

METODOLOGIA PARA DEFINIÇÃO DE CAMPOS EM BANCO DE DADOS

ANNA DA SOLEDADE VIEIRA

Dissertação apresentada ao Instituto Brasileiro de Bibliografia e Documentação/Universidade Federal do Rio de Janeiro para obtenção do grau de mestre em Biblioteconomia e Documentação.

Orientador:

Professor Doutor Bert Boyce

University of Missouri

Rio de Janeiro

1974

## SINOPSE

Partindo de dados obtidos no Arquivo da FINEP (Fi  
nanciadora de Estudos e Projetos) e com a finalidade de montar  
um banco de dados sobre documentação típica de administração de  
projetos, duas metodologias de bases estatísticas são apresenta  
das para a definição de campos de informação. Uma, baseada na  
ordenação dos campos segundo sua frequência nas consultas dos  
usuários, seria cabível quando estes tivessem interesses e ati  
vidades comuns. A segunda metodologia - teste do  $\chi^2$  - seria a  
plicável quando os usuários tivessem interesses e atividades di  
versificados.

## SUMÁRIO

1. INTRODUÇÃO .....	p. 3-5
1.1 - Importância dos Arquivos	
1.2 - Organização de Arquivos	
2. TEMA DA PESQUISA .....	p. 6-9
2.1 - Objetivo	
3. AMBIENTE DO SISTEMA .....	p.10-2
4. METODOLOGIA .....	p.13-25
4.1 - Coleta de Dados	
4.2 - Tratamento, Análise e Interpretação	
4.2.1. Potencialidade versus Uso do Sistema	
4.2.2. Identificação dos Componentes a partir das Questões dos Usuários	
4.2.2.1. Distribuição Geral	
4.2.2.2. Distribuição por Departamentos	
4.2.2.3. Distribuição por Tipo de Documento	
5. CONCLUSÕES .....	p.26-28
6. TABELAS .....	p. 29-49
7. REFERÊNCIAS .....	p. 50
8. BIBLIOGRAFIA .....	p.51.52

## 1. INTRODUÇÃO

Não obstante alguns dos mais sérios problemas da sociedade industrial ainda estarem associados a métodos desatualizados ou inadequados de tratar e interpretar informações, o interesse emergente pela formação de coleções de documentos relevantes ligadas a serviços de informação nas áreas de Ciência e Tecnologia são indicativos de que o registro dos eventos naquele domínio tornou-se um requisito básico da civilização. (1)

Cada elemento isolado que o sistema registra constitui um dado, o qual, ao ser elaborado ou conjugado a outros para fins de comunicação, transforma-se em informação. Poder-se-ia dizer que, em um sistema de recuperação de informações o dado é a matéria prima e a informação é o produto capaz de gerar uma ação, modificar um comportamento ou propiciar uma tomada de decisão

A responsabilidade do tratamento de dados e informações é compartilhada por entidades públicas e privadas, embora ao Governo caiba a maior parcela dessa tarefa, o que justifica a ênfase atualmente dada aos sistemas nacionais de informação científica e técnica.

Quanto às funções de um sistema nacional de informação científica e técnica, pode-se dizer, com base nas recomendações da Organização para a Cooperação e Desenvolvimento Econômico (OCDE), (2) que ele deve:

- a) permitir que a soma do conhecimento acumulado no domínio da Ciência, da Tecnologia, da Economia e das Ciências Sociais seja utilizada criteriosamente, de modo a que a nação possa alcançar seus objetivos relacionados com o maior bem-estar da sociedade;
- b) fornecer subsídio para avaliação do desenvolvimento da Ciência e da Tecnologia nacionais;

- c) colocar à disposição das pessoas encarregadas de tomar decisões administrativas e políticas, tanto no Governo quanto nas empresas privadas, o conjunto de dados necessários ao processo decisório;
- d) atrair a atenção dos poderes públicos e dos órgãos privados para os problemas de acesso e utilização da informação.

No processamento da informação tendo em vista a pesquisa científica, o desenvolvimento econômico e o bem-estar social, arquivos, bibliotecas e centros de documentação têm igual responsabilidade e importância. Entretanto, estas três instituições não têm recebido idêntico tratamento, seja em âmbito oficial, seja no setor privado, ficando os arquivos relegados a um plano inferior.

O depoimento de Henri B.Branche (3) sobre uma situação particular - o Arquivo Nacional na década passada - é bem representativo do panorama geral brasileiro que ainda hoje perdura, conforme se pode comprovar pela denúncia de J.Esposel (4) do descasso no tratamento da documentação histórica e administrativa no Brasil. Apesar disso, esforços isolados vem sendo feitos no sentido de modernização do tratamento de arquivos.

#### 1.1 - Importância dos Arquivos

Por se constituírem de documentos gráficos, geralmente não publicados ou de publicação limitada, os arquivos representam uma importante parcela dos registros nacionais, seja no aspecto da documentação histórica, econômica e social, seja no que concerne à documentação científica e técnica.

Do ponto de vista empresarial, a organização gera documentos e, dentro de uma perspectiva de administração científica, as possibilidades de êxito se baseiam na programação do trabalho, na precisão das informações e na rapidez com que elas são fornecidas.

As funções dos arquivos em geral evidenciam sua importância:

- coleccionar documentos básicos, os quais constituem a memória de uma instituição ou de uma cultura;
- preservar os documentos de importância permanente, seja visando a demanda atual, seja para transmissão às gerações futuras;
- analisar a informação de modo a facilitar seu uso;
- tornar acessíveis as informações que irão propiciar uma tomada de decisão administrativa.

## 1.2 - Organização de Arquivos

Para que um arquivo desempenhe as funções citadas, são necessárias algumas definições básicas dos seguintes pontos:

- área de atuação ou especialização, uma decorrência natural das características do organismo a que está vinculado e das restrições naturais impostas ao sistema;
- critérios de seleção dos documentos a serem arquivados, com base nos assuntos e materiais da área de especialização da Entidade, bem como na vida-média dos documentos;
- sistema de arquivamento a ser utilizado, visando à conservação dos documentos e à facilidade de consulta;
- sistemas de análise e indexação dos documentos;
- métodos adequados de recuperação de informações;
- meios de tornar o documento ou informação disponível ao usuário.

## 2. TEMA DA PESQUISA

Embora existam possibilidades de generalização so bre todos os aspectos acima enumerados e para todos os tipos de arquivos, o presente trabalho enfoca:

- arquivos especializados em administração de projetos;
- definição de campos de informação para um banco de dados em computador.

O material experimental da pesquisa foi coletado na Financiadora de Estudos e Projetos - FINEP, empresa pública cujas finalidades são o apoio e a administração de projetos na linha do desenvolvimento sócio-econômico nacional.

Banco de dados, no presente trabalho, é tomado como uma coleção de informações interrelacionadas de maneira coerente e que podem ser recuperadas sob quaisquer chaves de classificação ou condições lógicas, embora armazenadas de maneira não redundante. (5) Ressalve-se apenas quanto a esse conceito que, embora intencional, a não-redundância é difícil de ser atingida, sendo esse controle muitas vezes mais dispendioso em custo e esforços que a duplicação de dados no sistema.

Cada área do banco de dados, destinada a receber dados com idênticas características, é denominada campo.

