

SALF: UM ALGORÍTMO PARA INDEXAÇÃO AUTOMÁTICA UTILIZANDO
VOCABULÁRIO CONTROLADO

LUIS FERNANDO SAYÃO

Comissão Nacional de Energia Nuclear/
Centro de Informações Nucleares -CNEN/CIN

Dissertação apresentada ao curso de Mestrado em Comunicação: Ciência da Informação, da Escola de Comunicação da Universidade Federal do Rio de Janeiro e do Instituto Brasileiro de Informação em Ciência e Tecnologia - IBICT para obtenção do Grau de Mestre em Ciência da Informação.

ORIENTADORA: Prof.^a GILDA MARIA BRAGA, PhD - Instituto Brasileiro de Informação em Ciência e Tecnologia.

RIO DE JANEIRO

1985

AGRADECIMENTOS

À Gilda Braga, pela orientação.

À CNEN/CIN e ao IBICT, pelo apoio concedido.

A todos os meus amigos, pela força que deram.

SUMÁRIO

	Pág.
LISTA DE ILUSTRAÇÕES	iv
RESUMO	vi
1 <u>INTRODUÇÃO</u>	1
1.1 Indexação manual versus indexação automática	3
1.2 Problemas da indexação manual	5
1.3 Problemas da indexação automática	6
1.4 Justificativas	8
2 <u>OBJETIVO</u>	12
2.1 Objetivo principal	12
2.2 Objetivos secundários	12
3 <u>REVISÃO DA LITERATURA</u>	13
4 <u>AMBIENTE DE ESTUDO</u>	45
4.1 Centro de Informações Nucleares - CIN	45
4.1.1 Origens do CIN	45
4.1.2 Serviços e sistemas do CIN	48
4.1.2.1 Processamento Automático da Informação-PAI	48
4.1.2.2 Disseminação Seletiva de Informação-SONAR	48
4.1.2.3 Sistema de Busca Retrospectiva-SUPRIR	49
4.1.2.4 Acesso ao Documento Primário-SERVIR	50
5 <u>MATERIAL</u>	51
5.1 FONTE: Base de Dados Bibliográficos em Fontes de Energia	51
5.1.2 Principais características	51

5.2	A escolha da FONTE	52
5.3	A escolha da Área N13-Combustíveis a partir de hidrocarbonetos	53
5.4	A escolha da Amostra	53
6	<u>METODOLOGIA</u>	57
6.1	Descrição do método	57
6.1.1	Preparação do texto-base	57
6.1.2	Formação de expressões	61
6.2	Fluxograma do método SALF	68
6.3	Exemplo de indexação pelo método SALF	71
6.4	Testes de avaliação	75
7	<u>RESULTADOS</u>	79
7.1	Indexação da amostra	81
7.2	Conteúdo informativo do título e resumo	102
7.3	Exaustividade	102
7.4	Teste de recuperação	110
7.4.1	Análise das diferenças	110
7.4.1.1	Problemas de avaliação	112
7.4.1.2	Especificidade	113
7.5	Formação de expressões	115
7.6	Teste de consistência	116
8	<u>CONCLUSÃO</u>	129
9	<u>BIBLIOGRAFIA</u>	134
10	<u>ANEXOS</u>	146
	<u>ANEXO 1</u> - Folha de entrada da FONTE	147
	<u>ANEXO 2</u> - Tesouro DOE	153
	<u>ANEXO 3</u> - Tabela de classificação DOE	154

<u>ANEXO</u> 4 - Dicionário histórico da área N13 da FONTE	155
<u>ANEXO</u> 5 - Dicionário negativo (INIS/STAIRS stopword list)	159
<u>ANEXO</u> 6 - Formulação das consultas para o teste de recuperação	163
<u>ANEXO</u> 7 - Estatística dos descritores	166

LISTA DE ILUSTRAÇÕES

- FIGURA 1 - Reconhecimento do hífen
- FIGURA 2 - Estrutura do tesouro de trabalho
- FLUXOGRAMA 1 - Fluxograma do método SALF
- QUADRO 1 - Resumo dos resultados obtidos por tipo de indexação
- QUADRO 2 - Resumo dos resultados obtidos por documento
- GRÁFICO 1 - Indexação SALF-S - porcentagem de descritores gerados a partir do resumo, do título ou de ambos simultaneamente
- GRÁFICO 2 - Descritores Sugeridos Aceitos (DSA) - porcentagem dos DSA gerados a partir do resumo, do título ou de ambos simultaneamente
- GRÁFICO 3 - Indexação SALF-P - porcentagem de descritores gerados a partir do resumo, do título ou de ambos simultaneamente
- GRÁFICO 4 - Número médio de descritores assinalados (exaustividade média) por tipo de indexação
- GRÁFICO 5 - Número médio de descritores distintos por tipo de indexação

- GRÁFICO 6 - Precisão por revocação
- GRÁFICO 7 - Precisão e revocação por tipo de indexação
- GRÁFICO 8 - Número de palavras por descritor por tipo de indexação
- GRÁFICO 9 - Concordância entre as indexações
- GRÁFICO 10 - Concordância entre os termos de maiores frequências
- GRÁFICO 11 - Distorções nos descritores de maiores frequências
- GRÁFICO 12 - Distribuição de frequência dos descritores por tipo de indexação

RESUMO

Propõe-se um método de indexação automática para indexar títulos e resumos de documentos escritos, baseado em tesauro, capaz de assinalar descritores controlados formados preferencialmente por mais de uma palavra. O método proposto incorpora também dispositivo capaz de sugerir descritores ao indexador, com intuito de enriquecer tanto a indexação quanto o vocabulário da área. Testa-se o método, indexando-se uma amostra da base de dados em energia FONTE/CIN. Avaliam-se, em comparação com a indexação manual, os seguintes parâmetros: revocação, precisão, exhaustividade, especificidade, consistência e concordância dos descritores de altas frequências. Os resultados obtidos demonstram que o método proposto é viável.

ABSTRACT

A thesaurus based method of automatic indexing able to assign controlled descriptors formed preferentially by phrases, is developed. The method is designed to index titles and abstracts of written documents. It also includes device to suggest descriptors to the indexer in order to enrich the indexing and the specific vocabulary of the area. With the aim to test the method, a sample from the FONTE/CIN data base in energy, is indexed using the method. A comparative evaluation in relation to the manual indexing is carried out; the parameters investigated are: recall, precision, exhaustivity, specificity, coincidence of descriptors and concordance of the high frequency descriptors. The results confirm the feasibility of the proposed method.

"A indexação está se tornando uma arte inteiramente nova"

Keppel -1937

1 INTRODUÇÃO

A idéia de se ter a máquina trabalhando para o homem é ao mesmo tempo fascinante e controvertida e tem provocado muitas pesquisas e discussões.

Durante as três últimas décadas, uma quantidade muito grande de recursos humanos e materiais tem sido investido em pesquisas tanto nos campos de Ciência da Informação quanto de Informática com o propósito de desenvolver métodos computarizados capazes de analisar textos escritos em linguagem natural. Durante todos esses anos as experiências em Ciência da Informação relacionadas à análise de textos orientaram-se prioritariamente no sentido de envolver o computador nos processos considerados os mais intelectuais da análise documentária - gerar termos de indexação capazes de descrever o conteúdo informativo de documentos escritos (indexação automática); atribuir classes a documentos escritos (classificação automática); e elaborar resumos a partir de textos completos (resumo automático).

O método mais simples de gerar por máquina elementos representativos do conteúdo de um texto é conhecido como índice KWIC (Key Word In Context). Desenvolvida nos fins dos anos 50, a técnica KWIC é usada principalmente para criação de índices baseados em títulos de itens bibliográficos.

Com o desenvolvimento dos computadores e das técnicas de processamento automático de textos, a indexação automática deixou de ser mera manipulação de títulos; os novos algoritmos de indexação auto-

mática baseiam-se em textos completos ou resumos e envolvem nas suas operações elementos morfológicos, sintáticos e semânticos. Segundo Salton, nos dias de hoje não há razões técnicas óbvias que justifiquem a não substituição dos métodos manuais de indexação de documentos por métodos automáticos.^{1,2}

Um dos objetivos de um sistema de recuperação de informação é selecionar, de uma grande coleção de documentos, um sub-conjunto manipulável que seja capaz de satisfazer uma demanda de informação. Para desempenhar esta missão o sistema necessita de algum tipo de representação para cada documento incluído na coleção, i.e. os documentos precisam ser indexados. Lancaster³ define a indexação como a atividade na qual um especialista, o indexador, delibera sobre o conteúdo de um documento e constrói um perfil para este documento que deve assegurar a sua posterior recuperação. Associar um conjunto de termos de indexação a cada documento é a forma usual de se estabelecer este perfil. No processo de recuperação, o sistema geralmente calcula o grau de associação entre esses termos e um conjunto de termos originados da representação da consulta; o grau de associação estabelece as bases para o sistema decidir se o documento deve ou não ser recuperado. Desta forma, os procedimentos usados para indexar um documento determinam a capacidade de recuperação do sistema, assim como os custos e benefícios para seus usuários.⁴

Quando os termos de indexação usados para representar o conteúdo de um documento são retirados de um vocabulário controlado a indexação é dita controlada. As indexações controladas são realizadas por intermédio de uma linguagem artificial

orientada para a recuperação da informação. Esta linguagem deve ser projetada de forma a apresentar uma estrutura suficientemente formal e deve ser semanticamente bem definida. Uma linguagem de indexação completa consiste dos seguintes elementos^{3,5}:

- a) um conjunto de termos de indexação;
- b) um conjunto de termos de entrada;
- c) recursos especiais (por exemplo: controle de sinônimos, estrutura hierárquica, ponderação)
- d) regras de utilização

No processo de transferência de informação, a indexação ocorre em duas fases distintas: na análise do documento e na análise da consulta.

O objetivo da indexação é em última análise, representar adequadamente o conteúdo de documentos ou consultas através de um pequeno número de termos de indexação. Manualmente uma pessoa com conhecimento tanto da área específica do documento quanto das técnicas de indexação, lê o texto, identifica conceitos significante e assinala então termos de indexação selecionados de um vocabulário controlado ou Tesouro⁶. Na indexação automática o computador assume o papel do indexador.

1.1 Indexação Manual versus Indexação Automática

Na indexação manual o conteúdo do texto em linguagem natural do documento ou consulta é submetido a uma análise intelectual, para então ser representado por alguma linguagem de

indexação; o indexador faz uma análise particular para cada documento. Por outro lado, na indexação por máquina, um algoritmo substitui o indexador e é aplicado repetida e indiscriminadamente. Naturalmente o computador não lida com conceitos, mas unicamente com palavras. Os métodos de indexação automática devem portanto ser capazes de selecionar palavras significantes do texto que representem os conceitos importantes aí expressos. A máquina "vê" um texto escrito em linguagem natural como uma seqüência de símbolos - um deles é o "espaço" (blank). Ela pode identificar cadeias (strings) de caracteres separados por espaços e decidir de forma binária (sim/não) segundo a ausência ou presença de cadeias de caracteres particulares. Os algoritmos de indexação automática são desenvolvidos dentro destes princípios, i.e., baseados na capacidade do computador de reconhecer sinais e seqüências de sinais, características do texto tais como freqüência de ocorrência, co-ocorrência, posição relativa na frase, etc., de palavras, raízes de palavras ou expressões. Estes elementos são usados como pistas para a máquina proceder a análises do texto a ser indexado¹.

As análises que se seguem são geralmente fundamentadas em modelos matemáticos de textos escritos - uma representação formal que enfatiza características particulares do texto por meio de expressões matemáticas. Esses modelos se enquadram em duas grandes categorias:

- a) modelos determinísticos - usam, para representar textos, teorias da lógica, análise, álgebra ou geometria;
- b) modelos estocásticos - usam para representar textos teorias da probabilidade ou estatística.

A modelagem matemática de textos pode ser conduzida considerando-se um ou mais dos quatro níveis lingüísticos seguintes:⁷

- a) nível de grafema - o objeto de interesse são letras, sinais de pontuação e espaços;
- b) nível morfológico - o principal objeto de interesse é a estrutura da palavra - raízes e afixos (prefixos e sufixos);
- c) nível sintático - o principal objeto de interesse são as sentenças: análise da estrutura gramatical e a geração de frases gramaticalmente corretas;
- d) nível semântico - o principal objeto de interesse é o significado das palavras, frases, parágrafos, etc.

1.2 Problemas da Indexação Manual

Na maioria dos sistemas de recuperação de informação existentes a indexação é normalmente realizada manualmente. Nestes sistemas geralmente exige-se um alto nível de qualidade das indexações, pois este nível de qualidade tem reflexos diretos na recuperação. A necessidade de um treinamento específico é essencial: o indexador deve possuir conhecimento das características da área a que pertencem os documentos, do vocabulário de indexação e prática; além do mais o indexador deve ter também conhecimento das características da coleção e dos usuários do sistema, pois uma indexação de qualidade deve refletir

o tipo de consultas que o sistema espera processar no futuro⁸.

Outro problema que deve ser considerado está relacionado à inconsistência da indexação manual: mesmo que uma indexação seja realizada cuidadosamente, é impossível aplicar consistentemente os procedimentos de indexação, ainda mais que, na prática, é sempre necessário mais de um indexador para se indexar uma coleção de documentos. Dessa forma, o mesmo tipo de documento será indexado distintamente por diferentes indexadores. "Esta inconsistência necessariamente afeta o desempenho da recuperação e diminui a utilidade da decisão intelectual que controla o processo de indexação".⁸

A estes problemas se somam mais dois outros: tempo e custo. A informação deve ser coletada, selecionada, analisada, catalogada, indexada e tornada disponível ao usuário final sempre com urgência e a custo baixo. Mas o custo por si só não é um dado significativo; é necessário computar o custo-benefício. Sabe-se que "aproximadamente metade dos documentos que constituem uma base de dados de um sistema de recuperação nunca foram recuperados por ninguém. Obviamente, a indexação (...) não parece ser muito eficiente nem os seus benefícios compensados".⁶ Deve ficar claro que o custo-benefício deve ser também avaliado para os sistemas de indexação automática.

1.3 Problemas da Indexação Automática

Os pesquisadores da área de análise de texto em linguagem natural e particularmente os da área de indexação auto

mática, ressentem-se da falta de teorias que levem em consideração os fenômenos relacionados à linguagem natural. A falta dessas teorias torna impossível a construção de modelos linguísticos que possam representar completa e precisamente a estrutura da linguagem natural.⁹

Outro obstáculo que deve ser considerado é a falta de características universais destes estudos - são necessários modelos específicos para cada particular idioma; não é possível generalizá-los por enquanto, pelo menos.¹⁰ A falta de controle preciso sobre as variáveis linguísticas leva os sistemas mais avançados, permeados por técnicas gramaticais sofisticadas, a apresentarem um desempenho mais baixo que os sistemas mais simples que só utilizam tesouro para controlar sinônimos e termos relacionados.¹¹ "Os procedimentos mais simples de indexação (automática), aqueles que identificam num dado documento um conjunto de termos (...), são também os mais eficazes; dos muitos procedimentos estudados na tentativa de aumentar a revocação ou a precisão, somente o uso de um dicionário de sinônimos que agrupe termos relacionados em classes gera resultados melhores", afirma Salton⁹. Mas os sistemas simples (que usam por exemplo raízes e critérios estatísticos), apesar de serem relativamente fáceis de implementar e apresentarem desempenho considerado satisfatório, têm poucas chances de melhoramento, pois operam no limite máximo que os recursos que utilizam oferecem.¹² Um outro problema da área está relacionado à formalização matemática: as pesquisas em indexação automática se desenvolveram quase sempre dentro de uma perspectiva empírica; este fato teve como consequência a falta de provas analíti

cas formais da eficácia dos muitos métodos propostos.¹³

Um obstáculo que deve ser considerado também quando se pretende implantar um sistema de indexação é o "fator emocional que emerge quando qualquer procedimento automático é sugerido como substituto para uma atividade intelectual desempenhada manualmente. Não surpreende portanto que as evidências sobre a viabilidade dos sistemas mecanizados de indexação tenham produzido até agora tão poucos resultados práticos"².

1.4 Justificativas

A justificativa de qualquer técnica de indexação está centralizada principalmente nos resultados obtidos com a recuperação, e neste caso os resultados são bastante claros: dentro dos limites de uma amostra de documentos pequena e selecionada é possível demonstrar que:⁸

- a) Pode-se obter produtos de indexação de, pelo menos, alguns sistemas de indexação automáticos existentes, com níveis de consistência maiores que os sistemas manuais.
- b) A eficácia da recuperação no que diz respeito a itens relevantes recuperados (e não relevantes rejeitados) dos sistemas automáticos é no mínimo igual à dos sistemas manuais equivalentes.

- c) Os custos da indexação automática estão normalmente abaixo dos custos dos sistemas manuais equivalentes.

- d) Os sistemas de indexação automática oferecem geralmente uma economia significativa de tempo em relação aos processos manuais.

CITAÇÕES E NOTAS

- ¹ ARTANDI, S. Machine indexing, linguistic and semiotic implications. Journal of the American Society for Information Science, 27 (4): 235-39. July-August 1976.
- ² SALTON, G. A new comparison between conventional indexing and automatic text processing. Journal of the American Society for Information Science, 23 (2): 75-84, Mar./Apr. 1972.
- ³ LANCASTER, F.W. Information retrieval systems characteristics, testing and evaluation. New York, J. Willey, 1968. 222p.
- ⁴ BOOKSTEIN, A.; SWANSON, D.R. A probabilistic model for automatic indexing. Journal of American Society for Information Science, 25(5):312-8, Sept./Oct.1974.
- ⁵ PERSHKE, S.; VERNIMB, C. Feasibility study of automatic indexing of INIS abstracts. IAEA-SM-189-24 p.409-15
- ⁶ BORKO, H. Automatic indexing: a tutorial. SIGIR tutorial session, in: Automatic Information Retrieval. ACM'81, Los Angeles, CA; Nov.9-11, 1981.

- ⁷ EDMUNDSON, H.P. Mathematical Models of Text. Information Processing & Management, 20 (1,2): 261-68, 1984.
- ⁸ SALTON, G. Automatic indexing and abstracting. In: Dynamics information and library processing. Chapt. 3. Englewood Cliffs, N.J., Prentice Hall, 1975. p.76-144
- ⁹ SALTON, G. Automatic text analysis. Science, 168 (3929): 335-43, April 1970.
- ¹⁰ JANOS, J. Theory of functional sentence perspective and its application for the purpose of automatic extracting. Information Processing & Management, 15: 19-25, 1979.
- ¹¹ SALTON, G. Automatic term class construction using relevance - a summary of work in automatic pseudoclassification. Information Processing & Management, 16:1-15,1980
- ¹² DILLON, M; GRAY, A.S. A fully automatic syntactically based indexing system. Journal of the American Society for Information Science, 34(2): 99-108, 1983.
- ¹³ YU, C.T.; SALTON, G. Precision weighting - an effective automatic indexing method. Journal of Association for Computing Machinery , 23: 76-88, 1976.
- ¹⁴ SALTON, G. A new comparison between conventional indexing (MEDLARS) and automatic text processing (SMART). Journal of American Society for Science, March-April 1972 p.75-83

2 OBJETIVO

2.1 Objetivo Principal

Estabelecer um algoritmo de indexação automática utilizando vocabulário controlado. Tal algoritmo deverá ser capaz de indexar títulos e resumos de documentos escritos, sem que para isso seja necessário a interferência de qualquer indexador especialista nos assuntos tratados nos documentos.

2.2 Objetivos Secundários

2.2.1

O algoritmo deve ser capaz de sugerir ao indexador descritores para enriquecimento da indexação, assim como do vocabulário específico da área.

2.2.2

A qualidade da indexação obtida por esta metodologia, medida por testes de recuperação, deve ser, pelo menos, equivalente à da indexação manual correspondente.

3 REVISÃO DA LITERATURA

As experiências no sentido de automatizar todo o ciclo da análise documentária, embora relativamente recentes -as primeiras pesquisas se iniciaram há pouco mais de 30 anos, deram origem a uma literatura considerada vasta. Grande parte da literatura produzida a partir dessas experiências está dedicada à indexação automática. A revisão da literatura que segue, limita-se, sob todos os aspectos, a trabalhos de indexação automática mais diretamente relacionados ao tema desta dissertação.

É consenso entre os pesquisadores da área de processamento automático de textos que os primeiros trabalhos "sérios" sobre esta matéria começaram a ser publicados em fins dos anos 50 por Luhn¹. No período de 1957 a 1958 Luhn publicou vários artigos provocativos que tratavam das possibilidades de temas como "auto-resumo" (auto-abstracting) e "auto-indexação" (auto-indexing)^{2,3,4} - o termo "auto-indexação", sugerido originalmente por Luhn, foi substituído mais tarde por "indexação automática" (automatic indexing)⁵.

Luhn argumentava em seus trabalhos que o vocabulário contido em textos de documentos tinha necessariamente que constituir a base para uma análise do conteúdo e para a classificação do documento. Vários métodos de indexação foram propostos por Luhn a partir desses argumentos, como este, por exemplo: "... um conceito que ocorre pelo menos duas vezes no mesmo parágrafo, deve ser considerado um conceito importante; um conceito

que ocorre precedendo ou sucedendo um parágrafo, mesmo que apareça somente uma vez, deve ser considerado também um conceito importante..."²

Os trabalhos de Luhn sobre rotação de palavras significativas no título, no resumo e no texto completo - o índice KWIC (Key Word In Context) - começaram também por volta de 1958⁶:

Neste mesmo ano Baxendale⁷ publicou os resultados de suas experiências em indexação automática. As pesquisas pioneiras de Baxendale envolviam os seguintes tópicos: busca de sentenças significativas, processos sintáticos automáticos e seleção automática de expressões.

Grande parte das pesquisas realizadas nesta época (1957-1958), na IBM ou em outras instituições, foi na verdade incentivada pela "International Conference on Scientific Information" - ICSI, que aconteceu em Washington, em novembro de 1958. Os textos impressos deste evento, os "preprints" e os anais, foram deliberadamente preparados, apesar da objeção dos tipógrafos, de maneira a facilitar o processamento por máquina⁸. Muitas das primeiras experiências em indexação automática foram realizadas com este material; uma dessas experiências consistia na preparação de índices de palavras-chave por rotação de palavras de títulos, sub-títulos, título de seções e tabelas, além de legendas de figuras e frases selecionadas ou expressões tiradas diretamente do texto. Os dados foram preparados usando-se cartões perfurados. A listagem resultante da experiência foi distribuída aos participantes da Conferência^{7,8}. Outro conjunto de experiências envolvia técnicas de "auto-resumo" e "auto-indexação" propostas por Luhn⁹.

A rápida proliferação de trabalhos no campo da indexação automática a partir dessa época foi inspirada direta ou indiretamente pelos resultados das experiências realizadas com o material da ICSI.

O clima de debates sugeridos com as novas e desafiantes possibilidades pode ser expresso pela observação feita por Stevens⁶ (1965) no seu levantamento do estado da arte de indexação automática: "tendo em vista as controvérsias sobre a viabilidade e a avaliação das técnicas de indexação automática, um levantamento do estado da arte sobre este assunto nos dias de hoje talvez seja prematuro". As controvérsias de que falava Stevens podiam ser resumidas em pelo menos cinco questões:

- a) pode realmente uma indexação ser feita totalmente por máquina?
- b) o que pode realmente ser feito por máquina no processo de indexação, classificação ou confecção de resumo?
- c) o trabalho realizado pela máquina será pior, igual ou melhor do que o realizado manualmente?
- d) como avaliar a aceitabilidade da indexação automática, ou como compará-la com a indexação manual?
- e) a máquina pode chegar a um resultado satisfatório baseando-se apenas em considerações estatísticas, ou é preciso levar em conta aspectos de natureza sintática, semântica e pragmática?

Um ramo das pesquisas em indexação automática se

orientava no sentido de localizar no texto "palavras-pista" (clue-words), i.e., palavras que identificassem um assunto no texto. Swanson¹⁰ (1961) realizou uma série de experiências que caracterizavam esta linha de pesquisa. A fase preliminar desses estudos tinha por finalidade estabelecer a capacidade de discriminação das palavras-pista. Num primeiro momento, um grupo de documentos da área de física nuclear era incluído num corpus de documentos que tratavam de assuntos variados; as palavras mais freqüentes ocorridas nos artigos de física haviam sido consideradas palavras capazes de discriminar esse assunto. A hipótese a ser testada era a seguinte: qualquer documento pertencente à área nuclear deveria conter pelo menos duas dessas palavras. Dos 27 documentos de física nuclear incluídos no corpus de 100 documentos, 25 foram recuperados. Chegou-se então à conclusão que a capacidade das palavras-pista de reconhecer artigos da área de física nuclear era satisfatória, mas esse não era bom método para se recuperar documentos em resposta a uma consulta específica.

O segundo conjunto de experiências representava a transição entre as pesquisas em texto completo e as pesquisas em textos curtos, que significava mais ou menos a transição entre a "indexação derivativa" (experiência anterior) e a "indexação atributiva". Swanson¹¹ (1962) começou a trabalhar em 1.200 recortes de jornal de notícias breves. Para que a experiência fosse levada a cabo, construiu-se manualmente um tesouro formado por grupos de palavras associados a cerca de 20 cabeçalhos de assunto; cada palavra não podia pertencer a mais de um grupo, de forma que cada palavra indicasse somente uma

categoria. O processamento automático se realizava de acordo com a seguinte regra: se uma palavra pertencente a um particular grupo do tesouro ocorresse, então o cabeçalho de assunto correspondente a este grupo seria assinalado para o ítem em que havia ocorrido a palavra. Os resultados encontrados usando-se esta técnica, para este tipo de material pelo menos, foram altamente promissores, superando em qualidade a classificação manual. Swanson¹² (1963), além disso, propôs algumas estratégias que diminuíssem a ênfase nas palavras-pista, tais como localização de sinônimos e associação de pesos à palavra de acordo com a sua freqüência.

Maron¹³ (1960) também projetou modelo de indexação automática baseado em palavras-pista, mas com um enfoque probabilístico. O modelo tentava estabelecer uma associação estatística entre palavras-pista e cabeçalhos de assunto manualmente assinalados.

Continuando na linha probabilística inaugurada por Maron, Trachtenberg¹⁴ (1963) propôs um modelo probabilístico para indexar e classificar documentos automaticamente. Ele sugeriu métodos teóricos que determinariam o valor de discriminação das possíveis palavras-pista em relação à diferentes categorias.

Com a proliferação de projetos de indexação e classificação por computador, podia-se detectar uma certa tendência, dos pesquisadores da área, de testarem seus métodos no mesmo material a fim de facilitar as avaliações comparativas. Isto foi o que fizeram Borko¹⁵ e Borko em colaboração com Bernick^{16, 17, 18}; além de propor novas soluções, eles realizaram testes experimentais no mesmo material usado por Maron em (13).

Edmundson¹⁹ (1969) propôs um novo método de extração de frases com grande conteúdo informativo via computador. Estas frases deviam ser capazes de indicar ao leitor a substância do trabalho. Ao contrário de trabalhos anteriores que focalizavam a análise num componente da frase significante, ou seja, investigavam a presença de palavras de alta frequência (palavras-chave), o método de Edmundson reunia adicionalmente outras estratégias:

- a) método-pistas: baseava-se na hipótese de que frases de provável relevância podiam ser identificadas por "palavras pragmáticas" (pragmatic words), tais como "significante" e "impossível". Estas palavras estavam reunidas num dicionário de pistas;
- b) método-palavras-chave: semelhante ao proposto por Luhn em ⁽⁴⁾, baseava-se na alta frequência de algumas palavras;
- c) método-literário: baseava-se na hipótese que o autor resume no título o que há de mais importante no texto do documento;
- d) método-localização: partia da hipótese de que algumas posições dentro do texto são estratégicas e que frases que ocorressem nestas posições seriam provavelmente relevantes.

As características ou pistas acima eram reconhecidas por máquina, que atribuía pesos numéricos em função da presença ou não dessas pistas. A soma dos pesos definia a re-

cuperação ou não da frase.

Segundo Lancaster²⁰ "provavelmente o maior avanço em direção aos sistemas de recuperação da informação totalmente automatizados está representado pelo sistema experimental SMART, desenvolvido por Salton em Havard". O SMART aceita tanto documentos quanto consultas em linguagem natural na forma legível por máquina; realiza análise automática de conteúdo e ordena os documentos de acordo com o grau de coincidência com a consulta; o sistema traz incorporados uma série de procedimentos para otimizar a precisão e revocação. O SMART deu origem às mais interessantes pesquisas na área de indexação automática; Salton detalhou o sistema, as experiências e os testes realizados em dois livros publicados em 1965²¹ e 1968²². Uma visão geral do SMART pode ser encontrada em artigo de Salton publicado em 1967²³.

As características mais importantes incorporadas ao sistema SMART para análise de documentos são relacionadas a seguir:

- a) um sistema para separar palavras em inglês em raízes e sufixos;
- b) um dicionário de sinônimos, ou tesouro, que pode ser usado para reconhecer sinônimos através de substituição de cada palavra (ou raiz de palavra), por um ou mais "número de conceito" (concept number); estes códigos servem como identificadores de conteúdo no lugar da palavra original;
- c) um arranjo hierárquico de conceitos incluído no tesau

ro que torna possível, dado um número de conceito, descobrir todas as suas ramificações hierárquicas e todas as referências cruzadas relacionadas a ele. Este arranjo hierárquico pode ser usado para obter-se identificadores de conteúdo mais geral ou mais específico que os originalmente dados;

- d) procedimentos de associação estatística usados para calcular coeficientes de similaridade entre palavras, raízes de palavras ou conceitos, baseados em padrões de co-ocorrência entre elementos da frase de um particular documento ou de documentos de uma coleção;
- e) métodos de análise sintática que tornam possível comparar frases de documentos com consultas;
- f) métodos de reconhecimento estatístico de expressões, que operam baseados em dicionário de expressões pré - construídas; neste método não se faz nenhum teste para se verificar se os condicionantes sintáticos entre os termos estão satisfeitos;
- g) procedimentos para comparação consulta/documento, que tornam possível o uso de uma grande variedade de métodos de correlação para comparar documentos e consultas;
- h) sistemas de dicionários, projetados para atualizar os vários dicionários incluídos no sistema, tais como dicionário de raízes e sufixos;
- i) classificação automática de documentos que classifica documentos com identificadores de conteúdo similares em "clusters", de forma que uma consulta seja submetida somente a alguns "clusters" e não a todo o acervo;

j) sistema de retroalimentação para o usuário que modifica a consulta com base nas informações fornecidas por ele durante a consulta.

Salton¹ (1970) lamentava a falta de teorias que explicassem os fenômenos relacionados à linguagem natural: "A falta de tais teorias torna impossível a construção de modelos lingüísticos que descrevam completa e precisamente a linguagem natural(. . .) . Mas apesar da carência de bases lingüísticas, os sistemas automáticos produzem uma recuperação pelo menos tão eficaz quanto os métodos convencionais. Contudo, infelizmente, todos os procedimentos de indexação sejam eles manuais ou automáticos produzem resultados medíocres".

Em 1971 Salton²⁴ deslocando-se um pouco dessa problemática, fez também estudos sobre o valor da citação bibliográfica como indicador direto do conteúdo de documentos. Ele demonstrou que as citações em conjunto com as palavras-chave usuais melhoravam o desempenho da recuperação e que em algumas circunstâncias, as citações seriam mais eficazes para a recuperação do que os termos convencionais.

Em 1976, Salton e seu colaborador Yu²⁵ retomavam os problemas relacionados à lingüística e a formalização matemática de situações concernentes à recuperação automática de informação. Os problemas abordados foram:

a) a falta de elementos lingüísticos confiáveis durante o processo de indexação e recuperação - os aspectos semânticos das consultas eram desprezados em favor de características tais como distribuição de frequência e localização no texto;

- b) a falta de provas analíticas formais, devido, provavelmente, à natureza experimental da área, da superioridade de um método sobre o outro.

Tentando equacionar estes problemas, Yu e Salton expressaram um método de ponderação baseado na precisão, que usava pesos para determinar a caracterização lingüística de cada termo. Desenvolveram também um formalismo matemático para provar o desempenho do método. Os aspectos mais importantes desse método foram as transformações sugeridas pelos autores para aumentar a frequência de atribuição dos termos raros e diminuir a dos termos freqüentes.

- a) os termos genéricos que ocorriam em muitos itens e contribuíam para produzir um índice de precisão insatisfatória, deviam ser combinados com termos apropriados para formar "expressões-termos" (term-phrases);
- b) termos específicos que ocorriam em muito poucos itens e contribuíam para um baixo índice de revocação, deviam ser incorporados a uma classe do tesouro, e estas classes deviam ser assinaladas como identificadoras de conteúdo no lugar do termo individual.

Salton publicou também trabalhos onde desenvolveu teorias relacionadas à indexação automática envolvendo elementos da álgebra vetorial²⁶ e de conjuntos nebulosos (fuzzy sets)²⁷.

Salton²⁸ (1982) resumiu o desenvolvimento das técnicas de indexação automática. O trabalho foi escrito de uma forma bastante interessante: Salton dividiu didaticamente os procedimentos de indexação automática em elementos que foram se justapondo, aumentando gradativamente o nível de sofisticação

das técnicas, i.e., partiu de um sistema simples até chegar a um algoritmo complexo, que incorporava todos os desenvolvimentos alcançados até aquela data.

Ainda em 1969, Moyne²⁹, usando o argumento de que tanto a ciência da computação quanto a lingüística estavam suficientemente desenvolvidas, defendia o uso da linguagem natural no planejamento de sistemas de recuperação como solução para se chegar aos sistemas gerais.

Graves e Helander³⁰ (1970) estudaram a viabilidade de se implantar um sistema de indexação automática para os resumos do Petroleum Abstracts. Esse estudo consistia na comparação entre as expressões e palavras em linguagem natural contidas no título e no resumo, com descritores controlados assinalados manualmente pelo indexador. O estudo indicou que apenas 40% dos descritores assinalados estavam presentes no título e no resumo. Este resultado inviabilizou o projeto.

Bookstein e Swanson³¹ publicaram em 1974 um trabalho bastante importante na medida em que caracterizava uma nova linha de extração automática de indexadores. As pesquisas desses autores se orientavam na direção de métodos puramente probabilísticos. O desenvolvimento do módulo parte da observação de que as palavras portadoras de conteúdo tendem a se concentrar em poucos documentos em relação às não portadoras, i.e., as portadoras tendem a se agrupar em "clusters". Bookstein e Swanson estudaram essas propriedades de agrupamento ("clustering"), padrões de ocorrência, distribuição de ocupação, etc. baseados unicamente em propriedades probabilísticas. A importância desse trabalho estava sobretudo nas novas possibilidades que ele

trazia para a compreensão intelectual do processo de indexação em si.

