

UNIVERSIDADE ESTADUAL PAULISTA

“Júlio de Mesquita Filho”

Pós-Graduação em Ciência da Computação

Filipe Marcel Fernandes Gonçalves

Uma Abordagem Interativa Guiada por Semântica para
Identificação e Recuperação de Imagens

UNESP

2016

Gonçalves, Filipe Marcel Fernandes.

Uma abordagem interativa guiada por semântica para identificação e recuperação de imagens / Filipe Marcel Fernandes Gonçalves. -- São José do Rio Preto, 2016

158 f. : il., tabs.

Orientador: Ivan Rizzo Guilherme

Coorientador: Daniel Carlos G. Pedronette

Dissertação (mestrado) – Universidade Estadual Paulista "Júlio de Mesquita Filho", Instituto de Biociências, Letras e Ciências Exatas

1. Computação - Matemática. 2. Processamento de imagens – Técnicas digitais. 3. Sistemas de recuperação da informação. 4. Web semântica. 5. Ontologias (Recuperação da informação) I. Universidade Estadual Paulista "Júlio de Mesquita Filho". Instituto de Biociências, Letras e Ciências Exatas.
II. Título.

CDU – 518.72:76

Ficha catalográfica elaborada pela Biblioteca do IBILCE
UNESP - Câmpus de São José do Rio Preto

Filipe Marcel Fernandes Gonçalves

Uma Abordagem Interativa Guiada por Semântica para
Identificação e Recuperação de Imagens

Orientador: Prof. Dr. Ivan Rizzo Guilherme

Co-Orientador: Prof. Dr. Daniel Carlos G. Pedronette

Dissertação de Mestrado elaborada junto ao Programa de Pós-Graduação em Ciência da Computação – Área de Concentração em Matemática e Inteligência Computacional, como parte dos requisitos para a obtenção do título de Mestre em Ciência da Computação.

UNESP

2016

Filipe Marcel Fernandes Gonçalves

Uma Abordagem Interativa Guiada por Semântica para
Identificação e Recuperação de Imagens

Dissertação de Mestrado elaborada junto ao Programa de Pós-Graduação em Ciência da Computação – Área de Concentração em Matemática e Inteligência Computacional, como parte dos requisitos para a obtenção do título de Mestre em Ciência da Computação.

Comissão Examinadora

Prof. Dr. Ivan Rizzo Guilherme
DEMAC – UNESP – Rio Claro
Orientador

Prof. Dr. João Paulo Papa
Departamento de Computação – UNESP - Bauru

Prof. Dr. Jurandy Gomes de Almeida Junior
ICT – UNIFESP – São José dos Campos

UNESP

2016

Aos meus pais,
meus avós e
meu irmão

Agradecimentos

Primeiramente agradeço aos meus orientadores Ivan Rizzo Guilherme e Daniel Carlos Guimarães Pedronette pela oportunidade, confiança, amizade, pelos conhecimentos transmitidos e por toda a orientação no decorrer desses anos que estivemos juntos. Agradeço também ao professor Ivan pelas oportunidades e trabalhos que desenvolvemos desde 2012.

Esse trabalho não teria sido possível se não tivéssemos sucessivas reuniões e discussões sobre o tema e também sobre n outras coisas.

Agradeço também ao professor Alexandro José Baldassin pela amizade, confiança e pela oportunidade de termos trabalhado juntos nos anos de 2011 e 2012.

Posso dizer que, além de meus professores, vocês são meus amigos.

Agradeço à CAPES (Coordenação de Aperfeiçoamento Profissional de Nível Superior) pelo suporte prestado no desenvolvimento desse projeto e no decorrer desses anos de mestrado.

Agradeço também aos professores da graduação e da pós, pelas aulas e pelo conhecimento que nos foi transmitido.

Obrigado ao meu amigo Rodrigo pelas inúmeras discussões e auxílio na modelagem da ontologia.

Agradeço também ao Lucas Valem pelo auxílio na extração de características visuais das imagens.

Gostaria de agradecer também à minha amiga de sala quando cursei Biologia, Mariana Naomi Saka (Mangá), pelo auxílio na modelagem da ontologia na questão da Biologia. Você é fera demais japa. Valeu mesmo.

Agradeço também ao meu amigo e irmão Renato (Bill) pelas conversas, discussões e por toda a amizade que temos. Valeu Bill, você é sensacional!

Também gostaria de agradecer aos meus amigos da turma de Ciências Biológicas, integral – 2005, por todo carinho e amizade que temos.

Agradeço também todos os meus amigos da turma de Ciências da Computação, Noturno – 2010, mas em especial ao Renato (Japonês), Matheuzão, Lucas, Fred, Xandão, Diogo (Grilo), Rafael (Brutus), Carina, Waldeilson, Gui, Daniel (Savoia), Aurélio, Carol e Ju. A graduação foi demais! E as noites de *Guitar Hero* foram sensacionais.

Também agradeço aos meus amigos do mestrado em Ciência da Computação, mas em especial ao Daniel, Diego, Leonardo e Rafa pela amizade e pelas discussões e ajudas no desenvolvimento dos trabalhos.

Agradeço também aos meus amigos e irmãos de Rio Claro, por todos esses anos de amizade.

Obrigado também à UNESP, principalmente o campus de Rio Claro. Universidade essa que tenho enorme carinho, já que foram 12 anos entre cursos, nessa que posso chamar de minha segunda casa.

Finalmente, agradeço à toda minha família pelo apoio em todos esses anos de estudo. Agradeço à minha prima Vanessa e ao Harri Lorenzi pela ajuda e pelas imagens cedidas.

Agradeço imensamente aos meus pais: Valdir e Maria por tudo que fizeram por mim até hoje. Sempre me apoiando nos estudos e em tantas outras áreas. Obrigado pelas orações, amor, carinho, apoio, confiança e tudo aquilo que sempre me proporcionaram. Sou eternamente grato a tudo isso e tenham certeza de que cada conquista não seria possível sem vocês.

Agradeço também ao meu irmão Lucas (Quinhas). A amizade, o amor e o carinho que temos um pelo outro é algo sensacional doutor, você é demais, cara.

Amo vocês.

Muito Obrigado!

Together we stand, divided we fall

Hey You – Pink Floyd

Sumário

1. Introdução.....	16
2. Trabalhos Relacionados.....	22
2.1. Aplicações de Recuperação de Imagens de Plantas.....	22
2.2. Recuperação de Imagens – Conteúdo e Interpretação.....	24
2.3. Abordagens Semânticas para Recuperação de Imagens.....	26
2.4. Recuperação de Imagens com Auxílio de Usuários.....	30
3. Recuperação de Imagens Baseada no Conteúdo.....	33
3.1. Modelo de Recuperação.....	33
3.2. Modelo Formal de Recuperação de Imagens.....	34
3.3. Extração das Características Visuais.....	35
3.3.1. LIRe.....	36
3.3.2. Características Visuais.....	38
3.3.3. <i>Deep Learning</i> em Sistemas CBIR.....	40
3.4. Pós-Processamento Baseado em Aprendizado Não Supervisionado.....	41
3.4.1. Aprendizado Não Supervisionado – RL-Sim.....	43
3.4.1.1. Métrica de Distância Contextual Baseada em Ranqueamento.....	44
3.4.1.2. Impacto do Pós-processamento nas Listas de Resultados.....	46
3.4.2. <i>Rank Aggregation</i>	50
4. Ontologia.....	51
4.1. Web Semântica.....	51
4.2. Ontologias – Definições.....	53
4.2.1. OWL.....	55
4.2.2. Methontology.....	56
4.2.3. Protégé.....	59
4.3. Especificação e Conceitualização do Conhecimento.....	59
4.3.1. Sistemática Vegetal.....	60
4.4. Características da Ontologia.....	65

4.5. Modelagem da Ontologia.....	66
4.5.1. Estruturas das Plantas.....	72
4.5.2. Propriedades das Plantas.....	77
5. Abordagem Interativa Guiada por Semântica para Identificação e Recuperação de Imagens.....	81
5.1. Visão Geral da Abordagem Proposta.....	81
5.2. Ontologia.....	84
5.3. Grafo Bipartido da Ontologia (BOG).....	85
5.4. Histogramas dos <i>Rankings</i> das Famílias.....	87
5.5. Grafo de Atributos Discriminativos (DAG).....	88
5.6. Seleção Semântica de Atributos.....	91
5.7. Recuperação Interativa de Imagens.....	93
5.8. Análise Visual da Abordagem Proposta.....	94
5.9. Considerações da Abordagem.....	97
6. Avaliação Experimental.....	99
6.1. Protocolo Experimental.....	99
6.2. Conjuntos de Dados (<i>Datasets</i>).....	102
6.3. Características Visuais e Aprendizado Não Supervisionado.....	104
6.3.1. Resultados <i>Oxford Flowers – 17 Classes</i>	104
6.3.2. Resultados <i>Oxford Flowers – 102 Classes</i>	111
6.4. Recuperação Interativa de Imagens.....	115
6.4.1. Resultados <i>Oxford Flowers – 17 Classes</i>	116
6.4.2. Resultados <i>Oxford Flowers – 102 Classes</i>	126
7. Conclusão e Trabalhos Futuros.....	135
8. Referências Bibliográficas.....	139
9. Apêndice A – Análises Estatísticas.....	153

Lista de Figuras

Figura 1 – Visão Geral da Abordagem Proposta.....	20
Figura 2 – Planta da Família “Heliconiaceae”.....	23
Figura 3 – Filotaxia.....	30
Figura 4 – <i>Ranked Lists</i> ‘0’ e ‘1’, SURF, Sem RL-Sim.....	47
Figura 5 – <i>Ranked Lists</i> ‘0’ e ‘1’, Sem RL-Sim, com as Imagens Corretas.....	47
Figura 6 – <i>Ranked Lists</i> ‘0’ e ‘1’, SURF, Com RL-Sim.....	49
Figura 7 – <i>Ranked Lists</i> ‘0’ e ‘1’, Com RL-Sim, com as Imagens Corretas.....	49
Figura 8 – Arquitetura da Web Semântica.....	52
Figura 9 – Processo de Desenvolvimento de Ontologias – Methontology.....	58
Figura 10 – Diferença Visual Entre Duas Plantas da Família “Acanthaceae”..	62
Figura 11 – Exemplo de Chave Dicotômica.....	63
Figura 12 – Planta da Família “Araceae”.....	67
Figura 13 – Modelagem da Flor Aclamídea para a Família “Araceae”.....	69
Figura 14 – Forma de Vida Vegetal.....	70
Figura 15 – Modelagem do Gineceu para a Família “Araceae”.....	71
Figura 16 – Estrutura do Pistilo.....	73
Figura 17 – Grafo de Flor, até Pistilo.....	74
Figura 18 – Classes de Estruturas.....	76
Figura 19 – Relacionamento de “Folha” com “Propriedades de Folha”.....	77
Figura 20 – Classes de Propriedades.....	79
Figura 21 – Fluxo do Sistema Proposto.....	83
Figura 22 – Cálculo da Matriz O	86
Figura 23 – Histograma para um <i>Ranked List</i> Hipotético.....	88
Figura 24 – Cálculo da Matriz D	90
Figura 25 – Matriz D Completa Após os Respetivos Cálculos.....	90
Figura 26 – Cálculo da Adjacência Acumulada.....	92
Figura 27 – Escolha da Pergunta para o Usuário.....	92
Figura 28 – Comportamento de um <i>R.L.</i> Após a Aplicação do SIIR.....	95
Figura 29 – Exemplos de Classes do <i>dataset Oxford 17 Classes</i>	103

Lista de Gráficos

Gráfico 1 – P@10 e MAP para o ACC no <i>dataset 17 Classes</i>	107
Gráfico 2 – P@10 e MAP para o SURF no <i>dataset 17 Classes</i>	107
Gráfico 3 – P@10 e MAP para o BIC no <i>dataset 17 Classes</i>	107
Gráfico 4 – P@10 e MAP para o CNN no <i>dataset 17 Classes</i>	108
Gráfico 5 – P@10 e MAP para o ACC+SURF no <i>dataset 17 Classes</i>	108
Gráfico 6 – P@10 e MAP para o CNN+BIC no <i>dataset 17 Classes</i>	108
Gráfico 7 – P@10 e MAP para o CNN+SURF no <i>dataset 17 Classes</i>	109
Gráfico 8 – P@10 e MAP para o SURF no <i>dataset 102 Classes</i>	113
Gráfico 9 – P@10 e MAP para o BIC no <i>dataset 102 Classes</i>	114
Gráfico 10 – P@10 e MAP para o CNN no <i>dataset 102 Classes</i>	114
Gráfico 11 – P@10 e MAP para o CNN+BIC+SURF no <i>dataset 102 Classes</i>	114
Gráfico 12 – <i>Precision</i> × <i>Recall</i> para ACC e SURF com SIIR – <i>17 Classes</i> .	117
Gráfico 13 – <i>Precision</i> × <i>Recall</i> para SURF, BIC e CNN com SIIR <i>17 Classes</i>	118
Gráfico 14 – P@5 para os métodos de extração com o SIIR – <i>17 Classes</i>	119
Gráfico 14.A (Apêndice A) – Intervalo de Confiança do P@5 para os métodos de extração com o SIIR – <i>17 Classes</i>	154
Gráfico 15 – P@10 para os métodos de extração com o SIIR – <i>17 Classes</i> ..	121
Gráfico 15.A (Apêndice A) – Intervalo de Confiança do P@10 para os métodos de extração com o SIIR – <i>17 Classes</i>	154
Gráfico 16 – MAP para os métodos de extração com o SIIR – <i>17 Classes</i>	123
Gráfico 16.A (Apêndice A) – Intervalo de Confiança do MAP para os métodos de extração com o SIIR – <i>17 Classes</i>	155
Gráfico 17 – <i>Precision</i> × <i>Recall</i> para SURF, BIC e CNN com SIIR <i>102 Classes</i>	127
Gráfico 18 – P@5 para os métodos de extração com o SIIR – <i>102 Classes</i> ..	128
Gráfico 18.A (Apêndice A) – Intervalo de Confiança do P@5 para os métodos de extração com o SIIR – <i>102 Classes</i>	156
Gráfico 19 – P@10 para os métodos de extração com o SIIR – <i>102 Classes</i>	129

Gráfico 19.A (Apêndice A) – Intervalo de Confiança do P@10 para os métodos de extração com o SIIR – <i>102 Classes</i>	156
Gráfico 20 – MAP para os métodos de extração com o SIIR – <i>102 Classes</i> ...	131
Gráfico 20.A (Apêndice A) – Intervalo de Confiança do MAP para os métodos de extração com o SIIR – <i>102 Classes</i>	157

Lista de Tabelas

Tabela 1 – Descritores com Melhor Precisão de Classificação.....	39
Tabela 2 – P@5, Métodos de Extração, RL-Sim – <i>Oxford 17 Classes</i>	105
Tabela 3 – P@5, Métodos de Extração, RL-Sim – <i>Oxford 102 Classes</i>	112
Tabela 4 – Acurácia - classificação 20-NN famílias – <i>Oxford 17 Classes</i>	124
Tabela 5 – Acurácia – classificação 20-NN. Abordagem proposta comparada com métodos do estado-da-arte – <i>dataset Oxford 17 Classes</i>	125
Tabela 6 – Acurácia - classificação 20-NN famílias – <i>Oxford 102 Classes</i>	132
Tabela 7 – Acurácia – classificação 20-NN. Abordagem proposta comparada com métodos do estado-da-arte – <i>dataset Oxford 102 Classes</i>	132

Resumo

O grande volume de imagens disponível na Web gerado em diferentes domínios requer um conhecimento especializado para sua análise e identificação. Nesse sentido, recentes avanços ocorreram com desenvolvimento de técnicas de recuperação de imagens baseadas nas características visuais. Entretanto, o *gap* semântico entre as características de baixo-nível das imagens e aquilo que a imagem representa ainda é um grande desafio. Uma solução para diminuir o *gap* semântico consiste em combinar a informação de características visuais das imagens com o conhecimento do domínio de tais imagens. Nesse sentido, ontologias podem auxiliar, já que estruturam o conhecimento. Desse modo, o presente trabalho apresenta uma nova abordagem denominada Recuperação Interativa de Imagens Guiada por Semântica (*Semantic Interactive Image Retrieval – SIIR*) que combina técnicas de recuperação de imagens baseadas no conteúdo (*Content Based Image Retrieval – CBIR*) e aprendizado não supervisionado, com o conhecimento definido em ontologias. Desse modo, o trabalho em questão propõe uma nova abordagem a fim de simular o papel dos biólogos na classificação de famílias de Angiospermas a partir de uma imagem e seu conteúdo. Para tanto, foi desenvolvida uma ontologia de estruturas e propriedades de plantas com flor e fruto, de modo a conceitualizar e relacionar tais atributos visando a classificação de famílias de Angiospermas. Para análise das características visuais foram utilizados métodos de extração de características de baixo-nível das imagens. Com relação ao aprendizado não supervisionado foi utilizado o algoritmo RL-Sim a fim de melhorar a eficácia da recuperação das imagens. A abordagem combina técnicas CBIR com ontologias ao utilizar um grafo bipartido e um grafo discriminativo de atributos. O grafo discriminativo de atributos permite a análise semântica utilizada para selecionar o atributo que melhor classifica a planta da imagem de busca. Os atributos selecionados são utilizados para formular uma interação com um usuário, de modo a melhorar a eficácia da recuperação e diminuir os esforços necessários na identificação da planta. O método proposto foi avaliado nos conjuntos de dados públicos *Oxford Flowers 17* e *102 Classes*, de modo que os resultados demonstram alta

eficácia para ambos os conjuntos de dados quando comparados com outras abordagens.

Palavras-chave: Aprendizado não supervisionado, *Gap Semântico*, Ontologias, CBIR, Angiospermas.

Abstract

A large amount of images is currently generated in many domains, thus requiring specialized knowledge on the identification and analysis. From one standpoint, many advances have been accomplished in the development of image retrieval techniques based on visual image properties. However, the semantic gap between low-level features and high level concepts still represents a challenge scenario. On another standpoint, knowledge has also been structured in many fields by ontologies. A promising solution for bridging the semantic gap consists in combining the information from low-level features with semantic knowledge. This work proposes a new approach denominated Semantic Interactive Image Retrieval (SIIR) which combines Content Based Image Retrieval (CBIR) and unsupervised learning with ontology techniques. We present a novel approach aiming to simulate the biologist's role in the classification of Angiosperm families from image sources and their content. In order to achieve this goal, we developed a domain ontology from plant properties and structures, hence relating features from the Angiosperm families. In regard to Unsupervised Learning, we used the RL-Sim algorithm to improve image classification. The proposed approach combines CBIR techniques with ontologies using a bipartite graph and a discriminative attribute graph. Such graph structures allow a semantic analysis used for the selection of the attribute that best classifies the plant. The selected attributes are used for formulating the user interactions, improving the effectiveness and reducing the user efforts required. The proposed method was evaluated on the popular Oxford Flowers 17 and 102 Classes datasets, yielding very high effectiveness results in both datasets when compared to other approaches.

Keywords: Unsupervised Learning. Semantic Gap. Ontology. CBIR. Angiosperms.

1. Introdução

A crescente disponibilidade de acesso a imagens através de diferentes tecnologias demanda o desenvolvimento de novos métodos de recuperação e reconhecimento. Nesse cenário, várias técnicas de processamento de imagens e recuperação de informações têm sido desenvolvidas e aplicadas ao conteúdo de mídia digital (ARVOR *et al.*, 2013). Nesse sentido, destacam-se a aplicação de duas abordagens consideradas antes distintas: a recuperação de imagens baseada em características de baixo-nível (VELTKAMP *et al.*, 2001; DATTA *et al.*, 2008; KURTZ *et al.*, 2014) e a recuperação baseada no conhecimento de um determinado domínio da imagem (BANDIKOLLA *et al.*, 2014; LI, 2015; MANZOOR *et al.*, 2015; PANDEY *et al.*, 2015).

Por um lado têm-se técnicas baseadas em sistemas de recuperação de imagens baseados no conteúdo (*Content-Based Image Retrieval* – CBIR), que utilizam características quantitativas extraídas por descritores visuais capazes de recuperar e indexar imagens (VELTKAMP *et al.*, 2001; DATTA *et al.*, 2008). Em geral, cada imagem é representada em termos de características quantitativas (KURTZ *et al.*, 2014) que são exploradas com o objetivo de se comparar com outras imagens e podem ser aplicadas também para abordagens de ranqueamento (LIU *et al.*, 2007a; PEDRONETTE e TORRES, 2013). Sistemas CBIR consideram várias propriedades visuais de uma imagem, tais como forma, textura e cor, extraídas a partir de métodos globais ou locais (VELTKAMP *et al.*, 2001; DATTA *et al.*, 2008; KURTZ *et al.*, 2014). Recentemente, redes neurais convolucionais (*Convolutional Neural Networks* – CNN) têm sido aplicadas para os mesmos objetivos com resultados significantes (HOI *et al.*, 2010; JIA *et al.*, 2014; RAZAVIAN *et al.*, 2014).

Mais recentemente, o aprendizado não supervisionado tem atraído a atenção da comunidade de recuperação de imagens e informações, já que tais métodos exploram a estrutura de conjuntos de dados de imagens para melhorar a eficácia de tarefas de recuperação de imagens. Abordagens baseadas em ranqueamento têm sido propostas de modo a alcançar ganhos significativos na precisão da classificação e recuperação de imagens

(PEDRONETTE e TORRES, 2013; PEDRONETTE *et al.*, 2014; BAI *et al.*, 2015; BAI e BAI; 2016).

Por outro lado, abordagens recentes baseadas no conhecimento de um determinado domínio da imagem, também demonstram eficácia no estudo de recuperação e identificação de imagens (BANDIKOLLA *et al.*, 2014; LI, 2015; MANZOOR *et al.*, 2015; PANDEY *et al.*, 2015). Nesse sentido, ontologias têm sido amplamente aplicadas como técnicas de representação e reuso do conhecimento, já que transcrevem um entendimento comum para uma área específica, declaram semântica explícita, realizam declarações significativas e suportam o compartilhamento de informações, além de relacionar diversos conceitos (GRUBER, 1993; GUARINO, 1998; LACY, 2005). Entretanto, apesar dos recentes avanços, ainda há uma grande disparidade entre técnicas que utilizam características quantitativas e técnicas que utilizam o conhecimento representado nas ontologias.

Apesar do desenvolvimento contínuo da extração de características visuais e do aprendizado supervisionado e não supervisionado, a recuperação de imagens relevantes ao usuário ainda constitui um grande desafio. O principal ponto de tal desafio é relacionar a informação semântica do domínio da imagem com os valores numéricos das características de baixo-nível recuperadas pelos algoritmos de reconhecimento de padrões. Esse problema se refere à lacuna (*gap*) semântica, que é definida pela falta de coincidência entre a informação que pode ser extraída de uma determinada imagem e a interpretação desse mesmo dado para o usuário em uma determinada situação (SMEULDERS *et al.*, 2000; DATA *et al.*, 2008). A diminuição do *gap* semântico ainda permanece como um dos maiores desafios a serem enfrentados na área de CBIR, afetando diretamente a eficácia da recuperação.

Além dos avanços nas extrações das características visuais e do conhecimento, houve também avanços em outras linhas de pesquisa que também analisam a recuperação de imagens. Abordagens que exploram interações com o usuário por meio de métodos de aprendizado também têm sido aplicadas com o intuito de aumentar a eficácia desses sistemas (LIU *et al.*, 2007b; CHENG *et al.*, 2009; THOMEE e LEW, 2012).

Salton e McGill (1998) afirmam que, sistemas baseados em recuperação de informação podem apresentar melhor desempenho ao se aplicar mecanismos de interação com usuários. Tal abordagem tem sido utilizada em sistemas CBIR de modo a aperfeiçoar as métricas de similaridade aplicadas à recuperação de imagens (GIACINTO, 2007). Entre outras diversas técnicas, uma combinação linear pode ser utilizada a fim de se alterar o peso de similaridades entre as imagens de modo a melhorar a eficácia da classificação e recuperação de imagens de um determinado sistema (GIACINTO, 2007).

Outra abordagem capaz de diminuir a lacuna semântica são técnicas baseadas em anotação de imagens, já que apresentam uma boa ferramenta para abstrair o que uma determinada imagem apresenta. Porém, anotações de imagens são suscetíveis à inconsistência entre uma entrada textual do usuário e aquilo que foi anotado ou descrito na imagem (ZHANG *et al.*, 2005). Nesse sentido, o uso de ontologias descrevendo um vocabulário rico, controlado e específico para um determinado domínio pode auxiliar no problema em questão. Ainda, ao selecionar atributos de interesse presentes na imagem para análise de um usuário, restringe-se o problema de uma provável avaliação errônea como entrada textual ao realizar uma busca.

Assim, fica claro que a análise de imagens e sua identificação requerem conhecimento especializado em diferentes domínios. Porém, alguns domínios do conhecimento carecem de informações relacionadas e de análise de imagens de coletas ou de espécimes de estudo. Nesse sentido, pode-se citar a linha de Ciências Biológicas, em que se destaca o estudo de Botânica, mais especificamente a área de Sistemática Vegetal, que estuda a organização das plantas em grupos, baseando-se em suas relações evolutivas (SOUZA e LORENZI, 2005).

Amostras e imagens de plantas são analisadas por meio de descrições sistemáticas, o que permite a identificação desses organismos e sua classificação em grupos separados. A identificação de Angiospermas (plantas com flor e fruto) requer um vasto conhecimento das estruturas e propriedades de um espécime analisado (SOUZA e LORENZI, 2007). A tarefa de identificação é ainda mais desafiadora quando realizada por meio de análise de

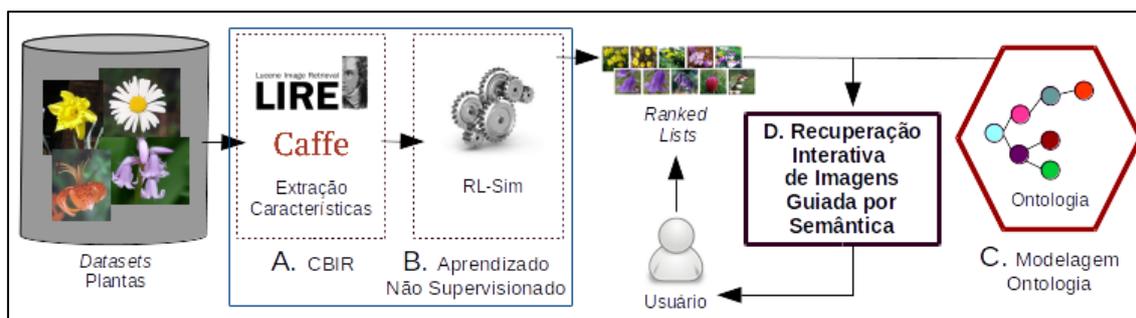
imagens, já que muitas estruturas e áreas das plantas estão ocultas. Nilsback e Zisserman (2006) afirmam que a classificação de imagens de ramos floríferos é difícil até mesmo para humanos, diferentemente de categorizações mais abrangentes como, por exemplo, classificar imagens de carros ou bicicletas. Desse modo, há um grande esforço para representar o conhecimento de muitas áreas em ontologias permitindo, assim, o processamento tanto por humanos quanto por máquinas.

Nesse sentido, o presente trabalho apresenta uma nova abordagem proposta baseada em recuperação interativa de imagens com o objetivo principal de suavizar o *gap* semântico para a identificação de plantas. O método proposto intitulado Recuperação Semântica Interativa de Imagens (*Semantic Interactive Image Retrieval* – SIIR), consiste em um módulo interativo que combina técnicas CBIR e de aprendizado não supervisionado (baseado em ranqueamento) com o conhecimento descrito em ontologias. A integração entre tais técnicas distintas é guiada pela recuperação de imagens e pela ontologia codificada pelo especialista de domínio. Desse modo, um grafo bipartido da ontologia e um grafo de atributos discriminativo são utilizados com o intuito de selecionar as características mais informativas capazes de diferenciar a planta da imagem de busca daquelas recuperadas com base nas características de baixo-nível.

Como o domínio da aplicação desenvolvida neste trabalho é referente à Sistemática Vegetal, o presente trabalho se propõe a identificar famílias de Angiospermas a partir de uma imagem de busca. Nesse sentido, um usuário envia uma foto de uma determinada planta para o sistema e, após um estágio de pós-processamento com a aplicação de um algoritmo de aprendizado não supervisionado, o usuário responde algumas perguntas referentes aos atributos que mais diferenciam a planta presente na foto enviada, das demais presentes nas posições iniciais do ranque gerado pelo sistema. Vale citar o fato de que a abordagem proposta pode ser aplicada em qualquer domínio, desde que haja uma ontologia que defina os conceitos e informações referentes ao conjunto de imagens analisadas.

A Figura 1 ilustra, de um modo geral, a organização da abordagem proposta e suas principais contribuições. Desse modo, são discutidas abaixo as principais contribuições do presente trabalho ao apresentar tal abordagem.

Figura 1: Visão geral da abordagem proposta.



- A. CBIR e Extração de Características:** (i) extração das características visuais utilizando abordagens recentes baseadas em CBIR e em *Deep Learning* (LIRe (LUX, 2013) e *CaffeNet* (JIA et al., 2014)); (ii) avaliação de várias características e identificação das mais efetivas para a tarefa de recuperação das imagens de plantas;
- B. Aprendizado Não Supervisionado:** (iii) utilização e avaliação de um métodos de ranqueamento baseado em aprendizado não supervisionado (RL-Sim (PEDRONETTE e TORRES, 2013)) na recuperação de imagens de plantas a fim de melhorar a eficácia dos resultados iniciais;
- C. Modelagem da Ontologia:** (iv) desenvolvimento da ontologia aplicada à classificação de famílias de Angiospermas;
- D. Recuperação Semântica Interativa de Imagens:** (v) a interação entre os resultados de saída do aprendizado não supervisionado e a codificação do conhecimento definido na ontologia; (vi) a proposta de uma abordagem baseada em grafos para a recuperação semântica interativa de imagens, em que um usuário responde perguntas selecionadas a partir de uma análise das estruturas morfológicas e propriedades das plantas.

A abordagem proposta foi avaliada por meio de experimentos em dois conjuntos públicos de imagens: *Oxford Flowers 17 Classes* e *Oxford Flowers 102 Classes*. Os resultados de tais experimentos demonstraram melhoras significativas na eficiência da recuperação dos resultados para ambas as bases de dados (*datasets*) quando comparados com outras abordagens descritas na literatura.

Assim, o presente trabalho encontra-se organizado conforme segue: o Capítulo 2 apresenta uma revisão bibliográfica e trabalhos relacionados ao tema de pesquisa. A fim de facilitar a explicação da abordagem proposta e dos aspectos conceituais envolvidos em cada um dos estágios de sua arquitetura, os Capítulos 3 e 4 definem os módulos de cada fluxo de dados presentes na Figura 1. De modo que no Capítulo 3 são definidos todos os aspectos conceituais e metodologias referentes à análise das imagens estudadas (Figura 1-A, Figura 1-B, Figura 1 – *ranked lists*), ou seja, o Capítulo 3 trata do fluxo da esquerda para a direita; e no Capítulo 4 são definidos os conceitos e a modelagem da ontologia (Figura 1-C) e parte dos principais conceitos necessários para a realização da modelagem, ou seja, o fluxo da direita para a esquerda da Figura 1. Já no Capítulo 5 são definidas as descrições minuciosas dos passos da abordagem proposta (Figura 1-D). O Capítulo 6 traz a avaliação experimental e os resultados envolvidos quando aplicado o novo método proposto e, finalmente, o Capítulo 7 apresenta a conclusão e os trabalhos futuros. Além desses capítulos, o presente trabalho traz o Apêndice A, referente à análise estatística dos intervalos de confiança dos gráficos apresentados no Capítulo 6.

2. Trabalhos Relacionados

Este Capítulo traz uma revisão bibliográfica de trabalhos relacionados ao tema do presente projeto, além de definições e dificuldades impostas pela Sistemática Vegetal ao ser utilizada com a abordagem proposta.

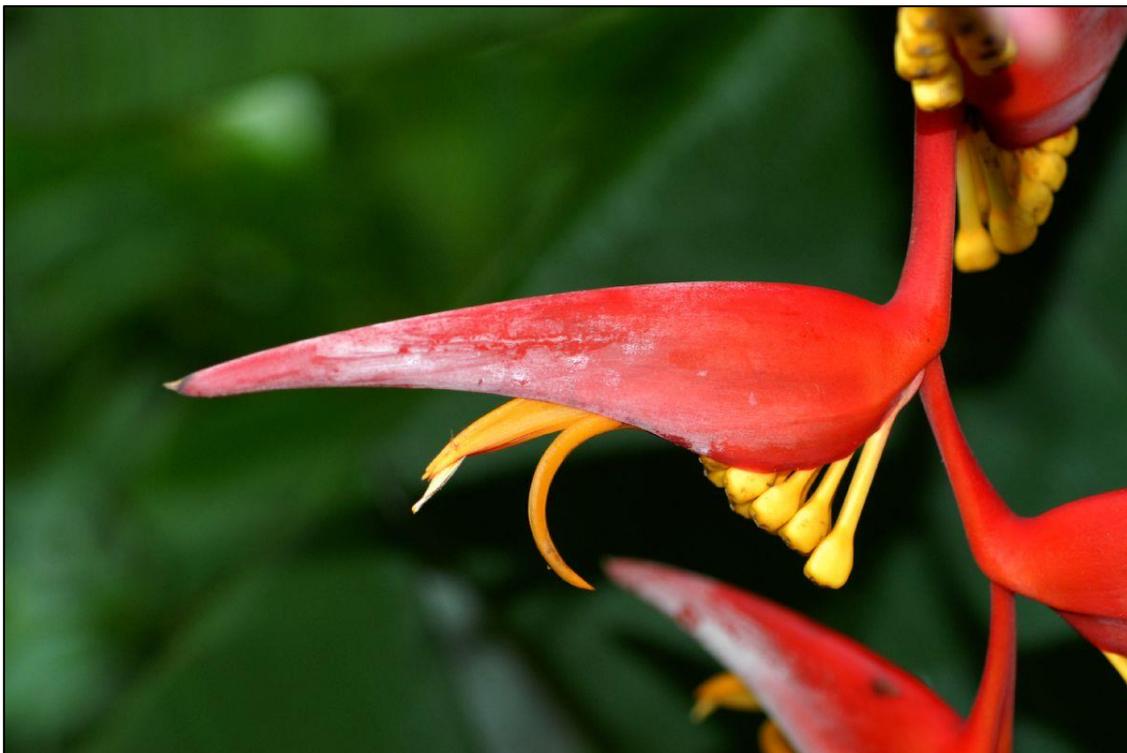
2.1 Aplicações de Recuperação de Imagens de Plantas

A identificação automática de espécies por meio do conteúdo digital é um desafio há anos, de modo que é importante citar o porquê da escolha da abordagem proposta. Devido ao fato da Botânica – mais precisamente da Morfologia Vegetal – apresentar muitas estruturas pequenas (em alguns casos tais estruturas estão ocultas, sendo necessário realizar uma análise de algum componente interno à flor), uma imagem de um ramo florífero geralmente não deixa visível tais partes com clareza (NILSBACK e ZISSERMAN, 2006). Ainda, algumas estruturas são facilmente confundidas com outras, já que são visualmente semelhantes.

Nilsback e Zisserman (2006), afirmam que a classificação de flores é uma tarefa muito difícil, mesmo para humanos, e não é nada semelhante a uma classificação mais abrangente, como a diferenciação de um carro e de uma bicicleta.

Um bom exemplo para a problemática citada acima é o caso da família “Heliconiaceae” (Figura 2), em que seus representantes podem apresentar uma bráctea vermelho-vivo (SOUZA e LORENZI, 2005). Brácteas são folhas usualmente modificadas que ocorrem no eixo floral, muitas vezes com forma, cor ou textura diferentes das folhas fotossintéticas. Algumas brácteas podem ser bastante vistosas e mais chamativas que pétalas e/ou sépalas (GONÇALVES e LORENZI, 2007).

Figura 2: Planta da família “Heliconiaceae”, destacando a bráctea vermelha e vistosa *.



Desse modo, a bráctea pode ser facilmente confundida com uma pétala ou com uma sépala por um indivíduo leigo em botânica, ou pela máquina, dependendo do tipo de abordagem utilizada para a avaliação.

Nilsback e Zisserman (2006, 2007, 2008 e 2009) demonstraram várias abordagens de segmentação de imagens de flores para se trabalhar na identificação de imagens de Angiospermas ao utilizar o conjunto de dados *Oxford Flowers 17 Classes* e *Oxford Flowers 102 Classes*.

Mattos (*et al.*, 2014) também realizaram trabalhos com imagens de Angiospermas ao utilizar o *dataset Oxford Flowers 102 Classes*. A fim de melhorar a precisão e a eficácia da classificação e recuperação de imagens de flores, Mattos (*et al.*, 2014) propuseram uma arquitetura que se baseia em uma interação com um usuário para realizar um recorte da imagem na região em que se encontra a flor.

* Fonte: http://www.plantsystematics.org/imgs/dws/r/Heliconiaceae_Heliconia_collinsiana_33753.html

Apesar dos avanços nas pesquisas com imagens de plantas, tanto Nilsback e Zisserman quanto Mattos não relacionaram tais análises com conceitos de alto-nível representados por um vocabulário controlado.

Um grande problema na identificação de plantas por imagens reside no fato de que não é de interesse identificar simplesmente imagens com flores, mas sim identificar imagens com flores de um determinado tipo.

Como a identificação de flores envolve a análise de seus verticilos florais, como pétalas e sépalas, além de sua simetria ou da disposição de algumas estruturas (como estames e gineceu) (GONÇALVES e LORENZI, 2007) e, tais imagens de plantas não demonstram algumas dessas estruturas com clareza, a escolha de uma abordagem que identifique automaticamente, por exemplo, imagens de plantas com flores monoclamídeas (flores que possuem somente um dos verticilos vegetativos (cálice – estrutura cujas unidades são representadas por sépalas; ou corola – estrutura cujas unidades são representadas por sépalas)), se torna muito mais desafiadora e de difícil desenvolvimento.

2.2 Recuperação de Imagens – Conteúdo e Interpretação

Muitos estudos da computação visam diminuir o *gap* semântico na avaliação de imagens (HUI *et al.*, 2010), e caracterizar a visão humana para o reconhecimento de objetos e recuperação de imagens (DENG e MANJUNATH, 2001). O grande problema reside na dificuldade de computadores em reconhecer determinadas categorias de objetos. A obtenção de um conhecimento específico para tal reconhecimento muitas vezes não é trivial nem mesmo para humanos (PANDEY *et al.*, 2015).

O desafio consiste em ligar a informação semântica com os valores numéricos das características de baixo nível, utilizados em algoritmos de reconhecimento de padrões.

Desse modo, uma abordagem que também trata de sistemas CBIR é a Interpretação de Imagens, que atribui tipos de objetos a objetos de imagens. Ambos os conceitos são discutidos por Castilla e Hay (2008), já que definem um objeto geográfico como um objeto de certo tamanho mínimo que esteja sobre ou próximo da superfície terrestre (por exemplo: florestas, lagos, montanhas). Nesse sentido, o processo de segmentação delinea regiões (ou segmentos) da imagem que apresentam atributos em comum (LIU *et al.*, 2008). Esse passo se baseia na hipótese de que ao particionar uma imagem em objetos está se relacionando à organização conceitual em que humanos compreendem uma determinada paisagem (CASTILLA e HAY, 2008).

O processo de classificação atribui segmentos às classes semânticas (conceitos de alto nível como vegetação). Esse processo é realizado por meio de análises de atributos dos segmentos e seus inter-relacionamentos entre segmentos para identificar rótulos geográficos (LIU *et al.*, 2008). Tal conceito destaca a importância da informação contextual em melhorar a classificação (BLASCHKE e STROBL, 2001).

Porém, a aplicação de uma abordagem baseada em segmentação para classificar as partes das plantas é um trabalho árduo que pode apresentar diversas inconsistências, já que muitas estruturas importantes para a identificação das plantas podem estar ocultas nas imagens, além do fato de que podem ser facilmente confundidas até mesmo por humanos.

O estado da arte de técnicas que reduzem o *gap* semântico pode ser classificado em diferentes pontos de vista (LIU *et al.*, 2007b). Nesse sentido, Liu (*et al.*, 2007b) definem 5 categorias principais que tratam das técnicas do estado da arte:

1. Uso de ontologia para definir conceitos de alto nível (CHIU *et al.*, 2001; MEZARIS *et al.*, 2003; REN *et al.*, 2003; STANCHEV *et al.*, 2003; LIU *et al.*, 2007a; BANDIKOLLA *et al.*, 2014).
2. Métodos de aprendizagem – supervisionado ou não supervisionado – a fim de associar características de baixo nível com conceitos de entrada

- de uma determinada consulta (SETH e COMAN, 2001; VAILAYA *et al.*, 2001; FENG e CHUA, 2003; SHI *et al.*, 2004).
3. Uso de *Relevance Feedback* em laço de recuperação para aprendizagem contínua (ZHANG *et al.*, 2001; JING *et al.*, 2003; MEZARIS *et al.*, 2003).
 4. Geração de um Modelo Semântico (*Semantic Template*) para suporte de recuperação de alto nível de imagens (SMITH e LI, 1998; CHANG e WANG, 1999).
 5. Uso de informação textual obtida na Web e conteúdo visual recuperado de imagens da Web (CAI *et al.*, 2004a; CAI *et al.*, 2004b; FENG *et al.*, 2004; BANDIKOLLA *et al.*, 2014).