### 2.1 - Objetivo

Esta é uma tentativa de desenvolvimento de metodologias alternativas, com vistas a estabelecer um modelo de sistema de recuperação de informações para arquivos de administração de projetos, podendo a metodologia aqui desenvolvida servir de orientação para arquivos de outras áreas.

No atual contexto, sistema de recuperação de informações deve ser compreendido como o conjunto ideal de campos que comporão o banco de dados e aos quais a indexação deverá se estender para caracterização exata de cada documento.

Desde que o propósito de qualquer sistema de informação é prover o usuário com documentos relevantes ao seu interesse, ele deve ser solicitado a estabelecer os parâmetros da recuperação e a julgar o produto recebido. As medidas mais comumente usadas para avaliar a relevância da recuperação são precisão e revocação ("recall"). Precisão refere-se à capacidade do sistema de rejeitar os documentos não-relevantes à pergunta, enquanto revocação mede sua capacidade de recuperar todos os documentos relevantes (6).

A observação do comportamento dos usuários do Arquivo da FINEP leva à conclusão de que a precisão é mais importante que a revocação na recuperação de documentos para efeitos de administração de projetos. Assim é que, para o administrador saber se uma atitude do mutuário x apoia-se nos termos do acordo firmado, somente através do contrato referente àquele projeto específico poderá ser esclarecida sua dúvida; nenhum outro documento do mesmo projeto ou qualquer contrato de outro projeto dar-lhe-á as informações necessárias. Solicitações de todos os documentos de um certo conjunto são pouco frequentes, o que confirma a menor importância da revocação para o sistema.

Partindo-se da premissa anterior e em se tratando de documentos com caracteres diferenciais (facetas) muito numerosos e diversificados, é pressuposto que a maior precisão está diretamente relacionada com três fatores:

- a exaustividade na definição do banco de dados, isto é, a criação de tal variedade de campos de informação que possibilite exaustividade na indexação e na estratégia de busca;
- a profundidade da indexação, isto é, cada documento deverá ser descrito sob todas as suas facetas (exaustividade) e, dentro de cada faceta,



aos de conteúdo e situacionais. Cada uma das características rejeitadas na definição do sistema redundaria em indexação e busca deficientes, resultando, portanto, em recuperação com baixa precisão.

Pesquisa interna do Centro de Processamento de Dados (CPD) da FINEP concluiu pela avaliação do custo de cada novo termo (unidades de informação armazenadas em um campo) de seu banco de dados em cerca de Cr\$0,19. Uma vez que a cada novo campo incluído no sistema corresponde um aumento de custo igual ao produto de Cr\$0,19 pelo total de seus termos a serem indexados, o equilíbrio na definição do banco de dados deverá ser encontrado através do grau de utilização dos campos pelos usuários.

O estabelecimento do presente modelo parte da análise de dados reais, sendo apresentados, a seguir, os resultados da observação do funcionamento da FINEP e do comportamento dos usuários de seu Arquivo quanto à busca de documentos. Em complementação, são apresentados também os dados relativos à análise da documentação circulante.

### 3. AMBIENTE DO SISTEMA

A fim de que quaisquer generalizações aqui feitas para arquivos de administração de projetos possam ser compreendidas dentro de suas limitações, é necessário que se descreva a FINEP e seu Arquivo, ambiente em que se desenvolveu a pesquisa e a partir do qual todas as conclusões foram extraídas.

A FINEP, órgão da Secretaria de Planejamento da Presidência da República, é constituída por treze setores, a saber: Presidência (PRES), Vice-Presidência (VICE-PRES), Departamento Jurídico (DEJ), Departamento Administrativo (DAD), Departamento Financeiro-Contábil (DFC), Centro de Processamento de Dados (CPD), Grupo de Ciência e Tecnologia (GCT), Grupo de Estudos e Projetos (GEP), Grupo de Pesquisa (GP), Grupo de Estudos de Fontes Alternativas de Energia (GE), Núcleo do Banco Interamericano de Desenvolvimento (N.BID), Grupo de Assessoria para o Gás Combustível (G.GAS) e Programa Nacional de Treinamento de Executivos (PNTE).

Foi instituída pelo Decreto nº 61.056 de 24 de julho de 1967, sendo hoje regida pelo Decreto nº 71.133, de 21 de setembro de 1972. Seu campo básico de atuação é o financiamento de estudos, projetos e programas de desenvolvimento econômico, social, científico e tecnológico, apresentados por entidades públicas e privadas.

Os programas atualmente em vigor são: pré-inversão, apoio à consultoria nacional e desenvolvimento da tecnologia industrial, financiados com recursos provenientes de empréstimos de organismos internacionais e com recursos orçamentários próprios ou oriundos do Fundo Nacional de Desenvolvimento Científico e Tecnológico (FNDOCT), do qual é a Secretaria Executiva. Os financiamentos são contratados diretamente com o mutuário ou através de Agentes Financeiros (Bancos de Desenvolvimento Regionais).

Embora cada programa tenha características próprias, em geral os projetos seguem duas fases distintas:

- a) anterior ao contrato, quando a entidade solicitante e o projeto apresentado são analisados em face do financiamento pretendido, resultando daí a aprovação ou o indeferimento da solicitação. A documentação deste período é em grande parte jurídica, isto é, comprobatória de exigências regulamentares, indo fornecer elementos para a preparação do contrato;
- b) posterior à assinatura do contrato, fase essa referente ao acompanhamento técnico do projeto e do controle da aplicação dos recursos. A documentação desta fase é predominantemente técnica e financeiro-contábil, embora decorrente de cláusulas contratuais.

Toda a documentação gerada desde a solicitação inicial e durante todas as etapas da vigência dos contratos vai agregar-se ao Arquivo, de uso reservado aos funcionários da Empresa.

A seleção dos documentos é iniciada no Protocolo e completada no Arquivo, onde também é feita a indexação.

O arquivamento físico dos documentos se faz de acordo com o código de identificação dos projetos. Assim, o código F/089P01D025 refere-se ao vigésimo quinto documento (D) da primeira pasta (P) componente do projeto número oitenta e nove do programa de pré-inversão (F). A documentação jurídica (contratos, certidões, procurações etc) é arquivada separadamente da documentação de acompanhamento do projeto (correspondência, relatório, prestação de contas etc). A sequência dos documentos dentro de cada pasta é cronológica.

Para otimizar as tarefas de administração de contratos foi criado um banco de dados, compreendendo quatro partes ou arquivos: um, referente aos eventos esperados e ocorridos sobre cada contrato (ADM-CONT); o segundo, relativo ao controle financeiro-contábil (CASH-FLOW); o terceiro, contendo as características de todos os projetos em andamento (CADASTRO) e, finalmente, o quarto, objeto desse estudo, referente à documentação do Arquivo (DOCUMENTOS).

As razões da escolha do sistema de banco de dados estão vinculadas a sua eficiência, uma vez que:

- o potencial informativo da empresa em questão fica centralizado;
- cada informação é armazenada uma só vez em um só registro, independentemente da variedade de usos posteriores;
- a recuperação é feita pela combinação lógica dos termos usados na indexação, de acordo com a necessidade particular do usuário;
- uma total independência pode existir entre programação ("software") e armazenamento físico;
- um sistema geral pode atender a todas as necessidades de informação da empresa, dispensando programas específicos para cada problema ou situação. (Entenda-se programa no duplo sentido de metas de ação e de "software" para o computador);
- a centralização dos dados leva à economia de tempo no processamento e na recuperação de informações, bem como propicia o melhor controle administrativo.