Sparck Jones³² (1973) investigou a influência das características de uma coleção sobre o desempenho de um sistema de classificação automática. Para isso Sparck Jones analisou o desempenho, através de testes de recuperação, de três coleções diferentes. O resultado foi satisfatório para apenas uma das coleções; desenhou então testes para investigar as causas deste resultado. Sparck Jones concluiu que a falha estava na própria coleção e não nos procedimentos de classificação automática. Em 1977 Sparck Jones³³ publicou os resultados de suas pesquisas em indexação automática. Os relatos se concentravam mais enfaticamente nas avaliações comparativas realizadas entre os sistemas em funcionamento nesta época, com atenção especial ao sistema SMART^{21, 22} de Salton. Todo o trabalho foi permeado por um tratamento estatístico. As conclusões mais importantes a que chegou Sparck Jones foram as seguintes:

- a) a ponderação estatística é uma estratégia eficiente e influencia sensivelmente os resultados; já a classificação, incorporada a sistema de indexação, não se mostrou tão eficiente;
- b) variações nas condições de contorno do sistema (por exemplo: exaustividade), não causam geralmente grande influência nos resultados;
- c) técnicas de retroalimentação estatística no processo de busca são altamente positivos; a saída ordenada é também um fator que deve ser considerado.
- d) os resultados dos testes comparativos a que foi subme

tido o SMART em relação aos outros projetos disponíveis na época são geralmente equivalentes.

Nos anais da 39ª Congresso da FID, realizado em setembro de 1978, Sparck Jones³⁴ analisou o papel da indexação automática nos sistemas de recuperação em linha operacionais. Sparck Jones defendeu a necessidade de posições mais realistas em relação aos sistemas totalmente automatizados e aos sistemas unificados. Este enfoque pragmático não admitia, pelo menos até a época, o uso de técnicas lingüísticas sofisticadas em sistemas de recuperação de informação; "...estamos tão distantes dos sistemas de alta qualidade totalmente automatizados quanto estávamos há 20 anos atrás" afirmava a autora; achava também necessário direcionar as pesquisas no sentido de melhorar as técnicas de formulação de consulta e não tanto a descrição do documento.

Field³⁵ (1975) projetou um sistema capaz de gerar automaticamente cabeçalhos de assunto controlados, descritores controlados e classes para documentos a partir de uma indexação livre. Os procedimentos usados pelo método envolviam a manipulação e truncamento automático de expressões livres assinaladas manualmente pelo indexador; usava também um tesouro construído manualmente. O método incorporava a geração automática dos dicionários usados no processo de transformação dos indexadores livres em controlados; incorporava ainda técnicas estatísticas para ponderação e ordenação. O grau de associação entre os diferentes elementos da indexação (por exemplo, entre cabeçalho de assunto e indexação livre), era expresso por um "coeficiente de adesão" (adhesion coefficient) gerado estatisticamente. Field testou o seu método usando uma amostra retirada do "input" da

base de dados do INSPEC. Em 1977 Field³⁶ anunciou uma versão multilíngüe do seu sistema que usava tesauro inglês (tesauro do INSPEC) e em francês (tesauro do INIS, pois não havia versão em língua francesa do tesauro do INSPEC) para gerar descritores controlados em uma única língua - o inglês, independente da língua em que tinha sido escrito o descritor livre original (inglês ou francês).

Artandi³⁷ (1976) publicou um artigo tecendo considerações sobre as implicações lingüísticas e semióticas nos processos de manipulação automática da linguagem natural. Artandi argumenta que o estudo mais profundo do significado pode contribuir para o desenvolvimento de sofisticados algoritmos de indexação por máquina: "...a compreensão do significado vai além do contexto lingüístico revelado pelo texto, pois ele não é um conceito lingüístico mas sim um conceito psicológico".

Van Rijsbergen³⁸ (1977) ofereceu também a sua contribuição ao estudo dos modelos de indexação probabilísticos, abordados de outra forma por Robertson e Sparck Jones³⁴. Van Rijsbergen desenvolveu uma teoria que buscava aumentar a eficácia dos sistemas de recuperação automática, explorando o conceito de dependência de co-ocorrências dos termos. Partia da suposição de que os termos de indexação não são independentes e calculava então funções não-lineares ponderadas para as quais cada parâmetro era estimado a partir da distribuição de "ocorrências" e "co-ocorrências"; o termo independente era considerado um caso particular do dependente.

Barnes e outros³⁹ relataram em 1977 as experiências realizadas em indexação automática utilizando como software bá

sico o SLC II Language Translator Package⁴⁰ - um sistema de múltiplas aplicações que analisava e identificava os elementos lingüísticos, especialmente os sintáticos. Os testes foram feitos com o INIS Atomindex e o tesouro INIS. O sistema de indexação possuía um módulo de "enriquecimento" do tesouro, através do qual incorporava novos termos ao tesouro, selecionados do texto processado.

Van der Meulen⁴¹ (1977) realizou avaliações comparativas entre o DIRECT (sistema bastante semelhante ao SMART) e a indexação manual da base de dados do INSPEC. Curiosamente Meulen usou para esta experiência apenas duas consultas; chegou à conclusão de que a indexação manual é 20% melhor que a obtida pelo sistema DIRECT. Concluiu então que a falha do sistema automático estava na formulação da consulta, já que a indexação automática era comparável em qualidade com a manual.

Dunham e outros⁴² (1978) conceberam um sistema de indexação automática destinado à indexação de diagnósticos patológicos. O sistema utiliza uma linguagem de indexação estruturada chamada SNOP (Systematized Nomenclature of Pathology). A indexação iniciava-se com a criação de um arquivo-texto em terminal de computador, através de um sistema interativo de edição de textos. O sistema fazia uso de análise morfológica e sintática e também de um dicionário auxiliar de múltiplas funções e formas capaz de identificar sinais de pontuação palavras e expressões vazias, reconhecimento de sinônimos, etc.

Dillon⁴³ publicou em 1981 os resultados de um trabalho inédito. Ele tentou mostrar a semelhança que existe entre os sistemas de indexação controlada automática e os sistemas

automáticos de geração de autoridades controladas (por exemplo nomes de autores). Para demonstrar essa semelhança, Dillon desenvolveu um sistema experimental fundamentado em um software de indexação automática baseada em tesouro. Neste sistema o texto do documento era examinado com a finalidade de se detectar alguns termos, como por exemplo, o nome do autor, que eram então remetidos à forma autorizada correspondente, através de um vocabulário controlado. Um exemplo: se o nome do autor "livre" (i.e., como ele aparece na página de rosto de um livro qualquer) era "Wilfrid Lancaster" o sistema assinalava a forma controlada, "Lancaster, F.W.", encontrada no vocabulário controlado.

Em 1982 Dillon⁴⁴ apresentou um sistema de indexação automática para livros baseado em tesouro. Esse sistema dependia de um sub-sistema que formatasse o texto convenientemente para a leitura por computador, usando para isso o sistema SCRIPT, desenvolvido na Universidade Waterloo, Canadá. Outra exigência do sistema era um tesouro construído manualmente. Dillon concluiu, após alguns testes, que o sistema funcionava bem, apesar do alto dispêndio de esforço humano na construção do tesouro, de cuja qualidade dependia o desempenho do sistema; concluiu também que o sistema funcionaria melhor em área de vocabulário altamente específico e com uma filosofia de indexação que desse ênfase à exaustividade. Dillon⁴⁵ descreveu num trabalho publicado em 1983 o sistema FASIT (Fully Automatic Syntactically based Indexing of Text). Este sistema foi desenhado para realizar indexações totalmente automáticas e de qualquer texto, sem entretanto considerar elementos semânticos. O FASIT está fundamenta

do em quatro sub-sistemas que desempenham as seguintes tarefas:

- a) categorização sintática das palavras individuais;
- b) seleção de conceitos baseada em um conjunto de regras pré-definidas;
- c) reunião dos conceitos em classes;
- d) geração de elementos de indexação a partir de classes.

Neste mesmo ano Dillon⁴⁶ publicou outro trabalho em que relatava os resultados da aplicação do FASIT para indexação de livros. Dillon utilizou-se para a experiência de um manual de processamento de dados para bibliotecas, por possuir um texto condensado e técnico, mas não muito especializado; outro elemento favorável era o índice remissivo da publicação construído manualmente, que poderia ser usado como padrão de comparação. Dillon⁴⁷ (1984) propôs ainda uma nova estratégia para melhorar a qualidade dos sistemas de recuperação que dependiam do uso de palavras truncadas ou raízes, através de termos chamados por ele de "termos tópicos" (topical terms) - "um segmento do texto que indicava sobre o que trata o texto". Dillon concluiu que estes termos tópicos podiam ser identificados por suas propriedades estatísticas.

Borko⁴⁸ (1981) apresentou palestra em que revisou rapidamente os procedimentos tradicionais e os mais recentes avanços na área de indexação automática. Borko defendeu o abandono definitivo da indexação manual em favor da automática, assim como a substituição "dos já antiquados conceitos de revocação e precisão por parâmetros mais apropriados para medir os benefícios que chegam ao usuário" baseado no conceito de medida de utilidade".

Aitchinson e Harding⁴ (1982) projetaram um sistema de indexação e classificação automáticas para a base de dados do INSPEC que é indexada manualmente por três elementos de indexação: códigos de categoria, descritores controlados e descritores livres. Os objetivos iniciais do projeto eram:

- a) desenvolver um sistema de análise de vocabulário de linguagem livre;
- b) partindo do título e do resumo do documento, gerar termos livres, termos controlados e códigos de classifi-
cação;
- c) avaliar a recuperação automática e comparar os custos da indexação automática versus indexação manual.

Durante o transcorrer do projeto, se configurou a possibilidade de incorporação de um sub-sistema que geraria automaticamente perfis. Este sub-sistema foi desenvolvido e os perfis resultantes foram mais tarde comparados com os construídos manualmente. Os autores chegaram à conclusão de que a geração de termos livres e de perfis era satisfatória e os custos eram realmente baixos; o desempenho na recuperação era comparável ao desempenho do processamento manual correspondente para algumas faixas de revocação. A atribuição automática de termos controlados e a classificação automática apresentaram custos menores que os do sistema manual e permitiam buscas com alto índices de revocação.

Stokolov⁵ (1982) descreveu uma técnica para assinalar cabeçalhos de assunto para o BIOSIS. Esta técnica era baseada no uso de uma linguagem especialmente formalizada para re-
presentação semântica de textos biológicos - a "linguagem dos conceitos primitivos", que tendo por finalidade diminuir o vo-

cabulário de biologia.

Nishida e outros⁵¹ (1983) publicaram trabalho descrevendo um método semi-automático de extração de informações de textos escritos em língua inglesa ou japonesa. Os textos eram analisados semi-automaticamente por métodos precisos de análise sintática e por métodos superficiais de análise semântica; os textos (escritos em ambos os idiomas) eram então normalizados para a mesma forma através de "expressões internas".

Brozowski⁵² (1983) desenvolveu para uso em sua companhia um sistema de armazenamento e recuperação de informação chamado MASQUERADE. O objetivo do autor era criar um sistema que usasse técnicas já dominadas e que agilizasse a manipulação das coleções de sua instituição, que eram constituídos de relatórios de geologia e exploração, mapas, dados de seções de choque e seções sísmicas, etc. O projeto tinha um caráter essencialmente prático. Brozowski classificou seu sistema de amigável (friendly); o sistema incorporava características tais como indexação automática, consulta livre e por lógica booleana, sistema de ponderação visando a ordenação da saída, retroalimentação das consultas por relevância e possibilidade de implantação em várias bases de dados.

Bernstein e Willianson⁵³ (1984) fizeram um relato sobre o sistema ANNOD - "A Navigator of Natural Language Organized Data". Este sistema de recuperação de informação combina o uso de elementos probabilísticos, lingüísticos e "empíricos" para ordenar cada parágrafo do texto completo em função da sua similaridade com a consulta formulada em linguagem natural. O sistema ANNOD incorpora em seu software rotinas de eliminação de

palavras comuns, separação de raízes de palavras, expansão de consulta via tesouro e um algoritmo de ordenação.

Edmundson⁵⁴ (1984) fez um levantamento sobre modelos matemáticos de textos escritos. Revisou os principais conceitos envolvidos na análise lingüística de textos tais como, gramática, morfologia, sintaxe e semântica e os relacionou com a Ciência da Informação.

O panorama da literatura brasileira sobre análise de texto por computador, mais especificamente sobre indexação automática, reflete o pouco desenvolvimento desta matéria no país. Os autores brasileiros que publicaram trabalhos sobre indexação automática preocuparam-se com maior intensidade na utilização, comparação e construção de sistemas KWIC. Bastos⁵⁵ (1984) na revisão da literatura de sua dissertação de mestrado, cita Zaher^{56, 57} afirma que a primeira experiência da utilização do programa KWIC no país foi para editar a Bibliografia Brasileira de Física em 1968. "Pode-se dizer que este fato marca o início da indexação automática no Brasil". Bastos comparou a qualidade da indexação manual com a indexação obtida por programa KWIC de artigos da revista Ciência da Informação. Braga⁵⁸ (1982) realizou uma análise numérica do conteúdo informativo do título e resumo de documentos da área de química tendo em vista o enriquecimento do título com palavras significativas do resumo. Martins⁵⁹ (1983), em sua dissertação, tinha como objetivo desenvolver um método de indexação automática baseado em análise sintática para os trabalhos enviados pelo CIN ao INIS; diante da impossibilidade de alcançar os objetivos previamente estabelecidos o autor desenvolveu então um sistema do tipo KWIC para in

dexar os títulos dos documentos.

Nos anais do III Congresso da Sociedade Brasileira de Computação, realizado em 1983, foi publicada a descrição de um sistema semi-automático de geração de índices remissivos em língua portuguesa - o PIXPAL, desenvolvido por Von Staa⁶⁰ na Pontifícia Universidade Católica do Rio de Janeiro. Este sistema está baseado no formatador de texto ATF disponível naquela universidade. O sistema PIXPAL foi desenvolvido em linguagem SPITBOL e consta de 9 programas. O PIXPAL foi utilizado para gerar o índice remissivo de um livro que deveria ser publicado comercialmente.

Haller⁶¹ (1982) fez um levantamento sobre os conceitos básicos relacionados ao processamento de texto em linguagem natural por computador. Descreveu também as aplicações comerciais da análise de texto. Em 1983, Haller⁶² publicou um relato sobre um sistema de indexação experimental baseado em regras lingüísticas que estava sendo desenvolvido na Universidade de Brasília. O sistema incorporava métodos de segmentação de texto, procura de palavras em dicionário, análise morfológica de palavras escritas em língua portuguesa, clareza de homografia sintática, construção de árvore de dependência, análise de pronomes.

Andreewsky e Ruas⁶³ (1982) desenvolveram na Pontifícia Universidade Católica do Rio de Janeiro um sistema de indexação automática de documentos escritos em linguagem natural. O sistema utilizava métodos lingüísticos combinados com métodos estatísticos que permitiam uma indexação ponderada. Este sistema é uma adaptação para a língua portuguesa do Système

Syntaxique et Probabiliste d'Informations Textuelles (SPIRIT) desenvolvido pelo "Centre National de la Recherche Scientifique" (CNRS) para a língua francesa. O SPIRIT é composto de um dicionário com 250 mil formas de vocábulos em francês e de algoritmos de análise sintática e semântica.

Robredo, utilizando a lei de Bradford-Zipf como instrumento de controle terminológico, desenvolveu vários trabalhos na área de indexação automática.^{64, 65, 66, 67}

CITAÇÕES E NOTAS

- ¹ SALTON, G. Automatic text analysis. Science. 168 (3929): 335-43, 17 Apr. 1970.
- ² LUHN, H.P. A statistical approach to mechanized encoding and searching of literature information. IBM Journal of Research and Development, 1(4):309-317, Oct.1957.
- ³ LUHN, H.P. The automatic creation of literature abstracts. IBM Journal of Research and Development, 2:159-165,1958
- ⁴ LUHN, H.P. Auto-encoding of documents for information retrieval system. In: BOAZ, M., ed.Modern Trends in Documentation nl, 1959. p.45-58.
- ⁵ LUHN, H.P. Automatic intelligence systems-some basic problems and prerequisites for their solution. In: TOMESKI, E.A. et al. Classification. Unification and Integration of Storage and Retrieval, nl, 1961. p.3-20.
- ⁶ STEVENS, M.E. Automatic indexing: a state-of-the-art report. National Bureau of Standards, Washington D.C., 1965.290p. (NBS-91).
- ⁷ BAXENDALE, P.B. Machine-made index for technical literature-an experiment. IBM Journal of Research and Development, 2:354-61, 1958.

- ⁸ CITRON, J.L.; HART, L.; OHLMAN, H. A permutation index to the "Preprints of the International Conference on Scientific Information". Santa Monica, Cal. System Development Corp., 1958. 140p. (SP-44).
- ⁹ LUHN, H.P. An experiment in auto-abstracting: auto-abstracts of area 5. In: INTERNATIONAL CONFERENCE ON SCIENTIFIC INFORMATION, 17 Nov 1958. Conference Papers, Yorktown Heights, N.Y., 18p.
- ¹⁰ SWANSON, D.R. Research Procedures for automatic indexing. In: _____ . Machine Indexing. American U., N.Y. , 1962. p.281-304
- ¹¹ SWANSON, D.R. Library goals and the role of automation. Spec.Libraries, 53: 466-71, 1962.
- ¹² SWANSON, D.R. Automation indexing and classification. In: NATO ADVANCED STUDY INSTITUTE ON AUTOMATIC ANALYSIS. Venice, 1963. Proceedings... N.Y., 1963, 125-128.
- ¹³ MARON, M.E. Automatic indexing: an experimental inquiry. (rev.2 Feb 1961}, 1960 31p. (P-2180)
- ¹⁴ TRACHTENBERG, A. Automatic document classification using information theoretical methods. In: LUHN, H.P., ed. Automation and Scientific Communication, 1963, Pt.2. p.349-50.

- ¹⁵ BORKO, H. The construction of an empirically based mathematically derived classification system. Santa Monica, Cal., System Development Corp., 26 Oct. 1961, 23p. (SP-585).
- ¹⁶ BORKO, H.; BERNICK, M.D. Automatic document classification. Santa Monica, Cal., System Development Corp., 13 Nov. 1963. 19p. (TM-771)
- ¹⁷ BORKO, H.; BERNICK, M.D. Automatic document classification, Part II - Adittional Experiments. Santa Monica, Cal., System Development Corp., 18 Oct 1963. 33p. (TM-771/001/001)
- ¹⁸ BORKO, H.; BERNICK, M.D. Toward establishment of a computer based classification system for scientific documentation. Santa Monica, Cal., System Development Corp., 19 Feb. 1964. 47p. (TM-1763)
- ¹⁹ EDMUNDSON, H.P. A new method in automatic extracting. Journal of ACM, 16(2): 264-85, April 1969.
- ²⁰ LANCASTER, F.W. Information retrieval systems: characteristics, testing and evaluation. J.Willey, New York, 1964, 222p.
- ²¹ SALTON, G. The SMART retieval systems - experiments in automatic document processing. New York, Prentice-Hall, Englewood Cliffs, 1981

- ²² SALTON, G. Automatic information, organization and retrieval.
New York, McGraw-Hill, 1968. 222p.
- ²³ SALTON, G. Designing automatic information system; results
obtained with the SAMRT programs. Social Science Information,
6(2):111-17, Feb 1967.
- ²⁴ SALTON, G. Automatic indexing using bibliographic citations
Journal of Documentation, 27(2): 98-110 , June 1971.
- ²⁵ YU, C.T.; SALTON, G. Precision weighting - an effective
indexing method. Journal of Association for Computing
Machiminery, 23:76-88, 1976.
- ²⁶ SALTON, G.; YANG, G.S.; YU, C.T. A theory of term importance
in automatic text analysis. Journal of the American
Society for Information Science, 26(1): 33-44, Jan - Feb
1975.
- ²⁷ SALTON, G. The measurement of the term importance in
automatic indexing. Journal of the American Society for
Information Science, 32 (3): 175-86, May 1981.
- ²⁸ SALTON, G. Automatic indexing: a summary. In: EURIM 5
CONFERENCE, Versailles, France, 1982. Proceedings...,
ASLIB, 1982. p.66-77.
- ²⁹ MOYNE, J.A. Information retrieval and natural language.
In: AMERICAN SOCIETY FOR INFORMATION SCIENCE. Proceedings

N.Y., nd. v.6, p.259-263.

- ³⁰ GRAVES, R.W.; HELANDER, D.P. A Feasibility study of automatic indexing and information retrieval. IEEE Transactions on Engineering Writing and Speech, 32 (2): 58-59, 1970.
- ³¹ BOOKSTEIN, A.; SWASON, D.R. A probabilistic models for automatic indexing. Journal of the American Society for Information Science, 25(5): 312-8, Sep/Oct 1974.
- ³² SPARCK JONES, K. Collection properties influencing automatic term classification performance. Information Storage and Retrieval, 9: 499-13, 1973.
- ³³ SPARCK JONES, K.; BATES, R.G. Research on automatic indexing 1974-1976. University of Cambridge, 1977.
- ³⁴ SPARCK JONES, K. The role of automatic indexing in operational on-line retrieval systems. In: FID CONGRESS, 39, Edinburg, 1978. New trends in documentation and information London, ASLIB, 1980, p.33-8.
- ³⁵ FIELD, B.J. Towards automatic indexing: automatic assignment of controlled-language indexing and classification from free indexing. Journal of Documentation, 31(4)Dec.1975.
- ³⁶ FIELD, B.J. Indexation automatique pour systèmes multilingues.

- In: CONGRÈS EUROPÉEN SUR LES SYSTÈMES ET RÉSEAUX DOCUMENTAIRES, 1, Luxemburgo, 1977. Proceedings... Luxemburgo, Verlag Dokumentation, 1977. p.479-501.
- ³⁷ ARTANDI, S. Machine indexing: linguistic and semiotic implications. Journal of the American Society for Information Science, p.235-39, July-August 1976.
- ³⁸ VAN RIJSBERGEN, C.J. A theoretical basis for the use of co-occurrence data in information retrieval. Journal of Documentation, 27(2):69-82, June 1971.
- ³⁹ BARNES, C.I.; CONSTANTINI, L.; PERSCHKE, S. Automatic indexing using the SLC-II system. Information Processing & Management, 14:107-19, 1978.
- ⁴⁰ KOLAR, W. The SLC-II language translation package user manual for input, Luxembourg, Commission for the European Communities, 1974 (EUR-5173).
- ⁴¹ VAN DER MEULEN, W.A.; JANSEN, P.J.F.C. Automatic versus manual indexing. Information Processing and Management, 13(1): 13-21, 1977.
- ⁴² DUNHAM, G.S.; PACAK, M.G.; PRATT, A.W. Automatic indexing of pathology data. Journal of the American Society for Information Science, p.81-90, 1978

- ⁴³ DILLON, M.; KNIGHT, C.; LOSPINUSO; ULMSCHENEIDER. The use of automatic indexing for authority control. Journal of Library Automation, 14(4): 268-77, 1981.
- ⁴⁴ DILLON, M. Thesauro-based automatic book indexing. Information Processing & Management, 18 (4):167-78, 1982.
- ⁴⁵ DILLON, M.; GRAY, A. Fully automatic syntax-based indexing. Journal of the American Society for Information Science, 34 (2):99-108, 1983.
- ⁴⁶ DILLON, M.; MC DONALD, L.K. Fully automatic book indexing. Journal of Documentation, 39(3):135-54 Sep.1983.
- ⁴⁷ DILLON, M.; FEDERHART, P. Statistical recognition of content terms in general texts. Journal of the American Society for Information Science, 34(1):3-10, 1984.
- ⁴⁸ BORKO, H. Automatic indexing: a tutorial. SIGIR tutorial session. In: ACM'81. Automatic Information Retrieval. Los Angeles, CA, Nov.9-11, 1981.
- ⁴⁹ AITCHISON, T.M.; HARDING, P. Automatic indexing and classification for mechanised information retrieval, snt.
- ⁵⁰ STOKOLOV, N.V. On automatic support to indexing a life science data base. Information Processing & Management, 18 (6):313-21, 1982.

- ^{5 1} NISHIDA, F.; TAKAMATSU, S.; FUJITA, Y. Semiautomatic indexing of structured information of text. J.Chem.Comput.Sci. , 24(1):15-20, 1984.
- ^{5 2} BROZOZOWSKI, J.P. MASCARADE: searching the full text of abstracts using automatic indexing. Journal of Information Science, 6: 67-73, 1983.
- ^{5 3} BERNSTEIN, L.M.; WILLIANSO, R.E. Testing of natural language retrieval system for a full text knowledge base. Journal of the American Society for Information Science, 35(4): 235-47, 1984.
- ^{5 4} EDMUNDSON, H.P. Mathematical models of the texts. Information Processing & Management, 20 (12): 261-68, 1984.
- ^{5 5} BASTOS, S.B. Análise comparativa entre indexação automática e manual da literatura brasileira de ciência da informação. Brasília. Universidade de Brasília, 1984.204p. (Dissertação)
- ^{5 6} ZAHER, C.L. et alli. Automação da informação em Física no Brasil. In: SEMINÁRIO SOBRE INFORMÁTICA, Rio de Janeiro, 1968. Anais... Rio de Janeiro, IBBD, 1969, p.39-52.
- ^{5 7} ZAHER, C.R.; DUARTE, Y.C. Sistema KWIC versus descritores. In: CONGRESSO SOBRE DOCUMENTAÇÃO, 2; FID/CLA Reunião 9, Rio de Janeiro, 1969. Anais... Rio de Janeiro, IBBD, 1969. p.195-206

- ⁵⁸ BRAGA, L.M. Palavras de títulos e resumos como acesso ao conteúdo do documento: uma análise numérica. Rio de Janeiro, UFRJ/IBICT, 1982. 181p. (Dissertação)
- ⁵⁹ MARTINS, A.V. Um método para indexação automática de textos. Rio de Janeiro, Instituto Militar de Engenharia, 1983. 100p. (Dissertação)
- ⁶⁰ VON STAA, A. PRAXPAL: um indexador semi-automático. In: CONGRESSO DA SOCIEDADE BRASILEIRA DE COMPUTAÇÃO, 3. Campinas, 1983. Anais... São Paulo, 1983. p.433-43.
- ⁶¹ HALLER, J. Processamento de textos em linguagem natural. In: CONGRESSO NACIONAL DE INFORMÁTICA, 15., Rio de Janeiro, out. 1982. Anais... Rio de Janeiro, 1982. 9p.
- ⁶² HALLER, J. Análise automática de textos em sistemas de informação. Revista de Biblioteconomia de Brasília, 11(1): 105-13 jan/jun 1983.
- ⁶³ ANDREWSKY, A.; RUAS, V. Indexação automática baseada em métodos lingüísticos e estatísticos e sua aplicabilidade à língua portuguesa. Rio de Janeiro, PUC-DI, 1982. 31p.
- ⁶⁴ ROBREDO, J. Otimização dos processos de indexação dos documentos e de recuperação da informação mediante o uso de instrumentos de controle terminológico. Ciência da In - formação, 11(1):3-18, 1982.

- ⁶⁵ ROBREDO, J. A indexação automática de textos: o presente já entrou no futuro. In: Machado, U.D., ed. Estudos avançados em Biblioteconomia e Ciência da Informação. Brasília, ABDF, 1982. v.1 p.236-74
- ⁶⁶ ROBREDO, J.; FERREIRA, J.A. Conceituação de um programa para indexação automática de textos. Revista de Biblioteconomia de Brasília, 8(2):254-63, jul./dez. 1980.
- ⁶⁷ ROBREDO, J. A indexação automática como mecanismo básico no processo de transferência da informação. In: CONGRESSO LATINO-AMERICANO DE BIBLIOTECONOMIA E DOCUMENTAÇÃO, 1., Salvador, 21-26 set 1980. Anais... Salvador, FEBAB, 1980. 19p.

4 AMBIENTE DE ESTUDO

Neste capítulo descreve-se o contexto em que se desenvolveu o presente trabalho - o Centro de Informações Nucleares, suas origens, os serviços e os sistemas mais ligados à informação bibliográfica.

4.1 Centro de Informações Nucleares-CIN

O CIN é um departamento da Comissão Nacional de Energia Nuclear e foi criado em junho de 1970, tendo por objetivo dar suporte bibliográfico a cientistas e técnicos brasileiros que atuam no campo nuclear.

4.1.1 Origens do CIN

Para se compreender as origens do CIN, é necessário conhecer um pouco da história do Sistema Internacional de Informações Nucleares-INIS, do qual o CIN faz parte.

A Agência Internacional de Energia Atômica (IAEA), órgão filiado à ONU, tem entre seus objetivos o de "promover a troca de informações técnico-científicas, sobre a utilização

de energia atômica para fins pacíficos".¹ Por este motivo, em fins de 1969, foi criado o INIS que, dentro da estrutura organizacional da Agência, seria o órgão responsável pelo intercâmbio das informações nucleares, a nível internacional.

Como seria impossível para o INIS tomar para si a responsabilidade do tratamento de toda a literatura nuclear, foi adotado como "modus operandi" o sistema de rede.

Dentro da orientação descentralizada, os estados-membros do INIS seriam os únicos alimentadores de informações. Para isso, uma das atribuições dos países seria a de coletar, processar e enviar o material bibliográfico da área nuclear, publicado dentro de suas fronteiras geográficas, ao órgão central em Viena. O INIS receberia estas informações e as reuniria num arquivo único, formando assim a base de dados do sistema.

Todos os países já membros da Agência foram convidados a participar no estabelecimento do INIS. A primeira medida a ser adotada por esses países seria a indicação de um órgão que atuaria como centro Nacional do INIS.

A Comissão Nacional de Energia Nuclear (CNEN), foi o órgão brasileiro a participar no estabelecimento do INIS.

Vendo a importância que esta participação traria para a comunidade nuclear do Brasil, a CNEN criou em junho de 1970, alguns meses após a criação do INIS, o Centro de Informações Nucleares. O CIN seria então o órgão nacional responsável pela participação bibliográfica do Brasil e pela disseminação das informações enviadas pelo INIS aos pesquisadores nucleares brasileiros.²

Entretanto, com o passar do tempo percebeu-se que havia por parte dos usuários do CIN uma demanda considerável de informações fora do escopo da base de dados INIS. O CIN então, para cobrir esta demanda, adotou duas linhas de ação levadas a cabo simultaneamente: a) aquisição de novas bases de dados contendo informações disponíveis no mercado internacional e b) formação de uma base de dados contendo informações que dificilmente seriam cobertas por bases de dados estrangeiras. A formação desta base de dados tinha também outra finalidade - desenvolver competência em construção de bases de dados bibliográficos. As bases de dados operadas pelo CIN até o fim de 1984 são:

- a) INIS (International Nuclear Information System) - cobre a área de ciência e tecnologia nucleares;
- b) ENDS (European Nuclear Documentation System) - acervo também na área nuclear, mas anterior ao INIS (documentos publicados de 1943 a 1970);
- c) INSPEC (Information Service-Physics, Electrical & Electronics, and Computer & Control) - acervo cobrindo as áreas de Física, Engenharia Elétrica e Eletrônica, Computação e Controle;
- d) FONTE - composta de documentos em fontes de energia, a maioria deles brasileiros. O método proposto foi projetado para indexar os documentos a serem incluídos nesta base. Detalhes sobre a FONTE são discutidos no capítulo 5.

4.1.2 Serviços e Sistemas do CIN

O CIN tem como diretriz básica a geração e desenvolvimento de "know-how" na área de sistemas de informação. Todos os sistemas operados pelo CIN foram criados com tecnologia desenvolvida por sua própria equipe. Aqui vamos falar das atividades mais ligadas à informação bibliográfica.

4.1.2.1 Processamento Automático da Informação-PAI

O PAI tem por finalidade captar publicações nas mais diversas fontes produtoras de literatura (instituições de pesquisa, universidades, revistas técnico-científicas, congressos etc.), para formação de bases de dados.

Este serviço compreende ainda a seleção, análise, catalogação, indexação, redação de resumo e gravação de dados, em fita magnética. Os dados em fita magnética constituem a matéria prima para outros serviços oferecidos pelo CIN, tais como impressão de bibliografias, buscas retrospectivas e DSI³.

4.1.2.2 Disseminação Seletiva da Informação-SONAR

O sistema de DSI SONAR dissemina quinzenalmente, de acordo com perfis profissionais individuais, informações ex -

traídas das bases de dados do INIS e do INSPEC, e bimestralmente da FONTE. Usa lógica de patamar⁴ como base de recuperação. O SONAR possui atualmente cerca de 1.800 usuários por todo Brasil e América Latina⁵.

4.1.2.3 Sistema de Busca Retrospectiva-SUPRIR

O SUPRIR é um sistema interativo, em linha, que recupera informações bibliográficas via lógica booleana. Interage com acervo de aproximadamente três milhões de informações⁵ pertencentes às quatro bases de dados atualmente disponíveis no CIN.

4.1.2.4 Acesso ao Documento Primário-SERVIR

Integrado ao SONAR e ao SUPRIR, o SERVIR fornece cópias de documentos de interesse que os usuários têm dificuldade de obter por seus próprios meios. Com este objetivo o CIN utiliza seus próprios arquivos de microficha, entra em contato com uma rede informal de bibliotecas no Brasil e ainda utiliza serviços de fornecimento de cópias no exterior⁶.

CITAÇÕES E NOTAS

- ¹ PELZER, C. International information networks: The International Nuclear Information System. ASLIB Proc. 24 (1): 38-55, Jan.1972
- ² BARREIRO, S.C. Avaliação do serviço de disseminação seletiva de informação do Centro de Informações Nucleares em relação à precisão, recuperação e novidade. Rio de Janeiro, UFRJ/IBICT, 1979. 160 p. (Dissertação)
- ³ BARREIRO, S.C.; QUEIROZ, G.G. Incorporação automática de informações brasileiras no INIS. Revista de Biblioteconomia de Brasília, 8(2):118-130, jun./dez. 1980.
- ⁴ Tradução do termo em inglês "Threshold logic". Critério usado na recuperação da informação fundamentado na soma algébrica de pesos atribuídos às palavras-chave de uma consulta. Se a soma dos pesos das palavras-chave da consulta coincidentes com as palavras-chave do documento que está sendo analisado for maior ou igual a um certo número real e positivo - o patamar, o documento é recuperado, caso contrário é rejeitado.
SAYÃO, L.F.; QUEIROZ, G.G. Otimização de perfis ponderados de DSI usando-se um sistema de busca retrospectiva por lógica booleana. In: CONGRESSO LATINO-AMERICANO DE BIBLIOTECONOMIA E DOCUMENTAÇÃO, 1., Salvador, 21-26 set., 1980. Anais, 23p.
- ⁵ CINFORME. Rio de Janeiro, Centro de Informações Nucleares, n.91, jan.1985.
- ⁶ QUEIROZ, G.G.; FRADE, M.E.; GONÇALVES, O.C. Acesso ao documento: experiência do CIN/CNEN. In: REUNIÃO BRASILEIRA DE CIÊNCIA DA INFORMAÇÃO, 2., Rio de Janeiro, 1979. Anais, 113p.

