Recuperação de Imagens na Web Baseada em Múltiplas Evidências Textuais

Tatiana A. S. Coelho Lamarque Vieira Souza Berthier Ribeiro-Neto

Departamento de Ciência da Computação
Universidade Federal de Minas Gerais
31270-901 Belo Horizonte - MG Brasil
{tasc,lamarque,berthier}@dcc.ufmg.br

Resumo

A crescente quantidade de bancos de dados de imagens disponíveis na Web e o fato das máquinas de busca atuais não satisfazerem a necessidade de informação do usuário sugerem o estudo de novos algoritmos de recuperação de informação. As máquinas de busca de imagens existentes normalmente utilizam técnicas de processamento digital de imagem, ou mesmo utilizam um processo de recuperação baseado em rótulos simples, tais como o nome do arquivo que contém a imagem. Como resultado, a tarefa de se encontrar as imagens que mais satisfazem a necessidade de informação do usuário continua sendo desafiadora. De forma diferente do que ocorre nos algoritmos hoje aplicados às máquinas de busca de imagens, acreditamos que a maioria das fontes de evidência disponíveis nos documentos HTML que contém as imagens podem ser utilizadas no processo de recuperação e ordenação das imagens no conjunto-resposta a uma consulta do usuário. Neste artigo, apresentamos várias alternativas para indexação de imagens e propomos um arcabouço Bayesiano que permite combiná-las para gerar um ranking de qualidade superior. Mostramos que esse arcabouço fornece um mecanismo efetivo de busca e ordenação de imagens na Web. Mostramos também que a combinação de várias fontes de evidência textuais aumenta a qualidade das respostas.

Abstract

The growing availability of image databases on the Web and the fact that available image search engines don´t satisfy the user information necessity suggest the study of new information retrieval algorithms. Current image search engines use image digital processing techniques, or use a retrieval process based on simple labels, such as the image file name. As a result, the task of finding the best images that satisfy the user information need continues to be a challenging problem. In a different form as occur in the current algorithms applied to the image search engines, we understand that most of the sources of evidence that are available in the HTML document that contains an image should be used to retrieve and rank images in the answer set of the user query. In this paper, we explore various alternatives for indexing images, and we propose a Bayesian framework which allows combine them to obtain an improved quality ranking. We show that this framework provides effective mechanism for searching and ranking images on the Web. We also show that many sources of textual evidences improve meaningfully the answers quality.

1 Introdução

A crescente quantidade de bancos de dados de imagens disponíveis na Web e o fato das máquinas de busca atuais freqüentemente não satisfazerem a necessidade de informação do usuário sugerem o estudo de novos algoritmos de recuperação de informação. Se por um lado a tarefa de recuperar documentos textuais espalhados pela Web conta com soluções variadas, a tarefa de recuperar imagens é ainda bastante desafiadora.

Existem basicamente duas técnicas distintas de indexação de imagens. A primeira delas diz respeito à indexação baseada em conteúdo. De acordo com essa técnica, características intrínsecas das imagens tais como cor, forma e textura são extraídas através de processamento digital e armazenadas em vetores de características (representações matemáticas do conteúdo visual das imagens). Durante o processamento da consulta, os vetores de características das imagens na coleção são comparados com o vetor de características da imagem alvo informada na consulta, imagem-exemplo que indica o que o usuário deseja receber como resposta à consulta. A similaridade obtida para cada imagem recuperada mede a distância visual entre essa imagem e a imagem alvo.

Esse tipo de técnica possui algumas desvantagens claras. Primeiro, o processamento digital de imagens implica em alto custo computacional. Segundo, o processo de formulação de consultas nesse tipo de técnica não é trivial, já que consultas são especificadas por meio de uma imagem alvo ou mesmo através de um esboço do que seria a imagem desejada. Em especial no caso em que a coleção possui um número muito grande de imagens, esse processamento se torna bastante caro. Apesar dessa técnica conseguir capturar características das imagens que não são possíveis sem o uso de processamento digital, ela não consegue capturar a semântica associada a uma imagem. Pelas razões já citadas, esse tipo de técnica não será estudada ao longo deste artigo.

A segunda técnica, conhecida como técnica baseada em texto, tem sido muito pouco estudada atualmente, sobretudo pela comunidade científica. No entanto, algumas máquinas de busca de imagens comerciais a utilizam, mas ainda de forma não satisfatória. Em geral, acredita-se que uma busca baseada somente nesse tipo de técnica não é suficiente para a obtenção de um bom resultado a uma consulta do usuário. Neste trabalho, investigamos até onde tal premissa é verdadeira e exploramos alternativas para melhorar a qualidade da resposta gerada.

Em uma abordagem simplista, a técnica de indexação de imagens baseada em texto é realizada com base nas anotações feitas por usuários. No entanto, esse processo manual é limitado, particularmente no caso da Web, onde a quantidade de imagens aumenta continuamente. Além disso, esse processo é subjetivo, já que o vocabulário exigido não é padrão, dado que cada usuário tem uma percepção distinta de uma imagem. Ainda assim, essa técnica é bastante utilizada em classificação de imagens em museus, por exemplo, obtendo bons resultados.

De acordo com uma abordagem mais realista, as imagens são indexadas pelas palavras encontradas nos documentos HTML correspondentes às páginas Web de onde foram extraídas. Essa técnica pode ser melhorada quando o processo de recuperação leva em consideração um número maior de evidências a serem extraídas dos documentos HTML e também o local no documento em que os termos ocorrem. Essa é a premissa básica do trabalho aqui apresentado.

Para que um sistema de recuperação atenda às necessidades de informação dos usuários, ele deve empregar bons algoritmos de indexação e ordenação. No caso da recuperação de imagens baseada em texto, acreditamos que o processo não depende exclusivamente da informação associada diretamente à imagem, que é uma informação bastante simples e nem sempre signi-

ficativa. Em especial, esse tipo de evidência se torna inadequada quando imagens são geradas automaticamente por sistemas, devido a questões de leiaute de página, e quando o nome dos arquivos que as contém é aleatório.

Máquinas de busca de imagens existentes atualmente, tais como Ditto [2] e Radix [3], parecem utilizar um número muito pequeno de evidências para melhorar a qualidade das respostas e, dessa forma, não conseguem atender às necessidades de informação dos usuários, uma vez que a precisão na resposta é baixa.

