

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
INSTITUTO DE INFORMÁTICA  
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO

LAURA SILVEIRA MASTELLA

**Um Modelo de Conhecimento Baseado em Eventos para Aquisição e  
Representação de Seqüências Temporais**

Dissertação apresentada como requisito parcial  
para a obtenção do grau de Mestre em Ciência  
da Computação.

Profa. Dra. Mara Abel  
Orientador

Prof. Dr. Luiz Fernando De Ros  
Co-orientador

Porto Alegre, janeiro de 2005.

## CIP - CATALOGAÇÃO NA PUBLICAÇÃO

Mastella, Laura Silveira

Um Modelo de Conhecimento Baseado em Eventos para Aquisição e Representação de Seqüências Temporais / por Laura Silveira Mastella. - Porto Alegre: Programa de Pós-Graduação em Computação, 2005.

163 f.: il.

Dissertação (mestrado) - Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Computação, Porto Alegre, BR - RS, 2005. Orientador: Mara Abel; Co-orientador: Luiz Fernando De Ros.

1. Engenharia de Conhecimento 2. Aquisição de Conhecimento 3. Ontologias Temporais 4. Petrografia Sedimentar. I. Abel, Mara; II. De Ros, Luiz Fernando. III. Título.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitor: Prof. José Carlos Ferraz Hennemann

Vice-Reitor: Prof. Pedro Cezar Dutra Fonseca

Pró-Reitora de Pós-Graduação: Profa. Valquíria Linck Bassani

Diretor do Instituto de Informática: Prof. Philippe Olivier Alexandre Navaux

Coordenador do PPGC: Prof. Flávio Rech Wagner

Bibliotecária-Chefe do Instituto de Informática: Beatriz Regina Bastos Haro

## AGRADECIMENTOS

*As pessoas entram em nossas vidas por acaso,  
mas não por acaso permanecem.*

Durante toda a elaboração do texto da dissertação precisei me conter para não ultrapassar um limite adequado de páginas. Nos agradecimentos não vou medir palavras. Mesmo assim não conseguirei agradecer a todos que contribuíram de alguma forma com este trabalho.

Agradeço primeiramente ao Instituto de Informática da UFRGS e ao Programa de Pós-Graduação em Computação por contribuírem para minha formação me proporcionando educação de qualidade e infra-estrutura ao estudo e à pesquisa desde a graduação. Aos funcionários que estavam sempre dispostos a ajudar, e aos professores, que servem de exemplo a ser seguido. Agradeço pelo orgulho que tenho de fazer parte dessa instituição. Ao CNPq, por dar o suporte financeiro necessário a esta pesquisa.

Aos colegas de pós-graduação, que estavam junto nos momentos de estudo e de lazer, nas reuniões de trabalho e nos almoços no RU. Aos colegas de mestrado que estudam comigo desde o início da graduação, em 1998. Ao Grupo de Estudantes de Bancos de Dados, por todo apoio dirigido aos pós-graduandos “mais novos”. Ao Grupo da Sala 228, que juntou uma turma cuja amizade se estende para além dos laboratórios.

Agradeço o apoio e amizade dos colegas de trabalho do grupo de pesquisa Bancos de Dados Inteligentes: Carlos Ramisch, Cláudio Fuzitaki, Eduardo Castro, Fabiana Lorenzi, Luiz Hermes Svoboda, Marina Minozzo, Maurício Schoenfelder, Patrícia Jaques, Roberto Torres, Sandro Fiorini ! Membros antigos ou novos, formam um grupo unido e eficiente no trabalho, mas que sabe fazer um churrasco animado. Ao Luis Álvaro Silva, por ser uma fonte de referência científica e um dos pilares do projeto *PetroGrapher*. À Renata Galante, por ser um exemplo de pesquisadora e pelas sugestões valiosas durante o mestrado. Um especial agradecimento ao Felipe Victoreti, por ser um colega de trabalho fantástico mesmo na posição difícil de gerenciar essa turma toda, e à Kelen Meregali, que se tornou uma amiga, mesmo para os momentos de dificuldade mais improváveis.

Agradeço ao Professor Luis Lamb pelas discussões elucidativas nos últimos meses da dissertação e pela disposição em dar continuidade ao trabalho.

Ao meu co-orientador e especialista deste trabalho, Professor Luiz Fernando De Ros, pelo tempo dedicado às inúmeras sessões de entrevistas, e pela disponibilidade e paciência em compartilhar seu conhecimento em geologia com uma estudante de computação.

Agradeço de maneira especial à pessoa que orientou cada passo do meu caminho na carreira acadêmica desde a iniciação científica até o título de mestrado. Mara Abel, orientadora, professora, colega de trabalho, companheira, amiga. Nesses quase cinco

anos desde o ingresso no grupo BDI, tive o privilégio de trabalhar com um exemplo de profissional, cujas preocupações com a equipe de trabalho vão dos recursos financeiros ao suporte emocional e espiritual. Agradeço a confiança inabalável, o incentivo à minha carreira, as oportunidades de experiências maravilhosas, a orientação rigorosa na hora certa. Espero ser capaz de retribuir o apoio que ela proporcionou à minha formação como profissional e como pessoa.

Aos amigos. Seria impensável percorrer esse caminho sem os momentos de leveza proporcionados por eles. Obrigada pela compreensão com minhas ausências nas festas, e por me darem garantia de que estariam me esperando no final para comemorarmos juntos. Às amigas do “clube da Lulu”, para as quais eu corria todas as vezes em que precisava esquecer que estava fazendo mestrado. É impossível citar o nome de todos os amigos que tornaram essa tarefa menos solitária. Mas citarei duas pessoas especiais: Mariângela Vanzin e Taisa Novello, que estiveram sempre ao meu lado nos últimos anos, dividindo os maiores momentos de crise e as grandes alegrias. Obrigada por deixarem a porta e o coração sempre abertos para me receberem.

Agradeço com carinho a Felipe Fehlberg pelo incentivo constante ao meu crescimento, pela amizade e por ser uma pessoa ainda presente em minha vida.

A toda minha família por sempre compartilharem minhas conquistas com alegria. O reconhecimento de minha família é o fio de prumo que me faz saber que estou no caminho certo. Agradeço principalmente a meus pais, Anita e Wilson, e a minha irmã Livia, por me apoiarem neste período, que é mais difícil para aqueles que estão mais próximos. Obrigada pela paciência com uma mestranda que nem sempre soube lidar com sua ansiedade. Agradeço principalmente a meu pai, que, mesmo discordando de algumas escolhas, me apoiou todas as vezes, e à minha mãe, que deu amor e atenção a uma filha por vezes eufórica e por vezes irritada. Obrigada pelo respeito aos meus momentos de estudo, pelos agrados nas horas certas, pelo apoio irrestrito e pelo papel fundamental na minha vida. Este trabalho é dedicado a vocês.

A Deus, pela inspiração para escrever, pela iluminação para as melhores escolhas e pela força para os momentos difíceis. Meu singelo agradecimento.

# SUMÁRIO

<b>LISTA DE ABREVIATURAS E SIGLAS .....</b>	<b>9</b>
<b>LISTA DE FIGURAS .....</b>	<b>10</b>
<b>LISTA DE TABELAS .....</b>	<b>12</b>
<b>RESUMO.....</b>	<b>13</b>
<b>ABSTRACT.....</b>	<b>14</b>
<b>1 INTRODUÇÃO .....</b>	<b>15</b>
<b>1.1 Domínio de Aplicação .....</b>	<b>15</b>
1.1.1 Petrografia Sedimentar .....	16
1.1.2 Sistema <i>PetroGrapher</i> .....	17
<b>1.2 Objetivo do Trabalho .....</b>	<b>18</b>
<b>1.3 Metodologia de Trabalho .....</b>	<b>20</b>
<b>1.4 Organização dos Capítulos .....</b>	<b>20</b>
<b>2 ENGENHARIA DE CONHECIMENTO.....</b>	<b>22</b>
<b>2.1 Classificação de Conhecimento .....</b>	<b>23</b>
2.1.1 Conhecimento Declarativo e Conhecimento Procedimental.....	23
2.1.2 Conhecimento Tácito e Conhecimento Focal .....	23
2.1.3 Conhecimento Implícito e Conhecimento Explícito .....	24
<b>2.2 Classificação de Técnicas de Aquisição de Conhecimento .....</b>	<b>24</b>
2.2.1 Conforme o Instrumento de Aplicação da Técnica .....	24
2.2.2 Conforme o Tipo de Conhecimento Eliciado .....	25
2.2.3 Conforme o Tipo de Interação .....	26
2.2.4 Conforme a Familiaridade do Agente com a Técnica .....	26
2.2.5 Outras Classificações .....	27
2.2.6 Aquisição e Eliciação de Conhecimento .....	27
<b>2.3 Técnicas de Eliciação de Conhecimento.....</b>	<b>28</b>
2.3.1 Imersão na Literatura .....	28
2.3.2 Entrevistas Não-Estruturadas .....	29
2.3.3 Entrevistas Estruturadas .....	29
2.3.4 Classificação de Conceitos ou de Fichas ( <i>Card Sorting</i> ) .....	30
2.3.5 Comparação Triádica .....	31
2.3.6 Grades de Repertório ( <i>Repertory Grids</i> ) .....	32
2.3.7 Técnicas de Observação .....	32
2.3.8 Limitação de Informações (Focalizando Contextos) .....	33
2.3.9 Relatórios Verbais .....	34
2.3.10 Análise de Protocolos .....	34
<b>2.4 Ontologias .....</b>	<b>35</b>

2.4.1	Ontologias para Representação de Conhecimento (RC).....	35
2.4.2	Meta-Ontologias (ou <i>Top-Level Ontologies</i> ) .....	36
2.4.3	Ontologias Lingüísticas.....	36
2.4.4	Ontologias de Domínio .....	36
<b>2.5</b>	<b>Metodologias para Desenvolvimento de Sistemas Baseados em Conhecimento .....</b>	<b>37</b>
2.5.1	A Suíte de Modelos CommonKADS para Representação de Conhecimento .....	37
<b>2.6</b>	<b>Resumo do Capítulo .....</b>	<b>39</b>
<b>3</b>	<b>MODELOS PARA REPRESENTAÇÃO DE TEMPO .....</b>	<b>41</b>
<b>3.1</b>	<b>Teorias de Tempo Absoluto e de Tempo Relativo.....</b>	<b>42</b>
<b>3.2</b>	<b>Classificação de Requisitos Temporais .....</b>	<b>44</b>
3.2.1	Requisitos Temporais Incondicionais .....	44
3.2.2	Requisitos Temporais Condicionais.....	44
<b>3.3</b>	<b>Questões para Representação Temporal.....</b>	<b>44</b>
3.3.1	A escolha de uma entidade primitiva de tempo. ....	45
3.3.2	Escolha da ordem de tempo. ....	45
3.3.3	Especificação da estrutura do tempo. ....	46
3.3.4	Limitações do tempo. ....	46
3.3.5	Introdução de uma métrica de tempo. ....	46
3.3.6	Raciocínio sobre o tempo. ....	47
<b>3.4</b>	<b>Principais Linguagens de Representação Temporal.....</b>	<b>47</b>
3.4.1	Lógicas de primeira-ordem com argumentos temporais (AT).....	48
3.4.2	Lógicas modais temporais (LMT) .....	49
3.4.3	Cálculo de eventos .....	49
3.4.4	Cálculo de situações.....	51
3.4.5	Álgebra de intervalos de Allen.....	53
<b>3.5</b>	<b>Resumo do Capítulo .....</b>	<b>55</b>
3.5.1	Sobre a Nomenclatura de Eventos .....	56
<b>4</b>	<b>ANÁLISE DE MODELOS TEMPORAIS PARA REPRESENTAR ORDEM DE EVENTOS NA PETROGRAFIA SEDIMENTAR .....</b>	<b>57</b>
<b>4.1</b>	<b>Lógicas de Primeira Ordem.....</b>	<b>57</b>
<b>4.2</b>	<b>Lógicas Modais Temporais (LMT) .....</b>	<b>57</b>
<b>4.3</b>	<b>Cálculo de Eventos .....</b>	<b>58</b>
<b>4.4</b>	<b>Cálculo de Situações.....</b>	<b>61</b>
<b>4.5</b>	<b>Álgebra de Intervalos.....</b>	<b>61</b>
<b>4.6</b>	<b>Resumo do Capítulo .....</b>	<b>62</b>
<b>5</b>	<b>ESTUDO DE CASO DE AQUISIÇÃO DE CONHECIMENTO .....</b>	<b>64</b>
<b>5.1</b>	<b>Planejamento das Sessões de AC .....</b>	<b>64</b>
<b>5.2</b>	<b>Processo de Aquisição de Conhecimento sobre Interpretação de Seqüência Diagenéticas .....</b>	<b>65</b>
5.2.1	Fase Anterior à Entrevista: Imersão na Literatura .....	65
5.2.2	Fase 1: Entrevista Não-Estruturada .....	66
5.2.3	Fase 2: Nova Imersão na Literatura .....	66
5.2.4	Fase 3: Entrevista Estruturada .....	67
5.2.5	Fase 4: Observação.....	67
5.2.6	Fase 5: Mudança da Saída da Interpretação .....	68
5.2.7	Fase 6: Classificação de Fichas .....	69
5.2.8	Fase 7: Observação: Detectar Associação entre Produtos e Processos .....	71

5.2.9 Fase 8: Entrevista estruturada: Detectar Associação entre Produtos e Processos .....	71
5.2.10 Fase 9: Técnica para Associar Relações Visuais a Relações Temporais .....	73
5.2.11 Fase 10: Eliciação Triádica .....	74
5.2.12 Fase 11: Grades de Repertório .....	75
5.2.13 Fase 12: Limitação de Informações (Focalizando Contextos) .....	76
5.2.14 Fase 13: Eliciar Propriedades das Relações .....	77
5.2.15 Fase 14: Análise de Protocolos .....	78
<b>5.3 Resultados do Processo de Aquisição de Conhecimento .....</b>	<b>80</b>
5.3.1 Conhecimento Resultante do Processo de AC .....	80
5.3.2 Avaliação Geral das Técnicas de AC .....	82
<b>5.4 Resumo do Capítulo .....</b>	<b>83</b>
<b>6 ONTOLOGIA DE EVENTOS .....</b>	<b>84</b>
<b>6.1 Avaliação dos Quesitos Temporais do Domínio .....</b>	<b>84</b>
6.1.1 Requisitos Temporais .....	85
6.1.2 Entidade Primitiva de Tempo .....	85
6.1.3 Escolha da Ordem de Tempo .....	85
6.1.4 Especificação da Estrutura do Tempo .....	85
6.1.5 Limitações do Tempo .....	86
6.1.6 Introdução de uma Métrica de Tempo .....	86
<b>6.2 Construtos da Ontologia de Eventos .....</b>	<b>86</b>
6.2.1 Conceitos .....	86
6.2.2 Eventos .....	87
6.2.3 Relações .....	87
6.2.4 Relação temporal .....	88
6.2.5 Tipos-Regra .....	88
<b>6.3 Resumo do Capítulo .....</b>	<b>89</b>
<b>7 PROPOSTA DE MODELO DE CONHECIMENTO PARA ORDENAÇÃO DE EVENTOS NA PETROGRAFIA SEDIMENTAR .....</b>	<b>90</b>
<b>7.1 Domínio de Aplicação .....</b>	<b>90</b>
<b>7.2 Modelo do Conhecimento de Domínio .....</b>	<b>91</b>
7.2.1 Conceitos e Eventos .....	91
7.2.2 Relação Temporal .....	92
7.2.3 Relações .....	93
7.2.4 Tipos-regra .....	96
<b>7.3 Modelo do Conhecimento de Inferência .....</b>	<b>100</b>
7.3.1 Inferências para Identificação de Eventos .....	101
7.3.2 Inferências para Ordenação de Eventos .....	102
<b>7.4 Modelo do Conhecimento de Tarefa .....</b>	<b>105</b>
7.4.1 Tarefa de Identificação de Eventos .....	106
7.4.2 Tarefa de Ordenação de Eventos .....	107
<b>7.5 Resumo do Capítulo .....</b>	<b>108</b>
<b>8 IMPLEMENTAÇÃO E VALIDAÇÃO DO MODELO DE CONHECIMENTO PARA ORDENAÇÃO DE EVENTOS .....</b>	<b>109</b>
<b>8.1 Implementação do Conhecimento de Domínio .....</b>	<b>109</b>
<b>8.2 Implementação do Conhecimento de Inferência e de Tarefa .....</b>	<b>111</b>
8.2.1 Entrada de Dados .....	111
8.2.2 Implementação da Tarefa de Identificação dos Eventos .....	112

8.2.3 Implementação da Tarefa de Ordenação dos Eventos .....	113
<b>8.3 Demonstração do Algoritmo com Dados Reais .....</b>	<b>115</b>
8.3.1 Entrada de Dados .....	115
8.3.2 Identificação dos Eventos.....	116
8.3.3 Ordenando os Eventos.....	117
<b>8.4 Apresentação Gráfica da Seqüência Diagenética .....</b>	<b>118</b>
<b>8.5 Validação do Algoritmo .....</b>	<b>118</b>
8.5.1 Interpretação da amostra <i>WINT01 SWC8 - 4616,5</i> .....	119
<b>9 CONCLUSÃO .....</b>	<b>122</b>
<b>9.1 Aquisição e Representação de Conhecimento para Interpretação de Seqüência Diagenética.....</b>	<b>122</b>
<b>9.2 Outros domínios .....</b>	<b>123</b>
<b>9.3 Sugestões para trabalhos futuros .....</b>	<b>124</b>
<b>REFERÊNCIAS.....</b>	<b>126</b>
<b>APÊNDICE A TRANSCRIÇÕES DE ENTREVISTAS COM O ESPECIALISTA EM PETROGRAFIA SEDIMENTAR .....</b>	<b>133</b>
<b>APÊNDICE B MODELO DE CONHECIMENTO PARA INTERPRETAÇÃO DE SEQÜÊNCIA DE EVENTOS DIAGENÉTICOS .....</b>	<b>144</b>
<b>ANEXO A REGRAS DE FORMAÇÃO DA SINTAXE DA LINGUAGEM CML .....</b>	<b>161</b>



## LISTA DE ABREVIATURAS E SIGLAS

AC	Aquisição de Conhecimento
AT	Argumentos Temporais
BD	Banco de Dados
BNF	Backus-Naur Form
CML	Conceptual Modeling Language
COBOL	Common Business Oriented Language
CYC	Cycorp Knowledge Server
CycL	CYC Language
DAML	DARPA Agent Markup Language
ES	Engenharia de Software
F-Logic	Frame Logic
FO	Frame Ontology
GUM	Generalized Upper Model
HRDM	Historical Relational Data Model
IA	Inteligência Artificial
IEEE	Institute of Electrical and Electronics Engineers
KL	Knowledge Levels
LISP	List Processing
LMT	Lógica Modal Temporal
OIL	Ontology Inference Layer
OKBC	Open Knowledge Base Connectivity
OWL	Ontology Web Language
PLN	Processamento de Linguagem Natural
P/OM	Production/Operation Management
PROLOG	Programming in Logic
RC	Representação de Conhecimento
RDF	Resource Description Framework
RDF(S)	RDF Schema
SBC	Sistemas Baseados em Conhecimento
SGBD	Sistema de Gerenciamento de Bancos de Dados
SI	Sistemas de Informação
TEER	Temporal Enhanced Entity-Relationship
UML	Unified Modeling Language

## LISTA DE FIGURAS

Figura 1.1: Ontologia de domínio do sistema <i>PetroGrapher</i> apresentando a partonomia de conceitos empregados para descrição e interpretação de amostras de rocha. ....	18
Figura 2.1: Técnicas de aquisição de conhecimento em relação ao tipo de conhecimento que se espera como resultado de sua aplicação. ....	25
Figura 3.1: Representação das propriedades mantidas antes e depois dos eventos <i>e1</i> e <i>e2</i> . ....	50
Figura 3.2: Representando o conhecimento sobre relações como uma rede de intervalos. ....	54
Figura 4.1: Propriedades que se mantém nos intervalos de tempo antes e depois do evento <i>Precipitação(mineral1)</i> . ....	59
Figura 4.2: Propriedades que se mantém nos intervalos de tempo antes e depois do evento <i>Dissolução(mineral2)</i> . ....	59
Figura 4.3: Representação da relação de ordem entre os eventos <i>Precipitação(mineral2)</i> e <i>Dissolução(mineral2)</i> . ....	60
Figura 5.1: Resultado da interpretação de seqüência diagenética feita pelo especialista para uma amostra de rocha. ....	68
Figura 5.2: Exemplo de interpretação de seqüência diagenética, demonstrando os processos associados aos constituintes. A descrição é feita como uma seqüência linear de processos. ....	69
Figura 5.3: Interface para aquisição das associações entre produtos e processos. ....	73
Figura 5.4: Imagem de parte da tela principal da ferramenta de análise de protocolos da suíte PCPack mostrando as marcações de regras temporais. ....	79
Figura 5.5: Imagem da tela principal da ferramenta de análise de protocolos PCPack da suíte mostrando as marcações de produtos e processos. ....	79
Figura 5.6: Imagem da da ferramenta de construção de diagramas da suíte PCPack. ...	80
Figura 6.1: Ontologia de representação de conhecimento que propõe construtos para representar eventos e relações de ordem. ....	86
Figura 7.1: Descrição dos principais conceitos e eventos da ontologia de eventos: <i>constituante-de-rocha</i> , <i>evento-diagenético</i> , e os subtipos de <i>constituante-de-rocha</i> , <i>constituante-mineral</i> , <i>constituente-macroporosidade</i> , <i>constituente-detrítico</i> e <i>constituente-diagenético</i> . ....	91
Figura 7.2: Descrição do conceito <i>constituente-de-rocha</i> e do evento <i>evento-diagenético</i> , empregando a sintaxe CML. ....	92
Figura 7.3: Descrição da relação temporal entre eventos diagenéticos. ....	92
Figura 7.4: Descrição da <i>relação-temporal</i> e da sub-relação <i>depois</i> , empregando a sintaxe CML. ....	93
Figura 7.5: Seqüência de eventos em ordem de tempo ramificada. ....	93

Figura 7.6: Descrição do relacionamento visual (relação paragenética) entre constituintes. ....	95
Figura 7.7: Descrição da <i>relação-paragenética</i> e da sub-relação <i>abrindo</i> , empregando a sintaxe CML. ....	95
Figura 7.8: Descrição da relação <i>produzido-por</i> empregando a sintaxe CML e no modelo gráfico. ....	96
Figura 7.9: Descrição do tipo-regra que define a associação entre constituintes e eventos. ....	97
Figura 7.10: Exemplos de instâncias do tipo-regra indicação-evento: regras que indicam o evento diagenético a partir de características do constituinte visualizado. ....	97
Figura 7.11: Definição do tipo-regra que retorna a relação de ordem entre constituintes a partir da relação paragenética entre eles. ....	99
Figura 7.12: Exemplos de instâncias do tipo-regra implicação-temporal: a relação temporal é implicada a partir da relação visual. ....	100
Figura 7.13: Modelo de inferência para identificação de eventos. ....	101
Figura 7.14: Modelo de inferência para ordenação de eventos. ....	103
Figura 7.15: Elementos C e D adicionados na seqüência. ....	104
Figura 7.16: Elemento E adicionado na seqüência depois do elemento C. ....	104
Figura 7.17: Elementos A e B adicionados na seqüência. ....	104
Figura 7.18: Elementos B e C colocados em ordem na seqüência. ....	105
Figura 7.19: Tarefa de interpretação de seqüência de eventos decomposta em sub-tarefas. ....	105
Figura 7.20: Método da tarefa de identificação de eventos. ....	107
Figura 7.21: Método da tarefa de ordenação de eventos. ....	108
Figura 8.1: Exemplo de descrição de dois tipos de constituintes de uma amostra de rocha – constituinte diagenético (clorita) e constituinte detrítico (esfeno). ....	112
Figura 8.2: Instância da classe que representa a relação paragenética <i>Substituindo</i> . ...	113
Figura 8.3: Instância da classe que representa a relação temporal DEPOIS entre os eventos. ....	113
Figura 8.4: Seqüência não linear de aparecimento de constituintes. ....	114
Figura 8.5: Algoritmo de construção do grafo dirigido a partir da lista de relações temporais. ....	114
Figura 8.6: descrição dos constituintes da amostra de rocha S-208 (tela capturada do módulo de descrição composicional do sistema <i>PetroGrapher</i> ). ....	115
Figura 8.7: Instâncias de <i>RelaçãoParagenética</i> - amostra S-208. ....	116
Figura 8.8: Instâncias de <i>RelaçãoTemporal</i> - amostra S-208. ....	117
Figura 8.9: configuração inicial da seqüência diagenética da amostra S-208. ....	117
Figura 8.10: configuração final da seqüência diagenética da amostra S-208. ....	118
Figura 8.11: Saída resultante do desenho do grafo dirigido montado para a amostra S-208. ....	118
Figura 8.12: Interpretação da seqüência diagenética feita pelo especialista (amostra SWC8). ....	119
Figura 8.13: Sugestão da seqüência diagenética pelo algoritmo de inferência (amostra SWC8). ....	120

## LISTA DE TABELAS

Tabela 2.1: Organização de técnicas de aquisição de conhecimento por similaridade entre os processos de aplicação do método.....	26
Tabela 2.2: Resultado da técnica de grade de repertório. ....	32
Tabela 3.1: Algumas noções que podem ser representadas na lógica modal e os significados dos operadores aplicados a uma fórmula F. ....	49
Tabela 3.2: Alguns dos operadores do cálculo de situações.....	52
Tabela 3.3: As treze possíveis relações temporais entre intervalos. ....	53
Tabela 3.4: Algumas restrições de transitividade resultantes da combinação de relações.....	54
Tabela 5.1: Grade de repertório para cotar eventos em relação aos modificadores dos constituintes que geram. ....	75
Tabela 5.2: Algumas propriedades de relações do domínio da petrografia. ....	78
Tabela 5.3: Vantagens e desvantagens de métodos de AC.....	82
Tabela 7.1: Propriedades definidas para as relações do domínio (S=sim, N=não). ....	94
Tabela 8.1: Tabela relacional <i>DiageneticProcessNomenclature</i> que representa a nomenclatura de eventos diagenéticos. ....	110
Tabela 8.2: Tabela relacional <i>ParageneticOrderRelation</i> que representa associações entre relações paragenéticas e relações de ordem. ....	110
Tabela 8.3: Tabela relacional <i>ModifierProcess</i> que representa as associações entre modificadores e eventos (chave estrangeira para a tabela de nomenclatura de eventos, representada na Tabela 8.1).....	111
Tabela 8.4: Tabela relacional <i>ParageneticRelationProcess</i> que representa as associações entre relações paragenéticas e eventos (chave estrangeira para a tabela de nomenclatura de eventos, representada na Tabela 8.1)..	111

## RESUMO

Ontologias de representação de conhecimento são modelos que agrupam e definem construtos de modelagem (tais como conceito, classe, frame). Uma ontologia temporal é uma especificação de uma conceitualização do domínio acrescida dos aspectos de tempo sobre os objetos do domínio. No caso de esta ontologia tratar somente de aspectos temporais ligados a eventos, ela é chamada ontologia de eventos. Este trabalho propõe uma *extensão de uma ontologia de representação de conhecimento (RC) com construtos temporais*, ou seja, estender um modelo de conceitualização de domínios com os construtos necessários para se representar eventos. A definição original de uma ontologia de RC foi estendida neste trabalho com os seguintes construtos: (i) o construto *evento* (que define um evento como um acontecimento que pode alterar objetos do domínio) e (ii) o construto *relação-temporal* (que define uma relação de tempo entre os eventos).

O objetivo deste trabalho é, baseando-se nos novos construtos propostos na ontologia de RC, criar um modelo de conhecimento específico para um domínio, o qual suporta a representação das seguintes primitivas: *características do domínio*, *eventos geradores* das características do domínio, *associações* entre características do domínio, *relações temporais* entre eventos. Além disso, o modelo tem o objetivo de suportar métodos de raciocínio para inferir a seqüência em que as características do domínio foram geradas pelos eventos, ou seja, inferir a ordem dos eventos.

O domínio de aplicação deste trabalho é a *petrografia sedimentar*. Esse domínio apresenta diversas tarefas de interpretação que exigem conhecimento especializado para sua solução. Dentre estas tarefas, o foco de estudo deste trabalho é a *interpretação de seqüência diagenética*, a qual procura identificar a ordem em que os processos diagenéticos atuaram sobre os constituintes das rochas sedimentares.

Para compreender e coletar o conhecimento necessário para a modelagem dessa tarefa, as principais técnicas de aquisição de conhecimento para sistemas baseados em conhecimento foram aplicadas em sessões com um especialista em petrografia sedimentar, gerando os principais objetos do domínio que precisavam ser modelados para se atingir o objetivo da ordenação de eventos. Esses objetos foram representados como um modelo de conhecimento específico para o domínio da petrografia, aplicando os construtos de eventos e relação temporal. Esse modelo foi validado através da implementação de um módulo de inferência no sistema *PetroGrapher*. A validação foi efetuada comparando as respostas do algoritmo com as interpretações do especialista, e foi possível notar que apesar de a apresentação gráfica da seqüência de eventos ainda ser diferente da original do especialista, o algoritmo efetua a ordenação de eventos corretamente.

**Palavras-chave:** Engenharia de Conhecimento, Aquisição de Conhecimento, Ontologias Temporais, Petrografia Sedimentar.

# An event-based knowledge model to support acquisition and representation of temporal sequences.

## ABSTRACT

Knowledge representation ontologies are models that group and define modeling constructs (such as concept, class, frame). Time ontologies are specifications of domain conceptualizations that deal with aspects of time. When the ontology is aware of representing just events, it is called event ontology. This work proposes an extension of a knowledge representation (KR) ontology with time constructs, that is, it extends a domain conceptualization model with the constructs necessary to represent events. The original definition of a specific KR ontology was extended in this work with the following constructs: (i) the construct *event* (that defines an event as an occurrence that can modify objects of the domain) and (ii) the construct *time-relation* (that defines a time relationship between the events).

The objective of this work is to create a knowledge model for a specific domain based on the new constructs proposed in the KR ontology. This knowledge model aims to support the representation of the following domain primitives: characteristics of the domain, events that generated the characteristics of the domain, associations among these characteristics, time relationships between events. Moreover, the model has the purpose of supporting reasoning methods to infer the sequence in that the characteristics of the domain were generated by the events; in other words, to infer the order of the events.

The application domain of this work is the *sedimentary petrography*. This domain presents several interpretation tasks that demand specialized knowledge for their solution. Among these tasks, the focus of this work is the *diagenetic sequence interpretation*, which aims to identify the order in that diagenetic processes occurred over the constituents of sedimentary rocks.

In order to understand and collect the necessary knowledge for modeling this task, the main techniques of knowledge acquisition for knowledge-based systems were applied to an expert in sedimentary petrography. These sessions of knowledge acquisition revealed the domain objects that needed to be modeled to reach the objective of the ordering of events. Those objects were represented in a knowledge model for the petrography domain, which applied the proposed constructs (events and time relation). That model was validated by the implementation of an inference module in the *PetroGrapher* system. The validation proceeded by comparing the outputs of the algorithm with the expert's interpretations, and it was possible to notice that despite the graphic presentation of the sequence of events being different from what the expert produces as a sequence of events, the algorithm is able to order the events correctly.

**Keywords:** Knowledge Engineering, Knowledge Acquisition, Time Ontologies, Sedimentary Petrography.

# 1 INTRODUÇÃO

Esse trabalho propõe um modelo de ontologia de eventos aplicado a uma tarefa específica do domínio da Petrografia Sedimentar. Essa tarefa abrange interpretações sobre a ordem em que determinados eventos geológicos ocorreram, e para que essas interpretações possam ser executadas por um sistema computacional, é necessário representar os aspectos que determinam as relações de ordem entre os eventos. O estudo pretende compreender o atual estado da arte em modelos de representação de tempo e coletar, através de técnicas de aquisição de conhecimento, os principais requisitos para esse tipo de interpretação. Dessa maneira, será possível indicar como as idéias dos modelos temporais existentes podem ser aplicadas para representar relações de ordem entre eventos e para construir sistemas de conhecimento capazes de executar inferência sobre ordenação de eventos. O resultado desse estudo será a proposta de construtos semânticos que representam eventos e relações temporais. Dessa forma será possível criar modelos de conhecimento baseados em eventos. Os modelos resultantes deste estudo poderão ser reutilizados para outros domínios que apresentem a mesma demanda: representar eventos que geraram característica do domínio, relações de ordem e raciocinar sobre a seqüência de eventos.

A motivação desse trabalho é que problemas clássicos que utilizam diferentes tipos de raciocínio envolvem primitivas de tempo: planejamento de ações, escalonamento de recursos com restrições de tempo, monitoração, raciocínio sobre históricos, entre outros que serão apresentados no capítulo 3 deste trabalho. Mas o estudo da representação de aspectos temporais não se restringe somente ao campo dos sistemas de conhecimento. Outras áreas estudam como representar informações temporais da maneira mais fiel, como a de processamento de linguagem natural que utiliza os modelos para, por exemplo, extrair informações temporais de textos. Além disso, a área de banco de dados se preocupa com a modelagem desses aspectos para armazenar e consultar históricos e dados evolutivos.

## 1.1 Domínio de Aplicação

O domínio de aplicação desse trabalho, a *petrografia sedimentar*, é uma subárea de pesquisa da Geologia que estuda rochas sedimentares, particularmente rochas-reservatório de petróleo através da descrição sistemática de amostras dessas rochas, extraídas de poços de exploração de petróleo. Esse estudo tem como objetivo investigar o potencial econômico de poços de petróleo, através da medição da qualidade desses reservatórios.

### 1.1.1 Petrografia Sedimentar

A qualidade de um reservatório de petróleo é determinada pelas características originais das rochas - os minerais que a formam, o volume e forma dos poros (espaços que armazenam fluidos dentro da rocha) - e os processos físico-químicos que possam ter modificado essas características depois da deposição dos sedimentos que originaram as rochas, que são chamados *processos diagenéticos*.

A petrografia é uma atividade desenvolvida em laboratório, onde são realizadas as diferentes análises sobre as rochas. O geólogo gera uma *descrição petrográfica*, que integra os resultados coletados de todas essas análises, trazendo informações importantes sobre todos os aspectos da rocha. Entre outros dados mais sofisticados, são descritos, individualmente, cada um dos constituintes que formam a rocha, incluindo minerais, fósseis e poros. Esses constituintes são descritos quantitativamente e qualitativamente. Para cada constituinte, são detalhadas todas as suas características, tais como cor, aspectos texturais, aspectos de fábrica, localização, hábito. São descritas também as informações de paragênese, que são relações espaciais e temporais possíveis de serem observadas entre os constituintes, chamadas *relações paragenéticas*.

A composição mineral, os aspectos texturais, de fábrica e de paragênese reconhecidos visualmente em uma amostra retratam as condições físico-químicas atuantes durante o processo de formação das rochas, dando pistas sobre a origem dos grãos, o ambiente deposicional e processos diagenéticos que as originaram. A interpretação dessas informações, a partir dos dados de descrição de uma amostra, é realizada em geral por geólogos especialistas em petrografia. A interpretação dos processos que originaram a rocha é o principal objetivo da sua tarefa ao realizar a descrição petrográfica, pois essas informações têm estreita correlação com a qualidade de um reservatório de petróleo e a forma de produção de petróleo a ser adotada, permitindo fazer previsões sobre seu potencial econômico.

Conforme o petrógrafo especialista, a interpretação de amostras de rochas pode ser sintetizada pelas tarefas de *classificação composicional de arenitos* (FOLK, 1974), *caracterização de proveniência* (DICKINSON, 1970; ZUFFA, 1985), *ambientes deposicionais, e diagênese* (ANJOS et al., 2000; MORAD, KETZER e DE ROS, 2000) (*ambientes diagenéticos e seqüência diagenética*). Dentre estas, o objetivo de estudo deste trabalho é a *seqüência diagenética*, pois no projeto do qual este trabalho faz parte as demais interpretações já foram implementadas.

O objetivo da interpretação de seqüência diagenética é identificar a *ordem* em que os processos físicos e químicos (diagenéticos) atuaram sobre a rocha, modificando os sedimentos depositados, formando poros, precipitando ou dissolvendo minerais, compactando ou fraturando os constituintes, até transformar esse conjunto na rocha no estado atual. Em consequência da diagênese, a rocha pode apresentar os grãos mais compactados, pode ser mais porosa, ou pode ter os poros preenchidos por cimentos (constituintes diagenéticos precipitados quimicamente entre os grãos) entre muitas outras características que os processos diagenéticos deixam impressas na rocha.

Esses processos acontecem em grandes intervalos medidos em escalas de tempo geológico. A datação absoluta dos processos formadores ou modificadores da rocha, como em qualquer outro evento geológico, são de difícil medição e as datas obtidas (da ordem de milhares ou milhões de anos) tem pouca influência na determinação da situação atual de uma unidade geológica (tal como um reservatório). Por sua vez, a ordem dos eventos que aconteceram com os sedimentos, que permite compreender o processo de formação da rocha, permite inferir o comportamento espacial daquela rocha em dimensões regionais. Não é a descrição de uma amostra de poucos centímetros de



rocha que determina a qualidade atual de um reservatório, mas a compreensão do processo de formação daquela rocha que atuou em escala regional que oferece dados para a compreensão de um reservatório como um todo.

A informação que indica mais diretamente a ordem entre os produtos de processos (constituintes da rocha) são as relações paragenéticas entre esses constituintes. A relação paragenética é uma disposição visual entre os constituintes de uma rocha: um constituinte pode estar cobrindo outro, dentro de outro, engolfando outro, etc. Essa disposição visual depende da ordem em que os constituintes apareceram na rocha, por isso essa ordem pode ser inferida a partir da relação paragenética.

Por sua complexidade, a geologia de petróleo tem sido constante objeto da pesquisa em Inteligência Artificial nas últimas décadas. As dificuldades em modelar informações para suportar análise petrográfica foram bem equacionadas nos experimentos ligados ao projeto Sysiphus (GAPPA e PUPPE, 1998), que aplica Engenharia de Conhecimento e a metodologia CommonKADS (SCHREIBER et al., 2000) para representar conhecimento sobre Petrologia Ígnea em um sistema especialista.

O domínio da petrografia sedimentar foi estudado e modelado em trabalhos anteriores, resultando em um sistema que aplica conhecimento para a interpretação de amostras de rocha: o sistema *PetroGrapher*, descrito a seguir.

### 1.1.2 Sistema *PetroGrapher*

O sistema *PetroGrapher* é uma aplicação de bancos de dados inteligentes (PARSAYE et al., 1989; BERTINO e ZARRI, 1998) destinada à descrição e interpretação de amostras de rochas sedimentares como suporte para a avaliação de reservatórios de petróleo. Essa aplicação é produto dos trabalhos de pesquisa em Engenharia de Conhecimento de Abel (ABEL, 2001) e Silva (SILVA, 2001) e é detalhado em (ABEL et al., 2004).

O sistema é composto por uma interface visual que provê para o usuário o vocabulário técnico geológico necessário para suportar a descrição correta e completa de amostras de rochas sedimentares. Quando uma descrição é terminada, os dados da amostra são armazenados em um banco de dados relacional e o usuário pode solicitar o início de uma interpretação específica sobre a amostra. Das quatro interpretações de rocha descritas na seção anterior, o *PetroGrapher* implementa, atualmente, duas: a classificação composicional e a interpretação de ambiente diagenético. Essas duas tarefas são modeladas como métodos de solução de problemas sobre a ontologia de domínio do sistema.

O sistema foi desenvolvido a partir de uma ontologia do domínio da petrografia sedimentar, modelada a partir do conhecimento eliciado de um petrógrafo especialista. A ontologia de domínio do sistema representa o vocabulário aplicado por geólogos para realizar a descrição de amostras de rocha e as tarefas de interpretação de rochas, e é modelada como uma partonomia<sup>1</sup> de termos. Uma visão geral da estrutura da ontologia de domínio da petrografia sedimentar é dada na Figura 1.1, que mostra os conceitos instanciados pelo sistema para a tarefa de descrição (a), os conceitos que representam a nomenclatura do domínio (b) e os conceitos instanciados para a inferência (c).

O presente trabalho vai gerar como resultado modelos complementares ao modelo de conhecimento original do *PetroGrapher*: a ontologia de domínio será estendida com primitivas para representar aspectos temporais sobre os conceitos, e o modelo de inferência será adicionado de um novo método de solução de problemas: a interpretação de seqüência diagenética. Além disso, a implementação desses novos modelos vai gerar

---

<sup>1</sup> Partonomia é uma estrutura de objetos em que eles são organizados pela relação parte-de.

como resultado um novo módulo de inferência para o sistema. Os objetivos desse novo módulo do sistema *PetroGrapher* serão detalhados a seguir.

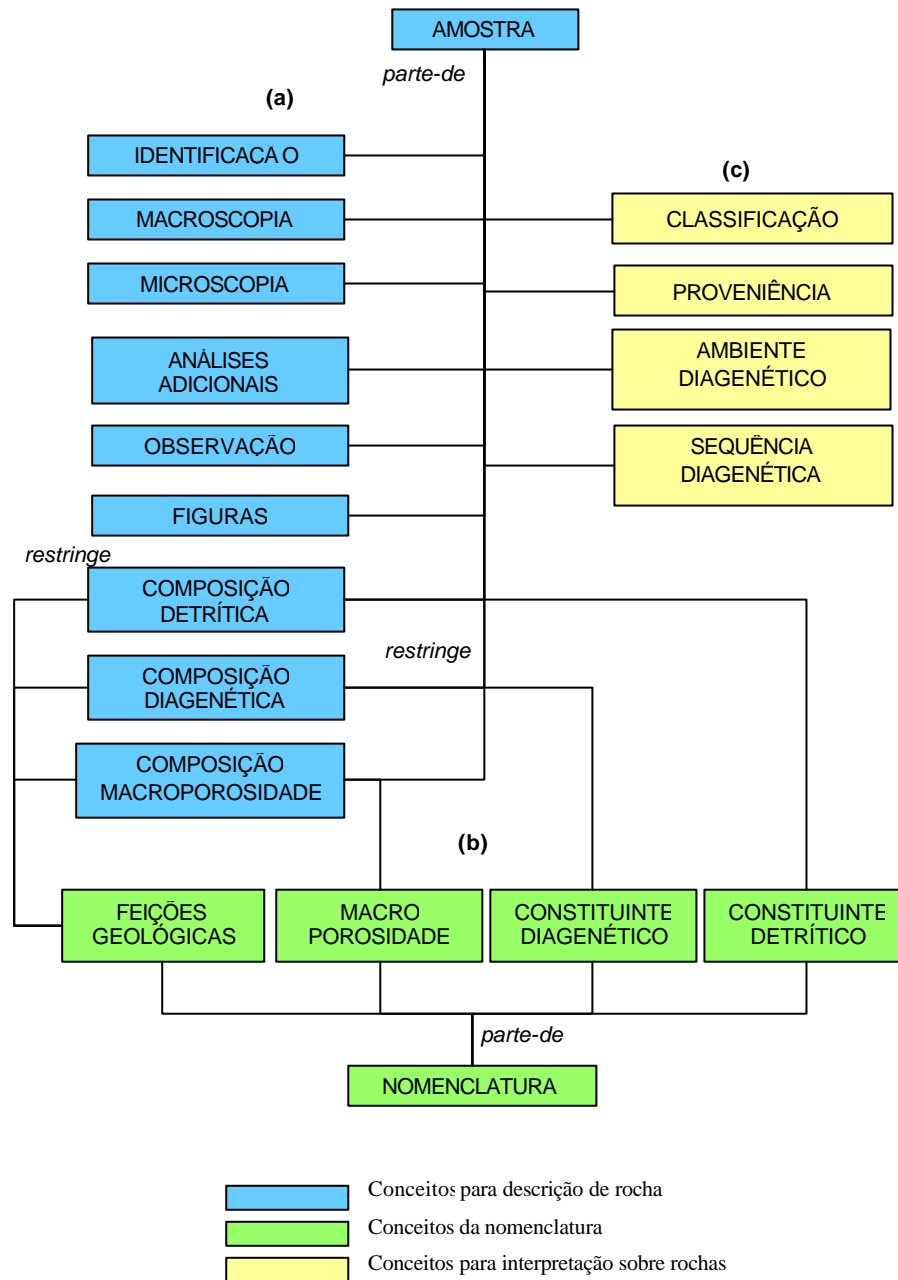


Figura 1.1: Ontologia de domínio do sistema *PetroGrapher* apresentando a partonomia de conceitos empregados para descrição e interpretação de amostras de rocha ((ABEL, 2001)).

## 1.2 Objetivo do Trabalho

O trabalho de conclusão de (MASTELLA, 2002) faz um estudo comparativo de formas de representação de conhecimento para modelar o tipo de conhecimento inferencial das heurísticas do especialista ao interpretar o *ambiente diagenético*, ou seja, as condições sob as quais o reservatório se formou.

A abordagem do presente trabalho é propor modelos de representação e raciocínio sobre a ordem em que ocorreram os eventos diagenéticos, ou seja, a *seqüência*

*diagenética*. Com esse fim, é proposto um modelo que suporta a representação dos eventos geradores das características do domínio, que permita representar os relacionamentos de ordem entre esses eventos, e que suporte a inferência para definir a seqüência em que as características geológicas foram impressas nas rochas. A interpretação de seqüência diagenética tem como objetivo (i) identificar os principais eventos diagenéticos que ocorreram à rocha e (ii) inferir sobre a ordem na qual esses eventos sucederam.

Para representar os aspectos temporais do domínio, serão estudados os principais modelos e linguagens de representação de tempo, com o objetivo de desenvolver uma *ontologia temporal* do domínio, utilizando uma linguagem de representação estendida por construtos temporais. Ontologias provaram ser um elemento essencial em muitas aplicações. Segundo (GÓMEZ-PÉREZ e CORCHO, 2002), ontologias são empregadas para declarar explicitamente o conhecimento de um domínio. Na seção 2.4 será apresentada um levantamento das principais classes de ontologias estudadas atualmente.

A representação de tempo é fundamental para uma base de conhecimento que precise inferir sobre processos, eventos ou mudanças nos dados. Uma ontologia temporal é uma especificação de uma conceitualização do domínio acrescida das primitivas que determinam os aspectos de tempo sobre os objetos da ontologia. Uma ontologia temporal é desenvolvida baseada em uma linguagem temporal. Existem diferentes linguagens que representam tempo na literatura (HAYES, 1995), e existem dificuldades associadas a cada uma delas. Por exemplo, teorias baseadas em instante (MCDERMOTT, 1982; SHOHAM, 1987; BACCHUS, TENENBERG e KOOMEN, 1989) não são naturais para representar eventos que possuem uma duração. Para esse fim, são empregadas as ontologias que se baseiam nas teorias de intervalos, como o cálculo de intervalos de Allen (Allen, 1983), que associa eventos aos intervalos de tempo em que eles ocorreram. Além disso, muito do nosso conhecimento temporal é relativo, ou seja, não está associado a um instante ou intervalo de tempo. São fatos que apresentam um aspecto temporal, tal como passado ou futuro, mas sem que se tenha associado o momento específico em que ocorreram, ou durante quanto tempo se estenderam. Para esse tipo de problema as lógicas modais temporais são mais adequadas. A ontologia temporal proposta neste trabalho será nomeada *ontologia de eventos*, por tratar somente de aspectos temporais ligados à representação de eventos.

Para se definir a ontologia do domínio é necessário primeiro capturar o conhecimento sobre os conceitos do domínio e sobre como eles se organizam. Esse processo é chamado *aquisição de conhecimento* (AC), e é entendido como a parte mais importante da engenharia de conhecimento, pois organiza e torna disponíveis informações estratégicas sobre a resolução de problemas em um domínio. O engenheiro de conhecimento precisa inicialmente compreender o domínio através das fontes (agentes do domínio, especialistas, documentos, etc.) e estruturar o conhecimento através de um modelo formal. Um levantamento das principais categorias de técnicas de aquisição foi organizado, focando principalmente no tipo de conhecimento que cada técnica é destinada a eliciar. Este trabalho pretende aplicar técnicas de AC com o objetivo de identificar as principais características do domínio da petrografia necessárias para se interpretar seqüência de eventos geológicos, e eliciar o conhecimento necessário para proceder com essa inferência.

As informações eliciadas serão representadas como um modelo de conhecimento, que vai compreender o conhecimento sobre os novos conceitos do domínio e as primitivas de tempo, e o conhecimento sobre os métodos de inferência para realizar a tarefa de interpretação de seqüência diagenética. O modelo proposto vai permitir que geólogos com pouca ou nenhuma experiência nessa tarefa tenham acesso a sugestões

sofisticadas e especializadas sobre os eventos diagenéticos e a seqüência em que eles ocorreram.

### 1.3 Metodologia de Trabalho

O objetivo principal desse trabalho é propor um modelo que permita associar as condições atuais do domínio aos seus eventos geradores, e representar relacionamentos de ordem entre esses eventos. Esse trabalho será desenvolvido através das seguintes fases:

- **Estudo de técnicas de aquisição de conhecimento.** As principais técnicas de aquisição de conhecimento para SBC foram organizadas como uma monografia de referência (MASTELLA, 2004), analisando as vantagens e desvantagens de se utilizar uma técnica em detrimento de outra, ou para que tipo de domínio e de conhecimento cada técnica deveria ser aplicada.
- **Aquisição de conhecimento com o especialista em Petrografia Sedimentar,** para eliciar os requisitos necessários. Essa fase gerou os principais objetos do domínio que precisavam ser modelados na ontologia para se representar os aspectos temporais e as heurísticas do especialista para ordenar eventos.
- **Estudo do atual estado da arte em linguagens temporais.** A partir desse estudo foi possível avaliar como as primitivas de eventos existentes seriam empregadas para permitir a representação de eventos específicos para este tipo de domínio.
- **Proposta de construtos semânticos** que representem os eventos e as relações de ordem entre eventos como uma extensão de uma ontologia de representação de conhecimento.
- **Proposta de um modelo de conhecimento para o domínio da petrografia,** que abrange a ontologia do domínio (desenvolvida com base nos construtos da ontologia de eventos) e os respectivos métodos de raciocínio.
- **Implementação do modelo proposto** em um módulo de inferência para suportar a tarefa de interpretação de seqüência de eventos diagenéticos, a ser integrado ao sistema *PetroGrapher*.

### 1.4 Organização dos Capítulos

A revisão sobre o estado da arte dos assuntos abordados neste trabalho será feita em dois capítulos: o capítulo 2 apresenta uma introdução sobre engenharia de conhecimento, abordando o processo de aquisição de conhecimento (um resumo do levantamento do estado da arte apresentado como uma monografia de referência), as ontologias (que são a principal abordagem de representação de conhecimento atualmente) e de metodologias de engenharia de conhecimento, focando a metodologia aplicada a este trabalho, CommonKADS. O capítulo 3 descreve as principais abordagens e formalismos de representação de aspectos temporais.

O processo de aquisição de conhecimento feito no domínio da petrografia sedimentar será detalhado no capítulo 5. Serão apresentadas as técnicas utilizadas, os resultados eliciados e uma avaliação comparativa da eficácia das técnicas. A proposta de construtos de representação de eventos e relações temporais é apresentada no

capítulo 6. Essa ontologia representa um modelo independente de domínio: os construtos modelados são aqueles identificados como necessários para a tarefa de ordenação de eventos. Esses construtos foram usados como base para definir o modelo de conhecimento para a petrografia, descrito no capítulo 7. O capítulo 8 demonstra a validação desse modelo pelo especialista do domínio através da utilização do módulo de inferência desenvolvido.

O anexo A mostra uma parte da BNF da linguagem usada para descrever o modelo de conhecimento, a CML. O apêndice A descreve os resultados mais significativos eliciados das entrevistas feitas com o especialista, tais como as regras de ordenação, as classificações, entre outras transcrições. O apêndice B apresenta o modelo de conhecimento completo do domínio, definido tanto na notação textual, a linguagem CML quanto na notação gráfica, a linguagem UML. O apêndice C mostra os diagramas de classes resultante da implementação do modelo de conhecimento em Java.

## 2 ENGENHARIA DE CONHECIMENTO

O objetivo principal da engenharia de conhecimento é transformar o processo *ad hoc* de construir sistemas baseados em conhecimento (SBC) em uma disciplina de engenharia baseada em métodos, linguagens e ferramentas especializadas (STUDER, BENJAMINS e FENSEL, 1998). Seguindo a idéia de Nível de Conhecimento como proposto por Newell (NEWELL, 1982) – que foi reinterpretado no contexto de desenvolvimento de SBC por Van de Velde em (VELDE, 1993) – o processo de aquisição de conhecimento começou a ser visto como um processo de *modelagem*, em oposição à visão tradicional da *transferência*, em que, para construir um sistema especialista, o conhecimento precisa somente ser transferido diretamente de um especialista para um computador.

Metodologias modernas de engenharia de conhecimento provêm estrutura para o processo de desenvolvimento de sistemas baseados em conhecimento, identificando modelos intermediários e definindo as linguagens e organização destes modelos. Porém, apresentado no trabalho de Abel (ABEL et al., 2005), estas abordagens ainda falham em prover o mesmo grau metodológico para suportar o processo de aquisição de conhecimento, especialmente quando o foco são *fontes humanas*.

A aquisição de conhecimento é o processo de extrair, estruturar e organizar o conhecimento de uma ou mais fontes (TURBAN, 1992). É a atividade inicial do processo de engenharia de conhecimento, realizada por um engenheiro de conhecimento, que é o agente que exerce o duplo papel de compreender o domínio da informação e de selecionar o melhor ambiente e forma de implementar o sistema. Ao mesmo tempo, a AC é o gargalo no desenvolvimento de SBC (HAYES-ROTH, WATERMAN e LENAT, 1983), pois ainda não existe uma metodologia confiável de aplicação que resulte em um modelo do domínio e dos processos cognitivos de seus agentes. Mesmo assim, é uma tarefa indispensável para que as informações essenciais sejam coletadas e o conhecimento chave fique disponível para ser organizado, representado, implementado e validado através de um sistema. Para este trabalho são estudadas e descritas as principais técnicas relacionadas à tarefa de aquisição de conhecimento de fontes humanas.

A importância crescente da abordagem de modelagem (em contrapartida à de transferência) na engenharia de conhecimento moderna, colocou em circulação diferentes formalismos para expressar a parte declarativa do conhecimento, que é capturada como a *ontologia do domínio*. Estes formalismos substituem os modelos da IA clássica anteriormente utilizados, tais como *frames* (FIKES e KEHLER, 1985), regras de produção (BUCHANAN e SHORTLIFFE, 1984), redes semânticas (REBOH, 1981), entre outros. Os principais paradigmas de linguagens usados atualmente para descrever ontologias são: lógicas de descrição (BAADER et al., 2003) e lógica de *frames* (*f-logic*) (ANGELE e LAUSEN, 2004). Estes paradigmas são complementados por RDF (*Resource Description Framework*) (BRICKLEY e GUHA, 2000) e sua mais

recente evolução, a linguagem OWL (*Ontology Web Language*) (ANTONIOU e HARMELEN, 2004) aplicada principalmente em conexão com a web semântica. Na seção 2.4 são apresentadas as principais classes de ontologias reconhecidas no estado da arte. Metodologias de engenharia de conhecimento, tais como PROTÉGÉ (PUERTA et al., 1992) e CommonKADS (SCHREIBER et al., 2000) propõem tipos de conhecimento proprietários para descrever a ontologia de domínio. Na seção 0 é apresentada a metodologia CommonKADS, utilizada para a modelagem de conhecimento neste trabalho.

## 2.1 Classificação de Conhecimento

A seguir são descritos alguns dos principais tipos de conhecimento. Essa divisão não chega a ser taxonômica, pois os tipos não são disjuntos e ainda não há maneira de se fazer uma listagem extensiva de todas as maneiras diferentes nas quais o cérebro pode estruturar as informações. Mas é importante conhecer as formas básicas de organização de conhecimento, principalmente o conhecimento tácito, o qual é considerado o tipo de conhecimento mais complexo de ser coletado e modelado.

### 2.1.1 Conhecimento Declarativo e Conhecimento Procedimental

Basicamente, conhecimento declarativo descreve os aspectos estáticos do conhecimento, ou seja, *o que* são os objetos do mundo. O conhecimento factual é especificado, mas não o modo como aplicar esse conhecimento para solucionar problemas. São representados fatos, relacionamentos, objetos, atributos e outras características estáticas do domínio. É um conhecimento superficial. Já o conhecimento procedimental abrange o conhecimento que está na forma de conjuntos de instruções, ou seja, *como*. São os procedimentos que indicam como as informações declarativas podem ser utilizadas para realizar tarefas sobre domínio ou se fazer inferência sobre fatos gerando novas informações. Pode fornecer uma aplicação imediata para o conhecimento declarativo.

### 2.1.2 Conhecimento Tácito e Conhecimento Focal

Um indivíduo pode possuir e utilizar conhecimento em diversos níveis de consciência. Segundo os estudos de Michael Polanyi (POLANYI, 1974), em cada atividade humana existem duas dimensões de conhecimento que são mutuamente exclusivas: o conhecimento sobre o objeto ou fenômeno que está em foco – *conhecimento focal* – e o conhecimento que é usado como uma ferramenta para manipular ou incrementar o que está em foco – *conhecimento tácito*. Essas duas dimensões são complementares para a realização de tarefas: o conhecimento tácito funciona como um suporte de fundo, ativado de maneira quase inconsciente.

Habilidades que são difíceis de serem articuladas e transmitidas para outras pessoas devem ter uma grande proporção de conhecimento tácito, e é difícil para as pessoas tomarem consciência do conhecimento que suporta suas habilidades. Aquilo que pode ser expresso em palavras e números representa apenas a ponta do iceberg, do conjunto do conhecimento como um todo. Segundo Polanyi, “podemos saber mais do que podemos dizer”. Entretanto, a maneira como esse conhecimento foi organizado e indexado no cérebro dificulta o acesso *focado*, ou seja, a tomada de consciência desse conhecimento, e torna sua articulação quase impossível. Esse é o motivo pelo qual o estudo do conhecimento tácito é importante para o processo de eliciação de

conhecimento. Especialistas utilizam conhecimento que foi internalizado por anos de experiência para executar as tarefas mais complexas de seus domínios de trabalho. Passos de raciocínio que outras pessoas precisariam tomar consciência para resolver o problema são analisados em background por especialistas, o que torna seu método de solução de problemas mais eficaz e eficiente. E deve-se tentar capturar exatamente esse processo de solução para o desenvolvimento de sistemas baseados em conhecimento.

### **2.1.3 Conhecimento Implícito e Conhecimento Explícito**

O termo conhecimento tácito foi novamente citado no trabalho "*The knowledge-creating company*" de Nonaka e Takeuchi (NONAKA e TAKEUCHI, 1995). Esse trabalho, entretanto, imprime um sentido diferente ao termo cunhado por Polanyi. Para Nonaka e Takeuchi, o conhecimento pode ser *tácito* ou *explícito*. Devido a forte influência do trabalho de Nonaka à literatura de Gestão de Conhecimento, o significado mais recente do termo tácito foi adotado. Contudo, alguns autores, como Stenmark (STENMARK, 2002), dizem que teria havido menos confusão se Nonaka tivesse adotado o termo conhecimento *implícito* em vez de *tácito*.

O pressuposto de Nonaka e Takeuchi é de que o conhecimento humano é criado e expandido através da interação social entre o conhecimento implícito e o conhecimento explícito. Chamam essa interação de *conversão do conhecimento* e observam que essa conversão é um processo social entre indivíduos, e não confinada dentro de um indivíduo, que era a visão de Polanyi. Esses autores estabelecem quatro tipos de conversão do conhecimento que eles denominam de socialização, externalização, combinação e internalização, e que se referem à transformação de um tipo de conhecimento em outro.

## **2.2 Classificação de Técnicas de Aquisição de Conhecimento**

Não existe um consenso entre os autores sobre qual a melhor maneira de classificar técnicas de aquisição de conhecimento. Alguns trabalhos as classificam em relação ao instrumento de aplicação da técnica (pessoa ou computador), outros dividem as técnicas em relação à familiaridade do agente do domínio com a técnica. Muitos autores preferem classificá-las utilizando como critério o tipo de conhecimento que pretendem capturar, ou agrupá-las conforme a similaridade do processo de interação com o agente. A seguir serão discutidas algumas formas de classificação de técnicas de aquisição de conhecimento.