5 MATERIAL

Neste capítulo descrevem-se as principais características da base de dados FONTE; justifica-se o seu uso na metodologia proposta. Apresenta-se também o critério usado na escolha da área em que se trabalhou, assim como a definição da amostra.

5.1 FONTE: Base de Dados Bibliográficos em Fontes de Energia

O CIN, a partir de 1983, sensível à demanda de informação em energia por parte dos seus usuários, decidiu concentrar esforços na formação de um acervo de informações bibliográficas, em suporte legível por computador, composto de documentos em fontes de energia, na sua maioria brasileiros. Criou-se então a FONTE, que está sendo explorada pelo SUPRIR para buscas retrospectivas e pelo SONAR para DSI. Além disso a FONTE gera semestralmente uma bibliografia em energia - Energia: Bibliografia Seletiva (EBS)¹.

5.1.2 Principais Características

Na FONTE são incluídos artigos de periódicos, anais

de conferências, relatórios, livros (analisados capítulo a capítulo), patentes, teses, etc. Inclui-se prioritariamente a produção brasileira: trabalhos publicados por brasileiros no Brasil ou no exterior e por estrangeiros no Brasil. Documentos estrangeiros de interesse para os projetos em desenvolvimento no país são incluídos como segunda opção.

A indexação a que são submetidos os documentos incorporados na FONTE é controlada (ver exemplo no anexo 1); o tesauro² (anexo 2) adotado assim como o esquema de classificação³ (anexo 3) são do Department of Energy-DOE, dos Estados Unidos. O tesauro vem sendo adotado na íntegra, mas alguns termos ligados essencialmente à realidade brasileira são incluídos à medida que se fazem necessários - "national alcohol program" e "jojoba oils" são exemplos de termos que estão nesta categoria⁴. O esquema de classificação foi adotado sem a parte relativa à energia nuclear, isto para evitar superposição com a base de dados do INIS.

São incluídos na FONTE resumos em língua portuguesa e inglesa para os documentos em português e resumos apenas em língua inglesa para documentos em qualquer outra língua. A catalogação segue o formato INIS (anexo 1), já usado pelo CIN para a preparação da parte brasileira da base de dados INIS.

5.2 A escolha da FONTE

As razões que determinaram a escolha da base de dados FONTE para que fosse realizada a experiência aqui relatada fo-

ram duas:

- a) A FONTE é uma base de dados projetada e executada inteiramente pelo CIN; todos os serviços derivados da FONTE são operados também exclusivamente pelo CIN. Isto torna mais fácil a introdução de modificações nas rotinas da FONTE que porventura se façam necessárias, para a implementação e implantação da metodologia que será proposta.
- b) A FONTE é uma base de dados ainda pequena. Isto facilitou algumas operações, como por exemplo a compilação dos termos do dicionário histórico da área N13 (anexo 4), que é discutida no capítulo 6.

5.3 A Escolha da Área N13

Selecionou-se os documentos incluídos até 31/12/84 na área N13 da base de dados FONTE - combustíveis a partir de hidrocarbonetos, como conjunto universo para realização da experiência. A escolha recaiu sobre esta particular área pelo fato dela encerrar um número ainda reduzido de documentos (sessenta e oito documentos até a data desta seleção).

5.4 A Escolha da Amostra

A partir dos documentos da área N13, selecionou-se

aleatoriamente, através de um programa de computador, uma amostra mínima de trinta documentos⁵.

CITAÇÕES E NOTAS

¹ QUEIROZ, G.G. Fonte; base de dados bibliográficos em fontes de energia. In: CONGRESSO BRASILEIRO DE BIBLIOTECOLOGIA E DOCUMENTAÇÃO, 14., Balneário Camboriú, SC, out. 1983. trabalhos apresentados, 14p.

² USA DEPARTMENT OF ENERGY. Technical Information Center. Energy data base: subject thesaurus. Oak Ridge, Te., 1981 (DOE/TIC 7000 R5).

³ USA DEPARTMENT OF ENERGY. Technical Information Center. Energy data base: energy categories. Oak Ridge, Te., 1980 (DOE/TIC 4584 R4).

⁴ Os termos que não têm um correspondente consagrado em inglês são incluídos em português, de preferência dentro da hierarquia de um termo já existente.

Exemplo

OILS

Bt₂ Organic Compounds

Nt, Castor Oil

Nt₂ Jojoba Oils

⁵ De acordo com Kenneth Bailey, "... geralmente o tamanho da amostra depende do tamanho da população a ser estudada. Embora as regras gerais sejam difíceis de se elaborar sem

conhecimento da população específica, em torno de 30 casos parece ser o mínimo para estudos, nos quais deve ser feita a análise de dados estatísticos, embora algumas técnicas possam ser usadas em menos de 30 casos".

BAILEY, Kenneth D. Methods of social research. New York, The Free Press; London, Collier Macmillan Publishers, 1978. p.84. apud BRAGA, L.M. Palavras de títulos e resumos como acesso ao conteúdo do documento: uma análise numérica. Rio de Janeiro, UFRJ/IBICT, 1982. 181p. (Dissertação).

6 METODOLOGIA

Neste capítulo desenvolvem-se os procedimentos usados pelo algoritmo proposto - o método SALF, que se desdobra em duas versões:

- a) SALF-S: que gera descritores estritamente dentro da área específica em que se indexa;
- b) SALF-P: que além de gerar descritores dentro da área específica em que se indexa, sugere ao indexador novos descritores para enriquecer a indexação e o vocabulário da área.

O capítulo está dividido em quatro seções: em primeiro lugar apresenta-se o método, detalhando-se descritivamente os procedimentos adotados; em seguida, o método é apresentado em forma de fluxograma; como ilustração, indexa-se passo a passo o primeiro documento da amostra; finalmente descreve-se a metodologia usada para avaliar o método proposto.

6.1 Descrição do Método

6.1.1 Preparação do Texto-Base

Seguindo a tendência da grande maioria dos projetos de indexação automática, o método proposto utiliza apenas o título e o resumo como fontes de extração dos elementos informativos do documento¹. O título e o resumo, aqui chamados de "texto-base", constituem a entrada para o algoritmo. É a partir do texto-base que são gerados os descritores dos documentos que se pretende indexar.

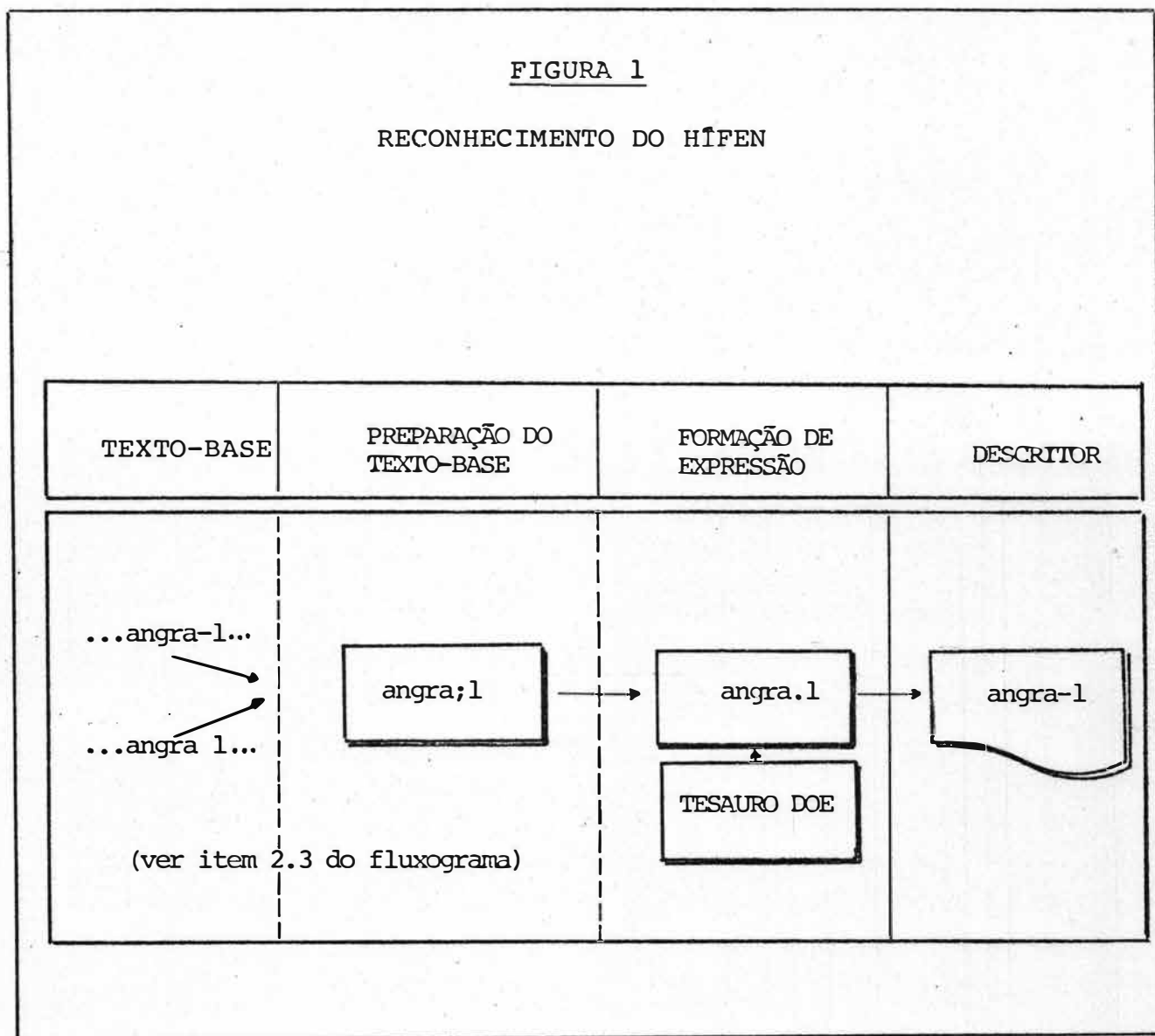
O texto-base deve obrigatoriamente estar escrito em língua inglesa, pois o tesauro utilizado pelo CIN para a indexação do acervo da FONTE, o tesauro do DOE-Department of Energy dos Estados Unidos, está escrito em inglês.^{2,3}

O método reconhece no texto-base palavras, i.e., grupos de caracteres (que podem ser letras ou números) limitados por espaços. Por exemplo; "uranium 235" é reconhecido pelo método como duas palavras "uranium" e "235". Os sinais gráficos que são reconhecidos são três:

- a) ponto final - identificado como limitador de frase;
- b) ponto e vírgula - também identificado como limitador de frase;
- c) hífen - é considerado pelo método por se tratar de um sinal gráfico que freqüentemente aparece nos termos do tesauro DOE. As palavras compostas encontradas no texto-base, cujos componentes estão unidos ou não por hífen, são tratadas, num primeiro momento, como palavras isoladas; na última fase, se pertinentes ao tesauro DOE são reconstituídos à forma originalmente registrada pelo tesauro (ver exemplo na figura 1).

FIGURA 1

RECONHECIMENTO DO HÍFEN



OBS: Outros símbolos (gráficos, acentos, matemáticos, de pontuação, etc) são ignorados pela metodologia.

Do texto -base expurgam-se palavras consideradas sem utilidade para recuperação: palavras de alta entropia⁴, vazias de sentido para a área particular em que se realiza a experiência; para esta operação utiliza-se integralmente, como dicionário negativo, o "INIS/STAIRS Stopword List" (anexo 5), publicado no "INIS Manual for Online Retrieval"⁵

O único aspecto de natureza morfológica levado em consideração pela metodologia foi o número - após o expurgo descrito no parágrafo anterior, todas as palavras que estão na forma plural são reduzidas à forma singular correspondente; esta transformação pode ser realizada por dois meios distintos: através de um dicionário de correspondência plural-singular ou por intermédio de um algoritmo de redução plural/singular.⁶

Após terem sido retiradas por intermédio do dicionário negativo, as palavras sem utilidade para a recuperação, o texto -base é submetido a outro expurgo. A razão pela qual o texto -base é submetido a mais esta "filtragem", está relacionada ao fato de que a indexação resultante da metodologia deve ser obrigatoriamente controlada; é inútil portanto manter no texto-base palavras cuja probabilidade de derivarem termos pertinentes ao tesouro DOE seja nula. Esta apuração é realizada com auxílio de um "tesouro de trabalho". Este tesouro é formado por palavras simples (i.e., não compostas), no singular, ordenadas alfabeticamente, resultantes do desmembramento dos termos, sejam eles preferenciais ou proibidos,⁷ do tesouro DOE. Por exemplo, o termo do tesouro DOE "abscopal radiation effects" gera os seguintes termos do tesouro de trabalho: "abscopal", "radiation" e "effect" (ver estrutura do tesouro de trabalho na

figura 2]. O expurgo descrito neste parágrafo baseia-se na seguinte hipótese: se uma palavra do texto-base não pertence ao tesauro de trabalho ela não terá nenhuma probabilidade de ser ou formar um termo do tesauro DOE; deve ser, portanto, eliminada, posto que a indexação que se deseja obter deve ser controlada.

Finalmente, o texto-base foi dividido em frases - o método reconhece o ponto final e o ponto e vírgula como os únicos sinais de pontuação divisores de frases. O título é analisado como uma frase comum; mantém-se entretanto a sua identidade, para que posteriormente possam ser distinguidos os descritores extraídos do título, do resumo ou de ambos.

6.1.2 Formação de Expressões

Após ter passado pelas operações da fase anterior, uma particular frase está formada por palavras potencialmente capazes de se transformarem em termos ou parte de termos do tesauro DOE.

Partindo do pressuposto que grupos de palavras são mais específicos do que palavras isoladas⁸ (por exemplo, a expressão "system analysis" tem um significado distinto e mais específico do que "system" e "analysis" isoladamente), a finalidade desta fase é combinar palavras da frase, de maneira que elas formem expressões pertinentes ao tesauro DOE. O método tenta formar prioritariamente expressões de níveis de complexida-

FIGURA 2

ESTRUTURA DO TESAURO DE TRABALHO

TERMOS DO TESAURO DOE		TERMOS DO TESAURO DE TRABALHO
FORMA ORIGINAL	FORMA SINGULAR	
.....
ABCESSSES	ABCESS	ABCESS
ABSCOPAL RADIATION EFFECTS	ABSCOPAL RADIATION EFFECT	ABSCOPAL
ABSOLUTE COUNTING	ABSOLUTE COUNTING	ABSOLUTE
ABSOLUTE ZERO TEMPERATURE	ABSOLUTE ZERO TEMPERATURE	ABSORBED
absorbed doses	absorbed dose	ABSORPTION
ABSORPTION	ABSORPTION	COUNTING
absorption (skin)	absorption (skin)	DOSE
.....	EFFECT
.....	INSTABILITY
.....	RADIATION
.....	SKIN
.....	TEMPERATURE
.....	ZERO
.....
.....

OBS: Os termos proibidos⁷ do tesauro DOE estão escritos em caixa baixa.

de maiores, i.e., composto por um maior número de palavras. No caso desta dissertação, o nível de complexidade máximo considerado é de três palavras, isto porque quando se levantou exaustivamente os descritores usados para indexar manualmente os documentos da área N13, constatou-se que o descritor mais complexo tinha quatro palavras e havia sido usado apenas uma vez (ver Dicionário Histórico da Área N13, anexo 4). Resumindo: nesta fase o método tenta formar expressões de três palavras, a seguir de duas e finalmente de uma palavra, nesta ordem.

A primeira dependência à formação de combinações mais complexas de palavras é o número de palavras da frase: se a frase possuir três ou mais palavras é possível formar combinações de três palavras (tripletos), de duas palavras (dupletos) e de uma palavra (singleto)⁹; caso a frase seja composta apenas por duas palavras, forma-se apenas dupletos e singletos; no caso extremo de uma palavra, só é possível formar um singleto.

Suponhamos uma frase F_i , formada por um número igual ou superior a três palavras ($n \geq 3$). A formação de tripletos se processa da seguinte forma:

$$F_i: P_1, P_2, \dots, P_j, \dots, P_n \quad (n \geq 3)$$

onde P_j é a palavra de ordem j da frase F_i .

Os tripletos formados são:

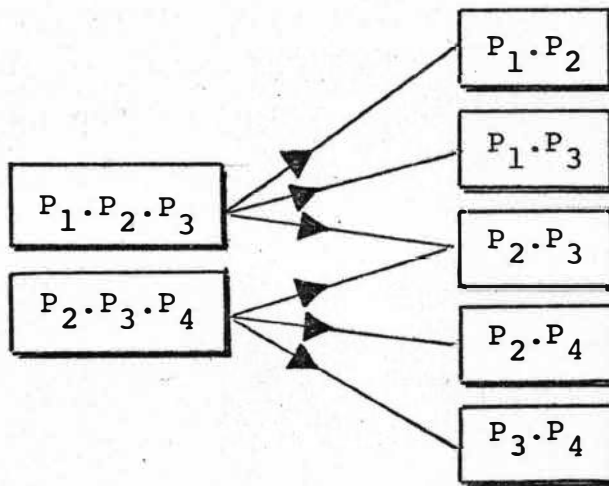
- 1) $P_1 P_2 P_3$
- 2) $P_2 P_3 P_4$
- ...

$$n-2) \quad P_{n-2}P_{n-1}P_n$$

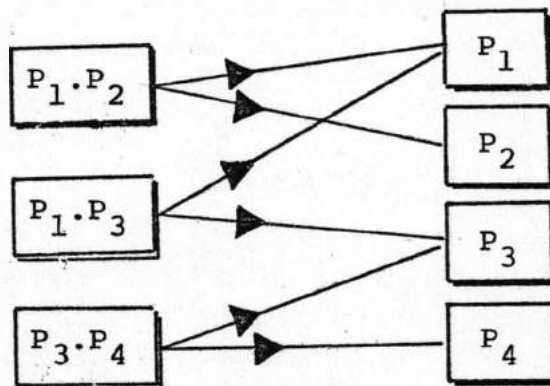
o que resulta num número de tripletos igual a $n-2$.

Estes $n-2$ tripletos são então comparados com o tesauro DOE para que se verifique a pertinência desses grupos; os tripletos não pertinentes devem gerar dupletos.

Supondo que os tripletos $P_1P_2P_3$ e $P_2P_3P_4$ não são pertinentes então eles devem formar os seguintes dupletos:






Estes dupletos, por sua vez, são comparados ao tesauro DOE; os que não são pertinentes, geram singletos. Se por exemplo, P_1P_2 , P_1P_3 e P_3P_4 não são pertinentes, geram os seguintes singletos:



Finalmente, os singletos que não são pertinentes são eliminados.

Os tripletos, dupletos e singletos pertinentes ao tesauro DOE devem submeter-se a mais alguns "controles" antes de se transformarem em descritores do documento que está sendo indexado.

- a) Controle de sinônimos - os grupos de palavras (tripleto, dupletos, singletos) pertinentes ao tesauro DOE, mas equivalentes a termos que não são termos preferenciais (termos alternativos, sinônimos, quase-sinônimos, termos com grafia alternativa), devem ser substituídos pelos termos preferenciais equivalentes. Esta operação é realizada através de lista de termos não preferenciais (termos proibidos) com ponteiros indicando o termo preferencial correspondente. Por exemplo:

termo proibido	substituir por	TERMO PREFERENCIAL
<div style="border: 1px solid black; width: 40px; height: 20px; margin: 0 auto; text-align: center; line-height: 20px;">x</div>		<div style="border: 1px solid black; width: 40px; height: 20px; margin: 0 auto; text-align: center; line-height: 20px;">X</div>
fusion fuels		THERMONUCLEAR FUELS
information centers		INFORMATION CENTERS

- b) Controle de expressões - os grupos de palavras que são pertinentes ao tesauro DOE, mas que eventualmente entram na composição de grupos de palavras também perti

entes, porém mais complexos (i.e., que são formados por mais palavras), devem ser eliminados. Por exemplo:

COAL GASIFICATION —————> eliminar

COAL GASIFICATION PLANTS —————> manter

- c) Controle do vocabulário da área N13 - os grupos de palavras pertinentes ao tesouro DOE devem, obrigatoriamente, para tornarem-se descritores, pertencer ao dicionário histórico da área N13. Este dicionário é constituído de todos os descritores usados, pelo menos uma vez, para indexar manualmente os documentos incluídos na área N13 da base de dados FONTE até 31 de dezembro de 1984. Os termos do dicionário histórico estão ordenados alfabeticamente e acompanhados da respectiva frequência de ocorrência (ver anexo 4). A função deste dicionário é evitar eventuais distorções no processo de atribuição de descritores pelo método. Resumindo: um grupo de palavras só é considerado um descritor válido se já foi usado pelo indexador pelo menos uma vez anteriormente, o que vale dizer que o grupo de palavras só é válido se pertencer ao dicionário histórico da área N13.

- d) Controle das palavras não pertinentes à área N13 - os grupos de palavras pertinentes ao tesouro DOE, mas não ao dicionário histórico da área N13, devem ser submetidos ao julgamento de um indexador especialista no assunto; se aceitos, são então incorporados à indexação do

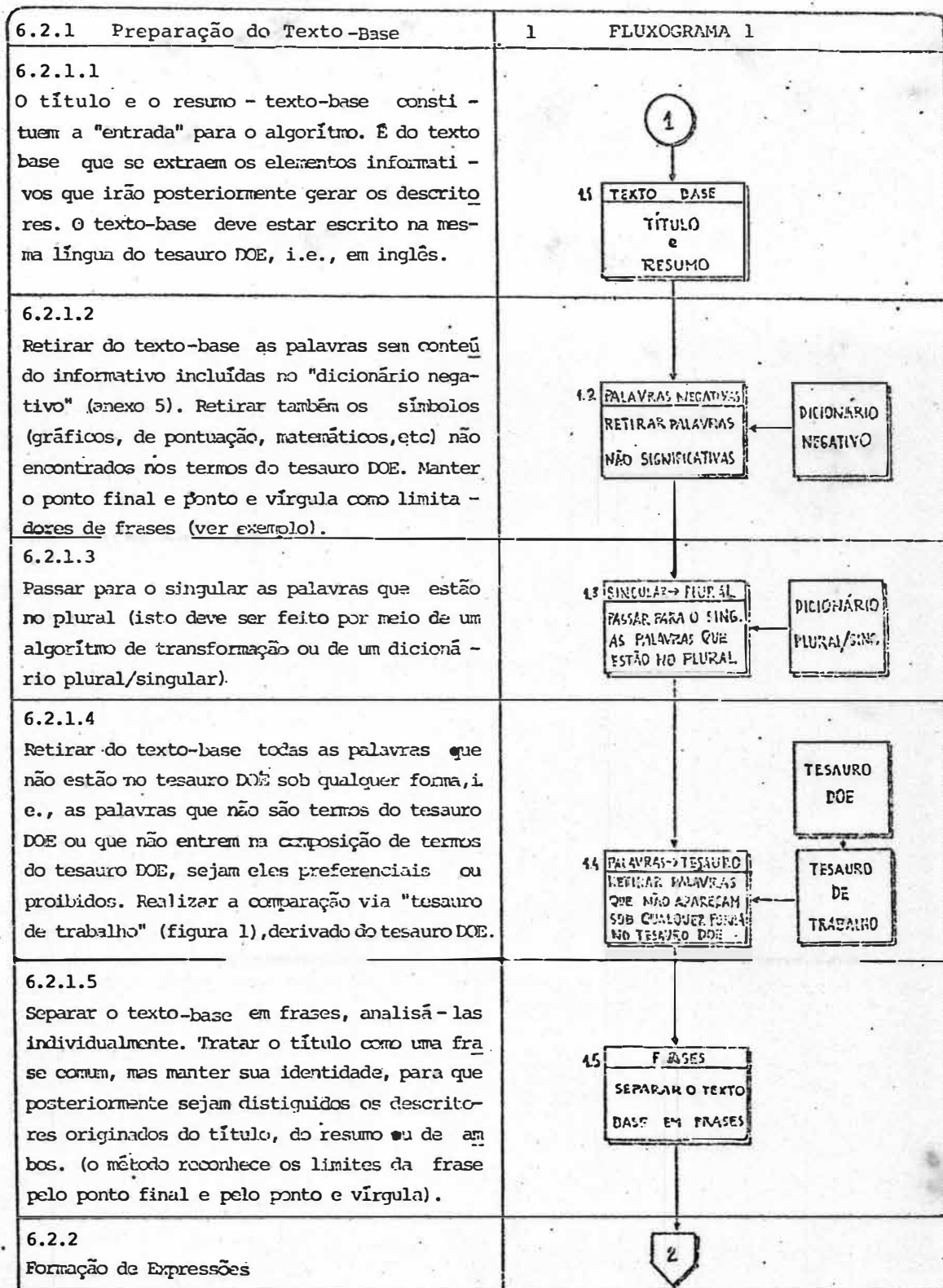
documento com o nome de "descritores sugeridos aceitos" (DSA); caso não sejam aceitos, são então eliminados. A indexação pelo método SALF, incluindo os descritores DSA, é denominada SALF-P, a versão que não inclui os descritores DSA denomina-se SALF-S.

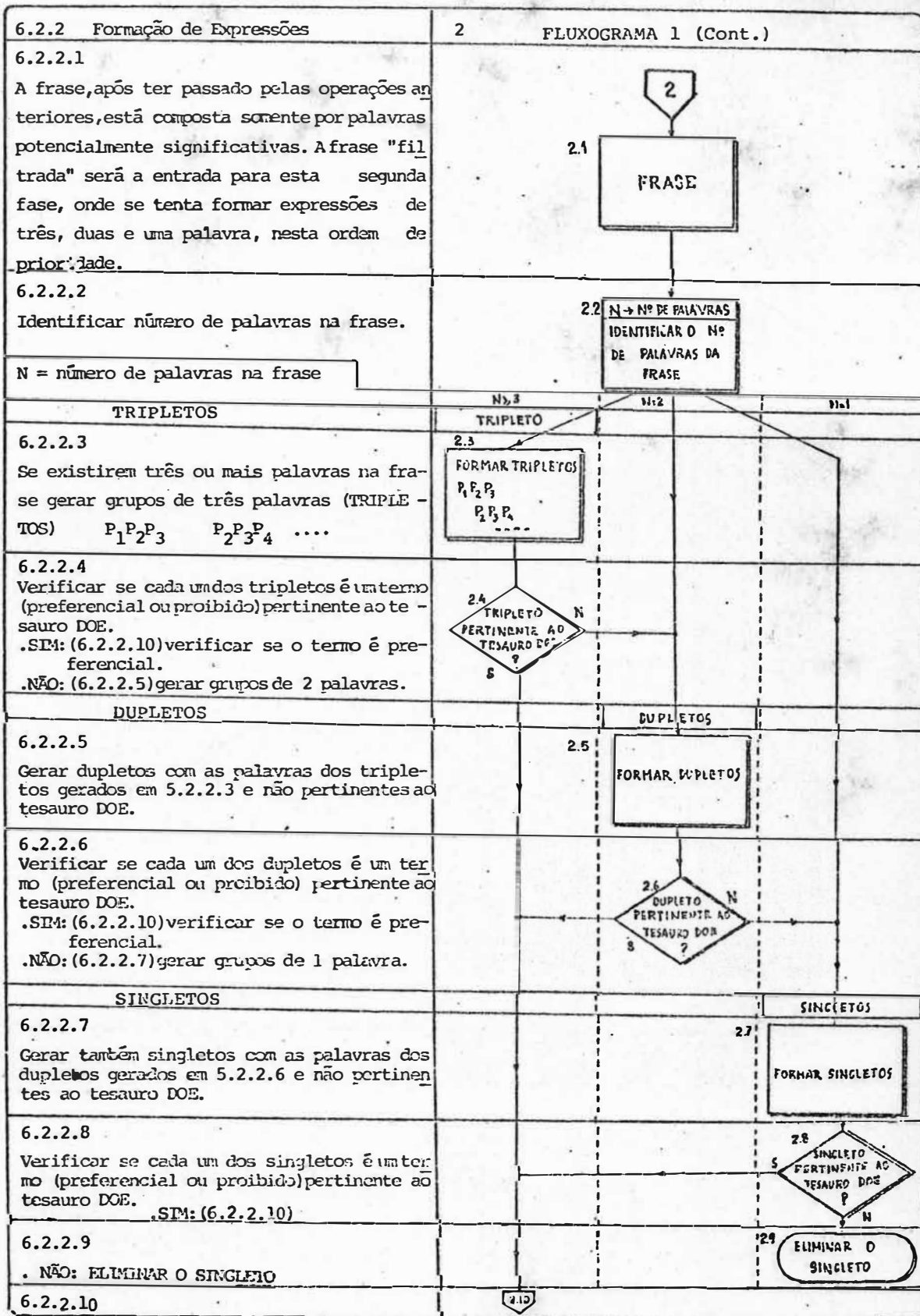
A formação de expressões para frases que possuam apenas duas palavras (i.e., $n = 2$), inicia-se a partir da geração do único duplete possível; para frases de apenas uma palavra ($n = 1$), a partir de um único singlete possível (para melhor compreensão ver itens 2.5 e 2.7 do fluxograma).

Como o tesouro DOE registra termos distintos formados por um mesmo grupo de palavras, mas em outra ordem, como por exemplo, "fuel gas" e "gas fuel", há a possibilidade de o mesmo grupo de palavras formado pelo método apontar para mais de um termo no tesouro DOE. Neste caso, é assinalada prioritariamente a forma que aparece na frase que está sendo analisada. Por exemplo:

<u>FRASE</u>	<u>DUPLETO</u>	<u>TERMOS DO TESAURO DOE</u>	<u>DESCRITOR ASSINALADO</u>
..gas fuel..	gas.fuel	fuel gas gas fuel(s)	GAS FUELS
..oil shale..	oil.shale	oil shale shale oil(s)	OIL SHALE

6.2 Fluxograma do Método SALF





6.2.2.10	FLUXOGRAMA 1 (Cont.)
<p>Verificar se cada um dos termos pertinentes é preferencial no tesouro DOE. .NÃO: (6.2.2.10.1) substituir pelo preferencial correspondente</p>	
<p>6.2.2.11 .SIM: Eliminar os termos pertinentes mais simples que entrem na composição de termos pertinentes mais complexos (por exemplo "fuel fabrication" e "fuel fabrication plants", o 1º deve ser eliminado).</p>	
<p>6.2.2.12 Verificar se cada um dos termos assinalados está incluído no dicionário histórico da área N13 (anexo 4) .NÃO: (6.2.2.14) gerar descritores sugeridos.</p>	
<p>6.2.2.13 .SIM: gerar descritores (os descritores gerados são listados na forma em que originalmente aparecem no tesouro DOE, i.e., se estiverem no plural ou se possuírem hífen devem ser listados assim).</p>	
<p>6.2.2.14 Gerar descritores sugeridos (listar na forma em que originalmente aparecem no tesouro DOE).</p>	
<p>6.2.2.15 Verificar para cada um dos descritores sugeridos se o indexador aceita o termo como descritor válido para o documento. .NÃO: Eliminar o descritor</p>	
<p>6.2.2.16 .SIM: Listar os descritores sugeridos aceitos (DSA) como válidos pelo indexador.</p>	
<p>F I M</p>	

6.3 Exemplo de Indexação pelo método SALF

Como ilustração, indexa-se pelo método proposto o primeiro documento da amostra. Mantem-se a mesma seqüência e numeração do fluxograma a fim de facilitar o acompanhamento.

1. PREPARAÇÃO DO TEXTO-BASE

RN 003091 DOCUMENTO 002 de 068 DOC. 01 DA AMOSTRA

1.1

TÍTULO:

BIOGAS LOW-COST FUEL.

RESUMO:

THE PROJECT THAT IS BEING DEVELOPED IN THE STATE OF PERNAMBUCO TO PRODUCE FUEL GAS FROM MANURES AND AGRICULTURAL WASTES IS PRESENTED. A VIABILITY ANALYSIS ON THE CONSTRUCTION OF 100 BIODIGESTERS TO PRODUCE BIOGAS IS DONE.

1.2

BIOGAS LOW COST FUEL.

PROJECT BEING DEVELOPED STATE PERNAMBUCO PRODUCE FUEL GAS MANURES AGRICULTURAL WASTES. VIABILITY ANALYSIS CONSTRUCTION BIODIGESTERS PRODUCE BIOGAS.

1.3

BIOGAS LOW COST FUEL.

PROJECT BEING DEVELOPED STATE PERNAMBUCO PRODUCE FUEL GAS MANURE AGRICULTURAL WASTE. VIABILITY ANALYSIS CONSTRUCTION BIODIGESTER PRODUCE BIOGAS.

1.4

BIOGAS LOW COST FUEL.

PROJECT DEVELOPED STATE FUEL GAS MANURE AGRICULTURAL WASTE VIABILITY ANALYSIS CONSTRUCTION BIODIGESTER BIOGAS.

1.5

T: BIOGAS LOW COST FUEL.