Geralmente essas máquinas de busca indexam a coleção de imagens com as palavras que ocorrem no nome do arquivo que contém cada imagem, no título da página Web de onde a imagem foi extraída, no elemento META do documento HTML associado e também no texto da página. Analisando as respostas geradas às inúmeras consultas submetidas às máquinas de busca Radix e Ditto, por exemplo, pudemos inferir as evidências utilizadas na indexação e, conseqüentemente, concluir que somente elas não são suficientes para atender à necessidade de informação do usuário.

Dessa forma, adotamos neste artigo a estratégia de que a maioria das fontes de evidência disponíveis nos documentos HTML dos quais as imagens são extraídas podem ser consideradas no processo de indexação de uma máquina de busca de imagens e, mais do que isso, podem aumentar a qualidade da resposta gerada para uma consulta do usuário.

O restante deste artigo encontra-se organizado da seguinte forma. A Seção 2 apresenta alguns trabalhos relacionados. A Seção 3 apresenta detalhes da máquina de busca utilizada nos testes e as abordagens de indexação propostas. A Seção 4 apresenta as abordagens combinadas através do arcabouço Bayesiano e a Seção 5 apresenta os resultados obtidos. Por fim, a Seção 6 apresenta nossas conclusões e discute trabalhos futuros.

2 Trabalhos Relacionados

A maioria dos sistemas de recuperação de imagens existentes utiliza a técnica de indexação baseada em conteúdo [4, 9, 11]. Outros sistemas ainda procuram formas de recuperar de maneira mais precisa as imagens de uma coleção, adotando técnicas de classificação de imagens, por exemplo, e mesmo combinando as duas técnicas de indexação descritas [5, 6, 10, 11].

Em [6], os autores apresentam uma abordagem que integra classificação e recuperação de imagens na Web, extraindo informação através de processamento digital de imagens e também por meio de palavras-chave encontradas nos documentos HTML relativos às páginas Web de onde as imagens são extraídas. As imagens são então classificadas em diversas classes (fotografias, gráficos, imagens coloridas) e a busca é realizada de forma distinta em cada classe. No entanto, essa abordagem, além de utilizar indexação baseada em conteúdo, que não é o foco do nosso trabalho, emprega um número muito pequeno de evidências textuais, as quais acreditamos não serem suficientes para se alcançar uma boa resposta à consulta do usuário.

Em [10, 11] é descrito o sistema WebSEEk, desenvolvido no Departamento de Engenharia Elétrica da Universidade de Colúmbia. Além de diretório, ele é uma máquina de busca de imagem e vídeo, suporta consultas por características de cor, leiaute espacial e textura, e também por palavras-chave. Os termos de indexação são extraídos do nome do arquivo que contém a imagem, do texto do atributo ALT a ela relacionado e também do texto encontrado entre as *tags* âncora $< A >$ e $< /A >$. Além do processamento digital de imagens, esse trabalho se difere do trabalho aqui apresentado pelo número de evidências textuais utilizadas na indexação. Por sua vez, em [5] Lu e William apresentam um sistema integrado de recuperação de imagens, que

adota tanto a técnica baseada em conteúdo quanto a técnica baseada em texto. Um usuário pode iniciar o processo de busca através de palavras-chave e, uma vez obtido o conjunto-resposta inicial, selecionar as imagens que mais se assemelham a uma imagem alvo, como se estivesse submetendo uma consulta baseada em conteúdo. A ordenação final das imagens no conjunto resposta leva em consideração as duas técnicas descritas.

A estratégia adotada pelos autores em [5], no caso da indexação baseada em texto, foi a extração dos termos de indexação a partir de oito grupos distintos de termos, definidos sobre áreas diferentes de um documento HTML. Cada grupo diz respeito ao lugar do documento no qual os termos encontram-se inseridos e está associado a um peso distinto, de acordo com a crença dos autores na importância de cada parte do documento.

Para que a similaridade final de cada imagem da coleção com a imagem alvo definida na consulta seja calculada, o sistema combina o resultado gerado pela abordagem baseada em conteúdo com o gerado pela abordagem baseada em texto. Na combinação, pesos distintos podem ser atribuídos a cada abordagem.

No entanto, essa atribuição de pesos se baseia somente na crença dos autores com relação à importância de cada local do documento HTML em que a imagem aparece e às duas abordagens utilizadas. Além disso, apesar dos autores concluírem que a técnica integrada apresenta a melhor solução para os sistemas de recuperação de imagens, o número de consultas utilizadas foi muito baixo (apenas 4 consultas para o caso da abordagem baseada em texto), assim como o tamanho da coleção (apenas 600 imagens, número insignificante se comparado com o tamanho da Web). Dessa forma, acreditamos que os resultados podem ser outros se a coleção e o número de consultas utilizados nos testes forem maiores.

Com relação à ordenação de documentos no conjunto-resposta gerado para uma consulta do usuário, em [8] é apresentada uma maneira de se aplicar o modelo de Redes de Crenças Bayesianas na combinação de diversas fontes de evidência em um único algoritmo de ordenação. Os resultados obtidos mostram que a combinação de várias fontes de evidência para recuperar documentos textuais na Web aumenta a qualidade da resposta gerada. Acreditamos que esse arcabouço pode também, de forma satisfatória, ser aplicado ao modelo de recuperação de imagens na Web. Essa é a premissa fundamental deste trabalho.

3 Solução Proposta

Nessa seção apresentamos uma visão geral da máquina de busca de imagens utilizada como ferramenta para a realização desse trabalho e as abordagens de indexação propostas.

3.1 A Máquina de Busca de Imagens

A arquitetura da máquina de busca de imagens utilizada nos testes encontra-se descrita na Figura 1. Distinguem-se quatro módulos: o coletor, o indexador, o processador de consultas e a interface.

O módulo coletor visita endereços da Web brasileira (domínio .br), coletando documentos HTML e suas imagens e armazenado-os localmente. Somente imagens cuja extensão é .gif, .jpg ou .png são coletadas, justamente por serem as mais comuns na Web. Páginas que não contém imagens não são consideradas.