#### 4. METODOLOGIA

O planejamento de qualquer sistema de informação deve levar em conta as condições impostas pelo ambiente e a demanda provável por parte dos usuários.

Assim sendo, foram considerados como pontos norteadores da definição de campos para o banco de dados os seguintes:

- as expectativas da Empresa em relação ao sistema, isto é, rapidez e precisão de informações capazes de assegurar a eficiência da administração dos projetos em andamento;
- as características da documentação com a qual o sistema terá de se envolver;
- os aspectos qualitativos e quantitativos dos pedidos atualmente dirigidos ao Arquivo.

##### 4.1 Coleta de Dados

Com a finalidade de identificar as necessidades dos usuários do Arquivo, bem como seu comportamento em relação à busca de documentos, foram coletadas todas as perguntas feitas por eles ao Arquivo, tanto pessoalmente quanto por telefone, durante o mês de julho do corrente ano.

Sendo a análise e a execução dos projetos atividades continuadas, cada mês se iniciam e se concluem contratos. Não há, portanto, épocas de pique, nem de baixa procura ao Arquivo. Daí se justificar uma amostra aleatória simples, tendo sido escolhido o mês de julho, após consulta à tabela de números equiprováveis de Hald (9), onde, começada a busca a partir da décima quinta linha e terceira coluna, foi encontrado o número aleatório 7.

Os pedidos, no total de 224, foram anotados exatamente conforme o solicitante se expressou. A seguir cada per

gunta foi analisada para identificação dos campos que a computam e registrada a ocorrência desses campos nas perguntas.

Paralelamente, foi feita a análise da documentação fornecida, identificados os campos de informação existentes e verificada a frequência de sua ocorrência nos 224 documentos.

Os cálculos foram parcialmente executados em computador IBM/360-40, do Centro de Computação da UFMG.

Uma visão crítica dos dados, face ao conhecimento da Empresa e do movimento do Arquivo, leva a concluir pela validade dos mesmos, uma vez que:

- no dia-a-dia os pedidos se referem realmente a 17 tipos de documentos bastante característicos da administração de contratos: carta, circular, ofício, balancete, fatura, guia, JE, recibo, "voucher", cadastro de consultoras, projeto, relatório, ato constitutivo, contrato, decisão, parecer e procuração;
- que os 4 Departamentos (PNTE, N.BID, GE e G. GAS) que não consultaram o Arquivo no período de observação ou aquele (GP) que consultou apenas uma vez, têm normalmente frequência nula, possuindo arquivos próprios e exercendo atividades especiais desvinculadas do problema de controle da aplicação de recursos da FINEP em projetos externos.

Os campos identificados nas perguntas foram reconhecidos também na documentação, diferindo apenas na frequência. São ao todo 29, a saber:

- tipo de documento: o aspecto formal do documento. Exemplo: carta, contrato, ofício etc.
- veículo da informação: o canal de comunicação. Exemplo: Diário Oficial da União onde se publicam os contratos aprovados.
- número do documento: número com que a instituição de origem caracteriza o documento;

- número do protocolo: número através do qual a FINEP incorpora o documento ao seu acervo.
- data do documento: data de origem.
- data do protocolo: data da incorporação do documento ao Arquivo da FINEP.
- instituição de origem: nome da entidade da qual provém o documento.
- instituição de destino: nome da entidade à qual o documento se destina.
- signatário: nome da pessoa que assina o documento.
- pessoa destinatária: nome da pessoa a quem o documento é endereçado.
- cargo do signatário: posto ocupado pelo signatário do documento.
- cargo do destinatário: posto ocupado pelo destinatário do documento.
- assunto: as ações administrativas com que o documento se relaciona ou o campo do conhecimento sobre o qual versa.
- referências: correlação de conteúdo entre documentos, um mencionando outro.
- anexos: correlação física entre documentos, um apenso a outro.
- código do projeto: código alfanumérico representativo do projeto.
- nome do projeto: nome oficial do projeto.
- variações do nome do projeto: apelidos que o projeto recebe internamente.
- mutuário: entidade responsável pelo projeto.
- executor: setor subordinado ao mutuário, onde é implantado o projeto.

- setor FINEP: nome do programa da FINEP ao qual o projeto está vinculado;
- classificação do projeto: área em que o projeto se enquadra, seja na classificação interna do GEP, seja na do Plano Básico de Desenvolvimento Científico e Tecnológico (PBDCT) utilizada pelo GCT;
- fonte de recursos: Instituições nacionais e internacionais de onde provêm os recursos aplicados ao projeto;
- Estado(s) do Brasil: local onde se realiza o projeto;
- agente financeiro (AF): Banco de Desenvolvimento regional responsável pelo repasse de verbas;
- agência do AF: subdivisão estadual dos Bancos regionais;
- nome da consultora: nome do escritório técnico que dá consultoria ao projeto;
- registro da consultora: número de registro que a consultora tem no cadastro da FINEP;
- valor: quantia a que o documento se refere. Exemplos: valor do financiamento, no contrato; ou valor pago, em um recibo.

#### 4.2 Tratamento, Análise e Interpretação

Os dados da amostra foram tratados estatisticamente, buscando elementos que permitissem a identificação de quais os campos ideais que a definição do banco de dados deveria incluir, a fim de cobrir todas as facetas da documentação útil.

Os dados referentes à documentação solicitada poderiam certamente fornecer elementos para uma decisão, desde que fosse simulado um julgamento semelhante ao de usuários, para

definição da importância de cada campo. Este seria, porém, um critério subjetivo, provavelmente mais sujeito a erros que qualquer outro elaborado em bases reais, envolvendo diretamente o usuário.

Assim sendo, as perguntas da amostra foram consideradas como representativas dos hábitos e necessidades de informação dos usuários e tomadas como base para o desenvolvimento da metodologia pretendida.

Os passos seguidos para cumprimento dos objetivos foram:

- a) comparação entre o potencial informativo existente nos documentos e seu uso efetivo pelos usuários;
- b) análise das perguntas para verificação da existência de um núcleo de campos, comum a todos os Departamentos.

Como um subproduto da pesquisa, foram também identificados os conjuntos de campos relacionados com cada tipo de documento, numa tentativa de possível generalização. (Tabelas 16-19).

Duas metodologias básicas foram seguidas utilizando instrumentos estatísticos para tratamento dos dados obtidos com as perguntas dos usuários:

- a) estudo da ocorrência dos campos nas perguntas, sem distinção de Departamento ou tipo de documento;
- b) teste do  $\chi^2$  (qui-quadrado) para identificação de necessidades comuns a todos os Departamentos com referência aos campos de informação.

#### 4.2.1 Potencialidade versus Uso dos Campos

A observação das Tabelas 1-2 permite a análise comparativa da ocorrência dos campos de informação nos documentos e nas perguntas.

Verifica-se pela simples observação dessas tabelas que a ocorrência dos campos nos documentos mostra-se bem maior que a dos campos nas perguntas em relação a todos os itens da série.

