Implementation and evaluation of a medical image management system with content-based retrieval support*

Implementação e avaliação de um sistema de gerenciamento de imagens médicas com suporte à recuperação baseada em conteúdo

Edilson Carlos Caritá[1], Enzo Seraphim[2], Marcelo Ossamu Honda[3], Paulo Mazzoncini de Azevedo-Marques[4]

Abstract **OBJECTIVE: The present paper describes the implementation and evaluation of a medical images management system with content-based retrieval support (PACS-CBIR) integrating modules focused on images acquisition, storage and distribution, and text retrieval by keyword and images retrieval by similarity. MATERIALS AND METHODS: Internet-compatible technologies were utilized for the system implementation with freeware, and C++, PHP and Java languages on a Linux platform. There is a DICOM-compatible image management module and two query modules, one of them based on text and the other on similarity of image texture attributes. RESULTS: Results demonstrate an appropriate images management and storage, and that the images retrieval time, always < 15 sec, was found to be good by users. The evaluation of retrieval by similarity has demonstrated that the selected images extractor allowed the sorting of images according to anatomical areas. CONCLUSION: Based on these results, one can conclude that the PACS-CBIR implementation is feasible. The system has demonstrated to be DICOM-compatible, and that it can be integrated with the local information system. The similar images retrieval functionality can be enhanced by the introduction of further descriptors.**
Keywords: CBIR; PACS; DICOM; Medical images; Health information system.

Resumo **OBJETIVO: Neste artigo são descritas a implementação e avaliação de um sistema de gerenciamento de imagens médicas com suporte à recuperação baseada em conteúdo (PACS-CBIR), integrando módulos voltados para a aquisição, armazenamento e distribuição de imagens, e a recuperação de informação textual por palavras-chave e de imagens por similaridade. MATERIAIS E MÉTODOS: O sistema foi implementado com tecnologias para Internet, utilizando-se programas livres, plataforma Linux e linguagem de programação C++, PHP e Java. Há um módulo de gerenciamento de imagens compatível com o padrão DICOM e outros dois módulos de busca, um baseado em informações textuais e outro na similaridade de atributos de textura de imagens. RESULTADOS: Os resultados obtidos indicaram que as imagens são gerenciadas e armazenadas corretamente e que o tempo de retorno das imagens, sempre menor do que 15 segundos, foi considerado bom pelos usuários. As avaliações da recuperação por similaridade demonstraram que o extrator escolhido possibilitou a separação das imagens por região anatômica. CONCLUSÃO: Com os resultados obtidos pode-se concluir que é viável a implementação de um PACS-CBIR. O sistema apresentou-se compatível com as funcionalidades do DICOM e integrável ao sistema de informação local. A funcionalidade de recuperação de imagens similares pode ser melhorada com a inclusão de outros descritores.**
Unitermos: CBIR; PACS; DICOM; Imagens médicas; Sistemas de informação em saúde.

Caritá EC, Seraphim E, Honda MO, Azevedo-Marques PM. Implementation and evaluation of a medical image management system with content-based retrieval support. Radiol Bras. 2008;41(5):331–336.

INTRODUCTION

The "picture archiving and communication systems" (PACS) has already become the preferred technological option for transmission, storage and visualization of images in the field of imaging diagnosis. Most recently, because of the increasing number of images produced by the currently available diagnostic imaging methods, the optimization of data storage and retrieval has stood out as a relevant topic in the study of radiological IT solutions. So, while the baseline requirement for a PACS would be the ability of visualizing related clinical data and images, a great interest in techniques utilizing "content-based image retrieval" (CBIR) has emerged in the last years. Based on data extracted from a de-

*Study developed at Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo (FMRPUSP), Ribeirão Preto, SP, Brazil.

1. PhD, Docent at the Center of Exact, Natural and Technological Sciences, Universidade de Ribeirão Preto (Unaerp), Ribeirão Preto, SP, Brazil.