Dentre as técnicas citadas acima, Halaschek-Wiener (*et al.*, 2005) mencionam alguns motivos pelos quais as ontologias podem auxiliar no processo de recuperação de imagens. O primeiro se baseia no fato de que as ontologias podem fornecer um modelo semântico daquilo que ocorre na imagem, como objetos, eventos, etc. Além disso, as ontologias também fornecem um mecanismo elegante para formalizar e organizar o conteúdo de imagens em pequenos grupos de conceitos. Ainda, permitem a possibilidade de se afirmar que muitas imagens se referem a um mesmo conceito por meio do uso de URI's. Consequentemente, o uso de ontologias fornece um padrão comum, permitindo que outros indivíduos processem o conteúdo de tais imagens, previamente anotadas (COTO, 2008).

2.3 Abordagens Semânticas para Recuperação de Imagens

Vários trabalhos focam o mapeamento de características de baixo-nível para conceitos definidos em uma ontologia (ADDIS *et al.*, 2003; DUPPLAW *et al.*, 2004; COTO, 2008; BANDIKOLLA *et al.*, 2014; MANZOOR *et al.*, 2015), e incluem o desenvolvimento de ferramentas, que são associadas ao domínio específico da ontologia, com o propósito de anotação de imagens

(SCHREIBER *et al.*, 2001; HOLLINK *et al.*, 2003; COTO, 2008; BANDIKOLLA *et al.*, 2014; LI, 2015).

A razão de se anotar um dado (ou seja, adicionar um metadado a esse) é simplificar o acesso a uma determinada informação. Um metadado adicionado a um documento ou imagem com o propósito de anotação permite uma busca mais efetiva (HANBURY, 2007).

Dentre as abordagens que tratam de anotação de imagens, destaca-se o “Metadado de Conteúdo Descritivo” (*Content-descriptive Metadata*), no qual se refere a um conceito semântico, uma entidade, geralmente presente nas imagens. A anotação pode seguir através de uma lista de palavras chave ou um vocabulário específico controlado (HANBURY, 2007).

Apesar de não apresentarem domínios semelhantes à Sistemática Vegetal, muitos trabalhos focam na extração de características de baixo-nível e relacionam tais dados com as informações definidas em uma ontologia. No estudo desenvolvido por Mezaris (*et al.*, 2003), uma ontologia foi desenvolvida levando em consideração uma abordagem para recuperação de imagens. Ao enviar uma imagem ao sistema, essa imagem passa pela avaliação de algoritmos de cor e segmentação, a fim de separá-la em várias regiões. A partir desse passo, um conjunto de indexadores de características de baixo-nível é calculado para cada região separada da imagem. Desse modo, tal avaliação de imagens descreve a ontologia desenvolvida no trabalho de Mezaris (*et al.*, 2003) que trata de conceitos como intensidade de cor, posição nos eixos horizontal e vertical, tamanho e forma.

O uso de ontologias também se mostrou eficaz no estudo de Manzoor (*et al.*, 2015). Os autores em questão utilizaram uma abordagem semelhante àquela apresentada por Mezaris (*et al.*, 2003), já que foi apresentada uma ontologia de domínio de imagens, capaz de identificar conceitos e características de alto-nível das imagens apresentadas. O trabalho de Manzoor (*et al.*, 2015) também avalia uma coleção de imagens, compara suas características de baixo-nível (cor, textura e forma) e infere conceitos, como coloração, dentre outros. Em seguida, avalia o fato de um determinado conceito, que está definido na ontologia, apresentar uma coloração, por

exemplo, avermelhada. Com o auxílio de outras características extraídas da imagem e com alguma entrada textual opcional do usuário, um ranque das imagens mais relevantes, que compartilham aquelas características e conceitos, é retornado ao usuário.

O trabalho de Coto (2008) também faz uso de ontologias para realizar anotações automáticas nas imagens avaliadas, a fim de melhorar a eficiência da recuperação de imagens e diminuir o *gap* semântico. Anotação automática de imagens é o processo de criação de um modelo que visa atribuir termos visuais às imagens (COTO, 2008). Além disso, a auto anotação consiste em um conjunto de técnicas que auxiliam a busca de uma correlação entre características visuais de baixo nível e semântica (COTO, 2008).

Outra abordagem que também trata semântica e recuperação de imagens está presente no trabalho de Pandey (*et al.*, 2015). Nesse estudo, utiliza-se um banco de dados de categorização hierárquica semântica de imagens, que auxilia na avaliação automática da semântica. De acordo com os autores, essa categorização baseada em semântica leva a uma estrutura hierárquica de árvore que contém imagens de diferentes categorias em vários níveis. Ainda, essa categorização pode auxiliar no entendimento da correlação entre as características visuais e semânticas (*e.g.* Animal, Vegetal, Fruta, etc), que poderá ser utilizado para prover semântica específica da imagem em questão (PANDEY *et al.*, 2015).

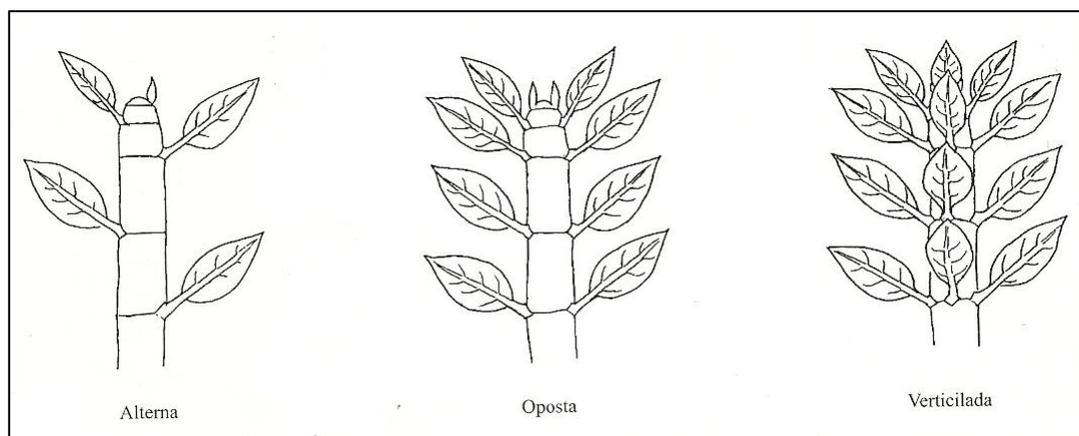
O trabalho desenvolvido por Vogel e Schiele (2007) apresentou um modelo semântico de cenas naturais para sistemas CBIR. Foram utilizados conceitos (também definidos por imagens) como: céu, água, grama, areia, rochas, dentre outros; além de conceitos como posição de um determinado conceito – por exemplo: o céu está na parte superior das imagens. Assim, uma determinada imagem é dividida em grade e cada segmento da grade é analisado separadamente e, posteriormente comparado com os conceitos já definidos. Desse modo, a partir de uma métrica, dependendo da quantidade de alguns conceitos na cena avaliada, é possível determinar se aquela imagem se refere, por exemplo, a uma praia, campo ou região montanhosa.

Como citado anteriormente, as imagens de ramos floríferos não demonstram todas as características necessárias para a avaliação integral e correta de uma planta. Logo, abordagens como a apresentada por Vogel e Schiele (2007) e Pandey (*et al.*, 2015), não seriam as mais adequadas para a problemática do presente projeto, já que, mesmo dividindo as imagens e avaliando as partes específicas, alguns conceitos não são (facilmente) visualizáveis, ou outros se confundem com alguma estrutura específica – como o caso da Bráctea.

Nesse sentido, Souza e Lorenzi (2007) afirmam que persiste a questão de que, observando as Angiospermas como um todo, verifica-se que algumas partes são mais constantes e outras mais variáveis. Na maioria das famílias, por exemplo, o gineceu – parte da flor que desempenha o papel feminino na reprodução sexuada (GONÇALVES e LORENZI, 2007) – é uma estrutura mais ou menos homogênea e, assim, não há como fazer uma identificação segura sem observar as suas características (posição do ovário, número de óvulos, placentação, etc.) e, dessa forma, se o usuário dispuser apenas de flores masculinas é provável que não obtenha sucesso usando a chave dicotômica. Ainda, a chave também não terá muita utilidade se o usuário não dispuser de materiais com flores (SOUZA e LORENZI, 2007).

Outro conceito da botânica que é muito utilizado para identificar determinada família é o tipo de Filotaxia (Figura 3), cuja definição de acordo com Gonçalves e Lorenzi (2007), é o arranjo das folhas ao longo do eixo caulinar, que tende a minimizar o sombreamento de uma folha por outra.

Figura 3: Filotaxia *.



Desse modo, os conceitos de botânica demonstrados confirmam a dificuldade em automatizar o processo de identificação de famílias de Angiospermas a partir de imagens ao utilizar técnicas de segmentação, localidade e divisão da imagem em grades, já que algumas estruturas morfológicas podem ser confundidas, ocultas ou, simplesmente, difíceis de rastrear ao se utilizar uma imagem.

2.4 Recuperação de Imagens com Auxílio de Usuários

Outro método muito utilizado para diminuir o *gap* semântico entre as características de baixo-nível das imagens e os significados que essas representam é a interação com o usuário (SALTON e MCGILL, 1998; GIACINTO, 2007; GUAN e QIU, 2007; KUNDU *et al.*, 2015). Tal abordagem apresenta um grande aumento na eficácia da recuperação de imagens já que em métodos interativos com um usuário, a busca é refinada interativamente de modo a se obter um resultado mais preciso (AREVALILLO-HERRAEZ e FERRI, 2013).

Desde meados da década de 1990, mecanismos interativos tem sido usados para recuperação de imagens a fim de otimizar as métricas de

* Retirado de: Morfologia Vegetal – GONÇALVES e LORENZI, 2007.

similaridades e corrigir erros cometidos por sistemas CBIR de maneira iterativa (KUNDU *et al.*, 2015). Assim, informações de recuperação interativas de imagens são utilizadas para modificar pesos de combinações específicas de modo a refletir diferenças relevantes para tal recuperação (GIACINTO, 2007). Quando métodos interativos são utilizados, a busca é considerada um processo interativo no qual a busca original é refinada iterativamente, a fim de obter de modo progressivo, um resultado mais preciso (AREVALILLO-HERRAEZ e FERRI, 2013).

Bandikolla e Reddy (2014) realizaram um trabalho sobre a recuperação de imagens utilizando combinações entre palavras-chave e características de imagens. Para tanto os autores desenvolveram uma ontologia sobre os jogadores e o campeonato mundial de críquete de 2007. Apesar de apresentar domínio diferente do foco deste trabalho, a abordagem proposta por Bandikolla e Reddy (2014) para diminuir o *gap* semântico ao relacionar as informações de alto-nível com as características de baixo-nível das imagens é a que mais se aproxima da abordagem proposta no presente trabalho.

No estudo desenvolvido por Bandikolla e Reddy (2014), os autores avaliaram imagens de diversas páginas na Web (relacionadas ao mundial de críquete de 2007) e extraíram suas características de baixo-nível, além de avaliarem os rótulos relacionados àquela imagem. Desse modo, conseguem inferir algumas informações a partir dos dados coletados. Por exemplo, se uma imagem tem um nome de um determinado jogador de críquete, pode-se inferir que aquele jogador é o capitão de determinada equipe. Além disso, ao comparar as características de baixo-nível de tal imagem com as demais analisadas, também pode-se recuperar imagens similares com seus respectivos dados anotados. Com o auxílio de um usuário ao realizar buscas através de entradas textuais e de uma determinada imagem, o sistema consegue recuperar informações relevantes à busca. O usuário ainda interage com o sistema através de um mecanismo de *Relevance Feedback*.

Neste sentido, levando em consideração a dificuldade de se identificar uma planta simplesmente ao analisar as características de baixo-nível da imagem, o presente trabalho apresenta uma abordagem que une técnicas de

recuperação de imagens (ao utilizar descritores de imagens e o algoritmo *Ranked List Similarity* (RL-Sim) (PEDRONETTE e TORRES, 2013)) e conceitos relacionados às famílias estudadas, definidos na ontologia, e tem como objetivos diminuir o *gap* semântico entre imagem e conteúdo, além de auxiliar biólogos ou indivíduos que se interessem por botânica, ao necessitarem identificar alguma família de Angiosperma.

Para tanto, o Capítulo 3 descreve os conceitos utilizados, assim como os métodos necessários para a análise das imagens estudadas.

3. Recuperação de Imagens Baseada no Conteúdo

Neste capítulo, são apresentados os aspectos conceituais relacionados à análise dos conjuntos de imagens, assim como os conceitos envolvendo o Modelo de Recuperação de Imagens Baseada no Conteúdo (Figura 1-A) e as ferramentas utilizadas para realizar a extração de características de baixo-nível (Figura 1-A). Este capítulo ainda apresenta o algoritmo de aprendizado não supervisionado (Figura 1-B, RL-Sim) e uma abordagem de agregação de ranques (*Rank Aggregation*) que foram utilizados a fim de melhorar a eficácia da recuperação de imagens.

3.1 Modelo de Recuperação

Dois importantes categorias podem ser definidas no processo de caracterização de imagens: (i) busca '*por exemplo*' (em que se busca exemplos similares ao levantado pelo usuário) e (ii) recuperação semântica utilizando a descrição de um determinado conceito (ex: encontrar imagens que contenham bicicletas). Buscas '*por exemplo*' geralmente são executadas comparando características de baixo nível da imagem, tais como cores, texturas e formas. Já a abordagem semântica, por outro lado, exige um alto nível de entendimento do conteúdo da imagem. Ainda, as buscas dos dois casos também podem ser subdivididas: (i) a busca '*por exemplo*' pode ser realizada a partir de um esboço (*sketch*) ou de um modelo, ao invés de uma imagem em si; de modo similar, (ii) a busca por recuperação semântica pode ser realizada em diferentes níveis de abstração (por imagem, ou pela descrição do que é uma bicicleta) (EAKINS, 2002).

Examinar imagens baseadas em cor é uma das técnicas mais utilizadas. A similaridade de cores entre duas imagens pode ser avaliada mediante a comparação de seus histogramas. O histograma de cor, que representa um

descritor de cores muito comum na linha de pesquisa, indica as frequências de ocorrência das cores na imagem (GONZALEZ e WOODS, 2001).

Outras maneiras também muito difundidas de se caracterizar imagens são técnicas de extração de formas e texturas. Em CBIR, são comuns os problemas na identificação de imagens com formas semelhantes, já que em sua maioria necessitam da segmentação da imagem. O processo de segmentação de imagens já é difícil por si só; na verdade, segmentação e reconhecimento de padrões geralmente se caracterizam por um problema circular do ovo-galinha, já que, uma vez segmentada, a forma de um objeto pode ser descrita e correspondida utilizando uma variedade de características da forma (KEBAPCI *et al.*, 2010).

Texturas podem ser descritas como padrões espaciais formados pela superfície característica de um objeto, que se manifesta como cores ou como variações de escalas de cinza na imagem. Apesar de cada superfície apresentar uma textura, alguns objetos se caracterizam por apresentarem distinções claras entre suas texturas (por exemplo: pele e areia) (MA e MANJUNATH, 1996; LIU e WECHSLER, 2001; HAN e MA, 2007).

Mais recentemente, ganharam grande destaque métodos de extração de características visuais baseados em redes neurais (JIA *et al.*, 2014).

Para realizar a extração das características de baixo-nível das imagens analisadas no presente trabalho, foram utilizados dois *frameworks* distintos: 1) o LIRe, que se baseia na aplicação de descritores de imagem, e 2) o *CaffeNet* (JIA *et al.*, 2014), uma rede neural convolucional (*Convolutional Neural Network* – CNN), com abordagem em *Deep Learning*.

3.2 Modelo Formal de Recuperação de Imagens

Um modelo geral de recuperação de imagens foi considerado para definir o presente passo. Seja $C = \{img_1, img_2, \dots, n\}$ uma coleção de imagens, em que cada imagem representa uma espécie de planta e n o

tamanho do conjunto de dados. Seja $p(i, j)$ a função de distância entre duas imagens img_i e img_j de acordo com a extração de uma determinada característica visual.

Baseado na função de distância p , um *ranked list* T_q pode ser computado em resposta a uma imagem de consulta img_q , que também define uma espécie de planta. Desse modo, é esperado que as primeiras posições dos *ranked lists* contenham as imagens mais similares às respectivas imagens de busca.

Um *ranked list* $T_q = (img_1, img_2, \dots, img_{n_s})$ pode ser definido como a permutação entre os subconjuntos $C_s \subset C$, que contêm as imagens mais similares à imagem de busca img_q , de modo que $|C_s| = n_s$. Uma permutação T_q é uma bijeção do conjunto C_s para o conjunto $[n_s] = \{1, 2, \dots, n_s\}$. Para a permutação T_q , interpreta-se $T_q(i)$ como a posição (ou o *rank*) da imagem img_i no *ranked list* T_q .

Desse modo, torna-se possível computar a matriz de distância A , contendo as distâncias entre todas as imagens da coleção. Também torna-se possível tomar toda imagem $img_i \in C$ como uma imagem de busca (*query*) img_q , com o intuito de se obter o conjunto $T = \{T_1, T_2, \dots, T_n\}$ de *ranked lists* para cada imagem do conjunto C .

O objetivo do uso do algoritmo de aprendizado não supervisionado consiste em explorar as informações contextuais presentes nas distâncias e nos *ranked lists* para melhorar a eficácia da recuperação dos resultados. Formalmente, pode ser definido como uma função f_r que computa uma nova e mais eficaz matriz de distâncias $\hat{A} = f_r(A, T)$.

3.3 Extração das Características Visuais

As imagens tiveram suas características visuais de baixo-nível extraídas pelo uso de descritores presentes no LIRe e de uma abordagem baseada em *Deep Learning* (Figura 1-A). Tais métodos extraíram características (como

textura, cor, forma e local) que foram responsáveis por quantificar a similaridade entre quaisquer pares de imagens da coleção de dados.

3.3.1 LIRe

A biblioteca LIRe emprega técnicas de extração de características globais e locais das imagens para sistemas CBIR (SEMANTIC METADATA, 2015) através do uso de descritores.

Os descritores estão disponíveis no LIRe (*Lucene Image Retrieval*), uma biblioteca de código aberto Java para CBIR (*Content Based Image Retrieval*), construída sobre a ferramenta (*engine*) de busca e recuperação textual *Lucene* – *Apache*. A biblioteca LIRe possibilita a extração de características de imagens e as armazena em um índice para posterior recuperação (LUX e CHATZICHRISTOFIS, 2008; LUX, 2013).

A biblioteca em questão visa a integração de imagens de um determinado conjunto de dados (*dataset*) por meio do uso de descritores de imagens, que realizam a extração de características como cor, textura e forma (LUX e CHATZICHRISTOFIS, 2008; SEMANTIC METADATA, 2015). Descritores de imagem podem ser definidos como os responsáveis por quantificar o quanto uma imagem é similar à outra (PENATTI *et al.*, 2012) em um determinado quesito.

Após o processo de indexação das imagens do conjunto de dados (*dataset*) analisado por um determinado descritor, um vetor de características de cada imagem é gerado e, em seguida, é calculada a distância entre cada imagem, de modo que todos os vetores de características são comparados entre si (SEMANTIC METADATA, 2015).

De acordo com Lux e Marques (2013), algumas métricas podem ser aplicadas para computar a distância entre as imagens – de modo que quanto menor essa distância, mais similares são as imagens.

Ainda de acordo com Lux e Marques (2013), típicas funções de distâncias são:

- *Manhattan ou Distância L1* – Distância entre dois pontos é a soma das diferenças absolutas das coordenadas. Por exemplo: Dado dois pontos $P_1(x_1, y_1)$ e $P_2(x_2, y_2)$, a distância L_1 entre os pontos P_1 e P_2 é calculada da seguinte maneira:

$$L_1 = |x_1 - x_2| + |y_1 - y_2|$$

- *Euclidiana ou Distância L2* – Soma da raiz quadrada da diferença entre x e y em suas respectivas dimensões. Por exemplo: Dado dois pontos $P_1(x_1, y_1)$ e $P_2(x_2, y_2)$, a distância L_2 entre os pontos P_1 e P_2 é calculada da seguinte maneira:

$$L_2 = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

- *Coefficiente Cosseno* – Medida da similaridade entre dois vetores que mede o cosseno do ângulo entre eles. Por exemplo: Dado dois vetores A e B, o coeficiente cosseno é calculado através do produto escalar de A e B, pela norma de A em B:

$$\cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|}$$

- *Divergência de Jensen-Shannon* – Método de medida da similaridade entre duas distribuições de probabilidades. Considere um conjunto de distribuição de probabilidades $M_+^1(A)$, a divergência de Jensen-Shannon (JSD) $M_+^1(A) \times M_+^1(A) \rightarrow [0, \infty)$ é uma versão suavizada e simetrizada da divergência de Kullback-Leibler $D(P || Q)$, e é definida da seguinte maneira:

$$JSD(P || Q) = \frac{1}{2} D(P || M) + \frac{1}{2} D(Q || M)$$

$$\text{Onde } M = \frac{1}{2} (P + Q)$$

As métricas de distância já estão implementadas na biblioteca LIRe, de modo que cada características visual utiliza a métrica mais adequada (LUX e MARQUES, 2013).

Tendo como base os índices (que armazenam os vetores de características) e as métricas de distância, uma matriz de distância A é calculada. Cada elemento A_{ij} corresponde à distância entre as imagens img_i e img_j analisadas. Por exemplo: o elemento $A_{1,2}$ seria referente à distância entre as imagens 1 e 2, de acordo com o índice gerado pela biblioteca LIRe e pelo descritor de imagens selecionado.

Cada linha (listas de ranques – *ranked lists*) da matriz de distâncias é ordenada de modo crescente, para que seja possível especificar quais imagens são mais similares com a imagem de busca, de acordo com o descritor utilizado, já que quanto menor as distâncias mais similares são as imagens. Assim, cada linha da matriz será uma lista de ranque (*ranked list*), cujo índice equivale ao número do índice da própria linha. Cada descritor utilizado gera uma matriz de distâncias diferente.

3.3.2 Características Visuais

No presente estudo foram avaliadas e utilizadas várias características visuais, a fim de se obter o melhor resultado da extração de características das imagens.

Para tanto, 16 descritores foram testados:

- **Descritores de cor:** **ACC** (*Auto Color Correlogram*, HUANG *et al.*, 2007), **BIC** (*Border/Interior Pixel Classification*, STEHLING *et al.*, 2002), **CEDD** (*Color and Edge Directivity Descriptor*, CHATZICHRISTOFIS e BOUTALIS, 2008a), **Color Histogram RGB** (•), **Color Layout** (•), **FCTH** (*Fuzzy Color and Texture Histogram*, CHATZICHRISTOFIS e BOUTALIS, 2008b), **JCD** (*Joint Composite Descriptor*, ZAGORIS *et al.*, 2010), **Opponent Histogram** (VAN DE SANDE *et al.*, 2009), **Scalable Color** (•).
- **Descritores de textura:** **CEDD** (*Color and Edge Directivity Descriptor*, CHATZICHRISTOFIS e BOUTALIS, 2008a), **Edge Histogram** (WON *et*

al., 2002), **FCTH** (*Fuzzy Color and Texture Histogram*, CHATZICHRISTOFIS e BOUTALIS, 2008b), **Gabor** (ZHANG *et al.*, 2012), **JCD** (*Joint Composite Descriptor*, ZAGORIS *et al.*, 2010), **LBP** (*Local binary patterns*, OJALA *et al.*, 2002), **PHOG** (*Pyramid Histogram of Oriented Gradients*, BOSCH *et al.*, 2007), **RLBP** (*Rotated Local Binary Pattern*, MEHTA, e EGIAZARIAN, 2013), **Tamura** (TAMURA *et al.*, 1978).

- **Descritor Local: SURF** (*Speeded Up Robust Features*, BAY *et al.*, 2008).

Vale citar que o modelo BOVW – *Bag Of Visual Words* – foi utilizado pelo descritor SURF (*Speeded Up Robust Features*). Nesse sentido, BOVW é uma representação de imagens, em que é utilizado uma grande quantidade de conjuntos de características locais. Esse modelo é inspirado no modelo *Bag Of Words* (BOW), utilizado para recuperação de textos, em que um documento é representado por um conjunto de palavras-chave distintas. Analogamente, em modelos BOVW, uma imagem é representada por um conjunto de palavras visuais derivadas de uma característica local (ZAGORIS *et al.*, 2011).

Desses 16 descritores, foram escolhidos apenas os que apresentaram as melhores classificações para cada um dos conjuntos de dados estudados. Desse modo, entre os dois *datasets* (*Oxford Flowers 17 e 102 Classes*) foram selecionados três descritores.

A Tabela 1 sumariza as características selecionadas para a utilização nesse trabalho:

Tabela 1: Descritores com melhor precisão de classificação.

Descritor (Sigla)	Descritor (Nome)	Função
ACC	<i>Auto Color Correlation</i>	Cor
BIC	<i>Border/Interior Pixel Classification</i>	Cor
SURF	<i>Speeded Up Robust Features</i>	Local

Outra ferramenta também utilizada para a extração de características das imagens foi a Rede Neural Convolutiva (*Convolutional Neural Network*) *CaffeNet*, um *framework* já treinado em uma das maiores bases de imagens da Web, o *ImageNet* (JIA *et al.*, 2014).

3.3.3 *Deep Learning* em Sistemas CBIR

Abordagens baseadas em *Deep Learning* têm sido muito aplicadas já que atingem a precisão de algoritmos do estado da arte na execução de várias tarefas (OOi *et al.*, 2015) como classificação de imagens (KRIZHEVSKY *et al.*, 2012) e análises multimodais (WANG *et al.*, 2014; WANG *et al.*, 2016). Além disso, são adequadas para o uso em grandes conjuntos de dados já que lida com certas nuances dos parâmetros e dos dados de um modo mais simples (OOi *et al.*, 2015; LUND´EM e KOIVUNEN, 2016).

Dentre as abordagens de implementação do conceito de *Deep Learning*, destaca-se o *Convolutional Neural Network* (CNN) (LeCUN *et al.*, 1998), que também tem sido aplicado devido à alta performance em reconhecer objetos (KRIZHEVSKY *et al.*, 2012) e localizá-los em um determinado contexto (JARRETT *et al.*, 2009). Além disso, CNN's são mais fáceis de treinar e apresentam menos conexões entre as camadas, quando comparadas a outras abordagens baseadas em *Deep Learning* (OUYANG *et al.*, 2015).

CNN's exploram a correlação espacial através da aplicação de um padrão de conectividade local entre os neurônios das camadas adjacentes. Em outras palavras, as entradas das unidades internas da camada m são um subconjunto das unidades da camada de $m - 1$ (LeNet, 2016). Além disso, CNN's são redes discriminativas, treinadas via retro propagação (*back-propagation*) através de camadas de filtros convolucionais e outras operações como retificação e *pooling* (JIA *et al.*, 2014).

Nesse sentido, destaca-se o *CaffeNet*, um *framework open-source* de livre acesso a arquiteturas baseadas em *Deep Learning*. Com o uso do

CaffeNet é possível ter acesso a um extenso leque de ferramentas do estado da arte de algoritmos *Deep Learning*, além do fato de que foi treinada com umas das maiores bases de imagens da Web – o ImageNet (JIA *et al.*, 2014).

Desse modo, é possível utilizar o *CaffeNet* como um extrator de características de baixo-nível das imagens dos conjuntos de dados analisados no presente trabalho.

O *framework CaffeNet* foi treinado para reconhecer 1000 categorias de objetos do *ImageNet* (RUSSAKOVSKY *et al.*, 2015) e é muito similar à *AlexNet* (KRIZHEVSKY *et al.*, 2012) com exceção do fato de que não foi utilizado reforço de dados (*data-augmentation*) e a ordem de *pooling* e normalização é comutada. Do *CaffeNet* foram consideradas as características da 7ª camada totalmente conectada (*fully connected layer – fc7*). As imagens foram redimensionadas para 256 x 256 *pixels* e os vetores de características apresentaram 4096 dimensões. Para o cálculo de distância foi considerado o espaço Euclidiano (função de distância L2).

Após a escolha dos parâmetros e a execução do *framework* em questão obtêm-se também uma matriz de distâncias que será utilizada como entrada para os próximos passos do sistema.

Após o estágio das extrações de características, foi utilizada uma abordagem de re-ranqueamento (*re-ranking*) como um pós-processamento baseado em aprendizado não-supervisionado, a fim de melhorar a eficácia da recuperação das imagens.

3.4 Pós-Processamento Baseado em Aprendizado Não Supervisionado

Ranquear (*Ranking*) e re-ranquear (*re-ranking*) são dois passos de extrema importância para a recuperação de informações, já que a maioria dos usuários faz uso dos resultados melhores ranqueados (TIAN *et al.*, 2008; WEI *et al.*, 2009; JIANG *et al.*, 2009; CHEN *et al.*, 2011). Ao realizar um *ranking* de

recuperação dos resultados garante-se que tal processo utiliza o conteúdo e/ou o contexto para descobrir os resultados mais relevantes (CHEN *et al.*, 2012). De acordo com Pedronette e Torres (2013), contexto pode ser definido como toda informação de uma determinada situação que seja relevante à aplicação ou a um conjunto de usuários.

Estudos utilizando *rankings* com o intuito de aumentar a eficácia dos resultados têm sido cada vez mais intensos nos últimos anos (QIN *et al.*, 2011; CHEN *et al.*, 2014). A principal motivação do uso de *rankings* se baseia no fato de que, enquanto uma função de distância relaciona apenas dois objetos, um ranque considera toda coleção em questão. Nesse cenário, ranquear de modo preciso as coleções de imagens é de extrema importância.

Como coleções de imagens são ranqueadas em ordem crescente de distância da imagem de busca, deve-se escolher uma medida de distância coerente e eficaz, já que esta é uma das partes mais críticas da construção de sistemas CBIR de qualidade (YANG *et al.*, 2008). Em geral, tais sistemas consideram análises de apenas pares de imagens, e tal comparação ignora informações codificadas nas relações entre o conjunto de imagens. Portanto, a distância entre duas imagens pode ser corretamente descrita somente se considerar o contexto de outras imagens similares. Esse processo requer um modelo para capturar a essência da similaridade entre imagens, ao invés de visualizar cada imagem como um conjunto de pontos ou características vetoriais (YANG *et al.*, 2008).

Para realizar a análise do pós-processamento para o presente trabalho, foram utilizados os algoritmos RL-Sim que se baseia na similaridade de *rankings*, e o *Rank Aggregation* que realiza a união de duas ou mais matrizes de distâncias para melhorar a eficácia da recuperação das imagens analisadas. Maiores detalhes desses algoritmos encontram-se disponíveis nas Seções 3.4.1 – Aprendizado Não Supervisionado – RL-Sim e 3.4.2 – *Rank Aggregation*, respectivamente.

3.4.1 Aprendizado Não Supervisionado – RL-Sim

O contexto de uma determinada informação pode ser aplicada em métodos de *re-ranking*. Nesse sentido, o uso de *re-ranking* pode ser definido como um processo de refinamento de buscas: o método em questão leva em consideração um *ranking* inicial e agrega outras informações adicionais a fim de aprimorar o processo de recuperação (PEDRONETTE e TORRES, 2013).

Uma abordagem baseada em *re-ranking* pode ser aplicada em sistemas CBIR, de modo explorar os relacionamentos entre imagens, a fim de melhorar a precisão de tais sistemas (PEDRONETTE e TORRES, 2011). Assim, as distâncias calculadas por uma determinada métrica, entre algumas imagens e *ranked-lists* podem ser utilizadas para analisar o contexto de uma informação (PEDRONETTE e TORRES, 2013). Sistemas CBIR que utilizam o *re-ranking* se baseiam em abordagens de agrupamento já que o contexto é modelado ao agrupar imagens similares (SCHWANDER e NIELSEN, 2010).

Desse modo, para realizar esse tipo de abordagem no presente trabalho, será utilizado o algoritmo RL-Sim (*Ranked Lists Similarities* – Figura 1-B), um algoritmo de *re-ranking* que considera a similaridade entre *ranked lists* para caracterizar informações contextuais em sistemas CBIR. O RL-Sim é um algoritmo de pós-processamento e sua motivação principal se baseia na conjuntura de que informações contextuais codificadas em similaridades entre *ranked lists* podem prover recursos para melhorar a efetividade de descritores CBIR. Em geral, apenas duas imagens são consideradas para a computação da distância e, se a medida de distância adotada não é precisa, as duas imagens serão alocadas em locais errados em seus *ranked lists* (PEDRONETTE e TORRES, 2013).

Sabe-se que as primeiras posições de um *ranked list* podem conter muitas imagens similares, mas por outro lado, também conter imagens não similares. As imagens posicionadas nas primeiras posições geralmente são similares entre si e, portanto, tendem a apresentar imagens em comum em seus *ranked lists*. Ainda, pode-se observar que esse conjunto de imagens

(similares à imagem de busca e similares entre si) aparece nos *ranked lists* de todas as imagens que compõem o conjunto de dados. O mesmo comportamento não pode ser observado ao analisar as primeiras posições dos *ranked lists* de imagens não similares (o mesmo conjunto de imagens não aparece nas primeiras posições). Nesse cenário, uma distância contextual é gerada, já que existem poucas imagens em comum nas primeiras posições dos *ranked lists* das imagens não similares (PEDRONETTE e TORRES, 2013).

3.4.1.1 Métrica de Distância Contextual Baseada em Ranqueamento

Nessa seção, o algoritmo RL-Sim (Figura 1-B) é descrito em termos da métrica de distância contextual dos *rankings* baseada em similaridade/dissimilaridade dos *ranked lists* (PEDRONETTE e TORRES, 2013).

Tal distância é iterativamente aprendida em um cenário não supervisionado, ao incorporar a informação contextual provida pela métrica correlacionando *rankings* (PEDRONETTE e TORRES, 2013).

Para tanto, considere os vizinhos mais próximos da imagem img_i como o conjunto $N(i; k)$, que contém as k imagens mais similares à img_i de acordo com uma dada distância p , definida pelo método de extração de imagem utilizado. O conjunto $N(i; k)$ pode ser obtido pela abordagem do “*K Vizinhos Mais Próximos*” (KNN – *K-Nearest Neighbour*), em que a cardinalidade do conjunto é denotada por $|N(i; k)| = k$.

A aplicação de tal medida propõe capturar a extensão da sobreposição entre *ranked lists*, de modo que esta possa ser estendida ao considerar não só a própria sobreposição na profundidade k , mas também a sobreposição cumulativa em profundidades cada vez maiores (PEDRONETTE e TORRES, 2013). Caso dois *ranked lists* apresentem as mesmas imagens nas primeiras

posições, o número de intersecções é maior, e a medida de similaridade de intersecções também será maior (PEDRONETTE e TORRES, 2013).

Como as imagens nas primeiras posições dos *ranked lists* geralmente são aquelas mais relevantes, no sentido de que geralmente representam o resultado no qual o usuário está interessado (PEDRONETTE e TORRES, 2013), deve-se considerar $d(T_i, T_j, k)$ a notação da medida de correlação de *ranks* entre os *ranked lists* T_i e T_j , considerando suas posições *top-k* dos conjuntos $N(i)$ e $N(j)$. A medida de correlação de *ranks* considerada é baseada na intersecção entre *ranked lists* (PEDRONETTE e TORRES, 2013). Uma distância contextual não iterativa $p_c(img_i, img_j)$ baseada na comparação dos *ranked lists* T_i e T_j pode ser definida conforme segue:

$$p_c(img_i, img_j) = d(T_i, T_j, k) \quad (1)$$

Baseando-se na conjectura de que a métrica de distância contextual p_c representa uma distância mais eficaz entre imagens, a distância entre todas as imagens de uma coleção pode ser recalculada baseada nessa métrica. Portanto, um novo conjunto de *ranked lists* pode ser obtido, de modo que a distância contextual pode também ser recalculada e o processo possa ser repetido de um modo iterativo. Seja (t) a iteração corrente e $T_i^{(t)}$ o *ranked list* na iteração t . Seja $p_c^{(0)}$ a distância contextual na primeira iteração, que é igual à distância definida pelo descritor de imagem, de modo que $p_c^{(0)}(img_i, img_j) = p(img_i, img_j)$ para todas as imagens $img_i, img_j \in C$. A medida iterativa contextual é definida como:

$$p_c^{(t+1)}(img_i, img_j) = d(T_i^{(t)}, T_j^{(t)}, k) \quad (2)$$

Desse modo, espera-se que a efetividade da métrica de distâncias melhore no decorrer das iterações, de modo que imagens não relevantes sejam movidas para fora das primeiras posições dos *ranked lists*. Nesse sentido, o tamanho dos k vizinhos pode ser aumentado a fim considerar mais imagens nas iterações. Portanto a métrica contextual pode ser redefinida como:

$$p_c^{(t+1)}(img_i, img_j) = d(T_i^{(t)}, T_j^{(t)}, k + t) \quad (3)$$

Após um determinado número de t iterações, uma nova distância \hat{p} é calculada baseada na métrica de distância contextual p_c .

$$\hat{p}(img_i, img_j) = p_c^{(t)}(img_i, img_j) \quad (4)$$

Finalmente, a distância \hat{p} propicia o cálculo da nova matriz de distâncias $\hat{A}_{ij} = \hat{p}(img_i, img_j)$, fornecendo maior eficácia na recuperação dos resultados.

Desse modo, Pedronette e Torres (2013) propuseram o RL-Sim, a fim de mover imagens não similares para posições inferiores dos *ranked lists* e, como resultado do processo, aumentar a eficácia dos resultados de recuperação dos *rankings*.

Como a aplicação do RL-Sim é realizada diretamente nos *ranked lists*, um exemplo do funcionamento e da utilização desses *rankings* será demonstrado na próxima subseção.

3.4.1.2 Impacto do Pós-processamento nas Listas de Resultados

A fim de exemplificar o funcionamento dos *ranked lists*, antes e depois da ação do RL-Sim, será demonstrado um caso real gerado pelo LIRe para a base de dados *Oxford Flowers 17 Classes* e com a extração de características realizada pelo descritor SURF.

Após a indexação de todas as imagens, cada linha da matriz equivale a um *ranked list*, cujo índice da linha referencia um determinado *ranking*. Para este exemplo, foram utilizados os *ranked lists* '0' e '1' (indexados pelo LIRe), ambos de imagens pertencentes à mesma classe.

A Figura 4 demonstra os *ranked lists* '0' e '1' (indexados pelo LIRe), ambos de imagens cuja primeira posição pertencem à mesma classe. A Figura 4 ainda mostra as sete primeiras posições dos respectivos *rankings*, sem a aplicação do algoritmo RL-Sim, ou seja, demonstra parte dos *rankings*, da maneira como o LIRe ordenou as imagens a partir da geração da matriz de distâncias.

Figura 4: *Ranked lists* '0' e '1', gerados a partir da extração de características do descritor SURF, sem a aplicação do RL-Sim.



Pode-se observar que, visivelmente é difícil notar erros, o que leva a crer que a maioria das imagens de cada um dos *ranked lists* está correta (ou seja, quando a imagem seguinte à imagem da primeira posição pertence à mesma classe da imagem de busca). Porém, a Figura 5 demonstra as imagens que de fato estão corretas para os respectivos *rankings*, de acordo com a classificação de espécies.

Figura 5: *Ranked lists* '0' e '1', com as respectivas imagens corretas.



Ao analisar a Figura 5 pode-se observar que as imagens na segunda posição dos *ranked lists* '0' e '1' equivalem à imagem de busca do *ranked list* oposto. Demonstrando que elas apresentam um alto grau de similaridade e sendo classificadas de modo correto. Ainda, observa-se também, que o *ranked list* '0' apresenta uma imagem recuperada corretamente na sétima posição, enquanto o *ranked list* '1' não apresenta mais nenhuma imagem recuperada de modo correto, a partir da segunda posição.

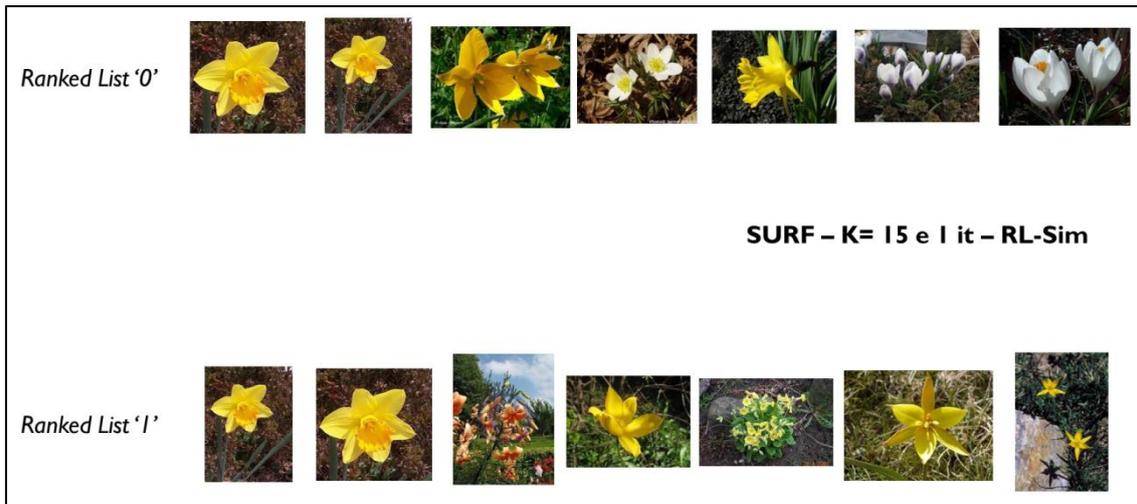
Após esta análise, inicia-se o processo de aplicação do RL-Sim. A princípio, o RL-sim varre as k primeiras posições, uma a uma, do primeiro *ranked list* analisado – nesse caso, o *ranked list* '0' – e então procura a próxima imagem, ou seja, aquela que apresenta a maior similaridade com a imagem de busca – nesse caso, a imagem '1'.

