



COPPE/UFRJ

PROSPECÇÃO TECNOLÓGICA E IDENTIFICAÇÃO DE ESPECIALISTAS
ATRAVÉS DA MINERAÇÃO DE DADOS DA PRODUÇÃO CIENTÍFICA

Rafael Nunes De Martino

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Engenharia de Sistemas e Computação, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Engenharia de Sistemas e Computação.

Orientadores: Jano Moreira de Souza

Jonice de Oliveira Sampaio

Rio de Janeiro

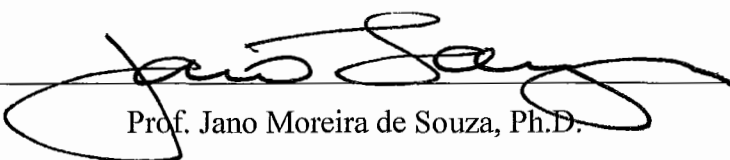
Junho de 2009

PROSPECÇÃO TECNOLÓGICA E IDENTIFICAÇÃO DE ESPECIALISTAS
ATRAVÉS DA MINERAÇÃO DE DADOS DA PRODUÇÃO CIENTÍFICA

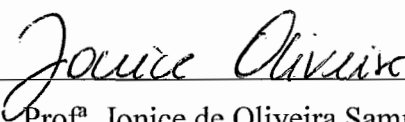
Rafael Nunes De Martino

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA (COPPE) DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM ENGENHARIA DE SISTEMAS E COMPUTAÇÃO.

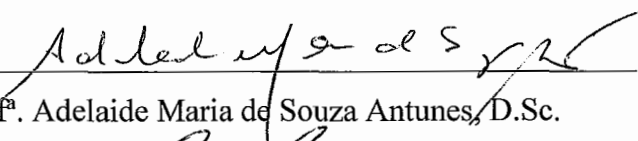
Aprovada por:



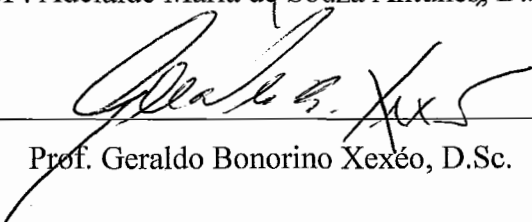
Prof. Jano Moreira de Souza, Ph.D.



Prof. Jonice de Oliveira Sampaio, D.Sc.



Prof. Adelaide Maria de Souza Antunes, D.Sc.



Prof. Geraldo Bonorino Xexéo, D.Sc.

RIO DE JANEIRO, RJ - BRASIL

JUNHO DE 2009

De Martino, Rafael Nunes

Prospecção Tecnológica e Identificação de Especialistas Através da Mineração de Dados da Produção Científica/ Rafael Nunes De Martino. – Rio de Janeiro: UFRJ/COPPE, 2009.

XVI, 139 p.: il.; 29,7 cm.

Orientadores: Jano Moreira de Souza

Jonice de Oliveira Sampaio.

Dissertação (mestrado) – UFRJ/ COPPE/ Programa de Engenharia de Sistemas e Computação, 2009.

Referências Bibliográficas: p. 112-121.

1. Prospecção Tecnológica. 2. Gestão do Conhecimento. 3. Mineração de Dados. I. Souza, Jano Moreira de *et al.* II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia de Sistemas e Computação. III. Título.

*Dedico esta dissertação
aos meus pais Sergio e Angela,
aos meus irmãos Marcia e Fabio
e aos meus avôs (in memoriam)
Nilza, Irineu, Itália e Lorenzo.*

Agradecimentos

Agradeço aos meus pais, por toda a dedicação na minha educação, pelo sacrifício de pagar bons colégios, por muitas vezes terem deixado de realizar os seus sonhos para que eu pudesse realizar os meus. Aos meus irmãos, que sempre foram e continuarão sendo meus exemplos. Aos meus avôs, que infelizmente não acompanharam parte das minhas vitórias, mas que com certeza estariam muito orgulhosos nesse momento. Aos meus tios, primos, e familiares não tão próximos, mas não menos importantes.

A Nathalia, pela ajuda fundamental durante a escrita dessa dissertação. Pelas palavras de incentivo, por revisar o texto e, principalmente, pela amizade.

Aos professores Jano Moreira de Souza e Jonice de Oliveira Sampaio, pela orientação nesse trabalho e por todas as oportunidades conferidas a mim até hoje. Espero ter o privilégio de continuar aprendendo com vocês.

Aos professores Geraldo Xexéo e Adelaide Antunes, por terem aceitado fazer parte da banca de avaliação, mesmo com tantos compromissos.

Ao Rodrigo Mesquita, por ter contribuído com esse trabalho através de seu projeto final.

A Viviane Kawamura, pela parceira no desenvolvimento do meu projeto final de curso, que acabou servindo como base para construção das idéias dessa dissertação.

Aos pesquisadores que participaram do estudo de caso, em especial aos que se interessaram pelo trabalho e deram importantes contribuições.

A algumas pessoas com quem tive oportunidade de trabalhar e, sobretudo, de aprender ao longo desses anos: Marcio Duran, Gustavo Pinto, Rafael Targino, Daniel Cardoso, Leonardo Azevedo, Raphael Vinagre e Blaschek (entre muitos outros) na Fundação COPPETEC; Flávio Rodrigo, Sergio Porto, Diogo Moreno, Marconi, Edmundo e Juan no CEPTEL; Carlos Frederico (Fred) e João Sergio Assis no NCE; Rosimary Almeida no PEB/COPPE; e finalmente, Rogério Garber e Sergio Santos na RGARBER.

Na Petrobras, agradeço os colegas de trabalho da TIC-E&P/GDIEP, em especial ao Vinícius Botelho, Henrique Bueno, Gelson Rodrigues, Eduardo Fonseca e Rafael Abrantes, além do meu gerente Francisco Aquino.

Agradeço aos amigos do Fundão, aos amigos do mestrado, aos amigos da Universidade Petrobras, aos amigos de longa data (Felipe Phil, Filipe Wins, Pati, Vitor, Carlinhos, Fernanda, entre outros) e a todos os outros que não foram citados, por proporcionarem momentos de alegria e tornarem toda trajetória mais fácil.

A Patrícia Leal, pela amizade e por toda ajuda durante o período do mestrado e dos projetos COPPETEC. A Carol e a Ana Paula pelo apoio dispensado.

Aos professores da graduação (DCC/IM/UFRJ) e do mestrado (PESC/COPPE/UFRJ), responsáveis pela minha formação e que contribuem para o ensino de excelência desta universidade.

A Capes, por ter me concedido bolsa de estudo nos dois primeiros anos do mestrado.

E finalmente, a Deus, por ter me dado forças para sempre seguir em frente.

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

PROSPECÇÃO TECNOLÓGICA E IDENTIFICAÇÃO DE ESPECIALISTAS ATRAVÉS DA MINERAÇÃO DE DADOS DA PRODUÇÃO CIENTÍFICA

Rafael Nunes De Martino

Junho/2009

Orientadores: Jano Moreira de Souza

Jonice de Oliveira Sampaio

Programa: Engenharia de Sistemas e Computação

Em um mundo em rápida evolução, marcado por constantes desenvolvimentos científicos e acirrada competição econômica, a antecipação no desenvolvimento tecnológico garante às organizações vantagem competitiva, pois possibilita a introdução de novos produtos no mercado e permite melhoria em processos existentes. Esse trabalho apresenta uma abordagem para prospecção tecnológica e identificação de especialistas que utiliza como insumo dados de publicações, teses e patentes. Foi desenvolvida uma ferramenta computacional, denominada KAIROS, capaz de minerar automaticamente esses dados na *web* e inferir as áreas de conhecimento associadas a cada uma dessas produções. Dessa forma, pode-se visualizar o comportamento das áreas científicas ao longo do tempo, como elas se relacionam e quais são consideradas promissoras. A partir de heurísticas propostas, são identificados pesquisadores e instituições especialistas ou visionários em determinado assunto. Com o objetivo de validar a abordagem proposta, foram realizados dois estudos de caso, aplicados no domínio de Computação, que apresentaram resultados satisfatórios.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

TECHNOLOGICAL FORECAST AND IDENTIFICATION OF EXPERTS
THROUGH MINING DATA FROM SCIENTIFIC PRODUCTION

Rafael Nunes De Martino

June/2009

Advisors: Jano Moreira de Souza

Jonice de Oliveira Sampaio

Department: Systems Engineering and Computer Science

Constant scientific development and hard economic competition sign the actual scenario. The advancement in the technological development ensures competitive advantage through new products introduced in the market and improvements in the existing processes. This work proposes an approach to perform technological forecast and identification of experts from publications, thesis and patents. A software named KAIROS was developed for mining the data to infer knowledge areas associated to each scientific production in the web. As a result, the behaviour of scientific areas may be seen over time, the way as they relationship themselves may be known and the promising areas may be identified. Expert or visionary researchers and institutions are identified in a domain by a set of proposed heuristics. Case studies were conducted in a Computational domain in order to validate the proposed approach. Satisfactory results were achieved in these studies.

Sumário

Capítulo 1 - Introdução	1
1.1 Problema e Motivação.....	1
1.2 Objetivos	2
1.3 Contexto	3
1.4 Organização.....	3
Capítulo 2 - Prospecção Tecnológica.....	4
2.1 Introdução.....	4
2.2 Definições.....	5
2.3 Ganhos e Aplicações	6
2.3.1 Investimento em Ciência e Tecnologia.....	7
2.3.2 Vantagem Competitiva e Conquista de Mercado.....	9
2.3.3 Instituições de Ensino e Pesquisa.....	9
2.3.4 Grandes Desafios em Computação no Brasil – SBC	10
2.4 Conclusão	11
Capítulo 3 - Técnicas de Prospecção Tecnológica	12
3.1 Introdução.....	12
3.2 Opinião de Especialistas.....	13
3.2.1 Delphi.....	13
3.2.2 Painel de Especialistas	14
3.2.3 Métodos para Estímulo da Criatividade	14
3.2.3.1 Brainstorming.....	15
3.2.3.2 Analogia Sistemática.....	15
3.3 Monitoramento Ambiental	15
3.3.1 Análise Bibliométrica.....	16
3.3.2 Mineração de Dados.....	18
3.4 Modelagem e Simulação	21
3.5 Cenários.....	22
3.5.1 Análise Estrutural Multivariada	23
3.5.2 Análise do Jogo de Atores.....	24
3.5.3 Formulação de Cenários.....	25
3.6 Análise de Tendências.....	26

3.7	Comparação entre as Técnicas	28
3.8	Trabalhos Correlatos	29
3.9	Conclusão	32
Capítulo 4 - A Proposta: Uma abordagem para Prospecção Tecnológica e Identificação de Especialistas.....		33
4.1	Introdução.....	33
4.2	GCC - Gestão do conhecimento Científico.....	35
4.2.1	Módulo de Gestão de Conhecimento Pessoal	36
4.2.2	Módulo de Gestão de Projetos	37
4.2.3	Módulo de Comunidades	37
4.2.4	Módulo de Navegação e Visualização	37
4.2.5	Máquina de Inferência.....	38
4.2.6	Módulo de Recomendação	38
4.2.7	Módulo de Inteligência Competitiva.....	39
4.2.8	Base de Conhecimento e Serviços de Busca.....	39
4.3	Arquitetura do KAIROS.....	39
4.3.1	Camada de Dados.....	40
4.3.2	Camada de Informação.....	41
4.3.3	Camada de Conhecimento.....	41
4.3.4	Tecnologias Utilizadas	42
4.4	Fonte de Dados e Extratores.....	43
4.4.1	Curriculum Lattes.....	43
4.4.1.1	Dados Importados.....	43
4.4.1.2	Extração	44
4.4.2	DBLP.....	45
4.4.2.1	Dados Importados.....	45
4.4.2.2	Extração	45
4.4.3	Banco de Teses e Dissertações Capes	46
4.4.3.1	Dados Importados.....	46
4.4.3.2	Extração	47
4.4.4	Base de Patentes Derwent Innovation Index (DII).....	47
4.4.4.1	Dados Importados.....	47
4.4.4.2	Extração	48
4.5	Identificação de Áreas.....	48

4.5.1	Inferência do Idioma	49
4.5.2	Pré-Processamento Textual	49
4.5.2.1	Retirada de Palavras Frequentes ou <i>Stop Words</i>	50
4.5.2.2	Radicalização ou <i>Stemming</i>	50
4.5.3	Dicionário de Radicais	51
4.5.3.1	Carga Manual	52
4.5.3.2	Carga Automática	53
4.5.4	Inferência das Áreas Associadas à Produção Científica	56
4.5.5	Inferência das Competências do Pesquisador	58
4.6	Mapeamento e Prospecção de Áreas	60
4.6.1	Consolidação	60
4.6.2	Prospecção	61
4.6.3	Desenvolvimento das Áreas de Conhecimento	62
4.7	Similaridade de Áreas	64
4.8	Identificação de Áreas Promissoras	67
4.9	Alinhamento com Áreas Promissoras	69
4.10	Identificação de Visionários	72
4.11	Busca por Conhecimento	76
4.12	Comparação com Trabalhos Correlatos	80
4.13	Conclusão	82
Capítulo 5 - Avaliação da Proposta		83
5.1	Introdução	83
5.2	Preparação dos Dados	84
5.3	Estudo 1: Áreas Promissoras e Pesquisadores Visionários	87
5.3.1	Definição	87
5.3.2	Planejamento	88
5.3.3	Execução	91
5.3.4	Análise de Resultados	91
5.3.4.1	Perfil dos Participantes	91
5.3.4.2	Análise das Respostas	94
5.4	Estudo 2: Identificação de Competências	98
5.4.1	Definição	98
5.4.2	Planejamento	99
5.4.3	Execução	101

5.4.4	Análise de Resultados	101
5.5	Conclusão	103
Capítulo 6 - Considerações Finais		105
6.1	Contribuições	105
6.2	Limitações	107
6.3	Trabalhos Futuros.....	109
6.4	Conclusão	111
Apêndice A – Modelo de Dados da Base KAIROS		122
Apêndice B – <i>StopWords</i> Utilizadas na Mineração		125
Apêndice C – Avaliação da Proposta.....		128
C.1	Estudo 1: Áreas Promissoras e Pesquisadores Visionários.....	128
C.2	Estudo 2: Identificação de Competências	133

Índice de Figuras

Figura 1: Modelo linear do processo de inovação (adaptado de CGEE, 2008)	8
Figura 2: Modelo paralelo de inovação (adaptado de UTTERBACK, 1994).....	8
Figura 3: Fequência de palavras em um texto (FARIA, 2001).	18
Figura 4: Técnicas de mineração textual (RODRIGUES, 2007).	20
Figura 5: Dependência direta das variáveis (LIPSOR, 2009).	23
Figura 6: Dependência indireta das variáveis (LIPSOR, 2009).	24
Figura 7: Matriz e diagrama de convergência de atores.....	25
Figura 8: Análise Morfológica aplicada para formulação de cenários.....	25
Figura 9: Exemplo da extrapolação de pontos futuros.	27
Figura 10: Exemplos de funções que podem ser utilizadas na interpolação.	28
Figura 11: Fluxo da aquisição de dados e processamento de informações.	34
Figura 12: Arquitetura do GCC (OLIEIRA <i>et al.</i> , 2005a)	36
Figura 13: Arquitetura do KAIROS	40
Figura 14: Tela inicial do GCC.	41
Figura 15: Tela inicial do KAIROS.....	42
Figura 16: Processo de identificação de áreas e competências	49
Figura 17: Exemplo da retirada de “ <i>Stop Words</i> ” (WIVES, 1997).....	50
Figura 18: Possibilidades de carga do dicionário de radicais.....	52
Figura 19: Tela de cadastro do dicionário de radicais.....	53
Figura 20: Palavras relevantes em um domínio (Adaptado de FARIA, 2001). ..	56
Figura 21: Mapeamento e prospecção de áreas (Consulta).....	63
Figura 22: Mapeamento e prospecção de áreas (Resultado)	63
Figura 23: Similaridade de áreas (Consulta)	66
Figura 24: Similaridade de áreas (Resultado)	66
Figura 25: Similaridade de áreas ano a ano.....	67
Figura 26: Identificação de áreas promissoras (Consulta)	68
Figura 27: Identificação de áreas promissoras (Resultado).....	69
Figura 28: Alinhamento com áreas promissoras (Consulta)	70
Figura 29: Alinhamento com áreas promissoras (Seleção do pesquisador).....	71
Figura 30: Alinhamento com áreas promissoras (Resultado).....	72
Figura 31: Exemplo do funcionamento da heurística.	74

Figura 32: Identificação de visionários (Consulta)	75
Figura 33: Identificação de visionários (Resultado).....	76
Figura 34: Busca por conhecimento (consulta)	77
Figura 35: Resultado da busca por produção científica.....	78
Figura 36: Resultado da busca por especialistas.	79
Figura 37: Dependência das funcionalidades e procedimentos internos.....	83
Figura 38: Tela inicial do formulário eletrônico do Estudo de Caso 1.	91
Figura 39: Gráfico das respostas sobre Envolvimento com a Pesquisa	92
Figura 40: Gráfico das respostas sobre Titulação dos Participantes.	92
Figura 41: Gráfico da relação entre o vínculo com a pesquisa e a titulação.	93
Figura 42: Gráfico das respostas sobre Experiência como Pesquisador.	93
Figura 43: Gráfico das respostas sobre Principal Área de Pesquisa.....	93
Figura 44: Universo de pesquisadores visionários e não visionários.....	104
Figura 45: Modelo de dados da base LATTES.	122
Figura 46: Modelo de dados da base DBLP.....	123
Figura 47: Modelo de dados da base de Patentes.....	123
Figura 48: Modelo de dados da base Teses.....	124
Figura 49: Tela de apresentação do formulário eletrônico do Estudo 1.....	128
Figura 50: Tela de preenchimento dos dados do participante.	129
Figura 51: Tela de avaliação da prospecção de áreas.....	129
Figura 52: Pergunta aberta sobre as áreas promissoras.....	130
Figura 53: Tela de avaliação de pesquisadores visionários (nomes omitidos)..	130

Índice de Tabelas

Tabela 1: Consolidação dos pontos fortes e fracos das famílias de técnicas.	28
Tabela 2: Tecnologias utilizadas pelo KAIROS.	42
Tabela 3: Exemplo da obtenção dos radicais das palavras.....	50
Tabela 4: Exemplo de preenchimento do dicionário de radicais.	52
Tabela 5: Informações coletadas para cada tipo de produção científica.	56
Tabela 6: Pesos de cada produção científica no cálculo de competências.....	59
Tabela 7: Exemplo da consolidação utilizando média.	61
Tabela 8: Produção científica de dado ano (dados fictícios).....	64
Tabela 9: Cálculo dos graus de similaridade entre as áreas	65
Tabela 10: Comparação entre os trabalhos correlatos.....	81
Tabela 11: Sumarização das respostas nominais da primeira etapa do Estudo 1.	94
Tabela 12: Consolidação das respostas da primeira etapa do Estudo 1.	95
Tabela 13: Áreas promissoras sugeridas pelos participantes do estudo.....	96
Tabela 14: Sumarização das respostas nominais da segunda etapa do Estudo 1.	97
Tabela 15: Consolidação das respostas da segunda etapa do Estudo 1.....	97
Tabela 16: Resultados consolidados do Estudo 2.	101
Tabela 17: Lista de <i>StopWords</i> no idioma português.....	125
Tabela 18: Lista de <i>StopWords</i> no idioma inglês.....	126
Tabela 19: Resultados detalhados da primeira etapa do Estudo 1.	131
Tabela 20: Resultados detalhados da segunda etapa do Estudo 1.....	132
Tabela 21: Resultados detalhados do Estudo 2.	133

Índice de Equações

Equação 1: Cálculo da relevância corrigida pelo peso e fator temporal.	58
Equação 2: Cálculo do fator temporal (Adaptado de KAWAMURA, 2006).	59
Equação 3: Cálculo da similaridade entre áreas.....	64
Equação 4: Cálculo da favorabilidade de uma área.	94

Capítulo 1 - Introdução

Este capítulo apresenta a motivação da dissertação e a caracterização dos problemas que serão estudados e tratados. São explicados, também, os objetivos do trabalho e o contexto dentro do qual a abordagem proposta está inserida. Por fim, para guiar o leitor e facilitar o manuseio deste trabalho, apresenta-se a organização dos capítulos da dissertação.

1.1 Problema e Motivação

Em um mundo em rápida evolução, marcado por constantes desenvolvimentos científicos e tecnológicos, integração de mercados e acirrada competição econômica, as organizações necessitam de informações que subsidiem o processo de tomada de decisão. A escolha pelas melhores estratégias possibilita que a organização sobreviva e se desenvolva no mercado (FARIA, 2001).

Nesse contexto, a antecipação no desenvolvimento tecnológico garante às organizações vantagem competitiva, pois possibilitam a introdução de novos produtos no mercado, além de permitir melhoria em processos existentes. Para a antecipação tecnológica, torna-se indispensável o investimento, no momento apropriado, em pesquisa de áreas estratégicas para a atuação da organização.

Os governos, por sua vez, possuem a difícil tarefa de direcionar seus, muitas vezes escassos, recursos públicos para fomentar a pesquisa e desenvolvimento (P&D) de áreas que produzam satisfatório retorno para os interesses estratégicos do país e para sociedade como um todo.

A prospecção tecnológica, definida por PORTER *et al.* (2004) como o processo de descrever as características e impactos de uma determinada tecnologia no futuro, surge, nesse cenário, como um meio de apoiar a tomada de decisão em nível estratégico.

O desenvolvimento tecnológico futuro é fortemente influenciado pela pesquisa científica que o antecede. Desse modo, os dados históricos da produção científica auxiliam o processo de prospecção, pois explicam o desenvolvimento de uma área de conhecimento ao longo do tempo e indicam os possíveis comportamentos no futuro.

O crescimento anual aproximado de 600.000 registros de patentes no mundo (MCT, 2009a) e o surgimento de novas conferências e periódicos evidenciam a grande disponibilidade de informação científica registrada. A proliferação na internet de bases eletrônicas de acesso público e o aprimoramento de técnicas computacionais de mineração de dados possibilitam novas análises e ampliam os horizontes da prospecção tecnológica.

O processo de prospecção tecnológica, apesar de freqüentemente ser apoiado por ferramentas computacionais, ainda necessita do envolvimento e da interação de especialistas, para validar e complementar os resultados.

Da mesma forma, o desenvolvimento da tecnologia, em si, depende fundamentalmente de capital humano. A competição entre as organizações, já acirrada no mercado, estende-se à captação de recursos humanos qualificados.

Portanto, pode-se afirmar que, para a prospecção tecnológica ser de boa qualidade e, principalmente, para gerar resultados práticos no desenvolvimento de uma nova tecnologia, torna-se necessária a identificação e o envolvimento de especialistas nas áreas estudadas.

1.2 Objetivos

O objetivo principal dessa dissertação, como o título sugere, é apresentar uma abordagem eficaz que possibilite a prospecção tecnológica e a identificação de especialistas. Essa abordagem utiliza dados relativos à produção científica, como publicações, patentes, teses e dissertações, disponibilizadas na *web* e é válida para qualquer domínio de conhecimento.

São propostos métodos e heurísticas capazes de analisar toda a informação minerada, de forma a inferir as áreas de conhecimento associadas a cada produção científica. Dessa forma, pode-se visualizar o comportamento das áreas científicas ao longo do tempo, como elas se relacionam e identificar quais são consideradas promissoras.

Pesquisadores e instituições, considerados especialistas ou visionários em determinado assunto, são devidamente identificados por heurísticas, propostas na abordagem, que analisam a autoria das produções científicas mineradas e ponderam sobre uma série de fatores.

Para validar esta abordagem, foi desenvolvido um ambiente computacional, denominado KAIROS, capaz de minerar automaticamente os dados disponibilizados na *web*, executar os métodos e heurísticas propostos, além apresentar funcionalidades para o usuário final da ferramenta.

1.3 Contexto

O ambiente computacional KAIROS, que suporta a abordagem proposta nesta dissertação, foi integrado ao GCC (Gestão de Conhecimento Científico), um ambiente *web* voltado ao público acadêmico, constituindo uma importante ferramenta para a Gestão do Conhecimento no ambiente científico (OLIVEIRA et al, 2005a). O GCC propicia uma centralização de dados, informações e conhecimento sobre o meio acadêmico, facilitando o compartilhamento e a disseminação do conhecimento gerado. Maiores informações são apresentadas no Capítulo 4.

1.4 Organização

Além deste capítulo introdutório, essa dissertação está organizada em outros cinco capítulos.

O Capítulo 2 introduz conceitos de prospecção tecnológica, apresentando definições e os possíveis ganhos em sua utilização. O Capítulo 3 apresenta diversas técnicas de prospecção, comparando seus pontos fortes e fracos.

O Capítulo 4 apresenta a proposta dessa dissertação: uma abordagem para prospecção tecnológica e identificação de especialistas, utilizando dados da produção científica. Um sistema computacional, denominado KAIROS, foi desenvolvido utilizando a abordagem proposta.

O Capítulo 5 apresenta um estudo de caso realizado com o objetivo de validar a proposta da dissertação. O estudo foi desenvolvido em duas etapas e contou com a participação de pesquisadores da área de computação. Ao final do capítulo, os resultados são apresentados em detalhe.

O Capítulo 6 conclui essa dissertação, apresentando as suas principais contribuições, relatando as limitações verificadas e enumerando possíveis trabalhos futuros.

Capítulo 2 - Prospecção Tecnológica

Esse capítulo introduz definições e conceitos associados à Prospecção Tecnológica, evidenciando sua importância e principais aplicações.

2.1 Introdução

A prospecção tecnológica nasceu na década de 50 em razão da crescente competição entre os países durante a guerra fria. A necessidade de estar à frente dos oponentes fez com que a inovação fosse uma meta constante a ser perseguida. Os governos da época buscavam formas de antecipar as necessidades e tecnologias futuras, e priorizavam ações de pesquisa e desenvolvimento para que a tecnologia estivesse disponível no momento necessário (PORTER, 1999).

Na década de 90, as empresas se tornaram os principais competidores por tecnologia. Na era da informação, os períodos de desenvolvimento tecnológico são menores, forçando com que as empresas apresentem resultados de aplicação imediata. Com isso, aumentam-se o interesse pela descoberta e antecipação do desenvolvimento tecnológico da concorrência, pois a informação sobre tecnologias emergentes é fundamental para introdução de novos produtos e melhorias em processos (PORTER, 1999, FARIA, 2001).

Dessa forma, desenvolveu-se a inteligência competitiva, cuja preocupação é identificar quem detém ou desenvolve quais conhecimentos, tecnologias, processos, produtos e mercados, para antecipar e auxiliar desenvolvimentos futuros em curto e médio prazo (BARANAUSKAS, 1998, FARIA, 2001). A inteligência competitiva neste contexto está associada à prospecção tecnológica e, através do monitoramento ambiental, auxilia a organização na atuação estratégica perante o mercado e a concorrência.

Sob outra perspectiva, a sociedade anseia por maior qualidade de vida e demanda ao Estado avanços tecnológicos que tragam prosperidade às suas vidas. Avanços nas áreas da saúde ou em áreas que geram emprego e renda beneficiam diretamente toda sociedade.

Na seção 2.2, são apresentadas definições de diversos autores para a prospecção tecnológica e, na seção 2.3, são apresentados os ganhos e algumas aplicações da prospecção tecnológica nas organizações e na sociedade como um todo.

2.2 Definições

Na literatura, encontra-se uma série de termos para prospecção tecnológica. No Brasil, utilizam-se os termos Prospecção, Prospectiva, Estudos do Futuro. Em inglês, os termos mais utilizados são *Forecast* e *Foresight*, que apesar de estarem associados à Prospecção, possuem semânticas levemente diferentes. *Forecast*, que pode ser traduzido como Previsão, está associada à prática de prospecção mais determinística, por exemplo, através de modelos matemáticos puros, enquanto que *Foresight*, traduzido como Visão, está associado a técnicas menos determinísticas como análise de cenários, opinião de especialistas, entre outras (MCT, 2009b).

Para COATES (1985) a atividade prospectiva está diretamente ligada ao planejamento. Ele define prospecção como um processo que leva a uma compreensão mais plena dos fatores que moldam o futuro de longo prazo e que devem ser considerados no planejamento, na formulação de políticas e na tomada de decisões.

HORTON (1999) define prospecção como um processo de desenvolvimento de visões de possíveis caminhos nos quais o futuro pode ser construído, entendendo que as ações do presente contribuirão com a construção da melhor possibilidade do amanhã.

MARTIN *et al.* (1998) definem prospecção tecnológica como um processo que sistemicamente examina o futuro de longo prazo da ciência e tecnologia, da economia e da sociedade, para identificar as áreas da ciência estratégicas e as tecnologias emergentes que tenham a propensão de gerar os maiores benefícios sociais e econômicos.

Segundo COELHO (2003), o resultado de um estudo prospectivo, de acordo com o grau de precisão, pode ser uma indicação sobre o futuro, uma indicação probabilística sobre o futuro ou uma indicação probabilística baseada em uma avaliação de possibilidades alternativas.

De acordo com SALLES FILHO *et al.* (2001), prospecção tecnológica pode possuir uma conotação próxima de predição, objetivando a definição de relações causais dos desenvolvimentos científicos para esboçar cenários probabilísticos do futuro. Ainda de acordo com autor, cada vez mais se entendem os desenvolvimentos futuros como consequência de múltiplos fatores e decisões, portanto, a prospecção não deve obedecer

somente a resultados técnicos, devendo levar em conta fatores de cunho político-sociais.

Prospecção tecnológica, segundo PORTER *et al.* (2004), é o processo de descrever as características e impactos de uma determinada tecnologia em algum momento no futuro. A prospecção tecnológica exige a compreensão da evolução da tecnologia, ou seja, o entendimento de como uma tecnologia se desenvolve, amadurece ou se extingue. Essa dissertação, particularmente, se apóia nessa definição.

2.3 Ganhos e Aplicações

A prospecção tecnológica pode ser aplicada em muitos cenários, através de diferentes técnicas e abordagens, e proporciona diversos ganhos para as organizações que a utilizam. Pode ser utilizada como um instrumento para auxílio na definição de políticas de investimento em ciência e tecnologia pelo governo, ou pode ser utilizada por uma organização com um instrumento de ganho de competitividade em um mercado disputado.

MENDES (2008) enumerou, de forma sintetizada, os principais ganhos da prospecção tecnológica:

- Entender as forças que orientam o futuro;
- Antecipar e entender o percurso das mudanças;
- Subsidiar e orientar o processo de tomada de decisão em ciência, tecnologia e inovação;
- Subsidiar decisões relativas ao estabelecimento de prioridades em P&D, gestão de risco das inovações tecnológicas, melhoria da competitividade tecnológica de produtos e processos;
- Organizar sistemas de inovação que correspondem aos interesses da sociedade;
- Identificar as oportunidades e as necessidades mais relevantes para a pesquisa no futuro.

A prospecção tecnológica está fortemente associada à estratégia, já que a antecipação suscita a ação e não existe sentido na ação sem finalidade específica (GODET, 2000). A falta da cultura de planejamento de longo prazo e da antecipação do futuro faz com que as organizações tenham problemas de decisão no seu cotidiano.

Nas seções 2.3.1, 2.3.2 e 2.3.3, são descritos alguns cenários de aplicação das técnicas de prospecção tecnológica, para investimento em ciência e tecnologia, para obtenção de vantagem competitiva e conquista de mercado e no caso específico de instituições de ensino e pesquisa.

Na seção 2.3.4, são apresentados os estudos realizados por uma iniciativa da Sociedade Brasileira de Computação (SBC), que buscaram enumerar os desafios futuros da área de computação. São destacados a aplicação da prospecção tecnológica e os ganhos oferecidos a esse tipo de estudo.

2.3.1 Investimento em Ciência e Tecnologia

A sociedade possui necessidades e vontades, demandando ao governo e ao mercado produtos que sejam capazes de satisfazê-la. A demanda por produtos pode implicar em demanda por tecnologia, uma vez que a tecnologia para desenvolvimento de determinado produto pode não estar disponível.

A sociedade, principal demandante por tecnologia, não é capaz de especificar o que deseja, pois não entende as possibilidades e nem possui familiaridade com a pesquisa científica.

Por outro lado, os pesquisadores não ofertam tecnologia que atenda todos os requisitos, pois não compreendem plenamente os problemas e as necessidades do demandante.

Existe um grande caminho a ser percorrido para que um produto inovador seja efetivamente disponibilizado para o mercado. O modelo de inovação linear apresenta as etapas de pesquisa básica, pesquisa aplicada, desenvolvimento, engenharia e produção, conforme pode ser verificado na Figura 1.

A etapa de pesquisa básica consiste na pesquisa de temas de grande relevância que poderão ser utilizados de maneira aplicada a diferentes problemas, servindo de base para a pesquisa efetivamente aplicada. Pesquisa aplicada, como o nome sugere, consiste na aplicação da pesquisa a problemas práticos. Os avanços, tanto da pesquisa básica como da pesquisa aplicada, são freqüentemente expostos em publicações ou conferências científicas.

Na etapa de desenvolvimento, os conhecimentos obtidos com a pesquisa aplicada tomam forma através de uma versão experimental do produto. Na etapa de engenharia, os produtos e processos são estruturados e melhorados, enquanto que, na etapa de produção, os produtos são disponibilizados em escala para o mercado. Os

conhecimentos desenvolvidos nessas etapas freqüentemente são formalizados em patentes industriais.

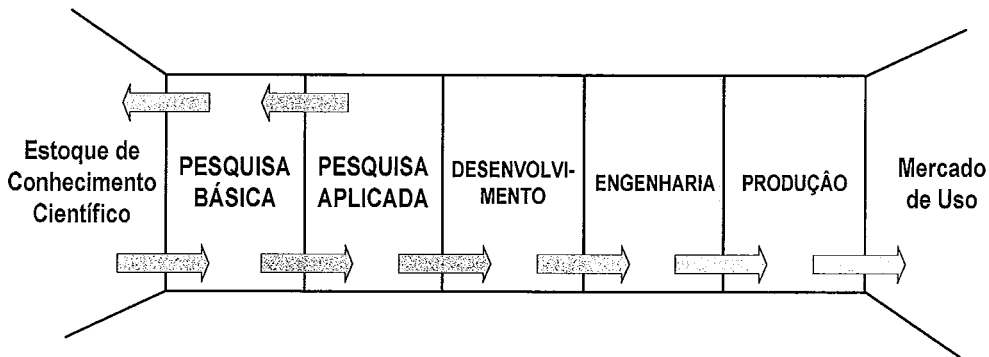


Figura 1: Modelo linear do processo de inovação (adaptado de CGEE, 2008)

O modelo de inovação paralelo, proposto por UTTERBACK (1994) e apresentado na Figura 2, apresenta a interação entre o mercado, demandante de tecnologia e os pesquisadores, responsáveis por desenvolver a tecnologia. A idéia para inovação nasce da junção da identificação de uma necessidade ou oportunidade, por parte do mercado, com a viabilidade técnica, verificada pelos pesquisadores e técnicos. Com a idéia amadurecida, pode ser necessário o desenvolvimento de tecnologia para suportar a produção, caso essa ainda não exista.

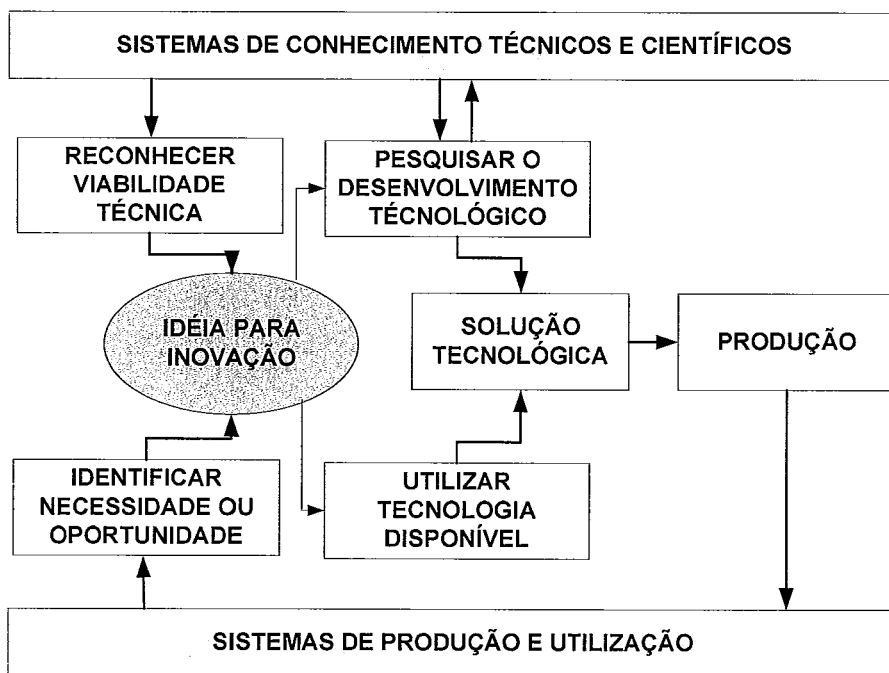


Figura 2: Modelo paralelo de inovação (adaptado de UTTERBACK, 1994).

Para adequar a oferta à demanda por tecnologia, o governo, as agências de fomento à pesquisa e as empresas, investem recursos no estudo de assuntos considerados estratégicos para o país, através da discussão com a sociedade, setor científico e produtivo.

A prospecção tecnológica pode auxiliar a formulação de políticas de fomento à pesquisa, identificando as áreas promissoras que podem ser do interesse do país. Com os investimentos ao longo do tempo, pode-se criar tecnologia necessária para suprir a demanda da sociedade.

2.3.2 Vantagem Competitiva e Conquista de Mercado

As organizações, atualmente, estão inseridas em um cenário de extrema competição. As organizações que possuem os melhores produtos, serviços e processos se sobressaem no mercado, enquanto que as outras são rapidamente extintas. A inovação tecnológica possui papel fundamental nesse contexto, pois permite a obtenção da vantagem competitiva frente aos concorrentes.

A prospecção tecnológica, através do monitoramento do ambiente externo, permite à organização visualizar as tendências científicas e investir recursos em pesquisas de temas promissores pertinentes às áreas de atuação. Permite inclusive, a construção de parcerias estratégicas com outras organizações que dominam ou se interessam em dominar tecnologias úteis aos seus negócios.

Com a prospecção, é possível antever adventos tecnológicos relevantes que podem ser utilizados pela concorrência ou que podem representar oportunidades se utilizados pela organização. A prospecção é, portanto, um instrumento de extrema valia para a formulação de estratégias das organizações que almejam obter vantagem competitiva e conquistar mercado.

2.3.3 Instituições de Ensino e Pesquisa

Instituições de Ensino e Pesquisa (IEP), assim como organizações comuns, estão inseridas em um mercado competitivo e tem como principal objetivo satisfazer plenamente seus clientes, que podem ser seus alunos, empresas e instituições parceiras, o governo e a sociedade. As IEP competem entre si basicamente por recursos financeiros e humanos. As instituições de destaque atraem mais investimentos do governo, empresas e órgãos de fomento, além de captarem os melhores alunos, técnicos e professores (KAWAMURA, 2006).

A reputação dessas instituições, que as fazem serem consideradas superiores à concorrência, decorre da percepção da sociedade, do mercado e dos órgãos governamentais responsáveis por avaliar a qualidade das mesmas.