Uma comparação global demonstra que:

- o somatório das ocorrências dos campos nos documentos é igual a 4069, enquanto que o total da frequência dos campos nas perguntas perfaz apenas 755;
- a frequência total dos campos nas perguntas equivale a apenas 18,55% do somatório de sua frequência na documentação;
- os campos de maior frequência nos documentos ocorrem em 100% da documentação analisada, isto é, em 224 documentos ( $f=224$ ,  $frp=1,000$  - Tabela 2), enquanto que nas perguntas a maior ocorrência corresponde ao campo que apareceu em 66,1% das questões, isto é, em 148 ( $f=148$ ,  $frp=0,661$  - Tabela 1).

Esta análise comparativa entre o potencial de informações existentes nos documentos e sua utilização pelos usuários leva a três conclusões parciais sobre a situação particular do Arquivo da FINEP:

- a) apenas uma parcela desse potencial é utilizada pelos usuários quando buscam um documento no Arquivo;
- b) poucos campos são muito utilizados, enquanto que muitos campos são sub-utilizados;
- c) a definição do banco de dados a partir da frequência dos campos nos documentos seria anti-econômica, uma vez que a baixa utilização de alguns campos não justifica o custo de sua manutenção no sistema. Quanto a este aspecto pode-se observar (Tabelas 1-2) a disparidade no uso de certos campos. Alguns, com baixa utili-

zação pelos usuários, apresentam alta frequência nos documentos. Fontes de recursos e classificação do projeto, por exemplo, foram usados uma única vez ( $fr=0,001$  e  $frp=0,004$ ), embora houvesse 207 possibilidades ( $fr=0,051$  e  $frp=0,924$ ). Isso equivale a dizer que esses campos, ainda que ocorrendo em 92,4% da documentação solicitada, foram explicitados em apenas 0,4% das solicitações. Fato semelhante ocorre com Estados do Brasil e setor FINEP. Já por outro lado, campos de baixa frequência nos documentos (Tabela 2) foram utilizados com frequência relativamente alta (Tabela 1), se considerada sua baixa disponibilidade. São exemplos: agente financeiro, veículo da informação, valor e referências.

#### 4.1.1 Identificação dos Componentes a partir das Questões dos Usuários

A partir dos dados originais levantados pela análise das perguntas, a ocorrência dos campos foi observada segundo a distribuição geral, por Departamentos solicitantes e por tipo de documento solicitado.

Nos diferentes casos, os campos foram ordenados decrescentemente segundo sua ocorrência nas perguntas e essa ocorrência estudada pelas frequências simples e relativa, bem como através do  $\chi^2$  (qui-quadrado). Desse estudo resultaram os elementos necessários para a avaliação dos dados.

#### 4.2.2.1 Distribuição Geral

Estando a série em ordem decrescente e analisando-se esta a partir da frequência simples ( $f$ , Tabela 1), pode-se verificar que os campos não foram igualmente preferidos pelos usuários: há uma concentração no uso de alguns campos e uma baixa procura de outros.

Em Estatística, por se trabalhar com grande número de dados, muitos valores se repetem, forçando uma ponderação. As frequências das observações funcionam como peso.

No presente caso, a frequência relativa ( $fr$ , Tabela 1) é capaz de expressar o peso ou valor relativo do campo no sistema, sendo esse valor determinado pela preferência do usuário.

Quando os campos escolhidos para o sistema forem um sub-conjunto ( $a$ ) dos campos utilizados ( $A$ ) pelos usuários e considerando-se que o somatório das frequências relativas de todos os campos usados equivale à unidade, o desempenho do sistema poderá ser previsto segundo esse critério básico: quanto mais o somatório dos pesos dos campos incluídos ( $a$ ) no sistema se aproximar de 1 ( $\sum frA$ ) melhor será seu desempenho, ocorrendo o inverso quando o somatório dos pesos tender a zero.

Isso leva a concluir que a escolha dos campos feita a partir da frequência deverá recair sobre os elementos mais frequentes na utilização, isto é, aqueles que obtiveram maior peso ( $fr$ ) segundo a preferência dos usuários.

Assim sendo, analisando-se a série a partir da frequência relativa ( $fr$ , Tabela 1) constata-se que 50% da soma dos valores ou pesos ( $fr$ ) abrangem apenas os 6 primeiros campos da série, e os 75% do total dos valores já correspondem a 16 campos. Os 25% restantes do total dos valores incorporam os 13 campos de valores ( $fr$ ) mais baixos da série.

Os 6 campos correspondentes aos 50% da soma dos pesos têm conceito muito extenso. São eles tipo de documento, código do projeto, assunto, mutuário, instituição de origem e data do documento. Tais campos devem ser incluídos; do contrário, sua extensão não limitada prejudicaria o desempenho do sistema. Porém, sua inclusão, com o objetivo de se obter recuperação precisa, implica em especificidade na indexação dos termos que cada campo abrigará.

Os 10 campos seguintes na série (do sétimo ao décimo sexto item) correspondem a mais 25% do total dos pesos. São eles: número do documento, nome da consultora, agente financeiro, instituição de destino, variações do nome do projeto, pessoa destinatária, número do protocolo, veículo da informação, número de registro da consultora, e nome do projeto. Sua posição mediana na série parece indicar que estes campos merecem uma preferência equilibrada por parte dos usuários e têm por conseguinte, seu lugar assegurado no sistema.

Os restantes 25% do total dos pesos correspondem àqueles 13 últimos campos da série. São eles: data do protocolo, valor, referências, signatário, anexos, cargo do signatário, cargo do destinatário, executor do projeto, setor FINEP, agência do AF, Estados do Brasil, classificação do projeto, e fontes de recursos. Tais campos, pouco frequentes nas perguntas, podem ser considerados de conteúdo informativo menor ou efêmero e, portanto, de importância secundária para o sistema.

Como já foi anteriormente exemplificado, a baixa frequência de uso de certos campos corresponde a alta frequência dos mesmos nos documentos. Considerando-se que quanto mais frequentes, maior é o custo de manutenção de um campo no sistema, é desejável que só sejam admitidos campos com alta frequência na documentação quando forem também frequentes na utilização.

A inclusão daqueles campos de baixa frequência - os últimos da série, cuja soma dos pesos perfaz 25% dos valores de todos os campos - deverá, então, ser considerada pelo administrador do banco de dados, do ponto de vista de custo/benefício: compensa para a Empresa investir em informações pouco utilizadas? Ou será admissível uma recuperação menos precisa em favor de mais baixo custo do sistema?

#### 4.2.2.2 - Distribuição por Departamentos

A Tabela 3 dá uma visão completa da utilização de cada campo pelos diversos Departamentos, bem como do volume de solicitações de cada um. As Tabelas 6-14 mostram as preferências de cada Departamento, apresentando os campos por ordem de importância.

A ausência de pedidos por parte do PNTE, do G. GAS do N. BID e do GE, bem como a baixa frequência do GP, já foram esclarecidas anteriormente. Não obstante essa justificativa, a frequência de utilização parece bastante dispersa. Essa dispersão é demonstrada pela baixa frequência dos campos em geral (Tabela 3), e pela ausência de alguns campos (frequência zero) nas Tabelas 6-14. Esse fato parece estar relacionado à diversificação de atividades e necessidades dos Departamentos. Ainda como apoio a essa suposição, pode-se verificar a semelhança existente

entre os dados referentes ao GEP e ao GCT, departamentos com atuação muito semelhantes, ambos trabalhando diretamente com análise técnica e avaliação de projetos.