2. PhD, Docent at Instituto de Engenharia de Sistemas e Tecnologias da Informação (IESTI) da Universidade Federal de Itajubá (Unifei), Itajubá, MG, Brazil.

3. PhD, Fellow PhD degree in the Department of Neurology – Sub-area of Biosignal and Imaging Technology, Friedrich-Schiller-Universität, Jena, Germany.

4. PhD, Associate Professor at Center of Imaging Sciences and Medical Physics, Department of Medical Practice, Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo (FMRPUSP), Ribeirão Preto, SP, Brazil.

Mailing address: Dr. Edilson Carlos Caritá. Rua Arnaldo Victaliano, 1404, ap. 34, Jardim Palma Travassos. Ribeirão Preto, SP, Brazil, 14091-22. E-mail: edcarita@uol.com.br

Received August 29, 2007. Accepted after revision February 25, 2008.

termined case, these techniques seek to localize similar cases previously diagnosed and stored in a database. CBIR systems utilize data extracted from images for representing them, with the primary objective of establishing a semantic description for these images[1]. In this type of approach[2–5], the images retrieval is based on the similarity of their attributes which are compared with attributes extracted from a reference image. Color, texture, shape, structures and spatial relationship are attributes most frequently utilized and serve as keys for images retrieval based on a previously established approximation by similarity. Content-based retrieval has shown to be a relevant alternative and an essential complement to the traditional text-based query systems[6–8]. A PACS with content-based images retrieval (PACS-CBIR) consists in a system that provides a connection with images modalities, visualization, storage, images retrieval and reports by means of alphanumeric or textual information and the studies/test retrieval based on queries by similarity of images characteristics.

The present study describes the implementation of an architecture for a medical images management system in a hospital environment integrated with the CBIR function, and the corresponding evaluation covering the functionalities of the modules dedicated to medical images management, content-based images retrieval support and associated images/data retrieval by means of textual data. The images utilized in the present study originate from digital imaging methods such as magnetic resonance imaging (MRI) and computed tomography (CT) in "digital imaging and communication in medicine" (DICOM) format, indexed, stored and linked to the hospital "radiology information system" (RIS)[9], for later visualization with their respective clinical reports. Queries for both images and reports can be performed by means of keywords (textual query) or by means of an images retrieval interface based on pictoric patterns similarity, quantified through texture attributes.

MATERIALS AND METHODS

The PACS-CBIR architecture modeling was based on the information and job stream system of the school-hospital where the present study was developed. Figure 1 illustrates the modeling of this architecture.

For the images server implementation the source codes of PACSOne version 3.1.8 were utilized. This software was developed in 2004 by Xiaohui Li in the United States of America and distributed by the company RainbowFish Software[10]. PACSOne is a DICOM-compliant server for short- and long-term medical images storage for later retrieval, transmission and visualization. The server configurations and images data are stored in a relational database such as the MySQL version 4.1.16 , the "database management system" (DBMS) utilized in the present study, that is a free, open source, portable software compatible with different programming languages[11].

The textual and similarity images retrieval interfaces were developed with network Technologies utilizing "hypertext markup language" (HTML) and "hypertext preprocessor" (PHP) version 4.0.4. For visualization of DICOM-compatible images on the browser, an application was developed utilizing the Java programming language[12], including functions for brightness and contrast adjustment.

The CBIR system was created for extracting the features of images stored in the PACS server, indexing them for later similarity retrieval. The application responsible by the images features was developed in C++ programming language and is executed through a command line. Upon initialization, the application searches for the studies/tests stored in the PACS server, extracts the images features utilizing texture descriptors developed by Haralick et al.[13] and stores them as feature vectors in the database of the CBIR module. The metric structure on which the images features indexation is based is the slim-tree proposed by Traina Jr et al.[14].

Performance, stability, main functionalities and DICOM-compatibility were taken into consideration in the evaluation of the implemented system. A total of 120 studies including 60,401 MRI and CT images were utilized for this evaluation.