### **2.2.1 Conforme o Instrumento de Aplicação da Técnica**

Segundo (BOOSE, 1990), as técnicas de aquisição de conhecimento podem ser classificadas em manuais e baseadas em computador, sendo essa última categoria dividida em semi-automáticas ou interativas, em que o especialista utiliza o computador com supervisão do engenheiro de conhecimento; ou totalmente automáticas, também chamadas de aprendizado de máquina, em que a aplicação de computador detém todo o processo de captura do domínio.

Essas duas últimas envolvem aplicações de redes neurais, ferramentas de modelagem como o *Protégé*, editores de base de conhecimento como o CYC, ou ainda ferramentas que implementam métodos manuais de aquisição. No entanto, ferramentas de aquisição automáticas ou semi-automáticas apresentam muitas desvantagens. A maior delas é o fato de serem generalizadas para qualquer domínio. O resultado da



aquisição podem ser informações muito superficiais sobre o domínio, pois a ferramenta não se adapta conforme os construtos vão sendo compreendidos. Ferramentas automáticas só podem ser aplicadas quando existe uma fonte de informação compreensível pelo computador, ou seja, é limitada a fontes documentadas. Quando a ferramenta semi-automática é projetada especificamente para um domínio particular, é necessário efetuar um processo de aquisição de conhecimento prévio, para identificar a melhor maneira de a ferramenta coletar informações dos usuários: descrição textual? Diagrama? Formulários dirigidos ou seqüência livre de descrição? Por esses motivos, este trabalho considera a etapa manual de aquisição de conhecimento a mais importante para se extrair as informações mais relevantes e significativas sobre o domínio.

## 2.2.2 Conforme o Tipo de Conhecimento Eliciado

A suposição que de diferentes métodos de eliciação derivam diferentes tipos de conhecimento foi chamada por (HOFFMANN, 1992) a *hipótese do acesso diferenciado*. As organizações de técnicas baseadas nessa hipótese tem o objetivo de facilitar e dirigir a escolha das técnicas mais adequadas, conforme as características do domínio do problema e da tarefa.

Nigel Shadbolt e colegas, no sítio de sua companhia *Epistemics* (SHADBOLT e SWALLOW, 1993), definem a seguinte classificação para as técnicas de AC, em relação ao tipo de conhecimento eliciado (Figura 2.1). Os tipos de conhecimento representado nos eixos horizontais e verticais serão mais bem detalhados na seção 2.1. As técnicas apresentadas na Figura 2.1 serão explicadas neste capítulo na seção 2.3.

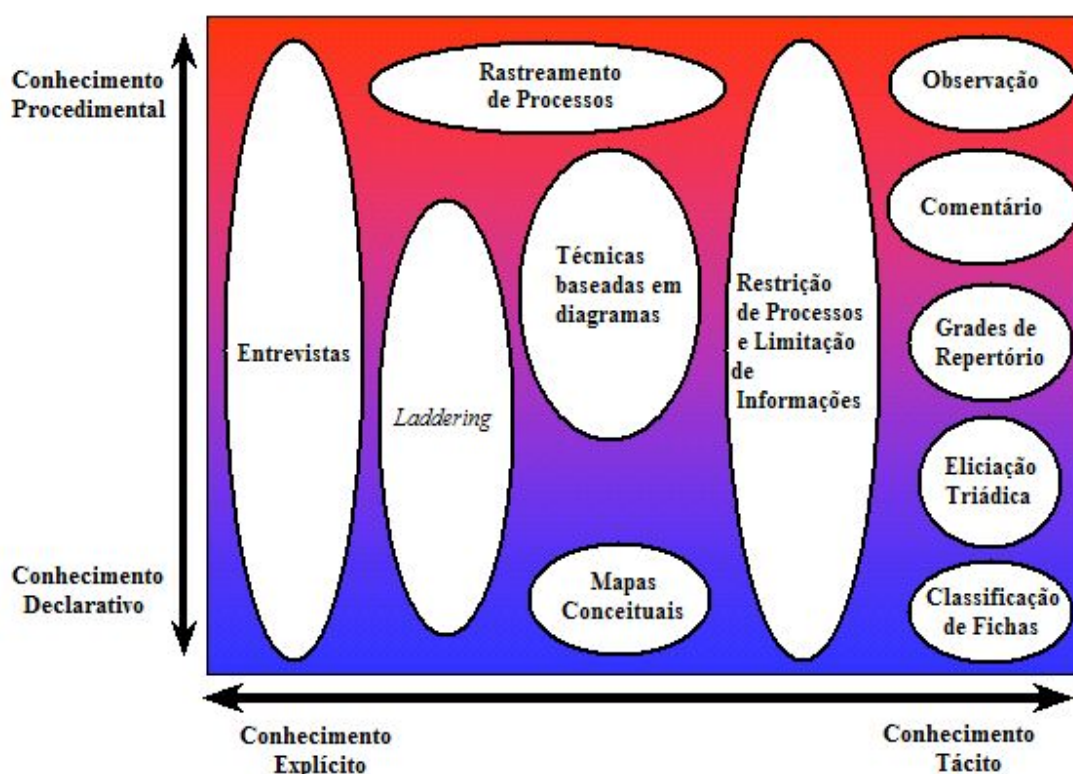


Figura 2.1: Técnicas de aquisição de conhecimento em relação ao tipo de conhecimento que se espera como resultado de sua aplicação (traduzido de (SHADBOLT e SWALLOW, 1993)).

### 2.2.3 Conforme o Tipo de Interação

O trabalho de (COOKE, 1994) afirma que a maneira mais útil de se classificar técnicas de AC está em agrupá-las pela similaridade do *método de interação com o agente*. Cooke sugere essa classificação em contrapartida à organização por *tipo de conhecimento eliciado*, pois muitas vezes não sabemos a priori qual o tipo de conhecimento que deve ser coletado do especialista. A organização por similaridade de técnicas pode ser mais prática para um catálogo a ser consultado para escolha de técnica mais apropriada para a eliciação. As técnicas de AC são classificadas por Cooke em três famílias, divididas em subcategorias, descritas como segue na Tabela 2.1.

Tabela 2.1: Organização de técnicas de aquisição de conhecimento por similaridade entre os processos de aplicação do método.

Família	Subcategorias	Exemplos de técnicas
Família 1 Observações e Entrevistas	1. Observações 2. Entrevistas 3. Análise de tarefa	1. Observação estruturada 2. Entrevistas estruturadas, estudo de caso, questionários 3. Análises de fluxo e seqüência de informação
Família 2 Rastreamento de Processos	1. Relatórios verbais 2. Relatórios não-verbais 3. Análise de protocolos 4. Análise de decisão	1. Pensar alto, discussão de grupo, relatórios retrospectivos 2. Rastreamento do movimento dos olhos 3. Análise de conteúdo, ferramentas de análise automáticas
Família 3 Técnicas conceituais	1. Eliciação de conceitos 2. Coleta de dados 3. Análise estrutural	1. Comparação triádica, <i>laddering</i> 2. Grades de repertório, classificação de fichas 3. Escalas multidimensionais, construção de grafos

No sítio *Epistemics* é definida a seguinte classificação para as técnicas de AC por seus métodos de aplicação:

- **Técnicas de geração de protocolos:** entrevistas, observações e relatórios.
- **Técnicas de análise de protocolos:** utilizadas para analisar os resultados das outras técnicas de AC. São a ponte entre a aquisição e a modelagem de conhecimento.
- **Técnicas baseadas em matrizes:** grades de repertórios e outras técnicas que envolvam a construção de grades para identificar novas informações.
- **Técnicas de classificação:** classificação de fichas, entre outras técnicas usadas para capturar a maneira como as pessoas comparam e ordenam conceitos do domínio.
- **Técnicas de limitação de informações e restrição de processos:** técnicas que limitam o tempo ou a informação disponível para o agente ao executar sua tarefa. Focalizando contextos, por exemplo, é uma maneira de limitar as informações sobre o problema para o agente.
- **Técnicas baseadas em diagramas:** incluem a geração e uso de mapas conceituais, redes de transição de estados, diagramas de eventos e mapas de processos.

### 2.2.4 Conforme a Familiaridade do Agente com a Técnica

Em relação à familiaridade do especialista do domínio com a técnica, elas podem ser divididas em *naturais* e *elaboradas* (no original, *contrived*). As primeiras são técnicas que poderiam ser adotadas naturalmente por um especialista para expressar sua perícia, como, por exemplo, entrevistas, ou relatórios verbais. As técnicas elaboradas são

aquelas não familiares para o especialista, tais como classificação de cartas ou grades de repertório.

As técnicas elaboradas tendem a revelar a maneira que o especialista relaciona os elementos de um domínio. Hoffman em (HOFFMAN et al., 1995) explica que experimentos psicológicos demonstram que modificações em tarefas já familiares podem revelar o raciocínio e o conhecimento do especialista, mesmo que inicialmente isso possa deixá-lo desconfortável. Ou seja, essas técnicas ajudam a descobrir conhecimento implícito, o qual não é facilmente verbalizável pelo especialista e por isso nem sempre é revelada nas técnicas naturais. Entretanto, é preciso analisar se o conhecimento eliciado por essas técnicas representa o *modus operandi* usual do especialista ou se representa uma estratégia de exceção.

### 2.2.5 Outras Classificações

O trabalho de (HOFFMAN et al., 1995), para efeito de discussão, classifica as técnicas de conhecimento em três categorias: (1) análise de tarefas que os especialistas geralmente executam (imersão na literatura, análise de tarefas, relatórios verbais), (2) vários tipos de entrevistas (estruturadas, não-estruturadas) e (3) técnicas elaboradas (análise de decisão, tomada de decisão de grupo, tarefas de classificação e de gradação, limitação de informações, construção de grafos). A explicação do autor para essa divisão é: o que os especialistas fazem? O que os especialistas dizem que fazem? O que eles fazem quando são restringidos de alguma maneira?

### 2.2.6 Aquisição e Eliciação de Conhecimento

As técnicas de AC são também classificadas conforme a fonte de onde se extrai o conhecimento. As fontes *documentadas* são, em geral, mapas, livros, filmes, manuais, bases de dados, diagramas, entre outras. A aquisição de conhecimento de fontes documentadas gera o conhecimento compartilhado, de senso comum, das pessoas sobre um assunto. É fácil de ser coletado pelo fato de já ser, por sua natureza, explícito, externalizado.

O conhecimento *não documentado*, ao contrário, reside na mente das pessoas que atuam no domínio, que representa sua visão de mundo e sua maneira pessoal de solucionar os problemas. É o mais difícil de ser coletado, pois exige interação direta com os agentes, os quais podem estar pouco disponíveis ou até mesmo não saberem se expressar com clareza. O conhecimento documentado, então, é adquirido de fontes humanas.

Quando são utilizadas para extrair informações de fontes humanas (em vez de fontes documentadas) elas são chamadas de técnicas de *eliciação*<sup>2</sup> de conhecimento. A maior parte das técnicas manuais de eliciação de conhecimento são originárias de estudos da psicologia e são todas caracterizadas por serem um tipo de entrevista, mas com uma estrutura bem definida e específica para o tipo de conhecimento que se deseja adquirir.

Em um domínio existem pessoas que são fonte de conhecimento e outras que são trabalhadores tradicionais, cuja atividade não gera nenhum tipo de nova informação. Os *agentes do domínio* são todas as pessoas que fazem parte do domínio e atuam sobre ele, mas nem todos são, necessariamente, fonte de conhecimento estratégico. Já o

---

<sup>2</sup> O termo em inglês "*elicitation*" é traduzido por "eliciação" em grande parte dos trabalhos que estudam técnicas de aquisição. Mas a melhor tradução seria "eliciação", do verbo "eliciar" que significa, conforme o dicionário Aurélio (HOLANDA, 1999): fazer sair; expulsar; **extrair uma resposta ou reação de;** extrair enunciados ou julgamentos lingüísticos de (informante).

*trabalhador de conhecimento* é todo o profissional que utiliza a informação como insumo, combina-a com seu conhecimento individual e gera nova informação como produto de sua atividade (RUAS, ANTONELLO e BOFF, 2005). São as pessoas de quem o conhecimento pode ser extraído e sua denominação muda conforme o ambiente de trabalho: em ambientes empresariais temos funcionários, gerentes, estagiários, alguns tipos de usuários. Em ambientes acadêmicos temos pesquisadores, estudantes, professores, assistentes de pesquisa. Mas em quase todos domínios existe um trabalhador de conhecimento que se destaca pela capacidade de desenvolver métodos particulares de lidar com os problemas: o *especialista*. Especialistas são trabalhadores de conhecimento especializados, que possuem a capacidade de aplicar habilidades cognitivas para resolver problemas em domínios estratégicos, com um desempenho e qualidade de solução superiores à média dos profissionais da área (ABEL, 2001). Isso porque desenvolvem estruturas de armazenamento de informação particulares ao domínio, onde agrupam informações relacionadas com índices de acesso muito eficientes. Estas estruturas tendem a ser mais enxutas e eficazes no reconhecimento de situações do domínio do que as que os novatos utilizam (LEÃO, 1988; ABEL, CASTILHO e CAMPBELL, 1998). Por esses motivos, o especialista é considerado a melhor fonte de conhecimento para uma tarefa de AC. O trabalho de (HOFFMAN et al., 1995) também se refere aos especialistas como *participantes* em vez de *sujeitos* do processo de eliciação, pois ele afirma que se estiverem na posição passiva de sujeitos, os especialistas logo procuram a saída mais fácil. O especialista, para ter participação ativa deve ser visto e tratado como um colaborador, ou ainda como um pesquisador associado<sup>3</sup>.

A próxima seção apresenta um levantamento das principais técnicas de *eliciação* de conhecimento, pois os métodos de aquisição utilizados foram aplicados sempre sobre uma pessoa, o especialista do domínio. No capítulo 5 as técnicas estudadas são testadas e comparadas com o objetivo de fazer uma avaliação de sua eficácia.

## 2.3 Técnicas de Eliciação de Conhecimento

A tarefa de se extrair informações de fontes humanas pode tornar-se um percalço, pois os problemas interpessoais que podem ocorrer entre engenheiro de conhecimento e agentes do domínio influenciam na interação necessária para se coletar conhecimento. Problemas como vocabulário inadequado, insegurança dos agentes ou do engenheiro, medo de expor seus métodos de solução de problemas para a avaliação de colegas e até capacidade de verbalização podem transformar o processo de AC no gargalo da engenharia de conhecimento.

As técnicas de eliciação de conhecimento são empregadas para tentar minimizar esses obstáculos no processo de AC. Nesta seção, são descritas as características de diversas técnicas de eliciação de conhecimento, com o objetivo de relatar os resultados de sua aplicação no capítulo 5.

### 2.3.1 Imersão na Literatura

O objetivo da imersão na literatura é conhecer um mínimo do domínio de trabalho, para que as entrevistas possam começar de um ponto mais adiantado. No entanto, não se deve tentar compreender o domínio, mas sim se familiarizar com os termos usados pelo especialista para poder saber a que ele se refere. Essa familiarização do engenheiro com

---

<sup>3</sup> No caso desta pesquisa, o especialista no domínio também é co-orientador do trabalho, assumindo assim seu posto de pesquisador associado.

o domínio permite que o especialista se sinta mais à vontade para usar os termos do seu jargão nas entrevistas subsequentes. Se o engenheiro mostra total incompreensão do domínio, o especialista tende a usar um vocabulário para novatos nas entrevistas. A consequência disso é que os termos adquiridos nas entrevistas não são aqueles que o especialista realmente usa na solução de problemas. Além disso, se o especialista percebe a incompreensão do engenheiro, ele já tende a considerar a aquisição uma tarefa difícil de concluir.

### **2.3.2 Entrevistas Não-Estruturadas**

Corresponde à técnica inicial de qualquer projeto de Engenharia de Conhecimento, onde o engenheiro de conhecimento conversa diretamente com o especialista, questionando diversos aspectos envolvidos na solução de problemas naquele domínio. Entrevistas não estruturadas apresentam poucas restrições. Isso significa que não existe uma ordem ou seqüência pré-definida de perguntas ou do rumo da conversa. O objetivo não é a aquisição de conhecimento sobre um tópico específico, mas sim que o agente do domínio forneça ao engenheiro uma visão geral do domínio em questão, por isso elas geralmente iniciam com o engenheiro perguntando “Como você resolve esse problema?” (LIOU, 1990).

Conforme Schreiber (SCHREIBER et al., 2000), esta abordagem deve ser usada quando for necessário que o agente e o engenheiro de conhecimento estabeleçam um bom relacionamento. Além disso, essa é a técnica ideal para que o especialista descreva o domínio da maneira que lhe é familiar. Entretanto, pelo fato de o agente poder falar quase livremente sobre seu domínio de trabalho, ele pode se estender em tópicos que não serão importantes para o desenvolvimento do sistema e dar pouca atenção aqueles realmente interessantes. Por isso esse tipo de entrevista não deve ser utilizado além das fases iniciais de identificação do conhecimento e de ambientação do engenheiro com o domínio.

### **2.3.3 Entrevistas Estruturadas**

Nesse tipo de entrevista, mais formal, as perguntas são preparadas para serem relevantes à tarefa sobre a qual se está adquirindo conhecimento. O engenheiro planeja e direciona a conversa utilizando as questões como um guia. Essa técnica é mais eficaz porque ela “força o especialista a ser sistemático ao participar das entrevistas” (HOFFMAN, 1987). Conforme (ABEL, 2001), o ser humano tem a tendência de trazer a entrevista para a sua zona de conforto, ou seja, trazer para o assunto que seja de maior compreensão o que acaba desviando o assunto. Por isso é uma vantagem definir poucas questões gerais sobre o assunto e utilizar somente essas perguntas. Outra estratégia é levar um mediador para acompanhar se os objetivos da entrevista estão sendo alcançados.

A entrevista estruturada é útil quando é necessário preencher as lacunas de conhecimento ainda existentes sobre o domínio para o engenheiro, tais como os principais processos envolvidos na aplicação, a definição do nível do conhecimento esperado no usuário do sistema, a definição do detalhamento (granularidade) das informações a serem modeladas, especificações gerais do sistema a ser proposto e expectativas de interface do sistema.

Entrevistas estruturadas são preparadas para serem específicas sobre o domínio em questão, por isso é difícil encontrar guias para esse processo de entrevista, de modo que o mais importante é a capacidade de análise e comunicação do entrevistador (TURBAN, 1992). Mesmo assim, existem instruções que são genéricas para qualquer tipo de

entrevista, tais como ‘Porque você faria isso?’, que converte uma afirmação em uma regra; ‘Como você faria isso?’, que gera regras de ordem mais baixa; ‘Quando você faria isso? Só existe esse caso?’, que revela a generalidade da regra e pode gerar outras regras; entre outras citadas em (SCHREIBER et al., 2000).

A desvantagem em todos os tipos de entrevistas é que o conhecimento adquirido é somente aquele que pode ser verbalizado pelo agente. O conhecimento tácito e/ou profundo sobre a tarefa, por não ser facilmente externalizado, não será eliciado nessa fase da aquisição. As perguntas mais específicas sobre como o agente soluciona determinados problemas serão respondidas com explicações no estilo *caixa-preta* (*black box*, no original (SCHREIBER et al., 2000) tais como “Eu não sei como faço isso...” ou “É óbvio que o certo a fazer é...”. Isso acontece porque descrições verbais sempre são tendenciosas e incompletas. O especialista explica o processo de solução de problema da forma como ele *acha* que acontece, ou da forma que seja mais fácil para o entrevistador compreender, não da forma como efetivamente acontece. O processo de comunicação não depende apenas do emissor, no caso, do especialista, mas também da capacidade do receptor, o engenheiro de conhecimento, de *decodificar* a informação.

Para grande parte dos domínios, uma parte do conhecimento simplesmente não tem tradução verbal, logo não é explicitado através de entrevistas. Essa dificuldade é tanto maior quanto mais sensório for o domínio. Nesse caso, o especialista desenvolve padrões de reconhecimento visual, tátil ou auditivo e os aplica de forma inconsciente para caracterizar o problema e identificar os caminhos de solução possíveis. Essas respostas serão obtidas mais precisamente com o uso de outras técnicas, específicas para a coleta desse tipo de conhecimento.

As entrevistas são úteis para aumentar o conhecimento do engenheiro sobre o domínio e devem ser utilizadas nas fases iniciais da aquisição de conhecimento. Sua vantagem sobre a imersão na literatura é que o engenheiro fica ciente do conhecimento pela visão do especialista, em vez de adquirir conhecimento de manual. Não é possível, no entanto, basear todo o processo de eliciação em entrevistas, embora pareça ser possível nas etapas iniciais. Com elas é possível identificar o escopo da aplicação, o tipo de domínio do conhecimento e planejar os próximos passos no processo de aquisição, bem como apontar para as técnicas de eliciação mais adequadas para as etapas subsequentes.

#### **2.3.4 Classificação de Conceitos ou de Fichas (*Card Sorting*)**

A técnica de classificação de fichas (WRIGHT e AYTON, 1987) visa a identificação e organização de termos ou conceitos e seus relacionamentos num domínio particular, tais como a classificação dos objetos, as hierarquias, a similaridade entre os conceitos e outras descrições estáticas dos objetos do domínio, segundo a visão do especialista. Uma lista de termos do domínio é obtida a partir das entrevistas iniciais ou extraída da análise de protocolos. Quando a análise de protocolos é feita, para eliciar nomes de conceitos para as fichas, é melhor não separar os substantivos de seus adjetivos. Dessa forma as fichas detêm mais significado para o especialista do que somente o objeto. Também é interessante diferenciar objetos das *ações* sobre esses objetos.

O nome dos objetos é escrito em cartões apresentados ao especialista, que é instruído a classificar os cartões que possuem alguma similaridade ou que são do mesmo tipo em pilhas, formando categorias de conceitos e nomeando-as sempre que possível. As cartas são embaralhadas novamente e o especialista deve agrupá-las novamente, utilizando outros critérios, de forma a criar novas categorias. Esse processo se repete até que não existam mais critérios de separação. Durante a classificação, o especialista irá retirar alguns objetos por serem sinônimos ou irrelevantes, e irá incluir

outros conceitos, que, segundo seu critério, estão ausentes. Esse processo se repete muitas vezes, mas podem ser feitas algumas variações. O resultado da aplicação da classificação de conceitos é um grupo de componentes que compartilham atributos comuns.

Além de ser fácil de ser aplicada e analisada pelo engenheiro, essa técnica pode levar o especialista a perceber certa estrutura no domínio da qual ele mesmo não estava consciente (SCHREIBER et al., 2000). Em geral, na primeira separação, o especialista gera imediatamente as informações mais importantes para fazer tomada de decisão. Costumam ser dois critérios, entre os quais ele separa a maior parte dos conceitos, e um critério chamado *lixo*, pois ele utiliza para agrupar os conceitos que não se adequaram nos outros dois critérios. Como o especialista pode adicionar novas fichas, é bom escrever as originais à mão, para não inibi-lo de adicionar novos conceitos à mão também. Outra boa prática é utilizar números nas fichas, para facilitar a anotação das organizações formadas, ou então fotografar as classificações.

Os resultados obtidos com a técnica de classificação de termos são o reconhecimento da hierarquia do domínio, a obtenção de termos não evidenciados através de entrevistas, o reconhecimento de conceitos que são sinônimos (um objeto mencionado com dois nomes diferentes) além de uma melhor compreensão global do domínio. A técnica é particularmente útil em domínios onde os métodos de solução de problemas são especialmente de classificação. Mesmo nesses casos, a aplicação torna-se difícil nos domínios muito complexos onde um número excessivamente grande de termos pode inviabilizar o uso de cartões. A prática aponta que os melhores resultados são obtidos quando a técnica é aplicada sobre menos do que uma centena de conceitos. Resultados medíocres são também obtidos em domínios excessivamente procedimentais, onde a explicitação das hierarquias de termos não parece fazer senso ao especialista. Essa técnica se mostra eficaz para extrair modelo relacional dos dados. Conforme o trabalho de (ZIULKOSKI, 2003) essa é uma ótima técnica para definir modelos de dados para data warehouse.

### 2.3.5 Comparação Triádica

Em geral essa técnica é usada em conjunto com a técnica mostrada a seguir, as grades de repertório. Empregada para eliciar as dimensões ou construtos. O engenheiro apresenta para o especialista todas as possíveis tríades de objetos do domínio, um por vez. Para cada tríade, o especialista deve escolher dois conceitos como 'similares' e um como o 'diferente' do grupo. Após isso ele deve indicar a razão pela qual diferenciou os três conceitos. Essa informação será usada como uma nova dimensão ou atributo associada aos objetos e que serão usados para diferenciar outros objetos do domínio. Outros autores sugerem uma modificação do método para que se usem escalas de diferenciação (escalas multidimensionais). Os objetos são colocados sobre essa escala, indicando o quanto são similares ou diferentes.

As dimensões eliciadas por essa técnica podem ser usadas em conjunto com outras técnicas para eliciar mais informações. Segundo Johnson (JOHNSON, 1992), o engenheiro deve ter muito cuidado na escolha dos três conceitos a serem apresentados para o especialista, pois pode haver alguma influência na comparação e no agrupamento. Comparar objetos com contrastes muito explícitos, por exemplo, pode induzir o agente a não escolher um atributo de comparação não tão óbvio, mas que, por esse mesmo motivo, poderia ser mais importante para a coleta de conhecimento. É a menos que o número de objetos no domínio do problema seja pequeno, um grande número de comparações precisa ser feito, o que torna esse processo muito demorado para algumas tarefas específicas.

### 2.3.6 Grades de Repertório (*Repertory Grids*)

Essa técnica tem sua origem em um modelo do pensamento humano chamado *teoria dos construtos pessoais* (*personal constructs theory* - PCT), desenvolvida por George Kelly (KELLY, 1955) em 1955, na qual conceitos ou elementos são categorizados sobre uma dimensão dicotômica. De forma similar às técnicas de classificação de conceitos, as grades permitem criar um mapeamento conceitual do domínio.

Para adotar essa técnica, é necessário que o engenheiro já tenha identificado muitos dos principais componentes de conhecimento associados à tarefa (usando eliciação triádica, por exemplo). O especialista estabelece então uma escala bipolar de valores para esse construto (em geral, uma característica e seu oposto). Os conceitos restantes são cotados<sup>4</sup> em relação a esse construto nas características associadas a eles.

Por exemplo, no caso da seleção de uma linguagem de programação (exemplo tirado de (TURBAN, 1992)). O especialista seleciona as linguagens LISP, PROLOG, COBOL e C. A aplicação anterior da eliciação triádica indica que as dimensões em que essas linguagens são similares ou diferentes são: a *orientação*, *facilidade de treinamento*, *tempo de treinamento* e *disponibilidade*. O especialista estabelece uma escala bipolar para as dimensões, por exemplo, indica que uma linguagem pode apresentar orientação *simbólica* ou *numérica*. As duas primeiras linguagens são simbólicas, enquanto COBOL é numérica. LISP e PROLOG são cotados com nota 3 e COBOL é cotado com nota 1. A escala utilizada para dar notas aos conceitos é aplicada pelo especialista dependendo do objetivo da cotação. O construto Tempo de Treinamento, por exemplo, recebe nota 1 se o tempo for alto e nota 3 caso contrário. Isso quer dizer que uma linguagem que possa ser rapidamente dominada é mais importante para o objetivo final do sistema que utilizará as informações coletadas. O resultado é uma matriz que relaciona elementos e seus construtos através de medidas de similaridade. Um exemplo de construtos, de suas características e das cotações resultantes da eliciação de conhecimento para a tarefa de seleção de uma linguagem de computador é mostrado na Tabela 2.2.

Tabela 2.2: Resultado da técnica de grade de repertório (modificado de (TURBAN, 1992)).

Atributo	Orientação	Facilidade de Programação	Tempo de Treinamento	Disponibilidade
Característica Oposto	Simbólica (3) Numérica (1)	Alta (3) Baixa (2)	Alta (1) Baixa (3)	Alta (3) Baixa (1)
LISP	3	3	1	1
PROLOG	3	2	2	1
C	2	3	2	2
COBOL	1	2	1	3

Algumas ferramentas de aquisição de conhecimento (implementadas para uso em computador) foram desenvolvidas baseadas em grades de repertórios, tais como ETS, KRITON e AQUINAS, as quais são detalhadas no trabalho de Boose (BOOSE, 1989).

### 2.3.7 Técnicas de Observação

O método consiste em o especialista proceder com a solução de problema e o engenheiro observar. Esta técnica permite que o especialista trabalhe no ambiente em que está acostumado sem interrupções e proporcione ao engenheiro uma visão da complexidade do problema a ser resolvido (LIOU, 1990). A observação pode ser usada

<sup>4</sup> Em geral é utilizada uma escala de 1 a 3 ou de 1 a 5 para cotar os conceitos em relação a um construto.



para identificar estratégias de solução de problemas que não são conscientes, ou para estudar habilidades motoras ou procedimentos automáticos, para identificar as tarefas envolvidas na solução de um problema, e as limitações e restrições dessas tarefas, para identificar as informações necessárias para uma tarefa, e para validar a descrição que o especialista deu sobre o que ele realmente faz.

O engenheiro coleta informações através de um estudo visual de como a tarefa é executada pelo especialista. Segundo Johnson (JOHNSON, 1992), as técnicas de observação podem ser *diretas*, quando o engenheiro presencia a execução da tarefa, o que pode influenciar no comportamento do especialista, e *indiretas*, em que um vídeo ou fita é gravado e assistido posteriormente, o que é menos intrusivo, mas requer mais tempo para análise.

Uma desvantagem da observação é que a linha de raciocínio do especialista não é revelada pelas suas ações, e não pode ser acompanhada pelo observador. Além disso, técnicas de observação não podem ser usadas de modo isolado. Ou seja, ao contrário de outras técnicas, em que se pode focar o processo de aquisição em um aspecto específico do domínio, é impossível pedir para o especialista para que execute somente uma determinada parte da tarefa total para ser observada. Além disso, é mais difícil de se interpretar as informações de uma observação do que de um relatório verbal. Em geral uma ação depende dos passos anteriores, sendo difícil separá-la do resto. Por esse motivo, as técnicas de observação são caras e consomem muito tempo, tanto do engenheiro quanto do especialista e são usadas como uma maneira de proceder com a análise de protocolos.

### **2.3.8 Limitação de Informações (Focalizando Contextos)**

A proposta da limitação de informações (também conhecida por *focalizando contextos*) é apresentar um problema fictício para o especialista, de modo que ele descreva seu método de solução. Pode-se adotar a abordagem de restringir progressivamente a disponibilidade de informações, dessa maneira, o especialista demonstra qual o mínimo de informações necessárias para a tomada de decisão.

O especialista pode também ser explicitamente instruído a adotar uma estratégia particular. Um experimento bastante comum é mostrar diversas fotos sobre assuntos que o especialista costuma trabalhar, mas com uma restrição bem maior de tempo, e pedir que ele externalize tudo o que consegue lembrar sobre as fotos e dê sua interpretação. Esse experimento foi realizado por (HOFFMAN, 1987), com analistas de terrenos inspecionando fotos aéreas por somente 2 minutos, quando essa interpretação leva geralmente horas ou dias. (ABEL, 2001) analisou o desempenho de geólogos novatos e especialistas ao analisar uma série de imagens de rochas. Os resultados desse processo restringido no tempo revelam até que ponto os especialistas adquirem percepção imediata do conteúdo das imagens e conseguem elaborar uma interpretação independente do contexto restrito.

Uma técnica bastante conhecida que ilustra tanto a restrição de processos quanto a limitação de informações é chamada *vinte perguntas* (*20 questions*, do original (GROVER, 1983)). O especialista recebe o mínimo de informações sobre um problema, e precisa perguntar para o engenheiro de conhecimento as informações necessárias para solucioná-lo. A informação pedida junto à ordem em que ela foi pedida revela ao engenheiro uma parte da estratégia de solução de problemas do especialista.

### 2.3.9 Relatórios Verbais

Relatórios verbais podem ser de dois tipos: aqueles em que o comentário sobre o procedimento é feito ao mesmo tempo em que o especialista executa a tarefa (*online* ou concorrente) e aqueles em que o especialista é gravado em ação e posteriormente a tarefa é explicada verbalmente (*offline* ou retrospectivo). Os protocolos retrospectivos são mais adequados quando o engenheiro precisa fazer questionamentos, pois o especialista não será interrompido. Se forem feitas perguntas durante a execução, a interrupção pode alterar significativamente a seqüência de raciocínio seguida pelo agente.

Uma característica da verbalização durante o processo de solução foi reconhecida em diversos experimentos que utilizaram relatórios verbais: o especialista interrompe a verbalização da solução quando a solução torna-se não trivial, ou quando há necessidade de buscar caminhos alternativos de solução. As heurísticas utilizadas nesses momentos são importantes e devem ser posteriormente investigadas através de entrevistas.

Relatórios verbais são criticados por algumas limitações. O especialista não conseguirá verbalizar seu raciocínio se a própria tarefa demanda dele algum tipo de comunicação ou alto processamento cognitivo. Um piloto de avião ou um motorista de carro apresentam seu melhor desempenho quando a tarefa é executada de maneira automática. Caso precisem tomar consciência de suas ações, a tarefa é executada da pior maneira.

Em segundo lugar, a própria definição de conhecimento especializado é que é um conhecimento compilado, procedimentalizado, ou empacotado de uma maneira particular para o especialista, e conseqüentemente, esses pacotes de informações agregadas não são verbalizáveis. Além disso, mesmo quando o especialista consegue verbalizar o processo, os especialistas falham em descrever aquelas ações óbvias ou rotineiras, pois quando uma ação é repetida muitas vezes ela se torna parte do conhecimento implícito/automático, ou seja, ele não toma mais consciência do que ele faz e se uma ação não é consciente, ela não será verbalizada. Outro problema é a própria habilidade do especialista em se comunicar. Algumas pessoas conseguem pensar mais rápido do que falar, outras dão respostas totalmente diferentes dependendo do momento, e alguns ainda precisam de um aquecimento para falar bem.

### 2.3.10 Análise de Protocolos

A análise do material coletado pelas técnicas de rastreamento de processos é, talvez, a parte mais tediosa da tarefa. Os problemas existem por causa da quantidade de dados gerados, da natureza qualitativa dos dados, da complexidade e da desordenação dos dados e da subjetividade das interpretações. Dois engenheiros de conhecimento podem dar interpretações completamente diferentes do mesmo relato verbal. Algumas técnicas são utilizadas para analisar os protocolos da maneira mais fidedigna possível ao domínio e aos processos mentais originais do especialista (COOKE, 1994).

- Análise de conteúdo: uma maneira de organizar uma quantidade muito grande de material identificando objetiva e sistematicamente algumas características específicas. Pode-se classificar frases ou parte delas pelo tipo de conhecimento que evidenciam: conceitos, regras, procedimentos, etc. A vantagem em identificar classes de sentenças está em facilitar a tradução para diferentes estruturas de representação. A dificuldade dessa tarefa é determinar as categorias apropriadas.

- Análise de interação: uma maneira de analisar quase que gramaticalmente a interação entre o entrevistador e o especialista para identificar categorias de sentenças. Nesse caso, as categorias criadas são orientadas à conversação, por exemplo: negociação, orientação, explicação, entre outras. Permite que o engenheiro de conhecimento elimine ou evidencie categorias de respostas do especialista que ele considera mais importantes em determinados momentos da transcrição.
- Ferramentas de análise de protocolos: ferramentas que utilizam técnicas para analisar automaticamente os protocolos. O principal foco das ferramentas de análise de protocolos é a identificação e categorização das proposições e organização semântica dos segmentos do texto. Algumas ferramentas são desenvolvidas para um contexto específico, e seu desempenho fica próximo ao da análise manual. Outras ferramentas são livres de contexto, seu uso é generalizado, mas elas exigem uma interação maior com o engenheiro de conhecimento.

## 2.4 Ontologias

Ontologias são usadas em sistemas de agentes, sistemas de gestão de conhecimento, plataformas de *e-commerce*, entre outros. Ainda, segundo (GÓMEZ-PÉREZ e CORCHO, 2002), ontologias são úteis para aplicações em que o conhecimento desempenha um papel principal. Elas podem gerar linguagem natural, integrar informações, prover acesso semântico à Internet, extrair informações de textos além de serem usadas em muitas outras aplicações para declarar explicitamente o conhecimento nelas embutido, assim como sugere sua mais conhecida definição: “uma ontologia é uma especificação explícita e formal de uma conceitualização compartilhada” (STUDER, BENJAMINS e FENSEL, 1998). Esta definição é bastante próxima da idéia principal de sistemas baseados em conhecimento, que são sistemas de inteligência artificial que otimizam a busca da solução pela aplicação de conhecimento explicitamente representado. Existem muitas outras formas de representação de conhecimento estudadas na literatura, tais como redes semânticas e *frames*; mas os pesquisadores dessa área têm focado em ontologias para modelar e compartilhar conhecimento sobre conceitualizações de domínio, estratégias de raciocínio e métodos de solução de problemas (CHANDRASEKARAN, JOSEPHSON e BENJAMINS, 1999). Uma ontologia permite descrever os conceitos de um domínio e seus atributos, os relacionamentos entre conceitos, axiomas, entre outros construtos importantes para a definição de uma representação semântica.

O desenvolvimento de ontologias para aplicação na ciência da computação segue quatro abordagens básicas (GÓMEZ-PÉREZ, FERNÁNDEZ-LÓPEZ e CORCHO, 2004):

### 2.4.1 Ontologias para Representação de Conhecimento (RC)

São ontologias que agrupam as primitivas de modelagem, tais como *classes*, *relações* e *atributos*, usadas para formalizar conhecimento como um modelo de representação. A ontologia de RC mais representativa é a *Frame Ontology* (FO) (GRUBER, 1993), construída para capturar convenções de RC em uma abordagem baseada em *frames*. Com a criação do OKBC (CHAUDHRI et al., 1998), baseado na FO, foi possível acessar ontologias em diversas linguagens, tais como Ontolingua (FARQUHAR, FIKES e RICE, 1997), LOOM (MACGREGOR, 1991), CycL (LENAT

e GUHA, 1990). As primitivas de modelagem das ontologias desenvolvidas nessas linguagens são similares à FO. São primitivas para construir taxonomias de classes, de relações, definir propriedades de relações, restringir atributos e suas facetas (domínio de valores dos atributos). Mais recentemente, outras linguagens de marcação de ontologias foram criadas para o contexto da representação de conhecimento na web semântica, todas sendo uma extensão de RDF (BECKETT e MCBRIDE, 2004): RDF(S) (BRICKLEY e GUHA, 2000), OIL (HORROCKS et al., 2000), DAML-OIL (HORROCKS, HARMELEN e PATEL-SCHNEIDER, 2001) e OWL (DEAN e SCHREIBER, 2003). Todas essas linguagens possuem suas respectivas ontologias de RC.

#### 2.4.2 Meta-Ontologias (ou *Top-Level Ontologies*)

Suas primitivas são conceitos muito gerais do mundo, utilizados depois para construir ou integrar ontologias de domínios específicos. Buscam ser *universais*, no sentido de serem supradomínios, e *articuladas* ao apresentarem as primitivas necessárias e suficientes para modelagem. Algumas das principais meta-ontologias são as seguintes: a ontologia dos conceitos universais e particulares (GUARINO e WELTY, 2000), disponível em WebODE; a meta-ontologia de Sowa (SOWA, 2000), que inclui as distinções e categorias básicas derivadas de uma variedade de fontes (lógica, lingüística, filosofia e inteligência artificial). Ainda a meta-ontologia CYC (LENAT e GUHA, 1990) implementada em CycL e em DAML-OIL, que reúne uma quantidade imensa de conhecimento de senso comum, e as ontologias desenvolvidas pelo grupo *IEEE Standard Upper Ontology* (PEASE e NILES, 2002), com participantes do governo, academia e indústria de vários países, com o objetivo de criar uma grande ontologia formal de propósitos gerais.

#### 2.4.3 Ontologias Lingüísticas

As ontologias lingüísticas oferecem construtos semânticos para suporte ao processamento de linguagem natural. Suas primitivas são unidades gramaticais e dão ênfase às regras de construção das unidades de representação. São utilizadas para representar dicionários, léxicos, sistemas de tradução, entre outras aplicações lingüísticas. Em algumas ontologias existe um mapeamento um-para-um entre os conceitos e palavras em uma linguagem natural, enquanto em outras, alguns conceitos podem até não ser mapeados para nenhuma palavra ou para várias na mesma língua. Algumas ontologias lingüísticas são: WordNet (MILLER et al., 1990), um banco léxico para o inglês criado na Universidade de Princeton e baseado em teorias psicolingüísticas; EuroWordNet (VOSSSEN, 1998), baseado no WordNet para o inglês, mas com o objetivo de ser uma base multilingüe para várias linguagens Européias (alemão, checo, espanhol, estoniano, francês, holandês, italiano); GUM (BATERMAN et al., 1990), construída para definir as semânticas dos constituintes da gramática da linguagem, tais como, frases preposicionais, grupos nominais, etc.; Mikrokosmos (MAHESH e NIRENBURG, 1995) e SENSUS (SWARTOUT et al., 1997), que são meta-ontologias que provêm uma estrutura para conceitos lingüísticos gerais para mecanismos de tradução.

#### 2.4.4 Ontologias de Domínio

Ontologias de domínio são vocabulários reusáveis de conceitos e relacionamentos de um domínio específico, os quais incluem as atividades daquele domínio e suas teorias e

princípios fundamentais. São potencialmente reusáveis em diversas aplicações no mesmo domínio. As mais conhecidas se referem a *comércio eletrônico*, pois o contexto das aplicações de *e-commerce* demanda uma comunicação eficaz entre as máquinas, usando termos consensuais; *medicina*, desenvolvidas para resolver problemas como o reuso, transmissão e compartilhamento de dados de pacientes e a necessidade de critérios para estatísticas médicas; *engenharia*, que contém modelos matemáticos usados para analisar o comportamento de sistemas físicos; e *gestão de conhecimento*, para criar memórias corporativas que promovam o crescimento, a comunicação e a preservação do conhecimento da organização.

Neste trabalho são definidas duas ontologias: a ontologia de representação de conhecimento baseada em eventos (capítulo 6), define os construtos necessários para se representar eventos e relações temporais independentes de domínio, sendo portanto descrita como uma ontologia de representação de conhecimento. Já a ontologia proposta no capítulo 7 é um modelo baseado nos construtos ontologia de eventos, e que usa os construtos da ontologia de RC para definir as primitivas específicas de um determinado domínio, o da petrografia. Essa última ontologia podem ser classificada como *ontologia de domínio*.

## 2.5 Metodologias para Desenvolvimento de Sistemas Baseados em Conhecimento

O objetivo das metodologias de desenvolvimento de SBC é estabelecer um padrão para o projeto de software baseado em conhecimento desde a coleta de requisitos junto ao usuário e especialista, até a etapa de implementação do sistema. As estruturas de modelagem especificam o sistema baseados na teoria do *nível do conhecimento* (NEWELL, 1982; VELDE, 1993).

A metodologia MIKE, proposta por Angele e colegas (ANGELE, FENSEL e STUDER, 1996) suporta a aquisição e representação de conhecimento utilizando um formalismo que permite a subsequente execução. Já PROTÉGÉ (ERIKSSOM et al., 1995) explora a noção de ontologias para descrever o domínio. Atualmente, a metodologia PROTÉGÉ está implementada como uma ferramenta de suporte à construção da ontologia. VITAL (O'HARA et al., 1992) fornece suporte para todas as fases do desenvolvimento de SBC através de quatro produtos: a especificação dos requerimentos, o modelo conceitual, o modelo funcional, o modelo de projeto técnico e o código executável.

KADS (WIELINGA, SCHREIBER e BREUKER, 1992; WIELINGA et al., 1992) e seu sucessor, CommonKADS, (SCHREIBER et al., 2000) são as mais conhecidas destas metodologias. Destacaram-se por definir uma suíte de modelos que permite organizar componentes reusáveis para representar conhecimento do domínio, da tarefa e de inferência, além de métodos de solução de problemas e conhecimento estratégicos, nunca anteriormente descritos como componentes isolados. Essa metodologia foi aplicada no desenvolvimento do modelo do nível de conhecimento deste trabalho, e será detalhada a seguir.

### 2.5.1 A Suíte de Modelos CommonKADS para Representação de Conhecimento

CommonKADS (SCHREIBER et al., 2000) é uma metodologia que dá suporte ao desenvolvimento de sistemas baseados em conhecimento em todas as suas fases e aspectos, ou seja, desde a aquisição do conhecimento, passando pela modelagem e organização desse conhecimento até o projeto de implementação do sistema.

Um dos modelos, o *modelo de conhecimento*, propõe uma descrição independente de implementação do conhecimento envolvido em uma tarefa crítica da organização, ou seja, as especificações para as funções de raciocínio do sistema. Ele descreve como solucionar as tarefas intensivas em conhecimento da organização. A metodologia CommonKADS propõe representar três categorias de conhecimento: conhecimento de *tarefa*, que é orientado aos objetivos do sistema e indicam que informações ou conhecimento devem ser obtidos para alcançá-los; de *domínio*, que representa as informações e conhecimentos estáticos, declarativos sobre o domínio e o conhecimento de *inferência*, que descreve os passos de raciocínio básicos que podem ser aplicados sobre os objetos do domínio para se chegar aos objetivos da tarefa.

O conhecimento de domínio é representado por dois componentes: o *esquema do domínio*, que é uma descrição esquemática dos tipos de objetos, comparável a um modelo de dados, mas usando primitivas mais semânticas; e a *base de conhecimento*, que são as instâncias dos objetos especificados no esquema do domínio.

Para representar os construtos das três categorias de conhecimento, CommonKADS propõe uma sintaxe BNF, chamada *CML*<sup>5</sup> (*CommonKADS Conceptual Modeling Language*). Os construtos de CommonKADS podem ser representados tanto textualmente, pela linguagem CML, quanto graficamente, pela notação UML (FOWLER e SCOTT, 2000). O modelo proposto como contribuição deste trabalho foi definido utilizando tanto a sintaxe textual quanto a notação gráfica, como será demonstrado nos apêndices. Os construtos propostos em CommonKADS serão demonstrados a seguir.

#### 2.5.1.1 *Conhecimento de domínio - esquema de domínio*

As primitivas são similares às de um modelo de dados, mas com algumas primitivas mais semânticas: uma primitiva que permite representar inferência no esquema, o que não é possível em outros modelos, o *tipo-regra*, e possibilidade de subtipos de relações.

- **Conceito:** classe, tipo de objeto, mas sem funcionalidades
- **Relação:** associações entre conceitos com qualquer aridade (relações binárias têm uma representação especial). Uma relação pode apresentar atributos e subtipos, assim como os conceitos.
- **Atributo:** característica do conceito que o diferencia dos outros conceitos. Os valores dos atributos são delimitados pelo tipo-valor, e diferenciam as instâncias de um mesmo conceito. Os atributos de um conceito não fazem referência a outro conceito. Para isso usa-se a relação.
- **Tipo-regra:** modela dependências entre expressões sobre conceitos. Os tipos-regra suprem a demanda de representação de regras, necessárias em um modelo que precise fazer inferências sobre conhecimento simbólico.

#### 2.5.1.2 *Conhecimento de domínio - base de conhecimento*

Contém as *instâncias* dos tipos definidos no esquema: instâncias de conceitos, instâncias de relações e instâncias de tipos-regra. Ao contrário dos bancos de dados, as instâncias do conhecimento do domínio não são instâncias do usuário do sistema, mas instâncias do próprio domínio, como nome de peças, de componentes, regras de diagnóstico, etc. As instâncias de regras são definidas através de *expressões do domínio*,

---

<sup>5</sup> A sintaxe CML é definida, originalmente, em inglês. Para fins de coerência com o texto da dissertação, os termos CML foram traduzidos para o português.

que são combinações de instâncias de conceitos e relações através de operadores lógicos (E, OU, NÃO) ou aritméticos (<, >, =, +, etc.).

### 2.5.1.3 *Conhecimento de inferência*

O conhecimento de inferência descreve a decomposição funcional do conhecimento através de unidades de processamento básico: as *inferências*. Uma inferência é um passo de raciocínio que recebe uma unidade de informação e gera outra. Por isso ela é totalmente descrita no modelo através da especificação de suas entradas e saídas, enquanto o funcionamento interno é uma caixa-preta. As informações de entrada e saída são definidas como entidades que capturam o papel de um objeto de domínio no processo de raciocínio, os *papéis*<sup>6</sup> *de conhecimento*. Os papéis têm mapeamento direto para os construtos definidos no conhecimento de domínio, e podem ser dinâmicos (quando podem ser assumidos por diferentes objetos a cada passo de raciocínio) ou estáticos (quando são assumidos por somente um objeto de domínio invariante). O conhecimento de inferência pode ser representado graficamente, demonstrando o fluxo de conhecimento dos papéis para as inferências. Mas nesse diagrama não existe indicação do caminho percorrido, ou de laços de repetição. Esse controle é definido no conhecimento de tarefa.

### 2.5.1.4 *Conhecimento de tarefa*

Descreve, de uma maneira hierárquica, os métodos realizados para atingir os objetivos do sistema. Os métodos são definidos por estruturas de controle, que são sub-programas descritos em uma pseudolinguagem independente de linguagem de programação. A estrutura de controle também pode ser representada por um diagrama de atividades UML e indica como os componentes do conhecimento de inferência (papéis e inferências) devem ser empregados através de primitivas de controle (repetir-até, enquanto-faça, se-então-senão) e operações entre papéis (especificar, adicionar, anexar, subtrair, etc.) para seguir a estratégia de raciocínio. O conhecimento de tarefa é que define se uma inferência será repetida até que se chegue a uma resposta, ou enquanto não se esgotarem as possibilidades de resposta. Essa estrutura de controle caracteriza o objetivo da tarefa.

A diferença entre uma tarefa e uma inferência é a definição do comportamento. Se o comportamento interno de alguma função é importante para defini-la, então essa função deve ser definida como uma tarefa e não como uma inferência.

## 2.6 **Resumo do Capítulo**

Esse capítulo revisa as principais abordagens da Engenharia de Conhecimento para transformar o processo de construir sistemas baseados em conhecimento em uma disciplina baseada em métodos, linguagens e ferramentas. As principais abordagens da engenharia de conhecimento são o processo de aquisição de conhecimento, que é a eliciação e captura das informações e requisitos necessários para a definição do sistema e a modelagem desse conhecimento como uma ontologia de domínio. Para formalizar o processo de desenvolvimento são empregadas metodologias, como CommonKADS, que é aplicada neste trabalho. O foco de estudo dessa metodologia está na linguagem de definição do conhecimento de domínio, que define construtos para se representar uma

---

<sup>6</sup> Papel (do original, *role*) tem o sentido de função do momento, como o papel de um ator.

ontologia de domínio clássica, definindo objetos do domínio como conceitos e especificando relações binárias entre eles.



### 3 MODELOS PARA REPRESENTAÇÃO DE TEMPO

Noções como ação, causalidade, mudança, necessárias para diversas tarefas que demandam inteligência são descritas em termos de tempo. Além disso, toda a percepção humana do mundo incorpora o conceito de tempo, pois o mundo não é estático. Por isso, a representação de aspectos temporais é essencial para a solução de problemas típicos que envolvem tempo, tais como os descritos a seguir (MAIOCCHI e PERNICI, 1991; VILA, 1994):

- Planejar ações futuras e predizer conseqüências;
- Trabalhar com restrições de tempo de resposta;
- Especificar requisitos para sistemas de tempo real;
- Extrair informações temporais de textos;
- Inferir diagnóstico médico e explicação;
- Supervisionar de processos industriais;
- Monitorar mudanças no domínio;
- Tratar compreensão de linguagem natural;
- Raciocinar sobre o período de validade dos valores de um dado;
- Reconhecer contradições e dados incorretos;
- Lidar com dados temporais incompletos;
- Raciocinar sobre históricos de eventos no passado, para analisar a situação presente;
- *Raciocinar sobre o domínio atual para inferir eventos no passado.*

O último tópico, inferir eventos no passado, é um tipo de problema em que a configuração atual do domínio deve ser analisada para inferir sobre os eventos que ocorreram e que determinaram essas características atuais.

Podem ser citados alguns exemplos de domínios em que essa tarefa é realizada.

- O registro médico de pacientes para diagnóstico de doenças a partir dos sintomas. O médico precisa não só diagnosticar as causas de uma doença, mas tentar identificar o histórico da doença, como saber que determinado sintoma só apareceu no paciente depois de um acontecimento específico.
- Os estudos dos movimentos tectônicos da Terra, que conseguem inferir quais eventos ocorreram há milhões de anos e em que ordem eles aconteceram para definir a estrutura das placas tectônicas como elas são hoje.

- A técnica de estudar o clima do passado no gelo. As geleiras funcionam como um arquivo natural da história ambiental do planeta. O gelo conserva a composição química e isotópica da atmosfera em cada época. A partir da análise da composição química e isotópica, é possível indicar os processos que deixaram mudanças climáticas no meio ambiente. A ordem desses processos está relacionada com a profundidade em que os componentes foram encontrados.
- A tarefa de inferir os eventos geológicos que transformaram os sedimentos nas rochas sedimentares como elas são hoje. A ordem em que esses processos aconteceram é determinante para que os sedimentos se transformem em rochas com diferentes características.

Por esse motivo, a pesquisa envolvendo representação de tempo abrange muitas áreas de pesquisa na ciência da computação, principalmente na área de pesquisa de bancos de dados (BD), na qual a datação da informação é importante para representar as características dinâmicas das aplicações e a interação temporal entre processos e para identificar tendências e comportamentos na evolução da informação (EDELWEISS e OLIVEIRA, 1994; JANSEN, SCHREIBER e WIELINGA, 1998). Em geral, a modelagem de aspectos temporais em sistemas de informação abrange a modelagem de dados, a linguagem de recuperação e manipulação de dados e a implementação do sistema de gerenciamento de bancos de dados (SGBD). Entretanto, alguns dos conceitos envolvidos na especificação de aspectos temporais são específicos para SGBD, tais como o tempo de transação e o tempo de validade, que são usados para garantir a integridade dos dados. Em resumo, a ênfase está em *prover uma maneira de armazenar e recuperar dados com informações de tempo associadas*.

Outra área de pesquisa que utiliza a representação de aspectos temporais é a Inteligência Artificial. Sistemas baseados em conhecimento (SBC) costumam empregar representações de tempo com o objetivo de *prover ao sistema a capacidade de realizar raciocínio temporal sobre os dados*. A diferença entre a aplicação de aspectos temporais no desenvolvimento dos tipos de sistemas está nos objetivos: enquanto SGBD usam tempo para solucionar problemas como atualização de dados ou gerenciamento de históricos, SBC pretendem detectar mudanças no domínio e responder perguntas sobre os relacionamentos temporais dos eventos do domínio (Allen, 1983). Existem ainda os sistemas de processamento de linguagem natural (PLN) que utilizam representação de tempos verbais para definir representações o mais fiéis possível à linguagem falada e para poder compreender textos que usam noções de tempo.

Esses são os tipos de problemas mais representativos da necessidade atual em representar informações temporais. A seguir serão discutidos alguns aspectos básicos que devem ser compreendidos para se representar tempo, tais como a natureza dos aspectos temporais, suas características e os requisitos principais necessários para se representar tempo em sistemas.

### 3.1 Teorias de Tempo Absoluto e de Tempo Relativo

Sentenças temporais geralmente assumem duas formas: sentenças que usam relações de tempo como *antes* e *depois*, e sentenças que identificam o tempo em que um evento ocorre usando pontos ou intervalos de tempo. A visão do tempo como um conjunto de relações antes/depois entre eventos é chamada *tempo relativo*, em contraposição ao *tempo absoluto*, definido por proposições que usam primitivas tempo (BOLOUR e DEKEYSER, 1983).

A dicotomia de tempo absoluto/relativo é baseada em uma controvérsia filosófica (RESCHER e URQUHART, 1971; MCARTHUR, 1976). Na visão absolutista, o tempo é anterior aos eventos, é visto como um contêiner de eventos. Nessa visão, os relacionamentos temporais entre eventos são simplesmente derivados dos relacionamentos temporais entre os tempos. Na visão relativista, os eventos e as relações temporais entre eles têm mais prioridade que o tempo absoluto. Afinal, o tempo não é possível de ser observado isoladamente, ele é definido pelos eventos que ocorrem, mesmo que esses eventos sejam o movimento do ponteiro do relógio.

Usar um esquema de datação absoluta é uma maneira prática de representar eventos. Associar cada evento com um tempo absoluto, por exemplo, o *clock* do computador, o horário ou a data do evento, dependendo da granularidade que se deseja.

A grande vantagem é que os algoritmos de comparação são reduzidos a simples comparações numéricas, além de termos informação sobre a duração de tempo entre eventos – só é preciso subtrair as datas. Esse tipo de representação, no entanto, só é utilizável em aplicações cuja informação sobre o tempo é sempre conhecida e pode ser identificada e associada ao evento. Por exemplo, aplicações de coleta de dados em tempo real, bancos de dados de transações não distribuídos, ou um sistema para se manter um calendário de compromissos.

Mas e se o tempo *exato* de cada evento pode não ser conhecido? Se, em vez disso, conhece-se a faixa na qual sabe-se que o evento ocorre ainda é possível usar um esquema de datação modificado (pseudo-datação) onde o tempo de cada evento é restringido pelo tempo mais cedo possível e pelo tempo mais tarde possível. Esse esquema é explicado em mais detalhes por (Allen, 1991).

Ainda assim, os eventos precisam ser associados a uma primitiva que indique os tempos possíveis em que ele ocorreu, e o cálculo será sobre os tempos. Mas como representar quando os eventos não possuem nenhuma datação precisa ou aproximada associada? Para sistemas que se baseiam, por exemplo, em informações coletadas da linguagem natural, o aspecto temporal é na maior parte das vezes indicado por relações *antes* ou *depois*. O problema citado no início do capítulo, sobre o registro médico de pacientes, é um exemplo de tarefa em que ambas as abordagens são importantes. O tempo absoluto (dia, hora, minuto, duração) é utilizado para registrar observações feitas pelos médicos nas clínicas e hospitais. Mas quando o paciente relata ao médico o histórico de uma doença, em geral utiliza muitas sentenças relativas, pois não se pode esperar que ele lembre com exatidão as datas dos eventos que somente depois são considerados significantes pelo médico.

Para permitir a representação de aspectos temporais relativos, Bolour (BOLOUR e DEKEYSER, 1983) propõe reduzir o volume de informações temporais criando uma abstração do tempo absoluto. As sentenças temporais se concentram somente em como os objetos se relacionam temporalmente com outros objetos, permitindo o uso de sentenças mais simples para representar informação temporal.

Os objetos temporais são qualquer objeto do domínio sobre o qual se possa afirmar alguma idéia de tempo ou que tenham uma existência temporal: ações, eventos, estados, entre outros. Conforme Bolour e Dekeyser, é suficiente restringir a atenção aos relacionamentos *antes* e *durante*, e considerar as outras relações de tempo como sendo derivadas destas. Por exemplo, a relação *mesmo-tempo* é definida pela equivalência  $mesmo-tempo(x,y) \Leftrightarrow (durante(x,y) \text{ e } durante(y,x))$ . Os axiomas dessa teoria são definidos pelas seguintes fórmulas:

$$durante(x,x).$$

$$(durante(x,y) \text{ e } durante(y,z)) \Rightarrow durante(x,z).$$

$(\text{antes}(x,y) \text{ e } \text{antes}(y,z)) \Rightarrow \text{antes}(x,z).$

$\text{antes}(x,y) \Rightarrow \text{não antes}(y,x).$

$(\text{antes}(x,y) \text{ e } \text{durante}(z,y)) \Rightarrow \text{antes}(x,z).$

$(\text{durante}(x,y) \text{ e } \text{antes}(y,z)) \Rightarrow \text{antes}(x,z).$

As fórmulas acima compõem a lista de axiomas *invariantes* de tempo (que valem para qualquer domínio), a qual pode ser acrescida dos fatos específicos do domínio de aplicação. A interpretação dessa teoria deriva uma visão temporal do domínio modelado.