Se for admitido que os Departamentos exercem atividades diversificadas, a metodologia mais indicada para identificação do núcleo de campos, comum a todos os Departamentos, parece ser o teste de significância do  $\chi^2$ , medida estatística capaz de estabelecer os limites de discrepância entre diferentes classes, a partir da comparação entre a frequência probabilisticamente esperada e aquela realmente obtida. Essa medida pode ser expressa através da seguinte fórmula:

$$\chi^2 = \sum \frac{(f_e - f_t)^2}{f_t}$$

sendo  $f_e$  a frequência empírica ou obtida e  $f_t$  a frequência teórica ou esperada.

Para efeitos da presente metodologia, para 12 G.L. foi adotado o nível de significância 0,01, rejeitando-se por conseguinte, os valores acima de 26,2 conforme distribuição do  $\chi^2_p(10)$ , uma vez que se buscavam os campos de uso generalizado e não aqueles estreitamente relacionados a Departamentos específicos.

Os resultados (Tabelas 4-5) apresentaram 23 campos aceitos e 6 campos rejeitados por não serem de importância no consenso geral e sim de interesse particular de algum Departamento.

Observando a tabela 3, verifica-se que 2 daqueles campos aceitos apresentaram-se homogêneos, porém em torno de zero, justificando sua exclusão. São eles: classificação do projeto e Estados do Brasil.

Os campos aceitos são: nome do projeto, referência, agência do AF, pessoa destinatária, agente financeiro, veículo da informação, anexos, variações do nome do projeto, número

ro do documento, mutuário, executor do projeto, instituição de origem, setor FINEP, nome da consultora, código do projeto, signatário, instituição de destino, data do protocolo, valor, tipo de documento e assunto.

Tais campos representam o núcleo comum a todos os Departamentos e constituem o conjunto mínimo capaz de satisfazer as necessidades básicas dos usuários. Os restantes serão incorporados ao banco de dados conforme as possibilidades da Empresa, na medida em que a satisfação das necessidades específicas de cada Departamento se constituir em uma meta.

#### 4.2.2.3 - Distribuição por Tipo de Documento

Relacionando-se a ocorrência dos campos nas perguntas com os respectivos tipos de documentos solicitados, podem ser estabelecidas certas relações que seriam, provavelmente constantes para documentos com idênticas características.

Com esse propósito, os dados referentes às perguntas foram trabalhados e identificados 17 tipos diferentes de documentos, aos quais se relacionou a frequência de cada um dos campos (Tabela 15). As ocorrências foram estudadas, constatando-se que:

- a frequência apresenta-se muito dispersa;
- apenas um pequeno número de campos está relacionado com cada tipo de documento.

Levando-se em conta estas considerações e o aspecto prático de caracterização de um acervo em caso de setorização de arquivos de projetos, segundo as características da documentação, os 17 tipos de documentos foram agrupados em quatro grandes classes, a saber:

- a) carta, circular, e ofício foram considerados correspondência;
- b) balancete, fatura, guia, JE, recibo e "voucher" foram agrupados como documentação financeiro-contábil;

- c) cadastro de consultoras, projeto e relatório foram reunidos como documentação técnica;
- d) ato constitutivo, contrato, decisão, parecer e procuração foram denominados documentação jurídica.

A ocorrência dos campos nas quatro classes de documentos pode ser observada nas tabelas 16-19, as quais permitem identificar os campos necessários à caracterização de cada classe de documentos, a saber:

- Correspondência

tipo de documento, instituição de origem, número de documento, assunto, data do documento, código do projeto, número do protocolo, instituição de destino, mutuário, data do protocolo, signatário, valor, cargo do signatário, anexos, executor do projeto, nome do projeto, variações do nome do projeto, fontes de recursos, agente financeiro, cargo do destinatário e nome da consultora.

- Documentação Financeiro-Contábil:

tipo de documento, código do projeto, instituição de destino, referências, instituição de origem, data do documento, valor, assunto, mutuário, número do documento, número do protocolo e Setor FINEP.

- Documentação técnica:

código do projeto, mutuário, tipo de documento, variações do nome do projeto, nome da consultora, assunto, número de registro da consultora, agente financeiro, anexos, nome do projeto, número do protocolo, data do protocolo, data do documento, executor do projeto, agência do AF, classificação do projeto, veículo da informação, Estados do Brasil e Setor FINEP.

- Documentação Jurídica:

tipo de documento, código do projeto, assunto, mutuário, instituição de origem, agente financeiro, data do documento, número do documento, pessoa destinatária, nome da consultora, veículo da informação, cargo do destinatário, instituição de destino, nome do projeto, signatário, agência do AF, cargo do signatário, executor do projeto, data do protocolo, referências, variações do nome do projeto, valor, setor FINEP e Estados do Brasil.

Feitas as devidas adaptações - eliminação dos campos diretamente relacionados a projetos e inclusão daqueles de interesse local - as relações acima estabelecidas podem servir de base para desenvolvimento de sistemas em arquivos diversos, cuja documentação seja de um dos quatro tipos mencionados.

## 5. CONCLUSÕES

Dois tipos de conclusões poderão ser extraídos do presente trabalho: um, particular, aplicável à FINEP, e outro geral, passível de ser utilizado para definição de qualquer sistema de banco de dados destinado a arquivos de administração de projetos. É mesmo provável que essas conclusões possam ser aplicáveis a quaisquer arquivos.

Com relação à FINEP, se se considerar que há atividades comuns como há também as específicas de cada setor, parece que a combinação das duas metodologias - teste do  $\chi^2$  corrigido pela ordenação simples de frequência - é o mais recomendado, para maior segurança.

Assim sendo, seriam considerados para inclusão no banco de dados os campos aceitos pelo  $\chi^2$  e aqueles rejeitados que estivessem situados dentre os 16 campos (tabela 1) de maior frequência na série ( $\Sigma fr = 0,752$ ). Desses 24 componentes aceitos, 5 poderiam ser ainda descartados pelas razões que se seguem:

- Setor FINEP: porque a Empresa é ainda relativa-mente pequena e todos os funcionários conhecem os diferentes programas da instituição, sendo capazes de com eles relacionar os projetos;
- nome do projeto: existe uma parte do banco de dados destinada ao cadastro, no qual todas as características dos projetos são registradas. Portanto, bastaria relacionar o arquivo DOCUMENTOS com o CADASTRO, para que as informações gerais se tornassem disponíveis.
- variações do nome do projeto: ficaria melhor no CADASTRO, onde este campo deveria ser introduzido;
- mutuário: Já existe no CADASTRO.

Os 19 campos restantes parecem satisfazer às necessidades de todos os Departamentos: tipo de documento, código do projeto, assunto, instituição de origem, data do documento, número do documento, agente financeiro, instituição de destino, número do protocolo, veículo da informação, número de registro da consultora, pessoa destinatária, signatário, data do protocolo, valor, referências, anexos, executor do projeto e agência do AF.

A adoção do sistema integral, visando a atingir o desempenho ótimo, ou a adoção de apenas parte dos campos ficará a critério da Administração da Empresa, segundo sua política interna.

