

UNIVERSIDADE DO VALE DO RIO DOS SINOS - UNISINOS  
CIÊNCIAS EXATAS E TECNOLÓGICAS,  
PROGRAMA INTERDISCIPLINAR DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO  
APLICADA - PIPCA  
NÍVEL MESTRADO

GIOVANI MANICA BARILI

**SisAC - Sistema de Auxílio à Classificação. Estudo de  
caso: Ostracode**

SÃO LEOPOLDO  
2011

GIOVANI MANICA BARILI

**SisAC - Sistema de Auxílio à Classificação. Estudo de  
caso: Ostracode**

Dissertação submetida à avaliação  
como requisito parcial para a obtenção  
do grau de Mestre em Computação  
Aplicada

Orientador: Prof Dr. João Francisco  
Valiati

## CIP — CATALOGAÇÃO NA PUBLICAÇÃO

B252s Barili, Giovani Manica

SisAC - Sistema de Auxílio à Classificação. Estudo de caso: Ostracode / por Giovani Manica Barili. — São Leopoldo, 2011.

105 f.: il. color. ; 30cm.

Dissertação (mestrado) — Universidade do Vale do Rio dos Sinos - UNISINOS, Ciências Exatas e Tecnológicas, Programa Interdisciplinar de Pós-Graduação em Computação Aplicada - PIPCA, 2011.

Orientação: Prof. Dr. João Francisco Valiati, Ciências Exatas e Tecnológicas.

1. Sistemas de Classificação. 2. Sistemas Especialistas (Computação). 3. Conhecimento Baseado em Regras. 4. Processamento de Imagens. I. Valiati, João Francisco. II. Título.

CDU 004.424.5  
004.891  
004.932

Catálogo na publicação:

Bibliotecária Carla Maria Goulart de Moraes - CRB 10/1252

## AGRADECIMENTOS

Ao meu pai por ter sempre me incentivado a conquistar meus objetivos e nunca desistir do que acredito. Agradeço a Deus, pelo pai que tive, pois ele foi e continua sendo a minha fonte de inspiração, espelho pelo qual procuro me espelhar; tentando ser tão bom ou melhor.

A minha mãe, por ter acreditado e me apoiado na realização desse mestrado.

A minha namorada, por todo apoio e pela paciência e inúmeros finais de semana que passamos em casa devido aos meus estudos.

Ao meu orientador, por ter acreditado e confiado no meu trabalho, e me guiado durante esses dois anos.

Ao meu ex-orientador do trabalho de conclusão, Adelmo Cechin, por ter me auxiliado e me apoiado a entrar no curso de mestrado.

A CAPES pelo apoio financeiro.

Dedico este trabalho à minha mãe, minha namorada e a meu pai em especial.

“"Sorte" é algo que decorre do encontro da preparação com a oportunidade. Esta é oferecida pelo meio, mas a preparação constante e contínua deve ser conduzida pela pessoa que quer ser bem-sucedida e deseja investir em seu próprio talento”

*(Dimitris N. Chorafas)*

“Jamais considere seus estudos como uma obrigação, mas como uma oportunidade invejável (...) para aprender a conhecer a influência libertadora da beleza do reino do espírito, para seu próprio prazer pessoal e para proveito da comunidade à qual seu futuro trabalho pertencer”

*(Albert Einstein)*

“Há dois tipos de sabedoria: a inferior e a superior. A sabedoria inferior é dada por tudo o que uma pessoa sabe, e a superior é dada pela consciência do que não se sabe. Os verdadeiros sábios são os mais convictos da sua ignorância. Desconfiem das pessoas auto-suficientes. O orgulho é um golpe contra a lucidez, um atentado contra a inteligência.

A sabedoria superior tolera; a inferior julga; a superior alivia, a inferior culpa; a superior perdoa, a inferior condena. Na sabedoria inferior há diplomas, na superior ninguém se diploma, não há mestres nem doutores, todos são eternos aprendizes”

*(Augusto Cury, O Mestre do Amor)*

## Resumo

O constante aumento no volume de informações em diversas áreas do conhecimento, via produção de material bibliográfico, multimídia e web, fruto dos avanços de pesquisas e estudos científicos, tem gerado um problema que é o gerenciamento e integração da informação de diferentes fontes. Com isso, a busca desses conhecimentos acaba tornando-se complexa e custosa, visto o volume de informações em alguns domínios de conhecimento. No entanto, esse conhecimento é de grande importância aos especialistas para raciocinar e chegar a conclusões que são utilizadas em tomadas de decisão ou como meio de solucionar problemas do domínio.

Com base nessas necessidades, essa dissertação propõe um sistema de auxílio à classificação que tem como meta, oferecer funcionalidades que permitam aos usuários especialistas realizar a manutenção de informações de um domínio na base de conhecimento e a disseminação dela entre os usuários do sistema, por meio de consultas e/ou sugestões acerca do conhecimento resultante. Onde os resultados apresentados pelo sistema, gerados por meio de um motor de inferência, são baseados em Sistemas Especialistas, que busca construir raciocínios a partir de informações que o usuário possui sobre as observações do cenário do domínio.

Juntamente com o sistema de inferência é proposto a associação de imagens relacionadas ao conhecimento, como forma de ilustrar as informações e descrições, e algoritmos de Processamento de Imagens para a redução da subjetividade nos casos de dúvida em relação às características visuais do estudo de caso, visto a falta de definições discretas de algumas informações que descrevem o conhecimento.

Como estudo de caso para demonstrar a aplicabilidade do sistema proposto, o trabalho foi focado no domínio da paleontologia, mais especificamente na classificação de espécies de Ostracodes, organismos que representam grande importância para a identificação de fontes petrolíferas.

Como resultado do trabalho obteve-se um sistema robusto e genérico, permitindo o armazenando de um grande volume de informações, separado pela área do conhecimento e sub-dividido por domínios de trabalho. Avaliações realizadas com usuários demonstram a efetividade da ferramenta e apontaram para a evolução de funcionalidades.

**Palavras-chave:** Sistemas de Classificação, Sistemas Especialistas (Computação), Conhecimento Baseado em Regras, Processamento de Imagens.

**TITLE:** “An Expert System Assisted by Image Processing for classification.Case: Ostracods”

## ABSTRACT

The constant increase in the volume of information of several areas of knowledge, through the production of bibliographic, multimedia and web material, is a result of advancements of research and scientific studies, and has led to issues regarding the management and integration of it. Therefore, the search for this knowledge becomes complex and difficult, considering the amount of information in some domains. However, this knowledge shows great importance to experts, who can analyze this information and use it to reach conclusions that are used in decision making or as a way for solving specific problems.

Based on these requirements, this dissertation proposes a system to aid classification task, aiming to provide functionalities that allow the expert to maintain domain information in a knowledge base. Also, it is possible to disseminate this information to the others users through searches and/or suggestions, presenting information about the resulting knowledge. The results presented by the system are generated by an inference engine based on Expert Systems, which seeks to produce reasoning from user’s information about the domain.

Along with the inference system is proposed a combination of images related to the knowledge as a way of illustrating the information and descriptions. Also, Image Processing algorithms are employed to reduce subjectivity in cases of uncertainty, regarding the visual characteristics of the case of study, considering the lack of discrete definitions for some information that describe the knowledge.

A system was proposed for integrating the Expert Systems and Image Processing techniques. As a way to demonstrate the system applicability, a pertinent problem of paleontology domain focused in classification of Ostracodes species, which have great importance to petroleum exploration, was developed.

The system was validated with users and experts that highlight its contribution: able to concentrate and store a big volume of information of many domains, incorporating support decision by image processing, and to be precise to map the expert knowledge. The experts cited as a major contribution that the system represents an application destined to inexperienced users, like students and novice researchers, used in the learning/training process.

**Keywords:** Expert System, Classification System, Rules Base, Image Processing.

## Sumário

<b>1</b>	<b>INTRODUÇÃO</b> . . . . .	<b>13</b>
1.1	OBJETIVO PRINCIPAL . . . . .	15
1.2	OBJETIVOS ESPECÍFICOS . . . . .	15
1.3	ORGANIZAÇÃO DO TRABALHO . . . . .	15
<b>2</b>	<b>PALEONTOLOGIA</b> . . . . .	<b>17</b>
<b>3</b>	<b>EMBASAMENTO TEÓRICO</b> . . . . .	<b>20</b>
3.1	SISTEMAS ESPECIALISTAS . . . . .	20
<b>3.1.1</b>	<b>Conhecimento Baseado em Regras</b> . . . . .	<b>21</b>
3.2	PROCESSAMENTO DE IMAGENS . . . . .	25
<b>3.2.1</b>	<b>Detecção de Cantos por K-Cossenos</b> . . . . .	<b>26</b>
<b>3.2.2</b>	<b>Matriz de Co-Ocorrência</b> . . . . .	<b>28</b>
<b>4</b>	<b>TRABALHOS CORRELATOS</b> . . . . .	<b>32</b>
4.1	ABORDAGEM DE SISTEMA ESPECIALISTA PARA INTERPRETAÇÃO DE EXAMES RADIOLÓGICOS . . . . .	32
4.2	ABORDAGEM DE SISTEMA ESPECIALISTA PARA CLASSIFICAÇÃO DE PLANTAS . . . . .	33
4.3	ABORDAGEM DE SISTEMA ESPECIALISTA PARA CLASSIFICAÇÃO DE PESTES, DOENÇAS E ERVAS DANINHAS . . . . .	34
4.4	COMENTÁRIOS SOBRE OS TRABALHOS CORRELATOS . . . . .	35
<b>5</b>	<b>ARQUITETURA DO SISTEMA DE AUXÍLIO À CLASSIFICAÇÃO</b> . . . . .	<b>37</b>
5.1	ARQUITETURA E FUNCIONALIDADES DO SISTEMA DE AUXÍLIO À CLASSIFICAÇÃO . . . . .	37
5.2	INTERFACE DE SUPORTE A IMAGENS . . . . .	43
<b>5.2.1</b>	<b>Detecção de Cantos por K-Cossenos</b> . . . . .	<b>44</b>
<b>5.2.2</b>	<b>Matriz de Co-Ocorrência</b> . . . . .	<b>48</b>
5.3	SISTEMA ESPECIALISTA . . . . .	49
<b>5.3.1</b>	<b>Representação do Conhecimento</b> . . . . .	<b>49</b>
5.3.1.1	Conhecimento, Atributos e Valores . . . . .	50
5.3.1.2	Regras . . . . .	50
5.3.1.3	Domínio de Trabalho . . . . .	51
<b>5.3.2</b>	<b>Motores de Inferência</b> . . . . .	<b>51</b>
5.3.2.1	Motor de inferência em regras . . . . .	52
<b>6</b>	<b>INTERFACES DO SISTEMA DE AUXÍLIO À CLASSIFICAÇÃO</b> <b>54</b>	
6.1	MENU DE CONHECIMENTO ESPECIALISTA . . . . .	55
6.2	MENU SUPORTE À IMAGEM . . . . .	64

6.3	MENU SISTEMA . . . . .	65
6.3.1	<b>Interfaces para Processamento de Imagem</b> . . . . .	71
7	<b>VALIDAÇÃO DO SISTEMA</b> . . . . .	74
7.1	FORMA DE VALIDAÇÃO . . . . .	74
7.2	RESULTADOS OBTIDOS . . . . .	75
7.2.1	<b>Classificação de espécimes</b> . . . . .	75
7.2.2	<b>Avaliação do sistema</b> . . . . .	76
7.2.3	<b>Observações e críticas</b> . . . . .	79
7.3	COMENTÁRIO GERAL DOS RESULTADOS . . . . .	81
8	<b>CONCLUSÕES</b> . . . . .	82
	<b>BIBLIOGRAFIA</b> . . . . .	84
	<b>APÊNDICE A - QUESTIONÁRIO UTILIZADO NA ATIVIDADE DE AVALIAÇÃO</b> . . . . .	87
	<b>APÊNDICE B RESPOSTAS DAS QUESTÕES DISSERTATIVAS</b> . . . . .	98
APÊNDICE B.1	QUAIS FORAM AS MAIORES DIFICULDADES DURANTE A CLASSIFICAÇÃO DE OSTRACODES UTILIZANDO O SEU CONHECIMENTO PRÓPRIO? . . . . .	98
APÊNDICE B.2	QUAIS FORAM AS MAIORES DIFICULDADES DURANTE A CLASSIFICAÇÃO DE OSTRACODES UTILIZANDO O SISTEMA SISAC? . . . . .	99
APÊNDICE B.3	POR MEIO DAS FUNCIONALIDADES OFERECIDAS PELO SISTEMA SISAC, COMO O SISTEMA AUXILIOU NO PROCESSO DE CLASSIFICAÇÃO DOS OSTRACODES? . . . . .	100
APÊNDICE B.4	ENTRE O PROCESSO DE CLASSIFICAÇÃO DE FORMA MANUAL E COM O SISTEMA, QUAL DAS DUAS ATIVIDADES CONSUMIU MAIS TEMPO PARA A REALIZAÇÃO DA MESMA? . . . . .	101
APÊNDICE B.5	DESCREVA/COMENTE COMO AS IMAGENS AUXILIARAM OU PODERIAM AUXILIAR NO PROCESSO DE CLASSIFICAÇÃO POR MEIO DAS FUNCIONALIDADES? . . . . .	102
APÊNDICE B.6	DESCREVE AS DIFERENÇAS OBSERVADAS EM RELAÇÃO AS DIFICULDADES ENCONTRADAS NA CLASSIFICAÇÃO DE OSTRACODES UTILIZANDO O SEU CONHECIMENTO EM RELAÇÃO AO AUXÍLIO OFERECIDO PELO SISTEMA? . . . . .	103
APÊNDICE B.7	ESCREVA SUAS CRÍTICAS, OBSERVAÇÕES, SUGESTÕES QUE POSSAM AUXILIAR E APRIMORAR O DESENVOLVIMENTO CONTÍNUO DO SISTEMA. . . . .	104

APÊNDICE C FOTOS TIRADAS DURANTE A AVALIAÇÃO DO SISTEMA . . . . . 105

## LISTA DE FIGURAS

Figura 2.1	Imagem de microscópio eletrônico de Ostracodes . . . . .	19
Figura 3.1	Exemplo de regras baseados na estrutura de sentenças do tipo <i>SE-ENTÃO</i> . . . . .	23
Figura 3.2	Fluxo de processos para PI, partindo da aquisição até a interpretação	26
Figura 3.3	Cálculo do K-Cosseno no ponto $P_i$ com $K = 40$ . . . . .	27
Figura 3.4	(a) Imagem de um trevo no formato binário e (b) Contorno do trevo.	27
Figura 3.5	Cosseno de cada pixel do contorno do trevo, com $K = 10$ . . . . .	28
Figura 3.6	A partir da imagem de um (a) trevo foram detectados os (b) cantos de ângulos maiores utilizando um limiar de -0.6 e os (c) cantos de ângulos menores utilizando um limiar de -0.99. . . . .	28
Figura 3.7	Ângulos e pixels vizinhos para cálculo da MCO . . . . .	29
Figura 3.8	Exemplo de uma imagem com dimensão de 5 x 5 e com 3 tons de cinza, onde (a) apresenta a imagem de forma visual e (b) apresenta a matriz de duas dimensões que representam a imagem de forma numérica . . . . .	29
Figura 3.9	MCO segundo a sua direção . . . . .	30
Figura 3.10	MCO da direção horizontal ( $0^\circ$ ) normalizada . . . . .	30
Figura 5.1	Diagrama de casos de uso do sistema de proposto . . . . .	38
Figura 5.2	Arquitetura cliente/servidor do sistema de auxílio à classificação . .	40
Figura 5.3	Diagrama de relacionamento do banco de dados do sistema de auxílio a classificação . . . . .	42
Figura 5.4	Arquitetura da parte cliente do sistema de auxílio à classificação . .	43
Figura 5.5	Arquitetura do componente de Interface de Suporte a Imagem . . .	44
Figura 5.6	(a) Imagem original capturada em um microscópio eletrônico e (b) Imagem com fundo preto . . . . .	45
Figura 5.7	(a) Imagem original e imagens binarizadas com limiar igual a: (b) 26, (c) 128 e (d) 192 . . . . .	46
Figura 5.8	(a) Imagem binarizada e (b) imagem do contorno do espécime . . .	46
Figura 5.9	Gráfico da variação do cosseno de um contorno . . . . .	47
Figura 5.10	Resultado obtido pelo algoritmo de detecção de cantos por K-Cossenos	47
Figura 5.11	(a) Imagem de um espécime limpo e inteiro, (b) imagem de espécime com rachaduras e cobertos por sedimentos e (c) imagem de espécime que sofreram processo de erosão . . . . .	48
Figura 5.12	Exemplo de uma regra do sistema de auxílio a classificação utilizando lógica booleana . . . . .	51
Figura 5.13	Arquitetura do componente Sistemas Especialistas . . . . .	52

Figura 5.14	Exemplo de configuração de memória de trabalho e conjunto de regras	53
Figura 6.1	Fluxo das interfaces do sistema implementado	54
Figura 6.2	(a) Interface de login e (b) Interface para seleção do domínio de trabalho.	55
Figura 6.3	Interface de gerência do conhecimento	56
Figura 6.4	Interface de visualização do conhecimento	57
Figura 6.5	Interface de gerência de atributos	58
Figura 6.6	Interface de gerenciamento do valor de um atributo	58
Figura 6.7	Interface de visualização de atributos e valores	59
Figura 6.8	Interface de gerenciamento de regras	60
Figura 6.9	Interface para associação de condições a regra gerenciada	61
Figura 6.10	Interface de visualização de regras	62
Figura 6.11	Interface de visualização das melhorias sugeridas na base de regras	62
Figura 6.12	Interface que apresenta as melhorias ou nova sugerida	63
Figura 6.13	Interface para manutenção dos conhecimentos que fazem parte de um respectivo domínio de trabalho	64
Figura 6.14	Interface de suporte a imagem para o algoritmo de MCO	65
Figura 6.15	Interface da inferência na base de conhecimentos	66
Figura 6.16	Atributos e valores possíveis de serem utilizados para inferir na base de conhecimento	67
Figura 6.17	Interface para sugestão de uma nova regra	68
Figura 6.18	Interface para exibição do conhecimento sugerido pelo sistema	69
Figura 6.19	Interface para sugestão de melhoria em uma regra já existente	70
Figura 6.20	Interface para validação dos conhecimentos sugeridos	70
Figura 6.21	Interface de exibição dos casos de conhecimentos validados	71
Figura 6.22	Interface para processamento de imagens utilizando o algoritmo de MCO	72
Figura 6.23	Interface para processamento de imagens utilizando o algoritmo de DCK	73
Figura 7.1	Resultado da classificação de espécimes referentes as atividades 1 e 2	76
Figura 7.2	Avaliação do manuseio do sistema	77
Figura 7.3	Avaliação do desempenho nos resultados de classificação	77
Figura 7.4	Avaliação de como o sistema apresentou as informações necessárias à classificação	78
Figura 7.5	Avaliação de como as imagens auxiliaram na classificação	78
Figura 7.6	Avaliação de como o sistema auxiliou na redução da subjetividade	79

## LISTA DE ABREVIATURAS

<b>CAD</b>	Computer Aided Diagnosis
<b>CBR</b>	Conhecimento Baseado em Regras
<b>DCK</b>	Detecção de Cantos por K-Cossenos
<b>FPI</b>	Ferramentas de Processamento de Imagem
<b>FTP</b>	File Transfer Protocol
<b>IA</b>	Inteligência Artificial
<b>LCC</b>	Limiar de Cantos Curvos
<b>MCO</b>	Matriz de Co-Ocorrência
<b>PI</b>	Processamento de Imagem
<b>SE</b>	Sistema Especialista
<b>SisAC</b>	Sistema de Auxílio à Classificação
<b>SMA</b>	Segundo Momento Angular
<b>SP</b>	Sistemas de Produção

## LISTA DE TABELAS

Tabela 3.1	Exemplo de um conjunto de regras de um Sistemas de Produção . .	22
Tabela 3.2	Transformação de uma cadeia de caracteres dado um conjunto de regras . . . . .	22

# 1 INTRODUÇÃO

A gerência do conhecimento vem a cada dia se tornando uma tarefa fundamental para especialistas que lidam com grande volume de informações, tanto em quantidade como em diversidade. Tais informações podem ser utilizadas em raciocínios futuros para tomada de decisões que visam a solução de problemas em um determinado domínio do conhecimento. No entanto, manter e organizar tanta informação torna-se algo difícil, sendo necessário utilizar estruturas para armazenamento e organização desse conhecimento. Como exemplos de domínios de conhecimento que possuem esse tipo de problema, pode-se citar: diagnose por imagem (PARK et al., 2009) (HUNG; CHEN, 2006); identificação de plantas, insetos, pestes, espécimes (CONTRERAS et al., 2003) (GONZALES-ANDUJAR et al., 2006) (GONZALES-ANDUJAR, 2009) (KALOUDIS et al., 2005); análise de dados e controle de sistemas (RODRÍGUEZ-SOLANO et al., 2006) (SRINIVAS et al., 2001) (QIAN et al., 2008) (JARMULAK et al., 2001).

Com isso, uma maneira de facilitar a recuperação desse conhecimento pode ser o uso de estruturas para organizar essa informação de forma classificada e catalogada, como exemplo poderia-se citar as árvores taxonômicas, chaves dicotômicas, atlas, entre outras. No entanto, essas formas de organização exigem que o especialista possua um conhecimento prévio, de quais informações necessita para concluir seu raciocínio e de onde deve procurar, continuamente, até chegar a uma conclusão de como solucionar os problemas do domínio.

Cada vez mais indústrias e setores da área acadêmica vem buscando ferramentas computacionais que auxiliem na solução de problemas e/ou tomada de decisão. Algumas dessas ferramentas são sistemas baseados em técnicas de Inteligência Artificial (IA), que buscam recriar comportamentos e ações semelhantes a de um humano em suas atividades de lidar com o conhecimento, como controlar um sistema ou tomar uma decisão. Dentre esses sistemas pode-se citar os Sistemas Especialistas (SEs), que visam basicamente recriar o raciocínio e organização do conhecimento especialista de um determinado domínio do conhecimento. Para isso utiliza-se componentes denominados de memória de trabalho e motor de inferência: responsáveis por representar a organização e armazenamento das informações de um domínio; e gerar um raciocínio para lidar com as informações contidas na memória, até chegar a uma conclusão, respectivamente. Ambos processos, semelhantes a organização e raciocínio intrínseco à mente de um especialista.

Com base nas necessidades de organização e recuperação dessas informações, este trabalho propõe o desenvolvimento de um sistema que auxilie na classificação de objetos de estudos com base nos conhecimentos armazenados no sistema, referentes a um determinado domínio, baseado em SEs com Conhecimento Baseado em Regras (CBR), auxiliando os especialistas a lidar com uma grande quantidade de informações. O sistema

proposto visa disponibilizar funcionalidades que permitam: a manutenção e organização de informações de forma espacial e categorizada; e a recuperação dessas informações de forma simples para especialistas e neófitos com vários níveis de conhecimento sobre o domínio.

Este trabalho também propõe a associação de informações visuais como forma de ilustração do conhecimento, por meio de imagens representativas ligadas a ele e através de algoritmos de Processamento de Imagens (PI). Essa abordagem é vista em trabalhos correlatos (PARK et al., 2009), (CONTRERAS et al., 2003), (GONZALES-ANDUJAR et al., 2006), (GONZALES-ANDUJAR, 2009) que buscam solucionar possíveis problemas de subjetividade acerca de domínios específicos, devido a falta de definições exatas de características que descrevem o conhecimento.

Um dos principais cenários onde o sistema foca-se é na recuperação de conhecimentos, referentes a objetos de estudo do domínio dos usuários especialistas, fornecendo auxílio na complementação do conhecimento dos usuários facilitando a tarefa de classificação e compreensão dos objetos de estudo. Para isso, assumindo que já exista uma base de conhecimento carregada com o conhecimento especialista, como entradas o usuário irá selecionar as características visíveis e conhecidas do objeto de estudo de um conjunto de valores e atributos que o sistema irá apresentar. A partir dessas características o sistema irá realizar inferências na base de conhecimento, retornando os conhecimentos resultante e um novo conjunto de atributos e valores para um possível aprimoramento dos resultados.

Como funcionalidade de apoio o sistema disponibiliza um mecanismo de PI, tal funcionalidade visa auxiliar na escolha dos valores de alguns atributos visuais, pois como mencionado a definição de algumas características que melhor descrevem visualmente o objeto de estudo pode apresentar subjetividades, o que pode ocasionar uma classificação incorreta. Essa funcionalidade possui como entrada uma imagem ou região do objeto de estudo a ser aplicado algum dos algoritmos de PI incluídos no sistema, onde o resultado será a sugestão de valor do atributo em questão.

Para demonstração da efetividade do sistema é realizado um estudo de caso que aborda o domínio da Paleontologia, focado nos microfósseis de Ostracodes<sup>1</sup>. Uma das maiores dificuldades encontradas pelos paleontólogos do estudo de caso é o grande número de espécimes presentes quando tratados Ostracodes, ultrapassando às milhares. A importância do estudo de Ostracodes na paleontologia está voltada no suporte à localização de possíveis fontes petrolíferas com base na identificação de espécimes de Ostracodes, os quais descrevem o período estatigráfico do local onde foram encontrados, o que sugere a possibilidade ou não da existência de reservas de petróleo.

As informações necessárias para a construção da base de conhecimento do estudo

---

<sup>1</sup>Crustáceos bentônicos utilizados na análise paleoambientais, colaborando para entender quando e como evoluiu uma determinada bacia

de caso são baseadas no conhecimento de especialistas do domínio Cristianini Trescastro Bergue, graduado em Ciências Biológicas, mestre em Geociências e doutor em Ciências pela Universidade do Rio Grande do Sul (UFRGS), e Gerson Fauth, graduado em Geologia, mestre em Geociências e doutor em Geociências pela *Heidelberg University*. Ambos atuam como pesquisadores do Laboratório de Micropaleontologia da Unisinos na área de taxonomia e bioestratigrafia de Ostracodes marinhos mesozóicos e cenozóicos.

## 1.1 OBJETIVO PRINCIPAL

Essa dissertação busca desenvolver um sistema que tem por objetivo auxiliar especialistas e neófitos na tarefa de classificação de objetos de estudo do domínio. Para isso, o sistema deverá oferecer funcionalidades que permitam armazenar grandes volumes de informações em forma de texto, imagens e documentos e permitir recuperar essas informações de forma simples e fácil para usuários especialistas e usuários com pouco conhecimentos na área.

Para a realização desses objetivos, essa dissertação propõem a utilização de SE, com motor de inferência de CBR, como mecanismos de filtragem e recuperação das informações de conhecimentos, juntamente com algoritmos de PI para auxiliar o usuário a definir características visuais que definem os objetos de estudo a serem classificados ou recuperados.

## 1.2 OBJETIVOS ESPECÍFICOS

A seguir serão apresentados os objetivos específicos referentes ao trabalho a ser realizado:

- a) Projetar o sistema, com módulos de SE e PI;
- b) Implementar o sistema projetado;
- c) Popular o sistema com o conhecimento do estudo de caso;
- d) Validar o sistema com usuário especialista e neófitos do estudo de caso por meio de questionário;
- e) Analisar o resultado dos questionários respondidos;

## 1.3 ORGANIZAÇÃO DO TRABALHO

Este trabalho está organizado em sete capítulos, sendo no capítulo 2 exposto uma explanação do estudo de caso utilizado no trabalho, apresentando a área de estudos e os microfósseis Ostracodes, espécimes da Micropaleontologia que auxiliam na identificação

de fontes petrolíferas. O capítulo 3 apresenta o referencial teórico sobre o tema que é abordado, como os SEs utilizados no sistema proposto e os algoritmos de PI que foram abordados como forma de redução da subjetividade. Esse capítulo também apresenta os trabalhos correlatos da literatura que se assemelham ao sistema proposto mostrando o estado da arte.

Na sequência, o capítulo 4 apresenta o sistema proposto expondo o projeto e seus sub-componentes, responsáveis pelo SE e PI. Também é descrito as partes participantes do sistema Cliente/Servidor que visam permitir a colaboração de usuários para a manutenção e ampliação da base de conhecimento. Já o capítulo 5 busca apresentar o sistema que resultou da implementação do projeto proposto, mostrando as suas interfaces e a forma de utilização.

Para validar o sistema proposto, no capítulo 6 é apresentada a metodologia abordada juntamente com os resultados obtidos e o comentário das análises realizadas. Por fim, no capítulo 7 são expostas as conclusões do trabalho, assim como suas contribuições e perspectiva de trabalhos futuros.

## 2 PALEONTOLOGIA

A Terra existe há 4,6 bilhões de anos e está em constante evolução provocando mudanças ambientais. Todas as modificações ocorridas afetaram diretamente a biota em nosso planeta e podem ser observadas nas rochas da superfície do mesmo. Animais e plantas que viveram na superfície da Terra podem ser encontrados na forma de fósseis e suas características e formas são preservadas nas rochas possibilitando aos paleontólogos estudá-las e compará-las. Sendo assim, hoje, o homem consegue reconstituir os ambientes e a geologia do passado através dos tipos de rochas e dos animais e plantas inseridas nas rochas na forma de fósseis.

Apenas há 600 milhões de anos que a vida na Terra passou a ocupar os espaços vazios, até então inóspitos, de nosso planeta. Durante o período inicial da escala geológica, no período Cambriano da era Paleozóica (entre 600 e 250 milhões de anos atrás), as formas de vida na Terra eram exclusivamente marinhas, pequenas e pouco especializadas. Posteriormente animais e plantas saíram do mar, ocuparam a Terra e evoluíram para formas maiores ocupando quase todos os nichos terrestres. Durante a era Mesozóica (entre 250 e 65 milhões de anos) os maiores animais em terra e no mar eram representados pelos dinossauros, que tiveram formas grandes e jamais observadas no registro fóssil. Durante o Cenozóico (entre 65 milhões de anos até hoje), o mais moderno, a Terra passou a ter flores, plantas com frutos e os mamíferos, que posteriormente originaram o homem (LEVIN, 2006).

Assim como a Terra registra em suas rochas a ocorrência de uma biota que viveu aqui em um passado geológico distante, animais invertebrados e plantas com tamanhos microscópicos também podem ser preservados nas rochas sedimentares, chamados de microfósseis. Os microfósseis mais estudados e conhecidos são: Foraminíferos (protistas, calcário, marinhos, bentônicos<sup>1</sup> ou planctônicos que vivem há 600 milhões de anos na Terra); Ostracodes (crustáceos, calcário, marinho e não-marinho, bentônico há 600 milhões de anos na Terra); Radiolários (protistas, silicosos, marinho, planctônicos que vivem há 600 milhões de anos na Terra); Nanofósseis (algas calcárias, marinho, planctônicas, que vivem há 200 milhões de anos na Terra); Dinoflagelados (algas, marinha, planctônica, que vivem há 200 milhões de anos na Terra). Todos estes grupos microfósseis podem ser importantes para datar as rochas, bem como colaborar para interpretações ecológicas e oceânicas (CARVALHO, 2004). Estas microscópicas formas de vida podem estar presentes no ar, lagos, rios ou mares do passado, sendo também encontrados nos dias atuais.

Os Ostracodes são pequenos crustáceos bentônicos cujo corpo mole, lateralmente comprimido, é envolvido por uma carapaça calcária bivalve de forma sub-ovalada ou

---

<sup>1</sup>Animais que vivem associados ao solo marinho, como por exemplo corais

reniforme, articulada dorsalmente, facilmente fossilizável. A maioria mede menos que 2 mm de comprimento e habita ambientes marinhos, mas com um grande número também em águas salobras e doces (HAQ; BOERSMA, 1998). Estes animais possuem mais de 65 mil espécies descritas e uma complexa estrutura de ornamentações e características que fazem com que sua descrição taxonômica seja complexa (HORNE et al., 2002). As ornamentações são estruturas externas das valvas que ocorrem de forma diversa e nomenclatura própria, e são excelentes feições diagnósticas, utilizadas para determinar os gênero, espécies e subespécies, de acordo com sua posição, tipo e presença ou ausência.

Os Ostracodes são muito utilizados como indicadores paleoambientais. Os estudos paleocológicos baseado nestes organismos apóiam-se em métodos como morfologia da carapaça e estrutura populacional. Quando trabalhamos com formas muito antigas fica um pouco mais difícil estabelecer relações com as formas viventes (COIMBRA; BERGUE, 2002). Outro exemplo de aplicação dos Ostracodes está na bioestratigrafia, na qual são usados para estabelecer zoneamentos dos estratos de determinadas bacias sedimentares, ou mesmo para colaborar na datação das rochas sedimentares. Na Figura 2.1 é apresentado um conjunto de imagens extraídas de microscópios eletrônicos de microfósseis de Ostracodes. Com base nessas imagens é possível notar a diversidade de espécimes presentes, do domínio e variedade de possíveis formas e ornamentações, que cada um pode possuir, permitindo que cada espécime possua diversos artigos na literatura descrevendo-o e classificando-o.