## 3.2 Classificação de Requisitos Temporais

Requisitos temporais são os aspectos de tempo que serão associados às entidades que demandam informação temporal (data, duração, condição, restrição, etc.). Conforme a classificação de (EDELWEISS e OLIVEIRA, 1994), existem dois tipos diferentes de requisitos temporais: requisitos representados por informações temporais *incondicionais* a respeito de eventos e requisitos *condicionais*.

### 3.2.1 Requisitos Temporais Incondicionais

São aqueles explicitamente definidos, representando um instante ou intervalo de tempo associado a uma informação.

- **Bem-definidos:** (i) um elemento temporal associado a um evento (data de um aniversário), (ii) o tempo de duração associado a um evento (tempo de duração de uma reunião), (iii) o tempo de validade de uma informação (tempo de validade de uma taxa de câmbio).
- **Parcialmente definidos:** (i) evento associado a um elemento temporal antes ou depois do qual ele deve ocorrer (entrega da declaração do imposto de renda que deve ser antes de determinada data), (ii) evento associado a um elemento temporal durante o qual ele deve ocorrer (matrícula de um aluno associada ao intervalo de tempo durante o qual ela pode ser efetuada).

### 3.2.2 Requisitos Temporais Condicionais

São aqueles expressos implicitamente por condições.

- **Controlados por restrições causais na execução de processos.** Por exemplo, o início de um processo só pode ocorrer com a ocorrência de um determinado evento.
- **Controlados por restrições temporais a valores.** Por exemplo, um novo valor de salário nunca deve ser inferior a um valor anteriormente definido.

## 3.3 Questões para Representação Temporal

Ao escolher uma forma de representação temporal, existem diversos tipos de questões sobre a representação de tempo que devem ser definidas para desenvolver a ontologia temporal:

### 3.3.1 A escolha de uma entidade primitiva de tempo.

Quando a definição do tempo é feita *explicitamente*, ou seja, através da associação de uma primitiva de tempo a uma informação (processo, evento, fato, ação, etc.) é necessário definir um elemento temporal primitivo, como o *instante* (ponto no tempo) ou o *intervalo* (par de pontos).

Alguns sistemas utilizam referências explícitas a *instantes de tempo* e a relações de ordem temporal entre esses instantes. Um instante é qualquer momento bem definido, como uma hora, ou dia específico (dependendo da granularidade do modelo). Instantes são, intuitivamente, como pontos de tempo, no sentido de que eles não apresentam pontos interiores. Alguns modelos podem definir um instante como tendo início e fim em si mesmos.

Outros sistemas fazem referências a *intervalos de tempo* (que pode ser visto como um par de pontos no tempo) e às relações temporais entre os intervalos (o que implicitamente restringe a ordenação dos pontos finais dos intervalos). Esses sistemas em geral precisam representar a duração das ações, que pode ser calculada pelos pontos de início e de fim de cada intervalo.

Os intervalos são definidos como pares de pontos de tempo. Os pontos são definidos como o limite inicial e final de um intervalo e identificam os lugares onde os intervalos se encontram. As relações de ordem sobre intervalos são expressas em termos das relações entre os limites. Ou seja, para  $i1 = [i1-, i1+]$  e  $i2 = [i2-, i2+]$ , temos (MONTANARI e PERNICI, 1993):

$$i1 \text{ antes } i2 \Leftrightarrow i1^+ < i2^-$$

A escolha entre ponto ou intervalo implica diferentes ontologias e capacidades de raciocínio. Independente da escolha, tanto a utilização de pontos quanto de intervalos implica que devem estar definidas no domínio as instâncias das informações de tempo sobre a entidade a ser representada (fatos, eventos ou ações). Se para o domínio não é significativo representar o momento em que esses eventos aconteceram ou o período de tempo durante o qual eles ocorreram, então não se deve decidir por representar essas primitivas de tempo.

Para o caso de o domínio não prover informações explícitas de tempo, existe uma segunda forma de representação, denominada representação *implícita*, que é definida pelo uso de uma linguagem de lógica temporal. A representação implícita é feita através da manipulação de conhecimento sobre a ocorrência de eventos ou do relacionamento entre intervalos de tempo. A mais importante contribuição desse enfoque é a possibilidade de tratar informações imprecisas, tais como *antes* e *depois*. Algumas dos formalismos que permitem representação implícita do tempo serão detalhados mais adiante nesse capítulo.

### 3.3.2 Escolha da ordem de tempo.

A escolha mais comum é que o tempo transcorra *linearmente*, ou seja, deve existir uma ordem total entre dois pontos no tempo, ou seja, somente uma das seguintes expressões pode ser verdadeira:  $t < t'$  ou  $t' < t$ , onde  $t$  e  $t'$  representam pontos distintos no tempo e a ordem de precedência é representada através do operador  $<$ . Se a primitiva de tempo for o intervalo, representado por  $[t1, t2]$  (intervalo fechado) então uma das duas fórmulas deve ser verdadeira:  $t1 < t2$  ou  $t1 = t2$ . Alguns sistemas, entretanto, precisam representar a característica do seu domínio de eventos que ocorrem não linearmente, ou seja, concomitantemente, gerando uma ordem parcial entre os pontos de

tempo (*linha de tempo ramificada*). Ainda pode ser necessário representar eventos recorrentes, através de uma *linha de tempo circular*.

### 3.3.3 Especificação da estrutura do tempo.

Podem ser consideradas duas formas diferentes: tempo contínuo (que representa mais fielmente a natureza) e tempo discreto (que simplifica grandemente a implementação). Assim, a linha de tempo pode ser mapeada para o conjunto de números racionais  $Q$ , números reais  $R$ , ou números inteiros  $Z$ , dependendo da variação.

### 3.3.4 Limitações do tempo.

No caso de teorias baseadas em pontos de tempo, a limitação é em relação a permitir que o tempo seja infinito, ou seja, todo ponto tem um predecessor/sucessor. No caso de teorias de intervalo de tempo, a limitação se refere a especificar se os intervalos são abertos/fechados em uma ou nas duas pontas.

### 3.3.5 Introdução de uma métrica de tempo.

A métrica de tempo inclui, por exemplo, a definição de uma duração temporal, de uma distância de tempo entre os pares de elementos de tempo, gerenciar diferentes granularidades de tempo e a mapear os pontos de tempo para um referencial externo, o qual pode ser extrínseco (referenciam sistemas de tempo convencionais, tais como calendários ou períodos astronômicos) ou intrínseco (sistema de tempo relativo). A duração temporal pode ser basicamente de dois tipos: *fixas*, que independe do contexto de sua definição (p. ex.: uma hora tem sempre 60 minutos) ou *variável*, que depende do contexto (p. ex.: a duração de um mês).

Segundo (EDELWEISS e OLIVEIRA, 1994) o conceito de *instante de tempo* depende da estrutura temporal, ou seja, da forma como o tempo varia. Quando é considerado tempo contínuo, um instante é um ponto de tempo de duração infinitesimal, e a linha de tempo pode ser mapeada para os números reais. Quando a variação é discreta, um instante é representado por uma unidade de tempo definida na linha de tempo suportada pelo modelo, ou *chronon*<sup>7</sup>. Além disso, quando a ordem de tempo é linear, é preciso definir um instante especial denominado *instante atual (now)*, que define o que é considerado passado e futuro e que se move constantemente no eixo de tempo.

O *intervalo temporal*, que é o tempo ocorrido entre dois instantes, depende também da estrutura definida para o tempo. Quando o tempo é contínuo, o intervalo consiste de infinitos instantes de tempo, quando é discreto o intervalo consiste de um conjunto finito de *chronons* consecutivos. Quando são considerados intervalos, a ordem de tempo é linear.

A especificação de cada um dos cinco quesitos anteriores adequadamente define a *ontologia* que representa informações temporais específicas para o domínio de aplicação.

---

<sup>7</sup> Um *chronon* é a menor duração de tempo suportada por um modelo temporal, uma unidade de tempo não decomponível (TANSEL et al., 1993).

### 3.3.6 Raciocínio sobre o tempo.

Em aplicações que utilizam modelos com datação precisa, ou em que todas as relações são homogêneas<sup>8</sup>, ou até mesmo modelos de pseudo-datação permitem um mecanismo de raciocínio incorporado diretamente no mecanismo de recuperação de dados. A única demanda do mecanismo é aplicar as restrições temporais sobre os dados recuperados.

Em SBC, o melhor é poder integrar o raciocinador temporal ao mecanismo de inferência disponível no sistema. Para isso, depois de formalizar as noções de tempo como uma ontologia, adicionalmente, deve-se modelar um método de solução de problemas que represente de maneira abstrata os principais passos de inferência sobre as entidades da ontologia. Esse método deve fazer deduções sobre as asserções descritas na ontologia temporal, suportando a geração de informações adicionais sobre aquelas representadas. O fato de modelar um método de solução de problemas sobre aspectos temporais sobre a ontologia de domínio permite que essa inferência seja integrada às outras inferências do SBC.

## 3.4 Principais Linguagens de Representação Temporal

Esta seção apresenta um levantamento das principais e mais conhecidas linguagens para se representar informações temporais. A literatura apresenta uma lista extensa de linguagens, lógicas, cálculos, representações, entre outros formalismos para se modelar a ontologia temporal de um domínio. Cada formalismo é voltado para a representação de um tipo de aspecto temporal, como por exemplo, linguagens que permitem representar o tempo de um modo absoluto ou relativo, linguagens que representam duração de tempo ou pontos de tempo, etc. O livro (EDELWEISS e OLIVEIRA, 1994) apresenta um levantamento das representantes clássicas de modelos de datação absoluta, que são os *modelos temporais para bancos de dados*, tais como o modelo de dados TEER (*Temporal EER*) (ELMASRI, WUU e KOURAMAJIAN, 1993), ou o modelo HRDM (*Historical Relational Data Model*) (CLIFFORD e CROKER, 1993).

No entanto, a fim de direcionar a pesquisa para os modelos relevantes para o domínio de aplicação deste trabalho (domínios em que os eventos não apresentam datação absoluta) decidiu-se por fazer um levantamento dos principais formalismos de representação temporal que utilizam a abordagem relativa de tempo.

São introduzidas, inicialmente, as *lógicas de primeira ordem com argumentos temporais* e as *lógicas modais temporais*. Contrariando o direcionamento dessa seção, o argumento temporal adicionado às lógicas de primeira ordem são informações absolutas. Mas essas lógicas são apresentadas somente com o objetivo de apresentar a dificuldade de se modelar tempo, argumentando sobre a vantagem dos outros formalismos. Mesmo assim, segundo (SHOHAM, 1985), seria necessário que todas as propostas de modelos temporais incluíssem tanto uma sintaxe quanto semântica formalizada por uma lógica. Isso ajudaria a clarificar em que a proposta consiste, como ela se comporta com diferentes problemas e compará-la com outras abordagens.

Os outros formalismos apresentados nessa seção são: *cálculo de eventos*, que modela e raciocina sobre eventos e relacionamentos entre intervalos de tempo; o *cálculo de situações*, que propõe um modelo do universo em que suas propriedades só mudam depois da ocorrência de ações; e por fim, a *álgebra de intervalos de Allen*, que permite raciocinar sobre intervalos de tempo nos quais se mantém alguma propriedade.

---

<sup>8</sup> Relações homogêneas: noção explicada mais adiante, na seção 3.4.1, sobre lógicas.

### 3.4.1 Lógicas de primeira-ordem com argumentos temporais (AT)

Esse método, chamado de Argumentos Temporais (AT) (HAUGH, 1987), consiste simplesmente em utilizar as lógicas de primeira ordem e capturar a dimensão temporal aumentando o predicado lógico com outro parâmetro, a ser preenchido com uma expressão que denota um aspecto temporal. Por exemplo, pode-se representar que “João estava feliz em 2 de Maio” por:

`feliz(João, 2.mai)`

A informação temporal é um argumento como os outros, não diferenciado. O terceiro argumento nesse predicado pode ser tanto um ponto de tempo, quanto um intervalo ou duração de tempo. Se for um intervalo ou duração, o predicado se mantém para todo o intervalo ou por toda a duração. Antes da chegada das lógicas modais temporais, o método do argumento temporal era a escolha natural de formalismo para expressar logicamente uma informação temporal.

Essa abordagem é com certeza mais simples de ser seguida, mas perde em expressividade. O problema de não dar um status diferente ao aspecto temporal é que nada impede alguém de usar a fórmula *feliz(João, Maria)*, e dessa forma a semântica é comprometida. Além disso, essa abordagem não vale para todos os predicados. Considere o predicado que diz que algum evento ocorreu, independente quando, por exemplo, “o sol se pôs”. Pode-se dizer que o sol se pôs durante um intervalo de tempo, o que não é a mesma coisa que afirmar que o sol se pôs em todos os pontos de tempo desse intervalo. Na verdade, o sol não se põe em um ponto de tempo.

Allen, em (VENDLER, 1957) apontou distinções entre esses tipos de predicados analisando os verbos da Língua Inglesa. Predicados como *feliz* são homogêneos – se o predicado se mantém durante um intervalo ele se mantém durante todos os subintervalos e pontos naquele intervalo. Predicados como *se-pôr* são não-homogêneos, pois tem a propriedade oposta: se ele se mantém durante um intervalo T, então ele não se mantém por nenhum subintervalo de T (são exclusivos). Processos, como o predicado *correr*, tem a propriedade que, se eles se mantém por dois intervalos T1 e T2, e T1 encontra T2, então ele se mantém pela união de T1 e T2. Essa discussão é a base da teoria formalizada por Allen em (Allen, 1983), que será apresentada adiante neste capítulo.

Para tentar representar alguns desses predicados, Galton (GALTON, 2003) sugere que se for introduzido um predicado binário “<” que denota a relação “mais cedo que” e a constante “agora” que denota o momento presente, então se pode simular a noção de tempo. A correspondência entre a informação de tempo e as fórmulas, podem ser montadas com o predicado “<” a constante *agora*, como no exemplo abaixo, que simula a noção de tempo usando predicados de lógica de primeira ordem. Nas fórmulas,  $p(t)$  um predicado p adicionado do argumento temporal.

passado:  $t < agora \wedge p(t)$

futuro:  $agora < t \wedge p(t)$

Dessa forma, pode-se representar “o sol se pôs” como  $t < agora \hat{U} se-pôr(sol, t)$ . Entretanto, os predicados montados usando a extensão temporal tornam a leitura não natural, de difícil compreensão. Para transpor essas desvantagens é que são estudadas lógicas em que os objetos temporais são distinguidos dos objetos não temporais, como a apresentada a seguir.



### 3.4.2 Lógicas modais temporais (LMT)

Nas lógicas proposicionais ou de predicados as fórmulas ou são verdadeiras ou são falsas, sem outras possibilidades. Mas isso pode ser não suficiente para representar o mundo. Como fazer para representar uma afirmação que não é verdadeira no momento presente, mas vai se tornar verdadeira no futuro? Ou a noção de que alguém acredita que algo é verdadeiro? Essas demandas confirmam a necessidade de se poder raciocinar sobre diferentes *modos da verdade*. Para isso emprega-se a *lógica modal*.

A lógica modal básica estuda o raciocínio que envolve a veracidade ou falsidade de uma fórmula em diferentes mundos. O termo lógica modal é utilizado em geral para cobrir uma família de lógicas em que é possível expressar diferentes modos de verdade, que provêm uma maior expressividade em relação à lógica clássica. Essas lógicas foram adicionadas de conectivos unários que permitem lidar com noções tais como necessidade, conhecimento ou tempo, o conectivo *box* ( $\Box$ ) e o conectivo *diamond* ( $\Diamond$ ). Esses conectivos são aplicados às fórmulas lógicas e sua interpretação muda dependendo da família de lógica modal em que se está trabalhando. Algumas das diferentes interpretações de lógicas modais são descritas na Tabela 3.1.

Tabela 3.1: Algumas noções que podem ser representadas na lógica modal e os significados dos operadores aplicados a uma fórmula F (modificado de (HUTH e RYAN, 2000)).

Noção	$\Box F$	$\Diamond F$
<b>Necessidade</b>	É necessariamente verdade que F	É possivelmente verdade que F
<b>Dever</b>	É obrigatório que F	É permitido que F
<b>Tempo</b>	Sempre será verdade que F	Em algum tempo no futuro será verdade que F
<b>Crença</b>	Agente Q acredita que F	F é consistente com a crença de Q

A abordagem da lógica modal temporal (LMT) é envolver operadores de modo de tempo de verbo. Em vez de apresentar um tempo explícito, são introduzidos operadores como o *box* ou o *diamond*, que expressam tempo FUTURO. Por exemplo,  $\Diamond(\text{recebe}(\text{João}, \text{carta}))$  significa que haverá um tempo no futuro em que João vai receber uma carta, mas não especifica o momento. Facilmente podem ser adicionar operadores modais de passado, e então o exemplo da seção anterior poderia ser representado facilmente por *PASSADO*(*se-Pôr*(*sol*)). Por esse motivo, a LMT é uma abordagem relativista de tempo, em contraste com a abordagem absolutista proposta pela lógica AT. Na proposta original das LMT não está prevista a representação de pontos ou intervalos de tempo.

LMTs são bastante úteis em certas aplicações, como para representações de linguagem natural, mas, segundo (Allen, 1991), pelo fato de não representarem explicitamente nenhum tipo de aspecto temporal, elas não são suficientes para representar eventos, mudança do mundo, ou relacionamentos temporais entre processos.

### 3.4.3 Cálculo de eventos (KOWALSKI e SERGOT, 1986)

O cálculo de eventos foi introduzido por Kowalski e Sergot como um formalismo baseado em lógica para representar eventos e seus efeitos, que são propriedades. Esse cálculo foi inicialmente desenvolvido baseado em um cálculo de predicados de primeira ordem, e é capaz de representar ações com efeitos indiretos, ações com efeitos não determinísticos, ações compostas, ações concorrentes e mudança contínua.

A descrição de um evento é usada para deduzir a existência de um período de tempo, assim como para deduzir o início e o fim desse período. Cada período é unicamente identificado por uma combinação da relação  $r$ , que se mantém durante o período, e do evento  $e$  que inicia o período. Por exemplo,  $depois(e,r)$  define o período de tempo iniciado pelo evento  $E$  em que a relação  $r$  se mantém. O átomo  $mantém(antes(e, r))$  é uma versão reduzida de  $mantém(r, depois(e,r))$ , e diz que a relação  $r$  se mantém no período que antecede o evento  $E$ . Os axiomas para deduzir inícios e fins de intervalos de tempo são como o seguinte:

$mantém(antes(e,r))$  se  $termina(e,r)$  (EC1)

$mantém(depois(e,r))$  se  $inicia(e,r)$  (EC2)

$Termina(e,r)$  significa que o evento  $e$  termina a relação  $r$ .  $Inicia(e,r)$  significa que o evento  $e$  inicia a relação  $r$ . Os predicados  $termina$  e  $inicia$  precisam ser definidos por regras específicas, porque eles dependem das características particulares do domínio de aplicação.

Por exemplo, seja  $e1$  o evento em que João deu um livro para Maria e  $e2$  o evento em que Maria deu o livro para José. Assumindo que  $e2$  aconteceu depois de  $e1$ , pode-se deduzir que existe um período de tempo iniciado pelo evento  $e1$  em que Maria possui o livro e um período terminado por  $e1$  em que João possui o livro. Uma representação dessas afirmações é mostrada na Figura 3.1.

É preciso representar como os eventos iniciam ou terminam a relação  $possui(x,y)$ . Essa representação é baseada em predicados específicos do domínio, tais como  $receptor(e,x)$  e  $doador(e,x)$ .

$inicia(e, possui(x, y))$  se  
ato( $e, dar$ ) e receptor( $e, x$ ) e objeto( $e, y$ )

$termina(e, possui(x, y))$  se  
ato( $e, dar$ ) e doador( $e, x$ ) e objeto( $e, y$ )

Depois de se deduzir os períodos de tempo, é necessário deduzir seu início e fim, através do átomo  $inicia(x,y)$ , que significa que o início do período de tempo  $x$ , é o evento  $y$ , e, similarmente,  $fim(x,y)$ , que significa que o fim do período  $x$  é indicado pelo evento  $y$ . Os axiomas para essa dedução são como segue:

$início(depois(e,r), e)$  (EC3)

$fim(antes(e,r), e)$  (EC4)

$início(antes(e',r), e)$  se  $depois(e,r) = antes(e',r)$  (EC5)

$fim(depois(e,r), e')$  se  $depois(e,r) = antes(e',r)$  (EC6)

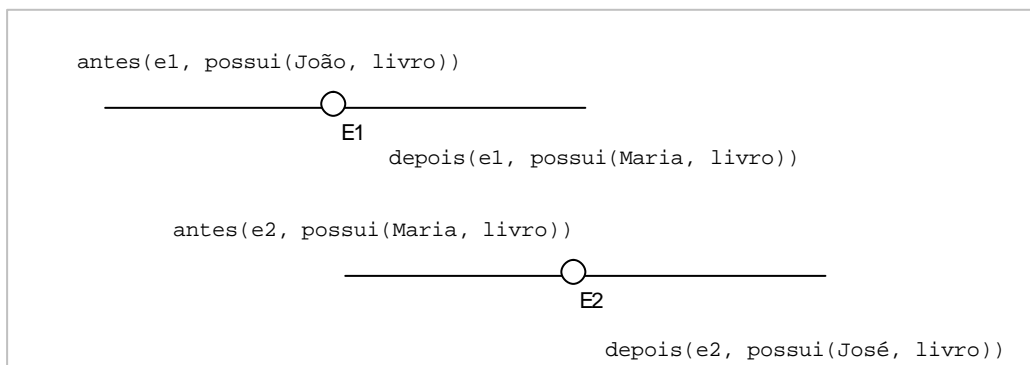


Figura 3.1: Representação das propriedades mantidas antes e depois dos eventos  $e1$  e  $e2$  (modificado de (SRIPADA, 1988)).

Esses axiomas permitem deduzir os intervalos de tempo, seus inícios e fins. Através da aplicação do axioma EC5, pode-se inferir que, se *depois*(*e1*, *possui*(*Maria*, *livro*)) e *antes*(*e2*, *possui*(*José*, *livro*)) são o mesmo período de tempo, então o período *antes*(*e2*, *possui*(*José*, *livro*)) é iniciado por *e1*. Entretanto, para deduzir que dois períodos de tempo são idênticos, deve-se ter certeza de que não existe nenhum evento no meio que possa iniciar ou terminar outro período de tempo, ou seja, tem que se verificar se uma propriedade se manteve durante todo o período. A regra *igual*(*e1*,*e2*) é apresentada abaixo:

$$\begin{aligned} & \text{igual}(\text{depois}(e1,r), \text{antes}(e2,r)) \text{ se} \\ & \text{mantém}(\text{depois}(e1,r)) \text{ e } \text{mantém}(\text{antes}(e2,r)) \text{ e} \quad (\text{EC7}) \\ & \text{precede}(e1,e2) \text{ e } \text{não}(\text{descontínuo}^9(e1,u,e2)) \end{aligned}$$

Para garantir que uma propriedade se manteve, utilizam-se o predicado *descontínuo*(*e1*,*u*,*e2*), que não será detalhado aqui, mas que verifica se algumas condições do domínio não estão sendo violadas. No exemplo acima, duas propriedades incompatíveis são duas pessoas possuírem um mesmo livro no mesmo período. Mas essas regras devem ser definidas em função das características específicas do domínio. Mais detalhes sobre essa regra são dados em (SRIPADA, 1988).

O predicado *precede*(*e1*,*e2*) significa que *e1* e *e2* são datados, e que a data de *e1* é menor que a data de *e2*. Isso seria uma desvantagem para modelos que não coletam datas absolutas sobre seus eventos. Mas, segundo (BERNARD, BORILLO e GAUME, 1990), essa característica pode ser facilmente mudada para que o modelo não utilize uma escala de tempo explícita.

O operador *não* é interpretado como *negação por falha*. Quando o predicado sobre o qual o operador de negação está sendo aplicado não se unifica com nenhum dos outros predicados, ele é considerado como tendo falhado. Essa é a situação em que se considera a negação do predicado.

O cálculo de eventos é uma linguagem de manipulação de intervalos de tempo, mas também utiliza a abordagem de tempo relativo. Não existe uma primitiva para representar os intervalos de tempo e suas relações. Todas as informações sobre intervalos de tempo estão implícitas na representação dos eventos e suas propriedades.

Um uso interessante do cálculo de eventos é armazenar todos os eventos relevantes para o domínio associados às propriedades que se mantêm antes e depois do evento. Os axiomas são utilizados para deduzir os relacionamentos entre os períodos de tempo em que eles se mantêm, deduzindo, indiretamente, o relacionamento entre os eventos.

### 3.4.4 Cálculo de situações (MCCARTHY e HAYES, 1969)

O cálculo de situações, baseado em lógicas de primeira ordem, foi definido por McCarthy e Hayes e assume que o universo é estável a menos que ações aconteçam para alterá-lo. O conhecimento é representado por uma série de situações, em que uma *situação* é a descrição de um estado completo do Universo em um determinado *instante de tempo*. Como é impossível uma descrição completa do Universo, descrevem-se somente fatos sobre as situações. Esses fatos são usados para deduzir outros fatos sobre a situação presente, sobre situações futuras e sobre situações que podem ser trazidas de outras situações. Ações e eventos são funções de uma situação para outra, modificando o estado do Universo.

---

<sup>9</sup> *Descontínuo*(*e1*,*r*,*e2*) significa a relação *r* não se mantém continuamente entre os eventos *e1* e *e2*. Tradução livre para o original, *broken*.

Para poder afirmar informações parciais sobre situações, usa-se a noção de *fluentes*. Fluentes são funções que descrevem fatos sobre situações. Por exemplo,  $chovendo(x)$  é um fluente tal que  $chovendo(x)(s)$  é verdade se e somente se está chovendo no lugar  $x$  na situação  $s$ .

A situação  $s$  em que a pessoa  $p$  está no lugar  $x$  e que está chovendo no lugar  $x$ , pode ser representada da seguinte forma:

$$em(p, x)(s) \wedge chovendo(x)(s)$$

A noção da causalidade é introduzida pelo operador  $F(p)(s)$ , que afirma que a situação  $s$  será seguida (após um período de tempo não especificado) por uma situação que satisfaz o fluente  $p$ . Por exemplo, para representar que se uma pessoa está em um lugar onde está chovendo ela vai se molhar escreve-se:

$$\forall x (\forall p (\forall s (em(p, x)(s) \wedge chovendo(x)(s)) \rightarrow F(molhada(p))(s)))$$

Esse tipo de construção é útil para representar leis da física e suas conseqüências. Baseado no operador  $F$ , são introduzidos mais três operadores modais, que aumentam a expressividade das sentenças no cálculo (Tabela 3.2):

Tabela 3.2: Alguns dos operadores do cálculo de situações.

Operador	Significado
$F(p,s)$	Para alguma situação $s'$ no futuro de $s$ , $p(s')$ se mantém.
$G(p,s)$	Para todas as situações $s'$ no futuro de $s$ , $p(s')$ se mantém.
$P(p,s)$	Para alguma situação $s'$ no passado de $s$ , $p(s')$ se mantém.
$H(p,s)$	Para todas situações $s'$ no passado de $s$ , $p(s')$ se mantém.

A noção de uma situação ser resultado de uma determinada ação é representada utilizando o fluente situacional  $resulta(? , s, s)$ , onde  $?$  é uma pessoa,  $s$  é uma ação,  $s$  é uma situação e o valor de  $resulta(? , s, s)$  é a situação que resulta quando a  $p$  conduz  $a$  começando na situação  $s$ . Com esse fluente, é possível expressar algumas leis, tais como, se uma pessoa  $p$  possui uma chave  $c$  que cabe na fechadura  $f$ , então a situação resultante de ela abrir a fechadura com a chave  $abrir(f, c)$  é que a fechadura estará aberta ( $aberta(f)$ ):

$$tem(p, c)(s) \wedge cabe(c, f) \wedge em(p, f)(s) \rightarrow \\ aberta(f, resulta(p, abre(f, c), s))$$

Pode-se notar que o cálculo de situações também lida com a noção de tempo de uma maneira relativista. O aspecto temporal é introduzido da mesma maneira como nas lógicas modais temporais: um operador unário que modifica a interpretação de uma fórmula para situá-la em algum momento relativo ao presente. Um aspecto dinâmico do domínio (indiretamente temporal) é indicado pelas modificações que os eventos fazem ao universo, que são resultados das ações.

Esta teoria é viável somente em domínios onde ocorre um evento por vez, e em que esse evento não possui duração. O cálculo de situações é útil para deduzir o novo estado do mundo quando determinados eventos ocorrerem. Entretanto, o cálculo não suporta inferências sobre os eventos em si. A transformação entre as situações não pode ser decomposta, pois o evento não ocupa tempo e nada se pode dizer sobre ele. O objetivo do cálculo é inferir sobre as características do mundo, mas não sobre as relações temporais entre elas. O trabalho de (MCCARTHY, 1999) apresenta um exemplo típico para o qual o cálculo de situações é útil: a modelagem do problema dos missionários que precisam levar os canibais de um lado a outro do rio, e que, caso o número de

canibais supere o de missionários, os últimos serão devorados. Alguns exemplos de proposições dessa modelagem são apresentados a seguir:

$$\begin{aligned} & \text{ruim(lado-do-rio)(s)} \Leftrightarrow \\ & 0 < \text{quant}(\{ x \mid x \in \text{Missionarios} \wedge \text{em}(x, \text{lado-do-rio})(s) \}) \\ & < \text{quant}(\{ x \mid x \in \text{Canibais} \wedge \text{em}(x, \text{lado-do-rio})(s) \}) \end{aligned}$$

Informa que a situação em que a quantidade de missionários é menor que a quantidade de canibais é ruim.

$$\begin{aligned} & \text{ruim(lado-do-rio)(s)} \rightarrow \\ & (\forall x)(x \in \text{Missionarios} \wedge \text{em}(x, \text{lado-do-rio})(s) \rightarrow \\ & \text{ocorre(comido}(x), s)) \end{aligned}$$

Informa que quando a situação de um lado do rio é ruim, todos os missionários que estiverem de um lado do rio serão comidos.

A vantagem do cálculo de situações é permitir acrescentar novos eventos que, se acontecerem, mudam a situação para outra configuração, logo, é possível criar diversas simulações. O exemplo que o trabalho cria é o cálculo da situação quando um dos missionários é Jesus.

### 3.4.5 Álgebra de intervalos de Allen (Allen, 1983)

A álgebra de intervalos de Allen permite representar eventos associados à noção de intervalos de tempo como primitiva temporal. O relacionamento entre dois intervalos de tempo é modelado através de relações temporais, descritas por um conjunto de 13 operadores: *antes*, *encontra*, *sobreposição*, *começa*, *durante* e *termina*, junto com seus inversos e o operador *igual* (Tabela 3.3). Por exemplo, o inverso da relação *durante* é *inclui*. Usando esses 13 operadores é possível definir relações temporais sem ambigüidade entre intervalos de tempo, pois as relações são mutuamente exclusivas.

Tabela 3.3: As treze possíveis relações temporais entre intervalos (modificado de (Allen, 1983)).

Relação	Símbolo	Símbolo para inverso	Exemplo figurativo
X antes Y	<	>	XXX YYY
X encontra Y	e	ei	XXXXYY
X sobreposição Y	s	si	XXX YYY
X durante Y	d	di	XXX YYYYY
X inicia Y	i	ii	XXX YYYYY
X termina Y	t	ti	XXX YYYYY
X igual Y	=	=	XXX YYY

A partir dessa tabela de relações temporais, Allen propõe uma tabela de transitividade para as relações. Para três eventos A, B e C, em que (A r1 B) e (C r2 B), essa tabela identifica qual a relação resultante entre A e C para cada possível combinação de relações temporais que r1 e r2 podem assumir. A tabela original mostra a combinação entre as 12 relações (omitindo “=”) A Tabela 3.4 mostra algumas delas.. Um exemplo de restrições encontradas na tabela é, por exemplo, quando A e B se relacionam por “é sobreposto por” (A si B) e B e C se relacionam por *antes* (B < C). Por

transitividade, as relações possíveis entre A e C são: *A antes C* ou *A sobrepõe C* ou *A encontra C* ou *A inclui C* ou *A é terminado por C*.

Tabela 3.4: Algumas restrições de transitividade resultantes da combinação de relações (modificado de (Allen, 1983)).

<b>A r1 B</b> \ <b>B r2 C</b>	<b>antes (&lt;)</b>	<b>durante (d)</b>	<b>...</b>
<b>depois (&gt;)</b>	sem informação	> si ei d t	
<b>é sobreposto por (si)</b>	< s e di ti	si d t	
<b>...</b>			

Os relacionamentos entre intervalos são mantidos em uma rede em que os nodos representam os intervalos individuais e cada arco é rotulado para representar os possíveis relacionamentos entre os dois intervalos. A Figura 3.2 mostra alguns exemplos dessa notação.

<b>Relação</b>	<b>Representação da Rede</b>
<i>i durante j</i>	$N_i \text{ -- (d) } \rightarrow N_j$
<i>i durante j</i> ou <i>i antes de j</i> ou <i>i inclui j</i>	$N_i \text{ -- (< d di) } \rightarrow N_j$
$(i < j)$ ou $(i > j)$ ou <i>i encontra j</i> ou <i>i é encontrado por j</i>	$N_i \text{ -- (< > e ei) } \rightarrow N_j$

Figura 3.2: Representando o conhecimento sobre relações como uma rede de intervalos (modificado de (Allen, 1983)).

Sempre que a leitura dos nodos for no sentido inverso, basta utilizar o nome inverso da relação. Para ler *i durante j* usa-se (*d*), para ler *j durante i* usa-se (*di*).

Considerando um exemplo simples, em que:

S sobrepõe ou encontra L

S é antes, encontra, é encontrado por ou depois de R

Estes podem ser fatos derivados de uma história tal como ‘João não estava na sala quando eu toquei o interruptor para acender a luz’, onde, S é o tempo em que o interruptor foi tocado, L é o tempo em que a luz foi acesa, e R é o tempo em que João estava na sala. A rede que representa o primeiro fato é apresentada como:

$N_s \text{ -- (s e) } \rightarrow N_l$ .

Quando o segundo fato é adicionado, a rede fica como segue:

$N_r \leftarrow (< e ei >) \text{ -- } N_s \text{ -- (s e) } \rightarrow N_l$ .

Cada fato adiciona uma nova restrição temporal aos relacionamentos entre os nodos, que podem, por sua vez, introduzir novas restrições por transitividade. Essas restrições precisam ser calculadas, pois a rede precisa representar todas as relações possíveis de serem deduzidas entre os nodos. Por exemplo, adiciona-se o fato *i é durante j* a uma rede em que se sabe que *i é antes de k*. É deduzida a restrição de que *i é antes de k*. Esse fato também é adicionado à rede o que possivelmente produz novas restrições para os

outros relacionamentos. O algoritmo de restrições procura na tabela de transições todas as possíveis restrições entre duas relações.

Para o exemplo acima, o algoritmo calcula as restrições entre as relações entre L e R passando por S. As relações passadas para que se calcule as restrições  $\{si\ ei\}$  e  $\{< e ei\}$ . Como a leitura é, de L para S e de S para R, usamos as relações inversas de S para L. Logo, para a combinação  $\{si\ <\}$  as restrições  $\{< s e di\ fi\}$  são adicionadas ao conjunto. As restrições são computadas para todas as outras combinações de relações. No final o conjunto de restrições entre L e R que deve ser adicionado à rede é  $\{< > s e di\ i ii\ ti =\}$ , o que resulta:

$$Nr \leftarrow (< e ei >) \text{ -- } Ns \text{ -- } (s e) \rightarrow Nl.$$

$$\left| \begin{array}{c} \text{--- } (< > s e di\ i ii\ ti =) \text{ ---} \\ \end{array} \right|$$

Se tentarmos adicionar o fato de que *L sobrepõe, inicia ou é durante R* (que pode ser uma continuação da historia acima, em que “João estava no quarto mais tarde, quando a luz apagou”) é preciso primeiro avaliar a intersecção das restrições já existentes entre L e R para eliminar relacionamentos impossíveis (tais como L é terminado por R e L inicia R). A nova relação fica:

$$Nl \text{ -- } (s i) \rightarrow Nr$$

E novamente, para adicioná-la as novas restrições devem ser calculadas. O algoritmo de cálculo de restrições continua a calcular enquanto produzir novas restrições entre os intervalos.

A álgebra de Allen lida com as noções de eventos e propriedades. Eventos ocorrem em um determinado intervalo de tempo (*ocorre*( $E,T$ )), enquanto propriedades são valores que se mantém durante um certo intervalo (*mantém*( $P,T$ )). Eventos são os ativadores das mudanças de estado das propriedades.

Um uso interessante de se testar se uma condição P se mantém durante um intervalo X é testar se ela se mantém em outro intervalo de tempo também. Por exemplo, se Y é durante X, e P se mantém em X, então P também se mantém em Y.

Segundo o estudo de (VENDLER, 1957), introduzido na seção 3.4.1, em que apresentamos o método AT, algumas propriedades podem valer para um intervalo de tempo, mas não para seus subintervalos ou para os pontos internos ao intervalo (propriedades não homogêneas). Esse argumento foi crucial para que Allen decidisse por não utilizar pontos para representar seus intervalos de tempo, ou seja, Allen não representa um intervalo como o conjunto de pontos dentro desse intervalo.

Se a lógica de pontos fosse escolhida, e um predicado fosse definido como verdadeiro em um intervalo somente se fosse verdadeiro para todos os pontos no intervalo, então essa lógica não poderia representar predicados não-homogêneos.

### 3.5 Resumo do Capítulo

Neste capítulo foram apresentados os principais assuntos relacionados à necessidade de representação de aspectos temporais em modelos computacionais. No capítulo 4, os modelos apresentados neste capítulo serão analisados visando decidir qual a melhor representação para o domínio de aplicação deste trabalho e para o tipo de problema de raciocínio.

### 3.5.1 Sobre a Nomenclatura de Eventos

Existe uma discussão sobre a definição mais adequada para eventos na literatura de modelagem de aspectos temporais. Para (TANSEL et al., 1993), um *evento* é definido como algo que acontece isoladamente em um instante de tempo, enquanto um *intervalo* é o tempo decorrido entre dois eventos. Em geral eventos são ocorrências usadas como ativadores do início ou do fim de processos ou como delimitadores de intervalos de tempo. Mas para (MCDERMOTT, 1982) essa definição não é boa porque implica que eventos não têm duração e porque não cobre eventos sem efeito, tais como “Maria está dançando”. A escolha de McDermott é identificar um evento com o conjunto de intervalos sobre os quais uma ocorrência do evento acontece. Já a classificação de Allen distingue *propriedades*, *eventos* e *processos*. Propriedades correspondem a estados que se mantêm homogeneamente durante um intervalo de tempo específico. Eventos ocorrem sobre o menor tempo possível e descrevem uma atividade que envolve um produto ou uma saída esperada. Processos ocorrem de maneira intermediária entre propriedades e eventos: eles devem ocorrer pelo menos durante um subintervalo de um intervalo de tempo específico. Na definição de evento no Glossário de Conceitos de Bancos de Dados Temporais (JENSEN et al., 1998) um *evento* é definido como um fato instantâneo. *Processo*, na nomenclatura, foi substituído por *macro-evento*, ou seja, um fato que ocorre com uma duração.

As entidades tratadas neste trabalho são processos geológicos que ocorreram durante um intervalo de tempo não especificado, em um ponto de tempo não especificado, mas que geraram produtos. Para evitar confusões de significado com os agentes do domínio, tomou-se a decisão de utilizar no modelo o mesmo nome utilizado no domínio de aplicação, ou seja, *eventos*, significando uma ocorrência do passado que teve duração de tempo mas que já está finalizada. Evento, para esse trabalho, é um *acontecimento que provoca uma alteração no domínio e gera um produto*. Essas restrições serão mais bem detalhadas no capítulo 7, sobre a modelagem de aspectos temporais no domínio da Petrografia.

No capítulo 4, os modelos de representação temporal estudados neste capítulo serão analisados e comparados para decidir qual o modelo mais adequado para a representação de eventos específicos do domínio de aplicação deste trabalho, a petrografia sedimentar. Assim será decidido qual destes modelos vai influenciar a representação de conhecimento para o presente trabalho.



## 4 ANÁLISE DE MODELOS TEMPORAIS PARA REPRESENTAR ORDEM DE EVENTOS NA PETROGRAFIA SEDIMENTAR

Este capítulo fará uma análise das características dos formalismos de representação temporal apresentados no capítulo 3, visando decidir qual a melhor representação para o domínio de aplicação deste trabalho e para o tipo de problema de raciocínio: inferência sobre a ordem em que eventos ocorreram no passado.

Segundo o trabalho de (VILA, 1994), considerando-se a grande quantidade de escolhas das quais uma proposta de modelo temporal é feita (entidades, estrutura, métricas, inferência), não é sensato imaginar que possa existir uma teoria geral do tempo, baseada em escolhas ideais que sejam capazes de atingir um desempenho satisfatório independente da aplicação para a qual é destinada. Para Vila, o fundamental é determinar claramente os requisitos para a aplicação, de forma a tomar as decisões apropriadas, e conhecer as implicações das diferentes escolhas de modelagem.

Os requisitos da presente aplicação estão claros: o modelo precisa suportar a representação das características do domínio, dos eventos, de relações temporais entre eventos e das regras que criam a relação de dependência entre características do domínio e relação temporal entre eventos.

### 4.1 Lógicas de Primeira Ordem

No capítulo 3, na seção em que são apresentadas as lógicas de primeira ordem, percebe-se que todos os formalismos estudados baseiam suas representações em uma estrutura de lógicas de predicados. Por isso, ao analisar cada um dos formalismos já estaremos avaliando a capacidade de se representar o conhecimento sobre seqüência de eventos utilizando lógicas de primeira ordem.

### 4.2 Lógicas Modais Temporais (LMT)

As LMT não representam explicitamente nenhum aspecto temporal. Utilizando LMT é possível indicar que determinados predicados ocorreram no passado, ou ocorrem sempre no futuro, ou seja, é possível mudar o *modo* de interpretação daquele predicado. No caso do problema da interpretação de seqüência de eventos diagenéticos, todos os eventos ocorreram no passado. Isso seria representado como segue usando um operador modal:

```
PASSADO(Dissolução(mineral))
```

```
PASSADO(Precipitação(mineral))
```

Mas a capacidade de afirmar que um predicado ocorreu no passado não é suficiente. Para ser eficaz em representar o conhecimento sobre os eventos, a LMT deveria ser capaz de representar alguma relação entre os eventos. Essa relação é essencial para que se possa inferir sobre a seqüência em que os eventos aconteceram. Por essa razão, a LMT pura não é capaz de expressar o conhecimento necessário para a ordenação de eventos diagenéticos.

### 4.3 Cálculo de Eventos

Para explicar a aplicação do cálculo de eventos na tarefa de seqüência de eventos diagenéticos, serão utilizadas as seguintes afirmações sobre o domínio:

```
mineral2 cobre mineral1
porosidade dentro-de mineral2
```

Essas afirmações podem ser representadas, utilizando um formalismo de lógica de primeira ordem, como segue:

```
cobre(mineral2, mineral1)
dentro-de(porosidade, mineral2)
```

O cálculo de eventos permite estabelecer relações entre eventos e as características do domínio ao qual estão associados utilizando regras do tipo *inicia(evento, propriedade) se (aqui são definidas as características do domínio que indicam essa relação)*. Pelas regras que foram adquiridas do especialista, quando um constituinte de rocha cobre o outro, isso quer dizer que ocorreu a *precipitação do primeiro constituinte*<sup>10</sup>. Então, a partir de uma característica do domínio (característica = *cobre(mineral1, mineral2)*) definimos o *evento* que ocorreu (*evento = precipitação(mineral1)*). Além disso, o produto gerado pela precipitação foi o primeiro constituinte. Esse produto pode ser visto como a *propriedade* que vai ser manter depois que o evento ocorreu (*propriedade = mineral1*). Essas associações podem ser definidas da seguinte forma:

```
inicia(precipitação(mineral2), mineral2)          (R1)
se cobre(mineral2, mineral1)
```

Essa regra significa que a propriedade *mineral2* é iniciada a partir do evento *precipitação(mineral2)*. Mas existe uma outra propriedade sobre a qual podemos definir uma regra. A característica *cobre(mineral2,mineral1)* se refere também ao *mineral1*. A precipitação pode ser vista como o evento que termina o período em que a propriedade *mineral2* se mantém. Isso pode ser representado usando uma regra similar à *inicia*:

```
termina(precipitação(mineral2), mineral1)        (R2)
se cobre(mineral2, mineral1)
```

Dessa forma, quando as regras R1 e R2 forem unificadas com a característica *cobre(mineral2, mineral1)*, será possível afirmar que o evento *precipitação(mineral2)* termina a propriedade *mineral1* e inicia a propriedade *mineral2*. O cálculo de eventos inclui axiomas, tais como EC1 e EC2:

```
mantém(antes(e,r)) se termina(e,r) (EC1)
```

<sup>10</sup> As regras do especialista foram simplificadas para que o exemplo não fosse muito extenso. Na verdade, para afirmar o evento que ocorreu, seria necessário testar outras características a mais, mas elas seriam representadas de uma maneira análoga. Em vez de nomear todas as características individualmente, elas serão representadas como um argumento. Por exemplo, *propriedade(cobre, mineral2, mineral1)*.

$\text{mantém}(\text{depois}(e,r)) \text{ se } \text{inicia}(e,r) \text{ (EC2)}$

Esses axiomas são usados para definir em quais os intervalos determinadas propriedades se mantêm. Esses axiomas são importantes para definir a relação de ordenação entre propriedades e eventos. Dessa maneira pode-se inferir que:

$\text{mantém}(\text{antes}(\text{precipitação}(\text{mineral2}), \text{mineral1})) \text{ (P1)}$

$\text{mantém}(\text{depois}(\text{precipitação}(\text{mineral2}), \text{mineral2})) \text{ (P2)}$

Ou seja, a propriedade *mineral1* se mantêm no intervalo anterior ao evento *precipitação(mineral2)* e a propriedade *mineral2* se mantêm no intervalo posterior ao mesmo evento. Essas deduções podem ser representadas graficamente da seguinte forma (Figura 4.1):

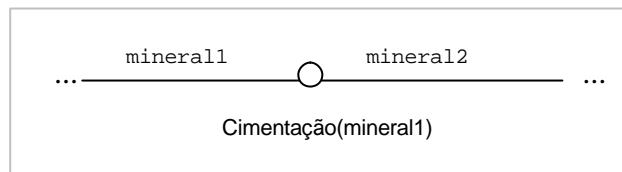


Figura 4.1: Propriedades que se mantêm nos intervalos de tempo antes e depois do evento *Precipitação(mineral1)*.

Ainda falta, neste exemplo, definirmos regras similares a R1 e R2, para a propriedade *dentro-de(porosidade, mineral2)*. Pelas regras que foram adquiridas do especialista, quando uma *porosidade está dentro de um mineral* (característica), isso quer dizer que ocorreu a *dissolução do constituinte* (evento). Além disso, o produto gerado pela dissolução foi a *porosidade* (propriedade). Essas associações podem ser definidas da seguinte forma:

$\text{inicia}(\text{dissolução}(\text{mineral2}), \text{porosidade}) \text{ (R3)}$

$\text{se } \text{dentro-de}(\text{porosidade}, \text{mineral2})$

Analogamente à regra R2, temos:

$\text{termina}(\text{dissolução}(\text{mineral2}), \text{mineral1}) \text{ (R4)}$

$\text{se } \text{dentro-de}(\text{porosidade}, \text{mineral2})$

Dessa forma, quando as regras R3 e R4 forem unificadas com a característica *dentro-de(porosity, mineral2)*, será possível afirmar que o evento *dissolução (mineral)* termina a propriedade *mineral* e inicia a propriedade *porosidade*. Aplicando os axiomas EC1 e EC2, teremos as seguintes afirmações a seguir, que podem ser representadas graficamente como na Figura 4.2.

$\text{mantém}(\text{antes}(\text{dissolução}(\text{mineral2}), \text{mineral2})) \text{ (P3)}$

$\text{mantém}(\text{depois}(\text{dissolução}(\text{mineral2}), \text{porosidade})) \text{ (P4)}$

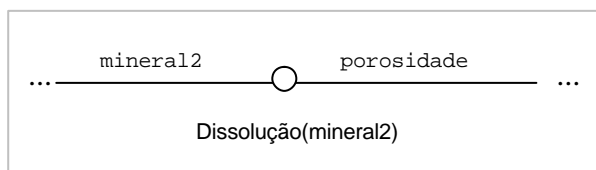


Figura 4.2: Propriedades que se mantêm nos intervalos de tempo antes e depois do evento *Dissolução(mineral2)*.

Os axiomas EC3, EC4, EC5 e EC6 são utilizados para definir os eventos de começo e fim dos intervalos de tempo. Por exemplo, o axioma:

```
início(antes(e',r), e) se (EC5)
depois(e,r) = antes(e',r)
```

Pode-se tentar unificá-lo com P3 e P2. Dessa forma, é preciso confirmar se:

```
depois(precipitação(mineral2), mineral2) =
antes(dissolução(mineral2), mineral2)
```

Ou seja, se o intervalo anterior ao evento *dissolução(mineral)* é o mesmo intervalo que se mantém depois do evento *precipitação(mineral2)*. O cálculo de eventos propõe algoritmos que garantem que o intervalo é o mesmo se algumas condições do domínio não estão sendo violadas. Por exemplo, se uma mesma propriedade se mantém entre o primeiro e o segundo evento, então o intervalo de tempo entre esses eventos é o mesmo.

Quando o especialista descreve uma rocha, ele descreve diversas características sobre diversos tipos específicos de minerais que podem pertencer a uma mesma classe (por exemplo, ele descreve k-feldspato, albita, plagioclásio, e todos fazem parte da classe Feldspato Detrítico). Mas o especialista considera todas essas especializações como sendo um mesmo mineral: o feldspato detrítico, ou seja, sua classe. Por isso esta representação vai utilizar o nome das classes de minerais. E quando uma classe de minerais aparecer em várias regras, todas estarão se referenciando ao mesmo mineral. Por isso, pode-se afirmar que as duas proposições acima – *antes(dissolução(mineral2), porosidade)* e *depois(precipitação(mineral2), mineral2)* – se referem ao mesmo mineral (*mineral2*). Logo, o intervalo de tempo em que a propriedade *mineral2* se mantém, nas duas proposições, é o mesmo. Essa lógica será a mesma para as outras comparações de igualdade entre intervalos de tempo. A partir da confirmação de que o intervalo antes da dissolução de *mineral2* e depois da precipitação *mineral2* é o mesmo, pode-se deduzir, pelo axioma EC5, que:

```
início(antes(dissolução(mineral2), mineral2),
precipitação(mineral2))
```

Ou seja, o início do período anterior à dissolução de *mineral2* é o evento precipitação(*mineral2*). A partir dessa confirmação, de que os períodos são iguais, pode-se também afirmar, baseado no axioma EC6, que:

```
fim(depois(precipitação(mineral2), mineral2),
dissolução(mineral2))
```

Ou seja, o fim do período posterior à precipitação de *mineral2* é o evento dissolução(*mineral2*). A partir do conhecimento dos eventos que delimitam o início e de fim de um intervalo, é possível chegar a uma relação de ordem entre os dois eventos. Essa relação pode ser representada como na Figura 4.3.

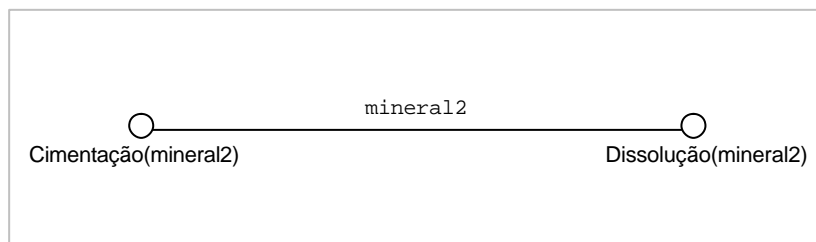


Figura 4.3: Representação da relação de ordem entre os eventos Precipitação(*mineral2*) e Dissolução(*mineral2*).

## 4.4 Cálculo de Situações

No domínio da petrografia, os eventos diagenéticos que ocorrem contribuem para a modificação da rocha de um estado inicial até o estado atual. Eventos como precipitação e infiltração adicionam novos minerais à rocha, eventos como dissolução e descompactação formam poros na rocha, entre outros eventos. O cálculo de situações poderia ser aplicado para representar as mudanças no domínio efetuadas pelos eventos. O seguinte exemplo representa o fato de que a dissolução de um mineral  $m$ , na situação  $s$  produz, no futuro, que exista porosidades ( $p$ ) dentro do mineral.

$$\forall m (\forall p (\forall s (\text{dissolução}(m)(s)))) \rightarrow F(\text{dentro-de}(p,m))(s)$$

Outro exemplo é o fato de que a precipitação de um mineral  $m2$  sobre um mineral  $m1$  pode fazer com que o primeiro mineral apareça cobrindo o outro.

$$\forall m1 (\forall m2 (\forall s (\text{precipitação}(m2)(s) \wedge \text{existe}(m1)(s)))) \rightarrow F(\text{cobrindo}(m2,m1))(s)$$

Esse cálculo poderia ser utilizado para inferir o que é modificado na rocha quando determinados eventos ocorrem. Entretanto, não é possível deduzir a relação temporal entre os eventos, pois os eventos em si não possuem uma representação. O cálculo é feito sobre a situação anterior e a posterior, mas o evento não possui representação explícita. Por esse motivo, esse cálculo não é suficiente para se representar o conhecimento necessário para ordenar os eventos.

## 4.5 Álgebra de Intervalos

A álgebra de intervalos de Allen é útil a partir do momento em que já são conhecidas as relações temporais entre os intervalos. Ou seja, já se sabe que um intervalo encontra o outro, ou vem antes de outro, ou sobrepõe o outro. Para a representação desse conhecimento a álgebra de Allen é extremamente eficaz: é possível representar uma extensa combinação de relações temporais entre intervalos, bem como seus inversos. Além disso, a álgebra provê algoritmos para se propagar por toda a rede de intervalos as restrições temporais que uma nova relação de tempo pode incluir. O objetivo da álgebra, então, é proporcionar uma maneira de manipular relações complexas de tempo entre intervalos e garantir que as relações de tempo entre todos os intervalos de uma rede se mantêm consistentes mesmo depois da inclusão de novos relacionamentos.

Entretanto, a álgebra de Allen supõe que os relacionamentos temporais são conhecidos no domínio, ela não suporta a *dedução* desses relacionamentos. Para uma tarefa em que os relacionamentos temporais não estão explícitos no domínio, e que precisam ser deduzidos a partir das características do domínio, a álgebra de intervalos não é suficiente.

No caso do domínio da Petrografia, temos, por exemplo, os seguintes eventos:

compactação ocorre *antes* da precipitação de feldspato

infiltração da anidrita ocorre *depois* da precipitação de feldspato

Essas afirmações podem ser representadas da seguinte forma:

$\text{antes}(C, Pf)$

$\text{depois}(Ia, Pf)$

Depois, monta-se a rede de nodos, adicionando um relacionamento temporal por vez e calculando as restrições de tempo, que neste caso não existirão porque não está sendo

adicionada mais de uma relação entre um mesmo par de intervalos. A rede fica como segue:

$$C \text{ -- } < \rightarrow Pf \leftarrow < \text{ -- } Ia$$

Uma observação a ser feita é que no caso dos eventos diagenéticos, da petrografia, as relações temporais são, predominantemente, *antes*, *depois* ou *durante*. Não foram encontradas ocorrências, nos relatos do especialista, de outras relações temporais entre os eventos (tais como, *sobrepõe*, *encontra* ou *termina*). Além disso, praticamente não existem casos em que seja utilizada mais de uma relação temporal entre dois eventos. O máximo que pode ocorrer é como no exemplo seguinte:

precipitação de feldspato veio antes ou junto com a precipitação de anidrita

Em que as características do domínio não foram suficientes para se afirmar com certeza se o primeiro evento aconteceu antes ou se ocorreu alternadamente com o segundo. Mas não existirão casos em que novas relações temporais entre os mesmos eventos virão a ser adicionadas com as próximas deduções.

Nesse caso, o algoritmo de propagação de restrições da lógica de Allen será subutilizado.

Resumindo-se os motivos pelos quais a lógica de Allen não poderá ser aplicada a esse domínio:

- Provê mais relações temporais do que o necessário. Só serão utilizadas as relações *antes*, *depois* e *durante*.
- O domínio, por natureza, não vai proporcionar mais do que uma relação temporal entre dois eventos, de forma que o algoritmo de propagação de restrições não seria usado.
- O motivo mais forte é o que o formalismo oferece *a menos* para a representação: ela não permite representar predicados de dedução dos eventos e de seu relacionamento temporal a partir de características específicas do domínio. É preciso supor que esses eventos e seus relacionamentos já estão explicitados.

Mesmo assim, a álgebra de Allen é sempre uma teoria na qual podemos nos basear no futuro para estender outros formalismos para representarem relações temporais mais sofisticadas.

## 4.6 Resumo do Capítulo

Por conta dessa análise, o formalismo de cálculo de eventos pode-ser considerado o mais apropriado para representar o conhecimento necessário para a inferência de dedução e ordenação de eventos diagenéticos do domínio da petrografia sedimentar. Entretanto, para a aplicação dessa estrutura de representar ordem temporal no domínio da petrografia, foi analisada uma maneira de simplificar as regras do CE.

O cálculo de eventos foi proposto para representar todas as condições que o domínio precisa satisfazer para que se considere que um determinado evento tenha sido iniciado ou terminado. Como no exemplo dado no capítulo anterior, em que o fato de alguém ser possuidor de um livro só pode ser considerado iniciado se algumas condições foram satisfeitas, como existir um *objeto*, um *receptor* e o ato de *dar*.

inicia(e, possui(x, y)) se  
ato(e, dar) e receptor(e, x) e objeto(e, y)

No domínio da petrografia, para a tarefa de ordenação de eventos diagenéticos, o fato de um evento ser iniciado ou terminado não depende de tantas e tão detalhadas condições do domínio. Pelas regras eliciadas do especialista, um processo em geral vai ser associado a uma feição do domínio, como apresentado no exemplo R2.

```
termina(evento_gerador(mineral2), mineral1) se      (R2)
cobre(mineral2, mineral1)
```

Pelo fato de as condições serem simples, foi possível unificar as regras que definem as *condições do domínio* nas quais os eventos ocorrem, com os axiomas que definem as *relações de ordem* (como o axioma EC1).

```
mantém(antes(e,r)) se termina(e,r) (EC1)
```

Essa unificação gerou regras mais simples, que relacionam diretamente as *condições do domínio* com as *relações de ordem*. O processo de unificar as regras é demonstrado a seguir.

Primeiro as regras foram trazidas para a ordem de implicação (SE-ENTÃO):

```
SE cobre(mineral2, mineral1)      (R2)
ENTÃO termina(evento_gerador(mineral2), mineral1)
SE termina(e,r) ENTÃO mantém(antes(e,r)) (EC1)
```

A estrutura SE-ENTÃO tem a propriedade da transitividade. Por isso, a partir de (R2) e de (EC1) é possível dizer que:

```
SE cobre(mineral2, mineral1)      (R5)
ENTÃO mantém(antes(evento_gerador(mineral2), mineral1))
```

Essa regra simplificada foi usada para definir a estrutura das regras que associam as condições do domínio com as relações de ordem. A decisão sobre a estrutura que vai representar as regras do domínio foi a principal contribuição do estudo do formalismo de cálculo de eventos. A definição dessas regras será detalhada no capítulo 7, que descreve o desenvolvimento do modelo de conhecimento para o domínio da Petrografia. As regras que indicam a ordem entre os eventos serão definidas na base de conhecimento do modelo.