As conclusões de ordem geral são as seguintes:

- a) a escolha dos componentes do banco de dados deve ser determinada em função de seu valor ou peso, (fr) e um campo só importa ao sistema porque importa ao usuário;
- b) só é economicamente justificável a definição dos campos a partir de sua ocorrência na documentação, quando esta frequência coincide com a de uso. Isto porque, sendo a manutenção de campos no banco de dados tanto mais dispendiosa quanto mais frequente é o campo nos documentos, a baixa utilização de um componente deveria ser o critério para sua exclusão do sistema;
- c) duas metodologias, baseadas nas necessidades expressas dos usuários, podem ser utilizadas para identificação dos campos fundamentais para o sistema atingir seu desempenho ótimo na recuperação de informações, dependendo das características dos usuários:
  - ordenação das frequências dos campos e eleição, a partir dos mais frequentes, quando o grupo de usuários é homogêneo;

- teste do  $\chi^2$ , quando as atividades e necessidades dos grupos diferem entre si.

d) o número de campos a serem eleitos após a apuração dos mais importantes dependerá da política da Empresa quanto à canalização de recursos para o sistema.

Em uma situação específica, qualquer das duas metodologias sendo adotada, seria recomendável que periodicamente fossem reavaliados os interesses dos usuários do sistema, a fim de corrigir o modelo proposto a partir da pesquisa inicial.

6. TABELAS

TABELA 1: CAMPOS IDENTIFICADOS NAS PERGUNTAS DOS USUÁRIOS  
DO ARQUIVO (FINEP, RIO DE JANEIRO, JULHO DE 1974)

Nome dos Campos	f	fr	frp
Tipo de documento .....	148	0,196	0,661
Código do projeto .....	119	0,158	0,532
Assunto .....	72	0,095	0,322
Mutuário .....	70	0,093	0,313
Instituição de origem .....	54	0,072	0,242
Data do documento .....	41	0,054	0,184
Número do documento .....	38	0,050	0,170
Nome da consultora .....	27	0,036	0,121
Agente financeiro .....	25	0,033	0,112
Instituição de destino .....	21	0,028	0,094
Variações do nome do projeto .....	16	0,021	0,071
Pessoa destinatária .....	14	0,019	0,063
Número do protocolo .....	14	0,019	0,063
Veículo da informação .....	11	0,015	0,049
Nº registro da Consultora .....	11	0,015	0,049
Nome do projeto .....	10	0,013	0,045
Data do protocolo .....	9	0,012	0,040
Valor .....	9	0,012	0,040
Referências .....	8	0,011	0,036
Signatário .....	7	0,009	0,031
Anexos .....	7	0,009	0,031
Cargo do signatário .....	5	0,007	0,022
Cargo do destinatário .....	5	0,007	0,022
Executor do projeto .....	4	0,005	0,018
Setor FINEP .....	3	0,004	0,013
Agência do AF .....	3	0,004	0,013
Estados do Brasil .....	2	0,003	0,009
Classificação do projeto .....	1	0,001	0,004
Fontes de recursos .....	1	0,001	0,004
<b>T O T A L</b>	<b>755</b>	<b>1,000</b>	<b>*</b>

Fonte: Pesquisa da autora no Arquivo da FINEP.

\* O somatório dessa coluna não é significativo.

fr : freqüência relativa ao somatório de f (755)

frp : freqüência relativa ao total de perguntas feitas ao Arquivo (224)

TABELA 2: CAMPOS IDENTIFICADOS NOS DOCUMENTOS DO ARQUIVO  
(FINEP, RIO DE JANEIRO, JULHO DE 1974)

Nome do Campo	f	fr	frp
Tipo de documento.....	224	0,055	1,000
Assunto.....	224	0,055	1,000
Instituição de origem.....	220	0,054	0,982
Setor FINEP.....	213	0,052	0,951
Nome do projeto.....	207	0,051	0,924
Classificação do projeto.....	207	0,051	0,924
Fonte de recursos.....	207	0,051	0,924
Data do documento.....	202	0,050	0,902
Signatário.....	202	0,050	0,902
Mutuário.....	202	0,050	0,902
Estado do Brasil.....	202	0,050	0,902
Código do projeto.....	200	0,049	0,893
Instituição de destino.....	184	0,045	0,821
Cargo do signatário.....	153	0,038	0,683
Pessoa destinatária.....	135	0,033	0,603
Número do documento.....	121	0,030	0,540
Variações do nome do projeto.....	106	0,026	0,473
Número do protocolo.....	97	0,024	0,433
Data do protocolo.....	97	0,024	0,433
Nome da consultora.....	97	0,024	0,433
Nº registro da consultora.....	97	0,024	0,433
Cargo do destinatário.....	90	0,022	0,402
Referências.....	72	0,017	0,321
Agente financeiro.....	70	0,017	0,313
Anexos.....	59	0,014	0,263
Valor.....	52	0,013	0,232
Executor.....	50	0,012	0,223
Agência do AF.....	45	0,011	0,201
Veículo da informação.....	34	0,008	0,152
<b>T O T A L</b>	<b>4069</b>	<b>1,000</b>	<b>- *</b>

Fonte: Pesquisa da autora no Arquivo da FINEP

\*O somatório dessa coluna não é significativo

fr : frequência relativa ao somatório de f (4069)

frp : frequência relativa ao total dos documentos fornecidos (224)



TABELA 4: CAMPOS ACEITOS SEGUNDO O TESTE DE SIGNIFICÂNCIA \*  
(RIO DE JANEIRO, JULHO DE 1974)

Nome dos Campos	$\chi^2$
Estados do Brasil.....	2,354
Classificação do projeto.....	2,809
Nome do projeto.....	3,239
Referências.....	3,642
Agência do AF.....	5,001
Pessoa destinatária.....	5,149
Agente financeiro.....	7,288
Veículo da informação.....	7,701
Anexos.....	8,306
Variações do nome do projeto.....	9,984
Número do documento.....	11,043
Mutuário.....	11,084
Executor do projeto.....	11,314
Instituição de destino.....	11,729
Setor FINEP.....	12,569
Nome da consultora.....	12,813
Código do projeto.....	17,483
Signatário.....	19,007
Instituição de origem.....	20,456
Data do protocolo.....	20,829
Valor.....	22,304
Tipo de documento.....	22,414
Assunto.....	25,262

\* 12 G.L. e nível de significância 0,01

Fonte: Pesquisa da autora no Arquivo da FINEP

TABELA 5: CAMPOS REJEITADOS SEGUNDO O TESTE DE SIGNIFICÂNCIA\*  
(RIO DE JANEIRO, JULHO DE 1974)

Nome dos Campos	$\chi^2$
Número do protocolo.....	27,606
Data do documento.....	30,478
Fonte de recursos.....	31,139
Número de registro da consultora.....	32,351
Cargo do signatário.....	50,937
Cargo do destinatário.....	52,538

\* 12 G.L. e nível de significância 0,001

Fonte: Pesquisa da autora no Arquivo da FINEP

TABELA 6: USO DOS CAMPOS PELA PRES  
(RIO DE JANEIRO, JULHO DE 1974)

Campos usados	Nº de Ocorrências
Tipo de documento.....	7
Assunto.....	5
Instituição de origem.....	3
Mutuário.....	3
Nome da consultora.....	3
Número do documento.....	2
Data do documento.....	2
Valor.....	2
Veículo da informação.....	1
Signatário.....	1
Cargo do signatário.....	1
Instituição de destino.....	1
Pessoa destinatária.....	1
Executor.....	1
Setor FINEP.....	1
Fonte de recursos.....	1
<b>T O T A L</b>	<b>35</b>

Fonte: Pesquisa da autora no Arquivo da FINEP

TABELA 7: USO DOS CAMPOS PELA VICE-PRES  
(RIO DE JANEIRO, JULHO DE 1974)

