

UNIVERSIDADE PRESBITERIANA MACKENZIE

Edson Tafeli Carneiro dos Santos

**GESTÃO ELETRÔNICA DE DOCUMENTOS:
GERENCIAMENTO DO CONHECIMENTO
PARA GRUPOS DE PESQUISAS**

São Paulo

2007

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

EDSON TAFELI CARNEIRO DOS SANTOS

**GESTÃO ELETRÔNICA DE DOCUMENTOS:
GERENCIAMENTO DO CONHECIMENTO
PARA GRUPOS DE PESQUISAS**

Dissertação apresentada à Universidade
Presbiteriana Mackenzie, como requisito
parcial para obtenção do título de
Mestre em Engenharia Elétrica.

Orientador: Prof Dr. Nizam Omar

São Paulo

2007

Aos meus pais, José(*in memorian*) e
Maria Lúcia (*in memorian*).
À minha esposa Adriana e
meu filho André
pela dedicação e amor.

Agradecimentos

Ao professor Dr. Nizam Omar pela sua orientação, compreensão e apoio, que tornaram este trabalho possível.

À minha esposa pelo apoio, amor e amizade que sempre recebi, até nos momentos mais difíceis. Ao meu filho, razão de meu viver.

À minha irmã e a todos meus familiares e amigos pelo carinho e incentivo.

Aos colegas da pós-graduação, especialmente os amigos Claudinei e Leonardo, pelo apoio e amizade nos vários fins de semana de estudos e resolução de listas de exercícios.

Aos professores e colegas da Universidade Presbiteriana Mackenzie, por terem contribuído para meu crescimento acadêmico e profissional, especialmente ao prof. Dr. Waldemar Monteiro por ser um grande incentivador e amigo nas horas difíceis.

Ao Instituto Presbiteriano Mackenzie pelo apoio financeiro por meio da bolsa auxílio cedida para o aperfeiçoamento constante de seus professores.

Aos meus alunos, que compartilharam com sua amizade ao longo desses anos.

A todos aqueles que mesmo não citados aqui, contribuíram para a realização desse trabalho.

Acima de tudo a Deus, proporcionado a mim força e perseverança ao longo destes anos e por conceder a benção da vida.

S237g Santos, Edson Tafeli Carneiro dos.

Gestão eletrônica de documentos : gerenciamento do conhecimento para grupos de pesquisas / Edson Tafeli Carneiro dos Santos – 2007.

183 f. : il. ; 30 cm.

Dissertação (Mestrado em Engenharia Elétrica) - Universidade Presbiteriana Mackenzie, São Paulo, 2007.

Bibliografia: f. 125-129.

1. Gerenciamento de documentos. 2. Recuperação de informação. 3. Ontologia. I. Título.

CDD 658.4038

RESUMO

A evolução dos sistemas de informação, dos sistemas gerenciadores de banco de dados, da inteligência artificial e das redes neurais, dentre outras tecnologias, contribuíram para que os sistemas computacionais pudessem manipular uma enorme quantidade de dados, e por consequência, a geração de informação aumentou após a última metade do século XX. A nova configuração do mundo globalizado e o surgimento da era da informação propiciaram, de uma forma singular e não vivenciada em outras épocas, uma vasta quantidade de informações. Assim, uma informação que necessite ser acessada, pode ser obtida de forma mais eficiente com o auxílio de ferramentas computacionais especializadas em gestão de documentos. Também, em um grupo de pesquisas, os dados que são trabalhados e as informações técnicas geradas, promovem uma massa de informações e conhecimentos que merece ser gerenciada. Dessa forma, neste trabalho, pretende-se adaptar às funcionalidades existentes de uma ferramenta de gerenciamento eletrônico de documentos, por meio de uma taxonomia de conceitos e áreas pertinentes, os documentos e o conhecimento de determinado grupo de pesquisa, para que possam ser gerenciados.

Palavras-chave: Gerenciamento de Documentos; Recuperação de Informação; Ontologia.

ABSTRACT

The evolution of information and data base management systems, artificial intelligence and neural networks, among other technologies, had contributed to the manipulation of an enormous amount of data and information by computational systems, what resulted in an increase of the information generated after the last half of XX century. The new configuration of the global world and the emerged information age had propitiated a vast amount of available information not ever seen at other times. Thus, information that needs to be accessed can be recovered more efficiently with the help of a specialized computational tool in document management. Also, in a research group, the data manipulated and the scientific and technical information generated by members of this group, promote a mass of information and knowledge that deserves to be managed. This work contemplates the management of documents and information of a determined research group, by means of the adaptation of concepts and pertinent areas taxonomy to the functionalities of an existing electronic document management tool, so that this information can be managed.

Keywords: Document Management, Information Retrieval, Ontology.

LISTA DE FIGURAS

Figura 1	Ciclo de vida de documentos, da criação ao descarte.....	12
Figura 2	Situações de trabalho em um grupo	14
Figura 3	Classificação de Ontologias	26
Figura 4	Arquitetura de um aplicativo servlet.....	32
Figura 5	Arquitetura de um aplicativo <i>servlet</i> juntamente com um servidor HTTP.....	32
Figura 6	Arquitetura Geral de um Mecanismo de Busca	34
Figura 7	Curva da Frequência das palavras x Ordem de Importância	36
Figura 8	Similaridade de documentos no modelo vetorial	40
Figura 9	Representação da similaridade entre documentos	41
Figura 10	Processo de Avaliação de um Sistema de Recuperação de Informação..	48
Figura 11	Arquitetura da ferramenta original EDMIS	50
Figura 12	Ontologia Base do Sistema REDMIS.....	56
Figura 13	Diagrama da classe docCientífico.....	58
Figura 14	Diagrama da classe Pesquisador.....	59
Figura 15	Diagrama da classe Obra.....	59
Figura 16	Diagrama da classe Autor.....	60
Figura 17	Diagrama da classe ÁreadePesquisa.....	61
Figura 18	Diagrama da classe Conceito.....	61
Figura 19	Diagrama da classe Livro.....	62
Figura 20	Diagrama da classe Instituição.....	62
Figura 21	Diagrama da classe Título.....	63
Figura 22	Modelo de dados do sistema REDMIS que alimenta os termos para a representação ontológica fuzzy.....	64
Figura 23	Visão da Arquitetura do sistema REDMIS sob ponto de vista de camadas.....	65
Figura 24	Mecanismo de indexação de documentos em REDMIS.....	66
Figura 25	Mecanismo de recuperação de documentos em EDMIS.....	67

Figura 26	Mecanismo de recuperação de documentos em REDMIS.....	67
Figura 27	Ontologia de representação de termos para elaboração da consulta.....	69
Figura 28	Cálculo da similaridade de rede de termos fuzzy.....	70
Figura 29	Mapa Conceitual não Hierárquico das Linhas de Pesquisa do Grupo CRAAM.....	76
Figura 30	Mapa Conceitual Hierárquico das Linhas de Pesquisa do Grupo CRAAM.....	77

LISTA DE GRÁFICOS

Gráfico 1	Documentos recuperados para o pesquisador P1 para os pontos de corte alto e baixo e com 70% da sua ontologia	94
Gráfico 2	Documentos recuperados para o pesquisador P1 para os pontos de corte alto e baixo e com 100% da sua ontologia.....	95
Gráfico 3	Documentos recuperados para a o pesquisador P1 para o ponto de corte baixo.....	96
Gráfico 4	Documentos recuperados para o pesquisador P2 para os pontos de corte alto e baixo e com 70% da sua ontologia	97
Gráfico 5	Documentos recuperados para o pesquisador P2 para os pontos de corte alto e baixo e com 100% da sua ontologia.....	98
Gráfico 6	Documentos recuperados para a o pesquisador P2 para o ponto de corte baixo.....	99
Gráfico 7	Documentos recuperados para o pesquisador P3 para os pontos de corte alto e baixo e com 70% da sua ontologia	100
Gráfico 8	Documentos recuperados para o pesquisador P3 para os pontos de corte alto e baixo e com 100% da sua ontologia.....	101
Gráfico 9	Documentos recuperados para a o pesquisador P3 para o ponto de corte baixo.....	102
Gráfico 10	Documentos recuperados para o pesquisador P4 para os pontos de corte alto e baixo e com 70% da sua ontologia	103
Gráfico 11	Documentos recuperados para o pesquisador P4 para os pontos de corte alto e baixo e com 100% da sua ontologia.....	104
Gráfico 12	Documentos recuperados para a o pesquisador P4 para o ponto de corte baixo.....	105
Gráfico 13	Documentos recuperados para o pesquisador P5 para os pontos de corte alto e baixo e com 70% da sua ontologia	106
Gráfico 14	Documentos recuperados para o pesquisador P5 para os pontos de corte alto e baixo e com 100% da sua ontologia.....	107

Gráfico 15	Documentos recuperados para a o pesquisador P5 para o ponto de corte baixo.....	108
Gráfico 16	Documentos recuperados para o pesquisador P6 para os pontos de corte alto e baixo e com 70% da sua ontologia	109
Gráfico 17	Documentos recuperados para o pesquisador P6 para os pontos de corte alto e baixo e com 100% da sua ontologia.....	110
Gráfico 18	Documentos recuperados para a o pesquisador P6 para o ponto de corte baixo.....	111
Gráfico 19	Documentos recuperados para o pesquisador P7 para os pontos de corte alto e baixo e com 70% da sua ontologia	112
Gráfico 20	Documentos recuperados para o pesquisador P7 para os pontos de corte alto e baixo e com 100% da sua ontologia.....	113
Gráfico 21	Documentos recuperados para a o pesquisador P7 para o ponto de corte baixo.....	114
Gráfico 22	Documentos recuperados para o pesquisador P8 para os pontos de corte alto e baixo e com 70% da sua ontologia	115
Gráfico 23	Documentos recuperados para o pesquisador P8 para os pontos de corte alto e baixo e com 100% da sua ontologia.....	116
Gráfico 24	Documentos recuperados para a o pesquisador P8 para o ponto de corte baixo.....	117
Gráfico 25	Documentos recuperados para o sub-grupo A.....	118
Gráfico 26	Documentos recuperados para o sub-grupo B	119

LISTA DE TABELAS

Tabela 1	Pesquisador P1 e seus termos associados	84
Tabela 2	Pesquisador P2 e seus termos associados.....	85
Tabela 3	Pesquisador P3 e seus termos associados.....	86
Tabela 4	Pesquisador P4 e seus termos associados.....	87
Tabela 5	Pesquisador P5 e seus termos associados.....	88
Tabela 6	Pesquisador P6 e seus termos associados.....	89
Tabela 7	Pesquisador P7 e seus termos associados.....	90
Tabela 8	Pesquisador P8 e seus termos associados.....	91
Tabela 9	Documentos recuperados por pesquisador e seu perfil ontológico.....	120

SUMÁRIO

Resumo	
Abstract	
Lista de Figuras	i
Lista de Gráficos	iii
Lista de Tabelas	v
Sumário.	vi
1	INTRODUÇÃO..... 1
1.1	Objetivo 3
1.1.1	<i>Objetivos Específicos</i> 3
1.2	Motivação..... 4
1.3	Resultados Alcançados 6
1.4	Organização do Trabalho 6
2	CONCEITOS ENVOLVIDOS 9
2.1	Gerenciamento Eletrônico da Informação por Ferramentas GED 10
2.2	Ambientes de Trabalho Colaborativo 13
2.2.1	<i>Trabalho Colaborativo</i> 13
2.2.2	<i>Características Funcionais de ambientes CSCW</i> 15
2.2.3	<i>Colaboração suportada por computador</i> 18
2.3	Conceito de Comunidade Virtual 20
2.4	Gerencia do Conhecimento 22
2.5	Ontologias 25
2.6	Tecnologias Associadas 30
2.6.1	<i>Modelo de Dados</i> 30
2.6.2	<i>Linguagem Java</i> 31
2.6.3	<i>Sistema de Serolets</i> 31
2.6.4	<i>Servidores JSP</i> 33

2.7	Recuperação da Informação	33
2.7.1	<i>Modelos de Recuperação da Informação</i>	37
2.7.1.1	<i>Modelo Booleano</i>	38
2.7.1.2	<i>Modelo Vetorial</i>	39
2.7.1.3	<i>Modelo Probabilístico</i>	42
2.7.1.4	<i>Modelo Fuzzy</i>	44
2.8	Avaliação de Relevância em Recuperação da Informação	46
3	O SISTEMA REDMIS	49
3.1	Arquitetura do Sistema Base “EDMIS”	50
3.2	O Sistema REDMIS	53
3.3	Metodologia	54
3.4	Ontologia do Sistema REDMIS.....	56
3.4.1	<i>Ontologia docCientifico</i>	57
3.4.2	<i>Ontologia Pesquisador</i>	58
3.4.3	<i>Ontologia Obra</i>	59
3.4.4	<i>Ontologia Autor</i>	60
3.4.5	<i>Ontologia Áreas de Pesquisa</i>	60
3.4.6	<i>Ontologia Conceito</i>	61
3.4.7	<i>Ontologia Livro</i>	61
3.4.8	<i>Ontologia Instituição</i>	62
3.4.9	<i>Ontologia Título</i>	62
3.5	Modelo de Dados	63
3.6	Arquitetura do Sistema REDMIS	65
3.7	Módulo de Classificação Ontológica.....	68
3.7.1	<i>Geração da Ontologia de Termos para Consulta</i>	68
3.7.2	<i>Cálculo da Similaridade</i>	69
4	ESTUDO DE CASO	72
4.1	Introdução	72
4.2	Abordagem do Problema	73

4.3	Características do Grupo Associadas ao Sistema	74
4.4	Taxonomia de Classificação por meio de Mapas Conceituais	75
4.5	Linhas de Pesquisa dos Membros do Grupo	78
4.6	Utilização da Ferramenta com Base nos Perfis e Documentos	79
4.7	Condições dos Ensaios	80
4.8	Busca do Conhecimento: Recuperação da Informação	81
4.8.1	<i>Elaboração dos Termos Ontológicos Fuzzy para Consulta</i>	82
4.9	Recuperação da Informação Obtida pela Ferramenta	91
4.9.1	<i>Resultados Obtidos pelo Sistema</i>	92
4.9.2	<i>Grupos de Pesquisa por Afinidade de Tema e Ontologia</i>	118
5	CONCLUSÃO	122
5.1	Conclusões	122
5.2	Contribuições	123
5.3	Perspectivas para Trabalhos Futuros	124
	REFERÊNCIAS BIBLIOGRÁFICAS	125
	ANEXO A - Banco de Dados da Ferramenta EDMIS	130
	APÊNDICE A - Lista dos Documentos Utilizados para Verificação da Ontologia Proposta.....	I
	APÊNDICE B - Conjunto de <i>Stop Words</i> incorporadas ao Sistema REDMIS .	VII
	APÊNDICE C - Mapas Conceituais de Cada Pesquisador do Grupo de Teste	XXII
	APÊNDICE D - Resultados das Consultas pelos Critérios de Ontologia e Pontos de Corte	XXXII

Capítulo

1

Introdução

Grupos de pesquisa necessitam, no que diz respeito às características e a complexidade das atividades desenvolvidas, de um conhecimento refinado e especializado. A atividade de pesquisa requer para seu desenvolvimento um acesso rápido e fácil de novas informações. Verifica-se que o sucesso de uma atividade de um grupo de pesquisa está vinculado à sua habilidade de gerenciar o conhecimento e reter experiências, e da mesma forma um dos fatores relevantes é a arte de “como fazer”, adquiridos ao longo do tempo de pesquisa (Lichtnow et al. , 2001).

A informação passa a ser um fator de sucesso em várias atividades, tanto no meio acadêmico, como no meio comercial. Dessa forma, por exemplo, o crescimento vertiginoso ocorrido na *Word Wide Web* disponibiliza a quem necessitar uma vasta quantidade de informações acerca dos mais variados temas. A dificuldade de recuperar a informação em documentos com relevância passa a ser um dos principais problemas dos sistemas de informação (Huberman et al., 1998). Devido aos fatos mencionados a necessidade da gerência de documentos é hoje observada

em todos os ramos de atuação, sejam científicos, empresariais ou pessoais. Atendendo essas exigências, neste trabalho foi elaborada e construída uma ferramenta para o gerenciamento do conhecimento, que possui mecanismos mínimos para o gerenciamento de documentos e recuperação da informação. A ferramenta EDMIS (*Electronic Document Management Intelligent System*) proposta por Alves (2005) tem como meta ser uma plataforma para a gestão documental, independente do tipo de documento e sua classificação acadêmica. É na realidade uma ferramenta GED (Gerência Eletrônica de Documentos) de propósito geral, com as funcionalidades necessárias para a gestão documental acadêmica.

Esta dissertação propõe uma adaptação do sistema EDMIS para sua adequação às necessidades de recuperação e gestão documental de grupos de pesquisa. Para estudo de caso dessa nova ferramenta e verificação de suas funcionalidades, serão utilizados como carga de teste os documentos do grupo de Rádio-Ciências Espaciais da Universidade Presbiteriana Mackenzie, denominado “Centro de Radio Astronomia e Astrofísica do Mackenzie” (CRAAM). Em relação às diversas áreas do conhecimento em um grupo de pesquisa, é natural que um ou mais pesquisadores atuem em mais de uma área ou em áreas correlatas. Isso proporciona uma grande variedade de possíveis documentos (artigos, relatórios, etc.) que poderão ser utilizados por outros pesquisadores do grupo. Para que a ferramenta possa administrar as pessoas, os conceitos e as áreas para a gerência dos documentos com os quais o sistema irá trabalhar, foi criado um banco de dados com informações sobre os pesquisadores, concebido por meio do uso de ontologias e de uma taxonomia de classificação.

1.1 Objetivo

O objetivo desse trabalho é incorporar à ferramenta de gestão eletrônica de documentos EDMIS (Alves, 2005) a capacidade de representação do conhecimento de grupos de pesquisa com base na sua representação ontológica. A gestão documental e a recuperação da informação decorrente classificarão os documentos com base na sua correspondência e relevância, levando-se em conta a classificação ontológica na qual o documento e o pesquisador tenham correspondência.

1.1.1 Objetivos Específicos

- Desenvolver uma ontologia de domínio, por meio do conceito de similaridade, para representar e classificar o conhecimento de um grupo de pesquisas;
- Elaborar na nova ferramenta um módulo "*Facilities*", para cada pesquisador usuário do sistema. Este módulo proporcionará ao pesquisador, via o sistema, uma integração de dados e documentos pertinentes ao grupo e da área de interesse dos pesquisadores com um grau de relevância, descartando os documentos não relevantes para seu perfil, bem como outras informações específicas relacionados à utilização do grupo de pesquisa;
- Verificar no sistema desenvolvido a sua capacidade de recuperação de informação, por meio de um estudo de caso, utilizando-se a ferramenta para os perfis e documentos de um grupo de pesquisadores.

1.2 Motivação

Com o aparecimento de novas tecnologias em relação aos sistemas gerenciadores de banco de dados (SGBD's), notadamente na última década do século passado, na área de banco de dados, apareceram ferramentas para realizar a recuperação e o armazenamento de elevados volumes de dados com segurança e eficiência. Outro fator relevante é que o conteúdo disponível na Internet duplica anualmente e as pessoas que incluem novos dados, em sua maioria, não sabem como funcionam os sistemas para recuperação de informações. Associado ao aumento do volume de conteúdo, vem também o aumento das fontes em outros idiomas que não o Inglês. Assim como nos primeiros anos da Internet, o inglês ainda é o idioma predominante, mas não tanto como no princípio. Atualmente, o número dos usuários da Internet que são falantes nativos do inglês já se restringe a 50% (Lyman et al., 2000). Estima-se que seja acima de 5 bilhões de palavras o tamanho da Internet em português (Aires e Santos, 2002).

A sociedade passa a ter um novo desafio, devido ao acúmulo de informação gerado por essas novas tecnologias. A necessidade de capturar e armazenar dados de todos os tipos aumentou de tal maneira, que atualmente não é mais possível com a capacidade humana, gerenciar grandes volumes de dados. Mecanismos de análise de dados, baseados na capacidade humana, não conseguem obter todo conhecimento possível. Nossa capacidade de analisá-los, resumi-los e deles extrair conhecimento é lenta e limitada. Este fato exige dos novos bancos de dados ferramentas mais inteligentes que possam auxiliar na descoberta e na extração de conhecimento destes

conjuntos volumosos de dados. Alves (2005, p.72) aponta alguns motivos, citados abaixo, que justificam a economia de tempo e valor que decorre da gestão de documentos:

- *“Dentre cada 20 documentos, um se perde, segundo Coopers & Lybrand”.*
- *“Executivos americanos gastam em média 150 horas/ano (aproximadamente 6 dias) procurando, localizando e solicitando documentos”.*
- *“As pessoas de modo geral perdem cerca de 20% a 30 %, de seu tempo na localização e recuperação de documentos”.*

Estamos inseridos em tempos em que a produção de informação nos últimos cinco milênios é equivalente à quantidade de informação produzidas na última metade do século XX, e em 2010 essa informação duplicará a cada meio dia (Andrade, 2002). Conforme Andrade (2002), as ferramentas GED (Gerenciamento Eletrônico de Documentos) são um mecanismo de apoio à busca da informação, uma vez que, é uma soma de tecnologias, que objetivam gerenciar a informação e também trabalhar com o documento desde sua aquisição até o seu posterior arquivamento, independente da mídia utilizada. A produção de conhecimento científico também tem como característica o elevado volume de dados. As ferramentas GED são uma forma de gerenciar eficientemente todos os documentos gerados, e que poderá possibilitar a geração de novos conhecimentos ou ainda disseminar dentro da organização os conhecimentos adquiridos.

1.3 Resultados Alcançados

Foi implementado um sistema de gerenciamento de documentos utilizando ontologias para descrição das características presentes em um grupo de pesquisas, propiciando a capacidade de recuperar documentos por similaridade e correspondência ao perfil do pesquisador. Esta pesquisa resultou na publicação do seguinte trabalho:

SANTOS, E.T.C; OMAR, N. REDMIS: Uma ferramenta para gerenciamento do conhecimento e documentos de um grupo de pesquisas em radioastronomia. In: 3º SIMPÓSIO MINEIRO DE SISTEMAS DE INFORMAÇÃO (3º SMSI-SBC), PUC-MINAS, Belo Horizonte-MG, Setembro, 2006. CD-ROM.

1.4 Organização do Trabalho

Este trabalho está organizado da seguinte forma: neste Capítulo foi apresentada uma visão geral do problema a ser investigado, o contexto associado ao tema, as motivações e os objetivos.

O capítulo 2 trata dos fundamentos teóricos sobre os conceitos envolvidos nessa dissertação tratados pela literatura sobre:

- Gerência eletrônica de documentos (GED), procurando mostrar seus princípios assim como a escolha e caracterização de um sistema GED;
- Ambientes Colaborativos (*Collaborative Environment*);
- Gerência do Conhecimento (*Knowledge Management*);
- Comunidades Virtuais (*Virtual Communities*);

- Uso de Mapas Conceituais para representação gráfica das relações do conhecimento;
- Uso de Ontologias para classificação, representação e detalhamento das entidades envolvidas pertencentes a um grupo de pesquisa, desde o perfil dos pesquisadores no que se refere as suas áreas de pesquisa, e seus documentos de uso freqüente e relacionados com a(s) sua(s) área(s) de pesquisa(s) (essa modelagem servirá como propósito para implementação de um sistema de recuperação de informação como base nessas informações);
- Recuperação da Informação; nesse tópico são analisados os modelos clássicos do tratamento e recuperação da informação assim como medidas de avaliação de relevância.

O capítulo 3 descreve o sistema REDMIS (*Rapid Electronic Document Management Intelligent System*). Apresenta a evolução e adaptação das funcionalidades da ferramenta de gestão eletrônica de documentos proposta por Alves (2005). Aborda-se a aplicação e as funcionalidades incorporadas a esta ferramenta adaptada a um grupo de pesquisas. A implementação dos módulos adicionados ao sistema original são discutidos nesse capítulo.

Na seqüência, o capítulo 4 apresenta a análise do comportamento da ferramenta no utilizando como estudo de caso a representação do grupo de pesquisadores em Radioastronomia da UPM, juntamente com os documentos gerados ou utilizados por estes. É realizada uma descrição comportamental e

funcional, por meio de ontologias aplicadas à área em estudo. São abordados aspectos dos componentes do grupo e conceitos sobre sua área de pesquisa, relacionados à uma calibração inicial da ferramenta.

O capítulo 5 apresenta a conclusão, a síntese das contribuições e as sugestões para trabalhos futuros.

Capítulo

2

Conceitos Envolvidos

Este capítulo é uma revisão da literatura a respeito dos princípios e técnicas utilizadas em sistemas de Gerência Eletrônica de Documentos e Recuperação da Informação. Inicialmente é feita uma explanação sobre o gerenciamento eletrônico de documentos, em seguida são abordados os conceitos sobre comunidades virtuais, ambientes colaborativos, gerência do conhecimento, mapas conceituais, ontologias e finalmente são apresentados os modelos de recuperação da informação. Esses conceitos são utilizados na descrição da arquitetura do sistema proposto nesta pesquisa.

2.1 Gerenciamento Eletrônico da Informação por ferramentas GED- “Gerenciamento Eletrônico de Documentos”.

Segundo Davenport e Prusak (1998), com o avanço da informática e evolução das ferramentas de administração do conhecimento, notadamente os pacotes de Software, foi possível lidar com os mais diversos tipos de mídias. Terra (2001) menciona que de acordo com a visão de Rheinner Group a Gerência do conhecimento pode ser feita de acordo com as seguintes abordagens:

- Gestão de Competências;
- Gestão de Conteúdo;
- Aprendizagem Organizacional;
- Inteligência Competitiva; e
- Educação Corporativa.

Esses são os fundamentos no qual a tecnologia GED (Gerenciamento Eletrônico de Documentos) pode auxiliar fornecendo os subsídios e sustentação à gestão de conteúdo. Davenport e Prusak (1998) mostram que os sistemas GED são essencialmente repositórios de conhecimento explícito, conceito este discutido mais adiante. Esses sistemas funcionam como “sistemas de gerenciamento de arquivos”, porém muito além de armazenar arquivos, eles devem garantir segurança e acesso e implementar a taxonomia de documentos, dentre outras funcionalidades (Andrade, 2002).

Andrade (2002) e Rios (2005) discutem sobre o ciclo de vida de um documento, que trata desde sua criação até o seu descarte. A frequência de uso (acesso) típica de um documento ao longo do tempo é discutida a seguir:

- **Captura e Criação:** Utiliza as tecnologias de T.I. para a digitalização da informação.
- **Armazenamento e Organização:** Diferentes tecnologias de informação impõem como o documento será armazenado e organizado. Sistemas de bancos de dados tradicionais, não atendem as exigências de armazenamento de dados multimídia.
- **Roteamento:** Re-direcionamento do documento para um outro usuário do grupo, antes de publicá-lo.
- **Recuperação e Síntese:** Uma vez no banco de dados o documento pode ser recuperado por meio da pesquisa de seu conteúdo. Tipicamente os recuperadores de texto utilizam lógica Booleana. Porém pode haver a combinação de palavras chave obtidas no documento ou todo o texto (atributos do texto). Técnica de Thesaurus para palavras próximas são também utilizadas.
- **Visualização e impressão:** Embora o documento esteja em formato eletrônico, há interesse, dependendo da utilização da informação de que o documento possa ser impresso. Para isso, há um conjunto de tecnologias disponíveis no mercado, que vão de impressoras de impacto até as impressoras laser ou ainda plotadoras para documentos CAD.

- **Descarte:** Esse é momento em que o documento deixa de ser útil, e deve ser descartado levando-se em conta sua importância e tempo de uso.

A figura 1 apresenta a frequência típica de utilização de um documento (acesso) em função do tempo.

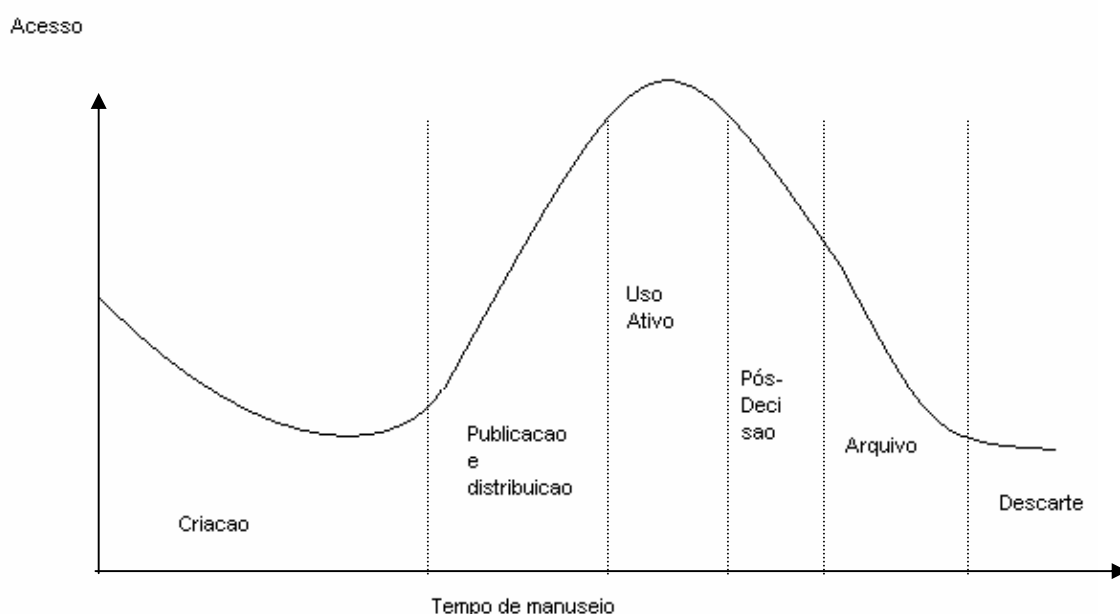


Figura 1 - Ciclo de vida de documentos, da criação ao descarte. Fonte: (CENADEM, 2004).

Durante a fase de criação do documento até o amadurecimento de seu conteúdo, seu acesso é moderado, mas uma vez pronto e disponibilizado para utilização por outras pessoas além seus autores (publicação e distribuição), verifica-se um acréscimo na sua utilização enquanto fonte de conhecimento (uso ativo). Extraídos seus conceitos e idéias principais declina-se o seu uso (pós-decisão) levando-se a etapa de arquivamento e verificado sua obsolência decide-se leva-lo ao descarte que é o termino da vida do documento.

2.2 Ambientes de Trabalho Colaborativo

São raras as atividades humanas que possam ser realizadas integralmente por uma só pessoa. A maioria dessas atividades requer a participação de grupos na sua execução. Nesse sentido, para que os indivíduos situados num mesmo local ou geograficamente distribuídos tenham um ambiente de suporte para sua interação, o auxílio de ferramentas computacionais que ofereçam meios de comunicação e interação, oferecem um ambiente virtual para uma maior aproximação e colaboração entre as pessoas. Esse ambiente, que propicia os mecanismos de interação, é um ambiente computacional denominado de *groupware* ou sistema de trabalho cooperativo apoiado por computador (CSCW- *Computer Supported Cooperative Work*) (Borges et al., 1995).

2.2.1 Trabalho Colaborativo

Como apresentado na seção anterior, os sistemas computacionais CSCW ou *groupware* são ambientes que permitem que processos sejam implementados por computador tendo em vista o apoio ao grupo de trabalho (Borges et al., 1995).

Grupos são uma coleção de pessoas que trabalham juntas numa tarefa comum e os grupos podem ser formados por um número fixo ou variável de indivíduos onde cada um tem diferentes conhecimentos, objetivos e habilidades, entretanto, contribuem na realização da tarefa comum (Macaulay, 1995 p.136). Devemos observar que nem sempre os grupos de trabalho estão no mesmo local para a realização das suas atividades. Nessa situação os sistemas CSCW tornam-se vitais,

fato este que não se observa em encontros face a face. Ainda há a possibilidade dos membros se encontrarem no sistema ao mesmo tempo (realização de tarefa síncrona) , ou trabalhar em momentos distintos (realização de tarefa assíncrona). Grupos de trabalho nem sempre estão no mesmo local para discutirem e trabalharem em encontros do tipo face-a-face. Eles podem estar geograficamente distribuídos e, nesse caso, a comunicação via sistemas CSCW torna-se essencial e não facultativa como nos casos de encontros face-a-face. Os membros de um grupo podem optar por se conectarem ao sistema ao mesmo tempo (trabalho síncrono), ou eles podem trabalhar numa mesma tarefa, mas em tempos diferentes (trabalho assíncrono) (Macaulay, 1995). A figura 2 apresenta as possíveis quatro situações de que um grupo pode estar em função da sua distribuição geográfica e temporal.

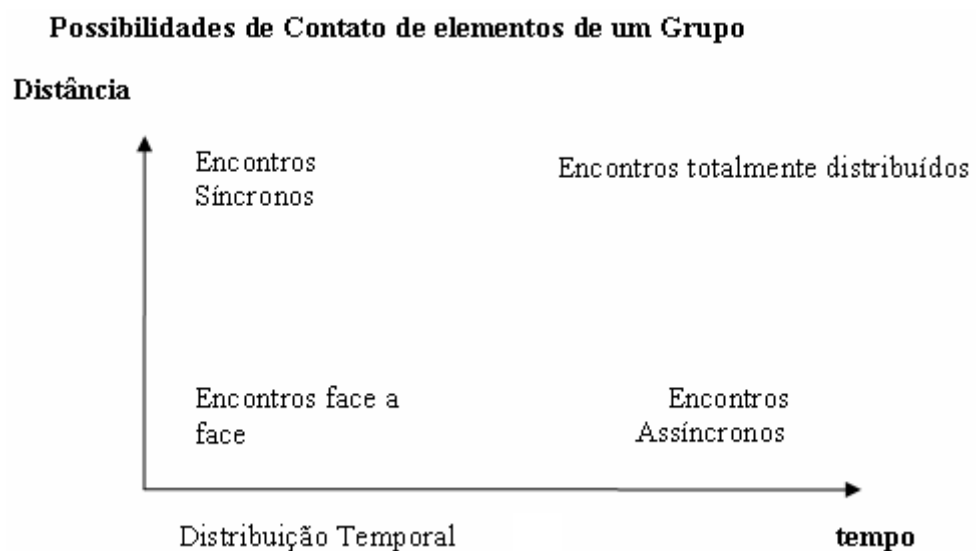


Figura 2 - Situações de trabalho em um grupo. Fonte: (Macaulay, 1995).

De acordo com a figura-2, encontros face-a-face ocorrem ao mesmo tempo em um mesmo lugar, em contrapartida, encontros assíncronos acontecem em tempos

diferentes, mas em um mesmo lugar. Por sua vez, os encontros síncronos distribuídos ocorrem ao mesmo tempo e em lugares distintos, e sendo que os encontros totalmente distribuídos possuem diferentes tempos e lugares para sua realização.

2.2.2 Características funcionais de ambientes CSCW

De acordo com Macaulay (1995) existem algumas características funcionais que podem estar presentes em sistemas CSCW, conforme mencionados a seguir:

- **Comunicação e Compartilhamento de Informações.**

A comunicação é de extrema importância em situações de trabalho em grupo, seja ela para transmissão de informações ou para tomada de decisões. Muitas vezes, decisões podem ser alcançadas num local onde as pessoas podem ver e sentir as reações e comportamento uma das outras. Por exemplo, uma pessoa bocejando pode estar indicando desinteresse, ou alguém balançando as pernas pode estar manifestando ansiedade. Enfim, qualquer tipo de manifestação é importante em uma discussão, decisão ou qualquer tipo de trabalho em grupo. Procurando manifestar todas as ações dos membros de um grupo em encontros, tem se tentado utilizar canais de áudio e vídeo nos qual o conteúdo de aparência, fala e escrita são transmitidos para todos os membros do grupo. Antes da popularização das *web cams*, a maioria das comunicações via computador eram restritas a canais textuais onde o usuário podia apenas ler e escrever mensagens. A comunicação podia ser síncrona ou assíncrona dependendo do tempo em que

os membros do grupo recebiam o conteúdo das mensagens enviadas. Atualmente, exemplos de sistemas CSCW que representam uma característica de comunicação áudio-visual mais ampla, possuem canais de vídeo e voz e sistemas de vídeo conferência .

- **Compartilhamento de informações**

Essa é uma característica essencial para grupos devido à necessidade de prevenir esforços repetitivos e assegurar que todos os membros do grupo estão utilizando a mesma informação. É importante notar que nem sempre todas as informações são compartilhadas pelos componentes do grupo, pois alguns dados podem ser de acesso exclusivo, por exemplo, de um coordenador do trabalho. As informações podem estar em formato de texto, números, gráficos, imagens ou vídeo. Grupos precisam de facilidades que lhes assegurem entrada, armazenamento, navegação e recuperação das informações úteis a todos os membros do grupo. Exemplos de tecnologias que facilitam o compartilhamento de informações e podem ser utilizadas por sistemas CSCW são recursos multimídia, hiperdocumentos multi-usuários, base de dados multi-usuários, sistemas de CD-ROM e outros.

- **Coordenação e controle de objetos.**

Em muitas situações, certos objetos devem ser compartilhados entre os membros de um grupo devido a algumas razões. Para trabalho cooperativo suportado por computador, as versões antigas do trabalho ou apenas as atualizações imediatas

devem ser apresentadas a todos os membros do grupo. Por exemplo, membros de um grupo de projetistas, ao desenvolver um projeto, estão isoladamente ou em grupo realizando modificações no projeto a todo instante. No caso de papéis, a cada nova versão do projeto, a versão anterior é jogada fora ou arquivada. Sendo de extrema importância que os outros membros do grupo estejam cientes de cada nova versão do projeto. Exemplos de tipos de sistemas CSCW que suportam a característica de coordenação e controle de objetos compartilhados são válidos para encontros síncronos e assíncronos, que incluem facilidades de planejamento, alocação de tarefas e troca de informações entre os membros do grupo.

- **Compartilhamento de espaços de trabalho.**

Encontros face-a-face são geralmente auxiliados por uma lousa (ou quadro), ou por computadores onde as pessoas escrevem suas idéias e os outros membros do grupo podem vê-las. Portanto, tanto a lousa ou a tela do computador são consideradas espaços de trabalho que podem ser compartilhados. O resultado de exercícios de “*brainstorming*” podem ser armazenados numa lista que pode ser trabalhada pelos membros do grupo. Uma pessoa pode ser responsável por escrever, mas todas devem ser capazes de ler, fazer sugestões e expor as críticas e seus pontos de vista. Exemplos de tecnologias que suportam a característica de compartilhamento de espaços de trabalho são os sistemas *WYSIWIS* (*What You See Is What I See*), projetores em salas de encontros face-a-face.

- **Organização e entendimento do processo de trabalho.**

Pessoas que trabalham juntas, normalmente, sabem o que e como elas devem fazer certas tarefas. As pessoas devem definir a tarefa a ser realizada, entrar em acordo sobre um conjunto de atividades a ser feito para a realização da tarefa e por fim comunicar a todos as decisões. Para isso, encontros devem ser agendados para que cada membro do grupo possa saber das intenções dos outros. Alguns sistemas CSCW que dão apoio à organização e entendimento do processo de trabalho são os co-editores, as ferramentas de gerenciamento de agendas e os sistemas de *Workflow*. Sistemas *WYSIWIS* exploram janelas públicas para implementar situações onde um escreve na tela do computador e os outros podem visualizar imediatamente na tela de seus computadores ou terminais o que o primeiro escreveu. Como exemplo de ambientes CSCW de domínio público disponíveis na *WWW* pode-se citar o IRC, o HIPERNEWS, o MSN, etc.

2.2.3 Colaboração Suportada por Computador

Conforme apresentado por Santos e Ferreira (1998), no final da década de 60, surgiram as primeiras redes de computadores conhecidas como Redes Locais (LANs - *Local Area Network*). As LANs tinham a função de permitir o compartilhamento de dados entre pequenos computadores separados por pequenas distâncias. Entretanto, estas redes eram problemáticas, pois haviam sido instaladas com diferentes tecnologias e, assim, não podiam ser interconectadas. Paralelamente ao desenvolvimento das LANs, surgiu um outro tipo de rede que era denominada de Redes de Longa Distância (WAN - *Wide Area Network*). Redes WAN conectavam computadores geograficamente distantes. Infelizmente LANs e WANs ainda

continuavam incompatíveis (Tanenbaum, 1997). Em 1970 nasceu o projeto ARPA (*Advanced Research Project*), que teve como um de seus resultados a possibilidade de interconexão de diferentes tipos de rede de computadores fossem elas LANs ou WANs. A implementação dessa interconexão foi definida como *internet work*, ou, Internet. A pesquisa na área de CSCW não é recente visto que desde o início da década de 60, Douglas C. Engelbart explorava o uso de computadores para o trabalho com grupos de alto desempenho.

Na década seguinte, com o advento dos minicomputadores, surgiu a Automação de Escritórios (*Office Automation*), que teve como objetivo adaptar e integrar as aplicações mono-usuário mais bem sucedidas, como os processadores de texto e as planilhas eletrônicas, para suportarem o trabalho em grupo. Somente o uso da tecnologia computacional não foi suficiente para esta transformação. Havia a necessidade de se estudar como as pessoas trabalham em grupo e como a tecnologia influi neste processo. Assim, os tecnólogos começaram a aprender sobre a atividade de grupo com economistas, sociólogos, antropólogos, cientistas organizacionais e educadores. Começaram a surgir neste ponto os primeiros trabalhos em direção ao Trabalho Cooperativo Suportado por Computador. Em 1984, o termo CSCW foi usado pela primeira vez quando Paul Cashman e Irene Greif organizaram um *workshop* para reunir pessoas de áreas diferentes, mas com um interesse comum em trabalho em grupo e como a tecnologia pode suportá-lo (Grudin, 1994). Os trabalhos em CSCW não param por aí. Além de vários outros eventos específicos da área, podem-se encontrar artigos da área em eventos de HCI, Hipermídia e Multimídia. Pesquisas, artigos e trabalhos em CSCW podem se encontrados em toda parte, porém

vale a pena ressaltar as grandes contribuições prestadas pelos pesquisadores japoneses, a partir de empresas de computadores e softwares como a NEC e a Toshiba, além das companhias de telecomunicações.

2.3 Conceito de Comunidade Virtual

A computação mediada por computador (CMC) vem transformando nossa maneira de agir e se comportar, pois comunidades surgiram com apoio da CMC, denominadas de “comunidades virtuais” (Rheingold, 1998). O termo comunidade leva a idéia de que um grupo compartilha fatos e experiências de interesse comum, por exemplo, os membros e profissionais pertencentes à um grupo científico de pesquisas. A diferença entre o conceito de “real” e “virtual” esta na idéia de territorialidade. O conceito de “virtual” vai além dos limites geográficos.

Para Jones (1997), o conceito de “comunidade virtual” pode ser visto sob dois pontos de vista:

- O primeiro no qual a “comunidade virtual” é originada e mantido pela CMC (ex: e-mails, IRC, etc);
- O segundo, no qual as “comunidades virtuais” são novas formas de comunidades que surgem pelo uso e apoio da CMC.

O primeiro conceito refere-se ao “*virtual settlement*” (estabelecimento virtual) e o segundo como sendo a própria “comunidade virtual”. Assim pode-se dizer que a comunidade virtual difere de seu “*virtual settlement*”, porém este é a base da

existência da comunidade virtual, mostrando que a base tecnológica é a condição necessária, mas não é suficiente para que haja a comunidade virtual. Por exemplo, um *chat* em si, não é uma comunidade virtual, embora ele possua vários canais de comunicação (salas), se não houver pessoas conectadas às salas nada ocorrerá.

Conforme Lemos (2002), nem todo canal de comunicação na Internet pode ser classificado como comunidade, pois segundo o autor os participantes não guardam qualquer vínculo de relação pessoal ou temporal, são somente modos de contato eletrônico. O *virtual settlement* é, portanto um ciber-lugar, que é simbolicamente caracterizado por um tópico de interesse, e onde uma porção significativa de interatividade ocorre. Os “*virtual settlement*” seriam caracterizados pelas ocorrências abaixo:

- Uma variedade de comunicadores, que é condição associada à primeira característica da interatividade;
- Um nível mínimo de interatividade, que, para Jones, trata-se da extensão em que essas mensagens em uma seqüência têm relação entre si e, especialmente, como as mensagens posteriores têm relação com as anteriores. É a expressão da extensão de uma série de trocas comunicativas;
- Um espaço público comum onde uma porção significativa do grupo de comunicação mediada por computador interativa de uma comunidade ocorre. O espaço público é um fator importante na existência da comunidade virtual, e diferencia o espaço público, onde está a comunidade, do espaço privado, onde ocorrem as trocas de mensagem individuais; e,

- Um nível mínimo de associação sustentada, ou ainda, uma quantidade de membros relativamente constante, necessária para o nível razoável da interatividade exposta pela primeira característica.

Esses conceitos trazem alguns pontos que podem ajudar a esclarecer um pouco a idéia de "comunidade virtual". Se agregarmos, como o próprio autor determina ao conceito de comunidade virtual o de *virtual settlement*, veremos que também existe como condição para a comunidade virtual, a existência de um espaço público, onde a maior parte da interação da comunidade se desenvolva. Este espaço, por si só não constitui a comunidade, mas a completa. A comunidade precisa, portanto, de uma *base* no ciberespaço: um lugar público onde a maior parte da interação se desenvolva. A comunidade virtual possui deste modo, uma base no ciberespaço, um senso de *lugar*, um *locus* virtual.

2.4 Gerência do Conhecimento

Segundo Polanyi (1996), o conhecimento é entendido como sendo um bem pessoal, criado inicialmente de um contexto social e individual, em outras palavras, o conhecimento não é propriedade de uma organização ou de um grupo. Para Davenport e Prusak (1998) o conhecimento existe internamente nas pessoas, faz parte da natureza humana e possui o seguinte conceito:

*“Conhecimento é uma mistura fluida de experiência condensada, valores, informação conceitual e **insight** experimentado, a qual proporciona uma infra-estrutura para avaliação e incorporação de novas experiências e informações. Ele tem origem e é aplicado na mente dos conhecedores. Nas organizações, ele costuma estar embutido não só em documentos ou repositórios, mas também em rotinas, processos, práticas e normas organizacionais”* (Davenport e Prusak, 1998 p.6).

Conforme Takeuchi (1998), o conhecimento pode ser classificado em dois tipos, a saber:

a) **TÁCITO:** este conhecimento é inerentemente pessoal e, portanto difícil de formalizar, criando uma dificuldade para comunicá-lo ou compartilhá-lo com outras pessoas. Inspirações subjetivas, intuições e pressentimentos são o núcleo dessa categoria do conhecimento. Além disso o conhecimento Tácito é profundamente vinculado nas experiências pessoais, como em suas idéias, valores ou emoções. Para ser mais preciso o conhecimento tácito é subdividido ainda em dimensão técnica (“know-how”) e dimensão subjetiva (intuição, “insight”,etc) .

b) **EXPLÍCITO:** este pode ser expresso em termos de palavras e números e compartilhado por meio de dados, fórmulas científicas, especificações de produtos, manuais, etc. Nesse caso, o conhecimento é transmitido por meio das pessoas formalmente ou sistematicamente. Essa é a forma dominante do conhecimento no ocidente. Por exemplo, os orientais (Japoneses) por sua vez, entendem o conhecimento como sendo primeiramente tácito, que por essência não é facilmente visível ou expresso. Os conhecimentos Tácitos também contem uma componente

cognitiva baseada nas percepções, convicções, ideais, valores, emoções e modelos mentais.