## 5 ESTUDO DE CASO DE AQUISIÇÃO DE CONHECIMENTO

Em geral, avaliações sobre a utilidade de técnicas de AC vêm de duas fontes principais (SHADBOLT e BURTON, 1990): a primeira são guias teóricos de como conduzir o processo de aquisição, baseados em um entendimento de senso comum sobre as combinações de variáveis importantes no processo. Esses trabalhos avaliam muitas técnicas, e apresentam classificações, mas não mostram estudos de casos e aplicações práticas; são trabalhos mais empíricos.

O segundo tipo de fonte descreve como o processo de AC foi conduzido efetivamente em um projeto em particular; não sendo a eliciação de conhecimento o objetivo do projeto, mas sim o conhecimento eliciado, organizado, modelado, e muitas vezes implementado como um sistema. Apesar de esse tipo de trabalho ser útil para prevenir futuros engenheiros de conhecimento sobre erros gerais e para dar algumas sugestões sobre a utilidade dos métodos usados, não é feita uma avaliação formal sobre as técnicas. Não existe garantia de que o experimento repetido funcionará da mesma maneira.

Este trabalho se enquadra parcialmente no segundo tipo de guia de AC, pois, independente do experimento de avaliação das técnicas, o processo de AC neste trabalho tem por objetivo *coletar e organizar o conhecimento sobre a interpretação da seqüência em que ocorreram determinados processos no domínio da petrografia*. O tipo de conhecimento a ser eliciado, então, é conhecimento sobre ordem de processos, ou seja, conhecimento sobre aspectos temporais do domínio. As informações coletadas nesse processo vão estender a ontologia de domínio já desenvolvida para o sistema *PetroGrapher*, detalhada no capítulo 1. O modelo original da ontologia da petrografia está em inglês e será estendido com os eventos diagenéticos e os aspectos temporais. À vista disso, o processo de aquisição de conhecimento também tem como objetivo preencher as lacunas de conhecimento do modelo já existente.

### 5.1 Planejamento das Sessões de AC

Antes da realização das sessões de aquisição de conhecimento em si, foi necessário planejá-las, estudando o material de referência, e escolhendo as técnicas mais adequadas para a tarefa.

Uma das decisões em relação à avaliação de técnicas de eliciação de conhecimento diz respeito à escolha das técnicas. Conforme mostrado no capítulo 2, existem diferentes técnicas de eliciação, e elas não podem ser testadas exaustivamente, por motivos de limitação do tempo destinado a esse trabalho. O critério utilizado para selecionar as técnicas, então, foi escolher pelo menos uma técnica representativa de cada categoria. Foi necessário escolher o maior número de técnicas que fossem possíveis de serem testadas nesse período.



Neste experimento, o melhor resultado seria poder indicar um método de aplicação mais eficaz em coletar conhecimento temporal, assim como poder dizer que classificação de fichas é útil para descobrir atributos taxonômicos (conhecimento declarativo), e que rastreamento de processos permite visualizar as heurísticas do especialista (conhecimento procedimental). Os seguintes métodos de AC foram escolhidos:

- Entrevistas não-estruturadas e estruturadas;
- Observação;
- Classificação de fichas;
- Técnica para eliciar relações temporais;
- Ferramenta computacional de aquisição de conhecimento
- Eliciação triádica;
- Grades de repertório;
- Limitação de informações (focalizando contextos);
- Análise de protocolos.

O material utilizado para as entrevistas foi basicamente o bloco de anotações e um gravador de áudio. Algumas sessões em que foram aplicadas técnicas mais elaboradas de AC necessitaram outros materiais, como filmadora, cartões, computador. Por conta da disponibilidade do especialista (que costuma ser de pouco tempo) as entrevistas aconteceram em uma frequência média de duas vezes por mês e duravam em torno de uma hora, pois era o tempo que o especialista tinha disponível em um dia de trabalho. Contando somente os meses do projeto em que foram realizadas entrevistas, o processo teve uma duração total de oito meses.

Todas as etapas de eliciação de conhecimento (a aquisição diretamente com o especialista) foram gravadas utilizando um mini-gravador de fita cassete (analogico). O áudio foi transformado em arquivo de som digital e transcrito para arquivos texto, para facilitar a procura das informações.

Na seção seguinte são relatadas as principais etapas do processo de aquisição de conhecimento, descrevendo as técnicas utilizadas, as dificuldades de eliciação, e os resultados obtidos.

## **5.2 Processo de Aquisição de Conhecimento sobre Interpretação de Seqüência Diagenéticas**

Um grande facilitador desse processo de aquisição foi o fato de o especialista já ter participado da primeira fase de aquisição de conhecimento do Projeto *PetroGrapher*, em que foram coletados todos os termos básicos da ontologia do domínio e as relações causais para uma das interpretações petrográficas, a interpretação de ambiente diagenético. Dessa forma, ele já reconhecia a necessidade da interação com o engenheiro de conhecimento.

### **5.2.1 Fase Anterior à Entrevista: Imersão na Literatura**

O fato de já existir o Projeto *PetroGrapher* evitou que fosse necessária uma imersão extensiva na literatura, visto que o vocabulário de geologia já era previamente

compreendido pelo engenheiro de conhecimento e que a maior parte da ontologia de domínio já havia sido eliciada.

A literatura indicada foi um material didático preparado pelo especialista demonstrando algumas das regras utilizadas por ele para definir uma ordem entre o aparecimento dos constituintes. Regras como “Um constituinte diagenético que ocorre entre a maior parte dos grãos foi precipitado antes da compactação (pré-compactacional)”. Esse material permitiu compreender o tipo de conhecimento que seria eliciado nas fases posteriores de entrevistas, pois já indica que o tipo de informação temporal que precisa ser eliciada é puramente seqüencial, ou seja, não é necessária uma quantificação do intervalo de tempo entre os processos que geraram constituintes. No exemplo acima, a informação temporal gerada é de que esse constituinte específico foi precipitado antes de os grãos da rocha se compactarem.

### 5.2.2 Fase 1: Entrevista Não-Estruturada

Esta foi uma entrevista não estruturada, ou seja, as perguntas não foram preparadas antes. No início as perguntas foram bem abrangentes, pedindo informações gerais sobre o processo em foco, tais como “O que é a seqüência diagenética?”, “Qual o objetivo dessa interpretação?” ou “Como você interpreta a seqüência diagenética?”. As outras perguntas que permearam a entrevistas foram motivadas por uma necessidade de maior detalhamento das respostas do especialista. Por exemplo, o especialista explicou que a principal informação visual a ser analisada para determinar a seqüência diagenética são as relações paragenéticas entre os constituintes diagenéticos e outros constituintes. Como existem constituintes detríticos também, foi solicitado o detalhamento desses processos entre constituintes, solicitando se “... existem outros processos entre constituintes detríticos?”.

Essa primeira entrevista também foi bastante útil para criar as conexões entre aquilo que já se conhecia previamente de Petrografia e o processo de interpretação seqüencial a ser representado. Foi bastante importante compreender que a maior parte das informações que é analisada para gerar uma seqüência já está sendo descrita atualmente no sistema *PetroGrapher*, o que significa que todas as entradas para o futuro algoritmo já estão sendo coletadas. Só é necessário identificá-las todas.

### 5.2.3 Fase 2: Nova Imersão na Literatura

Na primeira entrevista, o especialista indicou como leitura suplementar um Atlas geral de Petrografia Sedimentar. A leitura desse Atlas foi importante para criar uma ligação dos conceitos eliciados do especialista com as definições compartilhadas por vários autores proporcionadas por um material didático. Muitos termos que ficaram sem uma definição mais completa durante a entrevista (mesmo porque, não é interessante abrir demais o rol de perguntas) foram clarificados na leitura do Atlas. Ele permitiu o esclarecimento do significado de muitos termos utilizados na geologia. Se essas perguntas fossem programadas para a entrevista com o especialista, um tempo caro seria utilizado para dúvidas que podem ser tiradas com consultas aos livros.

A imersão na literatura anterior ao processo de eliciação é importante para o engenheiro de conhecimento definir perguntas utilizando um vocabulário não-leigo, mas a leitura de obras relacionadas ao assunto após as entrevistas é fundamental para concretizar os conceitos e preencher as lacunas semânticas que ficam entre as diversas informações adquiridas. A indicação de uma leitura complementar às entrevistas melhora a qualidade da informação adquirida simplesmente porque aumenta a compreensão do entrevistador sobre o assunto.

Além de filtrar aquelas perguntas que seriam desnecessárias, o engenheiro de conhecimento já consegue definir perguntas mais pertinentes para a próxima fase de entrevistas.

### 5.2.4 Fase 3: Entrevista Estruturada

Nessa entrevista foram preparadas algumas perguntas sobre as dúvidas que ficaram da última entrevista e para esclarecer algumas questões que sobraram sobre os conceitos explicados no Atlas de Petrografia. Foram oito perguntas planejadas, mas a entrevista se estende pelos esclarecimentos intermediários. As transcrições da entrevista são demonstradas no Apêndice A.

Já foram extraídas algumas regras de ordenação de processos utilizadas pelo especialista e foi detalhado mais claramente que tipos de informações realmente serão usadas como entrada para o algoritmo de seqüência, que são basicamente as relações paragenéticas entre os constituintes diagenéticos e todos os outros (p.ex.: constituinte 1 cobre constituinte 2), e as relações geométricas (p.ex.: tipo de contato, volume intragranular) entre os grãos.

### 5.2.5 Fase 4: Observação

A terceira etapa da eliciação de conhecimento foi feita através de um processo de observação e rastreamento de processos. Não foi realizada uma entrevista; em vez disso, o especialista foi convidado a executar uma descrição de amostra de rocha voltada para a interpretação de seqüência. Esse processo foi gravado em vídeo e áudio para permitir uma análise posterior do material. O processo foi filmado para que algumas informações demonstradas por um “apontar de dedos” na tela pudessem ser recuperadas posteriormente. Entretanto, o fato de ser filmado deixou o especialista em posição mais defensiva quanto à eliciação, pois ele achava que para capturar as imagens das amostras seria suficiente salvar cada uma delas em formato digital de foto. Foi preciso convencê-lo de que a informação significativa não estava somente na imagem, mas sim na combinação da imagem da amostra, do gesto de apontar uma região específica da amostra e da explicação em áudio sobre o significado daquela região da rocha.

Foram analisadas duas amostras de rocha ao microscópio. A descrição não foi feita formalmente, ou seja, analisando as características visuais ponto a ponto e relatando extensivamente. O especialista fez uma descrição dirigida para a interpretação de seqüência, começando pelas características que são procuradas em primeiro lugar (aquelas que indicam os processos que aconteceram mais cedo) e seguindo em busca de feições que demonstrem quais são os processos seguintes. O resultado da interpretação de seqüência de uma das amostras é mostrado na Figura 5.1.

Essa entrevista foi muito útil para identificar a estratégia de interpretação seguida pelo especialista, o que foi essencial para definir os passos do método de solução de problemas para ordenação de processos, descrito no capítulo 7. Ele aplica regras bem definidas para estabelecer os passos de inferência que levam até a ordenação final dos processos. Por exemplo, nas duas descrições de amostras, ele iniciou o processo procurando pelo material que estava depositado ao redor dos grãos, formando cutículas. Dependendo da maneira como esse constituinte se apresenta em relação aos grãos, é um indicativo forte de que ele foi depositado antes da compactação, e isso já indica que é um dos processos que ocorreram mais cedo. Depois disso ele procura por relações de cobertura ou engolfamento, e então por processos de dissolução e substituição.

Entretanto, a Figura 5.1 representa a maneira como o *engenheiro de conhecimento* anotou o resultado da interpretação, pois o especialista estava somente observando a

amostra e verbalizando as interpretações. A conseqüência disso é que a seqüência anotada foi entre *constituintes* (dolomita, quartzo, anidrita) e não entre os *processos* que geraram esses constituintes. Mais tarde percebeu-se que essa informação não estava sendo verbalizada pelo especialista (detalhado nas próximas seções). Por isso teria sido mais vantajoso solicitar que o especialista, ao final da observação, produzisse um rascunho do seu resultado (o que geraria uma seqüência de processos). Dessa maneira o engenheiro teria compreendido a lacuna entre o que o especialista falava e o que ele gerava de resultado.

- | <b>Amostra 1</b> |  |
|------------------|--|
| 1)               | Cutículas sobre os grãos, cutículas nos contatos entre os grãos .  |
| 2)               | Dolomita<br>recobrimdo a cutícula ( <i>covering</i> )<br>intergranular pré-compactacional (veio antes da compactação)              |
| 3)               | Compactação Mecânica/Química   |
| 4)               | Quartzo<br>crescimento de quartzo<br>engolfando a dolomita<br>posterior à dolomita<br>descontínuos<br>ausente das áreas de contato |
| 5)               | Anidrita<br>engolfando<br>substituindo / corroendo as bordas do Quartzo<br>Quartzo corroído pela Anidrita, então é anterior a ela. |
| 6)               | Porosidade<br>pelo menos posterior à dolomita (por causa da dissolução).   |

Figura 5.1: Resultado da interpretação de seqüência diagenética feita pelo especialista para uma amostra de rocha.

A seqüência formada é linear, ou seja, os processos são descritos como se ocorressem uns após os outros, sem processos concomitantes. Isso não representa fielmente a realidade, pois muitos processos ocorrem paralelamente na natureza. Além disso, existem muitos casos de processos recorrentes, ou seja, a rocha sofre processos que se repetem depois de um tempo. Como o especialista não dispõe de ferramentas adequadas para definir uma seqüência com ramos paralelos de processos, acaba sendo mais prático deixar a seqüência linear. Mas um algoritmo computacional pode gerar e disponibilizar uma série de alternativas de apresentação da seqüência para processos que tenham a possibilidade de terem ocorrido concomitantemente.

Foi possível perceber na descrição, pela necessidade que o especialista teve em expressar a descrição usando as informações que estão na ontologia de domínio do sistema, que nem todas as entradas necessárias são atualmente descritas no *PetroGrapher*. Não existem êições na ontologia para representar informações como “um mineral está ausente da área de contato entre os grãos” no atributo *localização* do constituinte, na ontologia. Para indicar, por exemplo, que a rocha sofreu compactação, o especialista precisa descrever que ela aparenta estar fraturada ou deformada. Os valores *fraturado* e *deformado* devem ser descritos na ontologia como um atributo de modificação do conceito *constituente detrítico*.

### 5.2.6 Fase 5: Mudança da Saída da Interpretação

Como comentado na sessão 5.2.5, o especialista até essa fase das entrevistas estava relatando as interpretações como uma seqüência de aparecimento de constituintes o que dava a impressão que a tarefa tinha como objetivo ordenar a seqüência em que os constituintes se formaram para a composição final da rocha. Isso porque para o

especialista, os processos associados a esses constituintes eram óbvios, e ele estava acostumado a fazer interpretações resumidas, em que o resultado era uma lista que diz qual constituinte apareceu antes do outro. Foram então definidas as regras que associavam constituintes através de uma relação temporal (sessão 5.2.10). Ao serem apresentadas essas regras para o especialista, ele detectou que faltavam os nomes dos processos. Nesse momento foi detectado que as regras conseguiam definir uma relação temporal entre os produtos finais de um processo diagenético, mas não entre os processos que geraram esses produtos, que deveriam ser o foco da aquisição. Então foi possível perceber que a saída esperada era uma seqüência, mas não de *constituintes* e sim de *processos*. Ou seja, em vez da resposta ser a seqüência “quartzo - pirita - dolomita” a saída deveria ser “*precipitação* de quartzo – *dissolução* de pirita - *dolomitização*”. Mas para o especialista, os processos sempre foram óbvios: quando ele simplesmente descrevia uma ordem de constituintes o processo implícito era claro. Entretanto, os nomes dos processos não estavam sendo explicitados.

O especialista mostrou então um relatório oficial de uma interpretação não resumida, ou seja, com os nomes dos processos listados na ordem em que ocorreram. Dessa forma foi possível entender a saída real da interpretação (Figura 5.2).

<b>Amostra 2</b>	
1)	Dissolução de grãos de minerais detríticos pesados instáveis
2)	Precipitação de óxidos de ferro e de titânio
3)	Infiltração mecânica de argilas detríticas
4)	Precipitação de nódulos e crostas de sílica
5)	Precipitação de calcita em nódulos, crostas e cimentando <i>lags</i> intraclásticos
6)	Compactação mecânica
7)	Compactação química
8)	Transformação das argilas infiltradas esmectíticas em I/S e C/S
9)	Cimentação por crescimentos secundários de quartzo
10)	Crescimentos secundários e neomorfismo de feldspato potássico
11)	Cimentação por calcita poiquilótica
12)	Dissolução parcial da calcita e de grãos de feldspatos
13)	Precipitação de caulinita
14)	Ordenação dos interestratificados illita-esmectita ou clorita-esmectita
15)	Neoformação de franjas de clorita
16)	Neoformação de illita fibrosa
17)	Nova precipitação de anatásio
18)	Crescimentos e cristais prismáticos de quartzo ( <i>outgrowths</i> )
19)	Albitização dos grãos de feldspatos
20)	Pirita grossa corrosiva

Figura 5.2: Exemplo de interpretação de seqüência diagenética, demonstrando os processos associados aos constituintes. A descrição é feita como uma seqüência linear de processos.

A partir daí, todas as interpretações feitas pelo especialista utilizaram a nomenclatura de processos, para que as regras pudessem ser definidas sobre processos.

### 5.2.7 Fase 6: Classificação de Fichas

Depois da transcrição e da análise de protocolos sobre as entrevistas anteriores, foram coletados alguns dos termos que o especialista associa aos constituintes e que dão a idéia da ocorrência de um processo sobre a rocha. Isso porque era preciso diferenciar os termos que realmente indicavam processos que ocorreram no passado daqueles termos que identificavam o produto dos processos, uma noção difícil de diferenciar para quem não trabalha no domínio da geologia. Um exemplo de frase transcrita de entrevistas com o especialista e que provê alguns termos dessa categoria é a seguinte:

*"(A porosidade) pode ser primária, ou seja, deposicional. Quando os grãos são depositados, fica um espaço entre os grãos, porosidade primária intergranular. A porosidade pode ser secundária, gerada na diagênese, por dissolução, seja dos grãos, seja do cimento; por contração, geralmente de material argiloso, que é hidratado, perde água, contrai, e forma gretas de contração; por fraturamento, pode quebrar. Esse é o tipo de porosidade secundária".*

Também foram coletados termos de artigos da área da petrografia que faziam descrições de seqüências diagenéticas, tais como (HIGLEY, 1997).

Os seguintes termos foram coletados das transcrições:

*substituição, precipitação, dissolução, dissolução por pressão, contração, descompactação, cobrindo, engolfando, sobrecrecimento, cimentação, fraturamento, pré-compactação, pós-compactação, crescimento para fora, expandindo, colocado, corroer, deformação mecânica, dolomitização, infiltração argila, compactação mecânica, compactação química, infiltração.*

Esses termos foram escritos em pequenos cartões, e numerados para facilitar a anotação da configuração como eles foram organizados. A técnica de classificação de fichas diz que os cartões devem ser apresentados para o especialista e esse define os critérios de separação.

Entretanto, nessa seção a técnica foi usada de uma maneira dirigida, ou seja, para diferenciar conceitos que foram coletados nas entrevistas mas cuja definição não ficou clara para o engenheiro, mas que para o especialista é óbvia. Por isso não foram feitos cartões de todos os termos coletados nas entrevistas, mas somente dos que representavam processos diagenéticos ou relações paragenéticas. Com esse objetivo, foi solicitado que o especialista separasse os termos que indicavam processos daqueles que indicavam produtos (relações paragenéticas).

O especialista começou a separar usando esse critério, mas se deu conta que existiam muitos termos descartados (conforme a idéia da categoria "lixo" descrita na seção 2.3.4). Assim, ele criou subcategorias de divisão para enquadrar os termos que não se enquadravam em processo ou produto. O resultado da classificação é demonstrado no Apêndice A.

O objetivo também era, tendo demonstrado para o especialista que existe um conjunto de termos usados como entrada (os produtos) e um conjunto de termos que farão parte da saída da inferência (os processos), coletar o restante dos termos de saída, ou seja, gerar uma lista completa dos processos que o especialista ordena na interpretação de seqüência diagenética.

Foram dois os resultados da classificação de fichas. A primeira foi no sentido da compreensão do domínio: foi possível entender a diferença entre duas informações importantes do domínio, as relações paragenéticas e os nomes de processos. O especialista consegue fazer essa separação naturalmente, mas as diferenças não são óbvias para quem não está habituado ao domínio. Por exemplo, *precipitação* é um processo, mas o fato de um constituinte estar *expandido* pela precipitação é uma relação paragenética.

O outro resultado foi em relação a eliciação das categorias de processos. Percebeu-se que o especialista agrupa os processos em diferentes categorias, por exemplo, processos químicos e físicos. Dessa maneira, foi possível identificar a necessidade de uma lista dos principais processos utilizados em uma interpretação de seqüência e de suas categorias. O especialista elaborou uma listagem de processos, que pode ser conferida

no Apêndice A. Ainda foi possível identificar quais conceitos não faziam parte de nenhuma dessas classes.

### **5.2.8 Fase 7: Observação: Detectar Associação entre Produtos e Processos**

Depois de os processos terem sido diferenciados das relações paragenéticas através da classificação de fichas, era necessário entender como eles poderiam ser inferidos. Isso consistia um problema porque:

1. Os constituintes e suas feições eram descritos pelo usuário do sistema como entrada;
2. As regras eliciadas do especialista obviamente conseguiam definir uma relação temporal entre os produtos visualizados na rocha, ou seja, uma feição foi gerada antes da outra, etc;
3. Entretanto, o especialista retornava como saída uma relação temporal entre os processos que geraram essas feições, e não entre os produtos, que são os constituintes e feições.

Logo, existe uma relação que o especialista faz entre os constituintes e os processos que os geraram, mas essa relação ainda não havia aparecido. Nesse ponto foi detectada uma lacuna entre o conhecimento eliciado e o que necessário para solucionar o problema. Estava faltando a associação entre os produtos e os processos que os geraram.

Foi realizada uma observação de uma descrição, dirigida para interpretação da seqüência de processos diagenéticos de uma amostra. Nessa observação, a cada processo inferido pelo especialista, eram feitas perguntas concorrentes para tentar compreender porque ele inferia aquele processo. Foi possível perceber que esse era o ponto mais difícil da aquisição, pois para o especialista a associação era óbvia e ele não conseguia explicitar as heurísticas que usava para associar.

Um exemplo que demonstra bem essa dificuldade de verbalização foi quando o especialista descreveu uma porosidade como tendo sido gerada por um processo de dissolução parcial de outro constituinte que ocupava aquele lugar previamente. Foi perguntado então porque aquela porosidade não poderia estar ali desde o princípio, ou seja, porque ela não seria uma porosidade primária. A resposta do especialista foi "como esse grão poderia ter sido carregado nesse estado deplorável?".

Isso demonstra que o especialista usou as características visuais da rocha para identificá-la como uma rocha com muitas porosidades, e que tinham um formato tão "deplorável" que se mostrava óbvio que não poderiam ter vindo desde o início da rocha, pois ela não teria sido transportada. Mas quando foi tentado traduzir essa noção que o especialista tem de um constituinte em "estado deplorável" para quais feições geológicas descritas representam esse estado, ele responde que nada do que ele descrevia poderia indicar isso. A técnica de observação comentada não foi suficiente então para eliciar as associações que o especialista faz entre produtos geológicos e os processos que os geraram.

### **5.2.9 Fase 8: Entrevista estruturada: Detectar Associação entre Produtos e Processos**

Como detectamos que a saída do algoritmo de inferência será a ordem entre os processos, e a informação que temos até o momento são somente os produtos desses processos, foi necessário descobrir qual a associação o especialista faz entre os produtos que ele visualiza na amostra de rocha e os processos que os geraram.

Nessa entrevista, foi apresentada para o especialista uma lista de informações retiradas de uma descrição de amostra de rocha. Ele deveria confirmar quais daquelas informações são realmente utilizadas quando ele define qual o processo gerou um determinado produto. As principais informações utilizadas pelo especialista para inferir que processo gerou uma feição geológica são as informações descritas sobre os constituintes: modificadores, hábitos, localizações e relações paragenéticas.

Sabendo quais informações ele usa, decidimos que ele precisava definir, para cada produto, qual o processo relacionado. Mas qual o momento mais adequado para definir essa associação:

1. Listar os processos durante uma descrição de rocha real?
2. Listar todas as combinações produtos-processo extensivamente?
3. Listar primeiro todos os processos e associá-los com os produtos que podem gerar?

Para o especialista, a opção (a) implica utilizar uma variedade enorme de lâminas que pudessem cobrir todas as possibilidades, logo, foi descartada. E a opção (c) não está na ordem natural para o especialista: ele deduz o processo que ocorreu a partir das características presentes na amostra, e não o contrário. Logo, foi decidido por listar todas as combinações de produtos exaustivamente, ou seja, apresentar para o especialista todas as possíveis combinações de feições geológicas que ele pode descrever, e associá-las com o processo que poderia ter gerado essas características.

A dificuldade então foi em apresentar essa lista extensiva de feições da maneira mais prática para o especialista. Essa definição poderia ser manual, mas sendo extensiva, levaria muito tempo. Então a escolha foi por utilizar uma versão modificada da própria interface de entrada do sistema *PetroGrapher*, a interface de contagem de pontos, com cuja usabilidade o especialista já estava acostumado. O *Contador de Pontos* é um módulo já usado pelo especialista para entrada de dados no sistema *PetroGrapher*, e permite que o especialista faça uma análise quantitativa e qualitativa dos constituintes que aparecem em uma amostra de rocha. Esse módulo foi reduzido, retirando a parte de análise quantitativa, para permitir somente a descrição das feições mais significativas para a interpretação de seqüência. As alterações seriam as seguintes:

- Eliminar da interface os componentes de entrada de informações não relevantes para a seqüência diagenética;
- Adicionar um componente visual para a escolha dos processos;
- Salvar as associações na base de conhecimento, como instâncias de conhecimento.

A interface tem a dinâmica de uma tabela: em uma coluna são descritas as feições geológicas que são a entrada natural do sistema; na outra coluna o especialista indica os processos que devem ser inferidos de cada descrição. Dessa maneira foi desenvolvida uma *ferramenta de aquisição de conhecimento para associar produtos aos seus processos geradores*. A

Figura 5.3 mostra a interface desenvolvida para a aquisição de processos diagenéticos.

A

Figura 5.3(a) mostra a coluna *Constituent Identification* onde o especialista pode descrever todas as informações de entrada, ou seja, todas os constituintes que podem ser visualizados em uma amostra de rocha. Ele pode descrever combinando todas as características de um constituinte extensivamente. A coluna *Process* (



Figura 5.3(b)) foi adicionada na interface para que o especialista selecione de uma lista aquele processo que é o provável gerador do produto descrito na coluna (a). Por exemplo, a linha demonstrada na

Figura 5.3(c) descreve que quando o constituinte for um quartzo detrítico e aparentar fraturamento, o processo que gerou essa fratura foi o de compactação.

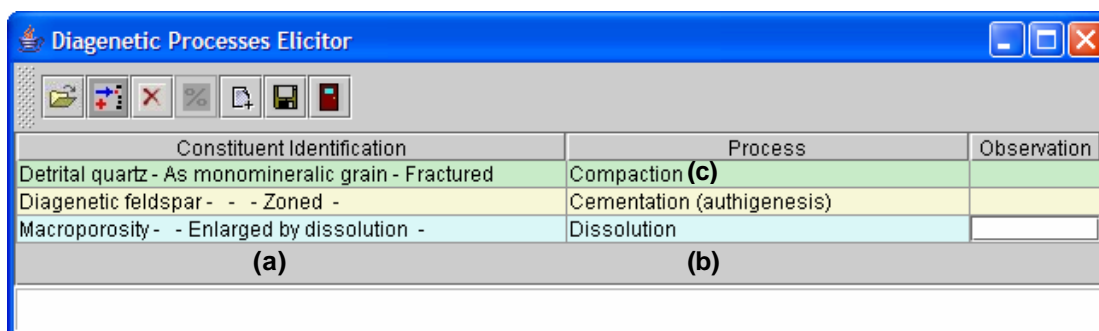


Figura 5.3: Interface para aquisição das associações entre produtos e processos.

O interessante dessa ferramenta era conseguir definir alguns regras gerais entre produtos e processos, ou seja, regras que digam que para toda uma determinada categoria de produtos que apresentam uma característica em comum, um processo pode ter sido o provável gerador daqueles produtos.

A princípio, pediu-se para o especialista deixar os constituintes não diferenciados, para não precisar escrever tão extensivamente. Isso funcionaria, pois pelo que se percebeu, ele repetia as mesmas informações para cada diferenciação do constituinte. O problema é que o especialista não consegue imaginar um domínio formado por constituintes não diferenciados, assim, ele não consegue descrevê-los dessa maneira. Então, essas regras terão que ser extraídas automaticamente a partir dos dados gerados pela ferramenta.

Uma parte da entrevista transcrita aqui mostra como o especialista tem dificuldade em imaginar padrões para os constituintes que descreve:

**Engenheiro de Conhecimento:** *qualquer constituinte detrítico que tenha o valor recristalizado como relação paragenética apareceu na rocha por processo de recristalização?*

**Especialista:** *obviamente...*

**Engenheiro de Conhecimento:** *Então, esse seria o caso de criar uma regra genérica para os constituintes detríticos?*

**Especialista:** *não porque nem todos os constituintes detríticos passaram por processo de recristalização!*

**Engenheiro de Conhecimento:** *Mas a regra definiria que QUANDO o constituinte detrítico apresentar essa informação, é porque ele passou por um processo de recristalização.*

A ferramenta continuou sendo utilizada pelo especialista, mas sem a presença do engenheiro de conhecimento.

### 5.2.10 Fase 9: Técnica para Associar Relações Visuais a Relações Temporais

Nas entrevistas anteriores, identificou-se que o especialista usa basicamente uma informação descrita sobre os constituintes da amostra para inferir a seqüência em que eles apareceram. Essa informação é a *relação paragenética* entre os constituintes. A

partir da análise de protocolos sobre as entrevistas, foram identificadas regras que permitem definir relações temporais, tais como:

*Se um Constituinte-1 cobre um Constituinte-2,  
Então o processo que gerou o Constituinte-1 veio depois do  
processo que gerou o Constituinte-2.*

A informação que um constituinte cobre o outro é a relação paragenética entre os constituintes. E a partir dela se pode inferir a ordem entre os processos que os geraram. Foram apresentadas para o especialista essas frases coletadas das entrevistas, em que ele descrevia uma relação paragenética e depois indicava a ordem entre os processos. Dessa maneira ele pôde compreender qual a informação era necessário eliciar.

Enfim, para poder eliciar todas essas regras de inferência sobre a ordem entre processos, foi apresentada uma listagem de todas as relações paragenéticas que podem ser descritas e pediu-se para o especialista indicar qual a relação de ordem inferida. Os resultados completos são mostrados no Apêndice A. Dessa forma, foi possível eliciar todas as relações temporais possíveis de serem inferidas pelo especialista.

### 5.2.11 Fase 10: Eliciação Triádica

As técnicas de eliciação triádica sugerem que sejam escolhidos conceitos do domínio de três em três. Se forem muitos conceitos, é necessário fazer um tipo de filtragem, pegando tríades de conceitos mais próximos. Se forem poucos conceitos, pode-se fazer extensivamente. Para cada conjunto de três conceitos, o especialista determina sua posição em uma escala, que geralmente vai de 0 a 10, conforme a similaridade e a diferença entre os conceitos *em relação a algum critério*. O trabalho do engenheiro é eliciar os critérios que o especialista usou para diferenciar os conceitos.

A técnica de eliciação triádica foi utilizada, então, utilizando os processos como os conceitos que deveriam ser posicionados na escala. Entretanto o especialista não conseguiu posicioná-los três a três, afirmando que não era essa a maneira de agrupar os processos. Ele acabou organizando todas as fichas de processos segundo uma hierarquia de categorias de processos, mostrada no Apêndice A.

Então a técnica foi aplicada novamente, mas de uma maneira mais dirigida: o especialista deveria selecionar os processos de três em três e indicar sua similaridade ou diferença em relação especificamente *ao produto que geravam*. Ou seja, se dois processos geravam o mesmo tipo de produto, era similares. Dessa maneira, seria possível eliciar alguns produtos que não haviam sido detectados até o momento com as outras entrevistas.

Entretanto, mais uma vez a técnica se mostrou inadequada, pois o especialista afirmou que não conseguiria, na verdade, selecionar três processos com produtos em comum, porque mais do que três processos pertenciam a uma 'categoria' que geraria mais ou menos o mesmo produto. Ele explicou que todos os processos de uma determinada categoria vão gerar produtos do mesmo tipo.

O problema da técnica é que ela serve para eliciar critérios, atributos de conceitos do domínio. Mas ela estava sendo aplicada com o objetivo de eliciar associações entre informações, ou seja, regras.

A técnica triádica foi abandonada, mas tentou-se por um outro caminho. Para cada processo de uma determinada categoria o especialista foi questionado sobre que tipo de produto aquele processo gerava. Foi possível, finalmente, achar um método para eliciar produtos ainda não conhecidos. Por exemplo, o especialista afirmou que o processo de Dissolução sob Pressão gera contatos intergranulares de determinado tipo, que são descritos no atributo Fábrica, que até o momento não estava sendo considerado como relevante para a inferência.

### 5.2.12 Fase 11: Grades de Repertório

As grades de repertório são técnicas usadas em conjunto com as triádicas. Primeiro se selecionam os critérios usando eliciação triádica, e depois todos os conceitos do domínio são cotados em relação a esses critérios, usando uma escala bipolar de valores (em geral, uma característica e seu oposto).

Primeiramente, o objetivo da técnica de grades de repertório tinha sido tentar eliciar associações entre os processos diagenéticos e um tipo específico de informação descrita sobre os produtos: a *localização*, que indica, como o próprio nome diz, o local em que o constituinte aparece na rocha (entre os grãos, nos contatos dos grãos, etc.). O especialista deveria utilizar um critério do tipo: quando determinado processo ocorre é possível gerar um constituinte com essa localização. Ou seja, a grade seria usada para coletar possíveis regras de associação entre as localizações dos constituintes e os eventos que tinham gerado esses constituintes.

Entretanto, durante o preenchimento da grade, foi descoberta uma incompreensão do domínio, por parte do engenheiro de conhecimento: os eventos diagenéticos não influenciam nas localizações dos constituintes. Um constituinte ocupa uma localização na rocha, antes de um evento acontecer. Essa localização em geral se mantém depois da ocorrência do evento, ou, muito raramente, é modificada por ele. Mas nunca uma localização é totalmente gerada por um processo. O que acontece é que um mesmo constituinte, em localizações diferentes da rocha passa por eventos diferentes, mas justamente porque sua localização era diferente. Então o nome do evento é diferente em função da localização do constituinte, e não o contrário.

A compreensão dessa informação foi importante, mas não fazia mais sentido aplicar a técnica de grades de repertório. Por isso foi montada uma nova experiência usando a informação de *modificadores* no lugar da localização. O modificador é a informação que indica de que maneira o constituinte em sua forma original foi modificado (*fraturado, dissolvido, etc.*). É uma informação que pode indicar quase diretamente o evento diagenético que gerou ou modificou o constituinte.

Então na primeira linha da grade são apresentados todos os eventos. Na primeira coluna da grade são apresentadas todas as possíveis modificadores de um constituinte na rocha, por exemplo, *deformado*. A grade ficou como mostrado parcialmente na Tabela 5.1:

Tabela 5.1: Grade de repertório para cotar eventos em relação aos modificadores dos constituintes que geram.

Eventos Diagenéticos	Dissolução	Cimentação	Substituição	Compactação	Dissolução por pressão	Recristalização	Fraturamento
Feições (Modificador)							
deformado							
fraturado							
dissolvido							
...							

Para coletar esse tipo de associação não foi possível usar a escala bipolar de valores, pois para o especialista a ligação é direta: um constituinte que apresenta um modificador *deformado* está associado ao processo de *compactação*.

A grade foi preenchida, mas as informações mais importantes foram eliciadas independente da grade. A grade completa preenchida na sessão de entrevista é mostrada no Apêndice A.

### 5.2.13 Fase 12: Limitação de Informações (Focalizando Contextos)

O especialista, quando realiza uma interpretação de seqüência de eventos diagenéticos, pode visualizar a amostra de rocha ao microscópio. Considerando que o mecanismo de inferência só vai poder utilizar as descrições das amostras, feitas pelo especialista, pensou-se em pedir que o especialista realizasse uma interpretação tendo disponível somente uma descrição de amostra de rocha.

Primeiro foi realizada a descrição da amostra *WINT-01 Plug 11* utilizando o software de contagem de pontos do sistema *PetroGrapher*. Sabendo que a descrição seria a única fonte de informação para ordenar os eventos, e especialista já pareceu descrever a amostra de maneira diferente da habitual. A descrição foi mais completa, no sentido que todos os constituintes foram descritos, o que geralmente não ocorre quando a descrição tem outros objetivos. Além disso, o especialista pareceu dar mais importância para a descrição de somente um dos três tipos de constituintes de rocha: os diagenéticos. Questionado sobre isso, o especialista argumentou que os constituintes detríticos não foram gerados por eventos diagenéticos. Todos os detríticos foram depositados antes dos eventos diagenéticos. O que ocorre é que eles podem ser modificados por evento, por isso os modificadores indicam tão fortemente os eventos diagenéticos. A macroporosidade primária, por ter vindo junto à deposição, também não indica diretamente nenhum processo. Uma ilustração da descrição dessa amostra é mostrada no apêndice A.

Para simular a interpretação da ordem dos eventos, foi utilizada a seguinte abordagem:

- os nomes dos constituintes descritos foram escritos em um papel separado, e recortados como fichas, em um método semelhante ao de classificação de fichas.
- a descrição da amostra foi lida procurando pelas relações paragenéticas, que é relação visual entre os constituintes que indica a ordem relativa em que eles apareceram na rocha. Por exemplo, para a relação paragenética <quartzo, cobrindo, clorita>, isso significa que o quartzo apareceu na rocha depois da clorita. Seguindo adiante na descrição, aparece a relação <esmectita, substituída-por, clorita>. Isso significa que a esmectita veio antes da clorita.
- Todas as relações paragenéticas foram lidas até que foram gerados alguns subconjuntos ordenados de todos esses constituintes, mas não se conseguiu chegar a uma ordem completa, uma seqüência linear.

A conclusão a que se chegou com a aplicação da técnica de limitação de informações foi que a descrição de uma amostra de rocha não fornece informações suficientes para a interpretação da seqüência completa. Algumas relações paragenéticas não foram visualizadas nessa amostra, por isso não foi possível relacionar alguns constituintes, tais como, entre a calcita e a clorita, entre a albita e a clorita e entre o feldspato e a clorita. O especialista chegou à conclusão que seriam necessárias mais

algumas amostras descritas para fornecer as relações paragenéticas que estavam faltando. Essa conclusão é importante para o funcionamento do software que vai realizar a interpretação, pois a interpretação feita com os dados de uma amostra vai ter um resultado diferente da interpretação baseada em várias amostras.

Esse procedimento foi repetido utilizando outras amostras para tentar coletar as relações paragenéticas que estavam faltando. Os resultados intermediários e a seqüência final da focalização do contexto sobre essas amostras são mostrados no apêndice A.

### 5.2.14 Fase 13: Eliciar Propriedades das Relações

Além da motivação explicada no capítulo 7 para se eliciar as propriedades das relações do domínio, uma vantagem dessa eliciação é que ao conseguir mapear o comportamento das relações do domínio para propriedades formais, o engenheiro de conhecimento aumenta sua compreensão sobre o domínio.

Mas para eliciar as propriedades das relações paragenéticas não era possível perguntar sobre elas diretamente para o especialista. Então foram criadas perguntas sobre o comportamento das relações paragenéticas que atingissem seu nível de vocabulário. Por exemplo, para descobrir se a relação *cobrindo(A,B)* é transitiva, foi elaborada a pergunta: “se um constituinte A cobre um constituinte B, e se esse constituinte B cobre outro constituinte C, isso necessariamente significa que o constituinte A também está cobrindo o constituinte C?”. Algumas das perguntas foram auxiliadas com o uso de cartões, representando os constituintes, e simulando visualmente a relação paragenética, para facilitar a compreensão do especialista. A transcrição de algumas dessas perguntas é mostrada no apêndice A. A partir dessas perguntas (entre outras), as seguintes observações foram coletadas:

1. Um constituinte A pode se relacionar com outros dois constituintes B e C, mas não é possível definir uma relação direta entre os constituintes B e C.
2. Uma porosidade pode ter sido gerada pela dissolução de mais de um constituinte (dissolução de(P, A)) ou pode estar dentro de (dentro de(P, A)) mais de um constituinte.
3. Um constituinte pode cobrir/substituir mais de um constituinte.
4. Um constituinte nunca vai cobrir ele mesmo, ou ter outras relações com ele mesmo. Ou ele é um só constituinte, e assim não justifica haver relações visuais, ou ele é um constituinte diferente. Isso indica que as relações não serão reflexivas.
5. Em geral, um constituinte sempre tem uma relação visual com outro constituinte da amostra. É raro um constituinte estar isolado na rocha.
6. Se um constituinte A cobre/engolfa um constituinte B, que por sua vez cobre/engolfa um constituinte C, isso não necessariamente significa que A esteja cobrindo/engolfando C.

A partir da observação (i), podemos definir que a propriedade Euclidiana não se aplica para relações visuais. As observações (ii) e (iii) mostram que um mesmo elemento do domínio pode se relacionar com mais de um outro elemento do domínio através dos relacionamentos específicos de *dissolver*, *dentro de*, *cobrir* e *substituir*. A partir da observação (iv) podemos identificar que as relações visuais provavelmente não serão reflexivas, pois os constituintes das relações são gerados pelos processos, e se cada processo sempre gera constituintes diferentes então as relações visuais também serão sempre entre constituintes diferentes. A observação (v) indica que as relações

visuais apresentam a propriedade de serem *seriais*, pois para cada constituinte de rocha ( $x$ ) sempre existe outro constituinte ( $y$ ) com o qual  $x$  está relacionado. A observação (vi) indica que as relações não são necessariamente transitivas.

Baseando-se nessas observações, foi elaborada a Tabela 5.2, que mostra um resumo das propriedades de algumas das relações visuais do domínio da petrografia, que são os relacionamentos entre os constituintes de uma rocha. São analisadas as propriedades das relações gerais (uma relação paragenética/temporal entre A e B, independente de qual for, é sempre reflexiva? É sempre simétrica? Ou não se aplica?) e da inferência entre relações paragenéticas e relações de ordem. Além disso, são detalhadas as propriedades das relações paragenéticas específicas (por exemplo, um constituinte A *cobre* um constituinte B, quais as propriedades?).

Tabela 5.2: Algumas propriedades de relações do domínio da petrografia.

Relação	Reflexiva	Simétrica	Serial	Transitiva	Euclidiana	Funcional	Linear	Total
<b>Relações Gerais</b>								
relação-paragenética(A, B)	N	S	S	N	N	N	N	S
relação-temporal(E1, E2)	S	S	S	S	N	N	N	S
produzido-por(C, E)	N	N	S	---	---	S	S	N
<b>Relações entre Porosidades e Constituintes</b>								
Dissolução de (P, A)	N	N	N	---	N	N	N	N
<b>Relações entre Constituintes</b>								
Cobrindo(A, B)	N	N	N	N	N	N	N	N
Alternado com(A, B)	N	S	N	N	---	N	---	N
<b>Relações Temporais entre Eventos</b>								
antes(E1, E2)	N	N	N	S	N	N	N	N
depois(E1, E2)	N	N	N	S	N	N	N	N
durante(E1, E2)	S	S	N	S	S	N	S	N

A tabela completa de propriedades de relações, descrevendo todas as relações paragenéticas, está representada no apêndice A.

### 5.2.15 Fase 14: Análise de Protocolos

Primeiramente, todas as sessões com o especialista foram transcritas do formato de áudio para documentos organizados por datas. As anotações feitas em papel durante as entrevistas foram passadas para o documento também, na posição do texto a que elas se referiam, como reflexões feitas entre os diálogos.

Procedeu-se com a análise de protocolos sobre as transcrições utilizando uma ferramenta de análise que faz parte da suíte *PCPack*, desenvolvida pelo sítio *Epistemics* (SHADBOLT e SWALLOW, 1993) para dar suporte à aquisição, análise e validação do conhecimento.

A de apoio à análise de protocolos, chamada *Protocol Tool*, é utilizada para analisar o texto de entrevistas utilizando marcadores que representam diferentes tipos de conhecimento, tais como, entidades, atividades, regras, etc., que podem ser adaptados para o tipo de texto específico. A análise procedeu então com a identificação e categorização das proposições e organização semântica dos segmentos do texto transcrito das entrevistas. A ferramenta ainda foi utilizada para identificar as regras de ordenação de processos. Utilizou-se um marcador específico, somente para marcar quando as sentenças formavam regras que permitiam inferir o aspecto temporal, como mostrado na Figura 5.4.

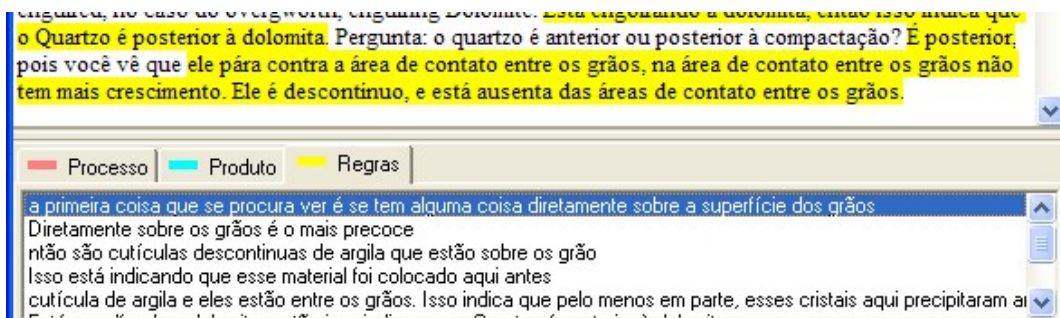


Figura 5.4: Imagem de parte da tela principal da ferramenta de análise de protocolos da suíte PCPack mostrando as marcações de regras temporais.

A ferramenta foi também adaptada para que os marcadores registrassem três tipos de construtos: *processos*, *produtos* e *regras*, como pode ser visto na Figura 5.5.

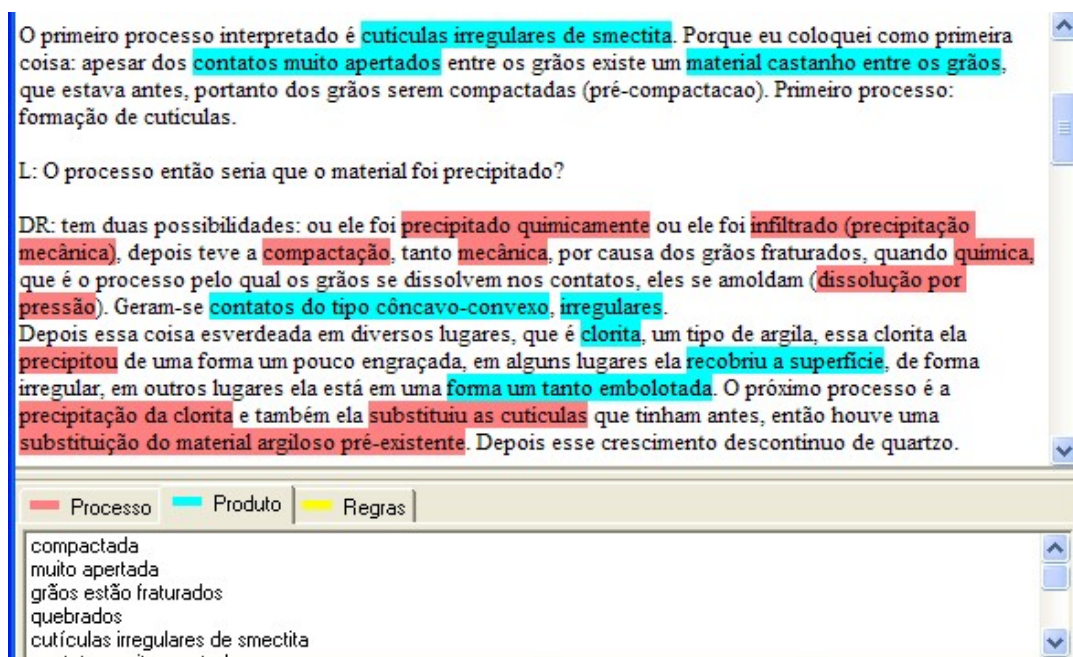


Figura 5.5: Imagem da tela principal da ferramenta de análise de protocolos PCPack da suíte mostrando as marcações de produtos e processos.

Dessa maneira foi possível separar facilmente os conceitos que foram utilizados posteriormente em outras sessões de aquisição de conhecimento. Por exemplo, antes da aplicação da classificação de fichas, essa ferramenta foi usada para coletar somente as palavras que representavam processos e relações paragenéticas dos textos, utilizando marcadores específicos para isso.

Outra função característica da suíte PCPack é que todos os construtos eliciados são salvos na base de conhecimento da ferramenta e podem ser usados em outros módulos da suíte, como por exemplo, a *Diagram Tool*, para criar diagramas em estilo esquema de domínio. A Figura 5.6 mostra que todos os construtos coletados com a ferramenta de análise de protocolos estão disponíveis para serem usados na elaboração do diagrama: os construtos disponíveis aparecem na parte direita da tela e os construtos já desenhados aparecem na parte esquerda da tela. Essa ferramenta foi importante para elaborar as fichas do processo de classificação de fichas.

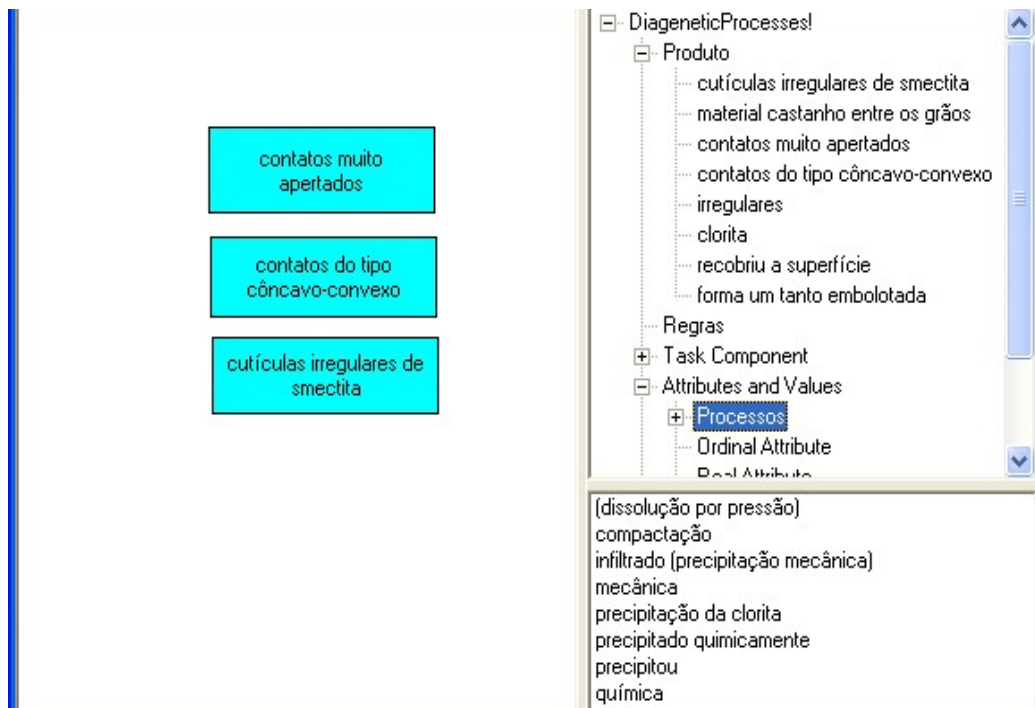


Figura 5.6: Imagem da da ferramenta de construção de diagramas da suíte PCPack.

### 5.3 Resultados do Processo de Aquisição de Conhecimento

Os resultados de um processo de aquisição de conhecimento são os métodos, objetos, relações que o especialista usa na tomada de decisão para resolver o problema. O processo de AC também permite avaliar dificuldades das próximas etapas da engenharia de conhecimento, como por exemplo, se as informações usadas na tomada de decisão são viáveis de serem eliciadas (o especialista usa cheiro, tato para solucionar o problema).

Essa seção descreve o conhecimento que foi efetivamente adquirido através do processo de aquisição de conhecimento e a viabilidade de modelagem desse conhecimento como um sistema. Além disso, mostra as conclusões tiradas sobre a avaliação de técnicas de AC.

#### 5.3.1 Conhecimento Resultante do Processo de AC

Através de entrevistas, observações e outras técnicas mais estruturadas de aquisição de conhecimento (AC) aplicadas ao especialista em Petrografia foram definidas as entradas para o método de interpretação de seqüência de processos diagenéticos e extraídas algumas regras de ordenação de processos.



As entradas são praticamente todas as informações qualitativas sobre os constituintes que formam a rocha. Além próprio constituinte, são os hábitos, as localizações, os modificadores e as *relações paragenéticas* entre os constituintes. Todas essas informações já são entradas atualmente pelo usuário do sistema *PetroGrapher* quando descreve uma amostra de rocha.

Relação paragenética é a relação no tempo e espaço entre as fases constituintes de uma rocha, normalmente entre os diferentes minerais que a constituem. As relações paragenéticas entre os constituintes minerais de um arenito são normalmente definidas com base na forma e na disposição espacial dos constituintes uns em relação aos outros e com relação à compactação e a porosidade, aspectos esses definidos na petrografia. É uma informação visualmente semi-interpretada pelo geólogo. Por exemplo, para descrever que “o mineral anidrita *cobre* o mineral dolomita”, o geólogo precisa identificar visualmente que um mineral aparece sobre o outro na amostra.

Ao final das etapas de aquisição de conhecimento, foi possível eliciar os principais construtos ontológicos que vão estar envolvidos na tarefa de interpretação computacional da seqüência diagenética:

- Feições do domínio: construto que representa as características possíveis de serem descritas objetivamente no domínio, que podem ser descritas como entrada na aplicação. Neste modelo as feições são os três principais constituintes da rocha, constituintes detríticos, constituintes diagenéticos e porosidades;
- Associações entre feições: construto que representa os relacionamentos entre as características do domínio que são essenciais para a interpretação da seqüência em que as características foram geradas. Neste modelo são as relações paragenéticas. Exemplo: <constituente diagenético> cobre <constituente detrítico>;

Os dois primeiros construtos são as entradas do algoritmo, ou seja, as informações descritas pelo usuário. Ao contrário dos construtos seguintes que são as saídas do método de raciocínio, ou seja, as informações interpretadas.

- Eventos: são os eventos que geraram as características atuais do domínio. Neste modelo, são os processos diagenéticos: compactação, dissolução, precipitação.
- Relações temporais: são as relações tais como antes, depois, durante que são inferidas com base nas características visuais do domínio.
- Regras extraídas: os textos das sessões de observação de interpretação de seqüência diagenética foram transcritos e as partes em que o especialista definia uma interpretação temporal foram traduzidas para regras em um formato quase de linguagem natural, mas estruturadas com premissas e conclusão e utilizando os termos do domínio da geologia, descritos na ontologia de domínio do *PetroGrapher*. Como verificado anteriormente, as informações de entrada do processo de interpretação de seqüência já estão descritas na ontologia de domínio do sistema. Por isso as premissas das inferências do especialista podem ser traduzidas diretamente para o esquema já descrito na ontologia. Mas nessa etapa da aquisição, a conclusão das regras (ou seja, da definição da seqüência temporal) ainda não tinha uma estrutura definida. Alguns parágrafos foram extraídos dos relatórios da observação e traduzidos em formato de regras. Essas regras são apresentadas detalhadamente no apêndice A. Pode-se notar que essas regras associam

temporalmente os constituintes, e não os processos, ainda como consequência da lacuna de não verbalização entre essas duas informações.

A modelagem desses construtos para o desenvolvimento de uma ontologia será detalhada no capítulo 6.

### 5.3.2 Avaliação Geral das Técnicas de AC

Na Tabela 5.3 são descritas e analisadas as vantagens e desvantagens das técnicas de AC aplicadas neste trabalho.

Tabela 5.3: Vantagens e desvantagens de métodos de AC.

Método	Vantagens	Desvantagens
Entrevistas não estruturadas	Útil para criar as conexões entre aquilo que já se conhecia previamente de Petrografia e o processo de interpretação seqüencial a ser representado Compreender que a maior parte das informações que é analisada para gerar uma seqüência já está sendo descrita atualmente no sistema <i>PetroGrapher</i> . Compreensão conceitual da tarefa de interpretação seqüencial A mais confortável para o especialista, pois ele podia falar livremente.	Trabalhosa de transcrever (transcrição de fitas para documentos e depois análise de protocolos)  Especialista perdia muito tempo em assuntos não essenciais
Entrevistas estruturadas	Levaram menos tempo que as não estruturadas Detalhado mais claramente que tipos de informações realmente serão usadas como entrada para o algoritmo de seqüência, que são basicamente as relações paragenéticas entre os constituintes.	Mesmo sendo mais eficiente, o trabalho de transcrição é o mesmo. Desvantagens de entrevistas em geral: só é possível eliciar as informações que são por natureza explicitáveis. Os principais problemas não aparecem durante as entrevistas.
Observação	O especialista se sentiu à vontade por trabalhar da maneira em que estava acostumado Foi possível ter uma visão total de como o especialista age, possibilitando definir os passos gerais do método de solução de problemas. Entender que a seqüência de processos diagenéticos formada é linear por falta de recursos para representar os processos que ocorreram em paralelo. Identificar informações importantes que não estão sendo descritas com entrada do sistema.	Pelo fato da observação ter pouca verbalização, foi necessário filmar, o que de certa maneira incomodou o especialista. Outra decorrência da pouca verbalização é que a linha de raciocínio do especialista não é revelada pelas suas ações, e por vezes é necessário interromper para fazer perguntas concorrentes. A segunda aplicação da técnica de observação, como descrito na seção 5.2.8, não obteve os resultados necessários, pois o conhecimento não era observável.
Classificação de fichas	A técnica foi rápida e fácil de ser aplicada. Foi possível compreender a diferenciação que o especialista faz entre duas informações fundamentais sobre o domínio (relações paragenéticas e processos). Foi possível também distinguir algumas das grandes categorias em que o especialista enquadra os processos.	A única desvantagem da classificação de fichas é sua limitação, pois só útil para eliciar esse tipo de conhecimento declarativo.
Ferramenta computacional de aquisição de conhecimento	As informações coletadas já estão em formato digital, o que facilita a análise. É possível gerar mais facilmente a combinação extensiva das informações. Em papel a complexidade seria maior. O especialista pode utilizar a ferramenta sem apoio do engenheiro, liberando tempo para outras entrevistas.	Perde-se mais tempo para preparar a técnica (deve ser implementada) É menos intuitiva para o especialista e sua aceitação depende de funcionar sempre corretamente. Quando ocorrem erros o especialista abandona o uso. Como o engenheiro não precisa estar presente, o especialista pode não ficar motivado a utilizar a ferramenta.