As controvérsias da taxonomia dos Ostracodes (e que, vale lembrar, existe em outros grupos também) remonta ao início da pesquisa com estes organismos, ainda na segunda metade do século XVIII (BERGUE, 2010). Alguns pesquisadores propuseram modelos de classificação de espécies atuais baseados na morfologia dos apêndices. Posteriormente, com a difusão do estudo paleontológico dos Ostracodes, a classificação passou a ser baseada na morfologia da carapaça, pois é esta em geral a única parte que fossiliza. Esta segregação gerou divergências entre os pesquisadores sobre quais seriam as características mais importantes na classificação. Apesar disso, é consenso que características como o arranjo das cicatrizes musculares centrais, a charneira e a ornamentação são os elementos básicos utilizados. Como outras ciências biológicas, a sistemática (da qual resultam as propostas taxonômicas) possui certa dose de subjetividade, sendo possível perceber linhas de pensamento entre os pesquisadores. Alguns tendem a limitar a proposição de novos gêneros e espécies, considerando a possibilidade de haver variação dentro de uma mesma espécie ("lumpers"). Outros veem cada modificação morfológica como o indicativo de uma divergência evolutiva e suficientes para a proposição de uma nova espécie ("splitters").

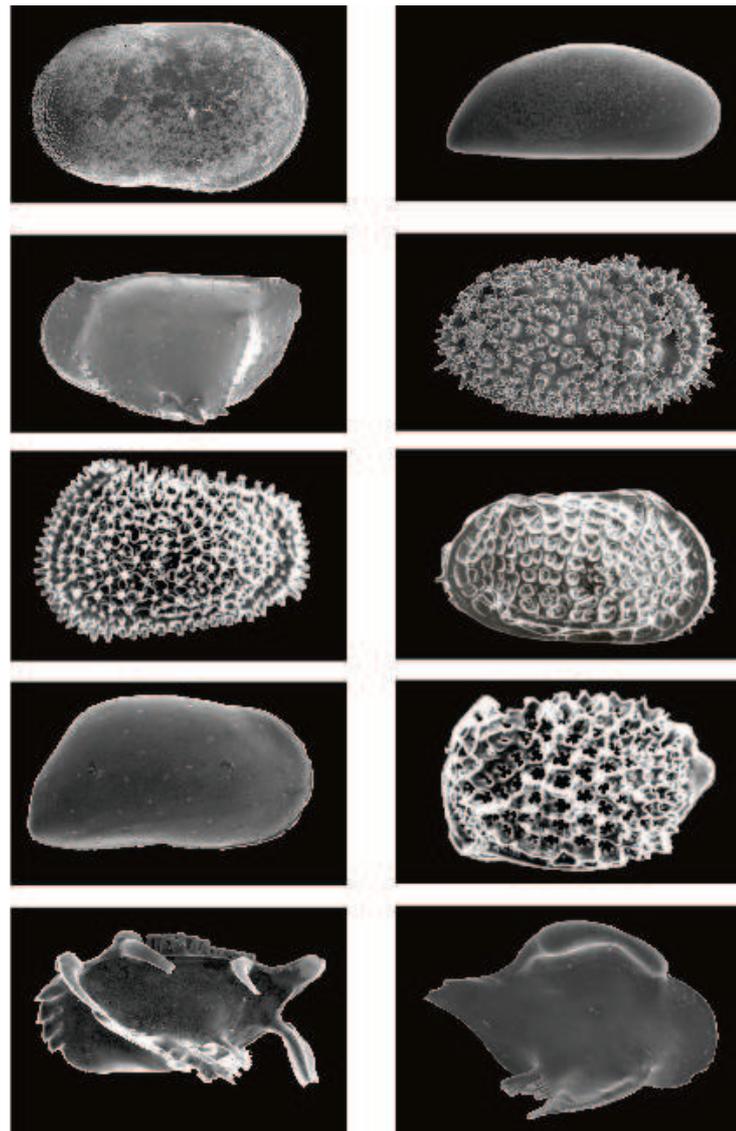


Figura 2.1: Imagem de microscópio eletrônico de Ostracodes

A importância da taxonomia é o fato dela constituir a base para as aplicações deste grupo nas geociências (especialmente a paleoecologia e bioestratigrafia) e biociências (zoogeografia, evolução). A identificação precisa de uma espécie define a acurácia de um zoneamento, a datação da ocorrência de um processo de dispersão ou da caracterização de um paleoambiente (o sedimento que originou a rocha depositando-se em um estuário ou em um lago, um mar raso ou em um mar profundo). Sendo esta a importância da correta identificação dos Ostracodes de uma fauna fóssil.

## 3 EMBASAMENTO TEÓRICO

Este capítulo tem como objetivo explorar o embasamento teórico, necessário para o desenvolvimento do trabalho. Conceitos teóricos da metodologia de SE, assim como trabalhos práticos, que apresentam a aplicabilidade de tais sistemas em soluções comerciais e científicas são apresentados.

O capítulo de embasamento está dividido em três seções. A primeira seção, Sistemas Especialista, apresenta a origem dos SEs e a teoria focada em CBR. Essa técnica é abordada nesse capítulo por ser utilizada no trabalho como mecanismo de recuperação de informações, baseando-se em um conjunto de características observadas pelo usuário, durante o processo de identificação de um objeto de estudo dentro de sua domínio de conhecimento.

A segunda seção apresenta uma breve descrição sobre PI, juntamente com os algoritmos descritores de imagem: Detecção de Cantos por K-Cossenos (DCK) e Matriz de Co-Ocorrência (MCO). A área de PI é abordada no trabalho como forma de minimizar o problema de subjetividade, o qual ocorre pela similaridade de algumas características, dos espécimes do estudo de caso, como contorno e a ornamentação. Ambos os métodos foram escolhidos por auxiliarem na escolha das característica de maior importância no processo de classificação e pela sua fácil compreensão e implementação, visto que o enfoque maior do sistema proposto por este trabalho está direcionado a um sistema baseado em SEs.

O método de DCK busca identificar regiões curvas do contorno de um objeto da imagem, sendo aplicado no sistema como forma de classificar o contorno do espécime da imagem entre uma conjunto de classes. Já a MCO visa classificar uma região de textura da imagem entre as possíveis classes de ornamentação dos espécimes do domínio, previamente definido pelos especialistas.

A terceira e última seção, Trabalhos Correlatos, apresenta trabalhos que baseiam-se de SEs como resolução de problemas, utilizando-se de abordagens em imagens como forma de auxílio no processo de classificação, demonstrando situações similares as abordadas pelo sistema proposto neste trabalho.

### 3.1 SISTEMAS ESPECIALISTAS

A IA, área do conhecimento de onde emerge os SEs, tem como meta o estudo e desenvolvimento de sistemas que apresentem ações ou comportamentos semelhantes a de humanos na execução de tarefas e tomadas de decisão. No entanto, diferentemente da IA que busca resolver problemas genéricos e complexos, os SEs buscam apenas resolver problemas de um conhecimento específico, simulando o raciocínio de um especialista do domínio (NIKOLOPOULOS, 1997).

A tomada de decisão de um especialista em um determinado problema, se dá através

de raciocínios compostos de fatos e hipóteses, previamente armazenados na memória, os quais foram adquiridos pelo especialista durante a sua formação acadêmica e por meio de experiências na sua vida profissional. O processo de raciocínio constitui-se da verificação de fatos, observados pelo especialista do problema a ser resolvido, comparando-os com as combinações de fatos e hipóteses de sua memória, criando assim novos fatos que poderão influenciar o fluxo do seu raciocínio, até gerar uma conclusão (RIBEIRO, 1987).

Os SEs direcionam o seu foco na quantidade de conhecimento específico e não em técnicas e implementações poderosas, como outras técnicas da área de IA. Para que esses sistemas sejam bem-sucedidos, eles devem conter um conhecimento bem-definido sobre uma determinada área e necessitam da ajuda de um ou mais especialistas, os quais precisam estar dispostos a dedicar tempo e esforços para transferir o seu conhecimento para o sistema (RICH; KNIGHT, 1993).

Um dos SEs mais conhecidos, amplamente apresentado na literatura (RICH; KNIGHT, 1993) (NIKOLOPOULOS, 1997) (BITTENCOURT, 2006), foi o MYCIN (SHORTLIFFE, 1976). Esse SE foi um dos pioneiros a ser desenvolvido e com um enfoque acadêmico e comercial, tendo como objetivo recomendar terapias apropriadas para pacientes com infecções bacteriológicas. O funcionamento desse SE era baseado na interação com um médico, a fim de adquirir dados clínicos, para a construção do seu raciocínio através de inferências em sua base de conhecimento, representada por regras, conduzindo o especialista ao uso de uma terapia apropriada, fornecendo justificativas para o resultado obtido (RICH; KNIGHT, 1993). Por meio da justificativa, o médico responsável pela decisão da terapia a ser indicada no tratamento, poderia verificar se a conclusão apresentada pelo SE era válida para os sintomas clínicos ou não. Caso o médico considerasse que a justificativa fosse incoerente com os sintomas, então ele poderia considerar a conclusão gerada pelo sistema como errada e ignorá-la (BITTENCOURT, 2006).

Ao longo dos anos, os estudos em desenvolvimento de sistemas computacionais que fossem capazes de encontrar soluções ou conclusões sobre um dado problema, geraram a publicação de diversos trabalhos. Em Liao (2005) é feita uma revisão desses trabalhos no período de uma década, que vai do início de 1995 ao final de 2004, onde o autor apresenta os trabalhos organizados conforme a técnica utilizada ou pela sua especificidade nas questões de arquitetura e armazenamento das informações.

### 3.1.1 Conhecimento Baseado em Regras

Os SEs foram inicialmente baseados em Sistemas de Produção (SP), idealizado por Post em 1943 (POST, 1943 apud BITTENCOURT, 2006). Um SP é constituído por pares de expressões, chamadas de regras, representadas por uma condição e uma ação. Sua arquitetura compõe-se por dois componentes passivos: *conjunto de regras* e uma *memória de trabalho*, e por um componente ativo: um *interpretador*, responsável por aplicar as regras nas cadeias de caracteres presentes na memória de trabalho.

As regras de Post representam o conhecimento de especificações sintáticas que quando aplicadas em uma cadeia de caracteres realizam transformações em caracteres conforme as ações das regras dada uma condição. Como exemplo de um SP, o conjunto de regras apresentadas na Tabela 3.1, aborda o problema da troca de caracteres de um e-mail.

Tabela 3.1: Exemplo de um conjunto de regras de um Sistemas de Produção

Regra N°	Condição	Ação
1	@	at
2	.	dot
3	>;<	,
4	>	Remove caracter
5	<	Remove caracter

Com o conjunto de regras definido, deve-se estipular em que conjunto de caracteres as regras serão aplicadas, como exemplo de sequência de caracteres <joao@email.com.br>;<maria@email.com.br>. Aplicando as regras da esquerda para a direita, a primeira transformação será baseada na regra 5, onde será removido o caracter <. Em seguida será aplicada a regra 1, onde será trocado o caracter @ pela sequência at, e assim sucessivamente até que não seja mais possível aplicar nenhuma regra, como apresentado na Tabela 3.2.

Tabela 3.2: Transformação de uma cadeia de caracteres dado um conjunto de regras

Pasos	Regra	Sequência de Caracteres
1	5	joao@email.com.br>;<maria@email.com.br>
2	1	joao at email.com.br>;<maria@email.com.br>
3	2	joao at email dot com.br>;<maria@email.com.br>
4	2	joao at email dot com dot br>;<maria@email.com.br>
5	3	joao at email dot com dot br, maria@email.com.br>
6	1	joao at email dot com dot br, maria at email.com.br>
7	2	joao at email dot com dot br, maria at email dot com.br>
8	2	joao at email dot com dot br, maria at email dot com dot br>
9	4	joao at email dot com dot br, maria at email dot com dot br

Os SEs mais recentes buscam trabalhar em uma quantidade maior de problemas de diversos domínios sendo capazes de apresentar comportamentos mais semelhantes ao de um especialista na tomada de decisão ou no raciocínio em seu domínio de conhecimento. Um SE possui uma arquitetura semelhante aos sistemas de Post sendo formado também por três componentes: uma *memória de trabalho*; uma *base de regras* e um *motor de inferência*. A base de regras e a memória de trabalho representam a base de conhecimento dos SEs, local onde é armazenado o conhecimento sobre o domínio do especialista (BITTENCOURT, 2006).

Entretanto, a base de conhecimento de um SE pode ser formalizada de diversas maneiras podendo conter qualquer tipo de estrutura de dados, contanto que sigam métodos de representação do conhecimento como uma linguagem formal. Essa generalização da base de conhecimento dos SEs, permite que os sistemas sejam utilizados para outros tipos de conhecimentos além das sequências de caracteres propostas por Post. Entre essas formas de representação podemos citar Redes Semânticas, Redes de Petri, Lógica, Quadros, Orientação a Objetos, entre outros (BITTENCOURT, 2006).

Uma das formas mais simples e predominante para o armazenamento do conhecimento especialista é a representação por meio de regras, semelhante as regras de Post, porém utilizadas com a estrutura de sentenças na forma SE (*IF*) <condição> ENTÃO (*THEN*) <conclusão ou ação>. Através desse formalismo é possível gerar regras mais genéricas, assim utilizando-as em diversas áreas do conhecimento e não somente restritas a sequências de caracteres (NIKOLOPOULOS, 1997). Como exemplo didático, são apresentadas regras no domínio de conhecimento para a manutenção de veículos como demonstrado na Figura 3.1.

<p><b>SE</b> carro não liga <b>E</b> luzes não ligam <b>ENTÃO</b> bateria morta;</p>	<p><b>SE</b> carro liga <b>E</b> luzes não ligam <b>ENTÃO</b> luzes queimadas <b>OU</b> problemas elétricos;</p>
<p><b>SE</b> não liga <b>E</b> luzes ligam <b>ENTÃO</b> bateria fraca;</p>	<p><b>SE</b> carro liga <b>E</b> problemas elétricos <b>OU</b> luzes queimadas <b>ENTÃO</b> luzes não ligam;</p>

Figura 3.1: Exemplo de regras baseados na estrutura de sentenças do tipo *SE-ENTÃO*

Dado tal conjunto de regras é possível verificar se a bateria ou as luzes de um veículo estão com problemas e qual é a origem desse problema. No entanto, para realizar essa busca ou inferência na base de regras, partindo de uma configuração inicial, torna-se necessário um motor de inferência como algoritmo de raciocínio, mais sofisticados e robustos que o interpretador de Post. As abordagens de raciocínio mais utilizadas em SEs, que possuem sua representação do conhecimento em regras, são fundamentadas em

um raciocínio com encadeamento para frente, encadeamento para trás ou uma abordagem híbrida das duas formas.

De um modo geral, o raciocínio de encadeamento para frente é dirigido pelos dados ou fatos que são observados pelo especialista sobre o problema e informados ou incluídos na memória de trabalho (RICH; KNIGHT, 1993) (BITTENCOURT, 2006). Essa forma de raciocínio tem início na recuperação de todas as regras que tenham suas premissas ou condições satisfeitas pelos fatos presentes na memória de trabalho, adicionando suas conclusões juntamente com os fatos já conhecidos na memória. Esse processo é repetido até que não sejam encontrados novos fatos/conclusões, ou quando a consulta seja respondida, no caso em que uma única resposta é suficiente (RUSSEL; NORVIG, 2004).

Em vez de dirigido por fatos, o raciocínio para trás é dirigido pelo objetivo (RICH; KNIGHT, 1993) (BITTENCOURT, 2006). Esse algoritmo, pelo fato de ter seu início numa lista de objetivos, tem seu funcionamento no sentido inverso do encadeamento para frente, buscando todas as regras na base de conhecimento cujas condições sejam positivas para o objetivo informado. Com isso, as condições aceitas são incluídas nas listas de objetivos e chamado novamente o algoritmo, de forma recursiva, até que não haja mais regras ou novas condições. Caso uma dessas regras possua todas as suas condições satisfeitas, então a sua conclusão também é incluída na lista de objetivos (RUSSEL; NORVIG, 2004).

A escolha da abordagem de encadeamento do motor de inferência deve ser definida de acordo com o problema a ser resolvido. Encadeamentos para frente mostram-se eficientes para problemas de classificação, pois há uma grande quantidade de objetivos a serem encontrados. Já encadeamentos para trás mostram-se melhores em problemas de diagnóstico, pois existem poucos objetivos mas uma grande quantidade de estados-iniciais (BITTENCOURT, 2006).

Como exemplo de seleção dos algoritmos de encadeamentos, conforme o raciocínio necessário para resolver o problema com base nas regras para verificação da seguinte situação como apresentado na Figura 3.1, onde são avaliados a bateria e as luzes do veículo. Cada tipo de encadeamento pode ser considerado mais eficiente conforme o ponto de vista sobre o problema:

- Se é visível que o carro e as luzes apresentam problemas, e procura-se saber qual é a origem do problema, o encadeamento para frente mostra-se muito eficaz.
- Se as luzes não estão acendendo, porém não se sabe o real problema, tornando-se necessária a busca da origem do problema, por meio do encadeamento para trás.

Porém os algoritmos de encadeamentos podem acabar resultando em um conjunto de regras, cujas cláusulas são verdadeiras, dadas as informações apresentadas pelo usuário. Com isso, o algoritmo de busca tem de apresentar seus resultados na ordem em que as ações devam ser aplicadas ao problema, processo denominado de resolução de conflito.

Dentre as estratégias de resolução de conflito existem três abordagens básicas: preferência com base na regra, preferência com base nas condições, preferência com base na ação que seria executada pela regra.

A preferência com base na regra consiste basicamente em duas maneiras de se ordenar as regras resultantes. A maneira mais simples é apresentar as ações das regras resultantes conforme elas tenham sido incluídas no sistema em uma determinada ordem. A outra forma é apresentar as ações das regras mais específicas e posteriormente as ações das regras mais genéricas. Onde as regras específicas são as que possuem em sua pré-condição todas as pré-condições das demais regras mais alguma condição, assim a segunda regra é mais genérica que a primeira.

Já a preferência com base nas condições consiste em definir um valor de importância e ordenar as ações, conforme as condições utilizadas em cada regra. Quanto maior ou mais condições importantes existam uma regra, maior a preferência das ações dessa regra. Por fim, a preferência com base na ação é definir o mérito de cada ação, ou seja, é basicamente serem realizadas as ações que possuem os maiores méritos (RICH; KNIGHT, 1993).

## 3.2 PROCESSAMENTO DE IMAGENS

Como mencionado na introdução desse capítulo, a área de PI é abordada neste trabalho como forma de amenizar problemas de subjetividade, os quais decorrem da semelhança entre valores que caracterizam um espécime de microfóssil. Segundo Gonzalez e Woods (1992), a área de PI está focada na solução de problemas com a ajuda da percepção das máquinas sobre as imagens. Problemas típicos que visam obter a percepção de máquinas na análise de imagens em tarefas como o reconhecimento automático de caracteres, visão para máquinas industriais em linhas de montagem e inspeção, reconhecimento militar, classificação de radiografias, impressões digitais entre outras tarefas .

A área de PI envolve diversos processos, que geralmente são expressos em forma de algoritmos, como ilustrado na Figura 3.2. A primeira parte desses processos é a aquisição da imagem. Processo que consiste em, através de sinais de sensores, transformar uma imagem real em uma matriz de duas dimensões representada por valores numéricos. A segunda etapa é o pré-processamento da imagem, o qual utiliza-se de técnicas que aumentem o nível de contraste, reduza ruídos, determine regiões de textura das imagens entre outros (GONZALEZ; WOODS, 1992).

Com a imagem digitalizada, a próxima etapa é a segmentação da imagem, uma das mais difíceis tarefas do PI, eventualmente pode-se a partir dela determinar o sucesso da análise de uma imagem. A segmentação tenta subdividir uma imagem em partes ou objetos. O nível de subdivisões pode ser contínuo dependendo do problema a ser resolvido ou quando o objeto de interesse for detectado e isolado, podendo ser realizado

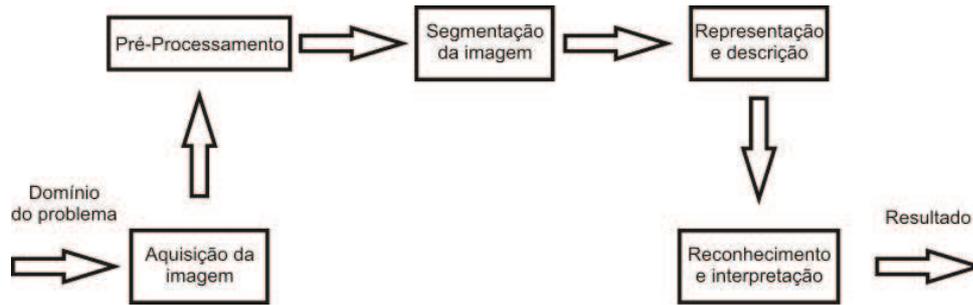


Figura 3.2: Fluxo de processos para PI, partindo da aquisição até a interpretação

através da detecção das bordas, a qual consiste em definir o contorno de duas regiões distintas (GONZALEZ; WOODS, 1992).

Após a detecção do contorno da imagem de um objeto, é possível representar e descrever o conjunto de pixels, que estão contidos na parte interna ou externa do contorno, de forma adequada para futuros processamentos computacionais. Basicamente a representação de uma região consiste na escolha de qual lado do contorno será descrito e que algoritmos serão utilizados. A região externa descreve o contorno do objeto, como por exemplo: características de tamanho, orientação de um trecho de linhas, o número de concavidades, cantos ou curvas do contorno. Já a região interna descreve propriedades refletivas, como cor e textura (GONZALEZ; WOODS, 1992).

Por fim, o último estágio do processamento envolve o reconhecimento e interpretação da imagem, processo que consiste em designar nomes a objetos identificados baseados nas descrições providas pelas etapas anteriores (GONZALEZ; WOODS, 1992).

O sistema de classificação somente irá utilizar-se do processo segmentação e do processo de representação e descrição das imagens. Entre os algoritmos implementados estão DCK, para descrição de curvas do contorno do objeto, e MCO, para descrição da textura do objeto da imagem. Ambos os métodos são descritos nas subseções seguintes.

### 3.2.1 Detecção de Cantos por K-Cossenos

Em Sun (2008) é apresentada uma técnica que descreve o contorno de uma imagem quanto aos seus cantos e/ou curvas do contorno de um objeto. Ela é baseada no cálculo dos  $K$ -Cossenos do contorno, onde quanto maior for o cosseno de um pixel maior a probabilidade da sua vizinhança ser uma curvatura do contorno.

Dado o contorno  $S$  de uma imagem, pode-se definir a sequência de pixels segundo a função

$$S = \{P_i = (x_i, y_i) | i = 1, 2, 3, \dots, m\}; \quad (3.1)$$

onde  $P_{i+k}$  representa o vizinho de  $P_i$   $K$  pixels distantes em um plano cartesiano  $(x_i, y_i)$ .

Com isso, para representar a curvatura de um contorno  $S$ , o valor do cosseno no pixel  $P_i$  é empregado como medida de quantificação de curvatura em uma determinada

região do contorno com um suporte  $K$ . Assim, o  $K$ -Cosseno de um contorno  $S = \{P_i | i = 1, 2, 3, \dots, m\}$  para cada ponto  $P_i$  é definido como

$$c_i(k) = \cos\theta_i = \frac{\vec{a}_i \cdot \vec{b}_i}{\|\vec{a}_i\| \cdot \|\vec{b}_i\|} \quad (3.2)$$

onde  $\vec{a}_i$  e  $\vec{b}_i$  são respectivamente os vetores entre um pixel  $P_i$  e os pixel  $P_{i+k}$  e  $P_{i-k}$ , de um plano cartesiano  $(x, y)$ .

A Figura 3.3 apresenta de forma ilustrada o cálculo do cosseno  $\theta$  do pixel  $P_i$  onde  $\theta$  é o ângulo entre  $\vec{a}_i(K) = P_{i+K} - P_i$  e  $\vec{b}_i(K) = P_{i-K} - P_i$  para todos  $K \in \mathbb{N}$ .

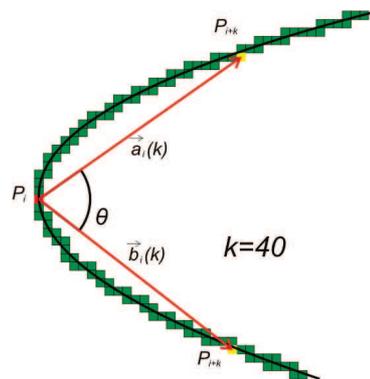


Figura 3.3: Cálculo do K-Cosseno no ponto  $P_i$  com  $K = 40$

Adaptado de Sun (2008)

Como exemplo, a Figura 3.4 (a) apresenta um trevo já no formato binário e a Figura 3.4 (b) apresenta o contorno, o qual foi calculado utilizando a ferramenta *OpenCV*, a qual fornece o método de busca de contornos.

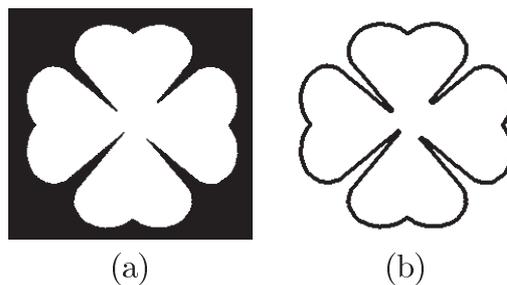


Figura 3.4: (a) Imagem de um trevo no formato binário e (b) Contorno do trevo.

Com o contorno  $S$  determinado é possível aplicar a equação 3.2 para todos  $P_i$  de  $S$ . A Figura 3.5 apresenta o resultado obtido para cada pixel do contorno, sendo o valor de  $K$  definido com o valor 10, levando em consideração o tamanho do contorno do objeto da imagem. É possível notar que há 4 picos com cosseno superior a 0.8 e mais 4 picos menores porém superiores a -0.6, sendo essas regiões as que apresentam curvas na imagem do trevo.

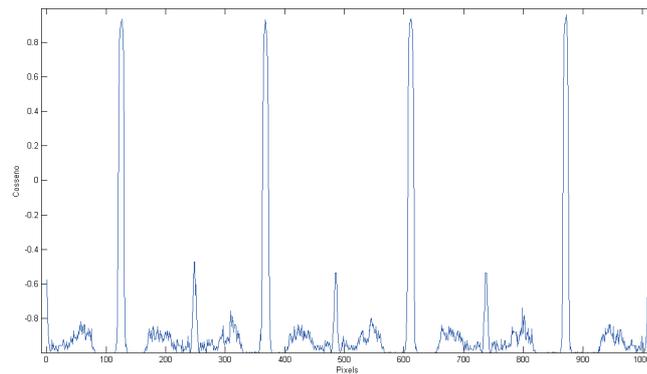


Figura 3.5: Cosseno de cada pixel do contorno do trevo, com  $K = 10$

Determinando um limiar, Limiar de Cantos Curvos (LCC), é possível definir se um conjunto de pixels está localizado em uma região curva do contorno ou se está numa região reta. A Figura 3.6 (b) apresenta os pixels que têm seu cosseno superior ao limiar de -0.6, já a Figura 3.6 (c) destaca os pixels que tem o seu cosseno superior a -0.99. Quanto maior for o cosseno, mais fechada é a curvatura de um contorno como demonstrado em ambas as imagens citadas, pois se o cosseno for igual a -1 então tais pixels estão localizados em uma reta do contorno. Para ambos foi utilizado  $K = 10$ .

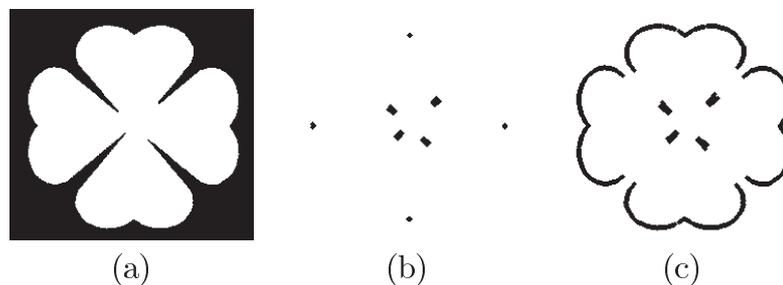


Figura 3.6: A partir da imagem de um (a) trevo foram detectados os (b) cantos de ângulos maiores utilizando um limiar de -0.6 e os (c) cantos de ângulos menores utilizando um limiar de -0.99.

### 3.2.2 Matriz de Co-Ocorrência

A MCO é uma técnica utilizada para descrever a textura de uma determinada região da imagem. Ela é baseada na frequência das variações dos níveis de cinza entre um pixel e seus vizinhos. A partir da MCO é possível extrair características estatísticas como Segundo Momento Angular (SMA), Energia, Contraste, Correlação, Variância entre outros. Em Haralick et al. (1973) são apresentados 14 tipos de cálculos que estatisticamente descrevem a textura.

Cada posição da matriz contém a quantidade de variações dos níveis de cinza, um dos eixos como o tom de cinza do pixel de origem e o outro eixo o tom de cinza do pixel

vizinho segundo uma distância e ângulo, sendo assim a matriz de co-ocorrência é uma matriz quadrada e simétrica de tamanho  $G$ , onde  $G$  é a quantidade de tons de cinza presentes na imagem.

Seja  $p$  um pixel de uma imagem, os seus vizinhos imediatos são aqueles que estão distantes a uma distância  $d$  nas direções horizontal, vertical e nas diagonais. Desta forma, os vizinhos de  $p = (x, y)$  são:

- na direção horizontal ou  $0^\circ \rightarrow (x + d, y)$  e  $(x - d, y)$
- na direção diagonal de  $45^\circ \rightarrow (x + d, y + d)$  e  $(x - d, y - d)$
- na direção vertical ou  $90^\circ \rightarrow (x, y + d)$  e  $(x, y - d)$
- na direção diagonal de  $135^\circ \rightarrow (x - d, y + d)$  e  $(x + d, y - d)$

Os vizinhos e as direções apresentados e que podem ser considerados na construção da MCO são ilustrados da Figura 3.7

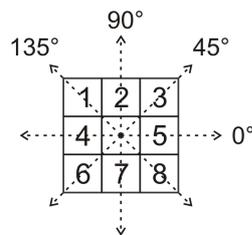


Figura 3.7: Ângulos e pixels vizinhos para cálculo da MCO

Como exemplo didático, se utilizada uma imagem de tamanho  $5 \times 5$  e com 3 tons de cinza, como representada na Figura 3.8, serão geradas 4 MCO, de dimensões  $3 \times 3$ . Onde cada uma delas irá representar a co-ocorrência segundo umas das possíveis direções, mencionadas anteriormente.

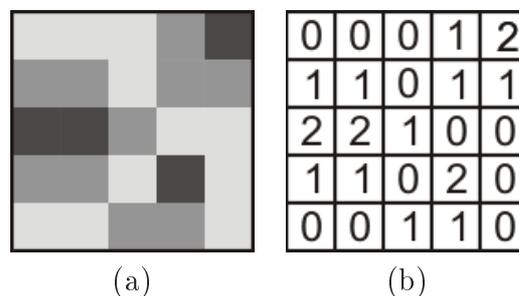


Figura 3.8: Exemplo de uma imagem com dimensão de  $5 \times 5$  e com 3 tons de cinza, onde (a) apresenta a imagem de forma visual e (b) apresenta a matriz de duas dimensões que representam a imagem de forma numérica

As matrizes de co-ocorrência resultantes são apresentadas na Figura 3.9, onde foi utilizado o valor de  $d$  igual a 1. Cada matriz possui dimensão  $3 \times 3$ , pois na imagem utilizada como exemplo havia uma escala de 3 tons de cinza.