Em seguida, o algoritmo analisa o *ranked list* da imagem '1', em busca da imagem '0' nas k primeiras posições do *ranked list* da imagem '1' e, conforme vai avançando no valor de k e a imagem '0' estiver relativamente distante da posição mais próxima da imagem '1', o RL-Sim diminui o valor de similaridade entre tais imagens – o que não é o caso desse exemplo, já que a imagem '0' está logo após a imagem '1' no *ranked list* da imagem '1'. Sendo assim, o RL-Sim aumenta a similaridade entre as imagens e os *rankings* analisados.

Após esse passo, que é realizado para todos os *rankings* de todas as imagens do conjunto de dados, a matriz de distâncias é novamente ordenada, para que um novo *ranked list* seja gerado. Esse processo é iterativo, e pode ser alterado para que diferentes análises sejam realizadas.

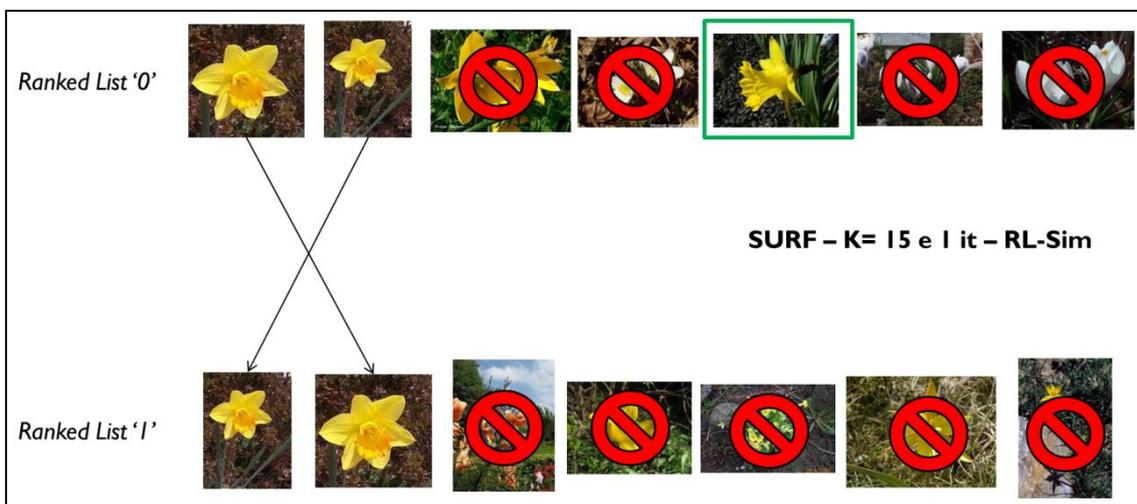
A Figura 6 já demonstra os *ranked lists* '0' e '1' após a primeira iteração do RL-Sim (Figura 1 – *Ranked Lists*) utilizando o parâmetro $k = 15$. Pode-se observar que a Figura 6 já é bem diferente da Figura 4, já que houve várias alterações no posicionamento das imagens de ambos os *rankings*.

Figura 6: *Ranked lists* '0' e '1', gerados a partir da extração de características do descritor SURF, com a aplicação do RL-Sim.



A Figura 7 demonstra os acertos e os erros cometidos pelo processo.

Figura 7: *Ranked lists* '0' e '1', com as respectivas imagens corretas.



Novamente, nota-se na Figura 7 que as imagens '0' e '1' aparecem nas segundas posições dos *ranked lists* opostos. Ainda, no *ranked list* '0', pode-se observar que a imagem que ocupa a quinta posição (destacada pelo quadro verde) é a mesma que na Figura 5 ocupava a sétima posição.

Desse modo, para esse exemplo, caso seja analisada a precisão de recuperação de imagens corretas para as cinco primeiras imagens do *ranked list* '0', pode-se observar que já houve um aumento dessa métrica, quando se compara o *ranked list* '0' das Figuras 7 e 5, já que na Figura 4, nas cinco primeiras posições, têm-se apenas duas imagens corretas, enquanto na Figura 7, têm-se três imagens corretas para as cinco primeiras posições.

Essa aplicação do RL-Sim pode aumentar ainda mais as métricas de eficácia quando utilizado o *Rank Aggregation* – um algoritmo em que ocorre a combinação ou união de diferentes descritores de imagens.

3.4.2 Rank Aggregation

O algoritmo RL-Sim (PEDRONETTE e TORRES, 2013) também pode ser utilizado para combinar diferentes características visuais, que podem complementar diferentes informações visuais, já que mais de um descritor é utilizado para gerar as matrizes.

Cada método de extração de características visuais gera uma matriz de distância, que pode compor um conjunto de matriz de distâncias $\{A_1, A_2, \dots, A_p\}$. De acordo com Pedronette e Torres (2013), o RL-Sim combina o conjunto de matriz de distância em uma única matriz A_c utilizando uma abordagem de multiplicação.

Cada posição da matriz A_c é calculada como segue:

$$A_{c_{ij}} = \prod_{l=1}^p (1 + A_{l_{ij}}) \quad (5)$$

Dado uma nova matriz combinada de distâncias A_c , um novo conjunto de *ranked lists* é computado e submetido ao algoritmo de distância não supervisionado original – RL-Sim.

4. Ontologia

Neste capítulo são apresentados os aspectos conceituais que envolvem as definições, a metodologia e a modelagem da ontologia. Além disso, o presente capítulo demonstra também os conceitos básicos do domínio da ontologia desenvolvida – a Sistemática Vegetal.

4.1 Web Semântica

A Web Semântica – proposta por Tim Berners-Lee, criador da Web – é uma extensão da Web atual e um meio universal para a troca de informação, de modo a atribuir significado (semântica) ao conteúdo dos documentos da Web, onde a informação estaria disponível não só para o consumo humano, mas também seria formatada de modo a permitir o processamento automático das fontes de informação por parte das máquinas computacionais (BREITMAN, 2005; BERNERS-LEE *et al.*, 2001).

Para que os computadores possam analisar e de alguma maneira utilizar o conteúdo disponível na Web, é preciso organizar, estruturar e adicionar semântica às informações da Web, de tal forma que agentes de *software* possam compreendê-las (GOÑI *et al.*, 2002). Na Web Semântica a estruturação da informação por meio de metadados e a adição da semântica à informação ao utilizar ontologias permitem o acesso automático à informação por parte das máquinas (BERNERS-LEE *et al.*, 2001).

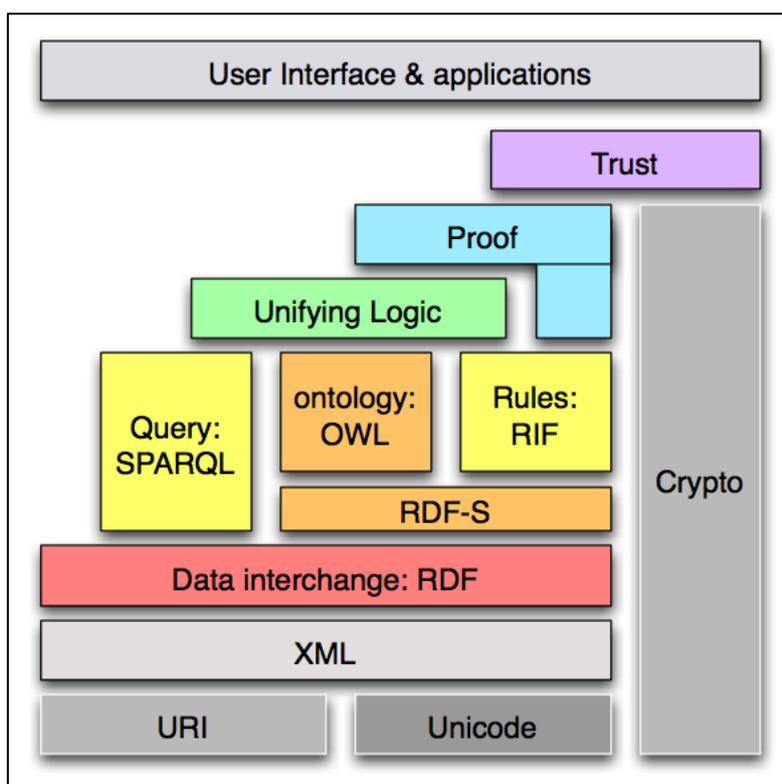
A semântica fornece uma descrição, um significado, para um determinado conceito e, ao expor a semântica de uma informação, o *software* que trata tal objeto, consegue manipular o dado em questão de maneira correta e coerente. Além disso, os computadores poderão inferir novos fatos relacionados àquela informação resgatada, ao invés de simplesmente demonstrar o resultado ao usuário final. A semântica explícita suporta a integração de dados e informações por meio de buscas semânticas, que irão

relacionar informações representativas que estão distribuídas na rede (LACY, 2005).

Uma arquitetura em camadas para a Web Semântica (Figura 8) foi proposta pelo W3C (*World Wide Web Consortium*), definindo os recursos e as linguagens para a Web em questão, na qual cada camada ou tecnologia é complementar e compatível com as camadas inferiores, e não é dependente de qualquer camada superior.

Os dados disponíveis podem ser acessados usando a arquitetura da Web Semântica (Figura 8) e estar relacionados uns aos outros da mesma forma que os documentos já o são. De certo modo, isso também significa criar uma plataforma comum que permita o compartilhamento e a reutilização dos dados e a recuperação de informações por meio das aplicações, empresas e comunidades. Assim, os dados podem ser processados automaticamente tanto por ferramentas, quanto de maneira manual, também revelando novos relacionamentos possíveis entre as porções de dados (W3C, 2014).

Figura 8: Arquitetura da Web Semântica *.



* Fonte: <http://www.w3.org/DesignIssues/diagrams/sWeb-stack/2006a.png>

Nesse sentido, o uso de ontologias – um dos pilares da arquitetura da Web Semântica – auxilia na criação de representações, visando conceitualizar e contextualizar determinado domínio de interesse (GAVA e MENEZES, 2003). Além disso, o conceito de ontologia, de acordo com a comunidade de Inteligência Artificial (IA), possibilita a descrição de uma semântica explícita (LACY, 2005).

4.2 Ontologias – Definições

O termo “ontologia” é extremamente sobrecarregado e pode apresentar diferentes significados entre as diferentes comunidades que utilizam tal termo. Historicamente, o termo ontologia se originou na filosofia e metafísica, e se referencia aos estudos da ciência do ser. Trata da natureza e da organização da realidade. A definição histórica de ontologia descreve as noções filosóficas abstratas de quais tipos de coisas existem. Ontologias fornecem definições e classificações de objetos e seus relacionamentos em todas as esferas do ser. Formalismos matemáticos e classificações também descrevem a ontologia histórica (LACY, 2005).

Ontologias vêm sendo amplamente utilizadas como técnica de representação e reutilização do conhecimento. De forma simplificada, pode-se dizer que uma ontologia é um conjunto de conceitos e termos usados para descrever determinado domínio, que se relacionam por atributos e relacionamentos, realizando declarações expressivas e suportando o compartilhamento de informações (GUARINO, 1998; LACY, 2005).

Na implementação da Web Semântica é necessário definir um modelo de representação das informações e do conhecimento. Nesse contexto, as ontologias são utilizadas para capturar o conhecimento do domínio de interesse, descrever seus conceitos e também as suas relações (HORRIDGE *et al.*, 2009).

O consórcio W3C define uma ontologia como “*a definição dos termos utilizados na descrição e na representação de uma área do conhecimento*”.

De acordo com Breitman (2005), a definição de ontologia encontrada mais frequentemente na literatura da Web Semântica é a proposta por Gruber (1993) em “*A translation approach to portable ontology specifications*”:

“Ontologia é uma especificação formal e explícita de uma conceitualização compartilhada”

Seguindo a definição de Gruber, “conceitualização” representa um modelo abstrato de algum fenômeno que identifica os conceitos relevantes para o mesmo; “explícita” significa que os elementos e suas restrições estão claramente definidos; “formal” significa que a ontologia deve ser passível de processamento automático; e “compartilhada” reflete a noção de que uma ontologia captura conhecimento consensual, aceito por um grupo de pessoas.

As ontologias suportam o compartilhamento ao descrever um domínio com semântica explícita que suporta declarações expressivas. Pelo fato de um *software* ser compatível com OWL (*Web Ontology Language*), este consegue interpretar ontologias, manipular tais informações internamente e interoperar com outros *software*. Além disso, permitem o mapeamento entre as fontes de informações.

Desse modo, uma ontologia é então uma descrição ou especificação formal dos conceitos e dos relacionamentos que podem existir e que podem ser usados por um agente ou uma comunidade de agentes, e cuja finalidade principal é permitir a descoberta, a partilha e a reutilização do conhecimento (STAAB e STUDER, 2004). Portanto, uma ontologia relaciona-se com um vocabulário específico, uma linguagem específica e a conceitualização de um determinado domínio (GONÇALVES, 2007).

Em ontologias, conceitos são categorizados em classes baseadas em características que eles têm em comum. A ideia de que categorias são

definidas por conjuntos de propriedades forma a base do que o linguista George Lakoff chama de “visão clássica” de categorização. Nessa visão, uma classe é definida por meio de uma série de propriedades. A condição básica para se pertencer a uma dada classe é possuir todas as propriedades (BREITMAN, 2005).

4.2.1 OWL

De acordo com a W3C (2012), OWL (*Web Ontology Language*) é uma linguagem da Web Semântica, projetada para representar conhecimentos complexos sobre alguma coisa, grupos de coisas, e relacionamentos entre essas coisas. A intenção da OWL é representar conceitos e seus relacionamentos na forma de ontologias (BREITMAN, 2005). OWL também se apresenta como um dos pilares da Web Semântica (Figura 8).

Além disso, OWL é uma linguagem baseada em lógica, de modo que o conhecimento expresso pela linguagem possa ser explorado por programas de computadores. Os documentos OWL, conhecidos como ontologias, podem ser publicados na *World Wide Web* e podem fazer referência, ou serem referenciados, por alguma outra ontologia OWL (W3C, 2012).

Tim Berners-Lee afirmou em um comunicado à imprensa no dia 19/08/2003 (LACY, 2005) que:

“OWL é um passo importante para que o dado disponível na Web seja muito mais processável pela máquina e reutilizável entre aplicações”

A linguagem OWL é utilizada para declaração, asserções e criação de classes, propriedades e indivíduos. Asserções podem ser declaradas em uma ontologia simples, ou em uma combinação de múltiplas ontologias relacionadas. Ainda com relação às asserções explícitas, fatos adicionais

podem ser derivados ou implicados logicamente como resultado da inferência (LACY, 2005).

A linguagem OWL é um dos alicerces da arquitetura da Web Semântica (Figura 8), que ainda inclui RDF, RDFS, SPARQL.

Nesse sentido, a modelagem da ontologia desenvolvida seguiu os padrões definidos no Capítulo 4, para a representação do conhecimento de Angiospermas no âmbito da Sistemática Vegetal.

Além disso, a modelagem da ontologia seguiu os padrões definidos pela Methontology (FERNÁNDEZ *et al.*, 1997).

4.2.2 Methontology

Para a definição, a modelagem e a criação da ontologia desse trabalho, foi utilizado o Methontology – uma metodologia para a construção de ontologias, baseada no ciclo de vida de prototipagem, o que permite voltar para qualquer estágio do desenvolvimento (especificação, conceitualização, formalização, implementação e manutenção) e inserir novas definições e/ou corrigir erros a qualquer momento (FERNÁNDEZ *et al.*, 1997; RAUTENBERG *et al.*, 2009).

Methontology é uma metodologia idealizada por um grupo de pesquisa em Inteligência Artificial da Universidade Politécnica de Madri, sendo fortemente influenciada por metodologias de Engenharia de *Software* e de Engenharia do Conhecimento (GÓMEZ-PÉREZ *et al.*, 2004).

Segundo Breitman (2005), o Methontology é baseado no processo-padrão IEEE para o desenvolvimento de *software*. Além disso, o processo de desenvolvimento de ontologias referencia quais as atividades que devem ser cumpridas ao se construir ontologias. Segundo seus autores, Mariano Fernández, Assunción Gómez-Perez e Natalia Juristo (1997), as atividades do Methontology podem ser classificadas em três grupos (Figura 9):

1. **Atividades de gerenciamento de ontologias:** Elaboração de cronogramas, controle, garantia de qualidade.
2. **Atividades ligadas ao desenvolvimento de ontologias:** Estudo do ambiente, estudo de viabilidade, especificação, conceitualização, formalização, implementação, manutenção e uso.
3. **Atividades de suporte:** Aquisição do conhecimento, avaliação, integração, documentação, gerência da configuração, alinhamento.

Ao se inspirar em preceitos do ciclo de vida de *software*, as atividades privilegiadas pelo Methontology são destinadas ao gerenciamento, desenvolvimento e suporte do ciclo de vida de uma ontologia, tendo como característica principal a utilização de um rico conjunto de artefatos de documentação (RAUTENBERG *et al.*, 2009). Além disso, o Methontology também trata-se de um *framework* que dá suporte à construção de ontologias no nível do conhecimento (GÓMEZ-PÉREZ, 1996; BREITMAN, 2005). Diferentemente das demais metodologias, o Methontology descreve a identificação do processo de desenvolvimento da ontologia, dividindo-o em tipos de atividades a serem desenvolvidas.

Ao iniciar o projeto de uma ontologia, o Methontology (Figura 9) julga necessário seguir alguns passos (BREITMAN, 2005):

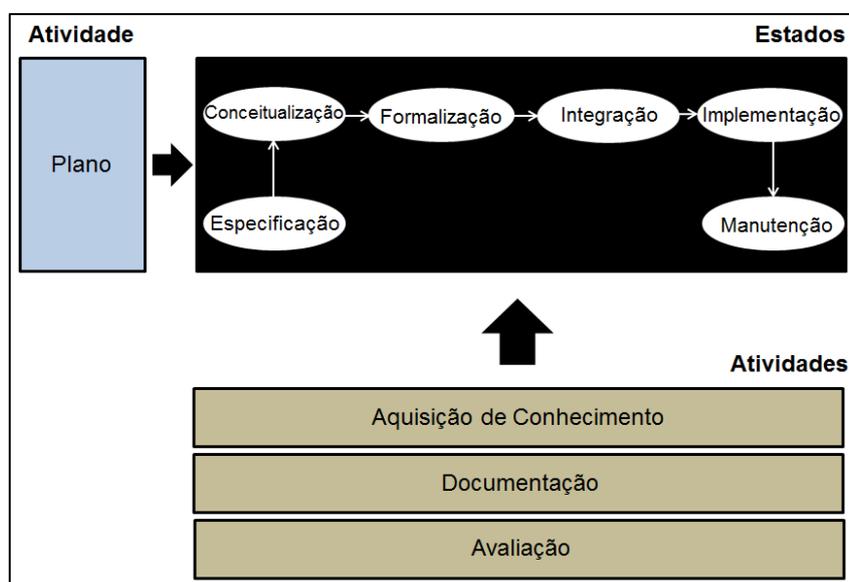
- **Plano:** Ao iniciar o projeto de uma ontologia é necessário planejar as tarefas que serão realizadas.
- **Especificação:** Nessa fase são definidos o escopo e os objetivos da ontologia.
- **Conceitualização:** Nessa etapa é realizado o levantamento dos termos da ontologia.
- **Formalização:** Após os demais passos, deve-se formalizar o modelo conceitual da fase anterior (conceitualização) por meio de

uma linguagem formal para descrição de ontologias (nesse caso, será utilizada a linguagem OWL – descrita na seção anterior).

- **Integração:** É realizada a integração do modelo em desenvolvimento com outras ontologias (caso haja necessidade).
- **Implementação:** Desenvolvimento da ontologia passível de processamento, mediante o uso de uma determinada linguagem (nesse projeto – OWL).
- **Avaliação:** Antes de disponibilizar a ontologia deve-se avaliá-la, para garantir sua qualidade e adequação aos padrões.
- **Documentação:** Essa etapa serve para garantir a evolução da ontologia, de modo a permitir que modificações e reusos sejam possíveis no futuro.
- **Manutenção:** Ontologias são teorias sobre o mundo ou parte do mundo (domínio). Da mesma forma que o mundo está em constante evolução, a ontologia também deve sofrer um constante processo de manutenção ou ficará obsoleta.

O processo de desenvolvimento de ontologias proposto pelo Methontology encontra-se ilustrado na Figura 9:

Figura 9: Processo de desenvolvimento de ontologias com o Methontology *.



* Retirado de: Web Semântica – A internet do Futuro – BREITMAN, 2005.

Após a idealização do modelo ontológico, seguiu-se para a implementação da ontologia, em que foi utilizado o *software* Protégé (versão 4.3.0) (STANFORD UNIVERSITY, 2014).

4.2.3 Protégé

O Protégé é um ambiente, de uso livre e com código-fonte aberto, utilizado para editar e criar ontologias com total suporte à OWL e conexões diretas de acesso à memória para descrição lógica de inferências (*reasoners*) como os *plugins* “HerMiT” e “Pellet” (STANFORD UNIVERSITY, 2014). O *framework* em questão foi desenvolvido pela *Stanford Medical Informatics*, e dá suporte ao desenvolvimento de aplicações baseadas em conhecimento.

Protégé suporta a criação e edição de uma ou mais ontologias em um mesmo ambiente de trabalho (*workspace*). Além disso, propicia ferramentas para navegar entre os relacionamentos das ontologias, auxilia no rastreamento de inconsistências, refatoramento de ontologias, renomear múltiplas entidades, entre outros (STANFORD UNIVERSITY, 2014).

Além disso, permite a extensão de objetos gráficos de interface para tabelas, diagramas e componentes, a fim de acessar outros sistemas baseados em conhecimento embutidos em aplicações. A construção de ontologias usando o Protégé é feita a partir da criação de uma hierarquia de conceitos ou classes. Além da definição dos conceitos/classes, também é possível definir propriedades e relações entre eles (NOY e MCGUINNESS, 2006).

4.3 Especificação e Conceitualização do Conhecimento

Após o levantamento das famílias de Angiospermas representadas na ontologia, um estudo sobre suas estruturas morfológicas foi realizado a fim de representar um modelo de domínio referente à sistemática vegetal.

Para tanto, a modelagem se baseou no sistema de identificação de plantas APG III (*Angiosperm Phylogeny Group*, 2009) – um sistema de taxonomia vegetal moderno utilizado na classificação de plantas com flor; e na 3ª edição do livro “Botânica Sistemática” – Souza e Lorenzi, 2009 – um guia ilustrado para identificação das famílias de Angiospermas da flora brasileira, baseado em APG III, que traz um conjunto de famílias descrevendo quais estruturas morfológicas estão presentes na planta, facilitando assim, sua identificação mediante o uso de uma chave dicotômica.

4.3.1 Sistemática Vegetal

A Botânica Sistemática é o ramo da ciência que estuda a diversidade das plantas, mediante sua organização em grupos, com base em suas relações evolutivas (SOUZA, LORENZI, 2005).

Somente por meio da filogenia pode-se começar a entender a diversificação, as regularidades em padrões evolutivos, ou simplesmente sugerir mudanças evolucionárias dentro de um determinado clado (MOBOT, 2014).

Não há como identificar de forma adequada as famílias de Angiospermas sem possuir o conhecimento de alguns termos morfológicos. Deve-se levar em consideração que as famílias reconhecidas aparecem como grupos naturais nas análises filogenéticas e que, portanto, não foram reconhecidas de forma aleatória ou no intuito de facilitar a identificação. Mesmo assim, por terem um histórico evolutivo em comum, compartilham diversas características morfológicas, o que permite a identificação das famílias, apenas observando a olho nu ou com o auxílio de uma lupa (SOUZA e LORENZI, 2007).

Desde a última década do século XX, houve uma grande mudança nos paradigmas da sistemática de plantas, especialmente devido ao avanço de estudos moleculares, o que tem gerado a modificação dos sistemas de

classificação de Angiospermas. Além disso, alguns posicionamentos das famílias ainda não são estáveis (SOUZA e LORENZI, 2005).

Já que as famílias reconhecidas em publicações de maior aceitação entre os botânicos aparecem como grupos naturais nas análises filogenéticas, também apresentam um histórico evolutivo em comum e compartilham diversas características morfológicas (SOUZA e LORENZI, 2007).

Vale citar o fato de que a identificação de família na taxonomia proposta por Linnaeus encontra-se em um nível superior de gênero e espécie, sendo mais generalista do que os dois últimos citados. É importante citar que a classe de imagens não faz nenhuma referência com o táxon “Classe” da taxonomia. Nesse sentido pode-se traçar um paralelo da classe de imagens com a classificação de “Espécie”. Abaixo um exemplo da taxonomia aplicada:

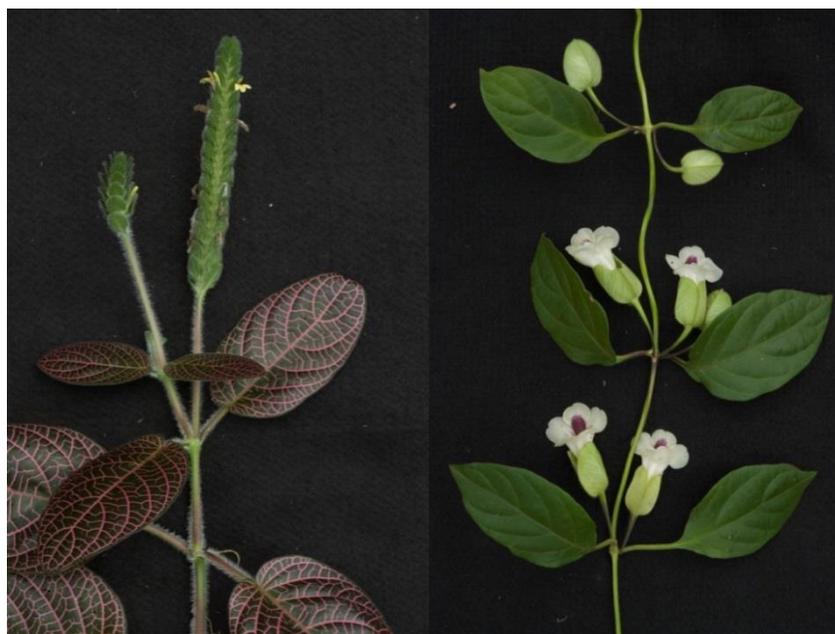
Reino → Filo → Classe → Ordem → **Família** → Gênero → Espécie

A dificuldade da identificação de famílias de Angiospermas ainda é decorrente da questão visual. Muitas famílias têm como representantes espécies visualmente diferentes, com relação à cor, forma e textura. Por exemplo, a família “Acanthaceae”. Pode-se observar tal diferença na Figura 10, já que a planta da esquerda (*Fittonia albivenis*), apesar de ser da mesma família da planta da direita (*Mendoncia puberula*), são visualmente diferentes. Por isso, a importância do conhecimento das estruturas e das propriedades. Os autores Souza e Lorenzi (2007) afirmam que:

“Não há como identificar de forma adequada as famílias de Angiospermas sem possuir o conhecimento de alguns termos morfológicos”

Tal afirmação ajuda a dimensionar o quão desafiador é a pesquisa relacionada à identificação automática de espécies e famílias de Angiospermas.

Figura 10: Diferença visual entre duas plantas da família “Acanthaceae” *.



Em um país com as dimensões do Brasil, que abriga a maior biodiversidade do mundo, os estudos sobre nossa flora são estratégicos, tanto no que se refere ao seu melhor aproveitamento econômico, conciliado com as necessidades das populações, quanto em relação à necessidade de conservação para garantir que estes recursos naturais também estejam disponíveis para as gerações que estão por vir (SOUZA, FLORES e LORENZI, 2013).

Nomear partes de uma planta é uma tarefa recorrente e atemporal em toda e qualquer cultura humana, principalmente pela necessidade de discriminar as partes usadas como remédio, alimento, etc. Entretanto, o início da formulação da terminologia descritiva em plantas coube ao filósofo grego Teofrasto de Ereso (378-287 a.C.), discípulo de Aristóteles e comumente denominado “Pai da Botânica” (GONÇALVES, LORENZI, 2007).

* Crédito das Imagens: Harri Lorenzi.

* Imagens cedidas por: Harri Lorenzi, Instituto Plantarum – Nova Odessa / SP.

Grande parte da morfologia atualmente utilizada teve suas bases na *Philosophia botanica* de Linnaeus (1753). Tal obra, escrita em um latim acessível, lançou as bases da morfologia moderna. Apresentava-se na forma de organografia, tendo sido reeditada várias vezes. As estruturas descritas (e muitas delas ilustradas) já eram apresentadas divididas em partes vegetativas e reprodutivas, e ocasionalmente grupos taxonômicos em que tais estruturas poderiam ser encontradas eram citados (GONÇALVES, LORENZI, 2007).

Em Ciências Biológicas, a taxonomia, ramo que trata da descrição, classificação, nomenclatura e agrupamento dos seres vivos, segundo suas relações filogenéticas, utiliza uma ferramenta denominada chave dicotômica.

Chave dicotômica é um conjunto de descrições sistemáticas que permitem identificar os nomes dos organismos pertencentes a um grupo. Essa chave, em geral, oferece duas alternativas em cada destinação, e a escolha de uma das alternativas determina a etapa seguinte (Figura 11).

Figura 11: Exemplo de chave dicotômica *.

1. Flores aclamídeas.....	2
1. Flores monoclamídeas	12
2. Folhas opostas ou verticiladas.....	3
2. Folhas alternas, escamiformes ou ausentes..	5
3. Ovário unilocular	Piperaceae
3. Ovário tri-quadrilocular	4
4. Plantas em geral com látex; ovário trilocular; fruto geralmente com deiscência explosiva (tricoca).....	Euphorbiaceae
4. Plantas sem látex; ovário tetralocular pelo desenvolvimento de um falso septo; fruto com deiscência não explosiva.....	Casuarinaceae
5. Inflorescência do tipo espádice	Araceae
5. Inflorescência de outros tipos.....	6

Quando há a necessidade de identificar um determinado organismo, muitas vezes o biólogo utiliza uma chave dicotômica. Porém, o esforço para

* Retirado de: Chave de Identificação – SOUZA e LORENZI, 2007

percorrê-la de modo correto exige um grande conhecimento das estruturas e das propriedades daquela amostra a ser identificada.

A Figura 11 traz parte de uma chave dicotômica de flores aclamídeas ou monoclamídeas. Para cada entrada dessa chave, existem dois caminhos a serem seguidos, por exemplo: na entrada 1, se a flor for aclamídea deve ser seguido para a entrada 2, senão, se ela for monoclamídea deve ser seguido para a entrada 12. Ainda seguindo esse exemplo, se na entrada 2, a amostra for identificada com folhas opostas ou verticiladas, tem-se como resultado a família “Piperaceae”.

A chave exemplificada pela Figura 11 traz somente famílias que apresentam flores aclamídeas ou monoclamídeas, porém, a complexidade dessa chave é muito grande, já que apresenta 117 entradas. Os autores dessa chave também criaram mais duas chaves – uma para flores diclamídeas dialipétalas, e outra para flores diclamídeas gamopétalas, cada uma delas com respectivamente 188 e 97 entradas, ou seja, totalizando 402 entradas. Vale ressaltar que para a criação dessas chaves foi utilizado um processo de otimização, de modo a minimizar o número de entradas, mas mesmo assim, atinge um valor muito alto.

Pode-se afirmar que a área do conhecimento da sistemática vegetal está passando por uma das suas épocas mais produtivas, em termos de geração de novos conhecimentos, especialmente no que se refere às relações evolutivas entre os táxons. Tudo isso está acontecendo em um ritmo tão acelerado que mesmo pesquisadores da área têm tido dificuldade em acompanhar as mudanças, tal a velocidade que elas ocorrem. Paralelamente, há uma falta de livros-texto que permitam o acesso dos estudantes a esse conhecimento e, assim, também o processo de aprendizagem fica dificultado pela falta de uma centralização das informações (SOUZA, LORENZI, 2005).

A fim de suprir parte dessa necessidade, o presente trabalho, além da abordagem proposta, define conceitos e informações relacionadas às estruturas e propriedades necessárias para identificar uma família de Angiosperma. Ao definir uma ontologia também fica claro que a abordagem também se propõe a se estabelecer nos moldes da Web Semântica.

4.4 Características da Ontologia

Como os conjuntos de dados estudados apresentam diferentes imagens e conseqüentemente diferentes famílias, a recuperação do conteúdo da ontologia dependerá de qual *dataset* for analisado para os experimentos.

Desse modo, o conjunto de dados *Oxford Flowers 17 Classes* apresenta 1360 imagens de 17 classes, sendo que tais classes referenciam 8 diferentes famílias de Angiospermas. São elas: Amaryllidaceae, Asparagaceae, Asteraceae, Iridaceae, Liliaceae, Primulaceae, Ranunculaceae e Violaceae.

Para modelar as 8 famílias citadas acima, foram implementadas 250 classes, 45 *object properties* e 1 *datatype property*.

Já para o conjunto de dados *Oxford Flowers 102 Classes*, são referenciadas 47 famílias de Angiospermas, entre 8189 imagens de 102 classes de imagens.

As famílias das 102 classes de imagem são: Acanthaceae, Alstroemeriaceae, Amaranthaceae, Amaryllidaceae, Apiaceae, Apocynaceae, Araceae, Asparagaceae, Asteraceae, Bignoniaceae, Brassicaceae, Bromeliaceae, Campalunaceae, Cannaceae, Caryophyllaceae, Convolvulaceae, Dipsicaceae, Ericaceae, Euphorbiaceae, Fabaceae, Gentianaceae, Geraniaceae, Iridaceae, Lamiaceae, Liliaceae, Magnoliaceae, Malvaceae, Myrsinaceae, Nelumbonaceae, Nyctaginaceae, Nymphaeaceae, Onagraceae, Orchidaceae, Papavaraceae, Passifloraceae, Plantaginaceae, Polemoniaceae, Primulaceae, Proteaceae, Ranunculaceae, Rosaceae, Solanaceae, Strelitziaceae, Theaceae, Tropaeolaceae, Violaceae e Zingiberaceae.

Para mapear os conceitos vinculados às 47 famílias do maior conjunto de dados estudado foram modeladas na ontologia 429 classes, 66 *object properties* e 2 *datatype properties*.

Como a ontologia desenvolvida apresenta uma complexidade muito alta, a próxima seção demonstra alguns conceitos importantes para avaliar a escolha dessa modelagem.

4.5 Modelagem da Ontologia

Todo o processo de modelagem do domínio das famílias de Angiospermas envolve uma análise profunda das estruturas e das partes da planta. Para tanto, os conceitos foram categorizados em classes baseadas em suas características comuns. Nessa visão, uma classe é definida através de uma série de propriedades e a condição básica para se pertencer a uma dada classe é possuir tais propriedades.

A fim de relacionar as classes definidas na ontologia desenvolvida, foram criados vários relacionamentos específicos (*object properties* e *datatype properties*). Nesse processo é possível determinar o domínio (*domain*) - quem detém aquela relação; e o alvo (*range*) - quais serão as classes alvo dessa propriedade; aumentando assim a expressividade das relações semânticas envolvidas na ontologia.

Com essas restrições e relacionamentos de classes com suas propriedades e atributos específicos, torna-se possível modelar as características básicas de cada família estudada e, desse modo, definir os conceitos vinculados ao domínio da ontologia.

A ontologia em questão trata de conceitos referentes às estruturas morfológicas de cada família escolhida para o presente trabalho. Em sistemática vegetal, estruturas são utilizadas para diferenciar as famílias de Angiospermas.

Desse modo, esse trabalho seguiu as estruturas e propriedades definidas pelos livros “Botânica Sistemática” (SOUZA e LORENZI, 2009) e “Morfologia Vegetal” (GONÇALVES e LORENZI, 2007).

Vale ressaltar a dificuldade de se representar o conhecimento e codificá-lo para a máquina, já que a linguagem natural é mais direta em suas sentenças, enquanto que a modelagem conceitual exige que os relacionamentos entre os conceitos sejam mais detalhados.

Para dar início à modelagem da ontologia, partiu-se do princípio de que o domínio especificado são as Angiospermas – ou seja, plantas que apresentam flores e cujas sementes são protegidas por um fruto.

A fim de exemplificar um caso de uso, será demonstrada parte da definição dos conceitos da família “Araceae” (Figura 12), tanto do modo que é apresentado pelo livro “Botânica Sistemática” (SOUZA e LORENZI, 2009), quanto da maneira que foi modelado na ontologia.

Figura 12: Planta da família “Araceae” *.



“Araceae” é uma família que possui distribuição cosmopolita, incluindo cerca de 100 gêneros e 3000 espécies. No Brasil ocorrem 35 gêneros e cerca de 400 espécies (SOUZA e LORENZI, 2005).

Do ponto de vista econômico, destacam-se espécies utilizadas como ornamentais, principalmente devido à folhagem ou pelas brácteas vistosas (SOUZA e LORENZI, 2005). Incluem-se aí os antúrios (*Anthurium andraeanum* e outras espécies), o copo-de-leite (*Zantedeschia aethiopica*), entre outras (SOUZA e LORENZI, 2005).

Com relação às estruturas que definem sua classificação, os autores Souza e Lorenzi (2005) apresentam os seguintes conceitos sobre a família em questão:

Ervas, frequentemente escandescentes ou epífitas, às vezes aquáticas flutuantes (*Lemma Pistia*), ocasionalmente com látex; ramos simpodiais ou raramente monopodiais;

Folhas alternas espiraladas ou dísticas, às vezes pinatissectas ou palmatissectas, peniparalelinérveas, peninérveas ou paralelinérveas;

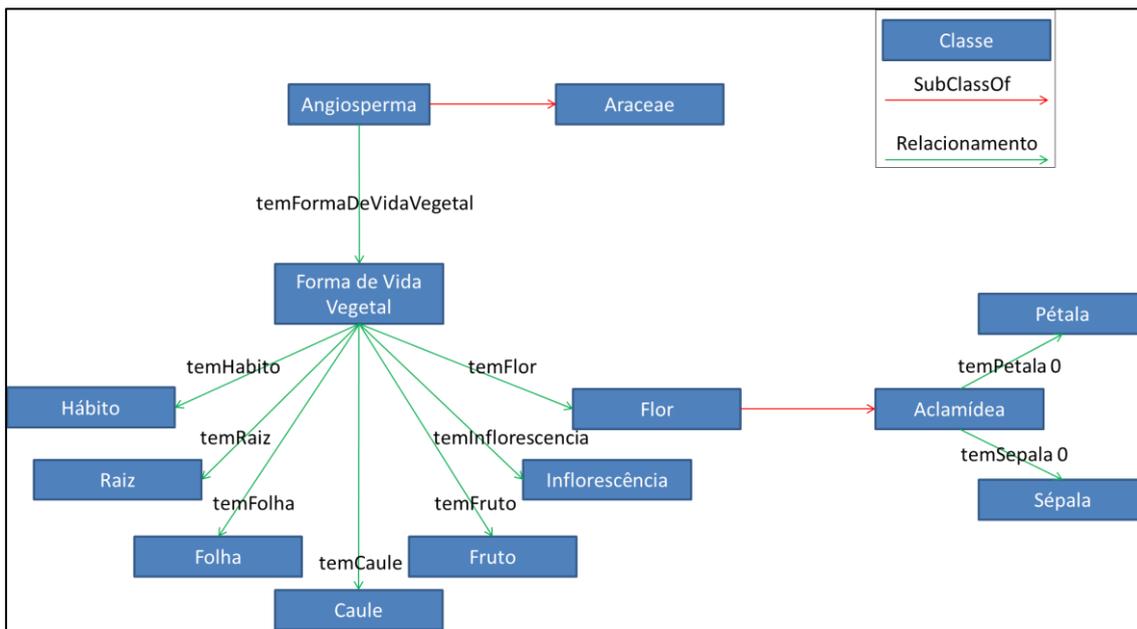
Inflorescência do tipo espádice simples, subtendido por uma bráctea (espata), comumente vistosa;

Flores pouco vistosas, bissexuadas ou unissexuadas (plantas geralmente monóicas), actinomorfas, **aclamídeas** ou monoclamídeas; Cálice 4-6(-8)-mero, gamossépalo ou dialissépalo, prefloração valvar ou imbricada; Estames (1-)4-8(-12), livres ou unidos entre si, anteras poricidas ou rimosas; **gineceu gamocarpelar**, ovário súpero, 2-3(-pluri)-carpelar, 1-3(-pluri)-locular, placentação axial, lóculos uni a plurióvulados.