Método	Vantagens	Desvantagens
Técnica para eliciar relações temporais	<p>Como as outras técnicas não estavam sendo úteis para esse fim, foi necessário definir uma técnica que permitisse eliciar essas relações. A técnica atingiu o objetivo, as relações temporais foram eliciadas.</p> <p>Foi fácil e prática de ser aplicada, e foi bem compreendida pelo especialista que, inclusive, fez algumas sugestões para melhorar a apresentação.</p> <p>Essa técnica, por ter um formato de grade, poderia ser implementada facilmente como uma ferramenta, e alterada para testar outras informações como indicadores de relação de ordem.</p>	<p>A técnica é muito específica: é preciso saber exatamente quais os tipos de informação do domínio podem gerar relações de ordem.</p> <p>Além disso, a técnica foi aplicada para eliciar relações de ordem de somente uma informação do domínio (relação paragenética). Se a relação de ordem for determinada por mais de uma informação descrita a técnica deve ser adaptada.</p>
Eliciação Triádica	<p>Quando utilizada em domínios adequados, é eficaz para eliciar construtos para discriminação de conceitos.</p>	<p>A eliciação triádica não foi bem recebida pelo especialista, que não conseguiu se adaptar à separação dos conceitos de três em três, mais ainda quando a técnica foi dirigida para eliciar um tipo de informação.</p> <p>Técnica bastante específica para coletar um tipo de conhecimento: atributos que o especialista usa para diferenciar objetos. Difícil de ser adaptada.</p>
Grades de repertório	<p>O uso de uma técnica elaborada (<i>contrived</i>) fez aparecer uma má-compreensão sobre o domínio que ainda não havia sido detectada depois de várias entrevistas: nem todas as informações que qualificam um constituinte são geradas pelos eventos ou tem alguma associação com estes.</p> <p>Ao ser aplicada sobre a informação apropriada, a técnica permite saber como o especialista avalia os conceitos em relação a vários critérios.</p> <p>Técnica bastante maleável: foi fácil adaptá-la para o objetivo específico: achar a associação entre os processos e uma das informações que qualifica um constituinte (o modificador).</p>	<p>Não foi possível aplicar a grade da maneira indicada, usando uma escala bipolar de valores, porque para o especialista não fazia sentido.</p>
Focalizando contextos	<p>Permite avaliar o mínimo de informação necessária para que o especialista tome uma decisão</p> <p>Explicita decisões em casos reais</p> <p>Fácil de aplicar, pois em geral é uma tarefa habitual do especialista, mas reduzida em quantidade de informações ou tempo.</p>	<p>A preparação da técnica é o mais trabalhoso, pois exige que alguns casos (situações) já tenham sido eliciados do especialista anteriormente.</p>

## 5.4 Resumo do Capítulo

Este capítulo demonstrou a aplicação de algumas técnicas de eliciação de conhecimento para identificar, coletar e organizar o conhecimento necessário para a realização de uma tarefa específica do domínio da petrografia sedimentar: a interpretação de seqüência diagenética. A aplicação das técnicas teve um caráter de avaliação informal, pois foram descritas as vantagens e desvantagens de se aplicar cada uma das técnicas.

Mas o objetivo principal foi mesmo a eliciação do conhecimento, cuja parte declarativa foi representada como uma ontologia para representar os conceitos relevantes para a tarefa de interpretação, e cuja parte procedimental foi modelada como um método de raciocínio no nível de conhecimento, que indica os passos de inferência para se atingir o objetivo da interpretação. Essa modelagem será detalhada nos próximos capítulos.

## 6 ONTOLOGIA DE EVENTOS

A modelagem do conhecimento, depois de sua aquisição, é ainda um dos problemas-chave na Engenharia de Conhecimento. Sistemas baseados em conhecimento dependem de suas bases de conhecimento para realizar inferências sobre os dados do usuário do sistema. Logo, o núcleo de um SBC está em sua base de conhecimento, e a modelagem e implementação desta dependem de compreensão do domínio de aplicação e familiaridade com técnicas que permitam representar o domínio da maneira mais fiel.

Ontologias estão para os sistemas baseados em conhecimento assim como os esquemas conceituais para os sistemas de bancos de dados. Uma ontologia permite descrever a conceitualização de um domínio de maneira semântica e formal. Alguns trabalhos, atualmente, buscam estender as ontologias de maneira a representar informações temporais sobre os conceitos, utilizando como base as diferentes teorias de tempo existentes.

Este capítulo descreve uma proposta de construtos<sup>11</sup> temporais para estender linguagens de ontologias com a capacidade de representar eventos e relações temporais entre eles. Esses construtos possibilitam definir ontologias de eventos para a tarefa de ordenação dos eventos que geraram as características atuais do domínio.

Os construtos da ontologia foram definidos utilizando uma linguagem para representação de conhecimento de domínio: a linguagem CML (*Conceptual Modeling Language*) da metodologia CommonKADS. Essa notação foi escolhida por permitir representar na ontologia, junto aos conceitos e relações, as regras que definem as dependências entre os objetos da ontologia. A linguagem CML atualmente permite representar objetos (conceitos) do domínio e relações binárias entre eles. Com a extensão dessa linguagem será possível representar eventos que modificam os objetos do domínio e relações temporais entre esses eventos.

A definição de novos construtos ontológicos enquadra essa proposta como uma *ontologia de representação de conhecimento*, como definido na sessão 2.4.1.

### 6.1 Avaliação dos Quesitos Temporais do Domínio

O uso do tempo absoluto é predominante nas principais aplicações. É necessário registrar o tempo preciso das informações e tratar esse aspecto como mais uma dimensão para ser armazenada. Entretanto, no uso da linguagem natural, geralmente a idéia de temporalidade é transmitida através do uso de relações de *antes* e *depois*. Por isso, é importante conhecer as características de domínios de tempo relativo, como

---

<sup>11</sup> Neste trabalho as palavras *construto* e *primitiva* são empregadas com semânticas diferentes: construtos são os termos definidos em uma linguagem ou ontologia de representação de conhecimento para construir os objetos do modelo (por exemplo, os construtos *conceito*, *relação-binária* e *evento*). Primitivas são os objetos específicos do domínio modelados usando os construtos (tais como um conceito *aluno* ou uma relação *filho-de*).

frisado no capítulo 3. A seguir, são definidos os principais quesitos para representação de tempo, seguindo os tópicos apresentados no capítulo 3.

### 6.1.1 Requisitos Temporais

Segundo as definições detalhadas no capítulo 3, pode-se classificar os requisitos temporais deste domínio como sendo *parcialmente definidos*, que são aqueles em que um evento está associado a um elemento temporal *antes* ou *depois* do qual ele deve ocorrer, sendo esse elemento temporal um outro evento. Esse requisito se diferencia de um requisito bem-definido por não demandar uma definição absoluta do tempo.

### 6.1.2 Entidade Primitiva de Tempo

O estudo feito sobre as lógicas e os modelos de representação de tempo mostrou que a abordagem adotada pela maior parte dos modelos é correlacionar os eventos aos *intervalos* de tempo no qual eles ocorrem, e definir a relação temporal sobre os intervalos e não sobre os eventos. Dessa maneira, a relação temporal entre os eventos pode ser descrita implicitamente através da relação temporal entre seus intervalos de tempo. É possível, por exemplo, representar que o evento  $e_1$  ocorreu no intervalo  $t_1$ , e que o evento  $e_2$  ocorreu no intervalo  $t_2$ . Logo, a relação de tempo entre  $t_1$  e  $t_2$  vai indicar, implicitamente, a relação de tempo entre os eventos.

Na ontologia de RC baseada em eventos proposta neste trabalho, os eventos estão diretamente associados a uma relação temporal, e *não é representada uma primitiva de intervalo de tempo*. Isso porque essa ontologia será empregada em domínios em que o aspecto temporal absoluto não é tão significativo quanto a definição da relação de ordem entre seus objetos (os domínios de tempo relativo). Então não são utilizadas unidades absolutas de tempo, como por exemplo, entidades para representar data, hora, ano ou intervalo de tempo. O aspecto temporal do modelo é dado pelos relacionamentos temporais entre os eventos: *antes*, *depois* e *durante*. Esses relacionamentos possuem a mesma semântica das relações entre intervalos de Allen correspondentes.

### 6.1.3 Escolha da Ordem de Tempo

A escolha mais comum é que o tempo transcorra *linearmente*, ou seja, somente uma das seguintes expressões pode ser verdadeira:  $t < t'$  ou  $t' < t$ , onde  $t$  e  $t'$  representam pontos distintos no tempo e a ordem de precedência é representada através do operador  $<$ . Alguns sistemas, entretanto, precisam representar eventos que ocorrem concomitantemente, gerando uma ordem parcial entre os pontos de tempo (*linha de tempo ramificada*). A ordem de tempo da tarefa de interpretação de seqüência diagenética transcorre como uma *linha de tempo ramificada*, pois esse modelo precisa representar eventos que podem ocorrer em paralelo com outros.

### 6.1.4 Especificação da Estrutura do Tempo

Podem ser consideradas duas formas diferentes: tempo contínuo (que representa mais fielmente a natureza) e tempo discreto. Entretanto, neste domínio os eventos não possuem informações absolutas de tempo, logo, não existem unidades de tempo a serem mapeadas para um conjunto discreto ou contínuo. Por isso, a especificação da estrutura de tempo não se aplica para esse modelo.

### 6.1.5 Limitações do Tempo

A limitação é em relação a permitir que todo ponto tenha um predecessor/sucessor ou a especificar se os intervalos de tempo são abertos/fechados em uma ou nas duas pontas. Como esse modelo não utiliza pontos ou intervalos de tempo, a limitação não se aplica.

### 6.1.6 Introdução de uma Métrica de Tempo

A métrica de tempo só se aplica para modelos que utilizam instantes ou intervalos de tempo, pois define quesitos como duração, distância, granularidade e mapeamentos para outros sistemas de tempo.

## 6.2 Construtos da Ontologia de Eventos

Essa proposta apresenta um modelo para representar características do domínio, as associações entre essas características e construtos temporais para que esse modelo possa representar os eventos geradores dessas características e as relações de ordem entre os eventos.

### 6.2.1 Conceitos

Os principais conceitos que devem ser representados no modelo são as características do domínio. Essas características podem ter sido geradas ou alteradas por eventos, mas o fundamental é que seja possível indicar o evento ocorrido a partir das características do domínio.

As feições do domínio são definidas utilizando um construto já existente na linguagem CML: o construto *conceito*. Esse construto é especificado pelas suas relações de hierarquia com outros conceitos, por seus atributos e por axiomas. No modelo apresentado (Figura 6.1), é representado o conceito *feição* (que representa os objetos do domínio que tem sua formação associada a um evento) usando o construto conceito.

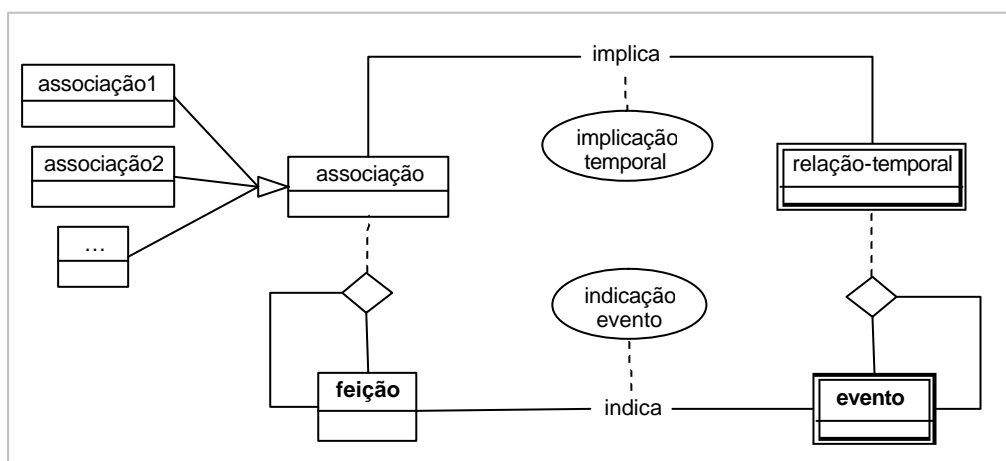


Figura 6.1: Ontologia de representação de conhecimento que propõe construtos para representar eventos e relações de ordem.

## 6.2.2 Eventos

O construto de representação de eventos está sendo proposto para que seja possível separar semanticamente os objetos do domínio dos eventos que os alteram. Eventos são acontecimentos que atuam sobre as feições do domínio mudando o estado do sistema através de produção de novas feições.

Para esse construto ser definido sobre a linguagem CML, foi necessário analisar a BNF dessa linguagem e estudar uma maneira de estendê-la. A extensão da BNF para aceitar o novo construto da linguagem é demonstrada a seguir:

```
evento ::=
    evento Evento ;
    [ super-tipo-de: Evento; ]
    [ sub-tipo-de: Evento; ]
    [ atributos ]
    fim evento [ Evento ; ] .
```

Cada evento tem um nome, um símbolo único que serve como identificador do evento e pode fazer parte de uma hierarquia de eventos

Dessa maneira é possível definir um evento específico do domínio utilizando um construto que o diferencia semanticamente de outros objetos do domínio. No capítulo seguinte esse construto será empregado para definir os eventos específicos do domínio.

Na Figura 6.1 é demonstrado a definição de um evento utilizando o construto *evento* proposto. A representação de classe UML foi alterada (com uma borda) para diferenciar o construto evento dos outros.

## 6.2.3 Relações

As relações expressam os relacionamentos estáticos entre os conceitos do modelo e são definidas através de seu nome e cardinalidade. No modelo, as relações associam conceitos e são utilizadas como primitivas que suportam a inferência. As relações propostas para o modelo são as associações entre feições e as relações de hierarquia entre os objetos. As associações entre feições são representadas como uma primitiva diferenciada das outras no modelo, e podem apresentar atributos e formar hierarquias. A relação de supertipo/subtipo não é representada como uma primitiva diferenciada no modelo, ela é indicada implicitamente: quando se define um conceito, é indicado de quais conceitos ele é subtipo ou supertipo. Essas relações são detalhadas a seguir.

### 6.2.3.1 Associações

É preciso representar no modelo as relações entre os objetos do domínio, pois em alguns domínios, a ordem de ocorrência dos eventos pode ser indicada através de algumas relações entre os objetos. Por exemplo, no estudo de testemunhos de gelo, apresentado no início do capítulo 3 deste trabalho, os componentes químicos tem uma relação de profundidade entre eles. No domínio da petrografia, os constituintes de rocha estão associados por distribuições espaciais na rocha. Essa relação é chamada *associação*. A associação entre feições foi definida usando um construto da linguagem CML: o construto *relação-binária*. A relação binária é usada para ligar quaisquer tipos de conceitos do domínio com outro conceito, formando uma dupla. Na Figura 6.1 é demonstrada a representação da relação de associação entre as feições do domínio usando o construto de relação-binária.

### 6.2.3.2 Supertipo/subtipo

São as relações que definem as hierarquias do domínio, e em outros modelos são conhecidas como relação *é-um*. Indica-se, na definição do objeto do domínio, de quais objetos ele é um supertipo ou um subtipo. Por exemplo, se existissem vários tipos de feições do domínio, seria necessário, na descrição de cada uma delas, indicar que seu supertipo é o conceito *Feição*; e na definição do conceito *Feição* é necessário indicar todas os conceitos que são subtipos de *Feição*. Na ontologia os relacionamentos de hierarquia são como os que ligam a *relação temporal* às sub-relações temporais *antes* e *depois*.

### 6.2.4 Relação temporal

Outro construto que precisa ser definido como extensão de CML é a *relação temporal* entre os eventos. A relação temporal precisa ser diferenciada semanticamente da relação binária para incorporar uma relação de precedência que equivale às relações intervalares de Allen (Allen, 1983) *antes* e *depois*.

As tuplas formadas a partir dessa relação serão a saída do algoritmo de inferência que vai ordenar os eventos. Os tipos de relação de ordem que precisam ser representados vão depender do domínio. Para alguns domínios será necessária somente a relação de *antes* e *depois*; para outros, pode ser necessário usar como referência as 13 relações intervalares de Allen. A seguir é definida a extensão da BNF da linguagem CML para aceitar o novo construto de relação temporal entre eventos.

```
relação-temporal ::=
    relação-temporal Relação-Temporal ;
    ordem: relação-ordem ;
    argumento-1 : Evento
    argumento-2 : Evento
    [ inverso : Relação-Temporal ; ]
    [ atributos ]
fim relação-temporal [ Relação-Temporal ; ] .

relação-ordem ::=
    antes |
    depois |
    durante .
```

A relação temporal é definida como um novo construto que cria uma ligação entre dois construtos *evento*. A ordem é definida pelo termo *relação-ordem*, que pode ser alterado conforme a necessidade de representação do domínio. Neste caso, as relações de ordem possíveis são *antes*, *durante* e *depois*. Dessa forma, é possível definir uma semântica particular para as relações temporais, com atributos diferentes e até mesmo indicar as relações temporais inversas de cada uma, assim como na lógica de Allen (*antes* é o inverso de *depois*, e etc.). A relação temporal entre os eventos também é demonstrada na Figura 6.1.

### 6.2.5 Tipos-Regra

Em praticamente todos os domínios em que se realiza inferência sobre eventos no passado, os eventos são indicados diretamente por características deixadas nos componentes do domínio. Por isso é necessário representar as dependências entre



eventos e conceitos que representam as feições geradas ou modificadas por esses eventos utilizando um tipo-regra, pois essa associação fará parte da inferência do sistema. As instâncias do tipo-regra indicação-evento, mostrado na Figura 6.1, vão identificar que, quando determinadas características do domínio foram observadas, isso significa que um evento específico ocorreu no passado para gerar aquela característica.

Além disso, o especialista consegue inferir a ordem em que os eventos ocorreram através das observações feitas sobre os componentes do domínio. No caso do estudo de testemunhos de gelo, a relação de ordem entre os processos climáticos é inferida através da profundidade em que os componentes químicos foram encontrados. No domínio da petrografia, as relações paragenéticas entre os constituintes indicam qual apareceu antes e qual apareceu depois. Essa dependência entre a *relação de associação entre feições* e a *relação temporal* também é representado como um tipo-regra (Figura 6.1).

### 6.3 Resumo do Capítulo

A ontologia de representação de conhecimento proposta tem o objetivo de servir como base conceitual para a definição de modelos de conhecimento para domínios específicos. Além dos conceitos e relações típicos do domínio, esta proposta define novos construtos semânticos para representar aspectos temporais do domínio, os quais são eventos e relações de ordem entre eventos. Para aplicar este modelo em um domínio específico, é necessário instanciar os construtos tipos-regra de indicação dos eventos a partir das feições, e de implicação da ordem de tempo a partir da associação.

Por esse motivo, a ontologia de RC baseada em eventos pode ser empregada tanto para definir a representação do conhecimento, quanto para dirigir o processo de aquisição de conhecimento, pois conhecendo os tipos de construtos que se deve identificar no domínio pode-se definir uma metodologia de eliciação focada para esse objetivo.

O capítulo seguinte mostra como a ontologia de eventos foi usada como base para definir um modelo específico para o domínio da petrografia, e como os construtos desta ontologia foram mapeados para os objetos específicos desse domínio.

## 7 PROPOSTA DE MODELO DE CONHECIMENTO PARA ORDENAÇÃO DE EVENTOS NA PETROGRAFIA SEDIMENTAR

Este capítulo descreve o processo de desenvolvimento de um modelo para representar o conhecimento necessário para a tarefa de *interpretação de seqüência diagenética a partir de descrições petrográficas de rochas siliciclásticas*. Esse modelo foi proposto baseado no *modelo de conhecimento* da metodologia CommonKADS, que propõe uma descrição independente de implementação do conhecimento envolvido em uma tarefa crítica de um domínio.

Como já detalhado na seção 2.5.1.4, o modelo de conhecimento é composto por três categorias de conhecimento: conhecimento de *tarefa*, conhecimento de *inferência* e conhecimento de *domínio*, sendo este último dividido em *esquema do domínio* e *base de conhecimento*.

Para definir as primitivas do conhecimento de domínio específico para esta tarefa da Petrografia foi usada como base a ontologia de representação de conhecimento baseada em eventos proposta no capítulo anterior, pois ela especifica os construtos necessários para representar e ordenar eventos. A seguir será detalhado como as três categorias de conhecimento do modelo de conhecimento foram instanciadas para o domínio da Petrografia.

### 7.1 Domínio de Aplicação

A *petrografia sedimentar* (apresentada no capítulo de introdução deste trabalho – capítulo 1) é o domínio de aplicação escolhido para esse trabalho. Como já explicado anteriormente, o foco de estudo deste trabalho é a *seqüência diagenética*, cujo objetivo é identificar a *ordem* em que os processos diagenéticos atuaram sobre a rocha. Para essa tarefa, a datação absoluta do instante em que o evento aconteceu não é tão relevante quanto a ordem. A interpretação de seqüência de eventos diagenéticos emprega as relações temporais de *antes* ou *depois* para imprimir o aspecto temporal aos eventos. Por esse motivo, podemos considerar essa interpretação como sendo sobre um domínio de tempo relativo.

As primitivas desse modelo complementam o modelo do conhecimento já definido para o sistema *PetroGrapher*: a ontologia de domínio original será estendida com primitivas para representar aspectos temporais sobre os conceitos, e o modelo de inferência será adicionado de um novo método de solução de problemas: a interpretação de seqüência diagenética.

## 7.2 Modelo do Conhecimento de Domínio

Esse modelo define os conceitos, eventos, relações-binárias, relações-temporais e tipos-regra específicos para o domínio da petrografia sedimentar, descritos na linguagem de CommonKADS. Foi definido um modelo para o *esquema do domínio* e a *base de conhecimento*, que juntos compõem a categoria chamada *conhecimento de domínio* do modelo de conhecimento.

### 7.2.1 Conceitos e Eventos

No caso do domínio da Petrografia Sedimentar, as características dos constituintes de rochas descritas pelos geólogos são as feições do domínio. Cada constituinte e seus aspectos foram produzidos ou modificados por um determinado evento, logo, todos os constituintes da rocha estão associados de alguma maneira aos eventos.

O primeiro objetivo do modelo é representar os eventos diagenéticos que ocorreram e as características visuais associadas a eles. Os constituintes são representados na ontologia usando o construto *conceito* de CML. Os eventos diagenéticos são definidos empregando o construto *evento* proposto no capítulo anterior. A Figura 7.1 mostra a definição gráfica dessas primitivas.

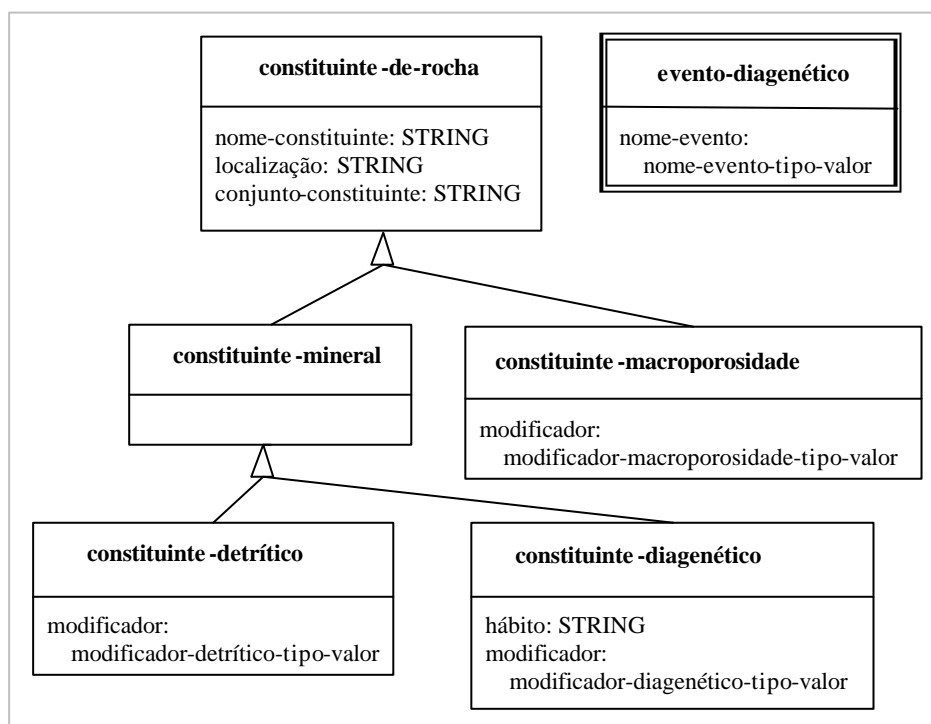


Figura 7.1: Descrição dos principais conceitos e eventos da ontologia de eventos: *constituente-de-rocha*, *evento-diagenético*, e os subtipos de *constituente-de-rocha*, *constituente-mineral*, *constituente-macroporosidade*, *constituente-detrítico* e *constituente-diagenético*.

No domínio da petrografia existem dois tipos de constituintes de rochas: os minerais e os poros das rochas (ou seja, espaços entre os minerais). Os minerais podem ser divididos em duas categorias: os *constituintes detríticos* (minerais que foram formados antes da deposição da rocha, e que, por isso, não são originados pelos eventos diagenéticos, mas sim, modificados por eles) e os *constituintes diagenéticos* (minerais formados exclusivamente por eventos diagenéticos). Os poros são chamados no domínio da

petrografia de *macroporosidade*. A Figura 7.1 mostra a modelagem dos constituintes minerais e constituintes de macroporosidade como subconceitos de um constituinte de rocha, e dos constituintes detríticos e diagenéticos como subconceitos de constituintes minerais, formando uma hierarquia de constituintes. A Figura 7.2 descreve o conceito *constituante-de-rocha* e o evento *evento-diagenético* usando a sintaxe CML estendida.

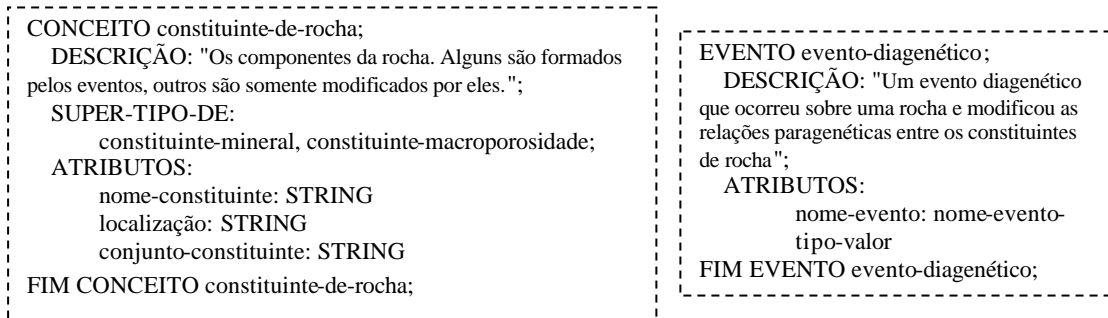


Figura 7.2: Descrição do conceito *constituante-de-rocha* e do evento *evento-diagenético*, empregando a sintaxe CML.

### 7.2.2 Relação Temporal

O resultado do processo de inferência é uma seqüência de eventos, apresentados na ordem em que ocorreram. Para que o sistema seja capaz de apresentar os eventos ordenados, é necessário definir a entidade que imprime o aspecto temporal sobre o modelo, que é o relacionamento de ordem entre os eventos. A Figura 7.3 demonstra como as entidades de ordem de tempo (antes, durante e depois) foram representadas no modelo gráfico do esquema do domínio.

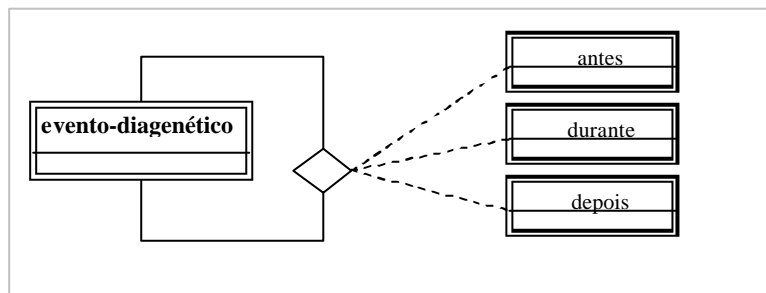


Figura 7.3: Descrição da relação temporal entre eventos diagenéticos.

Na Figura 7.4 é demonstrado como a relação temporal foi modelada usando o construto relação-temporal adicionado à sintaxe CML e como foi representada uma relação temporal específica, a relação de ordem *antes*.

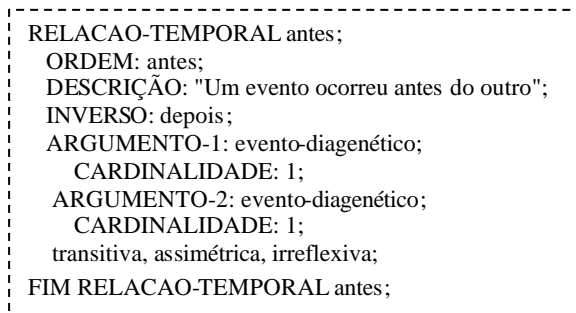


Figura 7.4: Descrição da *relação-temporal* e da sub-relação *depois*, empregando a sintaxe CML.

As relações antes e depois são definidas uma como a inversa da outra, como é possível fazer em CML. A relação de ordem antes é representada como uma relação temporal entre dois conceitos *evento-diagenético* ou seja, um evento tem uma relação de ordem com outro evento. Um trabalho futuro seria definir os argumentos das sub-relações temporais restringindo os eventos que podem vir antes, depois ou durante outros. Essas restrições de domínio não foram estudadas neste trabalho, mas a forma em que as relações temporais foram modeladas permite que essa semântica seja adicionada ao modelo.

As relações temporais *antes*, *depois* e *durante* possuem algumas propriedades lógicas. Por exemplo, a relação *antes*: um evento nunca acontece antes dele mesmo, logo ela não é reflexiva; se um evento acontece antes de outro isso significa que o segundo *não* ocorre antes do primeiro, logo não é simétrica; mas se um evento ocorre antes de um segundo, que ocorre antes de um terceiro, com certeza o primeiro ocorreu antes do terceiro, logo, é transitiva. As propriedades definidas são *transitiva*, *assimétrica* e *irreflexiva*.

As relações temporais modeladas (*antes*, *depois* e *durante*) permitem que seja definida como saída uma seqüência de eventos que pode apresentar ramificações, ou seja, uma seqüência não linear. Isso pode acontecer porque pode se chegar à conclusão de que um evento A ocorreu antes de um evento B ( $A \rightarrow B$ ) e que A também ocorreu antes de um evento C ( $A \rightarrow C$ ), mas sem saber o relacionamento temporal entre B e C. Isso gera uma seqüência como demonstrado na Figura 7.5.

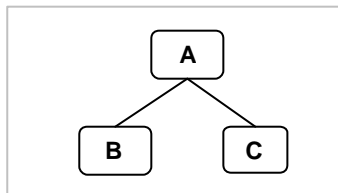


Figura 7.5: Seqüência de eventos em ordem de tempo ramificada.

### 7.2.3 Relações

As relações modeladas são as relações de supertipo/subtipo, as associações entre as feições de domínio e a relação produzido-por, que liga uma característica do domínio ao seu evento gerador. Como este modelo é uma extensão de uma ontologia previamente definida, é importante ressaltar que existem outras relações na ontologia completa, tais como as relações *parte-de*, que define uma partonomia entre as primitivas e a relação *restringe*, que limita a nomenclatura dos objetos do domínio. Depois das relações terem sido definidas na ontologia, foi necessário representar as *propriedades lógicas das relações*.

#### 7.2.3.1 Propriedades

Certas relações podem ser caracterizadas pelas suas propriedades. Uma *propriedade* é o construto pelo qual as relações podem ser distinguidas. As propriedades das relações de um domínio desempenham um papel fundamental na formalização da ontologia, pois indicam mais claramente a semântica daquela relação. Na Tabela 7.1 são resumidas as propriedades de algumas dessas relações somente para exemplificação. A tabela total é apresentada no apêndice A: são descritas as principais propriedades (propriedade

reflexiva, simétrica, linear, entre outras) das relações que serão representadas no domínio (todas as relações paragenéticas, relações temporais).

Tabela 7.1: Propriedades definidas para as relações do domínio (S=sim, N=não).

Relação	Reflexiva	Simétrica	Transitiva
relação-	N	S	N
dissolução-	N	N	---
coabrindo(A,	N	N	N
alternado-	N	S	N
produzido-	N	N	---

No exemplo da Tabela 7.1, o símbolo N é usado para indicar que a propriedade não se aplica à relação, enquanto o símbolo S indica que ela se aplica.

Algumas dessas propriedades são previstas na sintaxe de CommonKADS, para serem definidas nas relações, tais como, *transitividade*, *simetria*, *anti-simetria*, *assimetria*, *reflexividade* e *irreflexividade*. As outras propriedades, apresentadas na tabela completa no apêndice (Euclidiana, linear, serial, etc.) não são definidas na sintaxe.

A seguir são especificadas as relações do domínio, sendo definidas suas propriedades, quando possível, usando a sintaxe CML.

#### 7.2.3.2 *Supertipo/subtipo*

Definem as hierarquias de conceitos do domínio. Os constituintes de rocha formam uma típica hierarquia do domínio: existem constituintes minerais e porosidades. Os constituintes minerais podem ser detríticos ou diagenéticos. Na definição do conceito constituinte-de-rocha na Figura 7.2 está indicada a maneira de definir o relacionamento deste conceito com seus sub-conceitos: definindo que ele é um SUPERTIPO-DE constituinte-mineral e constituinte-macroporosidade.

#### 7.2.3.3 *Relação paragenética*

Quando um geólogo descreve uma amostra de rocha, ele faz algumas observações sobre como os constituintes se relacionam, indicando, por exemplo, que um constituinte foi formado cobrindo o outro, ou seja, ele aparece por cima do outro, o que um constituinte está substituindo o outro. Essa informação sobre a disposição visual entre os constituintes de uma rocha é essencial para a inferência sobre a ordem em que os constituintes foram gerados. A Figura 7.6 mostra a entidade criada para representar essa relação na ontologia.

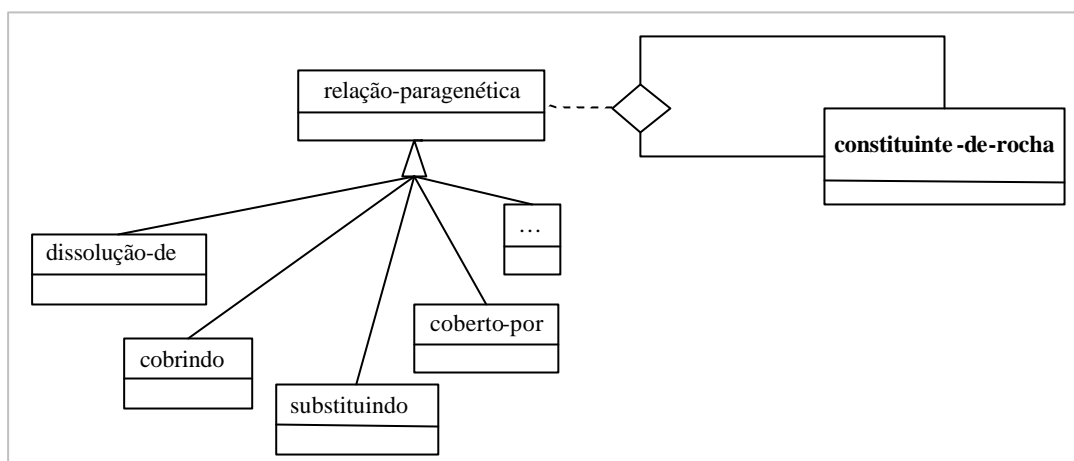


Figura 7.6: Descrição do relacionamento visual (relação paragenética) entre constituintes.

A relação paragenética possui diversos subtipos de relações, tais como os citados anteriormente (*substituindo*, *cobrindo*, etc.). Mas cada um desses relacionamentos ocorre no domínio entre diferentes constituintes. Por exemplo, a relação de *dissolução* só pode ocorrer entre um constituinte macroporosidade e um constituinte mineral, nessa ordem enquanto a relação *substituindo-grão-de* só ocorre entre um diagenético e um detrítico, também nessa ordem. Por isso foi necessário descrever os subtipos de relações paragenéticas especificando seus argumentos e indicando as relações inversas. Por exemplo, a relação *cobrindo* é inversa da relação *coberto-por*. A definição usando a sintaxe CML da relação paragenética e de um subtipo de relação paragenética (a relação *cobrindo*) é demonstrada na Figura 7.7.

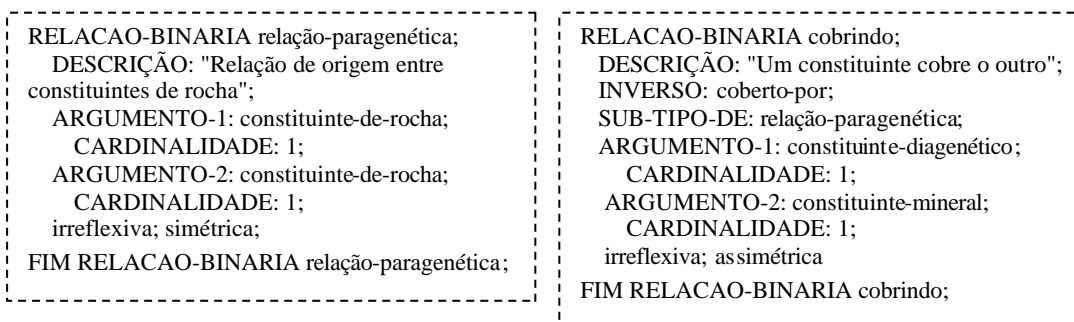


Figura 7.7: Descrição da *relação-paragenética* e da sub-relação *cobrindo*, empregando a sintaxe CML.

A relação paragenética não é reflexiva porque não existe relação paragenética entre um mesmo constituinte; é simétrica porque quando existe uma relação paragenética entre dois constituintes sempre existe uma no sentido inverso (em geral definida na voz passiva. Por exemplo, *cobrindo* e *coberto-por*); a transitividade da relação entre os constituintes não é garantida. Logo, as propriedades definidas na declaração CML da relação são: *irreflexiva* e *simétrica*. A relação *cobrindo* (sub-relação da relação paragenética) mantém a irreflexividade e a não-transitividade, mas, ao contrário de sua super-relação, ela não é simétrica, pois se um constituinte A está cobrindo outro constituinte B, com certeza B não está cobrindo A. As propriedades devem ser redefinidas para *irreflexiva* e *assimétrica*.

#### 7.2.3.4 Produzido-por

Essa é uma relação simples, sem atributos nem sub-relações, que objetiva somente concretizar a associação entre os constituintes e os eventos que os produziram. Essa informação é inferida através das regras de inferência (detalhadas na próxima seção, que explica os tipos-regra) mas é instanciada no modelo quando é criada a tupla que liga o constituinte ao evento. A Figura 7.8 mostra como a relação *produzido-por* é definida no modelo.

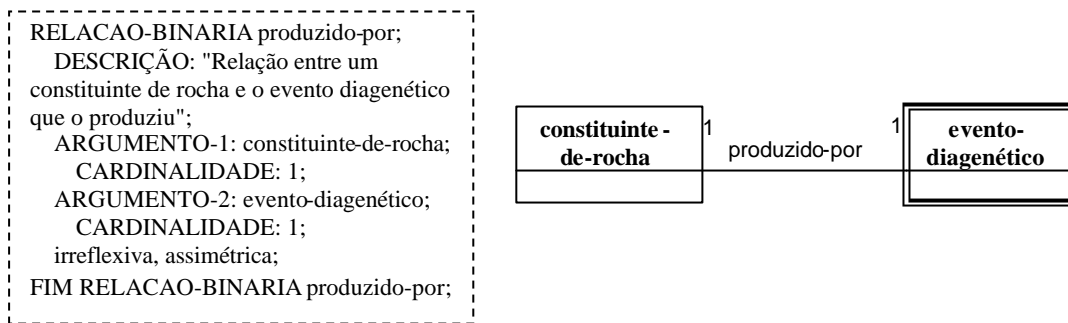


Figura 7.8: Descrição da relação *produzido-por* empregando a sintaxe CML e no modelo gráfico.

Para a relação *produzido-por* ser reflexiva, deveria ser possível dizer que um constituinte foi *produzido-por* ele mesmo, o que não ocorre. Para ser simétrica, ao dizer que um constituinte *C* foi produzido por um evento *E*, deveria ser possível dizer que *E* também foi produzido por *C*, o que não ocorre. Como consequência, a relação não é reflexiva (irreflexiva) e não é simétrica (assimétrica). Além disso, podemos afirmar que a transitividade não se aplica: para verificar a transitividade, deveria ser possível falar que quando *A produzido-por B* e *B produzido-por C* então *A produzido-por C*. Entretanto, a sentença *B produzido-por C* não faz sentido, pois neste domínio, *A* é um constituinte e *B* é um evento. E dizer que um evento foi produzido por algo não faz sentido. Logo, as propriedades definidas para essa relação na Figura 7.8 são *irreflexiva* e *assimétrica*.

### 7.2.4 Tipos-regra

A primitiva de tipo-regra será utilizada para representar o conhecimento sobre as regras que o especialista usa para inferir novas informações. Serão definidos dois tipos-regra: o que associa as feições do domínio com seus eventos geradores, e o que associa as associações entre feições com as relações de ordem.

#### 7.2.4.1 Regras de indicação de evento

O especialista em Petrografia, quando visualiza os constituintes de uma amostra de rocha, consegue saber quais foram os eventos geradores daqueles constituintes. Foi necessário representar as dependências entre eventos e constituintes gerados ou modificados por esses eventos utilizando um tipo-regra, pois essa associação fará parte da inferência do sistema. As instâncias dessas regras vão definir que quando determinado constituinte ou característica visual forem observados, isso indica que um evento específico ocorreu. A Figura 7.9 mostra a definição do tipo-regra que define a dependência entre constituinte e evento no esquema do domínio.



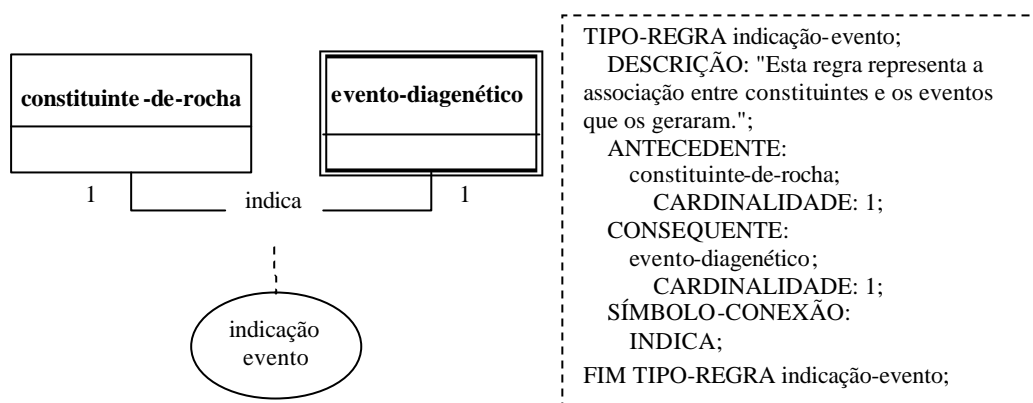


Figura 7.9: Descrição do tipo-regra que define a associação entre constituintes e eventos.

A Figura 7.10 mostra alguns exemplos de instâncias dessa regra na base de conhecimento. As instâncias dessas regras na base de conhecimento serão usadas como papéis de conhecimento estático pelo modelo de inferência para indicar os eventos que ocorreram. Essas instâncias relacionam o atributo modificador com o evento diagenético. Essas associações foram coletadas durante as seções de aquisição de conhecimento, e representam a maneira que o especialista infere qual o evento gerou o constituinte em questão.

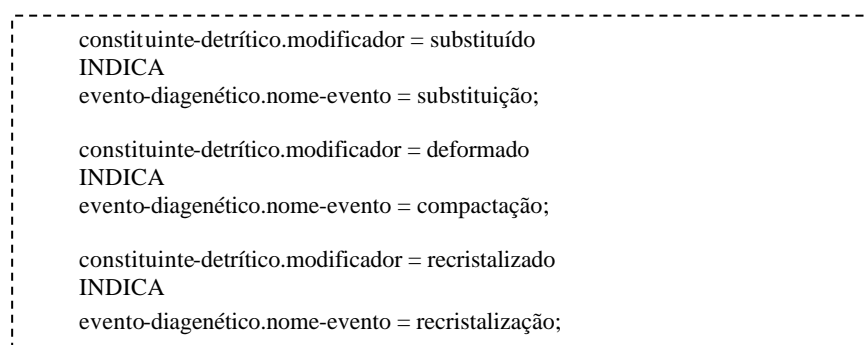


Figura 7.10: Exemplos de instâncias do tipo-regra indicação-evento: regras que indicam o evento diagenético a partir de características do constituinte visualizado.

A informação de qual evento gerou o constituinte é usada para definir uma relação *produzido-por* entre o constituinte de rocha e o evento diagenético em questão.

#### 7.2.4.2 Regras de implicação temporal

O especialista consegue inferir a ordem em que os eventos ocorreram através das observações feitas sobre os constituintes, tais como: “SE um mineral aparece cobrindo outro mineral, ENTÃO o evento que formou o primeiro mineral veio depois do evento que formou o segundo”. A primeira parte da observação (“um mineral aparece cobrindo outro mineral”) pode ser expressa no modelo através da relação paragenética entre constituintes. A segunda parte (“o evento que formou o primeiro mineral veio depois do evento que formou o segundo”) pode ser representada através da relação temporal entre eventos. A ontologia precisa poder representar essas regras, cujo antecedente é uma relação entre os constituintes, e cuja conclusão é a relação temporal. Elas são o conhecimento necessário para inferir a relação de ordem entre os eventos.

A dependência entre essas duas relações pode ser expressa através de um *tipo-regra*, em que uma instância do primeiro relacionamento implica em uma instância do segundo relacionamento, ou seja, é necessário representar *expressões de dependência entre relações*. Entretanto, por sua definição em CommonKADS, o tipo-regra representa uma dependência entre *conceitos* do domínio. Não é aceito pela BNF da linguagem CML a definição de uma expressão entre relações. E mesmo que fosse possível, como a relação temporal foi definida como um construto novo da linguagem CML, certamente não é possível definir um tipo-regra entre uma relação qualquer e uma relação temporal. Por isso, foi necessário modificar a BNF para suportar a nova relação proposta nesse trabalho. A BNF original que é utilizada para formar a sintaxe da notação CommonKADS é apresentada no anexo A<sup>12</sup>. Ela descreve as regras de formação do tipo-regra no esquema do domínio e das instâncias de um tipo-regra na base de conhecimento. Uma instância de tipo-regra em CML é formada basicamente por um termo do tipo *equação*, que pode derivar vários tipos de expressão, como mostrado a seguir:

```
equação ::=
  `( ` equação ` ) ` |
  operador-sinal equação |
  operador-negação equação |
  equação operador-artimético equação |
  equação operador-lógico equação |
  equação operador-comparação equação |
  equação operador-dereferência equação |
  equação operador-implicação equação |
  expressão-variável |
  expressão-função |
  expressão-condicional .
```

Entretanto, nenhuma dessas expressões possibilita derivar uma implicação entre relações. Por isso, para que essa BNF aceite definir instâncias de um tipo-regra entre relações, é necessário acrescentar as seguintes definições:

(i) na definição do termo *equação*, incluir uma alternativa para uma equação formada de uma expressão entre relações, representado pelo termo *expressão-relação*:

```
equação ::=
  ...
  expressão-condicional |
  expressão-relação .
```

(ii) definir uma *expressão-relação* como: (i) uma *relação* entre vários *construtos de relação*, (ii) uma *relação binária* entre dois *construtos de relação*, (iii) uma relação temporal entre dois construtos temporais. Um construto de relação pode ser um *Conceito* ou uma *Variável*, analogamente, um construto temporal pode ser um *Evento* ou uma *Variável*. Os termos *Relação*, *Relação-binária*, *Conceito*, *Variável*, *Relação-temporal* e *Evento* são nomes terminais definidos pelo usuário. A estrutura é como segue:

```
expressão-relação ::=
  Relação(construto-relação, ... ) |
```

<sup>12</sup> A BNF apresentada refere-se somente à sintaxe do construto tipo-regra. A BNF original está definida em inglês, e foi traduzida para o português neste trabalho para manter-se condizente com a língua do restante do texto.

```

Relação-binária(construto-relação, construto-relação) |
Relação-temporal(construto-temporal, construto-temporal).
construto-relação ::= Conceito | Variável .
construto-temporal ::= Evento | Variável .

```

Ou seja, na base de conhecimento, uma relação (ou relação temporal) pode ser definida usando conceitos (ou eventos), expressando um comportamento geral entre todas as instâncias daqueles conceitos (ou eventos), ou, se for necessário, pode-se instanciar variáveis para as relações ou relações temporais.

A expressão entre relações pode ser definida derivando-se uma equação. Por exemplo (cada linha apresenta um passo de derivação usando a BNF estendida de CML):

- i. equação
- ii. equação operador-implicação equação
- iii. equação ? equação
- iv. expressão-relação ? equação
- v. Relação-binária(Conceito1,Conceito2) ? equação
- vi. Relação-binária(Conceito1,Conceito2) ? expressão-relação
- vii. Relação-binária(Conceito1,Conceito2) ?  
Relação-temporal(Evento1,Evento2)

Onde as relações *Relação-binária* e *Relação-temporal* são nomes definidos no modelo de conhecimento de domínio, assim como *Conceito1* e *Conceito2* são nomes de conceitos e *Evento1* e *Evento2* são nomes de eventos, também definidos no modelo de conhecimento de domínio. A expressão definida cria uma relação de implicação entre duas relações do domínio. Quando a instância da primeira for detectada isso implica que deve ser definida uma instância para a outra relação. A BNF estendida será utilizada para a definição do tipo-regra entre relações-binárias, como demonstrado na Figura 7.11, que mostra a definição do tipo-regra implicação-temporal no esquema do domínio.

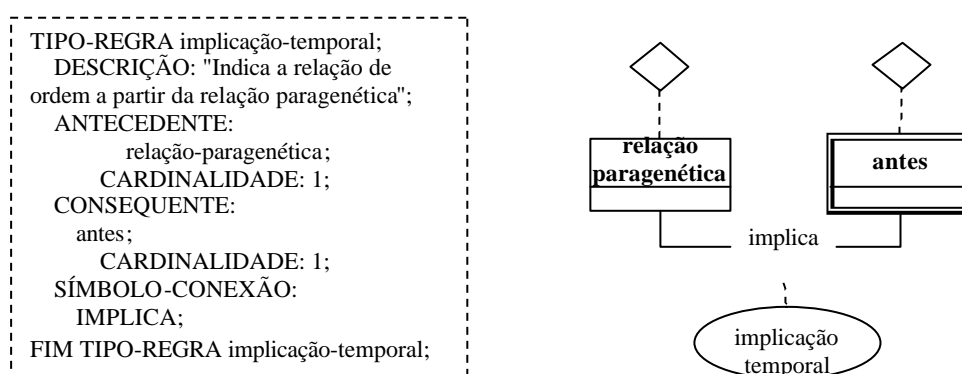


Figura 7.11: Definição do tipo-regra que retorna a relação de ordem entre constituintes a partir da relação paragenética entre eles.

A Figura 7.12 mostra algumas instâncias dessas regras na base de conhecimento, seguindo as regras de formação da sintaxe BNF modificada.

```

EXPRESSÕES:
dissolução-de(const-macrop, const-min)
E produzido-por(const-macrop, evento-diag1)
E produzido-por(const-min, evento-diag2)
IMPLICA depois(evento-diag1, evento-diag2);

cobrindo(const-diag, const-min)
E produzido-por(const-diag, evento-diag1)
E produzido-por(const-min, evento-diag2)
IMPLICA depois(evento-diag1, evento-diag2);

```

Figura 7.12: Exemplos de instâncias do tipo-regra implicação-temporal: a relação temporal é implicada a partir da relação visual.

As expressões apresentadas na Figura 7.12 podem ser lidas como (para a primeira expressão): se a relação paragenética entre um constituinte macroporosidade e um constituinte mineral é de dissolução, e se os eventos diagenéticos que produziram cada um dos constituintes é conhecido, então a relação temporal entre esses eventos é que o evento que gerou a macroporosidade é posterior ao evento que gerou o mineral. Uma interpretação similar pode ser dada para a segunda expressão, com a relação paragenética *cobrindo*.

O modelo completo do conhecimento de domínio proposto é demonstrado no Apêndice B. O modelo foi definido empregando a notação UML baseada na metodologia CommonKADS. A definição do conhecimento de domínio na sintaxe CML é apresentada na categoria CONHECIMENTO-DOMÍNIO, do Modelo de Conhecimento, também apresentado no Apêndice B. Nas próximas seções serão apresentadas as outras categorias de conhecimento do modelo de conhecimento: o conhecimento de tarefa e de inferência.

### 7.3 Modelo do Conhecimento de Inferência

Para representar um método de inferência que permita definir a seqüência de eventos que originou as características atuais de um domínio, foi especificado um modelo de inferência para a ordenação de eventos. Como já definido na seção 2.5.1.3, o conhecimento de inferência, descreve os passos de raciocínio básicos que podem ser aplicados sobre os objetos do domínio para se chegar aos objetivos da tarefa. O esquema de raciocínio para ordenação de eventos proposto neste capítulo é baseado no comportamento de solução de problemas de um especialista em Petrografia Sedimentar.

Como será mais bem detalhada na seção seguinte, a tarefa de interpretação de seqüência de eventos é subdividida em duas tarefas: a identificação e a ordenação desses eventos. Por isso foram definidos dois modelos de inferência: o modelo de inferência para *identificação de eventos* e o modelo de inferência para *ordenação de eventos*.

As estruturas devem ser percorridas progressivamente, ou seja, a inferência tem início com a entrada das feições do domínio, e segue o fluxo de dados e conhecimento até gerar uma ordem de eventos. Alguns modelos de inferência podem ser percorridos utilizando raciocínio regressivo, ou seja, a partir das hipóteses de solução avaliar as causas, mas não é este o caso. A seguir são detalhados os dois modelos do conhecimento de inferência, representados em modelos gráficos que mostram os papéis de conhecimento, os fluxos de informação e as inferências, que transformam um papel em outro.

### 7.3.1 Inferências para Identificação de Eventos

O processo de identificação de eventos é representado pelo modelo gráfico da Figura 7.13.

Esse modelo pode ser explicado da seguinte forma: um objeto do domínio é analisado e são identificadas algumas características relevantes para a interpretação: as *feições* (1) do domínio. Como as feições são analisadas individualmente, deve ser selecionada uma *feição* (2). Essa seleção em geral segue uma ordem seqüencial. Uma feição do domínio apresenta características que distinguem qual foi o seu evento gerador.

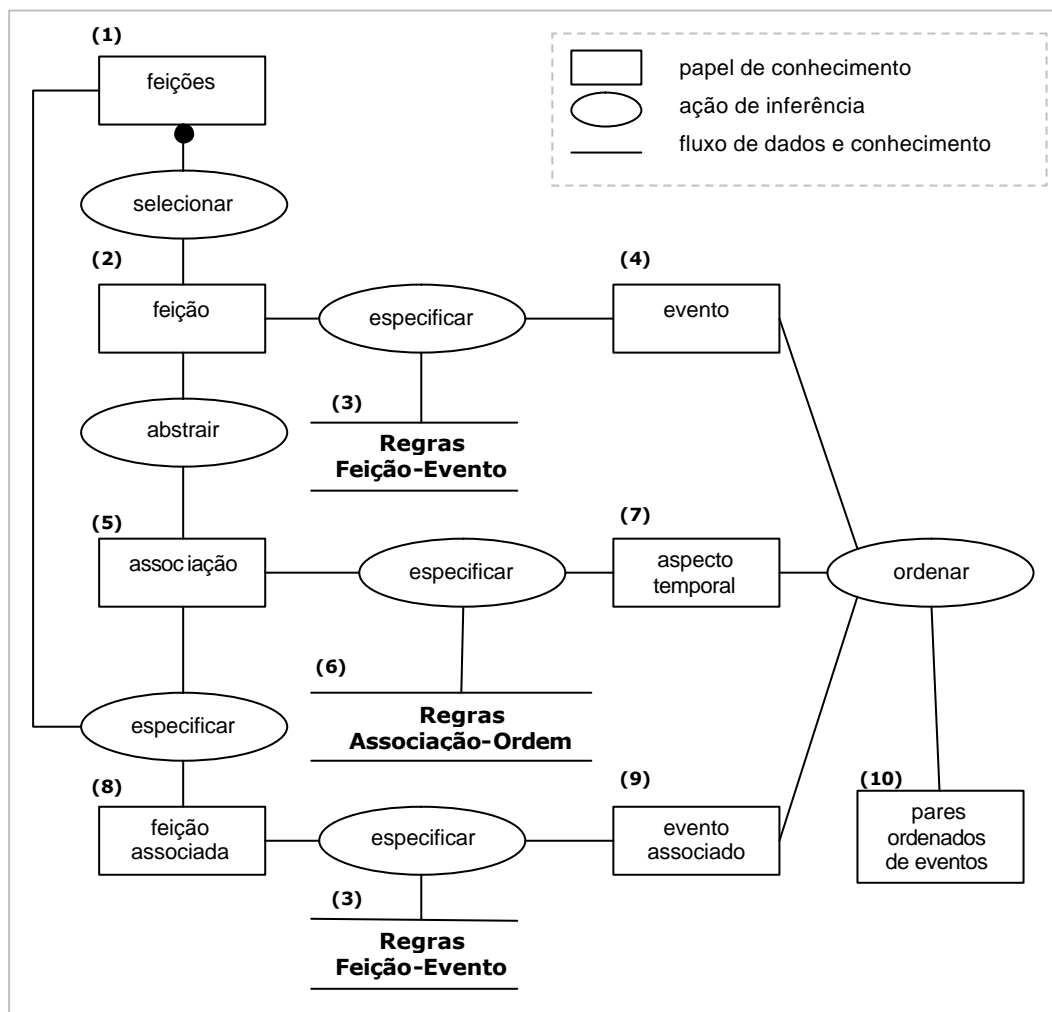


Figura 7.13: Modelo de inferência para identificação de eventos.

Essa dependência entre características e eventos é representada como conhecimento estático do domínio: as *regras feição-evento* (3). Essas regras especificam um *evento* (4) a partir de uma *feição*. Uma *feição* contém diversas informações em sua descrição. Pode-se abstrair da descrição da *feição* a *associação* (5) que essa *feição* apresenta com outra *feição* do domínio. São as associações (os relacionamentos) entre as *feições* do domínio que indicam a ordem em que essas *feições* foram geradas. As regras de dependência entre associações e relações de ordem são representadas como conhecimento estático do domínio: as *regras associação-ordem* (6). Essas regras

especificam um *aspecto temporal* (7) a partir de uma associação<sup>13</sup>. A informação de associação é utilizada para especificar, a partir da lista inicial de feições, a *feição* de domínio com a qual a primeira se associa (8). Essa feição também especifica um *evento* que a gerou (9), através das regras feição-evento. A partir do momento em que se tem um par de eventos e a relação de ordem entre eles (o evento original, o evento associado a ele, e o aspecto temporal), é possível ordenar esses eventos compondo um *par ordenado* (10).

Para melhor caracterizar o processo de raciocínio, ele será exemplificado a seguir por meio de conceitos de domínio da Petrografia Sedimentar.

Uma amostra de rocha é descrita qualitativamente pelo geólogo, ou seja seus principais constituintes (minerais e poros) são detalhados formando uma lista de constituintes de rocha (1). Cada um desses constituintes é selecionado para ser analisado individualmente (2). As informações que qualificam um constituinte dão indicações para especificar o evento gerador (4) daquelas características. Essas indicações são representadas como as regras feição-evento (3). Um exemplo dessas regras é:

```
SE constituinte-de-rocha.modificador = dissolvido
ENTÃO evento-diagenético.nome-evento = dissolução
```

Quando um constituinte é descrito, informa-se a relação paragenética que ele apresenta com outro constituinte da descrição. A relação paragenética (5) é então abstraída da descrição total do constituinte. A relação paragenética entre dois constituintes é a informação que implica na relação de ordem (7) entre eles. As regras que definem essa implicação são representadas como as regras associação-ordem (6). Essas regras são mapeadas no domínio para os tipos-regra implicação-temporal. Um exemplo dessa regra é:

```
SE cobrindo(constituente-1, constituinte-2)
E produzido-por(constituente-1, evento-1)
E produzido-por(constituente-2, evento-2)
ENTÃO depois(evento-1, evento-2)
```

A relação paragenética é usada para buscar na lista, aquele constituinte (8) que está relacionado com o primeiro constituinte selecionado. Esse constituinte também especifica o evento (9) que o gerou, usando as regras feição-evento. Tendo-se a relação de ordem em que os constituintes apareceram na rocha, e sabendo-se os eventos geradores desses constituintes, é possível compor um par ordenado dos eventos geradores (por exemplo, *antes(dissolução(constituente-1), compactação)*). Esse processo é repetido até todos os constituintes serem avaliados e ser gerada uma lista de pares de eventos

### 7.3.2 Inferências para Ordenação de Eventos

O processo de identificação de eventos é representado pelo modelo gráfico da Figura 7.14.