The textual retrieval system underwent two evaluation processes: a quantitative evaluation aimed at validating and analyzing the DICOM functionalities for guaranteeing the availability of all the images stored in the PACS server and linked to the RIS at the moment of retrieval in conjunction with the demographic data of the patient and the clinical study data, besides the measurement of the response time for each study; and a qualitative evaluation considering the user satisfaction in relation to the system. Both evaluations were performed by four physicians and a computation scientist. Chart 1 presents the observer's education and the experience in years. The system interface response time for studies, response time for images, images quality, browsing facility, and image settings were

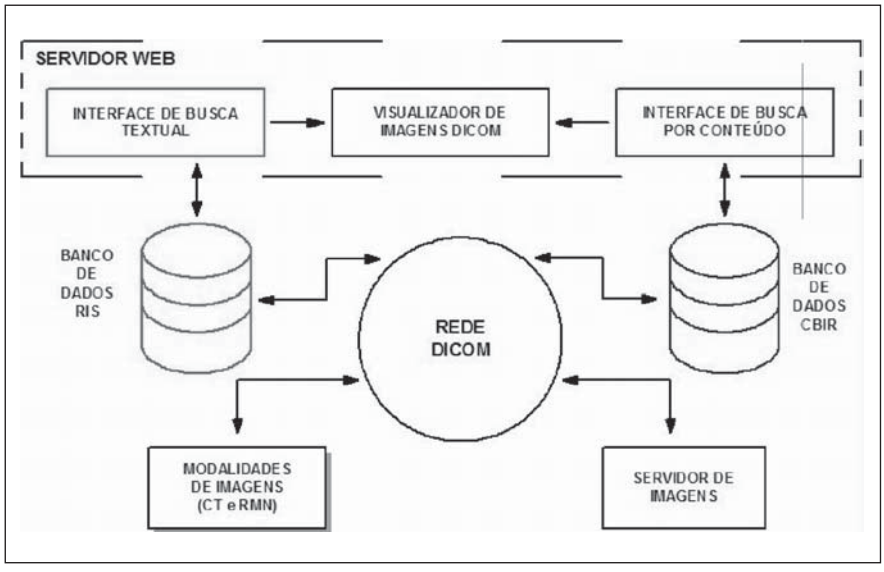

Figure 1. Architecture for a PACS-CBIR system.

Chart 1 Education and experience of observers involved in the system evaluation.

Identifier	Education	Experience
M1	MD, radiologist	15 years
M2	MD, radiologist	13 years
M3	MD, neuroradiologist	19 years
M4	MD, neuroradiologist	7 years
C1	Computation scientist	12 years

evaluated. Quality levels considered in the present evaluation were the following: 1 (very poor quality — inappropriate performance); 2 (poor quality — insufficient performance); 3 (satisfactory quality — it can be utilized although with some significant limitation); 4 (good quality — it can be utilized, although with some non-significant limitation); 5 (very good quality — appropriate performance).

Only MRI studies were considered in the evaluation of the system of query by similarity, including 61 cases and 17,099 images (39 cases with 12,939 brain images; 8 cases with 1,456 abdominal images; 10 cases with 2,448 spine images; and 4 cases with 256 pharynx images). Initially, the user selected a imaging diagnosis modality, defined a reference image and informed the number of similar images which should be retrieved by the system (Figure 2). Following the process of images retrieval, an image was considered as correct provided it corresponded to the same region of interest or diagnostic condition (normal or abnormal) of the reference image. For studies/tests with abnormal reports, the diagnostic descriptions compatibility was evaluated.

RESULTS

The system for textual retrieval has been implemented to allow the integration between the imaging server, the RIS server and the images viewer, assuming that the authorized user enters a query for patients' data, clinical studies/tests and respective images. Chart 2 presents the textual query options available in the system. Figure 3 presents

Chart 2 System options for textual query.