<table style="border-collapse: collapse; margin: auto;"> <tr><td style="border-right: 1px solid black; padding: 5px;"></td><td style="padding: 5px;">0</td><td style="padding: 5px;">1</td><td style="padding: 5px;">2</td></tr> <tr><td style="border-right: 1px solid black; padding: 5px;">0</td><td style="padding: 5px;">8</td><td style="padding: 5px;">7</td><td style="padding: 5px;">1</td></tr> <tr><td style="border-right: 1px solid black; padding: 5px;">1</td><td style="padding: 5px;">7</td><td style="padding: 5px;">6</td><td style="padding: 5px;">2</td></tr> <tr><td style="border-right: 1px solid black; padding: 5px;">2</td><td style="padding: 5px;">1</td><td style="padding: 5px;">2</td><td style="padding: 5px;">2</td></tr> </table> <p>(a) Resultante da direção de <math>0^\circ</math></p>		0	1	2	0	8	7	1	1	7	6	2	2	1	2	2	<table style="border-collapse: collapse; margin: auto;"> <tr><td style="border-right: 1px solid black; padding: 5px;"></td><td style="padding: 5px;">0</td><td style="padding: 5px;">1</td><td style="padding: 5px;">2</td></tr> <tr><td style="border-right: 1px solid black; padding: 5px;">0</td><td style="padding: 5px;">4</td><td style="padding: 5px;">6</td><td style="padding: 5px;">2</td></tr> <tr><td style="border-right: 1px solid black; padding: 5px;">1</td><td style="padding: 5px;">6</td><td style="padding: 5px;">4</td><td style="padding: 5px;">4</td></tr> <tr><td style="border-right: 1px solid black; padding: 5px;">2</td><td style="padding: 5px;">2</td><td style="padding: 5px;">4</td><td style="padding: 5px;">9</td></tr> </table> <p>(b) Resultante da direção de <math>45^\circ</math></p>		0	1	2	0	4	6	2	1	6	4	4	2	2	4	9
	0	1	2																														
0	8	7	1																														
1	7	6	2																														
2	1	2	2																														
	0	1	2																														
0	4	6	2																														
1	6	4	4																														
2	2	4	9																														
<table style="border-collapse: collapse; margin: auto;"> <tr><td style="border-right: 1px solid black; padding: 5px;"></td><td style="padding: 5px;">0</td><td style="padding: 5px;">1</td><td style="padding: 5px;">2</td></tr> <tr><td style="border-right: 1px solid black; padding: 5px;">0</td><td style="padding: 5px;">6</td><td style="padding: 5px;">9</td><td style="padding: 5px;">1</td></tr> <tr><td style="border-right: 1px solid black; padding: 5px;">1</td><td style="padding: 5px;">9</td><td style="padding: 5px;">2</td><td style="padding: 5px;">6</td></tr> <tr><td style="border-right: 1px solid black; padding: 5px;">2</td><td style="padding: 5px;">1</td><td style="padding: 5px;">6</td><td style="padding: 5px;">0</td></tr> </table> <p>(c) Resultante da direção de <math>90^\circ</math></p>		0	1	2	0	6	9	1	1	9	2	6	2	1	6	0	<table style="border-collapse: collapse; margin: auto;"> <tr><td style="border-right: 1px solid black; padding: 5px;"></td><td style="padding: 5px;">0</td><td style="padding: 5px;">1</td><td style="padding: 5px;">2</td></tr> <tr><td style="border-right: 1px solid black; padding: 5px;">0</td><td style="padding: 5px;">6</td><td style="padding: 5px;">5</td><td style="padding: 5px;">2</td></tr> <tr><td style="border-right: 1px solid black; padding: 5px;">1</td><td style="padding: 5px;">5</td><td style="padding: 5px;">5</td><td style="padding: 5px;">3</td></tr> <tr><td style="border-right: 1px solid black; padding: 5px;">2</td><td style="padding: 5px;">2</td><td style="padding: 5px;">3</td><td style="padding: 5px;">0</td></tr> </table> <p>(d) Resultante da direção de <math>135^\circ</math></p>		0	1	2	0	6	5	2	1	5	5	3	2	2	3	0
	0	1	2																														
0	6	9	1																														
1	9	2	6																														
2	1	6	0																														
	0	1	2																														
0	6	5	2																														
1	5	5	3																														
2	2	3	0																														

Figura 3.9: MCO segundo a sua direção

Com as matrizes definidas, a normalização dos dados de co-ocorrência é feita através da equação

$$P_{i,j} = \frac{M_{i,j}}{\sum_{i,j=0}^{N-1} (M_{i,j})}, \text{ onde } i, j, N \in \mathbb{N}, \quad (3.3)$$

sendo  $P_{i,j}$  a probabilidade do valor da célula  $(i, j)$  ocorrer e  $M$  o valor da célula  $(i, j)$ , onde  $i$  e  $j$  são os índices das linhas e colunas. A Figura 3.10 apresenta a matriz de co-ocorrência com direção horizontal da Figura 3.9 (a) normalizada. (STEIN, 2005)

	0	1	2
0	0,235	0,205	0,029
1	0,205	0,176	0,058
2	0,166	0,058	0,058

Figura 3.10: MCO da direção horizontal ( $0^\circ$ ) normalizada

A partir das MCOs normalizadas é possível extrair valores pontuais sobre a textura da região analisada (STEIN, 2005). Alguns dos cálculos para a extração dessas informações, e que foram abordados no desenvolvimento deste trabalho são:

- *Segundo Momento Angular ou Energia*: descreve a homogeneidade ou ordenação dos pixels da imagem. Quanto maior forem os valores das células próximas a diagonal principal, significa que não houveram grandes variações de tons de cinza de um pixel para seu vizinho. O cálculo do SMA é obtido por meio da equação 3.4 e a Energia é obtida através da equação 3.5.

$$SMA = \sum_i \sum_j p(i, j)^2 \quad (3.4)$$

$$Energia = \sqrt{SMA} \quad (3.5)$$

- *Contraste*: descreve a variação local dos níveis de cinza. Diferentemente do SMA, o

contraste é uma medida que relaciona os valores que estão mais distantes da diagonal da MCO, quanto mais longe maior é o contraste da região do pixels e seus vizinhos. O valor de contraste é obtido por meio da equação

$$\text{Contraste} = \sum_{n=0}^{N_g-1} n^2 \left\{ \sum_i \sum_j p(i, j), \text{ se } |i - j| = n \right\} \quad (3.6)$$

- *Correlação*: estatisticamente mede o grau de associação entre os vizinhos. A correlação da MCO é feita através da equação

$$\text{Correlação} = \frac{\sum_i \sum_j p(i, j) - \mu_x \mu_y}{\sigma_x \sigma_y} \quad (3.7)$$

- *Homogeneidade*: é uma característica inversamente proporcional ao contraste. Quanto mais distante da diagonal principal estiver a variação de tons de cinza, menor é a homogeneidade de um pixel em relação aos seus vizinhos. A homogeneidade é obtida por meio da equação

$$\text{Homogeneidade} = \sum_{i,j=0}^{N-1} \frac{P_{i,j}}{1 + (i - j)^2} \quad (3.8)$$

## 4 TRABALHOS CORRELATOS

A busca por ferramentas que automatizem processos humanos ou que possam auxiliar na tomada de decisões abre espaço para pesquisa de métodos ou técnicas de Inteligência Artificial que atendam tais necessidades. Como mencionado anteriormente, em uma década, inúmeros trabalhos foram publicados, apresentando o desenvolvimento de sistemas com abordagens em SE em várias áreas do conhecimento.

SEs são amplamente utilizados como abordagem de sistemas onde a quantificação numérica já não é mais suficiente, sendo necessário um raciocínio mais clínico e próximo ao raciocínio humano. Áreas do conhecimento como a agricultura e a botânica necessitam identificar pestes, insetos ou plantas (CONTRERAS et al., 2003) (GONZALES-ANDUJAR, 2009) (GONZALES-ANDUJAR et al., 2006) (KALOUDIS et al., 2005), e tais sistemas baseiam-se em conhecimentos de um especialista na área, sendo em muitos casos difícil uma representação em algoritmos ou de forma não simbólica.

A medicina também apresenta dificuldades em representar os conhecimentos do domínio, sendo aplicado SE para diversas áreas médicas. Como exemplo disso: auxílio ao médico sobre o encaminhamento do paciente para um especialista por apresentar sintomas de depressão (RODRÍGUEZ-SOLANO et al., 2006); no controle de aplicação de medicamentos para manutenção da pressão arterial (SRINIVAS et al., 2001); ou para auxiliar no diagnóstico de um exame radiológico (PARK et al., 2009) (HUNG; CHEN, 2006)

Essa busca por sistemas inteligentes não ocorre somente na área acadêmica. Na indústria, os sistemas inteligentes estão se tornando um produto comercial que auxiliam no controle de processos (QIAN et al., 2008) ou em análise dos resultados de testes de qualidade de produtos ou materiais (JARMULAK et al., 2001).

### 4.1 ABORDAGEM DE SISTEMA ESPECIALISTA PARA INTERPRETAÇÃO DE EXAMES RADIOLÓGICOS

Segundo Park et al. (2009), o incrível avanço da tecnologia está ajudando no domínio de imagens médicas, onde se faz necessário o auxílio computacional a radiologistas para lidar com uma grande quantidade de informações. Atualmente já existem sistemas de apoio a diagnose denominados de CAD (Computer Aided Diagnosis). Sistemas desse tipo somente operam com Ferramentas de Processamento de Imagens (FPI), fornecendo informações como: morfologia, geometria, topologia e outras características que descrevem anormalidades da imagem, não tratando tais características em relação ao conhecimento especialista do domínio.

Em Park et al. (2009), é apresentado um sistema com abordagem nas FPI integrados com um sistema de diagnoses, baseado em SEs. A primeira etapa do fluxo de diagnose

começa com a extração de características pelos FPI, processando os valores numéricos obtidos por valores simbólicos por meio de um sistema de Lógica Difusa. Os valores simbólicos então são aplicados em um SE que realiza as inferências baseado em regras. Por fim, caso o médico identifique que a interpretação realizada pelo sistema não seja válida, ele poderá acionar um mecanismo de aquisição de conhecimento. O novo conhecimento é adicionado a base de conhecimento sendo verificado e validado.

A base de conhecimento foi representada na forma de regras do tipo SE-ENTÃO, no entanto, foi abordado uma estrutura diferente de relacionamento das regras entre si. Em geral, os SE de CBR utilizam regras compostas de sentenças e conclusões, já nesse trabalho o autor optou por organizar as regras em estrutura de árvore, sendo as regras compostas por sentenças do lado esquerdo da regra e o lado direito composto por conclusões e/ou regras. Se na memória de trabalho houverem todas as condições para satisfazer a sentença de uma regra, que possui outras regras no lado direito da sentença, então essas regras são inseridas como regras a serem inferidas pelo sistema. O autor comenta que essa organização auxilia na verificação e validação dos conhecimentos da base.

O protótipo foi aplicado em 34 casos radiológicos da região torácica, onde foram extraídas 45 características pelo componente de processamento de imagem e com um sistema composto por 29 regras. Sendo os resultados satisfatórios para o autor, pois não foi possível comparar o sistema com outros pelo fato de outros sistemas CADs utilizarem outras metodologias para diagnósticos.

## 4.2 ABORDAGEM DE SISTEMA ESPECIALISTA PARA CLASSIFICAÇÃO DE PLANTAS

Também como a área médica, a taxonomia de plantas é um trabalho complexo e meticuloso. Há diversos caminhos para a identificação de plantas, umas das mais usadas são as chaves dicotômicas, porém necessitam de um grande conhecimento de botânica e organografia. Com isso, em Contreras et al. (2003) o autor propõem o desenvolvimento de um SE denominado *Gymnosperms Remote Expert Executec over Network* (GREEN). *Gymnosperms* é o grupo de plantas que o sistema aborda. Segundo o autor, o sistema foi um dos pioneiros no campo da botânica a utilizar técnicas de IA em sua metodologia.

O conhecimento foi representado no formato de regras do tipo SE-ENTÃO, sendo uma forma semelhante as chaves dicotômicas já utilizadas pelos especialistas para a classificação ou reconhecimento de plantas de forma manual. Como principais características para o grupo de plantas utilizadas no trabalho, foram informações como: aspectos gerais de taxo; folha; tronco; ramos; forma de reprodução; fruta; semente e ambiente.

Como resultado, foi desenvolvido um sistema de classificação de fácil manuseio pelos

usuários para o trabalho de classificação. Inicialmente o usuário define que conjunto de características serão utilizadas na classificação, sendo esse conjunto dividido pelo formato da folha e pelo tipo de semente. Com isso, o sistema exibe características por meio de texto e imagens ilustrativas, permitindo que o usuário informe quais delas são possíveis de serem observadas e o valor respectivo a cada uma delas. Também é permitido que o usuário informe valores de confiança da certeza de sua observação. Após a inferência, o sistema retorna um conjunto dos resultados obtidos, juntamente com informações sobre as plantas e uma justificativa descrevendo o raciocínio para as conclusões apresentadas.

#### 4.3 ABORDAGEM DE SISTEMA ESPECIALISTA PARA CLASSIFICAÇÃO DE PESTES, DOENÇAS E ERVAS DANINHAS

Semelhante ao sistema relatado anteriormente, Gonzales-Andujar (2009) também propõem um sistema de classificação, porém na identificação de pestes, doenças ou ervas daninha que prejudicam as plantações de oliva na Espanha.

A base de conhecimento foi representada no formato de regras do tipo SE-ENTÃO, sendo umas das representações mais utilizada segundo o autor. A base de conhecimento foi construída a partir de informações coletadas de bibliografias e por meio de entrevistas realizadas com especialistas da área. Por meio das entrevistas, o autor foi capaz de obter informações sobre o raciocínio de classificação do especialista humano, que não eram encontradas na literatura da área. Durante as entrevistas, também foram obtidas informações sobre as principais características que definem as famílias/gêneros/espécies abordadas pelos sistemas de classificação.

Semelhante a Contreras et al. (2003), esse trabalho também dividiu o conhecimento conforme a classe a ser identificada, como: peste; doença; ou ervas. Após essa seleção da classe, é apresentado um questionário solicitando que o usuário responda com sim ou não, apresentando uma imagem ilustrativa para o auxílio da definição da resposta. Por fim, o sistema apresenta uma conclusão e informações com imagens e descrição da peste classificada.

A validação do sistema foi realizada em 2 passos. Primeiramente foi realizado uma validação e verificação, que consistia em testar todas as possíveis combinações e verificar se havia um possível erro. Já o segundo passo tinha como finalidade validar o sistema com o usuário fim, foram compostos 2 grupos de 20 pessoas, um formado por técnicos e o outro por estudantes de agronomia, que puderam experimentar e classificar um conjunto de pestes.

Após a classificação, ambos os grupos responderam um questionário sobre a qualidade da interface, se era amigável ou complexa, e a relevância do sistema para atividades educacionais. As questões podiam ser respondidas com valores de 1 a 10, sendo 1 insatisfatório e 10 extremamente satisfatório. Todas as questões obtiveram uma

média superior a 8,94, sendo concluído pelo autor que os resultados foram satisfatórios.

#### 4.4 COMENTÁRIOS SOBRE OS TRABALHOS CORRELATOS

Por meio dos trabalhos correlatos é possível ver a abrangência de soluções que os SEs oferecem, passando por várias áreas do conhecimento, indo desde a pesquisa da área acadêmica até na utilização prática da indústria, um exemplo de trabalho com aplicação prática na busca de petróleo é apresentado em (GARCIA et al., 2000). Neste trabalho o autor apresenta uma abordagem de técnicas de IA para auxiliar geólogos na identificação de rochas que são utilizadas para definir a probabilidade de existir um poço de petróleo na região de interesse. Dentre as técnicas abordadas são utilizadas Redes Neurais, Sistemas Baseados em Conhecimento e Agentes.

Park et al. (2009) apresenta ferramentas para extração de informações em imagens radiológicas e uma nova abordagem de organização das regras em forma de árvore. Esse trabalho apresenta uma organização útil do conhecimento, pois as regras poderão ser estruturadas semelhantes as chaves dicotômicas, que são utilizadas atualmente pelos especialistas do domínio de microfósseis, no processo de classificação. Baseando-se nessa estrutura organizacional das regras, acredita-se que o sistema mostre-se mais amigável ao especialista, no momento da definição e construção das regras da base de conhecimento.

Em Park et al. (2009) também é apresentada uma ferramenta de PI, como parte integrante do fluxo de diagnose do sistema proposto. No entanto, as imagens que são utilizadas pelos especialistas do estudo de caso desta dissertação possuem uma grande quantidade de sedimentos, rachaduras e quebras, o que torna inviável o PI sem supervisão. Porém, a área de processamento de imagem pode ser útil como mecanismo de auxílio, em casos onde os valores de atributos possam ser de difícil distinção.

Os trabalhos apresentados em Contreras et al. (2003), Gonzales-Andujar et al. (2006) e Gonzales-Andujar (2009) mostram a utilização de imagens como forma de ilustração para o auxílio da definição de características que descrevam a planta ou peste a ser classificada. Essa abordagem será utilizada no sistema de classificação deste trabalho, pois os valores das características que definem um microfóssil são muito similares, dando espaço para a subjetividade<sup>1</sup> entre especialistas.

Dentre os trabalhos correlatos foi observado que há uma dificuldade no momento de validação dos sistemas propostos, onde os resultados que visam apresentar a eficiência do trabalho são baseados em análise extraídas de voluntários, que utilizam o sistema na prática ou simulação de domínios reais. Os trabalhos expostos também comentam a dificuldade de avaliação dos sistemas por falta de desenvolvimento de outros sistemas que utilizam as mesmas abordagens para uma comparação.

---

<sup>1</sup>Um mesmo objeto para duas ou mais pessoas podem apresentar inúmeros conceitos visuais conforme o seu conhecimento e informações da área de conhecimento em que esse objeto é abordado. Essa diferença de conceitos e valores é tratada como subjetividade neste trabalho.

De um modo geral, os trabalhos apresentados tentam suprir a necessidade de lidar com grandes quantidades de informação, além de poder ser utilizado como ferramenta de transição de conhecimento entre especialista e não especialista, por meio de sua utilização. O sistema apresentado neste trabalho, tem como objetivo construir um SE que contenha as informações do especialista na área de microfósseis e que elas possam ajudar na classificação de espécimes desconhecidas para outros especialistas ou estudantes da área.

## 5 ARQUITETURA DO SISTEMA DE AUXÍLIO À CLASSIFICAÇÃO

Esse capítulo apresenta a arquitetura e o desenvolvimento do sistema de auxílio à classificação proposto por este trabalho. Inicialmente serão apresentadas as funcionalidades oferecidas pelo sistema e a sua arquitetura juntamente com o SE e o PI. Após, será apresentado o SE descrevendo a forma de representação do conhecimento, o sistema de inferência na base de conhecimento. Por fim, será apresentado a integração do sistema a algoritmos de PI para auxiliar ao usuário na escolha dos valores de característica que definem o microfóssil que está sendo classificado.

### 5.1 ARQUITETURA E FUNCIONALIDADES DO SISTEMA DE AUXÍLIO À CLASSIFICAÇÃO

O sistema de auxílio à classificação é uma aplicação que oferece funcionalidades para a realização do trabalho de identificação de microfósseis, onde as principais funcionalidades do sistema são baseadas em técnicas de SE e PI.

As funcionalidades oferecidas pelo sistema são disponibilizadas conforme o tipo de usuário e suas permissões de acesso. Esses níveis de acesso foram implantados a fim de que o sistema mantenha a sua base de conhecimento consistentes com a realidade do domínio, não permitindo que qualquer usuário, exceto um especialista do domínio, tenha permissão para incluir, remover ou alterar informações sobre o conhecimento armazenado no banco de dados. Com isso, foram definidos três tipos de usuário: **Usuário Comum**; **Usuário Especialista**; e **Usuário Administrador**.

O usuário comum é qualquer pessoa que utilize o sistema para buscar sugestões de informações que auxiliem na identificação de um espécime ou até mesmo para estudos com fins de aprendizagem. Tal usuário além de poder realizar o **Login do Sistema**, também tem à disposição as funcionalidades de **Visualizar todo o conhecimento do domínio**; **Sugerir novas regras**; e **Realizar Inferências na base de Regras**.

Já um usuário especialista, além de ter acesso a todas as funcionalidades de um usuário comum, também tem permissão de interagir com as funcionalidades de: **Manter<sup>1</sup> informações de PI**; **Validar regras sugeridas**; **Manter os casos do SE**; **Manter as regras do SE**; **Manter o conhecimento sobre os microfósseis**; **Manter atributos e valores**; e **Manter imagens e documentos de suporte**.

Por fim, o usuário administrador, é uma generalização do usuário comum, porém com as funcionalidades adicionais de **Manter usuários do sistema de auxílio à**

---

<sup>1</sup>Entende-se por manter, a possibilidade de incluir/selecionar/remover/atualizar informações do banco de dados

classificação; Gerenciar o banco de dados; Manter domínios de trabalho; e Manter a relação de usuários com seus respectivos domínios de trabalho. As funcionalidades mencionadas são ilustradas na Figura 5.1 juntamente com a relações dos tipos de usuários.

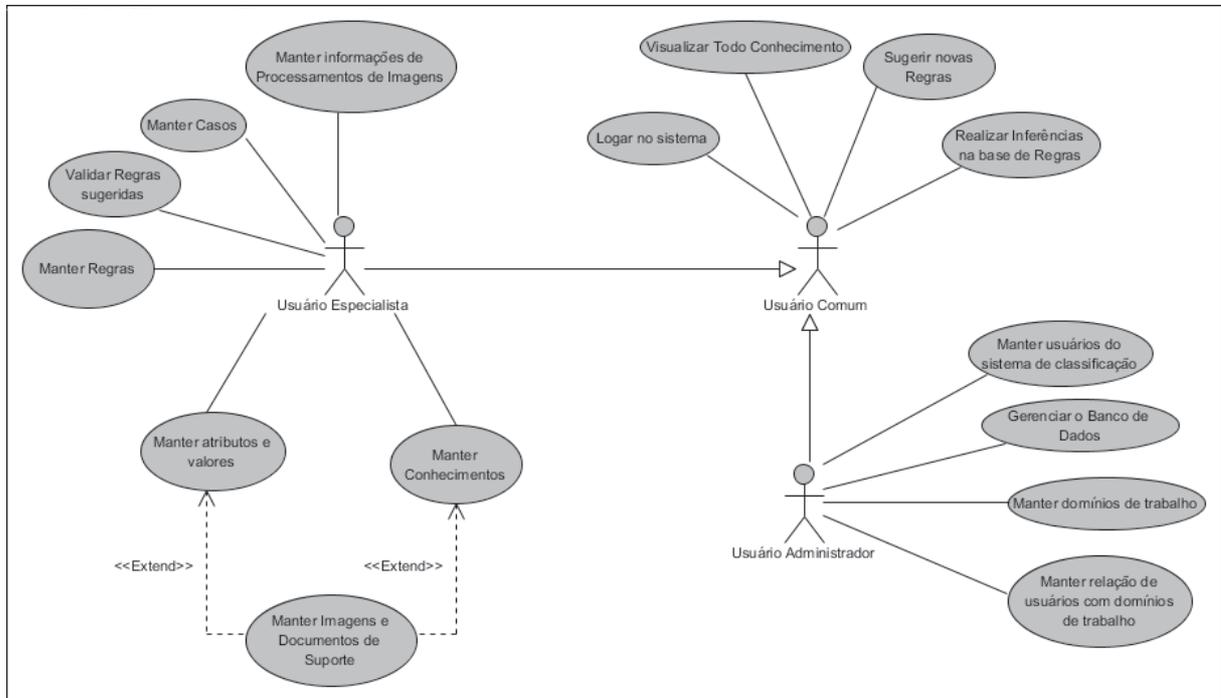


Figura 5.1: Diagrama de casos de uso do sistema de proposto

As principais funcionalidades apresentadas na Figura 5.1, que visam oferecer a sugestão de conhecimentos de microfósseis para auxiliar na tarefa de classificação dos mesmos são detalhadas da seguinte forma:

- **Manter imagens e documentos de suporte:** fornece a possibilidade do usuário a incluir e manter os documentos, como artigos e materiais bibliográficos, e imagens que representem o conhecimento, sendo acessível por meio das funcionalidades de manter atributos e valores, regras e conhecimentos. Esses arquivos podem ser visualizados a qualquer momento para consulta de conhecimento ou no momento da classificação, tanto para o auxílio da escolha de atributos e valores, como também para análise dos resultados apresentados pelos SEs.
- **Manter atributos e valores:** permite que o usuário gerencie os possíveis atributos que caracterizam o objeto de estudo e seus respectivos valores. Tais atributos são associados às regras e posteriormente são apresentados para seleção no momento da identificação de um objeto de estudo, juntamente com a imagem definida para os valores, como forma de representação dos mesmos. Assim, reduzindo as possíveis dúvidas do que tal atributo e valor significam de forma a minimizar a subjetividade para a definição deles durante a seleção.

- **Manter conhecimentos:** oferece ao especialista a possibilidade de construção da base de conhecimento com informações sobre os objetos de estudo. O conhecimento é a representação de um objeto de estudo dentro do sistema, o qual também é disponível ao SE, sendo utilizado nas regras. Os conhecimentos também são associados a imagens e documentos de suporte, os quais podem ser analisados durante o processo de classificação por meio dos resultados apresentados, auxiliando o especialista na validação do conhecimento apresentado pelo sistema.
- **Manter Regras:** permite que o usuário crie e mantenha a base de regras, a qual é inferida pelo SE durante o processo de recuperação de conhecimentos a serem sugeridos. Uma regra pode ser constituída pelos valores de atributos, conhecimentos e outras regras já incluídas na base de conhecimento.
- **Manter Casos:** permite que especialistas forneçam situações em que uma classificação tenha sido bem sucedida ou não, realizar a adição de um caso de classificação que posteriormente será utilizado por demais usuários, como forma de resolução de possíveis conflitos. Apresentando uma possibilidade do caso que esteja sendo sugerido possuir erros em sua regra.
- **Sugerir novas Regras e Validar novas regras:** são funcionalidades que dão a liberdade a um usuário comum de sugerir uma nova regra para aprimorar a base de conhecimento ou uma regra já existente. Porém, também torna-se necessário a confirmação dessas modificações por um especialista do domínio, para manter a consistência das informações com o mundo real. Somente após a validação da regra pelo especialista, ela poderá ser inferida pelo SE.
- **Manter informações de PI:** oferece a possibilidade de persistência de informações extraídas de PI, para futuras comparações com outras imagens. Por meio dessa funcionalidade, o especialista pode inserir as informações da imagem de determinados objetos de estudo.
- **Realizar Inferências na base de Regras:** permite ao usuário realizar a identificação de um determinado objeto de estudo por meio da sugestão de conhecimentos especialistas apresentados pelo sistema. Por meio dessa funcionalidade é permitido a inclusão de valores na memória de trabalho, que posteriormente serão utilizadas para inferir na base de conhecimento pelos SEs de CBR. Também é oferecido ao usuário a possibilidade de realizar um PI básico sobre as imagens do espécime, oferecendo auxílio na decisão dos valores de um determinado atributo.

Algumas dessas funcionalidades possibilitam a interação de diversos usuários, como meio de colaboração entre eles, visando a manutenção e melhoria das informações da

base de conhecimento. Por meio da colaboração, o sistema proposto pode ser aprimorado com informações de diversos especialistas de inúmeros domínios do conhecimento, com o objetivo de melhorar os seus resultados no auxílio à classificação.

Assim, torna-se necessário a centralização das informações em um local. Um exemplo pode ser um servidor com acesso a diversos usuários e de diversos locais, por meio de uma rede local ou até mesmo pela internet. Com isso, o sistema de auxílio à classificação foi desenvolvido com base em uma arquitetura cliente/servidor, como apresentado na Figura 5.2.

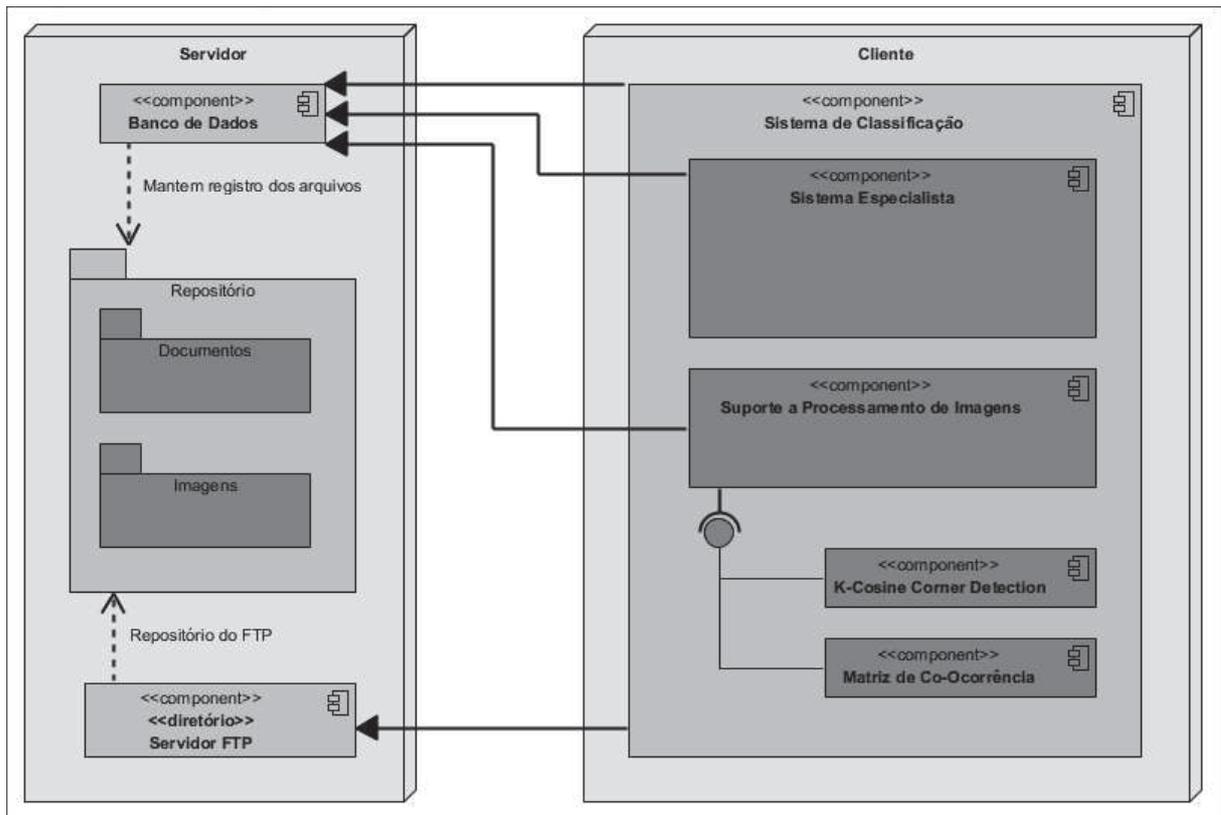


Figura 5.2: Arquitetura cliente/servidor do sistema de auxílio à classificação

O servidor é utilizado como forma de centralização de informações, disponibilizando serviços de transferência de arquivos, para armazenamento de arquivos físicos em disco, e de banco de dados, para persistência da base de conhecimento. A transferência de arquivos é realizada por um servidor FTP<sup>2</sup>, responsável pelo recebimento e envio dos arquivos entre o servidor e o cliente. Os arquivos a serem transferidos, nesse caso as imagens e documentos que representam os atributos, valores e conhecimento, que foram definidos pelos especialistas do domínio. Os arquivos, quando recebidos pelo servidor, são armazenados em um determinado diretório da máquina servidora, denominado de repositório, onde são organizados entre subdiretórios de documentos e imagens. O sistema

<sup>2</sup>(File Transfer Protocol) Protocolo para troca de arquivos entre máquinas de uma rede local ou até mesmo pela internet

cliente também possui uma estrutura igual ao servidor para armazenar os arquivos quando enviados pelo servidor FTP.

O banco de dados, como mencionado, é responsável por armazenar a base de conhecimento. Entre as informações da base estão os atributos, valores, conhecimentos, regras, casos e a referência dos documentos em disco. O banco de dados também persiste as informações referentes ao sistema de auxílio à classificação, para atender algumas regras de negócio relativas ao usuário, quanto ao seu tipo de acesso ao sistema, e aos domínios de trabalho, para controle de acesso às informações para domínios específicos de conhecimento.

A entidade usuário tem como finalidade manter dados para *login*, como usuário e senha e a que tipo de usuário ele está associado, conforme apresentado na Figura 5.1. Já a entidade de domínio de trabalho surge pelo fato do sistema de auxílio à classificação ser genérico, permitindo a construção do conhecimento de diversos áreas de estudo. Com isso, há a necessidade que algumas funcionalidades restrinjam o acesso de usuários apenas ao conhecimento de seus domínios, como forma de garantir a qualidade das informações na base de conhecimento.

O domínio de trabalho está relacionado diretamente com o tipo de objeto a ser trabalhado e a sua região espacial, visto a possibilidade de trabalhar-se com diversos tipos de objetos da mesma área de estudo. Cada usuário pode ter acesso a diversos domínios de trabalhos, conforme for definido pelo usuário administrador. Essas relações podem ser vistas na Figura 5.3, onde é apresentado o relacionamento entre as entidades do banco de dados. As relações e as entidades são melhores descritas na Seção 5.3.1 onde serão apresentadas as formas em que a base de conhecimento foi representada.

A parte cliente da arquitetura do sistema proposto, apresentado anteriormente, é composta por três componentes principais mais uma interface. A interface é constituída por um conjunto de telas que oferecem ao usuário as funcionalidades apresentadas na Figura 5.1. É por meio da interface que o usuário poderá realizar a manutenção de atributos, valores, conhecimentos, documentos, regras entre outros, até efetuar a classificação do objeto a ser estudado. A interface disponibiliza tais funcionalidades por meio de componentes internos do sistema. Na Figura 5.4, todos os componentes do sistema proposto são apresentados, juntamente com a relação dos componentes entre si, com a especificação de quais fazem acesso aos componentes do servidor.