Esse modelo pode ser explicado da seguinte forma: a entrada da inferência é a *lista de pares ordenados de eventos*, resultante do processo de identificação de eventos (1). Desta lista um *par* (2) é selecionado por vez, de forma sequencial. Os eventos precisam ser analisados em separado, por isso, o par de eventos ordenados é decomposto em:

---

<sup>13</sup> Um aspecto temporal pode ser uma primitiva de tempo relativa ou absoluta. Neste caso, a primitiva de tempo é relativa, pois se refere à relação de ordem entre os eventos que geraram as feições.

*ordem* entre os eventos (3), primeiro *evento* (4) e segundo *evento* (5). Cada um dos eventos é casado com a seqüência final de eventos (8) para verificar se ele já foi adicionado na seqüência, em função de os eventos poderem aparecer em diferentes pares de eventos. Essa comparação é feita por casamento de padrão dos atributos do evento. Se casamento de padrões encontra o evento na seqüência, este é retornado como *resultado* (6) e (7). Caso contrário, o evento vindo do par ordenado será retornado como resultado. Se os eventos não tiverem sido adicionados na seqüência ainda, então eles serão ordenados na *seqüência de eventos* (8) dependendo da informação dada pelo papel *ordem*. A seqüência é a saída da estrutura de inferência.

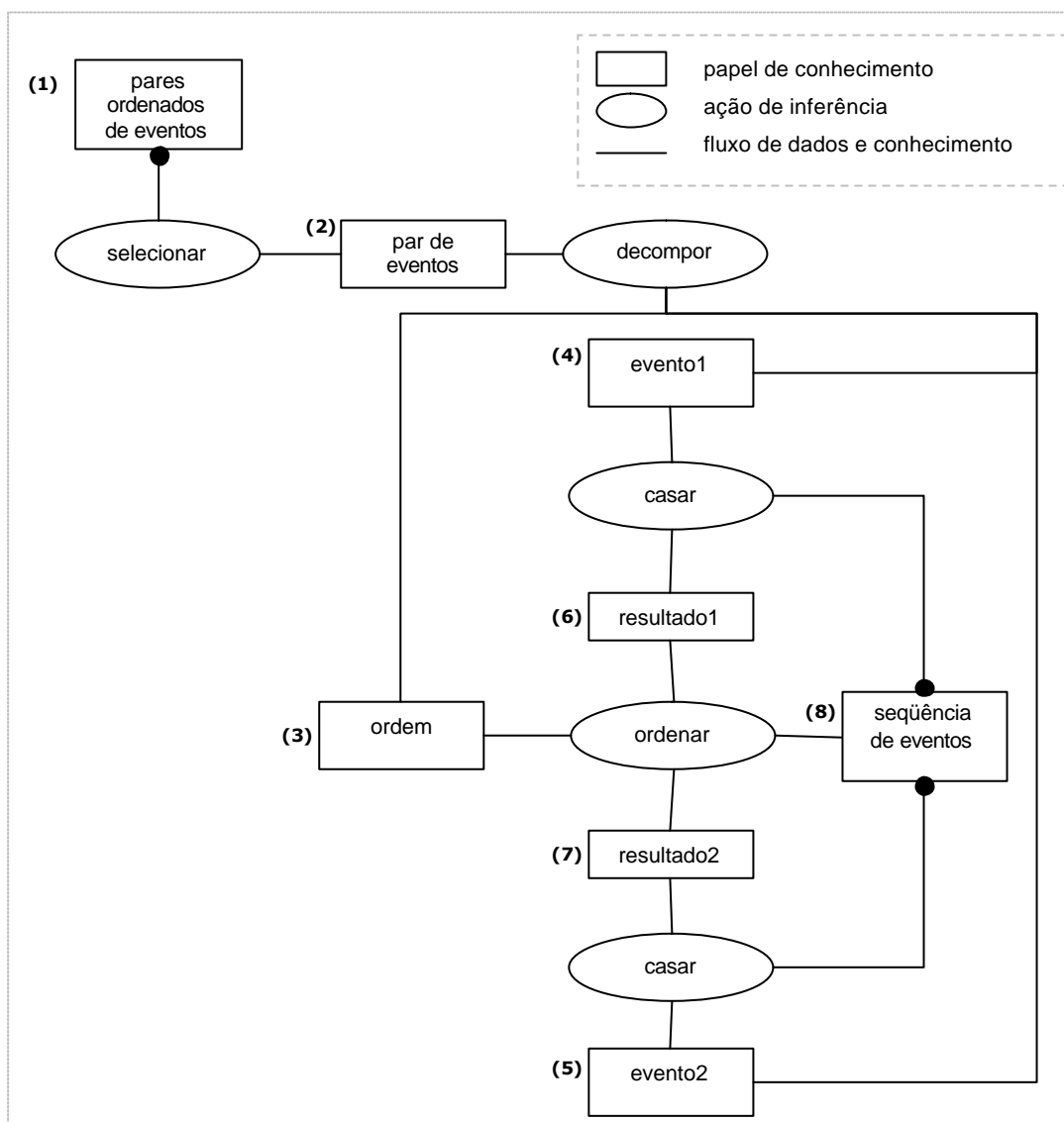


Figura 7.14: Modelo de inferência para ordenação de eventos.

Para melhor caracterizar o processo de raciocínio de ordenação de pares, ele será exemplificado a seguir. Supondo que a lista de entrada (os pares ordenados) sejam como segue:

[antes(C,D), antes(C,E), depois(B,A), antes(B,C)]

O primeiro par é selecionado (*antes(C,D)*) e a partir desse par é especificada a ordem (*antes*) (3), o primeiro elemento (C) (4) e o segundo elemento (D) (5). Como nenhum elemento foi adicionado ainda, o casamento não vai retornar nenhum elemento

da seqüência. Logo, C e D são adicionados na seqüência na ordem C antes de D (Figura 7.15):

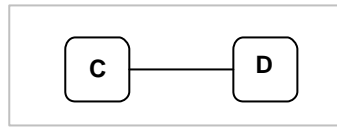


Figura 7.15: Elementos C e D adicionados na seqüência.

O segundo par é selecionado (*antes(C,E)*) e esse par é decomposto em: ordem (*antes*) (3), primeiro elemento (C) (4) e segundo elemento (E) (5). O casamento de C com a seqüência vai retornar o elemento C que já foi adicionado na seqüência. Isso significa que somente o elemento E será adicionado, e o C da seqüência deve vir antes dele (Figura 7.16):

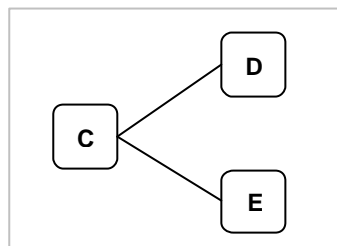


Figura 7.16: Elemento E adicionado na seqüência depois do elemento C.

O terceiro par é selecionado (*depois(B,A)*) e a partir desse par é especificada a ordem (*depois*) (3), o primeiro elemento (B) (4) e o segundo elemento (A) (5). Como nenhum desses elementos foi adicionado ainda, o casamento não vai retornar nenhum elemento da seqüência. Logo, B e A são adicionados na seqüência na ordem B depois de A (Figura 7.17):

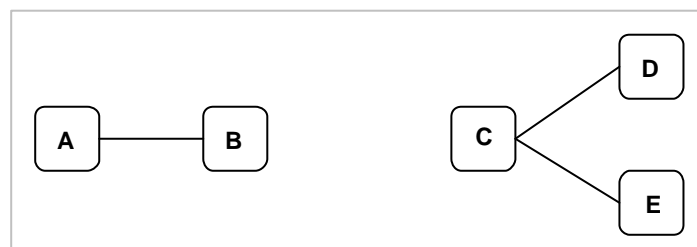


Figura 7.17: Elementos A e B adicionados na seqüência.

O último par é selecionado (*antes(B,C)*) e a partir desse par é especificada a ordem (*antes*) (3), o primeiro elemento (B) (4) e o segundo elemento (C) (5). O casamento de B e de C com a seqüência vai retornar os elementos B e C que já foram adicionados na seqüência. Isso significa que somente é preciso criar, na seqüência, a ordem entre esses elementos. A ordem é que B deve vir antes de C (Figura 7.18):

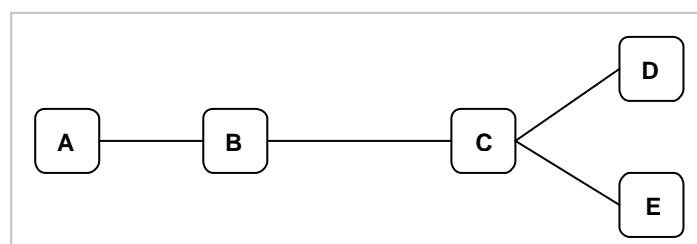




Figura 7.18: Elementos B e C colocados em ordem na seqüência.

Dessa maneira é definida uma seqüência entre os elementos da lista de pares ordenados. O fato de cada evento ser procurado na seqüência final antes de ser adicionado garante que eles serão colocados em ordem.

## 7.4 Modelo do Conhecimento de Tarefa

Como introduzido anteriormente, o conhecimento de tarefa indica os objetivos do sistema e como as inferências devem ser empregadas para alcançá-los. A tarefa de interpretação de seqüência de eventos pode ser dividida em duas sub-tarefas com objetivos diferentes. A primeira tarefa é chamada de *identificar eventos*, e tem como objetivo identificar eventos que já ocorreram e que são os responsáveis pela geração das características atuais no domínio. Essa tarefa recebe como entrada as informações sobre as características do domínio e devolve um conjunto de pares de eventos, ou seja, os eventos estão semi-ordenados (cada par de eventos tem associada a informação de ordenação entre eles). A segunda tarefa é chamada *ordenar eventos*, e tem como objetivo definir a ordem total em que os eventos identificados ocorreram. Essa tarefa recebe como entrada a lista de pares de eventos ordenados e devolve a seqüência completa de eventos, em uma ordem linear. Essas tarefas são mais bem detalhadas no modelo de conhecimento da tarefa, onde são definidos os objetivos, as informações de entrada e saída, as inferências necessárias para realizar a tarefa e a estrutura de controle da inferência, que é a representação de alto nível do futuro mecanismo de raciocínio do sistema. A Figura 7.19 mostra a decomposição da tarefa em suas sub-tarefas.

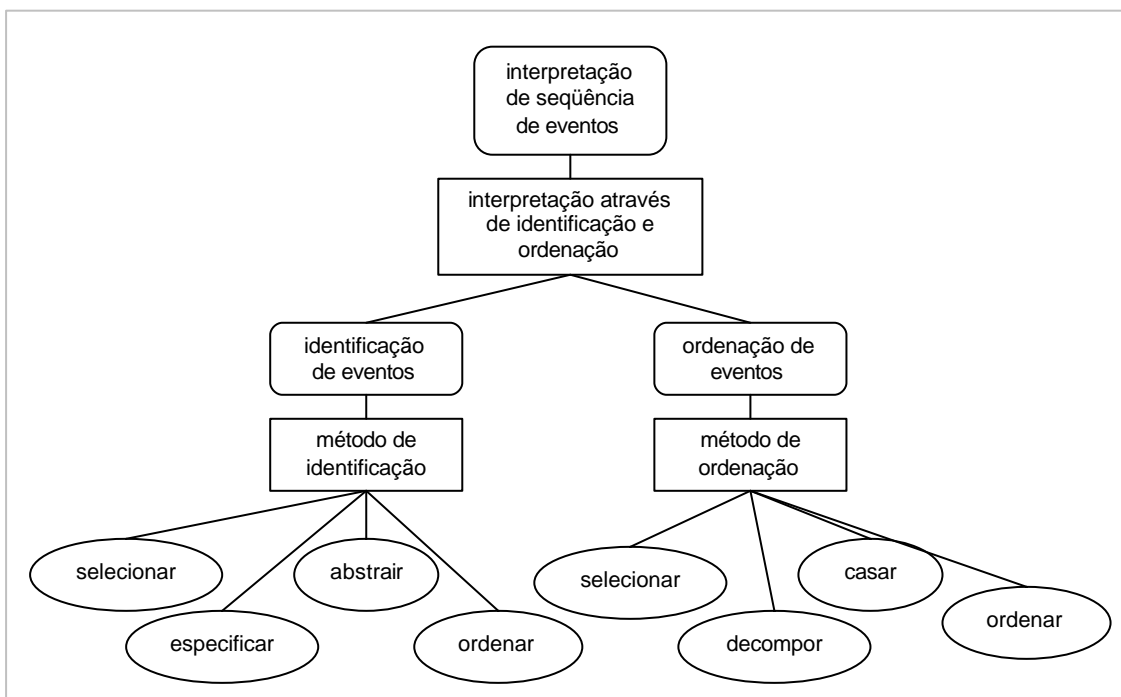


Figura 7.19: Tarefa de interpretação de seqüência de eventos decomposta em sub-tarefas.

As inferências utilizadas pela tarefa são definidas na seção a seguir. A maneira como as inferências e os papéis de conhecimento são empregados para realizar as tarefas e sub-tarefas são descritas no *método da tarefa*. O método tem a descrição da *estrutura de*

*controle*, que é um algoritmo em pseudo-linguagem de programação que emprega estruturas de laço, de teste, entre outras para indicar como chegar aos objetivos da tarefa.

#### 7.4.1 Tarefa de Identificação de Eventos

A tarefa de identificação de eventos será explicada utilizando o modelo de descrição de tarefa proposto no catálogo de tarefas da metodologia CommonKADS (SCHREIBER et al., 2000).

Caracterização geral da tarefa de Identificação de Eventos

**Objetivo:** identificar os eventos que geraram determinadas características do domínio.

**Exemplo típico:** identificar eventos geológicos que geraram as características de rochas.

Terminologia:

feição: Uma característica do domínio;

evento: O evento gerador da característica;

associação: Relacionamento entre feições;

aspecto-temporal: Relação de ordem entre os eventos;

feição-associada: Característica do domínio que se relaciona com a anterior;

evento-associado: O evento que gerou a feição associada;

par-ordenado-evento: Par de eventos dispostos na ordem em que ocorreram;

pares-de-eventos-ordenados: Conjunto de pares ordenados de eventos;

**Entrada:** características do domínio

**Saída:** eventos geradores

Método padrão:

Enquanto ainda existirem feições para serem avaliadas, o algoritmo repete a seguinte seqüência de ações: especificar o evento que produziu a primeira feição, abstrair a associação que essa feição tem com outra feição do domínio e buscar a feição associada na lista inicial de feições. Especificar o evento que gerou a segunda feição. Identificar a relação temporal através da relação de associação, e definir a relação de ordem entre os eventos, criando um par de eventos ordenados que é adicionado na lista que será gerada como saída do algoritmo. A estrutura de controle é representada na Figura 7.20. O laço é repetido até que não existam mais feições a serem avaliadas.

Inferências:

**selecionar:** seleciona um elemento da lista ou conjunto de entrada usando algoritmos de seleção padrão. Se a seleção é de um conjunto, a inferência deve produzir a saída em uma ordem randômica. Se a seleção é de uma lista, a inferência deve selecionar a saída seqüencialmente.

**especificar:** recebe como entrada um objeto e produz como saída um novo objeto associado de alguma maneira com o objeto de entrada. As regras de associação são recebidas como conhecimento estático.

**abstrair:** os dados de entrada são abstraídos para alguns aspectos relevantes. Se a entrada é a instância de uma classe, a saída pode ser um ou mais atributos dessa instância.

**ordenar:** entrada é um conjunto de elementos. Saída é uma lista ordenada contendo os mesmos elementos.

```

MÉTODO-TAREFA método-identificar;
REALIZA:
  identificar-eventos;
ESTRUTURA-CONTROLE:
  ENQUANTO NOVA-SOLUÇÃO selecionar(feições -> feição) FAÇA
    especificar(feição + regras-feição-evento -> evento);
    abstrair(feição-> associação);
    especificar(associação + feições -> feição-associada);
    especificar(feição-associada + regras-feição-evento -> evento-associado);
    especificar(associação + regras-associação-ordem -> aspecto-temporal);
    ordenar(evento + evento-associado + aspecto-temporal -> par-ordenado-evento);
    pares-de-eventos-ordenados := pares-de-eventos-ordenados ADICIONA par-ordenado-evento;
  FIM ENQUANTO
FIM MÉTODO-TAREFA método-identificar;

```

Figura 7.20: Método da tarefa de identificação de eventos.

## 7.4.2 Tarefa de Ordenação de Eventos

A tarefa de ordenação de eventos será explicada utilizando o modelo de descrição de tarefa proposto no catálogo de tarefas da metodologia CommonKADS.

Caracterização geral da tarefa de Ordenação de Eventos

**Objetivo:** ordenar uma lista de pares de eventos em uma seqüência de eventos.

Terminologia:

par-evento: par ordenado de eventos;

ordem: ordem em que os eventos se encontram no par;

evento1: primeiro evento do par;

evento2: segundo evento do par;

resultado1: se o primeiro evento já está adicionado na seqüência;

resultado2: se o segundo evento já está adicionado na seqüência;

**Entrada:** pares de eventos

**Saída:** seqüência de eventos na ordem em que eles ocorreram

Método padrão:

Enquanto ainda existirem pares de eventos para serem avaliados, o algoritmo repete a seguinte seqüência de ações: decompor o par nos dois eventos e na ordem em que eles se encontram; verificar se os eventos já foram adicionados na seqüência final de eventos. Se o evento não está lá, ele é adicionado. Ordenar (criar uma ligação entre os eventos) conforme a ordem especificada no par. No final, os eventos estarão em uma seqüência na ordem em que ocorreram. A estrutura de controle é representada na Figura 7.21.

Inferências:

**selecionar:** seleciona um elemento da lista ou conjunto de entrada usando algoritmos de seleção padrão. Se a seleção é de um conjunto, a inferência deve produzir a saída em uma ordem randômica. Se a seleção é de uma lista, a inferência deve selecionar a saída seqüencialmente.

**decompor:** recebe como entrada um objeto composto de várias informações e produz como saída cada uma das informações que compõe o objeto de entrada.

**casar:** compara um conjunto de entradas através de casamento das características.

**ordenar:** entrada é um conjunto de elementos. Saída é uma lista ordenada contendo os mesmos elementos.

```

MÉTODO-TAREFA método-ordenar
REALIZA:
ordenar-eventos;
ESTRUTURA-CONTROLE:
ENQUANTO NOVA-SOLUÇÃO selecionar(pares-de-eventos-ordenados-> par-evento) FAÇA
decompor(par-evento -> ordem + evento1 + evento2);
casar(evento1 + sequência-eventos -> resultado1);
SE resultado1 == falso
ENTÃO sequência-eventos := sequência-eventos ADICIONA evento1;
FIM SE
casar(evento2 + sequência-eventos -> resultado2);
IF resultado2 == falso
ENTÃO sequência-eventos := sequência-eventos ADICIONA evento2;
FIM IF
ordenar(evento1 + evento2 + ordem -> sequência-eventos)
FIM ENQUANTO
FIM MÉTODO-TAREFA método-ordenar;

```

Figura 7.21: Método da tarefa de ordenação de eventos.

A definição completa do modelo de tarefa através da linguagem CML é apresentada no Apêndice B, no modelo de conhecimento.

## 7.5 Resumo do Capítulo

O modelo de domínio proposto neste capítulo foi baseado na ontologia de eventos proposta no capítulo 6, e tem o objetivo de estender a conceitualização de um domínio específico, o da petrografia sedimentar, com os objetos de domínio necessários para a realização de uma nova tarefa de raciocínio: a interpretação de seqüências diagenéticas. Além disso, o modelo se propõe a representar as primitivas temporais que essa tarefa demanda: as relações de ordem de tempo entre os eventos. O modelo de inferência representa os papéis de conhecimento e as inferências necessárias para se realizar a interpretação de seqüência diagenética, enquanto o modelo da tarefa indica como esses papéis e inferências devem ser empregados para atingir os objetivos da interpretação: identificar os eventos que geraram as características do domínio e ordená-los na seqüência em que ocorreram. No capítulo seguinte esse modelo é validado através da implementação de um novo módulo de inferência para o sistema *PetroGrapher*: o módulo de interpretação de seqüência diagenética.

## 8 IMPLEMENTAÇÃO E VALIDAÇÃO DO MODELO DE CONHECIMENTO PARA ORDENAÇÃO DE EVENTOS

Para apresentar uma validação prática do modelo de conhecimento para ordenação de eventos diagenéticos foi proposta uma arquitetura de inferência para interpretação de seqüência diagenética, desenvolvida como um módulo do sistema *PetroGrapher* (apresentado no capítulo de introdução deste trabalho), que é desenvolvido na linguagem Java®. Essa arquitetura implementa as três categorias de conhecimento, modeladas no capítulo 7: domínio, inferência e tarefa. O conhecimento de domínio foi implementado como o sistema de classes e de tabelas da base de conhecimento. O conhecimento de inferência e de tarefa foi implementado como algoritmos de inferência sobre as classes e tabelas. A implementação é descrita a seguir.

### 8.1 Implementação do Conhecimento de Domínio

O mapeamento das primitivas no nível de conhecimento para o nível de implementação foi feito da seguinte forma:

- O conceito *constituente-de-rocha* e seus subconceitos e o evento *evento-diagenético* foram mapeados para classes da linguagem Java: *Constituente*, *ConstituenteMineral*, *ConstituenteDiagenético*, *ConstituenteDetritico*, *ConstituenteMacroporosidade*, *EventoDiagenético*. Essas classes apresentam como atributos as mesmas informações do modelo de conhecimento.
- A nomenclatura dos eventos diagenéticos foi representada como uma tabela na base de conhecimento (Tabela 8.1).
- O relacionamento *relação-paragenética*, por apresentar sub-relações (ou seja, é parte de uma hierarquia) foi implementado como uma classe Java (*RelaçãoParagenética*). Essa classe apresenta como atributos os dois constituintes que relaciona, o nome textual da relação paragenética e a informação de qual relação é a sua inversa.
- As subrelações de *relação-paragenética* (p.ex., *dissolução-de*, *substituindo*) foram implementadas por classes Java que herdam todos os atributos da classe *RelaçãoParagenética* (*DissoluçãoDe*, *Substituindo*, *Cobrindo*, etc.). A diferença é que cada sub-relação apresenta um método construtor específico, o qual determina quais tipos de constituintes cada relação paragenética pode relacionar.
- O construto *relação-temporal* foi implementado como uma classe Java (*RelaçãoTemporal*). Essa classe apresenta como atributos os dois argumentos (eventos) que a relação temporal associa (que podem ser dois constituintes

ou dois eventos) e nome textual da ordem da relação temporal (*antes, depois, etc.*).

Tabela 8.1: Tabela relacional `DiageneticProcessNomenclature` que representa a nomenclatura de eventos diagenéticos.

IdEvento	NomeEvento
1	Dissolution
2	Cementation (authigenesis)
3	Replacement (authigenesis)
4	Compaction
5	Pressure dissolution (chemical compaction)
6	Rearranjement (mechanical compaction)
7	Fracturing (mechanical compaction)
8	Elastic deformation (mechanical compaction)
9	Plastic deformation (mechanical compaction)
10	Mechanical infiltration
11	Hydration
12	Dehydration
13	Oxidation
14	Reduction
15	Recrystalization
16	Stabilization/neomorphism

- Para diferenciar as classes `ConstituinteDiagenético` e `ConstituinteMacroporosidade` da classe `ConstituinteDetritico`, pelo fato de esta última não apresentar a informação de relação paragenética, foi criada uma interface Java (`ParageneticRelationConstituent`) a qual as duas primeiras classes implementam. Ou seja, `ConstituinteDiagenético` e `ConstituinteMacroporosidade` são classes que implementam a informação de relação paragenética.
- As instâncias dos tipos-regras foram implementadas como tabelas relacionais na base de conhecimento. As associações entre relações paragenéticas e relações de ordem foram definidas como mostrado na Tabela 8.2.

Tabela 8.2: Tabela relacional `ParageneticOrderRelation` que representa associações entre relações paragenéticas e relações de ordem.

RelaçãoParagenética	Relação Ordem	Observação
Dissolução de <Constituinte-Diagenético>	>	P formado por dissolução de A --> P depois A --> P > A
Dentro de <Constituinte-Detrítico>	>=	poro pode ter vindo junto de A
Dentro de <Constituinte-Diagenético>	>=	poro pode ter vindo junto de A
Contração de <Constituinte-Detrítico>	>	P formado por contração de A --> P depois A --> P > A

- As associações entre constituintes e os eventos geradores foram implementadas como duas tabelas, dependendo do tipo de informação usado para indicar o evento. A implicação entre modificadores de constituintes detríticos e os eventos foi implementada como na Tabela 8.3.

Tabela 8.3: Tabela relacional ModifierProcess que representa as associações entre modificadores e eventos (chave estrangeira para a tabela de nomenclatura de eventos, representada na Tabela 8.1).

Modificador	IdEvento
Substituído	3
Alterado	3
Fraturado	4
Deformado	4
Recristalizado	15
Dissolvido	1

- A implicação entre relações paragenéticas e eventos foi implementada como na Tabela 8.4.

Tabela 8.4: Tabela relacional ParageneticRelationProcess que representa as associações entre relações paragenéticas e eventos (chave estrangeira para a tabela de nomenclatura de eventos, representada na Tabela 8.1).

Relação Paragenética	IdEvento
Substituindo grão de <Constituinte-Detrítico>	3
Engolfando <Constituinte-Detrítico>	2
Dentro porosidade intragranular primária	2
Cobrindo <Constituinte-Diagenético>	2
Cobrindo <Constituinte-Detrítico>	2

O modelo completo de classes desenvolvido a partir da ontologia de ordenação de eventos é mostrado no Apêndice B.

## 8.2 Implementação do Conhecimento de Inferência e de Tarefa

Os modelos de inferência e de tarefa para interpretação da seqüência de eventos propostos nesse trabalho foram usados como modelos descritos no nível de conhecimento. Para apresentar uma validação prática desses modelos, foi desenvolvido um *algoritmo de inferência para interpretação da seqüência de eventos diagenéticos*. Esse algoritmo fornece um recurso de inferência sobre a seqüência diagenética para o sistema *PetroGrapher*. O algoritmo de interpretação de seqüência foi desenvolvido a partir do modelo da tarefa de interpretação da seqüência de eventos, construindo a seqüência de eventos que formaram uma determinada amostra de rocha. As subtarefas nas quais a tarefa de interpretação se divide foram implementadas como dois métodos JAVA na classe principal.

Os métodos e funções internas do algoritmo de inferência foram baseados nas inferências utilizadas pelo modelo de tarefa e descritas no modelo do conhecimento de inferência. A seguir são descritos os passos de execução do algoritmo de inferência.

### 8.2.1 Entrada de Dados

A amostra de rocha em questão é descrita na entrada de dados do sistema *PetroGrapher*. O sistema possibilita diversas análises e interpretações de uma amostra descrita, entre elas, a interpretação da seqüência de eventos, que tem início com uma requisição dessa interpretação específica.

O primeiro passo do algoritmo é a *recuperação da descrição da amostra* do banco de dados do usuário. A descrição de uma amostra é formada pela descrição de todos os

constituintes de rocha e por informações texturais, visualizadas a olho nu (macroscopia) e com o auxílio de um microscópio ótico (microscopia). As descrições de constituintes são como as da Figura 8.1. O algoritmo cria listas internas de constituintes, uma para cada tipo de constituinte (detríticos, diagenéticos e macroporosidades).

<p><b><u>Constituinte Diagenético</u></b></p> <p><b>Nome:</b> clorita  <b>Conjunto:</b> argilas autigênicas  <b>Hábito:</b> revestimento  <b>Localização:</b> intergranular descontínuo revestindo poro  <b>Relação paragenética:</b> substituindo constituinte diagenético  <b>Constituinte da relação paragenética:</b> pirita</p>	<p><b><u>Constituinte Detrítico</u></b></p> <p><b>Nome:</b> esfênio  <b>Conjunto:</b> Minerais pesados  <b>Localização:</b> Como grão monomineralico  <b>Modificador:</b> dissolvido</p>
--	--

Figura 8.1: Exemplo de descrição de dois tipos de constituintes de uma amostra de rocha – constituinte diagenético (clorita) e constituinte detrítico (esfeno).

## 8.2.2 Implementação da Tarefa de Identificação dos Eventos

Essa tarefa foi implementada como o seguinte método:

```
List identifyEvents(constituentsList);
```

O método `identifyEvents()` recebe como parâmetro de entrada a lista de constituintes descritos e devolve como saída a lista de pares ordenados de eventos que geraram esses constituintes.

Os constituintes diagenéticos e as macroporosidades apresentam a informação que permite relacionar os constituintes: a *relação paragenética*. A lista de constituintes diagenéticos e de macroporosidades recebida pelo método `identifyEvents()` é então percorrida da seguinte maneira:

- A lista é lida por um laço, e é selecionado um constituinte por vez. Essa seleção sequencial é a implementação da inferência SELECIONAR do modelo de inferência.
- A relação paragenética do constituinte selecionado é identificada. Ao isolar a informação de relação paragenética do resto da descrição do constituinte o algoritmo está implementando a inferência ABSTRAIR do modelo de inferência. Identifica-se o tipo de constituinte com o qual essa relação cria uma associação.
- O constituinte associado é buscado na lista de constituintes da qual ele faz parte. Essa busca é feita comparando-se os atributos do constituinte descrito na relação paragenética com os constituintes da lista. Quando a comparação resultar em igualdade, o constituinte é dado como encontrado. Esse método implementa a inferência ESPECIFICAR do modelo de inferência.
- Instancia-se um objeto da classe `RelaçãoParagenética`, passando como parâmetro os dois constituintes e a relação paragenética específica entre eles, como mostrado na Figura 8.2.

<p><b><u>Substituindo: RelaçãoParagenética</u></b></p> <p>constituente1: Clorita  constituente2: Pirita  nomeRelacaoParagenética: substituindo</p>
--



Figura 8.2: Instância da classe que representa a relação paragenética *Substituindo*.

Quando o algoritmo termina de percorrer todas as listas de constituintes, ele dispõe de uma lista de instâncias de relações paragenéticas entre constituintes.

#### 8.2.2.1 Identificando eventos

A partir da lista de constituintes relacionados entre si através das relações paragenéticas, o algoritmo pode identificar os eventos que geraram os constituintes e inferir a relação temporal entre os eventos. Para isso o algoritmo percorre a lista de constituintes e, para cada um, chama o método (o qual implementa a inferência ESPECIFICAR, do modelo de inferência) que aplica as regras de inferência para identificar o constituinte. As regras de inferência são trazidas com uma consulta à tabela na base de conhecimento onde estão armazenadas essas associações (como demonstrado nas tabelas Tabela 8.3 e Tabela 8.4) a qual representa o papel de conhecimento estático REGRAS-FEIÇÃO-EVENTO. Esse método é chamado para os dois constituintes relacionados, gerando assim, os dois eventos que serão associados com uma relação temporal.

#### 8.2.2.2 Inferindo a relação temporal

Ainda dentro do par de constituintes, é selecionada a relação paragenética para inferir a relação temporal. O algoritmo aplica um método de classe que retorna a relação de ordem a partir da relação paragenética, fazendo uma consulta à tabela na base de conhecimento onde estão armazenadas essas associações (como demonstrado na Tabela 8.2). Esse método implementa a inferência ESPECIFICAR, do modelo de inferência, e o papel de conhecimento estático REGRAS-ASSOCIAÇÃO-ORDEM é implementado na tabela com as associações entre as relações paragenéticas e as relações de ordem.

O algoritmo cria uma instância da classe `RelaçãoTemporal`, passando a relação de ordem inferida e os dois eventos especificados a partir dos constituintes da `RelaçãoParagenética`, como mostrado na Figura 8.3. A instanciação de uma `RelaçãoTemporal` utilizando como informações de entrada os dois eventos e a relação de ordem é uma maneira de implementar a inferência ORDENAR do método de inferência.

O método implementa um laço, que repete essas ações até que toda a lista de constituintes da entrada tenha sido transformada em pares ordenados dos eventos geradores. Dessa forma, é criada uma lista de eventos relacionados através de uma `RelaçãoTemporal`.

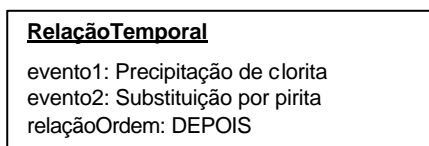


Figura 8.3: Instância da classe que representa a relação temporal DEPOIS entre os eventos.

### 8.2.3 Implementação da Tarefa de Ordenação dos Eventos

Essa tarefa foi implementada como o seguinte método:

```
DirectedGraph orderEvents(orderedEventsPairList);
```

O método `orderEvents()` recebe como parâmetro de entrada a lista de pares ordenados de eventos e devolve como saída uma estrutura interna da linguagem JAVA que representa a seqüência de eventos, a qual será explicada a seguir.

De forma a representar uma seqüência ordenada dos constituintes, foi escolhida uma estrutura de grafos dirigidos. A escolha óbvia seria por uma estrutura de lista, pois a saída do algoritmo deveria ser uma seqüência linear dos processos que formaram os produtos da rocha. Entretanto, algumas simulações (descritas na seção 5.2.13 do capítulo de aquisição de conhecimento) mostraram que algumas vezes pela falta de uma única relação paragenética (que pode não ser visível na lâmina descrita, ou pode não ter sido descrita por falta de detalhamento) não se consegue uma ordem completa entre os processos. Mesmo que se conseguisse todas as relações paragenéticas, alguns processos aconteceram simultaneamente no tempo, e precisariam ser representados no mesmo nível de tempo, impedindo o uso de uma lista que represente seqüência linear. A Figura 8.4 ilustra essa situação:

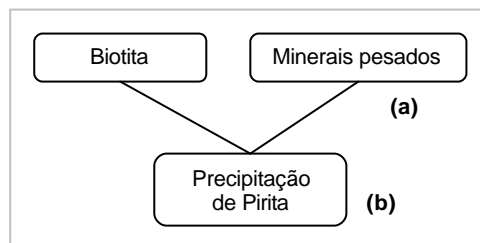


Figura 8.4: Seqüência não linear de aparecimento de constituintes.

Um grafo é uma estrutura adequada para essa representação. Os vértices do grafo representam os eventos. As arestas entre os vértices representam a relação de tempo entre os eventos. O vértice-origem de uma aresta (Figura 8.4 (a)) representa o evento que ocorreu antes, enquanto o vértice-destino da aresta (Figura 8.4 (b)) representa o evento que ocorreu depois. Se for necessário colocar dois eventos como tendo ocorrido em paralelo, cria-se dois nodos que partem do mesmo nodo origem (ou que se destinam a um mesmo nodo destino, como na Figura 8.4). Se os grafos fossem não-dirigidos, não seria possível representar a ordem entre os nodos do grafo. O algoritmo para a construção do grafo a partir das triplas de constituintes foi desenvolvido como mostrado na Figura 8.5 e é explicado a seguir.

```

Cria um grafo dirigido g;
Para cada tripla  $\langle e1, rel\_temporal, e2 \rangle$  da lista de pares faça
  Se e1 não existe como vértice em g
    Adiciona vértice e1 em g;
  Se e2 não existe como vértice em g
    Adiciona vértice e2 em g;
  Caso rel_ordem seja
    antes:
      se não existe aresta  $e1 \rightarrow e2$ ,
    adiciona aresta  $e1 \rightarrow e2$  em g;
    depois:
      se não existe aresta  $e2 \rightarrow e1$ ,
    adiciona aresta  $e2 \rightarrow e1$  em g;
Retorna g;

```

Figura 8.5: Algoritmo de construção do grafo dirigido a partir da lista de relações temporais.

- O algoritmo recebe como entrada a lista de pares ordenados de eventos (esses pares são representados pela classe `RelaçãoTemporal`).
- O algoritmo percorre essa lista identificando uma relação temporal a cada iteração.
- Cada um dos eventos descritos no objeto `RelaçãoTemporal` é procurado na lista de vértices já adicionados ao grafo. Se o evento não existe no conjunto de vértices do grafo, então ele é adicionado. Essa ação é efetuada para os dois eventos.
- Depois que os dois eventos são garantidamente representados por dois vértices no grafo, a relação temporal entre os eventos é testada. Caso a relação seja, por exemplo, (Evento1 ANTES Evento2), é adicionada uma aresta do vértice de Evento1 para o vértice de Evento2 (antes de se adicionarem as arestas, o algoritmo testa se elas já não existem). Caso a relação seja (Evento1 DEPOIS Evento2), é adicionada uma aresta do vértice de Evento2 para o vértice de Evento1.
- Depois que toda a lista é percorrida, o grafo é devolvido como retorno do algoritmo.

No caso em que a descrição gere uma seqüência, esse algoritmo terá deixado os eventos na ordem em que eles ocorreram, como demonstrado na próxima seção.

### 8.3 Demonstração do Algoritmo com Dados Reais

Uma simulação do algoritmo de seqüência diagenética pode ser feita baseada na amostra S-208, descrita pelo especialista exclusivamente para essa interpretação.

#### 8.3.1 Entrada de Dados

A Figura 8.6 mostra a entrada de dados: a descrição dos constituintes detríticos, diagenéticos e macroporosidades da amostra de rocha.

Constituent Identification
Detrital quartz monocrystalline - As monomineralic grain
Detrital quartz monocrystalline - As monomineralic grain - Dissolved
Hematite - Coating - Intergranular continuous pore-lining - - Covering <Detrital-Constituent> - Detrital quartz
Hematite - Coating - Intergranular continuous pore-lining - - Within intergranular primary porosity - Macroporosity Intergranular
Quartz - - Intergranular discontinuous pore-lining - Dissolved - Covering <Diagenetic-Constituent> - Hematite
Quartz - - Intergranular discontinuous pore-lining - Dissolved - Covered by <Diagenetic-Constituent> - Illite
Quartz - - Intergranular discontinuous pore-lining - Dissolved - Within intergranular primary porosity - Macroporosity Intergranular
Illite - Rim - Intergranular discontinuous pore-lining - - Covering <Diagenetic-Constituent> - Hematite
Illite - Rim - Intergranular discontinuous pore-lining - - Covering <Diagenetic-Constituent> - Quartz
Illite - Rim - Intergranular discontinuous pore-lining - - Within intergranular primary porosity - Macroporosity Intergranular
Macroporosity Intergranular - Interstitial - Reduced by cementation - Primary - Quartz
Macroporosity Intergranular - Interstitial - Reduced by cementation - Primary - Illite
Macroporosity Intergranular - Interstitial - Reduced by compaction - Primary -

Figura 8.6: descrição dos constituintes da amostra de rocha S-208 (tela capturada do módulo de descrição composicional do sistema *PetroGrapher*).

Essa descrição é feita pelo geólogo usuário do sistema *PetroGrapher* e é armazenada na base de dados de amostras do sistema. O algoritmo de seqüência diagenética carrega essa descrição através de consultas SQL ao banco de dados. É possível identificar, na descrição da Figura 8.6, vários exemplos de relações paragenéticas entre constituintes, como na quinta linha, que mostra a relação

paragenética *cobrindo* (*covering*) entre quartzo e hematita. A descrição dos constituintes é carregada em listas, e é chamado o método `identityEvents()` que vai identificar os eventos diagenéticos que geraram os constituintes.

### 8.3.2 Identificação dos Eventos

O algoritmo `identifyEvents()` recebe as listas de constituintes e percorre essa lista selecionando um constituinte por vez (por exemplo, o constituinte da terceira linha, *Hematite*). A relação paragenética do constituinte selecionado é identificada (neste exemplo, *Covering [Detrital-Constituent]*). Identifica-se o tipo de constituinte com o qual o primeiro está se relacionando (neste caso, um constituinte detrítico). O constituinte associado (*Detrital Quartz*) é buscado na lista de constituintes da qual ele faz parte. É criada, então, a instância de relação paragenética (*[Hematite] [Covering <Detrital-Constituent>] [Detrital Quartz]*). O algoritmo percorre a lista de constituintes até que seja criada uma lista de relações paragenéticas, que é mostrada na Figura 8.7.

<u>Amostra S-208 – Relações Paragenéticas</u>
[Hematite] [Covering] [Detrital quartz]
[Quartz] [Covering] [Hematite]
[Quartz] [IsCoveredBy] [Illite]
[Illite] [Covering] [Hematite]
[Illite] [Covering] [Quartz]

Figura 8.7: Instâncias de *RelaçãoParagenética* - amostra S-208

#### 8.3.2.1 Identificando eventos

O algoritmo percorre a lista de relações paragenéticas da Figura 8.7 e para cada objeto *RelaçãoParagenética* ele chama o método que aplica as regras de inferência para identificar o evento que gerou o constituinte. Esse método cria uma consulta SQL à tabela que armazena as associações entre os atributos dos constituintes e os eventos. Por exemplo, para o constituinte *hematita*, o método procura o evento relacionado à relação paragenética *covering* (cobrindo). O evento associado é o de *cimentação* (*cementation* (*authigenesis*)). Esse método é aplicado a todos os constituintes, retornando seus eventos geradores. Depois disso, é necessário inferir a relação de ordem entre esses eventos.

#### 8.3.2.2 Inferindo a relação temporal

Para cada relação paragenética, o algoritmo aplica o método que retorna a relação de ordem a partir da relação paragenética. Esse método monta uma consulta à tabela que guarda as associações entre relações paragenéticas e relações de ordem. Por exemplo:

Covering → AFTER

O algoritmo cria uma instância de *RelaçãoTemporal* passando a relação de ordem inferida e os dois eventos especificados (pelo método de identificação de eventos). Por exemplo:

[Cementation of Hematite] [AFTER] [Deposition of Detrital quartz]

O algoritmo percorre toda a lista de relações paragenéticas até criar uma lista de relações de ordem entre eventos, como mostrado na Figura 8.8.

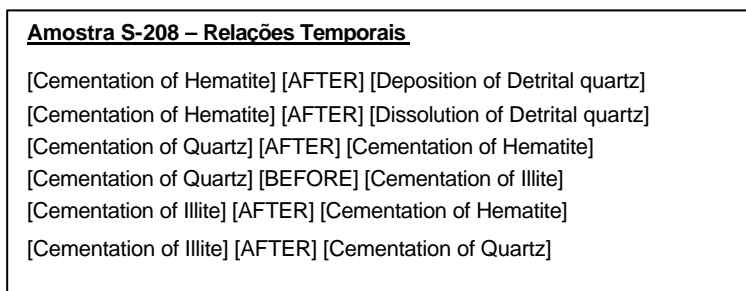


Figura 8.8: Instâncias de `RelaçãoTemporal` - amostra S-208

### 8.3.3 Ordenando os Eventos

A partir das relações temporais, o algoritmo percorre toda a lista e monta o grafo dirigido, criando os vértices para os eventos e os arcos para a relação de ordem.

Para o caso da lista de relações temporais da Figura 8.8, o algoritmo identifica a uma relação temporal da lista (por exemplo, [*Cementation of Quartz*] [AFTER] [*Cementation of Hematite*]) e verifica se o primeiro elemento ([*Cementation of Quartz*]) já existe como vértice no conjunto de vértices do grafo. Como não existe, ele adiciona. A mesma coisa ocorre para o segundo elemento ([*Cementation of Hematite*]).

A relação temporal entre os eventos é testada: a ordem é DEPOIS ([AFTER]), então o algoritmo verifica se existe uma aresta com origem no vértice [*Cementation of Quartz*] e destino em [*Cementation of Hematite*]. Como nenhuma aresta foi adicionada no grafo, essa aresta não existe ainda, logo, ela é adicionada. A configuração inicial do grafo é como mostrada na Figura 8.9.

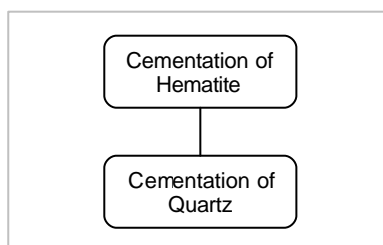


Figura 8.9: configuração inicial da seqüência diagenética da amostra S-208.

Ao recuperar a relação temporal ([*Cementation of Hematite*] [AFTER] [*Deposition of Detrital quartz*]) para ser adicionada, o algoritmo tenta adicionar o vértice [*Cementation of Hematite*] novamente, mas verificando que este já existe como vértice no grafo, não adiciona. Seguindo o algoritmo, o vértice [*Deposition of Detrital quartz*], por ainda não existir no grafo, é incluído. A relação entre [*Cementation of Hematite*] e [*Deposition of Detrital quartz*] é DEPOIS ([*Cementation of Hematite*] [AFTER] [*Deposition of Detrital quartz*]). O algoritmo verifica se existe uma aresta adicionada entre [*Cementation of Hematite*] e [*Deposition of Detrital quartz*] e adiciona a aresta.

O algoritmo continua percorrendo a lista até que todas as relações temporais tenham sido mapeadas para o grafo dirigido. A configuração final do grafo dirigido construído para essa amostra é como na Figura 8.10.

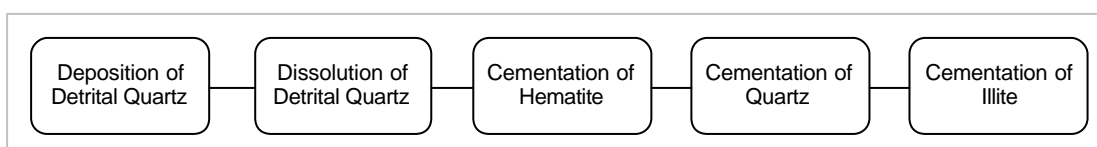


Figura 8.10: configuração final da seqüência diagenética da amostra S-208.

## 8.4 Apresentação Gráfica da Seqüência Diagenética

Para que a estrutura de dados do grafo dirigido montado pudesse ser apresentada como uma saída visual, foi necessário desenvolver um algoritmo de desenho de grafo, pois é preciso calcular a posição na tela de cada vértice do grafo, e isso depende do tamanho da tela, da quantidade de vértices, de como os vértices estão ligados. Um algoritmo foi desenvolvido para desenhar o grafo da seguinte maneira: ele divide os nodos do grafo em diferentes níveis, e depois desenha os nodos na tela calculando a posição mais adequada pela proporção entre o nível do nodo e o tamanho da tela.

O cálculo dos níveis a que cada nodo pertence depende do relacionamento deste nodo com os outros. Os nodos que não recebem ligações (arestas) de outros, são colocados no primeiro nível. Os nodos que recebem arestas dos nodos do primeiro nível são colocados no segundo nível, e assim por diante.

Quando todos os nodos foram ajustados em níveis, o algoritmo calcula o espaçamento entre os níveis dividindo a altura total da tela pelo número de níveis. O espaçamento entre os nodos em cada nível é calculado dividindo a largura da tela pelo número de nodos em cada nível.

No caso do grafo dirigido criado para representar a seqüência diagenética da amostra S-208, cada nodo ficou em um nível, formando uma seqüência linear. O desenho resultante do grafo dirigido montado para a amostra S-208 é mostrado na Figura 8.11.

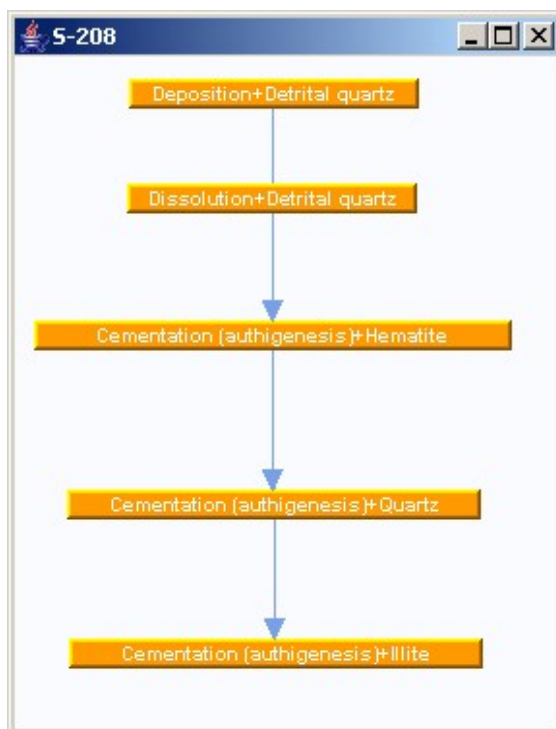


Figura 8.11: Saída resultante do desenho do grafo dirigido montado para a amostra S-208.

## 8.5 Validação do Algoritmo

Para a validação do algoritmo de interpretação se seqüência diagenética, foram utilizadas três amostras de rochas encontradas em profundidades próximas. O

especialista fez a descrição dessas amostras utilizando a interface de entrada de dados do sistema *PetroGrapher* (a interface de contagem de pontos). Depois, efetuou a interpretação de seqüência diagenética para cada uma das amostras, utilizando os cartões com os nomes dos constituintes para realizar a ordenação (experiência relatada na sessão 5.2.13).

A descrição das amostras, feita pelo especialista, foi então carregada no módulo de interpretação de seqüência diagenética, com o objetivo de comparar a sugestão de interpretação feita pelo algoritmo com a interpretação do especialista. Os resultados da interpretação da amostra WINT-01 SWC8 são descritos a seguir. Os resultados de outras amostras analisadas foram resumidos em uma tabela no fim da seção.

### 8.5.1 Interpretação da amostra WINT01 SWC8 - 4616,5

A interpretação feita pelo especialista para essa amostra, mostrada na Figura 8.12, não resultou em uma ordenação total, ou seja, os eventos diagenéticos não puderam ser dispostos em uma seqüência linear. Isso se deve a um fato já explicado na transcrição da entrevista, na sessão 5.2.13, que a amostra não foi descrita com feições suficientes para que pudessem ser inferidas todas as relações de ordem entre os eventos.

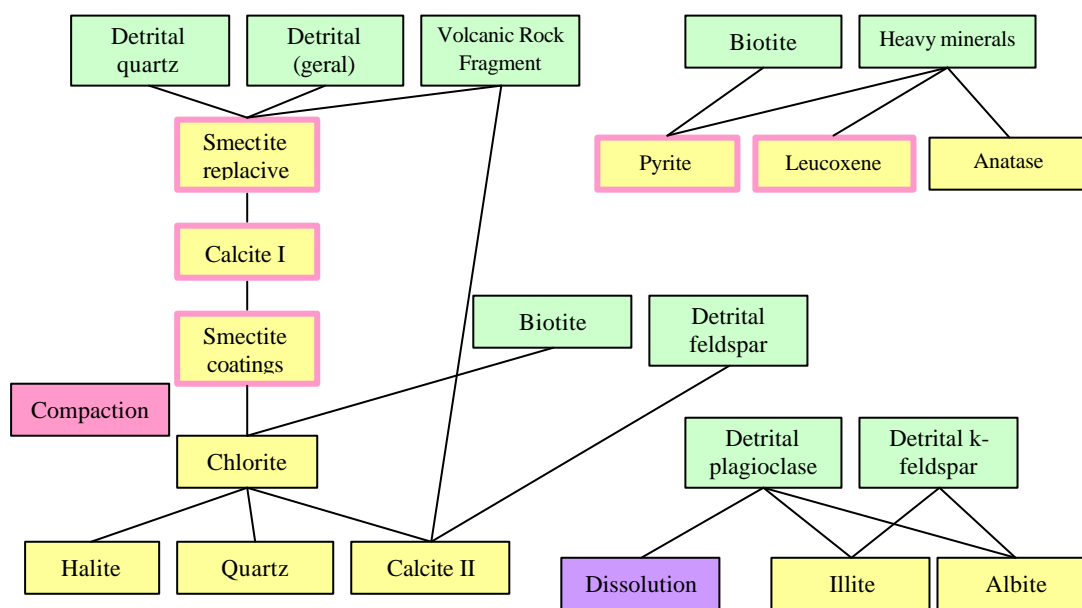


Figura 8.12: Interpretação da seqüência diagenética feita pelo especialista (amostra SWC8)

A interpretação da seqüência de eventos dessa mesma amostra, efetuada pelo algoritmo de inferência é mostrada na Figura 8.13.

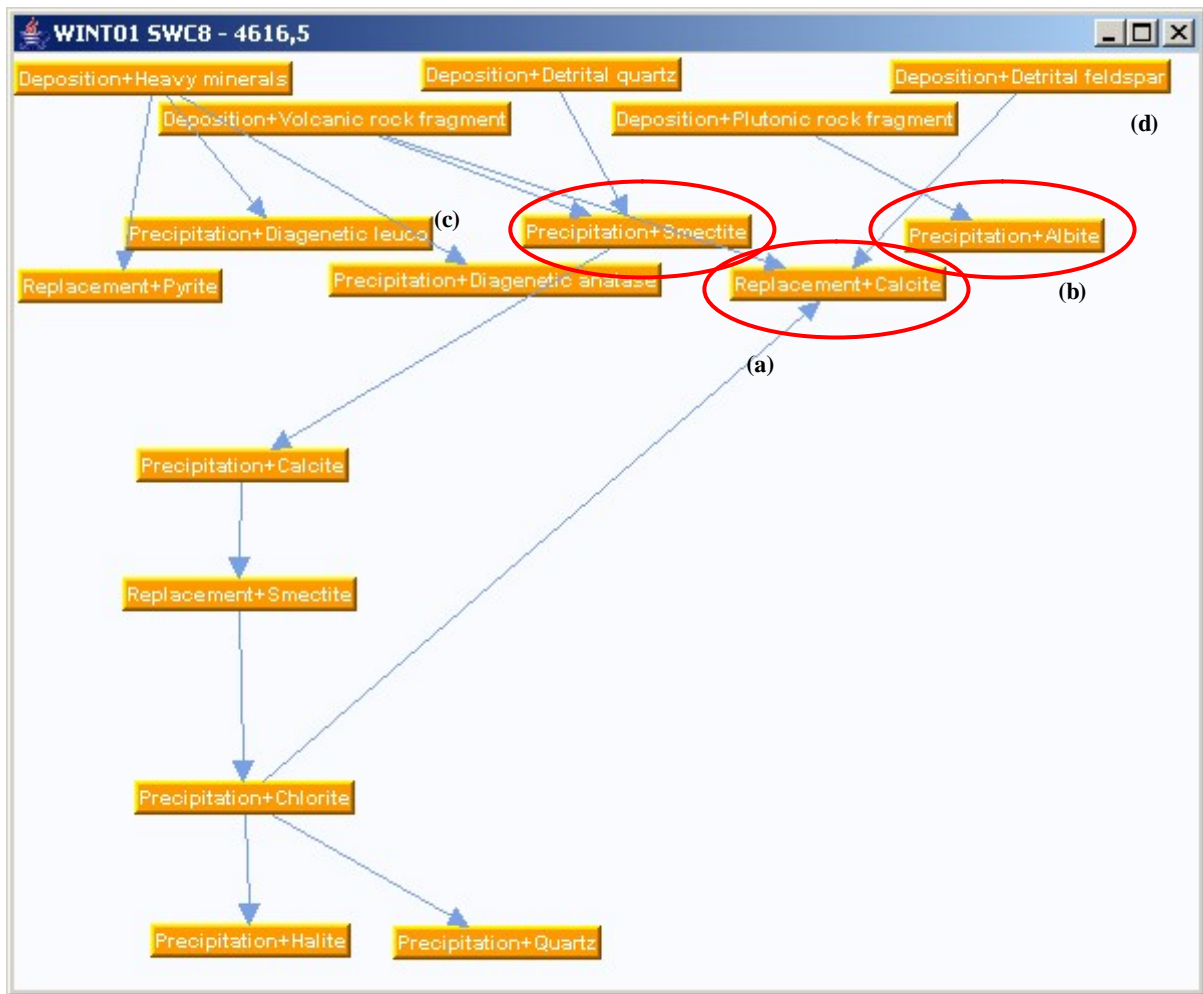


Figura 8.13: Sugestão da seqüência diagenética pelo algoritmo de inferência (amostra SWC8)

Comparando a interpretação feita pelo especialista com a interpretação sugerida pelo algoritmo de inferência é possível notar que a disposição dos eventos não é a mesma. O primeiro motivo é o fato de ainda não ter sido encontrado um algoritmo que desenhe o grafo na tela da maneira que o especialista faria. Pode-se pegar como exemplo a ligação entre a Clorita e a Calcita (Figura 8.13 (a)). A ordenação entre os eventos está correta (a precipitação da clorita veio antes da substituição por calcita), mas o algoritmo desenhou o evento da calcita no segundo nível do grafo, enquanto o especialista colocou esse evento no último nível.

Outro exemplo é o fato de o evento gerador da Albita (Figura 8.13 (b)) estar no mesmo nível dos eventos geradores do Leucoxênio e do Anatásio (Figura 8.13 (c)) (ao contrário do que é observado na interpretação do especialista). Isso é porque todos esses eventos estão ligados aos eventos de deposição dos minerais detriticos, que se encontram sempre no primeiro nível (Figura 8.13 (d)) por serem os primeiros constituintes das rochas. Em consequência disso, os eventos que são ligados somente a eles acabam ficando no segundo nível.

Mas é possível verificar que quando o grafo representa uma seqüência linear, essa seqüência é representada fielmente pelo algoritmo de desenho na tela. A amostra S208, que é apresentada como exemplo do capítulo para a descrição do algoritmo de



ordenação, e cujo resultado de ordenação pode ser conferido na Figura 8.11, é um caso em que o desenho do grafo segue fielmente a interpretação do especialista.

As outras amostras descritas também tiveram a sugestão de interpretação do algoritmo comparada com a interpretação do especialista. Existem algumas conclusões a respeito dessa comparação:

A parte do algoritmo descrita em 8.2.2, em que o constituinte associado é buscado na lista de constituintes da qual ele faz parte: como explicado anteriormente, essa busca é feita comparando-se os atributos do constituinte descrito na relação paragenética com os constituintes da lista. Mas às vezes o constituinte não é descrito de maneira completa e o casamento de padrões dos atributos pode resultar errado. Foram detectados alguns constituintes que não foram encontrados, assim, o evento gerador não aparece na ordenação.

O algoritmo de ordenação: para aqueles constituintes que foram encontrados o algoritmo consegue identificar os eventos e ordená-los corretamente no grafo.

O desenho dos grafos na tela: é preciso melhorar o algoritmo que desenha a ordenação dos eventos na tela, para que esse desenho fique o mais parecido possível com a interpretação que o especialista daria.

A ordenação dos eventos a partir das relações paragenéticas é solucionada pelo algoritmo, mas a identificação dos eventos a partir das feições ainda precisa ser melhorada. Mas isso depende do especialista descrever as feições necessárias para que o constituinte seja identificado por casamento de padrões.

## 9 CONCLUSÃO

A exploração de petróleo é uma atividade que demanda grandes quantidades de conhecimento especialista para ser realizada. A inteligência artificial vem buscando compreender e automatizar o conhecimento estratégico envolvido na solução das tarefas referentes à caracterização de depósitos de petróleo. Uma dessas tarefas é a *interpretação da seqüência de eventos diagenéticos*, que é realizada no domínio da Petrografia Sedimentar.

Neste trabalho foi desenvolvido um estudo sobre o domínio da petrografia sedimentar e o conhecimento envolvido para realizar a tarefa de interpretação de seqüência diagenética. A partir da compreensão do domínio assim obtida foram propostos novos construtos como uma ontologia de representação de conhecimento que possibilitam a representação de aspectos temporais em ontologias de domínio. Essa ontologia foi usada como base para a definição de um modelo de conhecimento que estende a ontologia da petrografia com primitivas temporais. Esse modelo foi validado através da implementação de um módulo de inferência no sistema *PetroGrapher*.

### 9.1 Aquisição e Representação de Conhecimento para Interpretação de Seqüência Diagenética

O estudo de caso de aquisição de conhecimento demonstrou a aplicação de algumas técnicas de eliciação de conhecimento para identificar, coletar e organizar o conhecimento necessário para a realização de uma tarefa específica do domínio da petrografia sedimentar: a interpretação de seqüência diagenética. A aplicação das técnicas teve um caráter de avaliação informal, pois foram descritas as vantagens e desvantagens de se aplicar cada uma das técnicas. A conclusão da aplicação dessas técnicas é de que não existe uma técnica de eliciação que seja específica para eliciar conhecimento sobre aspectos temporais. As técnicas existentes precisam ser modificadas para atingir esse objetivo, assim como foi feito no estudo de caso, para eliciar as regras que indicam a relação de ordem entre os eventos.