Fruto baga, raramente utrículo ou drupa, às vezes concrecido e formando um sincarpo.

Dentre os vários conceitos utilizados, pode-se destacar a classe “Flor”, já que está presente em todas as Angiospermas. No exemplo, a família “Araceae” apresenta flor “Aclamídea” (Figura 13), um tipo específico de flor que não possui pétalas nem sépalas (GONÇALVES e LORENZI, 2007).

Figura 13: Modelagem da flor Aclamídea para a família “Araceae”.



Desse modo, a classe “Aclamídea” foi definida da seguinte maneira no Protégé:

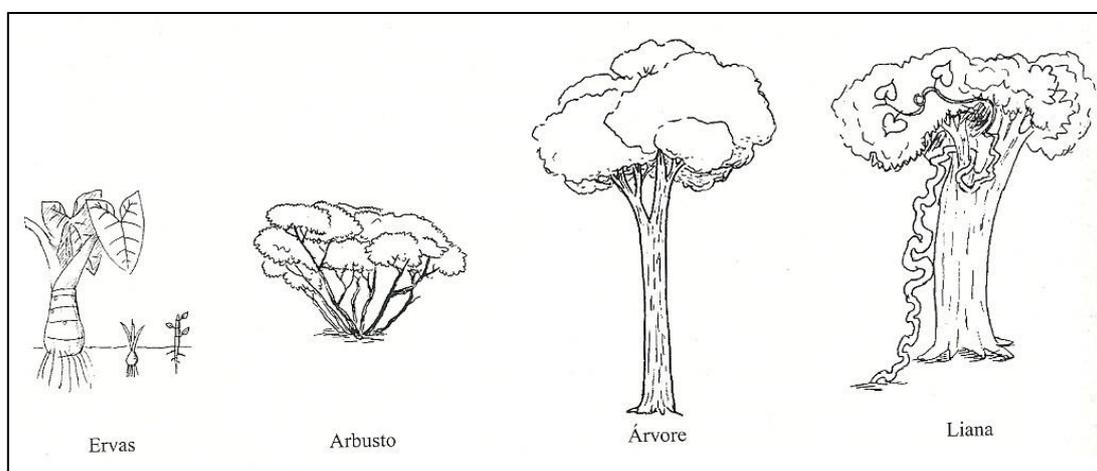
Flor and (*temPetala* exactly 0 **Pétala**) and (*temSepala* exactly 0 **Sépala**)

Ou seja, Aclamídea é um tipo de flor que tem zero pétala e zero sépala.

Assim, ao definir o tipo de flor da família “Araceae”, trata-se do fato de que esta família pode apresentar uma flor do tipo aclamídea, que por sua vez é definida seguindo a declaração acima.

Outro conceito importante que vale ser ressaltado é a “Forma de Vida Vegetal” (Figura 14). Conforme Gonçalves e Lorenzi (2007), ao invadirem o ambiente terrestre, as plantas especializaram-se em diferentes formas de vida. Tanta variação mostrou-se extremamente útil para a montagem de sistemas gerais de classificação das plantas em forma de vida. Uma forma de vida vegetal leva em consideração o organismo vegetal como um todo; a forma que se apresenta.

Figura 14: Forma de vida vegetal *.



Nesse sentido, a classe “Forma de Vida Vegetal” se destaca pela importância dentro da ontologia, já que esta classe se relacionará com os demais órgãos da planta. Além disso, essa classe traz como subclasses: Árvore, Arbusto, Subarbusto, Erva e Liana.

Assim, partindo desses conceitos definidos anteriormente, o fato da família “Araceae” – já definida como subclasse de Angiosperma – apresentar uma flor Aclamídea segue a seguinte declaração:

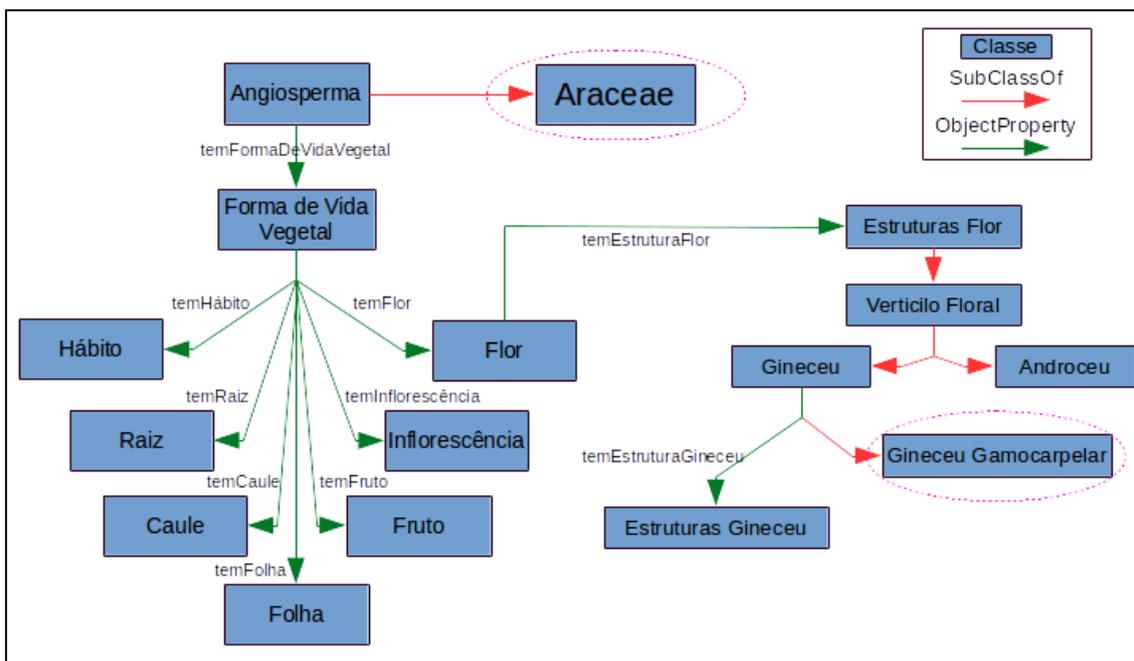
temFormaDeVidaVegetal some (*temFlor* some **Aclamídea**)

* Retirado de: Morfologia Vegetal – GONÇALVES e LORENZI, 2007.

Ou seja, “Araceae” tem dentre outras características, o fato de apresentar uma forma de vida vegetal – nesse caso, uma erva – que por sua vez apresenta uma flor do tipo Aclamídea (Figura 14).

Seguindo a modelagem da família “Araceae”, a Figura 15 exemplifica o fato de tal família de Angiosperma apresentar um “Gineceu Gamocarpelar”. Nesse sentido, “Gineceu” é a estrutura da flor que desempenha o papel feminino na reprodução sexuada (GONÇALVES e LORENZI, 2007). Cada unidade do gineceu é chamada de carpelo, que podem estar livres ou fundidas entre si, formando o pistilo (Figura 16). No caso do Gineceu Gamocarpelar, tais unidades do gineceu encontram-se fundidas em uma única estrutura (GONÇALVES e LORENZI, 2007).

Figura 15: Modelagem do Gineceu para a família “Araceae”.



Pode-se observar na Figura 15 que a classe “Flor” apresenta um *object property* “temEstruturaFlor” (cujo *domain* é a classe “Flor” e o *range* a classe “EstruturaFlor”), que se relaciona com a classe “EstruturaFlor”. A classe “Gineceu”, modelada como subclasse de “VerticiloFloral” implica que o “Gineceu” é um “VerticiloFloral”. A família “Araceae” apresenta um “Gineceu

Gamocarpelar”, implementado com um tipo particular de “Gineceu” (logo com subclasse de “Gineceu”).

Nesse exemplo, o fato da família “Araceae” apresentar “Gineceu Gamocarpelar” é considerado uma restrição, ou um atributo de tal família, de modo que esse atributo é utilizado para identificar tal família de Angiosperma.

Assim, o fato da família “Araceae” apresentar um Gineceu Gamocarpelar é definido na ontologia conforme a seguinte declaração:

temFormaDeVidaVegetal some (*temFlor* some
(*temEstruturaFlor* some **GineceuGamocarpelar**))

Esse tipo de modelagem seguiu para outras estruturas e propriedades, de modo que a Ontologia para o conjunto de dados *Oxford Flowers 17 Classes*, que contém 8 famílias de Angiospermas, determinou 132 atributos utilizados para definir tais famílias. Já para o *dataset Oxford Flowers 102 Classes* foram modelados 350 atributos para 47 famílias.

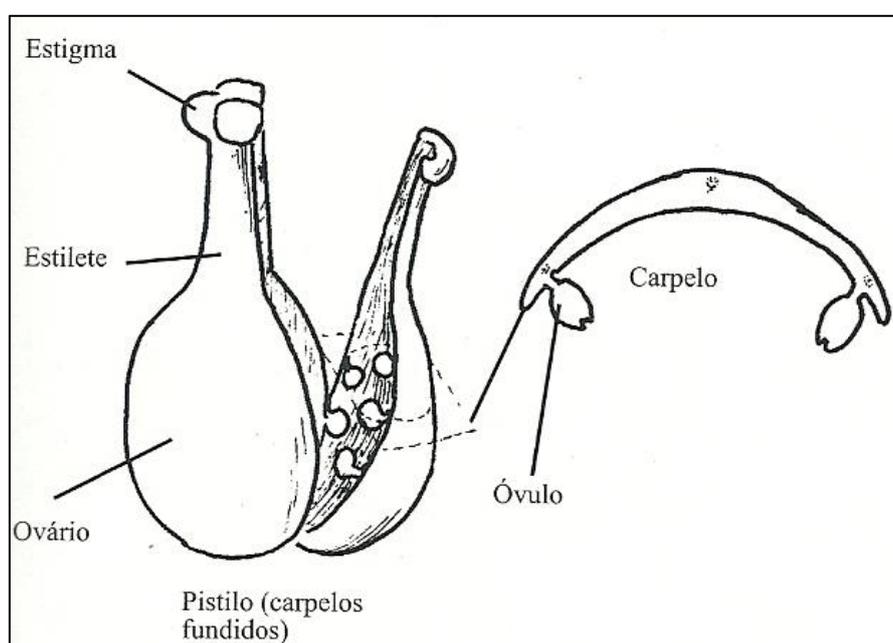
Como alguns órgãos das plantas apresentam estruturas e propriedades específicas que fazem parte desse órgão em questão, foram criadas classes da ontologia que designam estruturas e propriedades de determinado órgão. Essa abordagem se baseia nas definições apresentadas no livro “Morfologia Vegetal” (GONÇALVES e LORENZI, 2007) e na ontologia *Plant Ontology* – uma ontologia que descreve anatomia e morfologia de estágios de desenvolvimento das plantas e também define partes da flor como estruturas (WALLS, 2012).

4.5.1. Estruturas das Plantas

Um bom exemplo para explicar como foi realizada a modelagem, é o ‘Gineceu’ – parte da flor que desempenha o papel feminino na reprodução sexuada. O *Gineceu* apresenta várias estruturas, dentre elas, o *pistilo* (Figura

16) (GONÇALVES, LORENZI, 2007), uma estrutura formada pela soldadura de um ou mais carpelos composto de *ovário* (região basal do pistilo, onde desenvolvem-se os óvulos), *estilete* (porção do capelo definida entre a região que recebe o grão de pólen e a região mais volumosa que porta os óvulos) e *estigma* (porção distal de um pistilo, onde recebe os grãos de pólen trazidos pelo agente polinizador) (GONÇALVES, LORENZI, 2007).

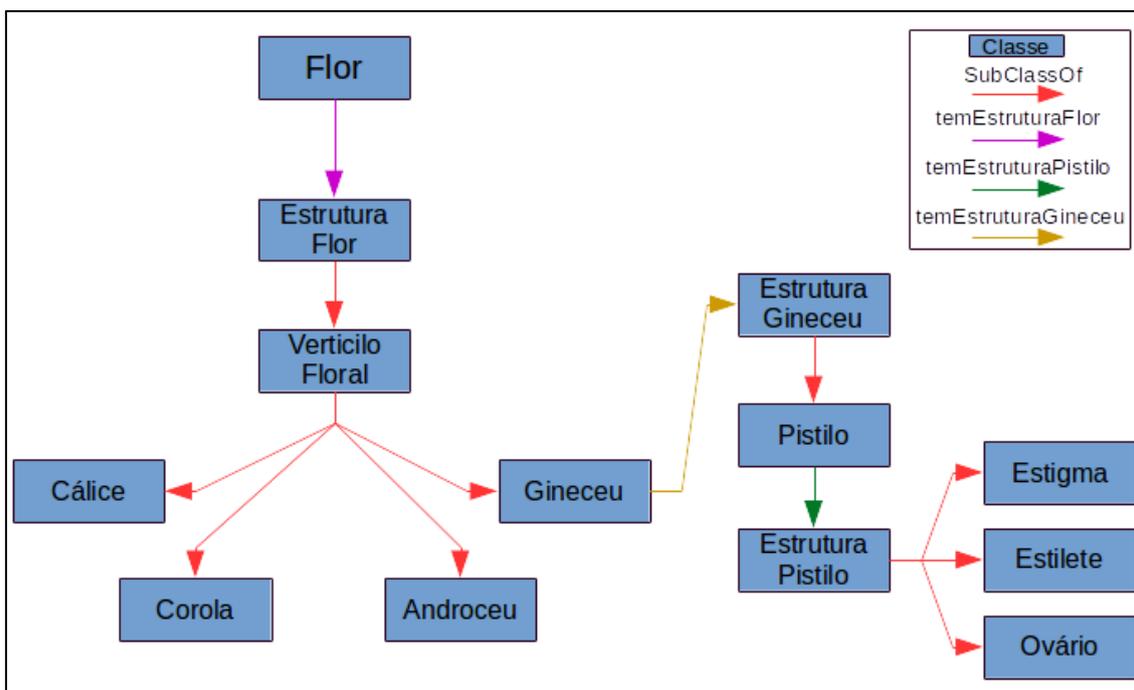
Figura 16: Estrutura do pistilo *.



Muito é dito que uma flor tem um pistilo, mas, quem tem o pistilo é o gineceu, ou seja, o pistilo é uma estrutura do gineceu. Sendo assim, uma flor tem um pistilo de modo indireto, pois uma flor tem um gineceu, que por sua vez apresenta um pistilo. Ainda, uma flor apresenta pistilo se ela for feminina ou hermafrodita, já que, para se ter um pistilo é obrigatória a presença de um gineceu – parte feminina da flor. Essas relações encontram-se sumarizadas na Figura 17.

* Retirado de: Morfologia Vegetal – GONÇALVES e LORENZI, 2007.

Figura 17: Grafo de Flor, até Pistilo.



No caso anterior (Figura 17), observa-se que a classe “Flor” apresenta um *object property* que a relaciona com a classe “Estrutura de Flor”, que por sua vez apresenta um “Gineceu” (subclasse de “Estrutura de Flor”), que é um “Verticilo Floral”, ou seja, “Gineceu” é subclasse de “Verticilo Floral”. Em seguida a classe “Gineceu” é relacionada com “Estrutura de Gineceu”, que por sua vez, apresenta a classe “Pistilo”, como uma de suas subclasses.

A maioria dos *object properties* implementados apresentam um *domain* e um *range* específicos. É importante explicitar que poderia ter sido criada apenas uma propriedade “temEstrutura”, mas desse modo, essa propriedade não poderia ter *domain* nem *range* definidos (já que seria mais abrangente) e isso impactaria diretamente na decidibilidade das inferências, ou seja, no desempenho do *reasoner*. Para resolver essa questão, foram criadas propriedades específicas para cada tipo de relação, e desse modo, tornar possível a restrição de *domain* e *range*. Essa avaliação contribui, além da questão da decidibilidade, para definir restrições, como por exemplo, Flor e Fruto. Com uma propriedade livre, a classe “Fruto” poderia (de maneira errônea) “terEstrutura” a classe “Flor”.

Outras classes que seguem o mesmo princípio das estruturas citadas acima também foram criadas e parte delas encontram-se sumarizadas na Figura 18:

- “Estrutura_Androceu” – Define as estruturas de um Androceu.
- “Estrutura_Antera” – Define as estruturas de uma Antera.
- “Estrutura_Estames” – Define as estruturas de um Estame.
- “Estrutura_Gineceu” – Define as estruturas de um Gineceu.
- “Estrutura_Pistilo” – Define as estruturas que formam um Pistilo.
- “Estrutura_Estigma” – Define as estruturas que formam um Estigma.
- “Estrutura_Ovario” – Define as estruturas de um Ovário.
- “Estrutura_Calice” – Define as estruturas que compõem um Cálice.
- “Estrutura_Corola” – Define as estruturas que compõem uma Corola.
- “Estrutura_Flor” – Define as estruturas de uma Flor.
- “Estrutura_Fruto” – Define as estruturas de um Fruto.
- “Estrutura_Semente” – Define as estruturas de uma Semente.
- “Estrutura_Folha” – Define as estruturas de Folha.
- “Estruturas_Gerais” – Estruturas que podem se relacionar com mais de uma parte ou órgão da planta.

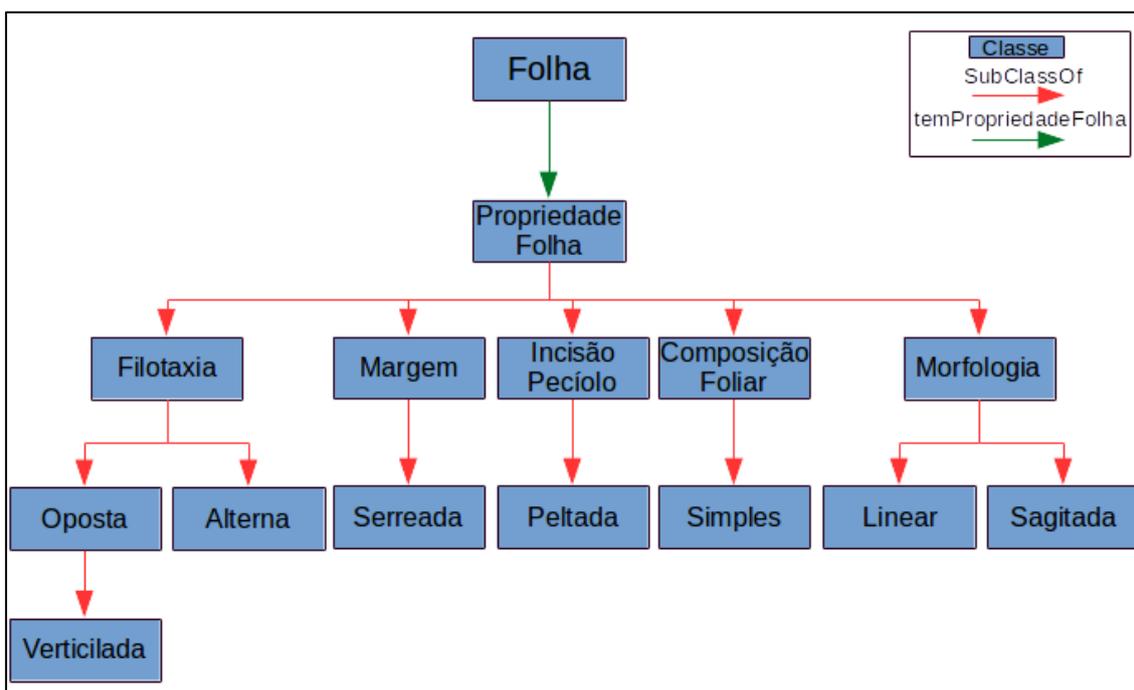
Figura 18: Classes de Estruturas. Laranja – Classes definidas com equivalências. Amarelo – Classes não definidas com relacionamentos.



4.5.2. Propriedades das Plantas

Seguindo o mesmo princípio apresentado pela seção anterior, as classes que definem propriedades de determinado órgão também foram criadas, como por exemplo: a classe “Propriedade de Folha”, em que se definem as subclasses “Composição Foliar”, “Faces”, “Filotaxia”, “Morfologia”, “Incisão do Pecíolo”, “Tipo de Margem” e “Tipo de Venação”. A Figura 19 sumariza parte das subclasses da classe “Propriedade de Folha”.

Figura 19: Relacionamento da Classe “Folha” com a Classe “Propriedades de Folha” e parte de suas Subclasses.



A mesma analogia de relacionamentos pode ser feita da folha. Uma folha alterna, como é explicitado na literatura científica de botânica (GONÇALVES e LORENZI, 2007; SOUZA e LORENZI, 2009), é uma folha, que apresenta uma propriedade de filotaxia alterna (Figura 19), de modo que uma

folha alterna é aquela na qual apenas uma folha é produzida em cada nó do caule (GONÇAVES e LORENZI, 2007).

Assim, uma folha alterna, é uma folha que apresenta uma filotaxia do tipo alterna, que por sua vez é uma propriedade de folha e apresenta um relacionamento exclusivo entre as classes 'Folha' e 'Propriedade_Folha'.

Outras classes de propriedades relacionadas a determinados órgãos de uma planta também são definidas na ontologia "Sistemática Vegetal", e parte delas encontram-se sumarizadas na Figura 20:

- "Propriedade_Caule" – Define as propriedades de um caule.
- "Propriedade_Flor" – Define as propriedades de uma flor.
- "Propriedade_Fruto" – Define as propriedades de um fruto.
- "Propriedade_Inflorescencia" – Define as propriedades de uma inflorescência.
- "Propriedade_Raiz" – Define as propriedades de uma raiz.
- "Propriedades_Gerais" – Define as propriedades que se relacionam com mais de um órgão, ou com a própria planta.

Dentre as subclasses de "Propriedades_Gerais", pode-se citar "Latescente" – cuja definição de acordo com Gonçalves e Lorenzi (2007), é a estrutura, órgão ou planta inteira que produz látex e é capaz de liberá-lo instantaneamente com o rompimento dos tecidos.

Figura 20: Classes de Propriedades. Laranja – Classes definidas com equivalências. Amarelo – Classes não definidas com equivalências.



As restrições e os relacionamentos com as respectivas propriedades e atributos permitiram a modelagem das características morfológicas das famílias estudadas e, portanto, definiram-se conceitos na ontologia, relacionados com o domínio da Sistemática Vegetal.

Vale ressaltar que cada um dos conceitos utilizados para a classificação das famílias tratadas no presente estudo, é modelado na ontologia.

Assim, seguindo a ideia apresentada por Bandikolla *et al.* (2014), na qual rótulos das imagens e informações do texto de páginas da Web foram retiradas, as imagens dos conjuntos de dados de Angiospermas estão relacionadas com suas respectivas famílias, justamente para tornar possível o acesso à ontologia e novas informações relacionadas possam ser investigadas.

Desse modo, as perguntas chave (Figura 1-D e Figura 1 – Usuário) serão realizadas levando em consideração as diferentes estruturas entre as famílias que se encontram nas posições iniciais do *ranking*. Para tanto, o capítulo seguinte apresenta a abordagem proposta capaz de realizar tais análises de diferenciação entre as famílias presentes nas posições iniciais dos *rankings* e selecionar o atributo mais relevante para tal discriminação.

5. Abordagem Interativa Guiada por Semântica para Identificação e Recuperação de Imagens

Este capítulo trata da abordagem proposta utilizada para realizar a integração das abordagens CBIR com os conceitos definidos pela ontologia.

Levando em consideração a dificuldade de se identificar uma planta simplesmente ao analisar uma imagem, é proposta uma nova abordagem, denominada Recuperação Semântica Interativa de Imagens (*Semantic Interactive Image Retrieval – SIIR*). A Figura 1 ilustra a abordagem proposta de forma mais geral, incluindo todas as etapas do processo de recuperação, enquanto a Figura 21 ilustra o módulo de recuperação semântica interativa de forma mais detalhada.

A fim de unir os conceitos definidos na ontologia com as características baixo-nível extraídas das imagens dos conjuntos de dados, a abordagem proposta aplica grafos para analisar o quanto um atributo pode diferenciar as imagens de plantas melhores ranqueadas. Assim, uma pergunta é retornada a um usuário, com o atributo selecionado após análises da ontologia e dos *ranked-lists*.

Essa abordagem traz como inovação a composição de um módulo de interação automático que integra técnicas CBIR e Aprendizado Não Supervisionado, com as técnicas baseadas em conhecimento definidos pela Ontologia (Figura 1 e Figura 21).

5.1 Visão Geral da Abordagem Proposta

O presente trabalho apresenta vários passos de execução e essa seção visa discutir o processo de um modo geral, sendo que, mais adiante, uma explicação mais detalhada sobre cada passo será apresentada.

Nessa seção será descrita a maior contribuição do presente trabalho, já que representa a união dos conceitos das características de baixo-nível com o processo da modelagem do conhecimento, ao guiar um usuário com questões discriminativas a fim de melhorar a eficácia da recuperação de imagens pelo sistema.

O maior desafio da nova abordagem proposta foi utilizar informações relevantes ao usuário que seja relacionada à imagem de busca e seus resultados parciais. Essa abordagem também permite a seleção do melhor atributo que distingue as famílias entre os resultados recuperados dos *ranked lists*, seguido pela resposta do usuário de modo a melhorar a classificação (Figura 21). A fim de resolver a seleção do melhor atributo, o problema em questão foi dividido em diferentes estágios.

A princípio as imagens tiveram suas características de baixo-nível extraídas mediante a análise de descritores – métodos de extração de características como: textura, cor e forma (Figura 1-A). Tais descritores estão disponíveis na ferramenta LIRe (*Lucene Image Retrieval*), uma biblioteca Java para recuperação de imagens. Além do LIRe, também foi utilizada a CNN Caffe (JIA, *et al.*, 2014) para realizar as extrações de características das imagens. Após esses passos, cada imagem é tratada de acordo com seu vetor de características – extraído pelo método utilizado. Em seguida, tais vetores são comparados com todos os outros vetores, também referentes às outras imagens do conjunto de dados, e assim, calcula-se a matriz de distâncias entre as imagens. Quanto menor o valor da distância, mais similares são as imagens comparadas.

Em seguida, a matriz de distâncias é reordenada pelo algoritmo *RL-Sim re-ranking*, que considera a similaridade entre *ranked lists* para decodificar informações contextuais de sistemas CBIR (Figura 1-B). Tal método visa melhorar a precisão da classificação entre as distâncias na matriz, de modo iterativo.

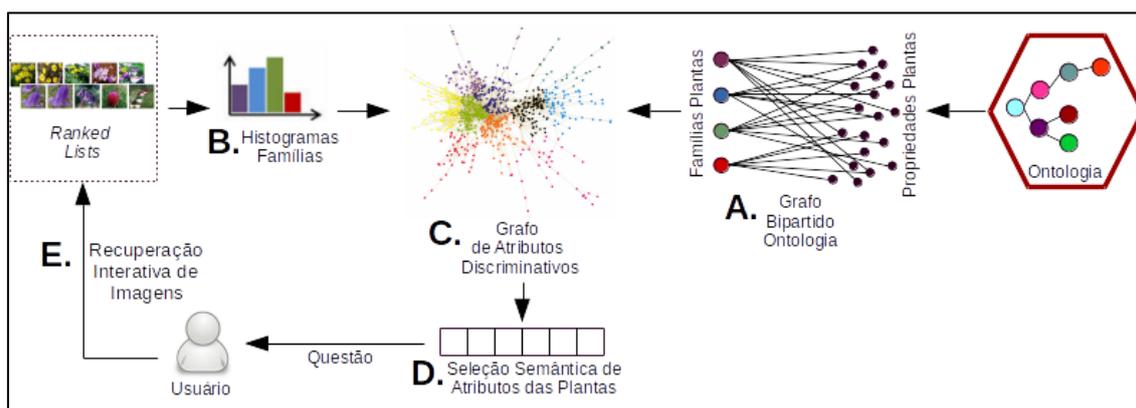
Também foram consideradas tarefas de *Rank Aggregation* (Seção 3.4.2) com o intuito de combinar matrizes geradas por diferentes descritores. A matriz gerada por essa combinação também foi submetida ao processamento do

algoritmo RL-Sim. Após esses passos, espera-se que um conjunto significativo das imagens mais similares esteja presente nas k primeiras posições do *ranked list* avaliado.

Como cada imagem está nomeada e anotada com a devida família para tornar possível a extração de atributos morfológicos do grupo que pertence, foi utilizada a ontologia desenvolvida para diferenciar as imagens que se encontram nas k primeiras posições, a fim de aumentar ainda mais a precisão nesses casos. Para tanto, foi desenvolvido um grafo bipartido (Figura 21-A, Seção 5.3) de modo que de um lado os nós representam as famílias, enquanto do outro as características morfológicas (os atributos) das plantas modeladas. As arestas entre as famílias e os atributos representam a presença daquela estrutura ou propriedade para a família em questão.

A ontologia desenvolvida trata das estruturas morfológicas referentes à classificação das famílias de Angiospermas, e seu uso tem o intuito de codificar o conhecimento da Sistemática Vegetal das famílias presentes nos conjuntos de dados avaliados.

Figura 21: Fluxo do sistema proposto.



A partir desse ponto, também foi analisada a frequência das famílias das imagens presentes nas k primeiras posições de um determinado *ranked list*.

Para tanto, foi criada uma estrutura representando um histograma das famílias (Figura 21-B, Seção 5.4).

Após esse passo, foi desenvolvido um grafo de atributos discriminativos (Figura 21-C, Seção 5.5), em que considera a proporção das famílias de Angiospermas calculadas pelo histograma das famílias (Figura 21-B, Seção 5.4) e a relaciona com o grafo bipartido dos atributos dessas famílias (Figura 21-A, Seção 5.3). O grafo de atributos discriminativos – a maior contribuição desse trabalho – é de extrema importância, já que representa a relação entre as áreas de extração de características de baixo-nível das imagens, com as informações e conceitos definidos na ontologia. Além disso, esse grafo tem como objetivo prover informações para identificação dos atributos com maior potencial para discriminar a família da imagem de consulta em questão.

Após a construção do grafo de atributos discriminativos, é realizada uma análise de modo a selecionar a característica (o atributo) que mais diferencia as famílias nas k primeiras posições (Figura 21-D, Seção 5.6), com o intuito de retornar ao usuário algumas perguntas-chave.

Ao responder a questão (Figura 21-E, Seção 5.7), a matriz de distâncias é atualizada com os novos pesos, de modo que as famílias que se encontram nas primeiras posições e apresentam aquela característica, sejam movidas para o topo do *ranking*, aprimorando assim, a classificação da planta avaliada.

No decorrer do presente capítulo, cada passo do processo será detalhado.

5.2 Ontologia

Conforme demonstrado no Capítulo 4, a modelagem da ontologia (Figura 1-C) foi definida pela criação de um vocabulário específico que trata de conceitos das estruturas morfológicas, relações e restrições para cada família de Angiospermas em um nível suficiente para diferenciar todas as plantas estudadas.

É importante ressaltar o fato de que a abordagem proposta pode ser aplicada em diferentes domínios, não só especificamente no domínio da Sistemática Vegetal. Basta a disponibilidade de uma ontologia que descreva o domínio que se deseja estudar a classificação de um determinado conjunto de imagens pertencentes ao domínio escolhido.

Após a modelagem da ontologia, utilizou-se a OWL API (SOURCE FORGE, 2016) de modo a recuperar as restrições definidas na ontologia para a geração de um grafo bipartido.

5.3 Grafo Bipartido da Ontologia (BOG)

Uma abordagem baseada em grafo é proposta com o objetivo de representar o conhecimento codificado na ontologia. Desse modo, um Grafo Bipartido da Ontologia (*Bipartite Ontology Graph* – BOG) é proposto para definir os relacionamentos entre as famílias de Angiospermas e os atributos das estruturas biológicas que as compõem. A Figura 21-A ilustra tal abordagem proposta para o BOG. Baseado na representação em grafo, o conhecimento codificado na ontologia pode ser explorado no processo de recuperação, guiando as interações com o usuário.

Formalmente, o BOG pode ser definido com um grafo não orientado $G_o = (V_o, E_o)$. Seja $F = \{f_1, f_2, \dots, f_r\}$ o conjunto de famílias de Angiospermas analisado. Seja $A_t = \{a_1, a_2, \dots, a_m\}$ o conjunto de atributos que representam as propriedades e estruturas das plantas modeladas na ontologia. Os nós V_o do grafo são definidos com a união dos conjuntos citados, $V_o = F \cup A_t$.

Uma matriz O , $|F| \times |A_t|$, da ontologia, foi utilizada como uma matriz adjacência para definir o conjunto de arestas E_o do grafo. Desse modo, a matriz O é calculada como (exemplo na Figura 22):

$$O_{ij} = f_o(f_i, a_j) \quad (6)$$

em que f_o é uma função binária calculada pela informação da ontologia, em que retorna 1, caso a família f_i apresente o atributo a_j , caso contrário, retorna 0.

Uma aresta e_{ij} entre a família f_i e um atributo a_j indica a presença daquele atributo para a família analisada. Tecnicamente, a aresta e_{ij} é determinada pela OWL API (SOURCE FORGE, 2016), responsável por realizar a busca do conhecimento relacionado às famílias na ontologia.

Figura 22: Cálculo da matriz O .

Lista Atributos Angiospermas		Atributos (A)			
		0 – Filotaxia Oposta	1 – Inflorescência Capítulo	2 – Folha Margem Serreada	3 – Flor Diclamídea
Famílias Ontologia					
Famílias (F)	F\A	0	1	2	3
0 – Amaryllidaceae	0	0	0	0	1
1 – Asparagaceae	1	1	0	0	1
2 – Asteraceae	2	1	1	1	1
3 – Iridaceae	3	0	0	0	1
4 – Liliaceae	4	0	0	0	1
5 – Primulaceae	5	1	0	1	0
6 – Ranunculaceae	6	1	0	0	1
7 – Violaceae	7	1	0	0	1

Matriz O

A Figura 22 demonstra um exemplo de como é computada a matriz O a partir dos cálculos definidos acima e das informações contidas na ontologia. Com as referências de uma lista de famílias e da lista de atributos das Angiospermas, são traçadas as relações de presença ou ausência para determinada estrutura ou propriedade. Por exemplo, o elemento $O_{2,1}$ apresenta o valor 1, determinando que a família “Asteraceae” apresenta “inflorescência” do tipo “capítulo”.

5.4 Histograma dos *Rankings* das Famílias

Enquanto o grafo da ontologia representa a informação semântica, uma estrutura para modelar as características visuais de baixo-nível também se torna necessária. Os *ranked lists* computados pelo aprendizado não supervisionado podem auxiliar nesse processo.

Para tanto, uma análise foi conduzida considerando as *top-k* posições dos *ranked lists* com o objetivo de calcular a frequência de cada família nos resultados recuperados (Figura 21-B).

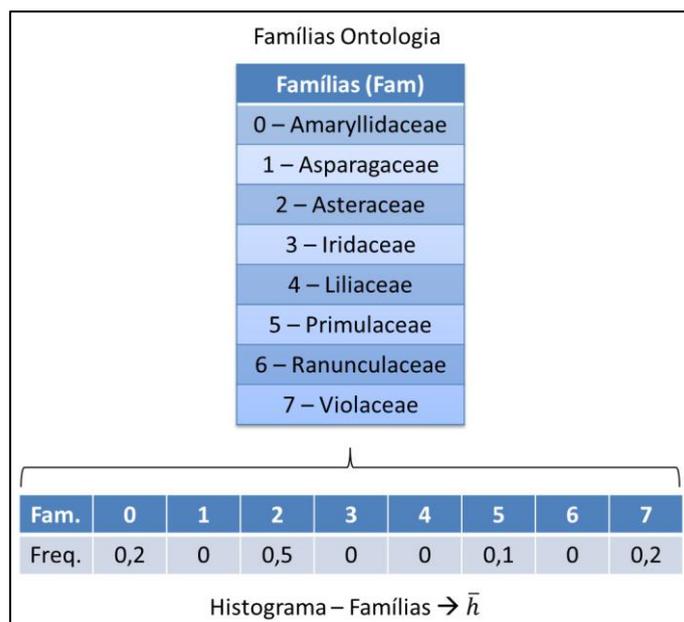
Como as imagens estão anotadas com suas respectivas famílias e, como é possível também identificar o índice de cada imagem na matriz e nos *rankings*, um histograma foi gerado com tal intuito de avaliar a frequência das famílias nas posições iniciais dos *rankings*. O número de índices foi definido de acordo com o número de famílias analisadas no conjunto de dados, como $|F|$.

Seja img_q a imagem de busca. Seja $N(q, k)$ o conjunto de vizinhança que recupera as k imagens mais próximas da imagem img_q , de modo que $|N(q, k)| = k$. Seja $h(i)$ o número de imagens da família f_i no conjunto da vizinhança $N(q, k)$. Seja $\bar{h}(i)$ a frequência normalizada da família f_i , em que pode ser definido com $\bar{h}(i) = h(i) \div k$. Desse modo:

$$\sum_{i=1}^r \bar{h}(i) = 1 \quad (7)$$

em que r representa o número de família, de modo que $r = |F|$. Assim, o histograma \bar{h} é proporcionalmente definido de acordo com frequência de cada família nas posições iniciais e fornece uma informação sumarizada das características visuais extraídas.

A Figura 23 demonstra um exemplo de um histograma para as k primeiras posições.

Figura 23: Histograma para um *ranked list* hipotético.

A Figura 23 deve ser interpretada de modo que, por exemplo, para um valor de $k = 10$, tem-se nas 10 primeiras posições, 5 imagens da família “Asteraceae” (representando uma frequência de $\bar{h}[2] = 0,5$), 2 imagens da famílias “Amaryllidaceae” (frequência de $\bar{h}[0] = 0,2$), 2 imagens da família “Violaceae” (frequência de $\bar{h}[7] = 0,2$) e, finalmente 1 imagem da família “Primulaceae” (frequência de $\bar{h}[5] = 0,1$).

5.5 Grafo de Atributos Discriminativos (DAG)

Cada atributo (estrutura morfológica ou propriedade) das plantas apresenta um potencial distinto para identificar a família do espécime. Por exemplo, um atributo que ocorre em muitas famílias pode ser inútil para propósitos de diferenciação e identificação das famílias. Ainda pode-se afirmar que tal potencial pode variar de acordo com as famílias mais frequentes nas posições iniciais das imagens recuperadas.

Nesse sentido, o Grafo de Atributos Discriminativos (*Discriminative Attribute Graph* – DAG) fornece a estrutura que determinará o atributo mais

discriminativo para propósitos de identificação (Figura 21-C). A ideia principal consiste em combinar a semântica e as informações de baixo-nível em um único grafo. Na verdade, tal passo consiste em uma das tarefas mais desafiadoras da abordagem proposta, em que o problema do *gap* semântico é abordado e minimizado.

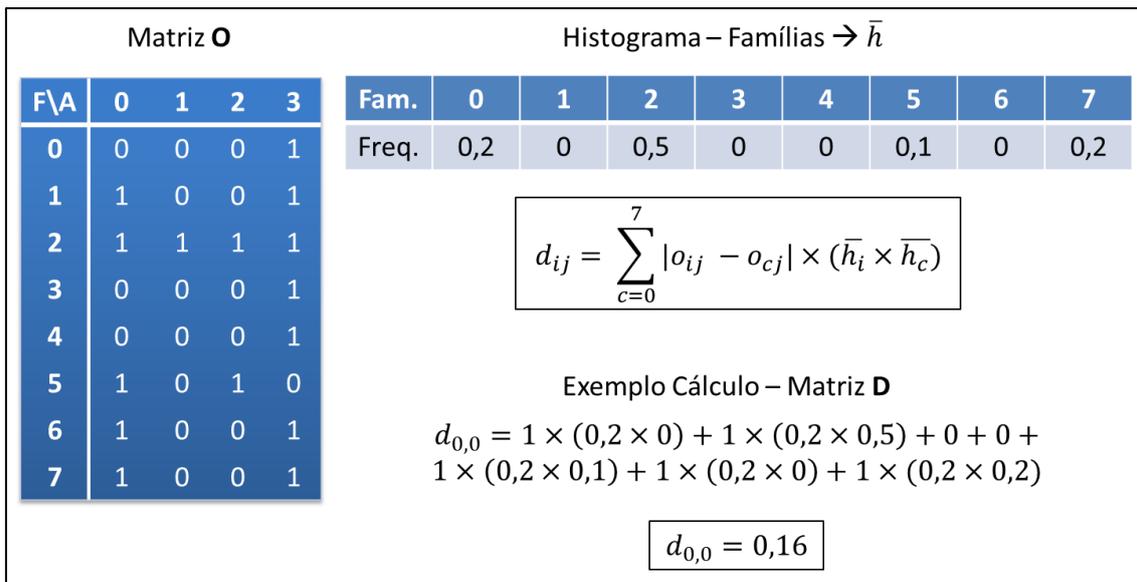
O Grafo de Atributos Discriminativos (DAG – Figura 21-C) combina informações do Grafo Bipartido da Ontologia (BOG – Figura 21-A) e do Histograma dos *Rankings* das Famílias (Figura 21-B). Enquanto o Histograma identifica as famílias mais frequentes nas *top* imagens recuperadas, o BOG é analisado com o intuito de descobrir o atributo mais discriminativo para tais famílias.