Options	Description
Name/Surname	The user should inform the patient's name or surname enabling the system to find his/her records and subsequently retrieving his/her clinical data and respective images
Region of interest/clinical test	The user will enter the region of interest and/or type of clinical study/test and a period of time, and the system will retrieve the clinical reports and respective images
Patient's login	The user will enter the patient's login and the system will retrieve the patient's data, as well as his/her clinical reports and respective images through a single query
Clinical report description	The user should inform keywords of the clinical report conclusion and a period of time, and the system will retrieve the clinical studies/tests and respective images
Complex query	Through this option, the user can mix attributes such as name, surname, region of interest, type of study and clinical report description. The system will retrieve the clinical study(ies)/test(s) with the characteristics informed

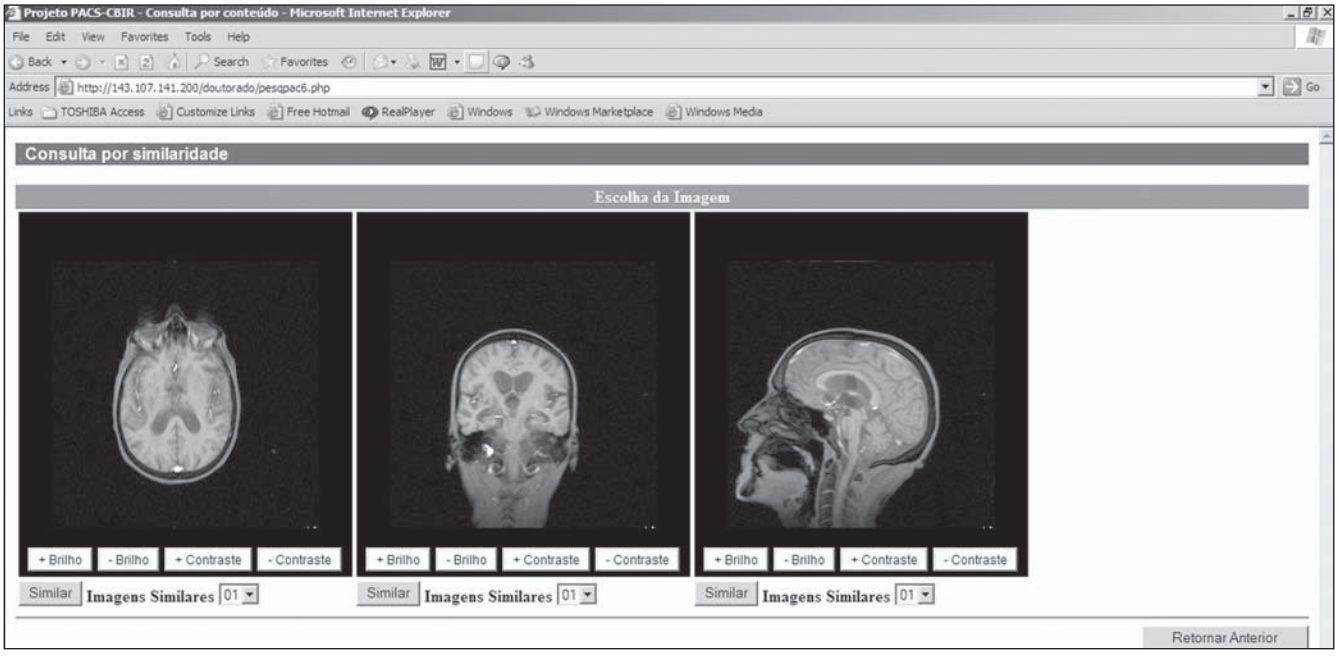

Figure 2. Interface for selection of a reference image for retrieval by similarity.

the screen showing the complex query option. Based on the patient's data and the list of studies/tests performed, one can access and visualize the images (Figure 4).

Results of the qualitative evaluation of the textual retrieval system are shown on Table 1. Table 2 shows the results of queries performed to evaluate the similarity between images retrieved through texture descriptors, taking the region of interest into consideration. The column "Average" presents mean percentages of hits in queries for images for each region of interest. Overall, the mean rate of hits reached 72% with 0.84 standards deviation.