A partir das sessões de AC foi possível entender que tipo de conhecimento era preciso ser modelado para permitir a inferência sobre seqüência diagenética. Os aspectos temporais do domínio são as ordens de tempo entre eventos (*antes, depois, etc.*) e essa ordenação é sugerida a partir das características atuais do domínio. As regras que indicam a relação de ordem a partir de feições do domínio são o principal conhecimento a ser representado para possibilitar a inferência sobre a seqüência de eventos.

Para sugerir uma representação adequada para essas regras, foram estudados os principais modelos para representação de tempo. A partir do modelo de *cálculo de eventos* chegou-se na idéia de representar as regras como uma implicação entre relações do domínio: uma relação entre as características do domínio indica uma relação de

ordem entre os eventos que geraram essas características. No caso do domínio da petrografia, a relação paragenética entre constituintes indica a relação de ordem entre os eventos diagenéticos que geraram esses constituintes.

Além dessas regras, foram identificados os outros objetos do domínio necessários para realizar essa inferência, os quais são: as feições, os eventos, as relações entre feições (associação) e as relações temporais. Para representar essa conceitualização, foi proposta uma *ontologia de representação de conhecimento baseada em eventos*, que tem o objetivo de servir como base conceitual para a definição de ontologias de domínio incrementadas com aspectos temporais. A ontologia de eventos pode ser empregada tanto para definir a representação do conhecimento, quanto para dirigir o processo de aquisição de conhecimento, pois conhecendo os tipos de construtos que se deve identificar no domínio pode-se definir uma metodologia de eliciação focada para esse objetivo. No sentido da representação de conhecimento, o objetivo é que os conceitos da ontologia de eventos sejam mapeados para os conceitos específicos dos domínios de aplicação, enquanto os construtos propostos (evento e relação-temporal) sejam usados para definir os eventos e relações de ordem específicos do domínio. Por exemplo, o conceito *feição* deve ser mapeado para um objeto que represente a descrição de características relevantes do domínio.

Para formalizar o conhecimento específico do domínio da petrografia, eliciado das sessões de AC, foi empregada uma metodologia de engenharia de conhecimento (metodologia CommonKADS), gerando um *modelo de conhecimento para a petrografia sedimentar*. A ontologia de eventos proposta foi usada como base para definir a parte declarativa do conhecimento: os construtos da ontologia foram mapeados para os objetos específicos do domínio da petrografia, definindo o *modelo de conhecimento de domínio*. Os passos de inferência e os papéis de conhecimento necessários para se realizar a inferência foram representados como o *modelo de conhecimento de inferência*. A parte procedimental (os passos de raciocínio para a interpretação da seqüência diagenética) foi representada como um método de raciocínio, o *modelo de conhecimento de tarefa*, que indica como os passos de inferência devem ser empregados para se atingir o objetivo da interpretação.

As propostas aqui apresentadas foram validadas através da implementação, teste e validação de um módulo de interpretação para o Sistema *PetroGrapher*, uma aplicação de banco de dados inteligente para apoio à descrição e interpretação de rochas sedimentares siliciclásticas. O módulo desenvolvido aplica o conhecimento representado no modelo sobre descrições de amostras de rocha para inferir a seqüência de eventos diagenéticos que gerou os constituintes dessa amostra. O algoritmo que infere a ordenação dos eventos a partir das relações paragenéticas consegue solucionar essa tarefa. Os grafos analisados apresentam a ordem de eventos da mesma maneira como ordenada pelo especialista. A única diferença se encontra na disposição na tela dessa ordem. O algoritmo ainda precisaria ser melhorado para disponibilizar os eventos da mesma maneira que o especialista faria, e isso requer mais algumas entrevistas para coletar do especialista alguns critérios de organização dos eventos.

## 9.2 Outros domínios

Além do domínio da Petrografia, para o qual é proposta uma ontologia no capítulo seguinte, outros domínios podem se enquadrar na conceitualização da ontologia de eventos. Um dos exemplos é o domínio da glaciologia, em que os componentes químicos encontrados no gelo podem ser vistos como feições do domínio, e a partir do

relacionamento entre os componentes (de profundidade) é possível inferir os eventos climáticos que ocorreram na terra e a ordem em que eles ocorreram.

Foram feitas algumas tentativas para enquadrar a ontologia de eventos em outros domínios, como, por exemplo, o da engenharia química. Nessa tentativa, verificou-se que esse domínio não era adequado porque os processos da engenharia química são previsíveis e controlados, podendo-se fazer analogia com uma receita. Ou seja, não é necessário identificar os processos químicos que unidos formaram um determinado composto químico, pois esses processos são sempre os mesmos e ocorrem na mesma ordem.

Dessa experiência pode-se imaginar que os domínios mais aplicáveis para essa ontologia são aqueles em que os eventos ou processos não são previsíveis ou controlados por fonte humana, mas sim, aleatórios. Por isso a ontologia de eventos tende a ser mais bem aplicada para domínios naturais em vez de domínios exatos, em que os eventos ocorrem de forma controlada.

### 9.3 Sugestões para trabalhos futuros

A ontologia de representação de conhecimento baseada em eventos foi proposta para definir construtos necessários para a interpretação de seqüência de eventos geradores de características do domínio. Ela foi validada no domínio da petrografia, sendo usada como base conceitual para a instanciação de um modelo de conhecimento específico, o qual foi implementado como um módulo de inferência do sistema *PetroGrapher*. Entretanto, essa ontologia de eventos precisa ser aplicada em outros domínios, para que seja validada sua capacidade de representar genericamente a conceitualização necessária para esse tipo de tarefa.

Os construtos propostos nesse trabalho, os quais foram o construto para definir eventos e o construto para definir relações temporais, foram descritos usando a BNF de uma linguagem de representação de conhecimento de domínio, a linguagem CML da metodologia CommonKADS. Mas essa linguagem é definida somente por sua BNF, ou seja, somente a sintaxe dos construtos pode ser empregada. Não é possível definir uma semântica. Por isso, ainda não é possível garantir que a semântica dos construtos é diferente. O trabalho a ser feito no futuro é estudar uma possibilidade de definir a semântica dos construtos de uma maneira genérica, sem estar associada a uma linguagem específica, como a CML.

As regras de implicação que criam uma dependência entre as relações paragenéticas entre constituintes e as relações de ordem entre eventos são o conhecimento principal que deve ser coletado do domínio. Neste estudo de caso, o conhecimento foi eliciado em entrevistas com o especialista, pois ainda não era conhecida a estrutura das regras e quais os conceitos faziam parte da relação de implicação. Mas para o processo de aquisição de conhecimento em outros domínios, é sugerido que seja elaborada uma ferramenta de aquisição de conhecimento que crie uma dependência entre associações de características do domínio e relações de ordem. Essa ferramenta deve se basear nos conceitos da ontologia de eventos, caso seja uma ferramenta genérica. Mas para ser uma ferramenta de AC específica para um domínio é preciso que os conceitos da ontologia genérica sejam mapeados para os conceitos do domínio específicos. A ferramenta então deve suportar a definição de uma implicação entre as associações de características do domínio e as relações de ordem entre os eventos que geraram essas características.

Este trabalho proporcionou uma melhor compreensão da Engenharia de Conhecimento como um processo de desenvolvimento de um sistema baseado em conhecimento desde a identificação e aquisição do conhecimento essencial para a

resolução do problema, passando pela sugestão da melhor abordagem de representação para as informações eliciadas até a implementação e validação do conhecimento adquirido como um sistema.

## REFERÊNCIAS

- ABEL, M. **Estudo da perícia em petrografia sedimentar e sua importância para a engenharia de conhecimento**. 2001. 239 f. Tese (Doutorado em Ciência da Computação) - Instituto de Informática, UFRGS, Porto Alegre.
- ABEL, M.; CASTILHO, J. M. V.; CAMPBELL, J. A. Analysis of expertise for implementing geological expert systems. In: WORLD CONFERENCE IN EXPERT SYSTEMS, 4., 1998, Mexico City, Mexico. **Application of Advanced Information Technologies**: proceedings. New York: Cognizant Communication Office, 1998. v.1, p. 170-177.
- ABEL, M.; SILVA, L. A. L.; CAMPBELL, J. A.; ROS, L. F. D. Knowledge acquisition and interpretation problem-solving methods for visual expertise: A study of petroleum-reservoir evaluation. Aguardando publicação em: Journal of Petroleum Science and Engineering, 2005.
- ABEL, M.; SILVA, L. A. L.; ROS, L. F. D.; MASTELLA, L. S.; CAMPBELL, J. A.; NOVELLO, T. Petrographer: Managing petrographic data and knowledge using an intelligent database application. **Expert Systems with Applications**, [S.l.], v. 26, n. 1, p. 9-18, 2004.
- Allen, J. F. Maintaining knowledge about temporal intervals. **Communications of the ACM**, New York, v. 26, n. 11, p. 832-843, 1983.
- Allen, J. F. Time and time again: The many ways to represent time. **International Journal of Intelligent Systems**, [S.l.], v. 6, p. 341-355, 1991.
- ANGELE, J.; FENSEL, D.; STUDER, R. Domain and task modeling in MIKE. In: SUTCLIFFE, A. **Domain knowledge for interactive system design**. [S.l.]: Chapman & Hall, 1996.
- ANGELE, J.; LAUSEN, G. Ontologies in logic. In: STAAB, S. et al. **Handbook on ontologies**. Berlin: Springer-Verlag, 2004. p. 29-50.
- ANJOS, S. D. et al. Depositional and diagenetic controls on the reservoir quality of lower cretaceous pendencia sandstones, potiguar rift basin, brazil. **AAPG Bulletin**, Tulsa, v. 84, n. 11, p. 1719-1742, 2000.
- ANTONIOU, G.; HARMELEN, F. V. Web ontology language: OWL. In: STAAB, S. et al. **Handbook on ontologies**. Berlin: Springer-Verlag, 2004. p. 67-92.
- BAADER, F. et al. **The description logic handbook: theory, implementation and applications**. Cambridge: Cambridge University Press, 2003. 555 p.
- BACCHUS, F.; TENENBERG, J.; KOOMEN, J. A non-reified temporal logic. In: INTERNATIONAL CONFERENCE ON PRINCIPLES OF KNOWLEDGE REPRESENTATION AND REASONING, KR, 1., 1989, Toronto, Canada. **Proceedings...** Toronto: Morgan Kaufmann, 1989. p. 2-10.

BATERMAN, J. A. et al. **A general organization of knowledge for natural language processing: the penman upper model**. Marina del Rey: USC/Information Sciences Institute, 1990. Technical Report.

BECKETT, D.; MCBRIDE, B. **RDF/XML syntax specification (revised)**. Disponível em: <<http://www.w3.org/TR/rdf-syntax-grammar/>>. Acesso em: dez. 2004.

BERNARD, D.; BORILLO, M.; GAUME, B. Cognitive temporal model for the planning in aircraft maintenance. In: INTERNATIONAL CONFERENCE ON INDUSTRIAL AND ENGINEERING APPLICATIONS OF ARTIFICIAL INTELLIGENCE AND EXPERT SYSTEMS, IEA/AIE, 3., 1990, Charleston, USA. **Proceedings...** New York: ACM, 1991. p. 318-324.

BERTINO, E.; ZARRI, G. P. Intelligent database systems: State of art. In: WORLD CONGRESS CONFERENCE ON EXPERT SYSTEMS, 4., 1998, Mexico City, Mexico. **Proceedings...** [S.l.]: Instituto Tecnológico y de Estudios Superiores de Monterrey, 1998.

BOLOUR, A.; DEKEYSER, L. J. Abstractions in temporal information. **Information Systems**, [S.l.], v. 8, n. 1, p. 41-49, 1983.

BOOSE, J. H. A survey of knowledge acquisition techniques and tools. **Knowledge Acquisition**, [S.l.], v. 1, n. 1, p. 3-37, 1989.

BOOSE, J. H. Knowledge acquisition tools, methods, and mediating representations. In: JAPANESE KNOWLEDGE ACQUISITION FOR KNOWLEDGE-BASED SYSTEMS WORKSHOP, JKAW, 1., 1990, Kyoto, Japan. **Proceedings...** Amsterdam: IOS Press, 1991.

BRICKLEY, D.; GUHA, R. V. **Resource description framework (RDF) schema specification 1.0**. Disponível em: <<http://www.w3.org/TR/2000/CR-rdf-schema-20000327/>>. Acesso em: maio 2004.

BUCHANAN, B.; SHORTLIFFE, E. **Rule-based expert systems: the MYCIN experiments**. Reading: Addison-Wesley, 1984. 748 p.

CHANDRASEKARAN, B.; JOSEPHSON, J. R.; BENJAMINS, V. R. What are ontologies, and why do we need them? **IEEE Intelligent Systems**, [S.l.], v. 14, n. 1, p. 20-26, 1999.

CHAUDHRI, V. K. et al. **Open knowledge base connectivity 2.0.3**. Disponível em: <<http://www.ai.sri.com/~okbc/okbc-2-0-3.pdf>>. Acesso em: dez. 2004.

CLIFFORD, J.; CROKER, A. The historical relational data model (HRDM) revisited. In: TANSEL, A. U. et al. **Temporal databases: theory, design, and implementation**. Redwood City: Benjamin/Cummings Pub., 1993. p. 6-27.

COOKE, N. Varieties of knowledge elicitation techniques. **International Journal of Human-Computer Studies**, [S.l.], v. 41, p. 801-849, 1994.

DEAN, M.; SCHREIBER, G. **OWL web ontology language reference: W3C working draft**. Disponível em: <<http://www.w3.org/TR/owl-ref/>>. Acesso em: dez. 2004.

DICKINSON, W. R. Interpreting detrital modes of graywacke and arkose. **Journal of Sedimentary Petrology**, [S.l.], v. 40, p. 695-707, 1970.

EDELWEISS, N.; OLIVEIRA, J. P. M. D. **Modelagem de aspectos temporais de sistemas de informação**. Recife: UFPE, 1994. 163 p.

ELMASRI, R.; WUU, G. T.; KOURAMAJIAN, V. A temporal model and query language for EER databases. In: TANSEL, A. U. et al. **Temporal databases: theory, design, and implementation**. Redwood City: Benjamin/Cummings Pub., 1993. p. 212-229.

- ERIKSSOM, H. et al. Task modeling with reusable problem solving methods. **Artificial Intelligence**, [S.l.], v. 79, n. 2, p. 293-326, 1995.
- FARQUHAR, A.; FIKES, R.; RICE, J. Ontolingua server: A tool for collaborative ontology construction. **International Journal of Human-Computer Studies**, [S.l.], v. 46, n. 6, p. 707-727, 1997.
- FIKES, R.; KEHLER, T. The role of frame-based representation in reasoning. **Communications of the ACM**, [S.l.], v. 238, n. 9, p. 904-920, 1985.
- FOLK, R. L. **Petrology of sedimentary rocks**. Austin: Hemphill Pub. Co, 1974. 170 p.
- FOWLER, M.; SCOTT, K. **UML essencial: um breve guia para a linguagem padrão de modelagem de objetos**. Porto Alegre: Bookman, 2000. 169 p.
- GALTON, A. **Temporal logic**: entry on the stanford encyclopedia of philosophy. Disponível em: <http://plato.stanford.edu/entries/logic-temporal/#1>. Acesso em: out. 2004.
- GAPPA, U.; PUPPE, F. A study of knowledge acquisition: Experiences from the sisyphus III experiment for rock classification. In: WORKSHOP ON KNOWLEDGE ACQUISITION, MODELING AND MANAGEMENT, KAW, 11., 1998, Banff, Canada. **Proceedings...** [S.l.:s.n.], 1998.
- GÓMEZ-PÉREZ, A.; CORCHO, O. Ontology specification languages for the semantic web. **IEEE Intelligent Systems**, [S.l.], v. 17, n. 1, p. 54-60, 2002.
- GÓMEZ-PÉREZ, A.; FERNÁNDEZ-LÓPEZ, M.; CORCHO, O. The most outstanding ontologies. In: WU, X. et al. **Ontological engineering: with examples from the areas of knowledge management, e-commerce and the semantic web**. Berlin: Springer-Verlag, 2004. p. 47-106.
- GROVER, M. D. A pragmatic knowledge acquisition methodology. In: INTERNATIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE, IJCAI, 8., 1983, Karlsruhe, West Germany. **Proceedings...** [S.l.]: William Kaufmann, 1983. v.1, p. 436-438.
- GRUBER, T. R. A translation approach to portable ontology specifications. **Knowledge Acquisition**, [S.l.], v. 5, p. 199-220, 1993.
- GUARINO, N.; WELTY, C. A. A formal ontology of properties. In: EUROPEAN WORKSHOP ON KNOWLEDGE ACQUISITION, MODELING AND MANAGEMENT, EKAW, 12., 2000, Juan-les-Pins, France. **Proceedings...** Berlin: Springer-Verlag, 2000. p. 97-112.
- HAUGH, B. A. Non-standard semantics for the method of temporal arguments. In: INTERNATIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE, IJCAI, 10., 1987, Milan, Italy. **Proceedings...** [S.l.]: Morgan Kaufmann, 1987. v.1, p. 449-455.
- HAYES, P. **A catalog of temporal theories**. Chicago: Beckman Institute/University of Illinois, 1995. Technical Report. Disponível em: <http://www.coginst.uwf.edu/ephayes/TimeCatalog1.ps>. Acesso em: dez. 2004.
- HAYES-ROTH, F.; WATERMAN, D. A.; LENAT, D. B. **Building expert systems**. Reading: Addison-Wesley, 1983.
- HIGLEY, D. **3-d reservoir characterization of the house creek oil field: diagenetic history of the sussex "b" sandstone**. Disponível em: [http://pubs.usgs.gov/dds/dds-033/USGS\\_3D/ssx\\_txt/diagenes.htm](http://pubs.usgs.gov/dds/dds-033/USGS_3D/ssx_txt/diagenes.htm). Acesso em: jul. 2004.
- HOFFMAN, R. et al. Eliciting knowledge from experts: A methodological analysis. **Organizational Behavior and Decision Processes**, [S.l.], v. 62, n. 2, p. 129-158, 1995.



- HOFFMAN, R. R. The problem of extracting the knowledge of experts from the perspective of experimental psychology. **Artificial Intelligence**, [S.l.], v. 8, n. 2, p. 53-67, 1987.
- HOFFMANN, R. R. Doing psychology in an AI context: A personal perspective and introduction to this volume. In: HOFFMANN, R. R. **The psychology of expertise: cognitive research and empirical AI**. New York: Springer-Verlag, 1992. p. 3-11.
- HOLANDA, A. B. D. **Novo aurélio século XXI**. 3 ed. Rio de Janeiro: Nova Fronteira, 1999. 2128 p.
- HORROCKS, I. et al. OIL in a nutshell. In: EUROPEAN WORKSHOP ON KNOWLEDGE ACQUISITION, MODELING AND MANAGEMENT, EKAW, 12., 2000, Juan-les-Pins, France. **Proceedings...** Berlin: Springer-Verlag, 2000. p. 1-16.
- HORROCKS, I.; HARMELEN, F. V.; PATEL-SCHNEIDER, P. F. **Reference description of the DAML+OIL (march 2001) ontology markup language**. Disponível em: <<http://www.daml.org/2001/03/reference.html>>. Acesso em: dez. 2004.
- HUTH, M.; RYAN, M. **Logic in computer science : modelling and reasoning about systems**. Cambridge: Cambridge University Press, 2000. 387 p.
- JANSEN, M. G.; SCHREIBER, A. T.; WIELINGA, B. J. Rocky III - round 1: A progress report. In: WORKSHOP ON KNOWLEDGE ACQUISITION, MODELING AND MANAGEMENT, KAW, 11., 1998, Banff, Canada. **Proceedings...** [S.l.:s.n.], 1998.
- JENSEN, C. S. et al. **The consensus glossary of temporal database concepts**. Disponível em: <<http://www.cs.auc.dk/~csj/Glossary/>>. Acesso em: out. 2004.
- JOHNSON, P. **Human-computer interaction: psychology, task analysis and software engineering**. Maidenhead: McGraw-Hill, 1992.
- KELLY, G. A. **The psychology of personal constructs**. New York: WW Norton, 1955.
- KOWALSKI, R.; SERGOT, M. Logic-based calculus of events. **New Generation Computing**, [S.l.], v. 4, n. 1, p. 67-95, 1986.
- LEÃO, B. D. F. **Construção de uma base de conhecimento de um sistema especialista de apoio ao diagnóstico de cardiopatias congênitas**. 1988. 230 f. Tese (Doutorado em Medicina) - Escola Paulista de Medicina, UNIFESP, São Paulo.
- LENAT, D. B.; GUHA, R. V. **Building large knowledge-based systems: representation and inference in the CYC project**. Boston: Addison-Wesley, 1990.
- LIU, Y. I. Knowledge acquisition: Issues, techniques, and methodology. In: CONFERENCE ON TRENDS AND DIRECTIONS IN EXPERT SYSTEMS, SIGBDP, 1990, Orlando, USA. **Proceedings...** New York: ACM, 1990. p. 212 - 236.
- MACGREGOR, R. Inside the LOOM classifier. **SIGART Bulletin**, [S.l.], v. 2, n. 3, p. 70-76, 1991.
- MAHESH, K.; NIRENBURG, S. Semantic classification for practical natural language processing. In: CLASSIFICATION RESEARCH WORKSHOP: AN INTERDISCIPLINARY MEETING - ADVANCES IN CLASSIFICATION RESEARCH, ASIS SIG/CR, 6., 1995, Chicago, USA. **Proceedings...** Medford: Learned Information, Inc, 1995. p. 79-94.
- MAIOCCHI, R.; PERNICI, B. Temporal data management systems: A comparative view. **IEEE Transactions on Knowledge and Data Engineering**, [S.l.], v. 3, n. 4, p. 504-524, 1991.
- MASTELLA, L. S. **Análise de formas de representação para modelar conhecimento inferencial**. 2002. 72 f. Projeto de Diplomação (Bacharelado em Ciência da Computação) - Instituto de Informática, UFRGS, Porto Alegre.

- MASTELLA, L. S. **Técnicas de aquisição de conhecimento para sistemas baseados em conhecimento**. 2004. 38 f. Trabalho Individual I (Mestrado em Ciência da Computação) - Instituto de Informática, UFRGS, Porto Alegre.
- MARTHUR, R. P. **Tense logic**. Dordrecht: Reidel Publishing Company, 1976.
- MCCARTHY, J. **Elaboration tolerance**. Stanford: Computer Science Department/Stanford University, 1999. 24 f. Technical Report. Disponível em: <<http://citeseer.ist.psu.edu/mccarthy99elaboration.html>>. Acesso em: nov. 1999.
- MCCARTHY, J.; HAYES, P. Some philosophical problems from the standpoint of artificial intelligence. In: MELTZER, B. et al. **Machine intelligence 4**. Edinburgh: Edinburgh University Press, 1969. p. 463-502.
- MCDERMOTT, D. A temporal logic for reasoning about processes and plans. **Cognitive Science**, [S.l.], v. 6, p. 101-155, 1982.
- MILLER, G. A. et al. Introduction to wordnet: An on-line lexical database. **International Journal of Lexicography**, [S.l.], v. 3, n. 4, p. 235-244, 1990.
- MONTANARI, A.; PERNICI, B. Temporal reasoning. In: TANSEL, A. U. et al. **Temporal databases: theory, design, and implementation**. Redwood City: Benjamin/Cummings Pub., 1993.
- MORAD, S.; KETZER, J. M.; DE ROS, L. F. Spatial and temporal distribution of diagenetic alterations in siliciclastic rocks: Implications for mass transfer in sedimentary basins. **Sedimentology**, [S.l.], v. 47, p. 95-120, 2000.
- NEWELL, A. The knowledge level. **Artificial Intelligence**, [S.l.], v. 18, n. 1, p. 87-127, 1982.
- NONAKA, I.; TAKEUCHI, H. **The knowledge-creating company: how japanese companies create the dynamics of innovation**. New York: Oxford University Press, 1995.
- O'HARA, K. et al. **The vital knowledge acquisition methodology**. 1992. 26 f. ESPRIT II Project Intermediary Report.
- PARSAYE, K. et al. **Intelligent databases: object oriented, deductive hypermedia technologies**. New York: John Wiley & Sons, 1989. 479 p.
- PEASE, R. A.; NILES, I. IEEE standard upper ontology: A progress report. **The Knowledge Engineering Reviews**, [S.l.], v. 17, n. 1, p. 65-70, 2002.
- POLANYI, M. **Personal knowledge**. Chicago: The University of Chicago Press, 1974.
- PUERTA, A. R. et al. A multiple-method knowledge-acquisition shell for the automatic generation of knowledge-acquisition tools. **Knowledge Acquisition**, [S.l.], v. 4, n. 2, p. 171-196, 1992.
- REBOH, R. **Knowledge engineering techniques and tools in the PROSPECTOR environment**. Palo Alto: Computer Science Laboratory/Stanford Research Institute International, 1981. Technical Report.
- RESCHER, N.; URQUHART, A. **Temporal logic**. New York: Springer-Verlag, 1971.
- RUAS, R.; ANTONELLO, C. S.; BOFF, L. H. Autodesenvolvimento e competências: O caso do trabalhador de conhecimento como especialista. In: RUAS, R. et al. **Os novos horizontes da gestão: aprendizagem organizacional e competências**. Porto Alegre: Bookman, 2005. p. 70-86.
- SCHREIBER, G. et al. **Knowledge engineering and management: the CommonKADS methodology**. Cambridge: The MIT Press, 2000. 465 p.

- SHADBOLT, N.; BURTON, M. Knowledge elicitation techniques: Some experimental results. In: MCGRAW, K. L. et al. **Readings in knowledge acquisition: current practices and trends**. [S.l.]: Ellis Horwood, 1990.
- SHADBOLT, N.; SWALLOW, S. **Epistemics**. Disponível em: <<http://www.epistemics.co.uk/>>. Acesso em: jun. 2003.
- SHOHAM, Y. Ten requirements for a theory of change. **New Generation Computing**, [S.l.], v. 3, n. 4, p. 467-477, 1985.
- SHOHAM, Y. Temporal logics in AI: Semantical and ontological considerations. **Artificial Intelligence**, [S.l.], v. 33, n. 1, p. 89-104, 1987.
- SILVA, L. A. L. **Aplicando métodos de solução de problemas em tarefas de interpretação de rochas**. 2001. 157 f. Dissertação (Mestrado em Ciência da Computação) - Instituto de Informática, UFRGS, Porto Alegre.
- SOWA, J. F. **Knowledge representation: logical, philosophical, and computational foundations**. Pacific Grove: Brooks Cole Publishing Co., 2000. 594 p.
- SRIPADA, S. A logical framework for temporal deductive databases. In: INTERNATIONAL CONFERENCE ON VERY LARGE DATA BASES, VLDB, 14., 1988, Los Angeles, USA. **Proceedings...** [S.l.]: Morgan Kaufmann, 1988. p. 171-182.
- STENMARK, D. Information vs. Knowledge: The role of intranets in knowledge management. In: INTERNATIONAL CONFERENCE ON SYSTEM SCIENCES, HICSS, 35., 2002, Big Island, Hawaii. **Proceedings...** [S.l.]: IEEE Computer Society Press, 2002. p. 928-937.
- STUDER, R.; BENJAMINS, V. R.; FENSEL, D. Knowledge engineering: Principles and methods. **Data & Knowledge Engineering**, [S.l.], v. 25, n. 1/2, p. 161-197, 1998.
- SWARTOUT, B. et al. Toward distributed use of large-scale ontologies. In: SPRING SYMPOSIUM SERIES ON ONTOLOGICAL ENGINEERING, AAAI, 1997, Stanford University, USA. **Proceedings...** [S.l.]: AAAI Press, 1997. p. 138-148.
- TANSEL, A. U. et al. **Temporal databases: theory, design, and implementation**. Redwood City: Benjamin/Cummings Pub., 1993.
- TURBAN, E. **Expert systems and applied artificial intelligence**. New York: Macmillan Publishing Company, 1992. 804 p.
- VELDE, W. V. Issues in knowledge level modeling. In: DAVID, J. M. et al. **Second generation expert systems**. Berlin: Springer-Verlag, 1993.
- VENDLER, Z. Verbs and times. In: **Linguistics in philosophy**. Ithaca: Cornell University Press, 1957. p. 97-121.
- VILA, L. Survey on temporal reasoning in artificial intelligence. **AI Communications**, [S.l.], v. 7, n. 1, p. 4-28, 1994.
- VOSSSEN, P. (Ed.). **Eurowordnet: a multilingual database with lexical semantic network**. Dordrecht: Kluwer Academic Publishers, 1998.
- WIELINGA, B. J.; SCHREIBER, A. T.; BREUKER, J. A. KADS: A modelling approach to knowledge engineering. **Knowledge Acquisition**, [S.l.], v. 4, n. 1, p. 5-53, 1992.
- WIELINGA, B. J. et al. The KADS knowledge modeling approach. In: JAPANESE KNOWLEDGE ACQUISITION FOR KNOWLEDGE-BASED SYSTEMS WORKSHOP, JKAW, 2., 1992, Hatoyama, Japan. **Proceedings...** [S.l.]: Advanced Research Laboratory, 1992. p. 23-42.
- WRIGHT, G.; AYTON, P. **Eliciting and modeling expert knowledge**. [S.l.]: Elsevier, 1987. 26 p.

ZIULKOSKI, L. C. C. **Coleta de requisitos e modelagem de dados para data warehouse:** um estudo de caso utilizando técnicas de aquisição de conhecimento. 2003. 63 f. Projeto de Diplomação (Bacharelado em Ciência da Computação) - Instituto de Informática, UFRGS, Porto Alegre.

ZUFFA, G. G. **Provenance of arenites.** Dordrecht: Reidel Publishing Company, 1985. 408 p.

# APÊNDICE A TRANSCRIÇÕES DE ENTREVISTAS COM O ESPECIALISTA EM PETROGRAFIA SEDIMENTAR

Nesse apêndice serão demonstradas as transcrições das sessões de entrevistas feitas com o especialista em Petrografia Sedimentar e alguns materiais (tais como gráficos, tabelas, listas, entre outros) resultantes das aplicações de técnicas de aquisição de conhecimento constritas, apresentada na seção 2.3.

## A.1 Regras Temporais

Uma das entrevistas que serviu de base para a proposta do modelo de interpretação de seqüência foi a que o especialista deixou transparecer as regras utilizadas para ordenar o aparecimento de constituintes baseado nas relações paragenéticas visualizadas na rocha. Essas regras foram coletadas durante a observação da interpretação de uma amostra, em que o especialista foi convidado a externalizar o motivo da ordenação. Segue a transcrição das partes dessa entrevistas que geraram as regras de ordenação do modelo<sup>14</sup>.

*Especialista: "O que está sobre as cutículas? O que está na seqüência seguinte? Nós temos as cutículas de argila e em cima das cutículas de argila nós temos esses cristais. Esses cristais me parecem ser o seguinte na seqüência, porque em muitos lugares eles estão diretamente sobre essas cutículas."*

SE constituinte.RelacaoParagenética = Cobrindo <outro-constituinte>  
ENTÃO (other\_constituent < constituinte)

*Especialista: "Nós temos uma espécie de uma expansão do grão, está além do limite original, que está marcado pela cutícula, overgrowth, crescimento sobre crescimento de quartzo. Qual é a relação desse crescimento de quartzo com a dolomita? A dolomita está dentro do crescimento. No caso do crescimento está engolfando a dolomita, então isso indica que o quartzo é posterior à dolomita."*

SE constituinte.RelacaoParagenética = Engolfando <outro-constituinte>  
ENTÃO (other\_constituent < constituinte)

SE constituinte.RelacaoParagenética = Engolfado por <outro-constituinte>

---

<sup>14</sup> Nas regras de ordenação serão utilizados os símbolos < e > para indicar respectivamente a noção de antes e depois.

ENTÃO (other\_constituent > constituinte)

***Especialista:** "A anidrita parece estar engolfando a dolomita, aqui é a mesma coisa. Ela não só está engolfando, mas também está corroendo ou substituindo as margens dos cristais de dolomita, então isso indica que essa anidrita é posterior à dolomita."*

SE constituinte.RelacaoParagenética = Engolfando <outro-constituinte>  
 OU constituinte.RelacaoParagenética = Cobrindo <outro-constituinte>  
 OU constituinte.RelacaoParagenética = Substituindo <outro-constituinte>  
 ENTÃO (other\_constituent < constituinte)

***Especialista:** "Tem bastante desse material fibroso. É uma argila diagenética chamada Ilita. Aqui tem bastante dolomita também, que também substituiu esse grão, e houve dissolução. Essa porosidade de dissolução foi preenchida por esse material, dolomita. Bom, então a geração de porosidade é pós-dolomita, a gente já viu a dolomita dissolvida também. Então a porosidade seria pelo menos posterior à dolomita. Você tem ilita onde? A ilita está dentro da porosidade intragranular. Então ela veio depois da porosidade."*

SE constituinte.RelacaoParagenética = dentro da porosidade intragranular  
 E porosidade.RelacaoParagenética = Dissolução-de <constituinte>  
 ENTÃO (constituinte > porosidade)

SE constituinte.Modificador = dissolvido  
 E porosidade.RelacaoParagenética = Dissolução-de <constituinte>  
 ENTÃO (porosidade > constituinte)

## A.2 Classificação de Fichas

A técnica de aquisição de conhecimento de classificação de fichas gera um conjunto de conceitos do domínio agrupados pelo seu significado para o especialista. A técnica de classificação de fichas foi empregada para diferenciar a semântica que o especialista dá para alguns conceitos, e o resultado pode ser visto na Figura 1.

## A.3 Processos Diagenéticos

O especialista elaborou uma listagem das principais categorias de processos diagenéticos, os quais serão os objetos a serem ordenados pelo mecanismo de inferência. A listagem é como segue:

- Dissolução (*dissolution*): destruição de grãos ou constituintes diagenéticos, com a geração de poros secundários.
- Autigênese (*authigenesis*): precipitação de novos minerais:
  - Cimentação (*cementation*): precipitação de minerais diagenéticos nos poros. Preenchimento dos espaços intersticiais promove a inibição ou interrupção da compactação.

- Substituição (*replacement*): dissolução de um mineral e precipitação simultânea de outro em seu lugar.

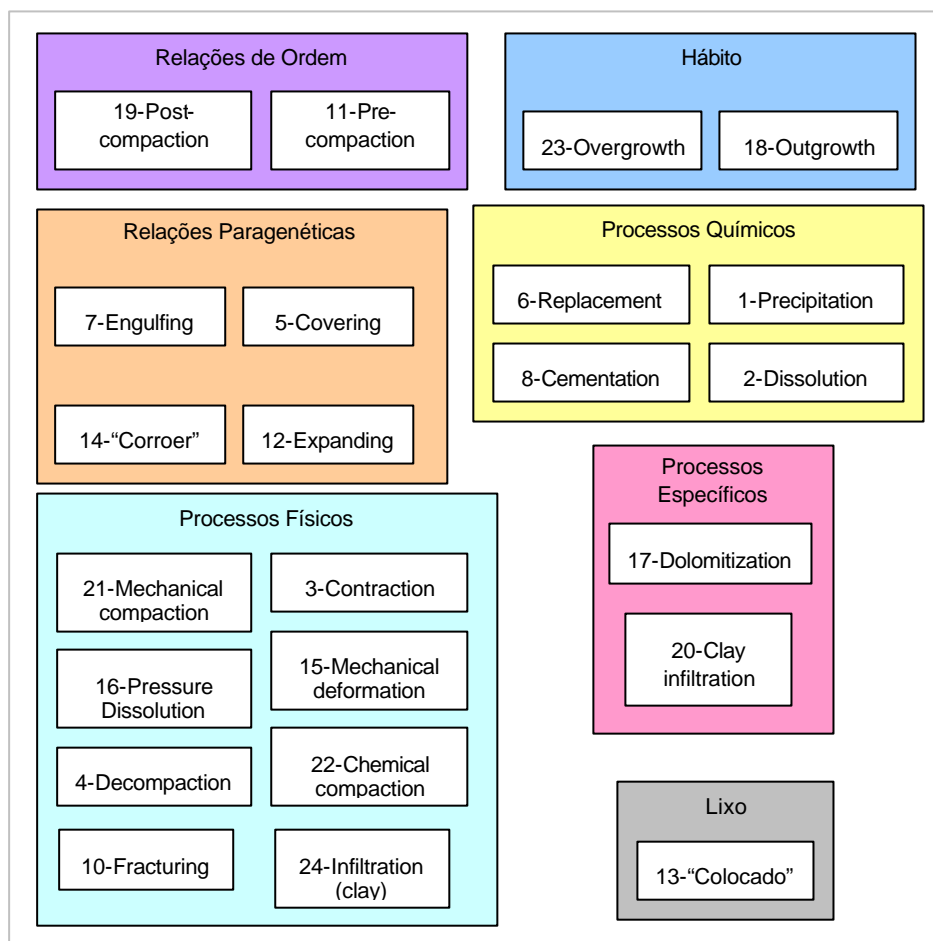


Figura 1: Resultado da classificação de fichas usando nomes de processos diagenéticos e relações paragenéticas.

- Hidratação (*hydration*) – Desidratação (*dehydration*): entrada ou saída de água da estrutura cristalina dos minerais. Desidratação de argilo-minerais comumente promove a geração de poros de contração (*shrinkage*).
- Oxidação (*oxidation*): na ou sob influência da superfície;  $O_2$ , bactérias aeróbicas. Formação de óxidos de ferro (hematita) substituindo grãos ou cimentos de minerais contendo  $Fe^{2+}$ .
- Redução (*reduction*): sob influência da matéria orgânica e de bactérias anaeróbicas. Minerais contendo  $Fe^{3+}$  são destruídos, podendo ser precipitados minerais contendo  $Fe^{2+}$  (ex.: pirita, clorita).
- Recristalização (*recrystallization*): crescimento ou diminuição do tamanho cristalino, mantendo-se a mesma composição mineralógica..
- Estabilização (*stabilization*) / neomorfismo (*neomorphism*): substituição por uma fase mineralógica de composição similar, mas mais estável nas condições diagenéticas.
- Compactação (*compaction*): causada pelo soterramento; redução do espaço entre as partículas ocupado por fluidos (poros); deformação de algumas partículas.

- Compactação Mecânica (Física) (*mechanical or physical compaction*): envolve quatro principais processos puramente físicos de rearranjo ou deformação das partículas:
  - Rearranjo (*rearrangement*): empacotamento ideal de esferas.
  - Fraturamento (*fracturing*): grãos rígidos, como quartzo e feldspatos podem ser fraturados na compactação.
  - Deformação Elástica (*elastic deformation*): praticamente todos minerais possuem certa elasticidade e podem ser bastante deformados na compactação.
  - Deformação Plástica (*plastic or ductile deformation*): grãos dúcteis, macios, são extensamente deformados pseudoplasticamente devido à compactação.
- Compactação Química – dissolução por pressão (*chemical compaction – pressure dissolution*): dissolução pela concentração de esforços ao longo dos contatos intergranulares.
- Infiltração mecânica (*mechanical infiltration*): processo de introdução de material fino, normalmente argila, nos poros de um sedimento ou rocha. O material decanta sobre os grãos formando cutículas (*coatings*) ou preenchendo os poros.

#### A.4 Relações Temporais

São demonstrados aqui os resultados da eliciação da associação entre as relações paragenéticas entre os constituintes de rocha e as relações de ordem entre o aparecimento desses constituintes. As tabelas 1 e 2 mostram todas as relações paragenéticas do domínio associadas às relações temporais correspondentes.

Tabela 1: Relações temporais derivadas da relação paragenética entre uma porosidade e outro constituinte (detritico ou diagenético).

Porosidade P - Constituinte A	P ? A	Observações
<b>P Dissolução de A</b>	P depois A	A é constituinte diagenético. Dissolução de constituinte detritico: nenhuma informação sobre seqüência diagenética. Mas se houver um constituinte diagenético dentro indica precipitação de constituinte diagenético. Dissolução depois da formação de A
<b>P Dentro de A</b>	P junto A P depois A	Porosidade pode ter vindo junto de A (formado junto) ou pode ter sido formado por dissolução de A
<b>P Contração de A</b>	P depois A	A é constituinte detritico ou diagenético Contração depois da deposição ou formação de A
<b>P Fraturamento de A</b>	P depois A	Fraturamento depois da deposição ou formação de A
<b>P Descompactação (da rocha)</b>	P depois A	Um tipo de fraturamento sobre a rocha inteira, formando poros entre os constituintes. Forma poros sobre a rocha toda.

Tabela 2: Relações temporais derivadas da relação paragenética entre dois constituintes (que não são porosidades).

Constituinte A - Constituinte B	A ? B	Observações
<b>A Cobrindo B</b>	A depois B	



<b>A Coberto por B</b>	B depois A	Inverso de <i>Cobrindo</i>
<b>A Substituindo grão de B</b>	A depois B	
<b>A Substituindo matriz de B</b>	A depois B	
<b>A Substituindo B</b>	A depois B	
<b>A Alternado com B</b>	A junto B	Dois constituintes precipitando-se alternadamente na rocha. Juntos na seqüência diagenética. Na lista de processos eles devem aparecer na mesma linha.
<b>A Engolfando B</b>	A depois B B depois A	B se formou e depois A engolfa B ou A se formou e B substitui A de forma que ele fica dentro de A no corte da lâmina delgada.
<b>A Engolfado por B</b>	B depois A A depois B	Inverso de <i>Engolfando</i>
<b>A Entrecrescido com B</b>	A junto B	Entrecrescidos, coprecipitacao, duas coisas precipitando juntas, formadas juntas.
<b>A Sobrecrescendo B</b>	A depois B	Crescendo em cima
<b>A Sobrecrescido por B</b>	B depois A	Inverso de <i>Sobrecrescendo</i>
<b>A Expandindo B</b>	A depois B	A precipitou dentro de B, expandindo ele, então A depois de B, expandindo B.
<b>A Compactado de B</b>	A depois B	Pseudomatriz formada da compactação de B
<b>A Dentro da porosidade primária intragranular</b>	A depois Deposição	A é posterior a deposição, não dá pra dizer nada mais, nem que ocorreu bem cedo.
<b>A Dentro da porosidade intragranular depois de B</b>	A depois B	B foi dissolvido para formar porosidade intergranular
<b>A Dentro da porosidade intragranular depois da matriz detrítica (de B)</b>	A depois B	matriz foi dissolvida ou contraída de forma a gerar poros.
<b>A Dentro da porosidade intragranular em B</b>	A depois B	Grão foi dissolvido e formou porosidade. A veio depois da formação e depois da dissolução de B. Ou A se formou em uma porosidade intergranular que já existia no grão, já veio com o grão. A é posterior à formação do poro.
<b>A Dentro da porosidade intracristalina em B</b>	A depois B	Cristal que foi dissolvido, formando-se um poro, no qual A se depositou. A é posterior à dissolução.
<b>A Dentro da porosidade móldica depois de B</b>	A depois B	constituente foi totalmente dissolvido e só ficou o buraco, e A veio depois da dissolução daquilo que foi dissolvido totalmente.
<b>A Dentro da porosidade de contração de B</b>	A depois B	Contração de argila detrítica ou diagenética, forma porosidade, A é posterior a contração de B.
<b>A Dentro da porosidade de fratura de grão em B</b>	A depois B	A é posterior ao fraturamento do grão B.
<b>A Dentro da porosidade de fratura de rocha em B</b>	A depois B	É uma rocha fraturada, e A é posterior ao fraturamento da rocha.

## A.5 Eliciação Triádica

Mostra a maneira como o especialista organizou os conceitos da eliciação triádica, a qual acabou gerando um resultado parecido com a classificação de fichas, pois o especialista acabou por organizar as fichas formando categorias de processos. Essa classificação é mostrada na Figura 2.

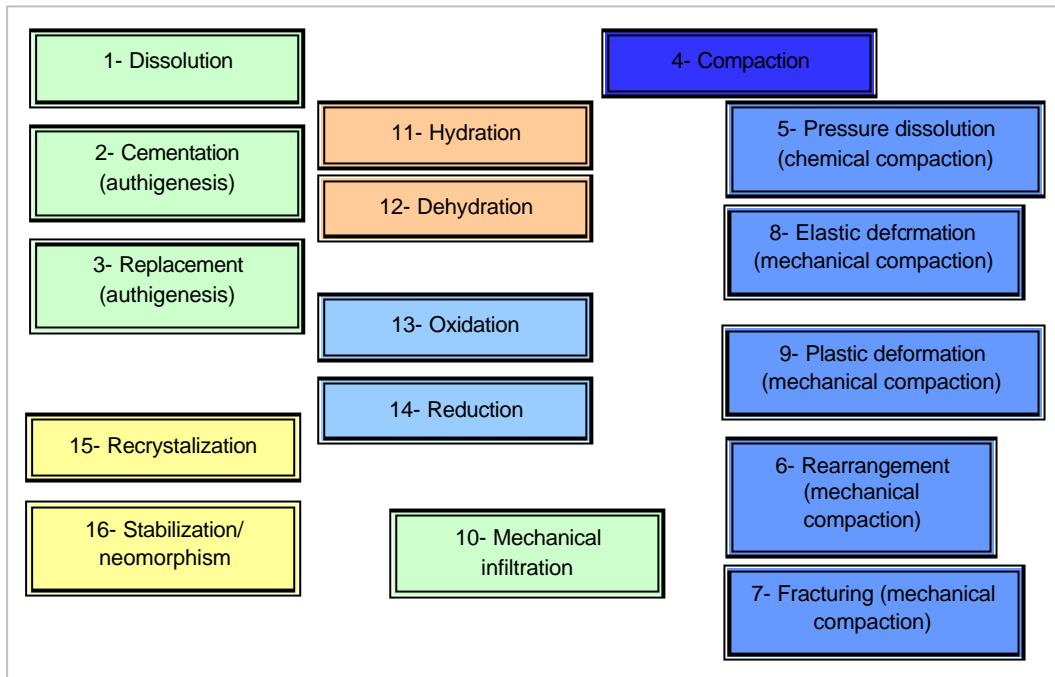


Figura 2: Processos classificados segundo categorias do especialista.

## A.6 Grade de Repertórios

A técnica de grade de repertórios (Tabela 3) foi aplicada para tentar eliciar relações entre os processos diagenéticos e outras feições do domínio que não fossem as relações paragenéticas, no caso, a *localização*. Os processos de *Hidratação*, *Desidratação*, *Oxidação*, *Redução* e *Estabilização/neomorfismo* não estão na tabela porque foram descartados pelo especialista por que não podem ser indicados pelo modificador.

Tabela 3: Resultados da aplicação da técnica de grades de repertório.

Eventos Diagenéticos		Dissolução	Cimentação	Substituição	Compactação Mecânica	Dissolução por pressão	Recristalização	Fraturamento
Modificadores Detriticos	<i>deformado</i>				X			
	<i>fraturado</i>				X			X
	<i>dissolvido</i>	X				X		
	<i>alterado</i>			X				
	<i>substituído</i>			X				
	<i>recristalizado</i>						X	
Modificadores Diagenéticos	<i>dissolvido</i>	X				X		
	<i>zonado</i>		X					
	<i>fraturado</i>				X			X
	<i>recristalizado</i>						X	

Feições (Modificador)		Eventos Diagenéticos						
		Dissolução	Cimentação	Substituição	Compactação	Dissolução por pressão	Recristalização	Fraturamento
Modificadores Macroporosidade	<i>reduzido por compactação</i>				X			
	<i>reduzido por cimentação</i>		X					
	<i>alargado por dissolução</i>	X				X		

## A.7 Propriedades de Relações

Segundo (HUTH e RYAN, 2000), uma relação entre dois elementos de um domínio  $W$  podem apresentar as seguintes propriedades:

- reflexiva: se, para cada  $x \in W$ , existe  $R(x,x)$  em  $W$ .
- irreflexiva: se, para cada  $x \in W$ , não existe  $R(x,x)$  em  $W$ .
- simétrica: se, para cada  $x, y \in W$ ,  $R(x,y)$  implica  $R(y,x)$ .
- assimétrica: se, para cada  $x, y \in W$ ,  $R(x,y)$  implica que não existe  $R(y,x)$ .
- serial: se, para cada  $x \in W$ , existe um  $y \in W$  tal que  $R(x,y)$ .
- transitiva: se, para cada  $x, y, z \in W$ , existe  $R(x,y)$  e  $R(y,z)$  implica  $R(x,z)$ .
- Euclidiana: se, para cada  $x, y, z \in W$  com  $R(x,y)$  e  $R(x,z)$ , existe  $R(y,z)$ .
- funcional: se, para cada  $x \in W$ , existe um único  $y \in W$  tal que  $R(x,y)$ .
- linear: se, para cada  $x, y, z \in W$ ,  $(R(x,y)$  e  $R(x,z))$  juntos implicam  $R(y,z)$ , ou  $y$  igual  $z$ , ou  $R(z,y)$ .
- total: se, para cada  $x, y \in W$ , existe  $R(x,y)$  ou  $R(y,x)$ .
- equivalência: se  $R$  é reflexiva, simétrica e transitiva.

A seguir é demonstrada a transcrição de algumas perguntas usadas para eliciar as propriedades das relações do domínio da petrografia.

**Engenheiro de Conhecimento:** *Uma porosidade pode estar dentro de mais de um constituinte ou somente um?*

**Especialista:** *A porosidade pode estar dentro mais de um constituinte. Não espacialmente, mas todos os poros de um tipo são uma porosidade. Por exemplo, pode existir porosidade intragranular dentro de vários constituintes. E a relação dentro de é utilizada para dizer que aquele poro está dentro de um constituinte, independente se o constituinte foi dissolvido (o que não se pode ter certeza) ou se o poro já estava lá desde o início.*

**Conclusão:** a relação *dentro de* não é funcional, pois pela sua definição, a porosidade deveria poder estar somente dentro de um constituinte.

**Engenheiro de Conhecimento:** Se a porosidade  $P$  estiver dentro de dois constituintes ( $A$  e  $B$ ), é possível dizer que existe alguma relação paragenética entre  $A$  e  $B$ ? (a mesma pergunta foi repetida, generalizando a relação dentro de para qualquer relação paragenética).

**Especialista:** Não é possível inferir relações visuais que não tenham sido descritas, pois os constituintes podem se organizar de maneiras não previsíveis. É preciso a descrição visual para dizer qual a relação visual está ocorrendo.

**Conclusão:** As relações paragenéticas não são Euclidianas, pois, pela sua definição, se existir uma relação paragenética entre  $P$  e  $A$  e entre  $P$  e  $B$ , deve-se poder inferir a relação paragenética entre  $A$  e  $B$ . Isso não ocorre.

**Engenheiro de Conhecimento:** Se um constituinte  $A$  cobre um constituinte  $B$ , e se esse constituinte  $B$  cobre outro constituinte  $C$ , isso necessariamente significa que o constituinte  $A$  também está cobrindo o constituinte  $C$ ? (a mesma pergunta foi repetida, generalizando a relação cobrindo para qualquer relação paragenética).

**Especialista:** Não necessariamente. Pode ocorrer transição, mas não é a regra. O exemplo de cobertura é típico: dependendo da configuração em que os minerais precipitaram  $A$  pode aparecer cobrindo  $C$  ou não. O dentro de até pode ser transitivo, mas é raro acontecer.

**Conclusão:** As relações visuais não são necessariamente transitivas, pois, pela sua definição, a transição deve ocorrer sempre.

**Engenheiro de Conhecimento:** Um constituinte pode ter sido precipitado junto com mais de um constituinte?

**Especialista:** Não, um constituinte precipita no máximo junto com mais um.

**Conclusão:** Não se aplica tentar verificar as propriedades Euclidiana e Linear para as relações *alternado com* e *entrecrescido com*, pois a primeira parte da definição (um constituinte  $C$  ter sido precipitado junto com  $A$  e junto com  $B$ ) não ocorre no domínio.

Na Tabela 4 é apresentado o resumo das propriedades lógicas entre relações específicas do domínio da petrografia, definidas com base nas observações coletadas nas sessões de aquisição de conhecimento.  $S$  significa que a propriedade é satisfeita;  $N$ , significa que não é satisfeita, --- significa que não se aplica.

Tabela 4: Propriedades de relações no domínio da petrografia.

Relação	Reflexiva	Simétrica	Serial	Transitiva	Euclidiana	Funcional	Linear	Total
<b>Relações Gerais</b>								
relação-paragenética( $A, B$ )	N	S	S	N	N	N	N	S
relação-temporal( $E1, E2$ )	S	S	S	S	N	N	N	S
produzido-por( $C, E$ )	N	N	S	---	---	S	S	N
Inferência: A relação- paragenética $B \rightarrow A$ relação-temporal $B$	S	N	S	---	N	N	N	N

Relação	Reflexiva	Simétrica	Serial	Transitiva	Euclidiana	Funcional	Linear	Total
<b>Relações entre Porosidades e Constituintes</b>								
Dissolução de(P, A)	N	N	N	---	N	N	N	N
Dentro(P, A)	N	N	N	---	N	N	N	N
Contração(P, A)	N	N	N	---	N	N	N	N
Fratramento(P, A)	N	N	N	---	N	N	N	N
Descompactação(P, A)	N	N	N	---	N	S	N	N
<b>Relações entre Constituintes</b>								
Cobrindo(A, B)	N	N	N	N	N	N	N	N
Coberto por(A, B)	N	N	N	N	N	N	N	N
Substituindo grão de(A, B)	N	N	N	N	N	N	N	N
Substituindo matriz de (A, B)	N	N	N	N	N	N	N	N
Substituindo (A, B)	N	N	N	N	N	N	N	N
Alternado com(A, B)	N	S	N	N	---	N	---	N
Engolfando(A, B)	N	N	N	N	N	N	N	N
Engolfado por(A, B)	N	N	N	N	N	N	N	N
Entrecrescido com(A, B)	N	S	N	N	---	N	---	N
Sobrecrescendo(A, B)	N	N	N	N	N	N	N	N
Sobrecrescido por(A, B)	N	N	N	N	N	N	N	N
Expandindo(A, B)	N	N	N	N	N	N	N	N
Compactado de(A, B)	N	N	N	N	N	N	N	N
Dentro de porosidade intragranular primária(A)	N	N	N	N	N	N	N	N
Dentro de porosidade intragranular depois de (A,B)	N	N	N	N	N	N	N	N
Dentro de porosidade intragranular depois de matriz detrítica de (A,B)	N	N	N	N	N	N	N	N
Dentro de porosidade intragranular em (A,B)	N	N	N	N	N	N	N	N
Dentro de porosidade intracristalina em(A,B)	N	N	N	N	N	N	N	N
Dentro de porosidade móldica depois de (A,B)	N	N	N	N	N	N	N	N
Dentro de porosidade de contração de(A,B)	N	N	N	N	N	N	N	N
Dentro de porosidade de fratura de grão em (A,B)	N	N	N	N	N	N	N	N
Dentro de porosidade de fratura de rocha em (A,B)	N	N	N	N	N	N	N	N
<b>Relações Temporais entre Eventos</b>								
antes(E1, E2)	N	N	N	S	N	N	N	N
depois(E1, E2)	N	N	N	S	N	N	N	N
durante(E1, E2)	S	S	N	S	S	N	S	N

Analisando a tabela de propriedades, é óbvio que quase nenhuma se aplica às relações, o que pode indicar um domínio em que o comportamento da relação entre os objetos não pode ser previsto com facilidade.

## A.8 Limitação de Informações (Focalizando Contextos)

A técnica de focalizar contextos/limitar informações foi aplicada solicitando que o especialista realizasse a interpretação de seqüência sem observar a amostra ao microscópio, usando como base somente as informações de descrição da amostra.

A Figura 3 mostra a descrição parcial da amostra WINT01-4615,5 onde se pode verificar a descrição das relações paragenéticas entre os constituintes.

Constituent Identification
Detrital quartz monocrystalline common plutonic - As monomineralic grain - Fractured
Detrital microcline - As monomineralic grain - Replaced
Detrital perthite - As monomineralic grain - Dissolved
Detrital plagioclase - As monomineralic grain
Biotite - As monomineralic grain - Deformed
Biotite - As monomineralic grain - Replaced
Garnet - As monomineralic grain
Garnet - As monomineralic grain - Dissolved
Sphene - As monomineralic grain
Sphene - As monomineralic grain - Dissolved
Sphene - As monomineralic grain - Replaced
Smectite - Microcrystalline - Intragranular replacive - - Replacing grain of <Detrital-Constituent> - Volcanic rock fragment
Calcite - Poikilotopic - Concretions/nodules - - Engulfing <Detrital-Constituent> - Detrital undifferentiated
Smectite - Coating - Intergranular continuous pore-lining - - Covering <Detrital-Constituent> - Detrital undifferentiated
Smectite - Coating - Intergranular continuous pore-lining - - Replaced by <Diagenetic-Constituent> - Chlorite
Diagenetic leuco - - Intergranular continuous pore-lining - - Covering <Detrital-Constituent> - Heavy minerals
Pyrite - Framboid - Intragranular discrete - - Within intergranular primary porosity -
Pyrite - Framboid - Intragranular pore-filling - - Replacing grain of <Detrital-Constituent> - Volcanic rock fragment
Pyrite - Framboid - Intragranular pore-filling - - Replacing grain of <Detrital-Constituent> - Heavy minerals
Pyrite - Framboid - Intragranular pore-filling - - Within intragranular porosity in <Detrital-Constituent> - Biotite
Chlorite - Rim - Intergranular discontinuous pore-lining - - Covering <Detrital-Constituent> - Detrital undifferentiated
Chlorite - Rim - Intergranular discontinuous pore-lining - - Covering <Diagenetic-Constituent> - Smectite
Chlorite - Rim - Intergranular discontinuous pore-lining - - Covering <Diagenetic-Constituent> - Calcite
Halite - Microcrystalline - Intergranular pore-filling - - Covering <Diagenetic-Constituent> - Chlorite
Illite - Microcrystalline - Intragranular replacive - - Replacing grain of <Detrital-Constituent> - Detrital K-feldspar
Illite - Microcrystalline - Intragranular replacive - - Replacing grain of <Detrital-Constituent> - Detrital plagioclase
Illite - Microcrystalline - Intragranular replacive - - Replacing grain of <Detrital-Constituent> - Plutonic rock fragment
Macroporosity Intergranular - Interstitial - Reduced by compaction - Primary -
Macroporosity Intergranular - Interstitial - Reduced by cementation - Primary - Chlorite
Macroporosity Intragranular - Framework - Enlarged by dissolution - Dissolution of <Detrital-Constituent> - Volcanic rock fragment
Macroporosity Intragranular - Framework - Enlarged by dissolution - Dissolution of <Detrital-Constituent> - Detrital feldspar

Figura 3: Descrição parcial dos constituintes da amostra WINT01-4615,5.

Alguns dos cartões gerados para a amostra WINT01-4615,5 são mostrados na Figura 4. Fichas amarelas são constituintes diagenéticos, fichas verdes são constituintes detriticos.

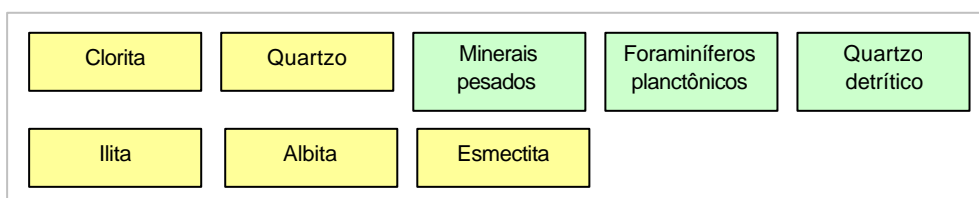


Figura 4: Fichas de alguns constituintes usados para simular a interpretação de seqüência diagenética.

A Figura 5 mostra como as fichas foram organizadas para representar a ordem de aparecimento relativo entre esmectita, clorita e quartzo (resultado intermediário).