No Brasil, a avaliação oficial desses órgãos baseia-se fortemente na capacidade das instituições desenvolverem pesquisa, isto é, na quantidade e qualidade de publicações em periódicos e conferências científicas. As instituições melhores avaliadas recebem mais recursos e investimentos governamentais e melhoram sua imagem no mercado.

A prospecção tecnológica auxilia as IEP a direcionarem seus esforços em temas considerados promissores pela comunidade científica. Esses temas, com tendência de crescimento, poderão representar grande espaço para produções científicas no futuro.

As IEP que se dedicarem a temas promissores terão grande vantagem competitiva frente às demais, pois serão as únicas no mercado a oferecer serviços de pesquisa, consultoria ou treinamento. Pelo mesmo motivo, terão acesso aos melhores recursos humanos do mercado, captando alunos, técnicos e professores mais qualificados.

2.3.4 Grandes Desafios em Computação no Brasil – SBC

Segundo PITAC (2005), a computação revolucionou o mundo científico ao ser considerada por alguns autores, junto com a teoria e a experimentação, como um dos pilares da pesquisa. A computação está relacionada com a pesquisa de várias áreas de conhecimento, participando diretamente das descobertas científicas associadas.

Dessa forma, o planejamento de longo prazo para a área de computação traz impactos para todas as áreas que fazem uso dela, sendo de extrema importância para o desenvolvimento da ciência como um todo (SBC, 2006).

Inspirada em iniciativas internacionais semelhantes, a Sociedade Brasileira de Computação (SBC) organizou, no ano de 2006, o Seminário de Grandes Desafios em Computação no Brasil, com o objetivo de listar as questões e assuntos de computação que terão grande importância para a ciência e para o país no longo prazo (período de 2006-2016). Esse seminário contou com a presença de renomados pesquisadores brasileiros, de diversas universidades, envolvidos diretamente com a pesquisa científica na área de computação.

O seminário, como resultado, deu origem a um relatório técnico que apresenta os cinco desafios destacados pelos pesquisadores (SBC, 2006):

1. Gestão da Informação em grandes volumes de dados multimídia distribuídos;
2. Modelagem computacional de sistemas complexos artificiais, naturais e socioculturais e da interação homem-natureza;
3. Impactos para a área da computação da transição do silício para novas tecnologias;
4. Acesso participativo e universal do cidadão brasileiro ao conhecimento;
5. Desenvolvimento tecnológico de qualidade: sistemas disponíveis, corretos, seguros, escaláveis, persistentes e ubíquos.

No contexto do tipo de estudo realizado, a prospecção tecnológica pode ser utilizada de duas formas: para auxiliar a definição dos desafios e para acompanhar a evolução e convergência da pesquisa científica com as metas e desafios propostos.

Na primeira forma, a prospecção tecnológica pode identificar as tecnologias emergentes que possivelmente estarão disponíveis no futuro e poderão modificar a forma como a população vive. Para listar os desafios da ciência para longo prazo, é importante que seja quebrado o paradigma das tecnologias atuais, e o futuro seja visto de uma forma diferente do presente. Em um futuro no qual as tecnologias disponíveis são diferentes das atuais, é possível que os anseios do país também sejam diferentes.

Na segunda forma, a prospecção tecnológica pode ser um meio de acompanhar a evolução das áreas científicas e das tecnologias relacionadas ao desafio proposto. Dessa forma, é possível direcionar investimentos em áreas importantes para a solução do desafio que por acaso não estejam evoluindo como esperado.

2.4 Conclusão

Este capítulo apresentou algumas definições existentes na literatura para prospecção tecnológica. Foi destacada a importância do tema e elencados alguns cenários de aplicação.

Uma vez definidos os propósitos de utilização, no próximo capítulo serão apresentadas algumas técnicas de prospecção. Ao final, serão apresentados cinco trabalhos correlatos a essa dissertação.

Capítulo 3 - Técnicas de Prospecção Tecnológica

Esse capítulo apresenta algumas técnicas de prospecção existentes na literatura. Essas técnicas foram agrupadas em cinco categorias ou famílias, cada qual apresentada em uma seção. Após apresentadas, são discutidos os pontos fortes e fracos de cada família de técnicas. Ao final, são listados alguns trabalhos correlatos à proposta dessa dissertação, disponíveis na literatura.

3.1 Introdução

Existe uma infinidade de técnicas de prospecção descritas na literatura. Elas se originam das mais diversas áreas de conhecimento e devem ser utilizadas de acordo com o problema que se propõem a resolver. Muitas vezes, opta-se pela aplicação de mais de uma técnica, para propiciar melhores resultados, ao invés da aplicação isolada de um único método.

Técnicas de prospecção se diferem em abordagens e em habilidades requeridas. Podem ser classificados como determinísticos, os métodos predominantemente quantitativos, numéricos ou empíricos, ou classificados como não-determinísticos, aqueles predominantemente qualitativos, baseados em conhecimento pessoais e julgamentos. Os métodos determinísticos são mais automatizados e dependem menos da interação humana, enquanto que os não determinísticos, mesmo apoiados por tecnologia, necessitam invariavelmente da interação humana.

Os métodos determinísticos ou quantitativos necessitam de séries históricas confiáveis e de dados preferencialmente estruturados, enquanto que os métodos não-determinísticos ou qualitativos são limitados ao conhecimento dos especialistas, suas experiências e preferências.

Uma classificação proposta por PORTER *et al.* (1991) divide os métodos de prospecção em cinco famílias: i) Opinião de Especialistas, ii) Monitoramento Ambiental, iii) Modelagem e Simulação, iv) Cenários e v) Análises de Tendências. As técnicas de Monitoramento Ambiental e Análise de Tendências devem receber uma atenção especial do leitor, já que são utilizadas na abordagem de prospecção tecnológica proposta por esse trabalho. As cinco famílias de técnicas são apresentadas nas seções a seguir.

3.2 Opinião de Especialistas

Essa família de métodos baseia-se no conhecimento, experiência, percepção e intuição de indivíduos com grande familiaridade com o assunto em questão.

Os métodos que usam a opinião de especialistas são qualitativos e devem ser usados sempre que a informação não puder ser quantificada ou quando não houver dados históricos disponíveis. Também deve ser utilizada combinada com outros métodos, principalmente para avaliar prospecções e resultados quantitativos.

Nas seções 3.2.1 e 3.2.2, são apresentados dois exemplos de técnicas freqüentemente utilizadas: Delphi e Painel de Especialistas. Na seção 3.2.3, são apresentados dois métodos para estímulo de criatividade, extremamente úteis para as abordagens qualitativas de prospecção.

3.2.1 Delphi

A técnica Delphi se disseminou no início dos anos 60 e tem como principal objetivo aprimorar o uso da opinião de especialistas (ESTES e KUESPERT, 1976).

Baseia-se na estruturação do conhecimento, da experiência e da criatividade de um grupo de especialistas e pressupõe que o julgamento coletivo, quando organizado adequadamente, é melhor que a opinião isolada de um indivíduo.

A técnica consiste no preenchimento iterativo, por especialistas no assunto estudado, de um questionário com questões quantitativas e qualitativas. Esse questionário é desenvolvido por uma equipe responsável por coordenar o processo. Existe garantia de anonimato dos respondentes do questionário.

Inicialmente, as perguntas são respondidas individualmente pelos especialistas e enviadas à equipe de coordenação para consolidação. Essa equipe tabula e realiza um tratamento estatístico simples nas respostas quantitativas e agrupa textualmente os argumentos e questões levantadas nas respostas qualitativas.

As respostas consolidadas são enviadas aos especialistas, que podem mudar suas avaliações à luz das respostas numéricas e das justificativas dos demais respondentes na rodada anterior. Os especialistas cujas respostas se diferirem muito das respostas dos demais, são convidados a justificar suas posições. O questionário novamente é enviado à equipe de coordenação, para nova consolidação e re-envio para os especialistas. O processo se repete até que a divergência de opiniões tenha se reduzido a um nível

satisfatório e a resposta da última iteração seja considerada oficialmente como a previsão do grupo.

O fato de não haver fisicamente reunião e o anonimato das respostas fazem com que seja menor a influência de atores psicológicos como, por exemplo, os efeitos da capacidade de persuasão, a relutância em abandonar posições assumidas e a dominância de grupos majoritários em relação a opiniões minoritárias (WRIGHT e GIOVINAZZO, 2000).

3.2.2 Painel de Especialistas

O painel de especialistas objetiva reunir pessoas com notório conhecimento da área estudada, tanto do meio acadêmico como do mercado. Devem investigar e estudar temas e dar conclusões e recomendações, tentando chegar a um consenso.

Os painéis podem ocorrer com os especialistas reunidos fisicamente ou, mais recentemente, com a evolução da tecnologia da informação, dispersos. Em ambos os casos, deve ser provida infra-estrutura necessária para apoiar os processos de discussão e de tomada de decisão.

Os painéis devem garantir uma representatividade equilibrada de todos os setores interessados, como, por exemplo, empresas de diferentes setores da economia, academia e governo (SANTOS *et al.*, 2004).

3.2.3 Métodos para Estímulo da Criatividade

Os métodos desse grupo têm o objetivo de evitar visões pré-concebidas dos problemas enfrentados, encorajando um novo padrão de percepção que amplia a visualização de futuros alternativos. Algumas características devem estar presentes naqueles que trabalham com técnicas que necessitam de criatividade. GUILFORD (1959) enumera algumas dessas características:

- Fluência: boa capacidade de expressão e associação de idéias;
- Flexibilidade: habilidade de abandonar formas antigas de pensar e buscar novas idéias;
- Originalidade: habilidade de ter idéias inovadoras;
- Percepção: habilidade de imaginar e perceber conexões e relações não óbvias;
- Vigor: motivação e força para concretizar as idéias;

A seguir são apresentados exemplos de métodos desse grupo.

3.2.3.1 Brainstorming

Um dos mais famosos métodos para auxiliar e estimular o processo de criação é o *Brainstorming* ou “Tempestade de Idéias”. Proposto por OSBORN (1953), o método objetiva a geração de idéias através de uma dinâmica, na qual um grupo de pessoas, a partir de um tópico ou problema alvo, sugere o maior número de idéias possível. Não são permitidas críticas ou justificativas, para que não exista inibição por parte dos participantes. Pensamentos e idéias incomuns são bem-vindos, já as idéias expostas por um participante podem despertar novas idéias em outros participantes. As idéias propostas são armazenadas e ao final do processo, são analisadas, selecionadas e aprimoradas.

3.2.3.2 Analogia Sistemática

O processo de analogia consiste na transferência de características originárias de dois domínios distintos em níveis similares de abstração. De acordo com LUHN e WETH (1999), a abstração obtida com a transferência de domínio, auxilia na reestruturação criativa do campo mental relacionado com o problema.

O processo para a geração de analogias parte da definição do problema. Algumas características são abstraídas e procura-se, então, transferir características do problema para áreas de analogia. Nesse processo, as características do problema são comparadas com características da área de analogia. Finalmente, são relatados pontos que ajudam na resolução do problema análogo e que também podem ajudar na resolução do problema proposto.

3.3 Monitoramento Ambiental

O monitoramento ambiental é um dos artifícios mais úteis para a prospecção tecnológica, pois fornece dados que são utilizados e complementados por outras técnicas. Ajuda, por exemplo, na definição de variáveis que devem ser incluídas em modelagens de sistemas e no fornecimento de dados para serem trabalhados nas técnicas de análises de tendências (FARIA, 2001). O monitoramento ambiental também é importante para a identificação de especialistas para participarem das técnicas qualitativas, como Delphi, painel de especialistas e prospecção por cenários (PORTER e DETAMPEL, 1995).

Para a prospecção tecnológica, o monitoramento ambiental coleta, analisa e valida dados sobre desenvolvimentos científicos e tecnológicos nas áreas de interesse do

estudo (FARIA, 2001). O monitoramento pode ser um estudo isolado de curto período ou pode ser um esforço contínuo e iterativo (HOFFMANN *et al.*, 1999). A base da informação deve ser exclusivamente pública, embora nem seja sempre gratuita. Como exemplos de dados que podem ser monitorados, podem-se listar registro de patentes, publicações científicas, projetos de pesquisa acadêmicos, lançamento de produtos no mercado, jornais e revistas, entre outros.

Segundo PORTER *et al.* (1991), alguns objetivos possíveis do monitoramento ambiental incluem:

- Identificar eventos técnicos, científicos ou sócio-econômicos importantes para a organização;
- Definir potenciais ameaças para a organização, implícitas nesses eventos;
- Identificar oportunidades para a organização de acordo com as mudanças no ambiente;
- Identificar a direção das tendências que estão convergindo, divergindo, ampliando, diminuindo ou interagindo.

As seções a seguir apresentam dois tópicos, relacionados com monitoramento ambiental, relevantes para o contexto desta dissertação: análise bibliométrica e mineração de dados.

3.3.1 Análise Bibliométrica

A bibliometria pode ser definida como a área científica que pretende quantificar os eventos e os processos de comunicação escrita, estudando a produção, disseminação e uso da informação registrada (VAN RAAN, 1997; FARIA, 2001). Segundo VAN RAAN (1997), uma das principais aplicações da bibliometria é a geração de indicadores de ciência e tecnologia, permitindo que os pesquisadores acompanhem a evolução das áreas científicas.

Dentre as informações registradas que podem ser analisadas para prospecção tecnológica, destacam-se as patentes e as publicações científicas.

Patente consiste na proteção concedida pelo Estado ao inventor, para que sua invenção não seja utilizada por terceiros. Essa proteção é temporária, variando de 15 a 20 anos de acordo com a modalidade da patente, e impede que qualquer outra pessoa, salvo com o consentimento do inventor, utilize-se comercialmente do bem protegido.

Como contrapartida, a invenção deve ser descrita para conhecimento público (OLIVEIRA *et al.*, 2005b).

Dentre as vantagens da análise bibliométrica de patentes, pode-se destacar a padronização e estruturação das informações em diferentes depósitos no mundo. As patentes depositadas em cerca de 70 países são classificadas detalhadamente segundo a Classificação Internacional de Patentes (CIP), que agrupa patentes em 8 seções, 21 subseções, 120 classes e 628 subclasses (MARTINS, 2008). Dessa forma, o levantamento de indicadores torna-se mais fácil e confiável.

Os periódicos e congressos científicos constituem em outra forma de divulgação de trabalhos. As universidades e empresas pelo mundo se utilizam desses meios para publicarem suas pesquisas. Cada meio possui uma reputação diferente, podendo ter abrangência regional, nacional ou internacional. No Brasil, a CAPES¹, com o objetivo de avaliar os programas de pós-graduação, classifica, através da iniciativa denominada WebQualis² (QUALIS, 2009), os meios científicos de publicação de acordo com a reputação dos mesmos.

Alguns pesquisadores da área de bibliometria estudaram os padrões das frequências de palavras nos textos científicos e propuseram uma série de teorias. Um estudo importante, que serviu como base para vários outros, foi realizado por Zipf, em 1949, que, de maneira empírica, constatou que, em um texto, obtém-se um valor constante ao multiplicar a frequência da palavra pela posição que ela ocupa em um ranking decrescente de frequência de palavras. O resultado desse estudo foi denominado Lei de Zipf (WORMELL, 1998).

Um estudo baseado na lei de Zipf, realizado por BOOTH (1967), constatou que, em textos escritos em linguagem natural, como uma publicação científica ou uma patente, as palavras podem ser separadas em quatro grupos de acordo com a frequência em que aparecem no texto (FARIA, 2001). A Figura 3 apresenta um gráfico que representa a frequência absoluta das palavras em um texto, onde as palavras aparecem ordenadas de forma decrescente por essa frequência. Os seguintes grupos podem ser identificados:

¹ Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) tem como missão avaliar, divulgar, formar recursos e promover a cooperação dos programas de pós-graduação brasileiros.

² A iniciativa WebQualis classifica os periódicos pela abrangência (regional, nacional ou internacional) e reputação (A, B ou C).

- Palavras-vazias: Este grupo contém palavras que não agregam nenhuma semântica importante ao texto, como pronomes, preposições, artigos ou palavras muito comuns no domínio de significado do texto.
- Informação trivial: Este grupo contém palavras de alta frequência que definem o tema central do texto.
- Informação útil ou interessante: Este grupo contém palavras de média frequência que definem temas periféricos. No caso de estudos de prospecção tecnológica, essas palavras podem definir temas potencialmente inovadores.
- Ruído: Este grupo contém palavras de baixa frequência que não possuem relevância para o texto. Nesse grupo encontram-se os erros de linguagem ou digitação.

Essa classificação sugere que, dependendo da aplicação da análise bibliométrica, os dados sejam tratados de forma que apenas as informações relevantes sejam consideradas, desconsiderando assim as palavras vazias e o ruído.

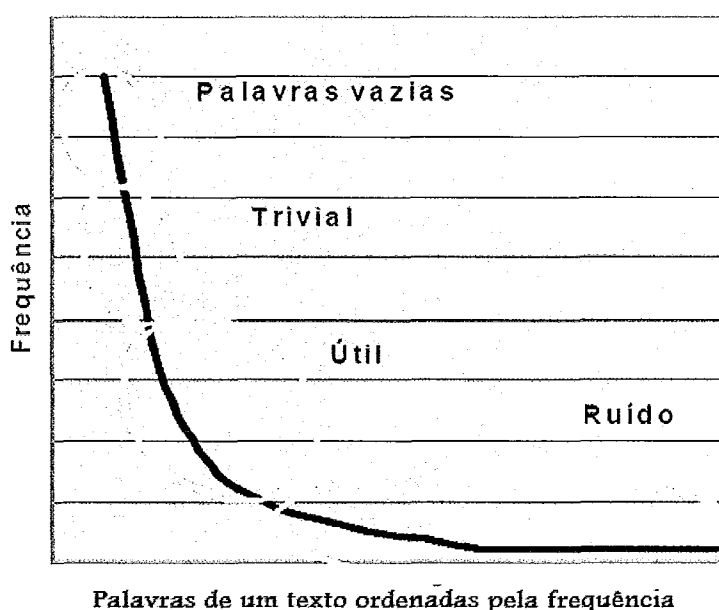


Figura 3: Histograma da frequência de palavras em um texto, com os grupos identificados (FARIA, 2001).

3.3.2 Mineração de Dados

Segundo WORMELL (1998), a disponibilização eletrônica de bases de dados, inclusive na internet, aumentou sensivelmente a oferta de dados sobre qualquer assunto. Porém, devido à grande quantidade de informação eletrônica disponível, a análise e síntese da informação tornaram-se mais complexa e motivaram o desenvolvimento de

metodologias de mineração e análise da informação auxiliada por computadores, objetivando a elaboração de informação de alto valor agregado para o tomador de decisão (FARIA, 2001).

O processo de mineração de dados ou *data mining* tem o objetivo de extrair conhecimento de grandes bases de dados, através de técnicas que procuram relações de similaridade ou discordância entre conjuntos de dados. Seu principal objetivo é encontrar e disponibilizar, de forma automática, padrões, anomalias, regras e outras informações ocultas em conhecimento útil para a organização (ZANASI, 2000).

Uma forma específica de mineração de dados realizada sobre textos, chamada de mineração textual ou *text mining*, é especialmente útil para a prospecção tecnológica, já que as informações científicas muitas vezes encontram-se sobre forma de texto na internet.

As técnicas de mineração textual, em um primeiro momento, preparam os textos para que, posteriormente, eles possam ser efetivamente analisados e minerados. A Figura 4 apresenta as etapas de preparação e análise da informação textual.

Na preparação do texto, podem ser destacadas as seguintes etapas:

- Retirada de *Stop Words*: são retiradas as palavras que não possuem significado para o texto, como artigos, preposições, advérbios, interjeições.
- Radicalização ou *stemming*: são retirados os sufixos das palavras, para que o processo de mineração textual trabalhe apenas com os radicais das palavras, que concentram seus significados.
- Tratamento semântico: para reduzir os problemas de homonímia e sinonímia dos textos minerados, pode ser utilizado um *Thesaurus*, que consiste em um dicionário de termos com suas inter-relações. Dessa forma, é possível unificar termos com o mesmo significado ou substituir termos por generalização da hierarquia – por exemplo, substituir “futebol” por “esportes”.

No processo de mineração textual, as seguintes análises podem ser destacadas:

- Categorização: através de processos de inferência, os textos podem ser automaticamente categorizados em classes previamente determinadas. Dentre as técnicas utilizadas para essa análise, podem ser citadas: árvores

de decisão, classificadores de Bayes, similaridades por vetores ou centróides, modelos conexionistas.

- Análise Lexicométrica: consiste na contagem dos radicais das palavras presentes no texto. A partir da lista ordenada da frequência dos termos, é possível verificar os radicais mais relevantes e conseqüentemente o tema central do texto.
- Sumarização: esse método apresenta as partes mais relevantes do texto sob forma de um resumo ou sumário.
- Agrupamento ou *clustering*: os textos que tratam de assuntos similares são armazenados em um mesmo grupo. A similaridade é verificada através da co-ocorrência das palavras mais freqüentes e relevantes dos textos.

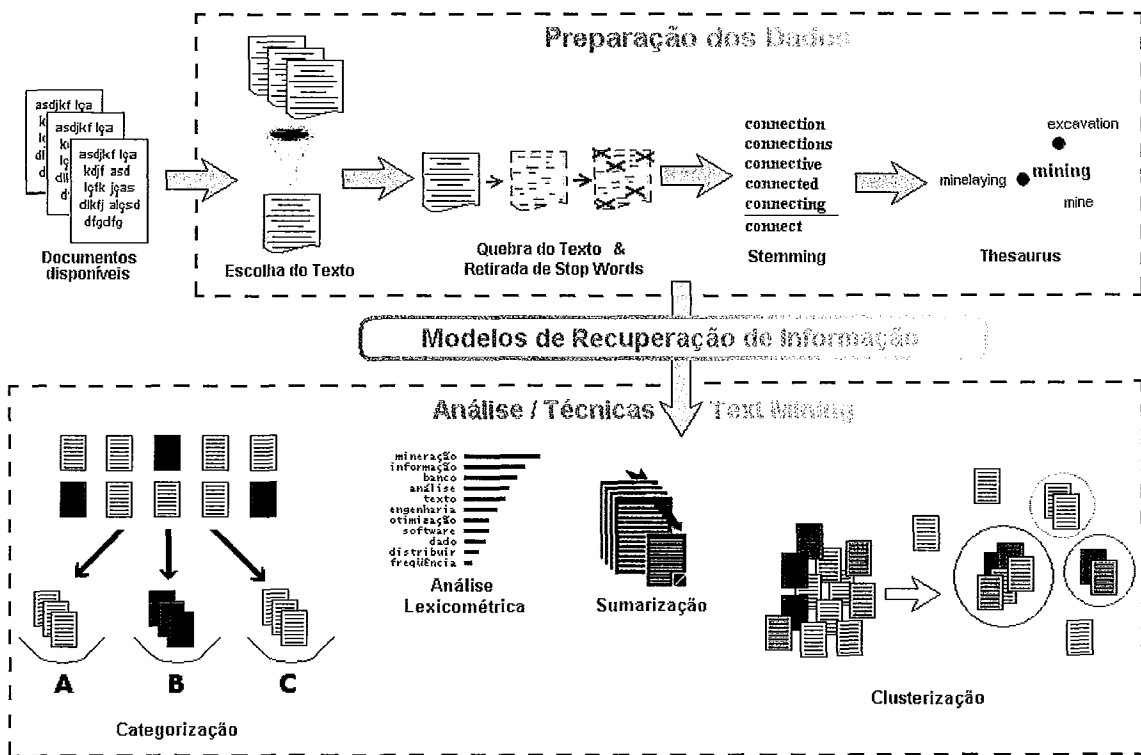


Figura 4: Técnicas de mineração textual (RODRIGUES, 2007).

Esses métodos agregam informação ao monitoramento ambiental devido à capacidade do processamento automático de grandes quantidades de dados e à inferência de informação não trivial e praticamente oculta aos olhos humanos.

3.4 Modelagem e Simulação

A modelagem e simulação computacional se baseiam em técnicas matemáticas associadas a programas de computadores, que objetivam imitar o funcionamento de sistemas ou processos do mundo real.

Segundo TAYLOR (1970), um sistema consiste em um conjunto de elementos que atuam e interagem com a intenção de alcançar um propósito lógico.

A partir da entrada de dados da operação do sistema, os modelos simulam os sistemas reais, oferecendo, como saída, o comportamento esperado pelo sistema.

Os modelos computacionais são construídos sobre equações matemáticas capazes de retratar a evolução temporal do sistema. Os elementos do sistema podem interagir e influenciar o comportamento do mesmo e servem de entrada para o modelo através das variáveis das equações.

SILVA (2009) enumerou as seguintes aplicações para a modelagem e simulação de sistemas:

- Descrever o comportamento dos sistemas modelados;
- Construir hipóteses e teorias, e considerando observações obtidas com os modelos;
- Evidenciar inter-relações entre os principais objetos do sistema;
- Prever o comportamento futuro dos sistemas, isto é, antecipar os efeitos produzidos por alterações em suas operações;
- Permitir ao analista analisar caminhos alternativos futuros ou que poderiam ter sido tomados.

Ainda de acordo com SILVA (2009), os modelos computacionais podem receber três classificações:

- Modelos voltados à previsão: A simulação pode ser usada para prever o estado de um sistema em algum ponto no futuro, com base no comportamento atual e passado;
- Modelos voltados à investigação: Busca de informações e desenvolvimento de hipóteses sobre o comportamento de sistemas. As variáveis de resposta servem para construir e organizar as informações sobre a natureza do fenômeno ou sistema sob estudo. Os experimentos recaem sobre as reações do sistema (modelo) a estímulos normais e anormais;

- Modelos voltados à comparação: Avaliar dos efeitos de mudanças sobre as variáveis de controle.

A modelagem e simulação computacional são muito utilizadas nas áreas de meteorologia, para previsão do tempo, e econômica, para diversos fins, como previsão de demanda, por exemplo.

Apesar da modelagem e simulação ser considerada uma técnica quantitativa e determinística, os modelos podem conter variáveis específicas para que especialistas possam influenciar o comportamento sistema com suas opiniões.

3.5 Cenários

A técnica de formulação de cenários leva em consideração a premissa de que o mundo real é complexo o bastante para que se possa descobrir seu eventual determinismo oculto através de técnicas puramente quantitativas. A incerteza, principalmente sobre a variável social, faz com que existam possibilidades de vários futuros distintos (GODET, 2000).

Cenários descrevem situações futuras e o conjunto de acontecimentos necessários para que se passe do momento atual para a situação futura (GODET, 2000). Não existe necessariamente um caminho único para a situação futura.

De acordo com COATES (2004) e GODET (2000), existem basicamente duas grandes categorias de cenários: exploratórios e antecipatórios. Os cenários exploratórios indicam o desdobramento das tendências futuras a partir do passado e do presente. Os cenários antecipatórios são construídos a partir de visões alternativas de futuros, indicando cenários desejáveis e cenários que devem ser evitados.

GODET (2000) enumera os objetivos fundamentais da prospecção utilizando cenários:

- Detectar as principais variáveis que caracterizam o sistema estudado;
- Determinar, a partir das variáveis identificadas, os principais atores e suas estratégias;
- Descrever, a partir de hipóteses, sob forma de cenários, a evolução do sistema, considerando as evoluções mais prováveis das variáveis e do comportamento dos atores.

GODET (2000) agrupou algumas técnicas de prospecção por cenários em uma “Caixa de Ferramentas”, fornecendo sistemas computacionais de apoio para a execução desses métodos.

Para identificação das variáveis do sistema, foi proposto por GODET (2000), o método de Análise Estrutural Multivariada, que será apresentado na seção 3.5.1.

Para determinar os atores e seus relacionamentos no sistema, foi proposto por GODET (2000), o método de Análise do Jogo de Atores, que será apresentado na seção 3.5.2.

Para formulação dos cenários, GODET (2000) propõe o levantamento de hipóteses utilizando Análise Morfológica, que será apresentado na seção 3.5.3.

3.5.1 Análise Estrutural Multivariada

Nesse método são identificadas, com a ajuda de especialistas, as variáveis mais importantes de um sistema que se quer estudar. A relação entre essas variáveis é ponderada, obtendo-se dessa forma uma matriz de análise estrutural. A partir da matriz de análise estrutural, são obtidas as relações de dependência diretas e indiretas entre as variáveis (GODET, 2000). A Figura 5 apresenta um exemplo das dependências diretas entre as variáveis de um sistema, enquanto que a Figura 6 apresenta as relações indiretas. As figuras foram obtidas com o software MICMAC (LIPSOR, 2009).

O entendimento das relações entre as variáveis serve para que o sistema seja mais bem compreendido e os possíveis cenários elencados sejam mais próximos da realidade.

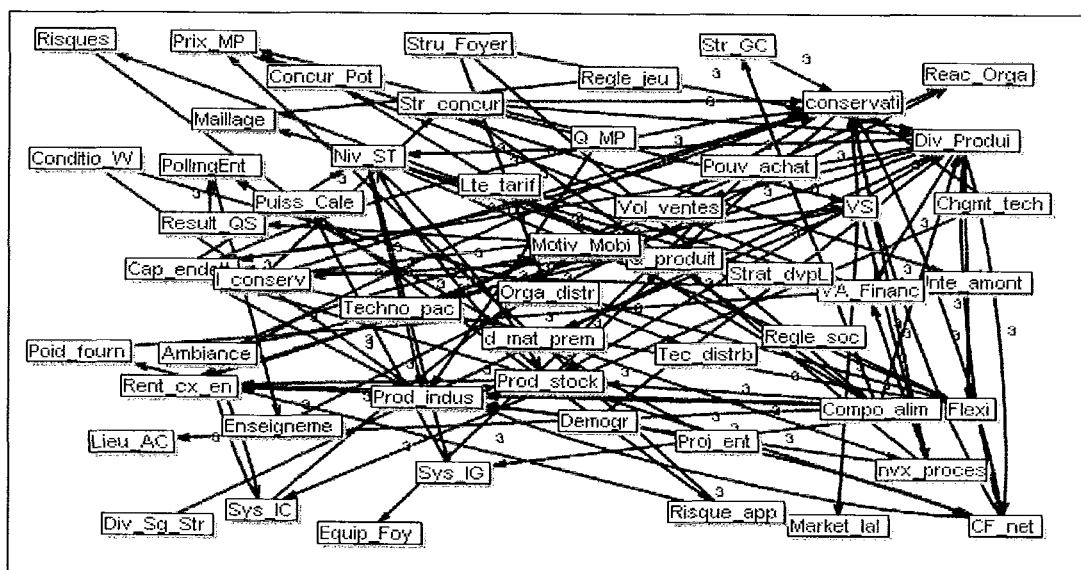


Figura 5: Dependência direta das variáveis através do software MICMAC (LIPSOR, 2009).

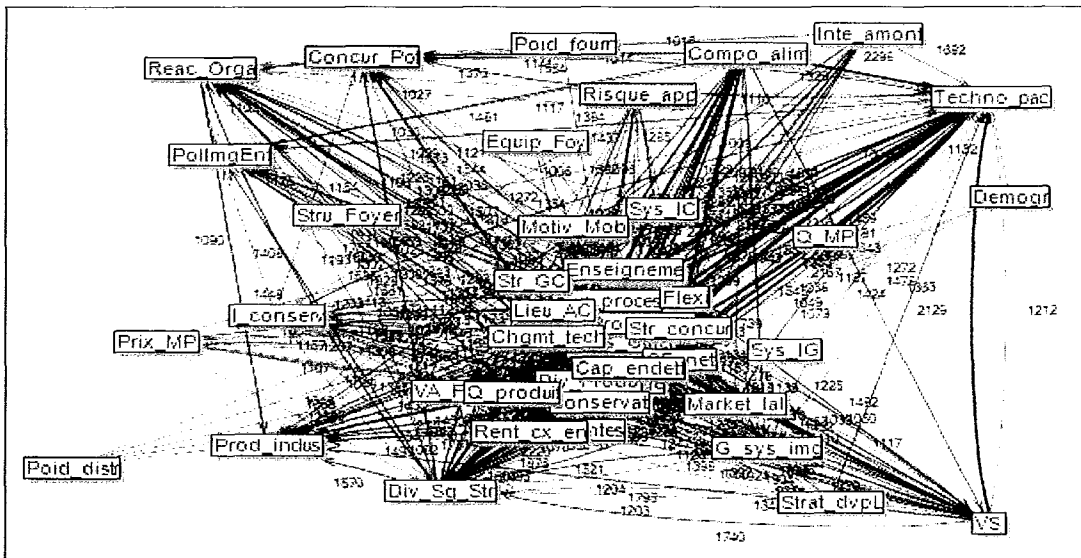


Figura 6: Dependência indireta das variáveis através do software MICMAC (LIPSOR, 2009).

3.5.2 Análise do Jogo de Atores

Nesse método, com ajuda de especialistas, são identificados os principais atores que influenciam as variáveis do sistema obtidas com a Análise Estrutural. Os atores não são considerados individualmente e sim pela classe que a pertencem. Como exemplo de ator, pode-se citar o Estado (nível municipal, estadual ou federal), determinado setor da economia ou da sociedade, entre outros (GODET, 2000).

Com a ajuda de especialistas e entrevistas com representantes dos próprios atores, são listadas os principais objetivos estratégicos de cada ator no sistema. O levantamento desses objetivos permite identificar a convergência e a divergência de interesses dos atores. GODET (2000) destaca que os atores, apesar de dificilmente divulgarem nas entrevistas suas estratégias, podem prever as estratégias de seus concorrentes.

A Figura 7 apresenta a matriz e o diagrama de convergência dos atores para os objetivos, respectivamente à esquerda e à direita. Cada linha da matriz representa um ator, enquanto que cada coluna é relativa a um objetivo. Os atores podem indicar concordância ou discordância com os objetivos, explicitado na matriz com os números +1 ou -1, ou podem indicar que não possuem envolvimento com o objetivo, explicitado com o número 0 na matriz. Os atores que concordarem ou discordarem com determinado objetivo, convergem em um ponto, conforme pode ser visto no diagrama de convergência.

	O1	O2	O3	O4	O5	Σ
A1	+1	+1	0	0	+1	3
A2	-1	0	+1	-1	-1	4
A3	-1	0	-1	+1	-1	4
A4	0	+1	+1	0	+1	3
A5	-1	0	-1	+1	-1	4
A6	0	0	0	0	+1	1

Acordo (Σ+)	+1	+2	+2	+2	+3
Desacordo (Σ-)	-3	0	-2	-1	-3
Posições	4	2	4	3	6

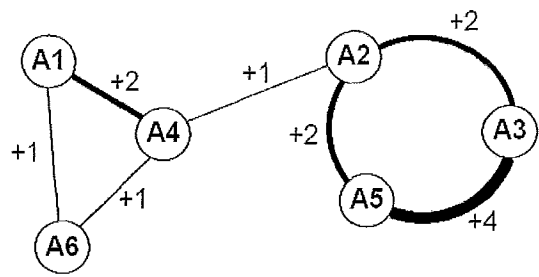


Figura 7: Matriz e diagrama de convergência de atores (adaptado de GODET, 2000).

3.5.3 Formulação de Cenários

Para a formulação dos cenários, os especialistas devem propor hipóteses baseadas nas tendências futuras, de acordo com as diferentes variáveis do sistema. As probabilidades de ocorrência dessas hipóteses devem ser ponderadas pelos especialistas.

GODET (2000) propõe que seja utilizada a técnica da Análise Morfológica para obtenção dos cenários possíveis. Essa técnica pretende combinar todas as hipóteses possíveis das questões-chave (ou variáveis) do sistema, formando, assim, os cenários possíveis. A Figura 8 exemplifica o uso da Análise Morfológica.

Os cenários obtidos podem ser classificados, não exclusivamente, em: cenários prováveis, cenários realizáveis, cenários desejados e cenários não desejados.

Os cenários realizáveis são aqueles cuja probabilidade de existência não é nula, enquanto que os cenários prováveis possuem alta probabilidade de ocorrerem. Os cenários desejados são aqueles almejados pela organização, que deve formular estratégias para que ele efetivamente se realize. Os cenários não desejados devem ser evitados de maneira pró-ativa pela organização.

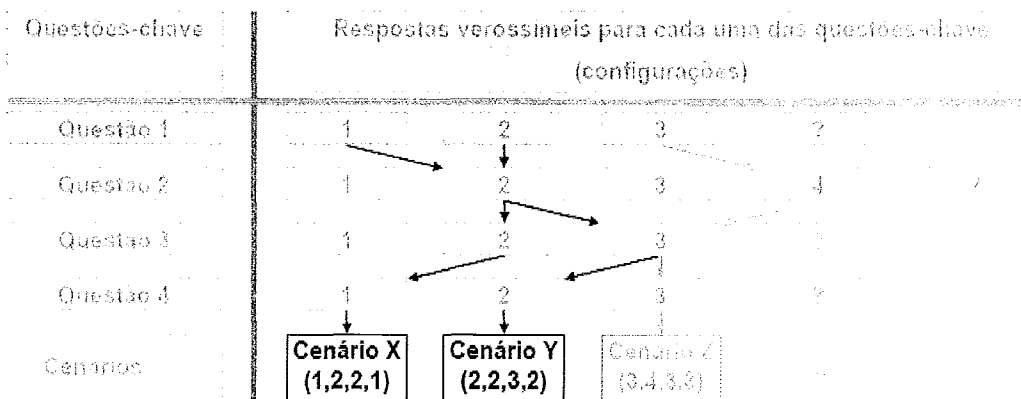


Figura 8: Análise Morfológica aplicada para formulação de cenários (GODET, 2000).

3.6 Análise de Tendências

A análise de tendências é um método de prospecção puramente quantitativo baseado na hipótese de que os padrões verificados no passado serão mantidos no futuro. A análise de tendências utiliza métodos matemáticos para extrapolar séries temporais.

A partir de pontos de uma série histórica real temporal, é obtida a função matemática que mais perto passa desses pontos. Essa função pode ser obtida através do método de mínimos quadrados, que obtém os parâmetros da função que propiciam o menor erro quadrático se comparados com os pontos reais da série (RUGGIERO e LOPES, 1996). Os pontos futuros são prospectados a partir da função obtida.

Por exemplo, a partir de n pontos (x_i, y_i) conhecidos de uma série histórica, onde $0 < i < n$, pode-se obter um polinômio de grau m que passa mais perto desses pontos. Seja o polinômio de grau m , onde $m < n - 1$, representado por:

$$P(x) = a_0 + a_1x + a_2x^2 + \dots + a_mx^m$$

Desse modo, a soma do quadrado da distância entre os n pontos reais (x_i, y_i) e os pontos do polinômio (x, y) , pode ser dado pela função q :

$$q = \sum (y_i - P(x_i))^2$$

A distância é elevada ao quadrado para que o sinal não influencie nos resultados, já que a distância é um número absoluto.

Deseja-se obter o valor mínimo da função q , isto é, o ponto onde a primeira derivada seja igual a zero:

$$\frac{\partial q}{\partial a_i} = 0 \quad \forall i = 0, 1, \dots, m$$

Dessa forma obtêm-se um sistema de $m+1$ equações. No caso específico da aproximação de um polinômio quadrático, representado por:

$$P(x) = a_0 + a_1x + a_2x^2$$

A minimização da função q origina um sistema com três equações:

$$\begin{cases} na_0 + \left(\sum_{i=1}^n x_i \right) a_1 + \left(\sum_{i=1}^n x_i^2 \right) a_2 = \sum_{i=1}^n y_i \\ \left(\sum_{i=1}^n x_i \right) a_0 + \left(\sum_{i=1}^n x_i^2 \right) a_1 + \left(\sum_{i=1}^n x_i^3 \right) a_2 = \sum_{i=1}^n x_i y_i \\ \left(\sum_{i=1}^n x_i^2 \right) a_0 + \left(\sum_{i=1}^n x_i^3 \right) a_1 + \left(\sum_{i=1}^n x_i^4 \right) a_2 = \sum_{i=1}^n x_i^2 y_i \end{cases}$$

A resolução do sistema acima fornece os valores dos parâmetros a_0 , a_1 e a_2 do polinômio $P(X)$.