As funcionalidades que lidam com a manutenção de conhecimentos e dos documentos utilizam-se do componente denominado de Gerenciamento do Conhecimentos Especialista. É a partir dele que o usuário do sistema pode incluir informações no banco de dados, desde informações pertinentes ao sistema até informações referentes a base de conhecimento dos SEs. Esse componente também gerencia a troca de arquivos entre o servidor e o cliente por meio do protocolo FTP, como mencionado anteriormente. Já o componente chamado de Sistemas Especialistas na Figura 5.4 é responsável pela inferência na base

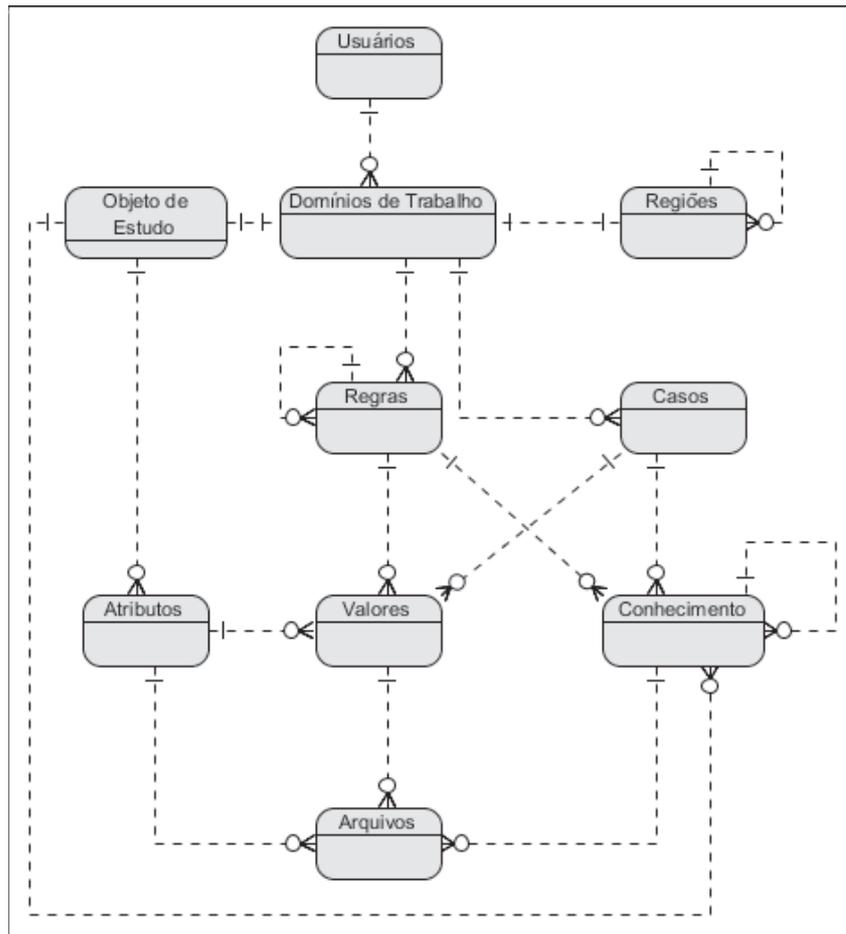


Figura 5.3: Diagrama de relacionamento do banco de dados do sistema de auxílio a classificação

de conhecimento, fornecendo ao usuário, por meio da interface, resultados de sugestão de conhecimentos, conforme as informações fornecidas, baseando-se em um SE de CBR.

Como pode ser visto na Figura 5.4, o componente de SE dependem do banco de dados do servidor para consultar as regras cadastrados, segundo o domínio de trabalho definido ao usuário. Com os resultados definidos, o sistema de interface pode exibir as informações dos conhecimentos obtidos, juntamente com os documentos e imagens, por meio do componente de Gerenciamento do Conhecimento Especialista, recuperando tais dados do servidor.

O componente Interface de Suporte a Imagem é responsável por integrar o sistema de auxílio à classificação com algoritmos de PI. Esse componente disponibiliza uma interface com métodos que os algoritmos de PI tem de implementar. É por meio deles que o sistema de auxílio à classificação fornece a imagem que deve ser processada e recupera as informações a serem exibidas na interface do usuário.

Por meio dessa abordagem é possível implementar inúmeros algoritmos de PI, sem que seja necessário realizar modificações no sistema. Como visto na Figura 5.4, esse componente é implementado com os algoritmos de DCK e MCO. A fundamentação desses

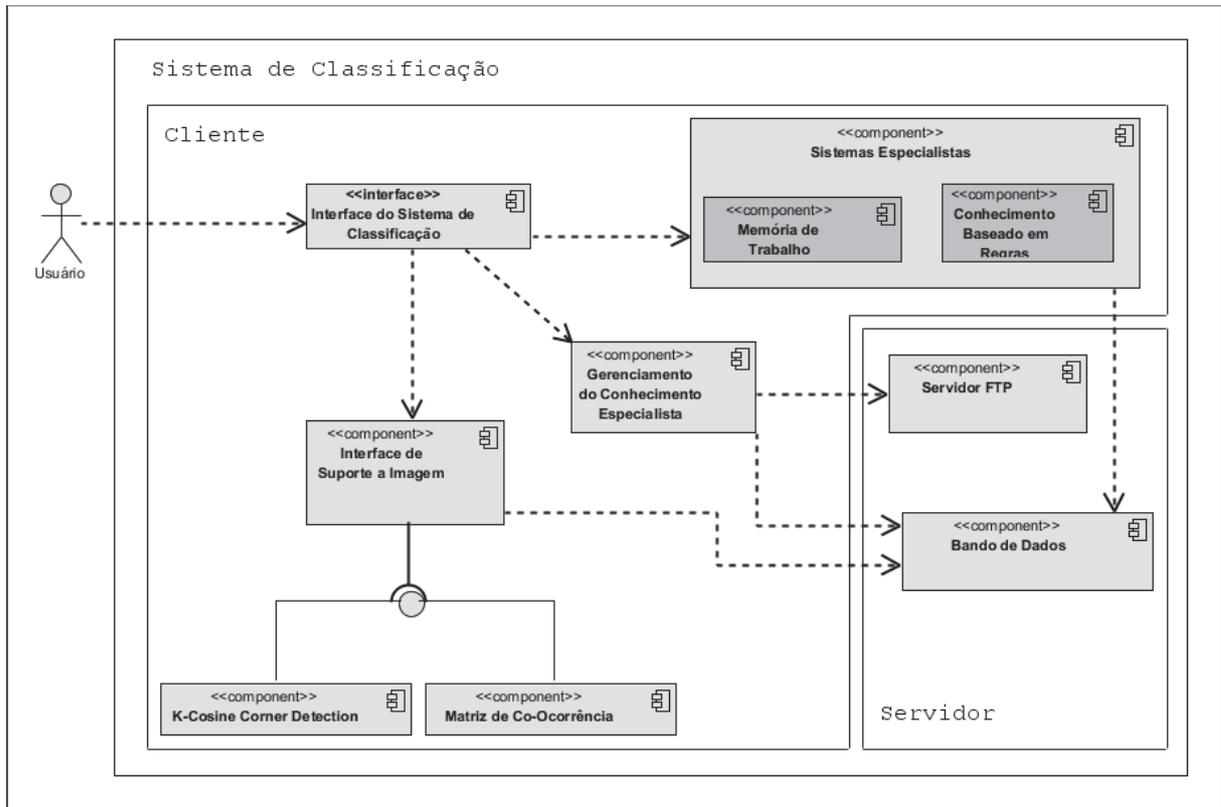


Figura 5.4: Arquitetura da parte cliente do sistema de auxílio à classificação

algoritmos é apresentada na Seção 3.2, que tem como objetivo auxiliar o usuário na decisão de que valores característicos melhor descrevem um espécime a ser identificado.

## 5.2 INTERFACE DE SUPORTE A IMAGENS

Os métodos que devem ser implementados para que um algoritmo de PI seja integrado ao sistema proposto estão separados em três grupos: ISuporteInferencia, ISuporteDataBase, ISuporteFormulario. Não é obrigatório a implementação de todas as interfaces, porém cada interface disponibiliza funcionalidades distintas.

A interface ISuporteDataBase permite que as implementações de algoritmos de PI tenham acesso ao banco de dados, caso seja necessário a persistência de informações específicas dos algoritmos. A interface ISuporteFormulario permite que seja implementado uma tela ou formulário, onde o usuário pode inserir informações sobre algum PI, a serem utilizadas posteriormente durante o processo de inferência na base de conhecimento como forma de auxílio ao usuário.

Já a interface ISuporteInferencia realiza a integração do algoritmo de PI com o ambiente de classificação do sistema. Essa interface tem como objetivo receber uma imagem para realizar um PI, retornando um conjunto de informações pertinentes a imagem conforme o objetivo do algoritmo. É por meio dela que o usuário irá interagir

com o Suporte a Imagem, com o objetivo de tentar reduzir a possível subjetividade em algum atributo que caracterize o espécime da imagem.

O componente de Detecção de Cantos por K-Cossenos somente necessita implementar a *ISuporteInferencia*, pois esse algoritmo visa apenas sugerir o tipo de contorno do espécime de uma imagem, conforme a quantidade de curvas. Esse processo é realizado no momento em que o usuário solicita tal ajuda, não havendo a necessidade de persistir informações ou do usuário fornecer informações prévias.

No entanto, o componente Matriz de Co-Ocorrência utiliza-se das três interfaces, pois necessita persistir valores estatísticos de outros grupos de ornamentação, inseridos pelos especialistas do domínio. Tais valores armazenados no banco de dados são utilizados como forma de classificar uma região de textura informada pelo usuário em relação as outras classes previamente definidas.

A Figura 5.5 apresenta as interfaces disponíveis e como elas estão relacionadas com os componentes que implementam os algoritmos de PI abordados neste trabalho. Ambos os componentes que implementam a interface são melhores detalhados nas seções seguintes.

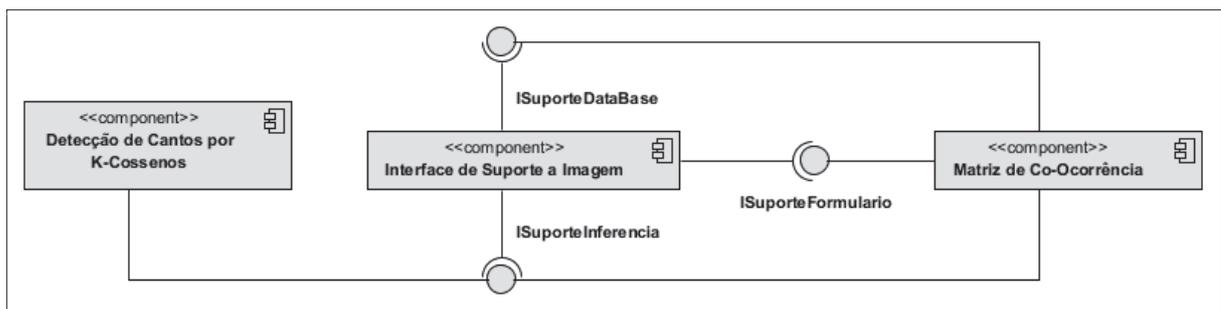


Figura 5.5: Arquitetura do componente de Interface de Suporte a Imagem

### 5.2.1 Detecção de Cantos por K-Cossenos

A aplicação desse algoritmo de detecção de cantos é aplicado sobre os pixels que representam o contorno ( $S$ ) do objeto como apresentado na Seção 3.2.1, necessitando de um conjunto de pré-processos a fim de determinar o contorno do espécime da imagem. Para realizar a detecção do contorno, esse componente utiliza-se do framework *OpenCV*, o qual implementa diversos algoritmos de PI em linguagem de programação.

Outro fator importante que auxilia na detecção do contorno é o padrão das imagens utilizadas pelos especialistas do domínio. Em geral, as imagens já possuem uma edição prévia, que é a remoção de sedimentos da imagem, o que poderia confundir pesquisadores que viessem a colaborar com a pesquisa. Mantendo assim apenas o espécime na imagem, como demonstrado na Figura 5.6.

Dentre todos os métodos de PI disponíveis no *OpenCV* (BRADSKI; KAEHLER, 2008), há o método *cvFindContours*, que tem como objetivo buscar o contorno dos objetos da

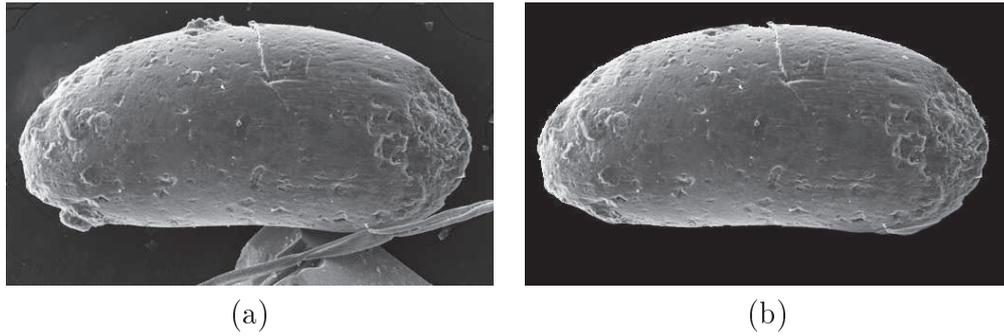


Figura 5.6: (a) Imagem original capturada em um microscópico eletrônico e (b) Imagem com fundo preto

imagem. Esse método retorna todos os possíveis contornos encontrados na imagem, dada a sua variação de cores, entretanto nenhum desses contornos abrange a real segmentação do espécime por questões de grandes variações de tons de cinza, devido a iluminação do espécime da imagem. Para contornar esse problema, o método *cvThreshold* foi associado ao processo de identificação do contorno para realizar a binarização de uma imagem em tons de cinza.

A binarização consiste em transformar uma imagem em tons de cinza, em uma imagem em preto e branco, a partir de um valor de tom de cinza denominado de limiar (*threshold*). Neste caso, todos os pixels que tiverem seu tom de cinza superior ao limiar passam a ser representados pela cor branca (1), caso contrário pela cor preta (0). Como ilustração, dada uma imagem padrão Figura 5.7 (a), do domínio do especialista com 255 tons de cinza, é possível analisar as imagens binarizadas com valores de limiar de tons de cinza de 26 (b), 128 (c) e 192 (d), respectivamente.

Como mencionado anteriormente, por padrão as imagens já possuem o fundo preto, então se utilizado um limiar um pouco superior a cor preta (0) (por exemplo, 26) já será suficiente para obter resultados como da Figura 5.7 (b), onde o espécime foi totalmente binarizado com o valor 1, e o restante da imagem, o fundo foi binarizado com o valor 0.

Assim, não havendo nenhum outro contorno possível de ser identificado pelo método *cvFindContours*, é possível determinar o conjunto de pixels que determinam o contorno  $S$  do espécime facilmente. A Figura 5.8 apresenta uma imagem binarizada (a), com o espécime em branco juntamente com a imagem dos pixels  $P_i$  do contorno (b) obtida pelo método de detecção de contorno do *OpenCV*.

Então, a partir desse contorno é possível aplicar o algoritmo de DCK, o qual resulta em uma sequência de cossenos que variam ao longo de todo o contorno. Essa variação pode ser expressa em um gráfico, onde no eixo das abscissas estão os pixels do contorno e no eixo das ordenadas o respectivo cosseno, dado um suporte  $K$ . A Figura 5.9 apresenta a variação do contorno da Figura 5.8 (b) com um suporte de 310 pixels, o que representa 10% dos pixels do contorno.

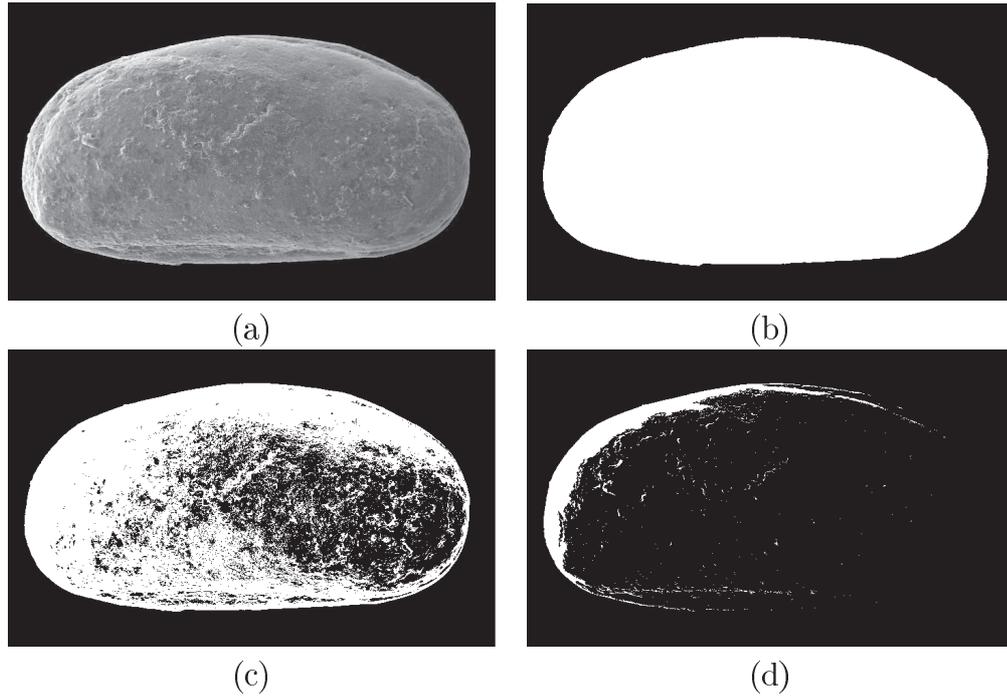


Figura 5.7: (a) Imagem original e imagens binarizadas com limiar igual a: (b) 26, (c) 128 e (d) 192



Figura 5.8: (a) Imagem binarizada e (b) imagem do contorno do espécime

O componente de Detecção de Cantos por K-Cossenos está baseado na variação dos cossenos do contorno obtida da imagem, para sugerir uma possível forma do espécime considerando a quantidade de cantos presentes. O número de cantos em uma imagem são obtidos a partir de um limiar de cossenos dos pixels que representam o contorno. Cada pico que estiver acima do LCC é considerado como canto ou região curva.

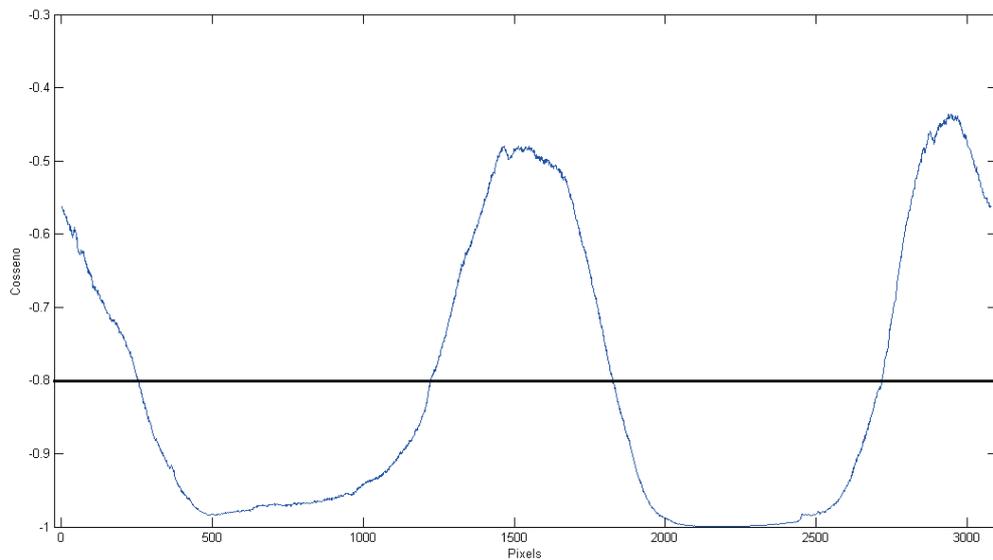


Figura 5.9: Gráfico da variação do cosseno de um contorno

A Figura 5.10 apresenta a imagem de um espécime com suas regiões curvas tem seus pixels destacados em branco. Esses pixels foram destacados com base em um LCC de -0.8, conforme detalhado na Figura 5.9, sugerindo que esse espécime tivesse forma arredonda ou oval.



Figura 5.10: Resultado obtido pelo algoritmo de detecção de cantos por K-Cossenos

A principal vantagem desse componente é a capacidade de sugerir formas que possuam entre duas à quatro regiões curvas. Duas regiões conduzem à formas sub-arredondadas ou sub-ovais; três regiões à formas sub-triangulares; e com quatro regiões à formas quadradas ou sub-trapezoidais. Uma desvantagem é a distinção entre formas que possuam a mesma quantidade de cantos. No entanto, cálculos que utilizem a

distância dos pixels de uma região entre outras regiões ou a distância desses pixels até o centro da área do espécime, podem ajudar a distinguir um contorno arredondado de um oval ou um quadrado de um trapezoidal, suprimindo essa desvantagem.

Outra desvantagem do componente de detecção de cantos, é que ele pode ser apenas utilizado em imagens de espécimes que não estejam quebrados ou que não possuam uma grande quantidade de sedimentos incrustados no seu contorno. No entanto, de um modo geral, os sedimentos são mascarados no momento em que é inserido o fundo preto na imagem.

### 5.2.2 Matriz de Co-Ocorrência

O componente de Matriz de Co-Ocorrência visa classificar uma região de textura da imagem em um conjunto de classes previamente cadastradas pelos especialistas do domínio, por meio de informações estatísticas baseadas em MCO, conforme a seção 3.2.2. As informações estatísticas extraídas são associadas diretamente aos valores de atributos da ornamentação que caracterizam um espécime do estudo de caso. Tal associação é realizada pelos usuários especialistas, selecionando regiões de texturas de uma imagem no sistema de auxílio à classificação.

Esse processo de seleção de uma região, visa apresentar ao sistema a ornamentação de um espécime o mais fielmente possível, pois as desvantagens do método de detecção de cantos, em relação a sedimentos e quebras que ocorrem na superfície do espécime, também prejudicam esse componente, não sendo possível a utilização de toda a região de textura da imagem para o cálculo das MCO. Na Figura 5.11 (a) é apresentada a imagem de um espécime com ótima preservação, já na Figura 5.11 (b) é possível notar que há quebras em grande parte de sua carapaça e na Figura 5.11 (c) o espécime está coberto por sedimentos e sofreu processo de erosão, ocultando feições de sua ornamentação.

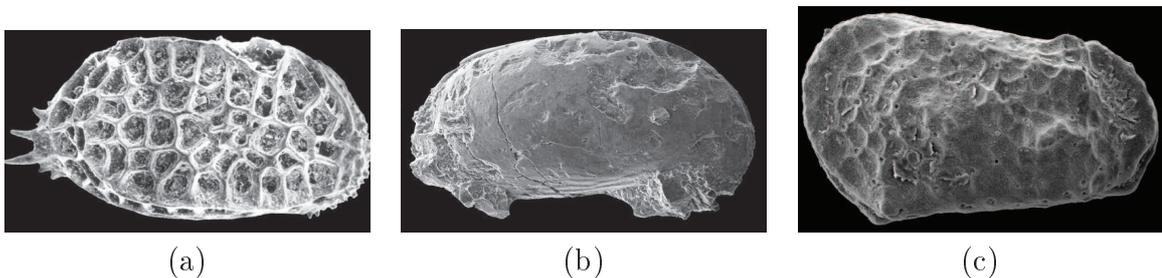


Figura 5.11: (a) Imagem de um espécime limpo e inteiro, (b) imagem de espécime com rachaduras e cobertos por sedimentos e (c) imagem de espécime que sofreram processo de erosão

Para a utilização desse algoritmo, inicialmente o usuário especialista deve informar ao sistema que atributo representa a ornamentação dos espécimes, para que seja possível construir um espaço amostral das texturas e posteriormente indicar a similaridade de uma nova textura. Tal funcionalidade é disponibilizada pela implementação da interface

de formulário (ISuporteFormulario) disponibilizado pela Interface de Suporte à Imagem como apresentado na Figura 5.5. Por meio dessa implementação é realizada a extração das informações das imagens associadas ao atributo selecionado pelo usuário e construído um espaço amostral, que posteriormente será utilizado para indicar que o valor que está mais próximo a nova textura apresentada.

A classificação da ornamentação que necessita ser identificada pelo usuário é feita por meio da implementação da ISuporteInferencia, a qual apresenta uma interface de interação com o usuário, permitindo a inserção da imagem do espécime de estudo no sistema e a seleção de uma região de ornamentação visível. É com base na região selecionada que o componente irá classificar a nova ornamentação em relação às classes inseridas no sistema anteriormente.

Como resultado, são apresentados os valores de possíveis ornamentações que mais se assemelham ao apresentado pelo usuário. O cálculo que define quanto uma ornamentação se assemelha a outra é feita a partir da distância euclidiana da nova ornamentação entre todos os outros exemplares já cadastrados.

### 5.3 SISTEMA ESPECIALISTA

Como mencionado anteriormente, o componente Sistema Especialista é responsável por inferir na base de conhecimento a busca por recomendações de espécimes ao usuário, dado as características informadas. Para isso, esse componente possui um mecanismo principal que implementa conceitos de um SE de CBR.

Tal componente fornece métodos ao componente de interface para que sejam realizadas inferências na base de conhecimento, devolvendo os conhecimentos resultantes. Utilizando para isso uma memória de trabalho, que é composta por representações de atributos, valores e conhecimentos que são representações do domínio do especialista, responsável pelo início do processo de inferência do CBR em busca de sugestões de conhecimentos.

As seções seguintes apresentam de forma detalhada de como foi projetada a representação do conhecimento do domínio e suas relações. Também como é realizada a inferência nesse conhecimento, detalhando a forma de comparação entre os valores e as regras do componente do SE.

#### 5.3.1 Representação do Conhecimento

A representação do conhecimento é uma das tarefas mais complexas e importantes da construção de um SE (BITTENCOURT, 2006) (RICH; KNIGHT, 1993), por ser necessário a extração de informações do especialista. Entre essas informações estão o conhecimento do domínio e o raciocínio utilizado pelo especialista para identificar um objeto de estudo, que em geral estão intrínsecas em sua mente.

Considerando o estudo de caso realizado, a extração de informações foi realizada por meio de entrevistas e estudos da área, juntamente com especialistas do estudo de caso. Como estudos complementares foram assistidas aulas ministradas pelo especialista, onde foi apresentado o conhecimento base para especificação e classificação de organismos, no caso Ostracodes.

#### *5.3.1.1 Conhecimento, Atributos e Valores*

O conhecimento do especialista foi organizado em três grupos principais de conceitos no sistema, denominados de Conhecimento, Atributo e Valor. O conceito de conhecimento é um formalismo do sistema para armazenar as informações sobre um objeto ou espécime o qual o sistema se propõem a trabalhar. Já o conceito de atributo é uma representação de um conjunto de valores característicos que descrevem os conceitos de conhecimento. Como exemplo, dado o conhecimento de um objeto, um atributo desse objeto poderia ser a sua cor, e os valores, a gama de cores que tal objeto pode possuir.

O conceito de conhecimento está organizado em forma de árvore, semelhante a estruturação atual do conhecimento do especialista que é em árvores taxonomicas. Já os atributos e valores foram organizados de maneira que possam haver inúmeros tipos de atributos no sistema, e esses atributos serem constituídos por diversos valores.

Esses conceitos podem ser constituídos por informações como: uma descrição, imagens, e documentos de suporte; para uma melhor representação do conceito que está sendo armazenado. A descrição tem como objetivo descrever o conhecimento de forma resumida, podendo ter referências ou trechos da literatura que melhor descrevem-no, sendo as imagens utilizadas como forma de ilustração dos conceitos e os documentos de suporte como fonte de informação mais abrangente.

#### *5.3.1.2 Regras*

As Regras são estruturas de informações que agregam os conceitos de conhecimento e valor para o mecanismo de raciocínio abordado neste trabalho. As regras são representadas em sentenças do tipo SE-ENTÃO como apresentado na Seção 5.3, onde as condições podem ser constituídas de valores e/ou conhecimentos e as conclusões de conhecimentos e/ou de outras regras, como apresentado em Park et al. (2009).

As condições do sistema de auxílio a classificação foram projetadas para utilizar a lógica booleana E, OU e/ou NÃO. Porém, as condições foram estruturadas de forma haver um único encadeamento principal unido pela lógica E, onde cada condição desse encadeamento pode estar ligado a um encadeamento secundário de condições unidos pela lógica OU. Com isso, se alguma condição do encadeamento principal for falsa, toda a regra será inválida; e se alguma condição do encadeamento secundário for verdadeira, todo o encadeamento é considerado como verdadeiro, tornando a condição do encadeamento principal também verdadeira. O uso dessa padronização das regras é devido ao modo como

é realizada a inferência, o que facilita a validação da regra caso o usuário não informe todos os valores das condições da regra.

A NEGAÇÃO apenas inverte o valor lógico de uma condição, caso uma condição seja falsa, se negada ela será verdadeira. A Figura 5.12 ilustra uma regra com condições utilizando lógicas E, OU e NÃO como descrito.

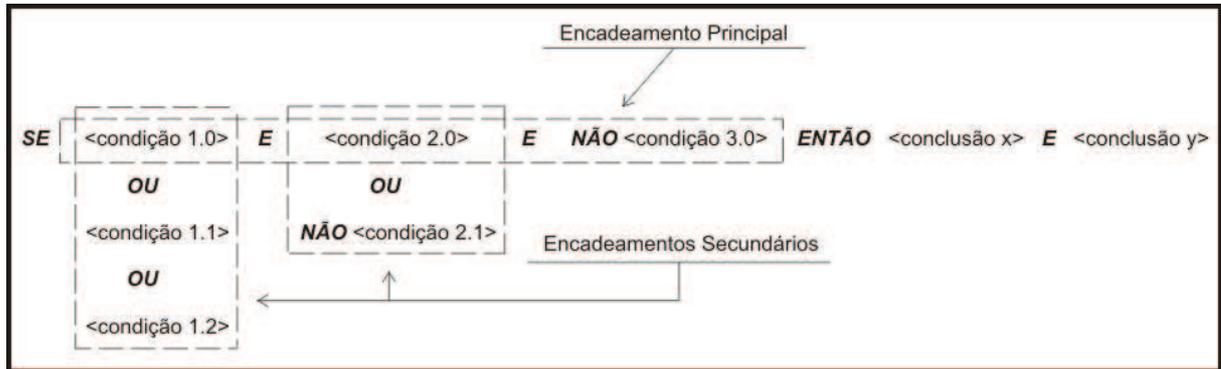


Figura 5.12: Exemplo de uma regra do sistema de auxílio a classificação utilizando lógica booleana

### 5.3.1.3 Domínio de Trabalho

O conceito de Domínio de Trabalho é utilizado como forma de gerenciamento e organização das informações do sistema. Essa organização é fundamental visto a quantidade de espécimes que cada tipo de microfóssil pode possuir, onde um crescimento da quantidade de regras pode ocasionar um aumento na quantidade de possíveis resultados a serem apresentadas pelo sistema, prejudicando a qualidade das informações para a identificação do objeto de estudo.

Nesse estudo de caso os elementos que caracterizam o conceito de domínio de trabalho foram definidos como "Região" e "Objeto de Estudo", onde a região é uma forma de organizar espacialmente um conjunto de informações e o objeto de estudo é uma distinção dos conhecimentos segundo o tipo de objeto que está sendo tratado. Com isso, um conjunto de regras é associado a um único domínio de trabalho. Assim, o usuário ao acessar o sistema deve selecionar o domínio de trabalho em que pretende inferir suas observações, sendo utilizadas apenas as regras desse domínio, não apresentando outros conhecimentos que poderiam prejudicar o raciocínio do usuário no processo de identificação.

### 5.3.2 Motores de Inferência

Com base nos conceitos de representação do conhecimento especialista apresentados é que o sistema realiza suas inferências em busca de sugestões de conhecimentos. Como apresentado na seção 3.1, um SE é composto por uma memória de trabalho, uma base de

conhecimento e um motor de inferência. Nesse caso, a memória de trabalho do componente de Sistema Especialista, seção 5.3, persiste conceitos do tipo valor, conhecimento e regra durante o processo de inferência.

Esses conceitos mantidos na memória de trabalhos são utilizados pelos componentes de CBR dado a sua forma de raciocínio, validando as regras de um domínio de trabalho segundo os valores contidos na memória de trabalho. A Figura 5.13 apresenta a relação da memória de trabalho com os sub-componentes do Sistema Especialista.

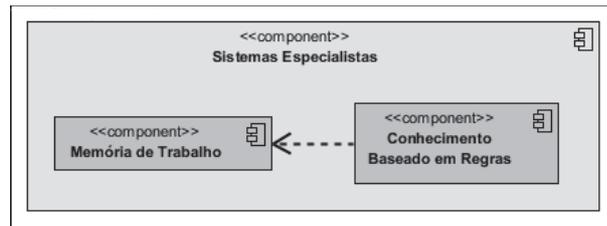


Figura 5.13: Arquitetura do componente Sistemas Especialistas

Com isso, cabe salientar que neste trabalho optou-se por implementar os próprios motores de inferência ao invés de utilizar ou integrar o sistema a uma API ou framework por questões da metodologia de inferência abordada, que busca uma melhor interação com o usuário e reduzindo as possíveis subjetividades. Também optou-se por não utilizar nenhum processo de resolução de conflitos, visto que o sistema poderia sugerir de forma errada algum conhecimento ao usuário, repassando essa responsabilidade, de determinar qual é o melhor resultado entre o conjunto de conhecimentos apresentados, ao usuário.