Campos usados	Nº de Ocorrências
Tipo de documento.....	7
Assunto.....	5
Código do projeto.....	4
Instituição de origem.....	4
Número do documento.....	2
Nome da consultora.....	2
Veículo da informação.....	1
Data do documento.....	1
Data do protocolo.....	1
Número do protocolo.....	1
Pessoa destinatária.....	1
Anexos.....	1
Instituição de destino.....	1
Mutuário.....	1
Agente financeiro.....	1
<b>T O T A L</b>	<b>33</b>

Fonte: Pesquisa da autora no Arquivo da FINEP

TABELA 8 : USO DOS CAMPOS PELO DEJ  
(RIO DE JANEIRO, JULHO DE 1974)

Campos usados	Nº de Ocorrências
Tipo de documento.....	30
Código do projeto.....	26
Mutuário.....	21
Assunto.....	16
Instituição de origem.....	10
Agente financeiro.....	8
Instituição de destino.....	5
Data do documento.....	4
Nome da consultora.....	4
Nome do projeto.....	4
Variações do nome do projeto.....	4
Pessoa destinatária.....	3
Número do documento.....	3
Cargo do destinatário.....	3
Veículo da informação.....	1
Número do protocolo.....	1
Cargo do signatário.....	1
Referências.....	1
Executor.....	1
Setor FINEP.....	1
Estados do Brasil.....	1
<b>T O T A L</b>	<b>148</b>

TABELA 9 : USO DOS CAMPOS PELO DAD  
(RIO DE JANEIRO, JULHO DE 1974)

Campos usados	Nº de Ocorrências
Tipo de documento.....	21
Data do documento.....	13
Assunto.....	12
Código do projeto.....	10
Instituição de origem.....	9
Número do documento.....	6
Mutuário.....	6
Instituição de destino.....	6
Veículo da informação.....	3
Signatário.....	3
Pessoa destinatária.....	3
Valor.....	3
Nome da consultora.....	3
Anexos.....	2
Cargo do signatário.....	1
Referências.....	1
Nome do projeto.....	1
Setor FINEP.....	1
Agente financeiro.....	1
<b>T O T A L</b>	<b>105</b>

Fonte: Pesquisa da autora no Arquivo da FINEP

TABELA 10: USO DOS CAMPOS PELO DFC  
(RIO DE JANEIRO, JULHO DE 1974)

Campos usados	Nº de Ocorrências
Tipo de documento.....	19
Código do projeto.....	14
Mutuário.....	6
Número do documento.....	6
Agente financeiro.....	4
Instituição de origem.....	4
Assunto.....	3
Variações do nome do projeto.....	3
Data do documento.....	2
Referências.....	2
número do protocolo.....	1
Data do protocolo.....	1
Anexos.....	1
Nome do projeto.....	1
Nome da consultora.....	1
Valor.....	1
Veículo da informação.....	1
Agência do AF.....	1
<b>T O T A L</b>	<b>71</b>

Fonte: Pesquisa da autora no Arquivo da FINEP

TABELA 11: USO DOS CAMPOS PELO CPD  
(RIO DE JANEIRO, JULHO DE 1974)

Campos usados	Nº de Ocorrências
Tipo de documento.....	10
Código do projeto.....	8
Instituição de origem.....	7
Assunto.....	4
Número do protocolo.....	4
Data do documento.....	3
Data do protocolo.....	3
Signatário.....	2
Mutuário.....	2
Número do documento.....	1
Instituição de destino.....	1
Pessoa destinatária.....	1
Cargo do destinatário.....	1
Nome do projeto.....	1
Nome da consultora.....	1
Referências.....	1
<b>T O T A L</b>	<b>59</b>

Fonte: Pesquisa da autora no Arquivo da FINEP

TABELA 12: USO DOS CAMPOS PELO GCT  
(RIO DE JANEIRO, JULHO DE 1974)

Campos usados	Nº de Ocorrências
Código do projeto.....	28
Tipo de documento.....	25
Assunto.....	14
Número do documento.....	10
Data do documento.....	10
Mutuário.....	10
Instituição de origem.....	9
Número do protocolo.....	7
Data do protocolo.....	4
Instituição de destino.....	4
Valor.....	3
Agente financeiro.....	3
Nome da consultora.....	3
Anexos.....	2
Pessoa destinatária.....	2
Referências.....	2
Executor do projeto.....	2
Veículo.....	1
Signatário.....	1
Nome do projeto.....	1
Variações do nome do projeto.....	1
<b>T O T A L</b>	<b>162</b>

Fonte: Pesquisa da autora no Arquivo da FINEP

TABELA 13: USO DOS CAMPOS PELO GEP  
(RIO DE JANEIRO, JULHO DE 1974)

Campos usados	Nº de Ocorrências
Código do projeto.....	29
Tipo do documento.....	28
Mutuário.....	21
Assunto.....	12
Nº registro da consultora.....	11
Nome da consultora.....	10
Agente financeiro.....	8
Variações do nome do projeto.....	8
Número do documento.....	8
Instituição de origem.....	7
Data do documento.....	6
Instituição de destino.....	3
Pessoa destinatária.....	3
Veículo da informação.....	3
Referências.....	2
Nome do projeto.....	2
Agência do AF.....	2
Cargo do signatário.....	1
Classificação do projeto.....	1
Anexos.....	1
Estados do Brasil.....	1
<b>T O T A L</b>	<b>167</b>

Fonte: Pesquisa da autora no Arquivo da FINEP

TABELA 14: USO DOS CAMPOS PELO GP  
(RIO DE JANEIRO, JULHO DE 1974)

Campos usados	Nº de Ocorrências
Assunto.....	1
Cargo do destinatário.....	1
Cargo do signatário.....	1
Instituição de origem.....	1
Tipo de documento.....	1
<b>T O T A L</b>	<b>5</b>

Fonte: Pesquisa da autora no Arquivo da FINEP



TABELA 16

USO DOS CAMPOS NAS PERGUNTAS SEGUNDO AS CLASSES DE  
DOCUMENTOS: CORRESPONDÊNCIA  
(RIO DE JANEIRO, JULHO DE 1974)

Nome dos Campos	Nº de Ocorrências
Tipo de documento .....	32
Instituição de origem.....	23
Número do documento.....	21
Assunto.....	21
Data do documento.....	18
Código do projeto.....	16
Número do protocolo.....	10
Instituição de destino.....	9
Mutuário.....	8
Data do protocolo.....	6
Signatário.....	4
Valor.....	3
Cargo do Signatário.....	3
Anexos.....	3
Executor do projeto.....	2
Nome do projeto.....	2
Variações do nome do projeto.....	1
Fontes de recursos.....	1
Agente financeiro.....	1
Cargo do destinatário.....	1
Nome da consultora.....	1
<b>T O T A L</b>	<b>186</b>

Fonte: Pesquisa da autora no Arquivo da FINEP

TABELA 17

USO DOS CAMPOS NAS PERGUNTAS SEGUNDO AS CLASSES DE  
DOCUMENTOS: DOCUMENTAÇÃO FINANCEIRO- CONTÁBIL  
(RIO DE JANEIRO, JULHO DE 1974)