A Figura 9 ilustra outro exemplo, onde, a partir de uma série histórica de dados de 2001 até 2008, obteve-se a função $f(x) = x^3 + 1$ que passava mais perto dos pontos reais. Os pontos futuros, relativos aos anos de 2009, 2010 e 2011 foram prospectados através da função obtida.

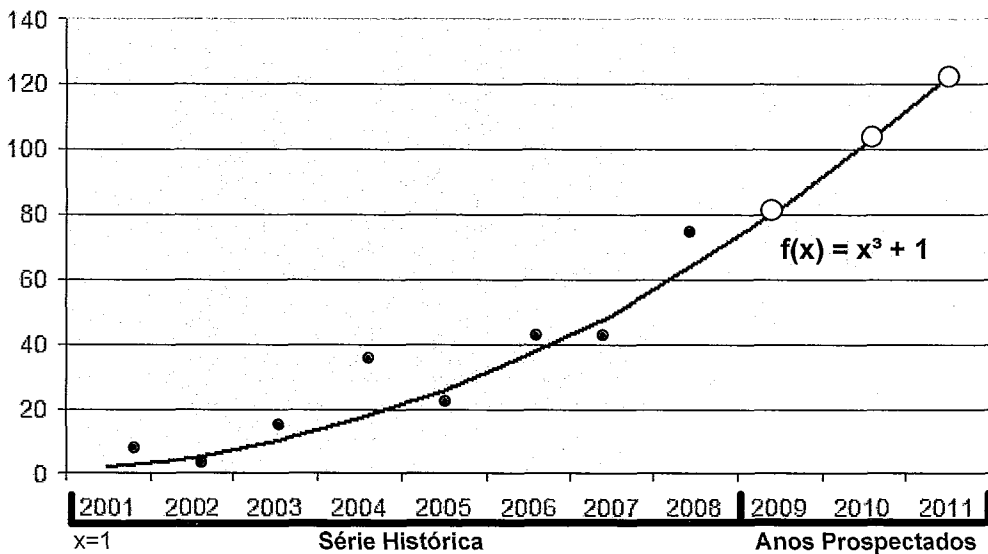


Figura 9: Exemplo da extrapolação de pontos futuros.

A função utilizada para interpolar os pontos pode ser de diversas classes: linear, polinomial, exponencial, logarítmica, entre outras. A classe de função escolhida, idealmente, deve ser próxima da forma com que os dados históricos se comportam. A Figura 10 apresenta exemplos de algumas classes de funções que podem ser utilizadas para interpolar os dados. Os parâmetros literais das funções podem ser obtidos através do método de mínimos quadrados.

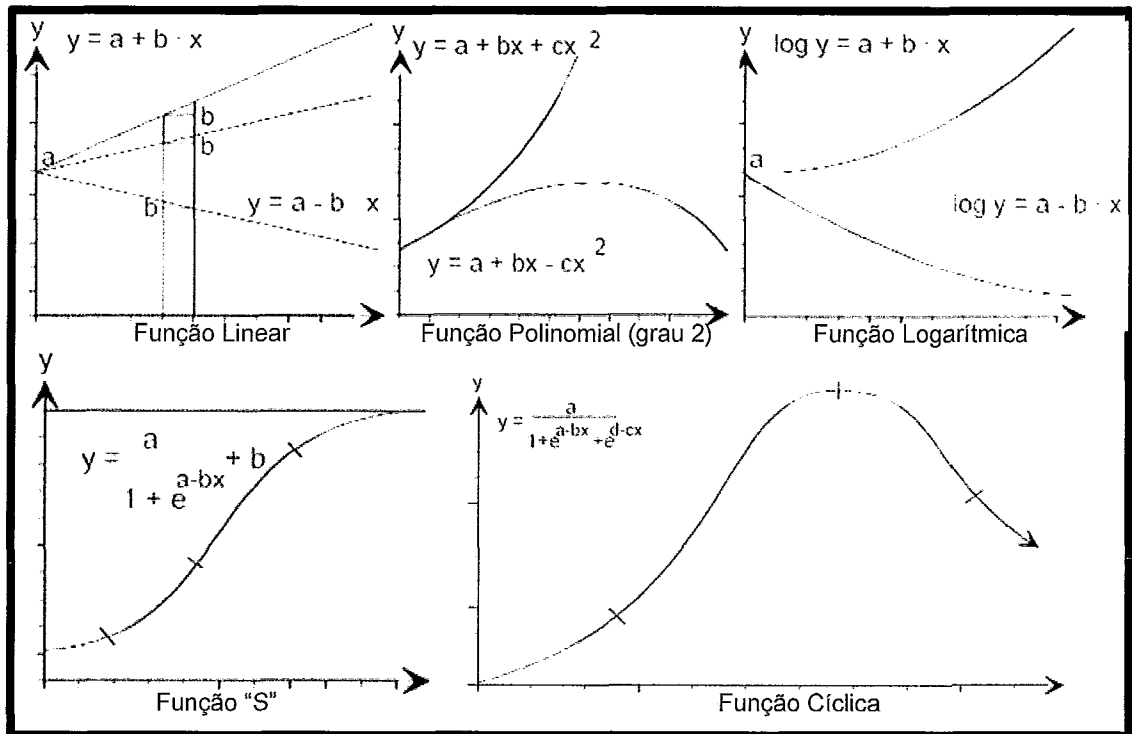


Figura 10: Exemplos de funções que podem ser utilizadas na interpolação.

3.7 Comparação entre as Técnicas

Todas as famílias de técnicas apresentadas, quando vistas isoladamente, possuem pontos fortes e fracos. Idealmente, elas devem ser utilizadas em conjunto de acordo com o problema que pretendem resolver.

As técnicas quantitativas podem fornecer dados importantes para subsidiar a decisão existente nas técnicas qualitativas, enquanto que as técnicas qualitativas podem avaliar e ampliar os resultados obtidos com as técnicas quantitativas.

A Tabela 1 apresenta a consolidação dos principais pontos fortes e fracos das técnicas apresentadas nesse capítulo.

Tabela 1: Consolidação dos pontos fortes e fracos das famílias de técnicas.

Família de Técnica	Pontos Fortes	Pontos Fracos
Opinião de Especialistas	<ul style="list-style-type: none"> — Utiliza-se de modelos e percepções internalizadas pelos especialistas, impossíveis de serem reproduzidas por outros métodos; — Permite que a intuição seja incorporada ao processo de prospecção. 	<ul style="list-style-type: none"> — Dificuldade de identificar especialistas em determinado assunto; — As projeções podem ser ambíguas ou divergentes entre especialistas de uma mesma área.

Família de Técnica	Pontos Fortes	Pontos Fracos
Monitoramento Ambiental	<ul style="list-style-type: none"> — Fornece grande quantidade e variedade de informação de diversas fontes; — Quando utilizado em conjunto com métodos computacionais, como a mineração de dados, pode fornecer informações não triviais aparentemente ocultas aos olhos humanos. 	<ul style="list-style-type: none"> — Pode resultar em excesso de informação não relevante; — As informações estão mais relacionadas ao passado e ao presente. Só a análise pode gerar perspectiva do futuro.
Modelagem e Simulação	<ul style="list-style-type: none"> — Podem fornecer excelentes análises do comportamento de sistemas complexos; — Algumas modelagens oferecem possibilidades de incorporação do julgamento humano. 	<ul style="list-style-type: none"> — Existem sistemas complexos demais para serem modelados; — O sucesso na previsão do comportamento histórico de um sistema não garante a boa previsão do futuro.
Cenários	<ul style="list-style-type: none"> — Apresentam retratos ricos em informação dos futuros possíveis; — Muitas vezes incorporam elementos que permitem ao tomador de decisão definir ações. 	<ul style="list-style-type: none"> — Algumas vezes, os cenários identificados não são realmente possíveis, pois há grande dificuldade em identificar todas as restrições e barreiras existentes.
Análise de Tendências	<ul style="list-style-type: none"> — Fornece previsões substanciais, que se baseiam em parâmetros quantificáveis; — Fornece resultados precisos no curto prazo. 	<ul style="list-style-type: none"> — Necessita de dados históricos consistentes coletados em um período razoável; — É vulnerável a descontinuidades da série e a mudanças bruscas; — Funciona apenas com parâmetros quantificáveis; — Não fornece bons resultados para projeções de longo prazo.

3.8 Trabalhos Correlatos

Essa seção apresenta cinco trabalhos correlatos a esta dissertação, encontrados na literatura. Foram pesquisados trabalhos sobre prospecção tecnológica ou identificação de especialistas que utilizem, como insumo, informações científicas.

Na literatura encontra-se um grande número de trabalhos se propõe a realizar prospecção tecnológica a partir de patentes. Muitos desses são aplicados em um domínio específico de tecnologia, como energia (POPP, 1998), nanotecnologia (COMPANO e HULLMANN, 2002), medicina (EILAND e GATLIN, 2008), entre outros.

ZHU e PORTER (2002) apresentam uma interessante abordagem para extração e visualização de informações para prospecção tecnológica. Através da importação de arquivos textuais estruturados relativos à produção científica, são apresentadas, em forma de mapas gráficos, as correlações entre assuntos e a proximidade da pesquisa de países, instituições e pesquisadores. Essas correlações são obtidas através do relacionamento dos termos contidos nas produções científicas. Para cada assunto são apresentadas a produção histórica e a tendência futura. Por fim, a abordagem fornece uma análise da migração de conceitos, listando os conceitos novos introduzidos a cada ano aos temas de pesquisa. Todas essas funcionalidades foram incorporadas a uma ferramenta comercial chamada *VantagePoint*³, comercializada pela empresa Search Technology, que também comercializa os arquivos estruturados com informações da produção científica, utilizados pelo *VantagePoint* (VANTAGEPOINT, 2009).

YOON e LEE (2008) apresentam uma abordagem para prospecção tecnológica a partir de dados da produção de patentes, levando em consideração o setor industrial ao qual a patente está relacionada. A abordagem, inicialmente, verifica o modelo de proteção intelectual adotado por cada um dos setores industriais, utilizando os dados da Community Innovation Survey⁴ (CIS, 2009). Posteriormente, nos setores que utilizam patentes como meio de proteção intelectual, são obtidos os dados relativos à produção de patentes da base da instituição USPTO⁵ (USPTO, 2009) e são prospectados os dados futuros por meio de extrapolação de uma função de crescimento, similar à “curva S”. É realizado um teste de regressão para medir o erro da extrapolação e, assim, verificar quais

³ A ferramenta *VantagePoint* existe desde o ano de 2000 e é vendida pelo valor de U\$\$7.500,00.

⁴ Community Innovation Survey (CIS) é uma série de pesquisas sobre inovação realizada pela EuroStat (órgão de levantamento de estatísticas da União Européia).

⁵ United States Patents and Trademarks Organization (USPTO) ou Organização Norte-Americana de Marcas e Patentes, é o órgão americano responsável por examinar e emitir patentes e marcas registradas.

setores respondem melhor à prospecção realizada. Por fim, a curva da produção de patentes é confrontada com a curva de investimentos em P&D e, para cada setor industrial, é mensurado o tempo médio decorrido entre o investimento realizado e o aumento de produção de patentes.

MARTINS (2008) apresenta um trabalho de prospecção tecnológica na área de Petróleo & Gás através da mineração de patentes industriais indexados na base da instituição USPTO. Em um primeiro momento, os resumos das patentes são obtidos, com auxílio de uma ferramenta computacional, para serem validados e separados manualmente em seis grupos. São determinadas, para cada grupo, a produção de patentes ao longo do tempo por classe (CIP) e por cessionário. Em um segundo momento, as patentes de cada um dos seis grupos são separadas em novas subclasses, através da ferramenta *RapidMiner* (RAPIDMINER, 2009), que implementa o algoritmo de agrupamento K-MEANS (MACQUEEN, 1967), e da ferramenta *PolyAnalyst* (POLYANALYST, 2009), que utiliza o algoritmo de agrupamento STC (WEINER, 1973). O autor apresenta as palavras mais relevantes de cada agrupamento obtido.

VELARDI *et al.* (2007) apresentam o *INTEROP KMap*, uma ferramenta computacional construída com o objetivo de fornecer um mapa de competências dos pesquisadores das 60 organizações europeias desenvolvedoras de tecnologia, participantes da *Network of Excellence* (CORDIS, 2009). Documentos de projetos, produções científicas, patentes, descrições de produtos e quaisquer textos técnicos relevantes das organizações são indexados na base de dados da ferramenta e recebem uma classificação automática. Essa classificação é inferida pela ferramenta com o uso de uma ontologia criada de forma semi-automática. A partir dos termos indexados na base de dados, a ferramenta procura a definição e correlação entre os termos na *web*, através da busca por padrões e expressões regulares. Essas definições e correlações formam a ontologia, que pode ser alterada ou validada pelos pesquisadores da rede. A ferramenta oferece uma série de possibilidades de busca por especialistas ou por documentos associados às competências.

RODRIGUES *et al.* (2006) apresentam uma ferramenta computacional, denominada *S-Miner*, capaz de identificar as competências de pesquisadores através da mineração de textos científicos. Especialistas em determinado domínio devem preencher um cadastro de palavras e competências relacionadas. Esse

cadastro é utilizado pelo *S-Miner* para inferir as competências associadas a cada texto científico e, com isso, inferir as competências dos pesquisadores autores do texto.

3.9 Conclusão

Nesse capítulo foram apresentadas algumas técnicas de prospecção tecnológica, divididas em cinco famílias: Opinião de Especialistas, Monitoramento Ambiental, Modelagem e Simulação, Cenários e Análise de Tendências. Todas as famílias de técnicas são comparadas através de seus pontos fortes e fracos.

Ao final do capítulo, foram apresentados cinco trabalhos correlatos a essa dissertação, relacionados com prospecção tecnológica ou identificação de especialistas.

No próximo capítulo será apresentada a proposta dessa dissertação: uma abordagem para prospecção tecnológica e identificação de especialistas, utilizando como insumo dados da produção científica de publicações, teses e patentes. Para suportar a abordagem, foi construído um ambiente computacional, denominado KAIROS, apresentado em detalhes.

Capítulo 4 - A Proposta: Uma abordagem para Prospecção Tecnológica e Identificação de Especialistas

Neste capítulo é apresentada a abordagem para prospecção tecnológica e identificação de especialistas. Simultaneamente, será apresentado o ambiente computacional desenvolvido para suportar essa abordagem, denominado KAIROS.

4.1 Introdução

Este trabalho propõe uma abordagem formada por heurísticas e métodos para prospecção tecnológica e identificação de especialistas. Para suportar a proposta, foi desenvolvido um sistema computacional, denominado KAIROS, integrado ao GCC (Gestão do Conhecimento Científico), um projeto mais amplo desenvolvido pelo Programa de Engenharia de Sistemas e Computação (PESC) do Instituto Alberto Luiz Coimbra de Pós-Graduação e Pesquisa de Engenharia (COPPE) da Universidade Federal do Rio de Janeiro (OLIVEIRA *et al.*, 2005a).

A partir da mineração de publicações científicas, teses e patentes, são descobertas as áreas da ciência associadas a cada uma dessas produções. Dessa forma, é possível estudar como uma área da ciência evoluiu ao longo do tempo e qual sua expectativa de comportamento no futuro. É possível estimar quais áreas são promissoras e determinar o quão alinhado estão os pesquisadores e instituições com essas áreas.

Através de heurísticas desenvolvidas neste trabalho, podem ser identificados pesquisadores e instituições especialistas ou visionários em determinada área. Foram considerados visionários os pesquisadores e as instituições que se anteciparam no estudo e no investimento de temas inovadores.

O KAIROS provê mecanismos inteligentes de busca em sua base de conhecimento. É possível consultar toda produção científica associada à determinada área e listar os pesquisadores ou as instituições que mais detêm conhecimento no assunto, considerados, portanto, especialistas.

A Figura 11 apresenta o fluxo de execução da aquisição e mineração de dados e do processamento de informações no KAIROS. Cada etapa especificada faz referência a uma seção contida nesse capítulo, responsável por apresentar a explicação em detalhes.

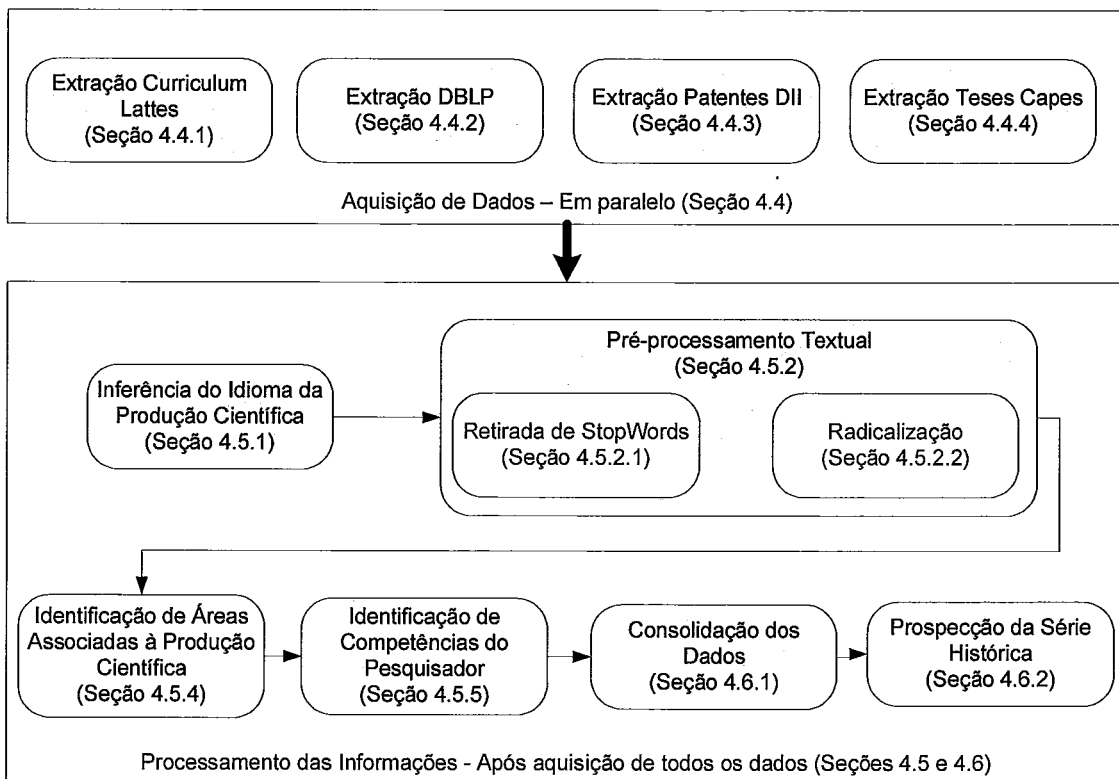


Figura 11: Fluxo de execução da aquisição de dados e processamento de informações.

Na seção 4.2, é apresentado o GCC, um sistema computacional desenvolvido com o objetivo de apoiar a gestão do conhecimento no ambiente científico. Conforme mencionado anteriormente, o KAIROS foi desenvolvido como um módulo do GCC.

Na seção 4.3, é apresentada a arquitetura em camadas do KAIROS, na qual se pode ter a visão geral de como o sistema foi concebido. Nessa seção, são especificadas as tecnologias de mercado utilizadas no desenvolvimento do KAIROS.

Na seção 4.4, são apresentadas as fontes de dados utilizadas na mineração da produção científica. Para cada fonte de dados, são apresentados o contexto, a importância das informações e o meio no qual as informações foram extraídas.

Na seção 4.5, são apresentados em detalhe o processo de identificação de áreas associadas à produção científica e a heurística de identificação de competências associadas aos pesquisadores e às instituições. A partir da identificação das competências dos pesquisadores e instituições é possível destacar os especialistas dentro de uma área de atuação.

Na seção 4.6, é apresentada a funcionalidade de mapeamento e prospecção das áreas da ciência, enquanto que na seção 4.7, é descrito o método de identificação da similaridade entre as áreas, na qual pode ser verificada a proximidade semântica ao

longo do tempo. A partir da prospecção das áreas, a seção 4.8 descreve a funcionalidade de identificação de áreas promissoras, em que são destacadas as áreas com maior tendência de crescimento. Na seção 4.9, é apresentada a funcionalidade desenvolvida para identificar o grau de alinhamento dos pesquisadores e instituições com as áreas da ciência consideradas promissoras.

Na seção 4.10, é apresentada uma heurística para identificação de pesquisadores e instituições visionários, enquanto que a seção 4.11 apresenta a busca por conhecimento na base KAIROS, que possibilita a identificação dos pesquisadores e instituições especialistas. A busca permite também a recuperação da produção científica associada à determinada área de conhecimento.

Por fim, na seção 4.12, é apresentado um estudo comparativo do KAIROS com outras propostas existentes na literatura, apresentadas no Capítulo 3.

No Capítulo 5, é apresentado um estudo de caso realizado com o objetivo de validar a abordagem do KAIROS, proposta por esse trabalho.

4.2 GCC - Gestão do conhecimento Científico

O GCC é um sistema “*web*”, desenvolvido inicialmente para atender ao público interno da COPPE e posteriormente expandido para atuar em todo cenário científico e acadêmico brasileiro. O GCC propicia a centralização de informações, facilitando a disseminação do conhecimento gerado. Além disso, esse ambiente disponibiliza toda a infra-estrutura para a criação e manutenção de comunidades virtuais de pesquisa e gerência de projetos, estimulando, assim, o desenvolvimento de novas idéias e a colaboração entre os pesquisadores (OLIVEIRA *et al.*, 2005a, OLIVEIRA, 2007). O GCC pode ser facilmente adaptado a outros contextos em que seja necessário gerir conhecimento complexo, como o conhecimento científico.

A Figura 12 apresenta uma visão geral da arquitetura do GCC. O KAIROS está inserido no módulo de Inteligência Competitiva, como um sub-módulo. Nas seções a seguir são apresentados todos os módulos do GCC.

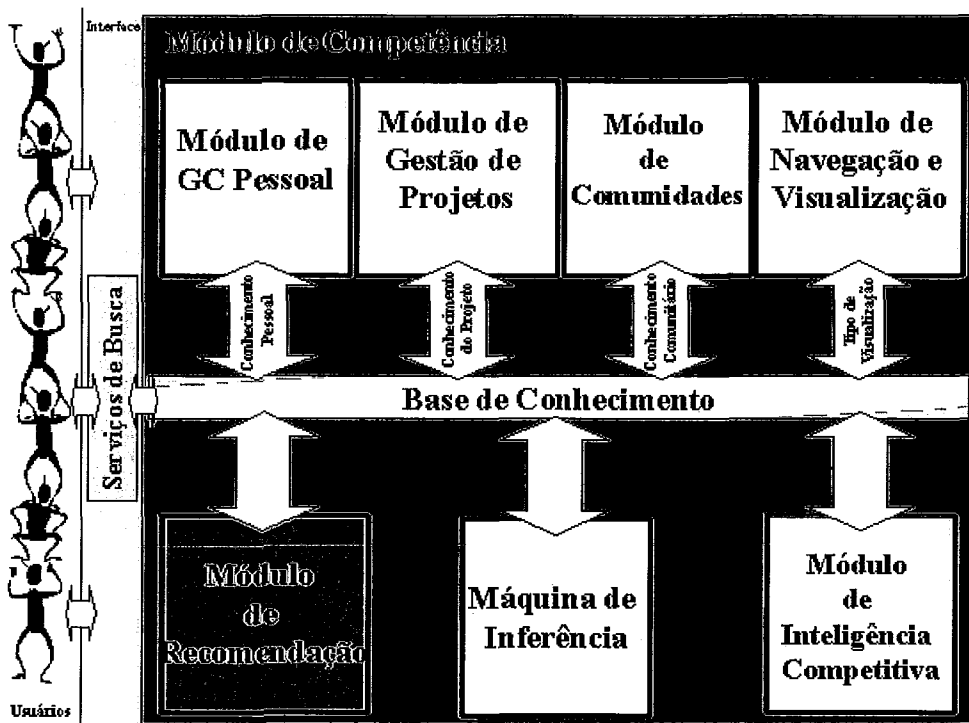


Figura 12: Arquitetura do GCC (OLIEIRA *et al.*, 2005a)

4.2.1 Módulo de Gestão de Conhecimento Pessoal

Este módulo provê funções para um pesquisador gerenciar seu próprio conhecimento. Destacam-se as seguintes funcionalidades:

- Curriculum vitae: o pesquisador pode cadastrar seu Curriculum Vitae completo, que será visível a todos os pesquisadores cadastrados no GCC. Pode também, declarar suas competências técnicas e atribuir, para cada uma, níveis distintos de conhecimento.
- Blog pessoal: Blogs podem ser vistos como páginas pessoais de conteúdo livre. No GCC, o blog representa mais uma ferramenta para propiciar a gestão e disseminação do conhecimento.
- Mapas mentais: Os pesquisadores podem construir mapas mentais que são utilizados para elucidar conceitos e simplificar as discussões realizadas entre pesquisadores geograficamente distribuídos.

4.2.2 Módulo de Gestão de Projetos

Neste módulo, são providos serviços para gerência, concepção e execução de projetos científicos. Os responsáveis pela coordenação de um projeto, através de uma ferramenta gráfica, criam o modelo de processo, com a seqüência de atividades e os insumos necessários para cada uma dessas atividades, como competências, documentos e outros artefatos.

A execução dos processos é controlada por uma máquina de workflow, que possibilita a criação de novos processos, o reuso de processos anteriores e a captura de conhecimento adquirido em atividades de processos já executadas (BOMFIM, 2005).

4.2.3 Módulo de Comunidades

Um dos focos de atuação do GCC é a utilização de comunidades que viabilizam o trabalho de forma colaborativa entre grupos de pesquisadores com interesses comuns. Esse módulo é composto por um conjunto de ferramentas para melhorar a interação entre as pessoas, como, por exemplo, fórum de comunicação, portal de novidades, votação, eventos, agendamento de encontros virtuais, entre outros. Permite comunicação rápida e fácil, propiciando colaboração assíncrona e síncrona e disseminação de informação e conhecimento entre membros de uma comunidade.

Todas as informações e conversações realizadas são armazenadas, já que representam uma forma de explicitação de conhecimento que pode ser consultada e analisada no futuro.

4.2.4 Módulo de Navegação e Visualização

Além das ferramentas usuais para visualização de informações, como relatórios e formulários, o sistema conta com ferramentas específicas desenvolvidas com propósito de facilitar a utilização do sistema, tornando o seu uso mais intuitivo. Destacam-se as seguintes ferramentas:

- **Árvore hiperbólica:** Esta é uma interface de visualização de uma estrutura hierárquica em árvore. Na árvore hiperbólica do GCC, é possível visualizar as áreas de conhecimentos propostas pelo CNPq. Um usuário pode navegar por essa árvore e consultar os projetos, as competências e outras informações associadas à área de conhecimento selecionada.

- Mapa conceitual do projeto: Provê a visualização do projeto como um mapa mental, facilitando a visualização das informações. Apresenta o conhecimento e seus relacionamentos de uma maneira visual e mais intuitiva, permitindo que o usuário interaja, navegue e acesse as informações e os conhecimentos.
- FBEE: A FBEE (Ferramenta de Busca de Especialista Através de Exemplo) foi desenvolvida por SOUZA (2008) e faz parte do submódulo denominado Busca de Substituto. A ferramenta provê busca de substitutos de contextos críticos de conhecimento na organização, através de técnicas de visualização de informação.

4.2.5 Máquina de Inferência

A máquina de inferência consiste em um conjunto de ferramentas construídas para apoiar o processo de identificação de competências dos pesquisadores. O SMiner (RODRIGUES *et al.*, 2006) – dispositivo capaz de minerar documentos criados, editados ou acessados pelos pesquisadores nas diferentes comunidades do GCC - apresentado no capítulo anterior, é uma das ferramentas que compõe a máquina de inferência.

O KAIROS construiu um método próprio de inferência de conhecimento que apresenta importantes vantagens em relação ao SMiner. Pelo fato do KAIROS ter sido construído de forma integrada ao GCC, outros módulos podem se beneficiar do método de inferência desenvolvido.

4.2.6 Módulo de Recomendação

O Módulo de Recomendação utiliza a máquina de inferência para descobrir a similaridade entre os conteúdos indexados no GCC e o perfil do usuário (preferências, comunidades e projetos). Dessa forma, o módulo pode recomendar documentos, pessoas, mapas mentais, comunidades e modelos de processos para novos usuários. Essas recomendações facilitam a interação do usuário com o sistema, já que as opções, provavelmente desejadas pelo usuário, são exibidas prontamente na tela inicial, sem que seja necessária a navegação pelo conteúdo do sistema (BARBOSA e MAIA, 2006).

Para realização da recomendação, foi utilizado o modelo em memória e o algoritmo de Correlação de Pearson (BREESE *et al.*, 1998).

4.2.7 Módulo de Inteligência Competitiva

Proposto por KAWAMURA (2006), o módulo de Inteligência Competitiva fornece diversas análises e indicadores sobre o ambiente interno e externo da organização. Os prováveis competidores são monitorados com intuito final de permitir que a organização obtenha vantagem competitiva frente às demais.

Destaca-se o submódulo Lattes, desenvolvido por DE MARTINO (2006), que minera informações do Curriculum Lattes dos pesquisadores da organização e de seus concorrentes, provendo uma série de consultas, comparações, relatórios e gráficos em diferentes granularidades: individual, por linha de pesquisa, setorial, departamental ou institucional.

4.2.8 Base de Conhecimento e Serviços de Busca

Todas as informações criadas e processadas no GCC, como processos, projetos, práticas, mensagens trocadas, características de pesquisadores e grupos, documentos, mapas mentais, entre outros, estão armazenadas na base de conhecimento. Esse conhecimento pode facilmente ser pesquisado e utilizado pelos usuários, através dos serviços de busca.

A base de dados KAIROS é completamente integrada à base de dados do GCC, fazendo com que outros módulos possam fazer uso das informações obtidas e inferidas pelo KAIROS.

4.3 Arquitetura do KAIROS

O KAIROS foi desenvolvido através de componentes dispersos, implementados em diferentes tecnologias. Esses componentes estão distribuídos dentro de uma arquitetura concebida em três camadas: camada de dados, de informação e de conhecimento, conforme pode ser verificado na Figura 13.

O fato dos componentes terem sido desenvolvidos de forma desacoplada permite que sejam utilizados ou adaptados facilmente em diferentes contextos. Ganhos de desempenho podem ser obtidos com a execução distribuída, em máquinas físicas diferentes, dos componentes com desempenho crítico.

As camadas da arquitetura e as tecnologias utilizadas para a sua construção serão apresentadas nas seções a seguir.

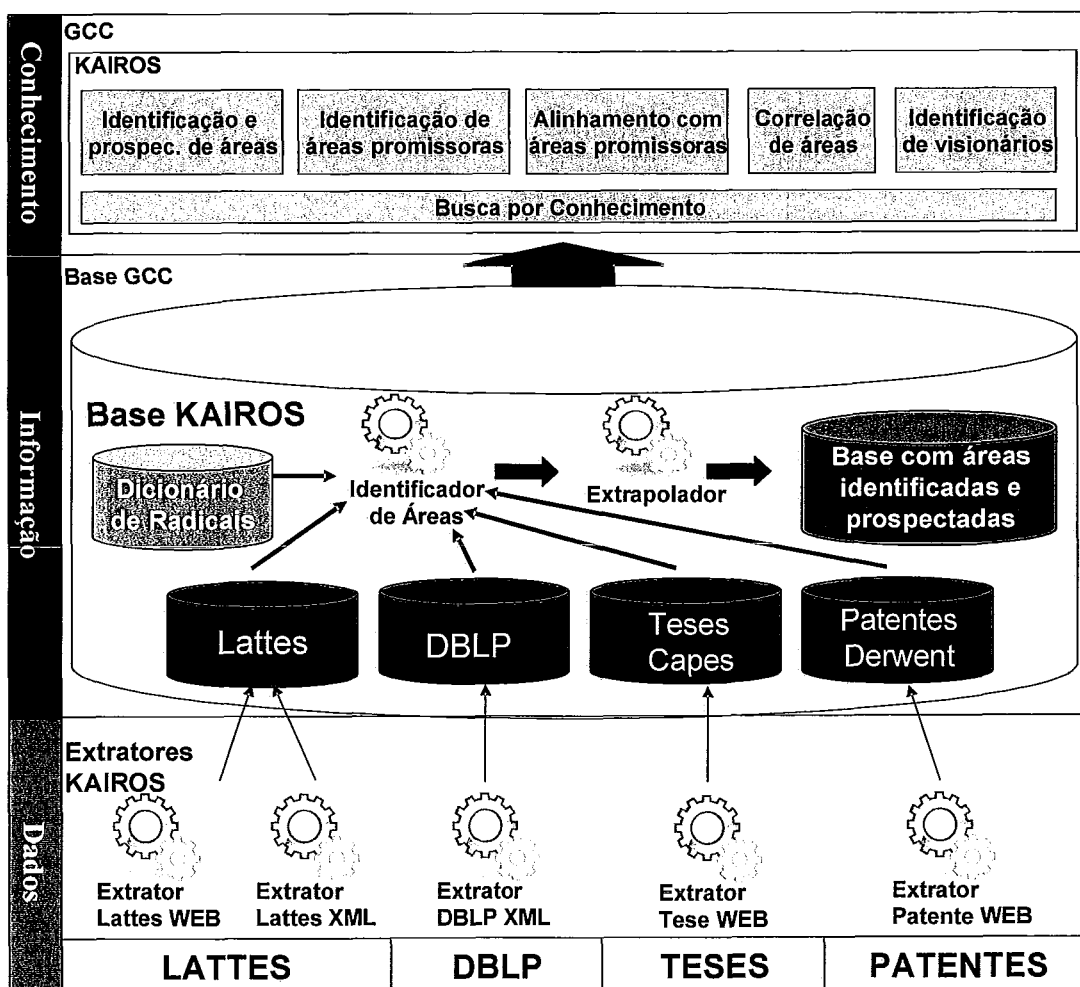


Figura 13: Arquitetura do KAIROS

4.3.1 Camada de Dados

A camada de dados provê mecanismos para captura e armazenamento dos dados que servirão de insumos para os estudos e inferências do KAIROS. As seguintes fontes são utilizadas: Curriculum Lattes, DBLP, Base de Teses da Capes e Base de Patentes da Derwent. Para cada fonte, existe pelo menos um aplicativo desenvolvido com o fim de extrair e armazenar os dados na base.

Com exceção do *Extrator Lattes Web*, que coleta informações sob demanda do usuário, todos os outros extratores adquirem grande quantidade de informação, consumindo considerável quantidade de recursos computacionais. Por esse fato, foram construídos de forma desacoplada, tornando possível a execução paralela em diferentes computadores.

A tarefa de extração de dados não é simples, pois, em geral, os dados estão distribuídos de forma não-estruturada na *web*, sem formatação nem padronização. As fontes de dados e os respectivos extratores serão apresentados em detalhe na seção 4.4.

4.3.2 Camada de Informação

A camada de informação tem como objetivo armazenar de forma estruturada e consolidada os dados provenientes da camada de dados. No apêndice A, é apresentado o modelo de dados da base KAIROS.

Nessa camada, através do *Identificador de Áreas*, que consome informações do *Dicionário de Radicais*, são associadas quais áreas da ciência estão relacionadas aos dados de produção científica. De forma similar, são identificadas as competências dos pesquisadores autores das produções científicas. O processo de identificação de áreas será explicado com detalhes na seção 4.5.

Além do tratamento de dados passados, os dados futuros são prospectados com ajuda do *Extrapolador*, que se utiliza de métodos matemáticos para estimar pontos futuros na série obtida com os dados históricos. Ainda nessa camada, os dados são consolidados e estruturados para otimizar o acesso e o consumo da informação pela camada de conhecimento. A consolidação e prospecção dos dados serão explicadas na seção 4.6.

4.3.3 Camada de Conhecimento

Essa camada tem como objetivo agregar conhecimento à informação armazenada na base de dados ao ser apresentada ao usuário final do KAIROS. Fazem parte dessa camada, relatórios e consultas configuráveis pelo usuário, completamente integrados ao GCC, no módulo de inteligência competitiva.

A Figura 14 apresenta a tela inicial do GCC, com destaque para o menu de acesso ao KAIROS. A tela inicial do KAIROS, onde o usuário escolhe a consulta que deseja trabalhar, é apresentada na Figura 15.

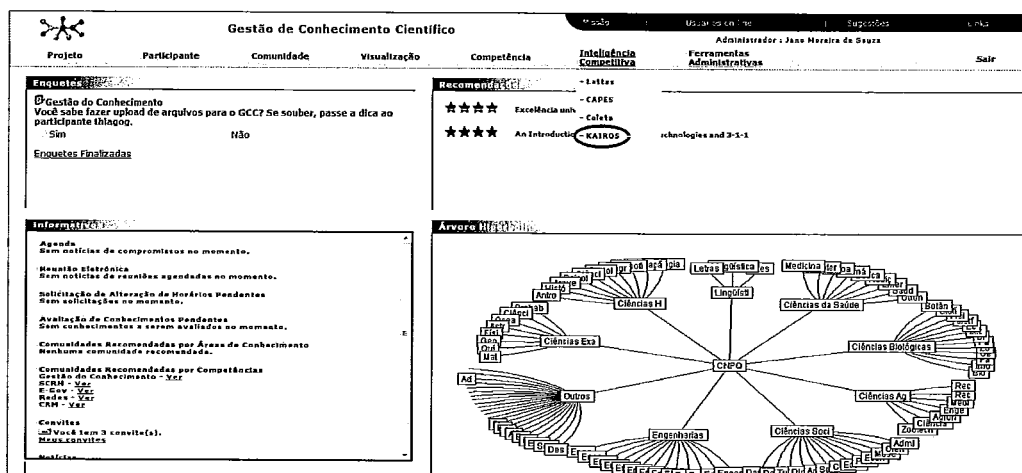


Figura 14: Tela inicial do GCC.



Projeto	Participante	Comunidade	Visualização	Competência
KAIROS				
Escolha o tipo de consulta:				
<ul style="list-style-type: none">• <u>Mapeamento e Prospecção das Áreas</u>• <u>Similaridade das Áreas</u>• <u>Identificação de Áreas Promissoras</u>• <u>Alinhamento de Pesquisadores e Instituições com Áreas Promissoras</u>• <u>Identificação de Pesquisadores e Instituições Visionários</u> • <u>Busca por Conhecimento (Áreas e Especialistas)</u>• <u>Dicionário de Radicais</u>				

Figura 15: Tela inicial do KAIROS.

4.3.4 Tecnologias Utilizadas

O KAIROS foi desenvolvido a partir de componentes desacoplados, que fazem uso de diferentes tecnologias. A escolha das tecnologias levou em conta duas questões fundamentais: integração e desempenho.

O GCC utiliza, predominantemente, as tecnologias ASP para desenvolvimento *web* e MS SQL Server para desenvolvimento da base de dados. Por esse motivo, todos os componentes do KAIROS integrados ao GCC também foram desenvolvidos com a mesma tecnologia. Por questões de desempenho, os demais componentes foram desenvolvidos em Java ou em procedimentos armazenados de banco de dados, conhecidos por *Stored Procedures*.