Formalmente, o DAG pode ser definido como um grafo não orientado $G_d = (V_d, E_d)$. O conjunto dos nós V_d é definido do mesmo modo que o BOG, como a união dos conjuntos das famílias e atributos, de modo que $V_d = F \cup A_t$. O conjunto de arestas E_d é definido por uma matriz adjacência \mathbf{D} .

Uma aresta e_{ij} entre uma família f_i e um atributo a_j indica a capacidade de tal atributo em discriminar a família f_i das demais famílias. Adicionalmente, o valor da aresta recebe um peso com relação à frequência das famílias presentes nas posições iniciais dos *rankings* das imagens recuperadas, dado pelo histograma \bar{h} . Nesse sentido, a matriz \mathbf{D} é calculada como segue:

$$D_{ij} = \sum_{c=1}^r |O_{ij} - O_{cj}| \times (\bar{h}(i) \times \bar{h}(c)) \quad (8)$$

A Figura 24 apresenta um exemplo de cálculo da matriz \mathbf{D} , baseando-se nos exemplos firmados da matriz \mathbf{O} (Figura 22) e no histograma das famílias demonstrado na Figura 23.

Figura 24: Cálculo da matriz **D**.

Nesse sentido, a Figura 24 demonstra o cálculo do elemento $d_{0,0}$ da matriz **D**. Após o cálculo dos demais elementos, apresenta-se na Figura 25 a matriz **D** completa.

Figura 25: Matriz **D** completa após os respectivos cálculos.

F\A	0	1	2	3
0	0,16	0,1	0,12	0,02
1	0	0	0	0
2	0,1	0,25	0,2	0,05
3	0	0	0	0
4	0	0	0	0
5	0,02	0,05	0,04	0,09
6	0	0	0	0
7	0,04	0,1	0,12	0,02

Matriz **D**

É importante notar que, os elementos que apresentam o valor 0 na matriz D (Figura 25) são referentes às famílias que não estavam presentes no *ranking* hipotético demonstrado na Figura 23.

A partir desse passo, calculam-se as informações adjacentes que podem ser recuperadas da matriz D para a realização da seleção dos atributos mais relevantes que podem diferenciar as famílias presentes no *ranking*.

5.6 Seleção Semântica de Atributos

O grafo DAG (Figura 21-C) combina semântica e informações visuais de baixo-nível para identificar o atributo mais discriminativo para cada família. Nesse sentido, a informação adjacente pode ser utilizada para identificar o atributo mais apropriado que será utilizado no passo de recuperação interativa de imagens. O atributo mais discriminativo é dado pelo nó a_j que apresenta a maior adjacência acumulada, *i.e.*, o atributo que pode fornecer a diferenciação entre as famílias mais frequentes (Figura 21-D).

Formalmente, a função $s(a_j)$ é calculada ao definir a soma das adjacências de um dado atributo a_j . A adjacência acumulada $s(a_j)$ é computada com base na matriz D , conforme segue abaixo:

$$s(a_j) = \sum_{i=1}^m d_{ij} \quad (9)$$

O atributo que apresentar a maior adjacência acumulada é selecionado para compor o passo de recuperação interativa. Uma pergunta é criada para o usuário com relação à presença do atributo selecionado para a imagem de busca.

Nesse sentido, a Figura 26 apresenta um exemplo do cálculo da adjacência acumulada com o intuito de selecionar o atributo mais relevante.

Figura 26: Cálculo da adjacência acumulada.

Matriz D				
F\A	0	1	2	3
0	0,16	0,1	0,12	0,02
1	0	0	0	0
2	0,1	0,25	0,2	0,05
3	0	0	0	0
4	0	0	0	0
5	0,02	0,05	0,04	0,09
6	0	0	0	0
7	0,04	0,1	0,12	0,02

$$s(a_j) = \sum_{i=0}^7 d_{ij}$$

Atrib.	0	1	2	3
Valor	0,32	0,50	0,48	0,18

Seleção Semântica dos Atributos - s

Após esse passo, os atributos são ordenados de maneira decrescente com relação ao seu valor de adjacência e, em seguida, o atributo que apresentar o maior valor é selecionado para retornar uma pergunta ao usuário (Figura 27).

Figura 27: Escolha da pergunta para o usuário.

s – ordenado decrescente					Lista Atributos Angiospermas
Atrib.	1	2	0	3	Atributos (A)
Valor	0,50	0,48	0,32	0,18	0 – Filotaxia Oposta
					1– Inflorescência Capítulo
					2– Folha Margem Serreada
					3– Flor Diclamídea

A família tem Inflorescência Capítulo?

Com isso, seguindo a bateria de exemplos apresentados nas seções anteriores (Figuras 22-27, Seções 5.3-5.6), a estrutura selecionada para a avaliação é se a imagem de busca apresenta a “Inflorescência Capítulo”.

5.7 Recuperação Interativa de Imagens

Finalmente, no último passo da abordagem proposta (Figura 21-E), dado um atributo selecionado a_j , o sistema compõe uma pergunta para o usuário. O usuário, por sua vez, responde “sim”, “não” ou “não sei” para as perguntas, desencadeando um retorno (*feedback*) diferente do sistema para cada situação. A resposta dada pelo usuário é utilizada para atualizar e melhorar os resultados recuperados.

Por exemplo, seja a_j o atributo selecionado para um experimento em particular que indica a presença de uma “Inflorescência Capítulo”. O usuário deve responder: “**A sua planta apresenta Inflorescência Capítulo?**”, supondo que a imagem de busca seja referente à família “Asteraceae”, a resposta deverá ser “sim”. Assim, todas as imagens das famílias que apresentam “Inflorescência Capítulo” terão suas respectivas distâncias atualizadas na matriz A . A distância de tais imagens deve diminuir de modo a serem movidas para as posições iniciais dos *ranked lists*, melhorando a classificação e recuperação dos resultados.

A abordagem utilizada para a atualização da distância se baseia na multiplicação de uma constante $\alpha < 1$.

Seja img_q a imagem de busca. Seja img_j uma imagem de uma planta que apresenta um determinado atributo a_j . A resposta “sim” para a presença do atributo a_j implica na diminuição da distância entre as imagens img_q e img_j . Desse modo, a matriz de distâncias A é atualizada da seguinte maneira:

$$A_{qj} = A_{qj} \times \alpha \quad (10)$$

Se a resposta com relação à presença do atributo for “não”, a atualização da distância segue um princípio similar. Seja img_j uma imagem que não contem o atributo analisado, desse modo, a matriz A deve ser atualizada da seguinte maneira:

$$A_{qj} = A_{qj} \times (1 + \alpha) \quad (11)$$

Nesse sentido, a resposta “não” pode ser inconclusiva, já que uma família pode apresentar mais de uma mesma característica para uma mesma categoria de propriedades ou estruturas morfológicas. Isso pode ocorrer, caso a planta da imagem de busca seja da família “Primulaceae”, por exemplo. Tal família pode apresentar inflorescências “Racemosa”, “Cimosa” ou “Paniculada”. Caso o atributo selecionado a_j seja inflorescência “Racemosa” e a imagem não apresentar tal característica (ou seja, apresente outro tipo de inflorescência), a distância não deve ser afetada de um modo brusco no *ranked list*. Portanto, a soma da constante com o valor 1.

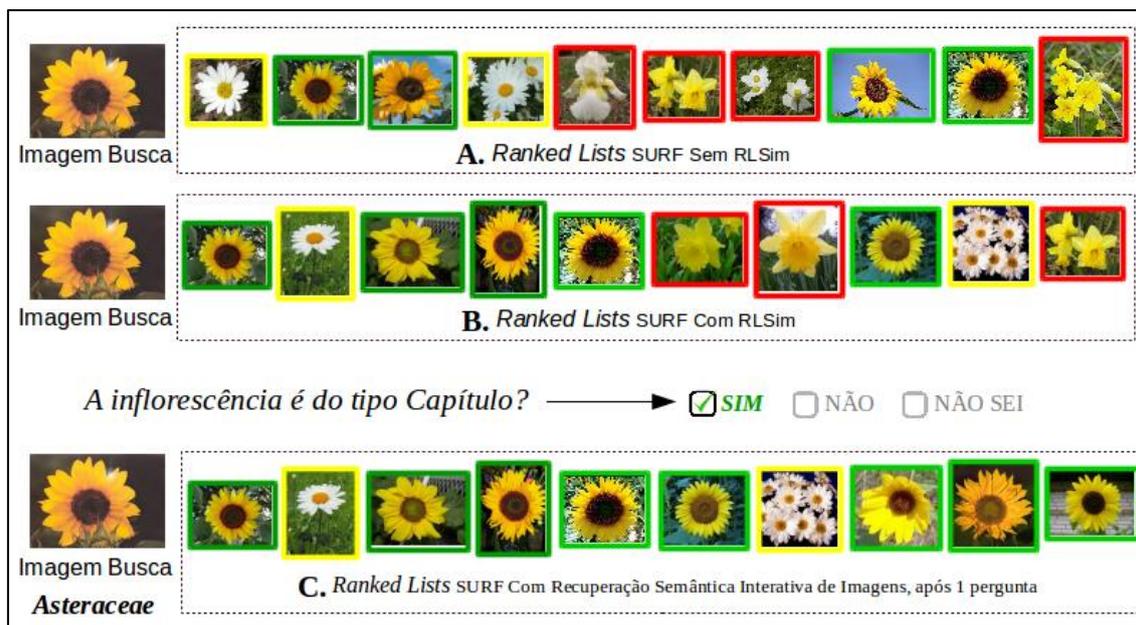
A situação final que pode ocorrer é quando o usuário não sabe a resposta. Nesse caso o próximo atributo com a maior adjacência acumulada é usado para compor uma nova questão.

5.8 Análise Visual da Abordagem Proposta

A Figura 28 demonstra um caso real da evolução do comportamento das dez primeiras imagens de um determinado *ranked list* do *dataset Oxford Flowers 17 Classes*.

Bordas em vermelho representam resultados incorretos das imagens recuperadas, enquanto bordas em amarelo representam resultados incorretos das classes das imagens recuperadas, porém, também demonstram resultados da mesma família da imagem de busca. Já as bordas em verde representam acerto na classe, e conseqüentemente na família, da imagem recuperada.

Figura 28: Comportamento de um *ranked list* após a aplicação da abordagem proposta.



As imagens demonstradas na Figura 28 tiveram suas características extraídas pelo descritor SURF e o ranque 'A' apresenta as *top-10* imagens similares de acordo com esse descritor. Pode-se observar na Figura 28-A que a precisão nas 10 primeiras posições ($P@10$) representa apenas 40%, já que 4 imagens pertencentes à classe da imagem de busca estão presentes nessas primeiras posições. Além disso, ao observar tanto as imagens com bordas em verde e em amarelo, a precisão da família da imagem de busca nas 10 primeiras posições ($PF@10$) atinge 60%.

Após a aplicação do RL-Sim com configuração de $k = 20$ e $it = 2$ para o SURF, a Figura 28-B apresenta um aumento em ambas as métricas analisadas até o momento, já que apresenta $P@10$ igual a 50% e o $PF@10$, 70%.

A Figura 28-C demonstra os resultados para o *ranking* da imagem de busca após a realização de 1 pergunta selecionada pelo sistema. Nesse sentido, a abordagem proposta atinge 80% no $P@$ já que apresenta 8 imagens pertencentes à classe da imagem de busca. Ainda, quando identificada a

família da imagem de busca, atinge-se 100% para o PF@10, já que todas as imagens nas 10 primeiras posições são referentes à família “Asteraceae”.

É importante frisar que com uma única pergunta, a abordagem proposta melhora a recuperação do sistema de modo eficaz quando compara-se os resultados finais (Figura 28-C) e iniciais (Figura 28-A).

A questão “A inflorescência é do tipo Capítulo?” foi escolhida pela *Seleção Semântica de Atributos* (Seção 5.6 – Figura 21-D) já que tal pergunta é a que mais diferencia as famílias das imagens presentes nas posições iniciais do *rank* B. Além do fato de ser uma questão simples de se resolver, já que o conceito “Inflorescência Capítulo” é bem conhecido por biólogos e entusiastas em botânica, tal conceito pode ser também observado na imagem. Ainda, se o usuário quiser consultar a descrição textual ou o grafo descrevendo o que é uma “Inflorescência Capítulo”, pode ser realizada uma consulta simples na ontologia a fim de se obter a informação relacionada.

Vale citar o fato de que esse processo de melhora na recuperação e classificação de imagens ocorre em todo o *ranking*, de modo que a precisão e a acurácia da análise também aumentam para as posições mais profundas do *ranking*.

Por exemplo, para a Figura 28, foram demonstradas apenas as 10 primeiras posições, mas ao analisar as 20 primeiras posições, observa-se que apenas aplicando o descritor SURF, tem-se uma precisão para as 20 primeiras posições (P@20) de 40% e uma precisão das famílias para as 20 primeiras posições (PF@20) de 50%. Quando aplica-se o RL-Sim o resultado do P@20 aumenta para 45% e o PF@20 para 55%.

Já ao utilizar a abordagem proposta, a eficácia é ainda maior, já que após a resposta correta, o P@20 atinge 85% de precisão, já que 17 imagens da mesma classe da imagem de busca estão presentes nas *top-20* posições, e o PF@20 apresenta um valor de 100%.

5.9 Considerações da Abordagem

O ponto principal da abordagem proposta é a relação entre a imagem e seu conteúdo. Ao unir conceitos presentes nas plantas definidas na ontologia com as imagens recuperadas nas primeiras posições dos *rankings* torna possível diminuir o *gap* semântico. A seleção de atributos visa também a menor interação possível do usuário com o sistema, já que tal seleção avalia o atributo que melhor diferencia tais plantas.

Deve-se citar também o fato de que algumas estruturas podem estar ocultas na imagem, mas o sistema eventualmente pode selecionar tal atributo para identificar tal planta presente na imagem. Desse modo, também é interessante o usuário ter a planta em mãos para resolver tal questão, ou simplesmente tratar uma próxima pergunta. Isso pode ocorrer quando a única diferença entre as famílias presentes nas posições iniciais calhar de ser um atributo interno à flor, havendo a necessidade de uma possível dissecação.

Vale citar o fato de que, quando um atributo é selecionado, geralmente, os atributos seguintes ao selecionado também apresentam um potencial semelhante de diferenciação, podendo ser implementada uma lista de atributos discriminativos para o usuário selecionar, ao invés de uma única pergunta por vez.

Apesar da ontologia apresentar uma descrição textual do significado das estruturas e das propriedades das Angiospermas, além de descrever formalmente tais conceitos e disponibilizar um grafo com tais restrições para a realização de consultas, muitas vezes o usuário pode desconhecer determinado conceito e/ou não entender o significado daquela informação. Nesse sentido, trabalhos futuros, também podem levar em consideração o nível de conhecimento do usuário, traçando diferentes perfis, de modo que perguntas mais simples possam ser retornadas, seguindo o mesmo padrão apresentado pela abordagem proposta, a fim de realizar a classificação corretamente. Porém, vale salientar que a eficácia do sistema talvez não atinja os valores que uma avaliação em um âmbito com mais atributos possa atingir.

Outro ponto que é importante salientar é que as questões não foram implementadas em linguagem natural. Porém, tal tarefa seria simples de ser

realizada devido ao fato de que o vocabulário desenvolvido na ontologia controla o *domain* e o *range* de várias relações, facilitando uma análise semelhante à proposta do RDF de “Sujeito → Predicado → Objeto”, aproximando assim o conhecimento codificado para a linguagem natural. Por exemplo: o *object property* “temDeiscencia” define a relação entre o tipo de deiscência que uma antera apresenta. Sendo assim, apresenta como *domain* da relação a classe “Antera” e como *range* o tipo de “Deiscência”. Desse modo, seria simples implementar algo como: “A antera tem deiscência poricida?”.

Como critério de parada para o sistema, o usuário pode decidir quando tem a classificação correta e parar; ou atribuir um número máximo de perguntas realizadas; ou quando o percentual de imagens de uma mesma família atinge um limiar, o sistema encerra aquela classificação. O problema com essa última hipótese é que não se tem a garantia da convergência. Assim, as melhores maneiras de adotar um critério de parada são: 1) Selecionar um número máximo de perguntas; e 2) O usuário decidir quando parar.

6. Avaliação Experimental

Este capítulo traz os resultados das avaliações experimentais conduzidas a fim de analisar a eficácia da abordagem proposta em ambos os conjuntos de dados (*datasets*) analisados.

6.1 Protocolo Experimental

Para analisar a efetividade da Abordagem Interativa Guiada por Semântica para Identificação e Recuperação de Imagens, foram utilizadas como métricas de avaliação a precisão em diferentes profundidades ($P@5$, $P@10$), o MAP (*Mean Average Precision*) e a taxa de acurácia de reconhecimento considerando uma classificação $k - NN$. Além disso, também foi utilizada a curva de *Precisão \times Revocação* (*Precision \times Recall*).

A precisão é a medida definida pela fração entre os objetos relevantes e o total de objetos recuperados em uma dada consulta. A precisão é comumente avaliada considerando os resultados obtidos nas primeiras posições das listas de resultados, de forma que a precisão de um *ranked list* até a n -ésima posição é denotada por $P@n$ (FUJIKAWA e PEDRONETTE, 2014). Por exemplo: Se nas cinco primeiras posições ($P@5$) são encontradas quatro imagens que compartilham a mesma classe da imagem de busca, tem-se uma precisão de 80% (4/5). Essa mesma ideia é aplicada à precisão nas dez primeiras posições ($P@10$), porém a diferença é a quantidade de posições analisadas no *ranked list*, nesse caso, dez.

A outra métrica de eficácia utilizada é o MAP (*Mean Average Precision*). Sabe-se que a medida de eficácia mais comumente utilizada para avaliar sistemas de recuperação de informações tem sido a precisão média (*Average Precision* - AP) (FUJIKAWA e PEDRONETTE, 2014). A fim de calcular a precisão média de um *ranked list*, calcula-se a precisão a cada vez que um

objeto relevante aparece na lista de resultados. Em seguida, calcula-se a média dos valores das precisões obtidas (FUJIKAWA e PEDRONETTE, 2014).

Mais formalmente, o cálculo de AP (*Average Precision*) é realizado da seguinte maneira: seja q um objeto de consulta e N_r o número de objetos relevantes à consulta q . Seja $(r_i | i = 1, 2, \dots, d)$ um vetor de relevância ordenado até a profundidade d , em que r indica a relevância da i -ésima pontuação que o item alcançou, sendo 0 (não relevante) ou 1 (relevante), a precisão média (AP) é definida como (FUJIKAWA e PEDRONETTE, 2014):

$$AP = \frac{1}{N_r} \sum_{i=1}^d \left(\frac{r_i}{i} \sum_{j=1}^i r_j \right)$$

O MAP, por sua vez, é definido ao calcular a precisão média (AP) para uma série de consultas (FUJIKAWA e PEDRONETTE, 2014). Outra característica importante dada pelo MAP consiste na aproximação da área média abaixo da curva de *Precisão* \times *Revocação*, outra medida de eficácia comumente utilizada na literatura. Para a realização dos experimentos do presente trabalho, tanto o MAP quanto a precisão consideram todas as imagens da coleção como sendo uma imagem de consulta.

A curva de *Precisão* \times *Revocação* descreve o quanto representa a precisão até a posição em que foi encontrada a n -ésima imagem mais relevante da classe em questão. Por exemplo, como o conjunto de dados *Oxford Flowers 17 Classes* apresenta 80 imagens por classe, tem-se que, quando a oitava imagem mais relevante de uma determinada classe for encontrada (ou seja, 10% de todas as imagens referentes à classe analisada, ou revocação igual a 0.1), avalia-se o quanto representa a precisão nessa posição. Caso essa oitava imagem tenha sido encontrada na posição 11, deve-se calcular então o P@11. A partir desses resultados, plota-se a curva de *Precisão* \times *Revocação*.

A acurácia é uma métrica na qual toma-se a classe ou a família predominante das imagens presentes nas k primeiras posições e compara-se com a classe da imagem de busca. Caso elas sejam iguais, a classificação é

dada como correta e posteriormente calcula-se o percentual de acerto dentre todas as imagens.

Após a extração de características realizada pelos descritores presentes no LIRe, dois descritores foram selecionados para o prosseguimento do trabalho (ACC e SURF). O BIC também foi selecionado, porém, por meio da utilização do “Eva”, um *framework* de extração de características (PENATTI e TORRES, 2010). O critério de escolha se baseou no fato desses descritores apresentarem os melhores resultados para as métricas testadas. Além disso, para a escolha dos descritores levou-se em consideração os melhores resultados gerados pela união de mais de um método de extração. Desse modo, foram escolhidos os seguintes descritores: ACC, SURF e BIC.

Além das matrizes geradas pelo LIRe, também foi utilizada a matriz de distâncias gerada pelo *CaffeNet*.

Para a realização dos testes, o algoritmo foi executado sem a aplicação do RL-Sim e, com o RL-Sim utilizando uma, duas e três iterações. Para cada iteração, o k (valor de comparação utilizado pelo RL-Sim, em que são escolhidos os k primeiros elementos do *ranked list*) variava de 5 a 35, de cinco em cinco. Desse modo, o k foi testado com, 5, 10, 15, 20, 25, 30 e 35.

As matrizes de distâncias, geradas pelos métodos de extração, foram unidas utilizando o algoritmo *Rank Aggregation* e avaliadas pelo mesmo protocolo experimental.

Para realizar a avaliação da efetividade das perguntas selecionadas, variou-se as questões de 1 a 10, de modo que a interação com o usuário foi simulada, sendo que a *matriz O* foi utilizada para avaliar as respostas, já que tal matriz apresenta todos os atributos vinculados às famílias. Como cada imagem encontra-se anotada com sua respectiva família (inclusive a imagem de busca), a cada pergunta realizada pelo sistema analisa-se a *matriz O* na linha referente à família da imagem de busca e na coluna do atributo presente na pergunta. Desse modo, o valor do elemento analisado da matriz simula a resposta do usuário com relação à presença ou ausência de determinada estrutura ou propriedade.

Tal abordagem de simulação foi selecionada já que foram utilizadas todas as imagens do conjunto de dados como imagens de busca. Nesse sentido, os testes com usuários ficariam desgastantes e difíceis de serem realizados, pois a cada imagem o usuário teria que responder um total de 55 perguntas, já que as perguntas variam de 1 a 10 para cada imagem. Para o conjunto de dados *17 Classes* a interação com o usuário seria 74800, pois são 1360 imagens. No caso do 102 classes, o número de interações com o usuário aumenta para 450395 vezes, pois esse *dataset* apresenta 8189 imagens. Ou seja, com valores altos de interação, seria praticamente impossível analisar todas as imagens e todos os *rankings* para avaliar a eficácia da abordagem proposta.

O protocolo experimental não levou em conta a aleatoriedade em que as perguntas podem ser respondidas. Por exemplo, quantas perguntas seriam respondidas com a resposta “Não Sei”, para que o usuário responda a uma resposta corretamente. A simulação do usuário para o protocolo experimental considera um usuário apto a responder todas as perguntas, de modo que as respostas para os experimentos do presente trabalho são “Sim” ou “Não”, apenas.

As próximas seções estão divididas entre: 1) Conjuntos de dados (*datasets*) utilizados para avaliação da abordagem proposta (Seção 6.2); 2) Valores das métricas aplicadas à extração de características, ao RL-Sim e ao *Rank Aggregation* para ambos os conjuntos de dados utilizados (Seção 6.3); 3) avaliação das métricas para a “Abordagem Interativa Guiada por Semântica para Identificação e Recuperação de Imagens” (Seção 6.4); e 4) Considerações da Abordagem (Seção 6.5).

6.2 Conjuntos de Dados (*Datasets*)

Para realizar o estudo de classificação de imagens, dois repositórios com diferentes resoluções e qualidades de imagens foram utilizados.

O primeiro *dataset* avaliado é o *Oxford Flowers 17 Classes* (Figura 29), disponível para *download* na Web e contém 17 “classes” de diferentes espécies de Angiospermas (VISUAL GEOMETRY GROUP, 2009). Cada classe apresenta 80 imagens, totalizando assim 1360 imagens pertencentes a 8 famílias de Angiospermas no *dataset* em questão.

O segundo conjunto de dados avaliado – também desenvolvido pelo mesmo grupo, é o *Oxford Flowers 102 Classes*, também disponível para *download* na Web e contém 102 “classes” de diferentes espécies de Angiospermas (VISUAL GEOMETRY GROUP, 2009). Nesse *dataset*, o número de imagens varia conforme a classe analisada, de modo que uma classe pode apresentar 40 imagens, enquanto outra 251. No total o conjunto de dados *102 Classes* apresenta 8189 imagens pertencentes a 47 famílias de Angiospermas.

A escolha dos conjuntos de dados se baseou no fato do tema ser semelhante ao estudado neste trabalho e de existir várias publicações na área (NILSBACK e ZISSERMAN, 2006; VARMA e RAY, 2007; NILSBACK e ZISSERMAN, 2008; ITO e KUBOTA, 2010; MATTOS *et al.*, 2014), possibilitando a comparação dos resultados gerados com outros publicados na literatura relacionada.

Figura 29: Quatro exemplos de classes do *dataset Oxford 17 Classes* *.



6.3 Características Visuais e Aprendizado Não Supervisionado

Essa seção apresenta os seguintes resultados: (i) Extração das características visuais; (ii) Extração das características visuais + Aprendizado Não-Supervisionado; (iii) União das características visuais através do *Rank Aggregation*.

Para facilitar a visualização dos resultados, a presente seção é subdividida em outras duas: 1) Resultados das métricas da extração e do RL-Sim para o conjunto de dados *17 Classes* (Seção 6.3.1) e 2) Resultados das métricas da extração e do RL-Sim para o conjunto de dados *102 Classes* (Seção 6.3.2).

6.3.1 Resultados *Oxford Flowers* – *17 Classes*

Após a extração de características, três descritores foram selecionados – ACC, SURF e BIC, além do *CaffeNet* (CNN). A Tabela 2 apresenta os resultados sumarizados para o P@5 inicial e o melhor P@5 de cada um desses métodos e de suas respectivas uniões (*Rank Aggregation*) após a execução do RL-Sim. O valor da precisão varia de 0 a 1.

Vale citar o fato de que o melhor P@5 para cada um dos métodos apresentou diferentes valores de k e de iteração.

Tabela 2: P@5 dos métodos de extração de características e do aprendizado não-supervisionado para o *Oxford Flowers 17 Classes*.

Métodos de Extração	P@5 sem RL-Sim	Melhor P@5
ACC	0,5310	0,5394
SURF	0,5204	0,5518
BIC	0,6001	0,6001
CNN	0,8569	0,8771
ACC+SURF	0,6085	0,6278
CNN+BIC	0,7635	0,7816
CNN+SURF	0,8859	0,9138

Pode-se observar na Tabela 2 que o framework *CaffeNet* (CNN), quando analisado isoladamente, apresentou o melhor resultado de P@5 para o processo de extração de características e para o aprendizado não-supervisionado, atingindo um valor de 87,71% de precisão de classificação dos *rankings* para as cinco primeiras posições, para $k = 15$ e iteração $it = 1$. Isso se deve ao fato de que o *CaffeNet* extrai as características que melhor agrupam as imagens de um determinado domínio e foi treinado com uma das maiores bases de imagens disponíveis na Web, o *ImageNet* - que inclui imagens de ramos floríferos.

Quando analisado os descritores de imagem, observa-se que o BIC apresenta melhor resultado (60,01%) quando comparado com o SURF e o ACC. Tal análise pode ser realizada pelo fato de que o BIC analisa tanto a parte interna quanto a externa das bordas dos *pixels*, enquanto o ACC analisa a imagem como um todo.

O SURF, por sua vez, apresenta melhor resultado (55,18%) do que o ACC (53,94%). Tal análise pode ser vinculada à questão do SURF ser um descritor local baseado em *Bag of Visual Words*, expandindo sua análise em mais de um ponto da imagem, enquanto o ACC é um descritor de cor e analisa a cor do *pixel* em questão. Como muitas flores de Angiospermas de diferentes

famílias apresentam coloração semelhante, o ACC não se mostra tão efetivo quanto os outros métodos utilizados.

Com a união de dois métodos de extração por meio do *Rank Aggregation*, observa-se que houve melhora nos resultados.

A união SURF+ACC demonstra um aumento de aproximadamente 16,39% para o P@5 ótimo (0,6278, $k = 35$ e $it = 1$), com relação ao melhor resultado apresentado pelo ACC nessa métrica (0,5394, $k = 10$ e $it = 1$).

Quando analisada a união CNN+BIC, observa-se que, apesar do CNN e do BIC serem os dois melhores métodos de extração, dentre aqueles testados, para o conjunto de dados *17 Classes*, não houve grande aumento quando comparado ao CNN+SURF. O CNN+BIC atinge um valor ótimo de P@5 de 78,16% com a configuração de $k = 30$ e $it = 1$, enquanto o CNN+SURF apresenta um P@5 ótimo de 91,38% para $k = 15$ e $it = 1$.

Apesar do SURF não ser o segundo melhor descritor de acordo com o P@5, quando unido com o CNN o aumento dos valores do P@5 e das outras métricas analisadas é mais significativo ainda. Isso se deve ao fato de que ambos apresentam características complementares, com acertos em comum e erros diferentes. Esse fato também pode ser observado para as outras métricas da união CNN+SURF (Gráfico 7).

Foram realizadas análises do CNN+ACC, ACC+BIC, CNN+SURF+BIC e do CNN+SURF+ACC, porém, os resultados apresentaram valores abaixo daqueles apresentados pela união CNN+SURF.

A mesma análise realizada para o P@5 foi feita para o P@10 e para o MAP, de modo que os resultados encontram-se sumarizados nos Gráficos 1–7.

Observa-se nos Gráficos 1–7 que a precisão nas dez primeiras posições (P@10) e o MAP também demonstram aumento para os métodos analisados. Porém, com valores menores àqueles apresentados pelo P@5, o que era de se esperar já que conforme se aprofunda no *ranking*, as distâncias aumentam e é possível encontrar maior quantidade de imagens que não pertencem à classe da imagem de busca.

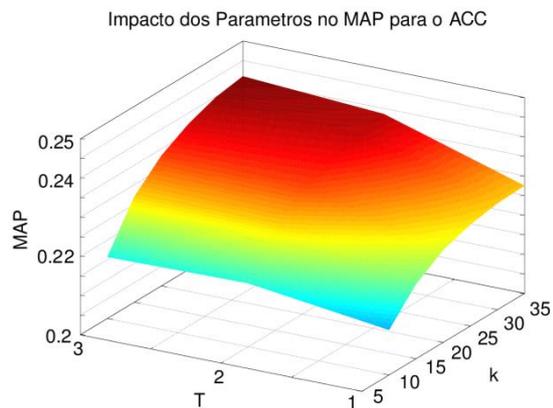
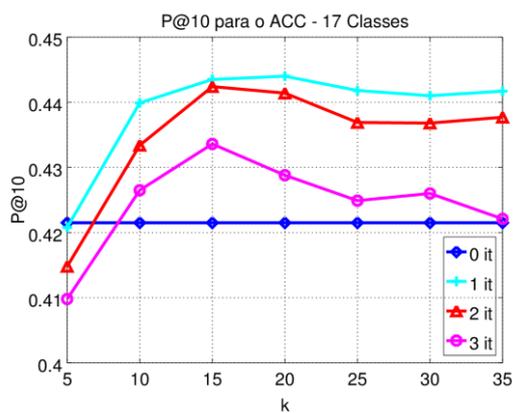
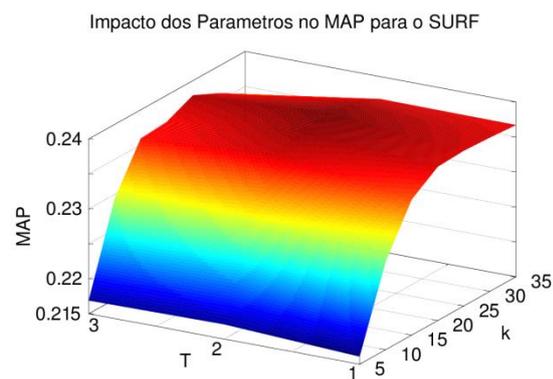
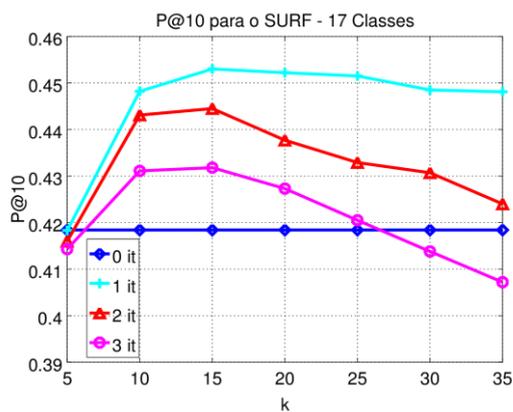
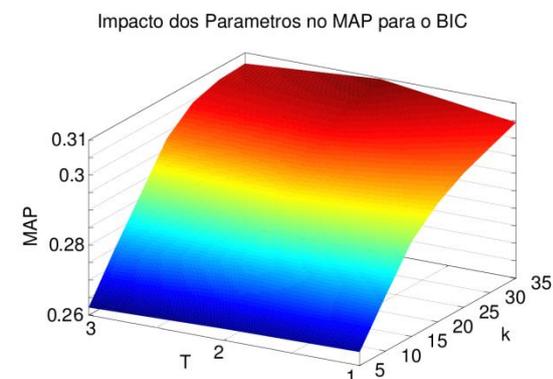
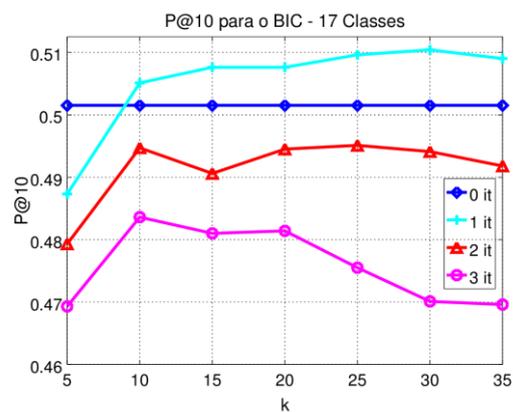
Gráfico 1: P@10 e MAP para o ACC no *dataset 17 Classes*.Gráfico 2: P@10 e MAP para o SURF no *dataset 17 Classes*.Gráfico 3: P@10 e MAP para o BIC no *dataset 17 Classes*.

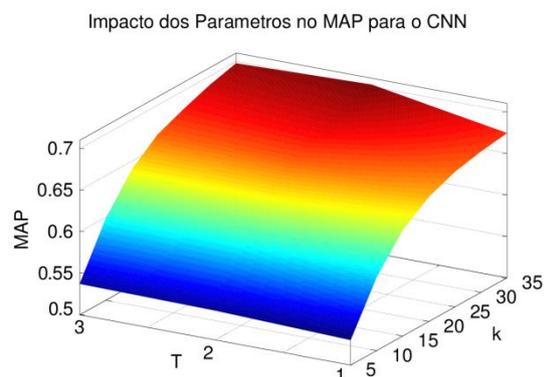
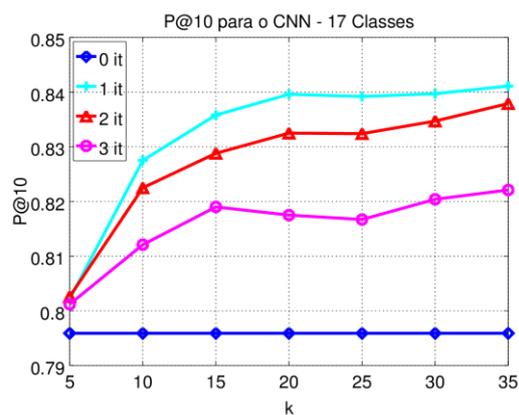
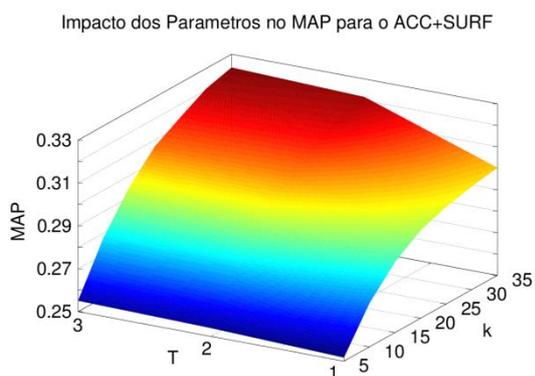
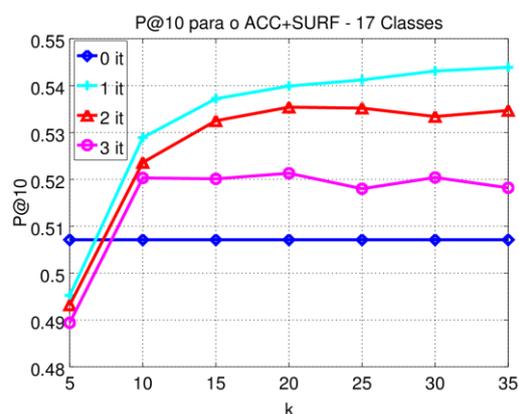
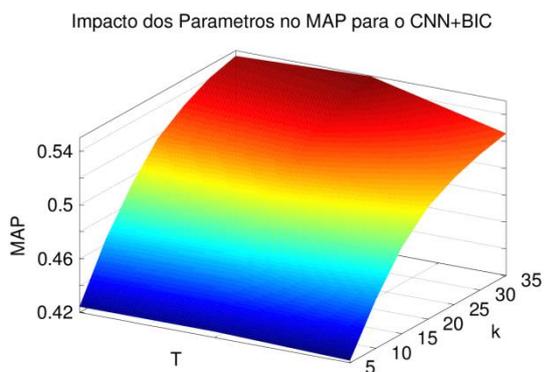
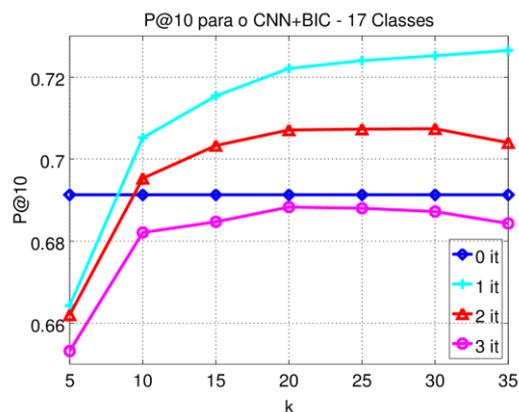
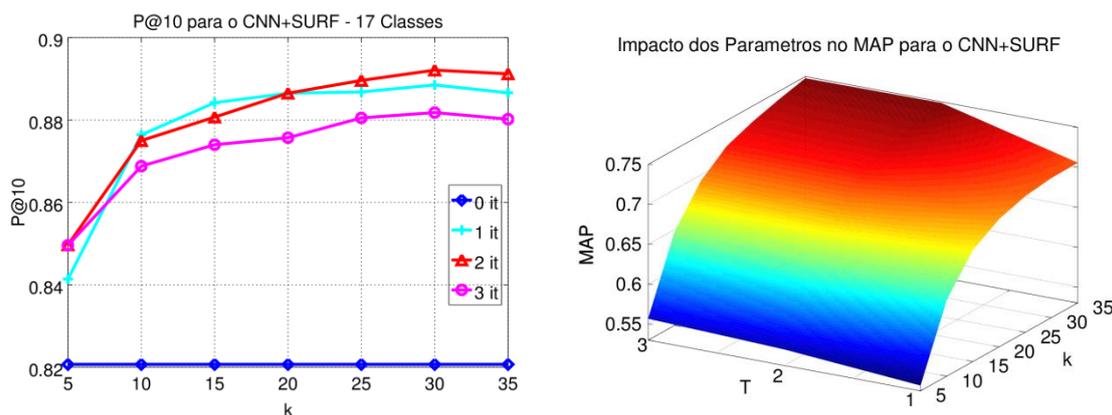
Gráfico 4: P@10 e MAP para o CNN no *dataset 17 Classes*.Gráfico 5: P@10 e MAP para o ACC+SURF no *dataset 17 Classes*.Gráfico 6: P@10 e MAP para o CNN+BIC no *dataset 17 Classes*.

Gráfico 7: P@10 e MAP para o CNN+SURF no *dataset 17 Classes*.

Ao comparar o P@10 inicial do ACC sem o RL-Sim (0 iterações – 0.4215) com o melhor P@10 do CNN+SURF ($k = 30, it = 2$; 0.8921) observa-se uma diferença de aproximadamente 111,65% no valor dessa métrica, somente utilizando os conceitos aplicados da extração de características baixo-nível, do RL-Sim e do *Rank Aggregation*. Pode-se analisar também o ganho do ponto ótimo obtido do P@10 do CNN+SURF com seu P@10 inicial (Gráfico 7), de modo que tal valor atinge aproximadamente 8,7%.

Vale ressaltar a importância de otimizar os *rankings* ao máximo, ou a um ponto próximo do máximo, já que, quanto maior a precisão da recuperação, melhor será a eficácia da aplicação do próximo passo do sistema – a abordagem guiada por semântica para identificação e recuperação de imagens. Com um *ranking* preciso, menor o número de perguntas retornadas ao usuário.