F₁: PROJECT DEVELOPED STATE FUEL GAS AGRICULTURAL WASTE

F₂: VIABILITY ANALYSIS CONSTRUCTION BIODIGESTER BIOGAS

2 FORMAÇÃO DE EXPRESSÕES

FRASE 02 - DOC. Nº 01 DA AMOSTRA

2.1		2.2		2.3		2.4		2.5		2.6		2.7		2.8		2.9		2.10		2.11		2.12	
VIABILITY	ANALYSIS	CONSTRUCTION	BIODIGESTER	BIOGAS	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
P_1	P_1	P_1	P_1	P_1	P_1	P_1	P_1	P_1	P_1	P_1	P_1	P_1	P_1	P_1	P_1	P_1	P_1	P_1	P_1	P_1	P_1	P_1	P_1
TRIPLETOS		DUPLÉTOS		SINGLETOS		DESCRITORES		DESCRITORES		DESCRITORES		DESCRITORES		DESCRITORES		DESCRITORES		DESCRITORES		DESCRITORES		DESCRITORES	
2.3		2.4		2.5		2.6		2.7		2.8		2.9		2.10		2.11		2.12		2.13		2.14	
VIABILITY ANALYSIS, CONSTRUCTION		VIABILITY ANALYSIS, CONSTRUCTION		VIABILITY ANALYSIS, CONSTRUCTION		VIABILITY ANALYSIS, CONSTRUCTION		VIABILITY ANALYSIS, CONSTRUCTION		VIABILITY ANALYSIS, CONSTRUCTION		VIABILITY ANALYSIS, CONSTRUCTION		VIABILITY ANALYSIS, CONSTRUCTION		VIABILITY ANALYSIS, CONSTRUCTION		VIABILITY ANALYSIS, CONSTRUCTION		VIABILITY ANALYSIS, CONSTRUCTION		VIABILITY ANALYSIS, CONSTRUCTION	
P_1, P_2		P_1, P_2		P_1, P_2		P_1, P_2		P_1, P_2		P_1, P_2		P_1, P_2		P_1, P_2		P_1, P_2		P_1, P_2		P_1, P_2		P_1, P_2	
CONSTRUCTION, BIODIGESTER, BIOGAS		CONSTRUCTION, BIODIGESTER, BIOGAS		CONSTRUCTION, BIODIGESTER, BIOGAS		CONSTRUCTION, BIODIGESTER, BIOGAS		CONSTRUCTION, BIODIGESTER, BIOGAS		CONSTRUCTION, BIODIGESTER, BIOGAS		CONSTRUCTION, BIODIGESTER, BIOGAS		CONSTRUCTION, BIODIGESTER, BIOGAS		CONSTRUCTION, BIODIGESTER, BIOGAS		CONSTRUCTION, BIODIGESTER, BIOGAS		CONSTRUCTION, BIODIGESTER, BIOGAS		CONSTRUCTION, BIODIGESTER, BIOGAS	
N		N		N		N		N		N		N		N		N		N		N		N	

INDEXAÇÃO MÉTODO SAIF

DESCRIPTORES	FREQ.	LOCALIZAÇÃO (TIT e/ou RES)
BIOGAS PROCESS	2	TIT (1) e RES (1)
COST	1	TIT
FUEL GAS	1	RES
MANURES	1	RES
AGRICULTURAL WASTES	1	RES
CONSTRUCTION	1	RES
BIODIGESTERS	1	RES

INDEXAÇÃO MANUAL

DESIGN; GASIFICATION; BIOMASS; EQUIPMENT; APPROPRIATE TECHNOLOGY; METHANE; FERMENTATION;

OBS: 1) nenhum descritor foi sugerido; 11) termos assinalados com asterisco (*) foram eliminados.

6.4 Testes de Avaliação

A amostra indexada pelas duas versões do método proposto - SALF-S e SALF-P, é submetida a uma análise comparativa em relação a indexação manual (MAN). Analisam-se os seguintes parâmetros:

- a) Exaustividade Média (\bar{E}) - número médio de descritores assinalados por documento, por tipo de indexação;
- b) Desempenho na Recuperação - avalia-se o desempenho das indexações SALF-S, SALF-P e MAN através dos índices de revocação e precisão.^{10, 11}
- c) Especificidade da Indexação - discute-se a capacidade das indexações de descrever um tópico precisamente;¹⁰
- d) Formação de Expressões - analisa-se a habilidade de cada tipo de indexação de assinalar descritores compostos, i.e., descritores formados por mais de uma palavra;
- e) Consistência - examina-se a concordância entre os três métodos de indexação; comparam-se os resultados com a literatura; examina-se também consistência dos descritores de maiores frequências.

Para se determinar o assunto das consultas usadas para avaliar a recuperação - as consultas-teste, escolheu-se aleatoriamente um documento da amostra. O conteúdo do título e do resumo deste documento serviu como "tema" para a formulação de um conjunto de 7 (sete) consultas-teste de níveis de especificidade variáveis. A seguir, submeteram-se as consultas-teste a cada uma das três indexações objeto da análise. Finalmente,

calcularam-se, a partir dos resultados obtidos, os índices de revocação e precisão para cada tipo de indexação.

A formulação das consultas-testes, assim como a seleção dos documentos pertinentes ao tema escolhido, foi realizada com a assistência do indexador especialista em energia. O tema, as estratégias de buscas e os números obtidos estão reunidos no anexo 6.

CITAÇÕES E NOTAS

- ¹ SALTON, G. Automatic indexing: a summary. In: EURIM 5 CONFERENCE, Versailles, France, 1982. Proceedings ..., ASLIB, 1982. p.66-77
- ² USA. DEPARTMENT OF ENERGY. Technical Information Center. Energy data base: Subject Thesaurus. Oak Ridge, Te., 1981 (DOE/TIC-7000-R5).
- ³ A necessidade do título e resumo em inglês não implica em nenhum trabalho extra, porque obrigatoriamente os documentos incluídos na FONTE devem ter resumos em inglês (ver capítulo 5, seção 5.1.2).
- ⁴ Termo utilizado por alguns autores em contraposição a "palavras de baixa entropia", i.e., palavras de alto conteúdo informativo.
ANDREEWSKI, A; RUAS, V. Indexação automática baseada em métodos lingüísticos e estatísticos e sua aplicabilidade à língua portuguesa. Rio de Janeiro, PUC-DI, 1982. 31p.
- ⁵ SCHIELE, J.; KURTEV, I.; RUCKENBAUER, E; STANIK, Z.; VARECKA, F.; SCHMID, H. INIS: Manual for online retrieval. Vienna, IAEA, 1984. 147p. (IAEA-INIS-17, rev.2).
- ⁶ LOVINS, J.B. Development of a stemming algorithm. Mech. Transl. and Comp. Linguistic II. 1968. p.22-31

⁷ Termos alternativos, sinônimos, quase-sinônimos, termos com grafia alternativa, etc. que são remetidos a termos preferenciais num tesouro. Por exemplo:

Fusion Fuels

USE TEHERMONUCLEAR FUELS

⁸ BORKO, H. Automatic indexing: a tutorial. SIGIR Forum (USA), 16 (3): 9-13, 1982.

⁹ Os termos "tripletos", "dupletos" e "singletos" estão registrados no dicionário Aurélio; são usados mais frequentemente em Física.

¹⁰ LANCASTER, F.W. Information retrieval systems: characteristics, testing and evaluation. New York, J. Willey, 1964, 222p.

¹¹ O índice de revocação indica a habilidade do sistema em recuperar documentos relevantes; é definido pela fórmula R/C , onde C é o total de documentos relevantes na base de dados, e R é o total de documentos relevantes recuperados. Já o índice de precisão mede a habilidade do sistema em não recuperar documentos irrelevantes; é expresso numericamente pela fórmula R/L , onde R é o total de documentos relevantes recuperados e L é o total de documentos recuperados.

CLEVERDON, C. Evaluation tests of information retrieval systems. J. Docum. 26 (1): 55-67, Mar. 1970.

7 RESULTADOS

Neste capítulo apresentam-se, em primeiro lugar, os resultados obtidos com a indexação da amostra. A listagem da amostra, apresentada na seção 7.1, gerada pelo sistema SUPRIR (ver seção 4.1.2.3 do capítulo 4) exhibe os seguintes itens:

- a) RN - nº do documento na base de dados FONTE;
- b) REFERÊNCIA - referência bibliográfica do documento;
- c) DESCRITORES - descritores assinalados manualmente;
- d) RESUMO - resumo do documento.

lista-se também os itens:

- e) INDEXAÇÃO SALF - descritores gerados pelo método SALF;
- f) DSA - descritores sugeridos aceitos.

Em seguida, analisam-se comparativamente as indexações manual e as obtidas pelas duas versões do método proposto, segundo os parâmetros exaustividade, desempenho na recuperação, especificidade, formação de expressões e consistência. O resumo dos resultados obtidos está nos quadros 1 e 2, no final do capítulo. O quadro 1, resumo dos resultados obtidos por tipo de indexação, exhibe, para cada tipo de indexação - manual da amostra (MAN_A), manual de toda área N13 (MAN_T), SALF-S e SALF-P da amostra e também para os descritores sugeridos aceitos (DSA), os seguintes itens:

- a) número de descritores assinalados;
- b) número médio de descritores assinalados por documento;
- c) número de descritores distintos⁴ assinalados;
- d) número médio de descritores distintos assinalados por documento;
- e) porcentagem de descritores retirados apenas do título;
- f) porcentagem de descritores retirados apenas do resumo;
- g) porcentagem de descritores retirados do título e do resumo simultaneamente;
- h) desempenho na recuperação (revocação e precisão)

O quadro 2 - resumo dos resultados obtidos por documento, apresenta, para cada documento da amostra, os seguintes itens:

- a) número de descritores da indexação manual;
- b) número de descritores da indexação SALF-S;
- c) número de descritores sugeridos;
- d) número de descritores sugeridos aceitos pelo indexador;
- e) número de descritores da indexação SALF-P.

No anexo 7, apresenta-se a estatística de ocorrência dos descritores por documento, tanto para a indexação manual quanto para as duas versões do método SALF.

7.1 Indexação da Amostra

Documento 002 de 068

RN 003091

REFERENCIA:

ANON. a. BIOGAS LOW-COST FUEL. a. BIOGAS COMBUSTIVEL DE BAIXO CUSTO. a. DIRIG. MUNIC. a. (MAIO-JUN 1981). V. 12(3) P. 12-13.

DESCRITORES:

BIOGAS PROCESS; COST; ANAEROBIC DIGESTION; MANURES; AGRICULTURAL WASTES; BRAZIL; BIOMASS; BIODIGESTERS;

RESUMO:

THE PROJECT THAT IS BEING DEVELOPED IN THE STATE OF PERNAMBUCO TO PRODUCE FUEL GAS FROM MANURES AND AGRICULTURAL WASTES IS PRESENTED. A VIABILITY ANALYSIS ON THE CONSTRUCTION OF 100 BIODIGESTERS TO PRODUCE BIOGAS IS DONE.

INDEXAÇÃO SALF:

BIOGAS PROCESS; COST; FUEL GAS; MANURES; AGRICULTURAL WASTES; CONSTRUCTION;
BIODIGESTERS

DSA: -

Documento 003 de 068

RN 002875

REFERENCIA:

GHOSH, S. (INSTITUTE OF GAS TECHNOLOGY, CHICAGO, IL (USA)). a. NOVEL BIOCONVERSION SYSTEMS AND INNOVATIVE BIOGASIFICATION PROCESS DESIGNS. a. 9500656. a. ASSET. a. (MAY 1983). V. 5(5) P. 19-24.

DESCRITORES:

DESIGN; GASIFICATION; BIOMASS; EQUIPMENT; APPROPRIATE TECHNOLOGY; FERMENTATION;
METHANE;

RESUMO:
 AN ANAEROBIC DIGESTION PROCESS FOR SOFT-DRINK BOTTLING INDUSTRY TO STABILIZE THE LOW-PH (2.3) AND HIGH-COD WASTES (12.000 TO 14.000 MG/L) IS DEVELOPED. THIS ALLOWS TO REDUCE THE INDUSTRY S SEWER CHARGES AND TO GENERATE METHANE TO DISPLACE PURCHASED FUELS.

INDEXAÇÃO SALT:

BIOCONVERSION; DESIGN; ANAEROBIC DIGESTION; INDUSTRY; METHANE

DSA:

Documento 005 de 068

RN 002873

REFERENCIA:

BEREZOVSKY, D.R. (COMPANHIA ENERGETICA DE SAO PAULO (BRASIL)). ASSESSORIA DE PESQUISA DE FONTES ENERGETICAS). THE POTENTIAL OF BIOMASS ENERGY GENERATION. SIMPOSIO SOBRE ALTERNATIVAS ENERGETICAS PARA TRANSPORTE E INDUSTRIA. SAO PAULO, SP, (BRASIL). AGO 1979. POTENCIAL DE GERACAO DE ENERGIA DE BIOMASSA. AV. 1 P. 70-99. 8950228. SECRETARIA DA INDUSTRIA, COMERCIO, CIENCIA E TECNOLOGIA, SAO PAULO (BRASIL). PROCEEDINGS ON ALTERNATIVE ENERGY SOURCES FOR TRANSPORTATION AND INDUSTRY - VOL. 1. ANAIS DO SIMPOSIO SOBRE ALTERNATIVAS ENERGETICAS PARA TRANSPORTE E INDUSTRIA - VOL. 1. AF-884. 1979. 669 P. 8950171.

DESCRIPTORES:

BIOMASS; BRAZIL; BIOCONVERSION; RESOURCES;

RESUMO:

DIFFERENT TYPES OF FUELS FROM BIOMASS ARE ANALYSED. A GENERAL VIEW OF BIOMASS USE, THE MAIN TYPES OF BIOCONVERSION, THE BIOMASS IN BRAZIL, THE ENERGY POTENTIAL OF BIOMASS IN SAO PAULO ARE PRESENTED. (A.A.).

INDEXAÇÃO SALT:

BIOMASS; BIOCONVERSION; BRAZIL

DSA:

Documento 006 de 068

RN 002872

REFERENCIA:

NIELSEN, M.J. (COMPANHIA DE SANEAMENTO DO PARANA (BRASIL)). UTILIZATION OF METHANE ORIGINATING IN THE BIOGAS PURIFICATION - PART 1. UTILIZACAO DO GAS METANO PROVENIENTE DA PURIFICACAO DO GAS DE ESGOTO - PARTE 1. PUBLICACAO REDUZIDA DE TRABALHO APRESENTADO DURANTE O 2 SUP(0) CONGRESSO BRASILEIRO DO GAS, REALIZADO NO RIO DE JANEIRO DE 3 A 7 DE OUTUBRO DE 1983. X. ENERG., FONTES ALTERN. AISSN 0100-6371. (JUL-SET 1983). V. 5(27-28) P. 12-20.

DESCRIPTORES:

METHANE; BIOGAS PROCESS; PURIFICATION;

RESUMO:

THE IMPORTANCE OF THE METHANE UTILIZATION IS SHOWN. SOME BASIC INFORMATION ABOUT THE BIOGAS IS ALSO GIVEN. AT THE END, SOME BIOGAS PURIFICATION PROCESSES ARE MENTIONED. (R.S.).

INDEXAÇÃO SAIF:

METHANE; PURIFICATION; BIOGAS PROCESS

DSA:

Documento 010 de 068

RN 002868

REFERENCIA:

REUSCH, D. APPARATUS WITH RECEPTACLE FOR DRY PURIFICATION, IN PARTICULAR OF BIOGAS. ADE 2/952,611/A/. 28 DEC 1979. 6 P.

DESCRITORES:

BIOGAS PROCESS;HYDROGEN SULFIDES;IRON HYDROXIDES;METHANE;PURIFICATION;
DESULFURIZATION;

RESUMO:

HYDROGEN SULPHIDE CONTAINED IN BIOGAS IS REMOVED WITH IRON HYDROXIDES. THE CONTAINER HAS A CUP-SHAPED FORM WITH AN INLET AND OUTLET FOR THE BIOGAS. SINCE THE GAS IS MOIST, WATER CONDENSES OUT AND COLLECTS AT THE BOTTOM OF THE CONTAINER WHERE IT IS LED OFF. THE INLET IS FITTED WITH A SCREEN TO PREVENT SPORADIC PRESSURE WAVES OF THE BIOGAS ENTERING THE CONTAINER.

INDEXAÇÃO SALF:

EQUIPMENT; PURIFICATION; BIOGAS PROCESS; HYDROGEN SULFIDES; IRON HYDROXIDES

DSA:

CONTAINERS

Documento 011 de 068

RN 002709

REFERENCIA:

LADEIRA, N.G. @INDUSTRIAL LIQUID FUEL. @SEMINARIO CONSERVACAO DE ENERGIA .
SAO PAULO, SP (BRASIL). 26-27 APR 1979. @COMBUSTIVEIS LIQUIDOS INDUSTRIAIS. @
P. 141-155. INSTITUTO BRASILEIRO DO GAS. PROCEEDINGS OF THE SEMINAR ENERGY
CONSERVATION . @ANAIS DO SEMINARIO CONSERVACAO DE ENERGIA . @f-883. @1979. 155
P. .

DESCRITORES:

LIQUID FUELS;RECOMMENDATIONS;SPECIFICATIONS;FUEL OILS;VISCOSITY;COMBUSTION;
 RESUMO:
 THE CHARACTERISTICS OF THE MAIN OILS FUEL USED IN BRAZIL AND THE SUITABLE
 USE OF THEM IN ORDER TO OBTAIN A GOOD COMBUSTION ARE PRESENTED. (A.R.).

INDEXAÇÃO SALF:

LIQUID FUELS; FUEL OILS; BRAZIL; COMBUSTION

DSA:

Documento 012 de 068

RN 002594

REFERENCIA:

VIAUD, P. TEISSEIRE, D. (UNIVERSITE DE DAKAR, DAKAR (SENEGAL)). @EUPHORBIA
 LATEX A POSSIBLE SOURCE OF HYDROCARBONS AND RUBBER. @X. @REV. GEN. CAOUTCH.
 PLAST. @ (1979). (593) P. 181-182, 184-185.

DESCRITORES:

HYDROCARBONS;RUBBERS;FEASIBILITY STUDIES;PLANTS;LATEX;ENERGY SOURCES;
 RESUMO:
 STUDIES ARE REPORTED ON THE HYDROCARBON CONTENT OF THE AERIAL PARTS OF
 EUPHORBIA TIRUCALLI. WORK IS ALSO IN PROGRESS ON E.BALSAMIFERA, WHICH
 CONTAINS ABOUT 10% RUBBER. THESE SPECIES GROW EASILY IN SENEGAL UNDER POOR
 CONDITIONS AND ARE OF POTENTIAL COMMERCIAL INTEREST.

INDEXAÇÃO SALF:

EUPHORBIA; LATEX; HYDROCARBONS; RUBBERS

DSA:

SENEGAL

Documento 013 de 068

RN 002492

REFERENCIA:

ANON. DA PROCESS TO OBTAIN GAS FROM NAPHTHA. @PROCESSO DE OBTENCAO DE GAS A PARTIR DA NAFTA. @ATUAL. CONS. NAC. PET. @ISSN 0100-1471. (MAY-JUN 1974). V. 4(36) P. 26-27.

DESCRITORES:

NAPHTHA; TOWN GAS; PILOT PLANTS; GASIFICATION; DESULFURIZATION; HYDROGENATION; CRG PROCESSES; GAS RECYCLE HYDROGENATION PROC; ODORIZATIION; EQUIPMENT; CATALYTIC REFORMING;

RESUMO:

A PROCESS TO OBTAIN GAS FROM NAPHTHA IS SUMMARIZED. THE DESULFURIZATION, PRODUCTION OF THE SO CALLED RICH GAS, THE CATALYTIC REFORMING THE CO CHEMICAL CONTENT REDUCTION, THE ENRICHMENT, THE HEAT RECOVERY FOR STEAM GENERATION, THE FINAL GAS AND THE ODORIZATIION, THAT ARE THE STAGES THE NAPHTHA IS SUBMITTED TO BEFORE THE GAS IS OBTAINED, ARE DESCRIBED. AT THE END, THE MASSINET SORCILLI PLANT WHICH BELONGS TO THE SAO PAULO GAS COMPANY, THAT PROVIDES WITH TOWN GAS THE POPULATION OF SAO PAULO STATE, IS CHARACTERIZED. (R.S.S.).

INDEXAÇÃO SALF:

NAPHTHA; DESULFURIZATION; PRODUCTION; CTALYTIC REFORMING; ODORIZATIION;
TOWN GAS

DSA:

REDUCTION; ENRICHMENT; HEAT RECOVERY; STEAM GENERATION

Documento 017 de 068

RN 002488

REFERENCIA:

ROUSSEAR, I. SHELEFF, G. MARCHAIM, U. (TECHNION - ISRAEL INSTITUTE OF TECHNOLOGY, HAIFA (ISRAEL)). @A SYSTEM FOR THE UTILIZATION OF AGRICULTURAL WASTES IN AN AGRO-INDUSTRIAL SETTLEMENT-KIBBUTZ AS A MODEL. @X. @RESOUR. RECOVERY AND CNSERV. @ (1979). V. 4 P. 159-168.

DESCRITORES:

AGRICULTURAL WASTES; ENERGY SUPPLIES; COMMUNITIES; METHANE;

RESUMO:

THIS ARTICLE DESCRIBES THE ENERGY POTENTIAL OF AGRICULTURAL WASTES IN ISRAEL. A SCHEME FOR UTILIZING AGRICULTURAL, DOMESTIC AND AGRO-INDUSTRIAL WASTES IN A KIBBUTZ IS PRESENTED. THE POSSIBILITY OF SUPPLYING THE ENERGY NEEDS OF A KIBBUTZ OF 1000 PERSONS USING BIOGAS IS CONSIDERED.

INDEXAÇÃO SALT:

AGRICULTURAL WASTES; BIOGAS PROCESS

DSA:

INDUSTRIAL WASTES; ISRAEL

Documento 020 de 068

RN 002485

REFERENCIA:

HILLEN, L.W. WAKE, L.V. WARREN, D.R. (AERONAUTICAL LABORATORY, MELBOURNE, VICTORIA 3001 (AUSTRALIA)). HYDROCARBON FUELS FROM PLANTS. OX. FUEL. ISSN 0016-2361. (1980). V. 59 P. 446-447.

DESCRIPTORES:

ALGAE; HYDROCARBONS; BIOMASS PLANTATIONS; TESTING; SYNTHETIC FUELS;

RESUMO:

THIS ARTICLE PRESENTS A STUDY OF ALGAE, BOTRYOCOCCUS BRAUNII, CONTAINING HYDROCARBONS HARVESTED IN AN AUSTRALIAN LAKE. TESTS WERE MADE TO DISTILL THE OIL EXTRACTED FROM THIS ALGAE.

INDEXAÇÃO SALT:

HYDROCARBONS; PLANTS; ALGAE; OILS

DSA:

LAKES

Documento 022 de 068

RN 002483

REFERENCIA:

YANG, P. Y. HUANG, C. J. GITLIN, H. M. (HAWAII UNIV., MANA, HONOLULU (USA)).
 AMERICAN SOCIETY OF AGRICULTURAL ENGINEERS (USA). ENHANCED ENERGY
 PRODUCTION AND WATER REUSE FOR SWINE OPERATION IN HAWAII. ASAE AND CANADIAN
 SOCIETY OF AGRICULTURAL ENGINEERING SUMMER MEETING. WINNIPEG, CANADA. 24
 JUN 1979. ASAE-TP--19. 21979. VP. X. .

DESCRIPTORES:

BIOGAS PROCESS; METHANE; ALGAE; MANURES; ENERGY SUPPLIES; ANAEROBIC DIGESTION;

RESUMO:

ENHANCEMENT OF BIOGAS PRODUCTION BY ENRICHMENT OF ALGAE GROWN FROM
 ANAEROBIC DIGESTION OF SWINE MANURE IS FEASIBLE. AN ADDITION OF 30% ALGAL
 TOTAL VOLATILE SOLIDS IN THE FEED WILL INCREASE GAS YIELD BY 17% COMPARED
 TO THE CONVENTIONAL ANAEROBIC DIGESTION PROCESS WITHOUT ALGAL ADDITION. A
 CONCENTRATION OF NH SUB(3) -N OF 80-100 MG/L WILL PROVIDE THE MAXIMUM
 SPECIFIC-GROWTH RATE OF ALGAL GROWTH WHICH CAN BE FURTHER DIGESTED TO
 METHANE GAS.

INDEXAÇÃO SALF:

PRODUCTION; BIOGAS PROCESS; ALGAE; ANAEROBIC DIGESTION; MANURES; METHANE

DSA:

SWINE; HAVAIT; ENRICHMENT; GAS YIELD; GROWTH

Documento 026 de 068

RN 002479

REFERENCIA:

CALVIN, M. (CALIFORNIA UNIV. (USA)). BIOMASS UTILIZATION AS AN ENERGY
 FACTOR. 7. ENCONTRO NACIONAL DOS PRODUTORES DE ACUCAR. CAMPOS, RJ (BRASIL).
 1979. DA UTILIZACAO DA BIOMASSA COMO FATOR ENERGETICO. AF-868. 21979. 18 P. X.

DESCRIPTORES:

FUEL SUBSTITUTION; ETHANOL FUELS; BIOMASS; RENEWABLE ENERGY SOURCES;

RESUMO:
THE PROBLEM OF LIQUID FUEL SUBSTITUTION BY ALCOHOL IS ANALYSED. THE ADVANTAGES OF SAO FRANCISCO VALLEY WATERING, THE IMPACT OF NEW TECHNOLOGIES AND LATEX RICH PLANTS ARE ANALYSED AS WELL. (A.A.).

INDEXAÇÃO SALF:

BIOMASS; LIQUID FUELS; FUEL SUBSTITUTION; LATEX; PLANTS

DSA:

ALCOHOL FUELS

Documento 027 de 068

RN 002478

REFERENCIA:

BUNGAY, H.R. (RENSSELAER POLYTECHNIC INST., TROY, NY (USA)). RESEARCH IN U.S.A. ABOUT FUELS FROM BIOMASS FERMENTATION. 27. ENCONTR. NACIONAL DOS PRODUTORES DE ACUCAR. CAMPOS, RJ (BRASIL). 1979. PESQUISA NOS ESTADOS UNIDOS SOBRE COMBUSTIVEIS DERIVADOS DE FERMENTAÇÃO DE BIOMASSA. 2F-866. 2 1979. 14 P. 5466000.

DESCRIPTORIOS:

BIOMASS; ETHANOL; CELLULOSE;

RESUMO:

THE RESEARCHES IN THE FIELD OF FERMENTATION IN U.S.A. ARE PRESENTED. CELLULOSE HYDROLYSIS AND GLUCOSE FERMENTATION TO PRODUCE ETHANOL AS WELL AS ITS COSTS ARE ANALYSED. (A.A.).

INDEXAÇÃO SALF:

BIOMASS; FERMENTATION; CELLULOSE; ETHANOL; COST

DSA:

USA; HYDROLYSIS; GLUCOSE

Documento 028 de 068

RN 002360

REFERENCIA:
 SCHWARZ, H. (FUNDICAO TUPY S.A., JOINVILLE (BRASIL)). CENTRO DE PESQUISA E DESENVOLVIMENTO). SOME CONSIDERATIONS RELATED TO AN INDUSTRIAL CHARCOAL GASIFICATION PLANT. CONSIDERACOES RELATIVAS A UMA INSTALACAO DE GASOGENIOS INDUSTRIAIS A CARVAO VEGETAL. af-473. ND . 13 P. 8950164.
 DESCRITORES:

CHARCOAL; COAL GAS; GASIFICATION; FUEL GAS; COAL GASIFICATION PLANTS;

RESUMO:

AT THE OUTSET THE PROCESSES OF COAL GAS GENERATION ARE REVIEWED WITH SPECIAL REFERENCE TO CHARCOAL, USING AIR AND WATER VAPOR MIXTURE FOR THE GASIFICATION. A CHARCOAL GAS GENERATOR IN CONTINUOUS ACTIVITY FOR 25 YEARS AT FUNDICAO TUPY S.A. IS DESCRIBED. THE LOW-BTU GAS OBTAINED IS USED IN INDUSTRIAL FURNACES FORMALLEABILIZING HEAT TREATMENT. THE BEST CHARACTERISTIC OF CHARCOAL AND A SHORT ANALYSIS OF THE SITUATION OF THIS LOW-BTU GAS AS AN ENERGETIC ALTERNATIVE ARE GIVEN. (AUTHOR).

INDEXAÇÃO SALF:

CHARCOAL; GASIFICATION; PLANTS; COAL GAS

DSA:

AIR; WATER VAPOURS; GAS GENERATOR; FURNACES; LOW BTU GAS

Documento 031 de 068

RN 002168

REFERENCIA:

FEDERACAO DAS INDUSTRIAS DO ESTADO DO RIO DE JANEIRO (BRAZIL) INSTITUTO DE DESENVOLVIMENTO ECONOMICO E GERENCIAL, RIO DE JANEIRO (BRAZIL). PROCEEDINGS OF THE 2. SEMINAR ON BIOMASS AS ENERGY IN INDUSTRY - VOL 2. RESIDUES ENERGY UTILIZATION. a2. SEMINARIO SOBRE BIOMASSA COMO ENERGIA NA INDUSTRIA. RIO DE JANEIRO, RJ (BRAZIL). APR 1982. ANAIS DO 2. SEMINARIO SOBRE BIOMASSA COMO ENERGIA NA INDUSTRIA - VOL 2. APROVEITAMENTO ENERGETICO DE RESIDUOS. af-823. 1982. 178 P. 8950191 8950192.

DESCRITORES:

REVIEWS:

RESUMO:

THE USE OF WOOD AS ENERGY SOURCE, THE TECHNIQUES OF GASIFICATION, PROBLEMS PRESENTED IN BIODIGESTION CONSIDERING THE RAW-MATERIAL USED, THE USE OF BIOMASS AS ENERGY ALTERNATIVE IN A CEMENT INDUSTRY ARE DISCUSSED. (A.R.).

INDEXAÇÃO SALF:

BIOMASS; INDUSTRY; WOOD; ENERGY SOURCES; GASIFICATION

DSA:

PROCEEDINGS; CURRENT

Documento 032 de 068

RN 002093

REFERENCIA:

SCHWERS, F. ANHYDROCARBONS PARTIALLY OXIDATED AND ITS USE AS FUEL. a1.
 CONGRESSO NACIONAL DE CARBURANTES. RIO DE JANEIRO, RJ (BRAZIL). 1942. a05
 HIDROCARBURETOS PARCIALMENTE OXIDADOS E SEU EMPREGO COMO CARBURANTES. a
 F-809. a1944. 37 P.

DESCRITORES:

HYDROCARBONS; OXIDATION; ETHANOL; FERMENTATION; POLYMERIZATION; ETHYLENE;
 REVIEWS; COMPARATIVE EVALUATIONS;

RESUMO:

A REVIEW OF THE CHEMICAL PREPARATIONS OF ETHANOL AND ETHYLENE ARE PRESENTED. THE POLYMERIZATION OF GASEOUS ETHYLENE AND ITS TRANSFORMATION INTO LIQUID HYDROCARBON AT ROOM TEMPERATURE IS SHOWN. (A.R.).

INDEXAÇÃO SALF:

HYDROCARBONS; REVIEWS; ETHANOL; ETHYLENE; POLYMERIZATION

DSA:

CHEMICAL PREPARATION

Documento 035 de 068

RN 002090

REFERENCIA:

PIRES, E.B. (INSTITUTO DE PESQUISAS DA MARINHA, RIO DE JANEIRO (BRAZIL)). a
 BIOGAS - INSTITUTO DE PESQUISAS DA MARINHA. a1. SIMPOSIO BRASILEIRO DE
 BIOGAS. BRASILIA, DF (BRAZIL). 23-26 MAR 1981. aBIOGAS - INSTITUTO DE
 PESQUISAS DA MARINHA. aP. 101-112. X. ANON. EMBRATER, BRASILIA (BRAZIL).
 PROCEEDINGS OF THE 1. BRAZILIAN SYMPOSIUM OF BIOGAS. aANAIS DO 1. SIMPOSIO
 BRASILEIRO DE BIOGAS. aF-746. a1983. 148 P. 8956058.

DESCRITORES:

BIOGAS PROCESS;BIOMASS;PERFORMANCE;RESEARCH PROGRAMS;PLANTS;FABRICATION;
 EQUIPMENT;

RESUMO:

RESEARCHES DEVELOPED AT INSTITUTO DE PESQUISAS DA MARINHA, RELATED TO THE
 BIOGAS PROGRAM ARE PRESENTED. THE MANUFACTURING OF A BIODIGESTOR AND ITS
 IMPROVEMENT ARE DESCRIBED. (A.R.H.).

INDEXAÇÃO SALLF:

BIOGAS PROCESS; RESEARCH PROGRAMS; FABRICATION; BIODIGESTERS

DSA:

Documento 038 de 068

RN 002087

REFERENCIA:

GORGATTI NETO, A. (EMPRESA BRASILEIRA DE PESQUISA AGROPECUARIA, CORONEL
 PACHECO. CENTRO DE GADO DE LEITE). aRESEARCH PROGRAM OF EMBRAPA IN THE
 ENERGY FIELD. a1. SIMPOSIO BRASILEIRO DE BIOGAS. BRASILIA, DF (BRAZIL).
 23-26 MAR 1981. aPROGRAMA DE PESQUISA DA EMBRAPA NA AREA DE ENERGIA. aP.
 55-69. 8950206. ANON. EMBRATER, BRASILIA (BRAZIL). PROCEEDINGS OF THE 1.
 BRAZILIAN SYMPOSIUM OF BIOGAS. aANAIS DO 1. SIMPOSIO BRASILEIRO DE BIOGAS. a
 F-746. a1983. 148 P. 8956058.

DESCRITORES:

BIOGAS PROCESS;ENERGY SOURCES;RESEARCH PROGRAMS;AGRICULTURE;RURAL AREAS;
 ETHANOL PLANTS;BRAZIL;

RESUMO:
FROM THE PROGRAM OF EMBRAPA (BRAZIL), IN THE AREA OF ENERGY, THE ESTABLISHMENT OF PROJECTS FOR BIODIGESTORS AND DEHYDRATED ETHANOL MICRODISTILLERIES ARE PRESENTED. (A.R.H.).

INDEXAÇÃO SALF:

RESEARCH PROGRAMS; BRAZIL; BIODIGESTERS; ETHANOL

DSA:

Documento 043 de 068

RN 002082

REFERENCIA:

CORREA, A.N. (SANTA CATARINA UNIV., FLORIANOPOLIS (BRAZIL)). FACULDADE DE AGRONOMIA). ADEVELOPMENT OF THE BIOGAS PROGRAM THROUGH SIBRATER. 1. SIMPOSIO BRASILEIRO DE BIOGAS. BRASILIA, DF (BRAZIL). 23-26 MAR 1981. ADEVOLUCAO DO PROGRAMA DE BIOGAS ATRAVES DO SIBRATER. p. 7-10. X. ANON. EMGRATER, BRASILIA (BRAZIL). PROCEEDINGS OF THE 1. BRAZILIAN SYMPOSIUM OF BIOGAS. ANAIS DC 1. SIMPOSIO BRASILEIRO DE BIOGAS. af-746. a1983. 148 P. 8956058.

DESCRIPTORES:

BIOGAS PROCESS; NATIONAL PROGRAM PLANS; RURAL AREAS; REVIEWS; BRAZIL;

RESUMO:

A REVIEW ON THE DEVELOPMENT OF THE BIOGAS PROGRAM IN BRAZIL AND THE ROLE OF THE BIODIGESTORS TO THE RURAL POPULATION ARE PRESENTED. (A.R.H.).

INDEXAÇÃO SALF:

BIOGAS PROCESS; REVIEWS; BRAZIL BIODIGESTER; RURAL POPULATIONS

DSA:

Documento 047 de 068

RN 001844

REFERENCIA:

EMBRATER, BRASILIA (BRAZIL). PROCEEDINGS OF THE 1. BRAZILIAN SYMPOSIUM OF BIOGAS. 21. SIMPOSIO BRASILEIRO DE BIOGAS. BRASILIA, DF (BRAZIL). 23-26 MAR 1981. ANAIS DO 1. SIMPOSIO BRASILEIRO DE BIOGAS. AF-746. 21983. 148 P. 8956058.