As diferenças apresentadas no item anterior entre conhecimento tácito e explícito são as chaves para entender as diferenças entre o aprendizado do conhecimento oriental e ocidental. Para os ocidentais o conhecimento explícito pode ser facilmente processado por um computador e transmitido eletronicamente, ou ainda, armazenado em um banco de dados. Contrapartida, o caráter subjetivo e a natureza intuitiva do conhecimento tácito impõem uma dificuldade para processar, adquirir e ou transmitir de forma sistemática e lógica. Para o conhecimento tácito ser comunicado ou compartilhado com a organização, ele deve ser convertido em palavras e números que possibilitem sua compreensão. Por sua vez, Davenport e Prusak (1999) indicam que para a criação de conhecimento a condição é que *“os dados sejam transformados em informação, e que após isso, seja transformado em conhecimento, por meio de contextualização, categorização, cálculos, etc”*. É nesta fase que o conhecimento é convertido de tácito para explícito e uma *“Organização de conhecimento”* é criada. Embora haja uma clara distinção entre conhecimento tácito e explícito, eles não estão desvinculados, mas são complementares. De acordo com Takeuchi (1998) na sua teoria do *“conhecimento criativo”* mostram que o conhecimento humano é a expansão entre o conhecimento tácito e explícito. Essa interação produz quatro modos de conversão de conhecimento:

- I. Tácito para Tácito, denominado Socialização;
- II. Tácito para Explícito ou Externalização;
- III. Explícito para Explícito ou Combinação; e,
- IV. Explícito para Tácito ou Internalização.

Segue uma descrição sobre cada um deles:

- I. A **Socialização** também denominada de conhecimento compartilhado, esta baseada no compartilhamento e observação de exemplos e experiências de outros indivíduos de forma tácita.
- II. A **Externalização**, também denominada de conhecimento conceitual, surge por meio da linguagem utilizada para expressar o conhecimento tácito de uma pessoa.
- III. A **Combinação** reside no fato de que a sistematização de conceitos em um sistema computacional de conhecimentos.
- IV. Finalmente a **Internalização** surge com o aprendizado pela experiência prática.

2.5 Ontologias

Uma ontologia pode ser definida com sendo uma conceitualização compartilhada de um determinado domínio (Gruber, 1993). Ela é composta por um conjunto de conceitos dentro desse domínio, sendo esses organizados como uma taxionomia e de relações entre esses conceitos. Uma ontologia possui também axiomas.

Uma das vantagens do uso de ontologias nos sistemas de computação está no fato de que:

- Ontologias fornecem termos para representação do conhecimento;
- Permitem compartilhamento do conhecimento;

- *Ontologias de Aplicação* descrevem conceitos que dependem tanto de um domínio específico como de uma tarefa específica, e geralmente são uma especialização de ambos.

Outra distinção que Guarino (1998) faz é entre ontologias refinadas e não-refinadas, ou ainda, *off-line* e *on-line*. Uma ontologia não-refinada tem um número mínimo de axiomas e seu objetivo é ser compartilhada por usuários que concordem sobre uma determinada visão de mundo. Uma ontologia refinada precisa de uma linguagem de alta expressividade e tem um grande número de axiomas. Ontologias não-refinadas têm maior probabilidade de serem compartilhadas e deveriam ser utilizadas *on-line* para dar suporte à funcionalidade de sistemas de informação. Já as ontologias refinadas deveriam ser usadas *off-line* e somente para referência.

Uma ontologia define os termos usados para descrever e representar uma área de conhecimento. As ontologias são usadas por pessoas e aplicações para a troca de informações sobre um determinado domínio ou área de conhecimento, como biologia ou medicina. Uma ontologia fornece definições de conceitos básicos de um domínio, apropriadas para o processamento automático. O termo “ontologia” é emprestado da filosofia, onde uma Ontologia é uma descrição da Existência. Na Ciência da Computação, vem sendo aplicado desde o início da década de 90 na área de inteligência artificial para representação computacional de conhecimento em áreas como engenharia de conhecimento e processamento de linguagem natural. Para sistemas de Inteligência Artificial, o que “existe” é o que pode ser representado. Neste contexto, podem ser entendidas como uma especificação formal e explícita de

uma conceitualização consensual (Grubber, 1993). A noção de conceitualização exige, entretanto, uma adequada formalização, visto que ela pode gerar algumas confusões.

Nesta definição de Tom Gruber (Gruber, 1993), a conceitualização é definida com uma estrutura $\langle D, R \rangle$: onde D é um domínio e R é um conjunto de relações sobre D.

O corpo de um conhecimento representado formalmente é baseado numa conceitualização: os objetos, conceitos e outras entidades que são assumidas para existirem em alguma área de interesse e os relacionamentos que são organizados dentro delas (Genesereth e Nilsson, 1987).

A representação do conhecimento é feito por técnicas que auxiliam a modelagem mais precisa de alguns domínios do conhecimento. (Elmasri e Navathe, 2005).

Uma conceitualização é uma visão abstrata e simplificada da realidade que se quer representar para algum propósito. Toda base de conhecimento e sistema baseado em conhecimento está compromissada com alguma conceitualização, explícita ou implícita. Dessa forma uma ontologia é uma especificação explícita de uma conceitualização. Usam-se ontologias comuns para descrever a representação de uma base de conhecimento para um conjunto de agentes. Desta forma, estes agentes podem passar informações sobre o domínio do discurso, sem necessariamente operar sobre uma teoria globalmente compartilhada. O Conhecimento é atribuído aos agentes através da observação de suas ações. Um agente “sabe” algo se ele age como se ele tivesse a informação e está agindo racionalmente para alcançar seus objetivos. As “ações” dos agentes - incluindo servidores de base de conhecimento e sistemas

baseados em conhecimento - podem ser vistos através de uma ordem e perguntas da interface funcional. Nesta interface, um cliente interage com um agente fazendo afirmações lógicas e propondo perguntas. De forma pragmática, uma ontologia comum define o vocabulário com a qual as pesquisas e afirmações são trocadas entre os agentes.

A elaboração de uma de uma ontologia é uma tarefa que demanda tempo e é complexa, pois há necessidade do conhecimento sobre o assunto em questão e seu respectivo domínio (Pezza, 2004).

Ainda segundo Pezza (2004) a concepção da ontologia requer as seguintes etapas:

- a) Especificar o domínio e o escopo da ontologia;
- b) Analisar a viabilidade do reuso de ontologias já existentes;
- c) Enunciar termos importantes;
- d) Definir as classes envolvidas;
- e) Identificar e montar a hierarquia das classes; e
- f) Definir as propriedades das classes;

Está descrita no capítulo 3 uma ontologia base que será utilizada no sistema REDMIS e seu respectivo Banco de Dados. Os diagramas de classe foram elaborados a partir do software livre para criação de diagramas DIA (*DIA gtk+ based diagram creation program*-2006). Possui interface simples e agradável de ser utilizada, e é oferecido gratuitamente sob licença GNU tanto para Windows com Linux.

2.6 Tecnologias Utilizadas

Nesta seção estão descritas as tecnologias utilizadas para a construção da ferramenta REDMIS. Foram utilizadas ferramentas tecnológicas gratuitas e de domínio público, proporcionando a elaboração sob o paradigma de software livre.

2.6.1 Modelo de Dados

Conforme Pezza (2004) menciona em seu trabalho, o banco de dados e seu respectivo sistema gerenciador devem propiciar de forma eficiente o gerenciamento dos dados, entenda-se por isso o armazenamento e a recuperação no banco de dados. O modelo lógico utilizado para a concepção da presente ferramenta foi o “modelo entidade – relacionamento” ou “modelo relacional”, por motivos de que nesse caso, concebem-se os dados externos do mundo real como sendo entidades e a relação que os envolve de relacionamento. Entretanto, Silberschatz et al (1999) mencionam que além desse tipo existem mais três modelos lógicos, perfazendo um total de quatro modelos. Fica mencionado que o modelo citado será utilizado por sua ampla aceitação em situações praticas acadêmica e ou comerciais.

A implementação do motor do sistema EDMIS foi feito com o Sistema Gerenciador de Banco de Dados (SGBD) MySQL, cujas características de disponibilidade para diversos sistemas operacionais, assim como seu desempenho e custo zero para licença não comercial (MySQL Licency Policy, 2006) levaram a sua escolha como a plataforma apropriada para ser parte da Ferramenta. Como o sistema

proposto será utilizado em ambientes acadêmicos, a licença de código livre deve ser distribuída juntamente com o pacote desenvolvido.

De acordo com Pezza (2004) o sistema apesar de ter como base a ontologia descrita nos itens anteriores , não é capaz de implementar o banco de dados real, uma vez que a ontologia proposta representa um nível de abstração que não pode ser representado no modelo entidade relacionamento.

2.6.2 Linguagem Java

Criada nos laboratórios da Sun Microsystems, essa linguagem de programação escrita inicialmente em C++, cujo codinome do projeto era Green, depois foi batizada de Java em homenagem a um café importado cuja cidade de origem era chamada de Java (Deitel e Deitel, 2003 pág 59). Como o sucesso da www em 1993 a Sun transforma o potencial dessa linguagem com objetivos de criar paginas da Web com conteúdo dinâmico.

Em essência, Java é uma linguagem orientada a objetos com suporte a programação Web e com suporte a banco de dados.

2.6.3 Sistema de Servlets

A tecnologia Servlet é a base do desenvolvimento de aplicações para Web usando linguagem de programação Java. Um servlet é uma classe Java que pode ser carregada e executada por um servidor Web. Damos o nome para esse servidor web cliente e é chamado de contentor de *servlet*, no início era chamado de máquina *servlet* (Kurniawan, 2002).

Pode-se então entender que os *servlets* são executados num servidor Web e que repondem às solicitações de clientes via Web browser. A figura 4 indica que os *servlets* interagem com seus clientes por meio da solicitação-resposta a um servidor HTTP. A figura 5 mostra um arranjo mais comum, pois as solicitações e respostas ficam agilizadas se o servidor http é colocado no *front-end*, deixando o mesmo cuidar das solicitações estáticas, passando para o contentor *servlet* as solicitações dinâmicas com um tratamento independente.

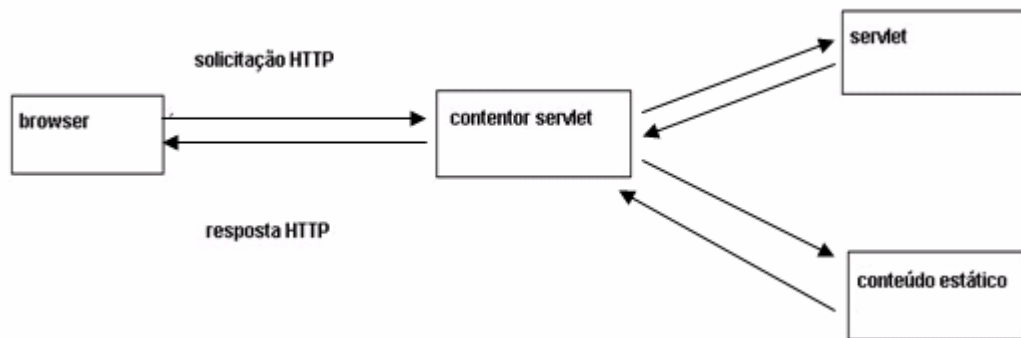


Figura 4 - Arquitetura de um aplicativo servlet. Fonte: Kurniawan (2002).

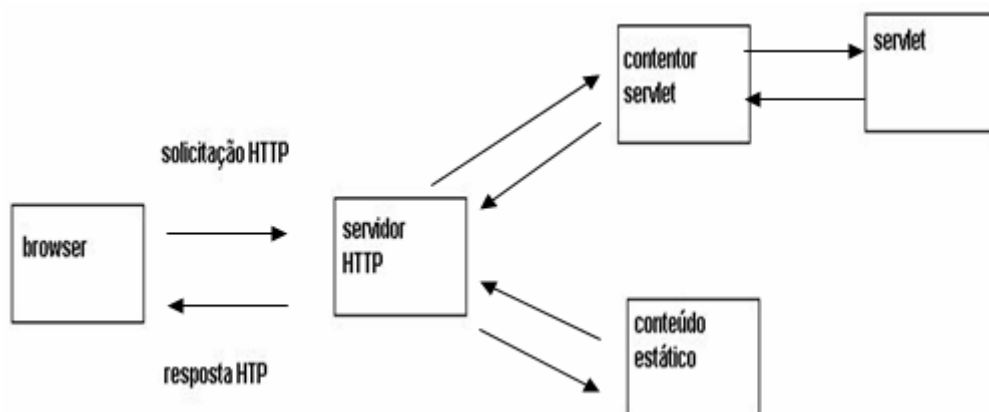


Figura 5 - Arquitetura de um aplicativo *servlet* juntamente com um servidor HTTP.

Fonte: Kurniawan (2002).

2.6.4 Servidores JSP

JavaServer Pages (JSP) é uma outra tecnologia Java para desenvolver aplicativos web. A JSP não objetiva substituir os *servlets*, na realidade JSP é uma extensão dessa tecnologia, permitindo misturar código estático com código dinâmico por meio da linguagem JAVA. Possui mecanismos de funcionamento simples, se uma página JSP for acessada, o código dinâmico escrito em JAVA é executado no servidor e a resposta para o cliente será um texto. Uma vez que o código JSP é processado no servidor, torna-se mais seguro, pois o cliente não tem acesso à fonte, somente conseguirá visualizar o código estático (HTML) criado (Kurniawan, 2002).

2.7 Recuperação da Informação

No início do século vinte alguns cientistas contribuíram para o desenvolvimento da Ciência da Informação, dentre eles Paul Otlet, que foi o primeiro a apresentar o termo “Documentação”, e também previu que deveriam haver tecnologias apropriadas a sua operacionalização, fato este apresentado em seu “*Traité de Documentation*” (1934), onde apresenta todo interesse por tecnologias que permitam o trabalho com a informação. Posteriormente, na década de 50, o pesquisador Calvin Mooers consagra o termo “*Information Retrieval*”, criando assim uma nova área de pesquisa que leva esse nome (Ferneda, 2003). Os primeiros resultados de importância no tratamento computacional da informação surgem com os trabalhos de Hans Peter Luhn, cujos trabalhos no campo da indexação automática e elaboração automática de resumos obtidos nos laboratórios da IBM trouxeram

importantes resultados na forma de tratar, armazenar e recuperar informação. Na década de 1960 Luhn já havia obtido por volta de 80 patentes nos Estados Unidos (Ferneda, 2003).

Um Sistema de Recuperação de Informação (S.R.I) pode ser entendido como sendo uma “caixa preta”, em que existe uma entrada, que dá lugar a um processamento, do qual resulta uma saída. Relativamente aos elementos de entrada, eles podem ser um conjunto de documentos ou uma consulta do usuário. Um modelo geral de arquitetura de R.I é apresentado na figura 6.

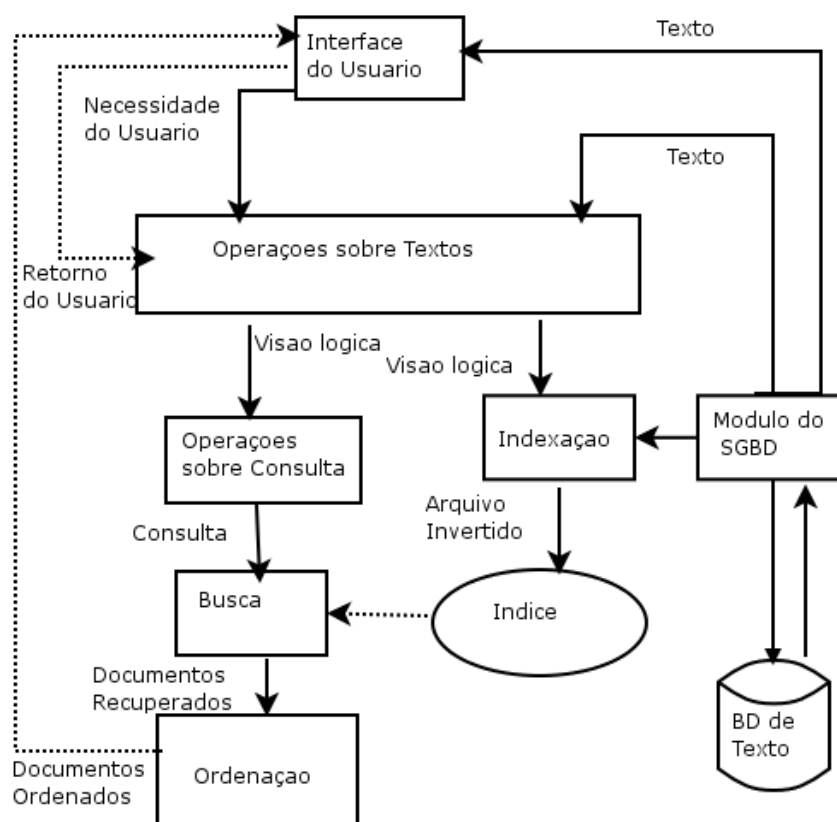


Figura 6 - Arquitetura Geral de um Mecanismo de Busca. Fonte: (Baeza-Yates e Ribeiro Neto, 1999).

É necessário encontrar uma representação, tanto para os documentos como para a consulta, que seja adequada para o processamento. Esta representação é normalmente obtida pela aplicação de técnicas de processamento de linguagem natural, e outros processos de Inteligência Artificial, sobre o texto dos documentos e da consulta. Por vezes aplicam-se métodos estatísticos sobre o texto, com o propósito de encontrar as palavras mais frequentes, pois entre elas podem estar as palavras-chave mais relevantes.

A parte do processamento de uma consulta tem a ver com a procura de informação ao longo dos documentos. Este processo da recuperação de informação pode envolver vários algoritmos de integração e classificação dos dados, no sentido de melhorar a eficiência do sistema, tornando o cálculo da relevância dos dados tão adequado quanto possível. A resposta ou saída de um sistema de recuperação de informação (S.R.I) consiste num conjunto de citações ou documentos em que foi encontrada informação sobre a consulta processada.

O pesquisador Luhn propõe que a frequência seja utilizada para extrair palavras e sentenças representativas de um documento. Dada uma frequência f de ocorrência e uma ordem r (*rank*) de sua frequência de ocorrência, então um gráfico relacionando r a f seria uma curva similar à mostrada na figura 7, que diz que o produto da frequência de uso de uma palavra e sua ordem de importância é aproximadamente constante (Ferneda, 2003). Luhn utiliza a denominada lei de Zipf, como uma hipótese inicial para estipular dois pontos de corte, um inferior e um superior (Pereira, 2004). As palavras que excedem o limiar superior são consideradas

comuns e as abaixo do limiar inferior são consideradas muito raras ou pouco freqüentes.

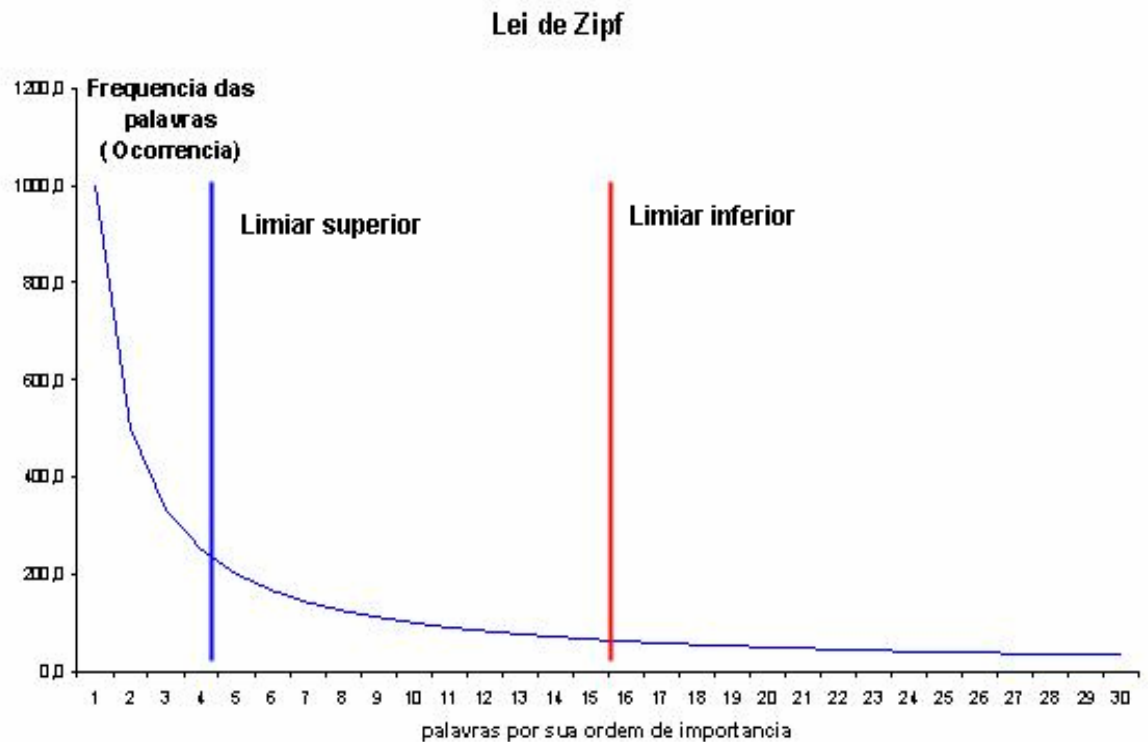


Figura 7 - Curva da Frequência das palavras x Ordem de Importância. Fonte: Adaptada de van Rijsbergen (1979, p.16).

Na literatura os S.R.I podem ainda ser classificados como:

- **Convencionais:** são baseados numa pesquisa de informação precisa ou “*exactmatch*”. Nestes são procurados documentos descritos exatamente pelos termos da consulta;
- **Não convencionais:** utilizados para encontrar os seus resultados em uma boa aproximação ou “*best match*”.

Para além desta classificação, os S.R.I podem utilizar diferentes modelos para representar e procurar informação sobre os documentos. A próxima secção descreve os principais métodos de RI.

2.7.1 Modelos para Recuperação da Informação

A recuperação de informação é uma área complexa, onde nem sempre se conseguem obter resultados perfeitos (Pereira, 2004). A eficácia do processo depende, por exemplo, da representação e estrutura escolhidas para os dados, dos algoritmos utilizados e de outros aspectos. Esses modelos de recuperação de informação baseiam-se no fato de que os documentos estão representados por um conjunto de termos de indexação. Isso significa que o termo de indexação é uma palavra que representa um conceito ou significado que está presente no documento. Mas nem sempre esses termos mostram-se representativos para descrever o seu conteúdo (Ferneda, 2003).

Existem alguns modelos em R.I. que são tomados como referencia e utilizados nas maiorias das situações são os modelos: Modelo Booleano (Paice, 1984), Modelo Booleano Estendido (*Extended Boolean models*) (Paice, 1984; Salton e McGill, 1983), Modelo Probabilístico (Maron e Kuhns, 1960), Modelo de Espaço Vetorial (Salton e McGill, 1983), Modelos de conjunto Fuzzy (*Fuzzy set models*) (Lee, 1995), Modelos Bayesianos (Ribeiro e Muntz, 1996) e Modelos Lingüísticos Estatísticos (*Statistical Language Models/Language Models*) (Ponte e Croft, 1998).

Um modelo de recuperação de informação prediz e explica o que um usuário irá considerar relevante dada sua consulta. São três os modelos clássicos seguidos por sistemas de RI para determinar a relevância de documentos: Vetorial, Booleano

(Lógico) e Probabilístico, mais detalhes podem ser obtidos em Baeza-Yates e Ribeiro-Neto (1999). Serão vistos nas próximas seções esses três modelos clássicos e o modelo de conjuntos *fuzzy*.

2.7.1.1 Modelo Booleano

O modelo booleano foi o primeiro modelo utilizado em RI e o modelo mais utilizado até o meio da década de 90, apesar das alternativas que surgiram desde o final dos anos 60 (Ferneda, 2003). É um modelo simples que considera a teoria dos conjuntos e a álgebra booleana como suporte teórico. Os documentos e consultas são considerados ou representados por um conjunto de palavras, onde uma consulta é representada na forma de uma expressão booleana convencional, que liga seus termos por meio de conectivos lógicos AND, OR e NOT (Pereira, 2004).

Neste modelo, um documento é considerado relevante ou irrelevante para uma consulta, não existe resultado parcial e não há informações que permitam a ordenação do resultado da consulta.

Um ponto desfavorável é o fato do modelo booleano não possibilitar a ordenação dos resultados por ordem de relevância. Isso é uma de suas principais desvantagens, já que esta classificação é uma característica considerada essencial em muitos dos sistemas de RI modernos, como por exemplo, nas máquinas de busca (Ferneda, 2003).

Outra característica deste modelo que pode ser considerada uma desvantagem no caso de usuários inexperientes é o uso de operadores booleanos. Para os usuários que conhecem bem álgebra booleana os operadores podem ser considerados como uma forma de controlar o sistema. Se o conjunto de resposta é

muito pequeno ou muito grande, eles saberão que operadores utilizar para produzir um conjunto de respostas maior ou menor. No entanto, para usuários comuns, os operadores booleanos não são intuitivos, pois seu uso é diferente do uso das palavras equivalentes a eles em língua natural.

Por exemplo, se um usuário se interessa por computação e por informática a consulta mais indicada seria:

“computação OR informática” e não “computação AND informática”

Outra desvantagem é que os termos de busca não podem ser ponderados, assumindo-se assim que todos os termos possuem igual ponderação.

2.7.1.2 Modelo Vetorial

No modelo de Espaço-vetorial, ou simplesmente modelo Vetorial, cada documento é representado por um vetor de termos e cada termo possui um peso associado que indica seu grau de importância no documento esse grau de relevância é um número ou peso que varia de 0 a 1. Quanto mais próximo de 1 estiver o documento, mais relevante para a consulta ele é considerado, o contrário para documentos que se aproximam de valores próximos de 0.

Em outras palavras, cada documento possui um vetor associado que é constituído por pares de elementos na forma $\{(palavra_1, peso_1), (palavra_2, peso_2), \dots, (palavra_n, peso_n)\}$. Cada elemento do vetor de termos é considerado uma coordenada dimensional. Assim, os documentos podem ser colocados em um espaço euclidiano de n dimensões (onde n é o número de termos) e a posição do documento em cada dimensão é dada pelo seu peso.

As distâncias entre um documento e outro indicam seu grau de similaridade, ou seja, documentos que possuem os mesmos termos acabam sendo colocados em uma mesma região do espaço e, em teoria, tratam de assuntos similares. Um exemplo é mostrado na Figura 8.

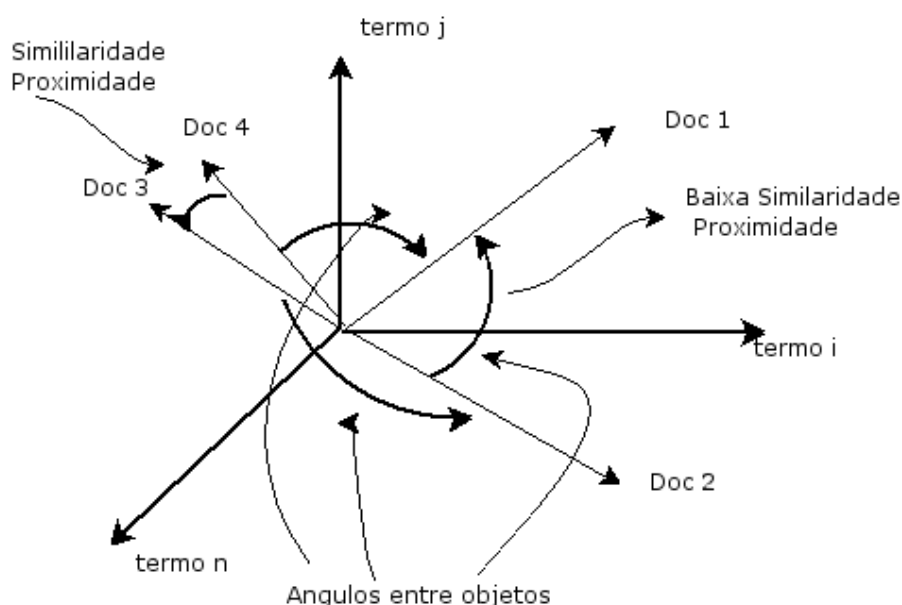


Figura 8 - Similaridade de documentos no modelo vetorial. Fonte: Ferneda (2003).

Consultas também são representadas por vetores. Desta forma, os vetores dos documentos podem ser comparados com o vetor da consulta e o grau de similaridade entre cada um deles pode ser identificado. Os documentos mais similares (mais próximos no espaço) à consulta são considerados relevantes para o usuário e retornados como resposta para ela. A equação 1 mostra o cálculo da similaridade entre a consulta e o documento.

$$sim (Q , D) = \frac{\sum_{k=1}^n w_{qk} w_{dk}}{\sqrt{\sum_{k=1}^n (w_{qk})^2 \cdot \sum_{k=1}^n (w_{dk})^2}} \quad (\text{Eq -1})$$

Onde os termos da equação são:

- Q é o vetor de termos da consulta;
- D é o vetor de termos do documento;
- w_{qk} são os pesos dos termos da consulta;
- w_{dk} são os pesos dos termos do documento.

Uma das formas de calcular a proximidade entre os vetores é testar o ângulo entre estes vetores. No modelo original, é utilizada a função cosseno (*cosine vector similarity*) conforme figura 10, que calcula o produto dos vetores de documentos através da fórmula da equação 1.

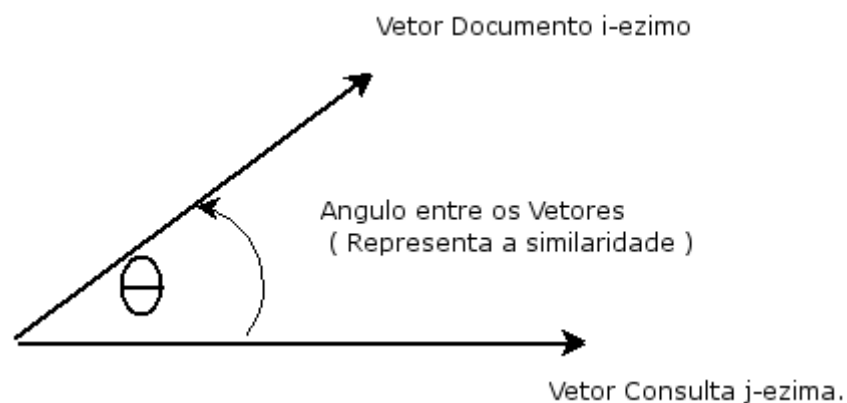


Figura 9 – Representação da similaridade entre documentos. Fonte: Pereira (2004).

Calculados os graus de similaridade, é possível montar uma lista ordenada de todos os documentos ordenados por seus respectivos graus de relevância à consulta (*ranking*).

Uma desvantagem do modelo Vetorial é que não é possível incluir dependências entre termos no modelo, para modelar, por exemplo, frases ou termos que aparecem perto um do outro. Este modelo traz ainda duas dificuldades: a associação de pesos aos termos, que nem sempre é uma tarefa simples e, a implementação propriamente dita.

Uma característica do modelo vetorial é de que os termos de indexação são independentes, ou seja, não é levado em conta o relacionamento entre eles. Uma das limitações desse modelo é não permitir que seja feita a formulação por buscas booleanas, esse fato restringe sua flexibilidade (Ferneda, 2003). Um de seus méritos é de que é considerado um modelo conceitual o que em uma teoria é de grande relevância.

2.7.1.3 Modelo Probabilístico

No modelo probabilístico os termos indexados dos documentos e das consultas não possuem pesos pré-definidos. A ordenação dos documentos é calculada pesando dinamicamente os termos da consulta relativamente aos documentos. É baseado no princípio da ordenação probabilística (*Probability Ranking Principle*) (Pereira, 2004).

Neste modelo, busca-se saber a probabilidade de um documento D ser ou não relevante para uma consulta Qa. Tal informação é obtida assumindo-se que a

distribuição de termos na coleção é capaz de informar a relevância provável para um documento qualquer da coleção. O modelo probabilístico é um dos poucos modelos que não necessita de algoritmos adicionais para associação de peso aos termos para ser implementado e os algoritmos de ordenação dos resultados são completamente derivados de sua teoria. O modelo assume que a relevância de um documento é independente da relevância de todos os outros, e que um documento D será dito relevante para uma consulta Q quando: $P(+R_q / D) > P(-R_q / D)$.

Onde a simbologia é representada por :

R_q — o documento é relevante para a consulta Q ;

$-R_q$ — o documento não é relevante para a consulta Q ;

$P(+R_q/D)$ — probabilidade de que o documento D seja relevante para a consulta Q ;

$P(-R_q/D)$ — probabilidade de que o documento D não seja relevante para a consulta Q .

Assim, dada uma consulta Q , o modelo probabilístico atribui a cada documento D (como medida de similaridade) um peso $W_{D/Q}$, como sendo a medida de similaridade entre a consulta Q e o documento D . w_{qk} é o peso para o termo k na consulta, enquanto w_{dk} é o peso para o termo k no documento. Por meio do teorema de Bayes e estimativas de relevância baseadas nos termos da consulta, a equação que calcula a probabilidade do documento D ser relevante a consulta Q é dada pela equação 2.

$$sim(D, Q) = \sum_{i=1}^t w_{qk} \cdot w_{dk} \cdot \left[\log \frac{P(k_i / R)}{1 - P(k_i / R)} + \log \frac{P(k_i / \bar{R})}{1 - P(k_i / \bar{R})} \right] \quad (Eq-2)$$

Os pesos w_{jk} e w_{dk} pertencem ao intervalo de 0 a 1.

As duas principais desvantagens deste modelo são: o fato de que para várias aplicações a distribuição dos termos entre documentos relevantes e irrelevantes não estará disponível; e, que o modelo define apenas uma ordenação parcial dos documentos.

2.7.1.4 Modelo *Fuzzy*

O objetivo da lógica *fuzzy* é representar, capturar e operar com a incerteza e as verdades parciais dos fenômenos naturais de forma sistemática e rigorosa (Ferneda, 2003). Para melhor compreender os conjuntos *fuzzy* pode-se recorrer a teoria clássica dos conjuntos, que define o conceito de pertinência de elemento. Ou seja, um elemento pode ou não pertencer a um conjunto.

Seja o conjunto $A(x)$ e U o universo do discurso então $A(x)$ é uma função fuzzy se :

$$A(x) : U \Rightarrow \{0, 1\}$$

Onde $A(x) = 1$, se x pertence a A e 0, caso contrário.

O modelo *fuzzy* foi proposto por Zadeh em 1965, e ele propõe uma nova teoria dos conjuntos em que não existe a descontinuidade, ou de outra forma, não há distinção abrupta entre elementos pertencentes e não pertencentes a um conjunto, e esse conjunto é denominado de conjunto nebuloso "*Fuzzy Sets*" (Ferneda, 2003).

Dessa forma um conjunto fuzzy A em U é definido por uma função de pertinência $A(x) : U \Rightarrow \{0, 1\}$, onde são representados por um conjunto de pares ordenados tais que :

$$A = \{ x / A(x) , x \text{ pertence a } U \}$$

U – é denominado de universo do discurso.

A - é um subconjunto de U .

O grau de imprecisão de um dado elemento pode ser interpretado como uma valoração de possibilidade, ou em outras palavras de que um termo seja membro de um conjunto. Dessa forma A é caracterizado por uma função $\mu(x)$ que associa a cada elemento x de U um numero $\mu(x)$ entre 0 e 1 . De forma que :

$A = \{ x, \mu(x) \} / x \text{ pertence a } U, \text{ e } \mu(x) \text{ é um valor numérico de quanto o elemento de } x \text{ pertence ao conjunto } A .$

Na teoria dos conjuntos *fuzzy* , existem as operações ou relações *fuzzy* , que são o complemento a união e interseção. Essas operações são apresentadas abaixo:

União $\mu_{A \cup B} = \max(\mu_A(x), \mu_B(x))$

Intersecção $\mu_{A \cap B} = \min(\mu_A(x), \mu_B(x))$

Complemento $\mu_{\bar{A}}(x) = 1 - \mu_A(x)$

Os conjuntos *fuzzy* têm sido utilizados na recuperação da informação uma vez que um documento pode ser entendido como sendo um *fuzzy set* de termos, que são representados por uma esperança $\mu(t) / t$, cujos pesos relativos são uma função do documento e do termo em análise. Esse peso que é associado ao termo mostra o grau de significância do termo em relação ao conteúdo do documento (Ferneda, 2003).

Dessa forma pode-se recuperar do documento aquela informação que for mais relevante para a consulta, pois a função $f(\text{doc}, \text{termo})$ é realizada pelo cálculo da importância do termo em cada uma das seções do documento, ou seja o usuário ajusta de acordo com seus interesses os documentos (Pereira, 2004).

2.8 Avaliação de Relevância em Recuperação da Informação

A noção de relevância é subjetiva e situacional, diferentes usuários podem ter opiniões diferentes sobre a relevância ou não de um documento. Um pesquisador pode julgar o resultado de uma consulta por vários níveis de importância, por exemplo :

“Se a consulta for sobre um assunto no qual o pesquisador obterá disponível um documento relacionado com o artigo que está escrevendo para um congresso, a ser enviado no dia seguinte, seu julgamento será possivelmente de *altamente relevante*. Porém, imaginemos o mesmo pesquisador, com a mesma consulta, mas relacionada a um trabalho de uma banca de doutorado para o mês seguinte, talvez o seu julgamento seria de *menor relevância* para aquele momento e situação”.

Portanto, a relevância possui um caráter cognitivo e dinâmico. Pois, além de julgar se é relevante ou não, a escala adotada para uma relevância padrão pode ser diferente para pessoas diferentes; e, ainda dinâmico no sentido de que não é so um resultado a ser avaliado, mas um conjunto que pode mudar, dependendo da consulta a ser feita e do estado da base de conhecimento naquele instante.

Apesar dessas ponderações, os resultados de uma consulta nem sempre são um documento, mas partes deste (título, resumo, citações, etc.). E nesse caso, o julgamento irá depender de quanto há de informação pertinente àquela área da pesquisa.

Os sistemas de recuperação de informação podem ser avaliados através de consultas que fazem parte de uma coleção de referência. Um exemplo é a conhecida coleção TIPSTER, usada na Text REtrieval Conference (TREC), descrita em Harman (1993). A TIPSTER é uma coleção de cerca de um milhão de documentos, obtidos de várias fontes, tais como o *Wall Street Journal*. Nesta coleção há um conjunto de consultas e para cada consulta é fornecido um conjunto ideal de documentos resposta, criado por especialistas nos temas envolvidos. Um SRI classifica os documentos recuperados para cada consulta, de acordo com uma ordem de relevância gerando um vetor resultado. Avalia-se o SRI através da comparação das respostas geradas por este sistema e o conjunto ideal de respostas. Para isso, o vetor resultado é examinado e comparado com o conjunto ideal, obtendo-se dois índices de avaliação: precisão e revocação.

Dessa maneira algumas métricas são apresentadas para avaliação de desempenho de sistemas S.R.I que são a cobertura(*recall*) e precisão(*precision*) (Baeza-Yates e Ribeiro-Netto, 1999). Segue a definição de cada conceito:

- Precisão é a fração dos documentos já examinados que são relevantes;
- Revocação é a fração dos documentos relevantes observada dentre os documentos examinados.

A avaliação do modelo de um SRI pode ser observada por um gráfico com as médias precisão versus revocação. O gráfico pode ser obtido calculando-se a precisão para níveis anteriormente estabelecidos de revocação. A figura 10 ilustra a forma geral de um gráfico “precisão versus revocação” em um sistema típico (Baeza-Yates e Ribeiro-Netto, 1999).

Seja, N o conjunto de resposta ideal, $|N|$ o número de documentos deste conjunto e R o vetor resultado recuperado pelo SRI, têm-se as equações 3 e 4.

- $\text{Revocação} = |N \cap R| / |N|$ (Eq- 3)

- $\text{Precisão} = |N \cup R| / |N|$ (Eq- 4)

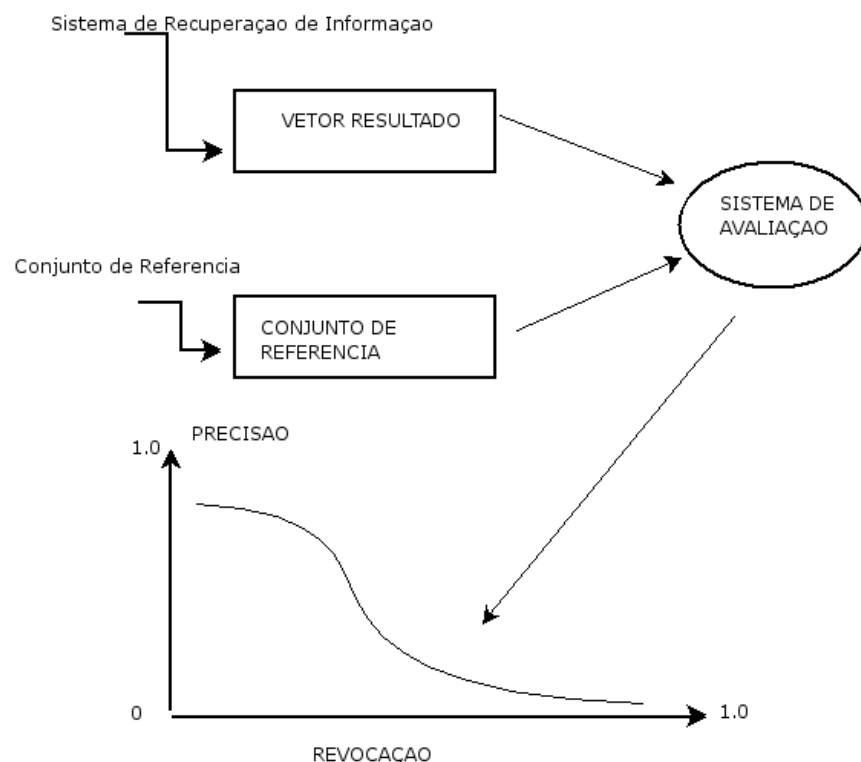


Figura 10 - Processo de Avaliação de um Sistema de Recuperação de Informação.

Fonte: Ferneda(2003).

Capítulo

3

O Sistema REDMIS

Este capítulo apresenta a descrição do sistema de Gerência de Documentos REDMIS (*Rapid Electronic Document Management Intelligent System*). Também são discutidas as implementações e adaptações do sistema EDMIS desenvolvido por Alves (2005), com base na utilização de ontologias de domínio para caracterização dos membros do grupo, assim, como as áreas de pesquisa relacionadas com seus pesquisadores.

3.1 Arquitetura do Sistema Base “EDMIS”

O sistema REDMIS é uma extensão da ferramenta desenvolvida por Alves (2005), cujo nome é EDMIS, e suas características em termos de arquitetura são brevemente descritas no próximo item.

A figura 11 esquematiza a arquitetura da ferramenta EDMIS. Existem na arquitetura do sistema cinco módulos dedicados de acordo com as seguintes funcionalidades:

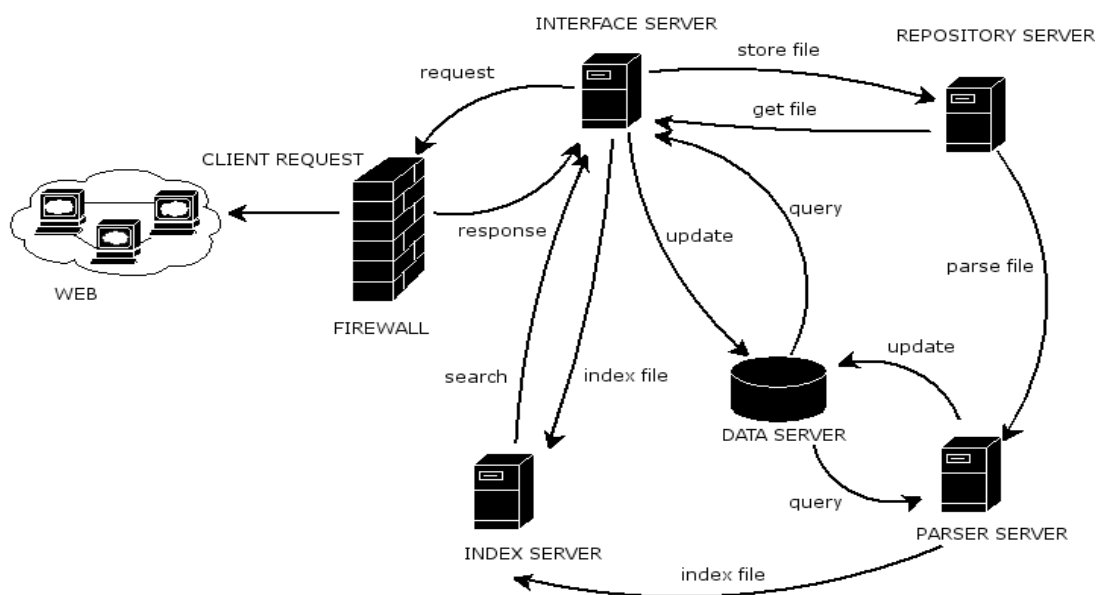


Figura 11 - Arquitetura da ferramenta original EDMIS.

a) Interface Server é o servidor responsável pela geração dinâmica das páginas que compõem a interface com o usuário.

b) Data Server é o servidor responsável pela execução do SGDB - Sistema Gerenciador de Banco de Dados, responsável pelo Banco de Dados do sistema.

c) *Parser Server* é responsável pela extração de texto e eliminação dos formatos dos documentos enviados pelos usuários ao sistema, permitindo assim a indexação dos arquivos pelo seu conteúdo, não somente pelos dados presentes no Banco de Dados, tais como Título, Autor, Editora, etc. O *Interface Server* envia a solicitação de indexação dos documentos ao *Parser Server* toda a vez em que um usuário adiciona ou remove um documento de um determinado grupo de documentos. Os formatos previstos de arquivo que são suportados nativamente pelos conversores do sistema são descritos abaixo:

- **c/cpp/h** : código-fonte de aplicativos escritos em linguagem C/C++.
- **chm** : Arquivo de ajuda do Windows em HTML codificado.
Atenção: devido às restrições existentes neste formato de arquivo, este conversor somente funciona em equipamentos rodando sobre o Windows.
- **doc** : Documento do Microsoft Word.
- **htm/html** : Páginas HTML.
- **java** : Código-Fonte de aplicações Java.
- **pdf** : *Portable Document Format*, formato de documentos digitais para leitura com o *Adobe Acrobat/Acrobat Reader*.
- **ppt** : Apresentação do Microsoft *PowerPoint*.
- **ps** : Arquivo *PostScript*.
- **rtf** : *Rich Text Format*, formato de documento compatível com diversos editores de texto.
- **txt**: Documento de Texto ASCII/ANSI.

- **xls**: Planilha de cálculos do Microsoft Excel.

Pela natureza dos documentos que um grupo de pesquisa utiliza normalmente, os formatos acima são suficientes para a maioria das situações.

d) Index Server é responsável pelo armazenamento e manutenção dos índices dos documentos presentes nos repositórios do sistema. Alves (2005) recomenda que se for o caso do sistema estar num ambiente de grande porte, onde o *Index Server* é instalado em um servidor dedicado, este necessita de um bom link de comunicação com o *Parser Server* e com o *Interface Server*, já que os índices armazenados no servidor são criados pelo *Parser Server* quando das atividades de indexação de documentos e solicitados pelo *Interface Server*, quando da ocorrência de consulta de documentos por parte dos usuários do sistema.

e) Repository Server é responsável pelo armazenamento e manutenção dos documentos presentes no sistema.

A ferramenta foi elaborada de forma que todos os elementos possam eventualmente ser instalados numa única máquina desktop, porém, por questões de desempenho, Alves (2005) recomenda que os servidores sejam distribuídos em máquinas distintas para um balanceamento da carga de processamento. Isso é relevante para o *ParserServer*, já que consome elevados recursos de Memória e CPU para a extração dos formatos e Indexação dos documentos por ele processado.