#### 5.3.2.1 Motor de inferência em regras

Dado os valores que caracterizam um objeto de estudo na memória de trabalho, o motor de inferência em regras busca o conhecimento por meio de regras válidas para tais valores apresentados pelo usuário. Diferentemente de outros SEs que interagem com o usuário por meio de questões, para realizar inferências na base de conhecimento, o sistema proposto neste trabalho utiliza uma abordagem onde o usuário apresenta as informações possíveis de serem analisadas em seu objeto de estudo. Isso permite que o usuário inclua na memória de trabalho somente informações que tenha certeza e que caracterizam o espécime a ser identificado. Essa forma de interação com o usuário dá-se de forma análoga aos trabalhos correlatos de Contreras et al. (2003) e Gonzales-Andujar (2009), apresentados na Seção 4. De um modo geral, o motor de regras está sendo executado em um encadeamento para frente, pelo fato de partir sempre de informações somente conhecidas (atributos e valores) para chegar a uma classe objetivo (conhecimento).

A inferência das regras, consiste primeiramente em verificar se, entre as condições das regras, há ao menos uma condição que seja de algum atributo presente na memória de trabalho. Com isso, são avaliadas apenas as condições desses atributos nas regras da base

de conhecimento. Com a estrutura do conceito de regras, apresentado anteriormente, fica fácil avaliar se uma regra é válida ou não.

Tomando como exemplo, o estudo de caso deste trabalho apresentado na Figura 5.14, o usuário apenas incluiu atributos de contorno e ornamentação, já eliminando a regra 2 do processo de inferência por não possuir nenhuma condição com esses atributos. A regra 1 é válida pela condição negada (ornamentação = estriada) validar todas condições encadeadas por OU (ornamentação = lisa OU NÃO ornamentação = estriada) e pela condição (contorno = sub-oval) também ser válida. Já a regra três apenas possui uma única condição válida pelo seu atributo e valor (contorno = sub-oval), tornando a cadeia de condições válida. No entanto, como não possui nenhuma condição unida por E e que seja inválida, a regra 3 é considerada verdadeira. Como resultado dessa inferência, são apresentados os conhecimentos das regras 1 e 3 para o usuário.

**Memória de Trabalho**

Atributo contorno : Valor sub-oval  
 Atributo ornamentação : Valor reticulado

**Regras**

Regra 1: **SE** ornamentação = lisa **E** contorno = sub-oval **ENTÃO** Gênero Bythocypris  
**OU**  
**NÃO** ornamentação = estriada

Regra 2: **SE** processo\_caudal = bazal **E** **NÃO** costelas = ausente **ENTÃO** Gênero Ambocythere

Regra 3: **SE** processo\_caudal = ausente **E** costelas = longitudinais **ENTÃO** Gênero Majungaella  
**OU**  
 denticulos\_marginais = ântero-ventral **OU** contorno = sub-oval

Figura 5.14: Exemplo de configuração de memória de trabalho e conjunto de regras

Caso uma dessas regras possua outra regra como condição, essa nova regra é avaliada segundo os atributos de condição e valores. Se essa nova regra for verdadeira, ela será apresentada ao usuário ao final do processo de inferência, junto com as demais. O processo terminará quando não houverem mais regras a serem validadas.

## 6 INTERFACES DO SISTEMA DE AUXÍLIO À CLASSIFICAÇÃO

Este capítulo visa apresentar as principais interfaces do sistema implementado, segundo as definições de arquitetura apresentadas. Inicialmente é apresentado o fluxo em que o sistema está organizado, permitindo um melhor entendimento da disposição das funcionalidades que serão apresentadas. Após, são apresentadas as interfaces e suas respectivas funções e características.

O sistema tem seu fluxo com origem na tela de *Login*, que o permite o acesso as demais interfaces. A partir da tela principal do sistema de auxílio a classificação (SisAC) tem-se acesso as demais funcionalidades, que estão divididas em quatro menus principais: Sistema, Conhecimento Especialista, Suporte a Imagem e Configurações. Como é demonstrado na Figura 6.1, com funcionalidades agrupadas distinguidas por cores.

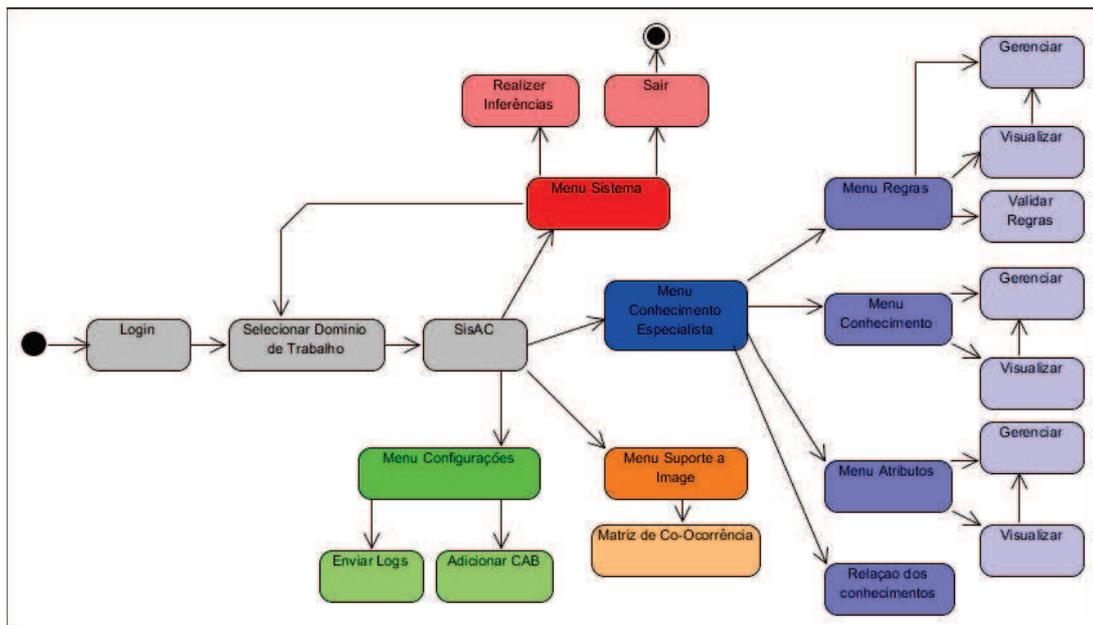


Figura 6.1: Fluxo das interfaces do sistema implementado

No menu Sistemas está agregada a funcionalidade principal que permite a realização de inferências na base de conhecimento. Já o menu Conhecimento Especialista reúne as funcionalidades referentes a manutenção da base de conhecimento, permitindo a inclusão/edição/remoção dos conceitos de regras, conhecimentos e atributos do domínio de trabalho atual. Nesse menu também é permitido que o administrador realize a manutenção do domínio de trabalho, adicionando ou removendo conhecimentos de um domínio. O menu Suporte a Imagem, disponibiliza as funcionalidades que implementam a interface de ISuporteFormulario, como mencionado anteriormente. E por fim, o menu Configurações

agrupa as funcionalidades que permitem configurar o sistema SisAC.

Como apresentado no fluxo de interfaces, o sistema tem seu início na tela de login, que é apresentada na Figura 6.2 (a). No momento em que o usuário realizar seu login com sucesso, ele é direcionado à interface de seleção do domínio de trabalho, que é exibida na Figura 6.2 (b). É a partir dessa tela que o usuário define em que informações ele terá acesso e poderá realizar inferências. Lembrando, que cada usuário terá acesso somente aos domínios de trabalho habilitados por um usuário administrador.

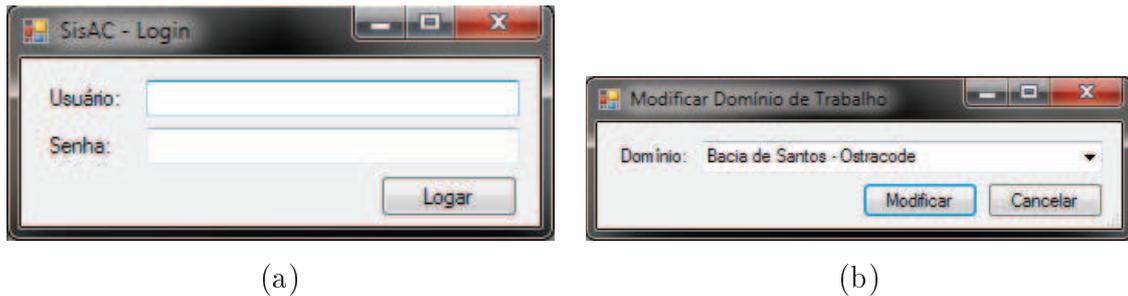


Figura 6.2: (a) Interface de login e (b) Interface para seleção do domínio de trabalho.

Com o domínio definido, o usuário é direcionado para a tela principal, e assim pode acessar todas as funcionalidades que lhe são oferecidas, segundo o seu tipo de usuário. Configuração que foi definida por um usuário administrador. Como detalhe, é exibido no título da interface o domínio atual que o sistema está trabalhando.

As seções seguintes apresentam os menus que são acessíveis do menu principal descrevendo sua funcionalidade e como a mesma pode ser utilizada.

## 6.1 MENU DE CONHECIMENTO ESPECIALISTA

Por meio desse menu é permitido aos usuários visualizarem os conceitos e informações de conhecimentos (total do domínio), regras e atributos. Porém, somente usuários especialistas tem acesso às interfaces de gerência de tais conceitos. Como forma de fluxo padrão, a partir de interfaces de visualização, os usuários especialistas têm acesso às telas de gerência de conceitos permitindo que os mesmos sejam alterados.

Como forma de construção da base de conhecimento, inicialmente é necessário que sejam inclusos os conceitos de conhecimento, atributo e valor iniciais, pois eles são a base para a construção do domínio de trabalho e das regras. Com isso, de um modo geral, as interfaces de gerência/visualização de conhecimentos, atributos e valor tem como padrão os campos:

- Nome: para definir o nome do conceito;
- Descrição: utilizada como forma de descrever o conceito, podendo ser uma descrição breve ou até mesmo um recorte de bibliografia;

- **Imagens:** que representam o conceito que está sendo trabalhado e que posteriormente serão utilizadas como forma de visualização. Tanto para a seleção de atributos que descrevam um espécime a ser classificado ou como forma de análise dos conhecimentos sugeridos;
- **Documento de Suporte:** que são basicamente artigos e/ou referência em forma digital que podem vir a ser utilizados por demais usuários do sistema como forma de organização e análise dos conhecimentos sugeridos.

Com características mais específicas, a interface de gerência do conhecimento, Figura 6.3, possui em seu canto superior direito uma estrutura de árvore onde são apresentados os conhecimentos já inseridos. Essa árvore de conhecimento é utilizada para que o especialista defina o conhecimento "pai" do conhecimento que está sendo inserido ou alterado. Pois como mencionado anteriormente, o conhecimento é organizado em forma de árvore na base de conhecimento.

Na interface de inserção de conhecimentos é possível visualizar as demais funcionalidades de interface, como imagens e arquivos. Sendo que para incluir/remover essas informações é apenas necessário clicar com o botão direito do *mouse* no respectivo espaço. Assim, no momento em que o conhecimento é salvo, essas informações são enviadas ao servidor e disponibilizado para os demais usuários do sistema.

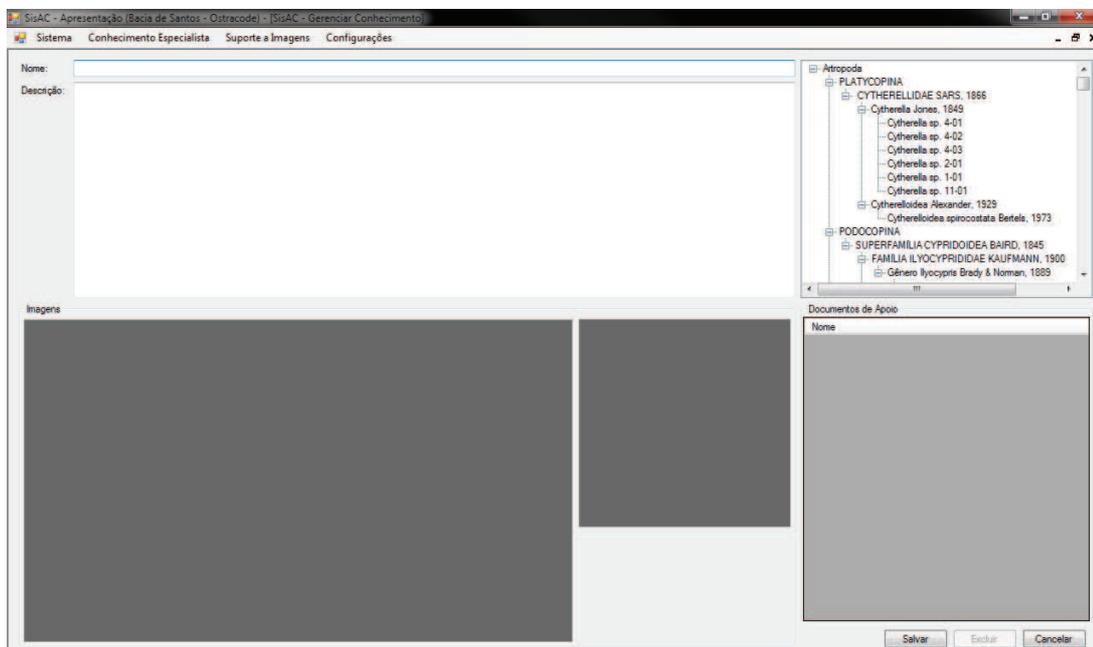


Figura 6.3: Interface de gerência do conhecimento

O conhecimento presente na base de dados para o domínio atual pode ser visualizado pelos usuários por meio da interface de visualização de conhecimento, Figura 6.4. Nessa interface, a esquerda, é apresentado todo o conhecimento, também estruturado em forma

de árvore. Assim, quando um conhecimento é selecionado o mesmo tem suas informações exibidas no restante da tela, apresentando as imagens, arquivos digitais e descrições a ele associado.

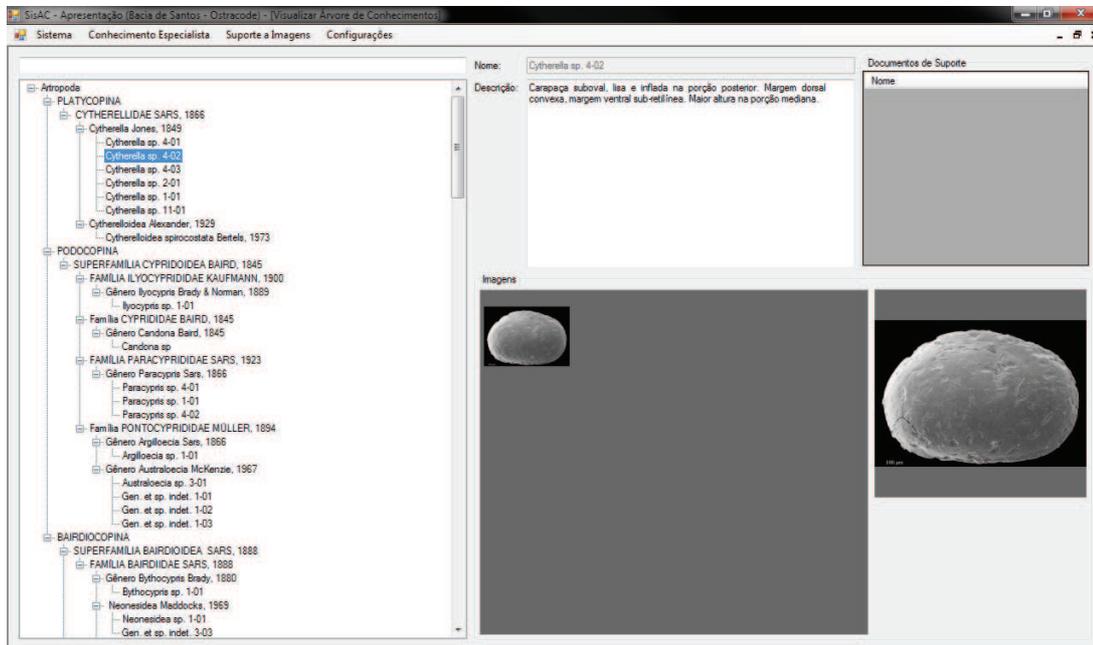


Figura 6.4: Interface de visualização do conhecimento

Quando é necessário alterar algum conhecimento já presente na base de dados, o mesmo pode ser realizado efetuando um duplo clique na informação, da interface de visualização, a ser atualizada. Com isso, um usuário especialista é redirecionado para a tela de gerenciamento de conhecimento permitindo que sejam realizadas as alterações necessárias. Para visualizar uma imagem ou arquivo via uma ferramenta externa é apenas necessário efetuar um duplo clique no respectivo arquivo, sendo aberto a ferramenta responsável pela visualização do arquivo.

A interface de gerenciamento de atributos possui as mesmas funcionalidades de manutenção de descrição, imagem e documentos digitais, mas também possui a manutenção do conjunto de valores que fazem parte de um determinado atributo. Tal funcionalidade está presente no canto superior direito da interface como demonstrado na Figura 6.5.

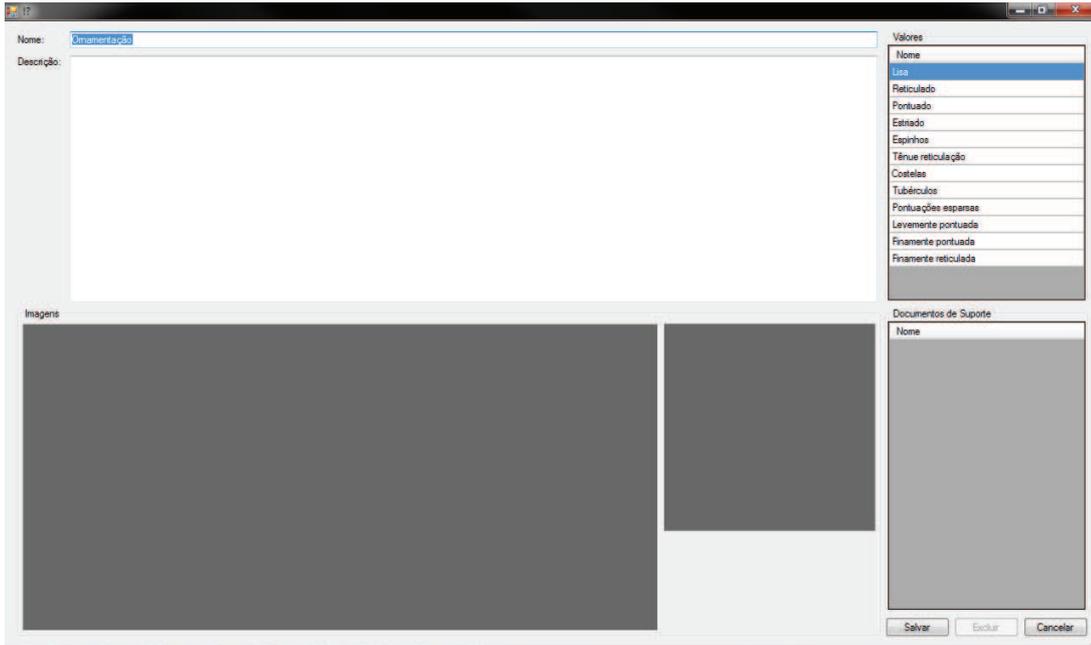


Figura 6.5: Interface de gerência de atributos

Para a manutenção dos conceitos de imagens, documentos digitais e valores associados a esse atributo também é apenas necessário realizar um clique com o botão direito no respectivo campo e informar se será incluído um novo conceito ou será removido o conceito selecionado. No caso da inclusão de um valor o usuário especialista é direcionado para a interface de gerenciamento de valores, apresentado na Figura 6.6. Semelhante às demais interfaces de gerenciamento, essa interface também oferece a possibilidade de associar uma descrição, imagem e/ou documentos digitais.

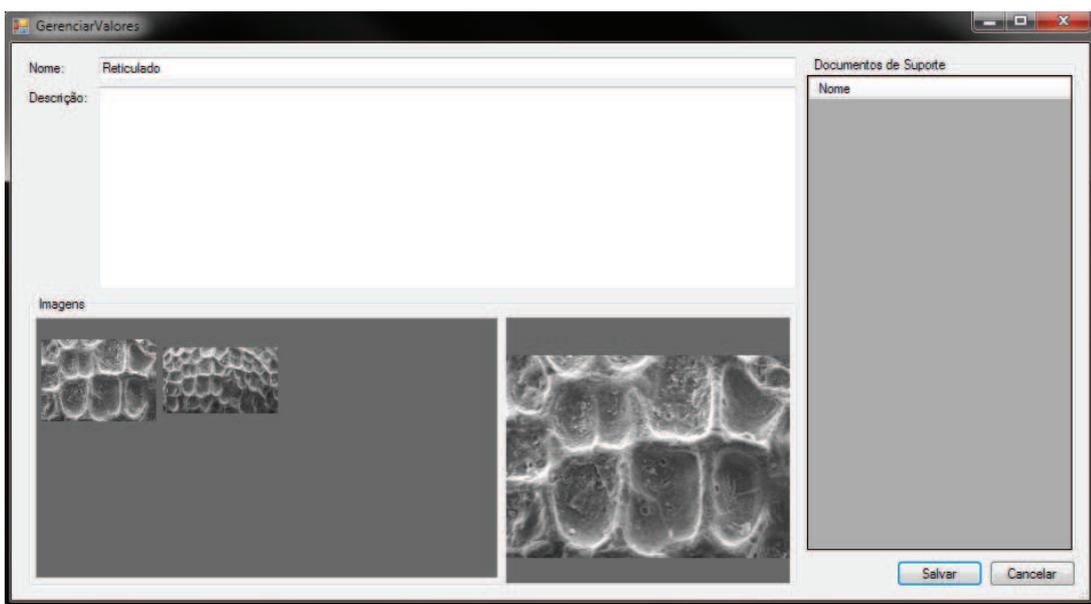


Figura 6.6: Interface de gerenciamento do valor de um atributo

Com a confirmação da manutenção do conceito de valor, o mesmo é temporariamente associado no respectivo atributo a ser trabalhado. Sendo salva a alteração do valor e atributo somente no momento em que o atributo é salvo. Visto que o valor é uma relação direta ao seu respectivo atributo permite que um conjunto de alterações, em um atributo e seus valores, sejam desprezadas caso o usuário cancele as manutenções realizadas.

Semelhante a abordagem de gerenciamento e visualização de conhecimentos, os conceitos de atributo e valor também podem ser visualizados por meio da interface de visualização, como apresentado na Figura 6.7. Essa interface possui as mesmas ações de visualização de imagem e arquivos e edição do conceito com base em duplo clique, semelhante à interface de visualização de conhecimento. Sendo que sempre que um usuário especialista solicitar a alteração de um respectivo atributo ou valor, o mesmo é direcionado para a interface de gerenciamento de atributo permitindo a manutenção dessas informações.

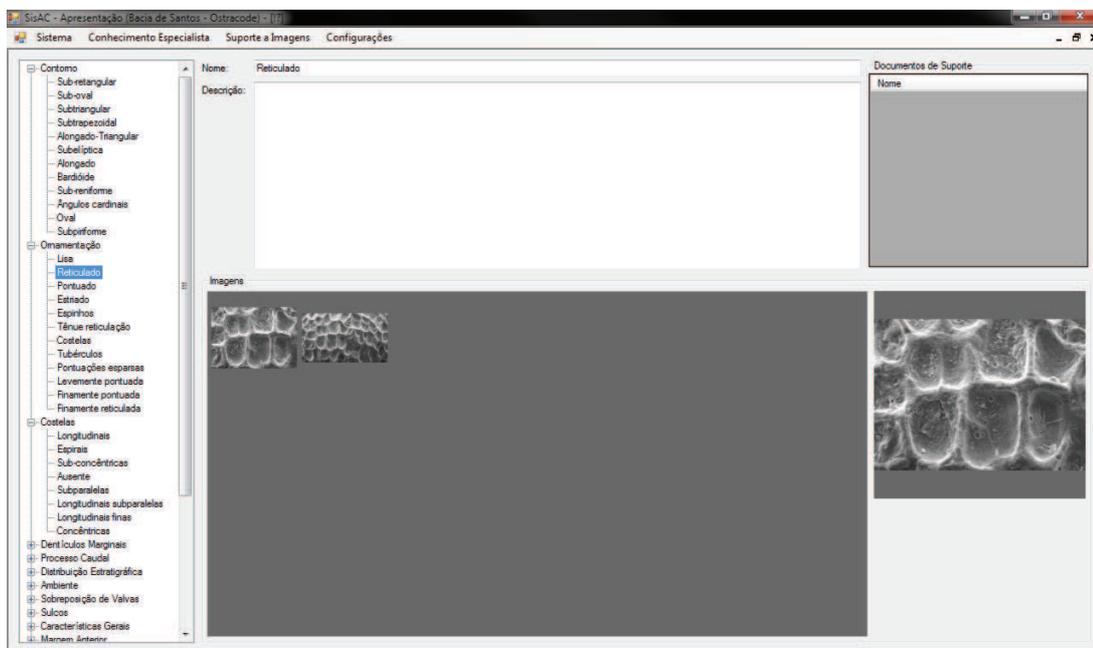


Figura 6.7: Interface de visualização de atributos e valores

Diferente das demais interfaces apresentadas, a interface para gerenciamento de regras não possui os conceitos de descrição, imagem ou documento digital. Nesta interface o usuário especialista deve informar o nome da regra e as condições que constituem a sentença também como a conclusão da regra se a mesma for válida. Essa tela, Figura 6.8, é dividida em 4 quadrantes, onde o quadrante superior esquerdo representa as cláusulas dos encadeamentos principais, que são unidos pela condição E. O quadrante superior direito apresenta as cláusulas dos encadeamentos secundários, do respectivo encadeamento principal selecionado, que são unidos pela condição OU. Já o quadrante inferior esquerdo define a conclusão da regra, informando que conclusões de conhecimentos e/ou regras são

verdadeiros caso a regra seja válida. Por fim, o quadrante inferior direito apresenta a regra em forma de texto.

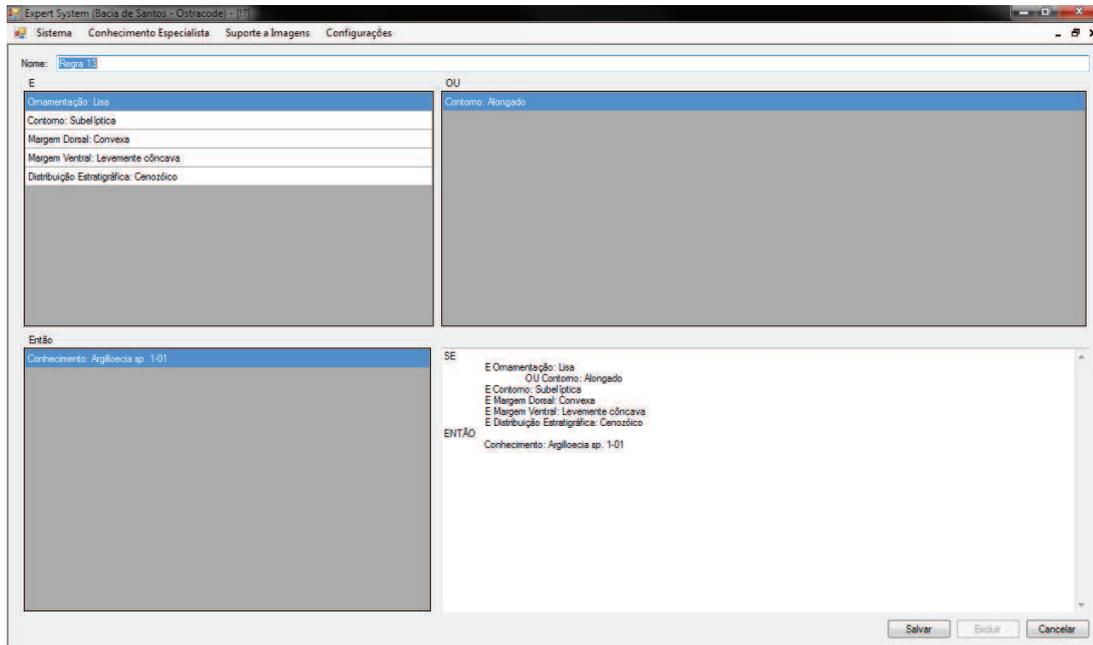


Figura 6.8: Interface de gerenciamento de regras

A inclusão ou remoção das condições ou conclusões de seus respectivos campos é realizada por meio de um clique com o botão direito no respectivo campo, onde quando o usuário for remover será apagado o valor selecionado do campo. Já se o usuário especialistas for incluir um novo valor, o mesmo será direcionado para uma tela onde ele deve informar que valor será inserido. Essa interface é apresentada na Figura 6.9.

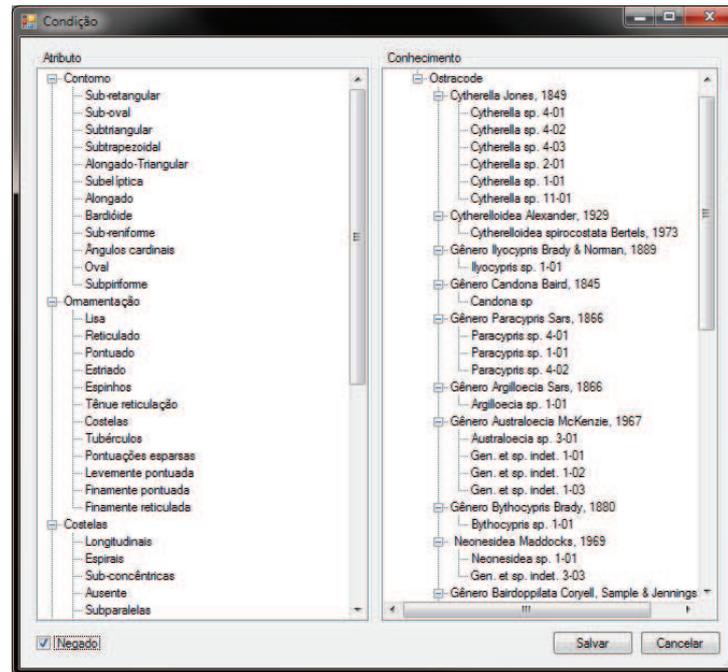


Figura 6.9: Interface para associação de condições a regra gerenciada

Nessa interface, são apresentados todos os atributos do domínio de trabalho e os conhecimentos que foram associados para este domínio. No lado esquerdo são apresentados os atributos e seus respectivos valores, sendo obrigatório o usuário selecionar um valor. No lado direito é apresentado todo o conhecimento do domínio, permitindo que o usuário informe o conhecimento a ser incluído na condição da regra. Além disso o usuário pode informar se a cláusula será negada, se a regra será somente válida caso o atributo e valor ou conhecimento não estejam presentes na memória de trabalho.

Do mesmo modo que os conceitos de conhecimento, atributo e valor, as regras também podem ser visualizadas na interface de visualização, apresentada na Figura 6.10. Diferente das demais telas, nessa interface é somente apresentada a sentença em forma de texto da regra. A manutenção das regras também é realizada por meio de um duplo clique na regra que o usuário especialista quer editar, sendo direcionado para a tela de gerenciamento de regras.