A Tabela 2 apresenta a tecnologia que cada componente da arquitetura foi desenvolvido.

Tabela 2: Tecnologias utilizadas pelo KAIROS.

Camada	Componente	Tecnologia
Dados	Extrator Lattes <i>Web</i>	ASP
	Extrator Lattes XML	Java
	Extrator DBLP XML	Java
	Extrator Tese <i>Web</i>	Java
	Extrator Patente <i>Web</i>	Java
Informação	Base KAIROS	MS SQL Server
	Identificador de Áreas	Procedimento MS SQL Server
	Extrapolador	Java
Conhecimento	Relatórios e Consultas	ASP

4.4 Fonte de Dados e Extratores

Para realização dos estudos e inferências, o KAIROS utiliza dados de quatro fontes: Curriculum Lattes (LATTES, 2009), DBLP (DBLP, 2009), Banco de Teses e Dissertações da Capes (CAPES, 2009) e Base de Patentes Derwent (DII, 2009). Para cada uma dessas fontes, foram desenvolvidos no KAIROS aplicativos com o objetivo de minerar e capturar os dados, chamados de Extratores. As seções a seguir apresentam cada fonte de dados, explicitando os dados importados e o método para extração.

4.4.1 Curriculum Lattes

O Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), agência do Ministério da Ciência e Tecnologia (MCT), desenvolveu no ano de 1999 o currículo eletrônico, denominado Currículo Lattes (LATTES, 2009), resultado da integração de esforços institucionais para a criação de um único instrumento de coleta de dados curriculares no país, cujo objetivo era eliminar a multiplicidade de cadastros assemelhados, reduzindo-se, de maneira significativa, o tempo gasto pela comunidade em atualizações repetitivas.

O currículo Lattes é um formulário eletrônico, que reúne informações sobre a vida científica e acadêmica de pesquisadores, estudantes, docentes, gestores, técnicos e profissionais ligados à ciência e tecnologia (C&T). Seus dados são utilizados para:

- Subsídio à avaliação da pesquisa e da pós-graduação brasileiras;
- Avaliação da competência de candidatos à obtenção de bolsas e auxílios;
- Seleção de consultores, de membros de comitês e de grupos assessores.

Os dados do Curriculum Lattes são de fundamental importância para a proposta do KAIROS, pois o Lattes é considerado referência nacional como repositório de dados acadêmicos.

4.4.1.1 Dados Importados

Um subconjunto de informações do Curriculum Lattes do pesquisador, considerado mais relevante, é importado para a base KAIROS:

- Dados do pesquisador (nome, endereço, bolsa de produtividade);
- Produção bibliográfica (título, ano, autores, meio, áreas de atuação);
- Formação acadêmica (nível, título, ano, palavras-chave);
- Atuação em projetos (título, descrição, co-autores, período);

- Orientações acadêmicas (aluno, nível, título, ano);
- Participação em bancas de concursos ou tese (aluno, nível, título, banca, ano);
- Linhas de pesquisa (nome da linha e palavras-chave);
- Área de atuação (grande área, área, subárea e especialidade);
- Prêmios de mérito científico (título, ano);
- Atuação profissional (nome da instituição que o pesquisador está associado)

Essas informações também são utilizadas pelo Módulo de Inteligência Competitiva do GCC (KAWAMURA, 2006, DE MARTINO, 2006). As produções bibliográficas são as informações mais utilizadas no KAIROS.

O Curriculum Lattes é a única fonte de dados utilizada que contém dados da instituição de pesquisa na qual o pesquisador faz parte.

4.4.1.2 Extração

O CNPq disponibiliza no sítio da plataforma Lattes, no formato HTML, os *Curricula* Lattes de todos os pesquisadores cadastrados. Para as instituições que os requererem formalmente, o CNPq disponibiliza um arquivo XML com os *Curricula* de todos os pesquisadores pertencentes à instituição.

Dessa forma, foram desenvolvidos dois extratores com o objetivo de importar dados do Curriculum Lattes: o *Extrator Lattes Web* e o *Extrator Lattes XML*.

Para extrair dos dados disponíveis abertamente na *web*, foi desenvolvido por DE MARTINO (2006) um extrator que processa o código HTML da página do Curriculum Lattes, convertendo-o em um documento XML estruturado. Esse XML, depois de validado, é processado e armazenado na base de dados relacional do KAIROS. O extrator importa um pesquisador por vez e recebe como entrada o nome do pesquisador que deve ser extraído.

Para extrair os dados pertencentes ao XML disponibilizado especificamente para uma instituição, foi desenvolvido um extrator capaz de ler e processar o formato definido pelo CNPq e, assim, armazenar os dados na base relacional do KAIROS. O arquivo importado contém os *curricula* de todos os pesquisadores pertencentes à instituição.

O *Extrator XML* tende a obter vantagem se comparado ao *Extrator Web*, pois o segundo é fortemente impactado por mudanças na página HTML do Curriculum

(mudanças de layout na página, por exemplo), enquanto não são esperadas mudanças no arquivo lido pelo *Extrator XML*. Portanto, para o correto funcionamento do *Extrator Web*, a manutenção do código do programa deve ser constante.

4.4.2 DBLP

Digital Bibliography & Library Project (DBLP) é um projeto mantido pela Universität Trier na Alemanha, que, até final de 2008, contava com uma base de mais de 1.000.000 de publicações da área de computação (DBLP, 2009).

O DBLP armazena dados relativos a publicações dos principais periódicos e conferências internacionais. Ao contrário do Curriculum Lattes, os pesquisadores não necessitam preencher e publicar seus dados, já que estes são enviados pelas editoras e conferências ao DBLP, e são publicados automaticamente de forma transparente aos autores. Essa característica faz com que o DBLP seja uma fonte rica de informações para o KAIROS, já que contém dados de pesquisadores de todo o mundo.

O DBLP desenvolveu um método funcional para identificar corretamente os autores que possuem diferentes assinaturas em suas publicações, provendo tratamento adequado para correta associação das publicações aos seus autores (LEY, 2002).

4.4.2.1 Dados Importados

As informações do DBLP relativas às publicações em periódicos e conferências são importadas para o KAIROS. São elas:

- Título da publicação;
- Meio da publicação (periódico, livro ou conferência);
- Local da publicação;
- Autores da publicação;
- Ano da publicação.

4.4.2.2 Extração

A equipe do DBLP disponibiliza periodicamente no sítio do projeto na *web*, todos os dados da base em um arquivo estruturado no formato XML. É fornecido também, um aplicativo de código aberto chamado DBLP Browser (KLINK *et al.*, 2004), apto a ler o arquivo XML e fornecer uma série de consultas e relatórios.

O extrator construído para importar os dados para a base do KAIROS utilizou-se de uma adaptação do código do DBLP Browser para importação do arquivo XML. Os

dados são mapeados, validados e armazenados nas tabelas da base de dados do KAIROS.

4.4.3 Banco de Teses e Dissertações Capes

A Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), subordinado ao Ministério da Educação (MEC), mantém o Banco de Teses e Dissertações, com o objetivo de facilitar o acesso a informações sobre teses e dissertações defendidas junto a programas de pós-graduação do país. O Banco de Teses faz parte do Portal de Periódicos da Capes/MEC (CAPES, 2009).

A Capes disponibiliza em seu portal, uma ferramenta de busca que consulta teses e dissertações defendidas a partir de 1987. As informações são fornecidas diretamente a Capes pelos programas de pós-graduação, que se responsabilizam pela veracidade dos dados. A ferramenta permite a pesquisa por autor, instituição, período, nível (mestrado ou doutorado) e assunto.

Essas informações são importantes para a proposta do KAIROS, uma vez que representam resultados de pesquisas desenvolvidas pelas universidades brasileiras, podendo indicar tendências e desafios locais.

4.4.3.1 Dados Importados

São importadas para a base KAIROS, as informações relativas às teses e dissertações das universidades brasileiras:

- Título do trabalho;
- Autor do trabalho;
- Orientadores do trabalho;
- Banca examinadora;
- Assuntos principais do trabalho;
- Palavras-chave do trabalho;
- Resumo do trabalho;
- Nível (mestrado acadêmico, mestrado profissionalizante ou doutorado);
- Ano de defesa do trabalho.

4.4.3.2 Extração

Foi desenvolvido por ABREU (2008), um extrator que acessa o sítio na *web* da Capes para importar os dados do Banco de Teses para a base do KAIROS. As informações são retornadas no formato HTML, sem estruturação alguma.

Como entrada, o extrator recebe condições de consulta disponíveis para filtro no sítio *web* da Capes, gravando os dados que satisfazem tais condições na base. O método de mineração utiliza expressões regulares para busca de atributos retornados na página HTML.

4.4.4 Base de Patentes Derwent Innovation Index (DII)

O Derwent Innovation Index (DII, 2009) é uma base de dados de patentes das áreas de química e engenharia, mantida pela empresa Thomson Reuters. Existem registros de patentes desde o ano de 1963, com referência aos depósitos de inúmeros países.

Patentes são resultados de pesquisas tanto de universidades quanto de empresas, e podem ter aproveitamento comercial. Pode-se dizer que a patente está presente no caminho entre a pesquisa teórica e o produto comercial. Os temas aos quais as patentes são relacionadas podem sugerir as tendências das pesquisas que são melhores absorvidas pelo mercado.

4.4.4.1 Dados Importados

São importadas para base KAIROS, as seguintes informações referentes às patentes da base Derwent:

- Número da patente;
- Título da patente;
- Resumo da patente;
- Área de conhecimento da patente (Classificação Derwent);
- Classificação internacional da patente (CIP);
- Inventores da patente;
- Data de publicação da patente;
- Data de registro da patente.

No KAIROS, o ano da patente considerado nas consultas é o menor ano entre o ano da publicação e o ano do registro da patente.

4.4.4.2 Extração

Os dados de patentes estão disponíveis no sítio da Derwent na *web*, através de consulta via formulário eletrônico. A consulta retorna os atributos da patente no formato HTML não-estruturado.

Foi desenvolvido por ABREU (2008) um extrator que acessa o sítio na *web* da Derwent para importar os dados das patentes para a base do KAIROS. Como entrada, o extrator recebe a área de conhecimento e o ano das patentes que devem ser importadas e, como resultado, grava as patentes que satisfazem tais condições na base de dados.

O extrator faz sucessivas requisições, uma para cada patente, ao formulário do sítio da Derwent e, para cada registro de patente obtido, procura expressões regulares para identificar os atributos da patente na página HTML. Dessa forma, consegue mapear e armazenar os dados na base KAIROS.

4.5 Identificação de Áreas

O processo de identificação de áreas tem como objetivo inferir as áreas de conhecimento associadas às produções científicas mineradas e, dessa forma, descobrir como essas áreas se desenvolveram ao longo do tempo e como elas tendem a se desenvolver no futuro.

Com a identificação dessas áreas é possível inferir as competências dos autores desses trabalhos. Como um pesquisador pode produzir uma série de trabalhos científicos ao longo de sua vida, em diferentes áreas, é proposto um método heurístico para inferir as competências do pesquisador, levando em conta as áreas de seus trabalhos e fatores como tempo e importância da produção.

O processo se dá através de mineração textual. Cada produção científica possui informações relevantes para serem mineradas. Essas informações devem sintetizar o conhecimento contido na produção científica.

Primeiramente, caso não exista a informação sobre o idioma da produção científica, a mesma deve ser inferida. Depois, ocorre um pré-processamento textual que limpa e prepara as informações para serem processadas na etapa de “Inferência das Áreas da Produção Científica”. Essa etapa utiliza o Dicionário de Radicais para associar o radical minerado à área de conhecimento. Por fim, as competências do pesquisador são inferidas. A Figura 16 apresenta, graficamente, o processo de identificação de áreas, que será detalhado nas seções a seguir.

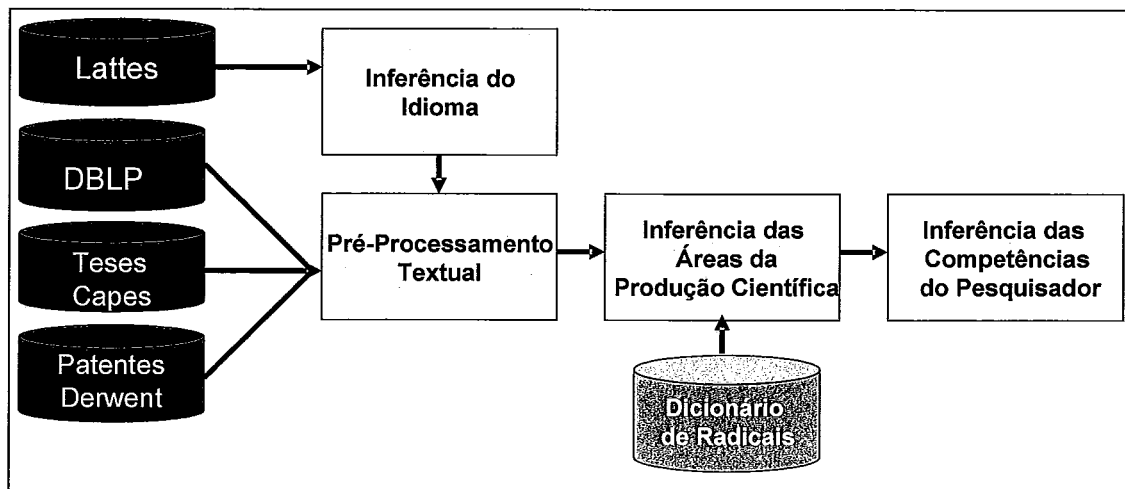


Figura 16: Processo de identificação de áreas e competências

4.5.1 Inferência do Idioma

O processo de identificação de áreas foi desenvolvido para os idiomas inglês e português, pois esses idiomas abrangem quase que a totalidade das produções científicas relevantes para os possíveis usuários do KAIROS.

As patentes da base Derwent e as publicações do DBLP estão em sua totalidade no idioma inglês, enquanto que as teses e dissertações da Capes estão no idioma português. As produções científicas cadastradas no Curriculum Lattes dos pesquisadores podem estar em inglês ou português, devendo o pesquisador opcionalmente cadastrar de forma manual o idioma.

Quando a informação do idioma não for explicitamente cadastrada, o KAIROS infere essa informação a partir da contagem de palavras nativas do idioma que, por si só, não possuem significado – chamadas de *stop words*, conceito detalhado na seção 4.5.2.1. Isto é, a partir de uma lista de palavras pré-definidas, como artigos, preposições, pronomes e advérbios, em ambos os idiomas, são contados o número dessas palavras que aparecem no texto que deve ser minerado. O idioma que tiver mais dessas palavras (*stop words*) contidas no texto, é considerado o idioma da produção científica.

4.5.2 Pré-Processamento Textual

A etapa de pré-processamento textual tem o objetivo de preparar o texto para a inferência. Nessa etapa, são retiradas as palavras frequentes ou *Stop Words* do texto e os sufixos das palavras. Como resultado, é obtido o conjunto dos radicais que serão efetivamente processados.

4.5.2.1 Retirada de Palavras Frequentes ou *Stop Words*

As *Stop Words* são palavras comuns e frequentes que, sozinhas, não apresentam significado relevante para o texto. As *Stop Words* variam para cada idioma, mas, normalmente, são artigos, preposições, pronomes, advérbios, locuções, algarismos isolados, entre outras palavras. A lista completa das *Stop Words* utilizadas pelo KAIROS pode ser obtida no Apêndice B.

Por não trazerem vantagens semânticas à identificação de áreas, nessa etapa essas palavras são eliminadas do texto. A Figura 17 exemplifica o processo de retirada das *Stop Words* de um texto.

... Na maioria das vezes os documentos retornados pelas ferramentas de recuperação de informações envolvem um contexto mais amplo fazendo com que o usuário tenha que garimpar, ou seja, especificar ou filtrar estes documentos, o que demanda tempo e conhecimento a fim de obter as informações que ele realmente necessita.

Figura 17: Exemplo da retirada de “*Stop Words*” (WIVES, 1997)

4.5.2.2 Radicalização ou *Stemming*

Nessa etapa, são extraídos os sufixos das palavras, descartando-se o gênero, número e conjugação verbal. Dessa forma, obtém-se o radical, que verdadeiramente concentra a semântica da palavra. O método para eliminação dos sufixos depende do idioma da palavra, e no KAIROS, devido às características das fontes de dados, foi previsto o processamento dos idiomas inglês e português. A Tabela 3 exemplifica o processo de obtenção dos radicais das palavras, ou radicalização.

Tabela 3: Exemplo da obtenção dos radicais das palavras.

Palavra	Idioma	Radical
CONNECTED	Inglês	CONNECT
CONNECTING	Inglês	CONNECT
CONNECTION	Inglês	CONNECT
CONNECTIONS	Inglês	CONNECT
EDUCAÇÃO	Português	EDUC
EDUCACIONAL	Português	EDUC
EDUCADOR	Português	EDUC

Para radicalização das palavras do idioma inglês, foi utilizado o Algoritmo de Porter, proposto por PORTER (1980), que serve como base para diversos algoritmos de

radicalização existentes em outros idiomas. O Algoritmo de Porter segue os seguintes passos:

- Remoção do plural, incluindo casos especiais como *'sses'* e *'ies'*;
- Substituição de sufixos por suas raízes, que poderão ser eliminadas nos passos seguintes. Por exemplo: *'ational'* por *'ate'*, *'tional'* por *'tion'*, *'enci'* por *'ence'*, *'ization'* por *'ize'*, entre outros;
- Transformações necessárias devido a palavras especiais, como por exemplo, *'icate'* por *'ic'*, *'ical'* por *'ic'*, *'alise'* por *'al'*, entre outras.
- Retirada de sufixos resultantes dessas transformações, como, por exemplo, *'able'*, *'al'*, *'ance'*, *'ant'*, *'ent'*, *'er'*, *'ible'*, *'ic'*, *'ive'*, *'ism'*, *'ize'*, *'ment'*, *'ous'*, *'tion'*, entre outros.

Para radicalização das palavras do idioma português, foi utilizado o Algoritmo Português Stemmer, proposto por ORENGO e HUYCK (2001), que se baseia no próprio Algoritmo de Porter. O algoritmo faz uso de mais de 200 regras, compostas pelos seguintes parâmetros: sufixo a ser removido, tamanho mínimo da palavra para que a regra seja executada, sufixo substituto (caso exista) e lista de exceções de palavras, que não são aplicáveis à regra. As regras fazem parte das seguintes classes:

- Redução de plural;
- Redução de feminino;
- Redução de advérbio;
- Redução de aumentativo e diminutivo;
- Redução de formas normais;
- Redução de terminações verbais;
- Redução de vogal temática;

4.5.3 Dicionário de Radicais

O dicionário de radicais consiste em um cadastro de radicais por área de conhecimento. Um peso é utilizado para pontuar a relevância de um radical em relação aos demais em uma determinada área. Esse cadastro existe para os idiomas inglês e português.

Não existe limitação em relação ao número de radicais de cada área de conhecimento, porém a soma da relevância de todos os radicais para determinada área deve ser igual a 10.

A Tabela 4 exemplifica o cadastro do dicionário de radicais, proposto para a área de conhecimento de *Banco de Dados*.

Tabela 4: Exemplo de preenchimento do dicionário de radicais.

Área de Conhecimento	Radical	Relevância	Idioma
Banco de Dados	banc	3	Português
Banco de Dados	dado	3	Português
Banco de Dados	bas	2	Português
Banco de Dados	inform	2	Português
Banco de Dados	databas	4	Inglês
Banco de Dados	data	3	Inglês
Banco de Dados	inform	2	Inglês
Banco de Dados	queri	1	Inglês

Existem duas formas de carregar dados nesse dicionário: manual ou automaticamente. O dicionário, no entanto, deve estar carregado antes do início do processo de identificação de áreas. A Figura 18 apresenta graficamente as possibilidades de carga no dicionário de radicais, que serão descritas nas seções a seguir.

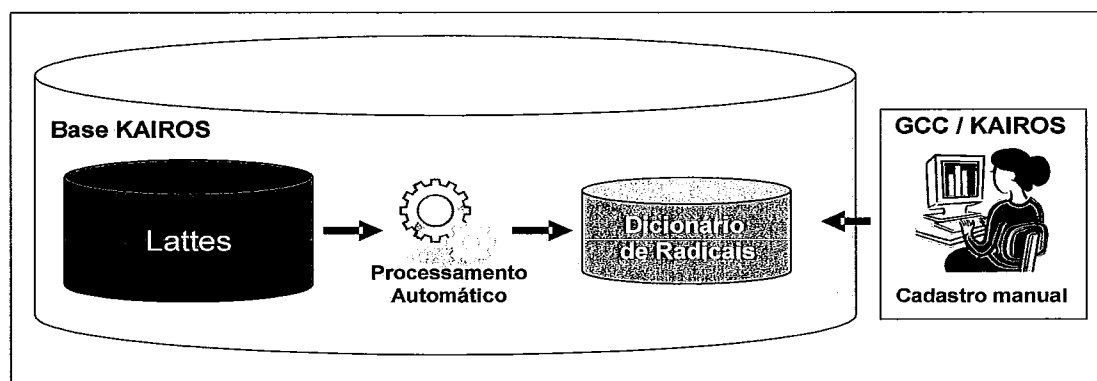


Figura 18: Possibilidades de carga do dicionário de radicais

4.5.3.1 Carga Manual

Através de uma funcionalidade no KAIROS, de forma integrada com o GCC, é possível cadastrar manualmente as áreas de conhecimento e os radicais associados, juntamente com a relevância da associação. Apesar do cadastro manual representar um

meio mais confiável e seguro de obtenção das informações, essa forma apresenta uma série de problemas.

Primeiramente, o cadastro demanda muito tempo de trabalho, pois, para uma única área de conhecimento, pode existir uma infinidade de radicais, cada qual com uma relevância. Além disso, para cada área de conhecimento, deve ser cadastrado radicais nos idiomas inglês e português, o que torna o trabalho ainda maior.

O trabalho para levantamento desses radicais deve ser conduzido por uma pessoa especialista na área de conhecimento. Caso contrário, o cadastro com qualidade torna-se inviável.

A Figura 19 apresenta a tela de cadastro do dicionário de radicais do KAIROS. O usuário cadastra a palavra, o idioma e a relevância do radical extraído da palavra para a área de conhecimento.

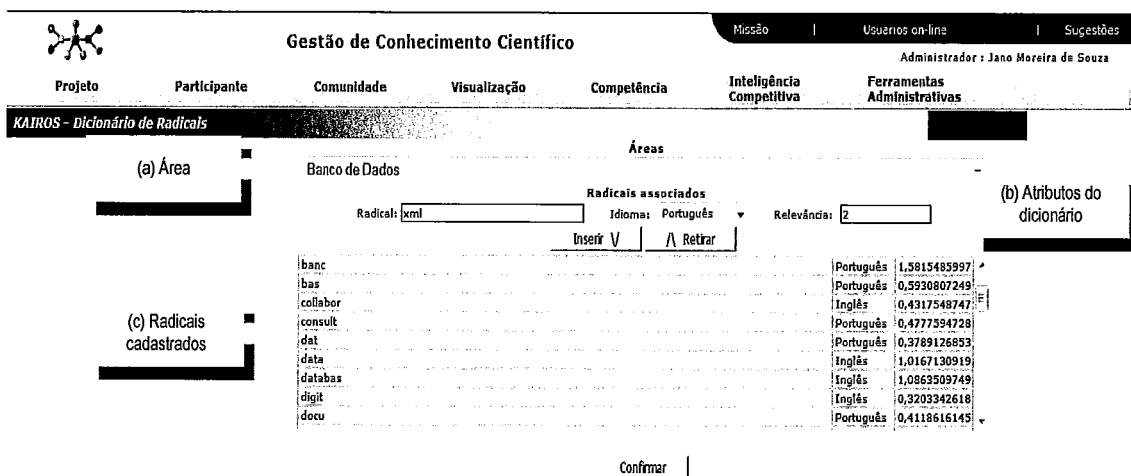


Figura 19: Tela de cadastro do dicionário de radicais.

4.5.3.2 Carga Automática

A carga automática tem o objetivo de preencher o dicionário de radicais de maneira eficiente e automatizada. Ela exclui a necessidade da mobilização de um profissional especialista para o cadastramento.

O processo utiliza dados presentes no Curriculum Lattes de pesquisadores, previamente importados para a base. Esses pesquisadores devem preencher o Curriculum com critério, para não haver distorções no cadastro de radicais. Quanto maior o número de pesquisadores importados, mais eficiente será o processo. Espera-se

que os pesquisadores considerados referência em uma área possuam o Curriculum Lattes preenchido de maneira criteriosa, pois os órgãos de fomento à pesquisa condicionam a concessão de bolsas e a liberação de recursos ao correto e completo preenchimento do Curriculum.

As produções bibliográficas presentes no Curriculum Lattes podem ter, a critério do pesquisador, as áreas de conhecimento declaradas nos seguintes níveis: Grande área, área, subárea e especialidade. Essas áreas são expressamente definidas pelo pesquisador e, por isso, possuem alto grau de confiabilidade. A carga automática se dá a partir dessas informações, mais precisamente da “especialidade” da produção bibliográfica.

A utilização dessas informações para o preenchimento do dicionário de radicais ocorreu por basicamente dois motivos: a liberdade do pesquisador digitar qualquer texto, permitindo que temas nascentes sejam identificados e o fato de existir o cadastro da produção científica nos idiomas português e inglês no Curriculum Lattes, ao contrário das outras fontes de dados.

O processo de carga automática consiste dos seguintes passos:

1. O título de todas as publicações com especialidade(s) declarada(s) são pré-processados (retirada das *stop words* e sufixos). Obtêm-se a lista dos radicais presentes nos títulos das publicações. Apenas os radicais com mais de duas letras são considerados.
2. Os radicais que aparecem em muitas publicações são excluídos dessa lista. Um radical pode se relacionar com muitas áreas de conhecimento dentro de um mesmo domínio. Nesse caso, o radical não pode ser usado para pontuar uma área específica. Na área de computação, por exemplo, palavras como ‘computação’, ‘software’, ‘hardware’, entre outras, originam radicais que estão relacionados à, praticamente, todas as áreas. No mesmo caso, encontram-se substantivos como ‘Brasil’ e adjetivos como ‘rápido’, ‘eficaz’, que não são úteis para o processo de identificação de áreas. Empiricamente, foi testada uma série de pontos de corte, com o objetivo de obter melhor resultado, e, no final dos testes, definiu-se por desconsiderar todos os radicais que se repetem em mais de 0.009% do número total de radicais levantados nas publicações no idioma português e mais que 0.028% do total no idioma inglês. Esses radicais, muito frequentes, são considerados palavras vazias pela classificação de FARIA (2001), conforme apresentado no Capítulo 3. Os

- pontos de corte 0.009% e 0.028% podem ser configurados livremente pelo usuário e servem apenas como sugestão, por terem apresentado bons resultados no domínio de computação, no qual os testes foram realizados.
3. O número de ocorrências de determinado radical é sumarizado para a respectiva especialidade declarada pelo pesquisador. As especialidades que possuírem cinco ou mais radicais relacionados (não necessariamente na mesma publicação, isto é, os radicais podem estar em publicações diferentes com a mesma especialidade) são cadastradas na base como uma área de conhecimento. Essa limitação faz com que não sejam cadastradas áreas de conhecimento com poucos radicais associados. Como todas as áreas de conhecimento, conforme exposto anteriormente, possuem soma total das relevâncias dos radicais igual a 10 para cada idioma, nas áreas com poucos radicais, os radicais teriam relevância excessivamente alta, podendo gerar distorções no processo de inferência de áreas.
 4. Os radicais relacionados com as áreas de conhecimento cadastradas na etapa anterior, cuja razão entre o número de ocorrências do radical para a área de conhecimento e o número total de ocorrência de todos os radicais associados à área de conhecimento seja maior ou igual a 0.006, são cadastrados no dicionário de radicais e associados com tais áreas de conhecimento. Essa razão foi obtida novamente de forma empírica, para limitar o número total de radicais associados à área de conhecimento. Os radicais associados poucas vezes à área de conhecimento são excluídos, por serem considerados ruídos de acordo com a classificação de FARIA (2001), apresentada no Capítulo 3. O idioma do radical cadastrado é igual ao idioma da publicação que lhe deu origem. A relevância do radical é obtida pelo número de ocorrências do radical para a área de conhecimento dividido pelo número total de ocorrências para todos os radicais cadastrados para a área de conhecimento multiplicado por 10, de forma que a relevância esteja sempre entre 0 e 10 e a soma da relevância de todos os radicais seja igual 10, conforme restrição imposta pelo dicionário de radicais.

Dessa forma, é possível carregar automaticamente o dicionário de radicais, mesmo que o usuário do KAIROS não tenha nenhum conhecimento sobre a área do domínio que deseja estudar. Para cada área de conhecimento cadastrada, a relevância total é igual a 10, fazendo com que todas as áreas sejam tratadas de forma igual, independente do número de radicais associados. A Figura 20 exemplifica a distribuição das palavras em um domínio, de acordo com a teoria apresentada no capítulo anterior.

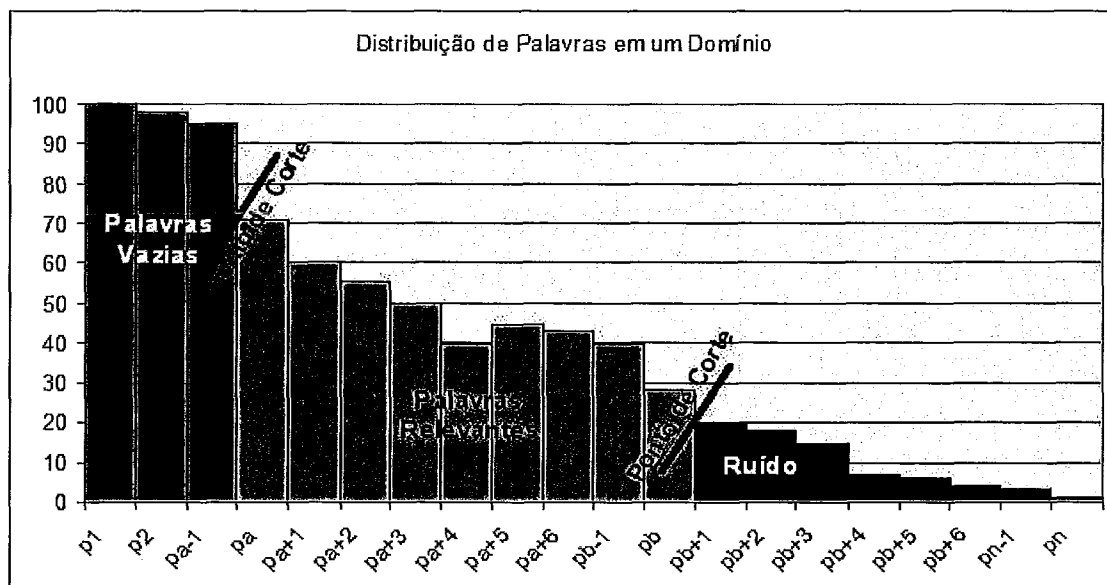


Figura 20: Palavras relevantes em um domínio (Adaptado de FARIA, 2001).

4.5.4 Inferência das Áreas Associadas à Produção Científica

Uma vez preenchido o dicionário de radicais é possível identificar as áreas de conhecimento associadas à produção científica. A Tabela 5 apresenta as informações coletadas de cada tipo de produção científica.

Tabela 5: Informações coletadas para cada tipo de produção científica.

Fonte de Dados	Tipo da Produção	Informação Coletada
Curriculum Lattes	Produção Bibliográfica	Título da publicação
	Produção Técnica	Título da publicação
	Projetos	Título e descrição
	Orientações	Título do trabalho
	Participações em Banca	Título do trabalho
	Formação Acadêmica	Título e palavras-chave
	Linha de Pesquisa	Nome e palavras-chave
DBLP	Publicação	Título da publicação
Teses e Dissertações Capes	Dissertações e Teses	Título da tese / dissertação
Patentes DERWENT	Patentes	Título da patente

Apesar de algumas fontes de dados oferecerem mais informações para serem mineradas, como é o caso dos resumos das teses e dissertações Capes e dos resumos das patentes Derwent, inicialmente optou-se por coletar apenas o título para a inferência, por se considerar que, nele, estão sumarizados as idéias e temas centrais abordados no texto. Além deste motivo, as informações das publicações do Curriculum Lattes e do DBLP não contêm resumo e, dessa forma, resolveu-se coletar informações semelhantes em cada fonte de dados.

O processo de inferência das áreas de conhecimento ocorre através dos seguintes passos:

1. Para cada informação coletada da produção científica, realiza-se o pré-processamento textual, obtendo-se uma lista de radicais;
2. Os radicais obtidos são confrontados com o dicionário de radicais, obtendo-se a relevância de cada área de conhecimento, de acordo com a soma da relevância dos radicais. Consideram-se apenas as cinco áreas de conhecimento mais relevantes para a produção científica;
3. As relevâncias dessas cinco áreas são normalizadas, de forma que a soma das relevâncias dessas áreas de conhecimento sejam igual a 10 para cada produção científica. Para isso, divide-se a relevância absoluta da área encontrada pela soma da relevância encontrada para a produção científica.

Dessa forma, independente do número de radicais existentes na informação coletada da produção científica, ou do número de áreas de conhecimento identificadas, a soma das relevâncias sempre será 10, evitando-se distorções.

Por exemplo, dado que os radicais ‘*banc*’ e ‘*xml*’ estejam associados à área de conhecimento *Banco de Dados* com relevância de 7 e 3, respectivamente, e que os radicais ‘*sistem*’ e ‘*xml*’ estejam associados à área de conhecimento *Engenharia de Software* com relevância de 8 e 2, respectivamente. Em uma publicação com o título fictício de “*Banco de Dados para arquivos Xml*”, os radicais obtidos na fase de pré-processamento textual seriam ‘*banc*’, ‘*dado*’, ‘*arqu*’ e ‘*xml*’, uma vez que *de* e *para* são eliminadas por serem *Stop Words*. Ao confrontar esses radicais com o dicionário, seriam obtidas as áreas de conhecimento *Banco de Dados*, com relevância absoluta igual a 10 (‘*banc*’ possui relevância igual a 7 e ‘*xml*’ igual a 3), e *Engenharia de Software*, com relevância absoluta igual a 2 (‘*xml*’ possui relevância igual a 2). Com a normalização, as

relevâncias obtidas seriam 8,3 (obtido pela divisão de 10 por 12) e 1,7 (obtido pela divisão de 2 por 12) para as áreas de *Banco de Dados* e *Engenharia de Software*, respectivamente.

Uma vez identificadas as áreas de conhecimento associadas às produções científicas é possível sumarizar a relevância das áreas de conhecimento ao longo do tempo, já que as produções científicas possuem a informação do ano em que ocorreram.

4.5.5 Inferência das Competências do Pesquisador

Com a identificação das áreas de conhecimento associadas às produções científicas é possível identificar as competências dos pesquisadores que as originaram. Para isso, o KAIROS desenvolveu uma heurística para pontuar as competências, levando em consideração o tipo de produção científica e a quantidade de tempo decorrido desde que ela foi produzida.

Como os dados são provenientes de fontes distintas, o cadastro do pesquisador não é único: em cada fonte, o pesquisador pode ter o nome levemente diferente, já que a informação minerada é textual e passível de erros ou abreviações. Além disso, uma mesma produção científica pode estar cadastrada em diferentes fontes. Por exemplo: uma publicação pode estar no Lattes e no DBLP, uma patente pode estar no Lattes e na base Derwent e uma dissertação pode estar no Lattes e na base da Capes.

Por isso, o KAIROS trata as bases de dados de cada fonte de maneira distinta. Nas consultas, o usuário poderá escolher em qual base de dados quer visualizar as informações ou, quando for possível, se deseja analisar as informações de todas as bases simultaneamente.

Para pontuar as competências do pesquisador, o KAIROS aplica a Equação 1, que obtém a relevância corrigida pelo peso e fator temporal da produção científica.

Equação 1: Cálculo da relevância corrigida pelo peso e fator temporal.

$$\text{Relevância Considerada} = \text{Relevância da Produção} * \text{PESO} * \text{FATOR_TEMPORAL}$$

O peso considerado na fórmula acima está descrito na Tabela 6. Como existe uma série de informações da vida acadêmica do pesquisador na base do KAIROS, o objetivo é pontuar os diferentes tipos de produção científica de acordo com a sua importância. Por exemplo: a orientação de um aluno de doutorado é uma atividade que, na maior parte das vezes, exige dedicação e conhecimento maiores do que a escrita de

um artigo. Da mesma forma, a linha de pesquisa declarada explicitamente pelo pesquisador é uma informação confiável e que deve receber um peso maior.

Tabela 6: Pesos de cada produção científica no cálculo de competências.

Fonte de Dados	Tipo da Produção	PESO
Curriculum Lattes	Produção Bibliográfica	1
	Produção Técnica	1
	Projetos	1
	Orientações Doutorado	3
	Orientações Mestrado	2
	Orientações (Outras)	1
	Participações em Banca de Doutorado	2
	Participações em Banca (Outras)	1
	Formação Acadêmica (Doutorado)	3
	Formação Acadêmica (Mestrado)	2
	Formação Acadêmica (Outras)	1
	Linha de Pesquisa	3
DBLP	Publicação	1
Dissertações e Teses Capes	Orientação Teses de Doutorado	3
	Orientação Dissertações de Mestrado	2
Patentes DERWENT	Patentes	2

O fator temporal (KAWAMURA, 2006, DE MARTINO, 2006) existe para penalizar os trabalhos que foram produzidos há muito tempo. Baseia-se na idéia de que quanto mais recente a produção científica, maior é o envolvimento do pesquisador com o tema. Por isso, descartam-se as produções científicas com mais de 35 anos e não se penalizam as que foram publicadas nos últimos 5 anos. As produções que têm entre 5 e 35 anos são proporcionalmente penalizadas de acordo com a Equação 2.

Equação 2: Cálculo do fator temporal (Adaptado de KAWAMURA, 2006).

$$\text{FATOR_TEMPORAL} = \frac{35 - (\text{Ano Atual} - \text{Ano da Produção})}{30}$$

A partir de todas as produções científicas do pesquisador nas diferentes bases, as relevâncias das áreas de conhecimento são somadas e corrigidas – com o fator de correção descrito acima – e atribuídas ao pesquisador na forma de competências pessoais.

Com isso, o KAIROS passa a ter um grande mapa de competências, que contém a lista de competências de todos os pesquisadores cadastrados no Lattes, orientadores de dissertações e teses nacionais ou autores de publicações ou patentes internacionais. A busca de áreas e competências, descrita na seção 4.11, provê, ao usuário do KAIROS, o acesso a esse mapa de competências, possibilitando uma rápida e eficiente busca por pesquisadores ou instituições especialistas em determinados assuntos.

4.6 Mapeamento e Prospecção de Áreas

O KAIROS possui a funcionalidade de mapeamento e prospecção das áreas de conhecimento, identificadas na seção anterior. É possível visualizar como uma área de conhecimento se comportou ao longo dos anos, comparar diferentes áreas e prospectar a tendência de comportamento de uma área no futuro.

Por motivos de desempenho, as informações são previamente trabalhadas e consolidadas na base de dados, para então serem prospectadas. As próximas seções mostram como se dá esse processo e apresentam as telas do sistema com essa funcionalidade.

4.6.1 Consolidação

Conforme mencionado anteriormente, os dados são previamente trabalhados e consolidados para não onerar o tempo de acesso à consulta no KAIROS. Para todas as áreas de conhecimento e para todos os anos, a soma da relevância da área em relação às demais é salva na base de dados.