Table 3 presents results validating the clinical report conclusion (normal or abnormal) and the compatibility of the clinical

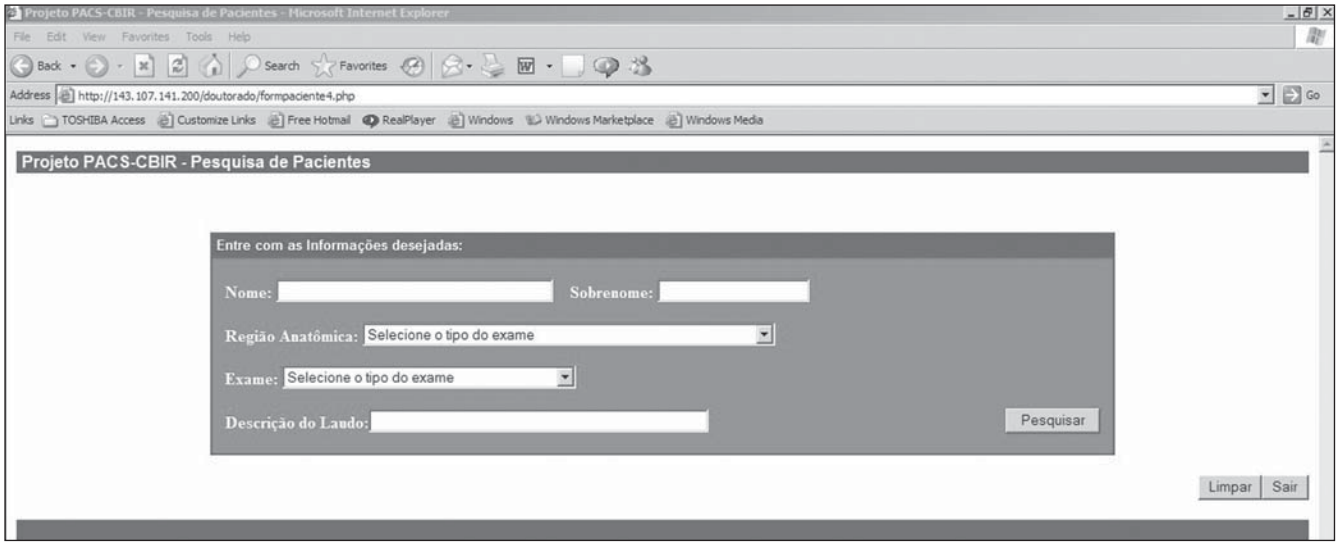

Figure 3. Screen showing the option "complex query" of the textual query module.

Figure 4. Study images shown by the implemented DICOM viewer.

Table 1 Results of qualitative evaluation of the textual retrieval system.

Item	M1	M2	M3	M4	C1	Mean	Final conclusion
System interface	4	5	5	5	5	4.8	Very good
Response time – studies/tests	3	5	5	5	5	4.6	Very good
Response time – images	3	5	4	5	4	4.2	Good
Visualization quality	4	4	4	4	4	4	Good
Browsing facility	4	5	5	5	5	4.8	Very good
Images settings	3	3	3	3	3	3	Satisfactory

M1 and M2: MDs, radiologists; M3 and M4: MDs, neuroradiologists; C1: computation scientist.

Table 2 Results according to regions of interest.

Region of interest	Average	Standard deviation
Brain	80%	1.41
Abdomen	100%	0
Spine	20%	0
Pharynx	80%	1.41

Table 3 Results considering the clinical reports conclusion and the compatibility of clinical reports description.

Region of interest	Average (report)	Standard deviation	Average (compatibility)	Standard deviation
Brain	72%	1.14	56%	0.70
Abdomen	80%	0	50%	0
Spine	100%	0	100%	0
Pharynx	70%	2.12	50%	2.12

reports description for abnormal cases, for each region of interest. Overall, the mean rate of hits reached 81% with 1.47 standards deviation for the clinical report conclusion, and 64% with 1.13 standard deviation for compatibility of clinical reports description.