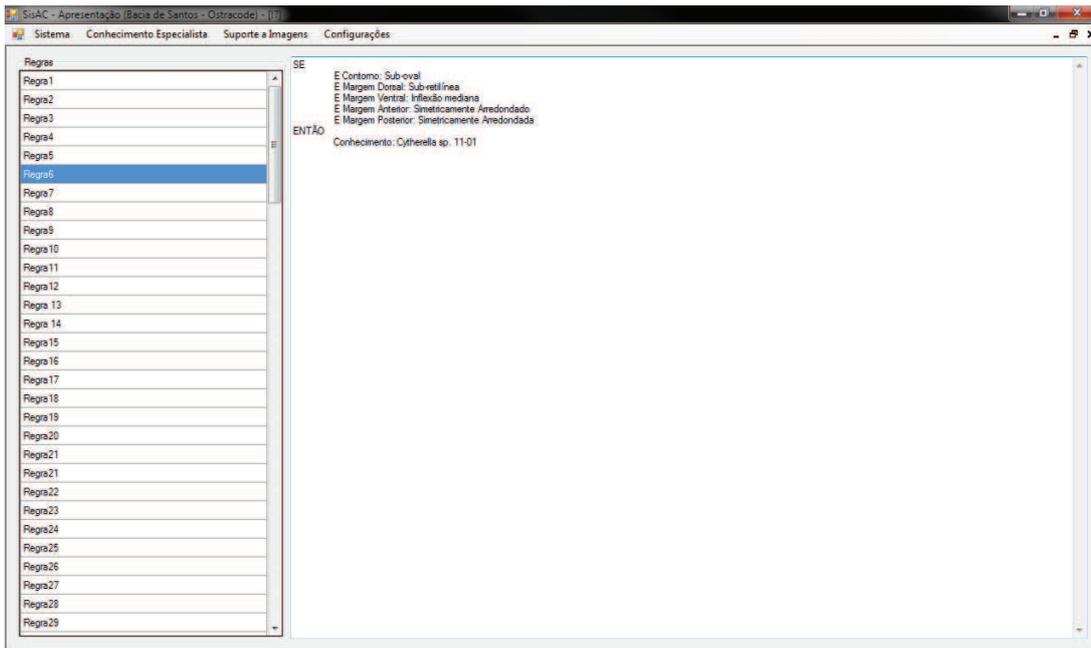


Figura 6.10: Interface de visualização de regras

Dentro do menu de conhecimento especialista também é oferecido aos usuários especialistas a funcionalidade de validação das melhorias e/ou novas regras sugeridas. Por meio dessa funcionalidade são apresentadas todas as melhorias na base de regras sugeridas dentro de um domínio de trabalho, como apresentada na Figura 6.11. Nessa interface são apresentadas as sugestões referentes a novas regras ou melhorias, a data em que foram criadas e o dia em que foram avaliadas juntamente com quem a gerou e quem a avaliou.

Melhoria	Regra	Data Criação	Data Avaliação	Usuário	Especialista
Melhoria	Regra93	28/12/2010 22:50:17		giovani	Especialista

Figura 6.11: Interface de visualização das melhorias sugeridas na base de regras

Para um usuário especialistas poder avaliar uma regra, é somente necessário dar um duplo clique na regra a ser avaliada. Assim o usuário será direcionado para a tela de melhoria ou sugestão de nova regra, onde será apresentada a regra e as respectivas cláusulas e/ou conclusões que foram inclusas ou removidas, juntamente de uma justificativa da mudança, como apresentada na Figura 6.12. Com isso, o usuário

responsável pela validação pode aceitar a melhoria ou nova regra, ou rejeitar a mesma caso não acredite que ela seja válida para o domínio.

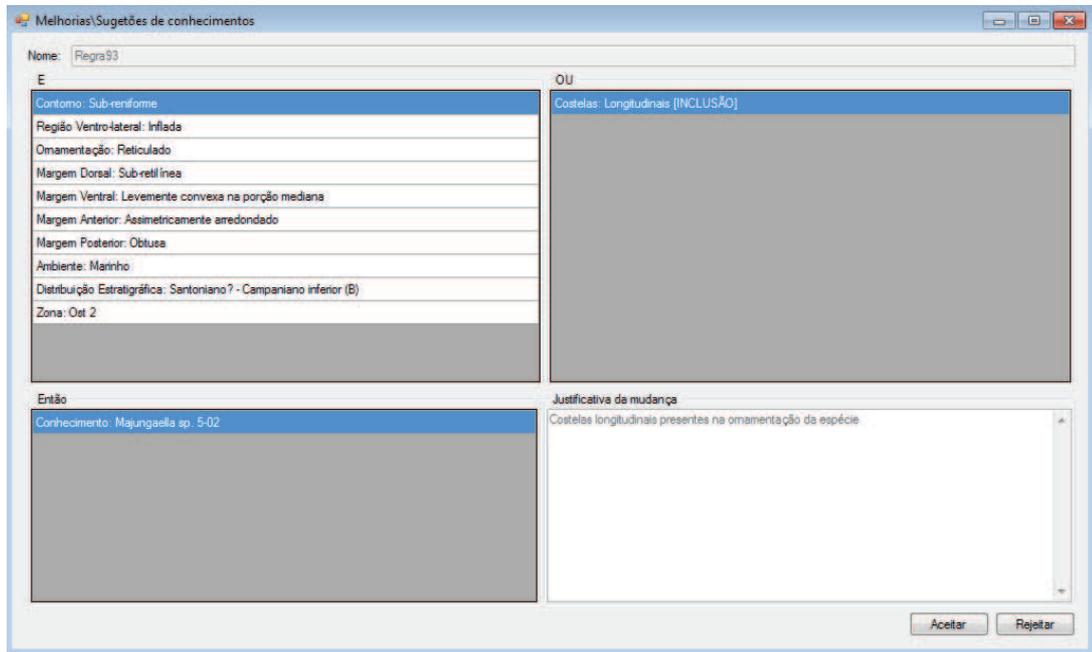


Figura 6.12: Interface que apresenta as melhorias ou nova sugerida

Por fim a última funcionalidade oferecida no menu de conhecimento especialista é a interface de manutenção dos conhecimentos que representam o domínio de trabalho. Por meio da interface, apresentada na Figura 6.13, um usuário administrador pode definir dentre os conhecimentos inseridos na base de conhecimento, quais deles fazem parte do domínio de trabalho. Com isso, é garantido que as regras serão baseadas somente nesses conhecimentos, permitindo uma melhor inferência na base de conhecimento.

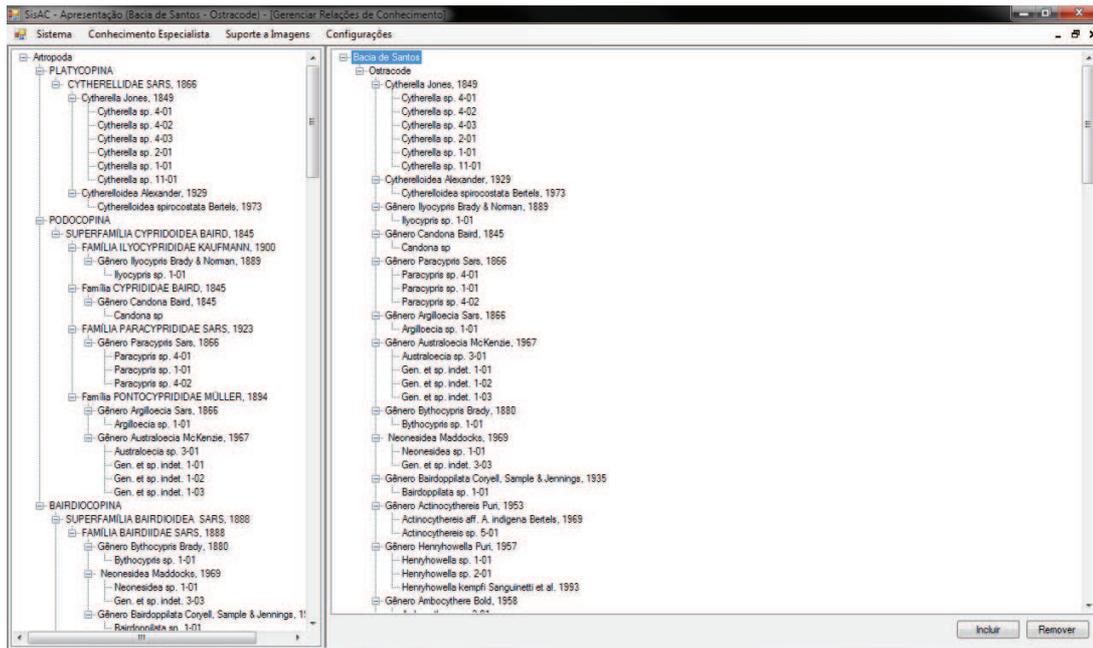


Figura 6.13: Interface para manutenção dos conhecimentos que fazem parte de um respectivo domínio de trabalho

No lado esquerdo da tela são apresentados os conhecimentos como um todo e no lado direito somente o conhecimento do domínio do trabalho. Para definir os conhecimentos do domínio de trabalho, o usuário seleciona o conhecimento do lado esquerdo e onde o mesmo deve ser inserido na árvore do conhecimento do lado direito e clicar em incluir, utilizando os botões localizados na parte inferior direita da tela. Para remover é somente necessário o usuário selecionar o conhecimento do domínio de trabalho e clicar no botão remover.

## 6.2 MENU SUPORTE À IMAGEM

Dentro do menu de suporte à imagem são apresentadas as funcionalidades que implementam a interface ISuporteFormulario do módulo de Suporte à Imagem apresentado anteriormente. Com base no estudo de caso foi apenas necessário utilizar essa interface para o algoritmo de MCO.

Essa interface tem como objetivo fazer com que o usuário especialista defina que atributo representa a ornamentação do espécime, a fim de indicar que imagens serão utilizadas para construir um espaço amostral e posteriormente sugerir qual é a ornamentação que mais se assemelha a nova textura apresentada pelo usuário. Como pode ser visto na interface apresentada na Figura 6.14, no lado esquerdo da tela são apresentados todos os atributos e no lado direito todas as imagens dos valores do atributo selecionado. Após a definição do atributo, então o usuário apenas solicita ao sistema a extração das características das imagens do atributo e construa um espaço amostral.

Com isso, quanto mais imagens são inseridas em cada valor do atributo, melhor será a classificação das texturas utilizadas durante o processo de interferência.

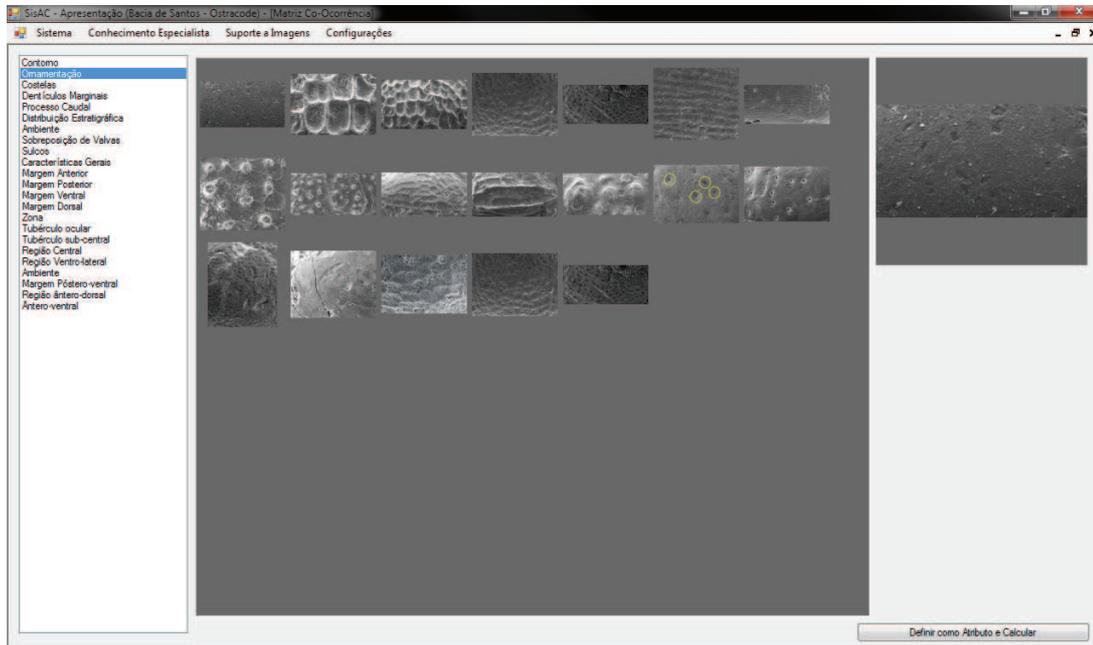


Figura 6.14: Interface de suporte a imagem para o algoritmo de MCO

### 6.3 MENU SISTEMA

Dentro do menu sistema, os usuários podem sair do sistema, modificar o domínio de trabalho e realizar a inferência na base de conhecimento. Caso o usuário opte por alterar o domínio de trabalho, o mesmo é redirecionado para a interface mencionada anteriormente e apresentada na Figura 6.2. Já a opção de realizar inferência permite ao usuário inferir suas observações em busca de conhecimentos para conclusão de seu raciocínio, o objetivo principal deste trabalho.

No momento em que o usuário acessa a funcionalidade de inferência, ele é direcionado para a interface apresentada na Figura 6.15. Nessa tela, inicialmente são disponibilizados todos os atributos que estão presentes pelo menos em uma regra, localizado na identificação 1 da Figura 6.15. Não são apresentados todos os atributos e valores presentes na base de conhecimento, somente os atributos e valores que foram utilizados ao menos em uma regra, a fim de evitar possíveis subjetividades, visto que os valores que não foram utilizados em nenhuma regra não irão oferecer nenhum resultado.

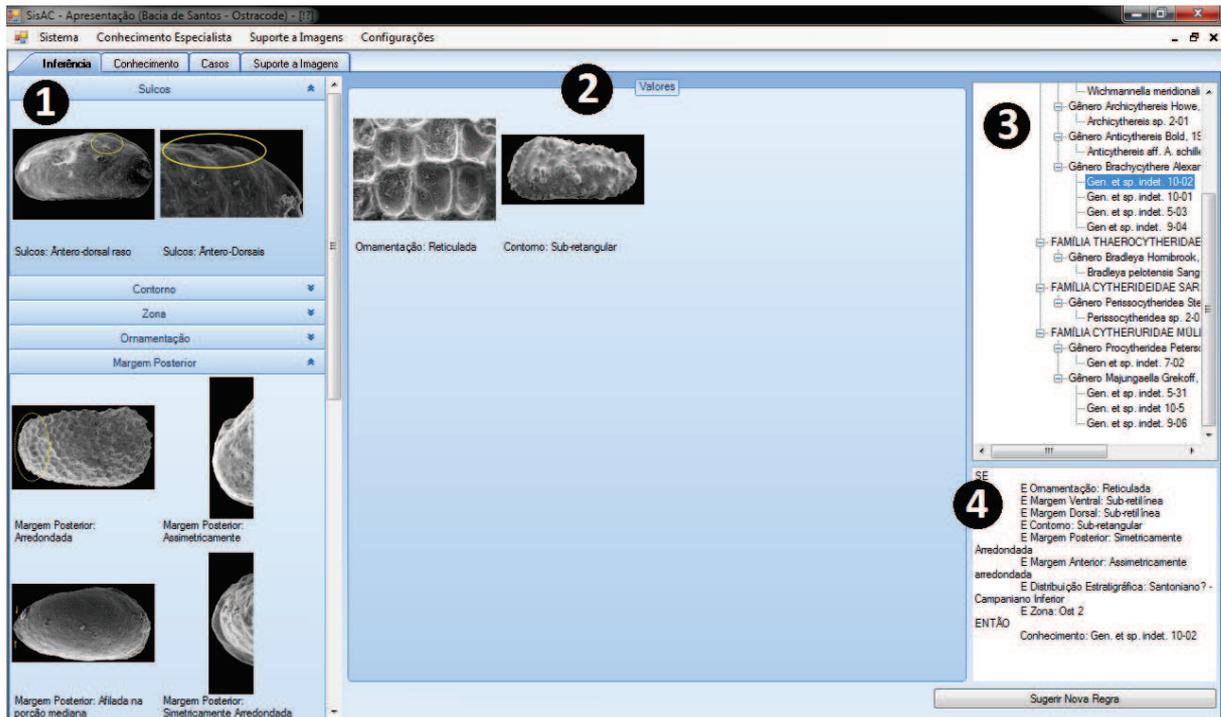


Figura 6.15: Interface da inferência na base de conhecimentos

Como pode ser visto de forma mais ampla na Figura 6.16, os valores disponíveis para o usuário inferir suas observações são agrupados conforme o seu atributo. Como forma de ilustração, caso o valor possua uma ou mais imagens, então a primeira imagem associada a ele será exibida. Por meio das escolhas, um duplo clique sobre a imagem, os valores são inseridos na memória de trabalho, permitindo que estes valores sejam utilizados para a realizar a inferência na base de conhecimento.

Com a inferência concluída os atributos são atualizados, passando apenas a exibir os atributos e valores presentes no conjunto de regras resultantes. Isso permite reduzir gradativamente as regras e os possíveis atributos até chegar a um conjunto pequeno de conhecimentos, que foram sugeridos pelas regras recuperadas no processo.



Figura 6.16: Atributos e valores possíveis de serem utilizados para inferir na base de conhecimento

Como mencionado, os valores selecionados são incluídos na memória de trabalho, e como forma de exibição, esses valores são apresentados no centro da interface de inferência como apresentado na região 2 da Figura 6.15. No exemplo apresentado, o usuário informou ao sistema que seu objeto de estudo possui (Atributo) ornamentação do tipo (Valor) reticulado e (Atributo) contorno do tipo (Valor) sub-retangular.

Para remover algum valor que esteja presente na memória de trabalho, o usuário necessita apenas dar um duplo clique no valor que deseja remover. Com isso será realizado uma nova inferência atualizando os conhecimentos sugeridos e os atributos e valores que podem ser utilizados.

A cada inferência realizada, os conhecimentos resultantes das regras são

apresentados aos usuários, na região 3 da Figura 6.15, organizados em forma de árvore. Onde os nodos folha são os conhecimentos sugeridos. Entretanto é permitido que o usuário tenha acesso ao conhecimento pai para uma melhor análise, verificando se o mesmo é válido ou não para o seu raciocínio. Já na região 4 da Figura 6.15 é apresentada a regra que levou o conhecimento selecionado a ser sugerido, mostrando a sentença que torna a regra verdadeira dado um conjunto de observações apresentadas pelo usuário.

Caso o sistema não informe nenhum conhecimento dado o conjunto de valores apresentado, é oferecido ao usuário a possibilidade de sugerir uma nova regra. Semelhante a interface de validação de regras. O usuário deve então informar o nome e as cláusulas da nova regra, por meio de interações iguais a interface de gerenciamento regras. Porém, no canto inferior direito, o usuário deve apresentar uma justificativa da nova regra. Como apresentado na Figura 6.17.

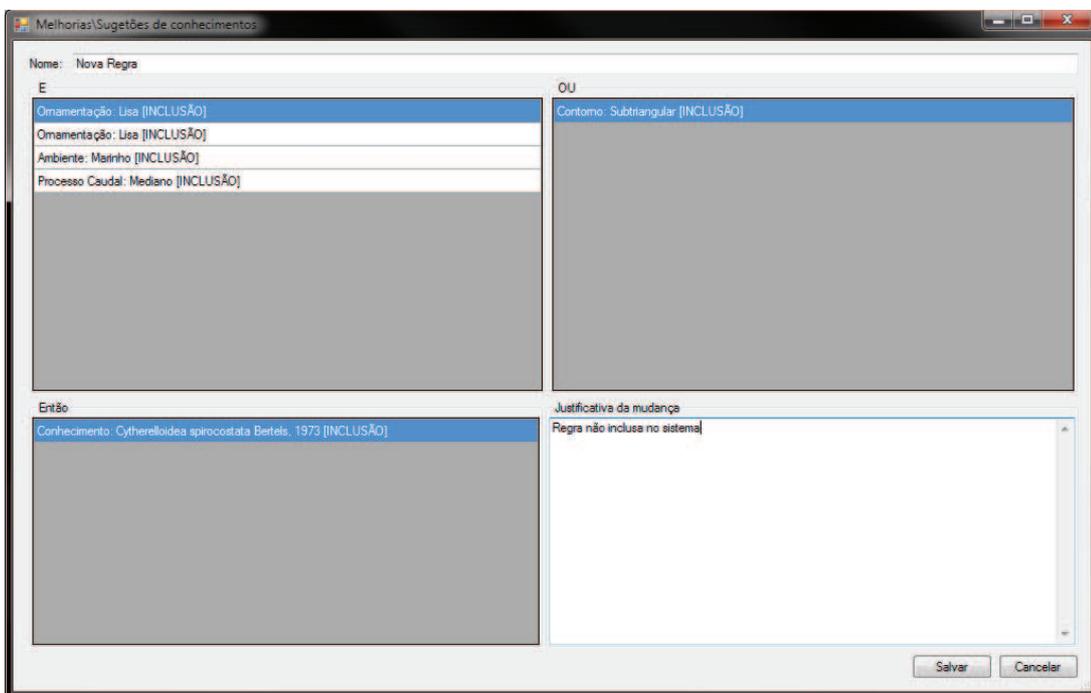


Figura 6.17: Interface para sugestão de uma nova regra

Já se o sistema sugeriu algum conhecimento, o usuário então tem acesso a todas as informações desse conhecimento por meio de um duplo clique sobre ele. Com isso, as informações como descrição, imagem e documentos digitais do conhecimento são carregados e o usuário é redirecionado para a aba de Conhecimento, onde as informações são apresentadas. Como apresentado na Figura 6.18.

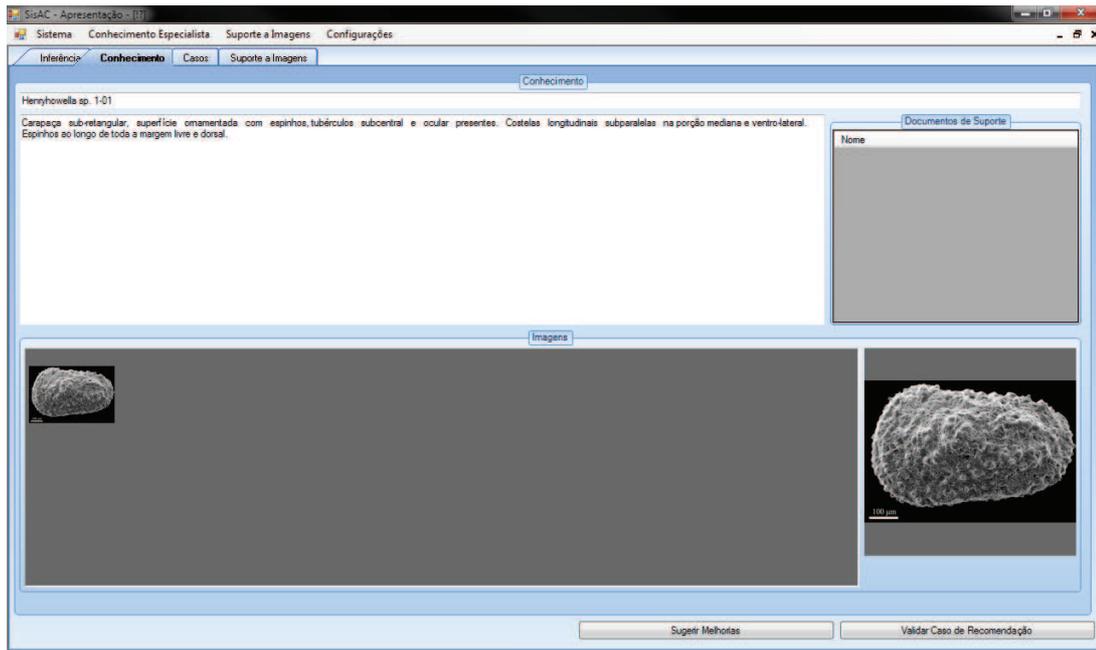


Figura 6.18: Interface para exibição do conhecimento sugerido pelo sistema

Por meio dessa interface, o usuário tem a opção de incorporar ao sistemas novas informações, como sugerir melhorias na regra que resultou no conhecimento visualizado, ou validá-lo. Tais funcionalidades são oferecidas por meio de dois botão na parte inferior direita da tela.

A melhoria de uma regra utiliza a mesma interface para a sugestão de uma nova regra, porém nesse caso todas as informações da regra já são apresentadas ao usuário. Permitindo que ele remova ou adicione novas cláusulas de condições ou conclusões. Como apresentado na Figura 6.19

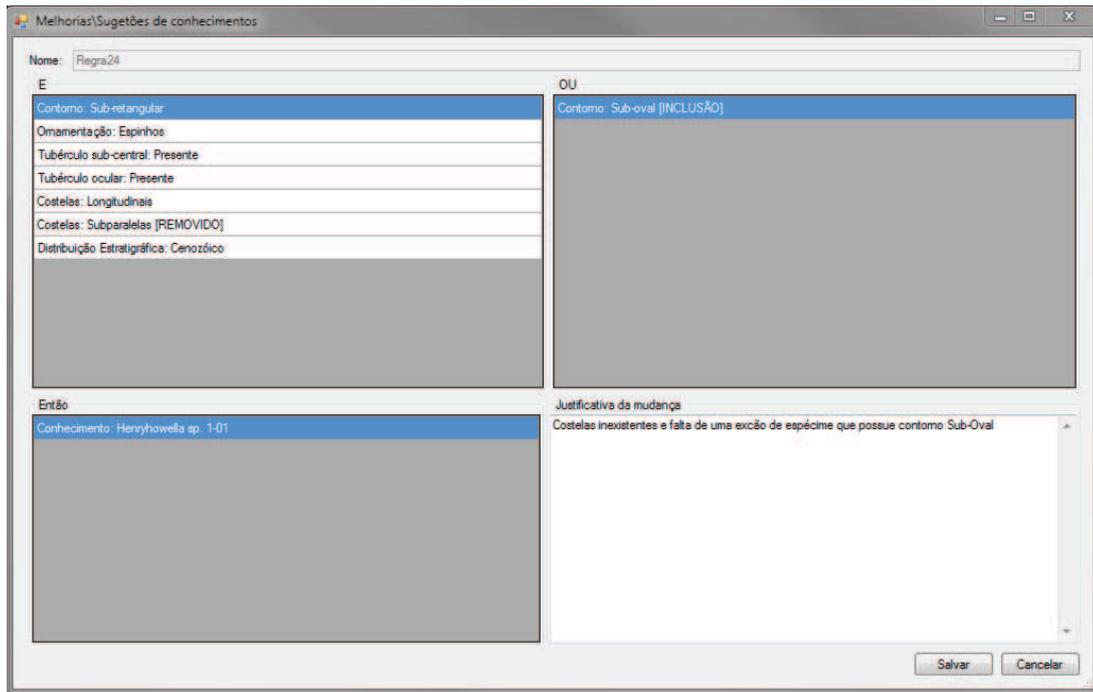


Figura 6.19: Interface para sugestão de melhoria em uma regra já existente

Já para a validação de um conhecimento sugerido, o usuário apenas necessita apresentar a sua opinião sobre a sugestão e a justificativa para tal, como apresentado na Figura 6.20. Permitindo que posteriormente outros usuários possam avaliar as opiniões expressadas sobre o conhecimento, dado o conjunto de valores selecionados.

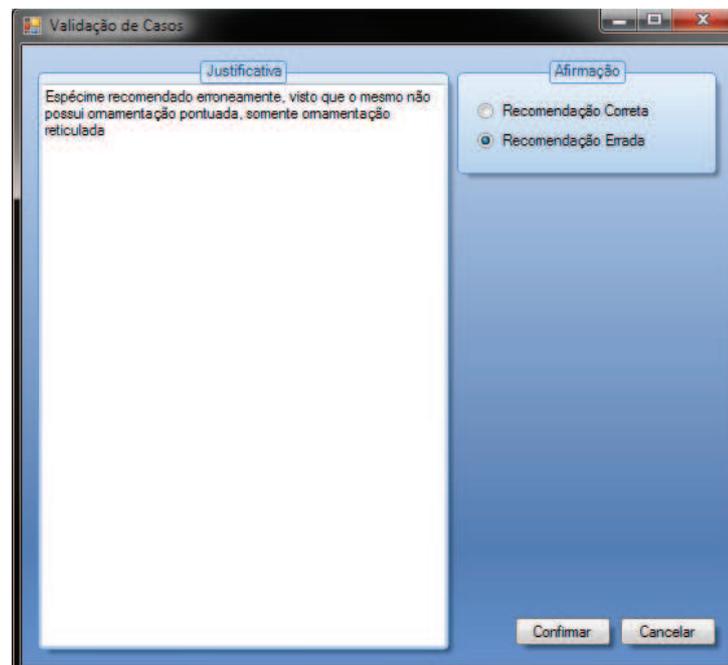


Figura 6.20: Interface para validação dos conhecimentos sugeridos

Na interface de inferências é oferecido aos usuários os casos inseridos no sistema para

um determinado domínio, permitindo a sua visualização. Para isso, o usuário após realizar uma inferência deve visualizar um conhecimento e as informações dos casos associados aos valores e ao conhecimento são apresentadas na aba de Casos, presente na parte superior esquerda da interface de inferência, como apresentado na Figura 6.21

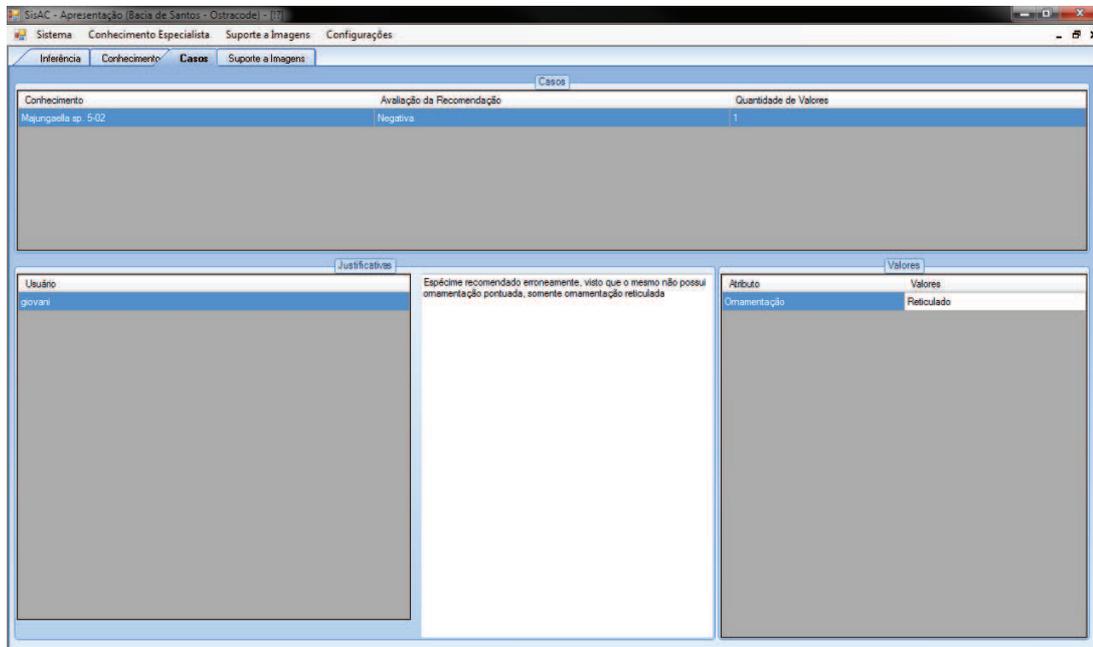


Figura 6.21: Interface de exibição dos casos de conhecimentos validados

Nessa tela, na parte superior, são exibidos todos os casos associados ao conhecimento, a avaliação recebida e a quantidade de valores utilizados para chegar ao conhecimento. Com isso, quando um caso é selecionado, são apresentados, na parte inferior esquerda, todos os usuários que avaliaram o conhecimento tendo como base o mesmo conjunto de valores. E se algum desses usuários é selecionado, então na parte inferior central é apresentada a respectiva justificativa. Na parte inferior direita são apresentados os valores utilizados como base para a sugestão do conhecimento e validação do mesmo.

### 6.3.1 Interfaces para Processamento de Imagem

Dentre as interfaces apresentadas para que o usuário possa inferir no sistema em busca de conhecimentos, também é oferecida uma interface para a utilização dos algoritmos de PI, implementados para auxiliar na definição sobre qual valor melhor representa o objeto de estudo a ser classificado. Essa tela é dividida verticalmente em dois lados, onde no lado esquerdo é possível selecionar o algoritmo a ser utilizado, a descrição e ajuda para utilização e o botão para carregar a imagem a ser processada. Já no lado direito é apresentada a implementação da interface ISuporteInferencia do respectivo algoritmo selecionado.

Na Figura 6.22 é apresentada a implementação do algoritmo de MCO. Para a utilização do algoritmo, após o usuário realizar a carga da imagem, o mesmo deve selecionar com o *mouse* a região da imagem que melhor define a ornamentação do espécime, e solicitar que ela seja calculada.