Como o número de conferências, publicações, orientações e patentes tem se observado crescente ao longo do tempo, parte-se do princípio de que o conhecimento e estudo sempre se mantiveram em crescimento uniforme, porém distribuído entre as áreas de conhecimento de maneira diferente ao longo do tempo. Esse pré-suposto existe para evitar distorções no estudo do desenvolvimento das áreas, pois como os registros das atividades científicas – em conferências, periódicos e outras fontes - têm se mostrado crescente, poder-se-ia achar que todas as áreas de conhecimento também se mantiveram crescentes.

O KAIROS sumariza sempre a razão entre a relevância da área de conhecimento e soma total das relevâncias de todas as áreas no ano, obtendo-se assim o percentual da área de conhecimento em relação às demais em cada ano.

Para oferecer mais possibilidades de estudo ao usuário do KAIROS é consolidado, também na base, o desenvolvimento da área ao longo do tempo, considerando-se a média dos cinco últimos valores absolutos da série histórica. Com isso, as variações abruptas são minimizadas e a curva é suavizada, permitindo que a prospecção seja de melhor qualidade. A Tabela 7 exemplifica a consolidação da média dos cinco últimos valores.

Tabela 7: Exemplo da consolidação utilizando média.

Ano	Valor Absoluto	Valor Média (5)
2000	6	0
2001	8	0
2002	10	0
2003	12	0
2004	14	$(14+12+10+8+6)/5 = 10$
2005	16	$(16+14+12+10+8)/5 = 12$
2006	25	$(25+16+14+12+10)/5 = 15.4$
2007	20	$(20+25+16+14+12)/5 = 17.4$
2008	22	$(22+20+25+16+14)/5 = 19.4$

4.6.2 Prospecção

Com base nos dados históricos do desenvolvimento das áreas de conhecimento, é possível prospectar como os dados tendem a se comportar no futuro, através da técnica de prospecção denominada Análise de Tendências. Essa técnica, apresentada no Capítulo 3, parte do princípio de que os dados futuros tendem a repetir o comportamento do passado.

A partir dos dados históricos dos últimos cinco anos, os pontos são interpolados por uma função polinomial de terceiro grau. Isto é, a função polinomial de terceiro grau que mais se aproxima dos pontos selecionados – os cinco últimos pontos da curva de dados históricos – extrapola o próximo ponto da curva, que é a previsão do valor do ano seguinte. Para obter o polinômio responsável pela interpolação, é utilizado o método de aproximação de mínimos quadrados, explicado no Capítulo 3.

Esses passos são repetidos por mais duas vezes, considerando o valor obtido no passo anterior para extrapolar o próximo ponto da seqüência. Com isso, são obtidas as previsões dos valores de três próximos anos.

Tomou-se o cuidado de utilizar um polinômio de grau maior que 2 para realizar as extrapolações, pois, caso contrário, as mudanças de tendências não seriam retratadas na extrapolação.

Para realização da extrapolação, foi utilizada a biblioteca matemática OpenForecast (OPENFORECAST, 2009), desenvolvida na linguagem Java e utilizada pelo programa *Extrapolador* presente na camada de dados da arquitetura do KAIROS.

4.6.3 Desenvolvimento das Áreas de Conhecimento

O KAIROS possui a funcionalidade de consulta e comparação, através de gráficos, do desenvolvimento histórico e das tendências das áreas de conhecimento.

Para exibir os gráficos, utilizou-se a biblioteca java JFreeChart (JFREE, 2009), que fornece uma série facilidades para visualização de informação. Os gráficos são exibidos em forma de Applet Java, garantindo interoperabilidade e correto funcionamento em qualquer navegador ou sistema operacional.

A Figura 21 apresenta a tela inicial da consulta, onde o usuário escolhe os parâmetros de entrada, conforme detalhado a seguir:

- (a) Seleção das áreas para visualização ou comparação: o usuário escolhe quais áreas de conhecimento quer visualizar ou comparar. Não existe limite de quantidade.
- (b) Seleção da fonte de dados: o usuário escolhe a origem da informação que deseja visualizar. Possibilidades: Curriculum Lattes (Produção Bibliográfica), DBLP, Base de Teses, Patentes ou todas as fontes somadas.
- (c) Seleção do tipo de gráfico: o usuário escolhe como quer visualizar as informações no gráfico. Possibilidades: média dos últimos 5 valores ou contagem absoluta.
- (d) Seleção do período de consulta: o usuário escolhe o ano inicial e final da consulta. Caso o período inicial não seja preenchido, é considerado o ano de início da base de dados. Caso o período final não seja preenchido, é considerado o último ano prospectado na base.

Práticas | Sugestões

Gestão de Conhecimento Científico

Administrador: J

Projeto | Participante | Comunidade | Visualização | Competência | Inteligência Competitiva | Ferramentas Administrativas

KAIROS - Mapeamento e Prospecção das Áreas

Configurações

Áreas

- CSCW (Trabalho Cooperativo Suportado por Computador)
- Engenharia de Requisitos
- Engenharia de Software
- Estatística Bayesiana
- Gestão do Conhecimento
- Groupware
- Hardware
- Implementação de Algoritmos Criptográficos
- Informática na Educação
- Inteligência Artificial

(a) Seleção das áreas para visualização ou comparação

Banco de Dados

Engenharia de Software

(b) Seleção da fonte de dados

(c) Seleção do tipo de gráfico

Fontes de dados: DBLP

Tipo de Gráfico: Média dos últimos 5 valores

Período: 1975 Até (Para comparar todo o período, deixe em branco)

Visualizar Gráficos

Figura 21: Mapeamento e prospecção de áreas (Consulta)

A Figura 22 apresenta o retorno da consulta, de acordo com os parâmetros selecionados no passo anterior. Pode-se destacar:

- (a) Informações do gráfico: o sistema informa qual fonte de dados foi utilizada e em quais anos foram prospectados.
- (b) Áreas estudadas: o sistema exibe uma cor para cada área selecionada e apresenta o gráfico da série temporal.
- (c) Anos prospectados: o sistema apresenta a previsão obtida para os três próximos anos da base de dados.

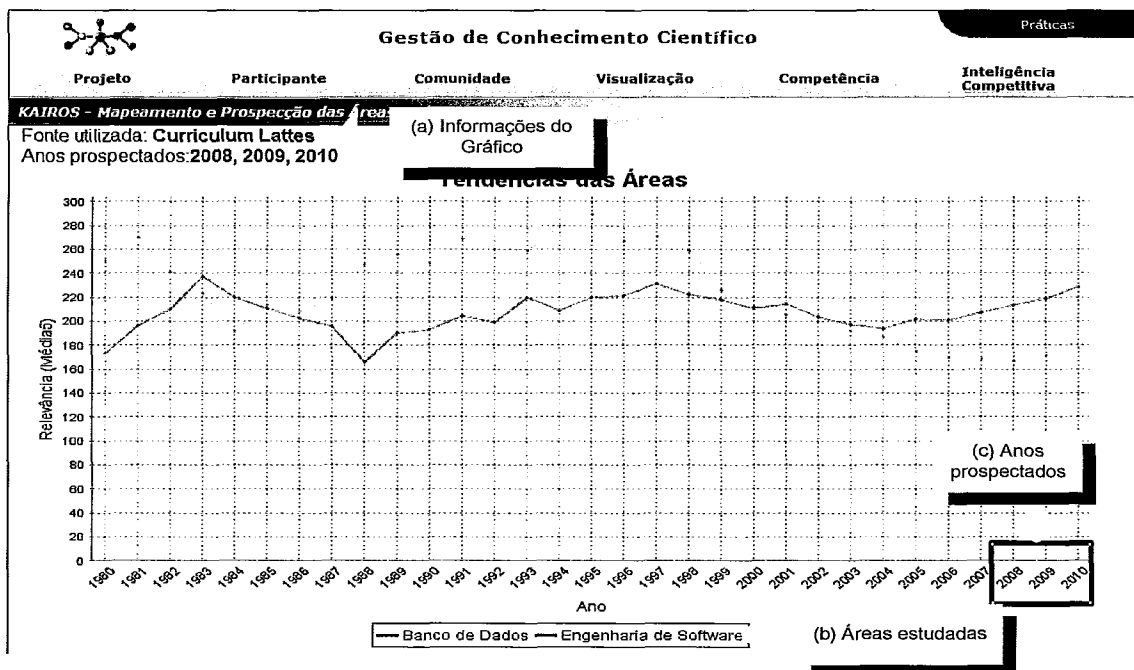


Figura 22: Mapeamento e prospecção de áreas (Resultado)

4.7 Similaridade de Áreas

Foi desenvolvido um método para identificação do grau de similaridade que as áreas possuem entre si. A idéia é identificar, em uma área estudada, em determinado ano, quais outras áreas foram identificadas em produções científicas concomitantemente com a área estudada. O grau de similaridade consiste na razão da soma das ocorrências das outras áreas em produções que a área estudada foi identificada pela soma das ocorrências da área estudada.

É importante verificar que, dada duas áreas A e B , estudando a área A , o grau de similaridade com a área B pode ser diferente do grau de similaridade com a área A , quando estudamos a área B . Ou seja, utilizando essa heurística, a mudança da ordem de comparação leva a resultados diferentes. Dada a área estudada A , o grau de similaridade com a área B pode ser obtido através da Equação 3.

Equação 3: Cálculo da similaridade entre áreas.

$$\text{Similaridade}(A,B) = \frac{\text{Número de vezes que A e B são identificadas na mesma produção científica}}{\text{Número de vezes que A é identificada em uma produção científica}}$$

Para exemplificar com dados fictícios, dado que a produção científica do ano de 2008 tenha sido conforme a Tabela 8, os graus de similaridade calculados pelo método seriam os descritos na Tabela 9.

Por exemplo, estudando a área 1, observa-se que o grau de similaridade com a área 2 é igual a 0.33, já que em 33% das publicações nas quais a área 1 é identificada, a área 2 também é verificada.

Tabela 8: Produção científica de dado ano (dados fictícios)

Ano	Produção Científica	Áreas identificadas
2008	Produção 1	Área 1
		Área 2
		Área 3
	Produção 2	Área 1
		Área 4
	Produção 3	Área 2
		Área 3
	Produção 4	Área 1
		Área 3

Tabela 9: Cálculo dos graus de similaridade entre as áreas

Ano	Área estudada	Num. de Produções	Áreas relacionadas	Relacionamentos	Grau de Similaridade
2008	Área 1	3	Área 2	1	0,33
			Área 3	2	0,66
			Área 4	1	0,33
	Área 2	2	Área 1	1	0,50
			Área 3	2	1,00
			Área 4	0	0,00
	Área 3	3	Área 1	2	0,66
			Área 2	2	0,66
			Área 4	0	0,00
	Área 4	1	Área 1	1	1,00
			Área 2	0	0,00
			Área 3	0	0,00

Para exemplificar a utilização da funcionalidade que identifica a similaridade de áreas do KAIROS, a Figura 23 apresenta a tela de consulta onde o usuário entra com os seguintes parâmetros:

- (a) Área que será estudada: o usuário seleciona a área que deseja estudar.
- (b) Ano em que os dados serão visualizados: o ano no qual a heurística levantará as áreas correlatas. A cada ano, os resultados são diferentes.
- (c) Fonte de dados: o usuário seleciona a fonte de dados sobre a qual o sistema processará os resultados.
- (d) Quantidade de áreas retornadas: o usuário seleciona quantas áreas semelhantes à área estudada devem ser retornadas pelo sistema.

A Figura 24 apresenta o retorno da consulta de identificação de similaridade de áreas. Pode-se destacar:

- (a) Área selecionada no passo anterior: o sistema exibe a área que o usuário selecionou para estudo.
- (b) Áreas relacionadas: o sistema exibe as áreas relacionadas à área estudada, ordenadas pelo grau de similaridade.

(c) Grau de proximidade: o sistema calcula o grau de similaridade da área estudada com cada área semelhante.

(d) Ano: o sistema exibe o ano em que o estudo foi realizado. Se o usuário desejar visualizar outro período, basta digitar o ano e clicar no botão “Atualizar”.

Figura 23: Similaridade de áreas (Consulta)

Ao selecionar uma área, o sistema exibe a similaridade entre a área estudada e a área selecionada ao longo do tempo, conforme pode ser verificado na Figura 25.

Áreas Semelhantes	Proximidade
Engenharia de Requisitos	0,088306
Interação Homem Computador	0,079196
programação paralela	0,053918
Implementação de Algoritmos Criptográficos	0,049763
Programação Linear, Não-Linear, Mista e Dinâmica	0,04513
Hardware	0,042933
Interação Humano Computador	0,040462
Linguagens de Programação	0,034806
Informática Aplicada à Educação	0,030665
Gestão do Conhecimento Em Equipes de Projeto	0,025878

*Clique sobre o nome da área para visualizar a série histórica

Figura 24: Similaridade de áreas (Resultado)

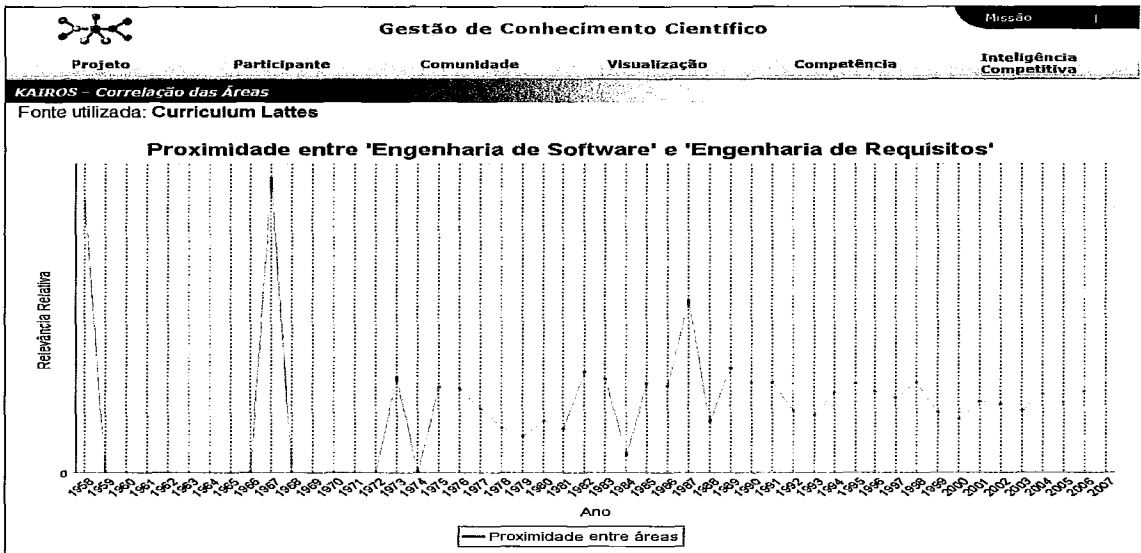


Figura 25: Similaridade de áreas ano a ano.

O gráfico da evolução da similaridade das áreas ano a ano permite algumas análises: a curva decrescente, por exemplo, indica o distanciamento das áreas, sugerindo que ocorreram mudanças na conceituação e definição de uma das áreas. A curva crescente, no entanto, indica um aumento da proximidade das áreas e, no caso de valores altos de similaridade, pode sugerir o início de um processo de fusão entre diferentes áreas.

4.8 Identificação de Áreas Promissoras

A partir da prospecção de dados futuros das áreas de conhecimento, é possível verificar quais áreas são mais promissoras, ou seja, verificar quais áreas possuem maior previsão de crescimento. O KAIROS apresenta uma lista dessas áreas, ordenada pelo percentual estimado de crescimento. O percentual é obtido pela razão entre a relevância da área de conhecimento no ano prospectado e a relevância da área no ano atual.

Para que áreas com relevância baixa não apareçam na lista de resultados, o usuário pode escolher um ponto de corte, a partir do qual serão selecionadas as áreas com relevância superior à relevância mínima determinada. Esse ponto de corte faz com que distorções não ocorram. Por exemplo, uma área que tem relevância 1 e que possua previsão de crescimento no próximo ano de relevância 2, tem taxa de crescimento estimado de 100%. Porém, o usuário pode estar mais interessado em uma área maior, que possua relevância no ano atual de 1000 e previsão de crescimento no próximo ano de 1500 (apesar da taxa de crescimento de 50% ser menor). O crescimento, mesmo que alto, de áreas muito pequenas, não tende a ser representativo.

A Figura 26 apresenta a tela da consulta de identificação de áreas promissoras, onde o usuário escolhe os seguintes parâmetros de entrada, conforme detalhado a seguir:

- (a) Seleção do ano considerado: o usuário escolhe em qual dos três anos prospectados deseja identificar as áreas promissoras.
- (b) Seleção da fonte de dados: o usuário escolhe a origem da informação que deseja visualizar: Curriculum Lattes (Produção Bibliográfica), DBLP, Base de Teses, Patentes ou todas as fontes somadas.
- (c) Seleção da relevância mínima considerada: o usuário escolhe a relevância mínima (ponto de corte) que as áreas devem possuir no ano atual para serem consideradas na consulta.

A Figura 27 apresenta o retorno da consulta, de acordo com os parâmetros selecionados no passo anterior. Pode-se destacar:

- (a) Informações da consulta: o sistema exibe no cabeçalho o ano e a fonte de dados considerados.
- (b) Áreas de conhecimento: o sistema exibe a lista de áreas de conhecimento, ordenadas pela taxa de crescimento. Clicando sobre o nome da área, o usuário visualiza o gráfico temporal do desenvolvimento da área.
- (c) Taxas de crescimento estimadas: o sistema exibe a taxa de crescimento estimado de cada área, obtida através da razão entre a relevância da área no ano prospectado e a relevância no ano atual.

Gestão de Conhecimento Científico Missão

Projeto: **KAIROS - Identificação de Áreas Promissoras**

Participante: [Redacted]

(a) Seleção do ano considerado: 2008

Visualização: [Redacted]

Competência: [Redacted]

Inteligência Competitiva: [Redacted]

Configurações

Ano Prospectado: 2008

Fonte de Dados: Curriculum Lattes

(b) Seleção da fonte de dados utilizada

Relevância Mínima em : 10

(c) Seleção da relevância mínima considerada

Visualizar

Figura 26: Identificação de áreas promissoras (Consulta)



Projeto	Participante	Comunidade	Visualização	Competência	Inteligência Competitiva
KAIROS - Identificação de Áreas Promissoras					
Áreas Promissoras 2008 - Curriculum Lattes					
Área		(a) Informações da consulta		Taxa Crescimento (%)	
1- Jogos 3D				17,8935	
2- Implementação de Algoritmos Criptográficos				14,7224	
3- Teoria dos Grafos				12,6083	
4- Processamento e Análise de Imagens Médicas		(b) Áreas de conhecimento		11,8828	
5- Sistemas de Informação Geográficos				10,3778	
6- Matemática Discreta e Combinatória				10,1218	
7- Algoritmos Distribuídos				9,0451	
8- Regressão e Correlação				8,0834	(c) Taxas de crescimento estimadas
9- Processamento de Imagens				7,6705	
10- Algoritmos Randômicos e Aproximativos				7,5249	
11- Processamento Paralelo e Distribuído				7,5068	
12- Mineração de Dados				6,4504	
13- Sistemas Hipermedia				6,3404	
14- Programação Não Linear				6,0339	
15- Sistemas Autônomos Inteligentes				5,854	
16- Computação Em Grids Ou Grades				5,8529	
17- Otimização				4,5626	
18- Visão Computacional				3,7558	
19- Sistemas Inteligentes				3,3356	
20- Engenharia de Software				3,0441	
21- Computação de Alto Desempenho				2,8618	
22- Robótica				2,8231	

Figura 27: Identificação de áreas promissoras (Resultado)

4.9 Alinhamento com Áreas Promissoras

O KAIROS permite consultar o grau de alinhamento dos pesquisadores e instituições em relação às áreas que tendem a ser mais promissoras. Para isso, são consultadas as competências dos pesquisadores e instituições, no último ano ou ao longo de todos os anos, e são verificadas se elas tendem a crescer ou não no futuro.

Quando consideradas as competências somente no último ano, o sistema verifica as áreas científicas associadas à produção científica do pesquisador ou da instituição no último ano. Quando consideradas as competências ao longo de todos os anos, utiliza-se a heurística descrita na seção 4.5.5 para verificar as competências.

O Curriculum Lattes é a única fonte de dados que contém informações sobre a instituição de pesquisa da qual o pesquisador faz parte. No KAIROS, quando for mencionado sobre o alinhamento da instituição, trata-se, na verdade, do alinhamento de todos os pesquisadores a ela subordinados, na fonte de dados do Curriculum Lattes.

A Figura 28 apresenta a tela da consulta de alinhamento com áreas promissoras, onde o usuário escolhe os seguintes parâmetros de entrada:

- (a) Tipo de consulta: o usuário escolhe se deseja consultar o alinhamento de pesquisadores ou instituições. No caso do alinhamento de instituições, serão considerados todos os pesquisadores subordinados.

- (b) Fonte de dados das competências: o usuário escolhe qual a fonte de dados na qual deseja levantar as competências do pesquisador, para posteriormente compará-las com as áreas promissoras. No caso das instituições, a única fonte de dados possível é o Curriculum Lattes.
- (c) Ano considerado: o usuário escolhe em qual dos anos prospectados deseja considerar as taxas de crescimento.
- (d) Fonte de dados das áreas: o usuário escolhe em qual fonte de dados as áreas promissoras deverão ser prospectadas.
- (e) Abrangência das competências: o usuário escolhe se deverão ser consideradas as competências somente do último ano ou de todos os anos dos pesquisadores/instituições. No caso da abrangência ser de todos os anos (Histórico Completo), o cálculo utilizado é o descrito na seção 4.5.5 (Identificação de Competência de Pesquisadores).
- (f) Número de competências para avaliação: o usuário escolhe quantas competências deverão ser consideradas para o estudo de alinhamento. Sempre são consideradas as principais competências, ordenadas pela relevância.
- (g) Taxa alvo: o usuário escolhe a taxa de crescimento que gostaria que as competências do pesquisador/instituição tivessem. Esse parâmetro é utilizado somente para configurar a visualização do relatório.

The screenshot shows the 'Gestão de Conhecimento Científico' interface. At the top, there is a navigation bar with 'Missão' on the right and a logo on the left. Below the navigation bar, there are several tabs: 'Projeto', '(a) Tipo de consulta', 'Comunidade', 'Visualização', 'Competência', and 'Inteligência Competitiva'. The 'Projeto' tab is active, showing 'KAIROS - Alinhamento c...'. Below the tabs, there is a search bar with the text 'Pesquisar alinhamento de Pesquisadores a partir da base de Curriculum Lattes'. To the right of the search bar, there is a dropdown menu labeled '(b) Fonte de dados das competências'. Below the search bar, there is a 'Configurações' section with several options: 'Ano Prospectado:' with a dropdown menu showing '2008' and a label '(c) Ano considerado'; 'Prospectar áreas a partir de:' with a dropdown menu showing 'Curriculum Lattes' and a label '(d) Fonte de dados das áreas'; 'Abrangência:' with a dropdown menu showing 'Histórico Completo (Relevância Calculada)' and a label '(e) Abrangência das competências'; 'Nº de Competências consideradas:' with a text input field showing '8'; and 'Taxa de Crescimento Alvo:' with a text input field showing '5 %' and a label '(g) Taxa alvo'. There is also a button labeled '(f) Número de competências para avaliação' and an 'Avançar' button.

Figura 28: Alinhamento com áreas promissoras (Consulta)

Após selecionar os parâmetros de configuração da consulta, o usuário deve selecionar qual pesquisador ou instituição será estudado, conforme mostra a Figura 29.

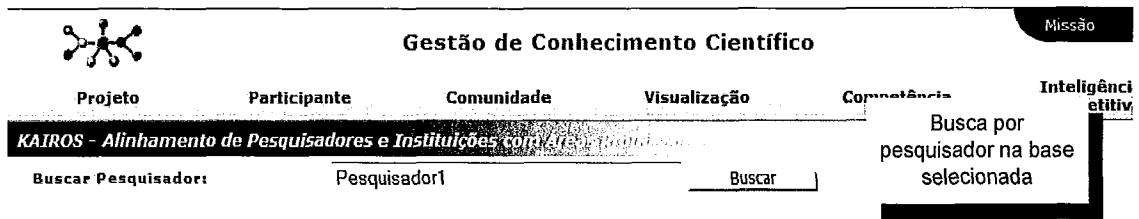
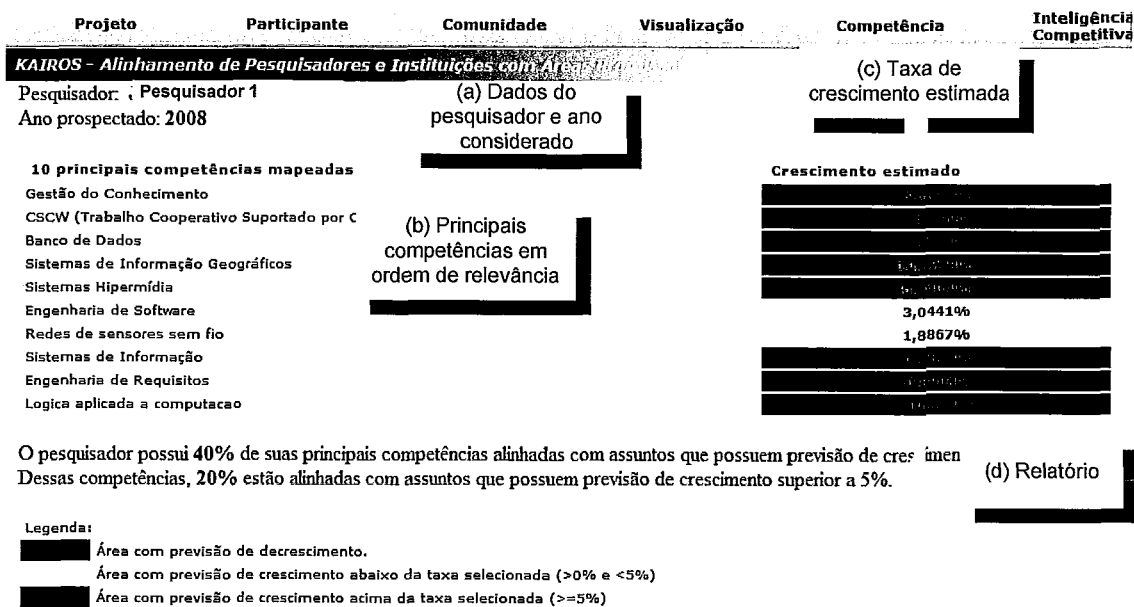


Figura 29: Alinhamento com áreas promissoras (Seleção do pesquisador ou instituição)

A Figura 30 apresenta o retorno da consulta do alinhamento de pesquisadores e instituições. Pode-se destacar:

- (a) Dados do pesquisador/instituição e ano considerado: o sistema exibe o nome do pesquisador/instituição e o ano considerado para obtenção da taxa de crescimento.
- (b) Principais competências em ordem de relevância: o sistema exibe, em ordem de relevância, as principais competências do pesquisador/instituição.
- (c) Taxa de crescimento estimada: o sistema exibe para cada competência do pesquisador / instituição, a taxa de crescimento estimado para o ano considerado. As cores seguem a legenda no final da página: Vermelho para taxas negativas, amarelo para taxas positivas porém menores que a taxa alvo e verde para taxas superiores à taxa alvo.
- (d) Relatório: o sistema exibe um relatório contendo o percentual de competências dos pesquisadores ou instituições que possui previsão de crescimento abaixo da taxa alvo e previsão de crescimento acima da taxa alvo.



As competências do pesquisador foram mapeadas através da base Curriculum Lattes.
As áreas foram prospectadas através de dados mapeados em Curriculum Lattes.

Figura 30: Alinhamento com áreas promissoras (Resultado)

4.10 Identificação de Visionários

Foi desenvolvida, no KAIROS, uma heurística para identificação dos pesquisadores e instituições visionários nas áreas científicas. Essa heurística identifica aqueles que desenvolveram pesquisas em uma determinada área antes de ela crescer e se tornar popular.

Essa heurística leva em conta dois parâmetros definidos pelo usuário: critério de seleção e percentual de início da área. O percentual de início da área é definido livremente pelo usuário, enquanto que o critério de seleção pode ser:

- Atividade no início da área (25%)
- Atividade no início da área (10%)
- Atividade no início da área (5%)

A heurística lista todos os pesquisadores ou instituições que produziram até determinado ano limite, sendo, assim, considerados visionários. O ano limite é calculado através dos critérios definidos pelo usuário. A idéia geral é considerar visionários todos que produziram até determinado percentual de tempo entre o início da

área e o pico da área (ano com maior produção da área). Esse percentual é definido no critério de seleção (atividade no início da área de 5%, 10% ou 25%). Como, inicialmente, uma área pode ter uma produção muito baixa por vários anos, existe outro parâmetro de configuração que faz com que o início da área seja considerado após ela atingir um percentual de produção em relação à produção total da área ao longo dos anos. Com isso, evita-se que o número de visionários nessas áreas seja muito reduzido.

A heurística para calcular o ano limite, dada uma área, segue os passos detalhados a seguir:

1. Calcula-se a relevância total da área científica ao longo de toda a série histórica (visualmente, é a área da superfície do gráfico da série histórica).
2. Obtém-se o ano inicial - ano em que área atingiu o percentual mínimo de relevância (parâmetro de entrada do usuário) em relação à relevância total.
3. Obtém-se o ano de pico da área, ou seja, o ano em que a área obteve maior relevância em sua série histórica.
4. Obtém-se o período válido, que consiste no período entre o ano inicial e o ano de pico.
5. Obtém-se o ano limite. Utiliza-se o percentual definido nos critérios de entrada do usuário (atividade no início da área de 5%, 10% ou 25%) para obter o ano limite. Esse percentual define até que ano limite o período válido é considerado para determinar se os pesquisadores ou instituições são visionários. No cálculo, o ano limite é arredondado para cima.
6. Caso o ano inicial considerado seja maior do que o ano de pico, considera-se o ano limite igual ao ano inicial considerado.

A partir do ano limite, são considerados visionários os pesquisadores ou instituições que produziram até esse ano. Para exemplificar o funcionamento do processo, a Figura 31 apresenta a curva histórica de uma área fictícia, com as marcações relevantes dos passos utilizados pela heurística.

A área do exemplo teve início em 1981 e possui dados até 2007. A soma total de relevância em todo período histórico é igual a 1639, representado pela área da superfície do gráfico destacada em cinza. O pico de produção da área ocorreu em 2006, com relevância igual a 107. Considerando, como parâmetro de configuração, o percentual de início da área igual a 5%, o ano início ocorre em 1985 (5% de 1639 é igual a 81,95 e

somente em 1985 a relevância acumulada chega a esse valor). Dessa forma, o período válido se dá entre 1985 (ano início) e 2006 (ano pico), totalizando 21 anos. Considerando o critério de seleção de atividade no início da área igual a 10%, o ano limite ocorre em 1988, já que o limite considerado no período válido é igual a 3 anos, pois 10% de 21 anos é igual a 2,1 anos (o algoritmo arredonda sempre para cima a quantidade de anos obtidos).

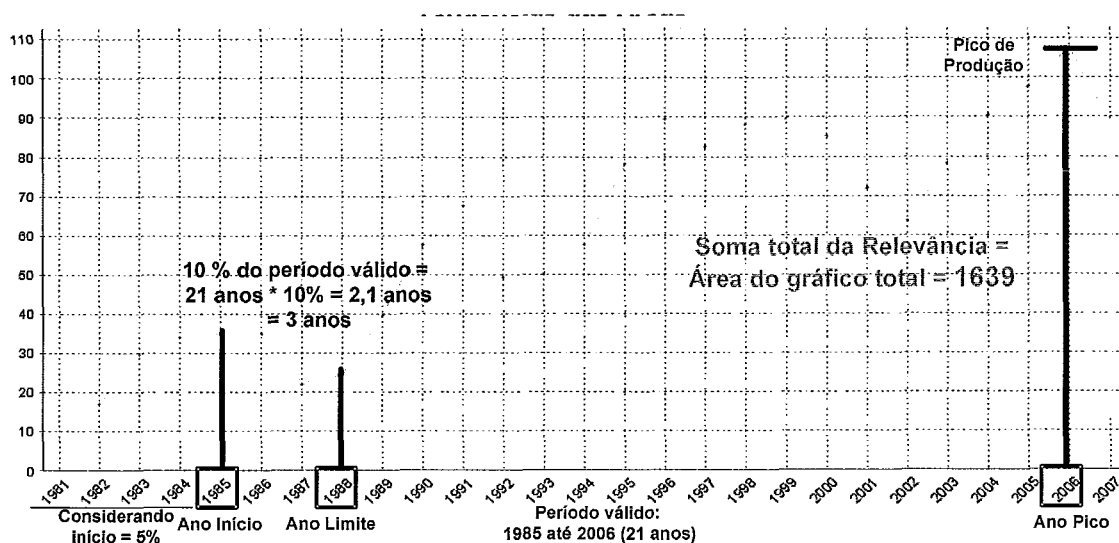


Figura 31: Exemplo do funcionamento da heurística.

Pelo exemplo, todos os pesquisadores ou instituições que produziram entre 1981 e 1988 são considerados visionários para a área fictícia.

Para exemplificar a utilização da funcionalidade de identificação de visionários do KAIROS, a Figura 32 apresenta a tela onde o usuário escolhe os seguintes parâmetros de consulta:

- (a) Tipo de consulta: o usuário escolhe se deseja identificar pesquisadores ou instituições visionários.
- (b) Seleção de áreas que serão estudadas: o usuário escolhe as áreas que deseja estudar. A heurística é aplicada para cada área separadamente.
- (c) Seleção dos parâmetros de configuração da heurística: o usuário escolhe os parâmetros de configuração da heurística.
- (d) Fonte de dados: o usuário seleciona a fonte de dados que será utilizada para a consulta. No caso de consulta de instituições, a única fonte de dados possível é o Curriculum Lattes.

Gestão de Conhecimento Científico						Missão	Usuarios on-line
Administrador : Jan							
Projeto	Participante	Comunidade	Visualização	Competência	Inteligência Competitiva	Ferramentas Administrativas	
KAIROS - Identificação de Pesquisadores e Instituições Visionários							
Configurações							
Identificação de		Pesquisadores		visionários(as).		(a) Tipo de consulta	
Áreas							
<input type="checkbox"/> Áreas cadastradas <input checked="" type="checkbox"/> Agentes Inteligentes <input type="checkbox"/> Álgebra Comutativa <input type="checkbox"/> Algoritmos Distribuídos <input type="checkbox"/> Análise de Algoritmos e Complexidade de Computação <input type="checkbox"/> Análise Numérica <input type="checkbox"/> Aprendizado de Máquina <input type="checkbox"/> Arquitetura de Sistemas de Computação <input type="checkbox"/> Automação Eletrônica de Processos Elétricos e Industriais <input type="checkbox"/> Avaliação de Desempenho				(b) Seleção das áreas que serão estudadas		Agentes Inteligentes	
Método de Seleção:				(c) Seleção da heurística que será utilizada		Atividade no início da área (25%)	
Fontes de dados:						Considerar ano inicial após 5 % das publicações.	
						Curriculum Lattes	
Visualizar Resultados						(d) Fonte dos dados	

Figura 32: Identificação de visionários (Consulta)

A Figura 33 apresenta o retorno da consulta de identificação de pesquisadores e instituições visionários. Pode-se destacar:

- (a) Área selecionada no passo anterior: o sistema exibe a área científica selecionada pelo usuário no passo anterior. Caso tenham sido escolhidas mais de uma área, o sistema exibe os resultados por blocos, área a área.
- (b) Lista de pesquisadores ou instituições: o sistema exibe a lista de pesquisadores ou instituições considerados visionários, ordenada pelo ano da primeira produção científica.
- (c) Ano do início da atividade: o sistema apresenta o ano da primeira produção do pesquisador ou instituição.

O nome dos pesquisadores retornados pela consulta foi omitido por questões de privacidade.

Gestão de Conhecimento Científico

Missão | Usuários on-line | Suggest

Administrador: Jairo Moreira de Souza

Projeto | Participante | Comunidade | Visualização | Competência | Inteligência Competitiva | Ferramentas Administrativas

KAIROS - Identificação de Pesquisadores e Agentes Inteligentes

(a) Área selecionada no passo anterior

Pesquisador	Ano Início
Pesquisador 1	1978
Pesquisador 2	1979
Pesquisador 3	1982
Pesquisador 4	1984
Pesquisador 5	1984
Pesquisador 6	1986
Pesquisador 7	1986
Pesquisador 8	1986
Pesquisador 9	1986
Pesquisador 10	1986
Pesquisador 11	1986
Pesquisador 12	1987
Pesquisador 13	1987
Pesquisador 14	1987
Pesquisador 15	1987
Pesquisador 16	1988
Pesquisador 17	1988
Pesquisador 18	1988
Pesquisador 19	1989
Pesquisador 20	1989
Pesquisador 21	1989
Pesquisador 22	1989
Pesquisador 23	1990
Pesquisador 24	1990
Pesquisador 25	1990
Pesquisador 26	1990

(b) Lista de pesquisadores ou instituições, ordenados pelo ano de início da atividade

(c) Ano do início da atividade

Figura 33: Identificação de visionários (Resultado).

4.11 Busca por Conhecimento

O KAIROS possui uma funcionalidade de busca, denominada *Busca por Conhecimento*, capaz de consultar na base de dados informações relativas às áreas de conhecimento cadastradas. A busca permite que sejam consultadas publicações, teses, patentes e demais produções científicas de determinada área de conhecimento ou especialistas identificados com atuação na área.

A Figura 34 apresenta a tela de consulta da busca por conhecimento, onde o usuário tem a opção de definir os seguintes parâmetros:

- (a) Tipo de busca: o usuário escolhe se deseja buscar por produção científica (publicações, teses, patentes e demais produções) ou por especialistas (pesquisadores ou instituições).
- (b) Área de conhecimento: o usuário escolhe a área de conhecimento relativa à busca.
- (c) Fonte de dados: o usuário seleciona a fonte de dados que será utilizada para a busca. A única fonte de dados que retorna instituições especialistas é o Curriculum Lattes.
- (d) Período da consulta: o usuário seleciona o período que deseja obter os resultados da busca. Esse período é utilizado apenas para a busca por

produção científica. Para a consulta de especialistas, o período é descartado, já que as competências são identificadas de acordo com a atuação do profissional ao longo de toda sua carreira.

Figura 34: Busca por conhecimento (consulta)

A Figura 35 apresenta um exemplo de retorno da busca por produção científica. Pode-se destacar:

- (a) Área de conhecimento: o sistema exibe a área de conhecimento consultada, selecionada na tela de consulta.
- (b) Tipo da produção científica: o sistema exibe o tipo da produção científica: publicação do DBLP, teses, patentes ou produção científica do Curriculum Lattes. No caso da fonte de dados Curriculum Lattes, o sistema exibe as produções bibliográficas, produções técnicas, projetos, formações acadêmicas, orientações concluídas e participações em banca relacionadas com a área de conhecimento pesquisada.
- (c) Título da produção científica: o sistema exibe o título completo da produção científica.
- (d) Relevância da área de conhecimento identificada na produção: o sistema apresenta a relevância obtida no processo de identificação de áreas da área de conhecimento para a produção científica.
- (e) Ano de publicação da produção científica: o sistema exibe o ano de publicação da produção científica.