O módulo indexador, por sua vez, indexa os dados coletados. A indexação ocorre com relação ao par imagem-página, e nunca somente com relação à imagem. Portanto, se uma

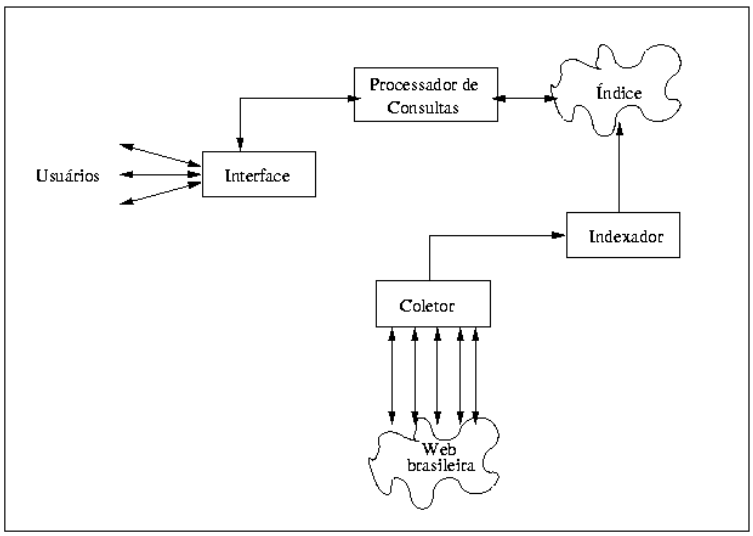

Figura 1: Arquitetura da máquina de busca de imagens.

mesma imagem ocorre em duas ou mais páginas distintas, ela é indexada em pares distintos de imagem-página. No entanto, se uma imagem ocorre mais de uma vez em uma mesma página Web, então ela é computada uma única vez e representada por um único par imagem-página. Imagens que não possuem atributos de altura e largura definidos explicitamente nos documentos HTML que as contém são sempre indexadas. Se esses atributos estão definidos, seus valores são analisados e somente são indexadas as imagens que são maiores do que 45 por 45 *pixels*. Esse valor foi definido depois da realização de testes que comprovaram que imagens menores provocam ruído no resultado das consultas dos usuários. No entanto, pode ser distinto quando outras coleções forem consideradas. Ainda, imagens do tipo papel de parede (que ocorrem na tag $< BODY\ BACKGROUND >$) e imagens que são inseridas via elemento $< INPUT\ type = \text{“}imagem'' >$ não são indexadas (maiores detalhes na Seção 5).

Os termos de indexação são extraídos dos documentos HTML relativos às páginas Web nas quais as imagens estão inseridas, conforme o local no documento em que aparecem. A estrutura de indexação é um arquivo invertido, composto de um vocabulário único (mantido em tabela *hash*) com todos os termos distintos da coleção e de quatro listas invertidas. Cada lista invertida se refere a uma abordagem de indexação proposta, conforme será apresentado na seção seguinte. Para as combinações das diversas fontes de evidência não são necessárias novas listas invertidas.

Essa estrutura de indexação possui o formato apresentado na Figura 2. A entrada relativa a cada termo distinto da coleção é uma quádrupla do tipo $< numPares; (f_{it}, qtd_i : i) >$, onde $numPares$ representa o número de pares imagem-página indexados pelo termo em questão, f_{it} é a freqüência de ocorrência desse termo t no par imagem-página i, qtd_i é a quantidade de pares imagem-página em que o termo ocorre com a freqüência f_{it} e i a lista de pares imagem-página associados a esse termo, com freqüência f_{it}.

O processador de consultas, por sua vez, funciona de forma semelhante ao de um sistema de recuperação de informação textual. Pela interface, o usuário informa ao sistema as palavras através das quais deseja realizar a busca e o sistema verifica se as mesmas são termos de indexação de algum par imagem-página. As imagens recuperadas são apresentadas aos usuários em ordem decrescente de relevância, de acordo com o modelo vetorial.

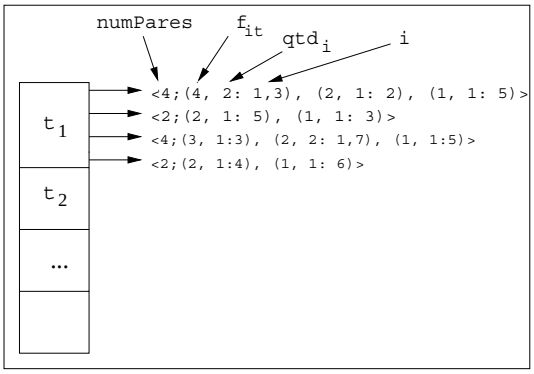

Figura 2: Estrutura do arquivo invertido utilizado na máquina de busca de imagens.

3.2 Abordagens de Indexação Propostas

Para se extrair as evidências dos documentos HTML contendo imagens e efetuar o processo de indexação, foram escolhidos campos desses documentos, que a princípio contém informação relevante para a indexação. Dessa forma, foram determinadas quatro abordagens de evidência única, três das quais foram posteriormente combinadas através do arcabouço Bayesiano, perfazendo um total de 7 abordagens testadas.

A primeira abordagem diz respeito à indexação a partir dos termos que ocorrem no nome relativo do arquivo (nome do arquivo, sem caminho de diretórios) que contém a imagem, no atributo opcional ALT associado a cada imagem e entre as *tags* âncora $< A >$ e $< /A >$. Não foi utilizado o nome absoluto (caminho de diretórios mais nome do arquivo) de cada imagem, porque nem sempre o nome do diretório impõe à imagem algum significado semântico. O atributo ALT, conforme anteriormente mencionado, normalmente contém uma descrição sucinta do conteúdo da imagem. Chamamos essa abordagem de *Indexação por Imagem*.

A segunda abordagem utiliza como termos de indexação para cada imagem de uma página Web as palavras que ocorrem no título da página (entre os elementos $< TITLE >$ e $< /TITLE >$) e no elemento META, quando seu conteúdo são palavras-chave associadas ao conteúdo da página, nome dos autores da página e descrição do conteúdo da mesma. Dessa forma, todas as imagens inseridas em uma determinada página Web são igualmente indexadas. Chamamos essa abordagem de *Indexação por TítuloMeta*.