3.2 O Sistema REDMIS

O Sistema REDMIS proposto nesse trabalho tem por objetivo proporcionar um ambiente de gerência de documentos para um grupo de pesquisas, oferecendo para isso um ambiente de colaboração. Uma vez que suas características são herdadas do sistema original EDMIS (Alves, 2005), todas as transações podem ser efetuadas via rede local ou mesmo Internet, estabelecendo assim a possibilidade da criação de uma comunidade virtual acadêmica. No objetivo de manter a característica de utilização de softwares de domínio público, foram mantidos a linguagem de programação Java, servidores JSP e o Banco de Dados MySql. O sistema REDMIS herda o conjunto das seguintes funcionalidades conforme discutido na seção 2.1:

- **Captura:** Presume-se que o documento já esteja num dos formatos eletrônicos válidos.
- **Armazenamento e Organização:** Os documentos são armazenados em sua forma original em arquivo de disco de duas formas conjuntamente: formato e tamanho originais, e compactados (zipados para backup).
- **Roteamento:** Re-direcionamento do documento para um outro usuário do grupo, conforme o grau de relevância para esse usuário com base no seu perfil acadêmico e de produção científicos.
- **Recuperação e Síntese:** Uma vez no banco de dados o documento pode ser recuperado por meio da pesquisa de seu conteúdo. Tipicamente os recuperadores de texto utilizam lógica Booleana. Porém pode haver a combinação de palavras chave obtidas no documento ou todo o texto

(atributos do texto). No caso do sistema REDMIS, foi implementada a técnica de Thesaurus-fuzzy para palavras próximas. Isto é feito por meio da extração de todos os formatos e a conversão para um arquivo de texto simples, no qual é feito todo o trabalho de mineração dos dados, com as técnicas discutidas no capítulo 2 seção 2.7, que são o modelo booleano, vetorial e *fuzzy*, que operam como base na descrição ontológica de seus usuários.

- **Visualização e impressão:** Embora o documento esteja em formato eletrônico, há interesse, dependendo da utilização da informação, de que o documento possa ser impresso. Para isso, a Ferramenta carrega o arquivo desejado utilizando-se da ferramenta nativa do sistema e formatos correntes para o arquivo específico (ex: Adobe-Reader®, Winzip®, MS-Word®, etc).
- **Descarte:** Esse é momento em que o documento deixa de ser útil, no caso do REDMIS, o documento é descartado do perfil do usuário, mas fica armazenado no diretório da ferramenta, de forma a permitir backup e recuperações futuras por parte de outros usuários.

3.3 Metodologia

Este sistema é projetado com base na descrição ontológica dos seus participantes, de suas atividades e de seus documentos. Foram adaptadas à base de dados da presente ferramenta informações que levam em consideração o perfil dos membros do grupo, tão como os conceitos a eles associados. Os mecanismos de indexação e busca personalizada de conteúdos, utilizam as informações disponíveis

num banco de perfis dos seus usuários, mantendo informações referentes a dados cadastrais e informações sobre o grau de interesse ou conhecimento do usuário em determinados assuntos. A modelagem do perfil do usuário serve para aprimorar o processo de indexação e busca personalizada de conteúdos, para que o sistema só recomende para o usuário itens que forem do seu interesse e adequados ao seu nível de conhecimento.

Middleton et al. (2001) afirmam que a modelagem do perfil do usuário é baseada em dois pontos:

- No seu comportamento (*behavior-based*) ou;
- No seu conhecimento (*knowledge-based*).

Caso seja feita a modelagem baseada no conhecimento, os usuários são associados aos chamados modelos estáticos de usuários do sistema. Em contrapartida a abordagem baseada no comportamento, parte do princípio que o próprio comportamento forma o modelo, sendo então aplicadas técnicas de *web mining* para descobrir padrões úteis de comportamento. O perfil pode ser estabelecido inicialmente no momento do cadastro do usuário, através da indicação direta de alguns assuntos que são do seu interesse. Na presente ferramenta foi adotada uma representação por conhecimentos, num perfil inicial (os pesquisadores são criados no momento da instalação da ferramenta), mas que pode evoluir por meio da alteração do cadastro inicial, à medida que os pesquisadores possam se interessar por outras áreas relacionadas, no momento de suas pesquisas ou projetos. Dessa forma a proposta do REDMIS é no sentido acima, híbrida.

Para isso, o sistema deve acompanhar a evolução do perfil do usuário, uma vez que este é dinâmico. A evolução poderá ser modelada pelo comportamento do usuário, por exemplo, analisando documentos eletrônicos que o mesmo acessa ou utiliza, palavras utilizadas para buscas, documentos que foram redigidos ou publicados pelo usuário, etc.

3.4 Ontologia do Sistema REDMIS

A descrição ontológica da ferramenta REDMIS, foi baseada nas informações que são pertinentes a um grupo de pesquisa, como descrito no trabalho de Pezza (2004), com a utilização da plataforma Lattes (CNPq, 2006) como modelo de dados para representação ontológica dos pesquisadores (Bonifácio, 2002).

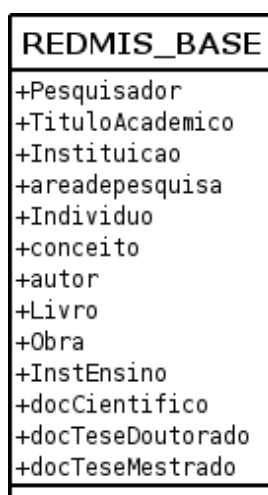


Figura 12 - Ontologia Base do Sistema REDMIS.

Como pode ser visto na figura 12 temos as classes básicas na representação do conhecimento, e estas classes são subdivididas em subclasses mais específicas, que

possam detalhar, por exemplo, no caso de um documento científico, conceitos envolvidos, palavras chaves, resumo, autor, etc.

São analisados também os atributos das classes mais importantes para que possa ser compreendido o relacionamento e o armazenamento do conhecimento assim como o relacionamento entre as classes.

3.4.1 Ontologia docCientifico

A classe doc_Cientifico (figura 13) possui os seguintes atributos :

- **anopublicacao** : ano da publicação;
- **autor**: autor da publicação;
- **titulo** : titulo da publicação;
- **categoria** : categoria a qual o documento pertence;
- **conceitos** : os principais conceitos destacados na obra;
- **palchave**: palavras chave presentes no documento;
- **resumo**: resumo da obra;
- **urlpublicacao** : lugar na www onde esta disponível o documento;
- **numpágs** : número de páginas do documento.

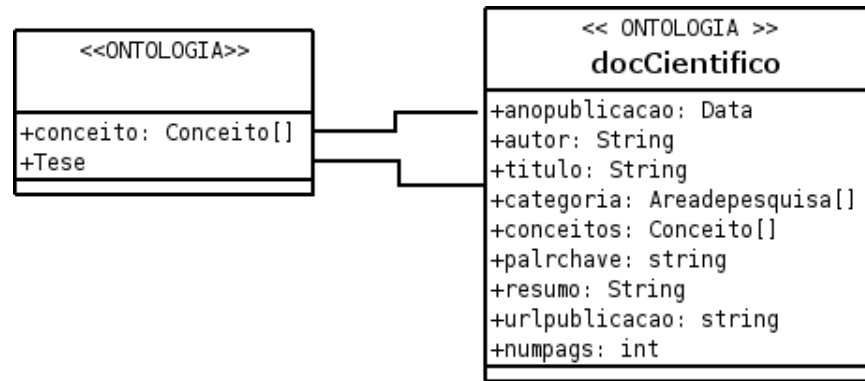


Figura 13 – Diagrama da classe docCientífico.

3.4.2 Ontologia Pesquisador

A classe Pesquisador (figura 14) possui os seguintes atributos :

- **titulo:** título acadêmico;
- **areadepesquisa:** área de atuação do pesquisador;
- **publicações:** publicações feitas pelo pesquisador ou de seu interesse;
- **telefone:** telefone do local de trabalho do pesquisador ou seu contato;
- **organização :** organização a qual o pesquisador pertence;
- **endereço :** endereço do pesquisador (local da instituição).

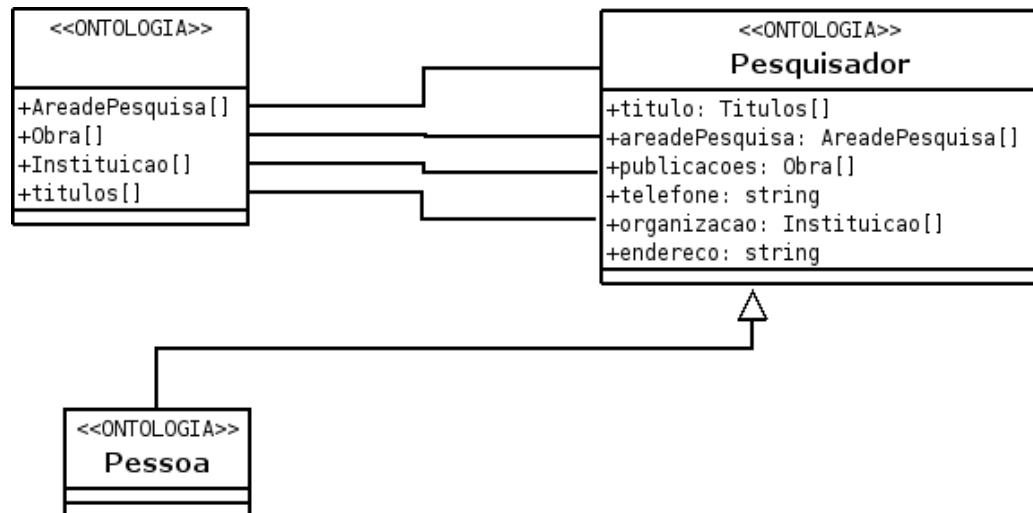


Figura 14 – Diagrama da classe Pesquisador.

3.4.3 Ontologia Obra

A classe Obra (figura 15) possui os atributos :

- **anodapublicacao** : ano em que a obra foi publicada;
- **autor** : autor que concebeu a obrar;
- **titulo** : título da obra.

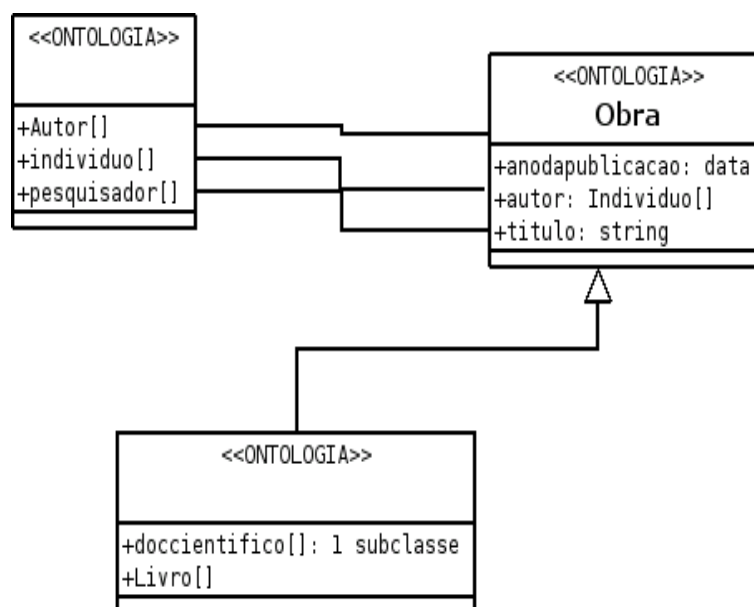


Figura 15 – Diagrama da classe Obra.

3.4.4 Ontologia Autor

A classe autor (figura 16) possui o atributo:

- **publicações** : publicações feitas pelo autor deriva da classe obra.

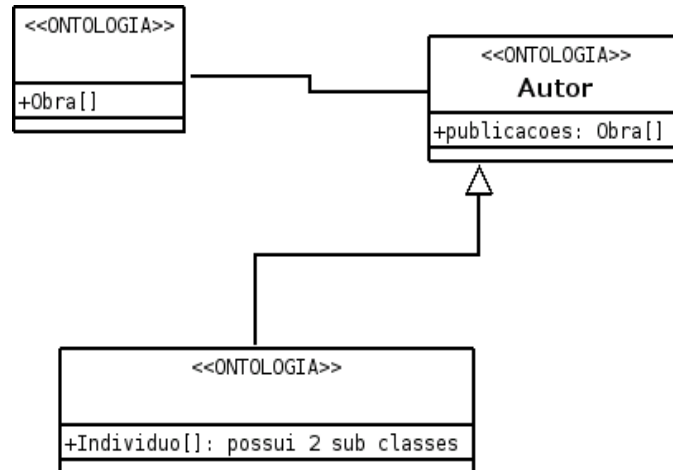


Figura 16 - Diagrama da classe Autor.

3.4.5 Ontologia Áreas de Pesquisa

A classe Áreas de Pesquisa (figura 17) possui os atributos:

- **grandearea** : grande área do conhecimento;
- **area** : área do conhecimento;
- **subarea** : subárea do conhecimento;
- **especialidade** : especialidade do conhecimento;
- **estuda** : o que se estuda nessa classe e depende de conceitos.

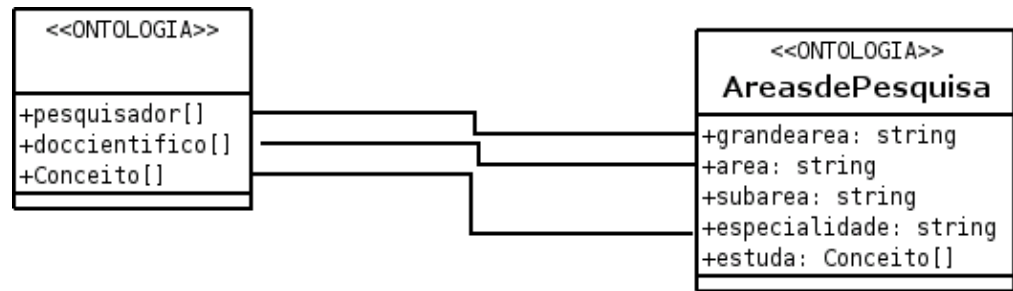


Figura 17 – Diagrama da classe Área de Pesquisa.

3.4.6 Ontologia Conceito

A classe Conceito (figura 18) possui os atributos:

- **dependede** : conhecimento que depende de algum conceito;
- **areadeestudo** : área do conhecimento que se estuda;
- **nome** : nome para o conceito;
- **prerequisitopara**: conhecimento que serve para outro conceito.

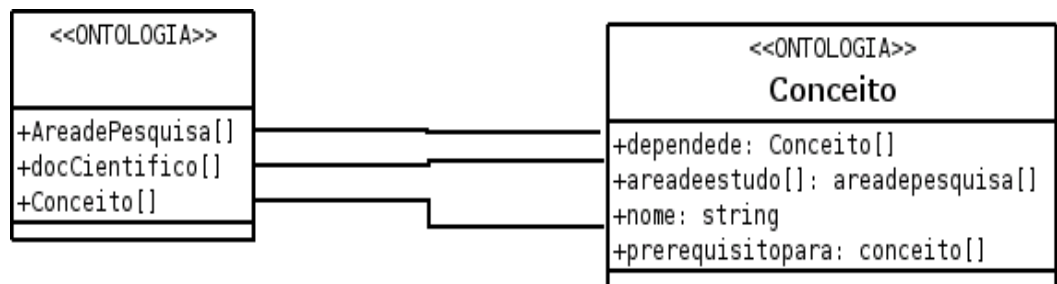


Figura 18 – Diagrama da classe Conceito.

3.4.7 Ontologia Livro

A classe Livro (figura 19) possui os atributos :

- **editora** : estabelecimento que editou o livro.

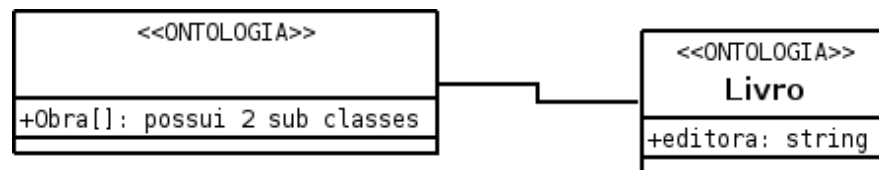


Figura 19 – Diagrama da classe Livro.

3.4.8 Ontologia Instituição

A classe Instituição (figura 20) possui os atributos:

- **endereco:** local do estabelecimento de pesquisa.
- **urldainstituicao:** lugar na rede www.
- **nome:** nome da instituição de pesquisa.

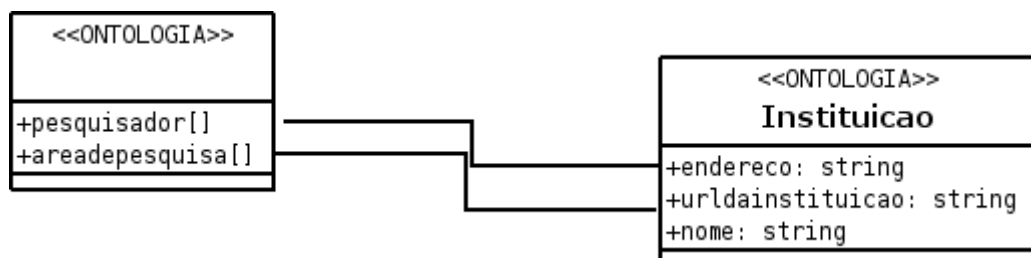


Figura 20 – Diagrama da classe Instituição.

3.4.9 Ontologia Título

A classe Titulo (figura 21) possui os atributos :

- **grau:** grau acadêmico do título (doutor, mestre, especialista, graduado);
- **Areadepesquisa:** área que originou ou pertence o título.



Figura 21 – Diagrama da classe Título.

3.5 Modelo de Dados

No item anterior foi descrito uma ontologia para os documentos, conceitos e pessoas que estão envolvidos no sistema REDMIS. Com base nisso, o passo seguinte foi a construção do banco de dados correspondente, para representar o conhecimento modelado por essa ontologia. Esse banco de dados é uma fonte de dados para o motor do sistema EDMIS, desenvolvido por Alves (2005). O banco de dados original da ferramenta encontra-se no Anexo A.

As entidades que fazem parte do banco de dados de REDMIS são:

- Conceitos
- Áreas de pesquisa
- Usuários
- Tipos de documentos

O modelo de dados apresentado na figura 22 foi a implementação da figura 17. Sua função é representar todos os termos que representem e forem cadastrados pelo pesquisador, no momento de seu cadastramento no sistema ou quando ele desejar alterar seus dados. Com base nessa representação, os cálculos de similaridade ontológica serão realizados, o que é discutido na seção 3.7.

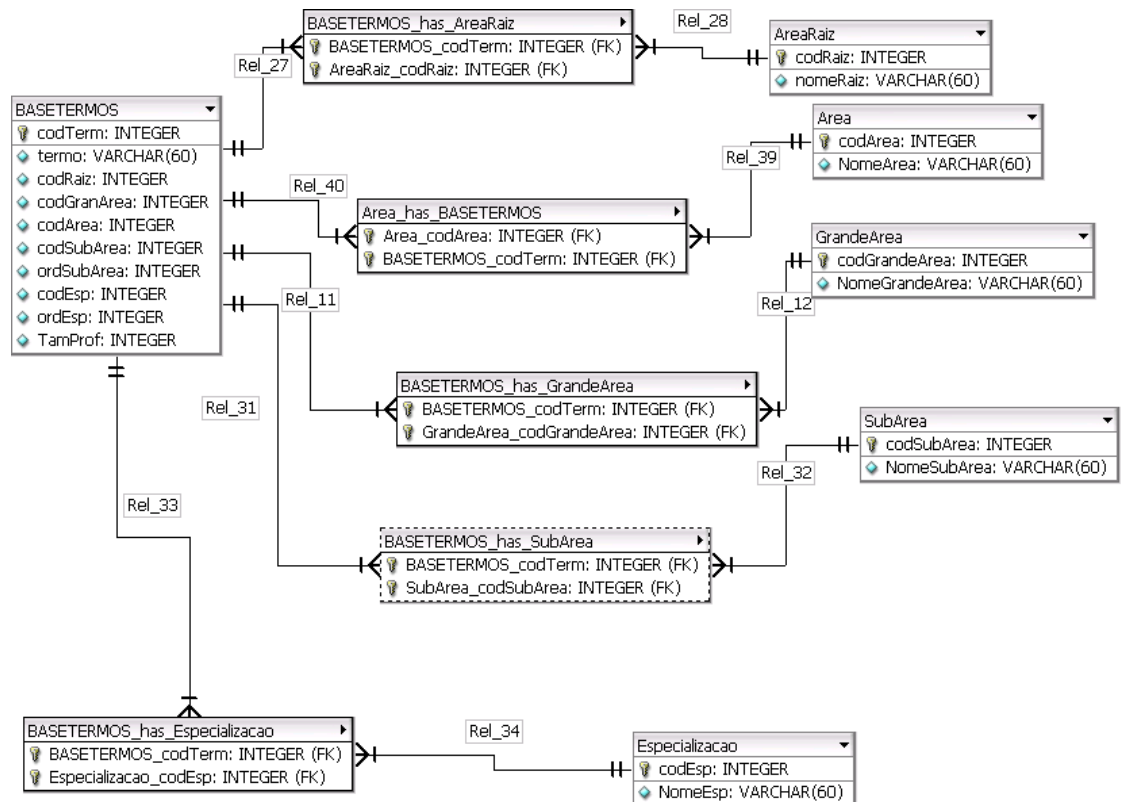


Figura 22 - Modelo de dados do sistema REDMIS que alimenta os termos para a representação ontológica fuzzy.

3.6 Arquitetura do Sistema REDMIS

O sistema EDMIS utiliza a tecnologia JSP, resultando no fato de que pode ser executado por meio de qualquer *browser* e por múltiplos usuários ao mesmo tempo. A arquitetura do sistema REDMIS, em termos de camadas, está esquematizada na figura 23.

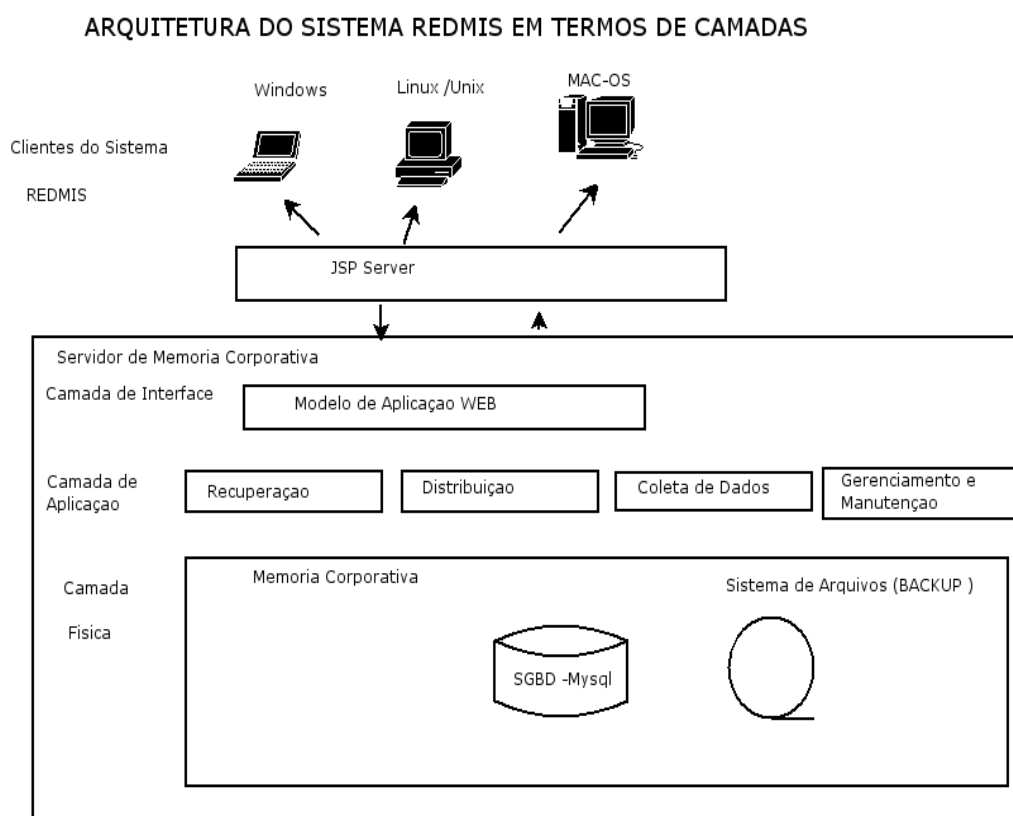


Figura 23 - Visão da Arquitetura do sistema REDMIS sob ponto de vista de camadas.

No item busca de documentos, estes são apresentados por meio de seu “*ranking*”, em ordem crescente de relevância. Esse mecanismo implementado, foi construído como sendo uma função nebulosa e com base no modelo de recuperação

de informação fuzzy e espaço vetorial. Neste mecanismo, um documento tem como saída um número de “*ranking*”, que está na escala de 0 a 1, que corresponde em relação aos termos de entrada (consultas), ao quanto o texto em análise tem correspondência de título ou conteúdo para aquele conjunto de entrada (Shamsfard et al, 2006). O mecanismo de indexação dos termos é ilustrada na figura 24.

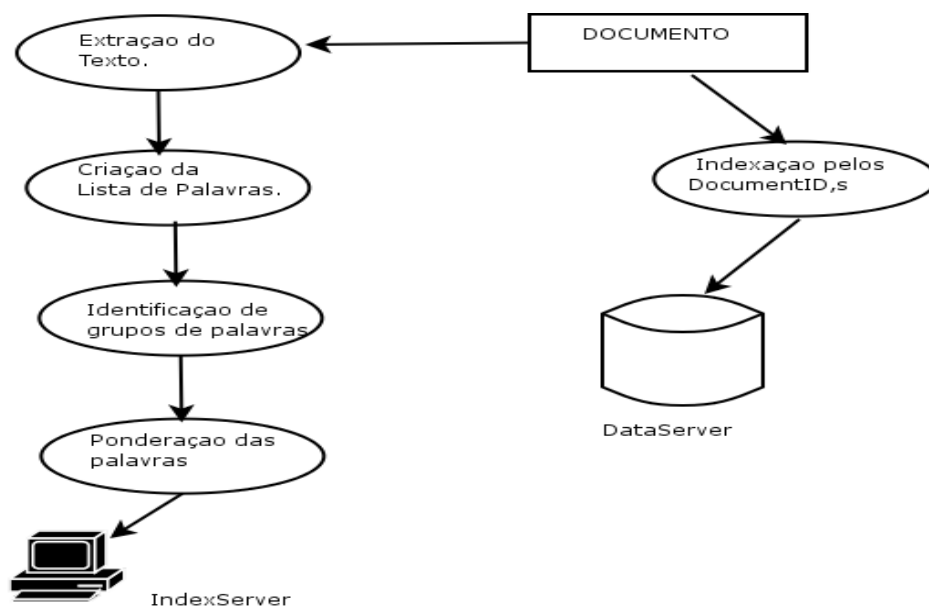


Figura 24 – Mecanismo de indexação de documentos em REDMIS.

A figura 25 apresenta o mecanismo do sistema EDMIS utilizado para a recuperação da informação, a partir de uma consulta, e a interação entre o *DataServer* e o *IndexServer*. A ferramenta REDMIS herda esse mecanismo e através da classificação do pesquisador usuário, por meio do módulo “Analisador Ontológico”, é feita de forma automática a recuperação da informação por meio da análise do perfil do pesquisador, como apresentado na figura 26.

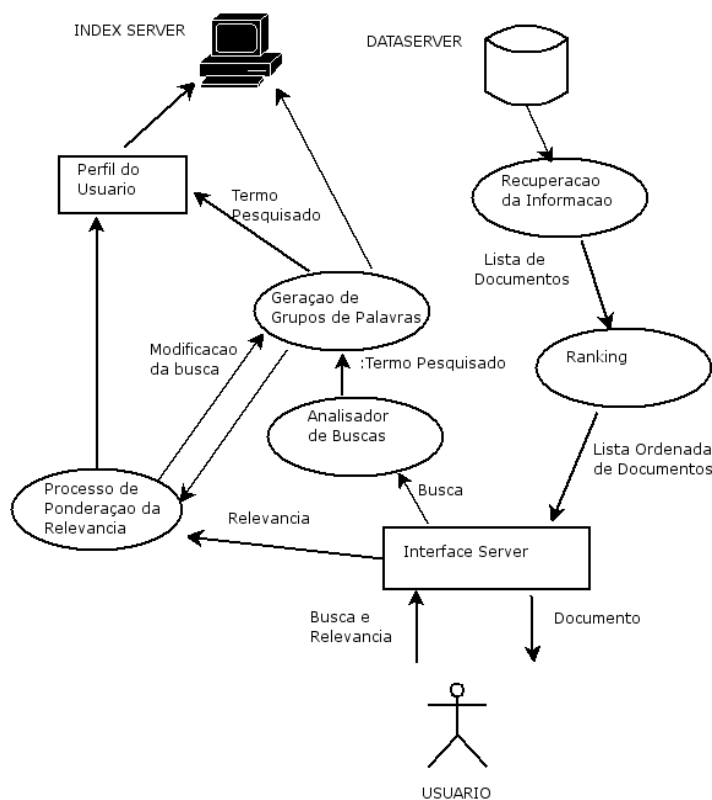
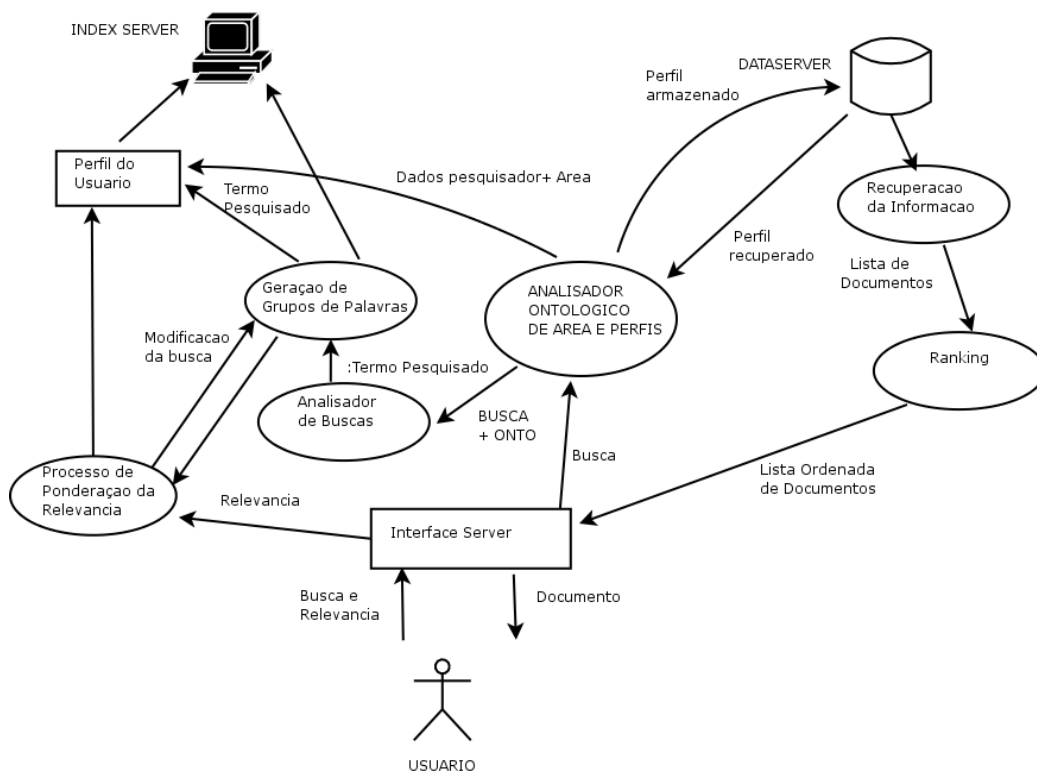


Figura 25 – Mecanismo de recuperação de documentos em EDMIS. Fonte: Alves (2005).



A figura 26 apresenta a modificação estrutural feita na ferramenta EDMIS, mostrando que a recuperação da informação é uma função não somente das consultas geradas pelo usuário pesquisador, mas também do seu perfil, baseado na representação ontológica de sua área de conhecimento. Esse módulo está disponível no *Interface Server* após o *login* e validação de usuário. O *IndexServer* do sistema REDMIS foi programado para possuir as *Stop Words* (palavras de uso comum) nos idiomas nativos do sistema, que são: inglês, português, espanhol, italiano e alemão, perfazendo um conjunto de 2774 palavras, que estão apresentadas no Apêndice B. Estes termos foram obtidos no portal da Universidade de Neuchatel disponível em: <<http://www.unine.ch/info/clef/>> (Acesso em: 15 Dez. 2006).

3.7 Módulo de Classificação Ontológica

Esse módulo aparece na figura 26 com sendo o “Analisador Ontológico de Área e Perfis”, sua função no sistema é receber os dados cadastrados que representam o perfil do pesquisador, assim como, os termos representativos seguindo-se a ordem: área raiz, grande área, área, sub área, especialidade e o termo.

3.7.1 Geração da Ontologia de Termos para Consulta

A ontologia para os termos representativa no sistema REDMIS, é representada por uma árvore de termos seguindo a hierarquia apresentada na figura 27. Cada termo é um registro da tabela BASETERMOS que está representada na figura 22.

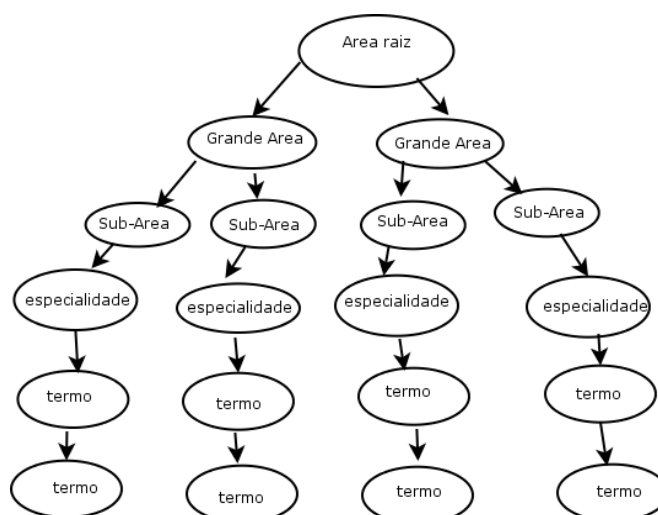


Figura 27 - Ontologia de representação de termos para elaboração da consulta.

3.7.2 Cálculo da Similaridade

A medida de similaridade ontológica é feita em nível de estruturas hierárquicas (Árvore) e, pode ser calculada por diferentes métodos (Cordì et al, 2005). Dentre esses métodos pode-se citar:

- a) *Zanobini's metric and Bouquet, Kouper e Scoz* – Essa métrica calcula a similaridade por meio de um conceito simples: seja $D(c1, c2)$ onde $c1$ e $c2$ são conceitos (nós da árvore) e D a distância entre eles. Essa métrica computa o caminho mínimo entre eles para o cálculo da similaridade.
- b) *van Harmelen's metric* – Essa métrica computa a similaridade por meio do comprimento do subtópico e do nível na árvore. É utilizada uma equação com componentes exponenciais naturais para esse cálculo.
- c) *Racca's metric* – Essa métrica avalia a similaridade dos termos em função de um *thesaurus*.
- d) *Palme's metric*- Essa métrica considera a distância entre os termos e o nó raiz , além do menor caminho entre eles (Bulskov et al., 2002).

O método mais utilizado atualmente, pela comunidade científica, é o método da métrica do caminho em comum (Cordi et al, 2004). Nesse trabalho, foi utilizado o conceito de similaridade fuzzy (Widyantoro e Yen ,2001) que utiliza um algoritmo para o cálculo da similaridade entre o termo e os termos que formam a ontologia, como apresentado na figura 28.

Algoritmo Cálculo da Similaridade de Rede de Termos Fuzzy

Início

Inicialização da Ontologia

$T = \{ t_1, t_2, t_3, \dots, t_m \}$ é uma lista de termos distintos extraídos de uma coleção C .

$0 < \alpha \leq 1$ onde α é um ponto de corte.

$O = \{ \}$ é uma ontologia de descrição vazia .

Fim da Inicialização.

Primeiro estagio – Para cada t_i e t_j pertencentes a T e t_i diferente de t_j faça:

Calcule a $\mu_{NT}(t_i, t_j)$ e $\mu_{NT}(t_j, t_i)$ como media de cada termo e relacao aos outros .

Selecione o correspondente $NT(t_p, t_q)$

$(t_p, t_q) = \max \{ \mu_{NT}(t_i, t_j) \text{ e } \mu_{NT}(t_j, t_i) \}$

$\mu_{NT}(t_i, t_j) \text{ e } \mu_{NT}(t_j, t_i) \geq \alpha$

adicione $\{ NT(t_p, t_q), \mu_{NT}(t_p, t_q) \}$ na ontologia

fim primeiro estagio.

Segundo estagio – Para cada $NT(t_i, t_j)$ pertencente a ontologia

Ache $P = \{ NT(t_i, t_{m1}), \mu_{NT}(t_{m1}, t_{m2}), \dots, NT(t_{mn}, t_j) \}$

$(t_p, t_q) = \min \{ \mu_{NT}(t_i, t_j) \text{ e } \mu_{NT}(t_j, t_i) \text{ pertencente a } P \}$

Se $\mu_{NT}(t_i, t_j) < \mu_{NT}(t_p, t_q)$ remova $NT(t_i, t_j)$ da ontologia.

Fim segundo estagio

Fim algoritmo.

Figura 28 - Cálculo da similaridade de rede de termos fuzzy. Fonte: Widyantoro e Yen (2001).

Cada par de termos possui um grau de similaridade que está entre [0,1], portanto, pode-se ter o seu cosseno correspondente da medida μNT , por meio do quantificador fuzzy gaussiano (Lee, 1995) apresentado na equação 3.

$$g(x, c, \sigma) = e^{-1/2((x-c)/\sigma)^2} \quad (\text{Eq- 3})$$

onde o valor de μNT , c é a quantidade de termos da amostra e sigma é o desvio padrão dos valores dessa amostra .

A aplicação da equação-3 nos valores obtidos pelo algoritmo da figura 28, comporta-se como um filtro, selecionando os termos de maior similaridade no centro da amostra e os menos similares nos extremos, como uma gaussiana em geral o faz. Isso permite selecionar o conjunto de termos mais significantes com base em um ponto de corte que representará os termos da consulta. Dessa forma, os termos que formarão a consulta poderão ser selecionados à medida que o pesquisador deseje recuperar os documentos por termos mais representativos do seu perfil. Essa opção é deixada a cargo do usuário do sistema, que fornece o valor da representação no intervalo de [0,1], o que implica em aumentar ou diminuir o conjunto de termos da consulta representativa.

Capítulo

4

Estudo de Caso

4.1 Introdução

O grupo alvo escolhido para a análise da adequação da ontologia e a capacidade de recuperação da informação por um grupo de pesquisas foi o grupo de pesquisas do CRAAM, que atualmente é formado por oito pesquisadores. No ano de 2006 foi incorporado à Escola de Engenharia da UPM. É constituído como grupo de pesquisas no Diretório de Pesquisas do Brasil do CNPq sob o nome de: “CRAAE e CRAAM (Centro de Radioastronomia e Astrofísica Mackenzie, convênio Mackenzie-INPE)” cujo ano de formação é de 1989, como um centro de excelência em pesquisa em Radioastronomia.

A ferramenta REDMIS é uma possível solução de T.I. (Tecnologia de Informação) que pode ajudar na gerência e na disponibilização de forma colaborativa dos conhecimentos, por vezes comuns, entre seus membros. Uma outra situação está no fato de que volume de publicações manipuladas pelos pesquisadores, tais como os recebidos por revistas, congressos ou fóruns, muitas vezes podem passar

despercebidos, e sem chamar a atenção a um, ou mais de um pesquisador como potencial de utilização, que tenha um possível interesse por essa informação. Pode ocorrer ainda a situação que, determinados assuntos que possuam área de conhecimento em comum, não sejam notados por um dado pesquisador. Restando apenas o mecanismo descrito anteriormente, da Socialização do conhecimento como canal de disseminação de tais assuntos.

4.2 Abordagem do Problema

O CRAAM é um centro de pesquisas em Radioastronomia cuja sede é em São Paulo Capital. É considerado um centro de estudos avançados na área de pesquisa e cujos membros pesquisadores atingem a mais alta classificação em termos de pesquisa segundo o CNPq-LATTES. Esse centro utiliza o radio-observatório situado na cidade de Atibaia-SP, denominado “ROI-Radio Observatório de Itapetinga”. Além deste, utiliza o radio-observatório no Ceará na cidade de Eusébio, denominado de “Radio Observatório Espacial do Nordeste”(ROEN); possui por meio de cooperação os sinais e a infra-estrutura do *Instituto de Astronomia y Física del Espacio* (Buenos Aires) o Telescópio Solar Submilimétrico-SST (El Leoncito) e também, finalmente, o Laboratório Antártico de Pesquisas Ionosféricas(Antártica) situado na Estação de Comandante Ferraz (apoio CNPq, Marinha do Brasil e Ministério do Meio Ambiente). Fonte: home-page da Universidade P. Mackenzie - Pós-Graduação em Engenharia Elétrica online (2006).

Dessa forma como pode ser observado, mesmo que um dado pesquisador não esteja em seu loco central, é possível que uma vez situado num desses Rádio Observatórios, ao realizar suas atividades de pesquisa e utilizando o sistema REDMIS via WWW, possa criar e estabelecer como mencionado anteriormente, uma comunidade virtual de trabalho colaborativo, perante seus pares distantes geograficamente. Qualquer pesquisador do grupo, mesmo que não presente geograficamente ou temporalmente, pode via esse ambiente de trabalho utilizar a ferramenta REDMIS como uma ferramenta de CSCW. Para sua implementação é feito nas próximas seções uma representação ontológica da área de Radioastronomia (ontologia de domínio).

4.3 Características do Grupo Associadas ao Sistema

A arquitetura do sistema REDMIS é definida segundo o modelo proposto no capítulo 3. Para isso utilizaremos uma descrição por meio do uso de ontologias de domínio. Os módulos descritos como “GRUPO-PEOPLE” aqui será caracterizado por “CRAAM-PEOPLE”. Para que os pesquisadores usuários desse sistema possam utilizá-lo, é descrito abaixo a classificação da especialidade Radioastronomia e seus vínculos com as Grandes Áreas da Ciência; de acordo com a classificação estabelecida pelo órgão federal brasileiro sob título de Fundação, o CNPq – Conselho Nacional de Desenvolvimento Científico e Tecnológico, subordinado ao Ministério de Ciência e Tecnologia.

Para visualizar as áreas e subáreas envolvidas foi utilizado a descrição por mapas conceituais, por meio do Software de domínio livre IHCM CmapTools (disponível em: <<http://cmcp.ihmc.us/download>>), juntamente com suas plataformas LATTES (CNPq, 2006).

4.4 Taxonomia de Classificação por meio de Mapas Conceituais

Os mapas conceituais, desenvolvidos por John Novak, são ferramentas para organizar e representar o conhecimento (Novak, 1977). Os mapas conceituais têm por objetivo apresentar, na forma gráfica, os conceitos considerados relevantes pelo autor para a compreensão de um novo conceito. Pode-se construí-los na forma de diagramas hierárquicos que indicam as inter-relações entre conceitos, os quais refletem a estrutura cognitiva do indivíduo sobre um determinado assunto. Essa técnica será utilizada posteriormente na pesquisa para caracterização das áreas correlatas do grupo de pesquisas utilizado no estudo de caso.

Para melhor compreensão, e posterior modelagem da estrutura do grupo de pesquisas, foi construído, com base nas informações disponíveis dos pesquisadores do grupo na plataforma LATTES do CNPq, o mapa conceitual, contendo as áreas, subáreas e especialidades dos pesquisadores do grupo. A figura 29 indica os mapas conceituais de modo não hierárquico e a figura 30 de forma hierárquica.



Figura 29 - Mapa Conceitual não Hierárquico das Linhas de Pesquisa do Grupo CRAAM.

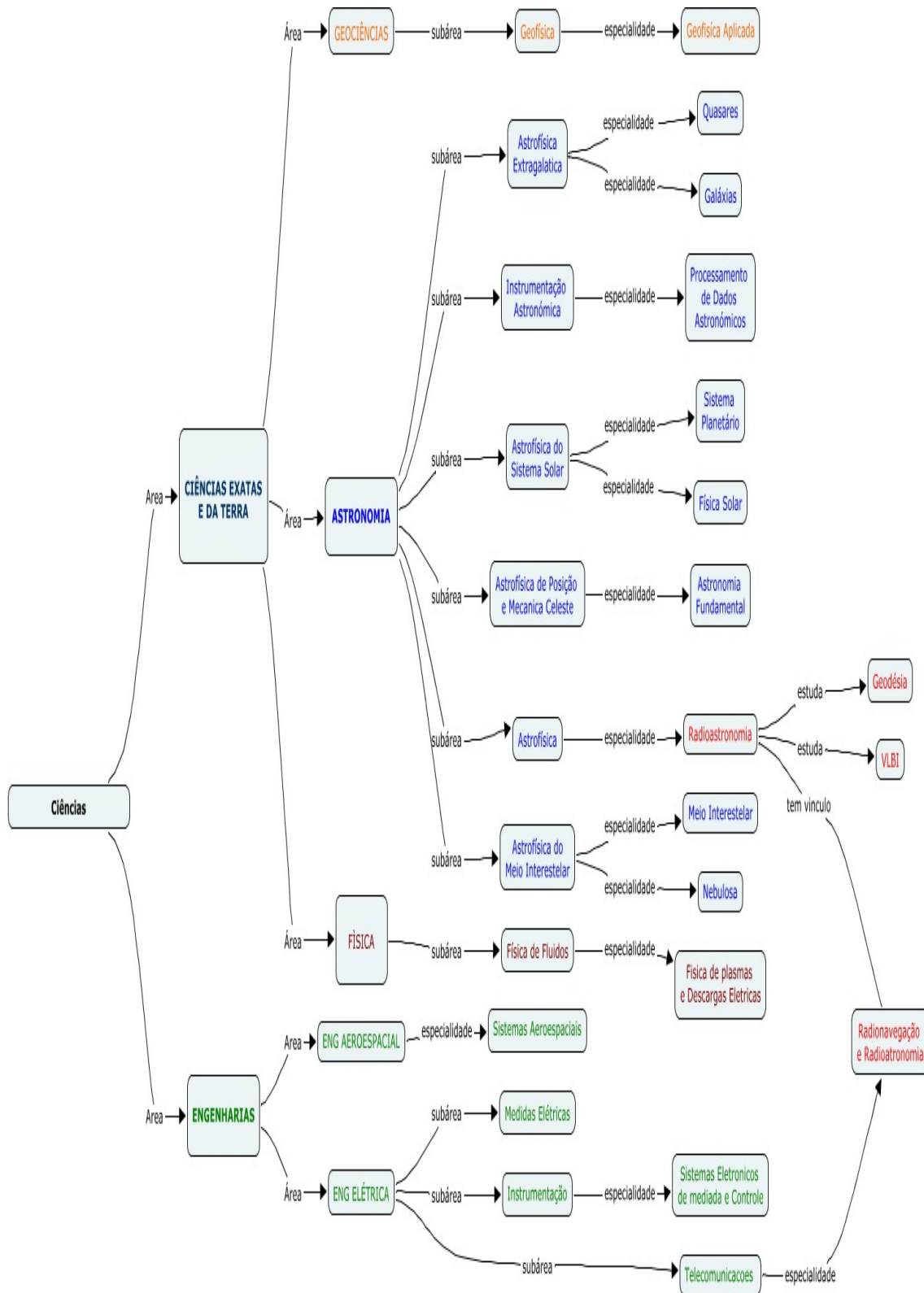


Figura 30 - Mapa Conceitual Hierárquico das Linhas de Pesquisa do Grupo CRAAM

Pode-se observar nas figuras 29 e 30 que a especialidade Radioastronomia está associada tanto à área de Engenharia Elétrica como à Astronomia e o vínculo é destacado, exceto o conceito de Radionavegação que só está presente na Engenharia Elétrica.