DESCRITORES:

METHANE;

RESUMO:

THE ROLE OF THE BIODIGESTORS IN RURAL AREAS, THE DEVELOPMENT OF KINDS OF BIODIGESTORS, ITS STANDARDIZATION AND THE NACIONAL PROGRAMS OF ENERGY RELATED TO THIS AREA ARE PRESENTED. (A.R.H.).

INDEXAÇÃO SALF:

BIOGAS PROCESS; BIODIGESTORS; RURAL AREA; STANDARDIZATION

DSA:

PROCEEDINGS

Documento 049 de 068

RN 001422

REFERENCIA:

THOMPSON, L.H. ENERGY FROM SLUDGE DIGESTION. 29 P. THOMPSON, L.H. CHEMICAL ENGINEERING IN A HOSTILE WORLD. ENGLAND. INSTITUTION OF CHEMICAL ENGINEERS. 1977. 1 V.

DESCRITORES:

ANAEROBIC DIGESTION; SEWAGE SLUDGE; METHANE; FERMENTATION; ENERGY SOURCES; OPTIMIZATION;

RESUMO:

THE ANAEROBIC DIGESTION OF SEWAGE SLUDGE WAS EVOLVED AS A PROCESS ABOUT 65 YEARS AGO, PRIMARILY AS A METHOD OF CONDITIONING THE SLUDGE PRIOR TO DISPOSAL OR UTILISATION. INCREASINGLY, AND PARTICULARLY SINCE FOSSIL FUELS HAVE BEEN SEEN TO BE OF FINITE QUANTITY THE METHANE BY-PRODUCT OF THE PROCESS HAS ASSUMED IMPORTANCE. MORE ATTENTION TO OPTIMISING AND EXTENDING THE USE OF SLUDGE DIGESTION IS NOW SUGGESTED. (AUTHOR).

INDEXAÇÃO SALF:

ANAEROBIC DIGESTION; SUVAGE SLUDGE; FOSSIL FUELS; METHANE; BY-PRODUCTS

DSA:

Documento 050 de 068

RN 000829

REFERENCIA:

RAUPP, A.A.A. CRDEIRO, D.S. PETRINI, J.A. PORTO, M.P. (EMPRESA BRASILEIRA DE PESQUISA AGROPECUARIA, PELOTAS (BRAZIL)) SANTOS FILHO, B.G. (UNIVERSIDADE FEDERAL DE PELOTAS (BRAZIL)) DIAS, J.M.C.S. (EMPRESA BRASILEIRA DE PESQUISA AGROPECUARIA, BRASILIA (BRAZIL)). SWEET SORGHUM AND SUGARBEET BIOMASS FOR THE PRODUCTION OF ALCOHOL AT RIO GRANDE DO SUL, BRAZIL. 2. BRAZILIAN ENERGY CONGRESS. RIO DE JANEIRO, RJ (BRAZIL). 6-10 APR 1981. ASCRGO, SACARINO E BETERRABA ACUCAREIRA BIOMASSAS PARA A PRODUCAO DE ALOOL NO ESTADO DO RIO GRANDE DO SUL. AP. 539-550. 8956021 8956024 8956022. RIO DE JANEIRO UNIV. (BRAZIL). COORDENACAO DOS PROGRAMAS DE POS-GRADUACAO DE ENGENHARIA CLUBE DE ENGENHARIA, RIO DE JANEIRO (BRAZIL). PROCEEDINGS OF THE 2. BRAZILIAN ENERGY CONGRESS. ANAIS DO SEGUNDO CONGRESSO BRASILEIRO DE ENERGIA. AF-246. A1981. 1246 P. 5529350 8956017.

DESCRIPTORES:

BRAZIL; SORGHUM; CROPS; ETHANOL; BEETS; BIOMASS; ECONOMIC DEVELOPMENT; PRODUCTION;

RESUMO:

THE AGROCLIMATIC ZONES OF RS INDICATE THAT THE AREA FOR GROWING ECONOMICALLY SUGAR CANE CROP IS RESTRICTED. HOWEVER, THE CLIMATE AND SOIL CONDITIONS IN THE STATE MAKE POSSIBLE THE ECONOMIC DEVELOPMENT OF SWEET SORGHUM CROP TOGETHER WITH SUGARBEET CROP, FOR A BETTER UTILIZATION OF ETHANOL DISTILLERIES. EMBRAPA/UEPAE-PELOTAS. PROPOSE A PROCESSING SYSTEM THAT INCLUDES SUGARBEET, AND SWEET SORGHUM GRAINS AND STALKS, WHEREBY THE ETHANOL PRODUCTION WOULD BE AROUND 6,000 L/HA. (AUTHOR).

INDEXAÇÃO SALF:

SORGHUM; PRODUCTION; BRAZIL; CROPS; ECONOMIC DEVELOPMENT; ETHANOL; BIOMASS

DSA:

SUGAR BEET; ALCOHOLS; SUGAR CANE; CLIMATES

REFERENCIA:

FILGUEIRAS, G. (ELETROBRAS, RIO DE JANEIRO (BRAZIL)). ETHANOL FROM BEET AT THE SOUTH STATES OF BRAZIL. 2. BRAZILIAN ENERGY CONGRESS. RIO DE JANEIRO, RJ (BRAZIL). 6-10 APR 1981. ETHANOL DE BETERRABA NOS ESTADOS DO SUL DO BRASIL. 2P. 527-537. 8956020. RIO DE JANEIRO UNIV. (BRAZIL). COORDENACAO DOS PROGRAMAS DE POS-GRADUACAO DE ENGENHARIA CLUBE DE ENGENHARIA, RIO DE JANEIRO (BRAZIL). PROCEEDINGS OF THE 2. BRAZILIAN ENERGY CONGRESS. ANAIS DO SEGUNDO CONGRESSO BRASILEIRO DE ENERGIA. 2P-246. 21981. 1246 P. 5529350 8956017.

DESCRITORES:

BRAZIL; ETHANOL; BEETS; POWER DEMAND; AGRICULTURE; CROPS; COST BENEFIT ANALYSIS;

RESUMO:

THE BRAZILIAN ENERGETICAL CRISIS MAY COMPEL ITS STATES TO BE SELF-SUFFICIENT IN ENERGY. ETHANOL WAS CHOSEN AS THE LIQUID CARBURANT IN BRAZIL AND HAVE RELEGATED SOUTH STATES, MAINLY SANTA CATARINA AND RIO GRANDE DO SUL, TO AN INFERIORITY LEVEL, DUE TO THE CLIMATE. THE SUCCESS IN THE BEET-SUGAR PLANTATIONS MAY SURPASS THIS DEFICIENCY WITH ECONOMICAL ADVANTAGES FOR THE LOCAL AGRICULTURES, SINCE THEY COULD MAKE GOOD USE OF THE PERIOD IN WHICH THE FIELD WOULD STAY WITHOUT ANOTHER CROP. (AUTHOR).

INDEXAÇÃO SAMP:

ETHANOL; BRAZIL; AGRICULTURE; BEETS; CROPS

DSA:

CLIMATES; SUGAR BEETS

REFERENCIA:

DIAS, C.A.B. (CLUBE DE ENGENHARIA, RIO DE JANEIRO (BRAZIL)). PRELIMINARY ANALYSIS ON THE POSSIBILITY FOR THE VINASSE UTILIZATION AS ENERGETIC RESOURCE. 2. BRAZILIAN ENERGY CONGRESS. RIO DE JANEIRO, RJ (BRAZIL). 6-10 APR 1981. ANALISE PRELIMINAR DA POSSIBILIDADE DE UTILIZACAO DO VINHOTO COM RECURSO ENERGETICO. P. 511-526. 8956017. RIO DE JANEIRO UNIV. (BRAZIL). COORDENACAO DOS PROGRAMAS DE POS-GRADUACAO DE ENGENHARIA CLUBE DE ENGENHARIA, RIO DE JANEIRO (BRAZIL). PROCEEDINGS OF THE 2. BRAZILIAN ENERGY CONGRESS. ANAIS DO SEGUNDO CONGRESSO BRASILEIRO DE ENERGIA. AF-246. 21981. 1246 P. 5529350 8956017.

DESCRITORES:

ETHANOL; METHANE; CARBON DIOXIDE; METHANOL; BY-PRODUCTS;

RESUMO:

ONE OF THE PROBLEMS ORIGINATED BY THE NATIONAL PROGRAM OF ALCOHOL IS CONSIDERED WHICH RELATED TO THE VINASSE AND ITS ENVIRONMENT IMPLICATIONS. A PRELIMINARY ANALYSIS WAS MADE WHICH RELATES TO THE POSSIBILITY OF ITS CONDITIONING AS PER AN ANAEROBIC PROCESS AND POSTERIOR UTILIZATION OF RESULTING GASES. METHANE WOULD BE USED AS AN ENERGETIC ELEMENT AND THE CARBON DIOXIDE AS RAW MATERIAL FOR A PROCESS WHICH WOULD HAVE, AS A RESULT, THE PRODUCTION OF METHANOL STARTING FROM SYNTHESIS GAS. (AUTHOR).

INDEXAÇÃO SALT:

NATIONAL ALCOHOL PROGRAM; METHANE; CARBON DIOXIDE; PRODUCTION; METHANOL

DSA:

ENVIRONMENT

Documento 053 de 068

RN 000826

REFERENCIA:

ASSUMPCAO, R.M.V. CAMARGO, A.A.S. (INSTITUTO DE PESQUISAS TECNOLOGICAS, SAO PAULO (BRAZIL)). CENTRO TECNICO EM CELULOSE E PAPEL). CONSIDERATIONS ON THE USE OF WOOD GAS AS FUEL FOR BOILERS. 2. BRAZILIAN ENERGY CONGRESS. RIO DE JANEIRO, RJ (BRAZIL). 6-10 APR 1981. CONSIDERACOES SOBRE A UTILIZACAO DE GAS DE MADEIRA COMO COMBUSTIVEL EM CALDEIRAS. AP. 323-333. 8950021. RIO DE JANEIRO UNIV. (BRAZIL). COORDENACAO DOS PROGRAMAS DE POS-GRADUACAO DE ENGENHARIA CLUBE DE ENGENHARIA, RIO DE JANEIRO (BRAZIL). PROCEEDINGS OF THE 2. BRAZILIAN ENERGY CONGRESS. ANAIS DO SEGUNDO CONGRESSO BRASILEIRO DE ENERGIA. AF-246. 1981. 1246 P. 5529350 8956017.

DESCRITORES:

PRODUCER GAS; FUEL OILS; FUEL GAS; COMPARATIVE EVALUATIONS; GASIFICATION; WOOD; BOILERS;

RESUMO:

A PRELIMINARY INVESTIGATION OF SOME WOOD GASIFICATION PROCESSES AS WELL AS THE RESULTS ONE CAN EXPECT FROM THE BURNING OF THIS LOW BTU GAS IN BOILERS ORIGINALLY DESIGNED TO BURN OIL ARE PRESENTED. BOTH ARE THEN COMPARED AGAINST THE DIRECT BURNING OF WOOD IN BOILERS PURPOSELY BUILT TO PERFORM THIS TASK. (AUTHOR).

INDEXAÇÃO SALF:

WOOD FUELS; FUEL GAS; BOILERS; GASIFICATION; OILS

DSA:

LOW BTU GAS

Documento 057 de 068

RN 000822

REFERENCIA:

ANJOS, J.R.S. DOS YIU, L.L. FRETZ, R. (INSTITUTO MILITAR DE ENGENHARIA, RIO DE JANEIRO (BRAZIL)). SECAO DE QUIMICA). CATALYTIC DEGRADATION OF VEGETABLE OILS. ADEGRADACAO CATALITICA DE OLEOS VEGETAIS. 8956008. 8901. TEC. PETROBRAS. CISSN 0006-6117. (APR-JUN 1981). V. 24(2) P. 139-147.

DESCRITORES:

SOYBEAN OIL;CRACKING;HYDROCARBONS;CHROMATOGRAPHY;INFRARED SPECTRA;
HYDROGENATION;PRODUCTION;

RESUMO:

THE FIRST RESULTS ARE PRESENTED OF THE CRACKING OF VEGETABLE OILS AIMING TO OBTAIN MIXTURES OF HYDROCARBONS SIMILAR TO DIESEL OIL. VAPORS OF RAW AND PREHYDROGENATED SOYBEAN OIL WERE PUT TOGETHER WITH OF A SOLID ACID (AL SUB(2) O SUB(3)) AND A BASIC ONE (MGO) IN A TUBULAR GLASS REACTOR AT 300,400 AND 500 SUP(O) C. THE LIQUID PRODUCTS FORMED WERE ANALYSED BY CHROMATCGRAPHY AND INFRARED SPECTROSCOPY. EFFICIENCIES OF 40% IN WEIGHT OF HYDROCARBONS SIMILAR TO DIESEL OIL INDICATE THAT THE SIMPLE HYDROGENATION OF THE ORIGINAL OIL IS THE MOST ADEQUATE PRE-TREATMENT OF THE VEGETABLE OILS FOR THIS PURPOSE. (AUTHGR).

INDEXAÇÃO SALF:

CRACKING; HYDROCARBON; SOYBEAN OILS; CHROMATOGRAPHY; EFFICIENCY; HYDROGENATION

DSA:

VEGETABLE OILS; VAPORS; SPECTROSCOPY

Documento 063 de 068

RN 000816

REFERENCIA:

ZATZ, J. (SAO PAULO UNIV. (BRAZIL)). INST. DE FISICA).APRODUCTION OF FUELS BY FLASH PYROLYSIS.APRODUCAO DE COMBUSTIVEIS POR PIROLISE RAPIDA.25662400.2 ATUAL. CONS. NAC. PET.aISSN 0100-1471. (NOV-DEC 1979). V. 11(69) P. 22-23.

DESCRITORES:

PYROLYSIS;ANAEROBIC DIGESTION;BIOGAS PROCESS;FUEL GAS;FUEL OILS;
ORGANIC WASTES;WASTE PROCESSING PLANTS;

RESUMO:

THE USE OF FLASH PYROLYSIS ON ORGANIC WASTES FOR PRODUCING FUEL OIL AND FUEL GAS IS DISCUSSED. (A.R.H.).

INDEXAÇÃO SALF:

PRODUCTION; PYROLYSIS; ORGANIC WASTES; FUEL OILS; FUEL GAS

DSA:

Documento 064 de 068

RN 000470

REFERENCIA:

BORGES, M.H. ALMEIDA, M.R. DE MAGALHAES, J.G.R. @THE CHARCOAL PRODUCTION IN BRAZIL. @A PRODUCAO DE CARVAO VEGETAL NO BRASIL. @F-28. @APR 1981. 38 P.

DESCRIPTORES:

BIOMASS;ECONOMICS;AGRICULTURAL WASTES;BRAZIL;WOOD;CARBONIZATION;CHARCOAL;
FORESTS;WOOD FUELS;

RESUMO:

THE IMPORTANCE OF CHARCOAL IN THE BRAZILIAN STEEL INDUSTRY-NOW USING NATIONAL TECHNOLOGY - AND ITS GROWING SHARE, AS WELL AS THAT OF OTHER WOOD PRODUCTS, IN THE COUNTRY S ENERGY CONSUMPTION STRUCTURE, ARE PRESENTED. CHARCOAL SOURCES AND CHARCOAL CARBONIZATION CHARACTERISTICS, AND THE SOCIOECONOMIC DATA CONCERNING ITS PRODUCTION ARE SHOWN. FUTURE PERSPECTIVES FOR A BETTER UTILIZATION OF THE ENERGETIC POTENTIAL OF EXISTING FOREST ARE ALSO PRESENTED. (E.G.).

INDEXAÇÃO SALF:

CHARCOAL; PRODUCTION; BRAZIL; WOOD; ENERGY CONSUMPTION; CARBONIZATION;
FORESTS

DSA:

STEEL INDUSTRY

Documento 066 de 068

RN 000464

REFERENCIA:

MACHADO, M.A.A.C. (CENIBRA, BELO HORIZONTE (BRAZIL)). @CHARCOAL, INEXPENSIVE AND PLENTIFUL ENERGY FOR THE INDUSTRY. @CARVAO VEGETAL, ENERGIA BARATA E ABUNDANTE PARA A INDUSTRIA. @8950036. @CULT. ENERG. - BIOMASSA. @ (JAN-FEB 1982). V. 1(1) P. 22-27.

DESCRIPTORES:

BIOMASS;BRAZILIAN ORGANIZATIONS;WOOD;METAL INDUSTRY;CARBONIZATION;
ECONOMICS;CHARCOAL;USES;WOOD FUELS;

RESUMO:
CONSIDERING THE SOLAR RADIATION OCCURRING IN TROPICAL COUNTRIES - PARTICULARLY BRAZIL - IT BECOMES ADVANTAGEOUS THE TRANSFORMATION OF FORESTIAL BIOMASS INTO FUEL. THE CHARCOAL AS A RENEWABLE ENERGY SOURCE IS STUDIED. THE PROCESSES WIDELY USED FOR CHARCOAL PRODUCTION AND ITS POTENTIAL USES ARE PRESENTED. (E.G.).

INDEXAÇÃO SALF:

CHARCOAL; INDUSTRY; BRAZIL; BIOMASS; PRODUCTION; RENEWABLE ENERGY SOURCE;
USERS

DSA:

Documento 067 de 068

RN 000156

REFERENCIA:

ASSUMPCAO, R.M.V. (INSTITUTO DE PESQUISAS TECNOLOGICAS, SAO PAULO (BRAZIL)). DIRECT GASIFICATION OF WOOD. GASEIFICACAO DIRETA DA MADEIRA. a. 3272680. aREV. BRAS. ENG. QUIM. a(SEP-DEC 1980). V. 4(3-4) P. 41-47.
DESCRITORES:

GASIFICATION; WOOD; GAS FUELS; BIOMASS;

RESUMO:

PROCESSES OF GASIFICATION OF WOOD ARE PRESENTED, TAKING INTO ACCOUNT THE VARIABLES WHICH AFFECT THEM. SEVERAL APPLICATIONS OF THE WOOD GAS OBTAINED ARE ALSO PRESENTED. (A.R.).

INDEXAÇÃO SALF:

GASIFICATION; WOOD; USERS

DSA:

7.2 Conteúdo Informativo do Título e Resumo

Os gráficos 1, 2 e 3 têm como finalidade mostrar quantitativamente o número de descritores que foram gerados pelo método SALF a partir do resumo, do título ou de ambos simultaneamente. Estes gráficos tentam mostrar o potencial informativo desses elementos do documento, tomando-se como referência a metodologia proposta.

A versão SALF-S da metodologia gerou cerca de 14% dos descritores somente a partir dos títulos e 65% a partir dos resumos; os outros 20% foram gerados a partir dos títulos e resumos simultaneamente. O fator de "riqueza" informativa dos resumos é, portanto, 4,6 vezes maior do que os dos títulos (gráfico 1); este número é sensivelmente menor do que o encontrado em estudos anteriores¹ - 12(doze). Isto se deve, provavelmente, ao fato dos resumos da amostra serem, na maior parte dos casos, curtos e apenas indicativos.

Exatamente 76,08% dos descritores DSA foram originados somente dos resumos da amostra, 17,30% somente dos títulos e 5,43% de ambos (gráfico 2). Os números para a versão SALF-P são os seguintes: 70% gerados somente do resumo, 15% somente do título e 14% de ambos (gráfico 3).

7.3 Exaustividade

Considerando a exaustividade como um parâmetro que

GRÁFICO 1

INDEXAÇÃO SALF-S - por centagem de descritores gerados a partir do resumo, do título ou de ambos simultaneamente.

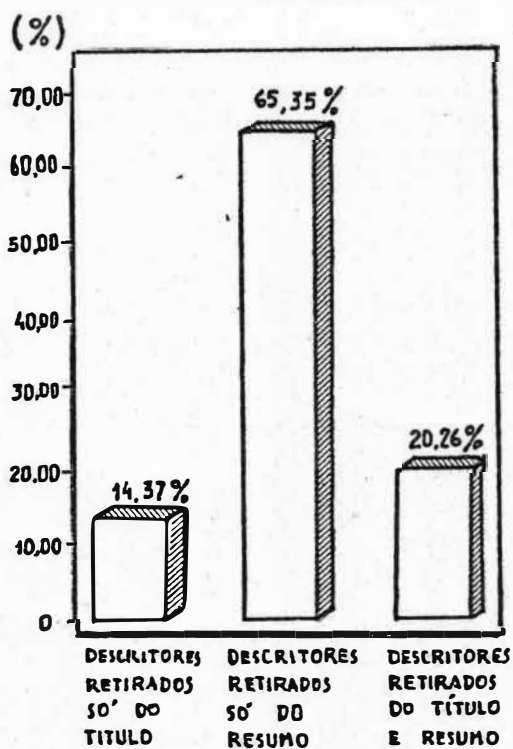


GRÁFICO 2

DESCRITORES SUGERIDOS ACEI
TOS (DSA) - porcentagem dos
DSA gerados a partir do re
sumo, do título ou de am -
bos simultaneamente

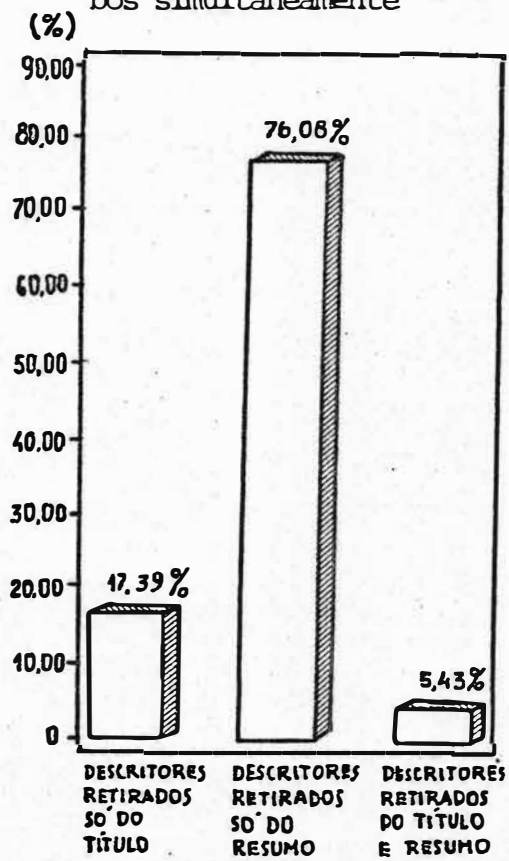
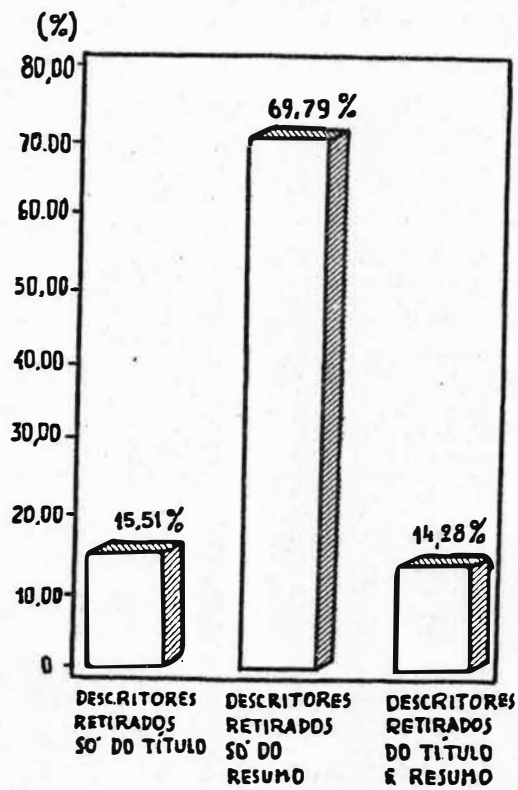


GRÁFICO 3

INDEXAÇÃO SALF-P - porcentagem de descritores gerados a partir do resumo, do título ou de ambos simultaneamente



mede a extensão com que todos os assuntos discutidos num particular documento são reconhecidos na operação e traduzidos para a linguagem do sistema, podemos expressar numericamente a exaustividade de uma indexação pelo número de descritores assinalados para um determinado documento. A exaustividade está diretamente ligada à revocação. Altos níveis de exaustividade tendem a gerar altos níveis de revocação; inversamente, baixos níveis de exaustividade levam a baixos níveis de revocação. Lancaster² afirma que a maioria dos usuários de sistemas de informação está interessada num alto índice de revocação, o que significa que as indexações devem apresentar um nível de exaustividade compatível com este interesse. "Revocação é obviamente a mais importante exigência de um usuário de um sistema de recuperação..."

O nível de exaustividade de uma indexação independe das propriedades da linguagem de indexação, sendo na verdade uma decisão administrativa que depende da filosofia adotada pela instituição que gerencia o sistema de recuperação². No caso aqui estudado, a linguagem de indexação é a mesma; o que se julga é a maneira mais econômica e mais conveniente para a instituição de operacionalizar esta linguagem.

Levando-se em conta estas considerações, o método é proposto em duas versões (SALF-S e SALF-P), que exibem entre si uma variação sensível de exaustividade. Isto significa que ter-se-á também alguma variação no índice de revocação entre as duas versões. Optar por uma das versões ou por ambas é uma possibilidade que o método oferece.

Determinaram-se os índices de exaustividade média^{2,3}

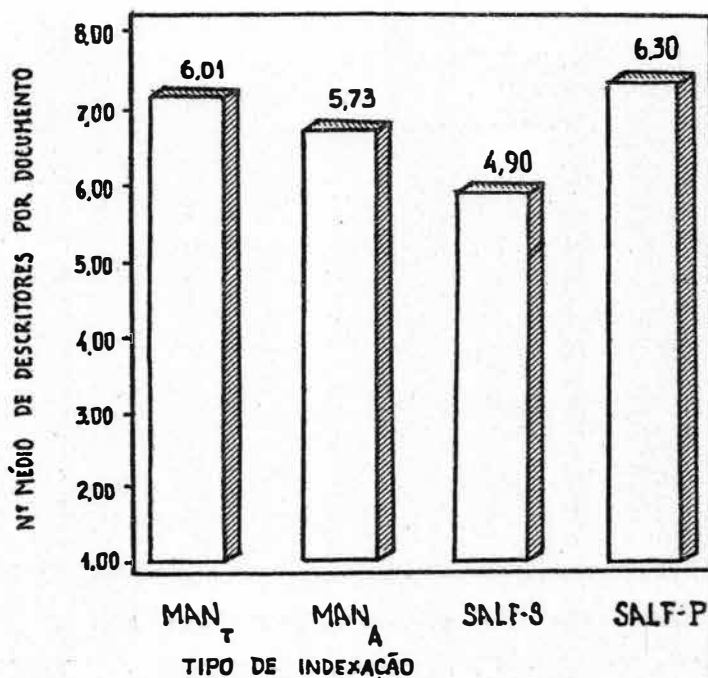
(\bar{E}) para a indexação manual de toda a área (\bar{E}_{MAN_T}), para a indexação manual da amostra (\bar{E}_{MAN_A}), para a indexação SALF-S (\bar{E}_{SALF-S}) e finalmente para a indexação SALF-P (\bar{E}_{SALF-P}). Os resultados estão representados no gráfico 4.

O índice \bar{E}_{MAN_A} , 5,73 descritores por documento, está compatível com o índice da área \bar{E}_{MAN_T} , que tem valor 6,01 desc/doc. O índice \bar{E}_{SALF-S} é igual a 4,90 desc/doc. que é cerca de 14% menor do que \bar{E}_{MAN_A} . Este resultado já era esperado e justifica-se pelo fato de haver um controle rígido sobre os termos candidatos a descritores - os descritores assinalados pelo método têm obrigatoriamente que pertencer ao dicionário histórico da área N13, i.e., têm que ter sido usados pelo menos uma vez para indexar documentos da área N13 da FONTE. A indexação SALF-P tem parte de seus descritores não submetidos a esse "filtro" e, portanto, exhibe um índice de exaustividade significativamente maior. O índice \bar{E}_{SALF-P} tem valor 6,30 desc/doc que é cerca de 9% maior que \bar{E}_{MAN_A} , e 22% maior que \bar{E}_{SALF-S} . (Para evitar defasagem do vocabulário da área N13, os descritores DSA devem ser incorporados a este vocabulário).

O gráfico 5 - "número médio de descritores distintos" assinalados por documento por indexação" - exhibe resultados parecidos com o gráfico anterior, excetuando-se a coluna de indexação manual de toda a área, cujo índice correspondente é agora menor do que o da indexação manual da amostra, quando no gráfico anterior era maior. Este comportamento está ligado ao fato de que, à medida que cresce o número de documentos indexados, diminui o aparecimento de descritores novos.

GRÁFICO 4

NÚMERO MÉDIO DE DESCRITORES AS
SINALADOS (exaustividade média)
POR TIPO DE INDEXAÇÃO



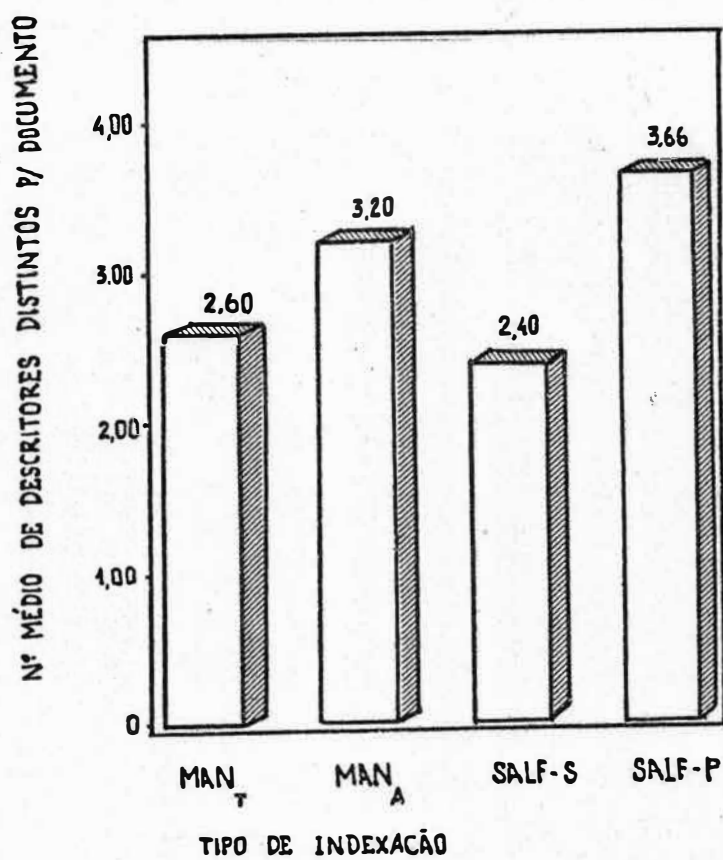
MAN_T - indexação manual de toda a área N13

MAN_A - indexação manual da amostra

SALF-S }
SALF-P } versões da indexação SALF da amostra

GRÁFICO 5

NÚMERO MÉDIO DE DESCRITORES DISTINTOS POR TIPOS DE INDEXAÇÃO



MAN_T - indexação manual de toda a área N13

MAN_A - indexação manual da amostra

SALF-S }
SALF-P } versões da indexação SALF da amostra

7.4 Teste de Recuperação

7.4.1 Análise das diferenças

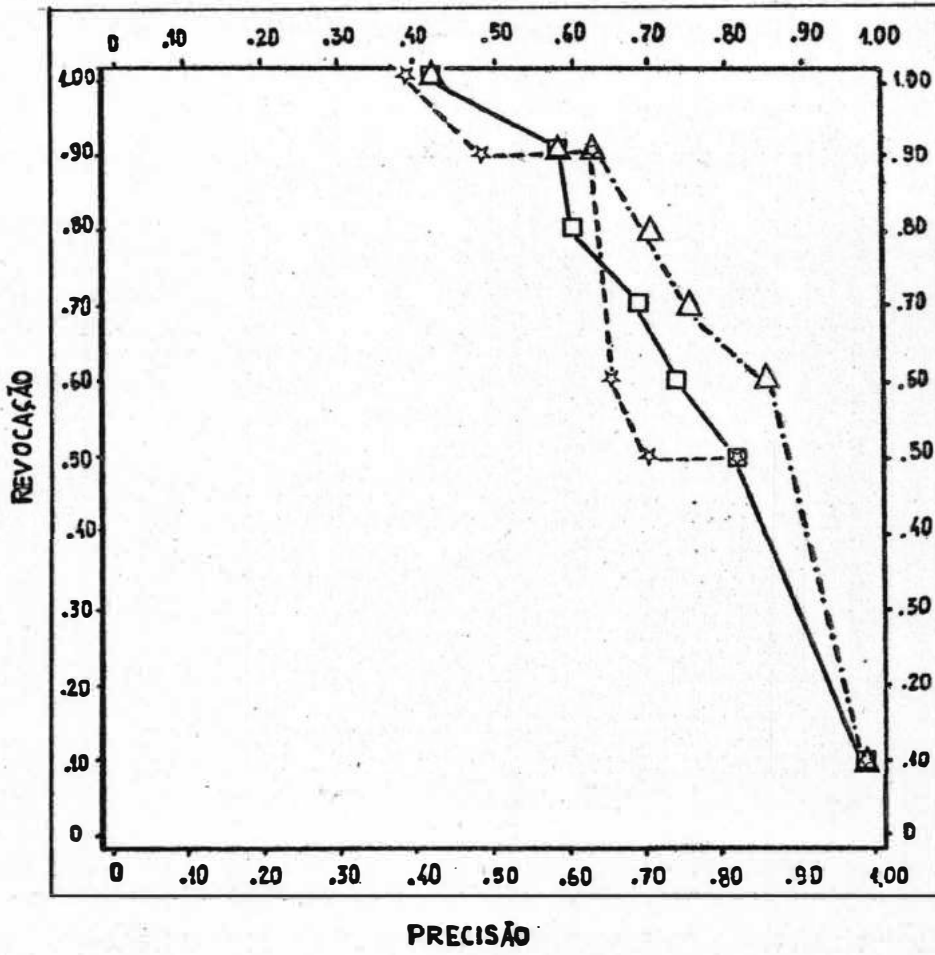
Karen Sparck Jones⁶ recomenda que, em testes de recuperação, diferenças inferiores a 10% não devem ser consideradas. Quando a média dos resultados dos testes de recuperação dos três métodos são comparadas, verifica-se que não há diferenças sensíveis e que, segundo a autora citada, elas são, para fins operacionais iguais. As indexações analisadas, em termos de consultas individuais, não apresentam diferenças significativamente profundas para que se possa caracterizar com facilidade qual delas exibe a "melhor" recuperação.