Nome dos Campos	Nº de Ocorrências
Tipo de documento.....	16
Código do projeto.....	12
Instituição de destino.....	9
Referências.....	7
Instituição de origem.....	7
Data do documento.....	6
Valor.....	5
Assunto.....	4
Mutuário.....	1
Número do documento.....	1
Número do protocolo.....	1
Setor FINEP.....	1
<b>T O T A L</b>	<b>70</b>

Fonte: Pesquisa da autora no Arquivo da FINEP

TABELA 18

USO DOS CAMPOS NAS PERGUNTAS SEGUNDO AS CLASSES DE  
DOCUMENTOS: DOCUMENTAÇÃO TÉCNICA  
(RIO DE JANEIRO, JULHO DE 1974)

Nome dos Campos	Nº de Ocorrências
Código do projeto.....	40
Mutuário.....	35
Tipo de documento.....	31
Variações do nome do projeto.....	14
Nome da consultora.....	13
Assunto.....	12
Número de reg. da consultora.....	11
Agente financeiro.....	8
Anexos.....	4
Nome do projeto.....	4
Número do protocolo.....	3
Data do protocolo.....	2
Data do documento.....	2
Executor do projeto.....	1
Agência do AF.....	1
Classificação do projeto.....	1
Veículo da informação.....	1
Estados do Brasil.....	1
Setor FINEP.....	1
<b>T O T A L</b>	<b>185</b>

Fonte: Pesquisa da autora no Arquivo da FINEP

## TABELA 19

USO DOS CAMPOS NAS PERGUNTAS SEGUNDO AS CLASSES DE  
DOCUMENTOS: DOCUMENTAÇÃO JURÍDICA  
(RIO DE JANEIRO, JULHO DE 1974)

Nome dos Campos	Nº de Ocorrências
Tipo de documento.....	69
Código do projeto.....	51
Assunto.....	35
Mutuário.....	26
Instituição de origem.....	24
Agente financeiro.....	16
Data do documento.....	15
Número do documento.....	15
Pessoa destinatária.....	14
Nome da consultora.....	13
Veículo da informação.....	10
Cargo do destinatário.....	4
Instituição de destino.....	4
Nome do projeto.....	3
Signatário.....	3
Agência de AF.....	2
Cargo do signatário.....	2
Executor do projeto.....	1
Data do protocolo.....	1
Referências.....	1
Variações do nome do projeto.....	1
Valor.....	1
Setor FINEP.....	1
Estados do Brasil.....	1
<b>T O T A L</b>	<b>313</b>

Fonte: Pesquisa da autora no Arquivo da FINEP

## 7. REFERÊNCIAS

- (1) DAMMERS, H.F. Information management systems: some views on problems and potentialities. In: DATA ORGANIZATION FOR MAINTENANCE AND ACCESS CONFERENCE, Keele, April 1970. Papers. Keele, The University, 1970.
- (2) ORGANIZAÇÃO PARA A COOPERAÇÃO E DESENVOLVIMENTO ECONÔMICO. Information dans une société en évolution; quelques considérations de caractère politique. Paris, 1971. p.30.
- (3) BRANCHE, Henri Boulrier. Relatório sobre o Arquivo Nacional do Brasil. Rio de Janeiro, Arquivo Nacional, 1960.
- (4) ESPDSEL, José Pedro Pinto. Editorial. Arquivos & Administração, 2 (2):5, Ago.1974.
- (5) SOUSA, Flávio Pereira. Introdução à recuperação da informação. /A ser publicado ainda em 1974 pelo Convênio MEC/PUC /.
- (6) SALTON, Gerard. Automatic information organization and retrieval. New York, McGraw-Hill, 1968. 514p.
- (7) CLEVERDON, Cyril. Information and its retrieval. Aslib Proceedings, 22 (11): 546, Nov. 1970.
- (8) LANCASTER, F. Wilfrid. MEDLARS: report on the evaluation of its operating efficiency. American Documentation, 20(2): 119-42, Apr.1969.
- (9) HALO. Statistical tables and formulas. New York, Wiley, 1952. p.96.
- (10) SPIEGEL, M. R. Estatística. Tradução de Pedro Cosentino São Paulo, McGraw-Hill do Brasil, 1974, 580p

## 8. BIBLIOGRAFIA CONSULTADA

- CUNHA, S. E. Estatística Descritiva (na Psicologia e Educação)  
Rio de Janeiro, Forense /s.d./243p.
- ENGELS, R.W. A tutorial on data-base organization; TR  
00.2004. In: IBM. Data base concepts; education guide.  
New York, 1972.
- FARRADANE, J. The evaluation of information retrieval  
systems. Journal of Documentation, 30 (2): 195-209,  
June 1974.
- GELLER, S.B. Archival data storage. Datamation, 20 (10):  
72-80, Oct. 1974.
- KEMP, D.A. Relevance, pertinence and information system  
development. Information Storage and Retrieval,  
10(2):37-47, Feb. 1974.
- KING, D.W. & BRYANT, E.C. The evaluation of information  
services and products. Washington, Information Resources  
Press, 1971. 306p.
- KÖNIGOVÁ, M. Mathematical and statistical methods of noise  
evaluation in a retrieval system. Information Storage  
and Retrieval, (6): 437-44, May 1971.
- LANCASTER, F.W. Information retrieval systems;  
characteristics, testing, and evaluation. New York,  
J.Wiley, 1968. 222p.
- \_\_\_\_\_ & FAYEN, E.G. Information retrieval on-line.  
Los Angeles, Melville Publishing, 1973. 597p.
- LANDAU, H. the proliferation of machine-readable data bases:  
current problems. Drexel Library Quarterly, 8(1):  
63-9, Jan. 1972.
- MARTYN, J. & VICKERY, B.C. The complexity of modelling  
of information systems. Journal of Documentation,  
26(3): 204-20, Sept. 1970.

- 52.
- NICK, E. & KELLNER, S.R.O. Fundamentos de estatística para as ciências do comportamento. Rio de Janeiro, Renes, 1971. 312p.
  - RIEGER, M. Le rôle des archives dans l'administration. Bulletin de l'Unesco pour les Bibliothèques, 27(1):43-5, Jan./Fev. 1973.
  - SAFFADY, W. A university archives and records management program: some operational guidelines. College & Research Libraries, 35 (3):204-10, May. 1974
  - SALTON, G. Evaluation problems in interactive information retrieval. Information Storage and Retrieval, 6(1):29-44, May 1970.
  - \_\_\_\_\_ & YANG, C.S. On the specification of term values in automatic indexing. Journal of Documentation, 29(4): 351-72, Dec. 1973.
  - SEELY, B.J. Indexing depth and retrieval effectiveness. Drexel Library Quarterly, 8(2):201-8, Apr. 1972.
  - TAKAHAMA, T. A model for a document retrieval system. Information Storage and Retrieval, 9(3):143-63, Mar.1973.
  - VICKERY, B.C. Information systems. London, Butterworths, 1973. 350p.
  - WILSON, P. Situational relevance. Information Storage and Retrieval, 9(8):457-71, Aug. 1973

## SYNDOPSIS

Based on data collected at FINEP's (Financiadora de Estudos e Projetos) Archives and having the aim of building up a data base on typical documents related to project administration, two alternative methodologies were designed, using statistical measures, to define which fields of information should be used at the system. The first methodology is based on the frequency order of the fields, according to their frequency at users' question, and should be useful when the users have common interests and activities. The second methodology - the  $\chi^2$  test - would be suitable when users have different interests and activities.