Estes mapas conceituais servem para a implementação do banco de dados representativo dos pesquisadores do grupo de teste, pois quando acessarem sua conta no sistema, após prévio cadastro, estarão personalizados. Como descrito no capítulo 3, tanto os documentos como conceitos afins serão devidamente “roteados” e “filtrados” para esse determinado pesquisador. As informações representadas nas figuras 13 a 21 passam a ser a alimentação inicial da base de conhecimento. Cumpre ressaltar que o pesquisador pode a qualquer momento alterar as informações de seu perfil. Essa capacidade permite ao sistema adaptar e melhor atender o pesquisador na evolução de sua pesquisa.

4.5 Linhas de Pesquisa dos Membros do Grupo

Nesse tópico são apresentadas as linhas de pesquisa dos membros do grupo. Esses dados são originados a partir da plataforma LATTES (CNPq, 2006), dessa forma é estabelecido um perfil inicial do pesquisador no momento do cadastramento do usuário no sistema. O grupo é formado por oito pesquisadores que serão denominados anonimamente pela representação de P1 a P8. Os dados iniciais em forma de mapas conceituais individuais estão disponíveis no Apêndice C.

4.6 Utilização da Ferramenta com Base nos Perfis e Documentos

Nessa etapa do trabalho foi utilizado para verificação das capacidades da ferramenta um conjunto de 100 documentos, pertinentes à área de Astronomia, Radioastronomia e Física. Esses documentos foram obtidos por meio de busca nos portais da WEB especializados na área, onde é frequente a publicação dos pesquisadores relativos às principais áreas do conhecimento. Para a coleta dos 100 documentos, foram obtidos todos os documentos de acesso gratuito, pois esses portais cobram de não sócios os documentos lá armazenados. Os portais utilizados foram das editoras: *ELSEVIER- ScienceDirect*, *Springer* e *Blackwell Synergy*. As fontes principais foram: *The Astrophysical Journal* e *The Royal Astronomical Society*.

O conjunto de documentos totalizam o tamanho de 92.1 Mb em formato PDF. Desse total, têm-se 11 arquivos de resumos perfazendo 586 Kb de tamanho e 89 arquivos com artigos completos perfazendo 91,5 Mb também em formato PDF. Os arquivos foram nomeados aleatoriamente por DocPDFXXX onde XXX varia de 001 a 100. A lista completa de todos os documentos encontra-se no Apêndice A onde a tabela A.1 mostra os arquivos com seus nomes e títulos, além da descrição de autoria no sentido de terem sido escritos por algum ou mais de um dos membros do grupo de pesquisa de teste. Nos ensaios foram estabelecidos como usuários da ferramenta, conforme a classificação ontológica, cada um dos oito integrantes do grupo de teste.

4.7 Condições dos Ensaios

A ferramenta REDMIS produz para cada usuário a opção de visão pelos tipos de documentos, ou seja, uma visão geral de todos os documentos que fazem parte do banco de dados assim com os que estão no grupo dos Favoritos do Usuário. Na ferramenta REDMIS, além disso, cada usuário tem uma visão de acordo com o perfil de sua pesquisa e área de interesse que é proveniente da sua representação ontológica no momento do seu cadastro no sistema. Também, há uma visão dos usuários que depositaram documentos na base de dados, e se estes lhe são úteis por grau de relevância. Também foi alvo a verificação de autoria e participação dos membros do grupo.

4.8 Busca do Conhecimento : Recuperação da Informação

Nessa seção são apresentados os conhecimentos obtidos da CM- Assets (Memória de Conhecimento Corporativa) armazenada no banco de dados de REDMIS. As consultas foram feitas automaticamente por meio do *login* do usuário, pois a consulta é elaborada automaticamente por meio da leitura dos termos que compõem e correspondem ao perfil do usuário, com base na avaliação ontológica dos termos que representam o usuário. Para cada um dos oito pesquisadores, o sistema possui um conjunto de palavras chaves que obedecem a uma hierarquia ontológica, discutida no capítulo 2. Foi elaborado um conjunto de termos que representam o pesquisador e que formam a consulta que alimenta o motor de busca. Este último, retorna os documentos que possuem maior similaridade com o pesquisador. Esses documentos são apresentados de forma decrescente de importância e relevância para a consulta, formando um “*ranking*” de documentos pelo grau de similaridade do documento com aquela consulta (Shamsfard et al, 2006).

Os ensaios foram efetuados de acordo com os seguintes critérios:

- a) Recuperação da Informação com 100% da representação ontológica pela média fuzzy e 100% dos documentos recuperados (0% da média do *ranking*);
- b) Recuperação da Informação com 70% (maior ou igual a) da representação ontológica pela média fuzzy e 100% dos documentos recuperados (0% da média do *ranking*);
- c) Recuperação da Informação com 100% (maior ou igual a) da representação ontológica pela média fuzzy e 70% da média do *ranking* dos documentos recuperados;
- d) Recuperação da Informação com 70% (maior ou igual a) da representação ontológica pela média fuzzy e 70% da média do *ranking* dos documentos recuperados.

4.8.1 Elaboração dos Termos Ontológicos Fuzzy para Consulta

Cada pesquisador tem seu perfil representado por um conjunto de termos coletados manualmente dos seus “Curriculum Lattes”, obedecendo a uma estrutura hierárquica conceitual, perfanzendo-se um total em média de 15 termos representativos. Essas informações fazem parte do banco de dados da representação ontológica do pesquisador. Aplicou-se o método de classificação e cálculo de similaridade discutido na seção 3.7.

Cada termo na estrutura hierárquica (árvore de termos) possui uma similaridade que foi obtida pelo algoritmo descrito na figura 28 e denominado aqui de similaridade média-fuzzy. Esse valor representa o grau de similaridade entre o

termo e os outros termos, respeitada a hierarquia ontológica elaborada e a presente no banco de dados (Área raiz, grande área, área, sub área, especialidade), esse valor associado é a similaridade do termo (Bulskov et al, 2002). Nesse sentido ele representa um ângulo entre o vetor do termo e o restante dos termos da ontologia, calculando-se o cosseno desse valor tem-se a similaridade do termo em relação à ontologia. Para cada pesquisador foi gerada uma tabela onde os termos e os cálculos dos seus pesos são apresentados na seguinte ordem: *simi_conj*(ângulo de similaridade), média fuzzy e o cosseno do ângulo.

Para a elaboração dos ensaios, as buscas foram realizadas pelo sistema com as opções:

- a) *Query Q1*- elaborada com o conjunto de termos que possuam similaridades iguais ou superiores a 70% da representação ontológica fuzzy;
- b) *Query Q2*- elaborada com o conjunto de todos os termos da representação ontológica fuzzy.

Cada pesquisador foi representado como uma consulta ao motor de busca de REDMIS conforme a figura 26. As tabelas 1 a 8 apresentam os resultados obtidos pela ferramenta, através do módulo ontológico que gera as consultas para a busca dos documentos. Os termos associados às sub-áreas e especialidades foram formadas pelo conjunto:

$T(\text{sub-áreas, especialidades}) = \{\text{Radioastronomy, Geophysics, Physics, Cosmology, Extragalactic, Astrophysics, Atmosphere, Solar-Physics}\}$

Tabela 1- Pesquisador P1 e os seus termos associados.

<i>Pesquisador P1</i>			
Termo	Simi_conj (ângulo)	media_fuzzy	Simil (cosseno)
<i>SST</i>	<i>0,00000</i>	<i>1,00000</i>	<i>1,00000</i>
<i>212 GHz 405 GHz</i>	<i>0,00000</i>	<i>1,00000</i>	<i>1,00000</i>
<i>Ionosphere</i>	<i>0,05669</i>	<i>0,98627</i>	<i>0,99839</i>
<i>Plasma</i>	<i>0,07706</i>	<i>0,95511</i>	<i>0,99703</i>
<i>Very Low Frequency</i>	<i>0,10963</i>	<i>0,86394</i>	<i>0,99400</i>
<i>Submillimeter</i>	<i>0,13112</i>	<i>0,73886</i>	<i>0,99142</i>
<i>SST submillimeter</i>	<i>0,14648</i>	<i>0,59708</i>	<i>0,98929</i>
<i>Extra Low Frequency</i>	<i>0,17918</i>	<i>0,36409</i>	<i>0,98399</i>
<i>Telescope</i>	<i>0,17796</i>	<i>0,28260</i>	<i>0,98421</i>
<i>SST solar</i>	<i>0,16632</i>	<i>0,25543</i>	<i>0,98620</i>
<i>Quasar</i>	<i>0,17314</i>	<i>0,16669</i>	<i>0,98505</i>
<i>Quasar</i>	<i>0,17881</i>	<i>0,10264</i>	<i>0,98406</i>
<i>Quasars</i>	<i>0,17814</i>	<i>0,07037</i>	<i>0,98418</i>
<i>Stars</i>	<i>0,17756</i>	<i>0,04687</i>	<i>0,98428</i>
<i>Planets</i>	<i>0,17706</i>	<i>0,03033</i>	<i>0,98437</i>

As consultas relativas a esse pesquisador pelo critério estabelecido são:

$Q1 = \{ SST, 212\text{ GHz } 405\text{ GHz}, Ionosphere, Plasma, Very\ Low\ Frequency, Submillimeter \}$

$Q2 = \{ SST, 212\text{ GHz } 405\text{ GHz}, Ionosphere, Plasma, Very\ Low\ Frequency, Submillimeter, SST\ submillimeter, Extra\ Low\ Frequency, Telescope, SST\ solar, Quasar, Quasar, Quasars, Stars, Planets \}$

Tabela 2 – Pesquisador P2 e os seus termos associados.

<i>Pesquisador P2</i>			
Termo	simi_conj	media_fuzzy	Simil (cos)
<i>SST</i>	<i>0,00000</i>	<i>1,00000</i>	<i>1,00000</i>
<i>212 GHz 405 GHz</i>	<i>0,00000</i>	<i>1,00000</i>	<i>1,00000</i>
<i>Ionosphere</i>	<i>0,05454</i>	<i>0,98725</i>	<i>0,99851</i>
<i>Plasma</i>	<i>0,07413</i>	<i>0,95830</i>	<i>0,99725</i>
<i>Very Low Frequency</i>	<i>0,10607</i>	<i>0,87185</i>	<i>0,99438</i>
<i>Submillimeter</i>	<i>0,12715</i>	<i>0,75204</i>	<i>0,99193</i>
<i>SST submillimeter</i>	<i>0,14220</i>	<i>0,61469</i>	<i>0,98991</i>
<i>Extra Low Frequency</i>	<i>0,17388</i>	<i>0,38578</i>	<i>0,98492</i>
<i>Telescope</i>	<i>0,17254</i>	<i>0,30452</i>	<i>0,98515</i>
<i>SST solar</i>	<i>0,15529</i>	<i>0,30401</i>	<i>0,98797</i>
<i>Thermal emission</i>	<i>0,16625</i>	<i>0,19145</i>	<i>0,98621</i>
<i>Flare</i>	<i>0,16588</i>	<i>0,14079</i>	<i>0,98627</i>
<i>X-ray</i>	<i>0,15312</i>	<i>0,14057</i>	<i>0,98830</i>
<i>X-ray</i>	<i>0,14218</i>	<i>0,14038</i>	<i>0,98991</i>
<i>solar minimum</i>	<i>0,14190</i>	<i>0,10579</i>	<i>0,98995</i>
<i>solar maximum</i>	<i>0,14165</i>	<i>0,07823</i>	<i>0,98998</i>

As consultas relativas a esse pesquisador pelo critério estabelecido são:

$Q1 = \{SST, 212\text{ GHz } 405\text{GHz}, Ionosphere, Plasma, Very Low Frequency, Submillimeter\}$

$Q2 = \{SST, 212\text{GHz} 405\text{GHz}, Ionosphere, Plasma, VeryLowFrequency, Submillimeter, SST$
Submillimeter, Extra Low Frequency, Telescope, SST solar, Thermal emission, Flare, X-ray, X-ray, solar
minimum, solar maximum\}

Tabela 3 – Pesquisador P3 e os seus termos associados.

<i>Pesquisador P3</i>			
Termo	simi_conj	media_fuzzy	Simil (cos)
<i>SST</i>	<i>0,00000</i>	<i>1,00000</i>	<i>1,00000</i>
<i>212 GHz 405 GHz</i>	<i>0,00000</i>	<i>1,00000</i>	<i>1,00000</i>
<i>Ionosphere</i>	<i>0,06439</i>	<i>0,98252</i>	<i>0,99793</i>
<i>Plasma</i>	<i>0,08752</i>	<i>0,94295</i>	<i>0,99617</i>
<i>Very Low Frequency</i>	<i>0,10865</i>	<i>0,86704</i>	<i>0,99410</i>
<i>Submillimeter</i>	<i>0,12238</i>	<i>0,76941</i>	<i>0,99252</i>
<i>CME Coronal Mass Ejection</i>	<i>0,18501</i>	<i>0,44099</i>	<i>0,98294</i>
<i>solar maximum</i>	<i>0,18223</i>	<i>0,35322</i>	<i>0,98344</i>
<i>energetic protons</i>	<i>0,20444</i>	<i>0,18986</i>	<i>0,97918</i>
<i>Solar Flares</i>	<i>0,20028</i>	<i>0,13914</i>	<i>0,98001</i>
<i>Geomagnetic Storm</i>	<i>0,21643</i>	<i>0,06136</i>	<i>0,97667</i>
<i>solar minimum</i>	<i>0,21196</i>	<i>0,04117</i>	<i>0,97762</i>

As consultas relativas a esse pesquisador pelo critério estabelecido são:

$Q1 = \{SST, 212 \text{ GHz } 405 \text{ GHz}, Ionosphere, Plasma, Very Low Frequency, Submillimeter \}$

$Q2 = \{SST, 212 \text{ GHz } 405 \text{ GHz}, Ionosphere, Plasma, Very Low Frequency, Submillimeter, CME, Coronal Mass Ejection, solar maximum, energetic protons, Solar Flares, Geomagnetic Storm, solar minimum\}$

Tabela 4 – Pesquisador P4 e os seus termos associados.

<i>Pesquisador P4</i>			
Termo	simi_conj	media_fuzzy	Simil (cos)
<i>SST</i>	<i>0,00000</i>	<i>1,00000</i>	<i>1,00000</i>
<i>212 GHz 405 GHz</i>	<i>0,00000</i>	<i>1,00000</i>	<i>1,00000</i>
<i>Nebula</i>	<i>0,06439</i>	<i>0,98252</i>	<i>0,99793</i>
<i>Data analysis</i>	<i>0,14381</i>	<i>0,85332</i>	<i>0,98968</i>
<i>Submillimeter</i>	<i>0,15325</i>	<i>0,75286</i>	<i>0,98828</i>
<i>Array radio astronomical</i>	<i>0,22118</i>	<i>0,42476</i>	<i>0,97564</i>
<i>Solar maximum</i>	<i>0,21284</i>	<i>0,33838</i>	<i>0,97744</i>
<i>Magnetohydrodynamic</i>	<i>0,21011</i>	<i>0,25071</i>	<i>0,97801</i>
<i>Solar Flares</i>	<i>0,20486</i>	<i>0,18856</i>	<i>0,97909</i>
<i>Stars</i>	<i>0,20369</i>	<i>0,13002</i>	<i>0,97933</i>
<i>Solar minimum</i>	<i>0,19997</i>	<i>0,09229</i>	<i>0,98007</i>
<i>Quasars</i>	<i>0,19941</i>	<i>0,05941</i>	<i>0,98018</i>

As consultas relativas a esse pesquisador pelo critério estabelecido são:

$Q1=\{SST, 212\text{ GHz } 405\text{ GHz}, Ionosphere, Plasma, Very Low Frequency, Submillimeter\}$

$Q2=\{SST, 212\text{ GHz } 405\text{ GHz}, Nebula, Data analysis, Submillimeter, Array radio astronomical, solar maximum, Magnetohydrodynamic, Solar Flares, Stars, solar minimum, Quasars \}$

Tabela 5 – Pesquisador P5 e os seus termos associados.

<i>Pesquisador P5</i>			
Termo	simi_conj	media_fuzzy	simil (cos)
<i>SST</i>	<i>0,00000</i>	<i>1,00000</i>	<i>1,00000</i>
<i>212 GHz 405 GHz</i>	<i>0,00000</i>	<i>1,00000</i>	<i>1,00000</i>
<i>Heliosphere</i>	<i>0,06368</i>	<i>0,98290</i>	<i>0,99797</i>
<i>Magnetic</i>	<i>0,09605</i>	<i>0,93169</i>	<i>0,99539</i>
<i>Submillimeter</i>	<i>0,11505</i>	<i>0,85216</i>	<i>0,99339</i>
<i>Planetary</i>	<i>0,12807</i>	<i>0,75046</i>	<i>0,99181</i>
<i>Stars</i>	<i>0,13737</i>	<i>0,63677</i>	<i>0,99058</i>
<i>Magnetohydrodynamic</i>	<i>0,14407</i>	<i>0,52180</i>	<i>0,98964</i>
<i>Solar Flares</i>	<i>0,14616</i>	<i>0,42775</i>	<i>0,98934</i>
<i>Solar Atmosphere</i>	<i>0,14782</i>	<i>0,34148</i>	<i>0,98909</i>
<i>solar minimum</i>	<i>0,14919</i>	<i>0,26548</i>	<i>0,98889</i>
<i>Astrophysics</i>	<i>0,15722</i>	<i>0,17289</i>	<i>0,98767</i>

As consultas relativas a esse pesquisador pelo critério estabelecido são:

$Q1 = \{SST, 212\text{ GHz } 405\text{ GHz}, Heliosphere, Magnetic, Submillimeter, Planetary\}$

$Q2 = \{SST, 212\text{ GHz } 405\text{ GHz}, Heliosphere, Magnetic, Submillimeter, Planetary,$

$Stars, Magnetohydrodynamic, Solar\ Flares, Solar\ Atmosphere, solar\ minimum, Astrophysics\}$

Tabela 6 - Pesquisador P6 e os seus termos associados.

<i>Pesquisador P6</i>			
Termo	simi_conj	media_fuzzy	Simil (cos)
<i>Quasars</i>	0,18477	0,98556	0,98298
<i>Quasar</i>	0,18477	0,93882	0,98298
<i>ejecta mass</i>	0,24501	0,77379	0,97014
<i>22GHz 43GHz</i>	0,27513	0,55851	0,96239
<i>Galactic nuclei</i>	0,29320	0,35284	0,95732
<i>non thermal radio sources</i>	0,35486	0,10982	0,93769
<i>extragalactic radio sources</i>	0,37309	0,03559	0,93121
<i>Stars</i>	0,34955	0,02160	0,93953
<i>3C 273</i>	0,31071	0,02142	0,95212
<i>super luminal</i>	0,31619	0,00728	0,95043
<i>Nucleous</i>	0,30424	0,00400	0,95407
<i>radio source</i>	0,31508	0,00086	0,95077
<i>Comets</i>	0,30506	0,00042	0,95383

As consultas relativas a esse pesquisador pelo critério estabelecido são:

$Q1 = \{Quasars, Quasar, ejecta\ mass\}$

$Q2 = \{Quasars, Quasar, ejecta\ mass, 22GHz\ 43GHz, Galactic\ nuclei, non\ thermal\ radio\ sources, extragalactic\ radio\ sources, Stars, 3C\ 273, super\ luminal, Nucleous, radio\ source, Comets\}$

Tabela 7 - Pesquisador P7 e os seus termos associados.

<i>Pesquisador P7</i>			
Termo	simi_conj	media_fuzzy	simil (cos)
<i>Quasars</i>	0,18477	0,98556	0,98298
<i>Quasar</i>	0,18477	0,93882	0,98298
<i>VLBI</i>	0,18477	0,86428	0,98298
<i>Galaxies</i>	0,18477	0,76896	0,98298
<i>galactic nuclei</i>	0,22091	0,55355	0,97570
<i>non thermal radio sources</i>	0,29463	0,21813	0,95691
<i>extragalactic radio sources</i>	0,32146	0,08405	0,94878
<i>Stars</i>	0,30437	0,05460	0,95404
<i>OJ287</i>	0,29108	0,03428	0,95793
<i>Astrometry</i>	0,28045	0,02081	0,96093
<i>Geodetic</i>	0,27175	0,01221	0,96330
<i>radio source</i>	0,28530	0,00307	0,95958
<i>Interferometry</i>	0,27757	0,00161	0,96172

As consultas relativas a esse pesquisador pelo critério estabelecido são:

$Q1 = \{ \textit{Quasars}, \textit{Quasar}, \textit{VLBI}, \textit{Galaxies} \}$

$Q2 = \{ \textit{Quasars}, \textit{Quasar}, \textit{VLBI}, \textit{Galaxies}, \textit{galactic nuclei}, \textit{non thermal radio}$

$\textit{sources}, \textit{extragalactic}, \textit{radiosources}, \textit{Stars}, \textit{OJ287}, \textit{Astrometry}, \textit{Geodetic}, \textit{radio source}, \textit{Interferometry} \}$

Tabela 8 - Pesquisador P8 e os seus termos associados.

<i>Pesquisador P8</i>			
Termo	simi_conj	media_fuzzy	simil (cos)
<i>SST</i>	<i>0,00000</i>	<i>1,00000</i>	<i>1,00000</i>
<i>212 GHz 405 GHz</i>	<i>0,00000</i>	<i>1,00000</i>	<i>1,00000</i>
<i>impulsive flare</i>	<i>0,14800</i>	<i>0,91104</i>	<i>0,98907</i>
<i>solar bursts</i>	<i>0,15171</i>	<i>0,83817</i>	<i>0,98851</i>
<i>17GHz 48GHz</i>	<i>0,19695</i>	<i>0,62573</i>	<i>0,98067</i>
<i>Magnetic structures</i>	<i>0,22711</i>	<i>0,40542</i>	<i>0,97432</i>
<i>SST submillimeter</i>	<i>0,22196</i>	<i>0,30776</i>	<i>0,97547</i>
<i>Polarization</i>	<i>0,21809</i>	<i>0,22526</i>	<i>0,97631</i>
<i>microwaves x-ray</i>	<i>0,21509</i>	<i>0,15896</i>	<i>0,97696</i>
<i>SST solar</i>	<i>0,19358</i>	<i>0,15842</i>	<i>0,98132</i>
<i>Flare</i>	<i>0,19917</i>	<i>0,09408</i>	<i>0,98023</i>
<i>22GHz</i>	<i>0,19849</i>	<i>0,06097</i>	<i>0,98037</i>

A consultas relativas a esse pesquisador pelo critério estabelecido são:

$Q1 = \{ SST, 212 \text{ GHz } 405 \text{ GHz}, \text{impulsive flare}, \text{solar bursts} \}$

$Q2 = \{ SST, 212 \text{ GHz } 405 \text{ GHz}, \text{impulsive flare}, \text{solar bursts}, 17\text{GHz } 48\text{GHz}, \text{Magnetic structures}, \text{SST submillimeter}, \text{polarization}, \text{microwaves x-ray}, \text{SST solar}, \text{flare}, 22\text{GHz} \}$

4.9 Recuperação da Informação Obtida pela Ferramenta

O pesquisador pode escolher a linha de corte de seu interesse no processo de recuperação da informação no sistema REDMIS. Ele pode selecionar os documentos segundo um grau de relevância (ponto de corte), levando-se em conta seu perfil

ontológico. Esse é o ponto de grande importância desta ferramenta em relação à ferramenta EDMIS.

Os pontos de corte utilizados para os ensaios foram “alto” e “baixo”, onde “alto” significa selecionar os documentos com valores de “*ranking*” superiores a 70% do valor médio obtido da amostra recuperada; e “baixo”, significa recuperar e apresentar 100% dos documentos recuperados (0% da média do *ranking*).

4.9.1 Resultados Obtidos pelo Sistema

Estão apresentados os gráficos representativos, obtidos para cada pesquisador, pelos valores numéricos gerados após as consultas (*query's*) elaboradas automaticamente pela ferramenta, e com os pontos de corte “alto” e “baixo” utilizados separadamente, aplicando-se o critério discutido na seção 4.8.1. As tabelas de origem estão apresentadas no Apêndice D.

As discussões apresentadas utilizaram para os nomes dos documentos os dígitos de seus nomes, por exemplo o DocPdf018 é referenciado no texto simplesmente como 18.

O gráfico 1 apresenta os documentos recuperados para a consulta Q1 e os dois pontos de corte 70% e 100% dos documentos recuperados. Para esse pesquisador, em 70% dos documentos recuperados observa-se uma concentração em torno dos 30 primeiros documentos. Utilizando-se, arbitrariamente para fins de análise, como ponto de corte o valor de 70% do “*ranking*” máximo obtido, e considerando apenas os documentos com “*ranking*” superior a 0,12, apenas, nove documentos mais significantes são obtidos: Doc 1, 2, 9, 10, 22, 23, 24, 25 e 26. Eles

possuem apenas um dos termos representativos da ontologia, retornando documentos que podem eventualmente ser relevantes para o pesquisador, uma vez analisado o perfil do seu mapa conceitual, que é o mais completo do grupo, e que a princípio tem interesse em todas aquelas áreas. Porém, quando se leva em conta todos os documentos recuperados, observa-se que esses documentos passam a ter um “ranking” reduzido e os documentos com maior significância passam a ser os da faixa nominal de 28 a 30 , 40 a 43 e 46 a 53, perfazendo doze documentos recuperados ou 12% da amostra. Analisando-se os seus conteúdos, observa-se mais de um termo da sua ontologia presente nos documentos e que o pesquisador é um dos autores nos doze documentos. Os últimos documentos também possuem uma pequena relevância para o pesquisador, porém não comparável à amostra central (documentos de 30 a 60).

No gráfico 2, têm-se todos os termos ontológicos representativos do pesquisador e por esse motivo observam-se alguns resultados diferentes da análise anterior:

- a) Os documentos iniciais com 70 % dos documentos retornados perdem representatividade (1 a 10) mas ressalta-se ainda os documentos 26, 27 e 28 com pouca representatividade para o pesquisador.
- b) Com 100% da amostra de documentos, os documentos mais relevantes são os 29 e 30, onde o pesquisador é um dos autores. Da mesma forma, a faixa de documentos situada entre 30 e 60, mais precisamente o conjunto {41,42,49,50,51,52,53}, retornou exatamente os documentos onde o pesquisador é um dos autores exceto o documento 52.

c) Os documentos finais passam a ser insignificantes para o interesse do pesquisador.

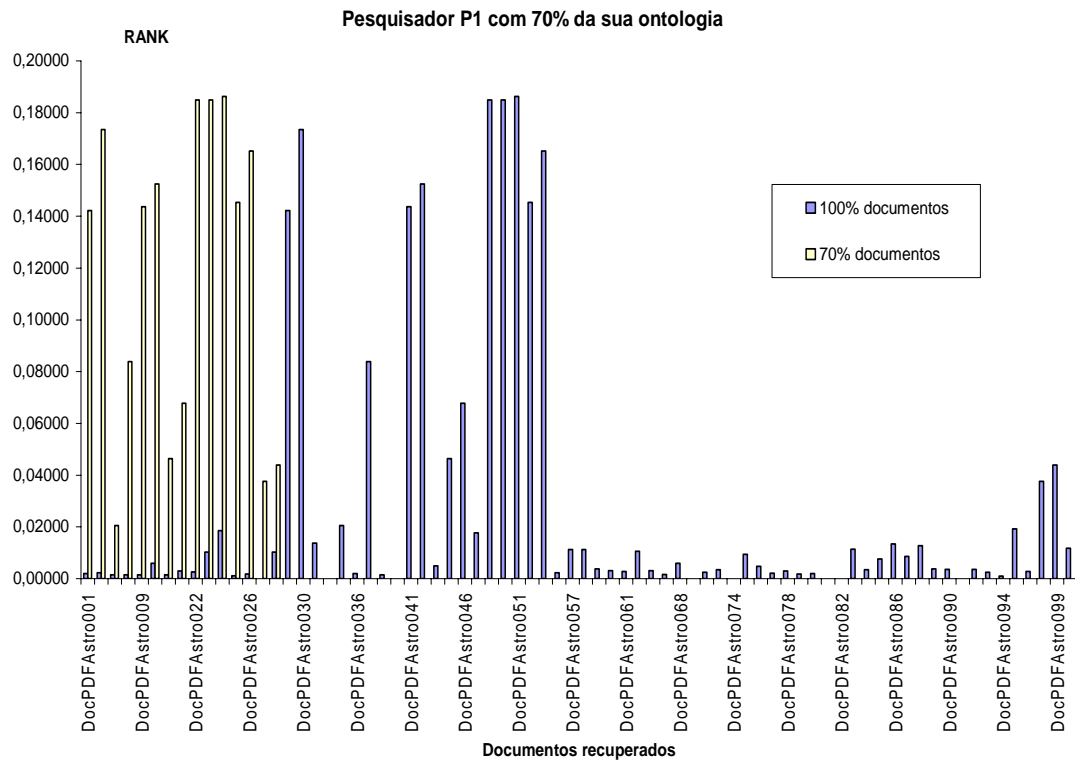


Gráfico 1- Documentos recuperados para a o pesquisador P1 para os pontos de corte alto e baixo e com 70% da sua ontologia.

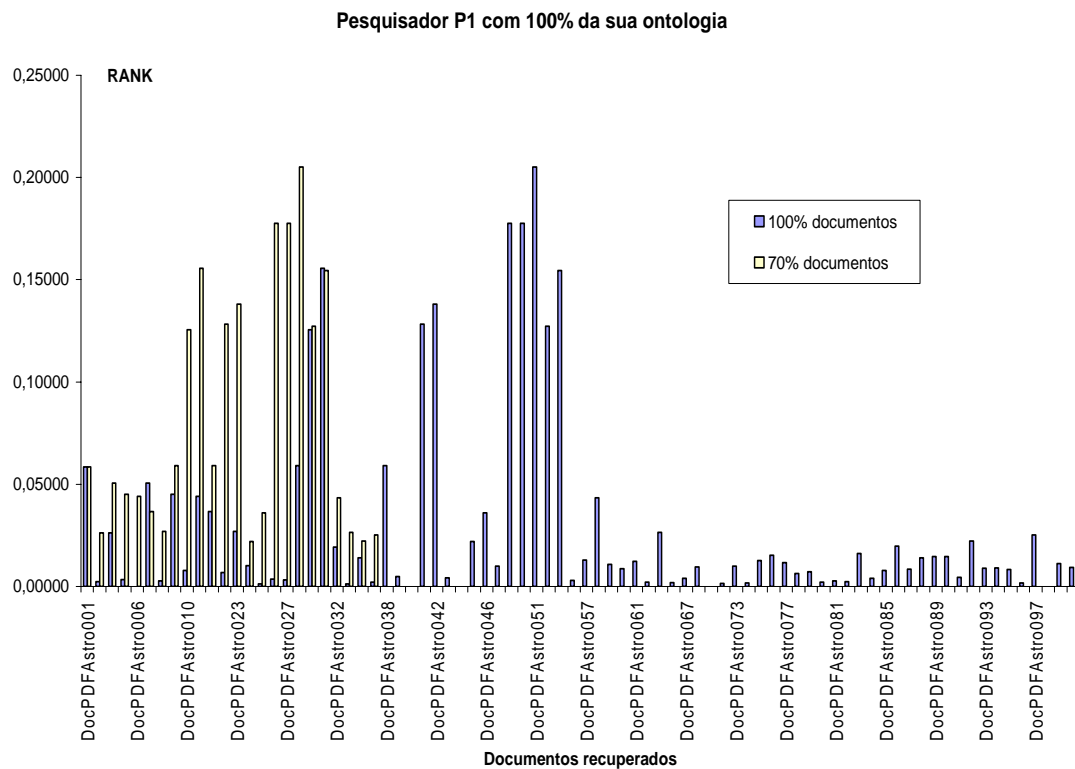


Gráfico 2- Documentos recuperados para o pesquisador P1 para os pontos de corte alto e baixo e com 100% da sua ontologia.

O gráfico 3 apresenta em ordem decrescente de “*ranking*” os documentos retornados pela ferramenta, indicando que os documentos mais representativos, em relação ao seu perfil, são aqueles em que o mesmo é um dos autores, no caso 80% dos documentos é autor e 20% não são de sua autoria, mas todos representando o perfil definido.

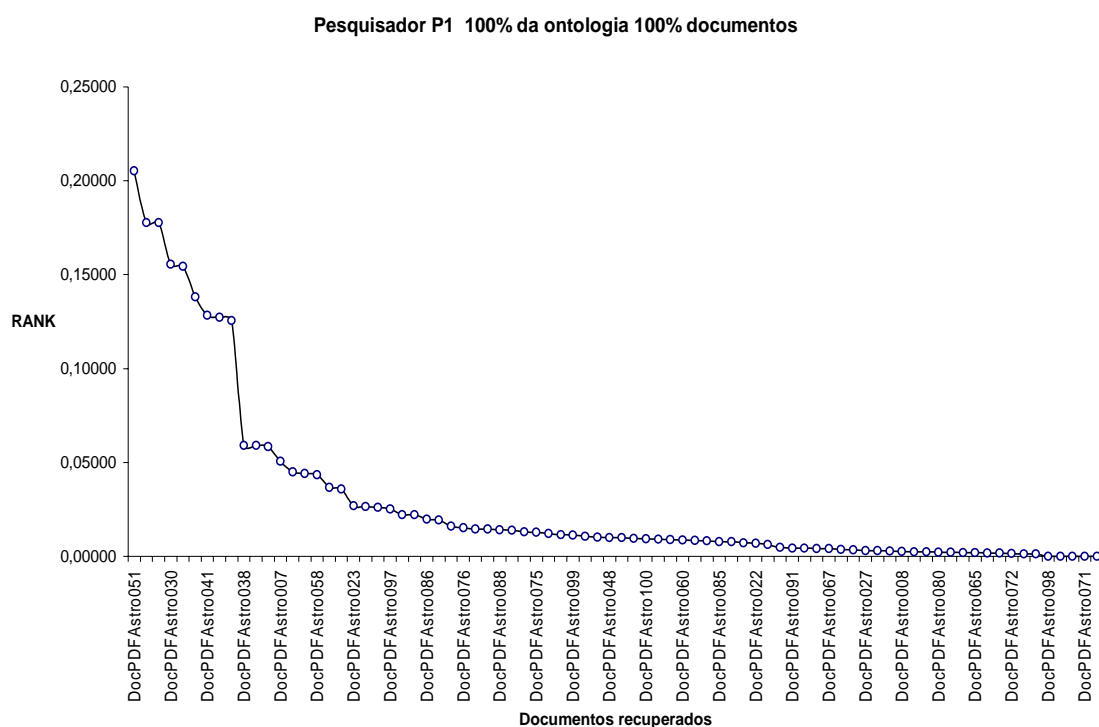


Gráfico 3 - Documentos recuperados para a o pesquisador P1 para o ponto de corte baixo.

O pesquisador P2 possui pelo gráfico 4 um perfil semelhante ao do pesquisador P1, e nesse sentido, para 70% da sua ontologia e com 70% dos documentos mais relevantes o comportamento de retorno de documentos na faixa de 1 a 26 se repete , ou seja, são documentos que pouco representam o perfil do pesquisador, porém, pelo seu mapa conceitual, alguns podem ser relevantes. Nesse caso todos os documentos superiores ao 26 são irrelevantes. Mas quando se consideram todos os documentos da amostra, observa-se que são retornados os documentos {29,30,38,41,42,49,50,51,52,53} onde é um dos autores, com excessão do documento 38 e 52 . Particularmente o documento 38 que não é de sua autoria possui quase todos os termos de seu perfil. Os documentos 88, 97 e 99 passam a representar

uma pequena relevância comparada aos documentos na faixa central , onde o documento 97 não possui correspondência ao pesquisador .

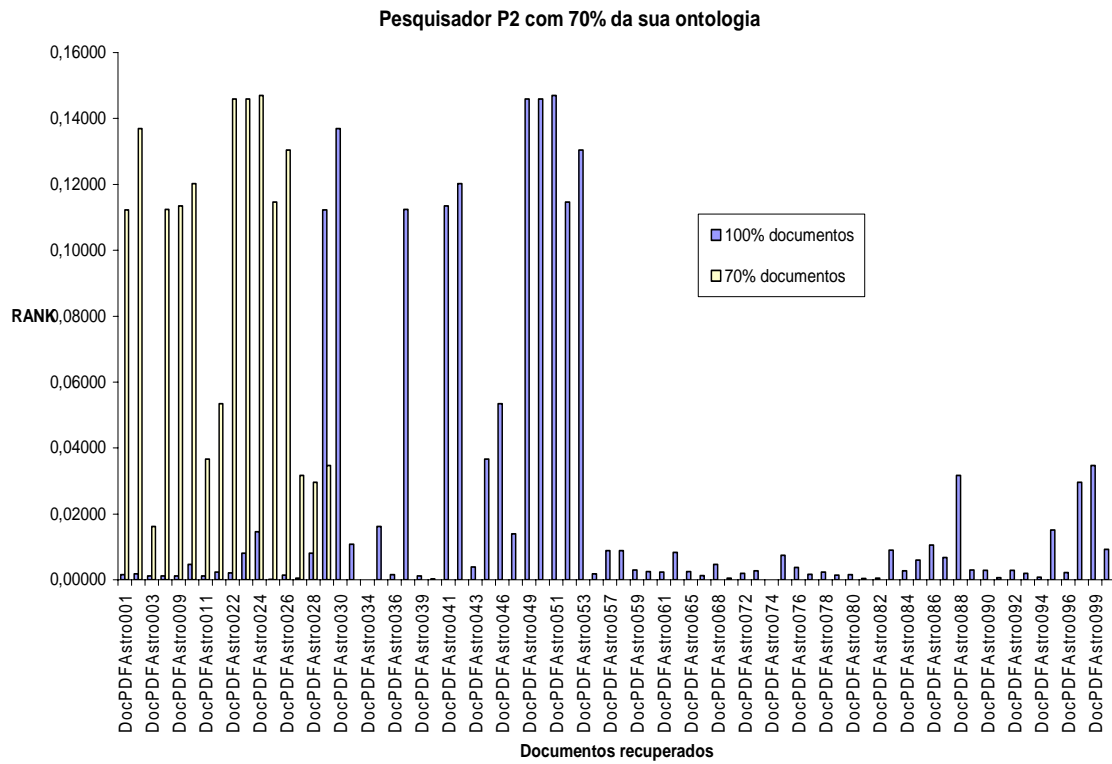


Gráfico 4 - Documentos recuperados para o pesquisador P2 para os pontos de corte alto e baixo e com 70% da sua ontologia.

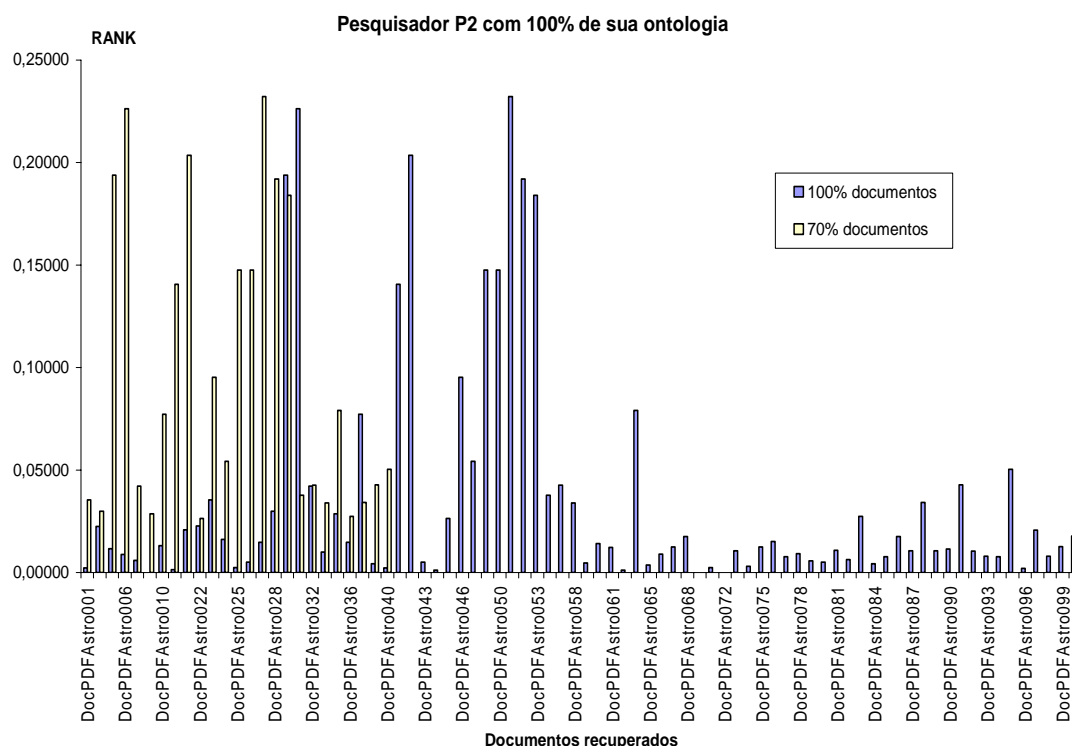


Gráfico 5 - Documentos recuperados para o pesquisador P2 para os pontos de corte alto e baixo e com 100% da sua ontologia.

O gráfico 5 apresenta para a ontologia completa e 0 % dos documentos, uma relevância comparável aos 100% dos documentos para o conjunto recuperado {3,6,11,21,25,26,27,28,29}. O pesquisador é autor do documento 29, porém, os documentos 25,26,27,28 não representam valor de conteúdo para o pesquisador, mas foram selecionados pela ferramenta por possuírem pelo menos dois termos da ontologia completa. Para os 100% dos documentos o conjunto retornado é semelhante ao apresentado no gráfico 4, ou seja independente da ontologia o sistema recupera o mesmo conjunto de dados mais significantes.

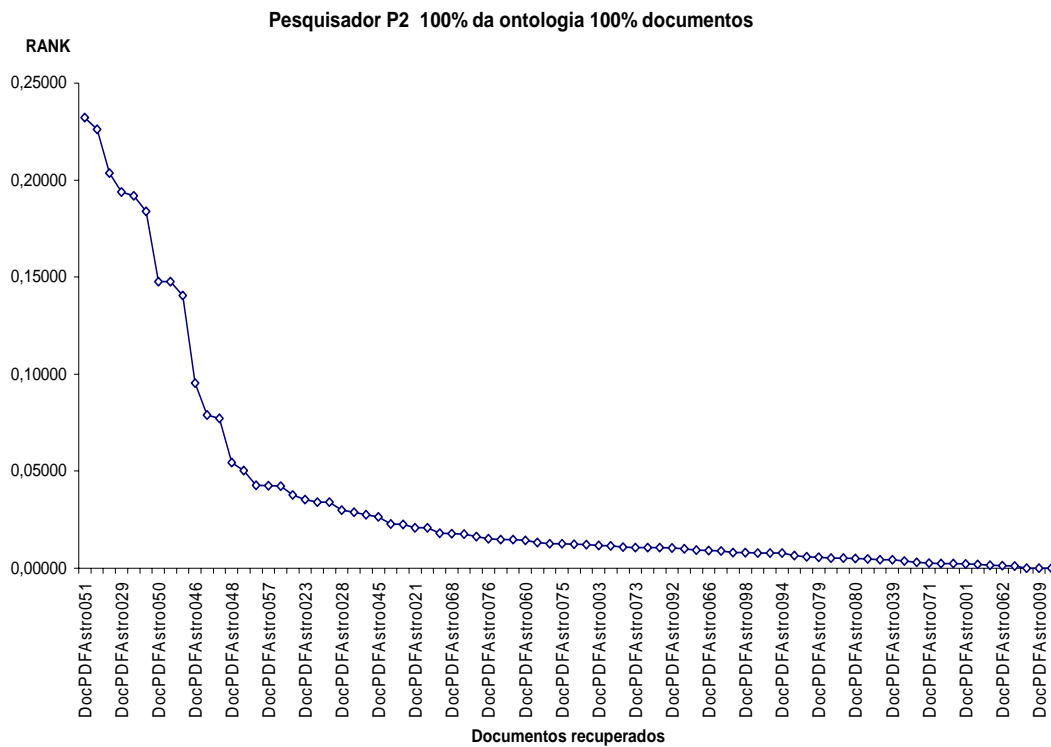


Gráfico 6 - Documentos recuperados para a o pesquisador P2 para o pontos de corte baixo.

O gráfico 6 apresenta em ordem decrescente de “*ranking*” os documentos retornados pela ferramenta indicando que os documentos mais representativos em relação ao seu perfil, são aqueles em que o mesmo é um dos autores, no caso 80% dos documentos é autor e 20% não são de sua autoria, mas todos representando o perfil definido. É um resultado semelhante ao obtido pelo pesquisador P1.

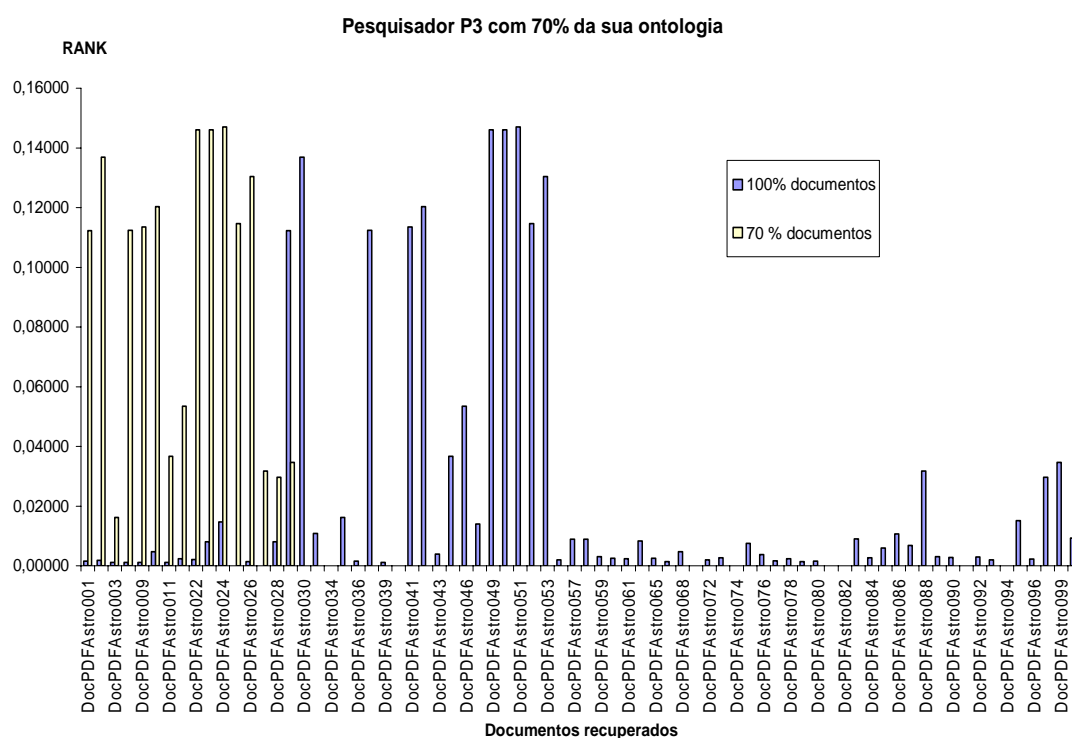


Gráfico 7- Documentos recuperados para o pesquisador P3 para os pontos de corte alto e baixo e com 70% da sua ontologia.