O gráfico "precisão x revocação" (gráfico 6), revela que os índices de recuperação da indexação SALF-P são ligeiramente superiores aos índices das outras duas indexações para todas as consultas submetidas, salvo para a faixa de revocação entre 1,00 e 0,90, em que a curva de desempenho coincide com a da indexação SALF-S, mantendo-se entretanto sempre superior aos índices da indexação manual.

Na faixa de revocação entre 0,90 e 0,75, os números da indexação manual são superiores aos da indexação SALF-S. Na faixa que vai da 0,50 a 0,10 as curvas das três indexações mantêm-se coincidentes. Nas outras regiões a indexação SALF-S apresenta uma ligeira "vantagem" sobre a indexação manual, que em média não é superior a 3%.

GRÁFICO 6

PRECISÃO POR REVOCAÇÃO



- ☆---☆ MAN - INDEXAÇÃO MANUAL
 - SALF-S
 - △---△ SALF-P
- } MÉTODO SALF

7.4.1.1 Problemas de avaliação

A linguagem de indexação usada para descrever os documentos da amostra é uma constante. A variável, na presente análise, é o modo de operação desta linguagem, ou seja, os procedimentos e critérios adotados para se atribuir descritores a documento, a fim de representá-lo adequadamente.

Na indexação manual, o conteúdo do texto é submetido a uma análise intelectual. O indexador faz para cada particular documento um julgamento, em que usa sua capacidade de discernimento. Além do mais, para o caso específico da amostra analisada, o indexador teve acesso ao texto completo, o que facilita bastante a determinação dos limites de abrangência do documento, assim como a análise da importância dos tópicos tratados nos documentos.

Por outro lado, na indexação automática, um algoritmo substitui o indexador e é aplicado repetida e indiscriminadamente para cada documento. No método proposto, assim como na maioria dos algoritmos de indexação automática, o título e o resumo são as fontes de onde são extraídos os descritores que representarão o documento.

O fato da análise de pertinência do documento, feita pelo especialista, estar fundamentada somente na leitura do título e do resumo de certa forma anula a "vantagem" do indexador de poder consultar o texto completo e tendência os resultados em favor dos métodos automatizados. Isto acontece porque um algoritmo indica o conteúdo informativo encontrado apenas no título e resumo de um particular documento, e o especialista julga a relevância deste documento analisando, igualmente, só o título e o resumo. A experiência do CIN na operação de sistema de recuperação, mostra que, numa busca real, o usuário tende a rejeitar os documentos recuperados relevantes que não apresentem indicação desta relevância no título e/ou no re-

sumo.⁵ Num teste, estes documentos seriam considerados como documentos recuperados impropriamente, mas se fossem indexados automaticamente, o algoritmo usado não assinalaria descritores para os tópicos relevantes que não se encontram no título e resumo, e assim esses documentos, provavelmente, não seriam recuperados.

O título e o resumo são, em última análise, os elementos que vão indicar o grau de relevância do documento ao usuário, não importando para sua apreciação o conteúdo do texto completo (que na maioria dos casos, não está disponível imediatamente). Este é, provavelmente, um dos fatores que contribuíram para que a indexação SALF obtivesse resultados ligeiramente melhores no teste de recuperação.

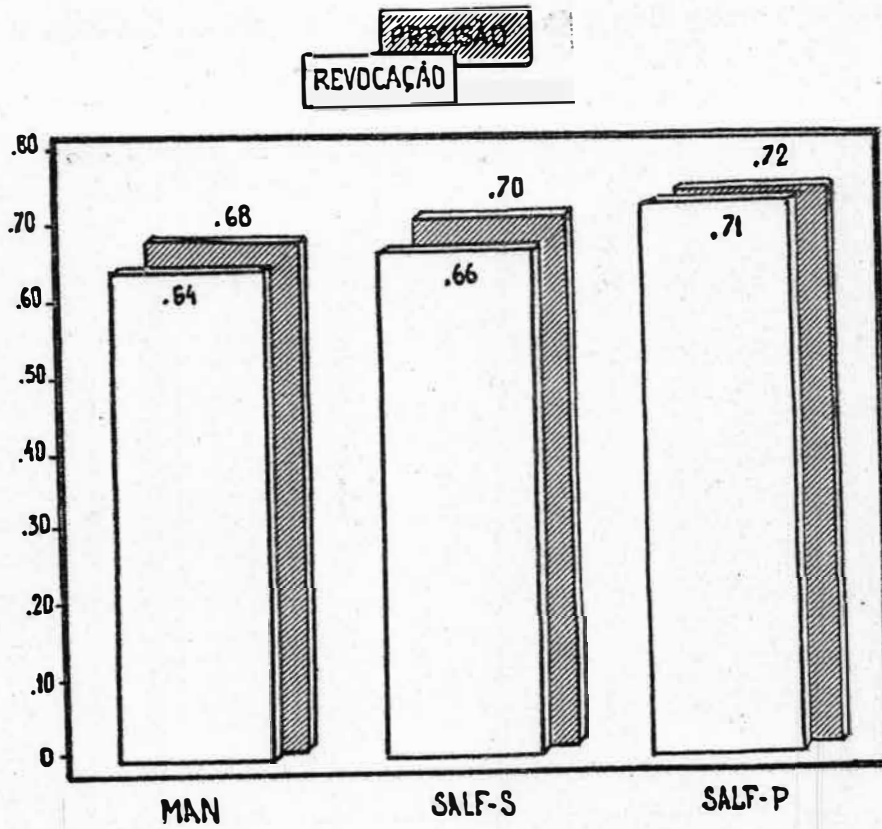
7.4.1.2 Especificidade

A especificidade da indexação tem influência direta no índice de precisão. Ter-se descritores disponíveis com grau de especificidade adequado para se indexar um particular tópico, depende da especificidade da linguagem de indexação - que no presente estudo é uma constante.

A variável que deve ser analisada, neste caso, é a habilidade relativa do indexador e da metodologia para reconhecer o nível correto da especificidade do assunto e associá-lo a um descritor, que para ambos os casos pertence à mesma linguagem de indexação. O gráfico 7 mostra que a indexação manual e as indexações SALF-S e SALF-P estão operando, em média, com o mesmo nível de especificidade.

GRÁFICO 7

PRECISÃO E REVOCAÇÃO POR TIPO DE INDEXAÇÃO



MAN - indexação manual

SALF-S }
 SALF-P } versões do método SALF

A constatação de que a indexação SALF-P apresenta níveis de precisão um pouco acima das outras duas, apesar de exibir o mais alto grau de exaustividade (gráfico 7 e ítem 7.3), aparentemente se configura como uma incoerência dos resultados. O que possivelmente acontece é que os descritores sugeridos pelo método e aceitos pelo indexador são muito específicos e possuem uma capacidade de discriminação bastante elevada, contribuindo, então, numa faixa de revocação bastante ampla (de 0,10 a 0,90), para índices de precisão maiores. A contribuição dos descritores sugeridos aceitos ficaria mais clara se os testes fossem realizados em amostras maiores: provavelmente teríamos índices mais elevados de precisão nas faixas mais baixas de revocação.

7.5 Formação de Expressões

As estratégias usadas para se elevar o nível de precisão da recuperação dependem do uso de termos específicos, tanto na indexação quanto na consulta. Para quem projeta um algoritmo de indexação automática, um dos problemas mais graves entre os que se apresentam é como aumentar a especificidade dos descritores gerados automaticamente. Uma das técnicas é combinar as palavras em expressões. Grupos de palavras são mais específicos que palavras isoladas. O termo "cement industry" tem um significado distinto e mais específico do que "cement" e "industry" isoladamente. Entretanto, é reconhecidamente difícil gerar por máquina expressões que sejam realmente úteis na

identificação do conteúdo dos documentos⁷.

Os procedimentos da metodologia proposta têm como objetivo gerar descritores compostos. O reconhecimento das expressões é realizado por intermédio do tesauro. Assim, a capacidade do método de estruturar expressões depende fortemente do nível de pré-coordenação do tesauro usado.

O controle de sinonímia que o método realiza é efetivado também via tesauro, que neste caso representa o papel de dicionário de sinônimos.

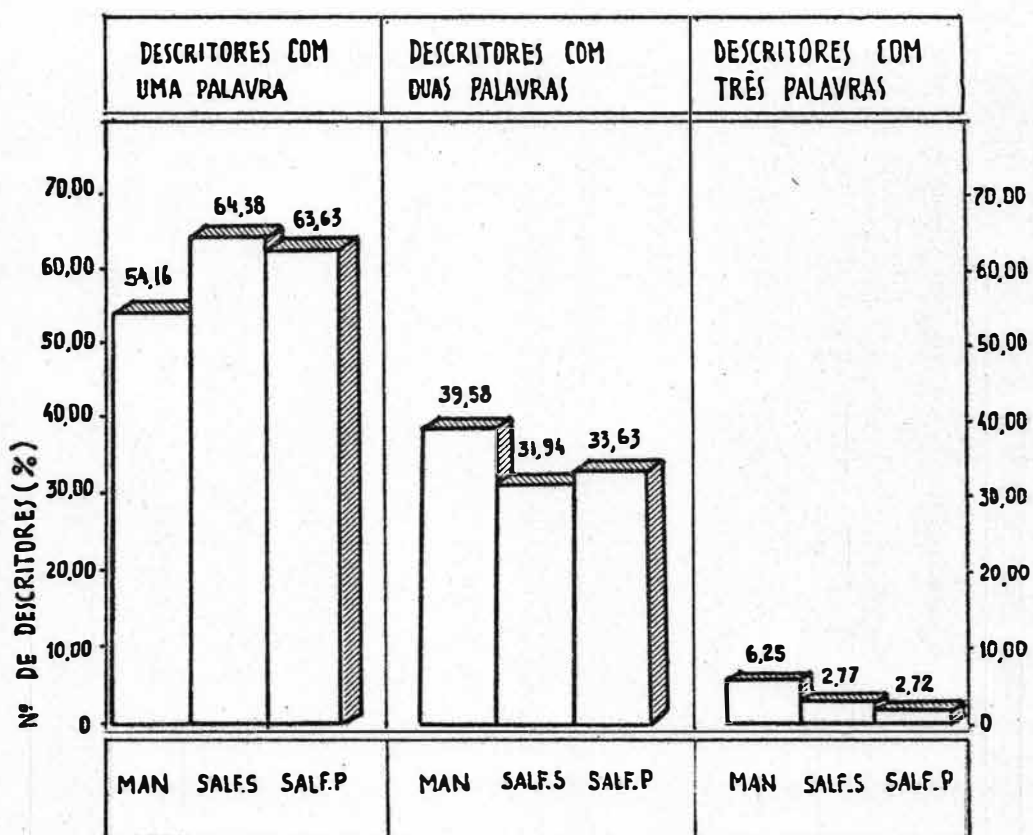
A capacidade de estruturação de expressões do método, quando comparado à indexação manual, apresenta resultados considerados satisfatórios. Comparando-se os números chega-se aos seguintes resultados: 40% dos descritores assinalados manualmente são compostos por duas palavras, a indexação SALF-S apresenta 32%, enquanto a indexação SALF-P tem 35%. A diferença é 20% e 12,5% respectivamente. Para expressões compostas por três palavras, a diferença é maior. No entanto, em números absolutos, a quantidade de expressões de três palavras é baixa para qualquer das três indexações. A indexação manual tem 6,25% dos seus descritores formados por três palavras, a indexação SALF-S 2,77% e a indexação SALF-P 2,72%. Os dados aqui discutidos estão representados no gráfico 8.

7.6 Teste de Consistência

Salton⁸ se posiciona contra a realização de testes

GRÁFICO 8

NÚMERO DE PALAVRAS POR DESCRITOR POR
TIPO DE INDEXAÇÃO



MAN - indexação manual

SALF-S }
SALF-P } versões do método SALF

de comparação, a nível de descritores individuais, entre indexações via computador e indexações manuais. Ele só considera como válidos os testes que avaliem a finalidade última da indexação: a recuperação. "Infelizmente todos os procedimentos de indexação - manuais ou automáticos - produzem resultados relativamente medíocres" afirma textualmente Salton, a propósito dos testes de consistência. Isto significa que a indexação manual não é um padrão confiável num teste de comparação descritor a descritor.

Embora se questione a validade, se testam, em termos de coincidência de descritores individuais, as indexações obtidas pela metodologia e a indexação manual, com a finalidade de se comparar com os números encontrados na literatura.

Nos vários testes já realizados, o consenso está, segundo Salton⁸, em torno de 60% de concordância entre os descritores obtidos manualmente e automaticamente. O coeficiente de avaliação normalmente usado para medir a coincidência de vocabulário é:

$$q = \frac{c}{a + m - c} \times 100$$

onde c representa o número de descritores comuns assinalados, a é o número de descritores gerados automaticamente e m é o número de descritores assinalados manualmente.

A indexação SALF-S e a indexação manual apresentam uma concordância de exatamente 63%, três pontos acima da média encontrada na literatura. Por outro lado, quando se comparam os

descritores da indexação manual com os do método SALF-P, tem-se um resultado obviamente mais baixo, pois os descritores sugeridos e aceitos nunca apareceram antes em nenhum outro documento da área, não podendo, portanto, coincidir com os descritores manuais. O resultado encontrado foi 46% de concordância entre a indexação SALF-P e a manual. A representação dos resultados está no gráfico 9.

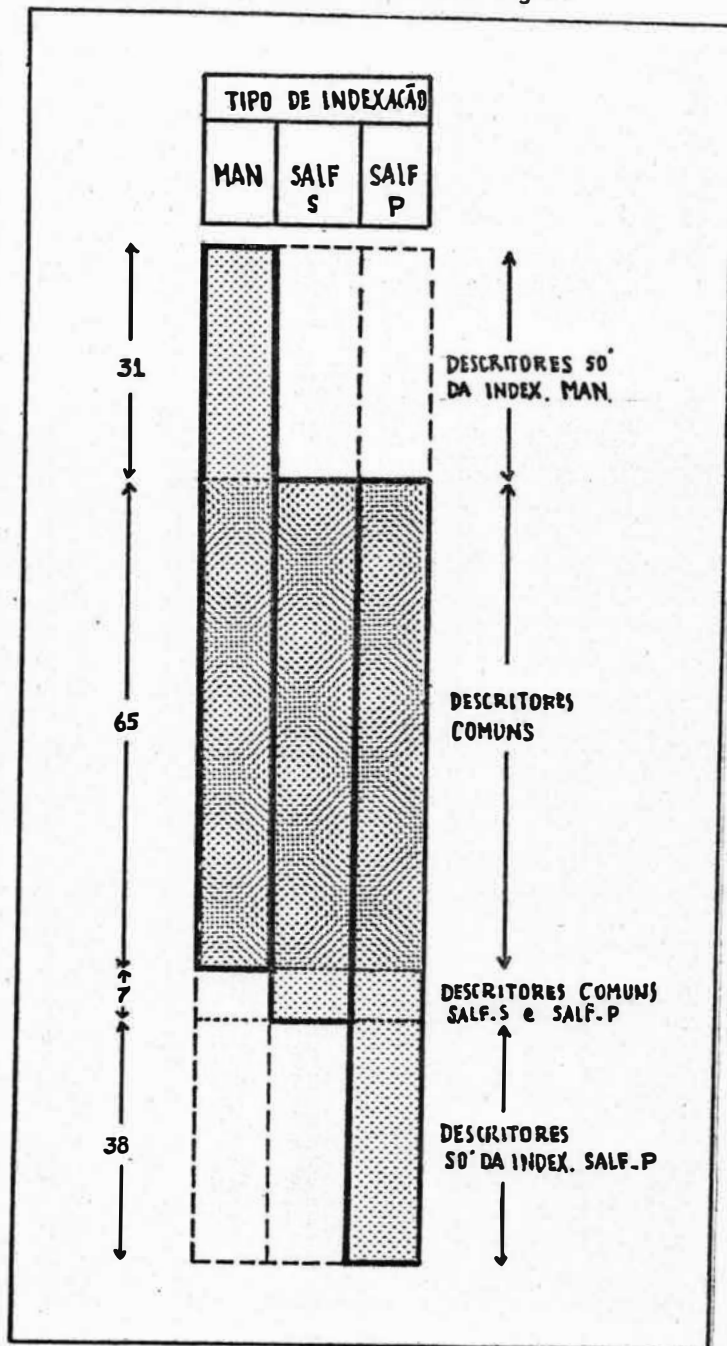
Avaliam-se também a concordância entre os descritores de maiores freqüências. O propósito desta avaliação é detectar possíveis distorções ocorridas na atribuição automática desses descritores em comparação com a atribuição manual. Para que isto fosse possível, selecionaram-se os descritores de freqüência igual ou maior que 4 (quatro)⁹ de ambas as indexações e compararam-se, a seguir, as freqüências dos descritores correspondentes. Considerou-se, para este fim, "aceitável" a diferença de freqüência de até 2 (dois)¹⁰. Diferenças maiores que esta foram consideradas possíveis "distorções" e foram analisadas individualmente (gráfico 10).

O gráfico 10 - construído tomando-se as freqüências das indexações MAN e SALF-S nesta ordem, como pares ordenados e assinalando-os sobre os eixos cartesianos x e y - permite uma visualização mais clara das possíveis distorções. Os pontos fora das linhas paralelas são três: BIODIGESTERS (1,5); BIOMASS (10,6) e PRODUCTION (2,7). Estes pontos são discutidos em seguida.

O descritor "biodigesters" foi assinalado somente uma vez pelo indexador em contraste com as 5 (cinco) vezes em que o método o assinalou. A explicação para esta diferença é sim -

GRÁFICO 9

CONCORDÂNCIA ENTRE AS INDEXAÇÕES



MAN - indexação manual

SALF-S }
 SALF-P } versões do método SALF

GRÁFICO 10

CONCORDÂNCIA ENTRE OS TERMOS DE MAIORES FREQUÊNCIAS

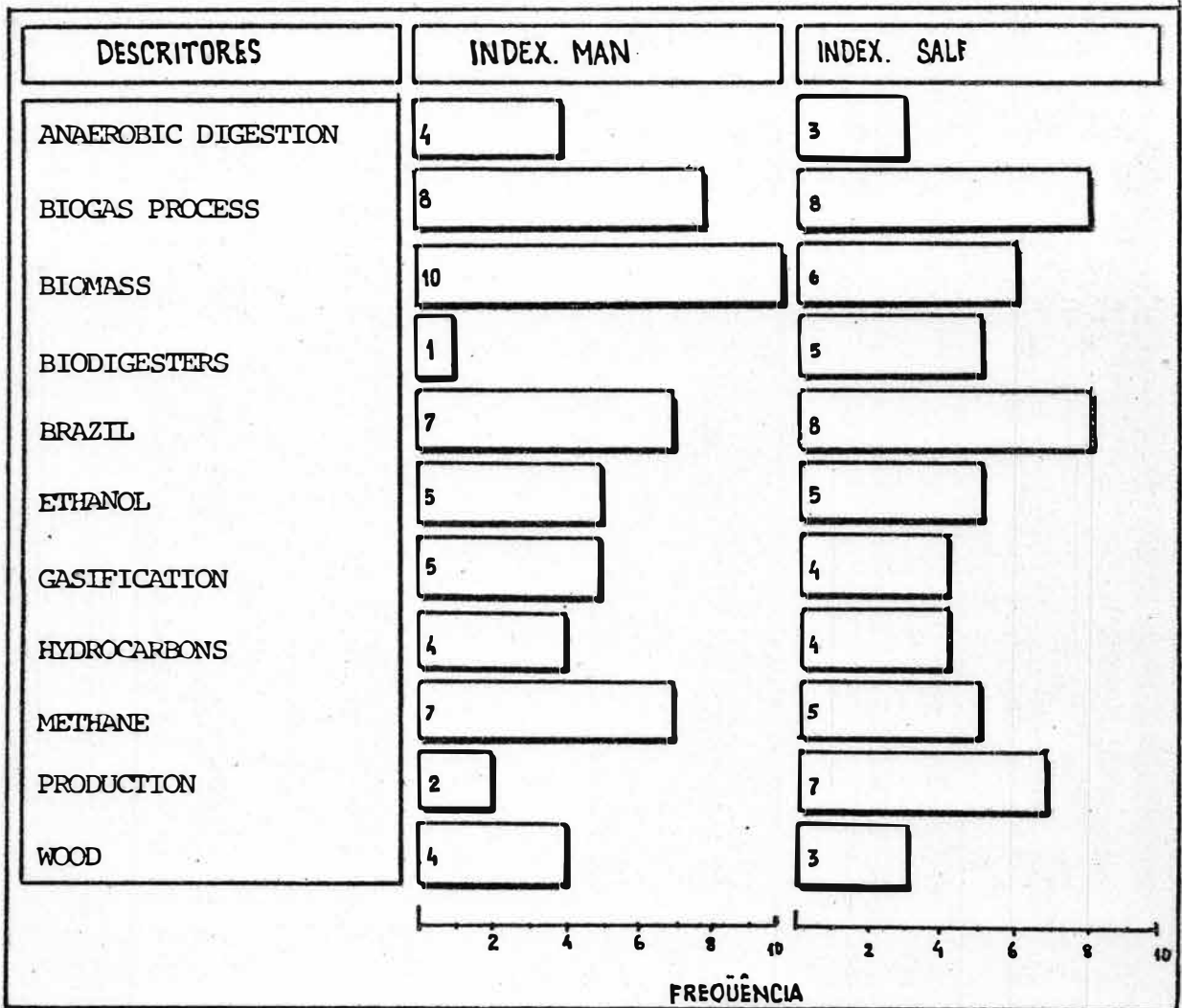
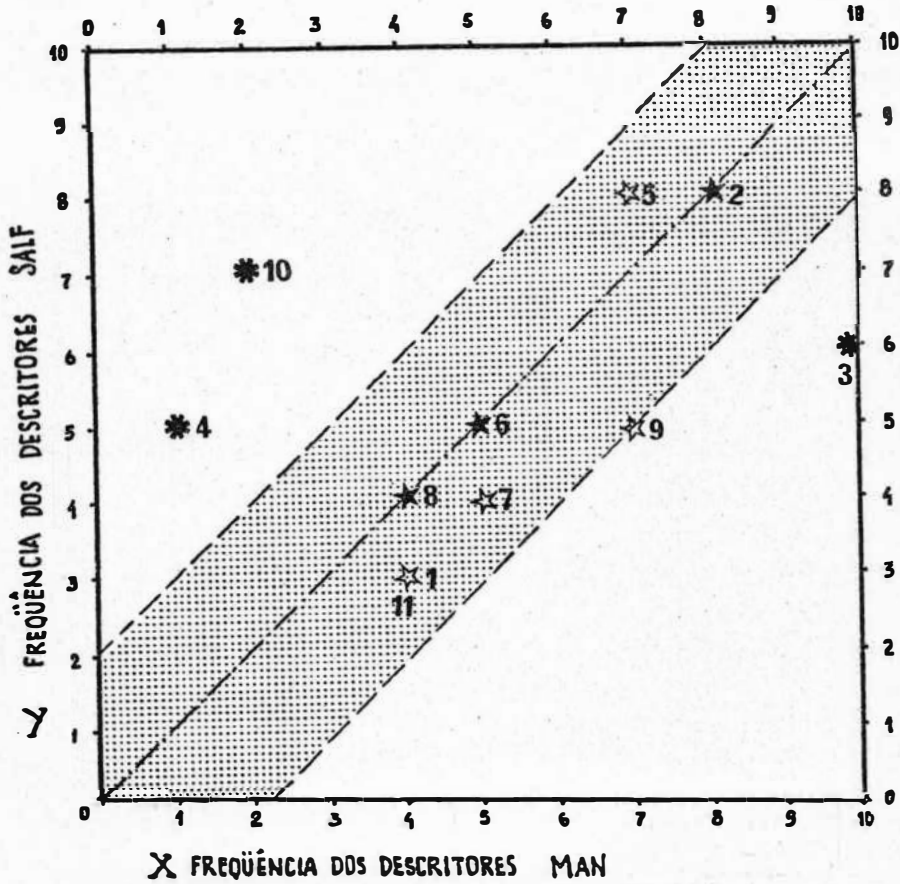


GRÁFICO 11

DISTORÇÕES NOS DESCRITORES DE MAIORES
FREQUÊNCIAS



LEGENDA

- ★ coincidência total: $x - y = 0$
 ☆ coincidência aceitável: $0 < |x - y| \leq 2$
 * possível distorção: $|x - y| > 2$

Nº descritor (x,y)

- 1 anaerobic digestion (4,3)
 2 biogas process (8,8)
 3 biomass (10,6)
 4 biogigesters (1,5)
 5 brazil (7,8)
 6 ethanol (5,5)

Nº descritor (x,y)

- 7 gasification (5,4)
 8 hydrocarbons (4,4)
 9 methane (7,5)
 10 production (2,7)
 11 wood (4,3)

ples: "biodigesters" é um descritor novo, só muito recentemente incorporado ao tesouro DOE, logo pouquíssimos documentos estão indexados com este termo. O fato citado não constitui, portanto, uma distorção.

Analisando a discordância na atribuição do descritor "biomass" verifica-se, com base no título e resumo, que: a) em 3 documentos o indexador assinalou o termo impropriamente, uma vez que "biomass" (segundo o tesouro DOE) é "toda matéria orgânica cultivável tais como plantas, árvores..." quando na realidade os documentos tratam de bioconversão de rejeitos (documentos 2875, 3091, 2090); b) em outros dois documentos, o método assinalou somente as fontes de biomassa ("wood" e "forests") referenciadas pelos documentos (0156 e 0470), o que torna a indexação mais específica; o indexador também assinalou estes descritores, aumentando assim a exaustividade da indexação manual em detrimento da especificidade; c) para o documento 2168, cujo título é "Anais do 2º Seminário sobre Biomassa como Energia na Indústria", só o método assinalou o descritor "biomass".

Quanto ao terceiro ponto sobre o gráfico 11, PRODUCTION (2,7), trata-se realmente de uma distorção da metodologia. O descritor "production" deve ser usado para indexar o conceito "produção industrial". No entanto, usando o algoritmo não se conseguiu distinguir satisfatoriamente o uso comum da palavra, que é bastante freqüente (e por isto ocorre a distorção), do uso estritamente dentro do limite estabelecido pelo tesouro.

QUADRO 1 - Resumo dos resultados obtidos por tipo de indexação

TIPO DE INDEXAÇÃO	Nº DE DESCRITORES ASSINALADOS	Nº MÉDIO DE DESCRITORES ASSINALADOS P/DOC. EXAUSTIVIDADE MÉDIA (\bar{E})	Nº DE DESCRITORES DISTINTOS ASSINALADOS	Nº MÉDIO DE DESCRITORES DISTINTOS ASSINALADOS P/DOCUMENTO	% DE DESCRITORES RETIRADOS APENAS DO TÍTULO	% DE DESCRITORES RETIRADOS APENAS DO RESUMO	% DE DESCRITORES RETIRADOS DO TÍTULO E DO RESUMO	DESEMPENHO		
								RECUPERAÇÃO	REVOC. PREC.	
A M O S T R A										
MANA	172	5,73	96	3,20	-	-	-	0,64	0,68	
	SALF-S	147	4,90	72	2,40	14,37	65,35	20,26	0,66	0,70
	SALF-P	189	6,30	110	3,66	15,51	76,08	5,43	0,71	0,72
ÁREA N13										
MAN _T	409	6,01	177	2,60	-	-	-	-	-	
DESCRITORES SUGERIDOS ACETOS (DSA)										
	42	1,40	38	1,26	17,39	69,79	14,28	-	-	

QUADRO 2 - Resumo dos resultados obtidos por documento

DOCUMENTO IDENTIFICAÇÃO		A	B	C	D	E
Nº	RN	Nº DESCRIT. INDEX MAN.	Nº DESCRIT. INDEX SALF.	Nº DESCRIT. SUGERIDOS	Nº DESCRIT. SUG.ACEITOS	B + D
02	003091	8	7	-	-	7
03	002875	7	5	2	-	5
05	002873	4	3	2	-	3
06	002872	3	3	2	-	3
10	002868	6	5	5	3	8
11	002709	6	4	-	-	4
12	002594	6	4	2	1	5
13	002492	11	6	4	4	10
17	002488	4	2	4	2	4
20	002485	5	4	2	1	5
22	002483	6	6	10	5	11
26	002479	4	5	3	1	6
27	002478	3	5	4	3	8
28	002360	5	4	8	5	9
31	002168	1	5	4	2	7
32	002093	8	5	4	1	6
35	002000	7	4	-	-	4
38	002087	7	4	1	-	4
43	002082	5	5	-	-	5
47	001844	1	4	2	1	5
49	001422	6	5	1	-	5
50	000890	8	7	6	4	11
51	000828	7	6	3	1	7
52	000827	5	5	4	1	6

QUADRO 2 - (Cont.)

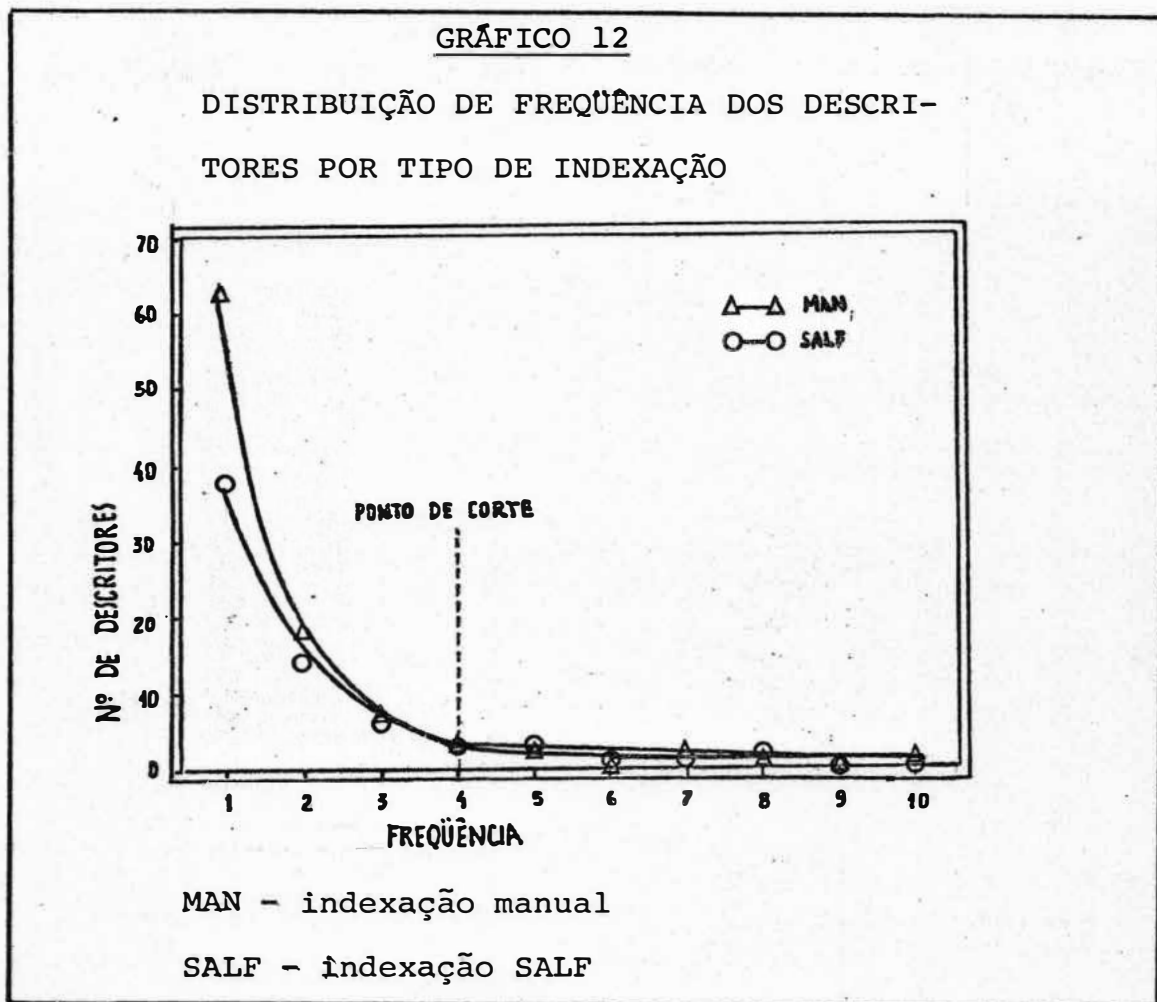
DOCUMENTO IDENTIFICAÇÃO		A	B	C	D	E
Nº	RN	Nº DESCRIT. INDEX MAN.	Nº DESCRIT. INDEX SALF	Nº DESCRIT. SUGERIDOS	Nº DESCRIT. SUG.ACETOS	B + D
53	000826	7	5	1	1	6
57	000822	7	6	10	3	9
63	000816	7	5	-	-	5
64	000470	9	7	3	1	8
66	000464	9	7	2	-	7
67	000156	4	3	1	-	3
TOTAL		172	147	92	42	189
MÉDIA		5.73	4.90	3.06	1.40	

REFERÊNCIAS E NOTAS

- ¹ BRAGA, L.M. Palavras de títulos e resumos como acesso ao conteúdo do documento: uma análise numérica. Rio de Janeiro, UFRJ/IBICT, 1982, 181p.(dissertação)
- ² LANCASTER, F.W. Information retrieval systems: characteristics, testing and evaluation. New York, John Wiley, 1968, 222p.
- ³ Considerou-se exaustividade média (\bar{E}) como a razão entre o número de descritores usados para indexar todos os documentos da amostra e o número de documentos da amostra.
- ⁴ Descritores distintos são aqueles computados apenas quando da primeira ocorrência. Não se consideram, portanto, as repetições do mesmo descritor.
- ⁵ Embora nada tenha sido escrito sobre o assunto, os indexadores do CIN que operam o sistema SUPRIR, têm observado este tipo de comportamento do usuário que acompanha a busca no terminal.
- ⁶ Sparck Jones, K. Progress in documentation-automatic indexing. Journal of Documentation. 30(4): 393-432, 1974.
- ⁷ BORKO, H. Automatic indexing: a tutorial SIGIR Forum (USA), 16 (3): 99-13, 1982.

⁸ SALTON, G. Automatic text analysis, SCIENCE, 165,17 April 1970.

⁹ Para a escolha da frequência 4 como patamar para análise das maiores distorções usou-se o gráfico de distribuição de frequência. Optou-se pelo corte no ponto em que a curva se tornava linear.



¹⁰ Calculando-se a média das discordâncias entre os descritores de maiores frequências, encontra-se o número 1,73. Decidiu-se então, considerar como patamar aceitável o nº inteiro mais próximo - 2 (dois).