Nota-se também que ao utilizar os parâmetros $it = 1$ e, $k = 15$ ou $k = 20$, os ganhos são bons para todos os descritores. Essa análise é importante, pois tal configuração seria relevante em uma aplicação real do método.

É interessante notar que, quando o CNN está combinado com o SURF (CNN+SURF – Gráfico 7) ou simplesmente está separado de qualquer outro método de extração analisado (CNN – Gráfico 4), pode-se observar que, independente do valor de k , toda iteração está acima do valor inicial (sem RL-Sim, ou seja, $it = 0$). Mesmo utilizando os valores iniciais (sem RL-Sim, $it = 0$) de P@5, P@10 e MAP, o CNN tem os maiores valores para tais métricas do

que qualquer outro método de extração quando analisado com outros métodos sem a aplicação do *Rank Aggregation*. Isso se deve ao fato de que o CNN apresenta os melhores valores iniciais para as métricas analisadas dentre os métodos estudados, de modo a facilitar a aplicação do RL-Sim já que muitas das imagens que estão presentes nas posições iniciais de um *ranking* podem estar presentes em um outro, e as intersecções entre os *rankings* aumentem, aumentando assim também a eficácia da recuperação e os valores das métricas analisadas.

Vale ressaltar também o fato de que conforme aumenta-se o número de iterações, o valor da precisão tende a diminuir para o conjunto de dados *Oxford Flowers 17 Classes* com os métodos de extração analisados. De modo que quando analisado o BIC (Gráfico 3), as curvas de P@10 que representam duas e três iterações não ultrapassam o valor inicial da extração de características de baixo-nível. Parte desse mesmo comportamento pode ser observado na união CNN+BIC, em que a curva de 3 iterações não ultrapassa os valores iniciais do P@10.

Tal fato isolado não explica o porquê das curvas de duas e três iterações não apresentarem ganho maior que os valores iniciais de P@10 do BIC (Gráfico 3). Isso ocorre devido ao fato de que, após aplicação da primeira iteração, houve melhora no P@10 para as k posições iniciais dos *rankings*. Porém, tal melhora não foi visualizada, em geral, a partir da segunda iteração, pois, provavelmente, muitas imagens que se encontram nas cinco e nas dez primeiras posições, após mais de duas iterações, não compartilhavam a interseção dos *rankings*, diminuindo assim os valores de P@10. Observa-se que tal comportamento não ocorreu com o MAP, já que tal métrica continuou demonstrando aumento nas demais iterações.

O algoritmo RL-Sim e *Rank Aggregation* demonstram que, mesmo com uma baixa taxa de precisão e MAP, como é o caso de ACC, SURF e BIC quando analisados em separados de qualquer união, é possível melhorar a recuperação das imagens, já que houve melhora em todos os casos analisados, quando se compara com o valor inicial de cada método de extração para cada métrica analisada.

Quando comparado o melhor MAP da união CNN+SURF (0,7485) com o MAP do SURF sem a aplicação do RL-Sim (0,2155), observa-se uma diferença de aproximadamente 247%. Tal diferença se deve ao fato de que o CNN e o SURF se complementaram na análise de extração de características. Além disso, a aplicação do RL-Sim representou um grande avanço para o ganho do MAP do CNN+SURF em 55,8%, quando comparados os valores do melhor MAP (0,7485) da união CNN+SURF, com o MAP sem a aplicação do RL-Sim para a mesma união (0,4804).

Outro ponto interessante a se analisar é que o MAP, em geral, aumenta conforme se aumenta o número de iterações (*it*) e o valor de *k*, para todos os métodos e uniões analisadas do *17 Classes*. Tal comportamento se deve ao fato de que o MAP ao computar a precisão de cada cálculo de *i*-ésima posição mais relevante, avalia o *ranking* em profundidades maiores do que aquelas iniciais. Desse modo, conforme o *k* é aumentado para a aplicação do RL-Sim, observa-se que a precisão nessas posições também vai aumentando e trazendo imagens relevantes para as posições iniciais, algo que simplesmente analisando o P@5 e o P@10 não se tem tal confirmação.

Tais resultados demonstram o quanto o RL-Sim pode aumentar a eficácia da recuperação de imagens, corroborando com a escolha de utilizá-lo como um passo intermediário para otimizar os *rankings*.

6.3.2 Resultados *Oxford Flowers* – 102 Classes

Essa seção apresenta os resultados dos métodos de extração de características visuais e do aprendizado não supervisionado para o *dataset 102 Classes*.

A Tabela 3 sumariza tanto os resultados iniciais (sem o aprendizado não supervisionado) quanto os melhores resultados apresentados para o P@5 após a aplicação do RL-Sim para o *102 Classes*.

Pode-se observar na Tabela 3 que os resultados do P@5 apresentados por esse conjunto de dados são menores do que aqueles apresentados para o *17 Classes* (Tabela 2). Isso ocorre pelo fato de que o *102 Classes* apresenta um número maior de classes quando comparado ao *17 Classes*, o que dificulta a recuperação de resultados corretos pela análise dos descritores, já que aumentam-se as chances de ocorrerem erros ao comparar imagens de classes diferentes e classificá-las como da mesma classe. Porém, a aplicação do RL-Sim e de técnicas de *Rank Aggregation* no *102 Classes* demonstram grande melhora na eficácia da recuperação de imagens.

Apesar de apresentar valores menores para o P@5, o *102 Classes*, em alguns casos, obteve melhor eficácia do que aqueles apresentados pelo conjunto de dados *17 Classes*.

Observa-se na Tabela 3 que o P@5 do CNN aplicado ao *102 Classes* apresenta um valor inicial de 0,5751. Com a união do CNN+BIC+SURF houve uma diferença de 39,45% quando comparado tal valor inicial (0,5751) com o melhor P@5 dessa união (0,8020).

No *17 Classes* (Tabela 2), ao comparar o valor inicial do P@5 do CNN com o melhor P@5 da melhor união (CNN+SURF) para o *dataset* em questão, observa-se uma diferença de 6,64%.

Tabela 3: P@5 dos métodos de extração de características e do aprendizado não-supervisionado para o *Oxford Flowers 102 Classes*.

Métodos de Extração	P@5 sem RL-Sim	Melhor P@5
SURF	0,3661	0,4201
BIC	0,5399	0,5432
CNN	0,5751	0,6009
CNN+BIC+SURF	0,7455	0,8020

A Tabela 3 apresenta os resultados do P@5 da agregação de três métodos de extração (CNN+BIC+SURF), pois tal combinação apresentou os

melhores resultados recuperados para as métricas testadas no conjunto de dados *102 Classes*. Também foram combinados os métodos CNN+SURF, CNN+BIC, BIC+SURF, além de uma combinação do ACC com esses três métodos, porém, os resultados apresentados foram abaixo do que aqueles apresentados pelo CNN+BIC+SURF.

Ao comparar o P@10 inicial do SURF sem o RL-Sim (0 iterações – 0,2623 – Gráfico 8) com o melhor P@10 do CNN+BIC+SURF ($k = 25, it = 2$; 0,7430 – Gráfico 11) observa-se um ganho de aproximadamente 183,26%, somente utilizando os conceitos aplicados da extração de características de baixo-nível, do RL-Sim e do *Rank Aggregation*. Pode-se analisar também o ganho do ponto ótimo obtido do P@10 do CNN+BIC+SURF com seu P@10 inicial (Gráfico 11), de modo que tal valor atinge aproximadamente 14,48%.

É interessante notar a relação entre os P@10 e MAP para o BIC (Gráfico 9). A curva com uma iteração é a única que apresenta melhora com relação ao resultado inicial do P@10. Esse tipo de avaliação impacta também na curva do MAP, de modo que fica claro que o ganho principal dessa métrica ocorre com o aumento do k , para uma iteração.

Gráfico 8: P@10 e MAP para o SURF no *dataset 102 Classes*.

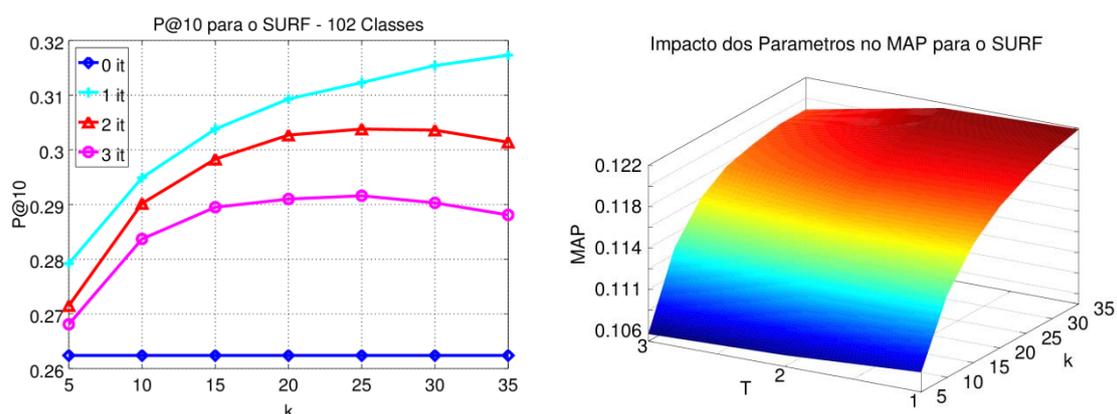
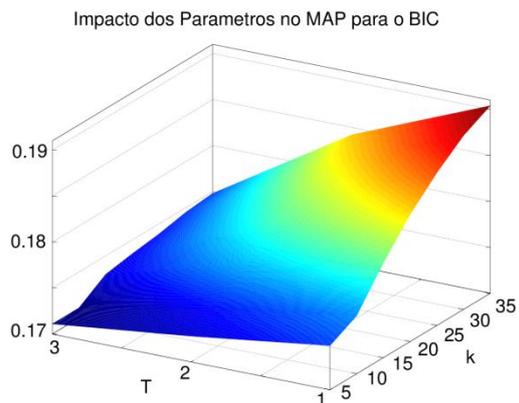
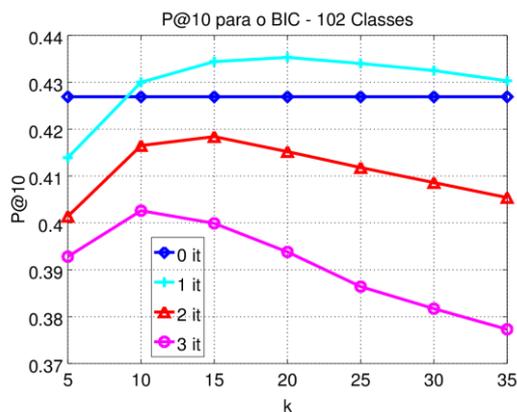
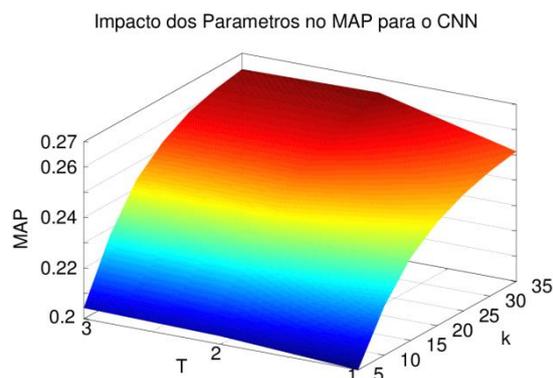
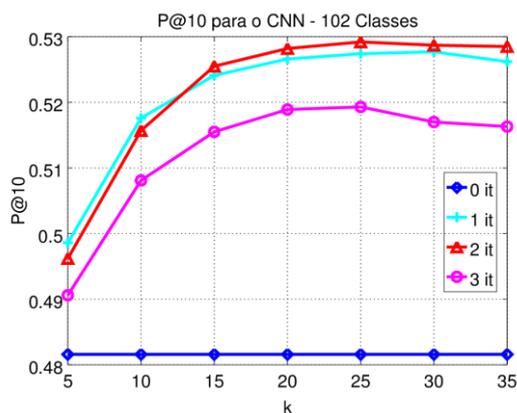
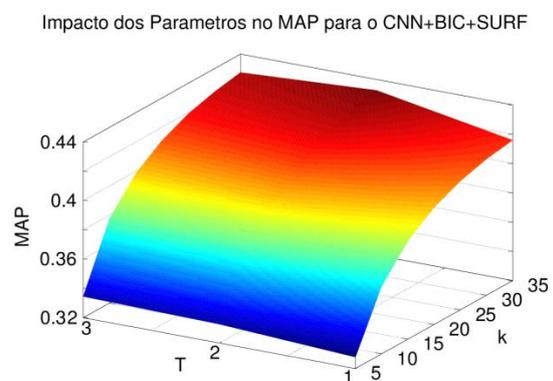
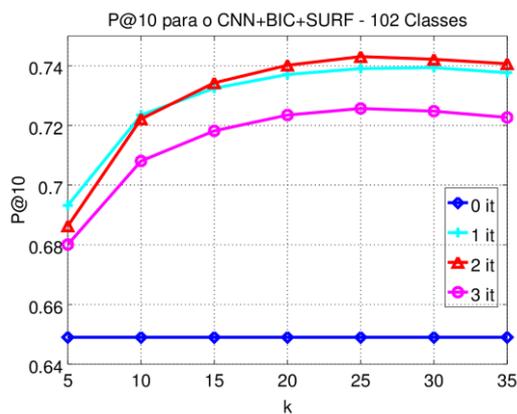


Gráfico 9: P@10 e MAP para o BIC no *dataset 102 Classes*.Gráfico 10: P@10 e MAP para o CNN no *dataset 102 Classes*.Gráfico 11: P@10 e MAP para o CNN+BIC+SURF no *dataset 102 Classes*.

Essa análise é importante já que quanto melhor o resultado das imagens recuperadas nos *rankings*, melhor será a seleção de atributos para a escolha da estrutura morfológica ou da propriedade da planta que melhor diferencia as imagens presentes nas posições iniciais. Com altas precisões dos *ranked lists*, mais efetivo será a recuperação de imagens e menor será o número de interações com o usuário.

Os resultados da agregação desses três métodos apresentam o quanto o aprendizado supervisionado pode auxiliar na melhora dos resultados para a abordagem proposta. O resultado inicial do MAP para a união CNN+BIC+SURF é de 0,2872 (sem RL-Sim), enquanto o resultado obtido após a aplicação do aprendizado não supervisionado é de 0,4326 (Gráfico 11), representando um ganho de 50,63%.

Um ponto interessante dessa união de descritores é que, apesar do SURF apresentar o menor resultado dentre os três métodos de extração escolhidos para o conjunto de dados *102 Classes*, quando unido com o CNN e o BIC, os resultados apresentam uma grande melhora. Isso se deve ao fato de que o CNN, o BIC e o SURF se complementaram nas análises de extração das características visuais. Em geral, isso significa que tais métodos apresentaram acertos em comum e erros diferentes para uma mesma imagem de busca.

6.4. Recuperação Interativa Semântica de Imagens

Os resultados apresentados nesta seção são referentes aos experimentos utilizando a abordagem interativa guiada por semântica para identificação e recuperação de imagens.

Após as análises realizadas na Seção 6.3 – Características Visuais e Aprendizado Não-Supervisionado – que teve como objetivo melhorar a eficácia da recuperação dos *rankings* para que quando estiverem nesse passo da análise, a seleção semântica de atributos seja otimizada de modo a selecionar

as perguntas que melhor diferenciem as famílias presentes nas posições iniciais.

Os resultados apresentados nesta seção demonstram aumento da precisão, do MAP e da acurácia quando aplicada a união das técnicas CBIR e técnicas de modelagem e conceitualização do conhecimento (Figura 21).

Como as configurações dos pontos ótimos atingidos por cada métrica foram diferentes para os métodos de extração estudados (Tabela 2 e Gráficos 1-7 para o *17 Classes*; e Tabela 3 e Gráficos 8-11 para *102 Classes*), foi padronizada a configuração dos parâmetros como $K = 20$ e $it = 2$ para a realização de todos os experimentos com o módulo de integração semântica proposto. Além disso, através de uma análise empírica foi definido o valor da constante $\alpha = 0.01$. As imagens presentes até a posição 700 do *ranking* são aquelas que podem ser analisadas para avaliar se apresentam ou não o atributo em questão.

O “apêndice A” apresenta a análise estatística dos gráficos, referentes às métricas P@5, P@10 e MAP, para ambos os conjuntos de imagens analisados.

6.4.1 Resultados *Oxford Flowers* – 17 Classes

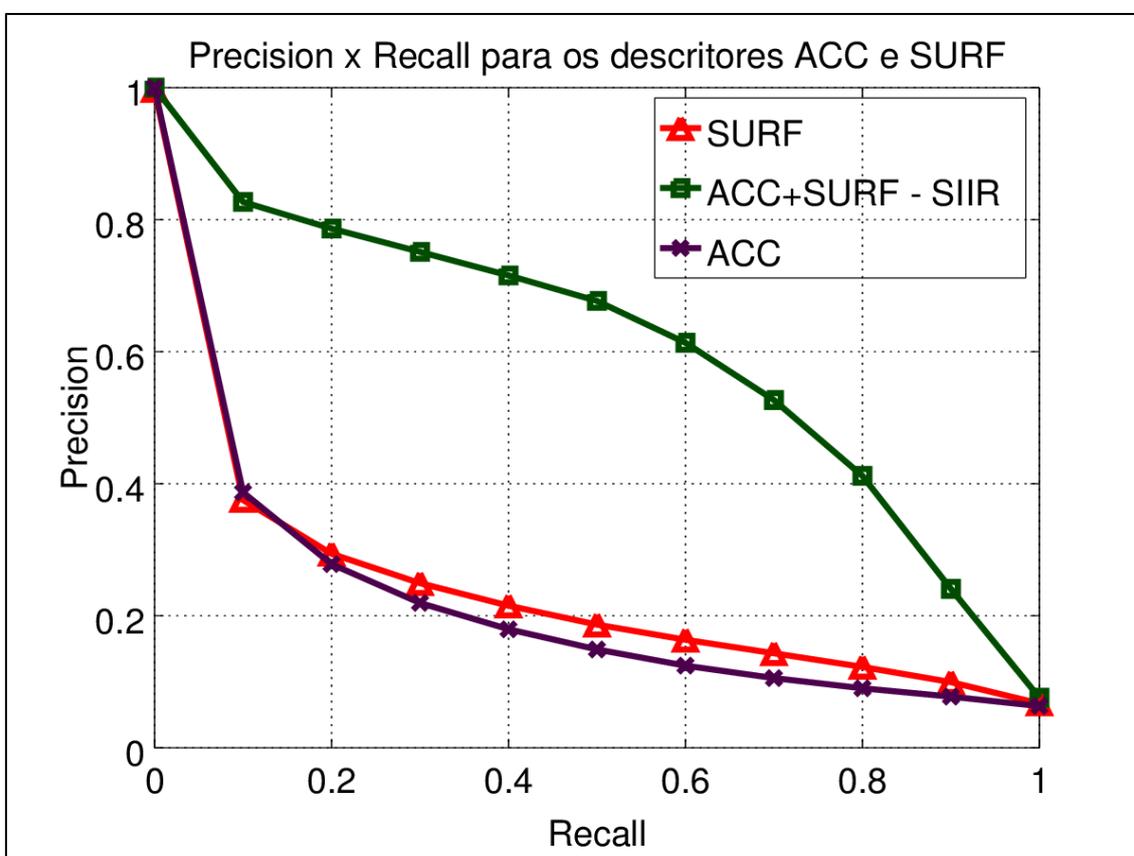
Pode-se observar pelo Gráfico 12 que houve melhora dos resultados recuperados de *Precision* \times *Recall* para os dois descritores (ACC e SURF) avaliados na configuração testada após 10 questões.

Os resultados apresentados no Gráfico 12 são referentes à configuração de $k = 20, it = 2, q = 10$ para a união ACC+SURF, enquanto que para os descritores isolados não foi aplicado o RL-Sim.

Observa-se que o ganho – representado pela distância entre as curvas dos dois descritores separados (ACC e SURF) e a curva da união ACC+SURF (*Rank Aggregation*), atinge um valor muito alto, demonstrando a eficácia da

abordagem proposta. É interessante visualizar também a diferença da curvatura entre as curvas dos descritores ACC e SURF (analisados em separado) com a curva da união ACC+SURF. A diferença na curvatura já demonstra uma melhor eficácia da precisão dos resultados recuperados com a abordagem proposta conforme se avança no *ranking* encontrando a *n*-ésima imagem relevante.

Gráfico 12: *Precision* × *Recall* para ACC e SURF com a abordagem proposta (SIIR) – 17 Classes.

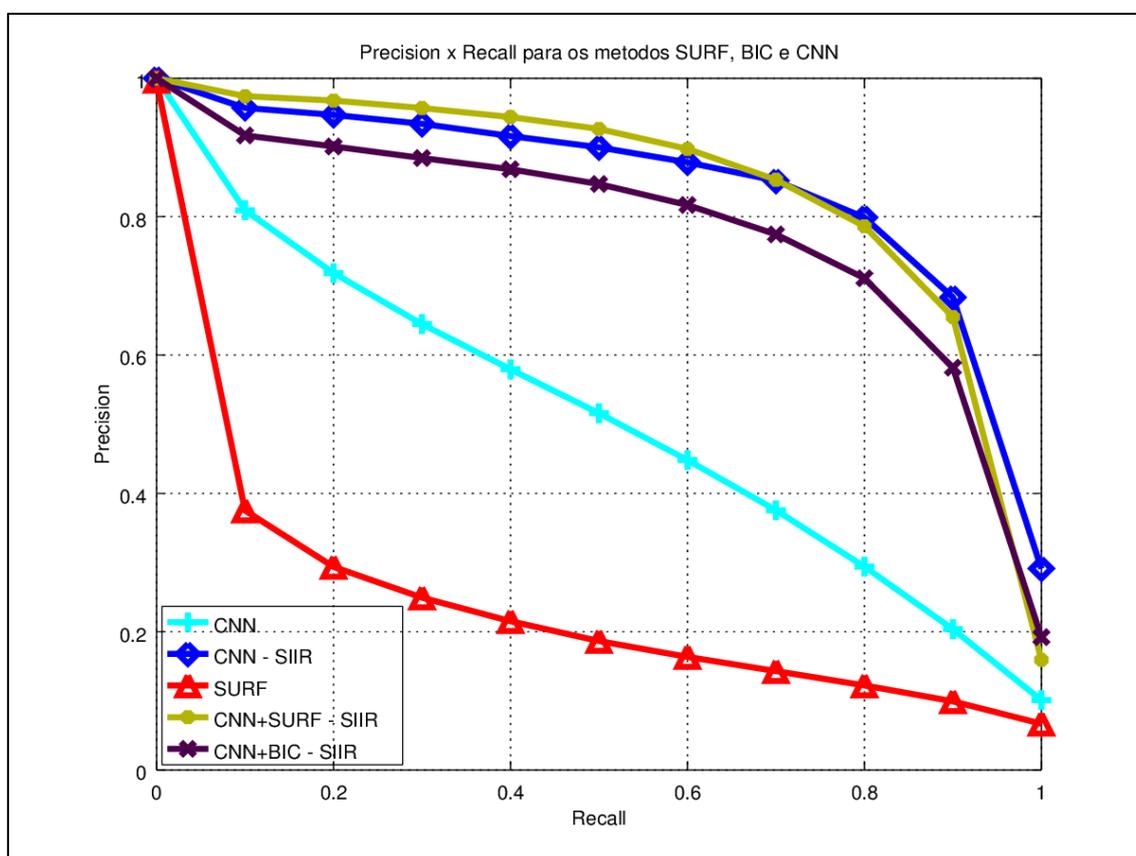


Os resultados apresentados no Gráfico 12 também demonstram a eficácia da abordagem proposta, já que mesmo para métodos que apresentam um baixo valor para as métricas avaliadas (como é o caso do ACC e do SURF – os dois descritores com menores valores para P@5, P@10 e MAP, dentre os descritores escolhidos para o estudo) há um grande aumento na precisão.

Esses resultados alcançam valores maiores quando comparadas as curvas de *Precision* \times *Recall* do *CaffeNet* (CNN) com o SURF (Gráfico 13).

A curva CNN+SURF do Gráfico 13 demonstra os melhores resultados apresentados para essa análise. Pode-se observar que a distância entre essa curva e a curva do SURF atinge um valor maior do que os demais observado no Gráfico 13. Esses resultados novamente corroboram com os objetivos apresentados pela abordagem proposta.

Gráfico 13: *Precision* \times *Recall* para SURF, BIC e CNN com a abordagem proposta (SIIR) – 17 Classes.

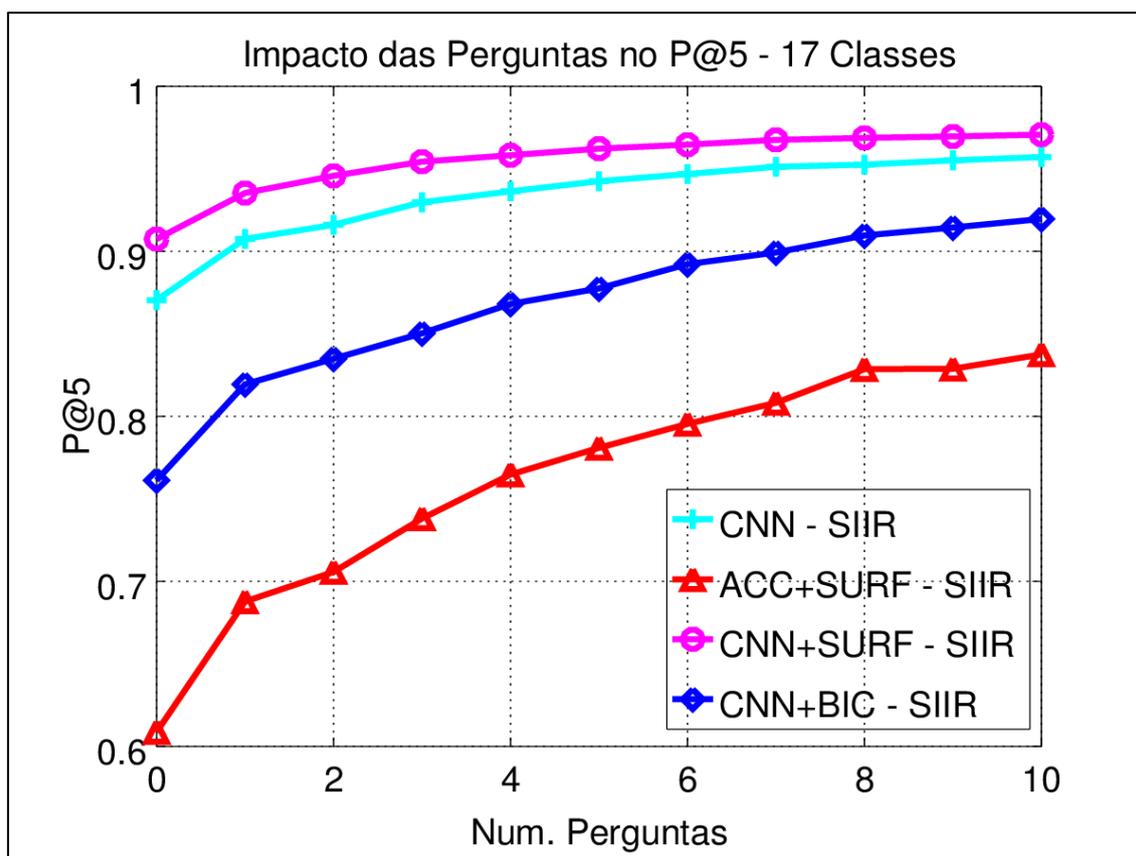


Quando se analisa as outras métricas estudadas, observa-se também que houve ganho. O Gráfico 14 apresenta os resultados da precisão nas cinco primeiras posições dos *rankings* para os quatro métodos de extração

analisados. O Gráfico 14.A (Apêndice A), além do P@5, também demonstra os intervalos de confiança para cada número de perguntas respondidas.

Observa-se que o valor do P@5 da união CNN+SURF é o maior atingido nesse experimento. O valor dessa métrica aumenta a cada pergunta respondida pelo usuário, atingindo seu valor máximo com 10 questões respondidas (0,9707). Esse valor representa um ganho de aproximadamente 6,97% com relação ao valor inicial do P@5 (0,9074) apresentado no Gráfico 14.

Gráfico 14: P@5 para os métodos de extração com a Recuperação Semântica Interativa de Imagens (SIIR) – 17 Classes.



Caso seja comparado o valor máximo do P@5 para o CNN+SURF (0,9707 – Gráfico 14) após aplicação do módulo de integração semântica, com

o P@5 da extração de característica do mesmo método sem aplicação do RL-Sim (0,8859 – Tabela 2), observa-se um ganho de aproximadamente 9,57%.

Se comparado o valor máximo do P@5 dessa junção (0,9707) com o valor inicial apresentado pelo ACC (0,5310; sem o RL-Sim – Tabela 2), o ganho atinge aproximadamente 79%.

É importante também realizar uma análise detalhada dos resultados da precisão das famílias nas cinco primeiras posições (PF@5). Nesse sentido, observa-se ganho nos resultados dessa métrica quando aplicado o SIIR. O PF@5 para o ACC+SURF, sem o RL-Sim, atingiu o valor de 0,6683; enquanto que quando aplicado o RL-Sim para a configuração de $k = 20, it = 2, q = 0$ observa-se uma leve queda do PF@5 com o valor de 0,6639. O interessante é que quando aplicado o SIIR a precisão das famílias aumenta. Com 1 questão o PF@5 atinge 0,7575 de precisão; com 5 questões, o valor para essa métrica representa aproximadamente 0,8784 de precisão, enquanto que para 10 questões esse valor aumenta para 0,9519.

O ganho máximo obtido para o PF@5 com o SIIR aplicado ao ACC+SURF atinge o valor de aproximadamente 43,38%, já que o PF@5 para $k = 20, it = 2, q = 0$ é igual a 0,6639 e o valor máximo do PF@5 para $k = 20, it = 2, q = 10$ é igual a 0,9519.

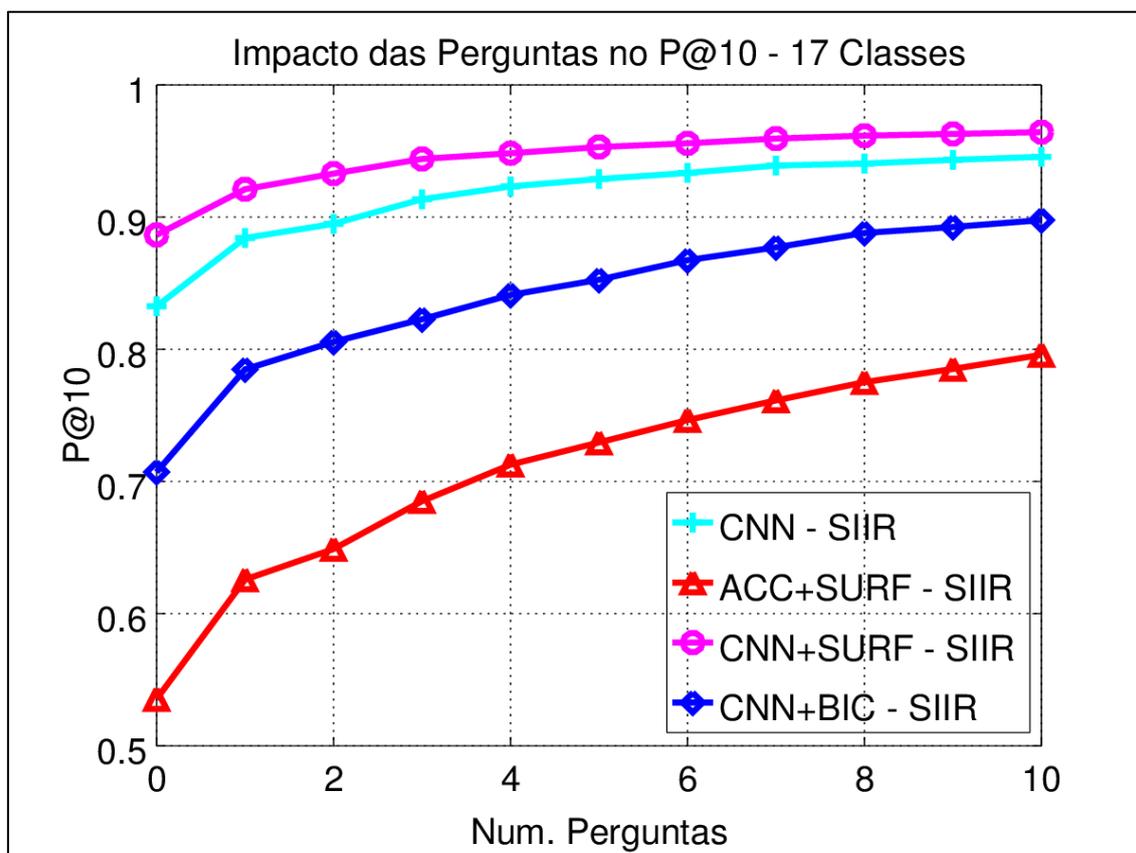
Quando comparado o PF@5 do CNN+SURF observa-se que o valor máximo atingido (para $k = 20, it = 2, q = 10$) foi de 0,9917 de precisão na classificação das famílias presentes nas cinco primeiras posições. Um ótimo valor próximo da corretude máxima. Com $k = 20, it = 2, q = 5$, o PF@5 do CNN+SURF apresenta valor de 0,9816; enquanto que para a mesma configuração, porém com 1 questão, o PF@5 é igual a 0,9356. Já quando não houve a aplicação do RL-Sim o PF@5 para essa união é de 0,9051.

É interessante notar que o ganho das precisões que apresentam menor valor acaba sendo maior do que aqueles apresentados pelos métodos com altos valores de precisão. Isso ocorre pelo fato de que, quando o *ranking* apresenta baixa taxa de recuperação correta de imagem é possível que grande parte das imagens, que pertençam àquela classe ou àquela família da imagem

de busca, não estejam tão próximas das posições iniciais, mas também não tão profunda no *ranking*, permitindo assim a mobilidade dessas imagens para posições iniciais após a resposta do usuário. Quando se tem altas precisões, a maioria das imagens pertencentes à classe ou à família da imagem de busca já estão nas posições iniciais, de modo que eventualmente, alguma imagem que esteja nas posições mais profundas do *ranking* não sejam afetadas pela aplicação do SIIR.

A análise da precisão nas dez primeiras posições também demonstra um aumento para os quatro métodos analisados (Gráfico 15), apresentando ganhos similares ao P@5. O Gráfico 15.A (Apêndice A), apresenta, além do P@10, os intervalos de confiança para cada número de perguntas respondidas.

Gráfico 15: P@10 para os métodos de extração com a Recuperação Semântica Interativa de Imagens (SIIR) – 17 Classes.



De maneira análoga, a precisão das famílias nas dez primeiras posições (PF@10) também segue um padrão daquele demonstrado para o PF@5, de modo que para o ACC+SURF, o ganho máximo do PF@10 foi de 62,20%, já que o PF@10 analisado para $k = 20, it = 2, q = 10$ foi de 0,9412, enquanto que o valor dessa métrica para essa união sem o RL-Sim e sem a aplicação do SIIR foi de 0,5803.

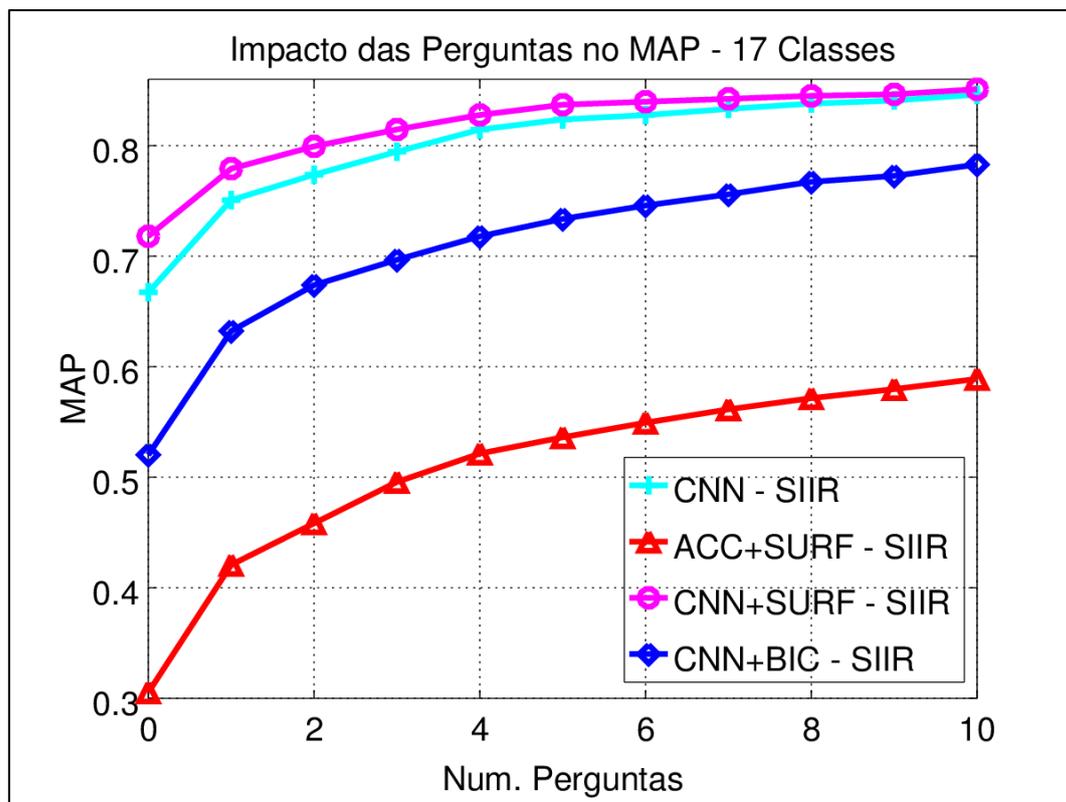
Para a união CNN+SURF, o valor de PF@10 atinge um ótimo resultado com 0,9911 para $k = 20, it = 2, q = 10$, representando um ganho de aproximadamente 16% para essa métrica, quando compara-se o PF@10 ótimo dessa união (0,9911) com o PF@10 sem a aplicação do método proposto (0,8647).

Esse aumento se deve ao fato de que, a cada resposta dada pelo usuário, a matriz de distâncias é atualizada, trazendo para as posições iniciais aquelas imagens das famílias que apresentam o atributo analisado pelo Seleção Semântica de Atributos (Seção 5.7).

Vale citar o fato de que o usuário, ao responder apenas uma pergunta, aumenta a eficácia do sistema de maneira positiva. Por exemplo, no caso do MAP (Gráfico 16), os quatro métodos de extração apresentam o maior aumento na primeira resposta dada pelo usuário. Esse resultado apresenta um grande avanço para a identificação de famílias de Angiospermas, já que uma chave dicotômica para Flores Dicotilédneas Dialipétalas atinge 188 entradas (SOUZA e LORENZI, 2007).

Os valores atingidos pelos pontos ótimos do MAP, quando comparados com os valores iniciais apresentados no Gráfico 16, demonstram ganhos de 18,52% para CNN+SURF, 26,81% para *Caffe* (CNN), 50,49% para CNN+BIC e 92,89% no caso do ACC+SURF.

Gráfico 16: MAP para os métodos de extração com a Recuperação Semântica Interativa de Imagens (SIIR) – 17 Classes.



O ganho atinge seu ápice quando comparados os valores do Gráfico 16 com os valores gerados pela extração de características (Tabela 2; sem RL-Sim). No caso do MAP para o CNN+SURF, é apresentado um ganho de 77,14%; para o CNN+BIC o ganho representa 87,97% (MAP inicial do CNN+BIC sem RL-Sim – 0,4165 – Tabela 2). Já para o CNN, o ganho é de 68,38%; enquanto que o ACC+SURF apresenta o maior valor com 138,81% de ganho.

Quando comparado o valor inicial do MAP para ACC+SURF (0,3053) apresentado pelo Gráfico 16 com o valor final (0,5889 – $k = 20, it = 2, q = 10$), observa-se um ganho de 92,89%. Esse ganho é referente unicamente ao processo de recuperação semântica interativa de imagens, já que o valor de 0,3053 representa o MAP para $k = 20, it = 2, q = 0$.

É interessante analisar que os ganhos atingidos pelo ACC+SURF apresentam os maiores valores. Isso se deve ao fato de que essa junção de descritores demonstrou precisões e MAP relativamente baixos no início do processo de recuperação das imagens, quando comparados com os outros experimentos realizados nesse estudo. Tal resultado demonstra que a abordagem proposta é efetiva em vários casos de análise de classificação e recuperação de imagens, tanto para métricas com baixo resultado, quanto para os melhores casos. O MAP calculado refere-se ao MAP das classes das imagens dos conjuntos de dados. Não foi calculado o MAP das famílias.

O Gráfico 16.A (Apêndice A) demonstra os intervalos de confiança para o MAP – *17 Classes*, para cada número de perguntas respondidas.