O gráfico 7 apresenta para 70% da ontologia e 70% dos documentos o retorno do seguinte conjunto de documentos { 1,2,7,9,10,22,23,24,25,26}, que não representam o perfil do pesquisador, mas foram recuperados por terem pelo menos um dos termos da ontologia presentes em seus conteúdos. Os resultados com a representação ontológica completa do autor tornam os documentos da amostra anterior irrelevantes em termos de “ranking”, e o conjunto resposta apresentado é formado pelos documentos { 29,30,38,41,42,49,50,51,52,53 }, que é um resultado semelhante ao obtido nessas mesmas condições de teste para os pesquisadores P1 e P2. Tal fato é explicado por terem um perfil ontológico semelhante.

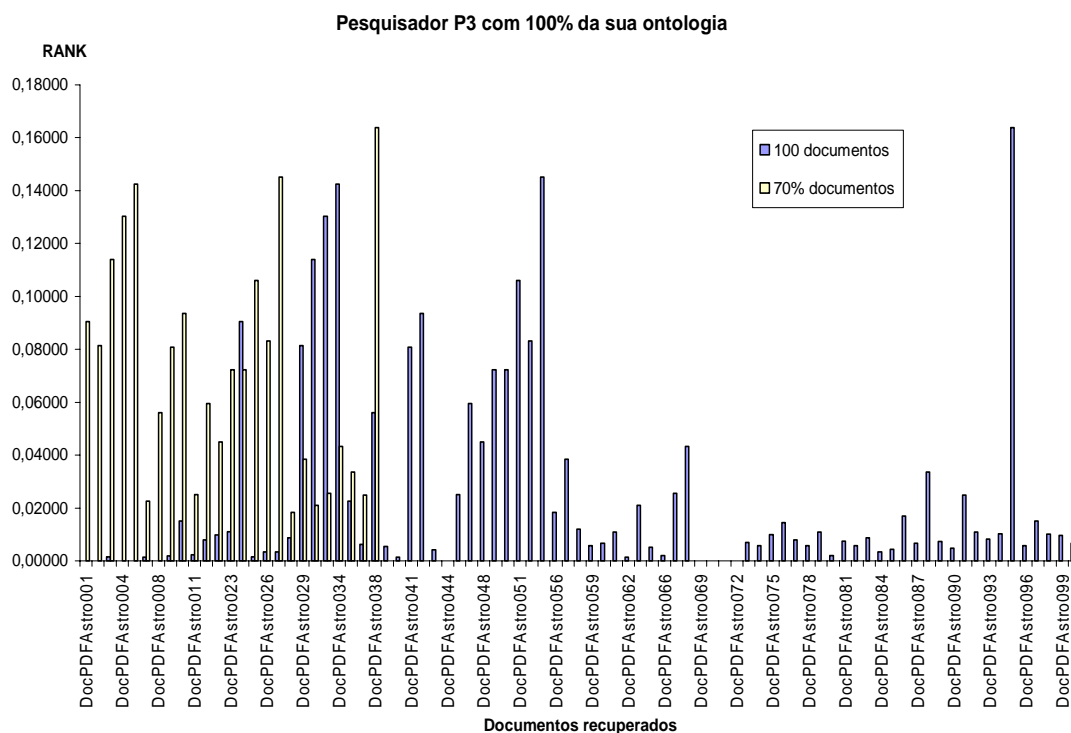


Gráfico 8 - Documentos recuperados para o pesquisador P2 para os pontos de corte alto e baixo e com 100% da sua ontologia.

O gráfico 8 apresenta para 100% da ontologia e 70% dos documentos o conjunto de documentos formados por {1,2,3,4,6,27 e 38} onde para seu perfil o documento 38 possui alguma significância. Levando-se em conta todos os documentos retornados, o perfil do pesquisador fica semelhante ao dos pesquisadores P1 e P2, mas destaca-se o conjunto de documentos {30,32,34,41,42,51,53,95}, nesses documentos o pesquisador é um dos autores, mais especificamente nos documentos 30,34,41,51,53 e nos documentos 32 e 95 a representatividade já esta incluída no próprio título do documento.

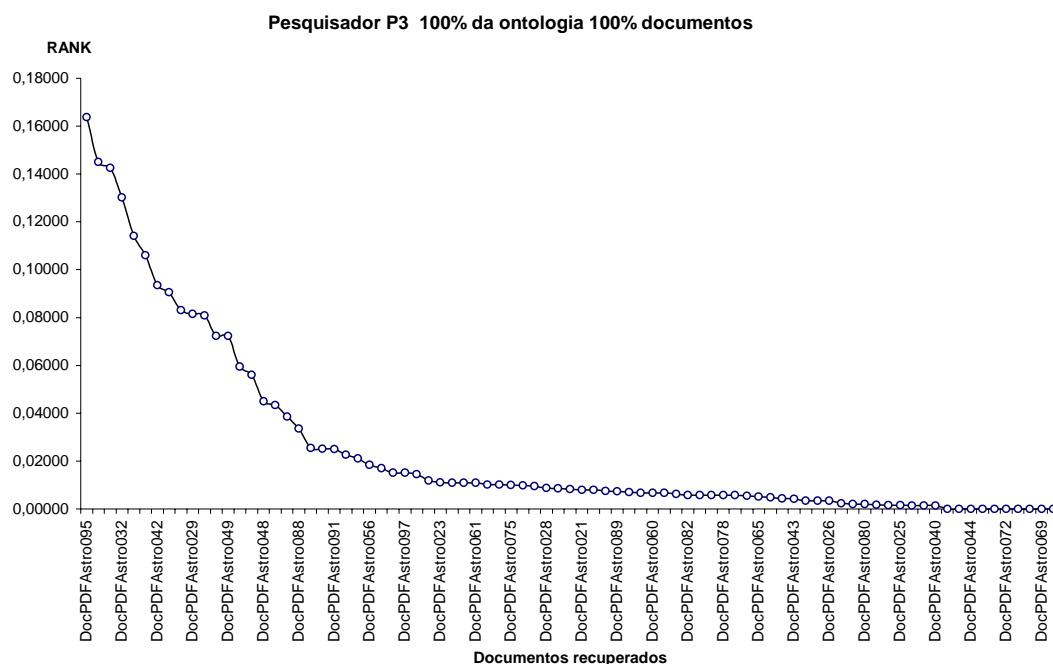


Gráfico 9- Documentos recuperados para a o pesquisador P3 para os pontos de corte baixo.

O gráfico 9 apresenta em ordem decrescente de “*ranking*” os documentos retornados pela ferramenta indicando que os documentos mais representativos, em relação ao seu perfil, são aqueles em que o mesmo é um dos autores, no caso 60% dos documentos recuperados é autor e 40% não são de sua autoria, mas todos representam o perfil afinado. É um resultado que indica que o perfil do pesquisador é similar ao conjunto de documentos que formam a base de teste.

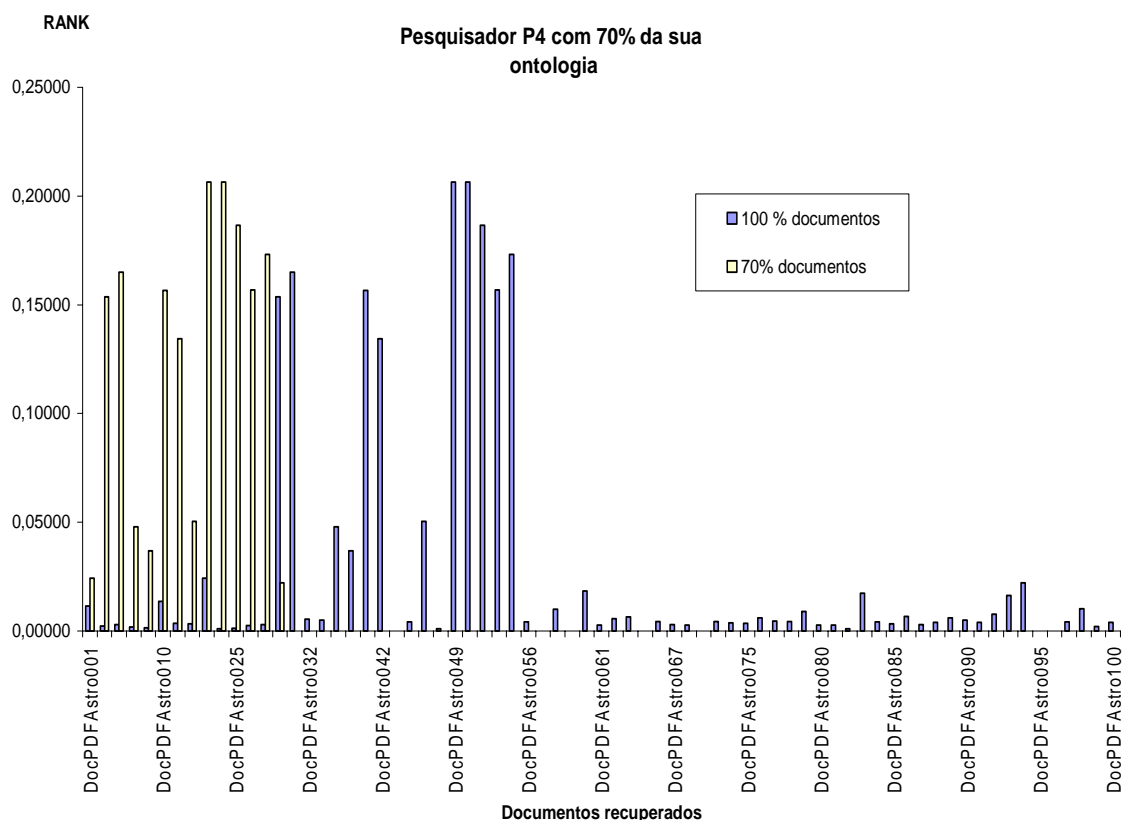


Gráfico 10 - Documentos recuperados para o pesquisador P4 para os pontos de corte alto e baixo e com 70% da sua ontologia.

O gráfico 10 apresenta para 70% da ontologia e 70% dos documentos o conjunto de documentos {2,5,10,23,24,25,26 e 28} e exceto o documento 28, todos possuem similaridade com o pesquisador para os termo "*nebula*", o que está de acordo com o seu perfil. Para 100% dos documentos o perfil do pesquisador se aproxima dos pesquisadores P1, P2 e P3, fato esse que se justifica pois possuem um conjunto de termos em comum (área de pesquisa em comum). O conjunto retornado é {29,30,41,49,50,51,52 e 53} onde o mesmo é um dos autores em todos os documentos.

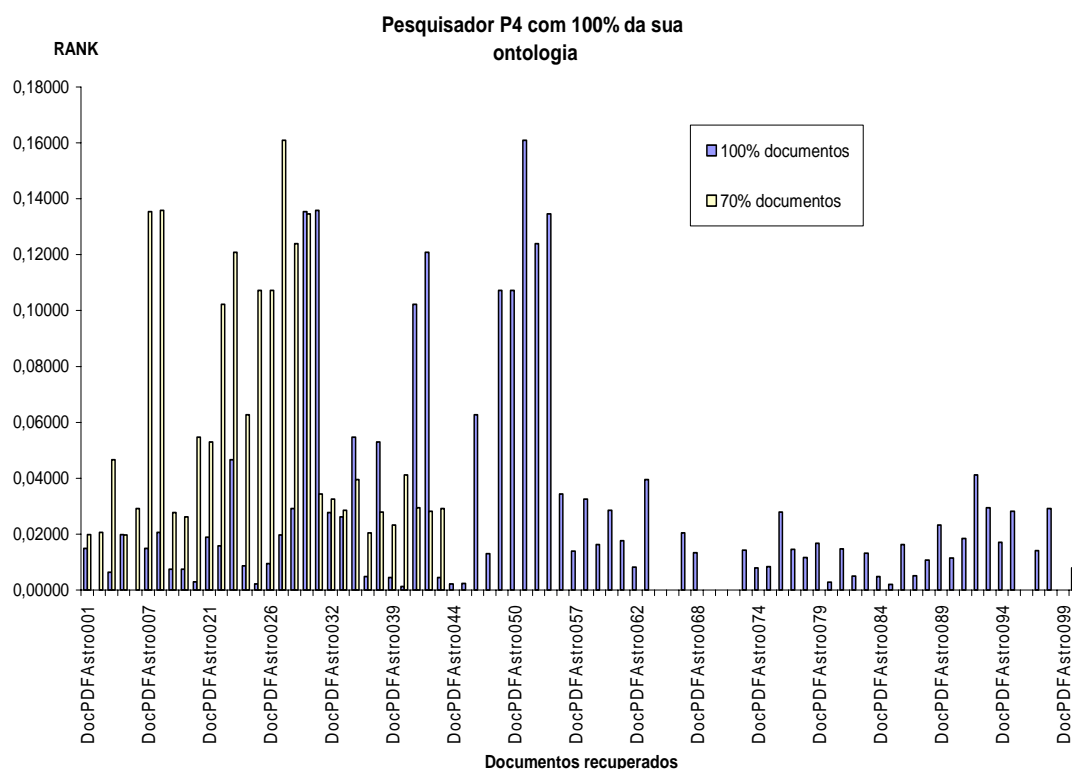


Gráfico 11- Documentos recuperados para o pesquisador P4 para os pontos de corte alto e baixo e com 100% da sua ontologia.

O gráfico 11 apresenta resultados semelhantes aos do gráfico 10 para 100% dos documentos e para 70 % dos documentos. Ressalta-se o conjunto {7,8,23 e 27}, todos documentos adequados ao perfil do pesquisador. No conjunto final de documentos aparece pelo menos um termo da sua ontologia, particularmente o documento 92.

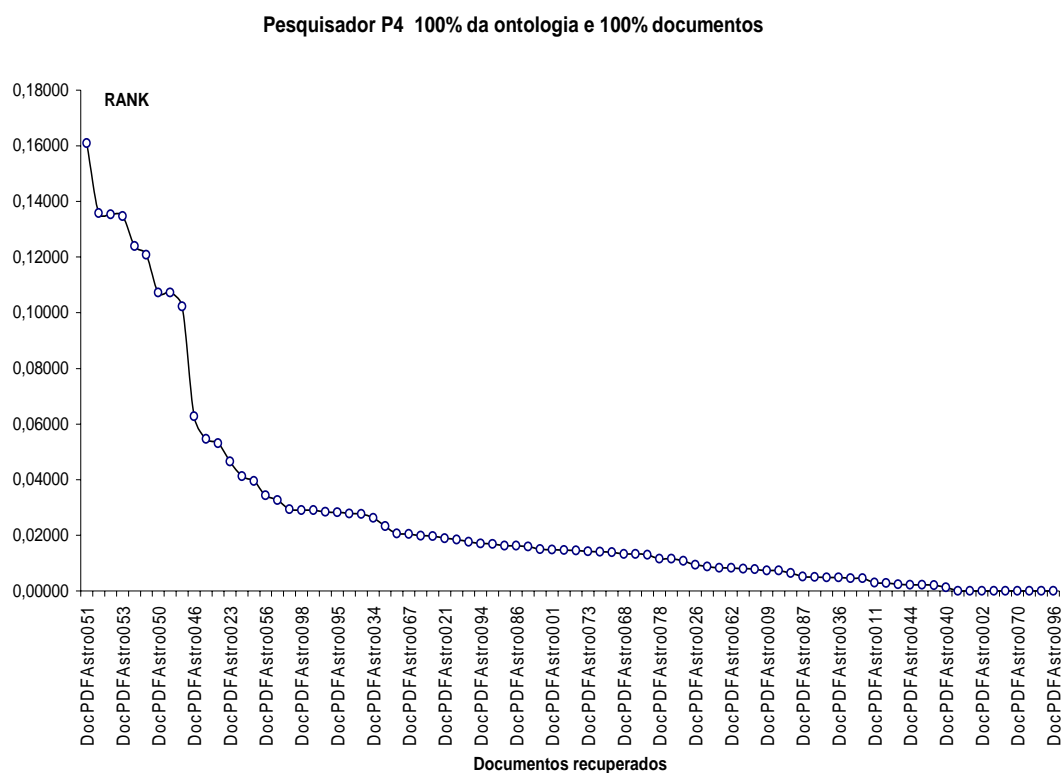


Gráfico 12- Documentos recuperados para a o pesquisador P4 para os pontos de corte baixo.

O gráfico 12 apresenta em ordem decrescente de “*ranking*” os documentos retornados pela ferramenta indicando que os documentos mais representativos em relação ao seu perfil, são aqueles em que o mesmo é um dos autores , no caso em 80% dos documentos recuperados é autor e 20% não são de sua autoria, mas todos representando o perfil definido. É um resultado que indica que o perfil do pesquisador é similar ao conjunto de documentos que formam a base de teste e sua ontologia melhor representa os documentos retornados .

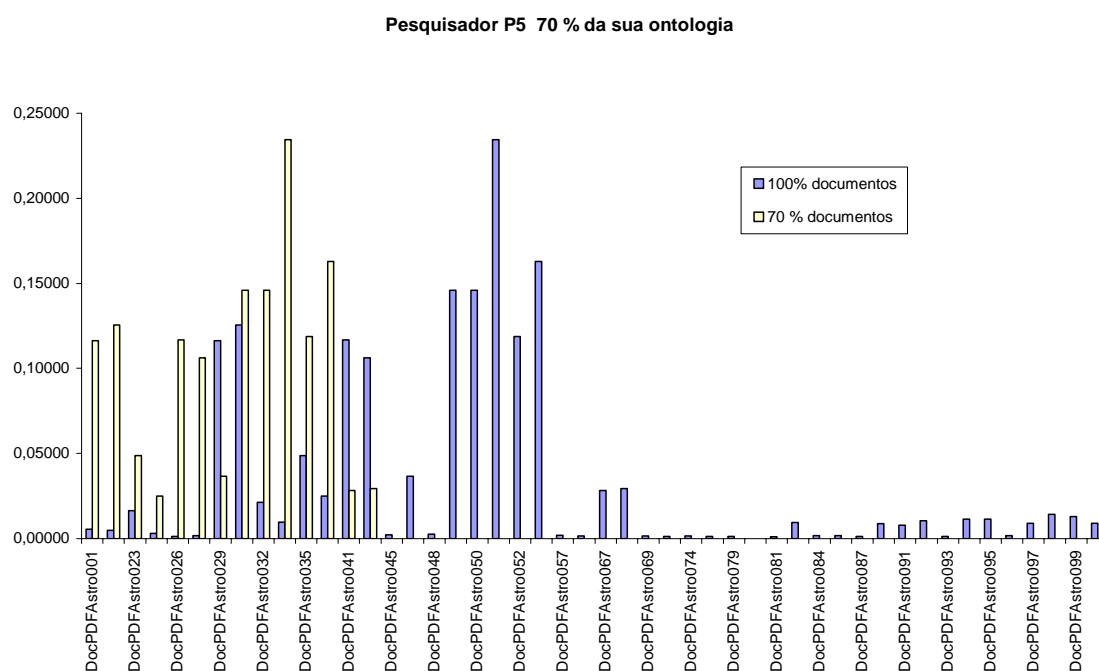


Gráfico 13 - Documentos recuperados para o pesquisador P5 para os pontos de corte alto e baixo e com 70% da sua ontologia.

O gráfico 13 apresenta para 70 % da ontologia os documentos recuperados e, observa-se que, para 70% dos documentos o conjunto de documentos retornado são {8,30,32,34 e 38} onde somente no documentos 34 o pesquisador é autor e possui relevância com a sua ontologia. Na situação de 100% dos documentos, o conjunto é formado pelos documentos {9,39,41,42,49,51,52 e 53}, onde o documento 53 possui o maior número de termos da ontologia e também onde o pesquisador é autor. Todos os documentos são de autoria do grupo com excessão do documento 52.

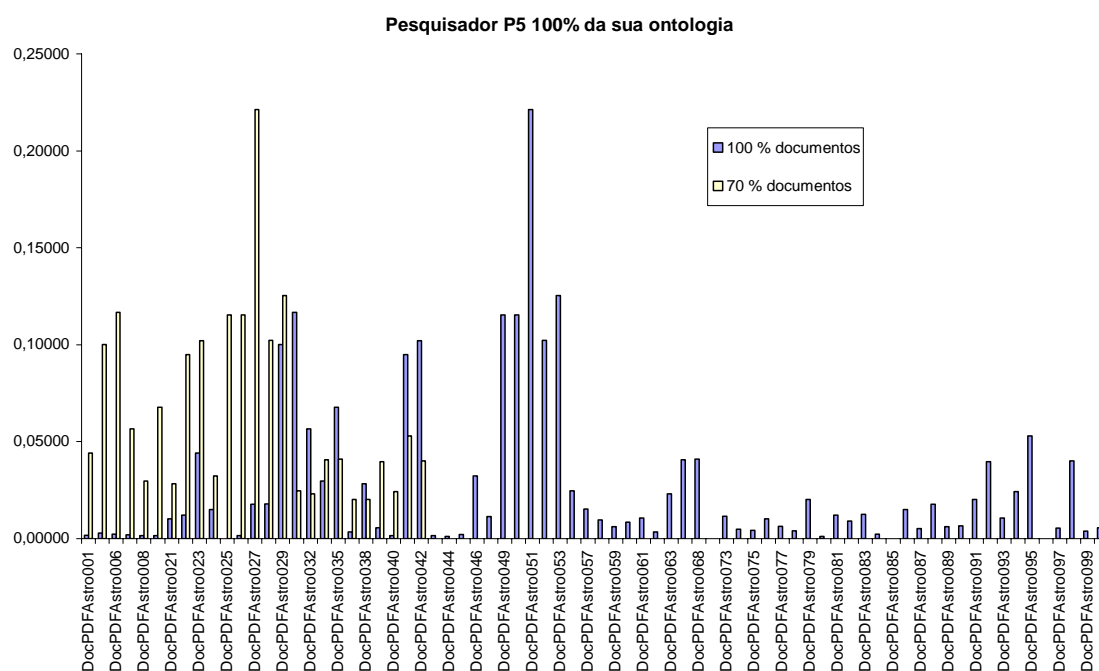


Gráfico 14- Documentos recuperados para o pesquisador P5 para os pontos de corte alto e baixo e com 100% da sua ontologia.

O gráfico14 apresenta para 100% da ontologia um comportamento semelhante ao gráfico 13 em 100% dos documentos retornados e para 70% dos documentos retornados o documento 34 aparece de forma destacada, como já mencionado, o pesquisador é um dos autores. A mesma tendência para os documentos finais ocorre de forma semelhante ao pesquisador P4, pois os mesmos possuem os termos “*solar flare*” e “*magnetohydrodynamic*” nas suas ontologias.

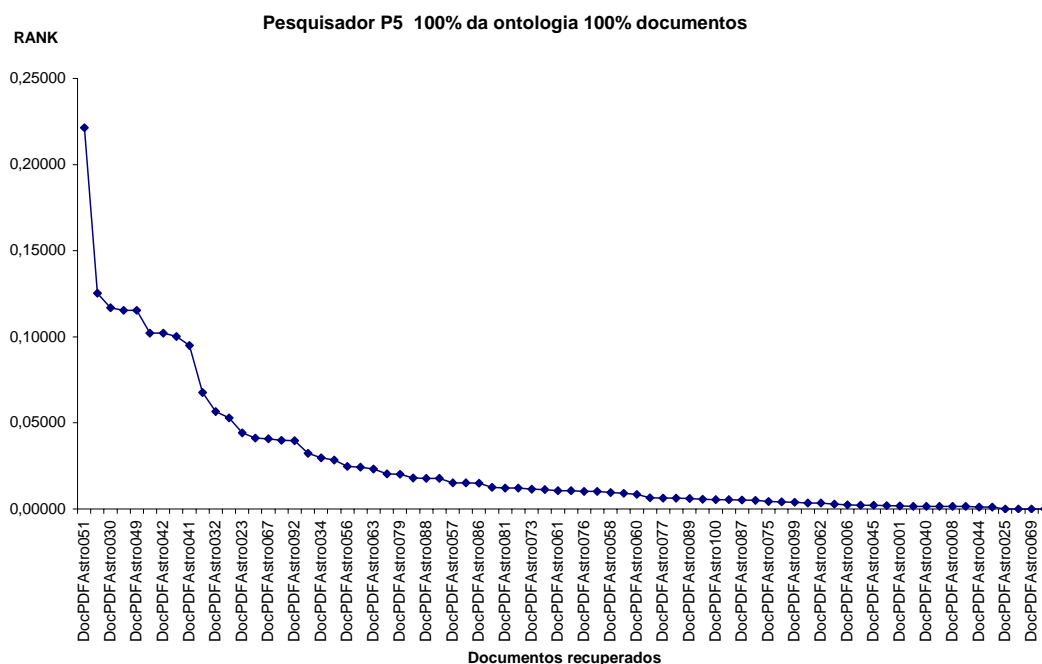


Gráfico 15- Documentos recuperados para a o pesquisador P6 para os pontos de corte baixo.

O gráfico 15 apresenta em ordem decrescente de “*ranking*” os documentos retornados pela ferramenta, indicando que os documentos mais representativos em relação ao seu perfil, são aqueles em que o mesmo é um dos autores , no caso 80% dos documentos recuperados é autor e 20% não são de sua autoria, mas todos representam o perfil definido. É um resultado que indica que o perfil do pesquisador é similar ao conjunto de documentos que formam a base de teste e sua ontologia melhor representa os documentos retornados .

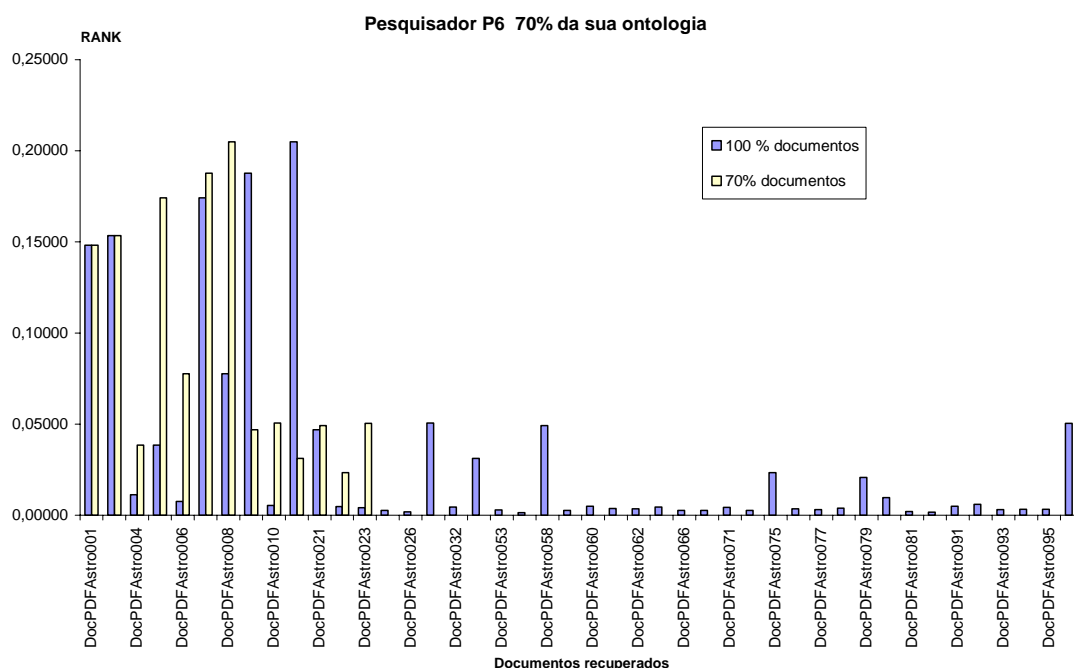


Gráfico 16 - Documentos recuperados para o pesquisador P6 para os pontos de corte alto e baixo e com 70% da sua ontologia.

O gráfico 16 apresenta para 70% da ontologia e 70% dos documentos retornados o conjunto { 1,8,30,32,34,35 e 38}, os quais são representativos com a ontologia do autor, de forma significativa para o 38 pois embora não seja autor, foi elaborado por membros do grupo de pesquisa. Analogamente para 100% dos documentos retornados {28,58,75,79 e 96} são correlatos ao pesquisador e os documentos 28 , 58 e 96 tratam da área específica representada pelo termo “*extragalactic*”.

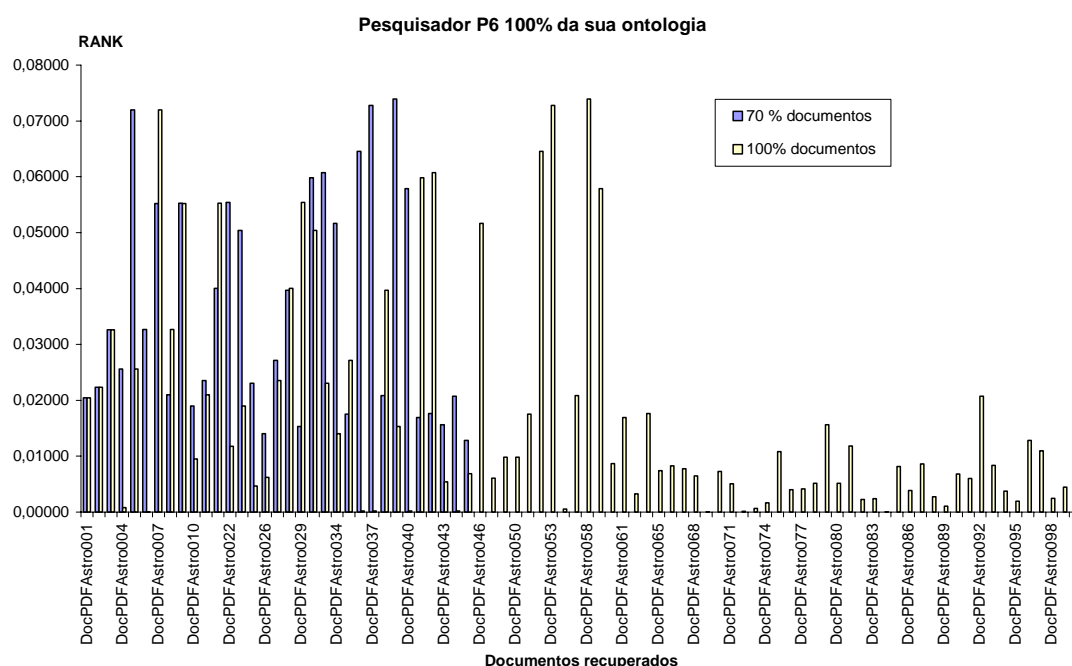


Gráfico 17- Documentos recuperados para o pesquisador P6 para os pontos de corte alto e baixo e com 100% da sua ontologia.

O gráfico 17 indica que para 100% da ontologia e 70% dos documentos os documentos retornados são relativos à área representada pelos termos “*extragalactic*”, “*quasars*” e que para 100% dos documentos retornados os documentos { 52,53,58 e 59 } correspondem totalmente ao seu perfil; os documentos 52 e 53 foram elaborados por membros do grupo de teste.

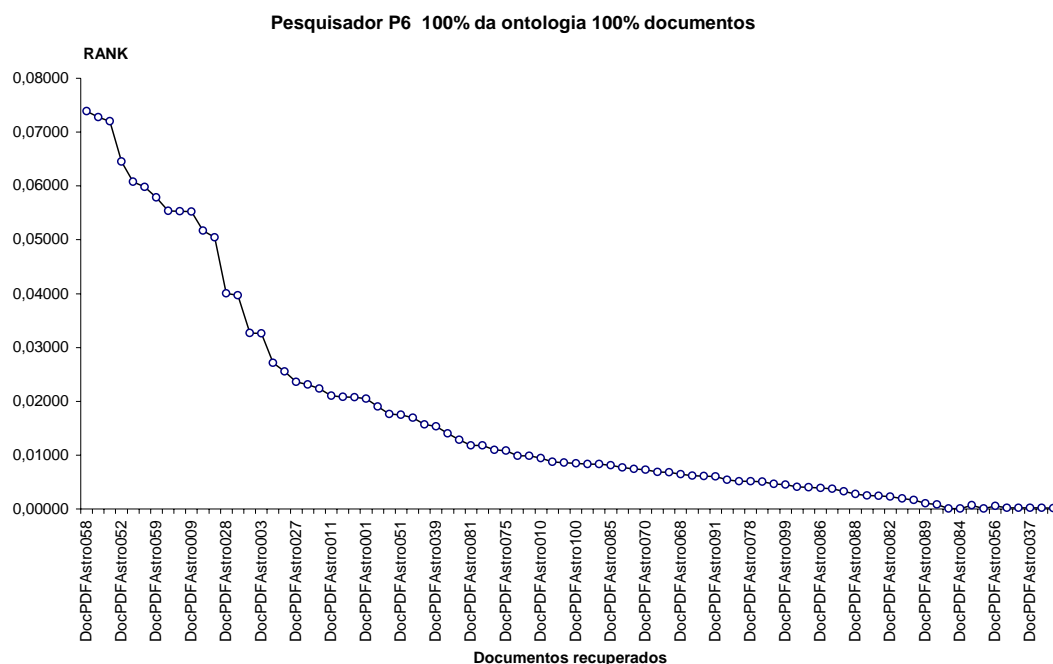


Gráfico 18 - Documentos recuperados para a o pesquisador P6 para os pontos de corte baixo.

O gráfico 18 apresenta em ordem decrescente de “*ranking*” os documentos retornados pela ferramenta e os documentos mais representativos, em relação ao seu perfil, são aqueles em que o mesmo é um dos autores, no caso 84% dos documentos recuperados é autor e 16% não são de sua autoria, mas todos representando o perfil definido. É um resultado que indica que o perfil do pesquisador é similar ao conjunto de documentos que formam a base de teste e sua ontologia melhor representa os documentos retornados.

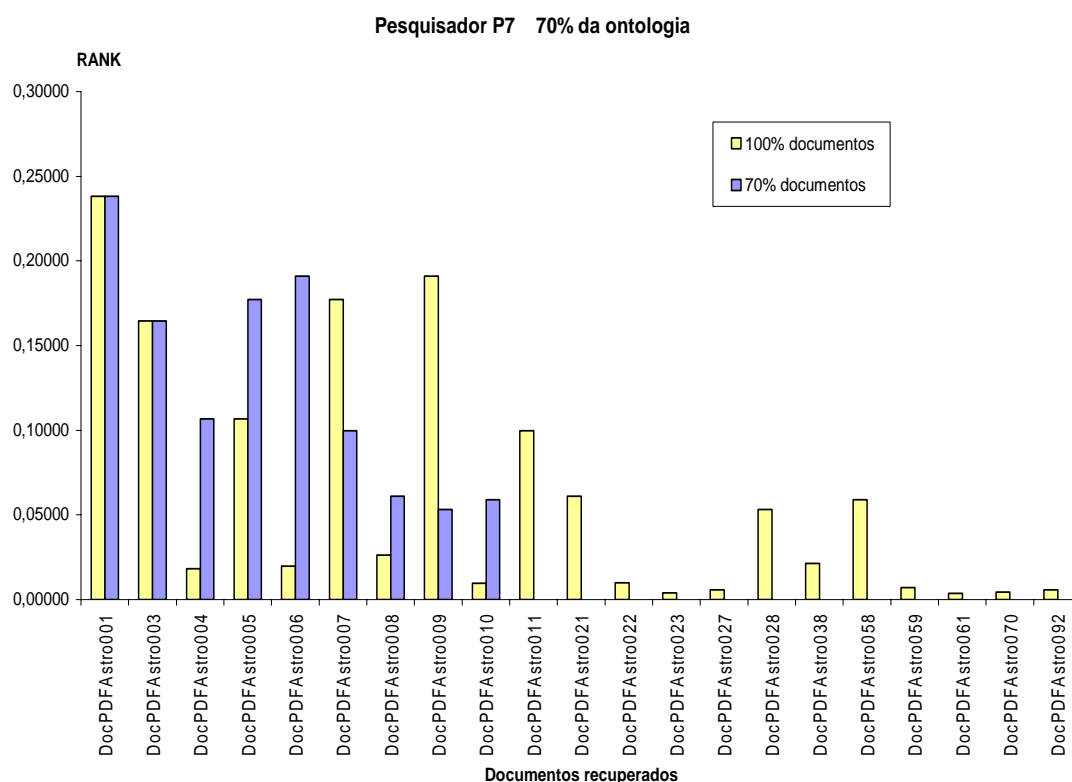


Gráfico 19 - Documentos recuperados para o pesquisador P7 para os pontos de corte alto e baixo e com 70% da sua ontologia.

O gráfico 19 apresenta para 70 % da ontologia e 70% dos documentos recuperados os documentos compatíveis com o perfil do pesquisador, representado pelos termos “*quasars*” , “*quasar*” , “*galaxies*” e “*geodetic*”, que representam a faixa de documentos 1 a 10. Para 100% dos documentos retornados os documentos continuam ainda nessa faixa e os de número 28 e 53 possuem compatibilidade com esses termos.

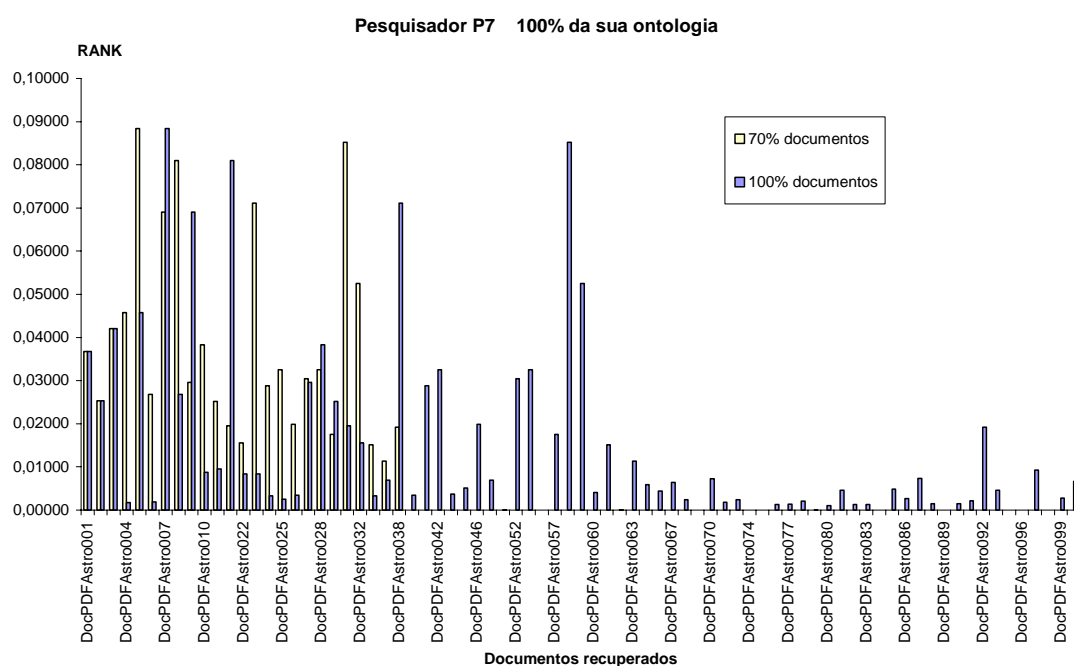


Gráfico 20 - Documentos recuperados para o pesquisador P7 para os pontos de corte alto e baixo e com 100% da sua ontologia.

O gráfico 20 apresenta o mesmo comportamento do gráfico 19, porém com um maior retorno na faixa de documentos de 1 a 40 e para o caso de 70% de documentos recuperados. Para 100% dos documentos recuperados o comportamento permanece semelhante aos pesquisadores P1, P2, P3, P4 e P5, na faixa de 1 a 40 e de 40 a 60. Nestes intervalos foram recuperados os mesmos documentos, pois os pesquisadores possuem ontologias semelhantes.

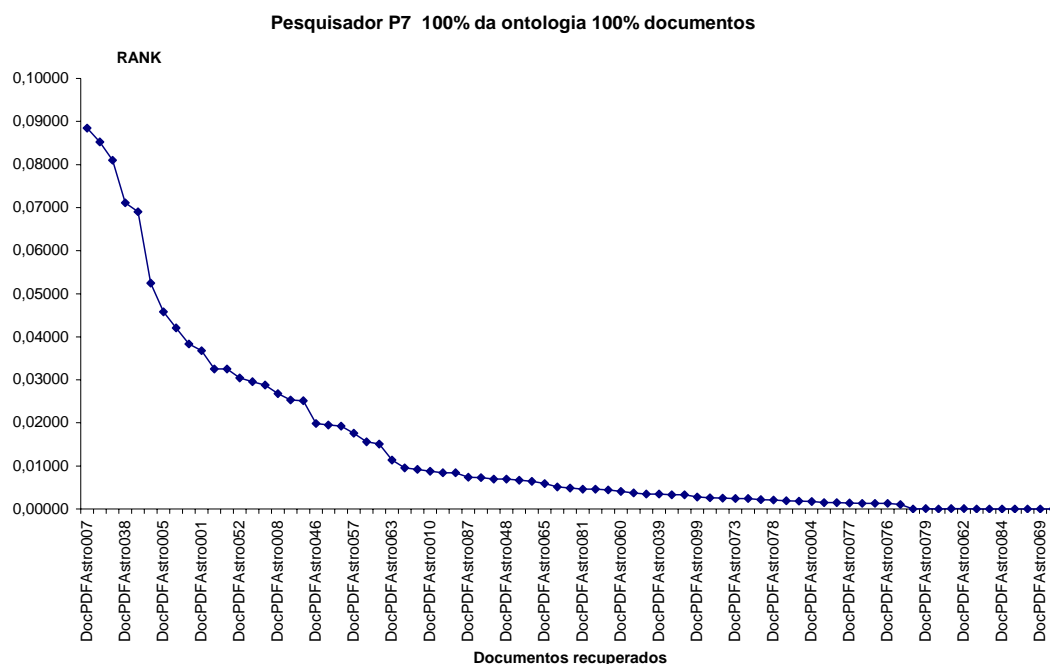


Gráfico 21- Documentos recuperados para a o pesquisador P7 para os pontos de corte baixo.

O gráfico 21 apresenta em ordem decrescente de “*ranking*” os documentos retornados pela ferramenta e indica que os documentos mais representativos em relação ao seu perfil são aqueles em que o mesmo é um dos autores , no caso 60% dos documentos recuperados é autor e 40% não são de sua autoria, mas todos representando o perfil definido. É um resultado que indica que o perfil do pesquisador é similar ao conjunto de documentos que formam a base de teste e sua ontologia melhor representa os documentos retornados .

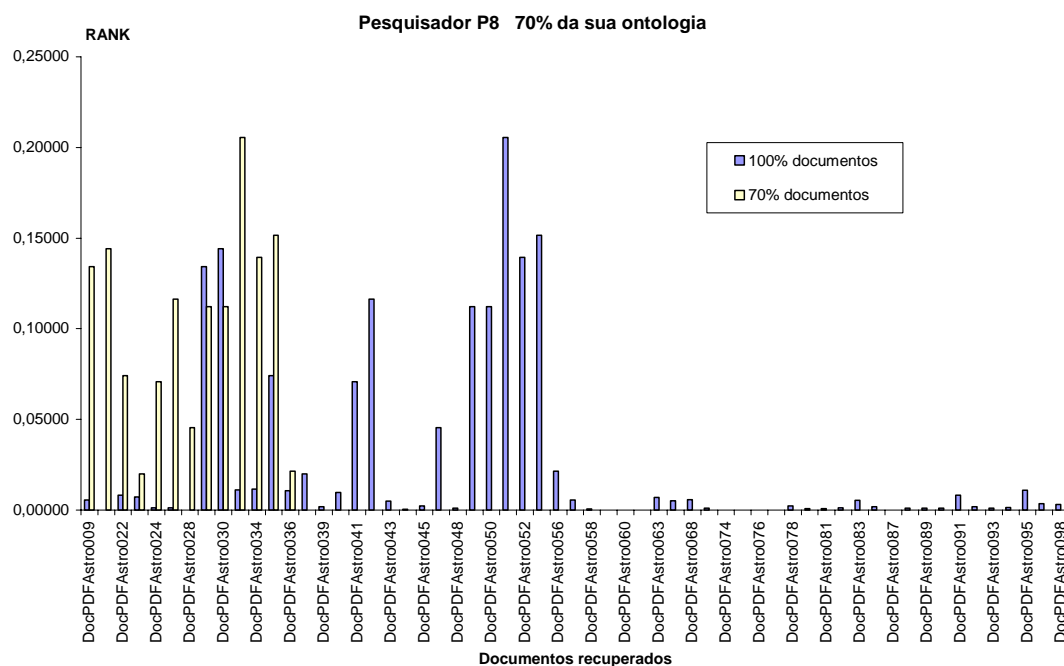


Gráfico 22- Documentos recuperados para o pesquisador P8 para os pontos de corte alto e baixo e com 70% da sua ontologia.

O gráfico 22 apresenta para 70% da sua ontologia e para 70% dos documentos comportamento análogo ao do pesquisador P6 e P7, pois seu perfil também é semelhante a estes nos termos “*quasars*”, “*quasar*”, “*galaxies*” e “*geodetic*”. Para 100% dos documentos o seu comportamento se assemelha ao dos pesquisadores P1 , P2, P3 , P4 e P5, visto que ele também é autor nesses documentos .

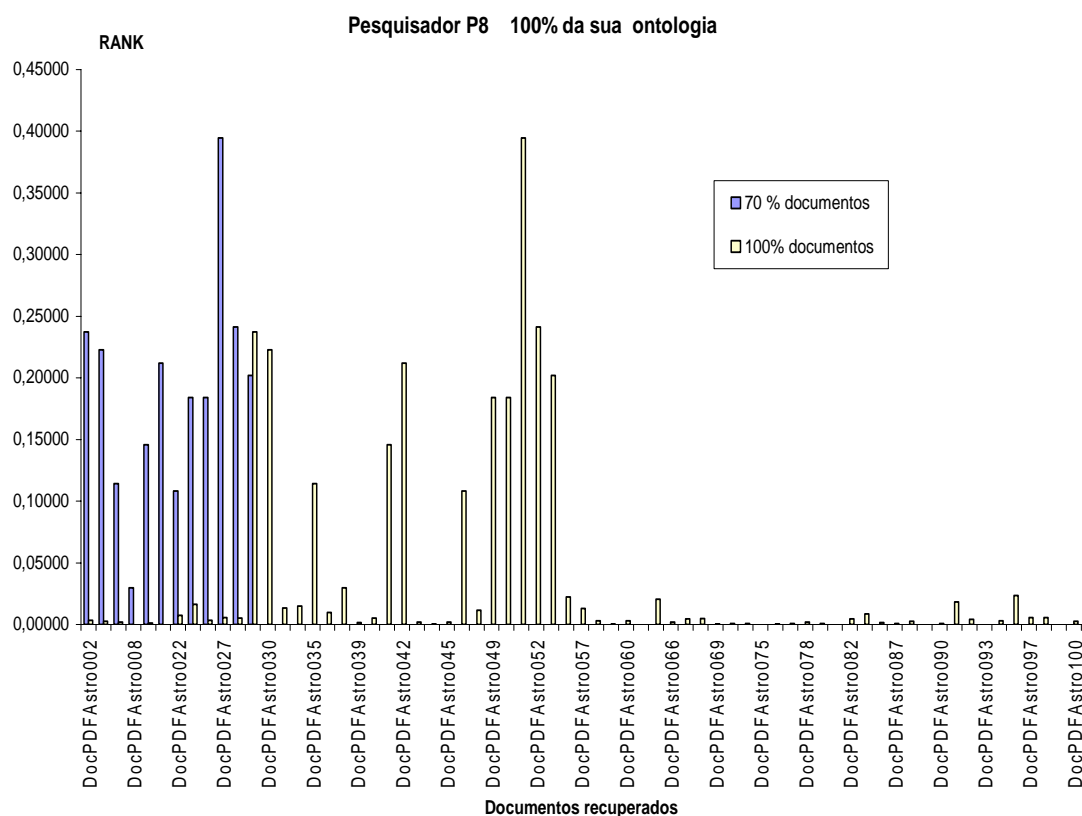


Gráfico 23 - Documentos recuperados para o pesquisador P8 para os pontos de corte alto e baixo e com 100% da sua ontologia.