A terceira abordagem considera como termos de indexação as palavras que compõem a página em que uma imagem ocorre, exceto as que pertencem à estrutura de vínculos da página (*links*). Nesse caso, também todas as imagens são indexadas igualmente. Chamamos essa abordagem de *Indexação por Texto da Página*. Uma outra alternativa é considerar como termos de indexação para um par imagem-página as palavras que ocorrem no parágrafo mais próximo à imagem (chamado aqui de passagem). Cada passagem é formada por no máximo 40 termos: os 20 termos anteriores e os 20 termos posteriores à imagem em questão. Se entre duas imagens não existe termo algum, então a primeira imagem é indexada pelo mesmo conjunto de termos que indexou a segunda imagem. Se entre elas existem 20 termos ou mais, os 20 mais próximos à cada uma são utilizados na indexação. Assim, todas as imagens podem ser recuperadas. Chamamos essa abordagem de *Indexação por Passagem*.

A Figura 3 ilustra um documento HTML e os termos de indexação em cada abordagem.

Testes realizados no sistema comprovaram que a abordagem *Indexação por Passagem* apre-

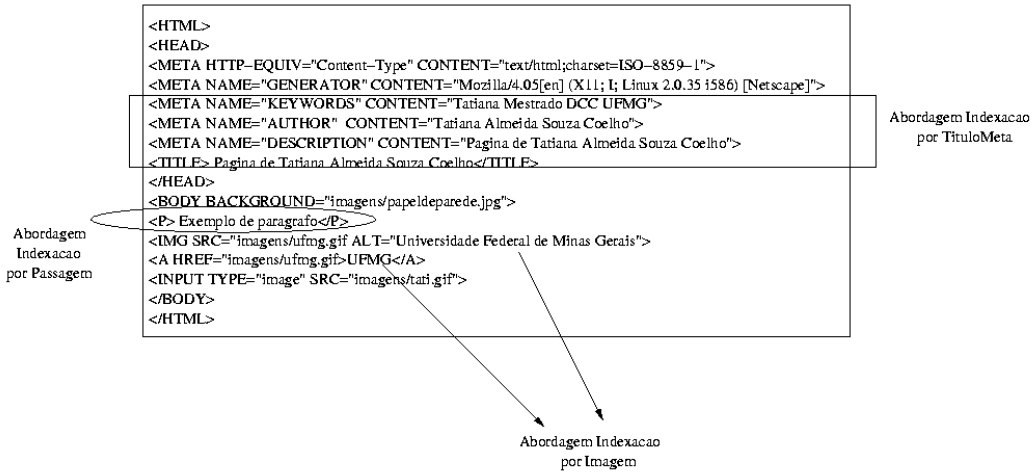

Figura 3: Documento HTML exemplo, com os termos de indexação indicados, conforme cada abordagem proposta.

senta resultado muito superior à *Indexação por Texto da Página*, o que já era esperado, uma vez que normalmente um texto que ocorre mais próximo a uma imagem tem mais significado semântico a ela associado. Assim, neste trabalho consideramos a abordagem *Indexação por Passagem* e desconsideramos a *Indexação por Texto da Página* e, por isso, no final são comparadas 7 abordagens: *Indexação por Imagem*, *Indexação por TítuloMeta*, *Indexação por Passagem* e as quatro combinações possíveis.

4 Ordenação Baseada em Redes de Crenças Bayesianas

De acordo com o modelo de Redes de Crenças, probabilidades são interpretadas como "graus" de crença, sem que experimentações sejam realizadas. Esse modelo adota o arcabouço de Redes Bayesianas, bastante útil por fornecer um formalismo gráfico que modela as dependências entre as variáveis da distribuição. "Ele permite combinar características de modelos distintos em um mesmo esquema representacional" ([7], página 253).

Redes Bayesianas são representadas por meio de grafos acíclicos direcionais, nos quais os nós representam variáveis aleatórias (no nosso caso, termos e imagens da coleção, e consulta do usuário) e arestas indicam o relacionamento existente entre essas variáveis. Probabilidades condicionais associadas aos pares de nós fornecem a "intensidade" de cada relacionamento. Nós-pais de um nó são a causa direta dele, ou seja, se um nó A é nó-pai de um nó B, então a probabilidade do nó B estar ativo depende da probabilidade do nó A estar ativo. Essa dependência é representada por uma seta direcional de A para B. Nós-raiz não são influenciados por nenhum outro nó da rede. No caso de um nó que representa uma imagem, por exemplo, ele está ativo se ela foi indexada, segundo a abordagem específica, com pelo menos uma das palavras que ocorrem na consulta.

De acordo com o modelo de Redes de Crenças, termos de indexação extraídos dos documentos HTML que definem as páginas Web da coleção formam o universo de discurso, denotado por U. Um conceito u é um subconjunto de U e pode representar um par imagem-página da coleção ou uma consulta do usuário. Os nós K_i modelam os termos da coleção e a eles estão associadas variáveis binárias, também chamadas K_i. Se a variável K_i possuir valor 1, então o termo k_i

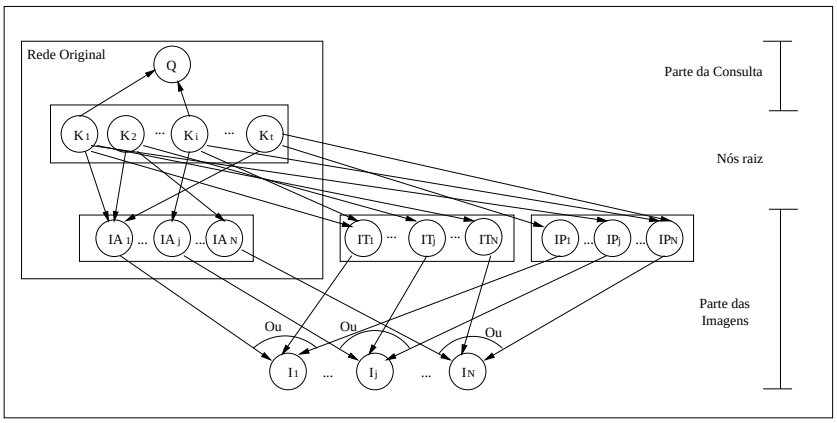

Figura 4: Rede Bayesiana expandida com múltiplas fontes de evidência.

correspondente é um membro do conceito u (está ativo nesse conceito). Essas variáveis são booleanas, porque valores 0 e 1 são expressivos o bastante em termos semânticos e não envolvem complexidade de cálculo.