8 CONCLUSÃO

A análise da literatura de indexação automática, deixa claro, sob o ponto de vista econômico, técnico, de tempo e sobretudo de qualidade da indexação, que, pelo menos dentro de um contexto experimental, é viável se desenvolver, implementar e implantar projetos de indexação automática.

Levando-se em conta que o objetivo de qualquer técnica de indexação está centralizado principalmente na recuperação, os resultados da análise comparativa entre a metodologia proposta (as duas versões do método SALF) e a indexação manual, confirmam os dados encontrados na literatura da área: é possível estabelecer algoritmos de indexação baseados em vocabulário controlado, implantável em computador, aplicáveis a títulos e resumos de documentos escritos cujos resultados, em função da recuperação, sejam pelo menos equivalentes às técnicas manuais correspondentes.

O método apresentado confirma também a hipótese de alguns autores que afirmam ser possível conseguir um desempenho satisfatório com algoritmos que não tenham envolvimento com elementos lingüísticos profundos e usem tesouro para controlar sinônimos.^{1, 2, 3}

O modelo de avaliação proposto para testar comparativamente o método SALF e o método manual, incluiu os parâmetros exaustividade, desempenho na recuperação (revocação e precisão), consistência e distorção dos termos de maiores frequências. Par

tiu-se do princípio que a linguagem de indexação era uma constante e o que se avaliou foi a capacidade do método SALF atribuir descritores capazes de descreverem todos os assuntos distintos discutidos num particular documento, precisamente, através de descritores preferencialmente compostos por mais de uma palavra, com o objetivo final de se obter graus de precisão e revocação pelo menos equivalentes aos da indexação manual.

A capacidade do método de sugerir ao indexador descritores, que, embora controlados, estão fora do vocabulário histórico da área, para enriquecer tanto a indexação quanto este próprio vocabulário fez com que a exaustividade variasse significativamente entre as duas versões do método proposto. Isto, como era de se esperar, acarretou uma variação detectável nos índices de revocação.

As curvas de precisão x revocação demonstram, para a amostra analisada, que o desempenho entre as indexações pode ser considerado, segundo a literatura⁴, equivalente. Mas na realidade, as duas versões do SALF exibem ligeira "vantagem" sobre o método manual. O índice de consistência (que foi calculado com o intuito de compará-lo com os números encontrados por outros estudos) para o método SALF-S em relação ao método manual, apresentou um resultado ligeiramente acima da média registrada por outros autores⁵. Quanto a versão SALF-P, exibiu um resultado abaixo da média consagrada, posto que usa um vocabulário que, em grande parte, é estranho à área que se indexa.

A investigação das possíveis distorções na atribuição de descritores de alta frequência (≥ 4), através da discor

dância do número de frequência, levou somente a um caso efetivo de distorção do método: o termo "production" estava sendo atribuído erroneamente pelo método SALF.

A formação de expressões pela metodologia apresentou um nível satisfatório em relação ao método manual: 32% e 35% dos descritores eram formados por duas palavras, contra 40% da indexação manual. Os descritores de três palavras eram poucos para ambos os métodos.

A grande vantagem do método SALF é, provavelmente, a fixação de um critério único de indexação para todos os documentos da coleção, o que conduz a um alto grau de consistência em relação aos métodos manuais, que possuem componentes subjetivos que influenciam sensivelmente a qualidade da indexação. Os erros de indexação automática são, em grande parte, sistemáticos, sendo portanto de fácil identificação.

Diante do exposto acima, consideram-se atingidos os objetivos previamente estabelecidos.

Algumas falhas encontradas na metodologia ficam como temas para aprimoramento do método bem como para desenvolvimento de algoritmos mais completos:

- a) forte dependência em relação ao vocabulário controlado - o método depende sob todos os aspectos das características do tesouro utilizado. Por exemplo, a capacidade de formação de expressões está diretamente relacionada ao nível de pré-coordenação do tesouro, assim como o reconhecimento de sinônimos está relacionado ao número de termos de entrada do tesouro;
- b) análise morfológica baseada somente em conversão plu -

ral-singular. Seria desejável uma análise morfológica mais complexa que envolvesse mecanismos para eliminação de prefixos e sufixos;

- c) não há atualização do tesouro: seria desejável que se incorporasse ao algoritmo algum dispositivo para enriquecimento do tesouro, i.e., procedimentos para reconhecimento de termos novos que refletissem o desenvolvimento da área;
- d) o método proposto é essencialmente derivativo (como a grande maioria dos métodos de indexação automática), is significa que a atribuição de descritores é altamente influenciada pelo estilo, tamanho e até a época em que o documento foi escrito.
- e) impossibilidade de se indexar o texto fonte em português e/ou de se assinalar descritores na língua portuguesa.

Como diretrizes para estudos futuros na área de indexação automática, sugerem-se os seguintes tópicos:

- a) desenvolvimento de modelos lingüísticos voltados para problemas ligados à recuperação de informação;
- b) por seu caráter experimental, as pesquisas na área de indexação automática tendem a ser empíricas. Deve-se, portanto, desenvolver métodos analíticos para a análise formal dos problemas e avaliação dos modelos de indexação automática;
- c) desenvolvimento de modelos automáticos que dêem maior atenção aos processos de formulação de consulta;
- d) estudos que considerem o custo-benefício da indexação automática.

CITAÇÕES E NOTAS

- ¹ SALTON, G. Automatic indexing and abstracting. In: Dynamics information and library processing. chapt 3. Englewood Cliffs, N.J., Prentice Hall, 1975. p.76-114.
- ² SALTON, G. Automatic term class construction using relevance - a summary of work in automatic pseudoclassification. Information Processing & Management.16; 1-15, 1980
- ³ DILLON, M; McDONALD, L.K. Fully automatic book indexing. Journal of Documentation, 39(3): 135-54, Sep. 1983.
- ⁴ SPARCK JONES, K. Progress in documentation- automatic indexing. Journal of Documentation. 3041: 393-432, 1974.
- ⁵ SALTON, G. Automatic text analysis, Science, 165, 17 April 1970.

9 BIBLIOGRAFIA

- ANDREWSKY, A.; RUAS, V. Indexação automática baseada em métodos linguísticos e estatísticos e sua aplicabilidade à língua portuguesa. Rio de Janeiro, PUC-DI, 1982. 31p.
- ARTANDI, S. Machine indexing, linguistic and semiotic implications. Journal of the American Society for Information Science, 27 (4):235-39
- BAILEY, K.D. Methods of social research. New York, The Free Press; London, Collier Macmillan Publishers, 1978. p.84.
- BARNES, C.I.; CONSTANTINI, L.; PERSCHKE, S. Automatic indexing using the SLC-II system. Information Processing & Management, 14: 107-19, 1978.
- BARREIRO, S.C. Avaliação do serviço de disseminação seletiva de informação do Centro de Informações Nucleares em relação à precisão, recuperação e novidade. Rio de Janeiro, UFRJ/IBICT, 1979. 160p. (Dissertação)
- BARREIRO, S.C.; QUEIROZ, G.G. Incorporação automática de informações brasileiras no INIS. Revista de Biblioteconomia de Brasília, 8(2):118-130, junh./dez.1980.

- BASTOS, S.B. Análise comparativa entre indexação automática e manual da literatura brasileira de ciência da informação. Brasília, Universidade de Brasília, 1984. 204p. (Dissertação)
- BAXENDALE, P.B. Machine-made index for technical literature-an experiment. IBM Journal of Research and Development, 2: 354-61, 1958.
- BERNSTEIN, L.M.; WILLIANSO, R.E. Testing of natural language retrieval system for a full text knowledge base. Journal of the American Society for Information Science, 35(4): 235-47, 1984.
- BOOKSTEIN, A.; SWANSON, D.R. A probabilistic models for automatic indexing. Journal of the American Society for Information Science, 25(5):312-8, Sep./Oct. 1974.
- BORKO, H. The construction of an empirically based mathematically derived classification system. Santa Monica, Cal., System Development Corp., 26 Oct. 1961, 23p. (SP-585).
- BORKO, H.; BERNICK, M.D. Automatic document classification, Part II - Adittional Experiments. Santa Monica, Cal., System Development Corp., 18 Oct. 1963. 33p. (TM-771/001/00).
- BORKO, H.; BERNICK, M.D. Automatic document classification, Santa Monica, Cal., System Development Corp., 13 Nov. 1963 19p. (TM-771).

- BORKO, H.; BERNICK, M.D. Toward establishment of a computer based classification system for scientific documentation. Santa Monica, Cal., System Development Corp., 19 Feb. 1964. 47p. (TM-1763).
- BORKO, H. Automatic indexing: a tutorial, SIGIR tutorial session, In: ACM'81. Automatic Information Retrieval. Los Angeles, CA, Nov. 9-11, 1981.
- BRAGA, L.M. Palavras de títulos e resumos como acesso ao conteúdo do documento: uma análise numérica. Rio de Janeiro, UFRJ/IBICT, 1982. 181p. (Dissertação).
- BROZOWSKI, J.P. MASCARADE: searching the full text of abstracts using automatic indexing. Journal of Information Science, 6:67-73, 1983.
- CERNY, B.A. & LAWRENCE, J.D. An interactive indexing-editing system for DOE Technical Information Center. Lawrence Berkeley Lab., nd Rp. (LBL--14049).
- CITRON, J.L.; HART, L.; OHLMAN, H. A permutation index to the "Preprints of the International Conference on Scientific Information". Santa Monica, Cal., System Development Corp., 1958. 140p. (SP-44).
- CLEVERDON, C. Evaluation tests of information retrieval systems J. Docum, 26(1):55-67, Mar. 1970.

- DILLON, M.; KNIGHT, C.; LOSPINUSO; ULMSCHENEIDER. The use of automatic indexing for authority control. Journal of Library Automation, 14(4):268-77, 1981.
- DILLON, M. Thesauro-based automatic book indexing. Information Processing & Management, 18(4):167-78, 1982.
- DILLON, M.; GRAY, A. Fully automatic syntax-based indexing. Journal of the American Society for Information Science, 34(2):99-108, 1983.
- DILLON, M.; McDONALD, L.K. Fully automatic book indexing. Journal of Documentation, 39(3):135-54, Sep.1983.
- DILLON, M.; FEDERHART, P. Statistical recognition of content terms in general texts. Journal of the American Society for Information Science, 34(1):3-10, 1984.
- DUNHAM, G.S.; PACAK, M.G.; PRATT, A.W. Automatic indexing of pathology data. Journal of the American Society for Information Science, p.81-90, 1978.
- EDMUNDSON, H.P. A new method in automatic extracting. Journal of ACM, 16(2):264-85, April 1969.
- EDMUNDSON, H.P. Mathematical models of the texts. Information Processing & Management, 20(12):261-68, 1984.

EUA.DEPARTMENT OF ENERGY. Technical Information Center. Energy information data base: energy categories. Oak Ridge, Te 1980 (DOE/TIC 4584-R4).

EUA.DEPARTMENT OF ENERGY. Technical Information Center. Energy data base: subject thesaurus. Oak Ridge, Te., 1981 (DOE/TIC 7000-R5).

FIELD, B.J. Towards automatic indexing: automatic assignment of controlled-language indexing and classification from free indexing. Journal of Documentation, 31(4) dec.1975.

FIELD, B.J. Indexation automatique pour systèmes multilingues. In: CONGRÈS EUROPEEN SUR LES SYSTÈMES ET RESEAUX DOCUMENTAIRES. 1, Luxemburgo, 1977. Proceedings...Luxemburgo, Verlag Dokumentation, 1977. p.479-501.

GRAVES, R.W.; HELANDER, D.P. A feasibility study of automatic indexing and information retrieval. IEE Transactions on Engineering Writing and Speech, 32(2): 58-59, 1970.

HALLER, J. Processamento de textos em linguagem natural. In: CONGRESSO NACIONAL DE INFORMÁTICA, 15., Rio de Janeiro, out. 1982. 9p.

HALLER, J. Análise automática de textos em sistemas de informação. Revista de Biblioteconomia de Brasília, 11(1):105-13 jan/jun 1983.

- JANOS, J. Theory of functional sentence perspective and its application for the purpose of automatic extracting. Information Processing & Management, 15:19-25, 1979.
- KEPPEL, F.P. Looking forward, a fantasy. In: DANTON, EM. ed: The Library of Tomorrow, 1939. p.1-11.
- KOLAR, W. The SLC-II language translation package user manual for input. Luxembourg, Commission for the European Communities, 1974 (EUR-5173).
- LANCASTER, F.W. Information retrieval systems: characteristics, testing and evaluation. New York, J.Willey, 1968. 222p.
- LOVINS, J.B. Development of a Stemming Algorithm. Mech.Transl and Comp.Lingustic II. 1968. p.22-31.
- LUHN, H.P. A statistical approach to mechanized encoding and searching of literature information. IBM Journal of Research and Development, 1(4):309-317, Oct.1957.
- LUHN, H.P. The automatic creation of literature abstracts. IBM Journal of Research and Development, 2:159-165, 1958.
- LUHN, H.P. An experiment in auto-abstracting: auto-abstracts of area 5. In: CONFERENCE PAPERS INTERNATIONAL CONFERENCE ON SCIENTIFIC INFORMATION, 17 Nov.1958, Conference Papers, Yorktown Heights, N.Y., IBM Research Center. 18p.

- LUHN, H.P. Auto-encoding of documents for information retrieval system. In: Boaz, M., ed. Modern Trends in Documentation . nl, 1959. p.45-58.
- LUHN, H.P. Automatic intelligence systems-some basic problems and prerequisites for their solution. In: TOMESKI, E.A. et al. Classification, Unification and Integration of Information Storage and Retrieval. nl, 1961. p.3-20.
- MARON, M.E. Automatic indexing: an experimental enquiry. nl, (rev.2 Feb.1961) 1960 3lp. (p-2180)
- MARTINS, A.V. Um método para indexação automática de textos. Rio de Janeiro, Instituto Militar de Engenharia, 1983.100p. (Dissertação).
- MOYNE, J.A. Information retrieval and natural language. In: AMERICAN SOCIETY FOR INFORMATION SCIENCE. Proceedings...v. 6, p.259-263.
- NISHIDA, F.; TAKAMATSU, S.; FUJITA, Y. Semiautomatic indexing of structured information of text. J.Chem.Comput.Sci,24(1): 15-20. 1984.
- PELZER, C. International information networks.The International Nuclear Information System. ASLIB Proc. 24(1):38-55. Jan. 1972.

PERSHKE, S.; VERNIMB, C. Feasibility study of automatic indexing of INIS abstracts. IAEA-SM-189-24. p.409-15.

QUEIROZ, G.G; FRADE, M.E.; GONÇALVES, O.C. Acesso ao documento: experiência do CIN/CNEN. In: 2a. REUNIÃO BRASILEIRA DE CIÊNCIA DA INFORMAÇÃO, 2, Rio de Janeiro, 1979. Anais, 113p.

QUEIROZ, G.G. Fonte: base de dados bibliográficos em fontes de energia. In: CONGRESSO BRASILEIRO DE BIBLIOTECONOMIA E DOCUMENTAÇÃO, 14., Balneário Camboriú, SC, out.1983. trabalhos apresentados, 14p.

ROBREDO, J.; FERREIRA, J.A. Conceituação de um programa para indexação automática de textos. Revista de Biblioteconomia de Brasília, 8(2):254-63, jul./dez.1980.

ROBREDO, J. A indexação automática como mecanismo básico no processo de transferência da informação. In: CONGRESSO LATINO-AMERICANO DE BIBLIOTECONOMIA E DOCUMENTAÇÃO, 1., Salvador, 21-26 set. 1980. Anais... Salvador, FEBAB, 1980. 19p.

ROBREDO, J. A indexação automática de textos: o presente já entrou no futuro. In: Machado, U.D., ed. Estudos avançados em Biblioteconomia e Ciência da Informação. Brasília, ABDF, 1982. v.1 p.236-74.

ROBREDO, J. Otimização dos processos de indexação dos documentos e de recuperação da informação mediante o uso de instrumentos de controle terminológica Ciência da Informação, 11(1):3-18, 1982.

- SALTON, G. Designing automatic information system, results obtained with the SMART programs. Social Science Information, 6(2):111-17, Feb. 1967.
- SALTON, G. Automatic information, organization and retrieval. New York, McGraw-Hill, 1968.
- SALTON, G. Automatic indexing using bibliographic citations. Journal of Documentation, 27(2):98-110, June 1971.
- SALTON, G. Automatic text analysis. Science, 168(3929):335-43, 17 Apr. 1970.
- SALTON, G. A new comparison between conventional indexing and automatic text processing. Journal of the American Society for Information Science, 23(2):75-84, Mar./Apr. 1972.
- SALTON, G.; YANG, G.S.; YU, C.T. A theory of term importance in automatic text analysis. Journal of the American Society for Information Science, 26(1):33-44, Jan-Feb. 1975.
- SALTON, G. Automatic indexing and abstracting CH.3 in his Dynamics information and library processing. Englewood Cliffs, N.J., Prentice Hall, 1975. p.76-114.
- SALTON, G. Automatic term class construction using relevance - a summary of work in automatic pseudoclassification. Information Processing & Management, 16:1-15, 1980.

- SALTON, G. The SMART retrieval systems - experiments in automatic document processing. New York, Prentice-Hall, Englewood Cliffs, 1981.
- SALTON, G. The measurement of the term importance in automatic indexing. Journal of the American Society for Information Science, 32(3):175-86, May 1981.
- SALTON, G. Automatic indexing: a summary. In: EURIM 5 CONFERENCE, Versailles, France, 1982. Proceedings... ASLIB, 1982. p.66-77.
- SAYÃO, L.F.; QUEIROZ, G.G. Otimização de perfis ponderados de DSI usando-se um sistema de busca retrospectiva por lógica booleana. In: CONGRESSO LATINO-AMERICANO DE BIBLIOTECONO - MIA E DOCUMENTAÇÃO, 1., Salvador, 21-26 set., 1980. Anais, 23p.
- SCHIELE, J.; KURTEV, I.; RUCKENBAURE, E.; STANIK, Z.; VARECKA, F.; SCHMID, H. INIS: manual for online retrieval. Vienna, IAEA, 1984. 147p. (IAEA-INIS-17, rev 2).
- SPARCK JONES, K. Collection properties influencing automatic term classification performance. Information Storage and Retrieval, 9:499-13, 1973.
- SPARCK JONES, K. Progress in documentation - automatic indexing. Journal of Documentation. 3041:393-432, 1974.

- SPARCK JONES, K.; BATES, R.G. Research on automatic indexing 1974-1976, University of Cambridge, 1977.
- SPARCK JONES, K. The role of automatic indexing in operational on-line retrieval systems. In: FID CONGRESS, 39, Edinburg, 25-28 Sep. 1978. New trends in documentation and information London, ASLIB, 1980, p.33-8.
- STEVENS, M.E. Automatic indexing: a state-of-the-art report. National Bureau of Standards, Washington D.C., 1965. 290p. (NBS-91).
- STOKOLOV, N.V. On automatic support to indexing a life science data base. Information Processing & Management, 18(6):313-21, 1982.
- SWANSON, D.R. Research Procedures for automatic indexing. In: _____ . Machine Indexing. American U., 1962. p.281-304.
- SWANSON, D.R. Library goals and the role of automation. Spec. Libraries, 53:466-71, 1962.
- SWANSON, D.R. Automation indexing and classification. In: NATO ADVANCED STUDY INSTITUTE ON AUTOMATIC DOCUMENT ANALYSIS, Venice, 7-20 July 1963. Proceedings...N.Y., 1963, 125-128.
- TRACHTENBERG, A. Automatic document classification using information theoretical methods, In: Luhn, H.P., ed. Automation and Scientific Communication, 1963, Pt.2. p.349-50.

VAN DER MEULEN, W.A.; JANSEN, P.J.F.C. Automatic versus manual indexing. Information Processing and Management, 13(1):13-21, 1977.

VAN RIJSBERGEN, C.J. A theoretical basis for the use of co-occurrence data in information retrieval, Journal of Documentation, 27(2):69-82, June 1971.

VON STAA, A. PRAXPAL: um indexador semi-automático. In: CONGRESSO DA SOCIEDADE BRASILEIRA DE COMPUTAÇÃO, 3. Campinas, 1983. Anais... São Paulo, 1983. p.433-43.

YU, C.T.; SALTON, G. Precision weighting-an effective automatic indexing method. Journal of Association for Computing Machinery, 23:76-88, 1976.

ZAHER, C.L. et alii. Automação da informação em Física no Brasil. In: SEMINÁRIO SOBRE INFORMÁTICA, Rio de Janeiro, 1968. Anais. Rio de Janeiro, IBBD, 1969. p.39-52.

ZAHER, C.R.; DUARTE, Y.C. Sistema KWIC versus descritores. In: CONGRESSO SOBRE DOCUMENTAÇÃO, 2.; FID/CLA Reunião 9., Rio de Janeiro, 1969. Anais. Rio de Janeiro, IBBD, 1969.p.195-206.

10 A N E X O S

ANEXO 1

FOLHA DE ENTRADA DA FONTE

(dados catalográficos e de indexação)

001 B R 85 0 4 2 6 3 002 1 / 4 003 A W 004 0 005

006 T / 007

008 T 33 009 02

RELACIONADOR CATEGORIA DE ESCRITÓRIO (PRIMARIAS) SECUNDARIAS TOTAL DE FOLHAS DE ENTRADA

009 A

TIPO DE REGISTRO: LIVRO, DISCO, FILME, MAPA, FOLHA, OUTROS, PERIÓDICO, PATENTE, FOLHA, COMPUT.

NÍVEL BIBL: A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S, T, U, V, W, X, Y, Z, AA, AB, AC, AD, AE, AF, AG, AH, AI, AJ, AK, AL, AM, AN, AO, AP, AQ, AR, AS, AT, AU, AV, AW, AX, AY, AZ, BA, BB, BC, BD, BE, BF, BG, BH, BI, BJ, BK, BL, BM, BN, BO, BP, BQ, BR, BS, BT, BU, BV, BW, BX, BY, BZ, CA, CB, CC, CD, CE, CF, CG, CH, CI, CJ, CK, CL, CM, CN, CO, CP, CQ, CR, CS, CT, CU, CV, CW, CX, CY, CZ, DA, DB, DC, DD, DE, DF, DG, DH, DI, DJ, DK, DL, DM, DN, DO, DP, DQ, DR, DS, DT, DU, DV, DW, DX, DY, DZ, EA, EB, EC, ED, EE, EF, EG, EH, EI, EJ, EK, EL, EM, EN, EO, EP, EQ, ER, ES, ET, EU, EV, EW, EX, EY, EZ, FA, FB, FC, FD, FE, FF, FG, FH, FI, FJ, FK, FL, FM, FN, FO, FP, FQ, FR, FS, FT, FU, FV, FW, FX, FY, FZ, GA, GB, GC, GD, GE, GF, GG, GH, GI, GJ, GK, GL, GM, GN, GO, GP, GQ, GR, GS, GT, GU, GV, GW, GX, GY, GZ, HA, HB, HC, HD, HE, HF, HG, HH, HI, HJ, HK, HL, HM, HN, HO, HP, HQ, HR, HS, HT, HU, HV, HW, HX, HY, HZ, IA, IB, IC, ID, IE, IF, IG, IH, II, IJ, IK, IL, IM, IN, IO, IP, IQ, IR, IS, IT, IU, IV, IW, IX, IY, IZ, JA, JB, JC, JD, JE, JF, JG, JH, JI, JJ, JK, JL, JM, JN, JO, JP, JQ, JR, JS, JT, JU, JV, JW, JX, JY, JZ, KA, KB, KC, KD, KE, KF, KG, KH, KI, KJ, KK, KL, KM, KN, KO, KP, KQ, KR, KS, KT, KU, KV, KW, KX, KY, KZ, LA, LB, LC, LD, LE, LF, LG, LH, LI, LJ, LK, LL, LM, LN, LO, LP, LQ, LR, LS, LT, LU, LV, LW, LX, LY, LZ, MA, MB, MC, MD, ME, MF, MG, MH, MI, MJ, MK, ML, MM, MN, MO, MP, MQ, MR, MS, MT, MU, MV, MW, MX, MY, MZ, NA, NB, NC, ND, NE, NF, NG, NH, NI, NJ, NK, NL, NM, NN, NO, NP, NQ, NR, NS, NT, NU, NV, NW, NX, NY, NZ, OA, OB, OC, OD, OE, OF, OG, OH, OI, OJ, OK, OL, OM, ON, OO, OP, OQ, OR, OS, OT, OU, OV, OW, OX, OY, OZ, PA, PB, PC, PD, PE, PF, PG, PH, PI, PJ, PK, PL, PM, PN, PO, PP, PQ, PR, PS, PT, PU, PV, PW, PX, PY, PZ, QA, QB, QC, QD, QE, QF, QG, QH, QI, QJ, QK, QL, QM, QN, QO, QP, QQ, QR, QS, QT, QU, QV, QW, QX, QY, QZ, RA, RB, RC, RD, RE, RF, RG, RH, RI, RJ, RK, RL, RM, RN, RO, RP, RQ, RR, RS, RT, RU, RV, RW, RX, RY, RZ, SA, SB, SC, SD, SE, SF, SG, SH, SI, SJ, SK, SL, SM, SN, SO, SP, SQ, SR, SS, ST, SU, SV, SW, SX, SY, SZ, TA, TB, TC, TD, TE, TF, TG, TH, TI, TJ, TK, TL, TM, TN, TO, TP, TQ, TR, TS, TT, TU, TV, TW, TX, TY, TZ, UA, UB, UC, UD, UE, UF, UG, UH, UI, UJ, UK, UL, UM, UN, UO, UP, UQ, UR, US, UT, UY, UZ, VA, VB, VC, VD, VE, VF, VG, VH, VI, VJ, VK, VL, VM, VN, VO, VP, VQ, VR, VS, VT, VU, VV, VW, VX, VY, VZ, WA, WB, WC, WD, WE, WF, WG, WH, WI, WJ, WK, WL, WM, WN, WO, WP, WQ, WR, WS, WT, WY, WZ, XA, XB, XC, XD, XE, XF, XG, XH, XI, XJ, XK, XL, XM, XN, XO, XP, XQ, XR, XS, XT, XU, XV, XW, XX, XY, XZ, YA, YB, YC, YD, YE, YF, YG, YH, YI, YJ, YK, YL, YM, YN, YO, YP, YQ, YR, YS, YT, YU, YV, YW, YX, YY, YZ, ZA, ZB, ZC, ZD, ZE, ZF, ZG, ZH, ZI, ZJ, ZK, ZL, ZM, ZN, ZO, ZP, ZQ, ZR, ZS, ZT, ZU, ZV, ZW, ZX, ZY, ZZ

FORNE
FOLHA DE ENTRADA
DADOS BIBLIOGRÁFICOS E DE INDEXAÇÃO

1 009 A [USE UMA FOLHA DE ENTRADA SEPARADA PARA CADA NÍVEL ASSIMILADO, COMEÇANDO PELO NÍVEL MAIS ESPECÍFICO, QUE DEVE COLOCADO NO ESPAÇO DOS PARA ENTRADA DE SÉRIES USE A SEÇÃO 2 DESTA FOLHA DE ENTRADA.]

	ETIQUETA	DADOS
AUTOR PESSOAL (S) INVENTOR (S) EDITOR (S) AFILIAÇÃO (S)	100	Fonseca, M.A.P. (ACESTA, Belo Horizonte (Brasil))
COLABORAÇÃO	107	
ENTIDADE	110	
ORGAO ACADÊMICO	111	
TÍTULO	200	The performance of a reactive power compensation system.
SUBTÍTULO PRIMÁRIO	201	
TÍTULO DA CONF.	210	Seminário de Eletrotermia
LOCAL DA CONF.	211	São Paulo, SP (Brasil)
DATA DA CONF.	213	20-21 Out 1982
TÍTULO SECUNDÁRIO	230	O desempenho de um sistema de compensação
SUBTÍTULO SECUNDÁRIO	231	
PRIMEIRO	250	
NR DE RELATÓRIO E PARTE	300	
NR SECUNDÁRIO	310	
ISBN/IPC	320	
LOCAL DE PUBLICAÇÃO	401	
EDITOR	402	
DATA DE PUBLICAÇÃO	403	
COLAÇÃO	500	p. 38-41
IDIOMA	600	(Em Português)
NOTAS	610	
DISPONIBILIDADE	611	
ALUNTO DE TÍTULO (OPCIONAL)	620	
CÓDIGO DE APLICAÇÃO	700	8950188
CODIGOS DE ENTIDADE	710	

2 009 S

SÉRIE / TÍTULO DO PERIÓDICO	230	
SUBTÍTULO SECUNDÁRIO	231	
PRIMEIRO	250	
DATA DE PUBLICAÇÃO	403	
COLAÇÃO	500	
NOTAS	610	

PAIS ANO NR SEQUENCIAL			REVISTA		NOVO		FONTE												
001	B R	8 5 0 4 2 6 3	2	4	B	N	H	C	B	005									
RELACIONADOR			CATEGORIA (PRIMARIA)		CATEGORIAS (SECUNDARIAS)		TOTAL DE FOLHAS DE ENTRADA		TIPO DE REGISTRO		NIVEL NIVL		INDICADOR LITERARIO						
006	Y																		

FONTE
FOLHA DE ENTRADA
DADOS BIBLIOGRÁFICOS E DE INDEXAÇÃO

1 009 M NIVEL [USE UMA FOLHA DE ENTRADA SEPARADA PARA CADA NIVEL ASSINELADO, COMEÇANDO PELO NIVEL MAIS ESPECIFICO, QUE DEVA COLOCADO NO ESPACO DOB. PARA ENTRADA DE SÉRIE USE A SEÇÃO 2 DESTA FOLHA DE ENTRADA]

ETIQUETA	DADOS	
AUTOR PESSOAL (S) INVENTOR (S) EDITOR (S) AFILIACÃO (S)	100	
COLABORAÇÃO	107	
ENTIDADE	110	Revista Eletricidade Moderna, Rio de Janeiro (Brasil)
ORAM ACADÊMICO	111	
TÍTULO	200	Proceedings of the Eletrothermia Seminar
SUBTÍTULO PRIMÁRIO	201	
TÍTULO DA CONF.	210	
LOCAL DA CONF.	211	
DATA DA CONF.	213	
TÍTULO SECUNDÁRIO	230	Anais do Seminário de Eletrotemia
SUBTÍTULO SECUNDÁRIO	231	
EDIÇÃO	250	
NR DE RELATORIO E PATENTE	300	F-982
NR SECUNDÁRIO	310	
ISBN / IAC	320	
LOCAL DE PUBLICAÇÃO	401	
EDITOR	402	
DATA DE PUBLICAÇÃO	403	1982
COLAÇÃO	500	50p.
NOTAS	600	
NOTAS	610	
DISPONIBILIDADE	611	F-982
ADJUNTOS DE TÍTULO (OPCIONAIS)	620	
PROJETO DE APLICAÇÃO	700	
CODIGO DE ENTIDADE	710	

2 009 S NIVEL

SÉRIE / TÍTULO DO PERIÓDICO	230	
SUBTÍTULO PERIÓDICO	231	
ISSN	320	
DATA DE PUBLICAÇÃO	403	
COLAÇÃO	500	
NOTAS	610	

PAIS	ANO	Nº SEQUENCIAL		
BR	85	04263	3	4
TRM		F DE EN- TRADA NA		TOTAL DE FOLHAS DE ENTRADA

FONTE
FOLHA DE ENTRADA
RESUMO

4	009	<input checked="" type="checkbox"/> X	<input type="checkbox"/> E	N	USE UMA FOLHA DE ENTRADA PARA CADA IDIOMA DO RESUMO!
		NIVEL	CODIGO DO IDIOMA		
	TIPOFITA	DADOS			
IDIOMA	850	English			
RESUMO	860	<p>The reactive power compensation system installed at ACESITA (Companhia de Aços Especiais Itabira, de Minas Gerais) is described. Its functioning, initial operation problems and its performance during their three years of functioning are presented. (RS)</p>			

INDEXADOR:

DIGITADOR:

PAIS	ANO	NR SEQUENCIAL	F. DE EN- TRADA		TOTAL DE PG. NAS DE ENTRADA
001	B.R.	8 5 0 4 2 6 3	000	4 / 4	4
728		F. DE EN- TRADA		TOTAL DE PG. NAS DE ENTRADA	

FONTE FOLHA DE ENTRADA RESUMO

4

009 X P T

INSERIR UMA FOLHA DE ENTRADA PARA CADA IDIOMA DO RESUMO

	ETIQUETA	DADOS
IDIOMA	850	Português
RESUMO	860	<p>Descreve-se, sucintamente, o sistema de compensação de reativos instalado na Companhia de Aços Especiais Itabira, de Minas Gerais. Mostra-se o seu funcionamento, problemas no início da operação e sua performance durante os seus três anos de funcionamento. (Autor)</p>

LIDE XADOR:

DIBITADDA:

DOE THESAURUS

A-BOMB SURVIVORS [01]

- RT Delayed Radiation Effects
- RT Epidemiology
- RT Hiroshima
- RT Man
- RT Nagasaki

A CENTERS

- DA February 1975
- BT1 Color Centers
- BT2 Vacancies
- BT3 Point Defects
- BT4 Crystal Defects
- BT5 Crystal Structure

A CODES [01]

- BT1 Computer Codes

A-15 Compounds

- DA May 1979
- USE Beta-W Lattices

A-1 Reactor (Bohunice)

- USE Bohunice A-1 Reactor

A-2 Reactor (Bohunice)

- USE Bohunice A-2 Reactor

A-1 Reactor (Calder Hall)

- USE Calder Hall A-1 Reactor

A-2 Reactor (Calder Hall)

- USE Calder Hall A-2 Reactor

A RESONANCES [01]

(For A3 resonances use PI-1640 RESONANCES.)