O gráfico 23 apresenta os documentos recuperados de forma semelhante e com maior valor de “*ranking*” do que no gráfico 22. Isso indica que a representação ontológica do autor é suficientemente representativa para o perfil do autor e foi talvez melhor elaborada.

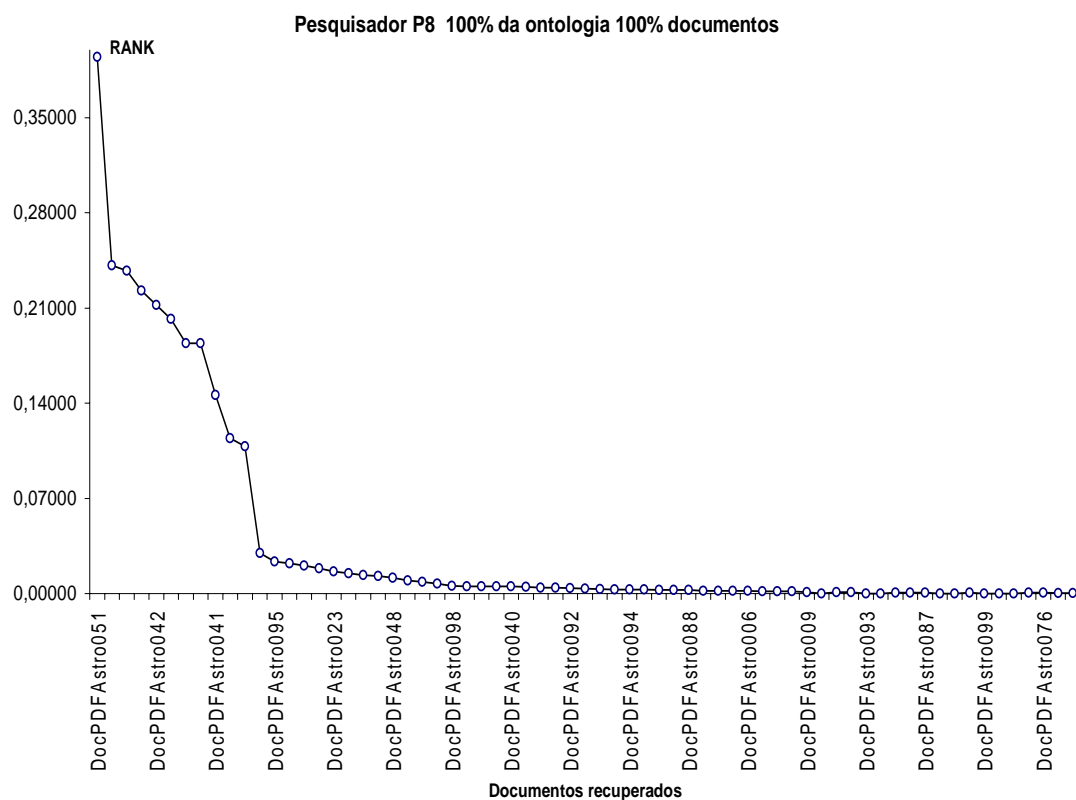


Gráfico 24 - Documentos recuperados para a o pesquisador P8 para os pontos de corte baixo.

O gráfico 24 apresenta em ordem decrescente de “*ranking*” os documentos retornados pela ferramenta, que indicam que os documentos mais representativos, em relação ao seu perfil, são aqueles em que o mesmo é um dos autores, no caso 70% dos documentos recuperados é autor e 30% não são de sua autoria, mas todos representam o perfil definido. É um resultado que indica que o perfil do pesquisador é similar ao conjunto de documentos que formam a base de teste e sua ontologia melhor representa os documentos retornados, e destaca que o mesmo possui duas linhas claramente distintas de pesquisa.

4.9.2 Grupos de Pesquisa por Afinidade de Tema e Ontologia

Pela análise dos dados do item 4.9.1 pôde-se observar a existência de dois sub-grupos de pesquisa no grupo de teste. O primeiro é o sub-grupo de pesquisa A, com os membros: P1,P2,P3,P4,P5 e P8; e, o segundo é o sub-grupo de pesquisa B, com os membros: P6 , P4 e P7.

Os gráficos 25 e 26 apresentam os perfis com “*ranking*” similares, obtidos por análise dos gráfico anteriormente discutidos. Esse gráficos foram elaborados com 100% das ontologias e 100% dos documentos recuperados.

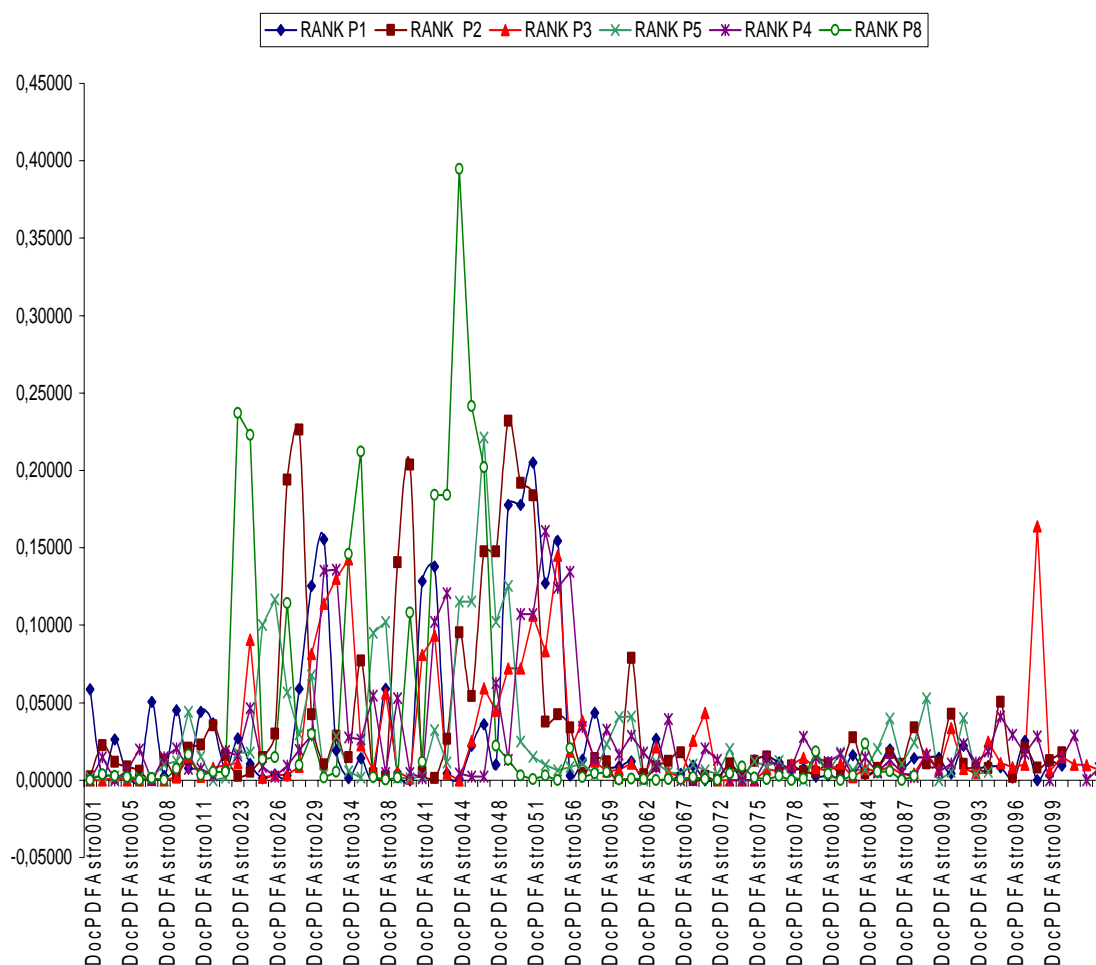


Gráfico 25 - Documentos recuperados para o sub-grupo A.

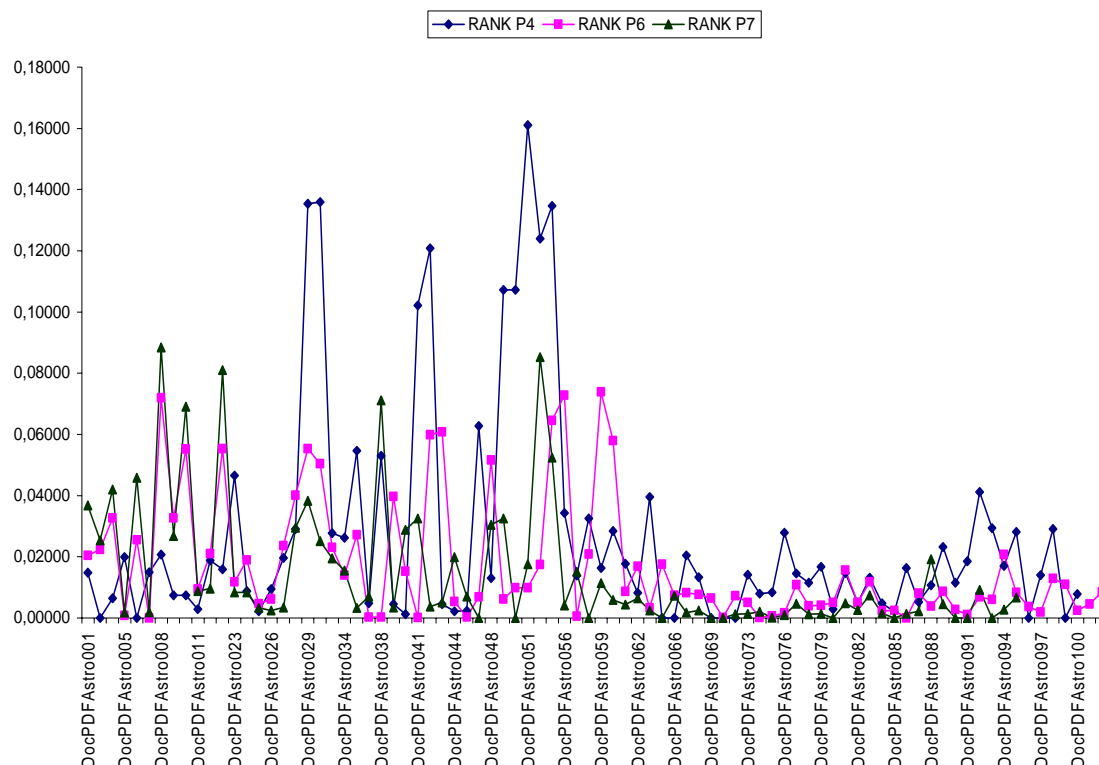


Gráfico 26 - Documentos recuperados para o sub-grupo B.

O gráfico 25 apresenta a correspondência de documentos e perfis dos pesquisadores do sub-grupo A e mostra a similaridade ontológica de seus membros, uma vez que os termos “*submillimeter*” e “*SST*” caracterizam esses pesquisadores. Destaca-se uma concentração de documentos situados na faixa de 40 a 60; documentos estes que foram elaborados por membros do sub-grupo assim como outros documentos de igual relevância para cada ontologia. A maioria desses documentos trata de estudos onde termos referentes ao tema Física Solar aparecem pelo menos uma vez.

O gráfico 26 apresenta de forma semelhante o perfil comum de pesquisadores do sub-grupo B, destacando-se que os mesmos mantêm compatibilidade com a linha geral do sub-grupo A, mas especificamente os documentos iniciais, que tratam sobre

os temas: Astrofísica do Meio Interestelar, Quasares e Astrofísica Extragalática, tem mais relevância aos seus perfis ontológicos.

Tabela 9 - Documentos recuperados por pesquisador e seu perfil ontológico.

	<i>100% Ontologia</i>		<i>70% Ontologia</i>	
	<i>100% doc</i>	<i>70% doc</i>	<i>100% doc</i>	<i>70% doc</i>
P1	80	24	74	15
P2	80	25	74	16
P3	83	25	74	16
P4	71	30	71	14
P5	48	25	48	14
P6	84	33	44	13
P7	77	25	21	9
P8	67	12	59	13
Média de Recuperação (Doc/Pesq)	73,75	24,875	58,125	13,75

A tabela 9 apresenta o número de documentos retornados pela ferramenta, por pesquisador, nos quatro ensaios que cada um teve. Como a base de documentos é composta por cem documentos, os resultados podem ser interpretados como dados percentuais. Dessa forma, observa-se que a taxa de recuperação média de documentos obtida pela ferramenta para os pesquisadores P1 a P3 é de 80 % com ontologia completa e 74% com ontologia 70%, ou seja, praticamente iguais, devido aos perfis de ontologia similares e de ampla cobertura nesta área do conhecimento. Para os pesquisadores P4, P5 e P8, que possuem ontologias mais específicas em relação à área do conhecimento, também foram verificadas taxas próximas de recuperação com ambas as ontologias testadas, porém, a ferramenta recuperou menor quantidade de documentos mas mais relacionados com suas ontologias. Para os pesquisadores P6 e P7, que possuem a ontologia mais específica do grupo e com maior similaridade à amostra de documentos, a ferramenta apresentou uma maior

taxa de recuperação com ontologia 100% do que com ontologia 70%, mas houve maior precisão com ontologia 70%, ao recuperar documentos relevantes aos perfis desses pesquisadores.

Capítulo

5

Conclusão

5.1 Conclusões

A representação do conhecimento é uma tarefa que demanda esforço e o uso das ontologias auxilia essa atividade. A Ontologia Fuzzy é um método que auxilia na elaboração da representação do conhecimento, porém, os resultados devem ser sempre acompanhados de uma análise humana. As informações manipuladas por grupos de pesquisa, assim como os perfis dos pesquisadores, podem ser representados por ontologias fuzzy, que combinadas com técnicas de recuperação de informação clássicas, tais como o modelo vetorial que foi utilizado nesta ferramenta, contribuiu para tornar a busca da informação mais representativa em relação ao perfil do pesquisador. Existem várias maneiras de se obter a similaridade em uma ontologia e neste trabalho, o uso do algoritmo de similaridade fuzzy mostrou-se eficiente para o domínio estudado. A dificuldade de uma abordagem automática e que contemple todos os conceitos de uma ontologia, exige em algum momento a

intervenção do usuário, ou no momento do cadastro ou ao longo do uso da ferramenta.

Na ferramenta implementada, foi possível detectar informações relevantes, mesmo com um número pequeno de termos representativos do perfil do pesquisador ou quando outros documentos sejam inseridos por outros pesquisadores. A precisão dependerá de como o pesquisador cadastrará sua ontologia no sistema. Os resultados obtidos apresentam um caráter de relevância subjetivo, mas que serviu ao propósito de verificar o módulo ontológico. Esse módulo mostrou-se eficiente na seleção de termos que refinam a pesquisa e, portanto aumentam as chances de recuperar documentos mais relevantes para um dado perfil.

5.2 Contribuições

A principal contribuição desse trabalho foi a proposta de representação do conhecimento de grupos de pesquisa por meio da Ontologia-Fuzzy. Essa técnica permite uma caracterização semi-automática dos termos representativos de um dado domínio considerando-se um grau de incerteza. Essa medida associada à uma técnica de cálculo de similaridade, auxilia na seleção de termos mais representativos da ontologia.

5.3 Perspectivas para trabalhos futuros

São sugestões para trabalhos futuros:

- Construir um módulo de geração gráfica dos documentos recuperados e seu *“ranking”*;
- Desenvolver um módulo para extração totalmente automática das informações da Plataforma Lattes dos pesquisadores;
- Estudar com a ontologia proposta outros grupos de pesquisa, de outras áreas do conhecimento;
- Implementar ao modelo ontológico fuzzy um módulo de aprendizado dinâmico de acordo com o perfil do usuário.

Referências Bibliográficas

AIRES, Rachel Virgínia Xavier; SANTOS, Diana. Measuring the Web in Portuguese. In: EUROWEB 2002 CONFERENCE, 2002. Oxford, UK, 17-18 December 2002. p.198-199. Disponível em: <<http://www.linguatca.pt/documentos/>>. Acesso em: 20 Out. 2006.

ALVES, Daniel Arndt. *Sistema de Gerenciamento Eletrônico de Documentos: uma ferramenta para otimizar a produtividade do conhecimento*. 2005. 189f. Trabalho de Graduação Interdisciplinar (Graduação em Ciência da Computação) - Universidade Presbiteriana Mackenzie, São Paulo, 2005.

ANDRADE, Marcos Vinicius Mendonça. Gerenciamento Eletrônico da Informação: Ferramenta para a Gerência Eficiente dos Processos de Trabalho. In: XIII SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS. 2002. Recife, 2002. Disponível em: <<http://www.ndc.uff.br/textostecnicos.asp>>. Acesso em: 18 Out. 2005.

BAEZA-YATES, R. e RIBEIRO-NETO, B. *Modern Information Retrieval*. Addison Wesley (EUA): ACM Press, 1999.

BONIFÁCIO, Ailton Sergio. *Ontologias e Consulta Semântica: Uma Aplicação ao Caso Lattes*. 2002. 85f. Dissertação (Mestrado em Ciência da Computação) - UFRGS, Porto Alegre, 2002.

BORGES, M.R.S.; CAVALCANTI, M.C.R.; CAMPOS, M.L.M. Suporte por Computador ao Trabalho Cooperativo. In: XV CONGRESSO DA SOCIEDADE BRASILEIRA DE COMPUTAÇÃO. 1995. CANELA, RS, 1995.

BULSKOV, H.; KNAPPE, R.; ANDREASEN, T. On Measuring Similarity for Conceptual Querying. In: PROCEEDINGS OF THE FLEXIBLE QUERY ANSWERING SYSTEMS 5th INTERNATIONAL CONFERENCE, FQAS 2002. T. Andreasen; A. Motro; H. Christiansen; H.L. Larsen, (Eds.). Copenhagen, Denmark, October 27-29, p. 100-111, 2002.

CENADEM (Centro Nacional do Desenvolvimento do Gerenciamento da Informação). *O portal GED (Gerenciamento Eletrônico de Documentos) do Brasil*. Disponível em: <<http://www.cenadem.com.br>>. Acesso em: 24 Mar. 2004.

CNPq (2006). *Plataforma Lattes*. Disponível em: <<http://lattes.cnpq.br/pl/>>. Acesso em 20 Jan. 2006.

CORDÌ, V.; LOMBARDI, P.; MARTELLI, M.; MASCARDI, V. An Ontology-Based Similarity between Sets of Concepts. In: PROCEEDINGS OF WOA 2005. F. Corradini, F. De Paoli, E. Merelli and A. Omicini (Eds.). Bologna: Pitagora Editrice, ISBN 88-371-1590-3, pag. 16-21, 2005. Disponível em: <http://www.disi.unige.it/person/CordiV/Research.php#Pubblications>. Acesso em: 22 Jan. 2007.

DAVENPORT, Thomas; PRUSAK, Laurence. *Conhecimento empresarial*. Rio de Janeiro: Campus, 1999.

DAVENPORT, Thomas; PRUSAK, Laurence. *Conhecimento Empresarial: como as organizações gerenciam o seu capital intelectual*. Tradução de Lenke Peres. 7. ed. Rio de Janeiro: Campus, 1998.

DEITEL, H. M. e DEITEL, P. J. *JAVA™ como programar*. 4ª. ed. Porto Alegre: ookman, 2003. 1386p.

DIA gtk + based diagram creation program. Disponível em: <<http://www.gnome.org/projects/dia/>> . Acesso em: 30 Jan. 2006.

ELMASRI, R e NAVATHE S.B. *Sistemas de Banco de Dados*. 4ª ed. São Paulo: Pearson, 2005. 723p.

FERNEDA, Edberto. *Recuperação da Informação: Análise da contribuição da Ciência da Computação para Ciência da Informação*. 2003. 147f. Tese (Doutorado em Ciências da Comunicação) - Universidade de São Paulo, São Paulo, 2003.

GENESERETH, M. R.; NILSSON, N. J. *Logical Foundation of Artificial Intelligence*. Los Altos, CA: Morgan Kaufmann, 1987.

GRUBER, T.R. *A Translation Approach to Portable Ontology Specification*. Technical Report KSL 92-71, Computer Science Department - Stanford University. 1993. Disponível em: <http://ksl-web.stanford.edu/KSL_Abstracts/KSL-92-71.html>. Acesso em: 3 Maio de 2006.

GRUDIN, J. CSCW: History And Focus. *IEEE Computer*. n.5, p.19-26, mai.1994.

GUARINO, N. Formal Ontology and Information Systems. *IOS Press*, Amsterdam, Netherlands, p. 3-15, 1998.

HARMAN, D. Overview of the Third Text Retrieval Conference (TREC-3). 1993. Disponível em: <http://trec.nist.gov/pubs/trec3/t3_proceedings.html> . Acesso em : 12 Dez. 2006.

HUBERMAN, B. A., PIROLI, P. L. T., PITKOW, J. E. e LUKOSE, R. M. Strong regularities in world wide web surfing, *Science*, v.280, n.5360, p. 95-97, 1998.

JONES, Steven G. (org). *Virtual Culture: Identity & Communication in Cybersociety*. Thousand Oaks, California: SagePublications, 1997.

KURNIAWAN, B. *JAVA para Web com Servlets, JSP e EJB*. Rio de Janeiro:Editora Ciência Moderna Ltda., 2002. 807p.

LEE, J. H. *Analyzing the effectiveness of extended Boolean models in information retrieval*. Technical Report TR95-1501, Cornell University. 1995. Disponível em: <<http://cs-tr.cs.cornell.edu/>>.

LEMOIS, André. *Agregações Eletrônicas ou Comunidades Virtuais? Análise das listas Facom e Cibercultura*. [S.l.], 2002. Disponível em: <<http://www.facom.ufba.br/ciberpesquisa/agregacao.htm>>. Acesso em: 16 Dez. 2005.

LINCHTNOW, D; WANGENHEIN, C.G.V; WANGENHEIN, A.V. Uma Abordagem para a Gerência do Conhecimento em Grupos de Pesquisa. In: ANAIS DO CONGRESSO BRASILEIRO DE COMPUTAÇÃO - CBCOMP, 2001. Itajaí, 2001.

LYMAN, P.; VARIAN, H. R.; DUNN, J.; STRYGIN, A.; SWEARINGEN, K. 2000. *How Much Information?* Disponível em: <<http://www.sims.berkeley.edu/research/projects/how-much-info/how-much-info.pdf>>. Acesso em: 15 out. 2006.

MACAULAY, L. *Human-Computer Interaction for Software Designers*. International Thomson Computer Press, 1995.

MARON, M. E. e KUHNS, J. L. On relevance, probabilistic indexing, and information retrieval. *Journal of the ACM*, 7(3), p.216-244, 71, Julho 1960.

MIDDLETON, S. E; DE ROURE, D. C.; SHADBOLT, N. R. (2001) Capturing knowledge of user preferences: ontologies in recommender systems. In: PROCEEDINGS FIRST INTERNATIONAL CONFERENCE ON KNOWLEDGE CAPTURE, 2001. Victoria, British Columbia, Canada. 2001. p. 100-107.

MySQL Licensing Policy (2006). Disponível em: <<http://www.mysql.com/company/legal/licensing/>>. Acesso em: 6 Fev.2006.

NOVAK, J. D. *A Theory of Education*. Ithaca, NY: Cornell University Press, 1977.

PAICE, C. P. Soft evaluation of boolean search queries in information retrieval systems. *Information Technology: Research and Development*, 3 (1), p.33-42., 1984.

PEREIRA, Rachel Carlos. *Modelo Ontológico relacional Fuzzy em sistemas de recuperação textual*. 2004. 82f. Dissertação (Mestrado em Engenharia Elétrica –Computação) – UNICAMP, São Paulo, 2004 .

PEZZA, André Bernardes. *Um Sistema Para Integração Disciplinar e Gestão Continuada da Aprendizagem*. 2004. 78f. Dissertação (Mestrado Engenharia Elétrica-Computação) - UNIVERSIDADE PRESBITERIANA MACKENZIE, São Paulo, 2004.

POLANYI. M. *The Tacit Dimension*. Londres: Routledgs & Kegan Paul, 1996.

PONTE, J. M.; CROFT, W. B. A language modeling approach to information retrieval. In: PROCEEDINGS OF THE 21ST ACM CONFERENCE ON RESEARCH AND DEVELOPMENT IN INFORMATION RETRIEVAL (SIGIR'98). 1998. Melbourne, Australia. 1998. p. 275-281.

RHEINGOLD, Howard. *The Virtual Community*. [S.l.], 1998. Disponível em: <<http://www.rheingold.com/vc/book/>>. Acesso em: 05 jan. 2006.

RIBEIRO, B. A. N.; MUNTZ, R. A belief network model for ir. In: PROCEEDINGS OF THE 19TH ACM CONFERENCE ON RESEARCH AND DEVELOPMENT IN INFORMATION RETRIEVAL (SIGIR'96), 1996. Zurich, Switzerland. 1996. p. 252-260.

RIJSBERGEN, C. J. VAN. *Information Retrieval*. Department of Computing Science. University of Glasgow. 1979. Disponível em: <<http://www.dcs.gla.ac.uk/Keith/Preface.html>>.

RIOS, Jocelma A. GED como Ferramenta na Gerência do Conhecimento Explícito Organizacional. In: VI CINFOM - ENCONTRO NACIONAL DE SISTEMAS DE INFORMAÇÃO, Salvador, BA, 2005. Disponível em: <http://www.cinform.ufba.br/vi_anais> . Acesso em: 16 Out. 2005.

SALTON, G. e MCGILL, M. J. (Eds.) *Introduction to Modern Information Retrieval*. McGraw-Hill, 1983.

SANTOS, N. e FERREIRA, H. M. C. Aprendizagem Cooperativa Distribuída na Biblioteca Kidlink-Brasil. *Revista Brasileira de Informática na Educação*, n. 02, pp. 35-42, abr. 1998.

SHAMSFARD, M.; NEMATZADEH, A.; MOTIEE, S. ORank: An Ontology Based system for Ranking Documents. *International Journal of Computer Science*. V.1, n.3, 2006.

SILBERSCHATZ, A., KORT, H.F., SUDARSHAN, S. *Sistema de Banco de Dados*. 3ª. Edição. Makron Books, 1999.

TAKEUCHI, Hirotaka. *Beyond Knowledge Management : Lessons from Japan*. 1998. Disponível em: <www.sveiby.com/articles/LessonsJapan.htm> . Acesso em: 19 Out. 2005.

TANEBAUM, A.S. *Redes de Computadores*. 3^a.ed Traduzida. Rio de Janeiro: Campus, 1997. 923p.

TERRA, J. *Gerência do conhecimento: o grande desafio empresarial*. 2. ed. São Paulo: Negócio Editora, 2001.

THE RHEINNER GROUP. Disponível em <<http://www.rheinner.com>>. Acesso em: 17 jul. 2006.

UNIVERSIDADE P. MACKENZIE - *Pós-Graduação Engenharia Elétrica*. online. Disponível em: <http://www.mackenzie.com.br/pos_graduacao/engenhariaeletrica/infraestrut.htm> . Acesso em: 10 Fev.2006.

WIDYANTORO, D.H; YEN, J. A Fuzzy Ontology-based Abstract Search Engine and Its User Studies. In PROCEEDINGS OF THE 10th IEEE INTERNATIONAL CONFERENCE ON FUZZY SYSTEMS, 2001. Melbourne, Australia, December 2-5, 2001. Disponível em: < <http://ist.psu.edu/yen/publications/index.html>> . Acesso em: 10 Jan. 2007.

ANEXO A

Banco de Dados da Ferramenta EDMIS

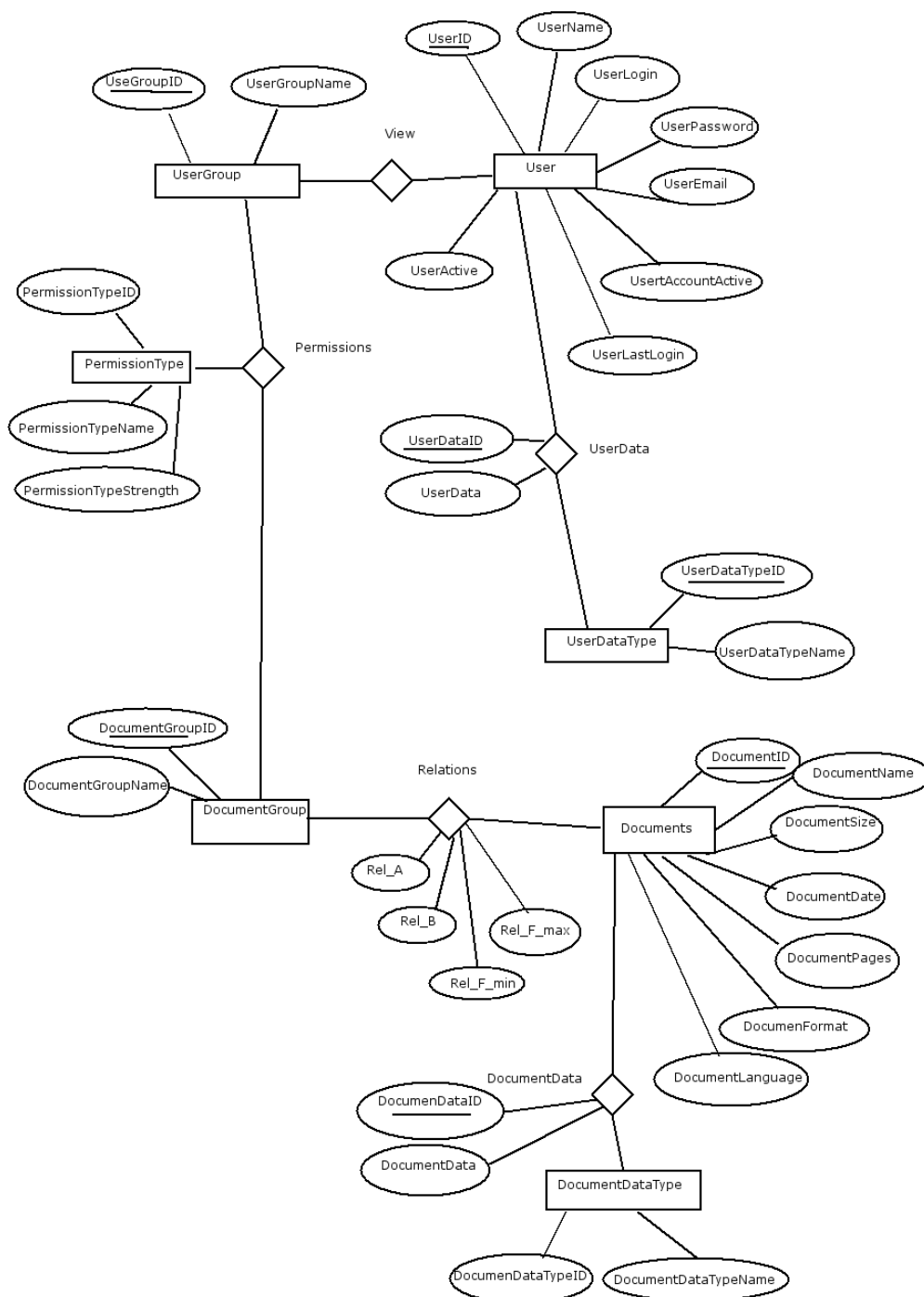


Figura A.1- Banco do Dados da Ferramenta EDMIS. Fonte: Alves (2005).

APÊNDICE A

Lista dos Documentos Utilizados para Verificação da Ontologia Proposta

Tabela A.1 - Lista dos Documentos Utilizados para Verificação da Ontologia Proposta.

Documento	Título do Documento	Membros CRAAM
DocPDF001	Faint quasar candidates from Hubble Space Telescope imaging: number counts from 31 new high-latitude fields	Não
DocPDF002	Modelling active galactic nuclei: ongoing problems for the faint-end of the luminosity function	Não
DocPDF003	The effect of stellar feedback and quasar winds on the active galactic nucleus population	Não
DocPDF004	The correlation between black hole mass and bulge velocity dispersion in hierarchical galaxy formation models	Não
DocPDF005	The host galaxies of luminous radio-quiet quasars	Não
DocPDF006	The spatial and kinematic distributions of cluster galaxies in a Λ CDM universe: comparison with observations	Não
DocPDF007	A NICMOS imaging study of high-z quasar host galaxies	Não
DocPDF008	The jet power extracted from a magnetized accretion disc	Não
DocPDF009	Chemical evolution in a model for the joint formation of quasars and spheroids	Não
DocPDF010	Infrared mergers and infrared quasi-stellar objects with galactic winds - I. NGC 2623: nuclear outflow in a proto-elliptical candidate	Não
DocPDF011	High-redshift quasars and the supermassive black hole mass budget: constraints on quasar formation models	Não
DocPDF012	Simulations of the population of Centaurus - II Individual objects	Não
DocPDF013	The narrow-line quasar NAB 0205 + 24 with XMM - Newton	Não
DocPDF014	Gravitational waves from hyper-accretion on to nascent black holes	Não
DocPDF015	Radiative heat conduction and the magnetorotational instability	Não
DocPDF016	On the observed disc temperature of accreting intermediate mass black holes	Não
DocPDF017	Present-day scaling relations for submillimetre galaxies : the origin of spheroidal systems	Não
DocPDF018	New Northern hemisphere common proper-motion pairs	Não
DocPDF019	On the detectability of a rotation-rate gradient in the solar core	Não
DocPDF020	Stellar population synthesis at the resolution of 2003	Não
DocPDF021	The host galaxies of galactic nuclei	Não

(continuação) **Tabela A.1** - Lista dos Documentos Utilizados para Verificação da Ontologia Proposta.

Documento	Título do Documento	Membros CRAAM
DocPDF022	Stellar masses and star formation histories for 10e5 galaxies from the Sloan Digital Sky Survey	Não
DocPDF023	Aurora gets the go-ahead	Não
DocPDF024	Look out, aurorae about	Não
DocPDF025	Satellites that see the Earth move	Não
DocPDF026	Small-scale seismic heterog	Não
DocPDF027	A short history of cosmology	Não
DocPDF028	The PhD and careers in astronomy in the UK	Não
DocPDF029	A new solar burst spectral component emitting only in the Terahertz range	Sim
DocPDF030	Properties of Fast Submillimeter time structures during a large Solar Flare	Sim
DocPDF031	Electron time-of-Flighth differences in Solar Flares	Não
DocPDF032	Solar cosmic rays in the near Earth space and atmosphere	Não
DocPDF033	The deduction of Energy spectra of non-thermal electrons in flares from the observed dynamic spectra of Hard x-ray bursts	Não
DocPDF034	Identification of solar sources of major geomagnetic storns	Sim
DocPDF035	Fast temporal variations of circular polarization degrrre during a microwave solar burst	Sim
DocPDF036	A High-Energy solar flare burst complex and the physical properties of its source region	Sim
DocPDF037	Primary and Derived parameters of common relevance of astronomy , geodesy , and geodynamics	Não
DocPDF038	Radiosatronomy and RadioInterferometry on the "Hidden" Lunar Surface	Não
DocPDF039	Meteoroid Enginning Model (MEM): A meteoroid model for the inner solar system	Não
DocPDF040	Interprtetation of fast ripple structure in solar impulse bursts	Sim
DocPDF041	Diffuse component spectra of solar active regions at submillimeter wavelengths	Sim
DocPDF042	Analysis of the impulsive phase of a solar flare at submillimeter wavelengths	Sim
DocPDF043	Decay time of type III solar bursts observed at kilometric wavelengths	Não
DocPDF044	Theory of solar bursts	Não
DocPDF045	Diffuse auroral scattering by electron cyclotron harmonic and whistler mode waves during an isolated substorm	Não
DocPDF046	Double loop configuration of a Flaring region from microwave , extreme -ultraviolet , and x-ray imaging data	Não
DocPDF047	The Position and Polarization of Type III solar Bursts	Não

(continuação) Tabela A.1 - Lista dos Documentos Utilizados para Verificação da Ontologia Proposta.

Documento	Título do Documento	Membros CRAAM
DocPDF048	Global-scale electron precipitation features seen in UV and X rays during substorms	Não
DocPDF049	Determination of submillimeter atmospheric opacity at el leoncito , argentina andes	Sim
DocPDF050	Determination of submillimeter atmospheric opacity at el leoncito.	Sim (outra fonte)
DocPDF051	A new solar burst spectral component emitting only in the Terahertz range	Sim
DocPDF052	Low-frequency waves upstream and downstream of the terrestrial bow shock	Não
DocPDF053	Recent results on solar activity at submillimeter wavelengths	Sim
DocPDF054	Gyrosynchrotron emission and absorption in a magnetoactive plasma	Não
DocPDF055	A multiwavelength analysis of na electron-dominated gamma-ray event associated with a disk solar flare	Não
DocPDF056	The origin of the solar flare waiting-time distribution	Não
DocPDF057	Electron acceleration in reconnecting curret sheets	Não
DocPDF058	Extragalactic Astronomy and Cosmology an Introduction	Não
DocPDF059	Extragalactic Astronomy and Cosmology	Não
DocPDF060	The lithium-rotation correlations for WTTS in Taurus-Auriga	Não
DocPDF061	The G-dwarf problem in the Galactic spheroid	Não
DocPDF062	The first lighth curves and period analysis of the eclipsing binary V821 Cas	Não
DocPDF063	Photometric monitoring of the ROSAT selected late type stars	Não
DocPDF064	Contribution of pair production process in full-energy peak efficiency calculation of semiconductor detectors for axial point sources with high-energy g-rays	Não
DocPDF065	MultiTrans SP3 code in coupled photon-electron transport problems	Não
DocPDF066	Kb/Ka X-ray intensity ratios for elements in the range 16pZp92 excited by 5.9, 59.5 and 123.6 keV photons	Não
DocPDF067	An overview of Jovian electrons during the distant Ulysses Jupiter flyby	Não
DocPDF068	Localized "Jets" of Jovian electrons observed during Ulysses' distant Jupiter flyby in 2003-2004	Não
DocPDF069	Entanglement Degree of Parasupersymmetric Coherent States of Harmonic Oscillator	Não

(continuação) Tabela A.1 - Lista dos Documentos Utilizados para Verificação da Ontologia Proposta.

Documento	Título do documento	Membros CRAAM
DocPDF070	Particle Production in 5-Dimensional Cosmological Relativity	Não
DocPDF071	Hawking Radiation of Charged Particles via Tunneling from Arbitrarily Dimensional Reissner-Nordström Black Holes	Não
DocPDF072	Competition Among Three Coupled Bose-Einstein Condensates due to Potential Deviation	Não
DocPDF073	Asteroid Models from the Pan-STARRS Photometry	Não
DocPDF074	Analysing the New Saturnian Rings, R/2004 S1 and R/2004 S2	Não
DocPDF075	Dust Production from the Hypervelocity Impact Disruption of Hydrated Targets	Não
DocPDF076	The Distinct Light Curve Shape of the Asteroid (469)	Não
DocPDF077	Orbit Determination of Binary Asteroids	Não
DocPDF078	Spectral Investigation of Two Asteroidal Fireballs	Não
DocPDF079	The origin and evolution of the atmospheres of venus, earth and mars	Não
DocPDF080	Mangala valles, mars: assessment of early stages of flooding and downstream flood evolution	Não
DocPDF081	The atmospheres of saturn and titan in the near-infrared: first results of cassini/vims	Não
DocPDF082	Titan's ground reflectance retrieval from cassini-vims data taken during the july 2nd, 2004 fly-by at 2 am ut	Não
DocPDF083	Cassini cirs observations of a roll-off in saturn ring spectra at submillimeter wavelengths	Não
DocPDF084	G-mode classification of spectroscopic data	Não
DocPDF085	A timewise kinematic method for satellite gradiometry: goce simulations	Não
DocPDF086	Determination of precipitation cycle in beijing area and comparison with solar activity cycle	Não

(continuação) Tabela A.1 - Lista dos Documentos Utilizados para Verificação da Ontologia Proposta.

Documento	Título do documento	Membros CRAAM
DocPDF087	Seas under ice: stability of liquid-water oceans within icy worlds	Não
DocPDF088	Effect of perpendicular a.c. Electric field on the oblique whistler mode instability in the earth's magnetosphere	Não
DocPDF089	Joint lightcurve observations of 10 near-earth asteroids from modra and ondřejov	Não
DocPDF090	Dust Activity in Comet 67P/Churyumov-Gerasimenko from February 20 to April 20, 2003	Não
DocPDF091	Radiation exposure and mission strategies for interplanetary manned missions (remsim)	Não
DocPDF092	Solar System Formation and Early Evolution: the First 100 Million Years	Não
DocPDF093	Dating Methods and Corresponding Chronometers in Astrobiology	Não
DocPDF094	Building of a Habitable Planet	Não
DocPDF095	Key features of intense geospace storms – A comparative study of a solar maximum and a solar minimum storm	Não
DocPDF096	Ejecta velocity distribution for impact cratering experiments on porous and low strength targets	Não
DocPDF097	Io-Jupiter interaction, millisecond bursts and field-aligned potentials	Não
DocPDF098	Solar and climate signal records in tree ring width from Chile (AD 1587-1994)	Não
DocPDF099	Effect of parallel electric field on Alfvén wave in thermal magnetoplasma	Não
DocPDF100	Electromagnetic measurements on Martian soil analogs: Implications for MARSIS and SHARAD radars in detecting subsurface water	Não

APÊNDICE B

Conjunto de *Stop Words* Incorporadas ao Sistema REDMIS

Tabela B.1 - *Stop Words* incorporadas ao sistema REDMIS.

INGLÊS	FRANCÊS	ALEMÃO	ESPAÑHOL	ITALIANO	PORTUGUES
a	a	a	él	a	a
a's	à	ab	ésta	abbastanza	à
able	â	aber	ésta	abbiamo	adeus
about	abord	aber	éste	accidenti	agora
above	afin	ach	éstos	ad	aí
according	ah	acht	última	adesso	ainda
accordingly	ai	achte	últimas	affinche	além
across	aie	achten	último	affinché	algo
actually	ainsi	achter	últimos	agli	algumas
after	allaient	achtes	a	ahime	alguns
afterwards	allo	ag	añadió	ahimè	ali
again	allô	alle	aún	ai	ano
against	allons	allein	actualmente	al	anos
ain't	après	allem	adelante	alcuna	antes
all	assez	allen	además	alcuni	ao
allow	attendu	aller	afirmó	alcuno	aos
allows	au	allerdings	agregó	all	apenas
almost	aucun	alles	ahí	alla	apoio
alone	aucune	allgemeinen	ahora	alle	após
along	aujourd	als	al	allo	aquela
already	aujourd'hui	als	algún	altri	aquelas
also	auquel	also	algo	altrimenti	aquele
although	aura	am	alguna	altro	aqueles
always	auront	an	algunas	altrui	aqui
am	aussi	andere	alguno	anche	aquilo
among	autre	anderen	algunos	ancora	área
amongst	autres	andern	alrededor	anni	as

(continuação) Tabela B.1 – *Stop Words* incorporadas ao sistema REDMIS.

INGLÊS	FRANÇÊS	ALEMÃO	ESPAÑHOL	ITALIANO	PORTUGUES
an	aux	anders	ambos	anno	às
and	auxquelles	au	ante	ansa	assim
another	auxquels	auch	anterior	assai	até
any	avaient	auch	antes	attesa	atrás
anybody	avais	auf	apenas	avanti	através
anyhow	avait	aus	aproximadamente	avendo	baixo
anyone	avant	ausser	aquí	avente	bastante
anything	avec	außer	así	aver	bem
anyway	avoir	ausserdem	aseguró	avere	bom
anyways	ayant	außerdem	aunque	avete	breve
anywhere	b	b	ayer	aveva	cá
apart	bah	bald	bajo	avuta	cada
appear	beaucoup	bei	bien	avute	catorze
appreciate	bien	beide	buen	avuti	cedo
appropriate	bigre	beiden	buena	avuto	cento
are	boum	beim	buenas	basta	certamente
aren't	bravo	beispiel	bueno	bene	certeza
around	brrr	bekannt	buenos	benissimo	cima
as	c	bereits	cómo	berlusconi	cinco
aside	ça	besonders	cada	brava	coisa
ask	car	besser	casi	bravo	com
asking	ce	besten	cerca	c	como
associated	ceci	bin	cierto	casa	conselho
at	cela	bis	cinco	caso	contra
available	celle	bisher	comentó	cento	custa
away	celle-ci	bist	como	certa	da
awfully	celle-là	c	con	certe	dá
b	celles	d	conocer	certi	dão
be	celles-ci	da	consideró	certo	daquela
became	celles-là	dabei	considera	che	daquele
because	celui	dadurch	contra	chi	dar
become	celui-ci	dafür	cosas	chicchessia	das
becomes	celui-là	dagegen	creo	cinque	de
becoming	cent	daher	cual	chiunque	debaixo
been	cependant	dahin	cuales	ci	demais
before	certain	dahinter	cualquier	ciascuna	dentro
beforehand	certaine	damals	quando	ciascuno	depois
behind	certaines	damit	cuanto	cima	desde
being	certain	danach	cuatro	cio	dessá
believe	certes	daneben	cuenta	ciò	desse
below	ces	dank	da	cioe	desta
beside	cet	dann	dado	cioè	deste
besides	cette	daran	dan	circa	deve
best	ceux	darauf	dar	citta	deverá

(continuação) Tabela B.1 – *Stop Words* incorporadas ao sistema REDMIS.

INGLÊS	FRANÇÊS	ALEMÃO	ESPAÑHOL	ITALIANO	PORTUGUES
better	ceux-ci	daraus	de	città	dez
between	ceux-là	darf	debe	codesta	dezanove
beyond	chacun	darfst	deben	codeste	dezasseis
both	chaque	darin	debido	codesti	dezassete
brief	cher	darüber	decir	codesto	dezoito
but	chère	darum	dejó	cogli	dia
by	chères	darunter	del	coi	diante
c	chers	das	demás	col	diz
c'mon	chez	das	dentro	colei	dizem
c's	chiche	dasein	desde	coll	dizer
came	chut	daselbst	después	coloro	do
can	ci	dass	dice	colui	dois
can't	cinq	daß	dicen	come	dos
cannot	cinquante	dasselbe	dicho	con	doze
cant	cinquante	davon	dieron	concernente	duas
cause	cinquantième	davor	diferente	consiglio	dúvida
causes	cinquième	dazu	diferentes	contro	e
certain	clac	dazwischen	dijeron	cortesia	é
certainly	clic	dein	dijo	cos	ela
changes	combien	deine	dio	cosa	elas
clearly	comme	deinem	donde	cosi	ele
co	comment	deiner	dos	cosí	eles
com	compris	dem	durante	così	em
come	concernant	dementsprechend	e	cui	embora
comes	contre	demgegenüber	ejemplo	d	entre
concerning	couic	demgemäss	el	da	era
consequently	crac	demgemäß	ella	dagli	és
consider	d	demselben	ellas	dai	essa
considering	da	demzufolge	ello	dal	essas
contain	dans	den	ellos	dall	esse
containing	de	denen	embargo	dalla	esses
contains	debout	denn	en	dalle	esta
corresponding	dedans	denn	encuentra	dallo	está
could	dehors	denselben	entonces	davanti	estar
couldn't	delà	der	entre	degli	estas
course	depuis	deren	era	dei	estás
currently	derrière	derjenige	eran	del	estava
d	des	derjenigen	es	dell	este
definitely	dès	dermassen	esa	della	estes
described	désormais	dermaßen	esas	delle	esteve
despite	desquelles	derselbe	ese	dello	estive
did	desquels	derselben	eso	dentro	estivemos
didn't	dessous	des	esos	detto	estiveram
different	dessus	deshalb	está	deve	estiveste

(continuação) Tabela B.1 – *Stop Words* incorporadas ao sistema REDMIS.