KAIROS - Busca por Áreas e Especialistas

Área: Gestão do Conhecimento

Produção Bibliográfica:

(a) Área de conhecimento

(b) Tipo da produção científica

(c) Título da produção científica

(d) Relevância da área de conhecimento identificada na produção

(e) Ano de publicação da produção científica

Título	Relevância	Ano
<u>K-ANNOTATIONS An Approach for Conceptual Knowledge Implementation using Metadata Annotations</u>	7,12799	2009
<u>K-ANNOTATIONS An Approach for Conceptual Knowledge Implementation using Metadata Annotations (aceito para publicação)</u>	7,12799	2009
<u>Ontologias como novas bases de conhecimento científico</u>	6,7033	2009
<u>Knowledge Reuse in Data Mining Projects and Its Practical Applications</u>	6,39116	2009
<u>MEK: Using Spatial-Temporal Information to Improve Social Networks and Knowledge Dissemination</u>	6,35328	2009
<u>Assessing the contribution of shallow and deep knowledge sources for word sense disambiguation</u>	6,34862	2009
<u>MEK: Using spatial temporal information to improve social networks and knowledge dissemination</u>	5,32462	2009
<u>Conceptual Model for Knowledge Management in Corporate Universities</u>	4,85278	2009
<u>Automatic detection of compaction level on petroleum reservoir rocks based on visual knowledge</u>	4,53588	2009
<u>A practical approach for using the knowledge chains exchange to improve the application of learning theories in learning communities</u>	4,50636	2009
<u>Knowledge Management Integrating Organizational Learning in a Military Context</u>	4,4138	2009
<u>Knowledge Management Integrating Organizational Learning in a Military Context: The 3M Model</u>	4,4138	2009
<u>Analysis and Balancing of Social Network to Improve the Knowledge Flow on Multidisciplinary Teams</u>	3,63399	2009
<u>Analysis and Balancing of Social Network to Improve the Knowledge Flow on Multidisciplinary Teams (aceito a ser publicado)</u>	3,63399	2009
<u>Quantifying Social and Opportunistic Behavior in Email Networks</u>	2,5875	2009

Figura 35: Resultado da busca por conhecimento (busca por produção científica).

A Figura 36 apresenta um exemplo de retorno da busca por especialistas. Pode-se destacar:

- (a) Área de conhecimento: o sistema exibe a área de conhecimento consultada, selecionada na tela de consulta.
- (b) Lista de pesquisadores e instituições: o sistema exibe a lista de pesquisadores e instituições que possuem competência relacionada à área de conhecimento. Os pesquisadores e instituições aparecem ordenados pela relevância.
- (c) Relevância calculada da competência: o sistema apresenta a relevância calculada dos pesquisadores e instituições. Essa relevância foi calculada segundo a heurística apresentada na seção 4.5.5.



Projeto

Participante

Comunidade

Visualização

Competência

Inteligência
Competitiv**KAIROS - Busca por Áreas e Especialistas**

Área: Gestão do Conhecimento

(a) Área de
conhecimento

Pesquisadores:

Pesquisador	(b) Lista de pesquisadores ou instituições	(c) Relevância calculada da competência	Relevância
Pesquisador 1			549,7737646667
Pesquisador 2			430,0246396667
Pesquisador 3			397,8825653333
Pesquisador 4			277,767847
Pesquisador 5			277,4844376667
Pesquisador 6			244,816837
Pesquisador 7			223,872603
Pesquisador 8			191,646453
Pesquisador 9			184,8821406667
Pesquisador 10			171,543183
Pesquisador 11			149,6739176667
Pesquisador 12			144,6663976667
Pesquisador 13			119,4456773333
Pesquisador 14			99,9929503333
Pesquisador 15			91,0278506667
Pesquisador 16			90,8087963333

Figura 36: Resultado da busca por conhecimento (busca por especialistas).

A busca por conhecimento traz grandes benefícios para as organizações que a utilizam. Em pouco tempo é possível ter acesso a vasto material de consulta sobre determinado assunto, ou ainda, identificar especialistas não antes conhecidos, com atuação nacional e internacional, uma vez que o KAIROS minera fontes de dados preenchidas por pesquisadores no Brasil e no mundo.

A identificação de especialistas é importante para fomentar a cooperação entre organizações e é necessária para o próprio processo de prospecção tecnológica, uma vez que a prospecção realizada por ferramentas automatizadas nem sempre apresenta resultados perfeitos e completamente confiáveis, sendo desejada a validação dos resultados por um especialista.

Por sua vez, o vasto material de consulta sobre determinado assunto, retornado pela busca por produção científica, pode auxiliar o processo de desenvolvimento de competências dos pesquisadores de determinada organização. Dessa forma, é possível formar especialistas dentro da própria organização.

4.12 Comparação com Trabalhos Correlatos

A proposta desta dissertação, apresentada nesse capítulo, possui algumas vantagens frente aos trabalhos correlatos citados no Capítulo 3, sobretudo por integrar, em uma única ferramenta, funcionalidades para prospecção tecnológica e para identificação de especialistas.

Nenhum dos trabalhos correlatos selecionados apresenta a funcionalidade de identificação de pesquisadores e instituições visionários e de identificação do alinhamento com áreas promissoras. Da mesma forma, nenhum dos trabalhos apresenta aquisição automática de dados de fontes variadas, como foi implementado no KAIROS.

A identificação de especialistas, proposta nessa dissertação, não se limita em sumarizar os trabalhos do pesquisador ou instituição, como é o caso dos trabalhos correlatos. No KAIROS, o tipo da produção científica e o tempo decorrido da publicação são considerados para pontuar as competências do pesquisador ou instituição.

ZHU e PORTER (2002) desenvolveram a análise de migração de conceitos, que possibilita a análise da evolução dos conceitos associados a cada área de conhecimento, ano a ano. A interface gráfica da ferramenta *VantagePoint*, originada a partir de trabalhos do grupo de pesquisa desses autores, possui diversas consultas, mapas e gráficos, sendo superior à interface gráfica do KAIROS.

O método automático de inferência de áreas de conhecimento proposto por VELARDI *et al.* (2007) apresenta vantagens ao método do KAIROS. Ao contrário do KAIROS, que infere as competências a partir de um dicionário simples de radicais, o método proposto por VELARDI *et al.* (2007) cria automaticamente uma ontologia, com relacionamentos entre as áreas de conhecimento, como especialização e generalização, e utiliza essa ontologia para inferir as áreas de conhecimento associadas aos documentos. A classificação do documento é mais completa, pois as áreas estão dispostas sobre uma hierarquia.

A Tabela 10 apresenta a comparação do KAIROS com todos os trabalhos correlatos citados no capítulo anterior. Conforme pode ser verificado, nenhum dos trabalhos possui a abrangência verificada no KAIROS.

Tabela 10: Comparação entre os trabalhos correlatos

Características	KAIROS	ZHU e PORTER (2002)	YOON e LEE (2008)	MARTINS (2008)	VELARDI <i>et al.</i> (2007)	RODRIGUES <i>et al.</i> (2006)
1- Abordagem						
1.1- Válida para qualquer domínio de conhecimento.	X	X	X		X	X
2- Fonte de dados						
2.1- Aquisições de dados automática.	X		X	X		
2.2- Informações de patentes.	X		X	X		
2.3- Informações de publicações científicas.	X					
2.4- Informações de teses e dissertações.	X					
2.5- Informações de arquivos (<i>Upload</i> direto).		X			X	X
3- Inferência de Conhecimento						
3.1- Automática.	X	X			X	
3.2- Semi-automática, com preenchimento manual de ontologia ou dicionário de dados.	X				X	X
3.3- Utilizando classificação da produção científica (por ex. CIP).			X	X		
4- Funcionalidades						
4.1- Análise da série histórica das áreas.	X	X	X	X		
4.2- Análise de tendências / extrapolação.	X	X	X			
4.3- Similaridade das áreas de conhecimento.	X	X				
4.4- Similaridade de pesquisadores e instituições.		X			X	
4.5- Alinhamento do pesquisador e instituição com áreas promissoras.	X					
4.6- Identificação de especialistas.	X	X			X	X
4.7- Identificação de visionários.	X					
4.8- Análise de migração de conceitos.		X				
4.9- Mecanismos de busca.	X	X			X	X

4.13 Conclusão

Este capítulo apresentou a proposta dessa dissertação: uma abordagem para prospecção tecnológica e identificação de especialistas. Essa abordagem é validada por um ambiente computacional, denominado KAIROS, também apresentado em detalhes junto com a abordagem. Para contextualizar o leitor, foi apresentado o ambiente GCC, um projeto de maior porte no qual o KAIROS está inserido.

Ao final do capítulo, foi apresentada uma comparação da ferramenta proposta com outros trabalhos correlatos existentes na literatura. A comparação mostrou que o KAIROS possui um grande número de funcionalidades, não verificadas em outros trabalhos.

No próximo capítulo são apresentados dois estudos de caso realizados com o objetivo de avaliar a proposta dessa dissertação. Os estudos de caso foram realizados no domínio de computação.

Capítulo 5 - Avaliação da Proposta

Para validar a abordagem proposta, foram realizados dois estudos de caso com o objetivo de avaliar dados obtidos pelo KAIROS. Neste capítulo, são detalhados os procedimentos utilizados e são apresentados os resultados.

5.1 Introdução

Foram realizados dois estudos de caso no domínio de computação. O primeiro, dividido em duas etapas, teve por objetivo validar o processo de prospecção de áreas e validar a heurística de identificação de pesquisadores visionários. Foi realizado com a participação de pesquisadores, que preencheram um formulário eletrônico disponibilizado internet. O segundo estudo teve por objetivo validar o processo de identificação de competências e foi realizado através da comparação das competências mapeadas no KAIROS e das áreas de atuação explicitamente informadas por pesquisadores selecionados.

O KAIROS, conforme apresentado no Capítulo 4, está disposto em uma arquitetura concebida em diferentes camadas. As funcionalidades, expostas ao usuário, dependem de procedimentos internos para apresentar os resultados da forma esperada. A Figura 37 apresenta as dependências entre alguns componentes da arquitetura do KAIROS e, em amarelo, a abrangência dos estudos de caso propostos nesse capítulo.

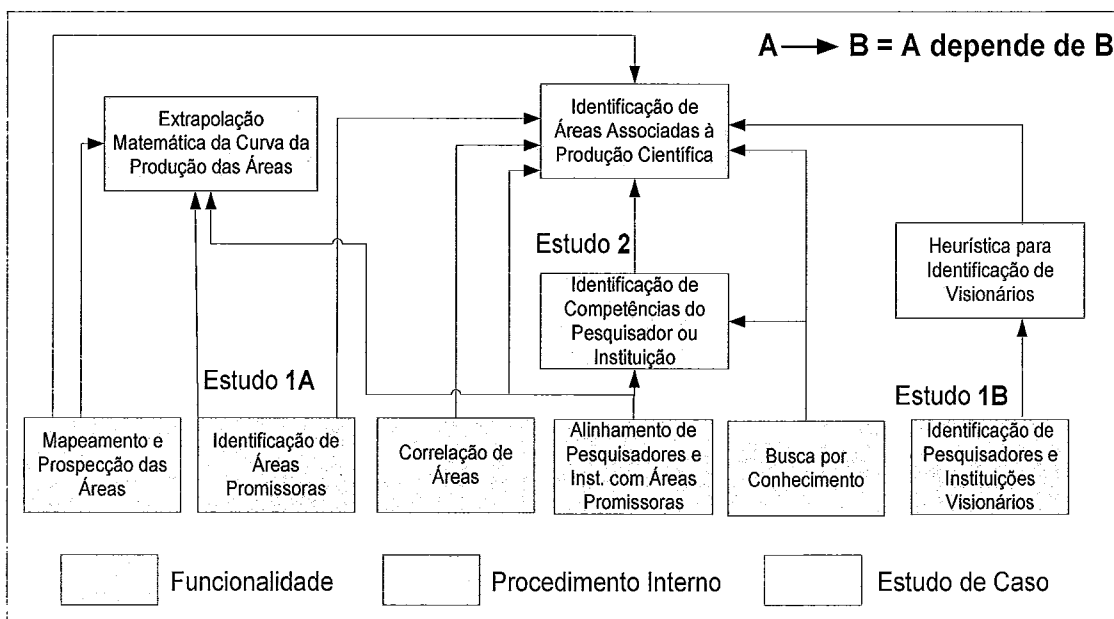


Figura 37: Dependência das funcionalidades e procedimentos internos do KAIROS.

Os estudos de caso propostos têm por objetivo validar as principais funcionalidades oferecidas pelo KAIROS. Porém, as outras funcionalidades e procedimentos internos são validados indiretamente, ao menos de forma parcial, devido à relação de dependência entre os componentes.

Isto é, o “Estudo 1A” ao avaliar a funcionalidade de “Identificação de Áreas Promissoras”, estará avaliando indiretamente os procedimentos internos “Extrapolação Matemática da Curva da Produção das Áreas” e “Identificação de Áreas Associadas à Produção Científica”. O “Estudo 1B” é aplicado sobre a funcionalidade “Identificação de Pesquisadores e Instituições Visionários”. Por consequência, os procedimentos internos “Heurística para Identificação de Visionários” e “Identificação de Áreas Associadas à Produção Científica” estarão sendo indiretamente avaliados. O “Estudo 2”, aplicado sobre o procedimento interno de “Identificação de Competências do Pesquisador ou Instituição”, avalia indiretamente o procedimento interno “Identificação de Áreas Associadas à Produção Científica”.

As funcionalidades “Mapeamento e Prospecção das Áreas”, “Similaridade de Áreas”, “Alinhamento de Pesquisadores e Instituições com Áreas Promissoras” e “Busca por Conhecimento” são, portanto, avaliadas indiretamente, uma vez que dependem de procedimentos internos validados anteriormente.

Segundo BARROS *et al.* (2002), a realização de um estudo experimental pode ser dividida em cinco fases: definição, planejamento, execução, análise e empacotamento. O planejamento descreve os participantes, os instrumentos, o processo de execução e uma avaliação dos problemas que existir haver ao longo da execução do estudo. A execução consiste na realização em si do estudo pelos participantes, enquanto que a análise consiste na organização dos resultados obtidos. Finalmente, o empacotamento tem o objetivo na organizar e armazenar os documentos construídos nas etapas anteriores, para facilitar a repetição do estudo experimental no futuro, se necessário. Os planos dos experimentos propostos seguiram o esse modelo.

5.2 Preparação dos Dados

Para realização dos estudos de caso, foi utilizado o domínio da área de computação. Para isso, foram importados dados relativos a essa área, nas fontes mineradas pelo KAIROS: Curriculum Lattes, DBLP, teses e dissertações da base Capes e patentes da base Derwent.

Na fonte de dados Curriculum Lattes, foram minerados, no período de 16/02/2009 até 18/02/2009, todos os 350 pesquisadores de instituições de pesquisa da área de Computação que obtiveram notas 5, 6 ou 7 na última avaliação trienal da Capes⁶ (CAPES, 2009), ocorrida em 2007, referente aos anos de 2004, 2005 e 2006. Essas instituições, portanto, são consideradas centros de excelência no cenário nacional:

- Pontifícia Universidade Católica do Rio de Janeiro (Nota 7)
- Universidade Federal do Rio de Janeiro – COPPE (Nota 7)
- Universidade Federal de Pernambuco (Nota 6)
- Universidade Federal de Minas Gerais (Nota 6)
- Universidade Federal do Rio Grande do Sul (Nota 6)
- Universidade Federal Fluminense (Nota 5)
- Universidade Estadual de Campinas (Nota 5)
- Universidade de São Paulo (Nota 5)
- Universidade de São Paulo – São Carlos (Nota 5)

Na fonte de dados DBLP, foram minerados os dados mais recentes, disponibilizados em 26/05/2008 pela Universität Trier. No total, foram importadas 1.042.256 publicações de 628.191 diferentes autores. Não foi necessário realizar nenhum tipo de seleção sobre os dados, já que a fonte do DBLP é específica do domínio de computação.

Na fonte de dados da base Capes, foram mineradas teses e dissertações da área de Computação das instituições com notas 5, 6 e 7 obtidas na avaliação trienal da Capes. A importação ocorreu no período de 02/05/2009 até 10/05/2009 e foram mineradas 6.904 teses e dissertações.

Na fonte de dados da base Derwent, foram mineradas 7.743 patentes da área de computação, com classificação Derwent igual a *T01-Digital Computers*. A mineração ocorreu de 12/01/2009 à 17/01/2009.

Com todas as informações mineradas, realizou-se o procedimento de carga automática do dicionário de radicais. Foram identificadas automaticamente 72 áreas de conhecimento, porém 7 não faziam parte do domínio de computação e foram excluídas manualmente do dicionário de radicais.

⁶ A avaliação trienal da Capes tem por objetivo medir a qualidade dos programas de pós-graduação brasileiros. A avaliação considera, entre outros fatores, a qualificação dos docentes, a quantidade e qualidade das publicações científicas e o impacto social do programa (CAPES, 2009).

Foi realizado o procedimento de inferência de áreas, que classificou automaticamente as áreas de conhecimento das produções científicas e em seguida as competências dos pesquisadores. Após a consolidação e prospecção dos dados, as seguintes áreas tiveram crescimento positivo estimado para os próximos 3 anos e foram consideradas promissoras pelo KAIROS:

- Agentes inteligentes
- Algoritmos distribuídos
- Análise numérica
- Aprendizagem de máquina
- Bioinformática
- Computação de alto Desempenho
- Computação em grids ou grades
- Computação gráfica
- CSCW (Trabalho Cooperativo Suportado por Computador)
- Implementação de algoritmos criptográficos
- Jogos 3D
- Linguagens de programação
- Lógica aplicada à computação
- Métodos ágeis de desenvolvimento de software
- Mineração de dados
- Otimização
- Processamento de imagens
- Processamento de sinais biológicos
- Processamento e análise de imagens médicas
- Processamento paralelo e distribuído
- Programação distribuída
- Programação não linear
- Redes de computadores
- Redes sem fio
- Robótica
- Sistemas autônomos inteligentes
- Sistemas de informação

- Software básico
- Teoria dos grafos
- Visão computacional

A funcionalidade de identificação de visionários do KAIROS, com parâmetros de seleção de 25% de início da área e ano inicial após 10% das publicações, identificou 9 pesquisadores visionários nas seguintes subáreas básicas da computação: Arquitetura e sistemas operacionais, algoritmos, banco de dados, engenharia de software, otimização, computação gráfica, inteligência artificial e redes de computadores. Os parâmetros de seleção da funcionalidade de identificação de visionários são explicados detalhadamente na seção 4.10. Essas subáreas foram escolhidas por serem consideradas linhas de pesquisa básicas de diversos programas de pós-graduação.

Por motivo de privacidade, os nomes dos pesquisadores considerados visionários não são apresentados no texto desta dissertação.

5.3 Estudo 1: Áreas Promissoras e Pesquisadores Visionários

5.3.1 Definição

O contexto global considerado nesse estudo é que ambientes de apoio à prospecção tecnológica até o momento não apresentam resultados claros e satisfatórios nem mecanismos para identificação de pesquisadores visionários.

O contexto local desse estudo tem por objetivo avaliar a qualidade da prospecção tecnológica realizada pelo KAIROS, bem como avaliar a heurística de identificação de pesquisadores visionários.

Seguindo as diretrizes definidas por BARROS *et al.*(2002), podemos ressaltar:

- Objeto de Estudo: resultado da consulta de áreas promissoras (horizonte de 3 anos) e pesquisadores visionários do KAIROS.
- Objetivo: identificar a viabilidade da utilização do KAIROS como ferramenta de apoio à prospecção tecnológica e identificação de pesquisadores visionários.
- Foco de Qualidade: mensurar a qualidade dos dados obtidos nas consultas do KAIROS e identificar áreas promissoras ou pesquisadores visionários não inferidos automaticamente.
- Perspectiva: o estudo será desenvolvido sob a ótica do profissional da ciência, professores de pós-graduações na área de Computação, cujas

instituições são avaliadas com nota 5, 6 ou 7 pela CAPES e pesquisadores associados a essas instituições, que executam atividades científicas, avaliando a qualidade dos dados obtidos.

- Contexto: A avaliação de resultados proveniente do processamento de dados reais.

Seguindo a notação GoalQuestionMetric (GQM) (SOLINGEN e BERGHOUT, 1999), a definição do estudo é:

Analisar a qualidade dos dados de prospecção tecnológica (áreas promissoras) e de identificação de pesquisadores visionários obtidos pelas consultas do KAIROS.

Com o propósito de recolher requisitos de melhorias e avaliar a viabilidade do uso do KAIROS no cenário científico.

Referente aos possíveis ganhos obtidos por seu uso.

Do ponto de vista do pesquisador.

No contexto de prospecção tecnológica.

5.3.2 Planejamento

- **Cenário Utilizado**

O estudo investigativo contou com a participação de professores de pós-graduações na área de Computação, cujas instituições são avaliadas com nota 5, 6 ou 7 pela CAPES e pesquisadores associados a essas instituições, que responderam um formulário que avaliou a qualidade dos dados de prospecção tecnológica (horizonte de 3 anos) e pesquisadores possivelmente visionários.

- **Participantes**

Os participantes do estudo foram professores, alunos de pós-graduação e pesquisadores associados às universidades e centros de pesquisa e empresas.

Receberam individualmente convite para participação no estudo, todos os 350 professores da área de computação de universidades avaliadas com nota 5, 6 ou 7 na avaliação trienal da Capes, que tinham seu endereço eletrônico divulgado publicamente na página *web* da instituição. O convite também foi enviado para alguns pesquisadores e colaboradores associados a essas instituições.

- **Treinamento**

Não foi necessário treinamento, já que a execução do estudo se deu com o preenchimento de um formulário simples no formato eletrônico.

- **Instrumentos**

Os participantes não necessitaram de nenhum instrumento especial para realização do estudo, salvo computador com acesso à internet para preenchimento do formulário.

- **Critérios**

O foco de qualidade do estudo exige critérios que avaliem os dados obtidos. As áreas promissoras e pesquisadores visionários identificados foram apresentados no formulário e avaliados, em uma escala de 1 a 5 (1 para ‘discordo completamente’ e 5 para ‘concordo completamente’, além da opção ‘Não apto a responder’) pelos participantes, segundo o grau de concordância das afirmações. Ou seja, o participante deveria avaliar em 5 uma área ou pesquisador, caso ele concorde completamente que a área é promissora ou que o pesquisador é visionário. Da mesma forma, deveria avaliar em 1 se não discordar completamente com a afirmação contida no formulário.

O formulário também continha espaço para que os participantes listassem as áreas promissoras ou os pesquisadores visionários não identificados pela pesquisa.

- **Hipótese Nula**

A hipótese nula é uma afirmativa que o estudo experimental tem como objetivo negar. No estudo atual, a hipótese nula determina que os dados obtidos no KAIROS são de baixa qualidade e não servem para apoiar a prospecção tecnológica e a identificação de visionários.

- **Hipótese Alternativa**

A hipótese alternativa é uma afirmativa que nega a hipótese nula. O estudo experimental tem como objetivo provar a hipótese alternativa, refutando assim a hipótese nula. No estudo atual, a hipótese alternativa determina que os participantes do estudo que avaliaram os resultados do KAIROS julgaram os resultados como sendo de boa qualidade, ou seja, que as áreas apresentadas são promissoras e os pesquisadores são visionários.

Mais especificamente, o estudo tem o objetivo de provar que o KAIROS pode ser utilizado como um instrumento de apoio ao processo de prospecção tecnológica.

- **Variáveis Independentes**

O vínculo do participante (professor, aluno ou pesquisador associado ao setor acadêmico, industrial ou do ramo de serviços), a experiência e a titulação dos participantes (medidas em uma escala nominal) e a área principal de pesquisa

(computação, interdisciplinar relacionada com computação ou outra área) são informações independentes coletadas durante o estudo.

- **Variáveis Dependentes**

A concordância das afirmações sobre áreas promissoras e pesquisadores visionários, é considerada variável dependente.

- **Análise Qualitativa**

Tem o objetivo de avaliar as áreas e pesquisadores que não foram listados nos resultados do KAIROS, mas que deveriam ser listados de acordo com o critério do participante.

- **Capacidade Aleatória**

Apesar de a pesquisa ter sido direcionada a professores e pesquisadores de instituições renomadas, os indivíduos que efetivamente responderam a pesquisa estão aleatoriamente dispersos no universo dos possíveis participantes do estudo de caso.

- **Mecanismo de Análise**

As variáveis dependentes são apresentadas utilizando-se a escala de 1 a 5. Além disto, os resultados são agrupados de forma que as funcionalidades possam ser avaliadas.

- **Validade Interna do Estudo**

A validade interna de um estudo é definida como a capacidade de um novo estudo repetir o comportamento do estudo atual com os mesmos participantes e objetos com que ele foi realizado. Considera-se um bom nível de validação interna, já que a percepção dos participantes não tende a mudar em um curto espaço de tempo.

- **Validade Externa do Estudo**

A validade externa do estudo mede sua capacidade de refletir o mesmo comportamento em outros grupos de participantes de mesmo perfil, ou seja, em outros grupos além daquele em que o estudo foi aplicado. A validade externa do estudo é considerada suficiente, dado a aleatoriedade de escolha e o número de participantes.

- **Validade de Construção do Estudo**

A validade de construção do estudo se refere à relação entre os instrumentos e participantes do estudo e a teoria que está sendo provada por este. Neste caso, todos os participantes são do meio científico. Os dados utilizados para demonstração são reais e dizem respeito à produção científica da área de computação.

• Validade de Conclusão do Estudo

A validade de conclusão do estudo mede a capacidade do estudo em gerar alguma conclusão. O estudo utiliza critérios objetivos, propiciando resultados claros.

5.3.3 Execução

Um formulário eletrônico, disponibilizado no período de 02/06/2009 até 06/06/2009, no endereço <http://kairos-coppe.no-ip.org:8000/>, foi respondido pelos participantes, que receberam convite individual por email. No total, participaram da pesquisa 57 pesquisadores.

Os formulários eletrônicos garantiam a confidencialidade e anonimato das respostas, uma vez que nenhuma identificação era exigida por parte do participante. Para evitar que um único participante respondesse o questionário mais de uma vez, o endereço enviado por *e-mail* continha uma senha de acesso, capaz de habilitar formulário apenas uma vez. Dessa forma, garantiu-se que nenhum participante votou mais de uma vez. A Figura 38 apresenta a tela inicial do formulário eletrônico disponibilizado. No Apêndice C, o mesmo formulário é apresentado em detalhes.


 PESC	Avaliação da Prospecção de Áreas Científicas e da Heurística para Identificação de Visionários														
Prezado pesquisador,															
Este questionário tem como objetivo avaliar resultados obtidos em pesquisas realizadas na Linha de Banco de Dados do Programa de Engenharia de Sistemas e Computação da COPPE UFRJ na área de Gestão do Conhecimento Científico.															
Este trabalho pretende identificar o comportamento das áreas científicas ao longo do tempo e prospectar o comportamento futuro a partir de dados históricos da produção científica. Além disso, prevê uma heurística para identificação de pesquisadores visionários nas diferentes áreas da ciência.															
Os questionários estão divididos em 2 etapas: a primeira etapa avalia a prospecção das áreas científicas, enquanto que a segunda etapa avalia uma heurística para identificação de pesquisadores visionários.															
As questões optativas de ambas as etapas seguem o seguinte padrão de resposta:															
<table border="1"><thead><tr><th colspan="2">Padrão de Resposta</th></tr></thead><tbody><tr><td>N/A</td><td>Não apto a responder (Desconheço)</td></tr><tr><td>1</td><td>Discordo completamente</td></tr><tr><td>2</td><td>Discordo parcialmente</td></tr><tr><td>3</td><td>Nem concordo, nem discordo</td></tr><tr><td>4</td><td>Concordo parcialmente</td></tr><tr><td>5</td><td>Concordo completamente</td></tr></tbody></table>		Padrão de Resposta		N/A	Não apto a responder (Desconheço)	1	Discordo completamente	2	Discordo parcialmente	3	Nem concordo, nem discordo	4	Concordo parcialmente	5	Concordo completamente
Padrão de Resposta															
N/A	Não apto a responder (Desconheço)														
1	Discordo completamente														
2	Discordo parcialmente														
3	Nem concordo, nem discordo														
4	Concordo parcialmente														
5	Concordo completamente														
Quaisquer dúvidas podem ser enviadas diretamente para o aluno responsável: Rafael De Marino (rmartino@gmail.com ou rmartino@cos.ufrj.br) Orientadores do trabalho: Prof. Jano Moreira de Souza (PESC/COPPE/UFRJ) e Prof. ^a . Joice de Oliveira Sampaio (DCC/IM/UFRJ).															
Agradecemos a colaboração.															
<input type="button" value="Iniciar Pesquisa"/>															

Figura 38: Tela inicial do formulário eletrônico do Estudo de Caso 1.

5.3.4 Análise de Resultados

5.3.4.1 Perfil dos Participantes

Na primeira etapa do estudo, os participantes foram convidados a preencher informações pessoais relativas à atuação profissional. Foram expostas 4 questões:

envolvimento com a pesquisa, titulação, tempo de experiência e principal área de atuação. As respostas são apresentadas em forma de gráfico. A Figura 39 apresenta o tipo de envolvimento com a pesquisa dos participantes. A Figura 40 apresenta a titulação dos participantes. A Figura 41 apresenta a relação entre o vínculo com a pesquisa e a titulação. A Figura 42 apresenta o tempo de experiência dos participantes. Por fim, a Figura 43 apresenta a área de atuação principal dos respondentes da pesquisa.

Observa-se, portanto, o perfil predominante do participante do estudo: pesquisador docente da área de computação, doutor e com mais de 15 anos de experiência no setor acadêmico.

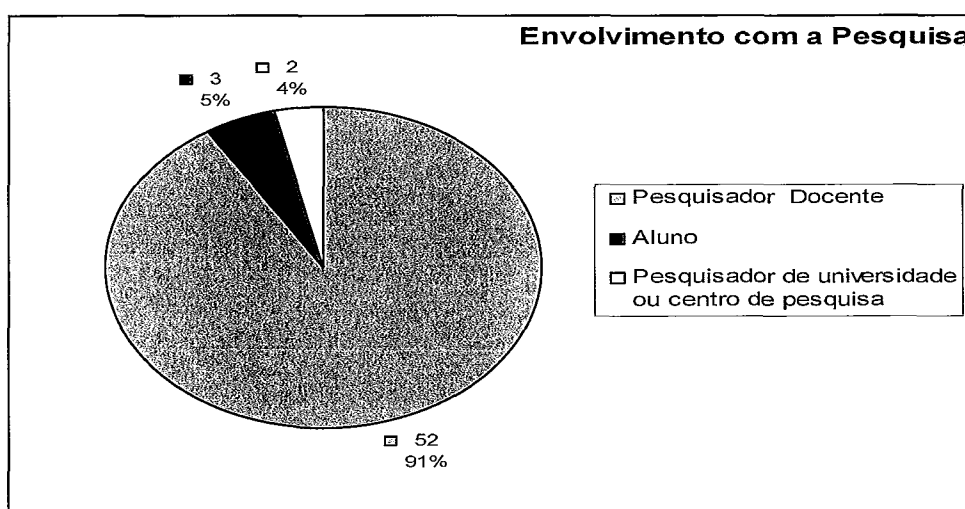


Figura 39: Gráfico das respostas sobre Envolvimento com a Pesquisa

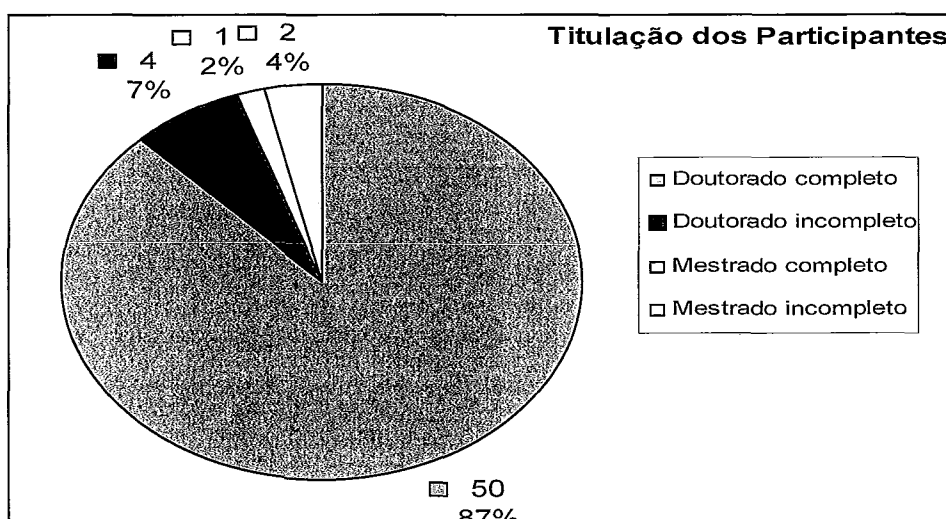


Figura 40: Gráfico das respostas sobre Titulação dos Participantes.

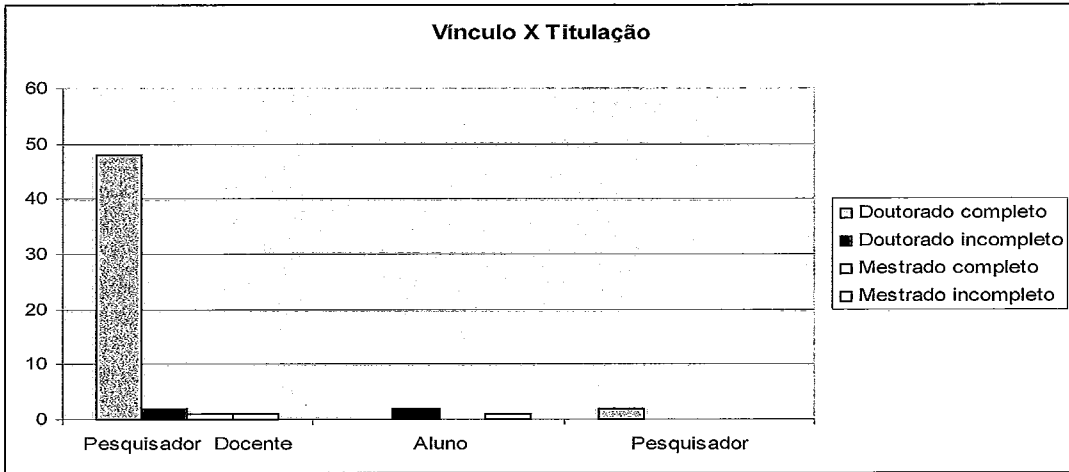


Figura 41: Gráfico da relação entre o vínculo com a pesquisa e a titulação.

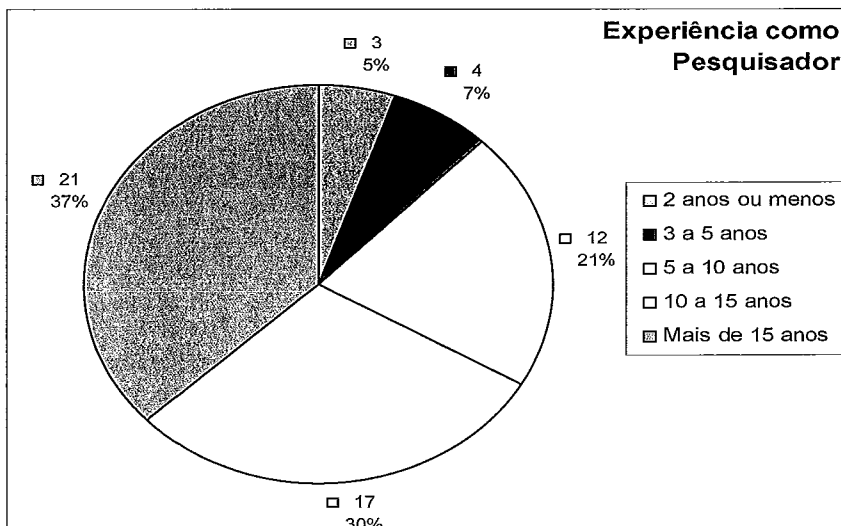


Figura 42: Gráfico das respostas sobre Experiência como Pesquisador.

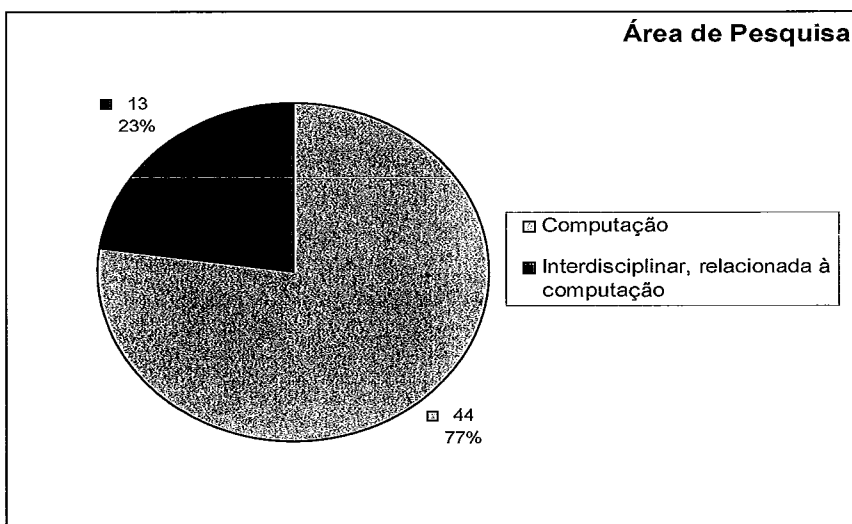


Figura 43: Gráfico das respostas sobre Principal Área de Pesquisa.

5.3.4.2 Análise das Respostas

As respostas nominais referentes ao grau de concordância do participante com as áreas promissoras e pesquisadores visionários foram tabuladas, obtendo-se a média e o desvio padrão da nota de cada área. Desse cálculo, foram excluídas as notas respondidas como “N/A” (Desconheço). No Apêndice C são apresentadas detalhadamente as respostas de cada área e pesquisador.

Além da nota média e do desvio padrão, definiu-se um índice, denominado “Favorabilidade”, que tem o objetivo de medir o percentual de pessoas que possuem opinião definida e concordam com a proposição da área ser promissora ou o pesquisador visionário. Esse critério mede o percentual de respostas 4 ou 5 (indicativas de concordância) sobre o universo de respostas 1, 2, 4 e 5 (todas as respostas explicitamente de concordância ou discordância).

Equação 4: Cálculo da favorabilidade de uma área.

$$\text{Favorabilidade da área A} = \frac{\text{Número de Notas 4 e 5 da área A}}{\text{Número de Notas 1, 2, 4 e 5 da área A}} \times 100$$

A Tabela 11 apresenta os resultados das respostas nominais da primeira etapa do estudo 1, referente à avaliação das áreas promissoras.

Tabela 11: Sumarização das respostas nominais da primeira etapa do Estudo 1.