- BT1 Meson Resonances
- BT2 Mesons
- BT3 Bosons
- BT3 Hadrons
- BT4 Elementary Particles
- BT2 Resonance Particles
- BT3 Hadrons
- BT4 Elementary Particles
- NT1 A1-1070 Resonances
- NT1 A2-1310 Resonances
- NT1 A4-1960 Resonances
- NT1 A2H-1320 Resonances
- NT1 A2L-1280 Resonances
- NT1 F-1540 Resonances

A1-1070 RESONANCES [01]

- BT1 A Resonances
- BT2 Meson Resonances
- BT3 Mesons
- BT4 Bosons
- BT4 Hadrons
- BT5 Elementary Particles
- BT3 Resonance Particles
- BT4 Hadrons
- BT5 Elementary Particles

A2-1310 RESONANCES [01]

- BT1 A Resonances
- BT2 Meson Resonances
- BT3 Mesons
- BT4 Bosons
- BT4 Hadrons
- BT5 Elementary Particles
- BT3 Resonance Particles
- BT4 Hadrons
- BT5 Elementary Particles
- BT1 Tensor Mesons
- BT2 Meson Resonances
- BT3 Mesons
- BT4 Bosons
- BT4 Hadrons
- BT5 Elementary Particles

- BT3 Resonance Particles
- BT4 Hadrons
- BT5 Elementary Particles

A3 Resonances

- USE Pi-1640 Resonances

A4-1960 RESONANCES [01]

- DA March 1975
- BT1 A Resonances
- BT2 Meson Resonances
- BT3 Mesons
- BT4 Bosons
- BT4 Hadrons
- BT5 Elementary Particles
- BT3 Resonance Particles
- BT4 Hadrons
- BT5 Elementary Particles
- BT1 Exotic Resonances
- BT2 Resonance Particles
- BT3 Hadrons
- BT4 Elementary Particles

A 285 Steel

- DA December 1978
- USE Steel-ASTM-A285

AABO CYCLOTRON [01]

- UF Turku Cyclotron
- BT1 Isochronous Cyclotrons
- BT2 Cyclotrons
- BT3 Cyclic Accelerators
- BT4 Accelerators

AAEC [01]

- DA April 1978
- UF Australian Atomic Energy Commission
- BT1 Australian Organizations
- BT2 National Organizations

AAPS

- DA May 1979
- UF Advanced Automotive Propulsion Systems
- RT Automotive Industry
- RT Electric-Powered Vehicles
- RT Gas Turbine Engines
- RT Internal Combustion Engines
- RT Stirling Engines

AARR REACTOR [01]

- UF Argonne Tank Research and Test Reactor-AARR
- BT1 Research Reactors
- BT2 Research and Test Reactors
- BT3 Reactors
- BT1 Tank Type Reactors
- BT2 Reactors
- BT1 Water Cooled Reactors
- BT2 Reactors
- BT1 Water Moderated Reactors
- BT2 Reactors

ABANDONED SHAFTS

- DA December 1977
- UF Disused Mineshafts
- BT1 Shaft Excavations
- RT Coal Mines
- RT Mines

ABANDONED SITES [01]

- DA October 1978
- RT Land Reclamation

ABANDONED WELLS

- DA August 1977
- BT1 Wells
- RT Natural Gas Wells

- RT Oil Wells
- DEF An oil or gas well abandoned because its yield has fallen below that necessary for profitable production.

Abashian-Booth-Crowe Effect

- DA November 1977
- USE ABC Effect

ABC EFFECT [01]

- DA November 1977
- UF Abashian-Booth-Crowe Effect
- RT Interactions
- RT Missing-Mass Spectra
- RT Pions

ABDOMEN [01]

- BT1 Body Areas
- BT2 Body
- RT Diaphragm
- RT Gastrointestinal Tract
- RT Liver
- RT Peritoncum
- RT Spleen

Aberdeen Maryland Reactor

- USE APRF Reactor

Aberration Yield

- USE Mutation Frequency

ABFST EQUATION [01]

- (Amati-Bertocchi-Fabini-Strangelloni-Tonin Equation.)
- BT1 Equations
- RT Multiperipheral Model
- RT Regge Poles
- RT Scattering Amplitudes

ABLATION [01]

(For the medical concept use SURGERY or RADIOTHERAPY.)

- RT Erosion
- RT Heat Transfer
- RT Reentry
- RT Refractories
- RT Sublimation Heat

Abnormalities (Chromosomal)

- USE Chromosomal Aberrations

Abnormalities (Developmental)

- USE Malformations

ABORTION [01]

- RT Pregnancy
- RT Reproductive Disorders

Abragam Model

- USE Abragam-Pound Theory

ABRAGAM-POUND THEORY [01]

- UF Abragam Model
- RT Angular Correlation
- RT Angular Distribution

ABRASION [01]

- RT Abrasives
- RT Erosion
- RT Wear

ABRASIVES [01]

- RT Abrasion
- RT Pumice

ABRIKOSOV THEORY [01]

- RT Superconductivity

ABS [01]

- UF Alkyl Benzenesulfonates

ANEXO 3 - TABELA DE CLASSIFICAÇÃO DOE

CONTENTS

Category	Name	Page No.
01	COAL AND COAL PRODUCTS	1
02	PETROLEUM	2
03	NATURAL GAS	2
04	OIL SHALES AND TAR SANDS	3
05	FISSION FUELS	4
06	FUSION FUELS	5
07	ISOTOPE AND RADIATION SOURCE TECHNOLOGY	6
08	HYDROGEN	6
09	OTHER SYNTHETIC AND NATURAL FUELS	8
13	HYDRO ENERGY	8
14	SOLAR ENERGY	9
15	GEOTHERMAL ENERGY	10
16	TIDAL POWER	13
17	WIND ENERGY	13
20	ELECTRIC POWER ENGINEERING	14
21	NUCLEAR POWER PLANTS	14
22	NUCLEAR REACTOR TECHNOLOGY	15
25	ENERGY STORAGE	16
29	ENERGY MANAGEMENT AND CONSERVATION POLICY	17
30	ENERGY CONVERSION	19
32	ENERGY CONSERVATION, CONSUMPTION, AND UTILIZATION	20
33	ADVANCED AUTOMOTIVE PROPULSION SYSTEMS	21
36	MATERIALS	23
40	CHEMISTRY	25
42	ENGINEERING	27
43	PARTICLE ACCELERATORS	29
44	INSTRUMENTATION	29
45	EXPLOSIONS AND EXPLOSIVES	30
50	ENVIRONMENTAL SCIENCES, ATMOSPHERIC	30
51	ENVIRONMENTAL SCIENCES, TERRESTRIAL	31
52	ENVIRONMENTAL SCIENCES, AQUATIC	31
55	BIOMEDICAL SCIENCES, BASIC STUDIES	32
56	BIOMEDICAL SCIENCES, APPLIED STUDIES	34
57	HEALTH AND SAFETY	38
64	PHYSICS RESEARCH	38
70	CONTROLLED THERMONUCLEAR RESEARCH	50
99	GENERAL AND MISCELLANEOUS	51

ANEXO 4 - DICIONÁRIO HISTÓRICO DA ÁREA N13 DA FONTE

<u>DESCRIPTOR</u>	<u>FREQUÊNCIA</u>	<u>DESCRIPTOR</u>	<u>FREQUÊNCIA</u>
AGRICULTURAL WASTES	4	BY-PRODUCTS	1
AGRICULTURE	4	CARBON DIOXIDE	1
ALGAE	2	CARBONIZATION	2
ALKANES	1	CASTOR OIL	1
ALKENES	1	CATALYSIS	1
ALKYLATION	1	CATALYTIC REFORMING	
ANAEROBIC CONDITIONS	1	CATTLE	1
ANAEROBIC DIGESTION	10	CELLULOSE	2
APROPRIATE TECHNOLOGY	1	CHARCOAL	5
AUTOMOBILES	1	CHEMICAL PROPERTIES	1
AUTOMOTIVE FUELS	2	CHINA	1
BAGASSE	1	CHROMATOGRAPHY	1
BABASSU	2	COAL	1
BEETS	2	COAL BURNING APPLIANCES	1
BIOCONVERSION	2	COAL GAS	1
BIODIGESTERS	3	COAL GASIFICATION PLANTS	1
BIOGAS PROCESS	25	COCONUTS	1
BIOELECTRICITY	1	COMBUSTION	1
BIOLOGICAL WASTES	1	COMMUNITIES	1
BIOMASS	19	COMPARATIVE EVALUATIONS	2
BIOMASS PLANTATION	1	COST	3
BOILERS	1	COST BENEFIT ANALYSIS	1
BRAZIL	12	COTTONSEED OIL	1
BRAZILIAN ORGANIZATIONS	2	CRACKING	1

ANEXO 4 - (Continuação)

<u>DESCRIPTOR</u>	<u>FREQUÊNCIA</u>	<u>DESCRIPTOR</u>	<u>FREQUÊNCIA</u>
CRG PROCESSES	1	ETHYLENE	1
CRAPS	2	EUPHROBIA	1
DECOMPOSITION	2	EVALUATION	1
DESIGN	2	FABRICATION	1
DESULFURIZATION	2	FARMS	1
DIESEL ENGINES	1	FEASIBILITY STUDIES	2
DIESEL FUELS	4	FERMENTATION	6
DOMESTIC ANIMALS	1	FERTILIZER INDUSTRY	2
ECONOMIC ANALYSIS	1	FERTILIZERS	2
ECONOMIC DEVELOPMENT	1	FLUIDIZED-BED COMBUST.	1
ECONOMIC IMPACT	1	FOREST	1
ECONOMICS	2	FRACTIONATION	1
EFFICIENCY	1	FUEL GAS	4
ELECTRICAL POWER	1	FUEL OILS	4
ENERGY CONSUMPTION	1	FUEL SUBSTITUTION	9
ENERGY POLICY	2	GASIFICATION	5
ENERGY SOURCES	9	GAS FUELS	1
ENERGY SOURCES DEVELOPMENT	1	GAS RECYCLE HYDROG.PROC.	1
ENERGY SUPPLIES	2	HEATING	1
EQUIPMENT	7	HEMICELLULOSE	1
ESSENTIAL OILS	1	HYDROCARBONS	5
ETHANOL	5	HYDROFLUORIC ACID	1
ETHANOL FUELS	1	HYDROGENATION	3
ETHANOL PLANTS	1	HYDROGEN SULFIDES	1

ANEXO 4 - (Continuação)

<u>DESCRIPTOR</u>	<u>FREQUÊNCIA</u>	<u>DESCRIPTOR</u>	<u>FREQUÊNCIA</u>
IGNITION	1	OXIDATION	1
INDIA	1	PAPER INDUSTRY	1
INDUSTRY	1	PARTICLE SIZE	1
INFRARED SPECTRA	2	PEANUT OIL	2
IRON HYDROXIDES	1	PERFORMANCE	5
IRRIGATION	1	PESTICIDES	1
LABORATORIES	1	PHYSICAL PROPERTIES	1
LATEX	1	PILOT PLANTS	1
LIQUID FUELS	2	PLANTS	3
LUBRICATING OILS	1	POWER DEMAND	1
LUBRICANTS	1	POWER PLANTS	1
MAMURES	4	POLYMERIZATION	1
MANUFACTURING	1	PRESSURE DEPENDENCE	1
METAL INDUSTRY	2	PRODUCER GAS	1
METHANE	23	PRODUCTION	3
METHANOL	1	PURIFICATION	2
MUNICIPAL WASTES	1	PYROLISIS	2
NAPHTHA	1	QUALITY ASSURANCE	1
NATIONAL PROGRAM PLANS	3	QUALITY CONTROL	1
ODORIZATION	1	RECOMMENDATIONS	2
OILS	1	REMOVAL	1
OPTIMIZATION	1	RENEWABLE ENERGY SOURCE	2
ORGANIC MATTER	1	RESEARCH PROGRAMS	7
ORGANIC WASTES	2	RESIDENTIAL SECTOR	1

ANEXO 4 - (Conclusão)

<u>DESCRIPTOR</u>	<u>FREQUENCIA</u>	<u>DESCRIPTOR</u>	<u>FREQUENCIA</u>
RESOURCE POTENTIAL	1	URBAN AREAS	1
RESOURCES	1	USES	5
REVIEWS	5	VAPOR PRESSURE	1
RUBBERS	1	VISCOSITY	1
RURAL AREAS	12	WASTE PROCESSING	2
SANITARY LANDFILLS	1	WASTE PROCESSING PLANTS	5
SEPARATION PROCESSES	1	WASTE PRODUCT UTILIZATION	1
SEWAGE	1	WIND POWER	1
SEWAGE SLUDGE	1	WOOD	7
SOLAR COOKERS	1	WOOD FUELS	2
SOLAR DRYERS	1		
SOLAR ENERGY	1		
SOLAR STILLs	1		
SORGHUM	1		
SOYBEAN OIL	3		
SPECIFICATIONS	6		
STANDARD INDUSTRIAL CLASSIFICA.	1		
STANDARDIZATION	1		
STILLAGE	3		
SYNTHETIC FUELS	2		
TEMPERATURE MEASUREMENT	1		
TESTING	1		
TOWN GAS	1		
TRANSPORTATION SYSTEMS	1		

ANEXO 5 - DICIONÁRIO NEGATIVO (INIS/STAIRS Stopword List)

a	annotations	been
ab	another	bei
able	any	beim
above	appear	below
about	appeared	better
according	appears	between
achieved	applying	both
affected	approx	briefly
affects	are	but
after	arise	by
against	arising	
allows	around	calculate
almost	arranged	calculated
alone	as	calculating
along	assumed	cannot
already	assuming	carried
also	at	caused
although	attributed	causes
always	authors	certain
among	avec	chosen
an		clearly
and	be	concerned
and/or	because	concerning
annotation	become	concluded

ANEXO 5 - (Continuação)

consider	describing	foer
consideration	description	for
considered	di	found
considering	did	from
contain	die	fuer
contained	discuss	further
containing	discussed	give
contains	discusses	given
could	discussion	gives
das	do	good
de	does	had
deal	done	has
dealing	du	have
deals	due	held
dei	during	hence
del	each	here
dem	easily	his
den	eine	how
der	either	however
degli	especially	if
della	et	implications
dello	etc	important
describe	examined	in
described	find	include
describes		

ANEXO 5 - (Continuação)

included	most	out
includes	move	overall
including	much	particular
investigate	must	particularly
investigated	narrative	per
investigation	necessary	possible
involved	no	pour
involves	notes	preliminary
involving	observed	present
into	obtain	presented
is	obtained	presents
issued	obtaining	previous
it	occurred	previously
its	occurring	proposed
itself	of	published
known	often	rather
likely	on	recent
lo	ones	regarding
made	only	relatively
mainly	or	remarks
make	organized	reported
making	other	result
meaning	others	resulted
might	our	resulting

ANEXO 5 - (Conclusão)

results	thereby	various
same	therefore	very
seen	these	von
should	this	voor
show	those	vor
showed	though	was
shown	through	we
sows	throughout	were
similar	thus	what
since	times	when
some	to	where
studied	together	whereas
subsequent	too	whether
such	toward	which
sufficient	towards	while
suitable	und	who
taken	until	why
taking	upon	will
than	use	with
that	used	within
the	useful	without
their	using	would
them	usually	you
there	van	your

ANEXO 6 - FORMULAÇÃO DAS CONSULTAS PARA O TESTE DE RECUPERAÇÃO
 ASSUNTO: SUBSTITUIÇÃO DE COMBUSTÍVEIS FÓSSEIS NA IN-
 DÚSTRIA, AGRICULTURA, ÁREAS RURAIS, ETC. POR COMBUS-
 TÍVEIS GERADOS A PARTIR DA BIOMASSA.

QUADRO 6.1 - FORMULAÇÃO DAS CONSULTAS (Descritores selecionados
 pertinentes ao tema)

FA- CE- TAS	DESCRITORES	F R E Q U Ê N C I A		
		INDEX.MAN.	IND.SALF-S	IND.SALF-P
A	BIOMASS	10	6	6
	GASIFICATION	5	4	4
	HYDROCARBONS	4	4	4
B	ALGAE	2	2	2
	CELLULOSE	1	2	2
	PLANTS	2	3	3
	SUGAR BEETS	-	-	2
	SUGAR CANE	-	-	2
	WOOD	2	3	3
	WOOD FUELS	2	1	1
C	ALCOHOL FUEL	-	-	1
	ALCOHOLS	-	-	1
	CHARCOAL	3	3	3
	COAL GAS	1	1	1
	ETHANOL	5	5	5
	ETHYLENE	1	1	1
	FUEL OILS	3	2	2
	ETHANOL FUELS	1	-	-
	LIQUID FUELS	1	2	2
	LOW BTU GAS	-	-	2
	METHANOL	1	1	1
	PRODUCER GAS	1	-	-

ANEXO 6 - QUADRO 6.1 (Conclusão)

FA- CE- TAS	D E S C R I T O R E S	F R E Q U Ê N C I A		
		INDEX.MAN.	IND.SALF-S	IND.SALF-P
D	AGRICULTURE	2	1	1
	BRAZIL	7	8	8
	CEMENT INDUSTRY	-	-	1
	FURNACES	-	-	1
	METAL INDUSTRY	1	-	-
	INDUSTRY	-	3	1
	RURAL AREAS	2	1	1
	STEEL INDUSTRY	-	-	1
E	ENERGY SOURCES	3	1	1
	FUEL SUBSTITUTION	1	1	1
	RENEWABLE ENERG.SOURCES	1	1	1
	RESEARCH PROGRAMS	2	2	2
	USES	1	2	2

QUADRO 6.2 - FORMULAÇÃO DAS CONSULTAS (Combinação dos grupos de descritores)

COMBINAÇÃO DAS FACETAS	INDEX.MANUAL		INDEX.SALF-S		INDEX.SALF-P	
	PRECISÃO	REVOCAÇÃO	PRECISÃO	REVOCAÇÃO	PRECISÃO	REVOCAÇÃO
A+B+C+D+E	0,40	1,00	0,43	1,00	0,43	1,00
A*(B+C+D+E)	0,64	0,90	0,60	0,90	0,60	0,90
A*B	0,66	0,60	0,70	0,70	0,72	0,80
A*B*(C+D+E)	0,71	0,50	0,75	0,60	0,77	0,70
A*B*C	0,83	0,50	0,83	0,50	0,87	0,60
A*B*C*D*E	1,00	0,10	1,00	0,10	1,00	0,10
MÉDIAS	0,68	0,64	0,70	0,65	0,72	0,71

ANEXO 7 - ESTATÍSTICA DOS DESCRITORES

QUADRO 7.1 - DESCRITORES DA INDEXAÇÃO MANUAL

1a. coluna: descritores assinalados pelo indexador

2a. coluna: número de vezes que o descritor foi assinalado

3a. coluna: para que documentos o descritor foi assinalado

DESCRITORES ASSINALADOS NA INDEXAÇÃO MANUAL	FREQUÊNCIA	O C O R R Ê N C I A
AGRICULTURAL WASTES	3	02; 17; 64;
AGRICULTURE	2	38; 50;
ALGAE	2	20; 22;
ANAEROBIC DIGESTION	4	02; 22; 49; 63;
APROPRIATE TECHNOLOGY	1	03;
BEETS	2	50; 51;
BIOCONVERSION	1	05;
BIODIGESTERS	1	02;
BIOGAS PROCESS	8	02; 06; 10; 22; 35; 38; 43;63;
BIOMASS	10	02;03;05;26;27;35;50;64;66;67;
BIOMASS PLANTATION	1	20;
BOILERS	1	53;
BRAZIL	7	02;05;38;43;50;51;64;
BRAZILIAN ORGANIZATIONS	1	66;
BY-PRODUCTS	1	52;
CARBON DIOXIDE	1	52;
CARBONIZATION	2	64;66
CATALYTIC REFORMING	1	13;
CELLULOSE	1	27;
CHARCOAL	3	28;64;66;
COAL GAS	1	28;
COAL GASIFICATION PLANTS	1	28;
COMMUNITIES	1	17;
COMPARATIVE EVALUATIONS	2	32;53;
COST	1	02;
COST BENEFIT ANALYSIS	1	51;
CRACKING	1	57;
CRG PROCESSES	1	13;

QUADRO 7.1 - (Continuação)

DESCRIPTORES ASSINALADOS NA INDEXAÇÃO MANUAL	FREQUÊNCIA	O C O R R Ê N C I A
CROPS	1	50;51;
CHROMATOGRAPHY	1	57;
DESIGN	1	03;
DESULFURIZATION	2	10;13;
ECONOMIC DEVELOPMENT	1	50;
ECONOMICS	2	64;60;
ENERGY SOURCES	3	12;38;49;
ENERGY SUPPLIES	2	17;22;
EQUIPMENT	3	03;13;35;
ETHANOL	5	27;32;50;51;52;
ETHANOL FUELS	1	26;
ETHANOL PLANTS	1	38;
ETHYLENE	1	32;
FABRICATION	1	35;
FEASIBILITY STUDIES	1	12;
FERMENTATION	1	03;32;49;
FORESTS	1	64;
FUEL GAS	3	28;53;63;
FUEL OILS	3	11;53;63;
FUEL SUBSTITUTION	1	26;
GAS FUEL	1	67;
GAS RECYCLE HYDROG.	1	13;
GASIFICATION	5	03;13;28;53;67;
HYDROCARBONS	4	12;20;32;57;
HYDROGENATION	2	13;57;
HYDROGEN SULFIDES	1	10;
INFRARED SPECTRA	1	57;
IRON HYDROXIDES	1	10;
LATEX	1	12;
LIQUID FUELS	1	11;
MANURES	2	02;
METAL INDUSTRY	1	66;

QUADRO 7.1 - (Continuação)

DESCRITORES ASSINALADOS NA INDEXAÇÃO MANUAL	FREQUÊNCIA	O C O R R Ê N C I A
METHANE	7	03;06;10;17;22;47;49;52;
METHANOL	1	52;
NAPHTA	1	13;
NATIONAL PROGRAM PLANS	1	49;
ODORIZATION	1	13;
OPTIMIZATION	1	49;
ORGANIC WASTES	1	63;
OXIDATION	1	32;
PERFORMANCE	1	35;
PILOT PLANTS	1	13;
PLANTS	2	12;35;
POWER DEMAND	1	50;
POLYMERIZATION	1	32;
PRODUCER GAS	1	53;
PRODUCTION	2	50;57;
PURIFICATION	2	06;10;
PYROLYSIS	1	63;
RECOMENTATIONS	1	11;
RENEWABLE ENERGY SOURCES	1	26;
RESEARCH PROGRAMS	2	35;38;
RESOURCES	1	05;
REVIEWS	3	31;32;43;
RUBBERS	1	12;
RURAL AREAS	2	38;43;
SEWAGE SLUDGE	1	49;
SORGHUM	1	50;
SOYBEAN OIL	1	57;
SPECIFICATIONS	1	11;
SYNTHETIC FUELS	1	20;
TESTING	1	20;
TOWN GAS	1	13;
USES	1	66;

QUADRO 7.1 - (Conclusão)

DESCRIPTORES ASSINALADOS NA INDEXAÇÃO MANUAL	FREQUÊNCIA	O C O R R Ê N C I A
VISCOSITY	1	11;
WASTE PROCESSING PLANTS	1	65;
WOOD	4	53;64;
WOOD FUELS	2	64;66;

ANEXO 7 - (Continuação)

QUADRO 7.2 - DESCRITORES DA INDEXAÇÃO SAI-F-S

- 1a. Coluna: Descritores assinalados pelo método SAI-F-S
- 2a. Coluna: Frequência única
- 3a. Coluna: Nº de vezes que o descritor foi gerado a partir do título
- 4a. Coluna: Nº de vezes que o descritor foi gerado a partir do resumo
- 5a. Coluna: Nº de vezes que o descritor foi gerado a partir do título e resumo simultaneamente
- 6a. Coluna: Total
- 7a. Coluna: Para que documentos o descritor foi assinalado

DESCRITORES ASSINALADOS NA INDEXAÇÃO SAI-F-S	ÚNICA	F R E Q U Ê N C I A				Nº DOCUMENTO (FREQ. TÍTULO, RESUMO)
		G E R A L				
		TÍTULO	RESUMO	TIT & RES	TOTAL	
AGRICULTURAL WASTES	2	1	2	1	3	2 (1,R); 17 (2,T,R);
AGRICULTURE	1	-	1	-	1	51 (1,R);
ALGAE	2	-	3	-	3	20 (1,R); 22 (1,R);
ANAEROBIC DIGESTION	3	-	3	-	3	3 (1,R); 22 (1,R); 49 (1,R);
BIOCONVERSION	2	1	1	-	2	3 (1,T); 5 (1,R);
BIODIGESTER	5	-	5	-	5	2 (1,R); 35 (1,R); 38 (1,R); 43 (1,R); 47 (1,R);
BIOGAS PROCESS	8	6	9	5	15	2 (2,R,T); 6 (3,T,R); 10 (3,T,R); 17 (1,R); 22 (1,R); 35 (2,T,R); 43 (2,T,R); 47 (1,T);

QUADRO 7.2 - (continuação)

DESCRIPTORES ASSIMILADOS PELO MÉTODO SAFE-S	F R E Q U Ê N C I A					O C O R R Ê N C I A
	ÚNICA	G E R A L				
		TÍTULO	RESUMO	TÍT&RES	TOTAL	
BIOMASS	6	5	6	2	11	5(5,T,R);26(1,T);27(1,T);31(2,T,R); 50(1,T);66(1,R);
BOILERS	1	1	2	1	3	53(3,T,R);
BRAZIL	8	3	6	1	9	5(1,R);11(1,R);38(1,R);43(1,R);50(1,T); 51(2,T,R);64(1,T,R);66(1,R);
BY-PRODUCTS	1	-	1	-	1	49(1,R);
CARBON DIOXIDE	1	-	1	-	1	52(1,R);
CARBONIZATION	1	-	1	-	1	64(1,R);
CATALYTIC REFORMING	1	-	1	-	1	13(1,R);
CELLULOSE	1	-	1	-	1	27(1,R);
CHARCOAL	3	3	7	3	10	28(4,T,R);64(3,T,R);66(3,T,R);
CHROMATOGRAPHY	1	-	1	-	1	57(1,R);
COAL GAS	1	-	1	-	1	28(1,R);
COMBUSTION	1	-	1	-	1	11(1,R);
COST	2	1	1	-	2	2(1,T);27(1,R);
CRACKING	1	-	1	-	1	57(1,R);
CROPS	2	-	4	-	4	50(3,R);51(1,R);
DESIGN	1	1	-	-	1	3(1,T);
DESULFURIZATION	1	-	1	-	1	13(1,R);
ECONOMIC DEVELOPMENT	1	-	1	-	1	50(1,R);

QUADRO 7.2 - (Continuação)

DESCRITORES ASSINALADOS PELO MÉTODO SALT-S	F R E Q U Ê N C I A					O C O R R Ê N C I A Nº DOCUMENTO (FREQ., TÍTULO, RESUMO)
	ÚNICA	G E R A L				
		TÍTULO	RESUMO	TIT & RES	TOTAL	
EFFICIENCY	1	-	1	-	1	57(1,R);
ENERGY CONSUMPTION	1	-	1	-	1	64(1,R);
EQUIPMENT	1	1	-	-	1	10(1,T);
ETHANOL	5	1	6	1	7	27(1,R); 32(1,R); 38(1,R); 50(2,R); 51(2,T,R);
ETHYLENE	1	-	2	-	2	32(2,R);
EUPHORBIA	1	1	1	1	2	12(2,T,R);
FABRICATION	1	-	1	-	1	35(1,R);
FERMENTATION	1	1	2	1	3	27(3,T,R);
FORESTS	1	-	1	-	1	64(1,R);
FUEL GAS	3	1	2	-	3	2(1,R); 53(1,T); 63(1,R);
FUEL OILS	2	-	2	-	2	11(1,R); 63(1,R);
FUEL SUBSTITUTION	1	-	1	-	1	26(1,R);
GASIFICATION	4	2	4	-	6	28(2,T,R); 31(1,R); 53(1,R); 67(2,T,R);
HYDROCARBONS	4	3	5	3	8	12(2,T,R); 20(2,T,R); 32(2,T,R); 57(2,R);
HYDROGENATION	1	-	1	-	1	57(1,R);
HYDROGEN SULFIDES	1	-	1	-	1	10(1,R);
INDUSTRY	3	2	3	1	5	3(2,R); 31(2,T,R); 66(1,T);
IRON HYDROXIDES	1	-	1	-	1	10(1,R);
LATEX	2	1	1	-	2	12(1,T); 26(1,R);
LIQUID FUELS	2	1	1	-	2	11(1,T); 26(1,R);

QUADRO 7.2 - (Continuação)

DESCRIPTORIOS ASSINALADOS PELO MÉTODO SALT-S	F R E Q U Ê N C I A						O C O R R Ê N C I A Nº DOCUMENTO (FREQ., TÍTULO, RESUMO)
	ÚNICA	G E R A L					
		TÍTULO	RESUMO	TÍT & RES	TOTAL		
MANURES	2	-	2	-	2	2(1,R); 22(1,R);	
METHANE	5	1	5	-	6	3(1,R); 6(2,T,R); 22(1,R); 49(1,R); 52(1,R);	
METHANOL	1	-	1	-	1	52(1,R);	
NAPHTHA	1	1	2	1	3	13(3,T,R);	
NATIONAL ALCOHOL PROGRAM	1	-	1	-	1	52(1,R);	
ODORIZATION	1	-	1	-	1	13(1,R);	
OILS	2	-	2	-	2	20(1,R); 53(1,R);	
ORGANIC WASTES	1	-	1	-	1	63(1,R);	
PLANTS	3	2	1	-	3	20(1,T); 26(1,R); 28(1,T);	
POLYMERIZATION	1	-	1	-	1	32(1,R);	
PRODUCTION	7	4	6	3	10	13(1,R); 22(2,T,R); 50(2,T,R); 52(1,R); 63(1,T); 64(2,T,R); 66(1,R);	
PYROLISIS	1	1	1	1	2	63(1,T,R);	
PURIFICATION	2	2	1	1	3	6(2,T,R); 10(1,T);	
RENEWABLE ENERGY SOURCES	1	-	1	-	1	66(1,R);	
RESEARCH PROGRAMS	2	-	2	-	2	32(1,R); 43(1,R);	
REVIEWS	2	-	2	-	2	32(1,R); 43(1,R);	
RUBBERS	1	1	1	1	2	12(2,T,R);	
RURAL AREAS	1	-	1	-	1	47(1,R);	
SEWAGE SLUDGE	1	-	1	-	1	49(1,R);	

QUADRO 7.2 - (Conclusão)

DESCRIPTORIOS ASSINALADOS PELO MÉTODO SALE-S	F R E Q U Ê N C I A					O C O R R Ê N C I A Nº DOCUMENTO (FREQ., TÍTULO, RESUMO)
	ÚNICA	G E R A L				
		TÍTULO	RESUMO	TIT & RES	TOTAL	
SORGHUM	1	1	2	1	3	50 (3,T,R);
SOYBEAN OIL	1	-	1	-	1	57 (1,R);
STANDARDIZATION	1	-	1	-	1	47 (1,R);
TOWN GAS	1	-	1	-	1	13 (1,R);
USES	2	-	2	-	2	66 (1,R); 67 (1,R);
WOOD	3	1	3	1	4	31 (1,R); 64 (1,R); 67 (2,T,R);
WOOD FUELS	1	1	-	-	1	53 (1,T);

ANEXO 7 - (Conclusão)

QUADRO 7.3 - DESCRITORES SUGERIDOS ACEITOS PELO INDEXADOR (DSA)

- 1a. Coluna: Descritores sugeridos aceitos pelo indexador (DSA)
- 2a. Coluna: Frequência única
- 3a. Coluna: Nº de vezes que o descritor foi gerado a partir do título
- 4a. Coluna: Nº de vezes que o descritor foi gerado a partir do resumo
- 5a. Coluna: Nº de vezes que o descritor foi gerado a partir do título e resumo simultaneamente
- 6a. Coluna: Total
- 7a. Coluna: Para que documentos o descritor foi assinalado

DESCRITORES SUGERIDOS ACEITOS PELO INDEXADOR (DSA)	ÚNICA	F R E Q U Ê N C I A				O C O R R Ê N C I A Nº DO DOCUMENTO (FREQ., TÍTULO, RESUMO)
		G E R A L				
		TÍTULO	RESUMO	TIT & RES	TOTAL	
AIR	1	-	1	-	1	28 (1,R);
ALCOHOL FUELS	1	-	1	-	1	26 (1,R);
ALCOHOLS	1	1	-	-	1	50 (1,T);
CONTAINERS	1	1	2	-	2	10 (2,R);
CEMENT INDUSTRY	1	-	1	-	1	31 (1,R);
CHEMICAL PREPARATION	1	-	1	-	1	32 (1,R);
CLIMATES	2	-	2	-	2	50 (1,R); 51 (1,R);
CONSTRUCTION	1	-	1	-	1	2 (1,R);

QUADRO 7.3 - (Continuação)

DESCRIPTORES SUGERIDOS ACETOS PELO INDEXADOR (DSA)	F R E Q U Ê N C I A					O C O R R Ê N C I A Nº DO DOCUMENTO (FREQ., TÍTULO, RESUMO)
	ÚNICA	G E R A L				
		TÍTULO	RESUMO	TÍT & RES	TOTAL	
ENRICHMENT	2	-	2	-	2	13(1,R);22(1,R);
ENVIRONMENT	1	-	1	-	1	52(1,R);
FURNACES	1	-	1	-	1	28(1,R);
GAS GENERATORS	1	-	1	-	1	28(1,R);
GAS YIELD	1	-	1	-	1	22(1,R);
GLUCOSE	1	-	1	-	1	27(1,R);
GROWTH	1	-	1	1	1	22(1,R);
HAWAII	1	1	1	-	2	22(2,T,R);
HEAT COVERY	1	-	1	-	1	13(1,R);
HYDROLYSIS	1	-	1	-	1	27(1,R);
INDUSTRIAL WASTES	1	1	1	1	2	17(2,T,R);
IRON SULFIDES	1	-	1	-	1	10(1,R);
ISRAEL	1	-	1	-	1	17(1,R);
LAKES	1	-	1	-	1	20(1,R);
LOW BTU GAS	2	-	2	-	2	28(1,R);55(1,R);
PROCEEDINGS	2	2	-	-	2	31(1,T);47(1,T);
REDUCTION	1	-	1	-	1	13(1,R);
RURAL POPULATIONS	1	-	1	-	1	43(1,R);
SCREENS	1	-	1	-	1	10(1,R);

QUADRO 7.3 - (Conclusão)

DESCRITORES SUGERIDOS ACETITOS PELO INDEXADOR (DSA)	ÚNICA	F R E Q U Ê N C I A				Nº DO DOCUMENTO (FREQ., TÍTULO, RESUMO)
		G E R A L				
		TÍTULO	RESUMO	TÍT & RES	TOTAL	
SENEGAL	1	-	1	-	1	57 (1,R) ;
SPECTROSCOPY	1	-	1	-	1	57 (1,R) ;
STEAM GENERATION	1	-	1	-	1	13 (1,R) ;
STEAL INDUSTRY	1	-	1	-	1	64 (1,R) ;
SUGAR BEET	1	1	2	1	3	50 (3,T,R) ; 51 (1,R) ;
SUGAR CANE	1	-	1	-	1	50 (1,R) ; 51
SWINE	1	1	1	1	2	22 (2,T,R) ;
USA	1	1	1	1	2	27 (2,T,R) ;
VAPOURS	1	-	1	-	1	57 (1,R) ;
VEGETABLE OILS	1	1	2	1	3	57 (3,T,R) ;
WATER VAPOURS	1	-	1	-	1	28 (1,R) ;