Finalmente, a Tabela 4 apresenta a acurácia da classificação 20-NN das famílias de Angiospermas. Utilizaram-se as 20 primeiras imagens dos *rankings* para avaliar a acurácia. Pode-se observar que após 10 questões, a taxa de reconhecimento das famílias atinge 98,97% de acurácia para as 20 primeiras posições recuperadas.

Tabela 4: Acurácia da classificação das 20-NN famílias – *Oxford 17 Classes*

Métodos	Acurácia Famílias
CNN+SURF 1 pergunta	93,97%
CNN+SURF 5 perguntas	97,20%
CNN+SURF 10 perguntas	98,97%

Como os resultados sumarizados na Tabela 4 não podem ser comparados com outros estudos que utilizam o conjunto de dados *17 Classes* devido à diferença de classificação da abordagem proposta, já que a maioria dos trabalhos relacionados utilizam esse tipo de métrica para realizar a avaliação de suas respectivas propostas, compara-se então, a acurácia da classificação das classes (Tabela 5) do *dataset* (ao invés das famílias) com outras abordagens do estado-da-arte.

Apesar do fato do protocolo experimental do método proposto no presente trabalho (busca “por exemplo”, comparando todas as imagens com todas) ser diferente dos demais, uma comparação geral foi realizada a fim de demonstrar a efetividade da abordagem proposta. Tabela 5 sumariza as acurácias de reconhecimento publicadas na literatura juntamente com as acurácias geradas pelos experimentos desenvolvidos nesse projeto.

Tabela 5: Acurácia da classificação 20-NN da abordagem proposta em comparação com métodos do estado-da-arte para o *dataset Oxford 17 Classes*.

Métodos	Acurácia Classes
Visual Vocabulary (NILSBACK e ZISSERMAN, 2006)	71,76%
Discrim. Power-Invar. Trade-Off (VARMA e RAY, 2007)	82,55%
Auto. Flower Classif. (NILSBACK e ZISSERMAN, 2008)	88,33%
Top-down color attention (KHAN <i>et al.</i> , 2009)	89,00%
Bin-ratio information (XIE <i>et al.</i> , 2010)	89,02%
BiCoS (CHAI <i>et al.</i> , 2011)	90,04%
<i>RL-Sim - CNN+SURF</i>	90,44%
Multi-scale fusion (HU <i>et al.</i> , 2014)	91,39%
<i>SIIR - CNN+SURF 1 pergunta</i>	92,06%
<i>SIIR - CNN+SURF 5 perguntas</i>	95,22%
<i>SIIR - CNN+SURF 10 perguntas</i>	96,84%

Observa-se que a recuperação semântica interativa de imagens obtêm resultados de maior acurácia se comparado com as demais abordagens e pode atingir 96,84% de acurácia para a classificação 20-NN.

Desse modo, os experimentos demonstram o quanto a abordagem proposta (Figura 1 e Figura 21) pode aumentar os resultados de classificação e recuperação de um conjunto de imagens, nesse caso, mais precisamente para o conjunto de dados *Oxford Flowers 17 Classes*.

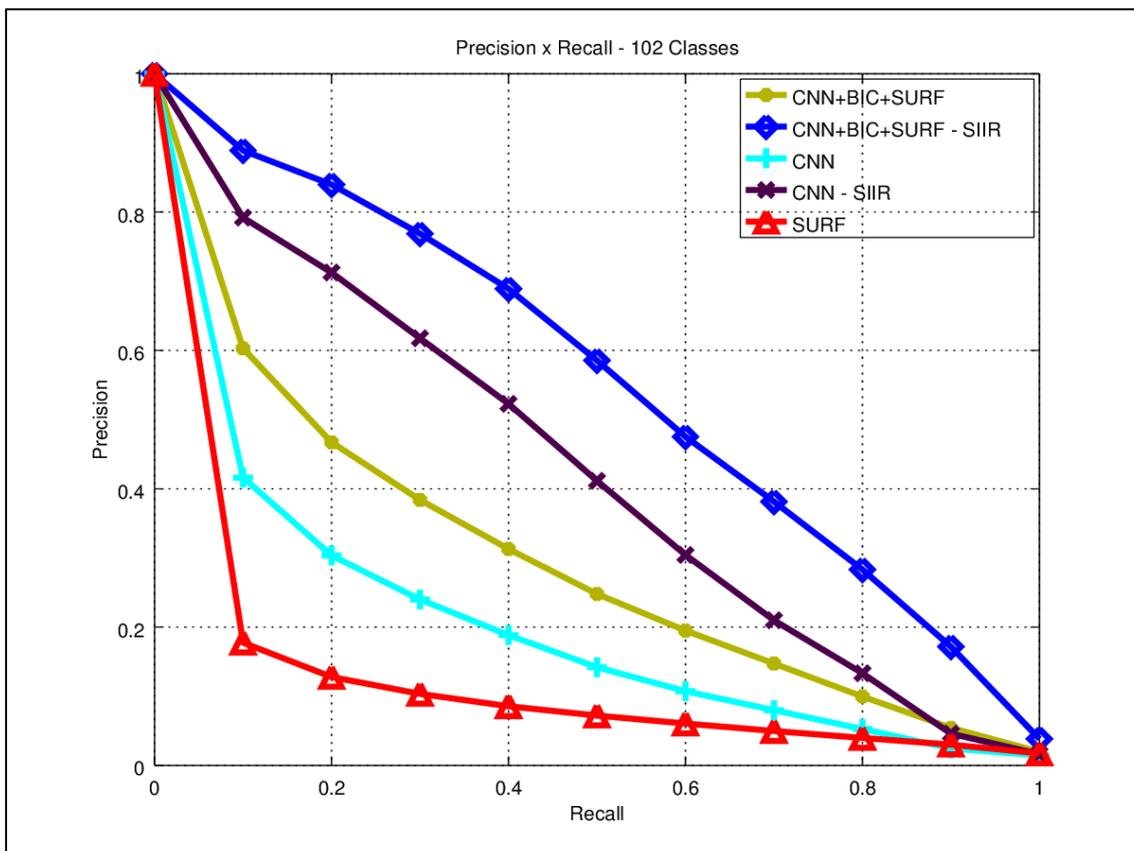
Resultados referentes à abordagem proposta aplicada ao *dataset 102 Classes* serão discutidos na próxima seção e também demonstram aumento em todos os resultados de recuperação.

6.4.2 Resultados *Oxford Flowers* – 102 Classes

Essa seção introduz os resultados da Recuperação Semântica Interativa de Imagens (*Semantic Interactive Image Retrieval* – SIIR) para o conjunto de dados *102 Classes*. A configuração utilizada para a realização dos testes foi a mesma daquela apresentada para o *17 Classes* ($k = 20, it = 2$).

Para tanto, o Gráfico 17 demonstra a melhora da abordagem proposta para o *dataset* em questão após 10 questões. Pode-se observar novamente ganho na efetividade do SIIR ao analisar a distância entre as curvas referentes aos métodos de extração – com e sem a aplicação da abordagem. Novamente é interessante observar a diferença da curvatura das curvas apresentadas no Gráfico 17. As duas curvas com melhor recuperação de imagens referem-se à aplicação do método proposto. A curva com melhor recuperação de resultados é aquela representando a união CNN+BIC+SURF.

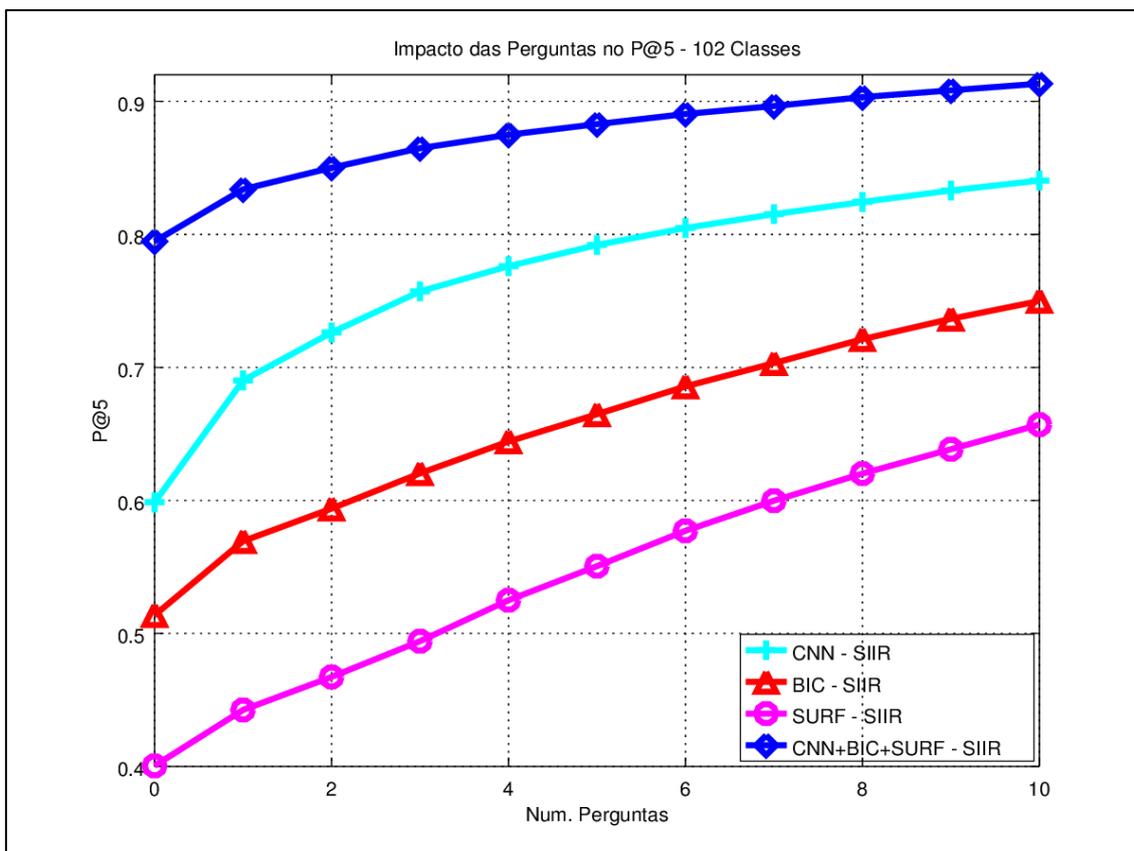
Gráfico 17: *Precision* × *Recall* para SURF, BIC e CNN com a abordagem proposta (SIIR) – 102 Classes.



O Gráfico 18 apresenta a precisão para as cinco primeiras posições dos *ranked lists* para o conjunto de dados 102 Classes. O Gráfico 18.A (Apêndice A), além do P@5, também demonstra os intervalos de confiança para cada número de perguntas respondidas para o conjunto de imagens 102 Classes.

Pode-se observar no Gráfico 18 que o CNN+BIC+SURF apresenta o melhor P@5 quando comparado aos outros métodos estudados. O ganho dessa união quando comparado ao valor inicial da curva presente no Gráfico 18 representa 14,92%. O melhor ganho dentre os apresentado no Gráfico 18 é o referente ao SURF (64,04%).

Gráfico 18: P@5 para os métodos de extração com a Recuperação Semântica Interativa de Imagens (SIIR) – 102 Classes.



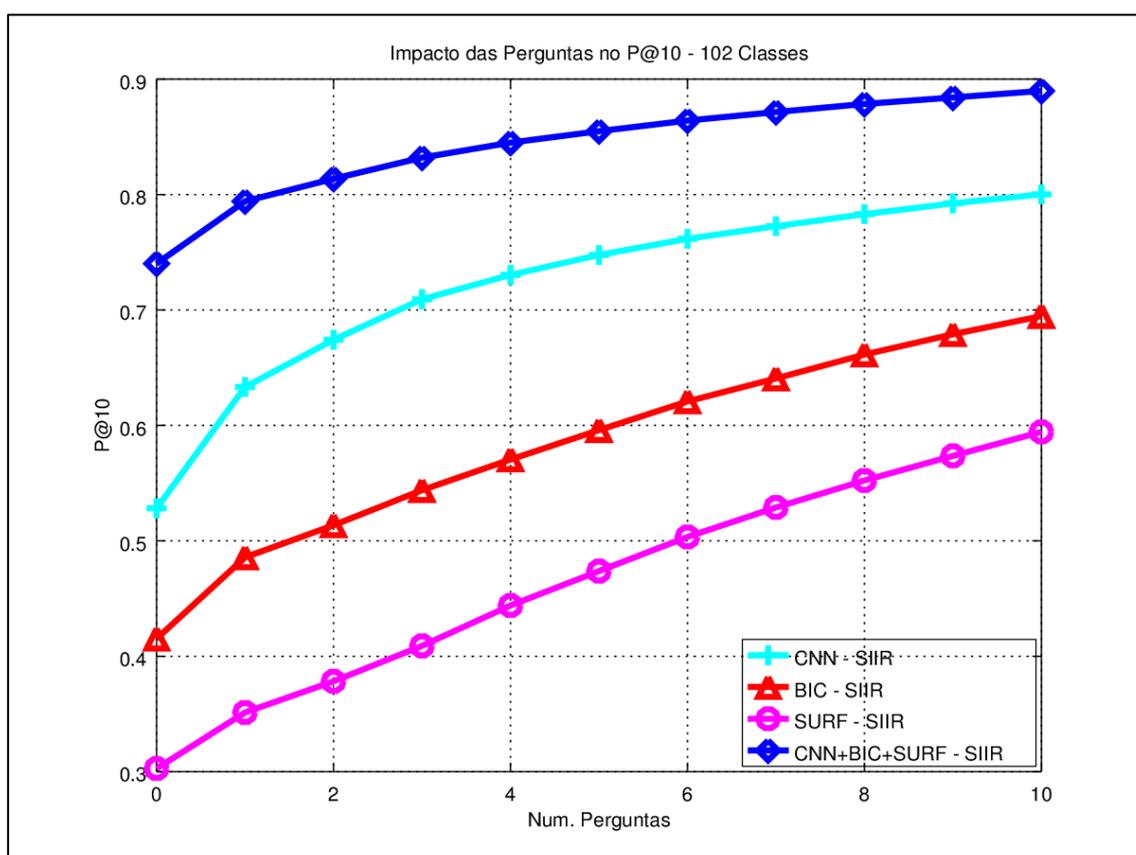
Também houve grande aumento na precisão das famílias nas cinco primeiras posições, de modo que para o SURF – método com os menores valores para as métricas testadas nesse conjunto de dados – o valor de PF@5 sem a aplicação do RL-Sim foi de 0,4052. Com a aplicação do RL-Sim para $k = 20, it = 2, q = 0$, o PF@5 foi de 0,4436. Já quando aplica-se as questões, para $k = 20, it = 2, q = 1$, o PF@5 para o SURF no 102 Classes atinge 0,5005. Para $k = 20, it = 2, q = 5$, o PF@5 apresenta valor de 0,6444; enquanto que para a mesma configuração com 10 questões, o PF@5 é igual a 0,7885.

Desse modo, o PF@5 para o SURF apresentou ganho máximo de 94,59% quando comparado o PF@5 de 0,7885 com o inicial de 0,4052. Aumentando consideravelmente a classificação das famílias das imagens de busca.

Para a junção CNN+BIC+SURF o PF@5 também apresentou altos valores, de modo que o valor máximo atingido por essa métrica para essa junção foi de 0,9320 para $k = 20, it = 2, q = 10$ e, tendo como ganho máximo o valor de aproximadamente 23,56%, quando comparado com o valor inicial do PF@5 – somente com a extração das características de baixo-nível dessa junção (0,7543).

O Gráfico 19 também apresenta altos valores para o P@10, após a aplicação das questões selecionadas pelo “Seleção Semântica de Atributos” (Seção 5.10).

Gráfico 19: P@10 para os métodos de extração com a Recuperação Semântica Interativa de Imagens (SIIR) – 102 Classes.



A precisão nas dez primeiras posições para o SURF apresenta novamente o maior ganho, com 96,36%. Quando analisado o BIC, observa-se

que o ganho é de 67,31%, demonstrando novamente que a abordagem proposta é efetiva para resultados com baixa taxa de recuperação.

O Gráfico 19.A (Apêndice A), demonstra os intervalos de confiança para cada número de perguntas respondidas com relação à precisão nas dez primeiras posições para o *102 Classes*.

A precisão das famílias também apresenta melhora nos resultados para as dez primeiras posições. Sendo que para o SURF o valor do PF@10 somente para a extração de características atingiu 0,3087; quando aplicado o RL-Sim com $k = 20, it = 2, q = 0$ o PF@10 apresenta valor de 0,3529. Quando utiliza-se a abordagem proposta para a configuração de $k = 20, it = 2, q = 1$, o PF@10 é igual a 0,4199. Com $k = 20, it = 2, q = 5$, o PF@10 é igual a 0,5882 e, finalmente, para 10 questões, o PF@10 do SURF atinge o maior valor testado, com 0,7541.

Nesse sentido, fica claro novamente o quão eficaz é a abordagem proposta na classificação de famílias de Angiospermas, já que se obteve um ganho de aproximadamente 113,68% para o PF@10 no SURF, somente com a aplicação das questões (comparando o valor de 0,3529 - $k = 20, it = 2, q = 0$, com 0,7541 - $k = 20, it = 2, q = 10$).

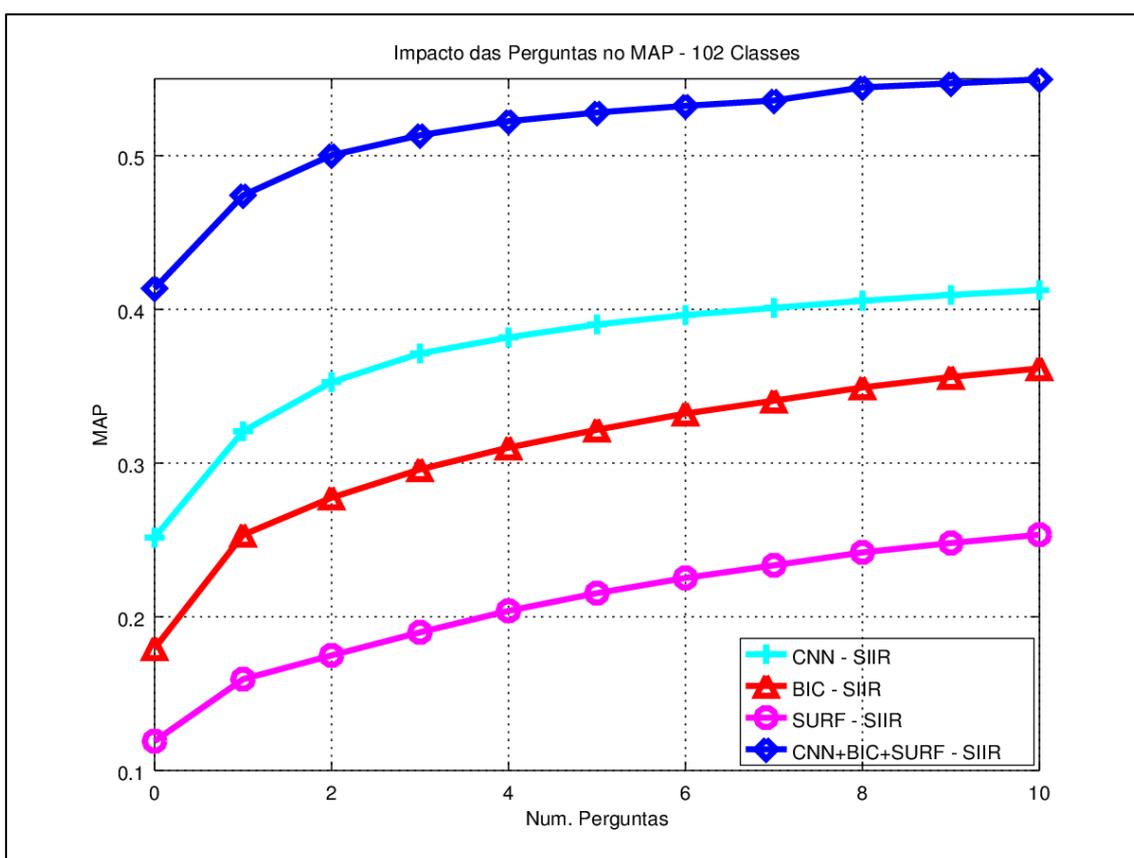
O valor de PF@10 para a união CNN+BIC+SURF também obteve bons resultados para essa junção, com o valor máximo atingido pela configuração de $k = 20, it = 2, q = 10$, com 0,9159; e valor inicial de 0,6645 (sem RL-Sim). Comparando esses dois valores observa-se um ganho de 37,83% com a aplicação da abordagem proposta.

O Gráfico 20 demonstra um aumento nos resultados recuperados não só nas posições iniciais do *ranking*, mas também naquelas posições mais profundas. Essa análise pode ser feita, já que, quando o usuário responde a uma determinada pergunta correspondente a um atributo específico, até mesmo as imagens de plantas que contenham aquele atributo e estão em posições mais afastadas das iniciais podem ser movidas para o topo do *ranking*. O Gráfico 20 demonstra que o SURF apresentou o melhor ganho para

essa métrica com 112,76%, após a aplicação do SIIR, seguido pelo BIC com 101,45%, enquanto a união CNN+BIC+SURF apresentou ganho de 32,88%.

O Gráfico 20.A (Apêndice A) demonstra os intervalos de confiança para o MAP – 102 Classes, para cada número de perguntas respondidas.

Gráfico 20: MAP para os métodos de extração com a Recuperação Semântica Interativa de Imagens (SIIR) – 102 Classes.



Finalmente, assim como demonstrado na seção anterior, a Tabela 6 apresenta a acurácia da 20-NN imagens recuperadas para as famílias de Angiospermas. Pode-se observar que após 10 questões, a taxa de reconhecimento das famílias atinge 85,39%.

Tabela 6: Acurácia da classificação das 20-NN famílias – *Oxford 102 Classes*

Métodos	Acurácia Famílias
CNN+BIC+SURF 1 pergunta	74,94%
CNN+BIC+SURF 5 perguntas	81,13%
CNN+BIC+SURF 10 perguntas	85,39%

Nesse sentido, era esperado que a taxa de acurácia das famílias fosse maior do que a acurácia de classes, porém, para o conjunto de dados *102 Classes*, a taxa de acurácia das classes apresentou maiores valores. Tais resultados podem ser observados na Tabela 7. A Tabela 7 ainda apresenta as comparações da aplicação do SIIR com outros métodos presentes na literatura.

Tabela 7: Acurácia da classificação 20-NN da abordagem proposta em comparação com métodos do estado-da-arte para o *dataset Oxford 102 Classes*.

Métodos	Acurácia Classes
Ito e Kubota (2010)	53,9%
Nilsback e Zisserman (2008)	72,8%
Khan <i>et al.</i> (2011)	73,3%
Kanan e Cottrell (2010)	75,2%
Nilsback (2009)	76,3%
<i>RL-Sim - CNN+BIC+SURF</i>	76,35%
Angelova <i>et al.</i> (2012)	76,7%
<i>SIIR - CNN+BIC+SURF 1 pergunta</i>	79,15%
Chai <i>et al.</i> (2011)	80,0%
Angelova e Shenghuo (2013)	80,6%
Mattos <i>et al.</i> (2014)	80,8%
<i>SIIR - CNN+BIC+SURF 5 perguntas</i>	84,89%
<i>SIIR - CNN+BIC+SURF 10 perguntas</i>	88,88%

É importante mencionar que a eficácia não pode ser comparada precisamente com outros métodos apresentados na Tabela 7 já que tal informação não é reportada em alguns dos trabalhos citados na tabela em questão, além do fato de que o protocolo experimental do presente trabalho difere dos demais. Entretanto, pode-se observar que a acurácia das *top* – 20 imagens com a aplicação do SIIR atingiram os mais altos valores de acurácia quando comparado com os outros métodos. Quando comparado o valor da acurácia das classes do SIIR para a classificação 1-NN no conjunto de dados *102 Classes* utilizando o CNN+BIC+SURF após 10 questões atinge o valor de 90,36%. A acurácia das classes para o 5-NN da mesma configuração testada é de 90,31%, enquanto a acurácia da classificação 10-NN é igual a 89,79%.

6.5. Considerações da Abordagem

Com a análise dos resultados e avaliando a eficácia da abordagem proposta, observa-se um grande ganho para todos os métodos estudados. Isso comprova os objetivos da “Recuperação Semântica Interativa de Imagens” demonstrados no início do presente trabalho.

Com o avanço das perguntas relacionadas às imagens, às famílias de Angiospermas, o conhecimento codificado na ontologia e a interação com o usuário, observa-se o quão robusto é a abordagem proposta, tanto para resultados com baixa taxa de recuperação, quanto para aqueles que apresentam um melhor valor para as métricas estudadas.

É importante realizar uma análise minuciosa do porquê os valores das métricas não atingiram uma taxa de classificação de 100%. Uma das razões são algumas classes de imagens apresentarem baixa taxa de recuperação, diminuindo os valores das métricas. Outro diagnóstico ocorre devido ao fato de que algumas perguntas são respondidas com “*Não*”, ou seja, aquela família da imagem não apresenta tal característica. Como a resposta “*Não*” teve um tratamento diferente da resposta “*Sim*”, os valores das métricas estudadas não obtiveram 100% de classificação, já que a resposta “*Não*” pode ser

inconclusiva, pois uma mesma família pode apresentar mais de uma mesma característica para uma mesma propriedade ou estrutura morfológica (conforme discutido na Seção 5.11).

Caso a análise fosse realizada para espécies ao invés de família, pode-se afirmar que a taxa de recuperação e classificação seriam maiores ainda, já que uma espécie, na classificação biológica, é um dos grupos mais específicos, enquanto família se apresenta mais generalista. Desse modo, as respostas “Sim” e “Não” poderiam apresentar o mesmo peso aumentando assim ainda mais a eficácia do sistema.

Um dos problemas nesse tipo de análise de espécies ao invés de famílias se depara no fato de que a literatura não apresenta grandes publicações das espécies presentes nas imagens dos conjunto de dados analisados, já que são espécies ornamentais conhecidas mundialmente. Além disso, muitas publicações dessas espécies estão em latim, o que dificulta a tradução e a codificação do conhecimento das estruturas morfológicas das espécies em questão, para a ontologia. Para realizar tal tarefa, o ideal seria obter as espécies em mãos, avaliar cada estrutura e dar início ao processo de descrição, porém, há a necessidade de despende grande tempo de trabalho para realizar tal classificação e só depois iniciar a estabilização desses conceitos e assim defini-los na ontologia.

7. Conclusão e Trabalhos Futuros

O presente trabalho apresentou uma nova abordagem para recuperação interativa de imagens. A abordagem proposta reduz o *gap* semântico entre as características de baixo-nível das imagens e os conceitos semânticos ao introduzir a Recuperação Semântica Interativa de Imagens (*Semantic Interactive Image Retrieval* – SIIR).

A ideia principal consiste em recuperar características visuais de um conjunto de imagens e relacionar tais imagens com os conceitos definidos por uma ontologia – também desenvolvida no trabalho em questão.

O método proposto também apresenta uma ferramenta com grande potencial de aplicação na área do ensino de botânica. Como os conceitos definidos na ontologia apresentam asserções claras, tanto para a máquina, quanto para os humanos, cada restrição pode ser consultada a fim de esclarecer qualquer dúvida do usuário. Além disso, os relacionamentos entre cada entidade da ontologia deixam claro qual estrutura ou propriedade pertencem a um determinado órgão, facilitando assim, o ensino de botânica através da ontologia desenvolvida.

Vale ressaltar a versatilidade que a abordagem proposta apresenta ao utilizar ontologia, já que o sistema também pode ser implementado de modo que o usuário insira as características que já conhece para identificar uma planta e, através de inferências realizadas por *reasoners* pode-se chegar à classificação de uma determinada família através de entradas com os atributos conhecidos.

Um conjunto de experimentos foi conduzido a fim de provar a eficácia da abordagem proposta e os resultados demonstraram que altas precisões podem ser obtidas em vários casos do estudo. Além disso, demonstrou a alta eficácia do SIIR tanto para resultados com baixo valor de recuperação, quanto para aqueles melhores classificados.

Trabalhos futuros visam aprimorar as métricas do SIIR a fim de melhorar a seleção de atributos, baseado na resposta do usuário. Por exemplo, caso o usuário responda de maneira positiva uma pergunta com relação a um tipo de inflorescência, o sistema não deve gerar uma nova pergunta sobre outro tipo de inflorescência.

Estender o *dataset* de modo a inserir novas espécies de plantas, ou gerar um novo conjunto de imagens com espécies brasileiras, também é um foco de trabalhos futuros. Nesse sentido, é provável que os resultados recuperados apresentem valores menores do que aqueles apresentados nesse trabalho. Isso pode ocorrer já que, quanto maior o número de imagens, maiores as chances de erros de classificação (ainda mais no caso de imagens de plantas que muitas vezes são semelhantes entre si, apesar de serem classificadas como espécies ou famílias diferentes). Ainda nesse sentido, a abordagem proposta apresentou bons resultados, já que também obteve aumento da recuperação de imagens de plantas em casos que a precisão foi considerada baixa.

Outro ponto que pode ser inserido como trabalho futuro é a aplicação de uma classificação para espécies de Angiosperma, ao invés de famílias. Como as espécies apresentam atributos específicos e não tão abrangentes como no nível de família, acredita-se que a seleção semântica de atributos possa ser mais efetiva, de modo a melhorar os resultados observados.

Além disso, trabalhos futuros também visam investigar técnicas similares e estudos que utilizam os mesmos conjuntos de dados e métricas, permitindo uma comparação dos resultados obtidos das métricas que não foram possíveis realizar tal comparação.

Dentre os trabalhos futuros, também pode-se citar a aplicação do RL-Sim a cada iteração do usuário. Visamos também estender o SIIR para diferentes perfis de usuários selecionando perguntas mais simples para aqueles inexperientes. Ainda, pode-se simplificar a ontologia de modo que a abordagem proposta realize perguntas com relação àquilo que se pode observar na imagem. Por exemplo: perguntas sobre as flores e folhas, e não

sobre a quantidade de óvulos (já que esse é um atributo oculto à imagem, sendo necessária a dissecação da flor).

Convém destacar como trabalhos futuros e desafios de pesquisa, a identificação automática pela imagem de tipos de flores e folhas descritas na ontologia. Ou seja, identificar automaticamente (sem a interação com um usuário) determinada imagem de planta como uma planta com flor do tipo “aclamídea”, por exemplo. A partir dessa identificação, fica fácil relacionar tal atributo identificado com as restrições das famílias (ou espécies) descritas na ontologia. De modo que, após sucessivas identificações automáticas de atributos, tenha-se como resultado final a identificação daquele espécime analisado.

Buscam-se também diferentes técnicas para utilizar as características de baixo-nível e as estruturas morfológicas para ambas serem utilizadas como entrada de uma abordagem baseada em aprendizado de máquina, com o intuito de distinguir e agrupar as imagens de plantas, com alta taxa de acurácia.

Uma análise aplicando recortes das imagens também pode ser realizada a fim de melhorar a eficácia. Essa avaliação pode ser aplicada de modo que recortes retangulares nas imagens são utilizados de modo a excluir áreas em que não ocorra a presença das flores, excluindo assim também as áreas que podem complicar a identificação e recuperação das imagens.

Fica claro que, além de possibilitar o desenvolvimento de novos trabalhos, conforme citado anteriormente, o presente método também pode ser aplicado em outras áreas do conhecimento, de modo a ser generalista, desde que haja a criação de uma nova ontologia e um conjunto de imagens para o novo domínio estudado.

Como a ideia é disponibilizar informações nos moldes de uma nova Web Semântica, pode-se afirmar que o desenvolvimento de vocabulários específicos deve crescer nas próximas décadas, de modo a relacionar vários domínios do conhecimento. Quando o desenvolvimento ocorre em conjunto por uma determinada comunidade de pesquisadores de diferentes áreas, seguindo

padrões estabelecidos, o conhecimento acaba se interligando em diversas áreas, já que pode ser compartilhado e reutilizado por estar disponível na Web.

Assim, o presente trabalho apresentou uma nova abordagem com alta eficácia na recuperação de imagens que une o conhecimento do domínio da Sistemática Vegetal com as características visuais das imagens.

8. Referências Bibliográficas

ADDIS, M.; BONIFACE, M.; GOODALL, S.; GRIMWOOD, P.; KIM, S.; LEWIS, P.; MARTINEZ, K.; STEVENSON, A. SCULPTEUR: Towards a New Paradigm for Multimedia Museum Information Handling. In: **Proceedings of the Second International Semantic Web Conference**, p. 582 -596. 2003.

AIGRAIN, P.; ZHANG, H.; PETKOVIC, D. Content-based representation and retrieval of visual media: a state-of-the-art review. **Multimedia Tools Appl**, v. 3, p. 179-202, 1996.

ANGELOVA, A.; ZHU, S.; LIN, Y.; WONG, J.; SHPECHT, C. Development and deployment of a large-scale flower recognition mobile app. **NEC Labs America Technical Report**, 2012.

ANGELOVA, A.; ZHU, S.; LIN, Y. Image segmentation for large-scale subcategory flower recognition. In: **Applications of Computer Vision (WACV), 2013 IEEE Workshop on**, p. 39-45, 2013.

AREVALILLO-HERRAEZ, M.; FERRI, F. J. An improved distance-based relevance feedback strategy for image retrieval. **Image and Vision Computing**, v. 31, n. 10, p. 704-713, 2013.

ARVOR, D.; DURIEUX, L.; ANDRÉS, S.; LAPORTE, M. Advances in Geographic Object-Based Image Analysis with ontologies: A review of main contributions and limitations from a remote sensing perspective. **Journal of Photogrammetry and Remote Sensing**, v. 82, p. 125-137, 2013.

BAI, X.; BAI, S.; WANG, X. Beyond diffusion process: Neighbor set similarity for fast re-ranking. **Information Sciences**, Elsevier, 325, 342-354, 2015.

BAI, S.; BAI, X. Sparse Contextual Activation for Efficient Visual Re-Ranking. **IEEE Transactions on Image Processing**, IEEE, 25, 1056-1069, 2016.

BANDIKOLLA, P.; REDDY, K.; REDDY, V. **Image retrieval using a combination of keywords and image features**. 2014. 45 f. Master Thesis (Computer Science) School of Engineering, Blekinge Institute of Technology. Sweden. 2014.

BAY, H.; ESS, A.; TUYTELAARS, T.; GOOL, L. V. "SURF: Speeded Up Robust Features". **Computer Vision and Image Understanding**, v. 110, n. 3, p. 346-359, 2008.

BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. **Scientific American**, v. 284, n. 5, p. 34-43, 2001.

BLASCHKE, T.; STROBL, J. What's wrong with pixels? Some recent developments interfacing remote sensing and GIS. **Interfacing Remote Sensing and GIS**, v. 6, p. 12-17, 2001.

BOSCH, A.; ZISSERMAN, A.; MUNOZ, X. Representing shape with a spatial pyramid kernel. In: **Proceedings of the 6th ACM international conference on Image and video retrieval**, p. 401-408, 2007.

BREITMAN, K. **Web Semântica**. A internet do futuro. Rio de Janeiro, Brasil: LTC, 2005. 190 p.

BURDESCU, D. D.; MIHAI, C. G.; STANESCU, L.; BREZOVAN, M. Automatic image annotation and semantic based image retrieval for medical domain. **Neurocomputing**, v. 109, p. 33-48, 2013.

CAI, D.; HE, X.; LI, Z.; MA, W.; WEN, J. Hierarchical clustering of www image search results using visual, textual and link information. In: **Proceedings of the ACM International Conference on Multimedia**, p. 952-959, 2004a.

CAI, D.; HE, X.; MA, W.; WEN, J.; ZHANG, H. Organizing www images based on the analysis of page layout and Web link structure. In: **Proceedings of the International Conference on Multimedia and Expo (ICME)**. 2004b.

CASTILLA, G.; HAY, G. J. Image objects and geographic objects. In: BLASCHKE, T.; LANG, S.; HAY, G.J. **Object-Based Image Analysis**. Springer Berlin Heidelberg, Berlin, Heidelberg, p. 91-110, 2008.

CHAI, Y.; LEMPITSKY, V. S.; ZISSERMAN, A. BiCoS: A Bi-level co-segmentation method for image classification. In: **ICCV, 2011**, p. 2579-2586, 2011.

CHANG, W.; WANG, J. Metadata for multi-level content-based retrieval. In: **Third IEEE Meta-Data Conference**. 1999.

CHATZICHRISTOFIS, S. A.; BOUTALIS, Y. S. CEDD: color and edge directivity descriptor: a compact descriptor for image indexing and retrieval. In: **Proceedings of the 6th international conference on Computer vision systems. ICVS'08**, p. 312-322, 2008a.

CHATZICHRISTOFIS, S. A.; BOUTALIS, Y. S. FCTH: fuzzy color and texture histogram – a low level feature for accurate image retrieval, In: **Ninth International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS'08)**, p. 191–196, 2008b.

CHEN, C.; LIN, L.; SHYU, M. Utilization of co-occurrence relationships between semantic concepts in re-ranking for information retrieval. In: **IEEE International Symposium on Multimedia (ISM2011)**, p. 53-60, 2011.

CHEN, C.; LIN, L.; SHYU, M. Re-ranking algorithm for multimedia retrieval via utilization of inclusive and exclusive relationships between semantic concepts. **International Journal of Semantic Computing**, v. 6, n. 2, p. 135-154, 2012.

CHEN, M.; CHEN, S. C.; SHYU, M. L.; WICKRAMARATNA, K. Semantic event detection via temporal analysis and multimodal data mining. **IEEE Signal Processing Magazine**. Special Issue on Semantic Retrieval of Multimedia, v. 23, n. 2, p. 38-46, 2006.

CHEN, Y.; LI, X.; DICK, A.; HILL, R. Ranking consistency for image matching and object retrieval. **Pattern Recognition**, v. 47, n. 3, p. 1349-1360, 2014.

CHENG, E.; JING, F.; ZHANG, L. A unified relevance feedback framework for web image retrieval. **IEEE Transactions on Image Processing**, v. 18, n. 6; p. 1350-1357, 2009.

CHIU, C.; LIN, H.; YANG, S. Texture retrieval with linguistic descriptors, In: **IEEE Pacific Rim Conference on Multimedia**, p. 308–315. 2001.

COTO, A. L. **The use of Ontologies for improving image retrieval and annotation**. Technical Report, Knowledge Media Institute (KMI). 73 p. 2008.

DATTA, R.; JOSHI, D.; LI, J.; WANG, J. Z. Image retrieval: Ideas, influences, and trends of the new age, **ACM Computing Surveys**, v. 40, n. 2, p. 1-60, 2008.

DENG, J.; DONG, W.; OCHER, R.; LI, L.; LI, K.; FEI-FEI, L. ImageNet: a large-scale hierarchical image database. In: **Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition**, p. 248-255, 2009.

DENG, Y.; MANJUNATH, B. Unsupervised segmentation of color-texture regions in images and video. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 23, n. 8, p. 800–810. 2001.

DUPPLAW, D.; DASMAHAPATRA, S.; HU, B.; LEWIS, P.; SHADBOLT, N. Multimedia Distributed Knowledge Management MIAKT. In: **ISWC 2004 Workshop on Knowledge Markup and Semantic Annotation**. 29 p. Japão, 2004.

EAKINS, J.P. Towards intelligent image retrieval. **Pattern Recognit**, v. 35, p. 3-14, 2002.

FARIA, F. F.; VELOSO, A.; ALMEIDA, H. M.; VALLE, E.; Das TORRES, R.; MEIRA JR, W.; GONÇALVES, M. A. Learning to rank for content-based image retrieval, **Multimedia Information Retrieval**, MIR'10, p. 285-294, 2010.

FENG, H.; CHUA, T. A bootstrapping approach to annotating large image collection. In: **Workshop on Multimedia Information Retrieval in ACM Multimedia**, p. 55-62. 2003.

FENG, H.; SHI, R.; CHUA, T. A bootstrapping framework for annotating and retrieving www images, In: **Proceedings of the ACM International Conference on Multimedia**, p. 960-967, 2004.

FERNÁNDEZ, M.; GÓMEZ-PÉREZ, A.; JURISTO, N. METHONTOLOGY: From Ontological Art Towards Ontological Engineering. In: **Spring Symposium Serie on Ontological Engineering**, p. 33-40, 1997.

FUJIKAWA, C. S.; PEDRONETTE, D. C. G. Análise e Combinação de Medidas Não-Supervisionadas de Estimativa de Eficácia. In: **XXVII SIBGRAPI Conference on Graphics, Patterns and Imagens, SIBGRAPI 2014**, p. 35-39. 2014

GAVA, T.; MENEZES, C. Especificação de Software Baseada em Ontologias. In. **III Escola Regional de Informática**, p. 167-205, 2003.

GIACINTO, G. A nearest-neighbor approach to relevance feedback in content based image retrieval. In: **CIVR'07, ACM**, p. 456-463, 2007.

GÓMEZ-PÉREZ, A. Towards a framework to verify knowledge sharing technology. **Expert Systems with Applications**, v. 11, n. 4, p. 519-529, 1996.

GÓMEZ-PÉREZ, A.; FERNÁNDEZ-PÉREZ, M.; CORCHO, O. **Ontological Engineering**. Springer Verlag, 2004.