INGLÊS	FRANÇÊS	ALEMÃO	ESPAÑHOL	ITALIANO	PORTUGUES
do	deux	desselben	están	di	estivestes
does	deuxième	dessen	esta	dice	estou
doesn't	deuxièmement	deswegen	estaba	dietro	eu
doing	devant	d.h	estaban	dila	exemplo
don't	devers	dich	estamos	dire	faço
done	devra	die	estar	dirimpetto	falta
down	différent	diejenige	estará	dopo	favor
downwards	différente	diejenigen	estas	dove	faz
during	différentes	dies	este	dovra	fazeis
e	différents	diese	esto	dovrà	fazem
each	dire	dieselbe	estos	due	fazemos
edu	divers	dieselben	estoy	dunque	fazer
eg	diverse	diesem	estuvo	durante	fazes
eight	diverses	diesen	ex	e	fez
either	dix	dieser	existe	è	fim
else	dix-huit	dieses	existen	ecco	final
elsewhere	dixième	dir	explicó	ed	foi
enough	dix-neuf	doch	expresó	egli	fomos
entirely	dix-sept	dort	fin	ella	for
especially	doit	drei	fue	eppure	foram
et	doivent	drin	fuera	era	forma
etc	donc	dritte	fueron	erano	foste
even	dont	dritten	gran	esse	fostes
ever	douze	dritter	grandes	essendo	fui
every	douzième	drittes	ha	esser	geral
everybody	dring	du	había	essere	grande
everyone	du	durch	habían	essi	grandes
everything	duquel	durchaus	haber	ex	grupo
everywhere	durant	dürfen	habrá	fa	há
ex	e	dürft	hace	fare	hoje
exactly	effet	durfte	hacen	fatto	horas
example	eh	durften	hacer	favore	isso
except	elle	e	hacerlo	fin	isto
f	elle-même	eben	hacia	finalmente	já
far	elles	ebenso	haciendo	finche	lá
few	elles-mêmes	ehrlich	han	finché	lado
fifth	en	ei	hasta	fine	local
first	encore	ei,	hay	fino	logo
five	entre	ei,	haya	forse	longe
followed	envers	eigen	he	fra	lugar
following	environ	eigene	hecho	frattanto	maior
follows	es	eigenen	hemos	fuori	maioria
for	ès	eigener	hicieron	gia	mais
former	est	eigenes	hizo	già	mal

(continuação) Tabela B.1 – *Stop Words* incorporadas ao sistema REDMIS.

INGLÊS	FRANÇÊS	ALEMÃO	ESPAÑHOL	ITALIANO	PORTUGUES
formerly	et	ein	hoy	giacche	mas
forth	etant	einander	hubo	giacché	máximo
four	étaient	eine	igual	giorni	me
from	étais	einem	incluso	giorno	meio
further	était	einen	indicó	gli	menor
furthermore	étant	einer	informó	gliela	menos
g	etc	eines	junto	gliene	mês
get	été	einige	la	glieli	meses
gets	etre	einigen	lado	glielo	meu
getting	être	einiger	las	gliene	meus
given	eu	einiges	le	governo	mil
gives	euh	einmal	les	grande	minha
go	eux	einmal	llegó	grazie	minhas
goes	eux-mêmes	eins	lleva	gruppo	momento
going	excepté	elf	llevar	ha	muito
gone	f	en	lo	hai	muitos
got	façon	ende	los	hanno	na
gotten	fais	endlich	luego	ho	nada
greetings	faisaient	entweder	lugar	i	não
h	faisant	entweder	más	ieri	naquela
had	fait	er	manera	il	naquele
hadn't	feront	Ernst	manifestó	improvviso	nas
happens	fi	erst	mayor	in	nem
hardly	flac	erste	me	infatti	nenhuma
has	floc	ersten	mediante	insieme	nessa
hasn't	font	erster	mejor	intanto	nesse
have	g	erstes	mencionó	intorno	nesta
haven't	gens	es	menos	invece	neste
having	h	etwa	mi	invere	nível
he	ha	etwas	mientras	io	no
he's	hé	euch	misma	l	noite
hello	hein	f	mismas	la	nome
help	hélas	früher	mismo	là	nos
hence	hem	fünf	mismos	lavoro	nós
her	hep	fünfte	momento	le	nossa
here	hi	fünften	mucha	lei	nossas
here's	ho	fünfter	muchas	li	nosso
hereafter	holà	fünftes	mucho	lo	nossos
hereby	hop	für	muchos	lontano	nova
herein	hormis	g	muy	loro	nove
hereupon	hors	gab	nada	lui	novo
hers	hou	ganz	nadie	lungo	novos
herself	houp	ganze	ni	ma	num
hi	hue	ganzen	ningún	macche	numa

(continuação) Tabela B.1 – *Stop Words* incorporadas ao sistema REDMIS.

INGLÊS	FRANÇÊS	ALEMÃO	ESPAÑHOL	ITALIANO	PORTUGUES
him	hui	ganzer	ninguna	macché	número
himself	huit	ganzes	ningunas	magari	nunca
his	huitième	gar	ninguno	mai	o
hither	hum	gedurft	ningunos	male	obra
hopefully	hurrah	gegen	no	malgrado	obrigada
how	i	gegenüber	nos	malissimo	obrigado
howbeit	il	gehabt	nosotras	me	oitava
however	ils	gehen	nosotros	medesimo	oitavo
i	importe	geht	nuestra	mediante	oito
i'd	j	gekannt	nuestras	meglio	onde
i'll	je	gekonnt	nuestro	meno	ontem
i'm	jusqu	gemacht	nuestros	mentre	onze
i've	jusque	gemocht	nueva	mesi	os
ie	k	gemusst	nuevas	mezzo	ou
if	l	genug	nuevo	mi	outra
ignored	la	gerade	nuevos	mia	outras
immediate	là	gern	nunca	mie	outro
in	laquelle	gesagt	o	miei	outros
inasmuch	las	gesagt	ocho	mieri	para
inc	le	geschweige	otra	mila	parece
indeed	lequel	gewesen	otras	miliardi	parte
indicate	les	gewollt	otro	milioni	partir
indicated	lès	geworden	otros	ministro	pela
indicates	lesquelles	gibt	para	mio	pelas
inner	lesquels	ging	parece	moltissimo	pelo
insofar	leur	gleich	parte	molto	pelos
instead	leurs	gott	partir	mondo	perto
into	longtemps	gross	pasada	nazionale	pode
inward	lorsque	groß	pasado	ne	pôde
is	lui	grosse	pero	né	podem
isn't	lui-même	große	pesar	negli	poder
it	m	grossen	poca	nei	põe
it'd	ma	großen	pocas	nel	põem
it'll	maint	grosser	poco	nell	ponto
it's	mais	großer	pocos	nella	pontos
its	malgré	grosses	podemos	nelle	por
itself	me	großes	podrá	nello	porque
j	même	gut	podrán	nemmeno	porquê
just	mêmes	gute	podría	neppure	posição
k	merci	guter	podrían	nessuna	possível
keep	mes	gutes	poner	nessuno	possivelmente
keeps	mien	h	por	niente	posso
kept	mienne	habe	porque	no	pouca
know	miennes	haben	posible	noi	pouco

(continuação) Tabela B.1 – *Stop Words* incorporadas ao sistema REDMIS.

INGLÊS	FRANCÊS	ALEMÃO	ESPAÑHOL	ITALIANO	PORTUGUES
knows	miens	habt	próximo	non	primeira
known	mille	hast	próximos	nondimeno	primeiro
l	mince	hat	primer	nondimento	próprio
last	moi	hatte	primera	nostra	próximo
lately	moi-même	hätte	primero	nostre	puderam
later	moins	hatten	primeros	nostri	qual
latter	mon	hätten	principalmente	nostro	quando
latterly	moyennant	heisst	propia	nulla	quanto
least	n	her	propias	nuovo	quarta
less	na	heute	propio	o	quarto
lest	ne	hier	propios	od	quatro
let	néanmoins	hin	pudo	oggi	que
let's	neuf	hinter	pueda	ogni	quê
like	neuvième	hoch	puede	ognuna	quem
liked	ni	i	pueden	ognuno	quer
likely	nombreuses	ich	pues	oltre	quero
little	nombreux	ihm	qué	oppure	questão
look	non	ihn	que	ora	quinta
looking	nos	ihnen	quedó	ore	quinto
looks	notre	ihr	queremos	osi	quinze
ltd	nôtre	ihre	quién	osì	relação
m	nôtres	ihrem	quien	ossia	sabe
mainly	nous	ihren	quienes	paese	são
many	nous-mêmes	ihrer	quiere	parecchi	se
may	nul	ihres	realizó	parecchie	segunda
maybe	o	im	realizado	parecchio	segundo
me	o	im	realizar	parte	sei
mean	ô	immer	respecto	partendo	seis
meanwhile	oh	in	sí	peccato	sem
merely	ohé	in	sólo	peggio	sempre
might	olé	indem	se	per	ser
more	ollé	infolgedessen	señaló	perche	seria
moreover	on	ins	sea	perché	sete
most	ont	irgend	sean	perchè	sétima
mostly	onze	ist	según	percio	sétimo
much	onzième	j	segunda	perciò	seu
must	ore	ja	segundo	perfino	seus
my	ou	ja	seis	pero	sexta
myself	où	jahr	ser	però	sexto
n	ouf	jahre	será	perque	sim
name	ouias	jahren	serán	perqué	sistema
namely	oust	je	sería	persone	sob
nd	ouste	jede	si	piedi	sobre
near	oultre	jedem	sido	pieno	sois

(continuação) Tabela B.1 – *Stop Words* incorporadas ao sistema REDMIS.

INGLÊS	FRANÇÊS	ALEMÃO	ESPAÑHOL	ITALIANO	PORTUGUES
nearly	p	jeden	siempre	piglia	somos
necessary	paf	jeder	siendo	piu	sou
need	pan	jedermann	siete	piú	sua
needs	par	jedermanns	sigue	più	suas
neither	parmi	jedoch	siguiente	po	tal
never	partant	jemand	sin	pochissimo	talvez
nevertheless	particulier	jemandem	sino	poco	também
new	particulière	jemanden	sobre	poi	tanto
next	particulièrement	jene	sola	poiche	tão
nine	pas	jenem	solamente	poiché	tarde
no	passé	jenen	solas	press	te
nobody	pendant	jener	solo	prima	tem
non	personne	jenes	solos	primo	têm
none	peu	jetzt	son	proprio	temos
noone	peut	k	su	puo	tendes
nor	peuvent	kam	sus	può	tenho
normally	peux	kann	tal	pure	tens
not	pff	kannst	también	purtroppo	ter
nothing	pfift	kaum	tampoco	qualche	terceira
novel	pfut	kein	tan	qualcuna	terceiro
now	pif	keine	tanto	qualcuno	teu
nowhere	plein	keinem	tenía	quale	teus
o	plouf	keinen	tendrá	quali	teve
obviously	plus	keiner	tendrán	qualsiani	tive
of	plusieurs	kleine	tenemos	qualunque	tivemos
off	plutôt	kleinen	tener	quando	tiveram
often	pouah	kleiner	tenga	quanta	tiveste
oh	pour	kleines	tengo	quante	tivestes
ok	pourquoi	kommen	tenido	quanti	toda
okay	premier	kommt	tercera	quanto	todas
old	première	können	tiene	quantunque	todo
on	premièrement	könnnt	tienen	quasi	todos
once	près	konnte	toda	quattro	trabalho
one	proche	könnte	todas	quel	três
ones	psitt	konnten	todavía	quella	treze
only	puisque	kurz	todo	quelli	tu
onto	q	l	todos	quello	tua
or	qu	lang	total	quest	tuas
other	quand	lange	tras	questa	tudo
others	quant	lange	trata	queste	um
otherwise	quanta	leicht	través	questi	uma
ought	quant-à-soi	leide	tres	questo	umas
our	quarante	lieber	tuvo	qui	uns
ours	quatorze	los	un	quindi	vai

(continuação) Tabela B.1 – *Stop Words* incorporadas ao sistema REDMIS.

INGLÊS	FRANÇÊS	ALEMÃO	ESPAÑHOL	ITALIANO	PORTUGUES
ourselves	quatre	m	una	riecco	vais
out	quatre-vingt	machen	unas	riecco	vão
outside	quatrième	macht	uno	salto	vários
over	quatrièmement	machte	unos	salvo	vem
overall	que	mag	usted	sara	vêm
own	quel	magst	va	sarà	vens
p	quelconque	mahn	vamos	sarebbe	ver
particular	quelle	man	van	scopo	vez
particularly	quelles	manche	várias	scorso	vezes
per	quelque	manchem	varios	se	viagem
perhaps	quelques	manchen	vecas	sé	vindo
placed	quelqu'un	mancher	ver	secondo	vinte
please	quels	manches	vez	seguinte	você
plus	qui	mann	y	sei	vocês
possible	quiconque	mehr	ya	sempre	vos
presumably	quinze	mein	yo	senza	vós
probably	quoi	meine		si	vossa
provides	quoique	meinem		sí	vossas
q	r	meinen		sia	vosso
que	revoici	meiner		siamo	vossos
quite	revoilà	meines		siete	zero
qv	rien	menschen		solito	
r	s	menschen		solo	
rather	sa	mich		sono	
rd	sacrebleu	mir		sopra	
re	sans	mit		sopra	
really	sapristi	mittel		sotto	
reasonably	sauf	mochte		sta	
regarding	se	möchte		staranno	
regardless	seize	mochten		stata	
regards	selon	mögen		state	
relatively	sept	möglich		stati	
respectively	septième	mögt		stato	
right	sera	morgen		stesso	
s	seront	muss		stresso	
said	ses	muß		su	
same	si	müssen		sua	
saw	sien	musst		successivo	
say	sienne	müsst		sue	
saying	siennes	musste		sugli	
says	siens	mussten		sui	
second	sinon	n		sul	
secondly	six	na		sull	
see	sixième	nach		sulla	

(continuação) Tabela B.1 – *Stop Words* incorporadas ao sistema REDMIS.

INGLÊS	FRANÇÊS	ALEMÃO	ESPAÑHOL	ITALIANO	PORTUGUES
seeing	soi	nachdem		sulle	
seem	soi-même	nahm		sullo	
seemed	soit	natürlich		suo	
seeming	soixante	neben		suoi	
seems	son	nein		tale	
seen	sont	neue		talvolta	
self	sous	neuen		tanto	
selves	stop	neun		te	
sensible	suis	neunte		tempo	
sent	suivant	neunten		ti	
serious	sur	neunter		torino	
seriously	surtout	neuntes		tra	
seven	t	nicht		tranne	
several	ta	nicht		trannefino	
shall	tac	nichts		tre	
she	tant	nie		troppo	
should	te	niemand		tu	
shouldn't	té	niemandem		tua	
since	tel	niemanden		tue	
six	telle	noch		tuo	
so	tellement	nun		tuoi	
some	telles	nun		tutta	
somebody	tels	nur		tuttavia	
somehow	tenant	o		tutte	
someone	tes	ob		tutti	
something	tic	ob		tutto	
sometime	tien	oben		uguali	
sometimes	tienne	oder		un	
somewhat	tiennes	oder		una	
somewhere	tiens	offen		uno	
soon	toc	oft		uomo	
sorry	toi	oft		uori	
specified	toi-même	ohne		va	
specify	ton	Ordnung		vale	
specifying	touchant	p		varia	
still	toujours	q		varie	
sub	tous	r		vario	
such	tout	recht		verso	
sup	toute	rechte		vi	
sure	toutes	rechten		via	
t	treize	rechter		vicino	
t's	trente	rechtes		vise	
take	très	richtig		visé	
taken	trois	rund		visto	

(continuação) Tabela B.1 – *Stop Words* incorporadas ao sistema REDMIS.

INGLÊS	FRANÇÊS	ALEMÃO	ESPAÑHOL	ITALIANO	PORTUGUES
tell	troisième	s		vita	
tends	troisièmement	sa		voi	
th	trop	sache		volta	
than	tsoin	sagt		vostra	
thank	tsouin	sagte		vostre	
thanks	tu	sah		vostri	
thanx	u	satt		vostro	
that	un	schlecht			
that's	une	Schluss			
thats	unes	schon			
the	uns	sechs			
their	v	sechste			
theirs	va	sechsten			
them	vais	sechster			
themselves	vas	sechstes			
then	vé	sehr			
thence	vers	sei			
there	via	sei			
there's	vif	seid			
thereafter	vifs	seien			
thereby	vingt	sein			
therefore	vivat	seine			
therein	vive	seinem			
theres	vives	seinen			
thereupon	vlan	seiner			
these	voici	seines			
they	voilà	seit			
they'd	vont	seitdem			
they'll	vos	selbst			
they're	votre	selbst			
they've	vôtre	sich			
think	vôtres	sie			
third	vous	sieben			
this	vous-mêmes	siebente			
thorough	vu	siebenten			
thoroughly	w	siebenter			
those	x	siebentes			
though	y	sind			
three	z	so			
through	zut	solang			
throughout		solche			
thru		solchem			
thus		solchen			
to		solcher			

(continuação) Tabela B.1 – *Stop Words* incorporadas ao sistema REDMIS.

INGLÊS	FRANÇÊS	ALEMÃO	ESPAÑHOL	ITALIANO	PORTUGUES
together		solches			
too		soll			
took		sollen			
toward		sollte			
towards		sollten			
tried		sondern			
tries		sonst			
truly		sowie			
try		später			
trying		statt			
twice		t			
two		tag			
u		tage			
un		tagen			
under		tat			
unfortunately		teil			
unless		tel			
unlikely		tritt			
until		trotzdem			
unto		tun			
up		u			
upon		über			
us		überhaupt			
use		übrigens			
used		uhr			
useful		um			
uses		und			
using		und?			
usually		uns			
uucp		unser			
v		unsere			
value		unserer			
various		unter			
very		v			
via		vergangen			
viz		viel			
vs		viele			
w		vielem			
want		vielen			
wants		vielleicht			
was		vier			
wasn't		vierte			
way		vierten			
we		vierter			

(continuação) Tabela B.1 – *Stop Words* incorporadas ao sistema REDMIS.

INGLÊS	FRANCÊS	ALEMÃO	ESPAÑHOL	ITALIANO	PORTUGUES
we'd		viertes			
we'll		vom			
we're		von			
we've		vor			
welcome		w			
well		wahr?			
went		während			
were		währenddem			
weren't		währenddessen			
what		wann			
what's		war			
whatever		wäre			
when		waren			
whence		wart			
whenever		warum			
where		was			
where's		wegen			
whereafter		weil			
whereas		weit			
whereby		weiter			
wherein		weitere			
whereupon		weiteren			
wherever		weiteres			
whether		welche			
which		welchem			
while		welchen			
whither		welcher			
who		welches			
who's		wem			
whoever		wen			
whole		wenig			
whom		wenig			
whose		wenige			
why		weniger			
will		weniges			
willing		wenigstens			
wish		wenn			
with		wenn			
within		wer			
without		werde			
won't		werden			
wonder		werdet			
would		wessen			
would		wie			

(continuação) Tabela B.1 – *Stop Words* incorporadas ao sistema REDMIS.

INGLÊS	FRANÇÊS	ALEMÃO	ESPAÑHOL	ITALIANO	PORTUGUES
wouldn't		wie			
x		wieder			
y		will			
yes		willst			
yet		wir			
you		wird			
you'd		wirklich			
you'll		wirst			
you're		wo			
you've		wohl			
your		wollen			
yours		wollt			
yourself		wollte			
yourselves		wollten			
z		worden			
zero		wurde			
		würde			
		wurden			
		würden			
		x			
		y			
		z			
		z.b			
		zehn			
		zehnte			
		zehnten			
		zehnter			
		zehntes			
		zeit			
		zu			
		zuerst			
		zugleich			
		zum			
		zum			
		zunächst			
		zur			
		zurück			
		zusammen			
		zwanzig			
		zwar			
		zwar			
		zwei			
		zweite			
		zweiten			

(continuação) Tabela B.1 – *Stop Words* incorporadas ao sistema REDMIS.

INGLÊS	FRANCÊS	ALEMÃO	ESPAÑHOL	ITALIANO	PORTUGUES
		weiter			
		weites			
		zwischen			
		zwölf			

APÊNDICE C

Mapas Conceituais de Cada Pesquisador do Grupo de Teste

Os dados para a construção dos mapas conceituais do perfil dos pesquisadores foram extraídos da Plataforma Lattes (CNPq, 2006) para o grupo de pesquisas do CRAAM – UPM.

Pesquisador P1

Projeto A- *Novos diagnósticos de anomalias geomagnéticas e seus efeitos na atmosfera terrestre global.*

Projeto B- *Telescópio solar para ondas submilimétrica cujo objetivo é conduzir pesquisas científicas, tecnológicas e ambientais, usando diagnóstico submilimétrico obtidos em 212 GHz e 405 GHz através do novo Telescópio Solar Submilimétrico (SST).*

Quadro C1- Projetos de pesquisa nos quais o pesquisador P1 atua nos últimos 10 anos.

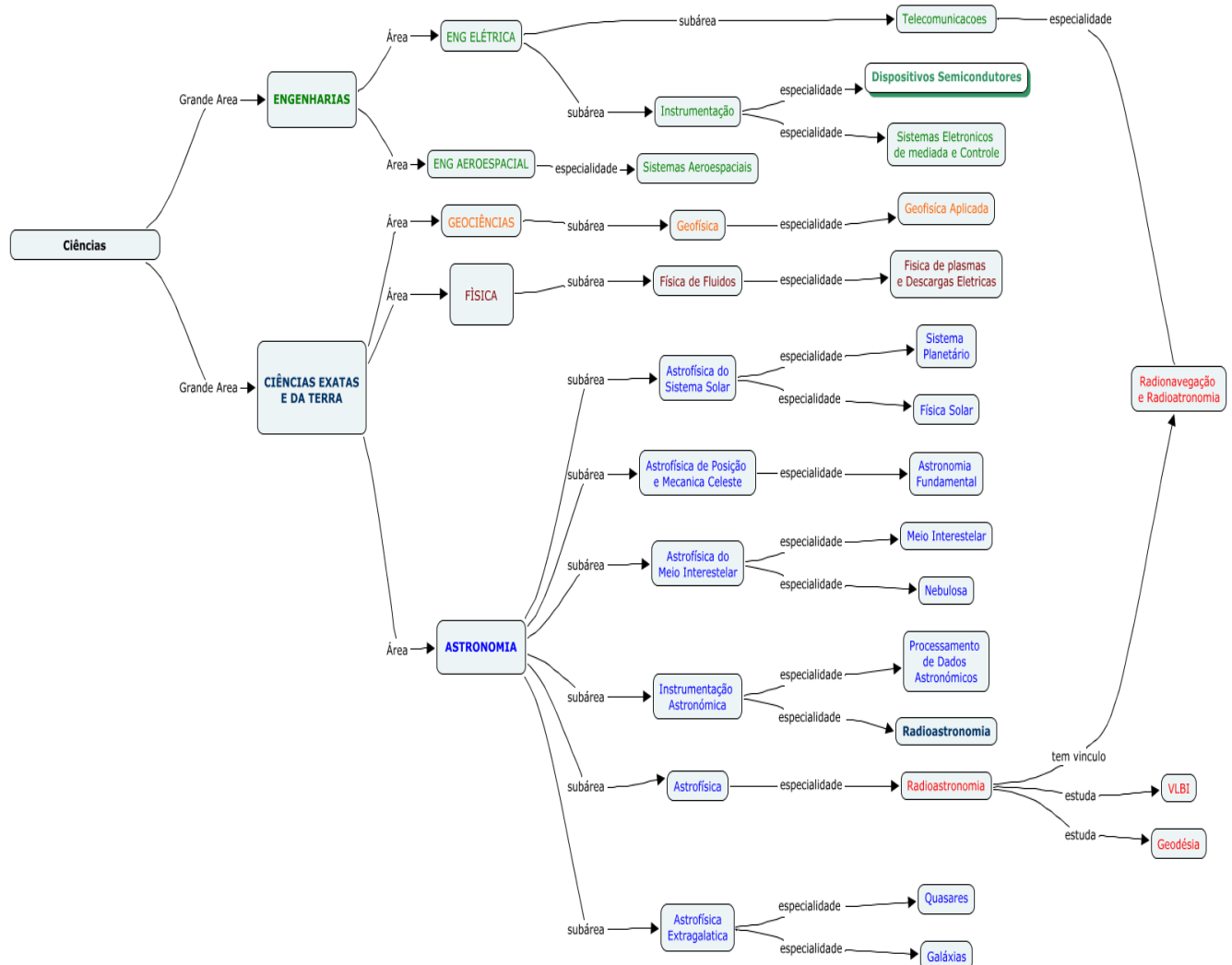


Figura C1 - Linhas de pesquisa do pesquisador P1.

Pesquisador P2

Projeto A- *Novos diagnósticos de anomalias geomagnéticas e seus efeitos na atmosfera terrestre global.*

Projeto B- *Telescópio solar para ondas submilimétrica cujo objetivo é conduzir pesquisas científicas, tecnológicas e ambientais, usando diagnóstico submilimétrico obtidos em 212 GHz e 405 GHz através do novo Telescópio Solar Submilimétrico (SST).*

Quadro C2- Projetos de pesquisa nos quais o pesquisador P2 atua nos últimos 10 anos.

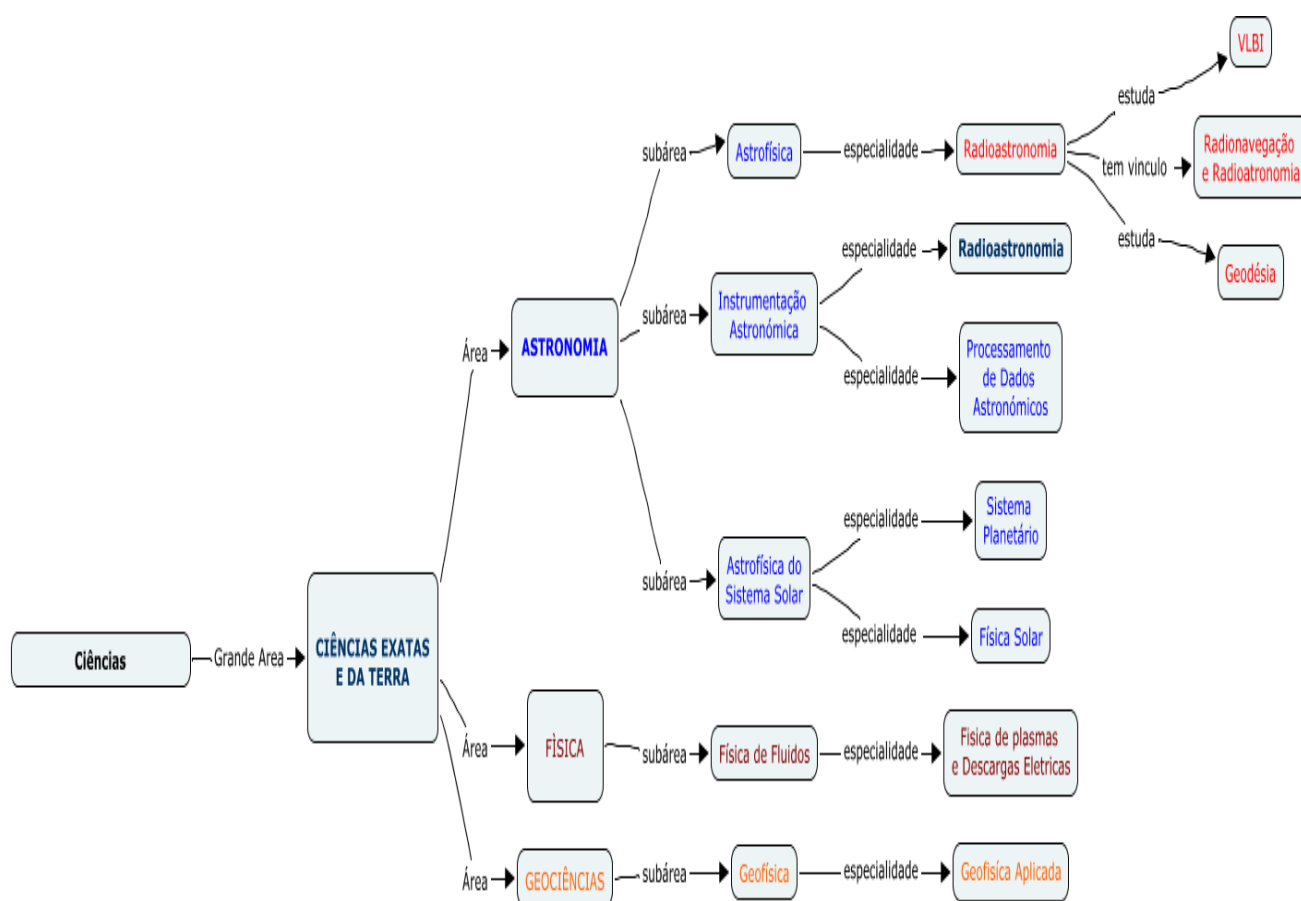


Figura C2 - Linhas de pesquisa do pesquisador P2.

Pesquisador P3

Projeto A- Caracterização dos fenômenos Sol-Terra. Tem como objetivo caracterizar as relações Sol-Terra e estudar alguns de seus domínios físicos, entre eles:

(a) Sol - atividade solar associada a ejeção de massa coronal (CME - Coronal Mass Ejection) com o objetivo de caracterizar mecanismos pulsantes que precedem o fenômeno na superfície do Sol;

(b) magnetosfera - a incidência de raios cósmicos, solares ou não, a serem detectados na região da Anomalia Magnética do Atlântico Sul (SAGA), e;

(c) ionosfera - o conteúdo total de elétrons (GPS), e efeitos de ionização em excesso causadas na baixa ionosfera na Antártica (VLF).

Projeto B- Radio Polarímetro Solar de 7 GHz. O projeto envolve o rastreamento contínuo do Sol nos dois modos de polarização circular para se estudar o comportamento da atividade solar. O objetivo é se caracterizar os sinais microondas para se estabelecer um método de previsão de atividade solar de curto e médio prazo.

Quadro C3 - Projetos de pesquisa nos quais o pesquisador P 3 atua nos últimos 10 anos.

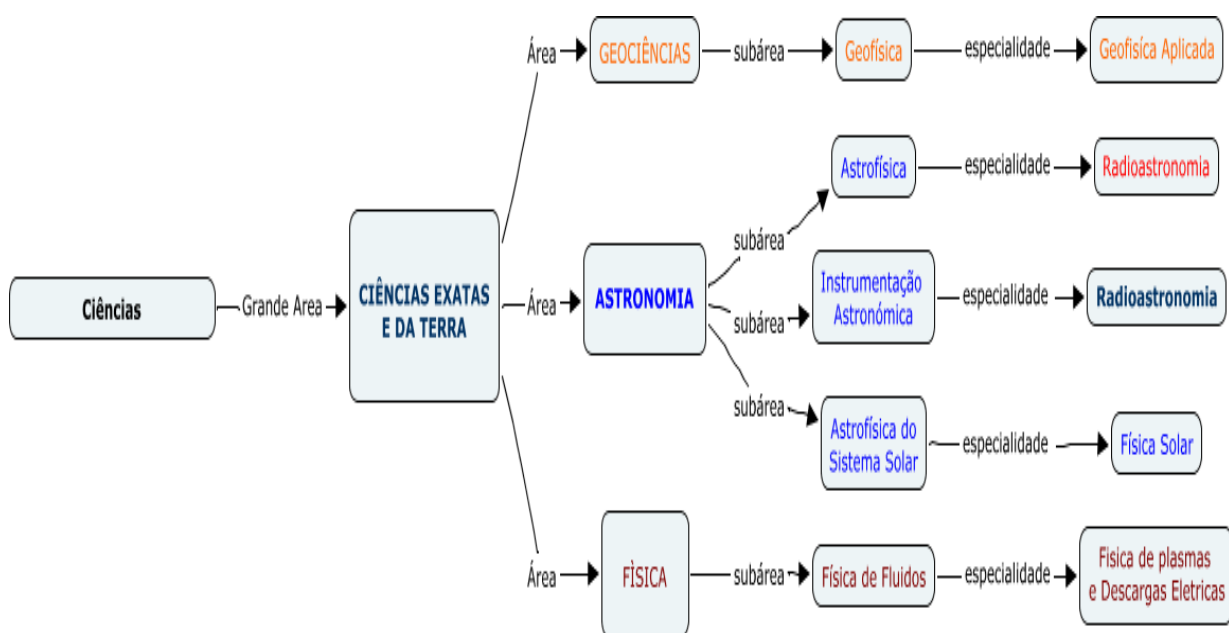


Figura C3 - Linhas de pesquisa do pesquisador P3.

Pesquisador P4

Não são informados projetos de pesquisa para esse pesquisador.

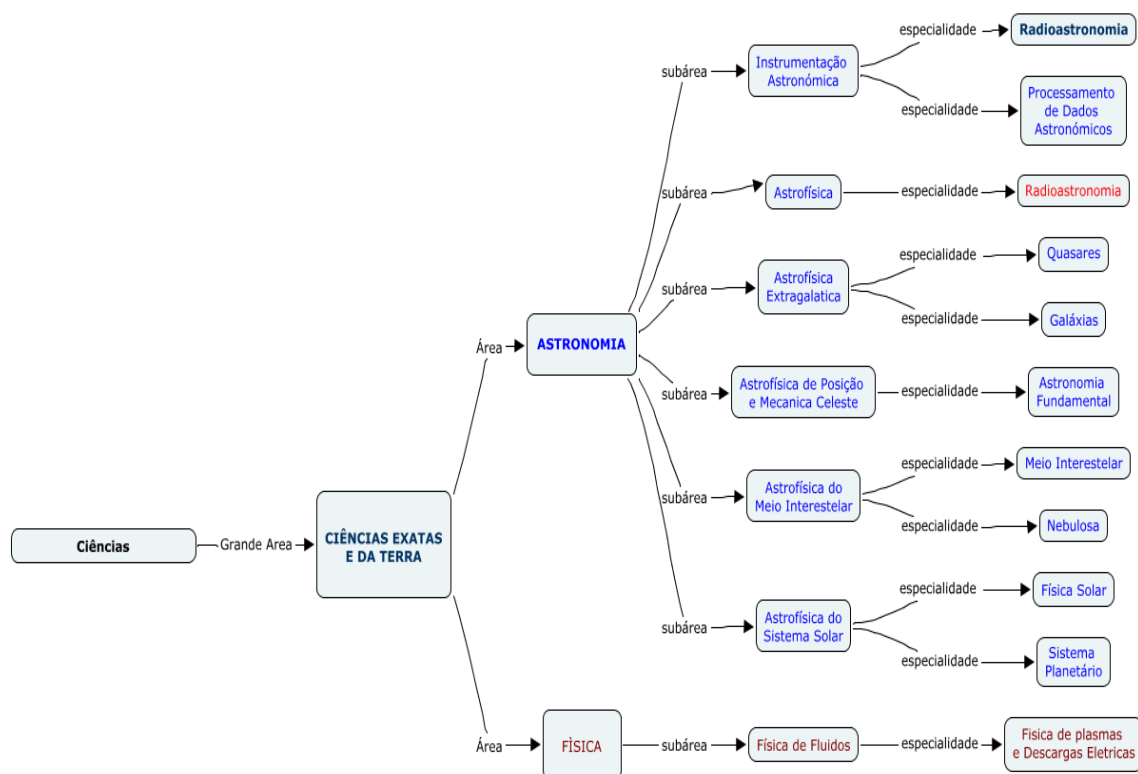


Figura C4 - Linhas de pesquisa do pesquisador P4.

Pesquisador P5

Não são informados projetos de pesquisa para esse pesquisador.

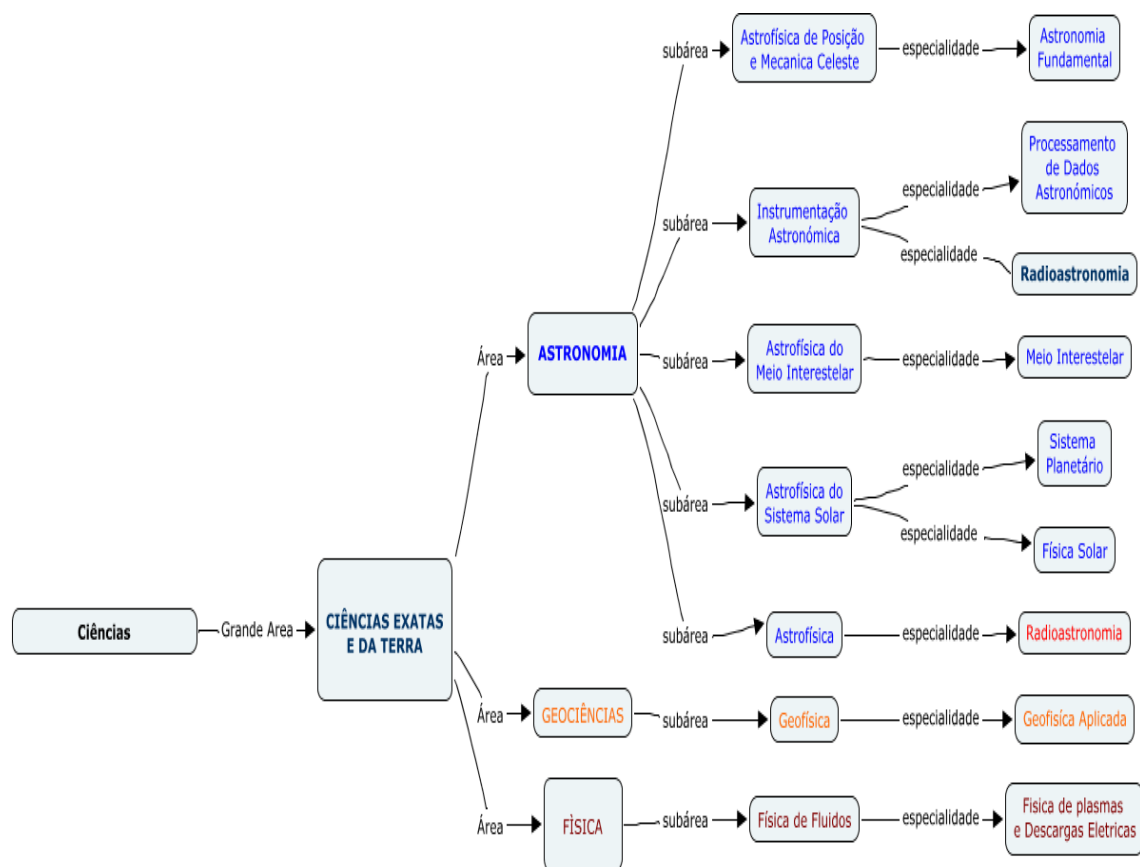


Figura C5 - Linhas de pesquisa do pesquisador P5.

Pesquisador P6

Não são informados projetos de pesquisa para esse pesquisador.

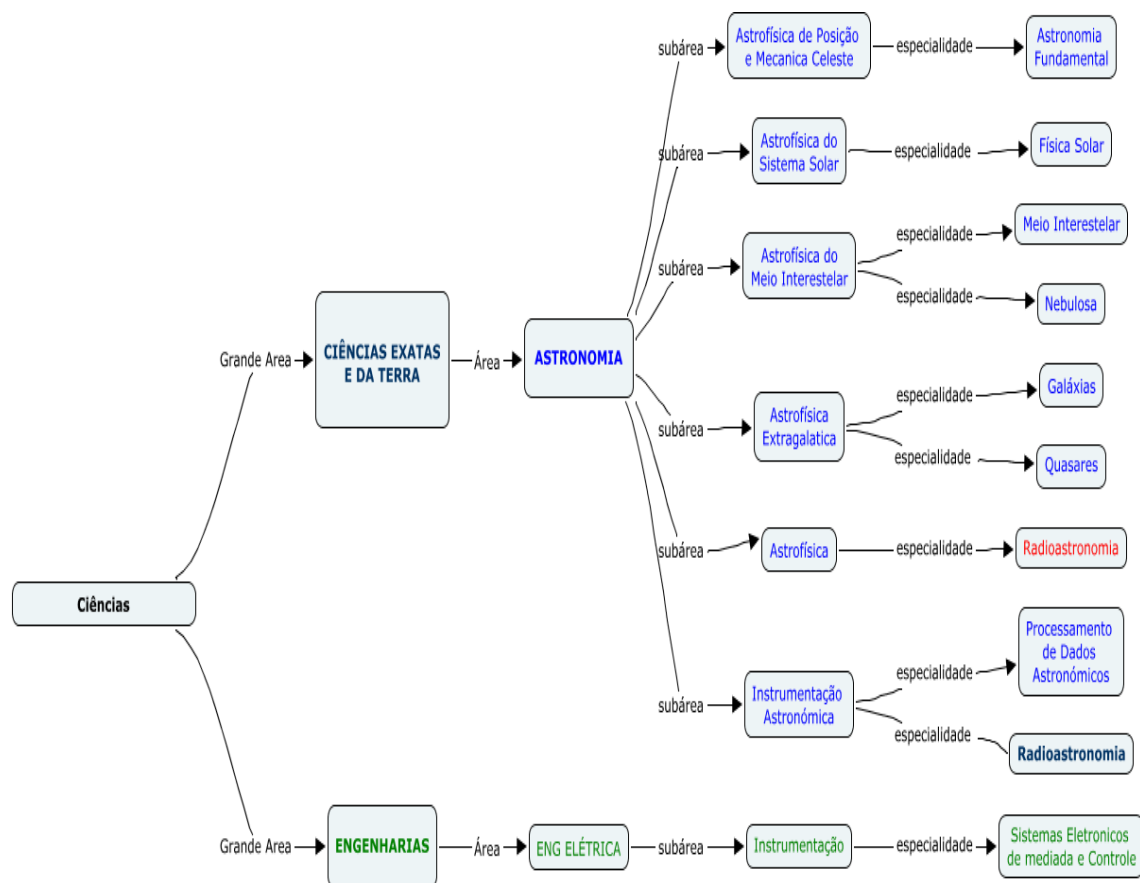


Figura C6 - Linhas de pesquisa do pesquisador P6.

Pesquisador P 7

Não são informados projetos de pesquisa para esse pesquisador.

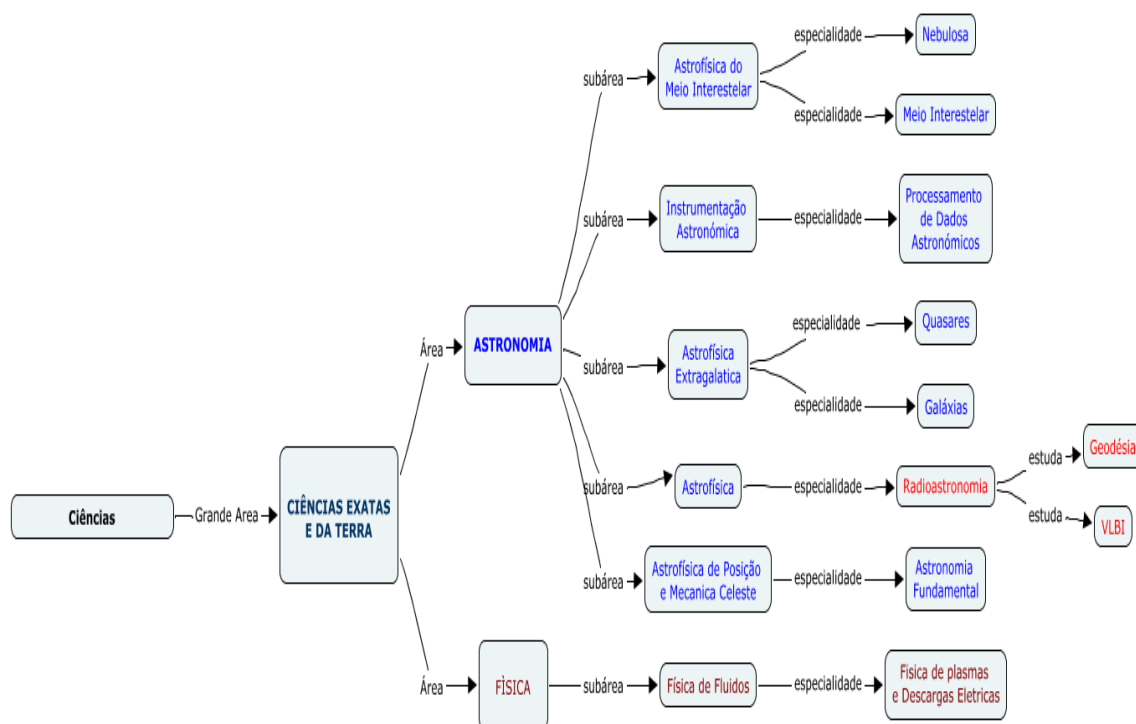


Figura C7 - Linhas de pesquisa do pesquisador P7.

Pesquisador P 8**Projeto A: Física Solar**

Pesquisa de traços de emissão de elétrons em volumes aprisionados ou jatos em precipitação na atmosfera solar. Objetivo: Propõe-se caracterizar deslocamentos e/ou confinamentos da região emissora com alta resolução angular (segundos de arco) e temporal (milissegundos). Estudo se propõe a quantificar e analisar a evolução temporal do número de partículas emissoras no espaço delimitado pela topografia magnética das regiões ativas e consequentes oscilações devidas ao espelhamento magnético ou perturbações no campo. Estudo da variabilidade da emissão quiescente solar.

Projeto B: Aplicações do Telescópio Solar para Ondas Submilimétricas – SST

Atividades concentradas nos testes e integração do sistema no sítio definitivo, e realização de observações pesquisas solares, otimização do sistema e pesquisas rádios-meteorológicas com o SST - Telescópio Solar Submilimétrico, Instalado no Complejo Astronômico El Leoncito, San Juan, Argentina, nos Andes Argentinos (Projeto FAPESP 99/06126-7).

Quadro C4 - Projetos de pesquisa nos quais o pesquisador P 8 atua nos últimos 10 anos.