Área	Nota Média	Desvio Padrão	Favorabilidade
Agentes Inteligentes	3,750	1,102	77,1%
Algoritmos Distribuídos	3,898	0,770	94,7%
Análise Numérica	2,872	1,196	44%
Aprendizagem de Máquina	3,863	0,938	90%
Bioinformática	4,536	0,712	100%
Computação de Alto Desempenho	4,043	0,832	100%
Computação Em Grids Ou Grades	4,000	0,834	94,6%
Computação Gráfica	3,911	0,820	94,1%
CSCW (Trabalho Cooperativo Sup. por Computador)	3,636	0,966	83,9%
Implementação de Algoritmos Criptográficos	3,622	1,072	75%
Jogos 3D	3,872	0,969	89,2%

Linguagens de Programação	2,478	0,887	20%
Lógica aplicada à computação	2,500	1,130	27,8%
Métodos Ágeis de Desenvolvimento de Software	3,468	1,039	73,3%
Mineração de Dados	4,154	0,801	97,6%
Otimização	3,511	0,952	76,6%
Processamento de Imagens	4,000	0,840	93%
Processamento de Sinais Biológicos	4,222	0,974	94,6%
Processamento e Análise de Imagens Médicas	4,128	0,991	92,3%
Processamento Paralelo e Distribuído	3,813	0,866	94,1%
Programação Distribuída	3,646	0,862	87,1%
Programação Não-Linear	2,912	0,865	42,1%
Redes de Computadores	3,455	0,975	75,8%
Redes sem fio	4,128	0,991	88,1%
Robótica	4,196	0,909	94,6%
Sistemas Autônomos Inteligentes	4,000	0,879	94,1%
Sistemas de Informação	3,102	1,176	55,9%
Software Básico	2,220	0,935	15,1%
Teoria dos Grafos	2,500	1,033	29%
Visão Computacional	4,000	0,910	91,2%

A Tabela 12 apresenta os resultados consolidados: a nota média de todas as áreas, o desvio padrão médio de todas as respostas e o percentual de áreas que a favorabilidade foi maior que 50%, ou seja, o percentual de áreas validadas pelos participantes com opinião formada.

Tabela 12: Consolidação das respostas da primeira etapa do Estudo 1.

Nota média de todas das áreas	3,65
Desvio padrão médio de todas as áreas	1,10
Percentual de áreas com Favorabilidade maior que 50%	80%

Os participantes também informaram quais áreas consideravam promissoras e que não foram listadas pelo KAIROS. No total, foram listadas 39 áreas, nas quais 26 diferentes. Em média 0,68 áreas foram listadas por cada pesquisador.

A Tabela 13 apresenta as áreas listadas pelos participantes com o número de citações na pesquisa. A última coluna indica se a área foi identificada pelo KAIROS

como uma área não-promissora, ou seja, uma área que apresentou estimativa de decrescimento no horizonte dos próximos 3 anos.

Tabela 13: Áreas promissoras sugeridas pelos participantes do estudo.

Área	Número citações	Identificada KAIROS?
Arquiteturas Orientadas a Serviços (SOA)	1	Não
Computação em nuvem (Cloud Computing)	1	Não
Computação móvel e ubíqua	5	Não
Desenvolvimento baseado em componentes	1	Não
Desenvolvimento de software confiável	3	Não
Engenharia de software experimental	1	Não
Gerenciamento de riscos em TI	1	Não
Gestão do conhecimento	2	Sim
Gestão do conhecimento distribuído	1	Não
Informática na educação	1	Sim
Interface humano computador	4	Sim
Métodos de desenv. orientados à modelos (MDA)	1	Não
Modelagem de sistemas complexos	2	Sim
Negociação em Projetos de Software	1	Não
Processamento de informação por sistemas naturais.	1	Sim
Programação orientada a aspectos	1	Não
Realidade virtual e aumentada	1	Não
Recuperação de informação	1	Não
Redes de sensores	1	Não
Redes neurais artificiais	1	Sim
Reuso de software	1	Não
Sistemas embarcados	2	Sim
Software livre	3	Não
Teoria de jogos	1	Não
Web semântica	1	Não
Workflows científicos	1	Não

A Tabela 14 apresenta os resultados das respostas nominais da segunda etapa do estudo 1, referente à avaliação dos pesquisadores visionários. Os nomes dos pesquisadores, por questão de privacidade, estão omitidos.

Os participantes, como na primeira etapa, informaram pesquisadores consideravam visionários e que não foram listadas pelo KAIROS. No total, foram listados 59 pesquisadores, nas quais 29 diferentes. Em média, foram listados 1,04 visionários por cada pesquisador respondente do questionário.

Dentre esses 29 pesquisadores, um pesquisador do foi citado 15 vezes, um pesquisador foi citado 5 vezes, um pesquisador foi citado 4 vezes, três pesquisadores foram citados 3 vezes, três pesquisadores foram citados 2 vezes e os 20 pesquisadores restantes foram citados apenas uma vez.

Tabela 14: Sumarização das respostas nominais da segunda etapa do Estudo 1.

Pesquisador	Nota Média	Desvio Padrão	Favorabilidade
Pesquisador 1	4,142857143	1,276	88,9%
Pesquisador 2	3,461538462	1,066	81,3%
Pesquisador 3	3,142857143	1,152	60%
Pesquisador 4	4,361111111	0,833	94,1%
Pesquisador 5	3,689655172	0,967	86,4%
Pesquisador 6	4,034482759	0,905	91,7%
Pesquisador 7	3,625	1,095	81,3%
Pesquisador 8	3,523809524	1,077	75%
Pesquisador 9	3,136363636	1,037	57,1%

A Tabela 15 apresenta os resultados consolidados: a nota média de todos os pesquisadores, o desvio padrão médio de todas as respostas e o percentual de pesquisadores que a favorabilidade foi maior que 50%, ou seja, o percentual de pesquisadores visionários validados pelos participantes com opinião formada.

Tabela 15: Consolidação das respostas da segunda etapa do Estudo 1.

Nota média de todos os pesquisadores	3,73
Desvio padrão médio de todos os pesquisadores	1,09
Percentual de pesquisadores com Favorabilidade maior que 50%	100%

5.4 Estudo 2: Identificação de Competências

5.4.1 Definição

O contexto global considerado nesse estudo é que sistemas para inferência automática de competências até o momento não apresentam resultados satisfatórios.

O contexto local desse estudo tem como objetivo avaliar a qualidade da identificação de competência de pesquisadores realizada pelo KAIROS.

Seguindo as diretrizes definidas por BARROS *et al.* (2002), podemos ressaltar:

- Objeto de Estudo: competências mapeadas para os pesquisadores do Programa e Engenharia de Sistemas e Computação (PESC) do Instituto Alberto Luiz Coimbra de Pós-Graduação e Pesquisa de Engenharia (COPPE) da Universidade Federal do Rio de Janeiro (UFRJ) no KAIROS.
- Objetivo: identificar a viabilidade da utilização do KAIROS como ferramenta de apoio à identificação de especialistas.
- Foco de Qualidade: mensurar a qualidade dos dados obtidos pelo KAIROS e identificar outras competências que deveriam ter sido identificadas e não foram.
- Perspectiva: o estudo será desenvolvido sob a ótica do profissional da Ciência, que executa atividades científicas, avaliando a qualidade dos dados obtidos.
- Contexto: A avaliação de dados proveniente do processamento de dados reais.

Seguindo a notação GoalQuestionMetric (GQM) (SOLINGEN e BERGHOUT, 1999), a definição do estudo é:

Analisar a qualidade dos dados de identificação de competências dos pesquisadores obtidos pelo KAIROS.

Com o propósito de recolher requisitos de melhorias e avaliar a viabilidade do uso do KAIROS para identificação de especialistas e mapeamento de competências.

Referente aos possíveis ganhos obtidos por seu uso.

Do ponto de vista do pesquisador.

No contexto de identificação de especialistas e mapeamento de competências.

5.4.2 Planejamento

- **Cenário Utilizado**

O estudo investigativo contou com a participação de um pesquisador da área de computação, que avaliou se as competências dos professores do Programa de Engenharia de Sistemas e Computação da COPPE/UFRJ, mapeadas pelo KAIROS, estão de acordo com a área de atuação informada pelos próprios pesquisadores na página *web* do PESC (PESC, 2009).

O pesquisador responsável por conduzir o estudo possuía experiência na área de computação, para ser capaz de mapear áreas eventualmente com nomes diferentes, porém com mesma semântica. Por exemplo: “Redes Neurais” X “Redes Neurais Artificiais” ou “Redes sem fio” X “Redes *wireless*”.

- **Participantes**

Esse estudo contou com a participação de apenas um pesquisador da área de computação.

- **Treinamento**

Não foi necessário treinamento, já que a execução do estudo consiste na simples verificação dos resultados obtidos pelo KAIROS e pelos dados cadastrados na página do PESC. O conhecimento prévio do pesquisador na área de computação foi o bastante.

- **Instrumentos**

Foi necessário um computador com acesso à internet. O pesquisador registrou o resultado da comparação em uma planilha simples.

- **Critérios**

O foco de qualidade do estudo exige critérios que avaliem os dados obtidos. O critério utilizado foi o percentual de áreas de atuação do pesquisador contidas na página *web* do PESC mapeadas satisfatoriamente pelo KAIROS.

Portanto, cada pesquisador obteve um percentual de mapeamento bem sucedido.

- **Hipótese Nula**

A hipótese nula é uma afirmativa que o estudo experimental tem como objetivo negar. No estudo atual, a hipótese nula determina que as competências mapeadas no KAIROS não correspondem à realidade.

- **Hipótese Alternativa**

A hipótese alternativa é uma afirmativa que nega a hipótese nula. O estudo experimental tem como objetivo provar a hipótese alternativa, refutando assim a

hipótese nula. No estudo atual, a hipótese alternativa determina que as competências mapeadas no KAIROS correspondem à realidade e, portanto, o KAIROS pode ser utilizado como um instrumento de apoio à identificação de especialistas.

- **Variáveis Independentes**

Nesse estudo não existem variáveis independentes.

- **Variáveis Dependentes**

O percentual de mapeamento entre as áreas declaradas na página *web* do PESC e as competências mapeadas no KAIROS é considerado uma variável dependente.

- **Capacidade Aleatória**

A capacidade aleatória é baixa, visto que não é necessária nesse estudo. A participação não exige esforço intelectual intenso e esperam-se resultados iguais mesmo com participantes diferentes.

- **Mecanismo de Análise**

A variável dependente é apresentada em forma de percentual. Além disto, os resultados são agrupados de forma que a funcionalidade de identificação de competência possa ser avaliada.

- **Validade Interna do Estudo**

A validade interna de um estudo é definida como a capacidade de um novo estudo repetir o comportamento do estudo atual com os mesmos participantes e objetos com que ele foi realizado. Considera-se altíssimo o nível de validação interna, já que a percepção do participante influi pouco no estudo.

- **Validade Externa do Estudo**

A validade externa do estudo mede sua capacidade de refletir o mesmo comportamento em outros grupos de participantes, ou seja, em outros grupos além daquele em que o estudo foi aplicado. A validade externa do estudo é considerada suficiente, dado que a tarefa executada pelo participante é trivial e sua percepção pouco influi nos resultados.

- **Validade de Construção do Estudo**

A validade de construção do estudo se refere à relação entre os instrumentos e participantes do estudo e a teoria que está sendo provada por este. Neste caso, o participante é do meio científico e os dados utilizados no estudo são reais e dizem respeito às competências de pesquisadores.

- **Validade de Conclusão do Estudo**

A validade de conclusão do estudo mede a capacidade do mesmo em gerar alguma conclusão. O estudo utiliza um critério objetivo, propiciando resultados diretos.

5.4.3 Execução

O estudo foi realizado conforme o planejamento, no dia 04/06/2009, durante cerca de duas horas, por um único participante, que utilizou um computador com acesso à internet para acessar a página *web* do PESC e o KAIROS.

Foram comparados dados de 35 pesquisadores docentes. Os resultados foram registrados em uma planilha simples, apresentada no Apêndice C.

5.4.4 Análise de Resultados

Os resultados da comparação entre as competências mapeadas pelo KAIROS e as áreas de atuação informadas na página *web* do PESC são apresentadas na Tabela 16. Para cada pesquisador, é apresentada a principal área de atuação, a confirmação se a principal área foi mapeada e o percentual de mapeamento de todas as áreas contidas na página *web* do PESC. A linha de pesquisa do PESC na qual o pesquisador atua, foi considerada a principal área de atuação. O nome dos pesquisadores, por motivo de privacidade, novamente foi omitido.

Tabela 16: Resultados consolidados do Estudo 2.

Pesquisador	Principal Área de Atuação	Principal área mapeada?	Percentual de acerto do mapeamento
Pesquisador 1	Otimização	Sim	60%
Pesquisador 2	Engenharia de Software	Sim	25%
Pesquisador 3	Computação Gráfica	Sim	100%
Pesquisador 4	Algoritmos e Combinatória	Sim	100%
Pesquisador 5	Engenharia de Software	Sim	33,3%
Pesquisador 6	Computação Gráfica	Sim	87,5%
Pesquisador 7	Arquitetura e Sist. Operacionais	Sim	100%
Pesquisador 8	Redes de Computadores	Sim	100%
Pesquisador 9	Redes de Computadores	Sim	71,4%
Pesquisador 10	Otimização	Sim	50%
Pesquisador 11	Algoritmos e Combinatória	Sim	100%
Pesquisador 12	Inteligência Artificial	Sim	100%

Pesquisador 13	Redes de Computadores	Sim	42,8%
Pesquisador 14	Otimização	Sim	60%
Pesquisador 15	Algoritmos e Combinatória	Sim	100%
Pesquisador 16	Inteligência Artificial	Não	50%
Pesquisador 17	Banco de Dados	Sim	33,3%
Pesquisador 18	Otimização	Sim	50%
Pesquisador 19	Arquitetura e Sist. Operacionais	Sim	66,6%
Pesquisador 20	Banco de Dados	Sim	50%
Pesquisador 21	Banco de Dados	Sim	50%
Pesquisador 22	Inteligência Artificial	Sim	100%
Pesquisador 23	Engenharia de Software	Sim	28,6%
Pesquisador 24	Informática e Sociedade	Não	25%
Pesquisador 25	Banco de Dados	Sim	50%
Pesquisador 26	Inteligência Artificial	Sim	87,5%
Pesquisador 27	Otimização	Sim	28,6%
Pesquisador 28	Computação Gráfica	Sim	40%
Pesquisador 29	Computação Gráfica	Sim	66,6%
Pesquisador 30	Redes de Computadores	Sim	50%
Pesquisador 31	Otimização	Sim	20%
Pesquisador 32	Algoritmos e Combinatória	Sim	40%
Pesquisador 33	Algoritmos e Combinatória	Sim	60%
Pesquisador 34	Otimização	Sim	16,7%
Pesquisador 35	Inteligência Artificial	Sim	60%
Média do percentual de acerto do mapeamento			60,1%
Desvio padrão do percentual de acerto do mapeamento			27,5
Percentual de acerto da área principal			94,3%

No Apêndice C são apresentados detalhadamente os resultados desse estudo. Para cada pesquisador, são apresentadas as áreas de atuação informadas na página *web* do PESC, junto com a confirmação de mapeamento pelo KAIROS.

Grande parte das áreas, informadas na página *web* do PESC, que não foram mapeadas, são muito específicas. O fato de alguns pesquisadores detalharem muito sua área de atuação, fez com que o desvio padrão de acerto do mapeamento das áreas fosse

alto. Ou seja, alguns pesquisadores, que detalharam muito sua área de atuação, tiveram mapeamento de poucas áreas no KAIROS, enquanto que pesquisadores que especificaram suas áreas em um nível de granularidade razoável, tiveram muitas áreas mapeadas.

Como exemplo, pode-se citar o Pesquisador 34, que informou as seguintes áreas de atuação, que estão em um nível de granularidade maior que as competências cadastradas no KAIROS: Otimização não-diferenciável, programação matemática de dois níveis, programação matemática com restrições de equilíbrio, desigualdade variacional e desigualdade variacional generalizada.

Os pesquisadores que atuam em áreas interdisciplinares não obtiveram grande êxito no mapeamento de competências do KAIROS, já que o dicionário de radicais, nos estudos de caso desse capítulo, foi carregado com áreas do domínio de computação apenas.

Como exemplo, pode-se citar o Pesquisador 31, que informou, na página *web* do PESC, as seguintes áreas de atuação: Teoria quântica de campo em espaços curvos, gravitação e cosmologia, astrofísica relativística e biologia matemática e computacional.

5.5 Conclusão

O Estudo 1 apresentou índices de favorabilidade considerados excelentes na avaliação de áreas promissoras e de pesquisadores visionários. Entretanto, o desvio padrão das respostas de ambas as etapas foi superior a 1, indicando que não existe um consenso alto entre os respondentes da pesquisa.

Na segunda etapa do Estudo 1, foram selecionados pela heurística do KAIROS apenas 9 pesquisadores visionários. O baixo número sugere que os critérios de seleção da heurística devem ser flexibilizados de forma que mais pesquisadores sejam relacionados como visionários. Outro ponto relevante verificado é que a heurística de identificação de visionários privilegia a escolha de pesquisadores mais antigos, já que estes tiveram mais chances de estudar as áreas antes que elas se tornassem popular.

A Figura 44 apresenta graficamente o universo de pesquisadores: os considerados visionários segundo senso comum, os considerados visionários pelos participantes do estudo, os identificados como visionários pelo KAIROS e os considerados não visionários pelo senso comum.

Os resultados sugerem que o conjunto dos pesquisadores identificados pelo KAIROS está contido no conjunto dos visionários segundo percepção dos participantes do estudo de caso, dado que 100% dos pesquisadores obtiveram favorabilidade maior que 50%.

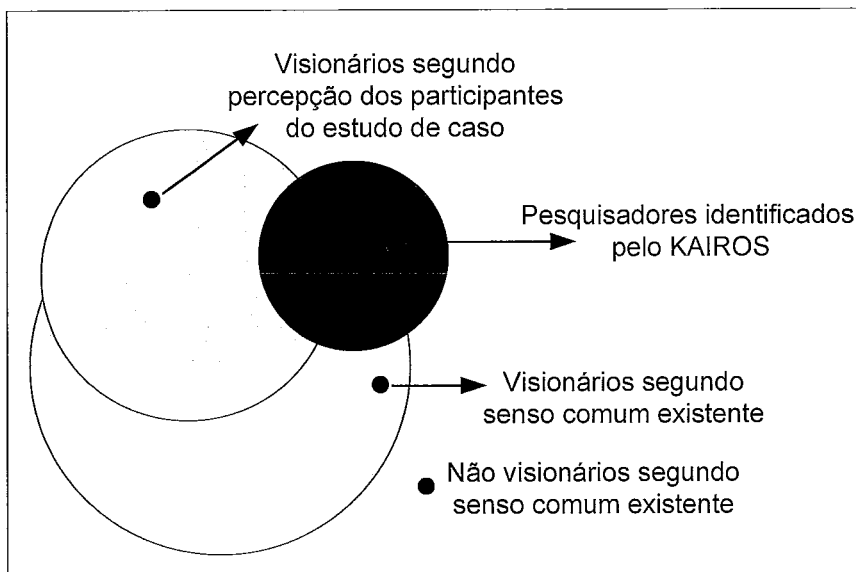


Figura 44: Universo de pesquisadores visionários e não visionários.

O Estudo 2 indicou que o processo de mapeamento de competências do KAIROS apresentou ótimos resultados. A área de atuação principal do pesquisador foi corretamente mapeada em 94,3% das vezes. Cerca de 60% das áreas de atuação informadas pelos pesquisadores foram mapeadas com sucesso.

No próximo capítulo serão apresentadas as considerações finais desta dissertação. Serão expostas as principais contribuições, as principais limitações e as sugestões de trabalhos futuro.

Capítulo 6 - Considerações Finais

Este capítulo finaliza esta dissertação expondo suas principais contribuições, as limitações enfrentadas, bem como sugestões de trabalhos futuros.

6.1 Contribuições

Este trabalho apresentou uma abordagem para prospecção tecnológica e identificação de especialistas, utilizando como insumo, dados de publicações científicas, patentes, teses e dissertações.

Para suportar essa abordagem, foi desenvolvida uma ferramenta computacional, denominada KAIROS, capaz de automatizar todo o processo aquisição e mineração de dados, inferência de áreas de conhecimento, consolidação e prospecção da produção científica, além de fornecer diversas consultas para o usuário final do sistema.

Os resultados obtidos pelas principais funcionalidades do KAIROS foram avaliados através de dois estudos de caso aplicados no domínio de Computação. Esses estudos indicaram que os resultados obtidos são satisfatórios.

As principais contribuições do trabalho são:

- Auxiliar o processo de decisão no investimento em Ciência e Tecnologia: a produção científica das áreas é prospectada e são apresentadas as áreas que possuem estimativa de crescimento no horizonte de três anos. Essas informações podem auxiliar os gestores no direcionamento de recursos de fomento à Ciência e Tecnologia.
- Permitir a identificação de especialistas e visionários: a partir de heurísticas propostas, são apresentados os pesquisadores e instituições considerados especialistas ou visionários em determinada área de conhecimento. Os pesquisadores especialistas ou visionários podem auxiliar a validação da prospecção tecnológica obtida pelo KAIROS ou podem trabalhar no desenvolvimento de novas tecnologias.
- Auxiliar a formação de equipes de trabalho: os pesquisadores identificados como visionários ou especialistas podem ser elencados como líderes de equipes ou chefes de pesquisa, caso possuam perfil para

isto. A experiência e conhecimento desses líderes podem otimizar a produtividade da equipe como um todo.

- Estimular a cooperação entre instituições: o mapeamento de competências de pesquisadores e instituições permite que novas oportunidades de parcerias estratégicas sejam identificadas e concretizadas.
- Identificar o alinhamento com áreas promissoras: os pesquisadores podem consultar se a sua área de atuação no presente está alinhada com as áreas consideradas promissoras no futuro. Dessa forma, podem direcionar suas pesquisas em determinados temas.
- Permitir o estudo da evolução das áreas: o KAIROS apresenta o comportamento das áreas científicas ao longo do tempo. É possível visualizar a série histórica da produção científica de determinada área e obter o grau de similaridade entre as áreas ao longo dos anos.
- Inferência automática das áreas de conhecimento associadas à produção científica: toda produção científica é minerada e classificada automaticamente, podendo ser recuperada através de mecanismos de busca concebidos. Esse conhecimento passa a estar disponível para toda a organização e auxilia o desenvolvimento da mesma.
- Auxiliar no desenvolvimento de novas tecnologias: a abordagem proposta apóia diretamente o processo de desenvolvimento tecnológico, uma vez que, as áreas com maior possibilidade de retorno são elencadas, os especialistas e visionários em determinado assunto em todo mundo, com capacidade de desenvolver pesquisa de excelência, são identificados, e um vasto material de pesquisa é automaticamente classificado e disponibilizado para consulta de toda a organização. O desenvolvimento de novas tecnologias traz importantes benefícios para o país: i) Alavanca a ciência como um todo, ii) Traz benefícios diretos, quando a tecnologia desenvolvida impacta positivamente na vida da população, iii) Traz benefícios indiretos, quando a tecnologia desenvolvida nacionalmente faz com que a economia local se desenvolva, gerando emprego e renda para a população, iv) Aumenta a arrecadação de impostos, uma vez que a empresa desenvolvedora da

tecnologia obtém vantagem competitiva no mercado, e por consequência, tende a aumentar seu lucro.

Na próxima seção são apresentadas algumas limitações encontradas ao longo do desenvolvimento deste trabalho.

6.2 Limitações

A partir de uma análise crítica sobre a abordagem proposta por este trabalho e sua implementação, foram identificadas dificuldades e limitações. Destacam-se as seguintes:

- Qualidade do dicionário de radicais depende da qualidade do preenchimento do Curriculum Lattes: o dicionário de radicais pode ser carregado automaticamente a partir de informações do Curriculum Lattes, mais precisamente da Especialidade das publicações bibliográficas. Caso o Curriculum Lattes seja preenchido de forma errada, o erro é propagado para o dicionário de radicais. Para contornar essa limitação, existe possibilidade do preenchimento manual do dicionário.
- Áreas de conhecimento cadastradas automaticamente no dicionário de radicais possuem diferentes níveis de granularidade: como a informação da Especialidade da produção bibliográfica do Curriculum Lattes é um texto livre, os pesquisadores podem preencher essa informação em diferentes níveis de granularidade, fazendo com que o dicionário de radicais carregue essa informação igualmente em diferentes níveis de granularidade. Por exemplo, alguns pesquisadores podem preencher a Especialidade de um artigo sobre a linguagem de programação Java, como “Java” ou “Linguagens de Programação”, que conceitualmente estão em diferentes níveis, já que Java é uma linguagem de programação. Para contornar essa limitação, o dicionário de radicais pode ser corrigido manualmente.
- Os idiomas português e inglês são tratados de forma isolada: o cadastro do dicionário de radicais é separado para os idiomas inglês e português. Ou seja, um radical em português pode não ter sua tradução cadastrada no dicionário de radicais no idioma inglês, ou pode ter com relevância

diferente. O ideal seria que o cadastro de radicais fosse único para todos os idiomas, e houvesse tradução das palavras de modo que, dado um radical cadastrado em um idioma, o mesmo radical existisse traduzido nos outros idiomas existentes.

- Os dados obtidos nas diferentes fontes de dados são armazenados em diferentes bases: apesar de existirem consultas que sumarizam os dados minerados em todas as fontes, os dados estão guardados fisicamente em diferentes tabelas. Ou seja, existem pesquisadores cadastrados no Curriculum Lattes e no DBLP, que na base de dados do KAIROS aparecem em duplicidade. Essa dificuldade decorre do fato de que o nome dos pesquisadores pode estar cadastrado de forma diferente em cada fonte de dados (por causa de abreviaturas, por exemplo). Uma solução para esse problema seria confrontar partes dos nomes nos diferentes cadastros, confrontando os nomes abreviados com os nomes escritos por extenso.
- Os extratores *web* desenvolvidos nesta abordagem dependem da apresentação da página: os extratores *web* capturam informações contidas em páginas *web* no formato HTML. Quando o *layout* de apresentação dessas páginas é modificado, os extratores podem parar de funcionar. Por isso, é importante que os extratores sejam constantemente verificados e, caso necessário, atualizados.
- Identificação de visionários privilegia pesquisadores mais antigos: a heurística proposta para identificação de visionários busca os pesquisadores que produziram em determinada área de conhecimento antes dos demais. Dessa forma, os pesquisadores mais antigos possuem mais chance de terem estudado uma área no início do desenvolvimento da mesma. Em contrapartida, um pesquisador novo é considerado visionário apenas quando a área estudada é igualmente nova.

Na próxima seção são apresentadas algumas sugestões de trabalhos futuros.

6.3 Trabalhos Futuros

Para continuidade dessa dissertação, foram elencados possíveis trabalhos futuros:

- Aprimoramento do dicionário de radicais: sugere-se que o dicionário de radicais passe a guardar uma taxonomia de termos, hierarquicamente dispostos em forma de árvore. Essa taxonomia deve ser construída automaticamente através das informações do Curriculum Lattes (como na versão atual) e de informações do relacionamento entre os termos obtidas na *web*, como realizado no trabalho correlato de VELARDI *et al.* (2007). Outra possibilidade é a utilização da base de dados do *WordNet* (FELLBAUM, 1998) para identificação de sinônimos e dos relacionamentos entre termos em inglês.
- Utilização da função cíclica na interpolação dos dados: sugere-se que a interpolação matemática realizada sobre a série histórica de produção científica seja realizada utilizando a função cíclica, descrita na seção 3.6. Verificou-se que essa função é a que melhor representa o padrão apresentado pelas séries históricas de produção científica das áreas, que geralmente crescem até atingir um topo, em seguida decrescem. Essa função de interpolação possivelmente melhorará os resultados da prospecção.
- Estudo das redes sociais científicas: sugere-se que seja construída uma funcionalidade capaz de identificar a similaridade entre os pesquisadores e instituições, de acordo com a proximidade dos temas estudados por eles e da relação de co-autoria de suas produções científicas. Nesta direção, alguns trabalhos estão sendo desenvolvidos no cenário desta dissertação, como os propostos por MONCLAR *et al.* (2009) e STROELE *et al.* (2009a). O primeiro trata-se de uma abordagem para identificar redes sociais científicas, utilizando os dados do Curriculum Lattes e do GCC para identificar possíveis problemas na formação da rede que possam interferir no fluxo do conhecimento científico e para recomendar novos relacionamentos (MONCLAR *et al.*, 2009). O segundo trata-se de uma abordagem para minerar, identificar e visualizar redes sociais científicas,

inter e intra-organizacionais, podendo ainda identificar as áreas de conhecimento presentes em tais redes (STROELE *et al.*, 2009).

- Estudar como as áreas se fundem ou se especializam: sugere-se a construção de uma funcionalidade capaz de identificar o surgimento de áreas oriundas da fusão de outras áreas e do surgimento de áreas resultantes da especialização de outras áreas. Como exemplo de fusão de áreas, bioinformática surgiu como resultado da fusão de áreas do domínio de biologia, com áreas do domínio de informática. Como exemplo de especialização, banco de dados distribuídos e banco de dados orientado a objeto surgiram como especialização da área de banco de dados. Para identificação desses eventos, sugere-se guardar a evolução do dicionário de radicais ao longo do tempo, realizando a carga dos dados a cada ano. A partir do histórico do dicionário de radicais, deve-se verificar a migração dos termos entre as áreas de conhecimento ao longo dos anos. Caso os termos migrem de uma área para duas ou mais áreas, identifica-se a especialização áreas, caso os termos migrem de duas ou mais áreas para uma área, identifica-se a fusão de áreas. Essa idéia deve ser amadurecida e melhorada.
- Utilização da classificação QUALIS (QUALIS, 2009) para ponderar a qualidade das publicações científicas: sugere-se que a classificação da qualidade dos meios de publicação seja utilizada como um peso adicional no cálculo da relevância das competências dos pesquisadores. Dessa forma, os pesquisadores que publicassem em meios bem avaliados teriam relevância em dada competência maior do que pesquisadores que publicassem em meios pior avaliados.
- Avaliação da abordagem em outra área de conhecimento: essa dissertação avaliou os resultados obtidos pelo KAIROS no domínio de computação, porém é importante que outros domínios sejam avaliados para se ter certeza que a abordagem é válida para qualquer domínio.

6.4 Conclusão

Mais do que uma ferramenta computacional, esse trabalho propôs uma abordagem capaz de ser aplicada na vida real, e que possivelmente trará benefícios práticos para as organizações a utilizarem.

A proposta apresentada, objetivando à contínua melhoria, estará sempre passível de alterações. Acredita-se que seu uso em um cenário real, demandará modificações, adequações e desenvolvimento de novas funcionalidades.

Gostaria de agradecer a todos os leitores que chegaram até o fim. Todas as críticas, idéias e contribuições serão extremamente úteis para sucesso deste e dos próximos trabalhos.

Referências Bibliográficas

ABREU, R.M., 2008, Identificação e Análise Temporal de Competências Através de Informações Científicas Disponibilizadas na Web, Projeto Final do curso de Bacharelado em Ciência da Computação, Instituto de Matemática, UFRJ, Rio de Janeiro, Brasil

BARANAUSKAS, T., 1998. "Insights into brazilian competitive intelligence practices". In: Competitive Intelligence Magazine, v. 1, n. 1, p. 41-43.

BARBOSA, C.E., MAIA, L., 2006, MISIR: Módulo Integrado de Recomendação, Projeto Final de Curso, Instituto de Matemática, UFRJ, Rio de Janeiro, Brasil.

BARROS, M.O., WERNER, C.M.L., TRAVASSOS, G.H., 2002, "Um Estudo Experimental sobre a Utilização de Modelagem e Simulação no Apoio à Gerência de Projetos de Software". In: XIX Simpósio Brasileiro de Engenharia de Software (SBES), Uberlândia, MG, Brasil, Outubro de 2002.

BONFIM, E. L., OLIVEIRA, J., SOUZA, J. M., et al, 2005, "Thoth: Improving Experiences Reuses in the Scientific Environment through Workflow Management System", 9th International Conference on Computer Supported Cooperative Work in Design (CSCWD 2005) Coventry, UK.

BOOTH, Andrew D.A.A., 1967. "“Law” of occurrences for words of low frequency". Information and Control, [s.l.], v. 10, n.4, p.386-393, April 1967.

BREESE, J. S., HECKERMAN, D., KADIE, C., 1998, "Empirical Analysis of Predictive Algorithms for Collaborative Filtering". In: Proceedings of Fourteenth Conference on Uncertainty in Artificial Intelligence, pp. 4352, Madison, WI, USA.

CAPES, 2009. Disponível em: <http://www.capes.gov.br/avaliacao/avaliacao-da-pos-graduacao/>. Acessado em: 12/02/2009.

CIS, 2009. Disponível em: <http://www.statistics.gov.uk/STATBASE>. Acessado em: 12/02/2009.

COATES, J., 1985 “Foresight in Federal Government Policy Making”. In: *Futures Research Quarterly*, v. 1, p.29-53.

COELHO, G.M., 2003. “Prospecção tecnológica: metodologias e experiências nacionais e internacionais”. Rio de Janeiro: INT/Finep/ANP Projeto CT-Petro Tendências Tecnológicas.

COMPANO, R., HULLMANN, A., 2002. “Forecasting the development of nanotechnology with the help of science and technology indicators”. *Nanotechnology* 13, p.243-247.

CORDIS, 2009. Disponível em: http://cordis.europa.eu/fp6/instr_noe.htm. Acessado em: 12/02/2009.

DBLP, 2009. Disponível em: <http://www.informatik.uni-trier.de/~ley/db/>. Acessado em: 12/02/2009.

DII, 2009. Disponível em: <http://isiknowledge.com/diidw>. Acessado em: 12/02/2009.

DE MARTINO, R., 2006, *Análise Competitiva em Ambientes de Ensino ou Pesquisa Utilizando o Currículo Lattes*, Projeto Final do curso de Bacharelado em Ciência da Computação, Instituto de Matemática, UFRJ, Rio de Janeiro, Brasil

EILAND, E.H., GATLIN, D., 2008. "Forecast of Antibiotic Development in an Era of Increasing Bacterial Resistance". *Journal of Pharmacy Practice* 2008; 21; 313 originally published online Aug 28, 2008.

ESTES, G.M. e KUESPERT, D., 1976. “Delphi in industrial forecasting”. In: *Chemical and Engineering News*, EUA, p. 40-47.

FARIA, L.I.L., 2001, "Prospecção Tecnológica em Materiais: Aumento da Eficiência do Tratamento Bibliométrico. Aplicação na Análise de Tratamentos de Superfície Resistentes ao Desgaste", Tese de D.Sc., Universidade Federal de São Carlos, São Carlos, Brasil.

FELLBAUM, C., 1998. "WordNet - An Electronic Lexical Database" Ed. Mit Press.

GODET, M, 2000. "A caixa de ferramentas" da prospectiva estratégica. Caderno n. 5. Lisboa: Centro de Estudos de Prospectiva e Estratégia.

GUILFORD, J.P., 1959. "Traits of creativity" Ed. H.H. Anderson.

HOFFMANN, F. T. e MAYR, P., 1990. "Nitriding and nitrocarburizing". ASM Handbook: Friction and wear of compressors, ASM International, v.18, p.878-883.

HORTON, A., 1999 "Foresight: how to do simply and successfully Foresight", v. 01, n. 01.

JFREE, 2009. Disponível em: <http://www.jfree.org/jfreechart/>. Acessado em: 12/02/2009.

KAWAMURA, V.; OLIVEIRA, J.; SOUZA, J.; DE MARTINO, R.; 2005. "Competitive Intelligence in Brazilian Educational and Research Institutions". In: *The First International Conference/Workshop on Business, Technology and Competitive Intelligence The Fifth on Science to Business Marketing*. pp. 98-109. Nihon University, Tokyo, Japan. October.

KAWAMURA, V., 2006, Inteligência Competitiva para Instituições de Ensino e Pesquisa, Dissertação de M.Sc., PESC, COPPE, UFRJ, Rio de Janeiro, Brasil.

KLINK, S., LEY, S., RABBIDGE, E., REUTHER, P., WALTER, B, WEBER, A., 2004. "Browsing and Visualizing Digital Bibliographic". In: *Data. Joint EUROGRAPHICS - IEEE TCVG Symposium on Visualization*. p. 237-242.

KREJCI, D., 2009, Inteligência Competitiva para Instituições de Ensino e Pesquisa, Dissertação de M.Sc., COPPE, UFRJ, Rio de Janeiro, Brasil..

LATTES, 2009. Disponível em: <http://lattes.cnpq.br/>. Acessado em: 12/02/2009.

LEY,2002. "The DBLP Computer Science Bibliography: Evolution, Research Issues, Perspectives". In: *String Processing and Information Retrieval - Lecture Notes in Computer Science*. V. 2476, p. 481-486.

LIPSOR - "Laboratory for Investigation in Prospective Strategy and Organization", 2009. Disponível em: <http://www.cnam.fr/lipsor/eng/software02.php>. Obtido em: 05/03/2009.

LUHN, G.; WETH, R.V.D., 1999. "Abstraction and experience: Engineering design in new contexts of cognitive and philosophical science". In: ICED, Munich, vol 2, p. 947-952.

MACQUEEN, J.B., 1967. "Some Methods for classification and Analysis of Multivariate Observations", Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability", Berkeley, University of California Press, 1:281-297

MARTIN, B.R.; ANDERSON, J.; MACLEAN, M. "Identifying Research Priorities in Public-Sector Funding Agencies: Mapping Science Outputs onto User Needs". In: Technology Analysis and Strategic Management, v. 10, 1998.

MARTINS, C. J. M., 2008, "Aplicação de ferramentas computacionais para prospecção tecnológica por mineração de dados não-estruturados sobre patentes industriais em idiomas inglês.", Dissertação de M.Sc., PEC, COPPE, UFRJ, Rio de Janeiro, Brasil.

MCGONAGLE, J. J.; VELLA, C. M., 1998. Protecting your company against competitive intelligence. Quorum Books.

MCT, 2009a. "Informação e Inovação: A Propriedade Industrial no Contexto da Informação e da Inovação". Disponível em: **Erro! A referência de hiperlink não é válida.** Acessado em: 12/02/2009.

MCT, 2009b. "Estudo Prospectar - Relatório Final". Disponível em: http://ftp.mct.gov.br/cct/Prospectar/Relatorio_3/cap1_1.htm. Acessado em: 12/02/2009.

MENDES, C.S., 2008. "Prospecção Tecnológica". Curso de Capacitação em PI para Gestores de Tecnologia - Módulo Avançado. Março/2008. Disponível em: http://www.tecpar.br/appi/Curso_Avançado_NITs/Cristina%20dUrso%20de%20Souza%20Mendes.pdf. Acessado em: 12/12/2008.

MONCLAR, R.S. ; OLIVEIRA, J. ; SOUZA, J. M., 2007. A New Approach to Balance Social Networks. In: UK Social Network Conference, 2007, Londres. Proceedings of UK Social Network Conference, 2007. p. 141-142.

MONCLAR, R.S. ; OLIVEIRA, J. ; SOUZA, J. M., 2008. The Future of Social Network Balancing. In: 12th International Symposium on the Management of Industrial and Corporate Knowledge, 2008, Niteroi. Proceedings of 12th International Symposium on the Management of Industrial and Corporate Knowledge.

MONCLAR, R.S. ; OLIVEIRA, J. ; SOUZA, J. M. , 2009. Analysis and Balancing of Social Network to Improve the Knowledge Flow on Multidisciplinary Teams (aceito, a ser publicado). In: Proceedings of 13th International Conference on Computer Supported Cooperative Work in Design (CSCWD 2009), 2009, Santiago. Proceedings of Proceedings of 13th International Conference on Computer Supported Cooperative Work in Design (CSCWD 2009).

OLIVEIRA, J.; SOUZA, J. M.; MIRANDA, R.; RODRIGUES, S., 2005a. "GCC: An Environment for Knowledge Management in Scientific Research and Higher Education Centers". In: I-Know - 5th International Conference on Knowledge Management, Special Track on "Knowledge Sharing in Research and Higher Education", Graz, Austria, Julho.