DISCUSSION

In the qualitative evaluation of the textual retrieval system, one can observe that four observers attributed grade 5 and one, grade 4 for "system interface"; so this item can be considered as very good. In the implementation of the PACS-CBIR project interfaces (PACS interface and textual retrieval and content-based retrieval interfaces) the global strategy involving the development of Internet-based systems for a safe, practical and fast provision of information was adopted. As a result, the users do not need to have the system installed in their own computers, considering that it is easily accessible from any place. Four observers attributed grade 5 and one, grade 3 for "response time for studies/tests"; most of them considered this response time as very good. It is important to note that this time has not exceeded five seconds, but variations in the transfer rate may occur as a function of the traffic in the local area network of the hospital.

In the evaluation of "response time for images", two observers attributes grade 5, two grade 4, and one, grade 3, so this item may be considered as good; however, the mean response time has always ranged between 10 and 15 seconds. Although this aspect has not been quantified in the present study, it is likely that the highest waiting time is associated to the process of image loading and displaying on the screen rather than to the process of query through the CBIR module, in compliance with the results reported by Harrison[15], who has been responsible for the implantation of a PACS-Web in the University of Mississippi Medical Center in Jackson, MS, USA. According to the author, the query is fast, and what is computationally expensive is retrieving the image from the database and loading and displaying the imaging on the screen.

In the topic "visualization quality", grade 4 was given by all the observers, corresponding to good, and the main limitation being associated to the windowing for images displaying. This limitation is confirmed in the evaluation of the topic "images settings" that was given grade 3 by all the observers, with the suggestion of adding some tools such as zoom, segmentation and binarization.

As regards "browsing facility", four observers attributed grade 5, and one, grade 4. The majority of the observers considered this topic as very good, a result compatible with other studies approaching the utilization of Internet technologies in PACS. According to Harrison[15], Peer et al.[16] and Cao et al.[17], the main advantages of the utilization of these technologies are costs reduction and studies/tests decentralization, allowing the radiologists to easily access images through hospital local network or Internet.

The results regarding the performance of the system for images retrieval by similarity utilizing texture attributes, with an overall mean rate of hits reaching 72% are compatible with results found by other studies in the literature[1,18,19]. Pereira Jr et al.[18] have developed a study demonstrating that texture attributes can be useful in the automatic differentiation between normal regions and regions with nodules or microcalcifications on digitized mammographic images, with more than 90% of hits. However, these attributes sensitivity for differentiating malignant from benign lesions decreases to as low as 50%. In a study involving the analysis of texture for images retrieval by similarity, Oliveira et al.[19] have found results for accuracy of about 54% for sagittal images of knee, and 40% for axial images of head. Kinoshita et al.[1] have presented results ranging between 78% and 83% of hits in a CBIR system based on texture and artificial neural networks developed to retrieve mammographic images by tissue density similarity.

CONCLUSIONS

The architecture implemented was based on non-proprietary, open-source

Technologies and the features of the information system and the flow of studies/tests in the hospital where the present project was implemented. However, its structure can be adapted for operating in conjunction with any DICOM-compatible management system and imaging diagnosis modality, basically by the addition of a content-based images retrieval module. The implementation of the similarity retrieval algorithm required texture descriptors which provide measurements of regions properties such as mildness, roughness and smoothness[20]. However, the system performance can be enhanced with the inclusion of further descriptors dedicated to the characterization of shape, for example.

As far as its application is concerned, the system implemented can be utilized in the clinical routine as tool for aiding in the decision-making process, as described in the study developed by Huang et al.[21], Who have described the implementation and evaluation of a diagnosis support tool based on similar images retrieval in the Children Hospital Los Angeles. Their study has involved 2500 consecutive MRI studies of children's brains, with results ranging between 10% and 60% of hits in queries by similarity, respectively for arachnoid cyst and neurofibromatosis. It is evident that the utilization of a CBIR system as an ancillary diagnostic tool will negatively affect the work flow and, consequently, the unit productivity. But the benefits of this system as a tool for supporting Radiology teaching and research are quite evident, as described in a review article published by Müller et al.[8]. Finally, it is important to note that there are still few reports in the literature approaching the integration of content-based retrieval tools into images management systems, reinforcing the innovative character of the present study.