Pares imagem-página e consulta são mapeados da mesma forma como nós da rede. Assim, um par imagem-página pode ser representado pelo conceito $i_j = \{k_1, k_2, ..., k_j\}$. Se um termo faz parte de um conceito, então ele está ativo nesse conceito. O nó Q representa a consulta do usuário submetida ao sistema e, portanto, estão ativos no conceito relativo à Q as palavras que ocorrem na consulta.

As três abordagens propostas foram então combinadas de acordo com o modelo de Redes de Crenças. A motivação para a utilização desse arcabouço foi o resultado obtido em [8] no âmbito da recuperação de documentos textuais, quando evidências textuais foram combinadas com evidências extraídas da estrutura de vínculos entre documentos. O arcabouço Bayesiano oferece um formalismo que nos permite combinar diversas evidências em uma única modelagem.

A Figura 4 apresenta a modelagem proposta, em que as diversas fontes de evidência são combinadas para a recuperação das imagens. No caso da figura, os termos k_1 e k_j estão ativos. As setas de um termo k_j para um par imagem-página I representa que esse termo foi utilizado na indexação dessa imagem, segundo a abordagem em questão.

Na rede apresentada, cada par imagem-página foi mapeado como um nó distinto. Nós identificados como IA representam pares indexados a partir da abordagem *Indexação por Imagem*. Nós cujo rótulo é IT representam pares indexados conforme a *Indexação por TítuloMeta*. Por sua vez, nós IP dizem respeito aos pares imagem-página indexados a partir da *Indexação por Passagem*. Cada nó I_j na parte das imagens representa um par imagem-página, de acordo com a combinação das múltiplas evidências extraídas. A abordagem *Indexação por Texto da Página* não foi utilizada, por ter tido os piores resultados dentre todas as abordagens testadas, conforme descrito na Seção 5.

Dessa forma, dado que um par imagem-página fará parte do conjunto resposta a uma consulta do usuário se for indexado pelas palavras-chave encontradas na consulta, então o operador *ou* pode ser aplicado à rede, de forma que todo par imagem-página recuperado tenha sido indexado pelo menos por uma das palavras da consulta, de acordo com alguma das abordagens. Assim, $P(I_j \mid k)$, que representa o valor da probabilidade do par imagem-página I_j estar ativo (estar presente no resultado da consulta q) dado o conceito k dos termos da coleção, é dada por:

$$P(I_j \mid k) = 1 - (1 - P(IA_j \mid k)) \times (1 - P(IT_j \mid k)) \times (1 - P(IP_j \mid k)) \quad (1)$$

onde $P(IA_j \mid k)$, $P(IT_j \mid k)$ e $P(IP_j \mid k)$ representam, respectivamente, a probabilidade do par imagem-página indexado estar ativo, conforme as abordagens *Indexação por Imagem*, *Indexação por TítuloMeta* e *Indexação por Passagem*.

No entanto, a probabilidade de um par imagem-página I_j estar ativo, dada a consulta q do usuário, é dada por:

$$P(I_j \mid q) = \eta \times \sum_k [1 - (1 - P(IA_j \mid k)) \times (1 - P(IT_j \mid k)) \times (1 - P(IP_j \mid k))] \times P(q \mid k) \times P(k) \quad (2)$$

onde η é uma constante de normalização, utilizada para ajustar o valor de $P(I_j \mid q)$ entre 0 e 1.

Se os termos ativos no conceito k são exatamente as palavras especificadas na consulta do usuário, então:

$$P(q \mid k) = \begin{cases} 1, & \text{se } \forall_i \; g_i(q) = g_i(k) \\ 0, & \text{caso contrário} \end{cases} \quad (3)$$

onde $g_i(u)$ é uma função que retorna o valor da i-ésima variável no vetor associado ao conceito u, ou seja, se $g_i(q) = g_i(k)$, então o termo k_i aparece na consulta e está ativo na coleção. As probabilidades apriori $P(k)$ são tomadas como constantes porque, antes da consulta do usuário, não preferimos nenhum termo k_i a outro k_l.

A probabilidade de um par imagem-página estar ativo de acordo com qualquer uma das abordagens, dado um conceito k, é dada pelo modelo vetorial:

$$P(IA_j \mid k) = \frac{\sum_{i=1}^{t} w_{ij} \times w_{ik}}{\sqrt{\sum_{i=1}^{t} w_{ij}^2} \times \sqrt{\sum_{i=1}^{t} w_{ik}^2}} \quad (4)$$

onde w_{ij} e w_{ik} representam, respectivamente, o peso do termo k_i com relação ao par imagem-página j e ao conceito k. Esses valores são dados por:

$$w_{ij} = (1 + \ln f_{ij}) \qquad w_{ik} = \ln\left(1 + \frac{N}{n_i}\right) \quad (5)$$

onde f_{ij} é a freqüência de ocorrência do termo i no par imagem-página j, N é o total de pares imagem-página da abordagem considerada e n_i é o número de pares imagem-página associados ao termo i na abordagem considerada.

Se os termos ativos no conceito k são exatamente as palavras da consulta, então podemos escrever a Equação 4 em função de q e chamar a probabilidade $P(IA_j \mid k)$ de RA_{jq}. Analogamente teremos RT_{jq} e RT_{jq}. Assim, podemos calcular os diversos valores de similaridade entre um par imagem-página da coleção e uma consulta do usuário, conforme a topologia da Rede Bayesiana apresentada.

De acordo com a abordagem *Indexação por Imagem*, somente termos associados com o nome do arquivo que contém a imagem, com o texto contido entre as tags âncora e com o atributo ALT são considerados termos de indexação para um par imagem-página. Ou seja, considera-se $P(IT_j \mid q) = 0$ e $P(IP_j \mid q) = 0$. Aplicando esses valores à Equação 2, obtemos:

$$P(I_j \mid q) = \eta \times \sum_k [1 - (1 - P(IA_j \mid k))] \times P(q \mid k) \times P(k) \tag{6}$$

Novamente, se apenas as palavras que ocorrem na consulta estão ativas na rede, então:

$$P(I_j \mid q) = \eta \times RA_{jq} \tag{7}$$

Analogamente podemos calcular $P(I_j \mid q)$ para as outras 6 abordagens propostas, a saber: *Indexação por Imagem, por Passagem, por TítuloMeta*, *por Imagem/Passagem*, *por Imagem/TítuloMeta*, *por Passagem/TítuloMeta*, e *por Imagem/Passagem/TítuloMeta*. Na seção seguinte, apresentamos os dados da coleção utilizada nos testes, as consultas executadas e os testes realizados com nossas 7 abordagens alternativas para recuperação de imagens na Web.