OLIVEIRA, L. G.; SUSTER, R.; PINTO, A. C.; RIBEIRO, N.; SILVA, R. B., 2005b. "Informação de Patentes: Ferramenta Indispensável para a Pesquisa e o Desenvolvimento Tecnológico". In: *Quim. Nova*, Vol. 28, Suplemento, S36-S40.

OLIVEIRA, J., 2007, *METHEXIS: Uma Abordagem de Apoio à Gestão do Conhecimento para Ambientes de "E-Science"*, Tese de D.Sc., COPPE, UFRJ, Rio de Janeiro, Brasil.

OPENFORECAST, 2009. Disponível em: <http://openforecast.sourceforge.net/>. Acessado em: 12/02/2009.

ORENGO, V.M., HUYCK, C.R., 2001. "A Stemming Algorithm for the Portuguese Language". In: *Proceedings of the SPIRE Conference*. Laguna de San Raphael, pg 13-15, 2001.

OSBORN, A. F. "Applied Imagination". Ed: Scribners, New York, 1953.

PESC, 2009. Disponível em: <http://www.cos.ufrj.br>. Acessado em: 12/02/2009.

PITAC - PRESIDENT'S INFORMATION TECHNOLOGY ADVISORY COMMITTEE, 2005. "Computational Science: Ensuring America's Competitiveness" In: *PITAC Report to the President, EUA*. Disponível em http://www.nitrd.gov/pitac/reports/20050609_computational/computational.pdf. Acessado em: 02/03/2009.

POLYANALYST, 2009. Disponível em: <http://www.megaputer.com/>. Acessado em: 12/02/2009.

POPP, D. C., 2001. "The effect of new technology on energy consumption," *Resource and Energy Economics*, Elsevier, vol. 23(3), pages 215-239, July.

PORTER, M.F., 1980. "An algorithm for suffix stripping", In: "Program", 14 no. 3, pp 130-137.

PORTER, A. L., DETAMPEL, M. J., 1995. Technology opportunities analysis. *Technological Forecasting and Social Change*, v. 49, p. 237-255.

PORTER, A.L.; ROPER, A.T.; MASON, T.W.; ROSSINI, F.A.; BANKS, J., 1991. *Forecasting and Management of Technology*. New York, Ed. John Wiley.

PORTER, A. L., 1991 “Tech forecasting”, In: *Technological Forecasting and Social Change*, v. 62, p. 19-28, 1999.

PORTER, A. L. *et al.*, 2004 “Technology futures analysis: towards integration of the field and new methods”. In: *Technological Forecasting and Social Change*, n.49, 2004.

QUALIS, 2009. Disponível em: <http://www.capes.gov.br/avaliacao/qualis/>. Acessado em: 12/02/2009.

RAPIDMINER, 2009. Disponível em: <http://rapid-i.com/>. Acessado em: 12/02/2009.

RODRIGUES, S; SAMPAIO, J. O ; SOUZA, J. M., 2004. “Competence mining for virtual scientific community creation”. *International Journal of Web Based Communities*, Grã-Bretanha, v. 1, n. 1, pp. 90-102.

RODRIGUES, S; SAMPAIO, J. O ; SOUZA, J. M., 2006. “Recommendation for Team and Virtual Community Formations Based on Competence Mining”. *Computer Supported Cooperative Work in Design II - Lecture Notes in Computer Science*, Berlin, v. 3865, n. 2, pp. 365-374.

RODRIGUES, S., 2007, Extração de Conhecimento em Processos de Certificação Profissional, Dissertação de M.Sc., PESC, COPPE, UFRJ, Rio de Janeiro, Brasil.

RUGGIERO, M. A., LOPES, V.L.R., 1996. *Cálculo Numérico - Aspectos Teóricos e Computacionais*. 2 ed., Editora Pearson Education.

SALLES FILHO, S. L. M., BONACELLI, M. B. M, MELLO, D. L., 2001. "Instrumentos de apoio à definição de políticas em biotecnologia". Brasília: MCT; Rio de Janeiro: FINEP.

SBC - Sociedade Brasileira da Computação. "Grandes Desafios da Pesquisa em Computação no Brasil – 2006 – 2016". Relatório sobre o Seminário realizado em 8 e 9 de maio de 2006. Disponível em http://www.ic.unicamp.br/~cmbm/desafios_SBC/RelatorioFinal.pdf. Acessado em: 02/03/2009.

SILVA, J. L. C. S, 2009. Disponível em: <http://www.larces.uece.br/~jlcs/>. Obtido em: 03/03/2009.

SOLINGEN, R.V., BERGHOUT, E., 1999, "The Goal / Question / Metric Method: A Practical Guide for Quality Improvement of Software Development", McGraw Hill.

STROELE, V. ; SILVA, R. ; SOUZA, M. ; OLIVEIRA, J. ; MELLO, C. E. ; SOUZA, J. M. ; ZIMBRAO, G., 2008a. Mining and Analyzing Organizational Social Networks Using Minimum Spanning Tree. In: 16th International Conference on Cooperative Information Systems (CoopIS 2008), 2008, Monterrey. Proceedings of 16th International Conference on Cooperative Information Systems, 2008.

STROELE, V. ; SILVA, R. ; SOUZA, M. ; OLIVEIRA, J. ; SOUZA, J. M. ; ZIMBRAO, G., 2008b . Mining and Analyzing Organizational Social Networks Using Minimum Spanning Tree. In: 12th International Symposium on the Management of Industrial and Corporate Knowledge, 2008, Niteroi. Proceedings of 12th International Symposium on the Management of Industrial and Corporate Knowledge, 2008.

STROELE, V. ; SILVA, R. ; OLIVEIRA, J. ; SOUZA, J. M., 2009. ZIMBRAO, G. . Mining and Analyzing Organizational Social Networks for Collaborative Design (aceito, a ser publicado). In: 13th International Conference on Computer Supported Cooperative Work in Design (CSCWD 2009), 2009, Santiago. Proceedings of 13th International Conference on Computer Supported Cooperative Work in Design (CSCWD 2009), 2009.

TAYLOR, F. W., 1970. "Princípios de administração científica". 7. ed. Tradução de Arlindo Vieira Ramos, São Paulo: Atlas.

UPSTO, 2009. Disponível em: <http://www.uspto.gov/>. Acessado em: 12/02/2009.

UTTERBACK, J. M., 1994. "Mastering the dynamics of innovation". Ed: Harvard Business School. Boston, EUA.

VAN RAAN, A. F. J., 1997. "Scientometrics: state-of-the-art". In: *Scientometrics*, v. 38, n. 1, p. 205-218.

VANTAGEPOINT, 2009. Disponível em: <http://www.thevantagepoint.com/>. Acessado em: 12/02/2009.

VELARDI, P., NAVIGLI, R., PETIT, M., 2007. "Semantic Indexing of a Competence Map to support Scientific Collaboration in a Research Community". In: *Proceedings of International Joint Conference on Artificial Intelligence*. p.2897-2902.

WEINER, P., 1973, "Linear pattern matching algorithm". 14th Annual IEEE Symposium on Switching and Automata Theory, pp 1-11, New York, USA.

WIVES, L. K., 1997. *Indexação de Documentos Textuais*. Porto Alegre: Universidade Federal do Rio Grande do Sul, Instituto de Informática, 1997.

WORMELL, I., 1998. "Informetria: explorando bases de dados como instrumentos de análise". *Ciência da Informação*, Brasília, DF, v. 27, n. 2, p. 210-216.

WRIGHT, J. T. C.; GIOVINAZZO, R. A. "Delphi uma ferramenta de apoio ao planejamento prospectivo", 2000. Disponível em: <http://www.iea.usp.br/iea/tematicas/futuro/projeto/delphi.pdf>. Obtido em: 03/03/2009.

YOON, B., LEE, S., 2008, "Patent analysis for technology forecasting: Sector-specific applications". *Engineering Management Conference, IEMC Europe 2008*. IEEE International. p. 1-5, 28-30 June 2008.

ZANASI, A., 2000. "Data mining and competitive intelligence through Internet". Disponível em: <http://www.cilea.it/collabora/GARR-NIR/nir-it-5/atti/zanasi/zanasi.htm>. Acessado em 10/12/2007.

ZHU, D., PORTER, A.L., 2002. "Automated extraction and visualization of information for technological intelligence and forecasting". In: *Technological Forecasting & Social Change*. V.69 (2002) 495–506

Apêndice A – Modelo de Dados da Base KAIROS

Foram construídos quatro modelos de dados para a base KAIROS: base LATTES, que guarda informações do Curriculum Lattes, base DBLP, que guarda informações do DBLP, base de Patentes, que guarda informações de patentes da Derwent e base de Teses, que guarda informações de teses e dissertações da Capes.

A Figura 45 apresenta o modelo de dados da base LATTES.

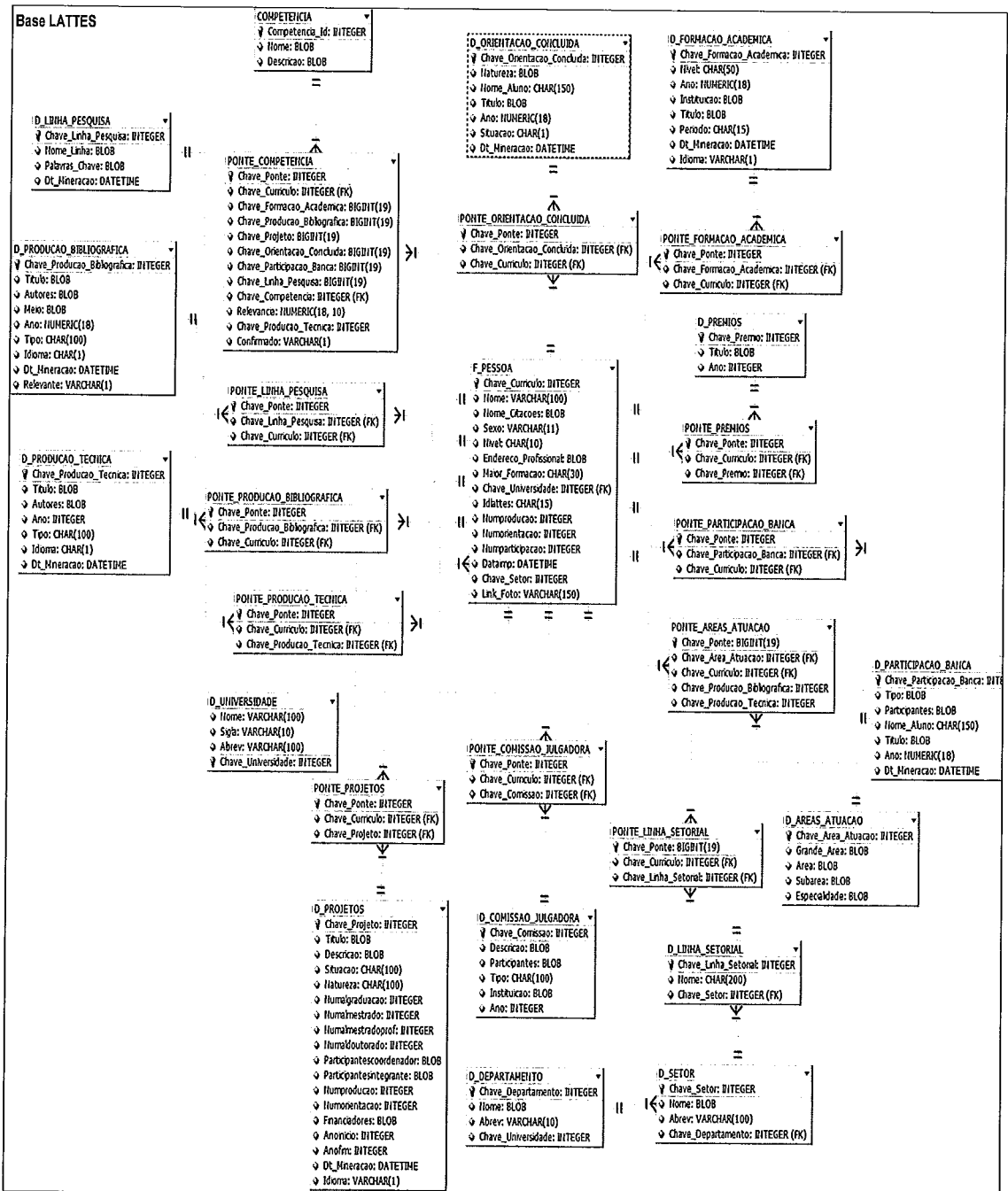


Figura 45: Modelo de dados da base LATTES.

A Figura 46 apresenta o modelo de dados da base DBLP.

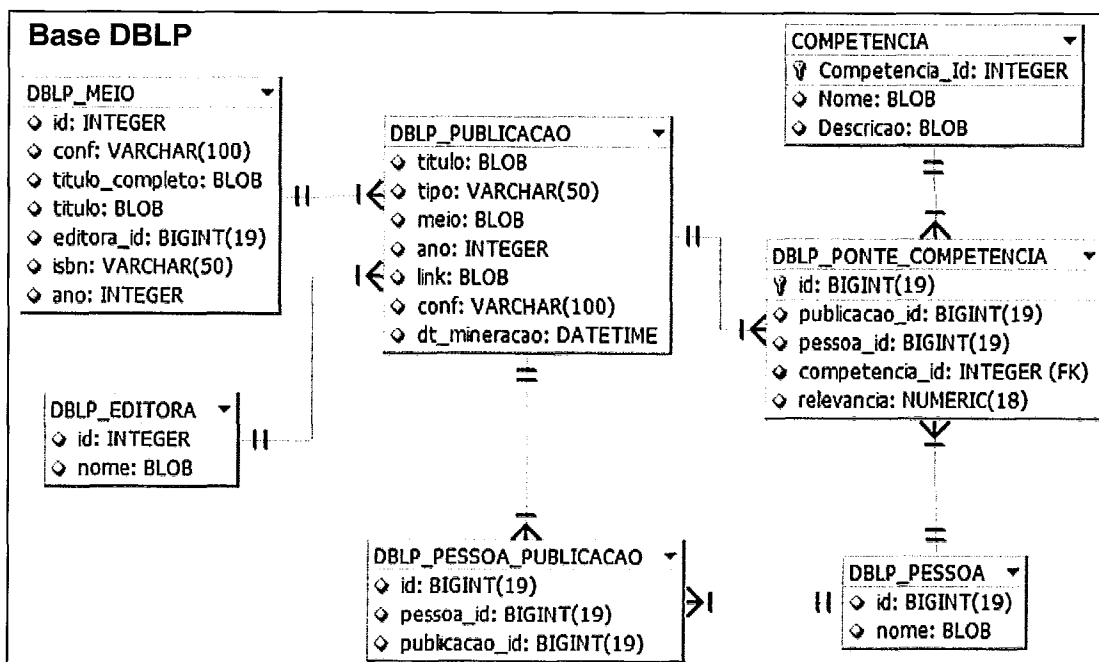


Figura 46: Modelo de dados da base DBLP.

A Figura 47 apresenta o modelo de dados da base de Patentes.

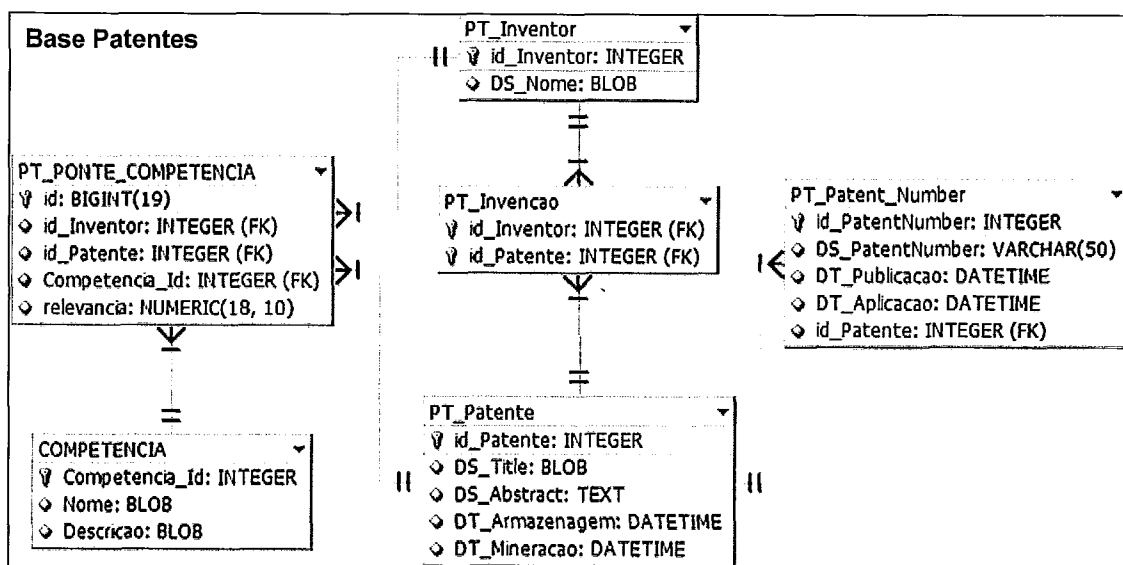


Figura 47: Modelo de dados da base de Patentes.

Apêndice B – StopWords Utilizadas na Mineração

Na etapa de pré-processamento textual, são retiradas do texto que está sendo minerado as palavras que não possuem significado relevante, chamadas de *StopWords*. A Tabela 17 apresenta as *StopWords* do idioma português, enquanto que a Tabela 18 apresenta as *StopWords* do idioma inglês.

Tabela 17: Lista de *StopWords* no idioma português.

a	calmamente	dumas	milesimos	ola	quarenta	tais
à	cedo	duns	milhao	onde	quarta	tal
absolutamente	cem	duzentos	milionesimo	ontem	quarto	tanta
agora	centesimo	e	milionesimos	onze	quatorze	tantas
ah	certa	eia	mim	opa	quatro	tanto
ai	certas	ela	muita	os	quatrocentos	tantos
ainda	certo	elas	muitas	ou	que	tarde
algo	certos	ele	muito	outra	quem	te
alguém	chi	eles	muitos	outras	quingentesimo	terceira
algum	cinco	em	na	outrem	quinhentos	terceiro
alguma	cinquenta	entao	nada	outro	quinguagesimo	teu
algumas	com	então	nao	outrora	quinta	teus
alguns	comigo	essa	não	outros	quinto	ti
alo	como	essas	nas	oxala	quinze	toda
alto	conosco	esse	nenhum	para	se	todas
amanha	consigo	esses	nenhuma	pára	segunda	todo
amanhã	contigo	esta	nenhumas	pela	segundo	todos
ante	contra	estas	nenhuns	pelas	seis	tras
anteontem	convosco	este	ninguém	pelo	seiscentesimo	trás
antes	coragem	estes	no	pelos	seiscentos	trecentos imo
ao	corretamente	eu	nogentesimo	perante	sem	tres
aonde	cuja	fielmente	nonagesimo	pior	sempre	treze
aos	cujo	hem	nono	por	senhor	trezentos
apos	da	hoje	nos	porem	senhora	trigesimo
após	das	hum	nós	porém	senhoria	trinta
aquela	de	ih	nosso	pouca	senhorita	tu
aquelas	debalde	isso	nossos	poucas	septingentesimo	tudo

aquele	decima	isto	nove	pouco	septuagesimo	ue
aqueles	decimo	ja	novecientos	poucos	sessenta	uh
aquilo	depois	já	noventa	primeira	sete	ui
as	depressa	jamais	num	primeira mente	setecentos	ultimame nte
às	desde	levemente	numa	primeiro	setenta	um
assim	devagar	lhe	numas	psit	setima	uma
ate	dez	lhes	nunca	psiu	setimo	umas
até	dezenove	logo	nuns	puxa	seu	uns
avante	dezesseis	mais	o	quadrages imo	seus	vamos
basta	dezessete	mal	oba	quadrigen tesimo	sexagesimo	varia
bem	dezoito	mas	octingentesimo	quais	sexcentesimo	varias
bilhao	do	me	octogesimo	quaisquer	sexta	vario
bilionesimo	dois	melhor	oh	qual	sexto	varios
bilionesimos	dos	menos	oitava	qualquer	si	vigesimo
bis	doze	meu	oitavo	quanta	silencio	vinte
bravo	ducentesimo	meus	oitenta	quantas	sob	viva
breve	dum	mil	oito	quanto	sobre	voce
cada	duma	milesimo	oitocentos	quantos	sua	você

Tabela 18: Lista de *StopWords* no idioma inglês.

about	beginnings	ever	in	not	she	throughout
above	begins	far	into	now	should	thus
accordingly	begone	ff	is	of	since	to
after	begun	following	it	off	so	too
again	being	for	its	often	some	toward
against	below	fro	itself	oh	such	unless
ah	between	from	last	on	than	until
all	but	further	lastly	only	that	up
also	by	get	less	or	the	upon
although	ca	go	ll	other	their	us
always	can	goes	many	otherwise	them	ve
am	cannot	going	may	our	themselves	very
an	come	got	me	ourselves	then	was
and	could	had	might	out	there	we
any	did	has	more	over	therefore	were

anymore	do	have	must	perhaps	these	what
anyone	doing	he	my	put	they	whatever
are	during	her	myself	puts	this	when
as	each	hers	nay	quite	those	where
at	either	herself	near	said	thou	which
away	else	him	nearly	saw	though	while
be	end	himself	never	say	throughout	who
been	et	his	new	see	thus	whom
begin	etc	how	next	seen	to	whomever
beginning	even	if	no	shall	too	whose

Apêndice C – Avaliação da Proposta

Esse Apêndice apresenta informações complementares à Avaliação da proposta dos Estudos 1 e 2.

C.1 – Estudo 1: Áreas Promissoras e Pesquisadores Visionários

Conforme apresentado no Capítulo 5, o Estudo 1 contou com um formulário eletrônico, que foi disponibilizado na internet para preenchimento dos participantes. A Figura 49 apresenta a tela inicial do formulário.

PESC Avaliação da Prospeção de Áreas Científicas e da Heurística para Identificação de Visionários

Prezado pesquisador,

Este questionário tem como objetivo avaliar resultados obtidos em pesquisas realizadas na linha de Banco de Dados do Programa de Engenharia de Sistemas e Computação da COPPE/UFRJ na área de Gestão do Conhecimento Científico.

Este trabalho pretende identificar o comportamento das áreas científicas ao longo do tempo e prospectar o comportamento futuro a partir de dados históricos da produção científica. Além disso, prevê uma heurística para identificação de pesquisadores visionários nas diferentes áreas da ciência.

Os questionários estão divididos em 2 etapas: a primeira etapa avalia a prospeção das áreas científicas, enquanto que a segunda etapa avalia uma heurística para identificação de pesquisadores visionários.

As questões optativas de ambas as etapas seguem o seguinte padrão de resposta:

Padrão de Resposta	
N/A	Não apto a responder (Desconheço)
1	Discordo completamente
2	Discordo parcialmente
3	Nem concordo, nem discordo
4	Concordo parcialmente
5	Concordo completamente

Quaisquer dúvidas podem ser enviadas diretamente para o aluno responsável: Rafael De Martino (rmartino@gmail.com ou rmartino@cos.ufrj.br)
Orientadores do trabalho: Prof. Jano Moreira de Souza (PESC.COPPE@ufrj.br) e Profª. Jonice de Oliveira Sampaio (DCC@ufrj.br).

Agradecemos a colaboração.

Figura 49: Tela de apresentação do formulário eletrônico do Estudo 1.

Após a leitura das instruções iniciais, o participante era direcionado para um questionário de perguntas relativas à sua atuação profissional, conforme pode ser verificado na Figura 50.

Etapa 1: Dados do Participante

1) Vínculo com a Pesquisa:

2) Titulação:

- Professor universitário
- Aluno de graduação ou pós-graduação
- Pesquisador associado à universidade ou centro de pesquisa
- Pesquisador associado à indústria
- Pesquisador associado à empresa de serviços

3) Tempo de experiência com:

4) Principal área de atuação:

Figura 50: Tela de preenchimento dos dados do participante.

Uma vez respondido o questionário, o participante era direcionado para a pesquisa propriamente dita. A Figura 51 apresenta o questionário de avaliação da prospecção de áreas. O questionário também continha uma pergunta aberta, na qual os participantes respondiam quais áreas consideravam promissoras que não haviam sido listadas, conforme pode ser visto na Figura 52.

Etapa 2: Avaliação da prospecção de áreas científicas

1) De acordo com sua percepção, avalie se as áreas científicas listadas abaixo podem ser consideradas promissoras nos próximos 3 anos, do ponto de vista do aumento (teses). Pontue de acordo com o padrão de resposta definido no início da pesquisa (1 para discordo completamente e 5 para concordo completamente). Caso não se si

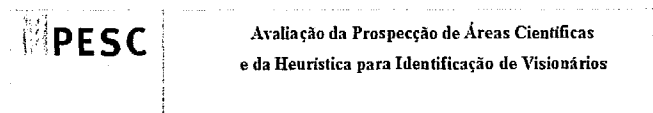
Área	N/A	1	2	3	4	5
Agentes Inteligentes	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Algoritmos Distribuídos	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Análise Numérica	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Aprendizagem de Máquina	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Bioinformática	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Computação de Alto Desempenho	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Computação Em Grids Ou Grades	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Computação Gráfica	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
CSCW (Trabalho Cooperativo Suportado por Computador)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Implementação de Algoritmos Criptográficos	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Jogos 3D	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Linguagens de Programação	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Lógica aplicada à computação	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Métodos Ágeis de Desenvolvimento de Software	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Mineração de Dados	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Otimização	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Processamento de Imagens	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Figura 51: Tela de avaliação da prospecção de áreas.

2) Quais áreas, que não apareceram na lista anterior, você considera promissoras nos próximos 3 anos? (preenchimento opcional)

Figura 52: Pergunta aberta sobre as áreas promissoras não listadas no questionário.

Após o preenchimento da primeira etapa de perguntas, o participante era encaminhado para a segunda etapa, na qual deveria avaliar a heurística de identificação de pesquisadores visionários, conforme apresentado na Figura 53. Os nomes dos pesquisadores foram omitidos por questões de privacidade. Como na primeira etapa, os usuários poderiam indicar pesquisadores que consideravam visionários e não foram listados pela heurística.



Etapa 3: Avaliação da heurística de identificação de pesquisadores visionários

1) De acordo com sua percepção, avalie se os pesquisadores abaixo listados, que atuam ou atuaram no Brasil, podem ser considerados visionários na área de computação (1 para discordo completamente e 5 para concordo completamente). Caso não se sinta apto a responder, selecione a opção N/A (não se aplica).

Pesquisador	N/A	1	2	3	4	5
Pesquisador 1 (Nome Omitido)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Pesquisador 2 (Nome Omitido)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Pesquisador 3 (Nome Omitido)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Pesquisador 4 (Nome Omitido)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Pesquisador 5 (Nome Omitido)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Pesquisador 6 (Nome Omitido)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Pesquisador 7 (Nome Omitido)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Pesquisador 8 (Nome Omitido)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Pesquisador 9 (Nome Omitido)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

2) Quais pesquisadores com atuação no Brasil, que não apareceram na lista anterior, você considera visionários na área de computação? (preenchimento opcional)

Figura 53: Tela de avaliação de pesquisadores visionários (nomes omitidos).

No Capítulo 5, foram apresentados os resultados consolidados das respostas dos participantes do Estudo 1.

Para o leitor ter uma visão aprofundada das respostas, a Tabela 19 apresenta os resultados detalhados da primeira etapa do estudo, enquanto que a Tabela 20 apresenta os resultados detalhados da segunda etapa.

Tabela 19: Resultados detalhados da primeira etapa do Estudo 1: avaliação de área promissoras.

Área	N/A	Nota 1	Nota 2	Nota 3	Nota 4	Nota 5
Agentes Inteligentes	13	0	8	9	13	14
Algoritmos Distribuídos	8	0	2	11	26	10
Análise Numérica	16	6	8	14	7	4
Aprendizagem de Máquina	6	1	3	11	23	13
Bioinformática	1	0	0	7	12	37
Computação de Alto Desempenho	10	0	0	15	15	17
Computação Em Grids Ou Grades	10	0	2	10	21	14
Computação Gráfica	11	0	2	11	21	11
CSCW (Trabalho Cooperativo Suportado por Computador)	13	1	4	13	18	8
Implementação de Algoritmos Criptográficos	12	0	8	13	12	12
Jogos 3D	10	1	3	10	20	13
Linguagens de Programação	11	6	18	16	6	0
Lógica aplicada à computação	11	9	17	10	8	2
Métodos Ágeis de Desenvolvimento de Software	10	1	7	17	13	9
Mineração de Dados	5	0	1	10	21	20
Otimização	10	0	7	17	15	8
Processamento de Imagens	5	0	3	9	25	15
Processamento de Sinais Biológicos	12	1	1	8	12	23
Processamento e Análise de	10	1	2	8	15	21

Imagens Médicas						
Processamento Paralelo e Distribuído	9	1	1	14	22	10
Programação Distribuída	9	0	4	17	19	8
Programação Não-Linear	23	1	10	15	7	1
Redes de Computadores	13	1	6	15	16	6
Redes sem fio	10	0	5	5	16	21
Robótica	11	0	2	9	13	22
Sistemas Autônomos Inteligentes	11	0	2	11	17	15
Sistemas de Informação	8	5	10	15	13	6
Software Básico	16	9	19	8	5	0
Teoria dos Grafos	19	6	16	7	9	0
Visão Computacional	15	0	3	8	17	14

Tabela 20: Resultados detalhados da segunda etapa do Estudo 1: avaliação de visionários.

Pesquisador	N/A	Nota 1	Nota 2	Nota 3	Nota 4	Nota 5
Pesquisador 1	28	2	0	3	4	12
Pesquisador 2	23	2	1	10	9	4
Pesquisador 3	28	2	4	6	7	2
Pesquisador 4	13	0	2	2	13	19
Pesquisador 5	20	1	2	7	14	5
Pesquisador 6	20	0	2	5	12	10
Pesquisador 7	25	1	2	8	7	6
Pesquisador 8	28	1	4	1	13	2
Pesquisador 9	27	1	5	8	6	2

C.2 – Estudo 2: Identificação de Competências

No Capítulo 5, foram apresentados os resultados consolidados das respostas dos participantes do Estudo 2.

Para o leitor ter uma visão aprofundada das respostas, a Tabela 21 apresenta os resultados detalhados. A primeira coluna apresenta a identificação do pesquisador, não identificado por questões de privacidade, a segunda coluna apresenta a área de atuação, cadastrada no PESC, e a terceira coluna apresenta o indicativo se a área foi ou não mapeada pelo KAIROS. Em negrito, são destacadas as áreas principais de atuação dos pesquisadores.

Tabela 21: Resultados detalhados do Estudo 2.

Pesquisador	Área de Atuação (obtida na página do PESC)	Mapeado?
Pesquisador 1	Métodos de Paralelização	Não
	Métodos Lagrangeanos	Não
	Modelagem Matemática	Não
	Suavização de Problemas não-diferenciáveis	Sim
	Otimização	Sim
Pesquisador 2	Controle de Qualidade de Software	Não
	Gerência de Projeto de Software	Não
	Software Direcionado à Medicina (Cardiologia)	Não
	Engenharia de Software	Sim
Pesquisador 3	Geometria Computacional	Sim
	Visão Computacional	Sim
	Processamento de Imagens	Sim
	Computação Gráfica	Sim
Pesquisador 4	Teoria dos Grafos: Algoritmos em Grafos e Problemas de Isomorfismo	Sim
	Teoria da Computação: Classes de Complexidade e Análise de Algoritmos	Sim
	Matemática Discreta relacionada com Ciência da Computação	Sim
	Otimização Combinatória: Coloração de Vértices e de Arestas em Grafos	Sim

	Algoritmos e Combinatória	Sim
Pesquisador 5	Reutilização de Software: Engenharia de Domínio, Desenvolvimento Baseado em Componentes, Padrões de Software e Aspectos Gerenciais	Não
	Ambientes de Desenvolvimento de Software: Ambientes Orientados a Domínio e Ferramentas de Suporte à Reutilização	Não
	Engenharia de Software	Sim
Pesquisador 6	Estruturas e bancos de dados espaciais	Sim
	Modelagem geométrica e de sólidos	Sim
	Animação baseada em física	Sim
	Visualização de terrenos	Não
	Linguagens de programação	Sim
	Banco de Dados	Sim
	Computação Gráfica	Sim
Pesquisador 7	Computação Paralela e Distribuída	Sim
	Computação Móvel em Redes de Comunicação sem Fio	Sim
	Arquitetura e Sistemas Operacionais	Sim
Pesquisador 8	Redes de Computadores	Sim
Pesquisador 9	Modelagem de Sistemas de Computação/Comunicação	Não
	Análise de Desempenho e Confiabilidade de Sistemas	Sim
	Redes de Computadores	Sim
	Redes de Alta Velocidade	Não
	Sistemas Distribuídos	Sim
	Arquitetura de Computadores	Sim
	Sistemas Operacionais	Sim
Pesquisador 10	Controle Ótimo Estocástico	Não
	Otimização	Sim
Pesquisador 11	Algoritmos	Sim
	Combinatória	Sim

	Grafos	Sim
	Estrutura de Dados	Sim
	Otimização	Sim
Pesquisador 12	Inteligência Artificial	Sim
	Redes Neurais	Sim
Pesquisador 13	Modelagem e Avaliação de Desempenho	Sim
	Redes Integradas de Faixa Larga	Não
	Redes Sem Fio e Sistemas Móveis de Comunicação -- Sensores; WiFi, WiMax e Mesh	Sim
	Caracterização de Tráfego em Redes	Não
	Qualidade de Serviços em Redes	Não
	Protocolos de Redes de Alta Velocidade	Não
	Redes de Computadores	Sim
Pesquisador 14	Programação Linear	Sim
	Programação Não Linear	Sim
	Programação Semidefinida	Não
	Algoritmos de Pontos Interiores	Não
	Otimização	Sim
Pesquisador 15	Teoria dos Grafos: Algoritmos em Grafos e Grafos em Interseção	Sim
	Teoria da Computação: Classes de Complexidade e Análise de Algoritmos	Sim
	Matemática Discreta Relacionada com Ciência da Computação	Sim
	Otimização Combinatória	Sim
	Algoritmos e Combinatória	Sim
Pesquisador 16	Especificação de Sistemas Distribuídos e Concorrência	Sim
	Lógicas Modais: Lógicas Dinâmicas, Semânticas Dinâmicas, Lógicas Deônticas e Lógicas Modais para Processos	Sim
	Aspectos Algébricos de Lógicas Modais	Não
	Inteligência Artificial	Não

Pesquisador 17	Projeto de Bases de Dados Distribuídos	Não
	Processamento de Consultas	Não
	Paralelismo e Distribuição em Bancos de Dados	Não
	Mineração de Dados e SGBDs	Sim
	SGBDs e XML	Não
	Banco de dados	Sim
Pesquisador 18	Programação Matemática	Não
	Otimização Combinatória	Sim
	Matemática da Computação	Sim
	Computação Biológica	Não
	Geometria Computacional	Não
	Otimização	Sim
Pesquisador 19	Computação Paralela	Sim
	Algoritmos Distribuídos	Sim
	Circuitos Digitais Assíncronos	Não
	Redes Neurais Artificiais	Sim
	Arquitetura de Computadores e S.O.	Sim
	Modelos de Sistemas Complexos	Não
Pesquisador 20	Busca, Recuperação e Extração da Informação	Não
	Aplicações P2P (Coppeer)	Não
	Mineração de Dados e Texto	Não
	Lógica Fuzzy	Sim
	WWW	Sim
	Banco de Dados	Sim
Pesquisador 21	Estruturas de dados espaciais e temporais	Sim
	Regras de Negócio	Não
	Datamining	Sim
	SGDBs Orientados a Objetos	Não
	Banco de dados	Sim
Pesquisador 22	Aprendizado de Máquina (Machine Learning) / Mineração de Dados (Data Mining)	Sim
	Lógica	Sim
	Inteligência Artificial	Sim

Pesquisador 23	Engenharia de Software Experimental	Não
	Ambientes para Experimentação e Ciência em Larga Escala	Não
	Modelos de Integração de Ferramentas e Componentes	Não
	Qualidade de Software Orientado a Objetos: Inspeção e Testes	Não
	Técnicas de Revisão de Artefatos de Software: Requisitos, Projeto, Arquitetura	Não
	Engenharia de Aplicações Web	Sim
	Engenharia de Software	Sim
Pesquisador 24	Estudos de Ciência e Tecnologia numa abordagem sócio-técnica	Não
	Análise do processo de adoção de novas tecnologias nos níveis gerencial e operacional e seus impactos	Não
	Informática na Educação	Sim
	Informática e Sociedade	Não
Pesquisador 25	Gestão do Conhecimento	Sim
	Gestão do Conhecimento Científico e da Engenharia	Não
	Sistemas de Suporte à Negociação	Não
	Trabalho Cooperativo Apoiado por Computador	Sim
	Computação Autônoma	Não
	SGDBs para Aplicações não Convencionais	Não
	Sistemas de Informação Geográfica	Sim
	Banco de dados	Sim
Pesquisador 26	Especificações Formais	Sim
	Tipos e Estruturas de Dados	Sim
	Metodologia de Programação	Sim
	Teoria de Problemas e de Algoritmos	Sim
	Fundamentos da Computação	Sim
	Lógica e Fundamentos da Matemática	Sim
	Teoria Matemática de Sistemas	Não
	Inteligência Artificial	Sim

Pesquisador 27	Programação Linear e Não-Linear	Não
	Pontos Interiores	Não
	Aplicações da Geometria Riemanniana à Otimização Contínua	Não
	Programação Semidefinida	Não
	Otimização	Sim
	Programação Linear e Não-Linear	Sim
	Pontos Interiores	Não
Pesquisador 28	Modelagem Geométrica e de Sólidos	Não
	Visualização	Sim
	Aplicações de CAD	Não
	Animação	Não
	Computação Gráfica	Sim
Pesquisador 29	Visualização científica	Sim
	Imagens médicas	Não
	Computação gráfica	Sim
	Computação de alta performance	Sim
	Computação Paralela	Sim
	Programação GPU	Não
Pesquisador 30	Modelagem e análise de desempenho de sistemas de computação/comunicação	Sim
	Redes de alta velocidade	Não
	Modelagem de tráfego multimídia	Não
	Redes de computadores	Sim
Pesquisador 31	Teoria Quântica de Campo em Espaços Curvos	Não
	Gravitação e Cosmologia	Não
	Astrofísica Relativística	Não
	Biologia Matemática e Computacional	Não
	Otimização	Sim
Pesquisador 32	Álgebra não comutativa	Não
	Criptografia	Sim
	Computação Algébrica	Não
	Geometria Algébrica	Não

	Algoritmos e Combinatória	Sim
Pesquisador 33	Teoria dos Grafos: Algoritmos em Grafos e Problemas de Isomorfismo	Sim
	Teoria da Computação: Classes de Complexidade e Análise de Algoritmos	Sim
	Matemática Discreta relacionada com Ciência da Computação	Não
	Otimização Combinatória: Coloração de Vértices e de Arestas em Grafos	Não
	Algoritmos e Combinatória	Sim
Pesquisador 34	Otimização Não Diferenciável	Não
	Programação Matemática de Dois Níveis	Não
	Programação Matemática com Restrições de Equilíbrio	Não
	Desigualdade Variacional	Não
	Desigualdade Variacional Generalizada	Não
	Otimização	Sim
Pesquisador 35	Computação Paralela e Distribuída	Sim
	Modelos de Sistemas Complexos	Não
	Algoritmos e Combinatória	Sim
	Arquitetura e Sistemas Operacionais	Não
	Inteligência Artificial	Sim