisões e testes.

Um ponto importante a ser ressaltado neste estudo foi a busca por definições atribuídas aos descritores. Por ser o VCF um instrumento que abarca diversas áreas do conhecimento é necessário que se tenha um acervo bastante diversificado de glossários, dicionários especializados, dicionários técnicos, manuais, etc. No presente trabalho, foi necessário, em alguns casos, recorrer a sites da internet, já que não se conseguiu encontrar as definições nas obras de referência existentes no acervo FINEP, o que no caso estudado não chegou a comprometer as análises. No entanto, cumpre ressaltar que a falta de credibilidade das fontes de onde são retiradas as definições dos descritores de um vocabulário controlado pode comprometer a estrutura em relação à consistência da categorização e dos relacionamentos entre os descritores.

Outra questão importante, ainda sobre a definição, é a validação por parte de especialistas das áreas das propostas financiadas abarcadas pelo VCF. O vocabulário em questão deve ter uma validação periódica das definições, uma vez que é sabido que a qualidade da definição está diretamente ligada à qualidade da categorização e, consequentemente, à qualidade da organização e recuperação da informação.



Um ponto relevante a ser destacado neste estudo foi a complexidade e a dificuldade da tentativa de conceituar o elenco das categorias do CRG, principalmente porque este processo exige um esforço intelectual considerável, além de um conhecimento de todo o referencial teórico da área de organização e representação do conhecimento. Porém, mesmo considerando que algumas decisões tomadas acerca dos conceitos de cada uma das categorias possam suscitar críticas ou questionamentos, os resultados da aplicação demonstraram que esses conceitos auxiliaram na diminuição da margem de erro no processo de categorização, o que não descarta a possibilidade de serem aprimorados futuramente.

Quanto à categorização, durante o desenvolvimento deste processo constatou-se que há categorias mais simples de se identificar e outras mais complexas que requerem uma interpretação mais cautelosa. Esta constatação fica bem visível na parte da análise das palavras e expressões significativas das definições. Destacamos que, numa definição bem estruturada, estas palavras ou expressões costumam aparecer logo no início da definição, em que é indicada a natureza do descritor: se Entidade, Processo, Sistema etc. O restante da definição, via de regra, descreve a função ou a composição do descritor. Nas categorias mais simples de se identificar, as palavras significativas que aparecem são, na maioria, quase sinônimas do nome da categoria, enquanto nas categorias mais abstratas essas palavras precisam/necessitam ser interpretadas a partir da definição do descritor e do conceito da categoria.

A partir da análise do processo de categorização foi possível verificar a importância da aplicação deste processo na construção de instrumentos de controle terminológico, visando à eficiência na representação do conhecimento para a recuperação da informação.

Apesar de no projeto inicial de implantação do VCF ter sido preconizado a utilização das categorias de Ranganathan (PMEST) e, por conjunturas variadas, ter sido adotada posteriormente a organização dos descritores pelas áreas do conhecimento do CNPq - que logo se mostrou insuficiente para atender às demandas da FINEP - verificou-se que a escolha da aplicação das categorias do CRG no VCF seria satisfatória, não sendo necessárias modificações nas categorias já existentes, uma vez que foram suficientes para agrupar os descritores da amostra estudada.

Além disso, no caso do VCF, os resultados revelaram que a categorização possibilitou uma estruturação sistemática dos descritores, qualificando ainda mais este instrumento e otimizando os processos de indexação e recuperação da informação na FINEP.

Desta forma, acredita-se também que a classificação do CRG apresenta potencial para ser aplicada às outras áreas do vocabulário. Para que isso ocorra, propõe-se um estudo mais aprofundado sobre os resultados desta aplicação na representação e recuperação da informação na FINEP. Além disso, é também desejável um estudo das necessidades dos usuários deste instrumento.

O processo de controlar/padronizar um vocabulário é dinâmico, como é dinâmica também a própria construção do conhecimento. Desta forma, acredita-se no encaminhamento metodológico de elaboração do VCF como proposto neste trabalho, principalmente porque esta metodologia é sustentada por teorias clássicas da organização e representação do conhecimento, já estabelecidas na Biblioteconomia e na Ciência da Informação.



Por outro lado, a metodologia e o produto gerado neste trabalho permitirão à FINEP avançar com qualidade no tratamento terminológico de temáticas de pesquisa no âmbito da Ciência, Tecnologia e Inovação.

Por último, é necessário reforçar que o tratamento informacional em uma organização é fator fundamental para a tomada de decisões e, no caso específico da FINEP, um elemento estratégico para o gestor.