5 Resultados Experimentais

Alguns testes foram realizados para verificar a precisão das 7 abordagens propostas, de acordo com o modelo de Redes de Crença. A Tabela 1 apresenta alguns dados da coleção.

Tamanho da Coleção (GB)	1,8
Tamanho do Índice (MB)	155,4
Número de Páginas	128.712
Número de Imagens Distintas	54.571
Número de Pares Indexados	631.942

Tabela 1: Dados da coleção de imagens testada.

O tamanho da coleção é dado em termos do volume de documentos HTML recuperados. O número de imagens distintas é o número de imagens na coleção que possuem endereços absolutos (URLs) distintos. Se uma mesma imagem aparece em páginas distintas, então ela é considerada imagens distintas. Somente por meio de processamento digital seria possível reconhecer que são a mesma imagem. Por sua vez, o número de pares indexados corresponde ao número de pares imagem-página indexados na coleção, dadas as diversas abordagens propostas.

À medida em que, a cada teste, variamos o tamanho das imagens indexadas (cada vez indexando imagens maiores) e deixamos de indexar imagens do tipo papel de parede ou aquelas inseridas via elemento INPUT do tipo imagem, o tamanho da coleção diminuiu.

Para a realização dos testes foram utilizadas 25 consultas de referência, a saber: pôr do sol, bola de futebol, Marisa Monte, Snoopy, igreja, coca cola, cavalo mangalarga, mapa Brasil, Corcovado, Basset, Serra da Canastra, Edson Arantes do Nascimento, Universidade Federal de Minas Gerais, turma da Mônica, Linux, Jesus, Pirenópolis, Fernando de Noronha, vaso de flores, cerveja Skol, Hotel Glória, fotos carnaval, Carrefour, tubarão e praia Rio de Janeiro.

Para a avaliação das imagens relevantes para cada consulta executada foi utilizada a técnica de *pooling*. De acordo com essa técnica, as imagens retornadas por cada uma de nossas 7 abordagens e para cada consulta do usuário são agrupadas e analisadas manualmente, sem que a pessoa que está avaliando saiba a partir de qual abordagem uma dada imagem foi obtida. Essa técnica permite imparcialidade. Para cada consulta, analisamos somente as 25 imagens mais relevantes retornadas por cada abordagem. Assim, como temos 7 abordagens distintas, o número máximo de respostas para cada consultas é 175. Uma vez definido o conjunto de imagens relevantes para cada consulta executada, o sistema foi avaliado segundo a técnica de precisão/revocação, a mais utilizada na análise de sistemas de recuperação de informação. Maiores detalhes podem ser obtidos em [1].

O primeiro teste foi realizado para investigar dentre as abordagens *Indexação por Passagem* e por *Texto da Página*, qual era aquela que retornava melhor resultado, já que ambas possuem o mesmo enfoque (o texto visível na página Web). Para que o resultado fosse efetivo, variamos o tamanho da Passagem (o número de termos que a define). No primeiro caso, pares imagem-página foram indexados segundo essa abordagem com no máximo 10 termos (os 5 anteriores e os 5 posteriores a cada imagem). No segundo caso, foram utilizados no máximo 20 termos e, por último, no máximo 40 termos. A Tabela 2 apresenta os valores de precisão média encontrados para as abordagens de evidência única, dadas as variações no tamanho da Passagem. Observamos que 40 termos é um bom tamanho para a passagem.

	10 Termos	**20 Termos**	**40 Termos**
Abordagem Indexação por Imagem	0,387792	0,403602	0,399518
Abordagem Indexação por TítuloMeta	0,129251	0,133563	0,133114
Abordagem Indexação por Passagem	0,182951	0,235012	0,258544
Abordagem Indexação por Texto da Página	0,171600	0,167181	0,159184

Tabela 2: Valores de precisão média encontrados nos testes para a obtenção do tamanho da Passagem. Os rótulos das colunas indicam a quantidade de termos utilizados na abordagem *Indexação por Passagem*.

As curvas de precisão/revocação apresentadas na Figura 5 mostram a superioridade da abordagem *Indexação por Passagem* com relação à abordagem por *Texto da Página*. No caso em que a Passagem foi definida com no máximo 10 termos, com um pouco mais de 30% de revocação essa abordagem apresentou uma curva pior do que a abordagem *Indexação por Texto da Página*. Isso se deve ao fato de não se ter eliminado *stopwords* e estas serem palavras muito comuns e sem significado semântico, fazendo com que uma passagem muito pequena contenha muito "ruído". No entanto, foi possível observar, conforme já esperado, que a abordagem *Indexação por Passagem* consegue uma melhor precisão do que a por *Texto da Página*, o que nos levou a desconsiderar essa última abordagem, chegando a um total de 7 abordagens. Ademais, os melhores resultados são obtidos quando se utiliza passagens com 40 termos.

Os testes seguintes foram realizados para se determinar um tamanho mínimo de imagem a ser indexada. Primeiramente, foram indexadas imagens maiores do que 15 por 15 pixels, depois maiores do que 45 por 45 e, finalmente, maiores do que 60 por 60 pixels. Os resultados obtidos mostraram que o melhor resultado é obtido quando as imagens indexadas têm no mínimo 45 por 45 pixels. No entanto, apesar desse ser o melhor valor para a coleção utilizada nos testes, ele poderá ser diferente em outra coleção. Em um segundo momento, foram realizados testes para

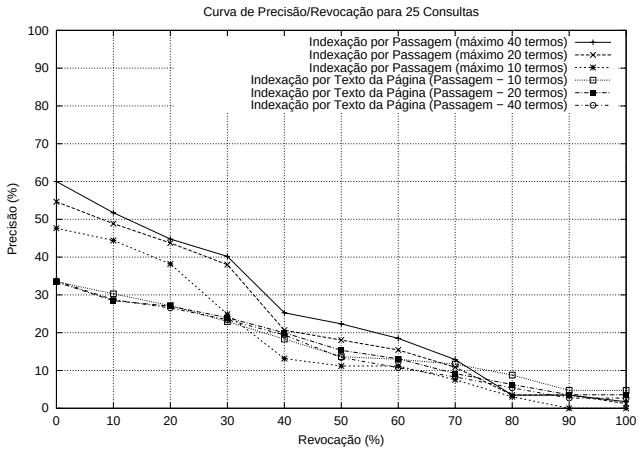

Figura 5: Curva de precisão/revocação para 25 consultas, mostrando os resultados das abordagens *Indexação por Passagem* e por *Texto da Página*, para os tamanhos de Passagem testados.