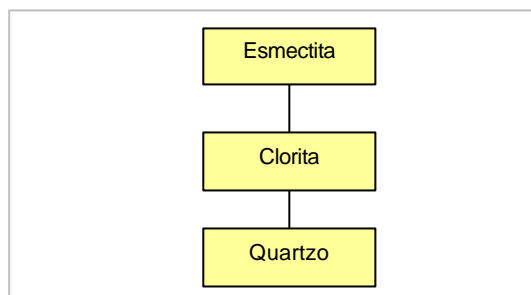


Figura 5: Resultado intermediário da ordenação de eventos diagenéticos.

A Figura 6 mostra o resultado da ordenação de aparecimento dos constituintes. A ordenação resultante não foi completa.

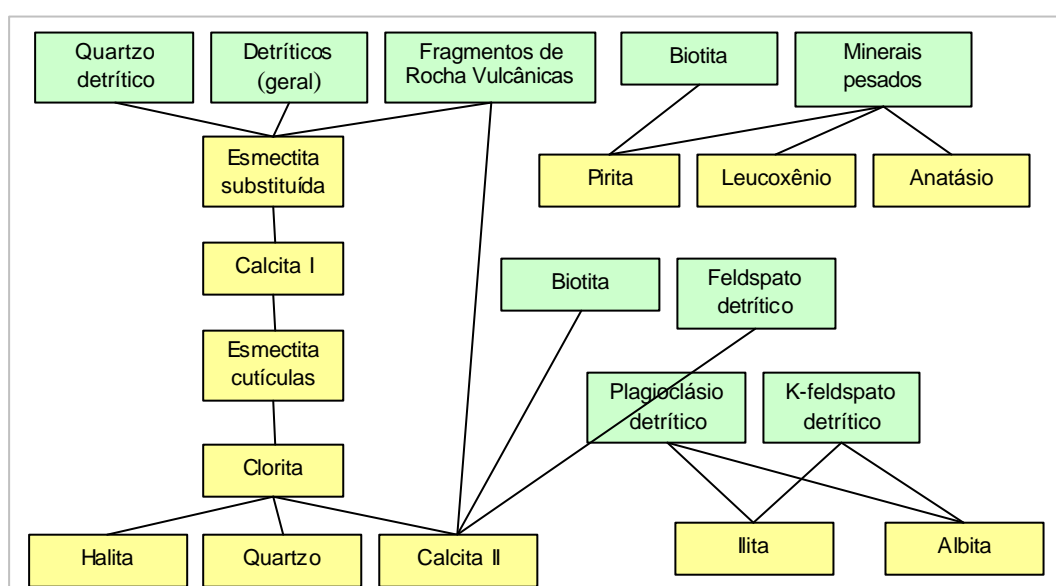


Figura 6: Ordenação de aparecimento dos constituintes na amostra WINT01-4615,5.

A seqüência de aparecimento dos constituintes da amostra S208 é mostrada na Figura 7. A ordenação resultante foi completa.

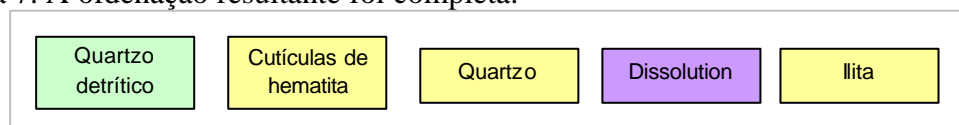


Figura 7: Ordenação de aparecimento dos constituintes na amostra S208.

## APÊNDICE B MODELO DE CONHECIMENTO PARA INTERPRETAÇÃO DE SEQÜÊNCIA DE EVENTOS DIAGENÉTICOS

Nesse apêndice é apresentado o modelo de conhecimento da suíte de modelos CommonKADS, desenvolvido para representar o conhecimento necessário para a tarefa de interpretação de seqüência de eventos diagenéticos. Foi utilizada a linguagem CML para formalizar as primitivas do modelo. Na Figura 1 é apresentada a representação gráfica em UML do modelo de inferência, que faz parte do conhecimento de inferência. Na Figura 2 é apresentado o modelo de classes Java desenvolvido para validar o modelo de conhecimento proposto para a tarefa de ordenação de eventos.

MODELO-CONHECIMENTO OntologiaEventos;

/\* Modelo de conhecimento para a tarefa de interpretação de seqüência de eventos no domínio da petrografia

AUTOR: Laura Mastella, UFRGS

MODIFICADO EM: Dez 2004

\*/

CONHECIMENTO-DOMÍNIO domínio-petrográfico;

ESQUEMA-DOMÍNIO esquema-eventos;

EVENTO evento-diagenético;

DESCRIÇÃO: "Um evento diagenético que ocorreu sobre uma rocha e modificou as relações paragenéticas entre os constituintes de rocha.";

ATRIBUTOS:

nome-evento: nome-evento-tipo-valor;

FIM EVENTO evento-diagenético;

TIPO-VALOR nome-evento-tipo-valor;

TIPO: NOMINAL;

LISTA-VALORES: {dissolução, cimentação-autigênese, substituição-autigênese, compactação, dissolução\_por\_pressão-compactação\_química, rearranjo-compactação\_mecânica, fraturamento-compactação\_mecânica, deformação\_elástica-compactação\_mecânica, deformação\_plástica-compactação\_mecânica, infiltração\_mecânica, hidratação, desidratação, oxidação, redução, recristalização, estabilização-neomorfismo};

FIM TIPO-VALOR nome-evento-tipo-valor;

RELACAO-TEMPORAL depois;

DESCRIÇÃO: "Um evento ocorreu depois do outro.";

ORDEM: depois;

INVERSO: antes;

ARGUMENTO-1: evento-diagenético;

CARDINALIDADE: 1;

ARGUMENTO-2: evento-diagenético;

CARDINALIDADE: 1;

FIM RELACAO-TEMPORAL depois;

RELACAO-TEMPORAL antes;

DESCRIÇÃO: "Um evento ocorreu antes do outro.";

ORDEM: antes;

INVERSO: depois;

ARGUMENTO-1: evento-diagenético;

CARDINALIDADE: 1;

ARGUMENTO-2: evento-diagenético;



CARDINALIDADE: 1;  
 FIM RELACAO-TEMPORAL antes;

RELACAO-TEMPORAL durante;  
 DESCRIÇÃO: "Os eventos ocorreram ao mesmo tempo.";  
 ORDEM: durante;  
 ARGUMENTO-1: evento-diagenético;  
 CARDINALIDADE: 1;  
 ARGUMENTO-2: evento-diagenético;  
 CARDINALIDADE: 1;  
 FIM RELACAO-TEMPORAL durante;

CONCEITO constituinte-de-rocha;  
 DESCRIÇÃO: "Os componentes da rocha. Alguns são formados pelos eventos, outros são somente modificados por eles.";  
 SUPER-TIPO-DE:  
 constituinte-mineral, constituinte-macroporosidade;  
 ATRIBUTOS:  
 nome-constituinte: STRING;  
 localização: STRING;  
 conjunto-constituinte: STRING;  
 FIM CONCEITO constituinte-de-rocha;

CONCEITO constituinte-mineral;  
 DESCRIÇÃO: "Um constituinte de rocha que não é uma porosidade.";  
 SUB-TIPO-DE:  
 constituinte-de-rocha;  
 FIM CONCEITO constituinte-mineral;

CONCEITO constituinte-detrítico;  
 DESCRIÇÃO: "Um constituinte originado fora do local atual da rocha e que pode ter sido modificado pela ocorrência de eventos diagenéticos.";  
 SUB-TIPO-DE:  
 constituinte-mineral;  
 ATRIBUTOS:  
 modificador: modificador-detrítico-tipo-valor;  
 FIM CONCEITO constituinte-detrítico;

TIPO-VALOR modificador-detrítico-tipo-valor;  
 TIPO: NOMINAL;  
 LISTA-VALORES: {deformado, fraturado, dissolvido, alterado, substituído, recristalizado};  
 FIM TIPO-VALOR modificador-detrítico-tipo-valor;

CONCEITO constituinte-diagenético;  
 DESCRIÇÃO: "Um constituinte formado pela ocorrência de eventos diagenéticos.";  
 SUB-TIPO-DE:  
 constituinte-mineral;  
 ATRIBUTOS:  
 hábito: STRING;  
 modificador: modificador-diagenético-tipo-valor;  
 FIM CONCEITO constituinte-diagenético;

TIPO-VALOR modificador-diagenético-tipo-valor;  
 TIPO: NOMINAL;  
 LISTA-VALORES: {dissolvido, zonado, fraturado, recristalizado};  
 FIM TIPO-VALOR modificador-diagenético-tipo-valor;

CONCEITO constituinte-macroporosidade;  
 DESCRIÇÃO: "Uma porosidade que pode ter sido formada no início da formação da rocha ou por um evento diagenético.";  
 SUB-TIPO-DE:  
 constituinte-de-rocha;  
 ATRIBUTOS:  
 modificador: modificador-macroporosidade-tipo-valor;  
 FIM CONCEITO constituinte-macroporosidade;

TIPO-VALOR modificador-macroporosidade-tipo-valor;  
 TIPO: NOMINAL;  
 LISTA-VALORES: {reduzido\_por\_compactação, reduzido\_por\_cimentação, alargado\_por\_dissolução};  
 FIM TIPO-VALOR modificador-macroporosidade-tipo-valor;

RELACAO-BINARIA produzido-por;  
 DESCRIÇÃO: "Relação entre um constituinte de rocha e o evento diagenético que o produziu";  
 ARGUMENTO-1: constituinte-de-rocha;  
 CARDINALIDADE: 1;  
 ARGUMENTO-2: evento-diagenético;

CARDINALIDADE: 1;  
irreflexiva, assimétrica;  
FIM RELACAO-BINARIA produzido-por;

RELACAO-BINARIA relação-paragenética;  
DESCRIÇÃO: "Relação de origem entre os constituintes de rocha."  
ARGUMENTO-1: constituinte-de-rocha;  
CARDINALIDADE: 1;  
ARGUMENTO-2: constituinte-de-rocha;  
CARDINALIDADE: 1;  
irreflexiva; simétrica;  
FIM RELACAO-BINARIA relação-paragenética;

RELACAO-BINARIA primária;  
DESCRIÇÃO:  
"Porosidade é primária."  
SUB-TIPO-DE: relação-paragenética;  
ARGUMENTO-1: constituinte-macroporosidade;  
CARDINALIDADE: 1;  
ARGUMENTO-2: constituinte-macroporosidade;  
CARDINALIDADE: 1;  
FIM RELACAO-BINARIA primária;

RELACAO-BINARIA dissolução-de;  
DESCRIÇÃO:  
"Porosidade resultante da dissolução de algum constituinte diagenético. A é constituinte diagenético. Dissolução depois da formação de A."  
SUB-TIPO-DE: relação-paragenética;  
ARGUMENTO-1: constituinte-macroporosidade;  
CARDINALIDADE: 1;  
ARGUMENTO-2: constituinte-mineral;  
CARDINALIDADE: 1;  
FIM RELACAO-BINARIA dissolução-de;

RELACAO-BINARIA dentro;  
DESCRIÇÃO:  
"Porosidade está dentro de um constituinte. Poro pode ter vindo junto de A (formado junto) ou pode ter sido formado por dissolução de A."  
SUB-TIPO-DE: relação-paragenética;  
ARGUMENTO-1: constituinte-macroporosidade;  
CARDINALIDADE: 1;  
ARGUMENTO-2: constituinte-mineral;  
CARDINALIDADE: 1;  
FIM RELACAO-BINARIA dentro;

RELACAO-BINARIA contração-de;  
DESCRIÇÃO:  
"Porosidade formada por contração de algum mineral. A é constituinte detrítico ou diagenético. Contração depois da deposição ou formação de A."  
SUB-TIPO-DE: relação-paragenética;  
ARGUMENTO-1: constituinte-macroporosidade;  
CARDINALIDADE: 1;  
ARGUMENTO-2: constituinte-mineral;  
CARDINALIDADE: 1;  
FIM RELACAO-BINARIA contração-de;

RELACAO-BINARIA fraturamento-de;  
DESCRIÇÃO:  
"Porosidade formada pelo fraturamento de algum mineral. Fraturamento depois da deposição ou formação";  
SUB-TIPO-DE: relação-paragenética;  
ARGUMENTO-1: constituinte-macroporosidade;  
CARDINALIDADE: 1;  
ARGUMENTO-2: constituinte-mineral;  
CARDINALIDADE: 1;  
FIM RELACAO-BINARIA fraturamento-de;

RELACAO-BINARIA descompactação-de;  
DESCRIÇÃO: "Porosidade formada pela descompactação da rocha. Contração depois da deposição ou formação";  
SUB-TIPO-DE: relação-paragenética;  
ARGUMENTO-1: constituinte-macroporosidade;  
CARDINALIDADE: 1;  
ARGUMENTO-2: constituinte-mineral;  
CARDINALIDADE: 1;  
FIM RELACAO-BINARIA descompactação-de;

/\* relações paragenéticas de constituintes diagenéticos\*/

- RELACAO-BINARIA cobrindo;  
 DESCRIÇÃO: "Um constituinte cobre o outro.";  
 INVERSO: é-coberto-por;  
 SUB-TIPO-DE: relação-paragenética;  
 ARGUMENTO-1: constituinte-diagenético;  
 CARDINALIDADE: 1;  
 ARGUMENTO-2: constituinte-mineral;  
 CARDINALIDADE: 1;  
 FIM RELACAO-BINARIA cobrindo;
- RELACAO-BINARIA é-coberto-por;  
 DESCRIÇÃO: "Um constituinte é coberto pelo outro.";  
 INVERSO: cobrindo;  
 SUB-TIPO-DE: relação-paragenética;  
 ARGUMENTO-1: constituinte-diagenético;  
 CARDINALIDADE: 1;  
 ARGUMENTO-2: constituinte-diagenético;  
 CARDINALIDADE: 1;  
 FIM RELACAO-BINARIA é-coberto-por;
- RELACAO-BINARIA substituindo -grão-de;  
 DESCRIÇÃO: "um constituinte substitui o grão de outro.";  
 SUB-TIPO-DE: relação-paragenética;  
 ARGUMENTO-1: constituinte-diagenético;  
 CARDINALIDADE: 1;  
 ARGUMENTO-2: constituinte-detrítico;  
 CARDINALIDADE: 1;  
 FIM RELACAO-BINARIA substituindo-grão-de;
- RELACAO-BINARIA substituindo -matriz-de;  
 DESCRIÇÃO: "Um constituinte substitui a matriz formada pelo outro.";  
 SUB-TIPO-DE: relação-paragenética;  
 ARGUMENTO-1: constituinte-diagenético;  
 CARDINALIDADE: 1;  
 ARGUMENTO-2: constituinte-detrítico;  
 CARDINALIDADE: 1;  
 FIM RELACAO-BINARIA substituindo-matriz-de;
- RELACAO-BINARIA substituindo;  
 DESCRIÇÃO: "Um constituinte substitui o outro.";  
 INVERSO: substituído-por;  
 SUB-TIPO-DE: relação-paragenética;  
 ARGUMENTO-1: constituinte-diagenético;  
 CARDINALIDADE: 1;  
 ARGUMENTO-2: constituinte-diagenético;  
 CARDINALIDADE: 1;  
 FIM RELACAO-BINARIA substituindo;
- RELACAO-BINARIA substituído-por;  
 DESCRIÇÃO: "Um constituinte é substituído pelo outro.";  
 INVERSO: substituindo;  
 SUB-TIPO-DE: relação-paragenética;  
 ARGUMENTO-1: constituinte-diagenético;  
 CARDINALIDADE: 1;  
 ARGUMENTO-2: constituinte-diagenético;  
 CARDINALIDADE: 1;  
 FIM RELACAO-BINARIA substituído-por;
- RELACAO-BINARIA alternado-com;  
 DESCRIÇÃO: "Um constituinte precipitou alternado com o outro. Dois constituintes se precipitando alternadamente sobre a rocha. Juntos no tempo, na seqüência diagenética. Na lista de processos eles devem aparecer na mesma linha.";  
 SUB-TIPO-DE: relação-paragenética;  
 ARGUMENTO-1: constituinte-diagenético;  
 CARDINALIDADE: 1;  
 ARGUMENTO-2: constituinte-diagenético;  
 CARDINALIDADE: 1;  
 FIM RELACAO-BINARIA alternado-com;
- RELACAO-BINARIA engolfando;  
 DESCRIÇÃO: "Um constituinte está engolfando o outro. B se formou e depois A engolfa B ou A se formou e B substitui A de forma que ele fica dentro de A no corte da lâmina delgada.";  
 INVERSO: engolfado-por;  
 SUB-TIPO-DE: relação-paragenética;  
 ARGUMENTO-1: constituinte-diagenético;  
 CARDINALIDADE: 1;

ARGUMENTO-2: constituinte-mineral;  
CARDINALIDADE: 1;  
FIM RELACAO-BINARIA engolfando;

RELACAO-BINARIA engolfado-por;  
DESCRIÇÃO: "Um constituinte é engolfado pelo outro.";  
INVERSO: engolfando;  
SUB-TIPO-DE: relação-paragenética;  
ARGUMENTO-1: constituinte-diagenético;  
CARDINALIDADE: 1;  
ARGUMENTO-2: constituinte-diagenético;  
CARDINALIDADE: 1;  
FIM RELACAO-BINARIA engolfado -por;

RELACAO-BINARIA entrecrescido-com;  
DESCRIÇÃO: "Um constituinte foi formado junto com o outro. Entrecrescidos, co-precipitação, duas coisas precipitando juntas, formadas juntas.";  
SUB-TIPO-DE: relação-paragenética;  
ARGUMENTO-1: constituinte-diagenético;  
CARDINALIDADE: 1;  
ARGUMENTO-2: constituinte-diagenético;  
CARDINALIDADE: 1;  
FIM RELACAO-BINARIA entrecrescido-com;

RELACAO-BINARIA sobrecrecendo;  
DESCRIÇÃO: "Um constituinte cresceu sobre o outro. Crescendo em cima.";  
INVERSO: sobrecrecendo-por;  
SUB-TIPO-DE: relação-paragenética;  
ARGUMENTO-1: constituinte-diagenético;  
CARDINALIDADE: 1;  
ARGUMENTO-2: constituinte-mineral;  
CARDINALIDADE: 1;  
FIM RELACAO-BINARIA sobrecrecendo;

RELACAO-BINARIA sobrecrecido-por;  
DESCRIÇÃO: "Um constituinte foi sobrecrecido por outro.";  
INVERSO: sobrecrecendo;  
SUB-TIPO-DE: relação-paragenética;  
ARGUMENTO-1: constituinte-diagenético;  
CARDINALIDADE: 1;  
ARGUMENTO-2: constituinte-diagenético;  
CARDINALIDADE: 1;  
FIM RELACAO-BINARIA sobrecrecido -por;

RELACAO-BINARIA expandindo;  
DESCRIÇÃO: "Um constituinte precipitou expandindo o outro. A precipitou dentro de B, expandindo ele, então A depois de B, expandindo B.";  
SUB-TIPO-DE: relação-paragenética;  
ARGUMENTO-1: constituinte-diagenético;  
CARDINALIDADE: 1;  
ARGUMENTO-2: constituinte-detrítico;  
CARDINALIDADE: 1;  
FIM RELACAO-BINARIA expandindo;

RELACAO-BINARIA compactado-de;  
DESCRIÇÃO: "Um constituinte (pseudomatriz) foi formado pela compactação do outro. Pseudomatriz formada da compactação de B.";  
SUB-TIPO-DE: relação-paragenética;  
ARGUMENTO-1: constituinte-diagenético;  
CARDINALIDADE: 1;  
ARGUMENTO-2: constituinte-detrítico;  
CARDINALIDADE: 1;  
FIM RELACAO-BINARIA compactado -de;

RELACAO-BINARIA dentro-de-porosidade-intergranular-primária;  
DESCRIÇÃO: "Um constituinte está dentro da porosidade primária. A é posterior à deposição, não dá pra dizer nada mais, nem que ocorreu bem cedo.";  
SUB-TIPO-DE: relação-paragenética;  
ARGUMENTO-1: constituinte-diagenético;  
CARDINALIDADE: 1;  
ARGUMENTO-2: constituinte-macroporosidade;  
CARDINALIDADE: 1;  
FIM RELACAO-BINARIA dentro-de-porosidade-intergranular-primária;

RELACAO-BINARIA dentro-de-porosidade-intergranular-depois-de;

DESCRIPÇÃO: "Um constituinte está dentro de uma porosidade formada de outro constituinte. B foi dissolvido para formar porosidade intergranular.";

SUB-TIPO-DE: relação-paragenética;  
 ARGUMENTO-1: constituinte-diagenético;  
 CARDINALIDADE: 1;  
 ARGUMENTO-2: constituinte-diagenético;  
 CARDINALIDADE: 1;

FIM RELACAO-BINARIA dentro-de-porosidade-intergranular-depois-de;

RELACAO-BINARIA dentro-de-porosidade-intergranular-depois-de-matriz-detrítica;

DESCRIPÇÃO: "Um constituinte está dentro de porosidade formada da matriz detrítica de outro constituinte. Matriz foi dissolvida ou contraída de forma que tem poros.";

SUB-TIPO-DE: relação-paragenética;  
 ARGUMENTO-1: constituinte-diagenético;  
 CARDINALIDADE: 1;  
 ARGUMENTO-2: constituinte-detrítico;  
 CARDINALIDADE: 1;

FIM RELACAO-BINARIA dentro-de-porosidade-intergranular-depois-de-matriz-detrítica;

RELACAO-BINARIA dentro-de-porosidade-intergranular-em;

DESCRIPÇÃO: "Um constituinte está dentro da porosidade formada de outro constituinte. Grão foi dissolvido e formou porosidade. A veio depois da formação e depois da dissolução de B. Ou A se formou em uma porosidade intergranular que já existia no grão, já veio com o grão. A é posterior à formação do poro.";

SUB-TIPO-DE: relação-paragenética;  
 ARGUMENTO-1: constituinte-diagenético;  
 CARDINALIDADE: 1;  
 ARGUMENTO-2: constituinte-detrítico;  
 CARDINALIDADE: 1;

FIM RELACAO-BINARIA dentro-de-porosidade-intergranular-em;

RELACAO-BINARIA dentro-de-porosidade-intracristalina-em;

DESCRIPÇÃO: "Um constituinte está dentro da porosidade intracristalina formada de outro constituinte. Tem um cristal que foi dissolvido e formou um poro, no qual A se depositou. A é posterior à dissolução.";

SUB-TIPO-DE: relação-paragenética;  
 ARGUMENTO-1: constituinte-diagenético;  
 CARDINALIDADE: 1;  
 ARGUMENTO-2: constituinte-diagenético;  
 CARDINALIDADE: 1;

FIM RELACAO-BINARIA dentro-de-porosidade-intracristalina-em;

RELACAO-BINARIA dentro-de-porosidade-móldica-depois-de;

DESCRIPÇÃO: "Um constituinte está dentro da porosidade móldica formada de outro constituinte. O constituinte foi totalmente dissolvido e só ficou o buraco, e A veio depois da dissolução daquilo que foi dissolvido totalmente.";

SUB-TIPO-DE: relação-paragenética;  
 ARGUMENTO-1: constituinte-diagenético;  
 CARDINALIDADE: 1;  
 ARGUMENTO-2: constituinte-mineral;  
 CARDINALIDADE: 1;

FIM RELACAO-BINARIA dentro-de-porosidade-móldica-depois-de;

RELACAO-BINARIA dentro-de-porosidade-de-contração-de;

DESCRIPÇÃO: "Um constituinte está dentro da porosidade formada da contração de outro constituinte. Contração de argila detrítica ou diagenética, forma porosidade, A é posterior a contração de B.";

SUB-TIPO-DE: relação-paragenética;  
 ARGUMENTO-1: constituinte-diagenético;  
 CARDINALIDADE: 1;  
 ARGUMENTO-2: constituinte-mineral;  
 CARDINALIDADE: 1;

FIM RELACAO-BINARIA dentro-de-porosidade-de-contração-de;

RELACAO-BINARIA dentro-de-porosidade-de-fratura-de-grão-em;

DESCRIPÇÃO: "Um constituinte está dentro da porosidade formada da fratura do grão de outro constituinte. A é posterior ao fraturamento do grão B.";

SUB-TIPO-DE: relação-paragenética;  
 ARGUMENTO-1: constituinte-diagenético;  
 CARDINALIDADE: 1;  
 ARGUMENTO-2: constituinte-detrítico;  
 CARDINALIDADE: 1;

FIM RELACAO-BINARIA dentro-de-porosidade-de-fratura-de-grão-em;

RELACAO-BINARIA dentro-de-porosidade-de-fratura-de-rocha-em;

DESCRIPÇÃO: "Um constituinte está dentro da porosidade formada da fratura da rocha. É uma rocha fraturada, e A é posterior ao fraturamento da rocha.";

SUB-TIPO-DE: relação-paragenética;  
 ARGUMENTO-1: constituinte-diagenético;  
 CARDINALIDADE: 1;

ARGUMENTO-2: constituinte-detrítico;  
 CARDINALIDADE: 1;  
 FIM RELACAO-BINARIA dentro-de-porosidade-de-fratura-de-rocha-em;

TIPO-REGRA indicação-evento;  
 DESCRIÇÃO: "Esta regra representa a associação entre constituintes os eventos que os geraram.";  
 ANTECEDENTE:  
 constituinte-de-rocha;  
 CARDINALIDADE: 1;  
 CONSEQUENTE:  
 evento-diagenético;  
 CARDINALIDADE: 1;  
 SÍMBOLO-CONEXÃO:  
 INDICA;  
 FIM TIPO-REGRA in dicação-evento;

TIPO-REGRA implicação-temporal;  
 DESCRIÇÃO: "Representa as regras que associam a relação paragenética à relação de ordem *antes*";  
 ANTECEDENTE:  
 relação-paragenética;  
 CARDINALIDADE: 1;  
 CONSEQUENTE:  
 antes;  
 CARDINALIDADE: 1;  
 SÍMBOLO-CONEXÃO:  
 IMPLICA;  
 FIM TIPO-REGRA implicação-temporal;

TIPO-REGRA implicação-temporal;  
 DESCRIÇÃO: "Representa as regras que associam a relação paragenética à relação de ordem *depois*";  
 ANTECEDENTE:  
 relação-paragenética;  
 CARDINALIDADE: 1;  
 CONSEQUENTE:  
 depois;  
 CARDINALIDADE: 1;  
 SÍMBOLO-CONEXÃO:  
 IMPLICA;  
 FIM TIPO-REGRA implicação-temporal;

TIPO-REGRA implicação-temporal;  
 DESCRIÇÃO: "Representa as regras que associam a relação paragenética à relação de ordem *durante*";  
 ANTECEDENTE:  
 relação-paragenética;  
 CARDINALIDADE: 1;  
 CONSEQUENTE:  
 durante;  
 CARDINALIDADE: 1;  
 SÍMBOLO-CONEXÃO:  
 IMPLICA;  
 FIM TIPO-REGRA implicação-temporal;

FIM ESQUEMA-DOMÍNIO esquema-eventos;

BASE-CONHECIMENTO conhecimento-ordenação-eventos;

USO:

indicação-evento DE esquema-eventos,  
 implicação-temporal DE esquema-eventos;

VARIABLES:

evento-diag1, evento-diag2: evento-diagenético;

const-diag.const-diag1,const-diag12 :constituente-diagenético;

const-macrop:constituente-macroporosidade;

const-min:constituente-mineral;

const-det:constituente-detrítico;

EXPRESSÕES:

/\* regras que indicam eventos sobre modificadores detríticos \*/

constituente-detrítico.modificador = substituído

INDICA evento-diagenético.nome-evento = substituição;

constituente-detrítico.modificador = alterado

INDICA evento-diagenético.nome-evento = substituição;

constituente-detrítico.modificador = fraturado

INDICA evento-diagenético.nome-evento = compactação;

```

constituente-detrítico.modificador = deformado
INDICA evento-dia-genético.nome-evento = compactação;

constituente-detrítico.modificador = recristalizado
INDICA evento-dia-genético.nome-evento = recristalização;

constituente-detrítico.modificador = dissolvido
INDICA evento-dia-genético.nome-evento = dissolução;

/* regras que indicam eventos sobre modificadores dia-genéticos */
constituente-dia-genético.modificador = dissolved
INDICA evento-dia-genético.nome-evento = dissolução;

constituente-dia-genético.modificador = zonado
INDICA evento-dia-genético.nome-evento = cimentação;

constituente-dia-genético.modificador = fractured
INDICA evento-dia-genético.nome-evento = compactação;

constituente-dia-genético.modificador = recrystallized
INDICA evento-dia-genético.nome-evento = recristalização;

/* regras que indicam eventos sobre modificadores de macroporosidade */
constituente-macroporosidade.modificador = reduzido_por_compactação
INDICA evento-dia-genético.nome-evento = compactação;

constituente-macroporosidade.modificador = alargado_por_dissolução
INDICA evento-dia-genético.nome-evento = compactação;

/* regras que indicam eventos sobre relações paragenéticas */
substituindo_grao_de(constituente-dia-genético, constituente-detrítico)
INDICA evento-dia-genético.nome-evento = substituição;

engulfing(constituente-dia-genético, constituente-detrítico)
INDICA evento-dia-genético.nome-evento = cimentação;

covering(constituente-dia-genético, constituente-dia-genético)
INDICA evento-dia-genético.nome-evento = cimentação;

/* regras que indicam a relação temporal pela relação-paragenética */
dissolução-de(const-macrop, const-min)
E produzido-por(const-macrop, evento-diag1)
E produzido-por(const-min, evento-diag2)
IMPLICA depois(evento-diag1, evento-diag2);

dentro(const-macrop, const-min)
E produzido-por(const-macrop, evento-diag1)
E produzido-por(const-min, evento-diag2)
IMPLICA depois(evento-diag1, evento-diag2);

contração-de(const-macrop, const-min)
E produzido-por(const-macrop, evento-diag1)
E produzido-por(const-min, evento-diag2)
IMPLICA depois(evento-diag1, evento-diag2);

fraturamento-de(const-macrop, const-min)
E produzido-por(const-macrop, evento-diag1)
E produzido-por(const-min, evento-diag2)
IMPLICA depois(evento-diag1, evento-diag2);

descompactação-de(const-macrop, const-min)
E produzido-por(const-macrop, evento-diag1)
E produzido-por(const-min, evento-diag2)
IMPLICA depois(evento-diag1, evento-diag2);

dentro(const-macrop, const-min)
E produzido-por(const-macrop, evento-diag1)
E produzido-por(const-min, evento-diag2)
IMPLICA durante(evento-diag1, evento-diag2);

cobrindo(const-diag, const-min)
E produzido-por(const-diag, evento-diag1)

```

E produzido-por(const -min, evento-diag2)  
 IMPLICA depois(evento-diag1, evento-diag2);

engolfando(const-diag, const-min)  
 E produzido-por(const-diag, evento-diag1)  
 E produzido-por(const-min, evento-diag2)  
 IMPLICA depois(evento-diag1, evento-diag2);

sobrecrescendo(const-diag, const-min)  
 E produzido-por(const-diag, evento-diag1)  
 E produzido-por(const-min, evento-diag2)  
 IMPLICA depois(evento-diag1, evento-diag2);

dentro-de-porosidade-móldica-depois-de(const-diag, const-min)  
 E produzido-por(const-diag, evento-diag1)  
 E produzido-por(const-min, evento-diag2)  
 IMPLICA depois(evento-diag1, evento-diag2);

dentro-de-porosidade-de-contração-de(const-diag, const-min)  
 E produzido-por(const-diag, evento-diag1)  
 E produzido-por(const-min, evento-diag2)  
 IMPLICA depois(evento-diag1, evento-diag2);

é-coberto-por(const-diag1, const-diag2)  
 E produzido-por(const-diag1, evento-diag1)  
 E produzido-por(const-diag2, evento-diag2)  
 IMPLICA antes(evento-diag1, evento-diag2);

substituindo-grão-de(const-diag, const-det)  
 E produzido-por(const-diag, evento-diag1)  
 E produzido-por(const-det, evento-diag2)  
 IMPLICA depois(evento-diag1, evento-diag2);

substituindo-matriz-de(const-diag, const-det)  
 E produzido-por(const-diag, evento-diag1)  
 E produzido-por(const-det, evento-diag2)  
 IMPLICA depois(evento-diag1, evento-diag2);

expandindo(const-diag, const-det)  
 E produzido-por(const-diag, evento-diag1)  
 E produzido-por(const-det, evento-diag2)  
 IMPLICA depois(evento-diag1, evento-diag2);

compactado-de(const-diag, const-det)  
 E produzido-por(const-diag, evento-diag1)  
 E produzido-por(const-det, evento-diag2)  
 IMPLICA depois(evento-diag1, evento-diag2);

dentro-de-porosidade-intergranular-depois-de-matriz-detrítica(const-diag, const-det)  
 E produzido-por(const-diag, evento-diag1)  
 E produzido-por(const-det, evento-diag2)  
 IMPLICA depois(evento-diag1, evento-diag2);

dentro-de-porosidade-intergranular-em(const-diag, const-det)  
 E produzido-por(const-diag, evento-diag1)  
 E produzido-por(const-det, evento-diag2)  
 IMPLICA depois(evento-diag1, evento-diag2);

dentro-de-porosidade-de-fratura-de-grão-em(const-diag, const-det)  
 E produzido-por(const-diag, evento-diag1)  
 E produzido-por(const-det, evento-diag2)  
 IMPLICA depois(evento-diag1, evento-diag2);

dentro-de-porosidade-de-fratura-de-rocha-em(const-diag, const-det)  
 E produzido-por(const-diag, evento-diag1)  
 E produzido-por(const-det, evento-diag2)  
 IMPLICA depois(evento-diag1, evento-diag2);

substituindo(const-diag1, const-diag2)  
 E produzido-por(const-diag1, evento-diag1)  
 E produzido-por(const-diag2, evento-diag2)  
 IMPLICA depois(evento-diag1, evento-diag2);

dentro-de-porosidade-intergranular-depois-de(const-diag1, const-diag2)  
 E produzido-por(const-diag1, evento-diag1)  
 E produzido-por(const-diag2, evento-diag2)



IMPLICA depois(evento-diag1, evento-diag2);

dentro-de-porosidade-intracristalina-em(const-diag1, const-diag2)

E produzido-por(const-diag1, evento-diag1)

E produzido-por(const-diag2, evento-diag2)

IMPLICA depois(evento-diag1, evento-diag2);

substituído-por(const-diag1, const-diag2)

E produzido-por(const-diag1, evento-diag1)

E produzido-por(const-diag2, evento-diag2)

IMPLICA antes(evento-diag1, evento-diag2);

engolfado -por(const-diag1, const-diag2)

E produzido-por(const-diag1, evento-diag1)

E produzido-por(const-diag2, evento-diag2)

IMPLICA antes(evento-diag1, evento-diag2);

sobrecrescido -por(const-diag1, const-diag2)

E produzido-por(const-diag1, evento-diag1)

E produzido-por(const-diag2, evento-diag2)

IMPLICA antes(evento-diag1, evento-diag2);

alternado -com(const-diag1, const-diag2)

E produzido-por(const-diag1, evento-diag1)

E produzido-por(const-diag2, evento-diag2)

IMPLICA durante(evento-diag1, evento-diag2);

engolfado -por(const-diag1, const-diag2)

E produzido-por(const-diag1, evento-diag1)

E produzido-por(const-diag2, evento-diag2)

IMPLICA durante(evento-diag1, evento-diag2);

entrecrescido-com(const-diag1, const-diag2)

E produzido-por(const-diag1, evento-diag1)

E produzido-por(const-diag2, evento-diag2)

IMPLICA durante(evento-diag1, evento-diag2);

engolfando(const-diag, const-min)

E produzido-por(const-diag, evento-diag1)

E produzido-por(const-min, evento-diag2)

IMPLICA durante(evento-diag1, evento-diag2);

dentro-de-porosidade-intergranular-primária(const-diag, const-macrop)

E produzido-por(const-diag, evento-diag1)

E produzido-por(const-diag2, evento-diag2)

IMPLICA depois(const-macrop, const-macrop);

FIM BASE-CONHECIMENTO conhecimento-ordenação-eventos;

FIM CONHECIMENTO-DOMÍNIO

CONHECIMENTO-INFERÊNCIA inferências-ordenação;

PAPEL-CONHECIMENTO feições;

TIPO: DINÂMICO;

MAPEAMENTO-DOMÍNIO:

LISTA-DE constituinte-de-rocha;

FIM PAPEL-CONHECIMENTO feições;

PAPEL-CONHECIMENTO feição;

TIPO: DINÂMICO;

MAPEAMENTO-DOMÍNIO:

constituinte-de-rocha;

FIM PAPEL-CONHECIMENTO feição;

PAPEL-CONHECIMENTO evento;

TIPO: DINÂMICO;

MAPEAMENTO-DOMÍNIO:

evento-diagenético;

FIM PAPEL-CONHECIMENTO evento;

PAPEL-CONHECIMENTO associação;

TIPO: DINÂMICO;

MAPEAMENTO-DOMÍNIO:

relação-paragenética;

FIM PAPEL-CONHECIMENTO associação;

PAPEL-CONHECIMENTO aspecto-temporal;  
 TIPO: DINÂMICO;  
 MAPEAMENTO-DOMÍNIO:  
 antes, durante, depois;  
 FIM PAPEL-CONHECIMENTO aspecto-temporal;

PAPEL-CONHECIMENTO feição-associada;  
 TIPO: DINÂMICO;  
 MAPEAMENTO-DOMÍNIO:  
 constituinte-de-rocha;  
 FIM PAPEL-CONHECIMENTO feição-associada;

PAPEL-CONHECIMENTO evento-associado;  
 TIPO: DINÂMICO;  
 MAPEAMENTO-DOMÍNIO:  
 evento-diagenético;  
 FIM PAPEL-CONHECIMENTO evento-associado;

PAPEL-CONHECIMENTO par-ordenado-eventos;  
 TIPO: DINÂMICO;  
 MAPEAMENTO-DOMÍNIO:  
 antes, durante, depois;  
 FIM PAPEL-CONHECIMENTO par-ordenado-eventos;

PAPEL-CONHECIMENTO pares-ordenados-eventos;  
 TIPO: DINÂMICO;  
 MAPEAMENTO-DOMÍNIO:  
 CONJUNTO-DE (PAR-DE evento-diagenético);  
 FIM PAPEL-CONHECIMENTO pares-ordenados-eventos;

PAPEL-CONHECIMENTO regras-feição-evento;  
 TIPO: ESTÁTICO;  
 MAPEAMENTO-DOMÍNIO:  
 indicação-evento DE conhecimento-ordenação-eventos;  
 FIM PAPEL-CONHECIMENTO regras-feição-evento;

PAPEL-CONHECIMENTO regras-associação-ordem;  
 TIPO: ESTÁTICO;  
 MAPEAMENTO-DOMÍNIO:  
 implicação-temporal DE conhecimento-ordenação-eventos;  
 FIM PAPEL-CONHECIMENTO regras-associação-ordem;

PAPEL-CONHECIMENTO par-de-eventos;  
 TIPO: DINÂMICO;  
 MAPEAMENTO-DOMÍNIO:  
 PAR-DE evento-diagenético;  
 FIM PAPEL-CONHECIMENTO par-de-eventos;

PAPEL-CONHECIMENTO ordem;  
 TIPO: DINÂMICO;  
 MAPEAMENTO-DOMÍNIO:  
 antes, durante, depois;  
 FIM PAPEL-CONHECIMENTO ordem;

PAPEL-CONHECIMENTO evento1;  
 TIPO: DINÂMICO;  
 MAPEAMENTO-DOMÍNIO:  
 evento-diagenético;  
 FIM PAPEL-CONHECIMENTO evento1;

PAPEL-CONHECIMENTO evento2;  
 TIPO: DINÂMICO;  
 MAPEAMENTO-DOMÍNIO:  
 evento-diagenético;  
 FIM PAPEL-CONHECIMENTO evento2;

PAPEL-CONHECIMENTO resultado1;  
 TIPO: DINÂMICO;  
 MAPEAMENTO-DOMÍNIO:  
 evento-diagenético;  
 FIM PAPEL-CONHECIMENTO resultado1;

PAPEL-CONHECIMENTO resultado2;  
 TIPO: DINÂMICO;  
 MAPEAMENTO-DOMÍNIO:

evento-diagnóstico;  
FIM PAPEL-CONHECIMENTO resultado2;

PAPEL-CONHECIMENTO sequência-eventos;  
TIPO: DINÂMICO;  
MAPEAMENTO-DOMÍNIO:  
LISTA-DE evento-diagnóstico;  
FIM PAPEL-CONHECIMENTO sequência-eventos;

INFERÊNCIA selecionar;  
PAPÉIS:  
ENTRADA:  
feições;  
SAÍDA:  
feição;  
ESPECIFICAÇÃO:  
"Nenhum conhecimento de domínio é usado como critério: a seleção é sequencial.";  
FIM INFERÊNCIA selecionar;

INFERÊNCIA especificar;  
PAPÉIS:  
ENTRADA:  
feição;  
SAÍDA:  
evento;  
ESTÁTICO:  
regras-feição-evento;  
ESPECIFICAÇÃO:  
"Entrada é uma feição do domínio. Saída é um evento.  
O evento é indicado pelas regras (conhecimento estático) que associam feições e eventos.";  
FIM INFERÊNCIA especificar;

INFERÊNCIA abstrair;  
PAPÉIS:  
ENTRADA:  
feição;  
SAÍDA:  
associação;  
ESPECIFICAÇÃO:  
"Uma feição é abstraída em alguns aspectos relevantes, neste caso, é abstraída na associação com a outra feição.";  
FIM INFERÊNCIA abstrair;

INFERÊNCIA especificar;  
PAPÉIS:  
ENTRADA:  
associação,  
feições;  
SAÍDA:  
feição-associada;  
ESPECIFICAÇÃO:  
"Busca na lista de feições aquela relacionada, através da associação, com a feição anterior.";  
FIM INFERÊNCIA specify;

INFERÊNCIA especificar;  
PAPÉIS:  
ENTRADA:  
associação;  
SAÍDA:  
aspecto-temporal;  
ESTÁTICO:  
regras-associação-ordem;  
ESPECIFICAÇÃO:  
"O aspecto temporal é indicado pela associação, usando as regras que relacionam associações e aspectos temporais.";  
FIM INFERÊNCIA especificar;

INFERÊNCIA especificar;  
PAPÉIS:  
ENTRADA:  
feição-associada;  
SAÍDA:  
evento-associado;  
ESTÁTICO:  
regras-feição-evento;  
ESPECIFICAÇÃO:  
"Entrada é uma feição do domínio. Saída é um evento.

O evento é indicado pelas regras (conhecimento estático) que associa feições e eventos.";  
 FIM INFERÊNCIA especificar;

INFERÊNCIA ordenar;

PAPÉIS:

ENTRADA:

evento,  
 evento-associado,  
 aspecto-temporal;

SAÍDA:

pares-ordenados-eventos;

ESPECIFICAÇÃO:

"Entrada é um par de eventos e a relação temporal entre eles. A inferência compõe um par ordenado de eventos.";

FIM INFERÊNCIA ordenar;

INFERÊNCIA selecionar;

PAPÉIS:

ENTRADA:

pares-ordenados-eventos;

SAÍDA:

par-evento;

ESPECIFICAÇÃO:

"Entrada é a lista de pares de eventos ordenados, da qual é selecionado um par.";

FIM INFERÊNCIA selecionar;

INFERÊNCIA especificar;

PAPÉIS:

ENTRADA:

par-evento;

SAÍDA:

ordem, evento1, evento2;

ESPECIFICAÇÃO:

"Os componentes do par de eventos são especificados.";

FIM INFERÊNCIA especificar;

INFERÊNCIA casar;

PAPÉIS:

ENTRADA:

evento, sequência-eventos;

SAÍDA:

resultado;

ESPECIFICAÇÃO:

"Tenta casar evento a qualquer evento já adicionado na sequencia.";

FIM INFERÊNCIA casar;

INFERÊNCIA ordenar;

PAPÉIS:

ENTRADA:

resultado1, resultado2, ordem;

SAÍDA:

sequência-eventos;

ESPECIFICAÇÃO:

"Tenta casar evento a qualquer evento já adicionado na sequencia.";

FIM INFERÊNCIA ordenar;

FIM CONHECIMENTO-INFERÊNCIA

CONHECIMENTO-TAREFA tarefas-interpretacao-sequência;

TAREFA interpretação-sequencia-eventos;

OBJETIVO: "Define a ordem de ocorrência dos eventos que geraram as feições de domínio atuais.";

PAPÉIS:

ENTRADA:

feições: "Dados sobre as características do domínio";

SAÍDA:

sequência-eventos: "Sequencia de eventos na ordem em que eles ocorreram";

FIM TAREFA interpretação-sequencia-eventos;

MÉTODO-TAREFA interpretação-atraves-identificação-e-ordenação;

REALIZA:

interpretação-sequencia-eventos;

DECOMPOSIÇÃO:

TAREFAS: identificar-eventos, ordenar-eventos;

PAPÉIS:

INTERMEDIÁRIO: pares-de-eventos: "Lista de pares de eventos ordenados especificados pelas feições";

ESTRUTURA-CONTROLE:  
 identificar-eventos(feições -> pares-de-eventos);  
 ordenar-eventos(pares-de-eventos -> sequencia-eventos);  
 FIM MÉTODO-TAREFA interpretação-atraves-identificação-e-ordenação;

TAREFA identificar-eventos;  
 OBJETIVO: "Descobre os eventos que geraram as feições do domínio.";

PAPÉIS:  
 ENTRADA:  
 feições: "Dados sobre as características do domínio";  
 SAÍDA:  
 pares-de-eventos-ordenados: "Lista de pares de eventos ordenados.";

FIM TAREFA identificar-eventos;

MÉTODO-TAREFA método-identificar;

REALIZA:  
 identificar-eventos;

DECOMPOSIÇÃO:  
 INFERÊNCIAS: selecionar, especificar, abstrair, ordenar;

PAPÉIS:  
 INTERMEDIÁRIO:  
 feição: "Uma feição selecionada de uma lista de feições";  
 evento: "O evento que gerou a feição";  
 associação: "Informação de associação abstraída da feição";  
 aspecto-temporal: "Informação temporal indicada pela associação entre as feições";  
 feição-associada: "Feição associada à anterior";  
 evento-associado: "O evento que gerou a feição associada";  
 par-ordenado-evento: "Par de eventos dispostos na ordem em que ocorreram";

ESTRUTURA-CONTROLE:  
 ENQUANTO NOVA-SOLUÇÃO selecionar(feições -> feição) FAÇA  
 especificar(feição + regras-feição-evento -> evento);  
 abstrair(feição-> associação);  
 especificar(associação + feições -> feição-associada);  
 especificar(feição-associada + regras-feição-evento -> evento-associado);  
 especificar(associação + regras-associação-ordem -> aspecto-temporal);  
 ordenar(evento + evento-associado + aspecto-temporal -> par-ordenado-evento);  
 pares-de-eventos-ordenados := pares-de-eventos-ordenados ADICIONA par-ordenado-evento;

FIM ENQUANTO

FIM MÉTODO-TAREFA método-identificar;

TAREFA ordenar-eventos;  
 OBJETIVO: "Ordena o conjunto de pares ordenados para criar uma lista ordenada.";

PAPÉIS:  
 ENTRADA:  
 pares-de-eventos-ordenados: "Conjunto de pares de eventos ordenados.";

SAÍDA:  
 sequencia-eventos: "Sequência de eventos na ordem em que eles ocorreram";

FIM TAREFA ordenar-eventos;

MÉTODO-TAREFA método-ordenar

REALIZA:  
 ordenar-eventos;

DECOMPOSIÇÃO:  
 INFERÊNCIAS: selecionar, especificar, casar, ordenar;

PAPÉIS:  
 INTERMEDIÁRIO:  
 par-evento: "par de eventos";  
 ordem: "ordem dos eventos";  
 evento1: "primeiro evento do par";  
 evento2: "segundo evento do par";  
 resultado1: "resultado do casamento do primeiro evento com a sequência";  
 resultado2: "resultado do casamento do primeiro evento com a sequência";

ESTRUTURA-CONTROLE:  
 ENQUANTO NOVA-SOLUÇÃO selecionar(pares-de-eventos-ordenados -> par-evento) FAÇA  
 especificar(par-evento -> ordem + evento1 + evento2);  
 casar(evento1 + sequencia-eventos -> resultado1);  
 SE resultado1 == falso  
 ENTÃO sequencia-eventos := sequencia-eventos ADICIONA evento1;  
 FIM SE  
 casar(evento2 + sequencia-eventos -> resultado2);  
 IF resultado2 == falso  
 ENTÃO sequencia-eventos := sequencia-eventos ADICIONA evento2;  
 FIM IF  
 ordenar(evento1 + evento2 + ordem -> sequencia-eventos)

FIM ENQUANTO

FIM MÉTODO-TAREFA método-ordenar;

FIM CONHECIMENTO-TAREFA tarefas-interpretação-sequência;

FIM MODELO-CONHECIMENTO

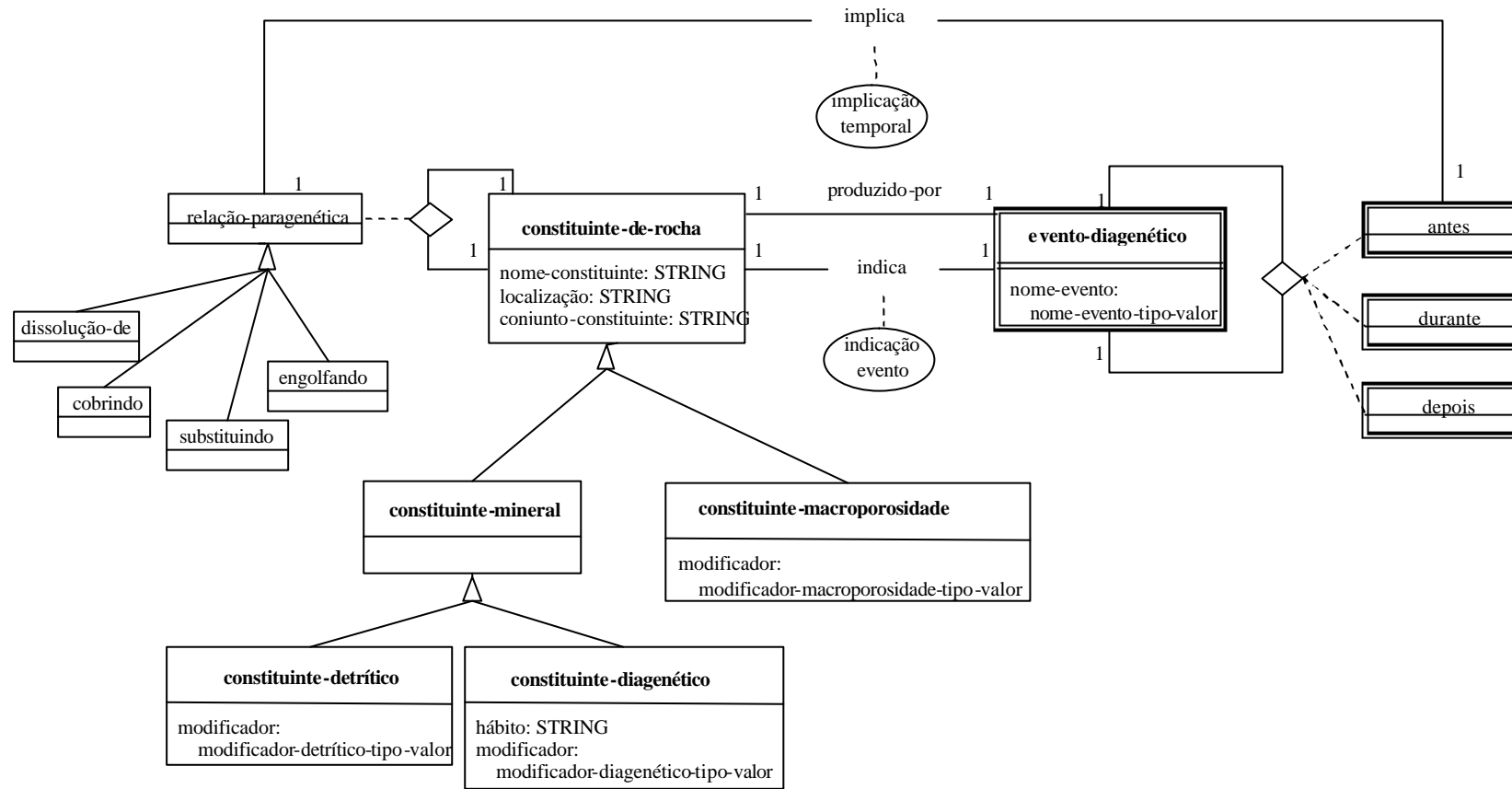


Figura 1: Modelo do domínio utilizando a linguagem UML.

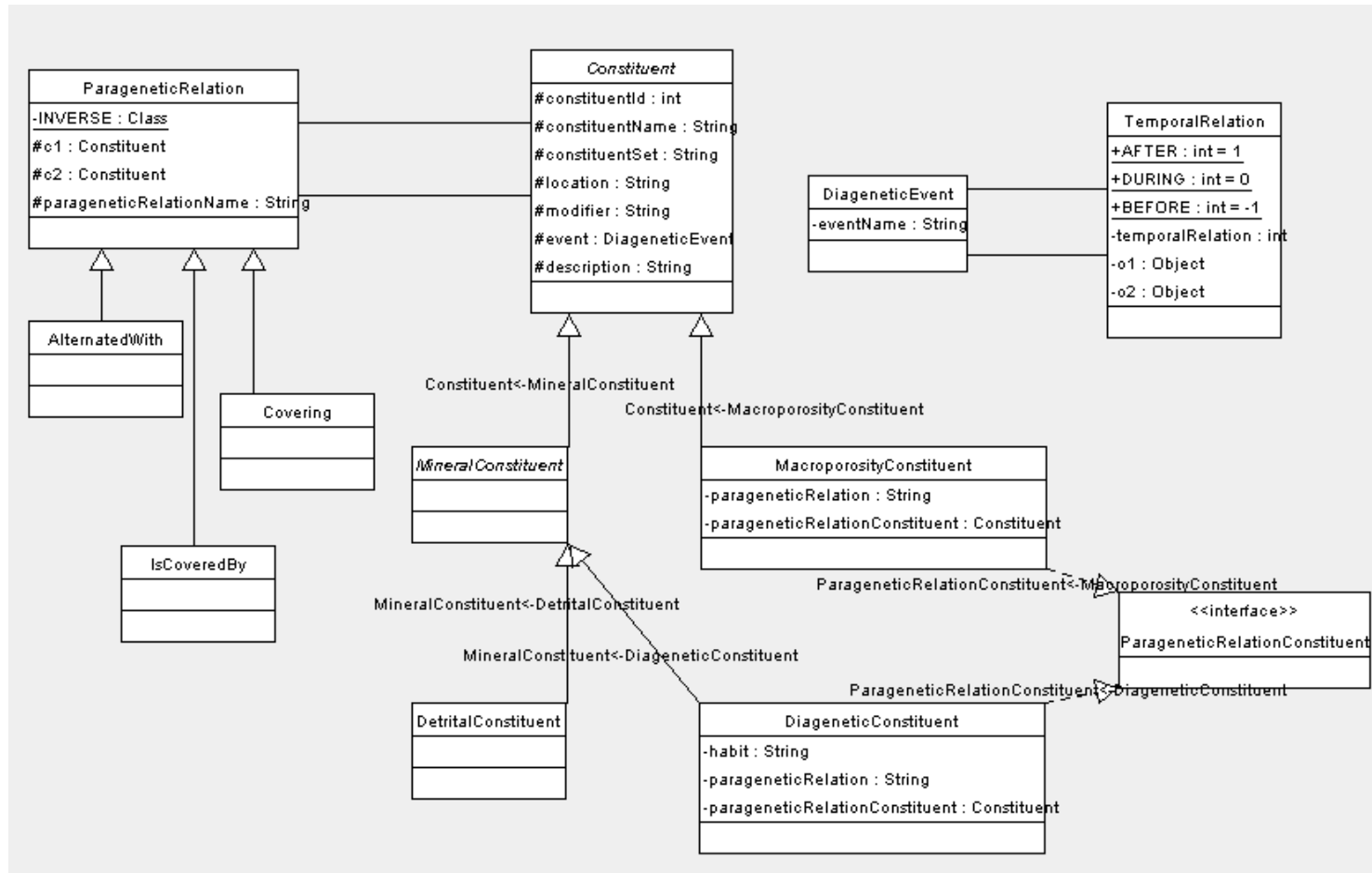


Figura 2: Modelo de classes Java que implementam a ontologia do domínio.



# ANEXO A REGRAS DE FORMAÇÃO DA SINTAXE DA LINGUAGEM CML

Nesse apêndice são demonstradas as convenções da sintaxe da linguagem CML e as regras de formação do construto *tipo-regra*, traduzidas de (SCHREIBER et al., 2000).

## A.1 Convenções da Sintaxe

As convenções da Tabela 1 são usadas na especificação da sintaxe de CML.

Tabela 1: convenções da sintaxe CML (modificada de (SCHREIBER et al., 2000)).

Construção	Definição
$X ::= Y$	A sintaxe de X (um não-terminal) é definido por Y.
$[ X ]$	Zero ou uma ocorrência de X.
$X^*$	Zero ou mais ocorrência de X.
$X^+$	Uma ou mais ocorrência de X.
$X Y \dots$	Uma ou mais ocorrência de X separadas por Y. Este construto é principalmente usado para abreviar listas separadas por vírgula. Por exemplo "Nome, ..." é uma abreviação para "Nome <,Nome>*".
$X / Y$	Um de X ou Y (ou exclusivo).
$< X >$	Construto para especificar o escopo dos operadores.
<b>símbolo</b>	Negrito: símbolos terminais pré-definidos da linguagem.
<i>Símbolo</i>	Itálico e capitalizado: símbolos terminais definidos pelo usuário.
símbolo	Caixa baixa: símbolos não terminais.
"Texto"	Texto arbitrário.
'X'	Escapa o símbolo operador (p.ex.: *) e denota o literal X.

## A.2 Sintaxe

A seguir é definida a sintaxe de construção da base de conhecimento, para chegar até a sintaxe de um tipo-regra. A proposta original permite que as instâncias de um tipo-regra sejam expressões entre conceitos do domínio.

```
tipo-regra ::=
    tipo-regra Tipo-regra ;
    corpo-tipo-regra
    fim tipo-regra [ Tipo-regra ; ] .

corpo-tipo-regra ::= tipo-regra-restrição | tipo-regra-implicação.

tipo-regra-restrição ::=
    restrição : tipo-definido-usuário [cardinalidade];

tipo-regra-implicação ::=
```

```

antecedente : tipo-definido-usuário [cardinalidade];
conseqüente : tipo-definido-usuário [cardinalidade];
símbolo-conexão : Nome-símbolo-conexão ;

```

```

tipo-definido-usuário ::= Conceito | Relação | Relação-binária |
Modelo-matemático .

```

```

base-conhecimento ::=
    base-conhecimento Base-conhecimento;
uso: uso-base-conhecimento, ...;
[ [instâncias :] <instância | tupla> + ]
[ variáveis : declaração-variáveis ; ... ; ]
[ expressões : expressão-base-conhecimento ... ; ]
fim base-conhecimento [ Base-conhecimento ; ] .

```

```

expressão-base-conhecimento ::= declaração-variável | expressão-
tipo-regra | "Texto" .

```

```

expressão-tipo-regra ::=
    equação |
    operador-tipo expressão-tipo-regra |
    expressão-tipo-regra operador-parte expressão-tipo-regra .

```

```

operador-tipo ::= sub-tipo-de | super-tipo-de | tipo-de .

```

```

operador-parte ::= tem-parte | dimensão | papel .

```

```

equação ::=
    `( ` equação ` ) ` |
    operador-sinal equação |
    operador-negação equação |
    equação operador-artimético equação |
    equação operador-lógico equação |
    equação operador-comparação equação |
    equação operador-dereferência equação |
    equação operador-implicação equação |
    expressão-variável |
    expressão-função |
    expressão-condicional .

```

```

operador-implicação ::= `?` | `?` | `?`

```