### REFERÊNCIAS

ARAÚJO, Vânia M. R. H. Sistemas de recuperação da informação: nova abordagem teórico conceitual. **Ciência da Informação**, Brasília, v. 24, n. 1, 1995.

BATTAGLIA, Maria Glória Botelho. A inteligência competitiva modelando o Sistema de Informação de Clientes – FINEP. **Ciência da Informação**, Brasília, v. 29, n. 2, p. 200-214, maio/ago. 1999.

CAMPOS, Maria Luiza de Almeida. **Linguagem Documentária**: teorias que fundamentam sua elaboração. Niterói: EdUFF, 2001.

CAMPOS, Maria Luiza de Almeida. **Vocabulário controlado**: instrumento para a gestão dos recursos informacionais dos acervos da FINEP. Apresentação. Rio de Janeiro, 2004.

CAMPOS, Maria Luiza de Almeida; GOMES, Hagar Espanha. Metodologia de elaboração de tesouro conceitual: a categorização como princípio norteador. **Perspectiva em Ciência da Informação**, Belo Horizonte, v. 11, n. 3, p. 348-359, set./dez. 2006. Disponível em: <[http://www.scielo.br/scielo.php?pid=S1413-99362006000300005&script=sci\\_arttext](http://www.scielo.br/scielo.php?pid=S1413-99362006000300005&script=sci_arttext)>. Acesso em: 20 jan. 2011.

CAMPOS, Maria Luiza de Almeida; GOMES, Hagar Espanha. Taxonomia e classificação: o princípio de categorização. **DataGramZero: Revista de Ciência da Informação**. v. 9, n. 4, ago. 2008. Disponível em: <[http://www.dgz.org.br/ago08/Art\\_01.htm](http://www.dgz.org.br/ago08/Art_01.htm)>. Acesso em: 22 out. 2010. Não Paginado.

CAMPOS, Maria Luiza de Almeida; GOMES, Hagar Espanha; MOTTA, Dilza (Coord.) Fonseca da. **Elaboração de tesouro documentário**: tutorial. 2006. Disponível em: <<http://www.conexaorio.com/biti/tesauro/>>. Acesso em: 20 fev. 2011. Não Paginado.

CINTRA, Ana Maria et al. **Para entender as linguagens documentárias**. 2. ed. rev. e ampl. São Paulo: Polis, 2002.

FINANCIADORA DE ESTUDOS E PROJETOS. Disponível em: <[www.finep.gov.br](http://www.finep.gov.br)>. Acesso em: 16 jun. 2010.

FINANCIADORA DE ESTUDOS E PROJETOS. **Glossário**. Disponível em: <<http://www.finep>>.



gov.br/o\_que\_e\_a\_finep/conceitos\_ct.asp>. Acesso em: 22 ago. 2010.

FINANCIADORA DE ESTUDOS E PROJETOS. **Portal da Informação**. Disponível em: <www.finep.gov.br/intranet>. Acesso em: 07 jan. 2011.

FINANCIADORA DE ESTUDOS E PROJETOS. **Proposta de consultoria para os acervos da FINEP**: indexação e recuperação da informação. Proposta. Rio de Janeiro, 2005.

FINANCIADORA DE ESTUDOS E PROJETOS. **Relatório das atividades da consultora Maria Luiza de Almeida Campos**. Relatório de atividades. Rio de Janeiro, 2003.

FINANCIADORA DE ESTUDOS E PROJETOS. **Vocabulário controlado FINEP (VCF)**: indexação e recuperação do acervo arquivístico. Relatório de atividades. Rio de Janeiro, 2007.

FINANCIADORA DE ESTUDOS E PROJETOS. **Vocabulário controlado**: uma nova metodologia visando o tratamento terminológico de documentos para fins de recuperação. Proposta. Rio de Janeiro, 2002.

FOSKETT, Antony Charles. **A abordagem temática da informação**. São Paulo: Polígozo; Brasília: Ed. Universidade de Brasília, 1973.

LANCASTER, F. W. **Construção e uso do tesauro**: curso condensado. Trad. de César Almeida de Menezes e Silva. Brasília: IBICT, 1987. 114p.

\_\_\_\_\_. **Indexação e resumos**: teoria e prática. Brasília: Briquet de Lemos, 1993. 347p.

\_\_\_\_\_. **Indexação e resumos**: teoria e prática. 2. ed. Brasília: Briquet de Lemos. 2004, 452p.

\_\_\_\_\_. **Vocabulary Control for Information Retrieval**. Virginia: Information Resources Press, 1986. 270p.

XII Encontro Nacional de Pesquisa em Ciência da Informação – ENANCIB 2011