se determinar se os resultados obtidos às consultas dos usuários são melhores quando imagens do tipo papel de parede e aquelas inseridas através do elemento INPUT são ou não indexadas. Como na coleção utilizada o número de imagens desse tipo era muito pequeno, a precisão obtida não foi muito diferente da obtida quando essas imagens foram desconsideradas. No entanto, o tamanho do índice e da própria coleção diminuiu, o que reduz o custo de processamento das consultas. Adotamos, então, os seguintes parâmetros para os testes realizados utilizando-se o arcabouço Bayesiano: imagens de no mínimo 45 por 45 pixels, desconsiderando imagens do tipo papel de parede e aquelas inseridas através do elemento INPUT do tipo imagem. Além disso, a Passagem foi considerada com no máximo 40 termos.

Para a avaliação da qualidade dos resultados gerados por cada uma de nossas 7 abordagens foram traçadas curvas de precisão/revocação, conforme ilustrado na Figura 4. Os valores respectivos de precisão média global são mostrados na Tabela 3.

	Rede Bayesiana
Abordagem Indexação de Imagem	0,400074
Abordagem Indexação de TítuloMeta	0,181678
Abordagem Indexação de Passagem	0,372204
Abordagem Indexação de Imagem/Passagem	0,601312
Abordagem Indexação de Imagem/TítuloMeta	0,475609
Abordagem Indexação de Passagem/TítuloMeta	0,415470
Abordagem Indexação de Imagem/Passagem/TítuloMeta	0,590051

Tabela 3: Valores de precisão média global encontrados nos testes realizados de acordo com o arcabouço Bayesiano.

É possível perceber que a abordagem *Indexação por TítuloMeta* é a pior das abordagens de fonte de evidência única. Isso ocorre por diversos fatores: primeiro, existem muitos casos em que autores de páginas Web repetem inúmeras vezes uma mesma palavra dentro do campo de palavras-chave no elemento META, no intuito de atribuir à página maior relevância com relação a essa palavra. Segundo, nem sempre as palavras encontradas no elemento META realmente

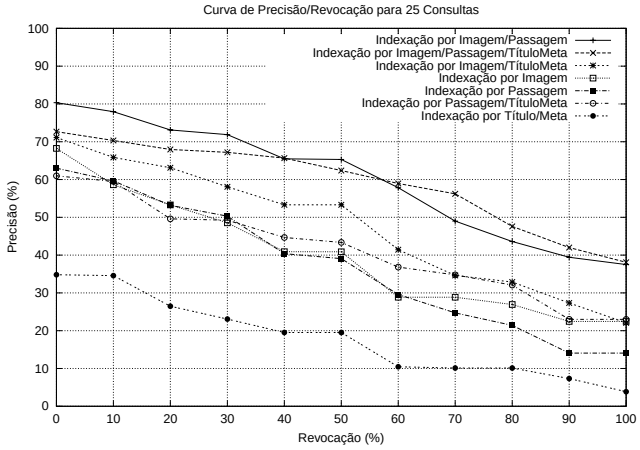

Figura 6: Curva de precisão/revocação média para as 7 abordagens, de acordo com o arcabouço Bayesiano.

dizem respeito ao conteúdo da página Web.

A Figura 7 apresenta as abordagens combinadas de acordo com a Rede Bayesiana. O gráfico mostra que as abordagens *Imagem/TítuloMeta* e *Passagem/TítuloMeta* são inferiores. Ademais, a abordagem *Imagem/Passagem/TítuloMeta* apresenta resultados comparáveis à abordagem *Imagem/Passagem*. Dessa forma, concluímos que a abordagem *Indexação por TítuloMeta* piora a qualidade das respostas, além de aumentar os gastos de memória e processamento.

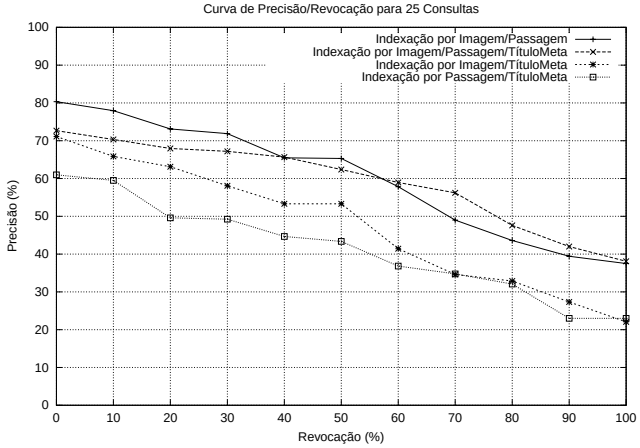

Figura 7: Curva de precisão/revocação para 25 consultas de acordo com a Rede Bayesiana, considerando somente as abordagens que combinam 2 ou 3 outras.

Os resultados encontrados sugerem que a abordagem *Indexação por Imagem/Passagem* pode ser vista como a melhor solução para o problema de recuperação de imagens Web, utilizando o formalismo descrito na seção anterior. No entanto, a abordagem que combina *Indexação por Imagem* e por *Passagem*, a partir dos 60% de revocação, apresenta precisão inferior a abordagem que combina as três fontes de evidência. Isso porque, como pôde ser observado na Figura 6, a partir desse ponto a abordagem *Indexação por Passagem* começa a obter resultados inferiores aos resultados da *Indexação por Imagem*, enquanto que a partir desse mesmo ponto a abordagem *Indexação por TítuloMeta* mantém-se praticamente constante. A *Indexação por*

Passagem apresenta uma piora devido ao número de termos de indexação utilizados.

Por sua vez, a combinação das abordagens *Indexação por Imagem* e *por TítuloMeta* não é uma boa solução, justamente pelo fato da *Indexação por TítuloMeta* não apresentar alta precisão. O mesmo ocorre com a combinação *Indexação por Passagem* e *Indexação por TítuloMeta*.