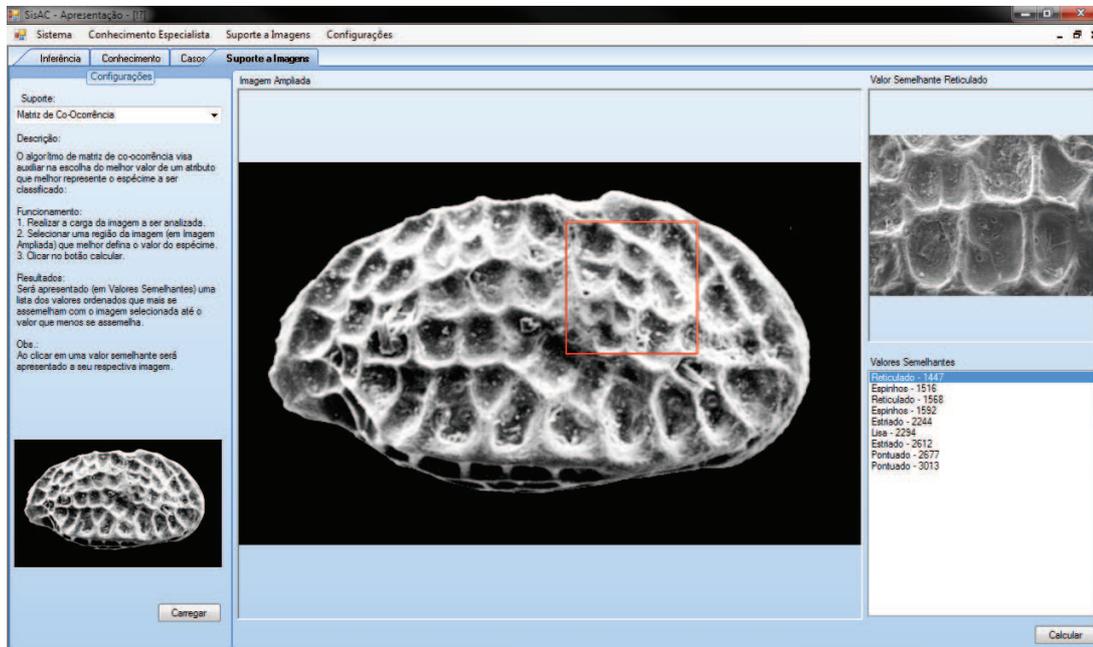


Figura 6.22: Interface para processamento de imagens utilizando o algoritmo de MCO

Após o cálculo, no canto extremo direito da tela é apresentado uma lista de valores que demonstra a distância da textura apresentada pelo usuário em relação as imagens dos valores utilizados para criar o espaço amostral. Caso o usuário selecione algum desses valores, na parte superior é apresentado a imagem do respectivo valor, permitindo uma avaliação do usuário.

No caso apresentado na Figura 6.22, o usuário apresenta um espécime com a sua ornamentação desgastada pela erosão, marcando em vermelho uma região mais preservada. Após o cálculo, o algoritmo sugeriu que os valores estatísticos extraídos da textura apresentada está próxima aos valores extraídos de uma imagem do valor reticulado. Sugerindo então que a imagem apresentada pelo usuário possua uma ornamentação reticulada.

Já a Figura 6.23 apresenta a implementação do algoritmo de DCK. A utilização desse algoritmos consiste em informar as variáveis de entrada K, LCC e Threshold, como apresentado na seção 3.2.1. A cada modificação o sistema realiza o processamento da image carregada exibindo os cantos curvos detectados destacados em vermelho.

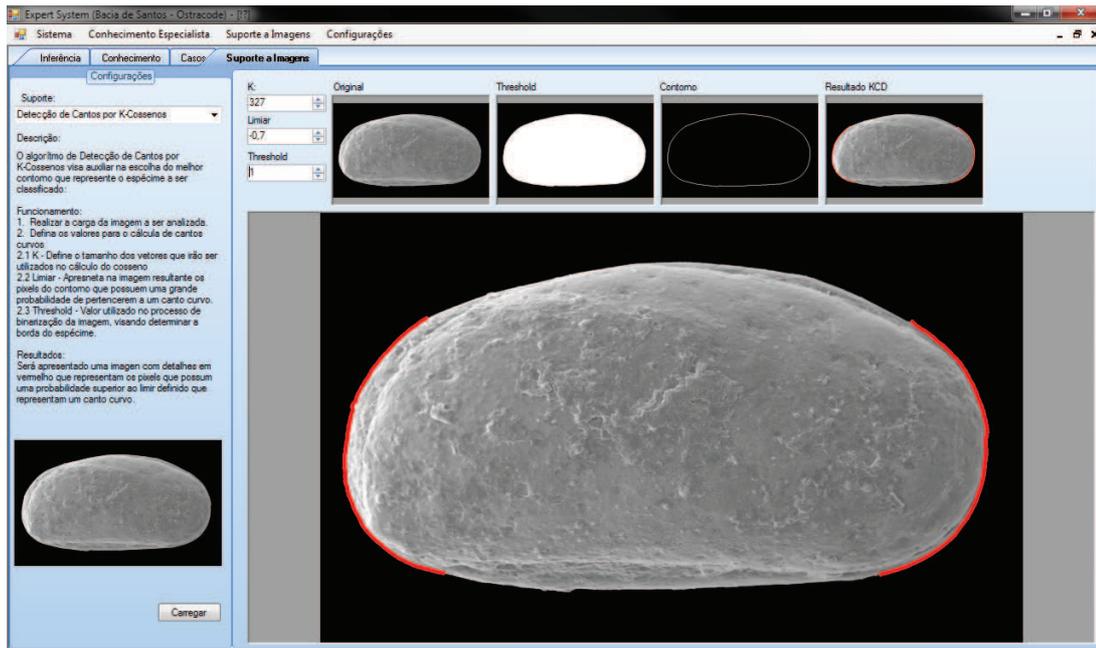


Figura 6.23: Interface para processamento de imagens utilizando o algoritmo de DCK

A cada processamento são apresentadas todas as imagens resultantes do processo, permitindo a um usuário mais técnico analisar e melhorar as variáveis de entrada, a fim de conseguir uma detecção melhor dos cantos. A definição do valor é com base na quantidade de cantos e pela avaliação do usuários. Sendo que se o sistema destacar 2 cantos curvos o espécime apresenta um contorno sub-ovalado, 3 cantos um contorno sub-triangular e 4 cantos um contorno sub-retangular.

Optou-se por deixar a definição do valor para o usuário, pois o processo pode gerar pequenas quebras na continuidade do canto detectado, inferindo que o espécime poderia possuir mais cantos do que realmente possui. Essa quebra pode ocorrer por diversos problemas, desde a segmentação da imagem, pouca qualidade até mesmo a detecção grosseira do contorno do espécime.

## 7 VALIDAÇÃO DO SISTEMA

Esse capítulo tem como objetivo apresentar a metodologia utilizada para validar o sistema. O capítulo está dividido em três seções, sendo inicialmente apresentado a forma de validação utilizada e os motivos que levaram a esta abordagem, e posteriormente os resultados obtidos através das atividades realizadas. Por fim, serão apresentados os comentários gerais sobre os resultados analisados.

### 7.1 FORMA DE VALIDAÇÃO

Como mencionado anteriormente, visto que o sistema apresentado neste trabalho teve a sua base de conhecimento construída com base em pesquisas pontuais e por especialistas humanos, os quais podem gerar divergências entre si, acaba sendo uma tarefa difícil validar de forma numérica e exata, sendo praticamente impossível apresentar resultados formais de correção de um SE (RICH; KNIGHT, 1993). Outro problema que dificulta a avaliação do sistema é a origem das informações utilizadas para a construção da base de conhecimento, que neste caso, teve-se como base o atlas de espécimes da bacia de Santos. Sendo assim, quando testadas as regras com base no mesmo atlas todos os conhecimentos sugeridos estão corretos, no entanto, quando o sistema é utilizado por usuários que não tem acesso ao atlas, algumas divergências de regras podem ocorrer visto a subjetividade que está presente nesse domínio de estudo.

Com isso, foi utilizado uma abordagem de validação do sistema com usuários fim, que nesse caso seriam especialistas e/ou neófitos que estudam e/ou trabalham com a área de conhecimento em questão. Com base no estudo de caso abordado neste trabalho, especialistas e neófitos da área de micropaleontologia foram convidados a participar da validação do sistema SisAC utilizando o sistema e respondendo um questionário. Metodologia semelhante a utilizada em trabalhos correlatos como em Gonzales-Andujar (2009).

O questionário foi composto por quatro atividades, onde as atividade 1 e 2 os participantes deveriam classificar um conjunto de espécimes sendo a primeira parte utilizando apenas o conhecimento próprio que cada um possuía e a segunda parte utilizando o sistema SisAC. Já as atividades 3 e 4 tinham como objetivo fazer com que os participantes expressassem suas observações e críticas, sendo a atividade 3 com questões fechadas com perguntas mais específicas ao sistema e a atividade 4 com questões abertas permitindo que os participantes pudessem dissertar sobre as questões. O questionário é apresentado no Apêndice A.

Inicialmente esperava-se ter como grupo de participantes que estivessem no curso de Micropaleontologia da Universidade do Vale do Rio dos Sinos (Unisinos). No entanto, no semestre em que este trabalho foi validado não havia nenhuma cadeira sendo

ministrada. Com isso, foi aberto o convite para o grupo de pesquisas do Laboratório de Micropaleontologia da Unisinos na área de taxonomia e bioestratigrafia de Ostracodes marinhos mesozóicos e cenozóicos. Assim, o grupo foi composto por 1 especialistas, 3 pessoas com conhecimentos avançados, 1 pessoa com conhecimento intermediário e 1 pessoa com conhecimento básico, e nenhum participante com conhecimento a nível de neófito sobre o domínio de estudo.

A validação do sistema teve início quando parte do questionário foi enviado aos participantes no dia 29 de novembro de 2010, solicitando que fosse realizada a atividade 1, visto que essa atividade pudesse consumir bastante tempo e talvez não fosse possível realizar em um único dia. Nessa parte foi especificado que a classificação fosse realizada de forma individual e sem o auxílio do atlas utilizado para a construção da base de conhecimento ou de outros colegas da área de estudos. Já a validação do sistema presencial com os participantes ocorreu nas dependências da UNISINOS, na sala 407 do prédio 6B, no dia 03 de dezembro de 2010 às 14:00.

Na validação presencial inicialmente foi apresentado o sistema aos participantes, descrevendo as funcionalidades que permitissem a realização de inferências na base de conhecimento e como os dados apresentados poderiam ser interpretados. Também foram apresentadas as técnicas de processamento de imagens e como elas poderiam ser utilizadas, auxiliando na decisão de escolha de algum atributo. Durante todo o período de validação do sistema, os participantes puderam realizar questionamentos e tirar possíveis dúvidas sobre o sistema e seu funcionamento. Junto ao Apêndice C é possível ver fotos tiradas durante o processo de validação do sistema.

## 7.2 RESULTADOS OBTIDOS

Com os questionários preenchidos, foram extraídas as respostas e analisadas as observações apresentadas pelos participantes. Sendo os resultados discriminados segundo o conhecimento que cada participante acreditava possuir sobre o domínio de estudos.

### 7.2.1 Classificação de espécimes

Essa atividade tinha como objetivo avaliar se o sistema poderia oferecer uma melhora de desempenho na classificação de espécimes de Ostracodes em um grupo de estudantes da Micropaleontologia. No entanto, essa atividade foi utilizada apenas como forma de inserir os participantes na tarefa de classificação de espécimes, utilizando o sistema como uma ferramenta de apoio. Permitindo que o restante do questionário pudesse ser analisado, porém com a visão de especialistas do domínio.

Com isso, não foi possível obter resultados significativos de acertos na classificação dos espécimes, visto que no grupo participante, exceto um, possuía um nível de

conhecimento na escala de intermediário a especialistas. Sendo assim, tais participantes não tiveram dificuldades na realização da atividade, como é apresentado no gráfico, da Figura 7.1, os resultados das classificações referentes a atividade 1, com o conhecimento próprio, e 2, com o auxílio do sistema.

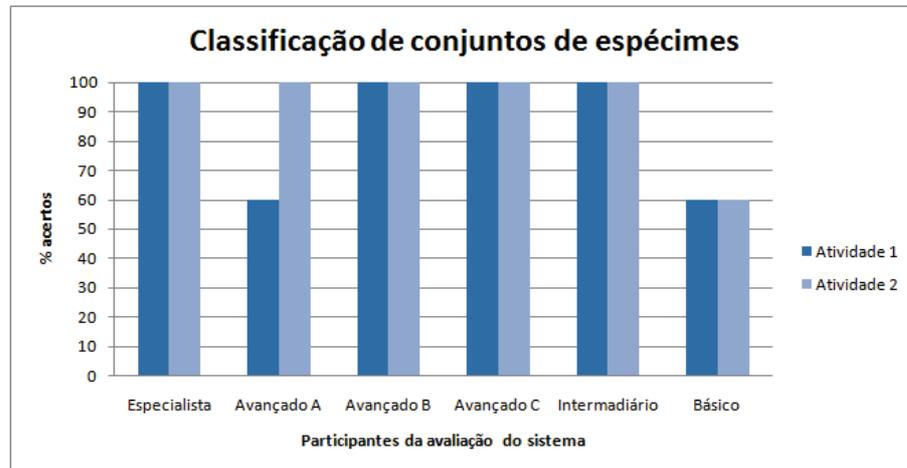


Figura 7.1: Resultado da classificação de espécimes referentes as atividades 1 e 2

### 7.2.2 Avaliação do sistema

Essa atividade tinha como objetivo extrair opiniões pontuais que os participantes observaram no sistema durante a sua utilização, questionando-os se os objetivos que o sistema se propõem a resolver foram atendidos. Entre essas questões foram observadas se o sistema: se mostrou de simples utilização, auxiliou os usuários a terem um desempenho melhor na classificação; se as imagens auxiliaram no processo de classificação; se o sistema apresentou todas as informações necessárias para chegar a uma classificação e se a subjetividade presente nos atributos e valores que caracterizam um espécime foram reduzidas com a utilização do sistema. Os resultados apresentados a seguir estão discretizados pelo conhecimento que cada participante acreditava possuir, sendo o eixo das abscissas as possíveis respostas para as questões e o eixo das ordenadas o número de participantes que responderam com a respectiva resposta.

Como respostas da primeira questão (O sistema mostrou-se de simples em sua utilização?), os resultados demonstram que o sistema se mostrou de fácil utilização segundo a avaliação dos participantes, não oferecendo maiores dificuldades, além das já presentes no processo de classificação. Esses resultados são apresentados na Figura 7.2.

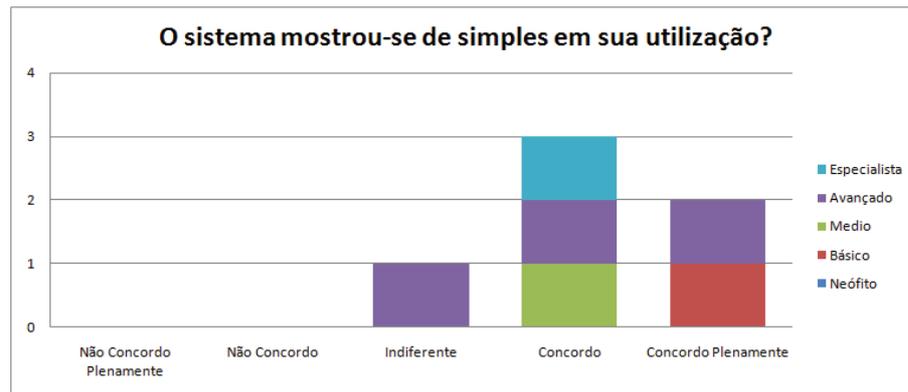


Figura 7.2: Avaliação do manuseio do sistema

Já em relação a desempenho da classificação (Acredita que o seu desempenho na classificação de Ostracodes aumentou com a utilização do sistema SisAC?) o sistema se mostrou insatisfatório na opinião dos participantes, como apresenta na Figura 7.3. Em parte esse resultado se deve ao fato dos participantes já possuírem conhecimento suficiente para classificarem os espécimes sem auxílio, como já mencionado anteriormente. No entanto, o participante de conhecimento básico, acredita que seu conhecimento tenha sido melhor utilizando o sistema como apresentado nos resultados, mesmo que isso não tenha se mostrado verdade no resultado anterior apresentado na Figura 7.1.

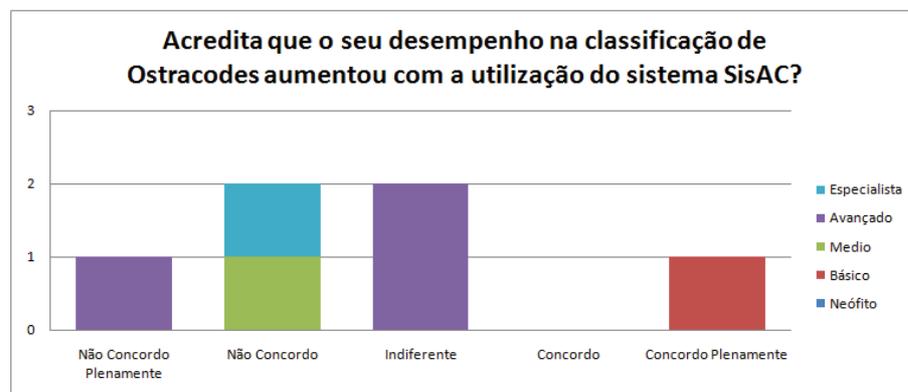


Figura 7.3: Avaliação do desempenho nos resultados de classificação

Em relação aos conhecimentos apresentados pelo sistema (O sistema apresentou todas as informações necessárias para um classificação de Ostracodes?), não foi possível obter uma análise visto que não houve consenso entre os participantes, com base nas respostas dos participantes, como apresentado na Figura 7.4. No entanto, essa situação já era esperada como mencionada anteriormente, pois o conhecimento apresentado está baseado no atlas utilizado para a criação da base de conhecimentos. Assim, para alguns participantes esse conhecimento pode não ter sido suficiente, diferente da opinião de outros participantes.

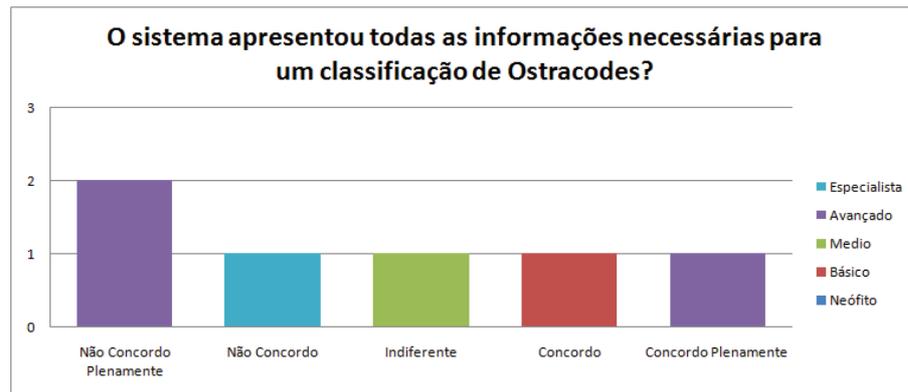


Figura 7.4: Avaliação de como o sistema apresentou as informações necessárias à classificação

Em relação a utilização de imagens como forma de auxílio à classificação os participantes opinaram que concordam ou concordam plenamente, como apresentado na Figura 7.5. Essa situação era esperada, pois em diversos trabalhos correlatos a utilização de imagens se mostrava de grande utilidade, e neste trabalho mostrou que a possibilidade de visualizar as opções a serem trabalhadas no processo de classificação de Ostracodes é de grande ajuda.

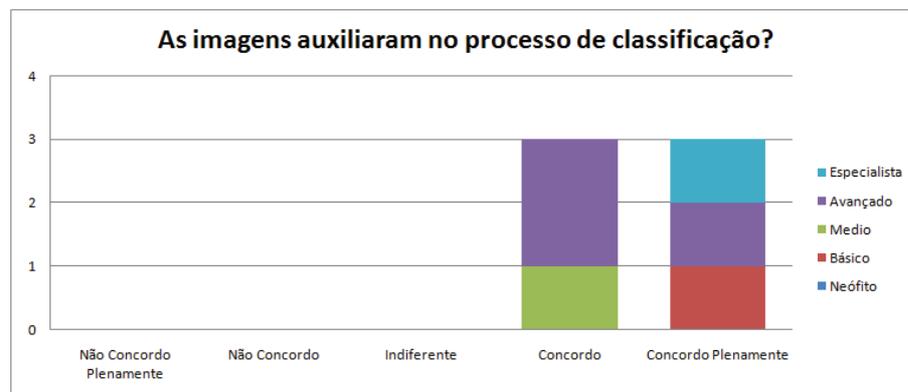


Figura 7.5: Avaliação de como as imagens auxiliaram na classificação

Em relação a redução da subjetividade não foi possível extrair nenhuma conclusão, visto que os participantes apresentaram respostas totalmente distribuídas na grade de respostas, como apresentado na Figura 7.6. No entanto é visível que quanto menos conhecimento o participante possui maior a probabilidade que o sistema auxilie na redução da subjetividade.

Isso deve-se ao fato de que especialistas já possuem conhecimento suficiente para lidarem com as subjetividades do domínio, sabendo distingui-las com maior facilidade. Resultado e análise semelhantes as respostas obtidas quando questionados se o sistema melhorou o desempenho na classificação com a utilização do sistema.

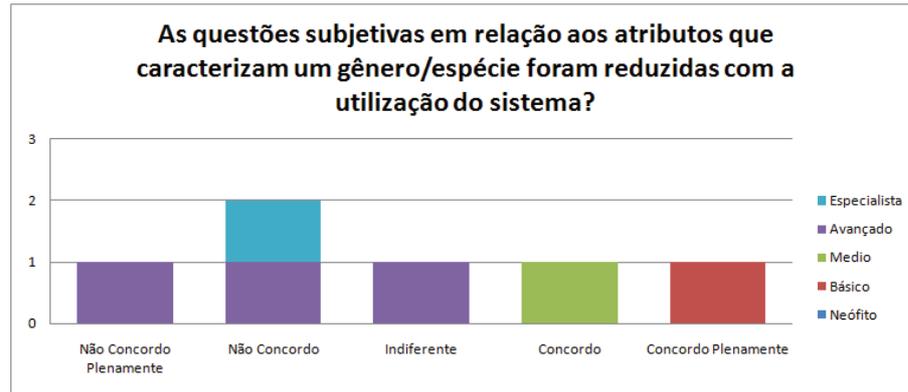


Figura 7.6: Avaliação de como o sistema auxiliou na redução da subjetividade

### 7.2.3 Observações e críticas

Esta atividade teve como objetivo obter as observações pessoais dos participantes em relação a utilização do sistema durante as atividades, permitindo que eles expressassem as suas observação e críticas com suas próprias palavras. Permitindo que fosse analisado como o sistema se comportou em relação aos objetivos que o SisAC se propunha a resolver. As respostas apresentadas por cada participante estão no Apêndice B, as quais estão divididas por questões e pelo tipo de conhecimento que cada um acredita possuir.

As questões, A (Quais foram as maiores dificuldades durante a classificação de Ostracodes utilizando o seu conhecimento próprio?) e B (Quais foram as maiores dificuldades durante a classificação de Ostracodes utilizando o sistema SisAC?) tinham como objetivo extrair dos participantes as principais dificuldades observadas entre cada forma de classificação. Como respostas da questão A, os participantes apresentaram dificuldades em relação a grande quantidade de espécimes, subjetividade e a falta de material físico, como já era esperado. Visto que esses eram os grandes objetivos que o sistema se propunha a tratar.

Em relação a questão B, os participantes apresentaram diversos pontos de dificuldades. Entre essas dificuldades foram apresentadas questões que se referem a base de conhecimento e utilização do sistema e outras questões apontadas em relação as dificuldades de classificação. Entre elas, estão "A exata relação entre as características do espécime em estudo e as possibilidades listadas no programa" como mencionado pelo participante especialista e "As maiores dificuldades estão nas definições de cada atributo (conceitos) que podem variar de acordo com o usuário" como mencionado por um participante de conhecimento avançado, demonstrando que o sistema ainda permite questões subjetivas. As demais respostas apresentaram a falta de conteúdo e/ou informação dos conhecimentos na base de dados.

Com a questão C (Por meio das funcionalidades oferecidas pelo sistema SisAC, como o sistema auxiliou no processo de classificação dos Ostracodes?) buscou-se analisar se o

sistema auxiliou e de que forma os participantes realizaram as classificações utilizando o sistema. Dentre os participantes, somente um de conhecimento avançado demonstrou que o sistema não o auxiliou na classificação, pois o mesmo já possuía conhecimento suficiente para classificar os espécimes com seu próprio conhecimento. Já os demais participantes demonstraram que o sistema apresentou pontos positivos como a organização e forma de apresentação dos atributos e valores. Entre os participantes de conhecimento intermediário e básico foi demonstrado que para eles o sistema reduziu a subjetividade da seleção das características.

Já com a questão D (Entre o processo de classificação de forma manual e com o sistema, qual das duas atividades consumiu mais tempo para a realização da mesma?) tentou-se extrair dos participantes dentre os dois modos de classificação, manual e com o sistema, entre qual deles eles acreditavam ter perdido maior tempo para realizar as atividades. No entanto, como a maioria dos participantes já conheciam os espécimes, alguns deles acreditam que de forma manual foram mais rápidos, no entanto descrevem que o sistema quando utilizado por usuário iniciantes ao estudo o mesmo deve ser muito útil.

A questão E (Descreva/comente como as imagens auxiliaram ou poderiam auxiliar no processo de classificação por meio das funcionalidades?) busca extrair dos participantes observações deles em relação aos imagens e como elas são exibidas no sistema durante a classificação dos espécimes. Como demonstrado nos resultados anteriores, as imagens se demonstraram úteis no processo de escolha dos atributos, como relatado por um participante de conhecimento avançado "As imagens são chave para identificação, sem elas o tempo para identificação irá aumentar".

Por meio da questão F (Descreva as diferenças observadas em relação as dificuldades encontradas na classificação de Ostracodes utilizando o seu conhecimento em relação ao auxílio oferecido pelo sistema?) tentou extrair dos participantes comparações entre a classificação de forma manual e a classificação com o auxílio do sistema. Por meio das respostas ficou visível que o sistema se destacou em relação a classificação manual, de modo geral, segundo os participantes que responderam a questão o sistema foi capaz de oferecer interfaces com os conjuntos de dados necessários, tornando mais rápido o processo de classificação facilitando a determinação do espécime, já que o sistema apresenta seus resultados de forma direta, não necessitando uma pesquisa em bibliografias e internet.

Por fim, a questão G (Escreva suas críticas, observações, sugestões que possam auxiliar e aprimorar o desenvolvimento contínuo do sistema.) procurava buscar dos participantes suas observações em relação ao sistema, permitindo que ele fosse validado e melhorado continuamente. Entre as respostas foi observado que vários participantes citam as imagens e sugerem que poderiam haver mais imagens específicas sobre características morfológicas e uma melhora nas informações presentes no banco de dados. Outros pontos apresentados pelos participantes, é que apesar do sistema não se mostrar muito eficiente

para os participantes que já possuíam um conhecimento elevado, mencionam que o sistema é interessante e que funcionou corretamente para o que se propunha. No entanto, é exposto que o estudo de caso utilizado é de difícil classificação pois entre os próprios pesquisadores não há um consenso sobre a taxonomia, visto a preservação pobre dos espécimes e pela subjetividade estarem intrínsecas ao estudo, e não aos pesquisadores, como citado pelos participantes.

### 7.3 COMENTÁRIO GERAL DOS RESULTADOS

Apesar da validação do sistema não ter sido realizada por um grupo maior de participantes com sua maioria de conhecimento básico ou neófito, o mesmo pode ser avaliado porém com olhares de especialistas. Apesar de não haver uma satisfação do grupo de usuários avaliados, os mesmos manifestaram que o sistema atendeu seus objetivos e que é de grande ajuda para usuários que não possuem grandes conhecimentos sobre o domínio.

Outras questões que foram abordadas pelos participantes são referentes ao conhecimento apresentado e que o mesmo poderia ser melhorado, inserindo mais informações como referencial bibliográfico e imagens. No entanto, o sistema permite que sejam incluídos documentos, trechos de bibliografias, descrições textuais e um grande número de imagens. Porém na construção da base de conhecimento foi utilizado apenas o atlas, visto a falta de um especialista que pudesse se dedicar a tarefa de alimentar ou dar suporte na construção da base de conhecimento.

Entre os resultados observados, algo que chamou a atenção foram os resultados quanto a utilização das imagens, que se mostraram de interesse dos participantes, apresentando grande utilidade no processo de classificação. Isso se deve ao fato que no processo de classificação manual, as imagens que poderiam auxiliar estavam distribuídas em diversas bibliografias e na internet e o sistema por sua vez agrupou diversas imagens que auxiliam a distinguir os diversos tipos de atributos e valor presentes no domínio.

De modo geral, a atividade de validação do sistema mostrou que os objetivos que o SisAC se propunha a atender foram alcançados. No entanto não com a eficiência para dar suporte a especialistas do estudo de caso como apresentado, porém com capacidade para auxiliar novos pesquisadores ou estudantes que estão ingressando nesse domínio do conhecimento. Utilizando o SisAC como uma ferramenta de ensino ou como suporte em um início de pesquisa.

## 8 CONCLUSÕES

O objetivo dessa dissertação era desenvolver um sistema que pudesse organizar e manter um grande volume de informações diversificadas e permitisse que tal conhecimento pudesse ser recuperado facilmente, com abordagens técnicas de Inteligência Artificial como Sistemas Especialistas, associada a utilização de Processamento de Imagens. Para que esse conhecimento fosse organizado e catalogado foi utilizado um motor de Conhecimento Baseado em Regras, permitindo que usuários especialistas construíssem uma base de conhecimento e montassem seus raciocínios lógicos em forma de regras. Permitindo assim, que posteriormente outros usuários tenham acesso as informações utilizando os raciocínios previamente trabalhados. Já a utilização de técnicas de PI como Detecção de Cantos por K-Cossenos e Matriz de Co-Ocorrências visam auxiliar os usuários do sistema na decisão de quais atributos e valores melhor representam o objeto de estudo que está sendo classificado. Na caso deste trabalho, Ostracodes.

Como resultado do desenvolvimento deste trabalho, obteve-se um sistema robusto e genérico que permite manter de forma organizada um conjunto de informações, com possibilidade de disseminação do conhecimento em uma rede local ou até mesmo na Internet. Por meio do SisAC é possível trabalhar com inúmeras áreas do conhecimento de forma única ou separada por domínios de trabalho, pois é possível criar os atributos e valores conforme as necessidades de evolução do raciocínio de regras, o qual deve se assemelhar ao conhecimento do especialistas. Em relação ao Processamento de Imagens é possível desenvolver, acoplar e utilizar outras abordagens de processamento e análise, conforme as necessidades da área, sendo apenas necessário possuir conhecimento em linguagem de programação e sobre a técnica desejada.

Com isso, para demonstrar tais funcionalidades foi utilizado um estudo de caso na área de Micropaleontologia, que teve como objetivo auxiliar estudantes e pesquisadores na atividade de classificação de Ostracodes. Como resultado da atividade de validação do sistema foi possível observar que o sistema não se mostrou eficaz para usuários que já possuem um grande conhecimento do domínio, visto que não há grandes dificuldades em reconhecer os espécimes utilizados. No entanto, tais especialistas mencionam que o sistema será de grande ajuda para estudantes e pesquisadores que estão iniciando nos domínios que o sistema possui em sua base de conhecimento. Dentre os resultados também destacou-se a facilidade com que o conhecimento pode ser encontrado utilizando o sistema já que o mesmo está agrupado em um único local e não distribuído em literaturas da área e na internet. Outra questão que chamou a atenção é em relação a utilização de imagens, o que foi amplamente destacado pelos participantes como uma forma de auxílio na classificação dos espécimes, já que dúvidas na sua definição são comuns.

Para o desenvolvimento do trabalho não houveram grandes problemas em relação

a definição, projeto e implementação. No entanto, os maiores problemas encontrados foram em relação a interação com os especialistas para a aquisição do conhecimento, visto que os mesmos possuíam projetos e demandas que deviam ser atendidas com maior prioridade. Isso acabou gerando uma base de conhecimento não tão rica como o esperado, contendo basicamente as informações de um atlas e não o conhecimento e o raciocínio dos especialistas como um todo.

Como apresentado pelos resultados os objetivos foram alcançados, porém não ao ponto de dar suporte a um especialista da área. Porém se mostrou uma grande contribuição visto o sistema como uma ferramenta de integração e disseminação do conhecimento, permitindo a união de vários especialistas da mesma área porém com focos de pesquisas distintos oferecendo a possibilidade de união de toda a informação em um único local. Além de servir como mecanismos de treinamento e aprendizado para novos pesquisadores e estudantes.

Como trabalhos futuros inicialmente será aprimorada a base de conhecimento incluindo novos conhecimentos e informações, assim como a melhoria do sistema em observações apresentadas pelos participantes na validação do sistema. Também deseja-se estender o sistema para outros domínios que utilizem imagens como auxílio na tarefa de classificação e realizar a validação utilizando um grupo maior de neófitos e especialistas que não estejam inseridos totalmente no domínio abordado.