REFERENCES

1. Kinoshita SK, Azevedo-Marques PM, Pereira RR Jr, et al. Content-based retrieval of mammograms using visual features related to breast density patterns. J Digit Imaging. 2007;20:172–90.
2. Song BC, Kim MJ, Ra JB. A fast multiresolution feature matching algorithm for exhaustive search in large image databases. IEEE Trans Circuits Syst Video Technol. 2001;11:673–8.
3. Cai W, Feng DD, Fulton R. Content-based retrieval of dynamic PET functional images. IEEE Trans Inf Technol Biomed. 2000;4:152–8.
4. Assfalg J, Pala P. Querying by photographs: a VR metaphor for image retrieval. IEEE Multimedia. 2000;7:52–9.
5. Wang Z, Chi Z, Feng D. Content-based image retrieval using block-constrained fractal coding and nona-tree decomposition. IEEE Proc Vis Image Signal Process. 2000;147:9–15.
6. Alto H, Rangayyan RM, Desautels JEL. Content-based retrieval and analysis of mammographic masses. J Electron Imaging. 2005;14:1–17.
7. Vailaya A, Figueiredo MAT, Jain AK, et al. Image classification for content-based indexing. IEEE Trans Image Processing. 2001;10:117–30.
8. Müller H, Michoux N, Bandon D, et al. A review of content-based image retrieval systems in medical applications – clinical benefits and future directions. Int J Med Inform. 2004;73:1–23.
9. Azevedo-Marques PM, Caritá EC, Benedicto AA, et al. Integração RIS/PACS no Hospital das Clínicas de Ribeirão Preto: uma solução baseada em "web". Radiol Bras. 2005;38:37–43.
10. PacsOne. [cited 2007 Jun 10]. Available from: http://www.pacsone.net
11. Milani A. MySQL – guia do programador. São Paulo: Novatec; 2006.
12. Caritá EC, Matos ALM, Azevedo-Marques PM. Ferramentas para visualização de imagens médicas em hospital universitário. Radiol Bras. 2004;37:437–40.
13. Haralick RM, Shanmugan K, Dinstein I. Textural features of images classification. IEEE Trans Systems Man Cybernetics. 1973;3:610–21.
14. Traina C Jr, Traina AJM, Faloutsos C, et al. Fast indexing and visualization of metric data sets using slim-trees. IEEE Trans Knowledge Data Engineering. 2002;14:244–60.
15. Harrison SW. Success with Web-based image access. Radiol Manage. 2003;25:36–8.
16. Peer S, Vogl R, Peer R, et al. Sophisticated hospital information system/radiology information system/picture archiving and communications system (PACS) integration in a large-scale traumatology PACS. J Digit Imaging. 1999;12:99–102.
17. Cao F, Huang HK, Zhou XQ. Medical image security in a HIPAA mandated PACS environment. Comput Med Imaging Graph. 2003;27:185–96.
18. Pereira RR Jr, Azevedo-Marques PM, Honda MO, et al. Usefulness of texture analysis for computerized classification of breast lesions on mammograms. J Digit Imaging. 2007;20:248–55.
19. Oliveira MC, Azevedo-Marques PM, Cirne Filho WC. Grades computacionais na otimização da recuperação de imagens médicas baseada em conteúdo. Radiol Bras. 2007;40:255–61.
20. Gonzalez RC, Woods RE. Processamento de imagens digitais. São Paulo: Edgard Blücher; 2000.
21. Huang HK, Nielsen JF, Nelson MD, et al. Image-matching as a medical diagnostic support tool (DST) for brain diseases in children. Comput Med Imaging Graph. 2005;29:195–202.