O gráfico da Figura 8 mostra como a abordagem Imagem/Passagem apresenta resultado muito superior a qualquer uma das abordagens baseadas em fontes de evidência únicas.

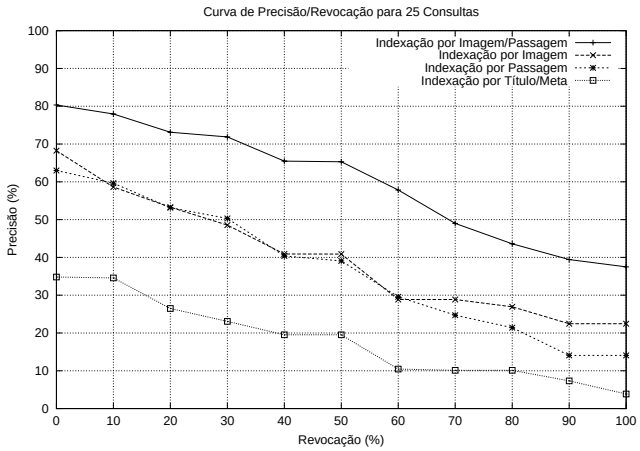

Figura 8: Curva de precisão/revocação para 25 consultas de acordo com a Rede Bayesiana, apresentando as abordagens de fontes de evidência únicas e a abordagem que combina *Indexação por Imagem* e *por Passagem* (melhor resultado de acordo com a rede Bayesiana).

Podemos perceber que a utilização da Rede Bayesiana na recuperação de imagens melhora os resultados das consultas dos usuários comparado aos métodos tradicionais, uma vez que estes utilizam um número menor de evidências e as combinam de forma pouco satisfatória.

6 Conclusões

A crescente quantidade de bancos de dados de imagens disponíveis na Web e o fato das máquinas de busca atuais não satisfazerem a necessidade de informação do usuário sugerem o estudo de novos algoritmos de recuperação de informação. Neste artigo, apresentamos várias alternativas para a indexação das imagens e propomos um arcabouço Bayesiano que permite combiná-las para gerar um *ranking* de qualidade superior. Mostramos que esse arcabouço fornece um mecanismo efetivo de busca e ordenação de imagens na Web. Mostramos também que a combinação de várias fontes de evidência textuais aumenta a qualidade das respostas.

Nossos resultados indicam uma melhora de cerca de 60% de precisão, quando o arcabouço Bayesiano foi utilizado. Evidência proporcionada pelo nome da imagem e as palavras que aparecem no atributo opcional ALT e entre as tags âncora nem sempre é suficiente para determinar a semântica das imagens. Evidência extraída de palavras do título, do nome do autor e de palavras-chave de uma página Web muitas vezes não é suficiente no processo de recuperação de informação, já que nem sempre associa semântica às imagens que nessa página ocorrem. Por outro lado, evidência também pode ser extraída do texto mais próximo a cada imagem.

A combinação das diversas fontes de evidência através de uma Rede Bayesiana, assim como no caso da recuperação de informação textual, nos permite alcançar alta precisão na resposta, o que não é possível se somente uma fonte de evidência for utilizada. A maior contribuição deste

trabalho é certamente a constatação de que múltiplas fontes de evidência, quando combinadas através do Modelo de Redes de Crenças, geram respostas de melhor qualidade e maior precisão.

Dentre os diversos trabalhos futuros, podemos citar: utilização da estrutura de vínculos entre as páginas, classificação das imagens em categorias (ícones, figuras, por exemplo), melhoria do algoritmo que indexa as imagens através de Passagens, realização de testes que utilizem fatores de multiplicação distintos na Rede Bayesiana (para que reflitam a importância de cada abordagem proposta) e também de testes com coleções maiores.

Referências

[1] BAEZA-YATES, R., AND RIBEIRO-NETO, B. *Modern Information Retrieval*, 1st ed. Addison-Wesley, Essex, England, May 1999.

[2] Ditto.com - the Leading Visual Search Engine. `http://www.ditto.com`.

[3] Radix: A Internet na sua Língua. `http://www.radix.com.br`, 1999-2000.

[4] FLICKNER, M., SAWHNEY, H., NIBLACK, W., ASHLEY, J., HUANG, Q., DOM, B., GORKANI, M., HAFNER, J., LEE, D., PETKOVIC, D., STEELE, D., AND YANKER, P. Query by Image and Video Content: The QBIC System. *IEEE Computer Magazine 28*, 9 (September 1995), 23–32.

[5] LU, G., AND WILLIAM, B. An Integrated WWW Image Retrieval System. In *Fifth Australian World Wide Web Conference* (Lismore, Australia, April 1999), `http://ausweb.scu.edu.au/aw99/papers/lu/`.

[6] N. ABBADENI, D. ZIOU, S. W. Image Classification and Retrieval on the World Wide Web. In *Proc. of the Fourth ACM Conference on Digital Libraries* (Berkeley, CA, August 1999), ACM Press, pp. 208–209.

[7] RIBEIRO, B. A. N., AND MUNTZ, R. A Belief Network Model for IR. In *Proc. of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (Zurich, Switzerland, August 1996), pp. 253–260.

[8] SILVA, I., RIBEIRO-NETO, B., CALADO, P., MOURA, E., AND ZIVIANI, N. Link-Based and Content-Based Evidential Information in a Belief Network Model. In *Proc. of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (Athens, Greece, July 2000), ACM Press, pp. 96–103.

[9] SMITH, J. R., AND CHANG, S.-F. VisualSEEk: a Fully Automated Content-based Image Query System. In *Proceedings of the Fourth ACM international Conference on Multimedia* (Boston, Masssachussets, November 1996), ACM Press, pp. 87–98.

[10] SMITH, J. R., AND CHANG, S.-F. An Image and Video Search Engine for the World-Wide Web. In *Symposium on Electronic Imaging: Science and Technology - Storage and Retrieval for Image and Video Databases V* (San Jose, CA, February 1997).

[11] VENTERS, C. C., AND COOPER, D. M. A Review of Content-Based Image Retrieval Systems. `http://www.jtap.ac.uk/reports/htm/jtap-054.html`. University of Manchester.