## BIBLIOGRAFIA

- BERGUE, C. T. *Agulhas e pincéis: as relações entre a paleontologia e a neontologia no estudo dos ostracodes. (no prelo)*. [S.l.]: Terrae Didatica, 2010.
- BITTENCOURT, G. *Inteligência Artificial. Ferramentas e Teorias*. 3<sup>a</sup>. ed. Florianópolis, RS - Brasil: Editora da UFSC, 2006. 371 p.
- BRADSKI, G.; KAEHLER, A. *Learning OpenCV: Computer Vision with the OpenCV Library*. Cambridge, MA: O'Reilly, 2008.
- CARVALHO, I. S. *Paleontologia*. Rio de Janeiro: Editora Interciência, 2004. 861 p.
- COIMBRA, J.; BERGUE, C. Paleontologia. In: \_\_\_\_\_. Rio de Janeiro: Editora Interciência, 2002. cap. Ostracode, p. 355–368.
- CONTRERAS, W. F.; GALINDO, E. G.; MORILLAS, A. B.; LORENZO, P. M. An application of expert systems to botanical taxonomy. *Expert System with Application*, Elsevier, n. 25, p. 425–430, 2003.
- GARCIA, A. C. B.; MACIEL, P. M.; FERRAZ, I. N. Addgeo: an intelligent agent to assist geologist finding petroleum in offshore lands. Springer-Verlag New York, Inc., Secaucus, NJ, USA, p. 316–321, 2000.
- GONZALES-ANDUJAR, J. Expert system for pests, diseases and weeds identification in olive crops. *Expert Systems with Application*, Elsevier, n. 36, p. 3278–3282, 2009.
- GONZALES-ANDUJAR, J.; FERNANDEZ-QUINTANILLHA, C.; IZQUIERDO, J.; URBANO, J. Simce: An expert system for seedling weed identification in cereals. *Computers and electronics in agriculture*, Elsevier, n. 54, p. 115–123, 2006.
- GONZALEZ, R. C.; WOODS, R. E. *Digital Image Processing*. [S.l.]: Addison-Wesley Publishing Company. Inc., 1992.
- HAQ, B. U.; BOERSMA, A. *Introduction to Marine Micropaleontology*. [S.l.]: Elsevier, 1998. 376 p.
- HARALICK, R. M.; SHANMUGAM, K.; DINSTEN, I. Tetural features for image classification. *IEEE Transactions on Systems*, IEEE, v. 3, n. 6, p. 610–621, 1973.
- HORNE, D.; COHEN, A.; MARTENS, K. The ostracoda: Applications in quaternary research. *Geophysical Monograph*, n. 131, p. 5–36, 2002.

- HUNG, S.; CHEN, C. Mammographic case base applied for supporting image diagnosis of breast lesion. *Expert Systems with Application*, Elsevier, n. 30, p. 93–108, 2006.
- JARMULAK, J.; KERCKHOFFS, E. J.; VEEN, P. Case-based reasoning for interpretation of data from non-destructive testing. *Engineering Applications of Artificial Intelligence*, Pergamon, n. 14, p. 401–417, 2001.
- KALOUDIS, S.; ANASTOPOULOS, D.; YIALOURIS, C.; LORENTZOS, N. A.; SIDERIDIS, A. Insect identification expert system for forest protection. *Expert Systems with Application*, Elsevier, n. 28, p. 445–452, 2005.
- LEVIN, H. *The Earth Through Time*. [S.l.]: John and Sons, Inc, 2006. 547 p.
- LIAO, S. Expert system methodologies and applications - a decade review from 1995 to 2004. *Expert Systems with Application*, Elsevier, n. 28, p. 93–103, 2005.
- NIKOLOPOULOS, C. *Expert Systems: Introduction to first and second generation and hybrid knowledge based systems*. [S.l.: s.n.], 1997.
- PARK, M.; KANG, B.; JIN, S.; LUO, S. Computer aided diagnosis system of medical images using incremental learning method. *Expert System with Application*, Elsevier, n. 36, p. 7242–7251, 2009.
- POST, E. Formal reduction of the general combinatorial problem. *American Journal of Mathematics*, p. 197–268, 1943.
- QIAN, Y.; XU, L.; LI, X.; LIN, L.; KRASLAWSKI, A. Lubres: An expert system development and implementation for real-time fault diagnosis of a lubrication oil refining process. *Expert Systems with Application*, Elsevier, n. 35, p. 1252–1266, 2008.
- RIBEIRO, H. da Cunha e S. *Introdução aos Sistemas Especialistas*. [S.l.: s.n.], 1987.
- RICH, E.; KNIGHT, K. *Inteligência Artificial*. 2<sup>a</sup>. ed. [S.l.]: McGraw-Hill Science/Engineering/Math, 1993. 640 p.
- RODRÍGUEZ-SOLANO, C.; LAITA, L. M.; ROANEZ-LOZANO, E.; LÓPEZ-CORRAL, L.; LAITA, L. A computational system for diagnosis of depressive situation. *Expert Systems with Application*, Elsevier, n. 31, p. 47–55, 2006.
- RUSSEL, S.; NORVIG, P. *Inteligência Artificial*. [S.l.]: Elsevir, 2004. 1021 p.
- SHORTLIFFE, E. *Computer-Based Medical Consultation: MYCIN*. [S.l.]: Elsevier Science Ltd, 1976. 286 p.

SRINIVAS, Y.; TIMMONS, W. D.; DURKIN, J. A comparative study of three expert systems for blood pressure control. *Expert Systems with Application*, Elsevier, n. 20, p. 267–274, 2001.

STEIN, T. *Avaliação de descritores de textura para segmentação de imagens*. Dissertação (Mestrado): Universidade Federal do Paraná, 2005.

SUN, T. K-cosine corner detection. *Journal Of Computers*, Academy Publisher, n. 3, p. 16–22, 2008.

## APÊNDICE A - QUESTIONÁRIO UTILIZADO NA ATIVIDADE DE AVALIAÇÃO

Universidade do Vale do Rio dos Sinos  
Programa Interdisciplinar de Pós-Graduação em Computação Aplicada  
Dissertação de Mestrado

SisAC - Sistema de Auxílio a Classificação  
Estudo de caso: Ostracode

Esse questionário tem como objetivo avaliar o sistema SisAC que visa auxiliar especialistas e neófitos na atividade de classificação de seus objetos de estudo. Como estudo de caso este trabalho propõe ajudar no processo de classificação de Ostracodes. Todas as informações preenchidas nesse formulário serão mantidas em anonimato, e serão utilizados como dados de validação do sistema descrito na dissertação de Mestrado de Giovani Manica Barili.

A base de conhecimento do sistema utiliza informações de espécimes encontrados na bacia de Santos e que já foram analisados e classificados por especialistas, podendo haver informações subjetivas. Com isso, será assumido que as informações utilizadas para a construção da base de conhecimento estão corretas.

Todas as imagens presentes nesse questionário estão disponíveis em <http://www.pilotoes.kinghost.net/questionario/>

### **Atividade 01 – Classificar utilizando conhecimento próprio**

Essa atividade tem como objetivo avaliar o conhecimento de cada participante sem o auxílio de qualquer fonte de informação.

As seguintes opções devem ser preenchidas com seus respectivos espécimes, que são apresentados na seqüência. Como exemplo: **X ( YA )**, onde X é o nome da espécie/gênero e o Y é o número da imagem que representa o espécime.

**Henryhowella ( )**

**Bradleya ( )**

**Procytheridea ( )**

**Fossocytheridea ( )**

**Buntonia ( )**

Imagem 1A:

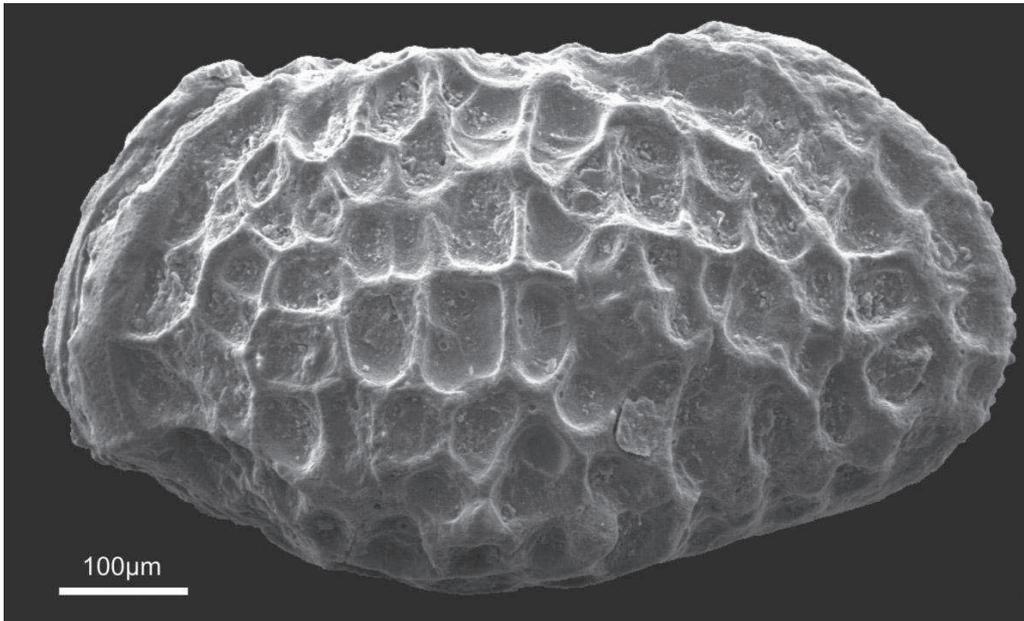


Imagem 2A:

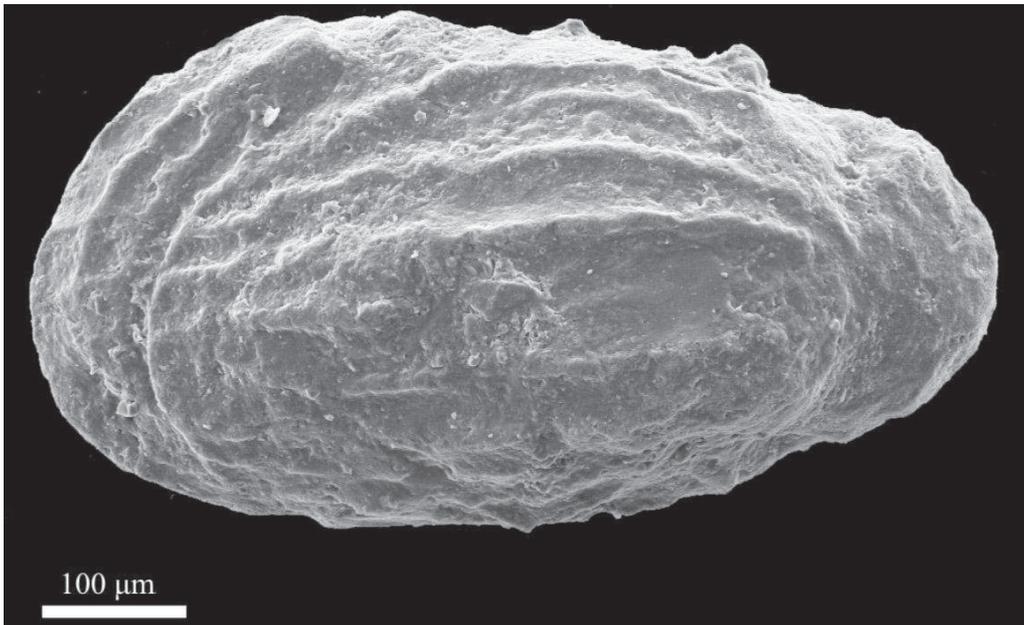


Imagem 3A:

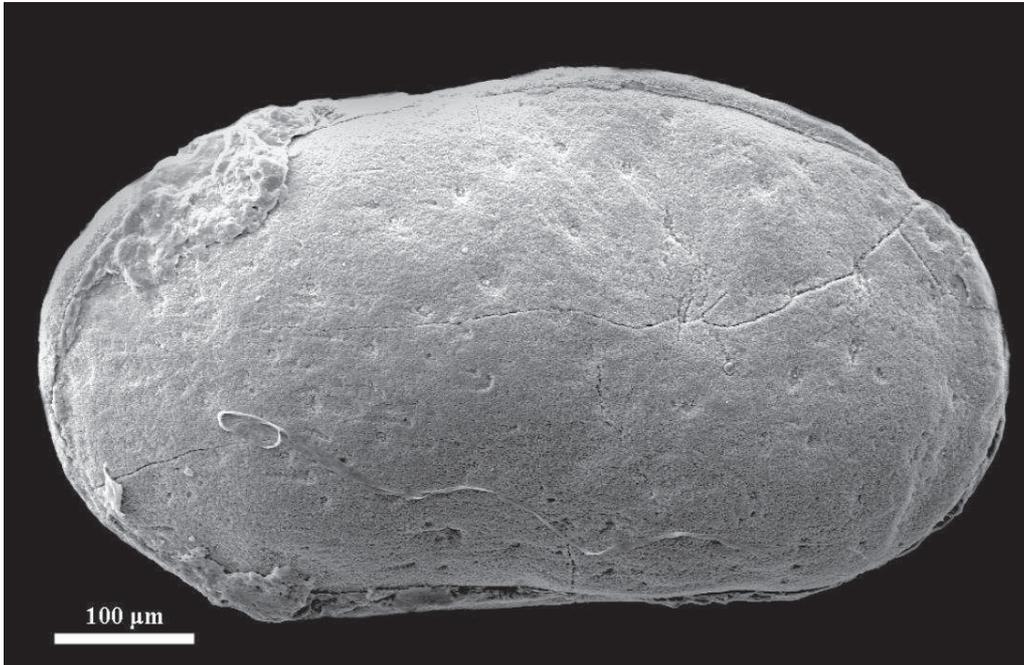


Imagem 4A:

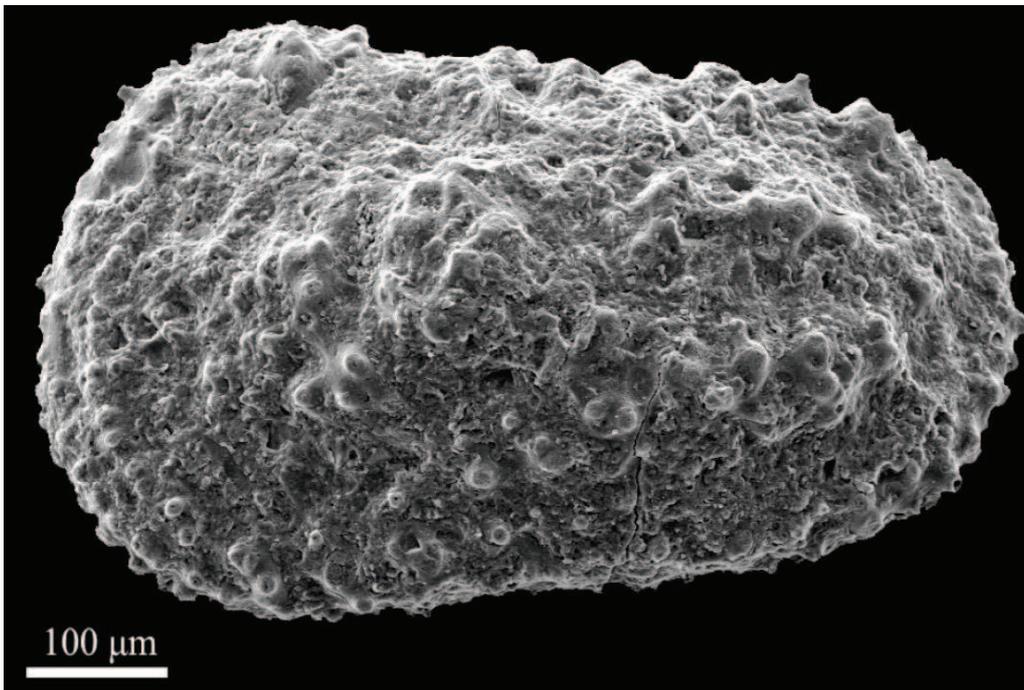
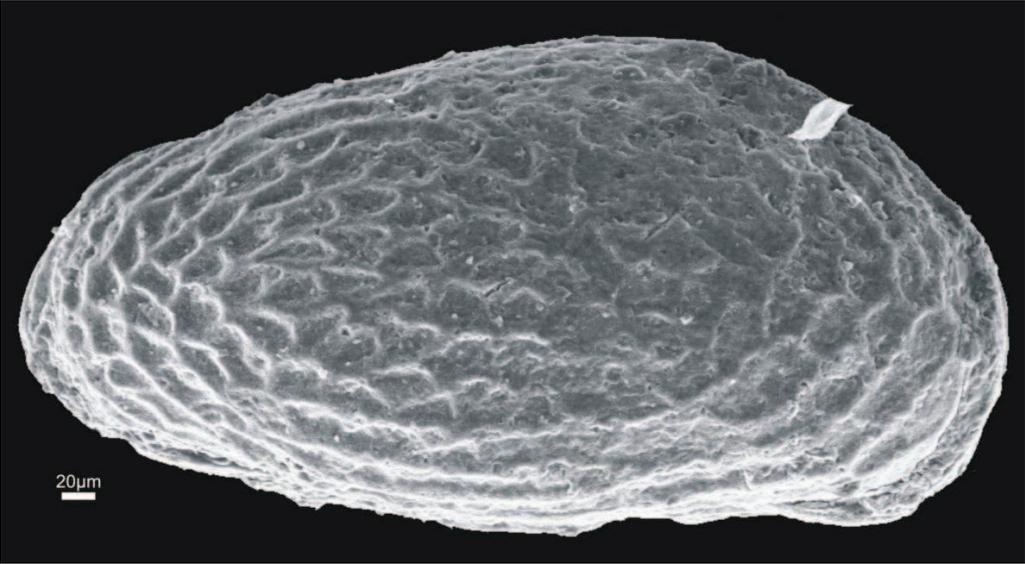


Imagem 5A:



## Atividade 2 – Classificar utilizando o conhecimento e funcionalidades oferecidas pelo sistema SisAC.

Essa atividade tem como objetivo avaliar o desempenho de usuários na utilização do sistema SisAC para a classificação de Ostracodes.

As seguintes opções devem ser preenchidas com seus respectivos espécimes, que são apresentados na seqüência. Como exemplo: **X ( YB )**, onde X é o nome da espécie/gênero e o Y é o número da imagem que representa o espécime.

Legitimocythere ( )

Jonesia ( )

Soudanella ( )

Cytherella ( )

Majungaella ( )

Imagem 1B:



Imagem 2B:



Imagem 3B:

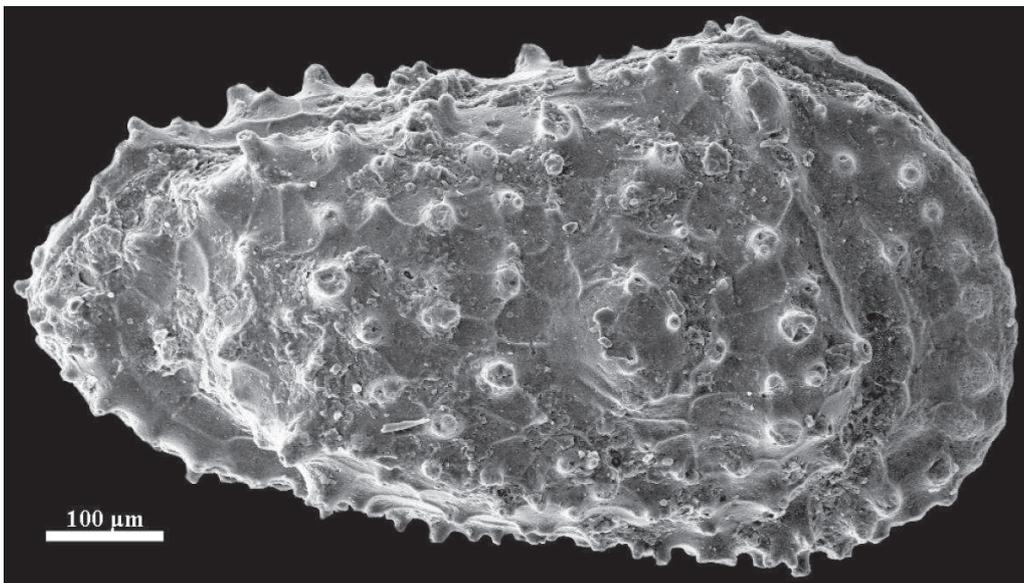
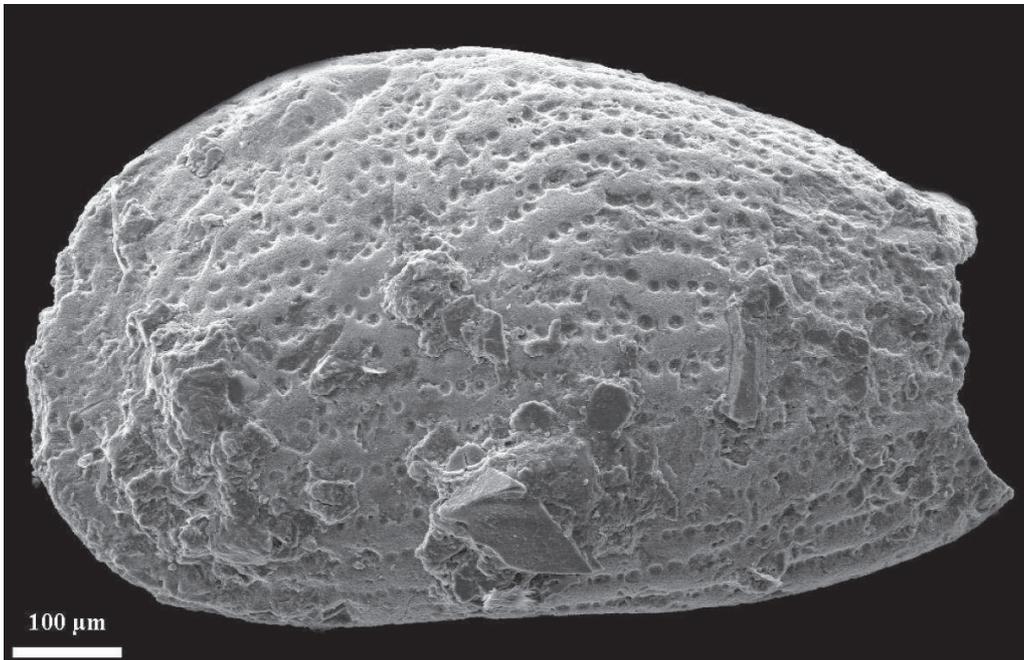


Imagem 4B



Imagem 5B:



### Atividade 3 – Questionário para avaliar o desempenho do sistema com ponto de vista do usuário

Esta atividade tem como objetivo verificar as críticas positivas e negativas em relação ao sistema e o desempenho que o sistema apresentou. Nessa atividade de resposta as seguintes questões marcando com um **X** na coluna/valor que melhor represente a sua opinião, sendo o valor 1 representado que você não concorda plenamente e 5 que você concorda plenamente.

	1	2	3	4	5
O sistema mostrou-se de simples em sua utilização?					
Acredita que o seu desempenho na classificação de Ostracodes aumentou com a utilização do sistema SisAC?					
O sistema apresentou todas as informações necessárias para um classificação de Ostracodes?					
As imagens auxiliaram no processo de classificação?					
As questões subjetivas em relação aos atributos que caracterizam um gênero/espécie foram reduzidas com a utilização do sistema?					

1 -> Não Concordo Plenamente

2 -> Não Concordo

3 -> Indiferente

4 -> Concordo

5 -> Concordo Plenamente

Em sua opinião você possui, marque com um **X** a resposta, o valor de conhecimento que você considere ter?

( ) Neófito, não possui nenhum conhecimento na área de domínio

( ) Conhecimento básico para identificar poucos espécimes

( ) Conhecimento médio para identificar alguns espécimes

( ) Conhecimento avançado para identificar diversos espécimes

( ) Especialista, possui conhecimento suficiente para identificar de forma fácil inúmeros espécimes

## **Atividade 4 – Questionário sobre observações e críticas em relação a atividade realizada**

Esta atividade tem como objetivo obter suas observações pessoais em relação a atividade e ao sistema, respondendo as questões com suas próprias palavras.

- a) Quais foram as maiores dificuldades durante a classificação de Ostracodes utilizando o seu conhecimento próprio?

---

---

---

---

---

---

- b) Quais foram as maiores dificuldades durante a classificação de Ostracodes utilizando o sistema SisAC?

---

---

---

---

---

---

- c) Por meio das funcionalidades oferecidas pelo sistema SisAC, como o sistema auxiliou no processo de classificação dos Ostracodes?

---

---

---

---

---

---

- d) Entre o processo de classificação de forma manual e com o sistema, qual das duas atividades consumiu mais tempo para a realização da mesma?

---

---

---

---

---

---



## APÊNDICE B RESPOSTAS DAS QUESTÕES DISSERTATIVAS

Nesse anexo são apresentadas as respostas de cada participantes de forma distintas segundo o grau de conhecimento que cada um acreditava possuir sobre o domínio de estudos trabalhado. Cada seção irá apresentar as respostas das perguntas presentes no questionário utilizado na validação do sistema.

### APÊNDICE B.1 QUAIS FORAM AS MAIORES DIFICULDADES DURANTE A CLASSIFICAÇÃO DE OSTRACODES UTILIZANDO O SEU CONHECIMENTO PRÓPRIO?

<b>Participante</b>	<b>Resposta</b>
<b>Especialista</b>	A subjetividade com que alguns autores descrevem os ostracodes
<b>Avançado A</b>	Falta de material físico e de outras vistas do espécime o que é o normal na rotina de identificação
<b>Avançado B</b>	Não houveram dificuldades na classificação uma vez que eram gêneros bem conhecidos
<b>Avançado C</b>	As maiores dificuldades na classificação de ostracodes esta no grande número de grupos e espécies sendo que as formas fósseis (as mais antigas) possuem péssima preservação, dificultando a identificação
<b>Intermediário</b>	Ter certeza sobre determinadas características analisadas (Atributos)
<b>Básico</b>	Inicialmente me interar do programa, para que houve-se uma melhor correlação dos dados que eu tinha sobe a organismo e o programa

APÊNDICE B.2 QUAIS FORAM AS MAIORES DIFICULDADES DURANTE A CLASSIFICAÇÃO DE OSTRACODES UTILIZANDO O SISTEMA SISAC?

<b>Participante</b>	<b>Resposta</b>
<b>Especialista</b>	A exata relação entre as características do espécime em estudo e as possibilidades listadas no programa
<b>Avançado A</b>	Atributos repetidos. Falta de alguns atributos. Objetividade exagerada na escolha dos atributos definidos das espécies.
<b>Avançado B</b>	Falta de características taxônomicas. Por exemplo, um espécime pode ser espinhoso e reticulado. Entretanto, considerando as dificuldades taxonômicas inerente aos ostracodes, a alimentação do banco de dados poderia auxiliar neste aspecto
<b>Avançado C</b>	As maiores dificuldades estão nas definições de cada atributo (conceitos) que podem variar de acordo com o usuário
<b>Intermediário</b>	Ter certeza sobre determinadas características analisadas (Atributos)
<b>Básico</b>	A relação de tamanho de algumas estruturas, como por exemplo tamanho das ornamentações

APÊNDICE B.3 POR MEIO DAS FUNCIONALIDADES OFERECIDAS PELO SISTEMA SISAC, COMO O SISTEMA AUXILIOU NO PROCESSO DE CLASSIFICAÇÃO DOS OSTRACODES?

<b>Participante</b>	<b>Resposta</b>
<b>Especialista</b>	Fornecendo um conjunto de opções ao pesquisador. Isto pode ser valioso para iniciantes.
<b>Avançado A</b>	Infelizmente o sistema não ajudou no processo de classificação, por já identificar o gênero analisando a imagem, a utilização do programa se tornou uma busca pelos atributos selecionados pelo programador para encontrar a espécie correta e em todas as tentativas esses atributos não foram os observáveis pela imagem fornecida para identificação
<b>Avançado B</b>	Foi possível visualizar as características agrupadas hierarquicamente
<b>Avançado C</b>	Acredito que o sistema organiza o conhecimento e pode auxiliar na identificação
<b>Intermediário</b>	Auxiliou "eliminando" características que não estavam de acordo com as imagens visualizadas
<b>Básico</b>	Através da redução de atributos desnecessários na classificação do organismo

APÊNDICE B.4 ENTRE O PROCESSO DE CLASSIFICAÇÃO DE FORMA MANUAL E COM O SISTEMA, QUAL DAS DUAS ATIVIDADES CONSUMIU MAIS TEMPO PARA A REALIZAÇÃO DA MESMA?

<b>Participante</b>	<b>Resposta</b>
<b>Especialista</b>	Para mim, o sistema "manual", contudo, deve-se considerar que já conhecia as espécies utilizadas no teste
<b>Avançado A</b>	Sistema. E em alguns casos não consegui encontrar a identificação correta com os dados do sistema
<b>Avançado B</b>	Considerando que tenho conhecimento anterior do grupo, a utilização do sistema consumiu mais tempo. Por outro lado, para iniciantes no estudo de Ostracodes pode ser muito útil
<b>Avançado C</b>	A manual
<b>Intermediário</b>	Acredito que as duas consumiram tempo parecido
<b>Básico</b>	Na forma manual, pois a obtenção de todas os parametros para classificação manual é mais demorada

APÊNDICE B.5 DESCREVA/COMENTE COMO AS IMAGENS AUXILIARAM OU PODERIAM AUXILIAR NO PROCESSO DE CLASSIFICAÇÃO POR MEIO DAS FUNCIONALIDADES?

<b>Participante</b>	<b>Resposta</b>
<b>Especialista</b>	As imagens podem exemplificar o que é uma determinada característica. Podem ainda levar ao pesquisador que ele inicialmente não consideraria.
<b>Avançado A</b>	As imagens foram uteis nas escolhas dos atributos
<b>Avançado B</b>	As imagens permitiram a visualização do significado de um determinado carácter morfológico e sua comparação com os demais
<b>Avançado C</b>	As imagens são chave para identificação sem elas o tempo para identificação irá aumentar
<b>Intermediário</b>	Podariam auxiliar melhor se houvesse um maior número de imagens para cada atributo
<b>Básico</b>	Através da visualização simultanea do exemplar estudado e o presente no programa

APÊNDICE B.6 DESCREVE AS DIFERENÇAS OBSERVADAS EM RELAÇÃO AS DIFICULDADES ENCONTRADAS NA CLASSIFICAÇÃO DE OSTRACODES UTILIZANDO O SEU CONHECIMENTO EM RELAÇÃO AO AUXÍLIO OFERECIDO PELO SISTEMA?

<b>Participante</b>	<b>Resposta</b>
<b>Especialista</b>	-
<b>Avançado A</b>	-
<b>Avançado B</b>	O sistema oferece na tela um conjunto de dados. Na rotina de pesquisa é necessário consultar uma série de livros e/ou artigos
<b>Avançado C</b>	O SisAC é um sistema mais rápido e ordena o conhecimento, facilitando a determinação
<b>Intermediário</b>	Com o sistema a classificação fica mais direta, não sendo necessária a consulta de outros meios (internet, livros)
<b>Básico</b>	A demora na busca dos dados e parâmetros que me levem a classificação do organismo

APÊNDICE B.7 ESCREVA SUAS CRÍTICAS, OBSERVAÇÕES, SUGESTÕES QUE POSSAM AUXILIAR E APRIMORAR O DESENVOLVIMENTO CONTÍNUO DO SISTEMA.

<b>Participante</b>	<b>Resposta</b>
<b>Especialista</b>	O sistema pode ser carregado com imagens mais específicas sobre determinadas características morfológicas. Estas poderiam ser tanto fatos como ilustrações de livros, por exemplo.
<b>Avançado A</b>	Acredito que o programa em si funciona corretamente para o que se pretende, penso que a escolha do tema (identificação de ostracodes) não foi a melhor para cumprir este objetivo, visto que nem mesmo entre os pesquisadores do grupo existe um conceito único sobre a sua taxonomia. O estudo dos ostracodes já está repleto de subjetividade, que são contornadas através de estudos mais minuciosos. Forçar uma classificação objetiva em cima disso é extremamente complicado, ainda mais tendo como base um estudo de espécime muito regionais e com uma preservação pobre como é o caso do atlas. Sugiro a utilização se for com ostracodes, de um estudo mais consagrado no meio científico, algum estudo clássico, e a identificação chegando no máximo ao nível genérico
<b>Avançado B</b>	Melhorar muito o banco de dados. O sistema é interessante. As dificuldades na classificação dos Ostracodes residem no fato de outros fatores subjetivos estarem implicados. No caso dos fósseis, temos ainda outros problemas advindos da preservação da integridade das formas. Acredito que as características devem ser organizadas de acordo com sua ordem de relevância. Seu projeto é ousado, te incentivo a seguir o aprimorando do banco de dados. Isso será possível através da interação com pesquisadores da área.
<b>Avançado C</b>	Sugestão aumentar os atributos (imagens) e descrições
<b>Intermediário</b>	O aprimoramento da base de dados seria bem vindo, já que congregaria outras espécies e atributos. A inserção de artigos científicos também seria de ajuda para o pesquisador interessado em utilizar o sistema.
<b>Básico</b>	Uma sugestão seria a acréscimo de mais imagens dos organismos para melhor comparação

## APÊNDICE C FOTOS TIRADAS DURANTE A AVALIAÇÃO DO SISTEMA