GONÇALVES, V. **A Web Semântica no Contexto Educativo**. 2007. Tese de Doutorado em Engenharia Electrotécnica e de Computadores - Universidade do Porto, Porto,

2008. Disponível em: < <https://bibliotecadigital.ipb.pt/handle/10198/620> >. Acesso em: 28 nov. 2014.

GONÇALVES, E. G.; LORENZI, H. **Morfologia Vegetal**: Organografia e dicionário ilustrado de morfologia das plantas vasculares. Instituto Plantarum, Nova Odessa – SP, Brasil, 2007.

GOÑI, J.; PFEIFFER, M.; DE LUCENA, C. E-learning e a Web Semântica. In: Congresso da Rede Ibero-americana de Informática Educativa (RIBIE), 6., 2002, Rio de Janeiro. **Anais eletrônicos**. Rio de Janeiro, 2002 Disponível em: < ftp://ftp.inf.puc-rio.br/pub/docs/techreports/02_12_goni.pdf >. Acesso em: 28 nov. 2014.

GONZALEZ, R. C.; WOODS, R. E. **Digital Image Processing**. Addison-Wesley, Boston, MA, USA, 2001.

GRUBER, T. A translation approach to portable Ontologies. **Knowledge Acquisition**, California, v.5, n. 2, p. 199-220, Apr. 1993.

GUAN, J.; QIU, G. Learning user intention in relevance feedback using optimization. In: **ACM MIR '07, ACM**. New York, NY, USA, p. 41-50, 2007.

GUARINO, N. Formal ontology, conceptual analysis and knowledge representation. **International Journal of Human-Computer Studies**, v. 43, p. 625-640, 1995.

GUARINO, N. **Formal ontology in information systems**. Amsterdam: IOS Press, 1998.

HALASCHEK-WIENER, C.; SCHAIN, A.; GROVE, M.; PARSIA, B.; HENDLER, J. Management of digital images on the semantic Web. In: **Proceedings of the International Semantic Web Conference**. 15 p. 2005.

HAN, J.; MA, K. Rotation-invariant and scale invariant gabor features for texture image retrieval. **Image and Visual Computing**, 25, 1474–1481, 2007.

HANBURY, A. A survey of methods for image annotation. **Journal of Visual Languages and Computing**, v. 19, p. 617-627, 2007.

HOI, S.; LIU, W.; CHANG, S. Semi-supervised Distance Metric Learning for Collaborative Image Retrieval and Clustering. In: **ACM Trans. Multimedia Comput. Commun. Appl.** USA. 2010.

HOLLINK, L.; SCHREIBER, G.; WIELEMAKER J.; WIELINGA, B. Semantic Annotation of Image Collections. In: **Proceedings of Knowledge Capture - Knowledge Markup and Semantic Annotation Workshop**. 8 p. 2003.

HORRIDGE, M.; KNUBLAUCH, H.; RECTOR, A.; STEVENS, R.; WROE, C., JUPP, S.; MOULTON, G.; DRUMMOND, N. **Practical guide to building owl ontologies using the protégé 4 and Co-ode tools**, v. 1.2, Manchester: The university of Manchester, 2009.

HU, W.; HU, R.; XIE, N.; LING, H.; MAYBANK, S. Image classification using multiscale information fusion based on saliency driven nonlinear diffusion filtering. In: **IEEE transactions on image processing**, v. 23, p. 1513-1526, 2014.

HUANG, J.; KUMAR, S. R.; MITRA, M. ; ZHU, W. J.; ZABIH, R. Image indexing using color correlograms, In: **IEEE Conference on Computer Vision and, Pattern Recognition (CVPR'97)**, p. 762-768, 2007.

HUI, H. W.; MOHAMAD, D.; ISMAIL, N. Semantic gap in CBIR: automatic objects spatial relationships semantic extraction and representation, **International Journal of Image Processing (IJIP)**, v. 4, n. 3, p. 192-204. 2010.

ITO, S.; KUBOTA, S. Object classification using heterogeneous co-occurrence features. In: **Computer Vision – ECCV 2010: Lecture Notes in Computer Science**, v. 6315, p. 701-714, 2010.

JARRETT, K.; KAVUKCUOGLU, K.; RANZATO, M.; LECUN, Y. What is the best multi-stage architecture for object recognition? In: **IEEE 12th International Conference on Computer Vision**, p. 2146-2153, 2009.

JIA, Y.; SHELFHAMER, E.; DONAHUE, J.; KARAYEV, S.; LONG, J.; GIRSHICK, R.; GUADARRAMA, S.; DARRELL, T. Caffe: Convolutional architecture for fast feature embedding. 2014. In: **Proceedings of the 22Nd ACM International Conference on Multimedia**. USA. 2014.

JIANG, Y.; WANG, J.; CHANG, S.; NGO, C. Domain adaptive semantic diffusion for large scale context-based video annotation. In: **IEEE International Conference on Computer Vision (ICCV09)**, p. 1420-1427, 2009.

JING, F.; LI, M.; ZHANG, L.; ZHANG, H.; ZHANG, B. Learning in region based image retrieval, In: **Proceedings of the International Conference on Image and Video Retrieval**, p. 206-215, 2003.

KANAN, C.; COTTRELL, G. Robust classification of objects, faces, and flowers using natural image statistics. In: **Computer Vision and Pattern Recognition (CVPR)**, p. 2472-2479, 2010.

KEBAPCI, H.; YANIKOGLU, B.; UNAL, G. Plant Image Retrieval Using Color, Shape and Texture Features. **The Computer Journal Advance Access**, 2010.

KHAN, F. S.; WEIJER, V. J.; VANRELL, M. Top-down color attention for object recognition. In: **IEEE 12th International Conference on Computer Vision**, p. 979-986, 2009.

KHAN, F. S.; WEIJER, J.; BAGDANOV, A. D.; VANRELL, M. Portmanteau vocabularies for multi-cue image representation. **Advances in neural information processing systems**, p. 1323-1333, 2011.

KOKAR, M. M.; WANG, J. Using ontologies for recognition: an example. In: **International Society of Information Fusion (ISIF)**, p. 1324-1330, 2002.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: **Neural Information Processing Systems (NIPS)**, p. 1106-1114, 2012.

KUNDU, M. K.; CHOWDHURY, M.; BULO, S. R. A graph-based relevance feedback mechanism in content-based image retrieval. **Knowledge-Based Systems**, v. 73, p. 254-264, 2015.

KURTZ, C.; DEPEURSINGE, A.; NAPEL, S.; BEAULIEU, C. F.; RUBIN, D. L. On combining image-based and ontological semantic dissimilarities for medical image retrieval applications. **Medical image analysis**, v. 18, p. 1082-1100, 2014.

LACY, L. **OWL**: Representing information using the Web Ontology Language. Victoria, Canada: Trafford, 2005. 285p.

LeCUN, Y.; BOTTOU, L.; BENGIO, Y.; HAFFNER, P. Gradient-based learning applied to document recognition. In: **Proceedings of the IEEE**, v. 86, p. 2278-2324, 1998.

LeNet. 2016. **Convolutional Neural Networks (LeNet)**. Disponível em: < <http://deeplearning.net/tutorial/lenet.html> >. Acesso em: 01 jul. 2016.

LEW, M. S.; SEBE, N.; DJERABA, C.; JAIN, R. Content-based multimedia information retrieval: State of the art and challenges. **ACM Transactions on Multimedia Computing, Communications, and Applications**, p. 1-19, 2006.

LI, J. Reachability based Ranking in Interactive Image Retrieval. In: **Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval**, p. 867-870, 2015.

LIN, L.; SHYU, M. L. Effective and efficient video high-level semantic retrieval using associations and correlations, **International Journal of Semantic Computing**, v. 3, n. 4, p. 421-444, 2009.

LIU, C.; WECHSLER, H. A Gabor Feature Classifier for Face Recognition. **ICCV**, v. 2, p. 270-275, 2001.

LIU, Y.; LIU, T.; QIN, Z. MA, H. L. Supervised rank aggregation. In: **International Conference on World Wide Web (WWW'2007)**, p. 481-490. 2007a.

LIU, Y.; ZHANG, D.; LU, G.; MA, W. A survey of content-based image retrieval with high-level semantics. **The journal of the pattern recognition society**, v. 40, p. 262-282. 2007b.

LIU, Y.; GUO, Q.; KELLY, M. A framework of region-based spatial relations for non-overlapping features and its application in object based image analysis. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 63, p. 461-475, 2008.

LUNDÉN, J.; KOIVUNEN, V. Deep learning for HRRP-based target recognition in multistatic radar systems. In: **2016 IEEE Radar Conference**, p. 1-6, 2016.

LUX, M.; CHATZICHRISTOFIS, S. A. LIRe: Lucene Image Retrieval – An Extensible Java CBIR Library. In: **Proceeding MM '08 – Proceedings of the 16th ACM international conference on Multimedia**, p. 1085-1087, 2008.

LUX, M. LIRe: Open Source Image Retrieval in Java. In: **Proceeding MM '13 Proceedings of the 21st ACM international conference on Multimedia**, p. 843-846, 2013.

LUX, M.; MARQUES, O. **Visual Information Retrieval using Java and LIRe**. Morgan & Claypool Publishers, 2013.

MA, H.; ZHU, J.; LYU, M. R. T.; KING, I. Bridging the semantic gap between images and tags. **IEEE Trans. Multimedia**, v. 12, p. 462-473, 2010.

MA, W.; MANJUNATH, B. Texture Features and Learning Similarity. **Proc. Conf. Computer Vis. Pattern Recognition (CVPR)**, p. 425–430, 1996.

MANZOOR, U.; BALUBAID, M. A.; ZAFAR, B.; UMAR, H.; KHAN, M. S. Semantic Image Retrieval: An Ontology Based Approach. **International journal of advanced research in artificial intelligence**, v. 4, n. 4, 2015.

MATTOS, A. B.; HERRMANN, R.; SHIGENO, K. K.; FERIS, R. A mission-oriented citizen science platform for efficient flower classification based on combination of feature descriptors. In: **EMR@ ICMR**, p. 45-52, 2014.

MEHTA, R.; EGIAZARIAN, K. Rotated Local Binary Pattern (RLBP): Rotation invariant texture descriptor. In: **2nd International Conference on Pattern Recognition Applications and Methods, ICPRAM 2013**, p. 497-502. 2013.

MEZARIS, V.; KOMPATSIARIS, I.; STRINTZIS, M. G. An ontology approach to object-based image retrieval. In: **Proceedings of the IEEE International Conference on Image Processing**, v. II, p. 511-514. 2003.

MOBOT (Missouri Botanical Garden). 2014. **ANGIOSPERM PHYLOGENY WEBSITE**. Disponível em: < <http://www.mobot.org/MOBOT/research/APWeb/> >. Acesso em 10 dez. 2014.

NILSBACK, M.; ZISSERMAN, A. A visual vocabulary for flower classification. In: **Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition**, v. 2, p. 1447-1454, 2006.

NILSBACK, M.; ZISSERMAN, A. Delving into the Whorl of Flower Segmentation. In: **Proceedings of the British Machine Vision Conference**, v. 1, p. 570-579, 2007.

NILSBACK, MARIA-ELENA; ZISSERMAN, A. Automated flower classification over a large number of classes. In: **Proceedings of the Indian Conference on Computer Vision, Graphics and Image Processing**, 8 p. 2008.

NILSBACK, MARIA-ELENA; ZISSERMAN, A. 2009. **17 Category Flower Dataset**. Disponível em: < <http://www.robots.ox.ac.uk/~vgg/data/flowers/17/index.html> >. Acesso em: 18 jul. 2015.

NILSBACK, M.-E.; ZISSERMAN, A. **An automatic visual flora-segmentation and classification of flower images**. Ph.D. thesis, Oxford University, 2009a.

NOY, N. F.; MCGUINNESS, D. L. 2006. **Ontology development 101: a guide to creating your first ontology**. Disponível em: < http://protege.stanford.edu/publications/ontology_development/ontology101-noymcguinness.html >. Acesso em: 11 dez. 2014.

OJALA, T.; PIETIKÄINEN, M.; MÄENPÄÄ, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. **Pattern Analysis and Machine Intelligence**, v. 24, n. 7, p. 971–987, 2002.

OOI, B. C.; TAN, K.-L.; WANG, S.; WANG, W.; CAI, Q.; CHEN, G.; GAO, J.; LUO, Z.; TUNG, A. K.; WANG, Y.; ZHONGLE, X.; ZHANG, M.; ZHENG, K. SINGA: A distributed deep learning platform. In: **Proceedings of the 23rd ACM international conference on Multimedia**, p. 685-688. 2015.

OUYANG, X.; ZHOU, P.; LI, C. H.; LIU, L. Sentiment Analysis Using Convolutional Neural Network. In: **Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing (CIT/IUCC/DASC/PICOM), IEEE International Conference**, p. 2359-2364, 2015.

PENATTI, O. A.; TORRES, R. S. Eva: an evaluation tool for comparing descriptors in content-based image retrieval tasks. In: **Proceedings of the international conference on Multimedia information retrieval**, p. 413-416, 2010

PANDEY, S.; KHANNA, P.; YOKOTA, H. An effective use of adaptive combination of visual features to retrieve image semantics from a hierarchical image database. **Journal of Visual Communication and Image Representation**, v. 30, p. 136-152, 2015.

PEDRONETTE, D. C. G.; TORRES, R. S. Exploiting clustering approaches for image re-ranking. **Journal of Visual Languages and Computing**, v. 22, p. 453-466, 2011.

PEDRONETTE, D. C. G.; TORRES, R. S. Image re-ranking and rank aggregation based on similarity of ranked lists. **Pattern Recognition**, v. 46, p. 2350-2360, 2013.

PEDRONETTE, D. C. G.; PENATTI, O. A.; TORRES, R. S. Unsupervised manifold learning using reciprocal knn graphs in image re-ranking and rank aggregation tasks. **Image and Vision Computing**, Elsevier, 32, 120-130, 2014.

PENATTI, O. A. B.; VALLE, E.; TORRES, R. S. Comparative study of global color and texture descriptors for Web image retrieval. **Journal of Visual Communication and Image Representation**, v. 23, p. 359-380, 2012.

QIN, D.; GAMMETER, S.; BOSSARD, L.; QUACK, T.; VAN GOOL, L. Hello neighbor: Accurate object retrieval with k-reciprocal nearest neighbors. In: **CVPR**, p. 777-784, 2011.

RAUTENBERG, S.; TODESCO, J. L.; GAUTHIER, F. A. O. Processo de desenvolvimento de ontologias: uma proposta e uma ferramenta. **Tecnologia**, Fortaleza, v. 30, n. 1, p. 133-144, 2009.

RAZAVIAN, A. S.; AZIZPOUR, H.; SULLIVAN, J.; CARLSSON, S. CNN features off-the-shelf: an astounding baseline for recognition. In: **Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops**, p. 806-813, 2014.

REN, J.; SHEN, Y.; GUO, L. A novel image retrieval based on representative colors. In: **Proceedings of the Image and Vision Computing**, p. 102–107. 2003.

RUSSAKOVSKY, O.; DENG, J.; SU, H.; KRAUSE, J.; SATHEESH, S.; MA, S.; HUANG, Z.; KARPATY, A.; KHOSLA, A.; BERNSTEIN, M.; A. C. BERG, L. FEI-FEI. ImageNet large scale visual recognition challenge. **International Journal of Computer Vision**, Springer, v. 115, p. 211-252, 2015.

SALTON, G.; MCGILL, M. J. **Introduction to modern information retrieval**. McGraw-Hill, New York, 1998.

SCHREIBER, G.; DUBBELDAM, B.; WIELEMAKER, J.; WIELINGA, B. Ontology-Based Photo Annotation. **IEEE Intelligent Systems**, v. 16, n. 3, p. 66-74, 2001.

SCHWANDER, O.; NIELSEN, F. Re-ranking with contextual dissimilarity measures from representational Bregman k-means, In: **Vision Theory and Application – VISAPP**, v.1,p.118–122, 2010.

SEMANTIC METADATA. **LIRe – An open source java content based image retrieval library**. Disponível em: < <http://www.semanticmetadata.net/lire/> >. Acesso em: 18 jul. 2015.

SETHI, I. K.; COMAN, I. L. Mining association rules between low-level image features and high-level concepts, In: **Proceedings of the SPIE Data Mining and Knowledge Discovery**, v. III, p. 279–290. 2001.

SHI, R.; FENG, H.; CHUA, T.; LEE, C. An adaptive image content representation and segmentation approach to automatic image annotation. In: **International Conference on Image and Video Retrieval**, p. 545–554, 2004.

SMEULDERS, A.W.M.; WORRING, M.; SANTINI, S.; GUPTA, A.; JAIN, R. Content-based image retrieval at the end of the early years. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 22, n. 12, p. 1349-1380, 2000.

SMITH, J. R.; LI, C. S. Decoding image semantics using composite region templates, In: **IEEE Workshop on Content-Based Access of Image and Video Libraries**, p. 9–13, 1998.

SOURCE FORGE. 2016. **OWL API**. Disponível em: < <http://owlapi.sourceforge.net/> >. Acesso em: 02 jul. 2016.

SOUZA, V. C.; LORENZI, H. **Botânica Sistemática**: Guia ilustrado para identificação das famílias de Angiospermas da flora brasileira, baseado em APG II. Instituto Plantarum, Nova Odessa – SP, Brasil, 2005.

SOUZA, V. C.; LORENZI, H. **Chave de Identificação**: Para as principais famílias de Angiospermas nativas e cultivadas do Brasil. Instituto Plantarum, Nova Odessa – SP, Brasil, 2007.

SOUZA, V. C.; LORENZI, H. **Botânica Sistemática**: Guia ilustrado para identificação das famílias de Fanerógamas nativa e exóticas no Brasil, baseado em APG III. Instituto Plantarum, Nova Odessa – SP, Brasil, 2009.

SOUZA, V. C.; FLORES, T. B.; LORENZI, H. **Introdução à Botânica**: Morfologia. Instituto Plantarum, Nova Odessa – SP, Brasil, 2013.

STAAB, S.; STUDER, R. Handbook on Ontologies, **International Handbooks on Information Systems**. Berlin: Springer-Verlag, 2004.

STANCHEV, P. L.; GREEN, D.; DIMITROV, B. High level color similarity retrieval. **The International Journal "Information Theories and Applications**, v. 10, n. 3, p. 363–369, 2003.

STANFORD UNIVERSITY. 2014. **Protégé**. Disponível em: < <http://protege.stanford.edu/products.php> >. Acesso em 11 dez. 2014

STEHLING, R. O.; NASCIMENTO, M. A.; FALCÃO, A. X. A compact and efficient image retrieval approach based on border/interior pixel classification. In: **Proceedings of the eleventh international conference on Information and knowledge management**, p. 102-109, 2002.

TAMURA, H.; MORI, S.; YAMAWAKI, T. Textural features corresponding to visual perception. **IEEE Transactions on Systems, Man, and Cybernetics**. v. 8, n. 6, p. 460–472, 1978.

THOMEE, B.; LEW, M. Interactive search in image retrieval: a survey. **International Journal of Multimedia Information Retrieval**, n. 1 v. 2, p. 71-86, 2012.

TIAN, X.; YANG, L.; WANG, J.; YANG, Y.; WU, X; HUA, X. Bayesian video search re-ranking. In: **ACM International Conference on Multimedia (MM08)**, p. 131-140, 2008.

VAILAYA, A., FIGUEIREDO, M. A. T.; JAIN, A. K.; ZHANG, H. J. Image classification for content-based indexing. **IEEE Transactions on Image Processing**, v. 10, n. 1, p. 117–130, 2001.

VAN DE SANDE, K. E. A.; GEVERS, T.; SNOEK, C. G. M. Evaluating Color Descriptors for Object and Scene Recognition. **Pattern Analysis and Machine Intelligence, IEEE Transactions on**, v. 32, n. 9, p. 1582-1596, 2009.

VARMA, M.; RAY, D. Learning the discriminative power-invariance trade-off. In: **IEEE International Conference on Computer Vision, IEEE International Conference 2007**, p. 1-8, 2007.

VELTKAMP, R. C.; BURKHARDT, H.; TANASE, M. **State-of-the-art in content-based image and video retrieval**. Kluwer Academic Publ. 2001.

VISUAL GEOMETRY GROUP. 2009. **17 Category Flower Dataset**. Disponível em: < <http://www.robots.ox.ac.uk/~vgg/data/flowers/17/index.html> >. Acesso em 01/09/2015.

VOGUEL, J.; SCHIELE, B. Semantic Modeling of Natural Scenes for Content-Based Image Retrieval. **International Journal of Computer Vision**, v. 72, n. 2, p. 133-157, 2007.

W3C (World Wide Web Consortium). 2012. **OWL**. Disponível em: < <http://www.w3.org/2001/sw/wiki/OWL> >. Acesso em 03 dez. 2014.

W3C (World Wide Web Consortium). 2014. **Semantic Web Frequently Asked Questions**. Disponível em: < <http://www.w3.org/2001/sw/SW-FAQ#What1> >. Acesso em 28 nov. 2014.

WALLS, R. L.; ATHREYA, B.; COOPER, L.; GANDOLFO, M. A.; JAISWAL, P.; MUNGALL, C. J.; PREECE, J.; RENSING, S.; SMITH, B.; STEVENSON, D. W. Ontologies as integrative tools for plant science. **American Journal of Botany**, v. 99, n. 8, p. 1263-1275, 2012.

WANG, W.; OOI, B. C.; YANG, X.; ZHANG, D.; ZHUANG, Y. Effective multi-modal retrieval based on stacked auto-encoders. In: **Proceedings of the VLDB Endowment**, v. 7, p. 649-660, 2014.

WANG, W.; YANG, X.; OOI, B. C.; ZHANG, D.; ZHUANG, Y. Effective deep learning-based multi-modal retrieval. **The VLDB Journal**, v. 25, p. 79-101, 2016.

WEI, X.; JIANG Y.; NGO, C. Exploring inter-concept relationship with context space for semantic video indexing. In: **ACM International Conference on Image and Video Retrieval**, 8 p., 2009.

WON, C. S.; PARK, D. K.; PARK, S. J. Efficient use of MPEG-7 edge histogram descriptor. **ETRI Journal**, v. 24, n. 1, p. 23–30, 2002.

XIE, N.; LING, H.; HU, W.; ZHANG, X. Use bin-ratio information for category and scene classification. In: **Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on**, p. 2313-2319, 2010.

YANG, X.; BAI, X.; LATECKI, L. J.; TU, Z. Improving shape retrieval by learning graph transduction. In: **European Conference on Computer Vision (ECCV'2008)**, v. 4, p. 788-801, 2008.

YANG, J.; ZHU, S. Narrowing Semantic Gap in Content-based Image Retrieval. In: **2012 International Conference on Computer Distributed Control and Intelligent Environmental Monitoring**, p. 433-438, 2012.

ZAGORIS, K.; CHATZICHRISTOFIS, S.; PAPAMARKOS, N.; BOUTALIS, Y. Automatic image annotation and retrieval using the joint composite descriptor, In: **14th Panhellenic Conference on Informatics (PCI)**, p. 143–147, 2010.

ZAGORIS, K.; CHATZICHRISTOFIS, S.; ARAMPATZIS, A. Bag-of-Visual-Words vs Global Image Descriptors on Two-Stage Multimodal Retrieval, In: **SIGIR '11 Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval**, p. 1251-1252, 2011.

ZHANG, L.; LIU, F.; ZHANG, B. Support vector machine learning for image retrieval. In: **International Conference on Image Processing**, p. 7–10, 2001.

ZHANG, C.; CHAI, J. Y.; JIN, R. User term feedback in interactive text-based image retrieval. **Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval**, p. 51-58, 2005.

ZHANG, L.; ZHOU, Z.; LI, H. Binary Gabor pattern: An efficient and robust descriptor for texture classification. In: **19th IEEE International Conference on Image Processing**, p. 81-84, 2012.

9. Apêndice A – Análises Estatísticas

Este apêndice apresenta os gráficos de P@5, P@10 e MAP, com seus respectivos intervalos de confiança, para cada pergunta respondida dos ranques analisado para os conjuntos de imagens estudados.

O intervalo de confiança foi calculado com o valor de confiança referente à 0.95.

9.1 *Oxford Flowers 17 Classes*

A presente seção apresenta os intervalos de confiança para as métricas avaliadas no conjunto de imagens *Oxford Flowers 17 Classes*. Os gráficos apresentados nessa seção diferenciam dos gráficos 14, 15 e 16 (Capítulo 6, Seção 6.4.1), pois demonstram o intervalo de confiança de cada um dos pontos analisados.

No Gráfico 14.A, observa-se que vários pontos não apresentam intersecção de seus respectivos intervalos de confiança, demonstrando assim, diferenças estatísticas entre alguns pontos analisados.

Pode-se observar um comportamento semelhante no Gráfico 15.A, já que vários pontos não apresentam intersecção entre seus intervalos de confiança, de modo que fica evidente as diferenças estatísticas entre tais pontos.

É interessante notar que, em curvas com valores menores de P@5, P@10 e MAP, as distâncias entre os intervalos de confiança (Gráficos 14.A, 15.A e 16.A) são maiores do que nas curvas que apresentam valores mais altos de P@5, P@10 e MAP (CNN e CNN+SURF). Isso demonstra a importância de se realizar mais perguntas para métodos com baixa taxa de precisão, além de corroborar também com a eficácia da abordagem proposta.

Gráfico 14.A: Intervalo de Confiança do P@5 para os métodos de extração com a Recuperação Semântica Interativa de Imagens (SIIR) – 17 Classes.

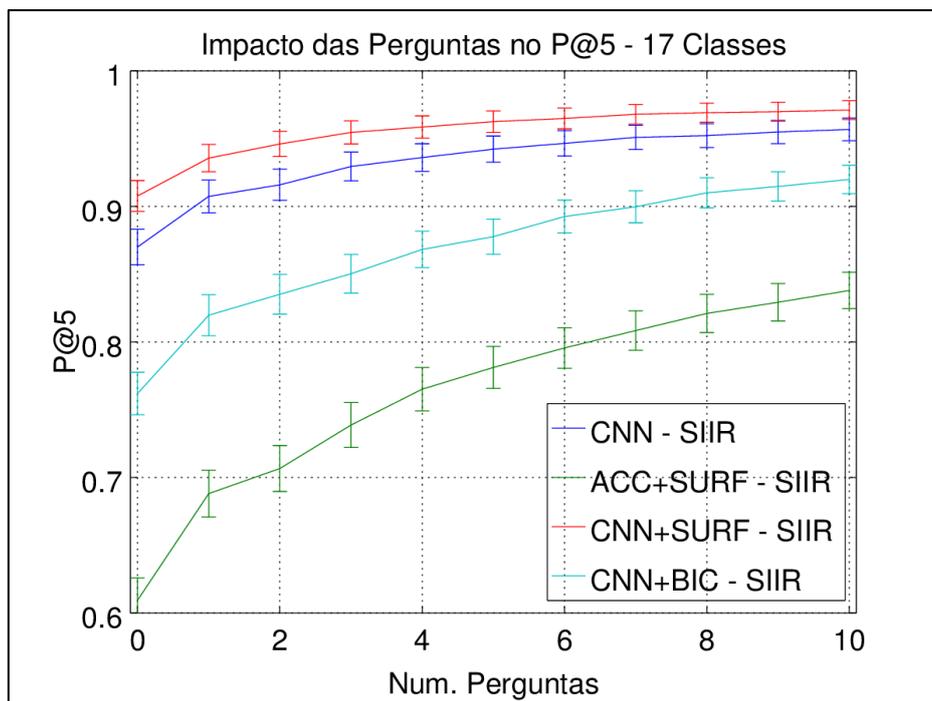


Gráfico 15.A: Intervalo de Confiança do P@10 para os métodos de extração com a Recuperação Semântica Interativa de Imagens (SIIR) – 17 Classes.

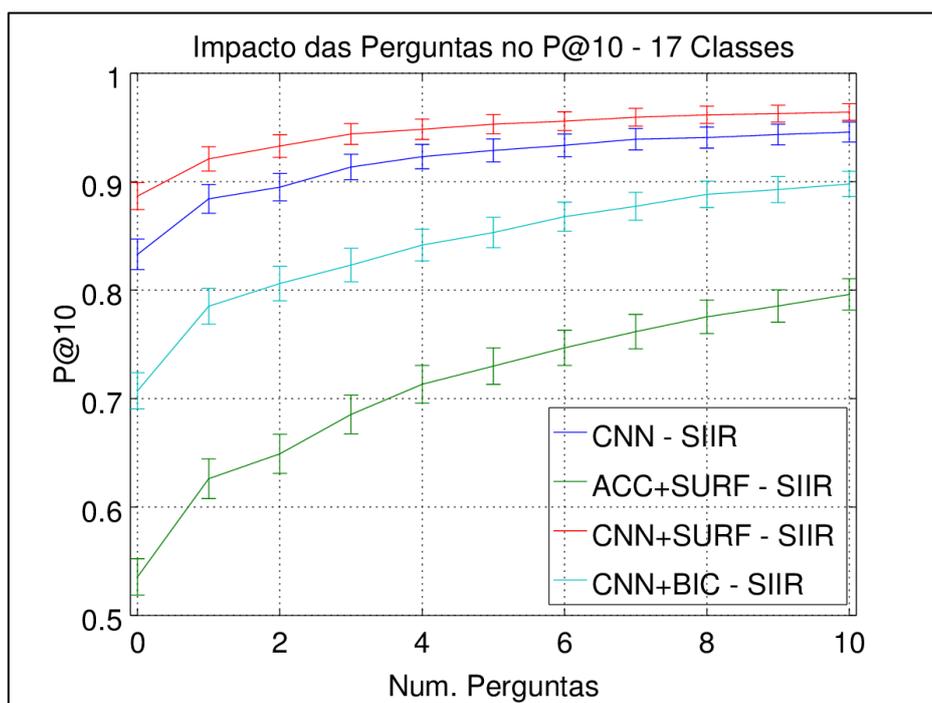
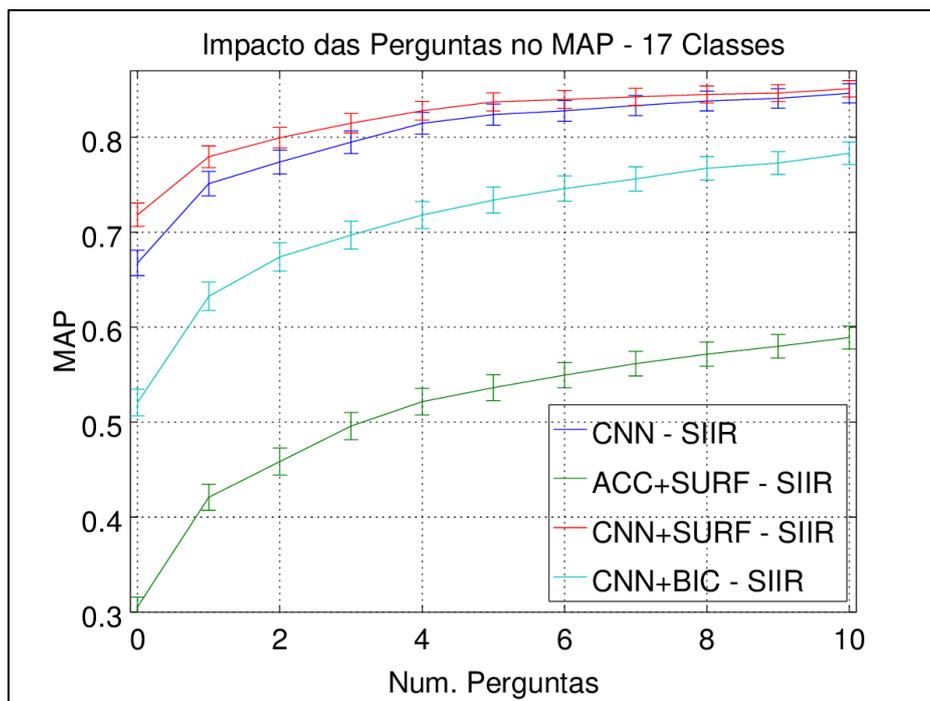


Gráfico 16.A: Intervalo de Confiança do MAP para os métodos de extração com a Recuperação Semântica Interativa de Imagens (SIIR) – 17 Classes.



9.2 Oxford Flowers 102 Classes

A presente seção apresenta os intervalos de confiança para as métricas avaliadas no conjunto de imagens *Oxford Flowers 102 Classes*. Os gráficos apresentados nessa seção diferenciam dos gráficos 18, 19 e 20 (Capítulo 6, Seção 6.4.2), pois demonstram o intervalo de confiança de cada um dos pontos analisados.

Pode-se observar no Gráfico 18.A a não intersecção entre os intervalos de confiança de alguns pontos, demonstrando assim, diferenças estatísticas entre tais pontos analisados.

No Gráfico 19.A, observa-se que vários pontos não apresentam intersecção de seus respectivos intervalos de confiança, demonstrando novamente, diferenças estatísticas entre tais pontos analisados.

Gráfico 18.A: Intervalo de Confiança do P@5 para os métodos de extração com a Recuperação Semântica Interativa de Imagens (SIIR) – 102 Classes.

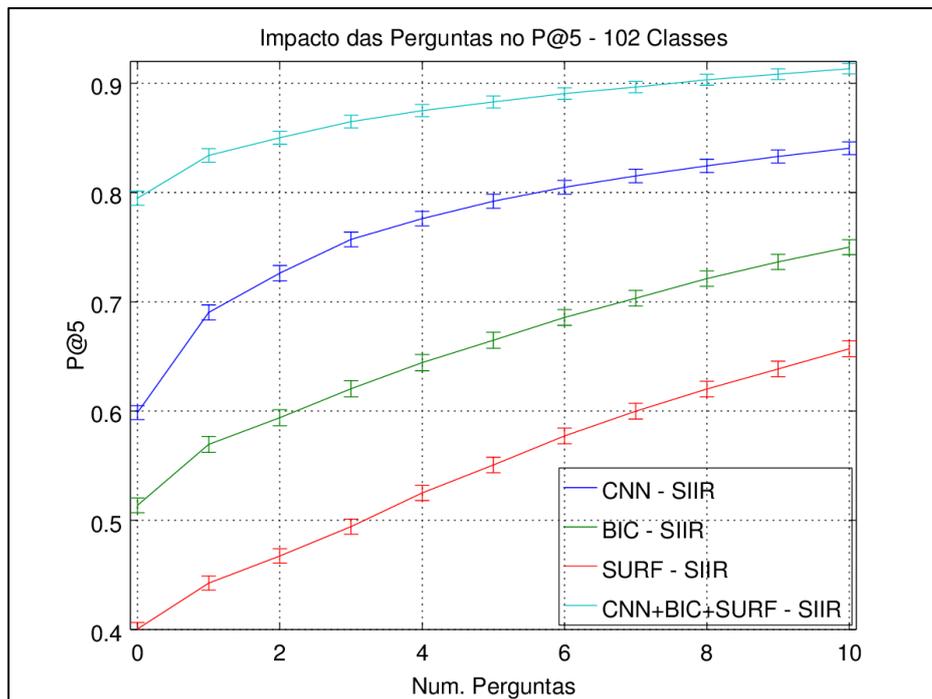


Gráfico 19.A: Intervalo de Confiança do P@10 para os métodos de extração com a Recuperação Semântica Interativa de Imagens (SIIR) – 102 Classes.

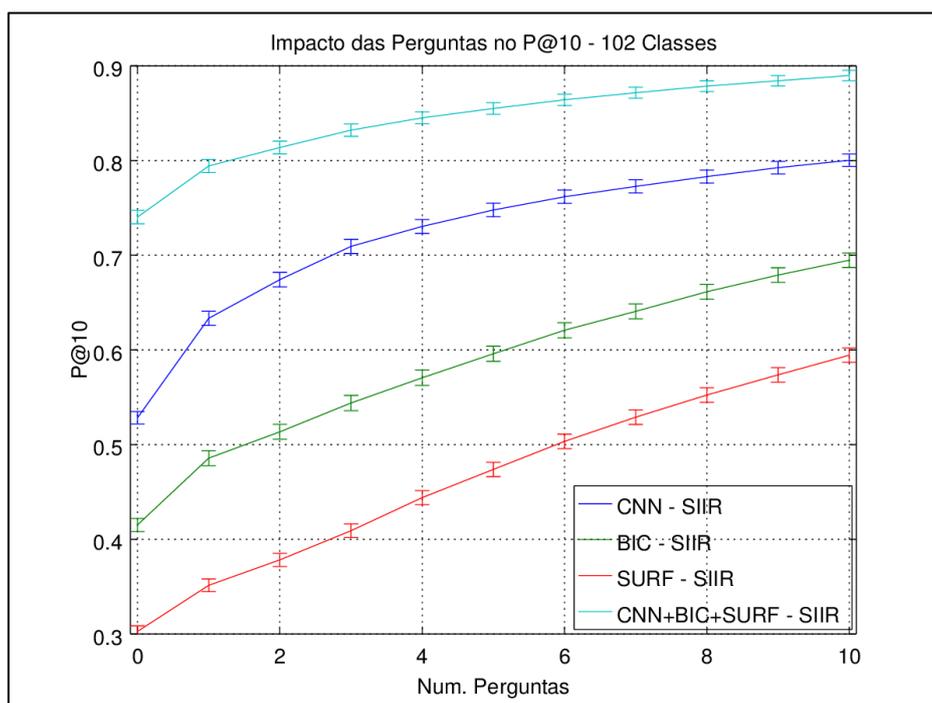
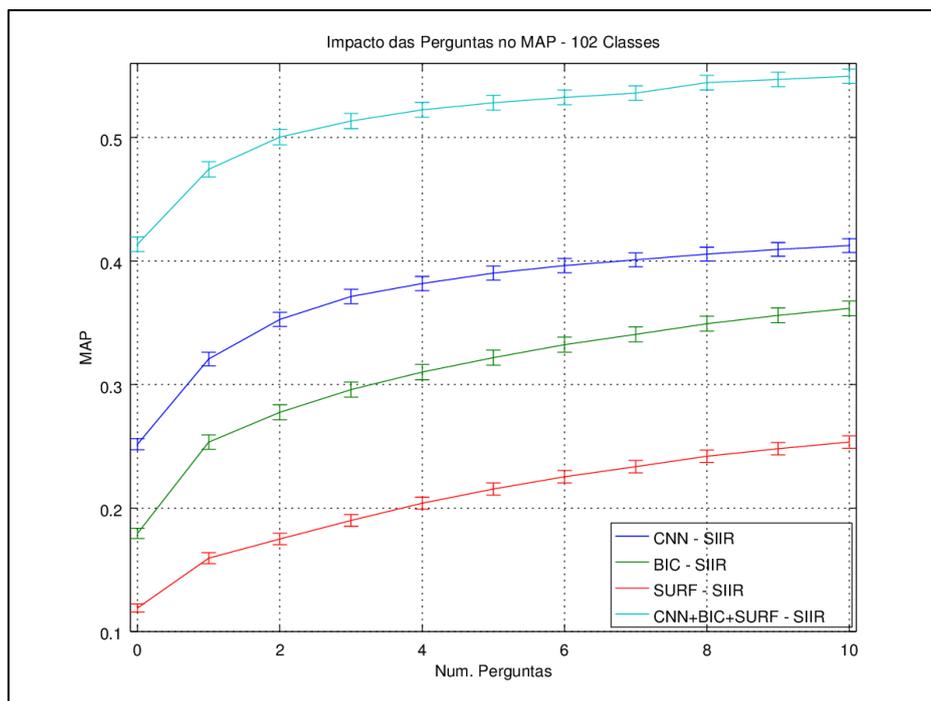


Gráfico 20.A: Intervalo de Confiança do MAP para os métodos de extração com a Recuperação Semântica Interativa de Imagens (SIIR) – 102 Classes.



Quando comparados os gráficos do intervalo de confiança do *102 Classes* (Gráficos 18.A, 19.A e 20.A – Apêndice A) com os gráficos do *17 Classes* (Gráficos 15.A, 16.A e 17.A – Apêndice A), pode-se observar que os intervalos de confiança dos pontos analisados para o *102 Classes* são menores do que aqueles apresentados para o *17 Classes*. Isso se deve ao fato de que o número de comparações entre os *rankings* e o número de imagens analisadas é maior no conjunto de imagens *102 Classes*.

Outro ponto interessante de se notar nos Gráficos 18.A, 19.A e 20.A – referentes ao *102 Classes* – é que as distâncias entre os intervalos de confiança dos pontos analisados também são maiores em curvas que apresentam valores mais baixos (SURF) de P@5, P@10 e MAP. Porém, também observa-se que as distâncias entre os intervalos de confiança para curvas com melhores valores para as métricas testadas também são altas, diferentemente daquilo observado nos Gráficos 15.A, 16.A e 17.A (Apêndice A) – referentes ao *17 Classes*. Esse comportamento ocorre, pois o conjunto de dados *102 Classes* apresenta maior número de imagens e maior número de

classes de imagens, de modo que esse aumento de imagens, quando comparado com o *17 Classes*, diminui a taxa de recuperação, já que os erros de classificação aumentam. De modo que, mesmo curvas com melhores valores para as métricas testadas demonstram a necessidade da aplicação da abordagem proposta.

Nesse sentido, a abordagem proposta, ao realizar perguntas de modo contínuo, demonstra que a eficácia da classificação e recuperação de imagens aumenta com diferença estatística entre alguns pontos analisados dos gráficos presentes no Apêndice A.