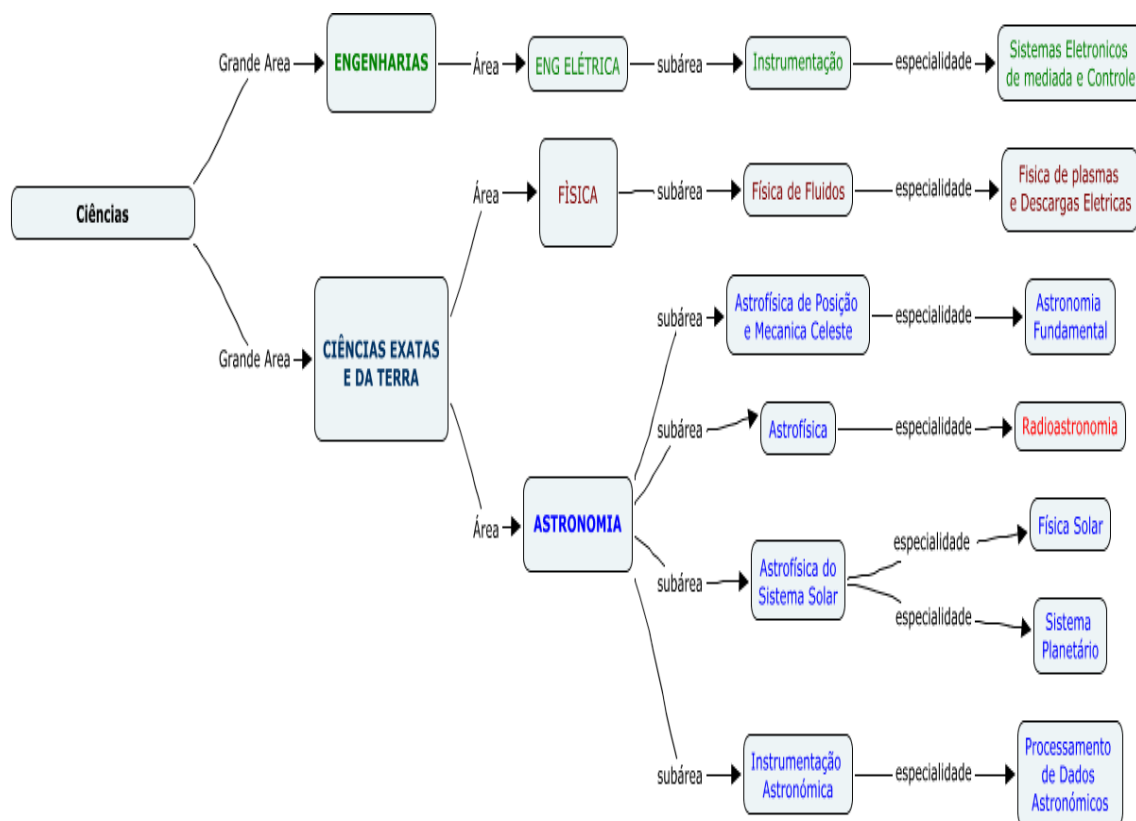


Figura C 8 - Linhas de pesquisa do pesquisador P8.

APÊNDICE D

Resultados das Consultas pelos Critérios de Ontologia e Pontos de Corte

Os quadros D1 a D8 representam as respostas obtidas na tela do usuário do sistema REDMIS, representativo dos perfis ontológicos de cada um dos oito pesquisadores do grupo em estudo. Estes dados serviram como base de dados para os gráficos do item 4.9 do texto. Os valores são apresentados por ordem decrescente de importância (*ranking*).

Pesquisador P1					
70% ont 100% doc		70% onto 70 % doc		100% onto 100%doc	
Recuper		Recuperados		Recuperados	
74		15		80	
Nome	RANK	Nome	RANK	Nome	RANK
DocPDFAstro051	0,18620	DocPDFAstro051	0,18620	DocPDFAstro051	0,20519
DocPDFAstro050	0,18493	DocPDFAstro050	0,18493	DocPDFAstro050	0,17757
DocPDFAstro049	0,18493	DocPDFAstro049	0,18493	DocPDFAstro049	0,17757
DocPDFAstro030	0,17350	DocPDFAstro030	0,17350	DocPDFAstro030	0,15549
DocPDFAstro053	0,16526	DocPDFAstro053	0,16526	DocPDFAstro053	0,15449
DocPDFAstro042	0,15237	DocPDFAstro042	0,15237	DocPDFAstro042	0,13804
DocPDFAstro052	0,14527	DocPDFAstro052	0,14527	DocPDFAstro041	0,12832
DocPDFAstro041	0,14372	DocPDFAstro041	0,14372	DocPDFAstro052	0,12729
DocPDFAstro029	0,14216	DocPDFAstro029	0,14216	DocPDFAstro029	0,12552
DocPDFAstro038	0,08384	DocPDFAstro038	0,08384	DocPDFAstro038	0,05914
DocPDFAstro046	0,06773	DocPDFAstro046	0,06773	DocPDFAstro028	0,05904
DocPDFAstro045	0,04644	DocPDFAstro045	0,04644	DocPDFAstro001	0,05850
DocPDFAstro099	0,04387	DocPDFAstro099	0,04387	DocPDFAstro007	0,05065
DocPDFAstro097	0,03754	DocPDFAstro097	0,03754	DocPDFAstro009	0,04500
DocPDFAstro035	0,02053	DocPDFAstro035	0,02053	DocPDFAstro011	0,04401
DocPDFAstro095	0,01918			DocPDFAstro058	0,04345
DocPDFAstro024	0,01853			DocPDFAstro021	0,03671
DocPDFAstro048	0,01771			DocPDFAstro046	0,03594
DocPDFAstro032	0,01372			DocPDFAstro023	0,02690
DocPDFAstro086	0,01345			DocPDFAstro063	0,02654
DocPDFAstro088	0,01271			DocPDFAstro003	0,02615
DocPDFAstro100	0,01171			DocPDFAstro097	0,02527
DocPDFAstro083	0,01136			DocPDFAstro092	0,02217
DocPDFAstro057	0,01123			DocPDFAstro045	0,02206
DocPDFAstro058	0,01123			DocPDFAstro086	0,01973
DocPDFAstro063	0,01053			DocPDFAstro032	0,01932
DocPDFAstro028	0,01024			DocPDFAstro083	0,01602
DocPDFAstro023	0,01022			DocPDFAstro076	0,01521
DocPDFAstro075	0,00949			DocPDFAstro089	0,01454
DocPDFAstro087	0,00859			DocPDFAstro090	0,01454
DocPDFAstro085	0,00758			DocPDFAstro088	0,01407
DocPDFAstro010	0,00598			DocPDFAstro035	0,01400
DocPDFAstro068	0,00593			DocPDFAstro057	0,01297
DocPDFAstro043	0,00495			DocPDFAstro075	0,01272
DocPDFAstro076	0,00474			DocPDFAstro061	0,01223
DocPDFAstro059	0,00386			DocPDFAstro077	0,01156
DocPDFAstro089	0,00383			DocPDFAstro099	0,01126
DocPDFAstro092	0,00372			DocPDFAstro059	0,01073
DocPDFAstro090	0,00356			DocPDFAstro024	0,01025
DocPDFAstro073	0,00341			DocPDFAstro048	0,00992
DocPDFAstro084	0,00341			DocPDFAstro073	0,00989
DocPDFAstro060	0,00316			DocPDFAstro068	0,00953
DocPDFAstro065	0,00309			DocPDFAstro100	0,00938
DocPDFAstro078	0,00302			DocPDFAstro094	0,00913
DocPDFAstro021	0,00293			DocPDFAstro093	0,00882
DocPDFAstro061	0,00289			DocPDFAstro060	0,00870
DocPDFAstro096	0,00286			DocPDFAstro087	0,00838
DocPDFAstro022	0,00268			DocPDFAstro095	0,00825
DocPDFAstro072	0,00252			DocPDFAstro085	0,00790
DocPDFAstro093	0,00249			DocPDFAstro010	0,00773
DocPDFAstro056	0,00239			DocPDFAstro079	0,00723
DocPDFAstro002	0,00236			DocPDFAstro022	0,00685
DocPDFAstro077	0,00217			DocPDFAstro078	0,00627
DocPDFAstro080	0,00197			DocPDFAstro039	0,00479
DocPDFAstro001	0,00197			DocPDFAstro091	0,00444
DocPDFAstro036	0,00194			DocPDFAstro043	0,00426
DocPDFAstro079	0,00183			DocPDFAstro084	0,00412
DocPDFAstro026	0,00179			DocPDFAstro067	0,00410
DocPDFAstro067	0,00172			DocPDFAstro026	0,00364
DocPDFAstro009	0,00148			DocPDFAstro005	0,00338
DocPDFAstro003	0,00148			DocPDFAstro027	0,00314
DocPDFAstro011	0,00148			DocPDFAstro056	0,00299
DocPDFAstro039	0,00148			DocPDFAstro081	0,00283
DocPDFAstro007	0,00148			DocPDFAstro008	0,00265
DocPDFAstro025	0,00118			DocPDFAstro002	0,00243
DocPDFAstro094	0,00104			DocPDFAstro082	0,00228
DocPDFAstro091	0,00000			DocPDFAstro080	0,00214
DocPDFAstro034	0,00000			DocPDFAstro036	0,00207
DocPDFAstro027	0,00000			DocPDFAstro062	0,00203

Quadro D1- Tabelas de Dados do Pesquisador P1.

Pesquisador P2							
70 % onto 100% doc		70% onto 70%doc		100% onto 70%doc		100% onto 100% doc	
Recuper 74		Recuper 16		Recuper 25		Recuper 80	
Nome	RANK	Nome	RANK	Nome	RANK	Nome	RANK
DocPDFAstro051	0,14698	DocPDFAstro051	0,14698	DocPDFAstro051	0,23220	DocPDFAstro051	0,23220
DocPDFAstro050	0,14597	DocPDFAstro050	0,14597	DocPDFAstro030	0,22617	DocPDFAstro030	0,22617
DocPDFAstro049	0,14597	DocPDFAstro049	0,14597	DocPDFAstro042	0,20356	DocPDFAstro042	0,20356
DocPDFAstro030	0,13695	DocPDFAstro030	0,13695	DocPDFAstro029	0,19379	DocPDFAstro029	0,19379
DocPDFAstro053	0,13045	DocPDFAstro053	0,13045	DocPDFAstro052	0,19200	DocPDFAstro052	0,19200
DocPDFAstro042	0,12027	DocPDFAstro042	0,12027	DocPDFAstro053	0,18395	DocPDFAstro053	0,18395
DocPDFAstro052	0,11467	DocPDFAstro052	0,11467	DocPDFAstro050	0,14762	DocPDFAstro050	0,14762
DocPDFAstro041	0,11345	DocPDFAstro041	0,11345	DocPDFAstro049	0,14762	DocPDFAstro049	0,14762
DocPDFAstro038	0,11241	DocPDFAstro038	0,11241	DocPDFAstro041	0,14055	DocPDFAstro041	0,14055
DocPDFAstro029	0,11221	DocPDFAstro029	0,11221	DocPDFAstro046	0,09531	DocPDFAstro046	0,09531
DocPDFAstro046	0,05346	DocPDFAstro046	0,05346	DocPDFAstro063	0,07901	DocPDFAstro063	0,07901
DocPDFAstro045	0,03666	DocPDFAstro045	0,03666	DocPDFAstro038	0,07730	DocPDFAstro038	0,07730
DocPDFAstro099	0,03463	DocPDFAstro099	0,03463	DocPDFAstro048	0,05436	DocPDFAstro048	0,05436
DocPDFAstro088	0,03174	DocPDFAstro088	0,03174	DocPDFAstro095	0,05039	DocPDFAstro095	0,05039
DocPDFAstro097	0,02963	DocPDFAstro097	0,02963	DocPDFAstro091	0,04280	DocPDFAstro091	0,04280
DocPDFAstro035	0,01621	DocPDFAstro035	0,01621	DocPDFAstro057	0,04259	DocPDFAstro057	0,04259
DocPDFAstro095	0,01514			DocPDFAstro032	0,04229	DocPDFAstro032	0,04229
DocPDFAstro024	0,01463			DocPDFAstro056	0,03775	DocPDFAstro056	0,03775
DocPDFAstro048	0,01398			DocPDFAstro023	0,03542	DocPDFAstro023	0,03542
DocPDFAstro032	0,01083			DocPDFAstro088	0,03413	DocPDFAstro088	0,03413
DocPDFAstro086	0,01062			DocPDFAstro058	0,03407	DocPDFAstro058	0,03407
DocPDFAstro100	0,00924			DocPDFAstro028	0,02999	DocPDFAstro028	0,02999
DocPDFAstro083	0,00896			DocPDFAstro035	0,02873	DocPDFAstro035	0,02873
DocPDFAstro057	0,00887			DocPDFAstro083	0,02756	DocPDFAstro083	0,02756
DocPDFAstro058	0,00886			DocPDFAstro045	0,02647	DocPDFAstro045	0,02647
DocPDFAstro063	0,00831					DocPDFAstro022	0,02273
DocPDFAstro028	0,00808					DocPDFAstro002	0,02256
DocPDFAstro023	0,00806					DocPDFAstro021	0,02088
DocPDFAstro075	0,00749					DocPDFAstro097	0,02073
DocPDFAstro087	0,00678					DocPDFAstro100	0,01792
DocPDFAstro085	0,00598					DocPDFAstro068	0,01769
DocPDFAstro010	0,00472					DocPDFAstro086	0,01753
DocPDFAstro068	0,00468					DocPDFAstro024	0,01628
DocPDFAstro043	0,00391					DocPDFAstro076	0,01526
DocPDFAstro076	0,00374					DocPDFAstro027	0,01483
DocPDFAstro059	0,00305					DocPDFAstro036	0,01475
DocPDFAstro089	0,00302					DocPDFAstro060	0,01422
DocPDFAstro092	0,00293					DocPDFAstro010	0,01319
DocPDFAstro090	0,00281					DocPDFAstro099	0,01265
DocPDFAstro073	0,00269					DocPDFAstro075	0,01249
DocPDFAstro084	0,00269					DocPDFAstro067	0,01245
DocPDFAstro060	0,00249					DocPDFAstro061	0,01224
DocPDFAstro065	0,00244					DocPDFAstro003	0,01163
DocPDFAstro078	0,00238					DocPDFAstro090	0,01156
DocPDFAstro021	0,00231					DocPDFAstro081	0,01080
DocPDFAstro061	0,00228					DocPDFAstro073	0,01073
DocPDFAstro096	0,00226					DocPDFAstro089	0,01067
DocPDFAstro022	0,00211					DocPDFAstro087	0,01056
DocPDFAstro072	0,00199					DocPDFAstro092	0,01037
DocPDFAstro093	0,00197					DocPDFAstro034	0,01004
DocPDFAstro056	0,00189					DocPDFAstro078	0,00928
DocPDFAstro002	0,00187					DocPDFAstro066	0,00901
DocPDFAstro077	0,00171					DocPDFAstro006	0,00882
DocPDFAstro080	0,00156					DocPDFAstro093	0,00795
DocPDFAstro001	0,00155					DocPDFAstro098	0,00794
DocPDFAstro036	0,00153					DocPDFAstro085	0,00784
DocPDFAstro079	0,00145					DocPDFAstro077	0,00773
DocPDFAstro026	0,00141					DocPDFAstro094	0,00770
DocPDFAstro067	0,00136					DocPDFAstro082	0,00641
DocPDFAstro009	0,00117					DocPDFAstro007	0,00584
DocPDFAstro003	0,00117					DocPDFAstro079	0,00574
DocPDFAstro011	0,00117					DocPDFAstro043	0,00521
DocPDFAstro039	0,00117					DocPDFAstro026	0,00521
DocPDFAstro007	0,00117					DocPDFAstro080	0,00504
DocPDFAstro025	0,00009					DocPDFAstro059	0,00469
DocPDFAstro094	0,00082					DocPDFAstro084	0,00431
DocPDFAstro091	0,00069					DocPDFAstro039	0,00426
DocPDFAstro034	0,00006					DocPDFAstro065	0,00363

Quadro D2- Tabelas de Dados do Pesquisador P2.

Pesquisador P3					
70 % onto 100%doc			70% onto 70% doc	100%onto 100% doc	100%onto 70%doc
Recuperados	74	Recuperados	16	Recuperados	83
					25
Nome	RANK	Nome	RANK	Nome	RANK
DocPDFAstro051	0,14698	DocPDFAstro051	0,14698	DocPDFAstro095	0,16382
DocPDFAstro050	0,14597	DocPDFAstro050	0,14597	DocPDFAstro053	0,14505
DocPDFAstro049	0,14597	DocPDFAstro049	0,14597	DocPDFAstro034	0,14249
DocPDFAstro030	0,13695	DocPDFAstro030	0,13695	DocPDFAstro032	0,13026
DocPDFAstro053	0,13045	DocPDFAstro053	0,13045	DocPDFAstro030	0,11404
DocPDFAstro042	0,12027	DocPDFAstro042	0,12027	DocPDFAstro051	0,10607
DocPDFAstro052	0,11467	DocPDFAstro052	0,11467	DocPDFAstro042	0,09352
DocPDFAstro041	0,11345	DocPDFAstro041	0,11345	DocPDFAstro024	0,09053
DocPDFAstro038	0,11241	DocPDFAstro038	0,11241	DocPDFAstro052	0,08310
DocPDFAstro029	0,11221	DocPDFAstro029	0,11221	DocPDFAstro029	0,08143
DocPDFAstro046	0,05346	DocPDFAstro046	0,05346	DocPDFAstro041	0,08083
DocPDFAstro045	0,03666	DocPDFAstro045	0,03666	DocPDFAstro050	0,07221
DocPDFAstro099	0,03463	DocPDFAstro099	0,03463	DocPDFAstro049	0,07221
DocPDFAstro088	0,03174	DocPDFAstro088	0,03174	DocPDFAstro046	0,05947
DocPDFAstro097	0,02963	DocPDFAstro097	0,02963	DocPDFAstro038	0,05606
DocPDFAstro035	0,01621	DocPDFAstro035	0,01621	DocPDFAstro048	0,04494
DocPDFAstro095	0,01514			DocPDFAstro068	0,04333
DocPDFAstro024	0,01463			DocPDFAstro057	0,03854
DocPDFAstro048	0,01398			DocPDFAstro088	0,03359
DocPDFAstro032	0,01083			DocPDFAstro067	0,02547
DocPDFAstro086	0,01062			DocPDFAstro045	0,02509
DocPDFAstro100	0,00924			DocPDFAstro091	0,02493
DocPDFAstro083	0,00896			DocPDFAstro035	0,02259
DocPDFAstro057	0,00887			DocPDFAstro063	0,02104
DocPDFAstro058	0,00886			DocPDFAstro056	0,01837
DocPDFAstro063	0,00831			DocPDFAstro086	0,01698
DocPDFAstro028	0,00808			DocPDFAstro010	0,01517
DocPDFAstro023	0,00806			DocPDFAstro097	0,01513
DocPDFAstro075	0,00749			DocPDFAstro076	0,01449
DocPDFAstro087	0,00678			DocPDFAstro058	0,01192
DocPDFAstro085	0,00598			DocPDFAstro023	0,01109
DocPDFAstro010	0,00472			DocPDFAstro092	0,01098
DocPDFAstro068	0,00468			DocPDFAstro079	0,01089
DocPDFAstro043	0,00391			DocPDFAstro061	0,01086
DocPDFAstro076	0,00374			DocPDFAstro094	0,01022
DocPDFAstro059	0,00305			DocPDFAstro098	0,01014
DocPDFAstro089	0,00302			DocPDFAstro075	0,01000
DocPDFAstro092	0,00293			DocPDFAstro022	0,00979
DocPDFAstro090	0,00281			DocPDFAstro099	0,00959
DocPDFAstro073	0,00269			DocPDFAstro028	0,00877
DocPDFAstro084	0,00269			DocPDFAstro083	0,00865
DocPDFAstro060	0,00249			DocPDFAstro093	0,00830
DocPDFAstro065	0,00244			DocPDFAstro021	0,00799
DocPDFAstro078	0,00238			DocPDFAstro077	0,00789
DocPDFAstro021	0,00231			DocPDFAstro081	0,00745
DocPDFAstro061	0,00228			DocPDFAstro089	0,00731
DocPDFAstro096	0,00226			DocPDFAstro073	0,00696
DocPDFAstro022	0,00211			DocPDFAstro100	0,00677
DocPDFAstro072	0,00199			DocPDFAstro060	0,00676
DocPDFAstro093	0,00197			DocPDFAstro087	0,00669
DocPDFAstro056	0,00189			DocPDFAstro036	0,00620
DocPDFAstro002	0,00187			DocPDFAstro082	0,00581
DocPDFAstro077	0,00171			DocPDFAstro059	0,00575
DocPDFAstro080	0,00156			DocPDFAstro096	0,00575
DocPDFAstro001	0,00155			DocPDFAstro078	0,00570
DocPDFAstro036	0,00153			DocPDFAstro074	0,00570
DocPDFAstro079	0,00145			DocPDFAstro039	0,00544
DocPDFAstro026	0,00141			DocPDFAstro065	0,00508
DocPDFAstro067	0,00136			DocPDFAstro090	0,00479
DocPDFAstro009	0,00117			DocPDFAstro085	0,00437
DocPDFAstro003	0,00117			DocPDFAstro043	0,00422
DocPDFAstro011	0,00117			DocPDFAstro084	0,00348
DocPDFAstro039	0,00117			DocPDFAstro027	0,00337
DocPDFAstro007	0,00117			DocPDFAstro026	0,00337
DocPDFAstro025	0,00000			DocPDFAstro011	0,00227
DocPDFAstro094	0,00000			DocPDFAstro066	0,00198
DocPDFAstro091	0,00000			DocPDFAstro080	0,00198
DocPDFAstro034	0,00000			DocPDFAstro009	0,00179

Quadro D3- Tabelas de Dados do Pesquisador P3.

Pesquisador P4							
70% onto 100% doc			70%onto 70%doc		100% onto 70%doc		100%onto 100%doc
Recuperados		71	Recuperados		30	Recuperados	
				14			71
Nome	RANK		Nome	RANK		Nome	RANK
DocPDFAstro050	0,20652		DocPDFAstro050	0,20652		DocPDFAstro051	0,16101
DocPDFAstro049	0,20652		DocPDFAstro049	0,20652		DocPDFAstro030	0,13587
DocPDFAstro051	0,18646		DocPDFAstro051	0,18646		DocPDFAstro029	0,13534
DocPDFAstro053	0,17307		DocPDFAstro053	0,17307		DocPDFAstro053	0,13468
DocPDFAstro030	0,16491		DocPDFAstro030	0,16491		DocPDFAstro052	0,12396
DocPDFAstro052	0,15681		DocPDFAstro052	0,15681		DocPDFAstro042	0,12088
DocPDFAstro041	0,15654		DocPDFAstro041	0,15654		DocPDFAstro050	0,10724
DocPDFAstro029	0,15366		DocPDFAstro029	0,15366		DocPDFAstro049	0,10724
DocPDFAstro042	0,13438		DocPDFAstro042	0,13438		DocPDFAstro041	0,10219
DocPDFAstro046	0,05036		DocPDFAstro046	0,05036		DocPDFAstro046	0,06271
DocPDFAstro035	0,04800		DocPDFAstro035	0,04800		DocPDFAstro035	0,05466
DocPDFAstro038	0,03692		DocPDFAstro038	0,03692		DocPDFAstro038	0,05303
DocPDFAstro023	0,02428		DocPDFAstro023	0,02428		DocPDFAstro023	0,04657
DocPDFAstro094	0,02213		DocPDFAstro094	0,02213		DocPDFAstro092	0,04116
DocPDFAstro060	0,01834					DocPDFAstro063	0,03949
DocPDFAstro083	0,01741					DocPDFAstro056	0,03433
DocPDFAstro093	0,01637					DocPDFAstro058	0,03256
DocPDFAstro010	0,01361					DocPDFAstro093	0,02940
DocPDFAstro001	0,01151					DocPDFAstro098	0,02909
DocPDFAstro098	0,01017					DocPDFAstro028	0,02908
DocPDFAstro058	0,00998					DocPDFAstro060	0,02849
DocPDFAstro079	0,00908					DocPDFAstro095	0,02820
DocPDFAstro092	0,00780					DocPDFAstro076	0,02786
DocPDFAstro086	0,00666					DocPDFAstro032	0,02771
DocPDFAstro063	0,00646					DocPDFAstro034	0,02621
DocPDFAstro076	0,00602					DocPDFAstro089	0,02325
DocPDFAstro089	0,00597					DocPDFAstro008	0,02067
DocPDFAstro062	0,00570					DocPDFAstro067	0,02043
DocPDFAstro032	0,00543					DocPDFAstro005	0,01989
DocPDFAstro090	0,00511					DocPDFAstro027	0,01965
DocPDFAstro034	0,00509						
DocPDFAstro077	0,00464					DocPDFAstro021	0,01882
DocPDFAstro073	0,00449					DocPDFAstro091	0,01850
DocPDFAstro066	0,00439					DocPDFAstro061	0,01765
DocPDFAstro078	0,00431					DocPDFAstro094	0,01707
DocPDFAstro084	0,00425					DocPDFAstro079	0,01679
DocPDFAstro045	0,00423					DocPDFAstro059	0,01630
DocPDFAstro056	0,00418					DocPDFAstro086	0,01628
DocPDFAstro097	0,00418					DocPDFAstro022	0,01586
DocPDFAstro100	0,00400					DocPDFAstro007	0,01494
DocPDFAstro088	0,00399					DocPDFAstro001	0,01483
DocPDFAstro091	0,00388					DocPDFAstro081	0,01466
DocPDFAstro074	0,00369					DocPDFAstro077	0,01450
DocPDFAstro075	0,00360					DocPDFAstro073	0,01419
DocPDFAstro021	0,00350					DocPDFAstro097	0,01406
DocPDFAstro085	0,00345					DocPDFAstro057	0,01390
DocPDFAstro022	0,00332					DocPDFAstro068	0,01325
DocPDFAstro067	0,00298					DocPDFAstro083	0,01321
DocPDFAstro005	0,00294					DocPDFAstro048	0,01301
DocPDFAstro028	0,00288					DocPDFAstro078	0,01158
DocPDFAstro087	0,00288					DocPDFAstro090	0,01151
DocPDFAstro081	0,00277					DocPDFAstro088	0,01074
DocPDFAstro061	0,00275					DocPDFAstro026	0,00942
DocPDFAstro080	0,00270					DocPDFAstro024	0,00872
DocPDFAstro068	0,00267					DocPDFAstro075	0,00831
DocPDFAstro026	0,00244					DocPDFAstro062	0,00825
DocPDFAstro002	0,00240					DocPDFAstro074	0,00791
DocPDFAstro099	0,00210					DocPDFAstro100	0,00783
DocPDFAstro006	0,00197					DocPDFAstro009	0,00741
DocPDFAstro007	0,00150					DocPDFAstro010	0,00740
DocPDFAstro025	0,00120					DocPDFAstro003	0,00639
DocPDFAstro048	0,00112					DocPDFAstro087	0,00519
DocPDFAstro082	0,00112					DocPDFAstro082	0,00496
DocPDFAstro024	0,00106					DocPDFAstro084	0,00482
DocPDFAstro095	0,00000					DocPDFAstro036	0,00480
DocPDFAstro072	0,00000					DocPDFAstro039	0,00454
DocPDFAstro059	0,00000					DocPDFAstro043	0,00448
DocPDFAstro065	0,00000					DocPDFAstro011	0,00294
						DocPDFAstro080	0,00284

Quadro D4- Tabelas de Dados do Pesquisador P4.

Pesquisador P5															
70% onto 100%doc				70% onto 70%doc				100%onto 70%doc				100%onto 100%doc			
Recuperados		48		Recuperados		14		Recuperados		25		Recuperados		74	
Nome	RANK			Nome	RANK			Nome	RANK			Nome	RANK		
DocPDFAstro051	0,23452			DocPDFAstro051	0,23452			DocPDFAstro051	0,22133			DocPDFAstro051	0,22133		
DocPDFAstro053	0,16288			DocPDFAstro053	0,16288			DocPDFAstro053	0,12539			DocPDFAstro053	0,12539		
DocPDFAstro050	0,14591			DocPDFAstro050	0,14591			DocPDFAstro030	0,11680			DocPDFAstro030	0,11680		
DocPDFAstro049	0,14591			DocPDFAstro049	0,14591			DocPDFAstro050	0,11528			DocPDFAstro050	0,11528		
DocPDFAstro030	0,12560			DocPDFAstro030	0,12560			DocPDFAstro049	0,11528			DocPDFAstro049	0,11528		
DocPDFAstro052	0,11884			DocPDFAstro052	0,11884			DocPDFAstro052	0,10218			DocPDFAstro052	0,10218		
DocPDFAstro041	0,11672			DocPDFAstro041	0,11672			DocPDFAstro042	0,10207			DocPDFAstro042	0,10207		
DocPDFAstro029	0,11643			DocPDFAstro029	0,11643			DocPDFAstro029	0,10018			DocPDFAstro029	0,10018		
DocPDFAstro042	0,10613			DocPDFAstro042	0,10613			DocPDFAstro041	0,09501			DocPDFAstro041	0,09501		
DocPDFAstro035	0,04868			DocPDFAstro035	0,04868			DocPDFAstro035	0,06774			DocPDFAstro035	0,06774		
DocPDFAstro046	0,03656			DocPDFAstro046	0,03656			DocPDFAstro032	0,05656			DocPDFAstro032	0,05656		
DocPDFAstro068	0,02933			DocPDFAstro068	0,02933			DocPDFAstro095	0,05294			DocPDFAstro095	0,05294		
DocPDFAstro067	0,02842			DocPDFAstro067	0,02842			DocPDFAstro023	0,04418			DocPDFAstro023	0,04418		
DocPDFAstro038	0,02500			DocPDFAstro038	0,02500			DocPDFAstro068	0,04109			DocPDFAstro068	0,04109		
DocPDFAstro032	0,02133							DocPDFAstro067	0,04070			DocPDFAstro067	0,04070		
DocPDFAstro023	0,01655							DocPDFAstro098	0,03998			DocPDFAstro098	0,03998		
DocPDFAstro098	0,01425							DocPDFAstro092	0,03967			DocPDFAstro092	0,03967		
DocPDFAstro099	0,01295							DocPDFAstro046	0,03238			DocPDFAstro046	0,03238		
DocPDFAstro094	0,01148							DocPDFAstro034	0,02967			DocPDFAstro034	0,02967		
DocPDFAstro095	0,01136							DocPDFAstro038	0,02836			DocPDFAstro038	0,02836		
DocPDFAstro092	0,01045							DocPDFAstro056	0,02473			DocPDFAstro056	0,02473		
DocPDFAstro034	0,00970							DocPDFAstro094	0,02423			DocPDFAstro094	0,02423		
DocPDFAstro083	0,00946							DocPDFAstro063	0,02320			DocPDFAstro063	0,02320		
DocPDFAstro097	0,00909							DocPDFAstro091	0,02030			DocPDFAstro091	0,02030		
DocPDFAstro100	0,00897							DocPDFAstro079	0,02023			DocPDFAstro079	0,02023		
DocPDFAstro088	0,00885											DocPDFAstro028	0,01796		
DocPDFAstro091	0,00798											DocPDFAstro088	0,01780		
DocPDFAstro001	0,00549											DocPDFAstro027	0,01771		
DocPDFAstro008	0,00487											DocPDFAstro057	0,01520		
DocPDFAstro024	0,00314											DocPDFAstro024	0,01508		
DocPDFAstro048	0,00264											DocPDFAstro086	0,01500		
DocPDFAstro045	0,00222											DocPDFAstro083	0,01258		
DocPDFAstro057	0,00201											DocPDFAstro081	0,01223		
DocPDFAstro084	0,00181											DocPDFAstro022	0,01212		
DocPDFAstro028	0,00172											DocPDFAstro073	0,01155		
DocPDFAstro086	0,00172											DocPDFAstro048	0,01129		
DocPDFAstro096	0,00172											DocPDFAstro061	0,01062		
DocPDFAstro074	0,00163											DocPDFAstro093	0,01055		
DocPDFAstro063	0,00152											DocPDFAstro076	0,01018		
DocPDFAstro069	0,00145											DocPDFAstro021	0,01015		
DocPDFAstro073	0,00137											DocPDFAstro058	0,00964		
DocPDFAstro087	0,00137											DocPDFAstro082	0,00914		
DocPDFAstro075	0,00135											DocPDFAstro060	0,00852		
DocPDFAstro079	0,00134											DocPDFAstro090	0,00657		
DocPDFAstro093	0,00128											DocPDFAstro077	0,00633		
DocPDFAstro026	0,00122											DocPDFAstro059	0,00620		
DocPDFAstro081	0,00102											DocPDFAstro089	0,00618		
DocPDFAstro080	0,00008											DocPDFAstro039	0,00555		
												DocPDFAstro100	0,00550		
												DocPDFAstro097	0,00542		
												DocPDFAstro087	0,00514		
												DocPDFAstro074	0,00489		
												DocPDFAstro075	0,00424		
												DocPDFAstro078	0,00411		
												DocPDFAstro099	0,00389		
												DocPDFAstro036	0,00356		
												DocPDFAstro062	0,00347		
												DocPDFAstro005	0,00281		
												DocPDFAstro006	0,00238		
												DocPDFAstro084	0,00224		
												DocPDFAstro045	0,00221		
												DocPDFAstro007	0,00200		
												DocPDFAstro001	0,00165		
												DocPDFAstro043	0,00162		
												DocPDFAstro040	0,00152		
												DocPDFAstro009	0,00151		
												DocPDFAstro008	0,00146		
												DocPDFAstro026	0,00146		

Quadro D5 - Tabelas de Dados do Pesquisador P5.

Pesquisador P6							
70% onto 100%doc		70%onto 70%doc		100%onto 70%doc		100% onto 100% doc	
Recuperado 44		Recuperado 13		Recuperado 33		Recuperado 84	
Nome	RANK	Nome	RANK	Nome	RANK	Nome	RANK
DocPDFAstro011	0,20495	DocPDFAstro001	0,14818	DocPDFAstro058	0,07392	DocPDFAstro058	0,07392
DocPDFAstro009	0,18776	DocPDFAstro003	0,15347	DocPDFAstro053	0,07282	DocPDFAstro053	0,07282
DocPDFAstro007	0,17426	DocPDFAstro005	0,03855	DocPDFAstro007	0,07201	DocPDFAstro007	0,07201
DocPDFAstro003	0,15347	DocPDFAstro007	0,17426	DocPDFAstro052	0,06454	DocPDFAstro052	0,06454
DocPDFAstro001	0,14818	DocPDFAstro008	0,07774	DocPDFAstro042	0,06077	DocPDFAstro042	0,06077
DocPDFAstro008	0,07774	DocPDFAstro009	0,18776	DocPDFAstro041	0,05983	DocPDFAstro041	0,05983
DocPDFAstro028	0,05069	DocPDFAstro011	0,20495	DocPDFAstro059	0,05789	DocPDFAstro059	0,05789
DocPDFAstro096	0,05042	DocPDFAstro021	0,04686	DocPDFAstro029	0,05538	DocPDFAstro029	0,05538
DocPDFAstro058	0,04919	DocPDFAstro028	0,05069	DocPDFAstro021	0,05529	DocPDFAstro021	0,05529
DocPDFAstro021	0,04686	DocPDFAstro034	0,03124	DocPDFAstro009	0,05521	DocPDFAstro009	0,05521
DocPDFAstro005	0,03855	DocPDFAstro058	0,04919	DocPDFAstro046	0,05166	DocPDFAstro046	0,05166
DocPDFAstro034	0,03124	DocPDFAstro075	0,02342	DocPDFAstro030	0,05042	DocPDFAstro030	0,05042
DocPDFAstro075	0,02342	DocPDFAstro096	0,05042	DocPDFAstro028	0,04003	DocPDFAstro028	0,04003
DocPDFAstro079	0,02062			DocPDFAstro038	0,03968	DocPDFAstro038	0,03968
DocPDFAstro004	0,01135			DocPDFAstro008	0,03266	DocPDFAstro008	0,03266
DocPDFAstro080	0,00973			DocPDFAstro003	0,03262	DocPDFAstro003	0,03262
DocPDFAstro006	0,00762			DocPDFAstro035	0,02715	DocPDFAstro035	0,02715
DocPDFAstro092	0,00589			DocPDFAstro005	0,02557	DocPDFAstro005	0,02557
DocPDFAstro010	0,00539			DocPDFAstro027	0,02356	DocPDFAstro027	0,02356
DocPDFAstro060	0,00496			DocPDFAstro032	0,02307	DocPDFAstro032	0,02307
DocPDFAstro091	0,00491			DocPDFAstro002	0,02233	DocPDFAstro002	0,02233
DocPDFAstro022	0,00468			DocPDFAstro011	0,02101	DocPDFAstro011	0,02101
DocPDFAstro063	0,00453			DocPDFAstro057	0,02084	DocPDFAstro057	0,02084
DocPDFAstro032	0,00449			DocPDFAstro092	0,02072	DocPDFAstro092	0,02072
DocPDFAstro071	0,00426			DocPDFAstro001	0,02047	DocPDFAstro001	0,02047
DocPDFAstro023	0,00401			DocPDFAstro023	0,01901	DocPDFAstro023	0,01901
DocPDFAstro078	0,00389			DocPDFAstro063	0,01762	DocPDFAstro063	0,01762
DocPDFAstro061	0,00375			DocPDFAstro051	0,01748	DocPDFAstro051	0,01748
DocPDFAstro062	0,00355			DocPDFAstro061	0,01691	DocPDFAstro061	0,01691
DocPDFAstro076	0,00351			DocPDFAstro079	0,01565	DocPDFAstro079	0,01565
DocPDFAstro094	0,00323			DocPDFAstro039	0,01530	DocPDFAstro039	0,01530
DocPDFAstro095	0,00320			DocPDFAstro034	0,01402	DocPDFAstro034	0,01402
DocPDFAstro093	0,00307			DocPDFAstro096	0,01285	DocPDFAstro096	0,01285
DocPDFAstro077	0,00305					DocPDFAstro081	0,01182
DocPDFAstro053	0,00284					DocPDFAstro022	0,01176
DocPDFAstro024	0,00269					DocPDFAstro097	0,01096
DocPDFAstro074	0,00256					DocPDFAstro075	0,01081
DocPDFAstro070	0,00256					DocPDFAstro050	0,00985
DocPDFAstro066	0,00256					DocPDFAstro049	0,00985
DocPDFAstro059	0,00256					DocPDFAstro010	0,00947
DocPDFAstro081	0,00201					DocPDFAstro060	0,00871
DocPDFAstro026	0,00192					DocPDFAstro087	0,00861
DocPDFAstro085	0,00160					DocPDFAstro100	0,00847
DocPDFAstro057	0,00135					DocPDFAstro093	0,00835
						DocPDFAstro066	0,00830
						DocPDFAstro085	0,00813
						DocPDFAstro067	0,00773
						DocPDFAstro065	0,00743
						DocPDFAstro070	0,00728
						DocPDFAstro045	0,00690
						DocPDFAstro090	0,00681
						DocPDFAstro068	0,00645
						DocPDFAstro026	0,00619
						DocPDFAstro048	0,00611
						DocPDFAstro091	0,00603
						DocPDFAstro043	0,00540
						DocPDFAstro080	0,00514
						DocPDFAstro078	0,00512
						DocPDFAstro071	0,00506
						DocPDFAstro024	0,00466
						DocPDFAstro099	0,00451
						DocPDFAstro077	0,00412
						DocPDFAstro076	0,00401
						DocPDFAstro086	0,00390
						DocPDFAstro094	0,00373
						DocPDFAstro062	0,00326
						DocPDFAstro088	0,00276
						DocPDFAstro098	0,00250

Quadro D6 - Tabelas de Dados do Pesquisador P6.

Pesquisador P7					
70% onto 100% doc		70%onto 70% doc	100%onto 70%doc	100% onto 100% doc	
Nome	RANK	Nome	RANK	Nome	RANK
DocPDFAstro001	0,23826	DocPDFAstro001	0,23826	DocPDFAstro007	0,08844
DocPDFAstro009	0,19109	DocPDFAstro009	0,19109	DocPDFAstro058	0,08522
DocPDFAstro007	0,17733	DocPDFAstro007	0,17733	DocPDFAstro021	0,08099
DocPDFAstro003	0,16455	DocPDFAstro003	0,16455	DocPDFAstro038	0,07109
DocPDFAstro005	0,10657	DocPDFAstro005	0,10657	DocPDFAstro009	0,06900
DocPDFAstro011	0,09961	DocPDFAstro011	0,09961	DocPDFAstro059	0,05247
DocPDFAstro021	0,06097	DocPDFAstro021	0,06097	DocPDFAstro005	0,04580
DocPDFAstro058	0,05892	DocPDFAstro058	0,05892	DocPDFAstro003	0,04205
DocPDFAstro028	0,05333	DocPDFAstro028	0,05333	DocPDFAstro028	0,03832
DocPDFAstro008	0,02623			DocPDFAstro001	0,03681
DocPDFAstro038	0,02128			DocPDFAstro042	0,03256
DocPDFAstro006	0,01981			DocPDFAstro053	0,03252
DocPDFAstro004	0,01805			DocPDFAstro052	0,03044
DocPDFAstro022	0,00977			DocPDFAstro027	0,02959
DocPDFAstro010	0,00949			DocPDFAstro041	0,02879
DocPDFAstro059	0,00706			DocPDFAstro008	0,02678
DocPDFAstro027	0,00569			DocPDFAstro002	0,02534
DocPDFAstro092	0,00569			DocPDFAstro029	0,02519
DocPDFAstro070	0,00451			DocPDFAstro046	0,01986
DocPDFAstro023	0,00379			DocPDFAstro030	0,01951
DocPDFAstro061	0,00355			DocPDFAstro092	0,01924
				DocPDFAstro057	0,01757
				DocPDFAstro032	0,01557
				DocPDFAstro061	0,01511
				DocPDFAstro063	0,01134
				DocPDFAstro011	0,00956
				DocPDFAstro097	0,00922
				DocPDFAstro010	0,00875
				DocPDFAstro023	0,00843
				DocPDFAstro022	0,00839
				DocPDFAstro087	0,00736
				DocPDFAstro070	0,00726
				DocPDFAstro035	0,00694
				DocPDFAstro048	0,00693
				DocPDFAstro100	0,00669
				DocPDFAstro067	0,00639
				DocPDFAstro065	0,00589
				DocPDFAstro045	0,00510
				DocPDFAstro085	0,00486
				DocPDFAstro081	0,00462
				DocPDFAstro093	0,00456
				DocPDFAstro066	0,00439
				DocPDFAstro060	0,00407
				DocPDFAstro043	0,00372
				DocPDFAstro026	0,00348
				DocPDFAstro039	0,00346
				DocPDFAstro034	0,00330
				DocPDFAstro024	0,00329
				DocPDFAstro099	0,00280
				DocPDFAstro086	0,00264
				DocPDFAstro025	0,00248
				DocPDFAstro073	0,00246
				DocPDFAstro068	0,00242
				DocPDFAstro091	0,00214
				DocPDFAstro078	0,00210
				DocPDFAstro006	0,00190
				DocPDFAstro071	0,00180
				DocPDFAstro004	0,00173
				DocPDFAstro088	0,00145
				DocPDFAstro090	0,00143
				DocPDFAstro077	0,00137
				DocPDFAstro083	0,00134
				DocPDFAstro082	0,00130
				DocPDFAstro076	0,00126
				DocPDFAstro080	0,00100
				DocPDFAstro098	0,00001
				DocPDFAstro079	0,00008
				DocPDFAstro075	0,00001

Quadro D7- Tabelas de Dados do Pesquisador P7.

Pesquisador P8					
70% onto 100% doc		70%onto 70%doc		100%onto 70%doc	
Recuperados 59		Recuperados 13		Recuperados 12	
100%onto 100%doc				Recuperados 67	
Nome	RANK	Nome	RANK	Nome	RANK
DocPDFAstro051	0,20557	DocPDFAstro051	0,20557	DocPDFAstro051	0,39456
DocPDFAstro053	0,15157	DocPDFAstro053	0,15157	DocPDFAstro052	0,24136
DocPDFAstro030	0,14414	DocPDFAstro030	0,14414	DocPDFAstro029	0,23728
DocPDFAstro052	0,13945	DocPDFAstro052	0,13945	DocPDFAstro030	0,22288
DocPDFAstro029	0,13429	DocPDFAstro029	0,13429	DocPDFAstro042	0,21215
DocPDFAstro042	0,11635	DocPDFAstro042	0,11635	DocPDFAstro053	0,20213
DocPDFAstro050	0,11215	DocPDFAstro050	0,11215	DocPDFAstro050	0,18413
DocPDFAstro049	0,11215	DocPDFAstro049	0,11215	DocPDFAstro049	0,18413
DocPDFAstro035	0,07416	DocPDFAstro035	0,07416	DocPDFAstro041	0,14597
DocPDFAstro041	0,07087	DocPDFAstro041	0,07087	DocPDFAstro035	0,11425
DocPDFAstro046	0,04551	DocPDFAstro046	0,04551	DocPDFAstro046	0,10837
DocPDFAstro056	0,02140	DocPDFAstro056	0,02140	DocPDFAstro038	0,02975
DocPDFAstro038	0,01992	DocPDFAstro038	0,01992		
DocPDFAstro034	0,01155			DocPDFAstro095	0,02350
DocPDFAstro032	0,01106			DocPDFAstro056	0,02228
DocPDFAstro095	0,01096			DocPDFAstro063	0,02056
DocPDFAstro036	0,01078			DocPDFAstro091	0,01839
DocPDFAstro040	0,00971			DocPDFAstro023	0,01636
DocPDFAstro091	0,00833			DocPDFAstro034	0,01488
DocPDFAstro022	0,00829			DocPDFAstro032	0,01342
DocPDFAstro023	0,00725			DocPDFAstro057	0,01296
DocPDFAstro063	0,00695			DocPDFAstro048	0,01158
DocPDFAstro068	0,00584			DocPDFAstro036	0,00959
DocPDFAstro009	0,00560			DocPDFAstro083	0,00868
DocPDFAstro057	0,00548			DocPDFAstro022	0,00738
DocPDFAstro083	0,00538			DocPDFAstro098	0,00564
DocPDFAstro067	0,00507			DocPDFAstro027	0,00546
DocPDFAstro043	0,00499			DocPDFAstro097	0,00544
DocPDFAstro097	0,00341			DocPDFAstro028	0,00528
DocPDFAstro098	0,00316			DocPDFAstro040	0,00524
DocPDFAstro078	0,00234			DocPDFAstro068	0,00493
DocPDFAstro045	0,00222			DocPDFAstro067	0,00439
DocPDFAstro039	0,00190			DocPDFAstro082	0,00437
DocPDFAstro086	0,00188			DocPDFAstro092	0,00403
DocPDFAstro092	0,00180			DocPDFAstro002	0,00349
DocPDFAstro094	0,00148			DocPDFAstro024	0,00338
DocPDFAstro024	0,00128			DocPDFAstro060	0,00313
DocPDFAstro027	0,00125			DocPDFAstro094	0,00307
DocPDFAstro082	0,00124			DocPDFAstro058	0,00288
DocPDFAstro073	0,00112			DocPDFAstro100	0,00277
DocPDFAstro090	0,00104			DocPDFAstro003	0,00261
DocPDFAstro093	0,00095			DocPDFAstro088	0,00252
DocPDFAstro048	0,00094			DocPDFAstro066	0,00199
DocPDFAstro089	0,00094			DocPDFAstro078	0,00193
DocPDFAstro088	0,00094			DocPDFAstro043	0,00188
DocPDFAstro079	0,00092			DocPDFAstro006	0,00183
DocPDFAstro077	0,00009			DocPDFAstro045	0,00176
DocPDFAstro087	0,00009			DocPDFAstro039	0,00156
DocPDFAstro061	0,00001			DocPDFAstro086	0,00155
DocPDFAstro081	0,00080			DocPDFAstro009	0,00115
DocPDFAstro060	0,00008			DocPDFAstro008	0,00009
DocPDFAstro074	0,00008			DocPDFAstro073	0,00093
DocPDFAstro075	0,00007			DocPDFAstro090	0,00086
DocPDFAstro021	0,00007			DocPDFAstro093	0,00001
DocPDFAstro028	0,00006			DocPDFAstro089	0,00008
DocPDFAstro059	0,00006			DocPDFAstro079	0,00076
DocPDFAstro076	0,00006			DocPDFAstro077	0,00074
DocPDFAstro058	0,00052			DocPDFAstro087	0,00074
DocPDFAstro044	0,00039			DocPDFAstro061	0,00007
				DocPDFAstro081	0,00007
				DocPDFAstro074	0,00062
				DocPDFAstro099	0,00006
				DocPDFAstro075	0,00006
				DocPDFAstro021	0,00005
				DocPDFAstro059	0,00052
				DocPDFAstro076	0,00052
				DocPDFAstro044	0,00033
				DocPDFAstro069	0,00028

Quadro D8- Tabelas de Dados do Pesquisador P8.

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)