

Daniel Zielas, Adam Pelikant
Wydział Informatyki i Zarządzania
Wyższa Szkoła Informatyki w Łodzi
e-mail: dzielas@wsinf.edu.pl; apelikan@wsinf.edu.pl

PRZETWARZANIE I ANALIZA DANYCH MEDYCZNYCH NA PLATFORMIE ORACLE Z ZASTOSOWANIEM DICOM

Streszczenie – Artykuł zawiera prezentację metod pozwalających na zarządzanie danymi medycznymi z poziomu centralnego serwera bazy danych. Jako nośnik wykorzystany został format DICOM, który jest standardem wymiany informacji w obrazowych urządzeniach medycznych – skanerach medycznych. Przedstawiono metody przechowywania oraz ekstrakcji metadanych na platformie serwera danych ORACLE, wykorzystując wbudowane typy obiektowe oraz biblioteki API. Poza bieżącym przetwarzaniem zbudowano bank filtrów graficznych pozwalających na obróbkę informacji wizualnej, w celu poprawy jakości oraz przede wszystkim ekstrakcję cech.

1 Wprowadzenie

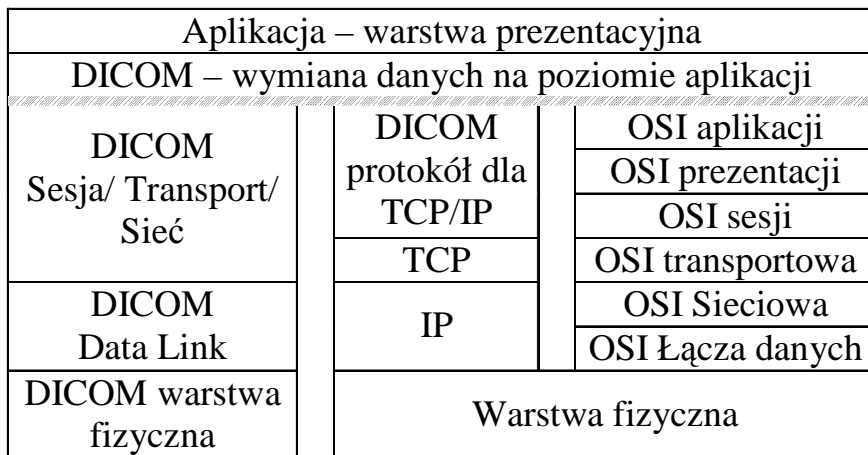
Standard DICOM (Digital Imaging Communications in Medicine) opracowany został przez ACR/NEMA (American College Radiology/ National Electrical Manufactures Association) dla potrzeb interpretacji i wymiany danych medycznych reprezentujących lub związanych obrazy diagnostyczne w medycynie. Tworzony i dopracowywany od roku 1983 stał się najbardziej uniwersalnym i podstawowym standardem w medycynie [1]. DICOM jest wykorzystywany przez większość systemów typu Picture Archiving and Communication Systems (PACS) [3], [4]. Dane w formacie DICOM mają przeważnie dużą objętość i wymagają specjalnego oprogramowania [5], [6] i łącz wysokiej przepustowości, dzięki czemu pozwalają zachować wysoką jakość przechowywanej informacji [8], [9]. Pliki DICOM mogą być wymieniane pomiędzy dwoma podmiotami, które są zdolne do odbierania danych i obrazów pacjenta w tym formacie. Prawa autorskie do tego standardu posiada National Electrical Manufactures Association. DICOM został opracowany przez komitet standardu DICOM, którego członkowie należą do NEMA.

Format DICOM stara się odwzorować obiekty świata realnego, takie jak: urządzenia, pacjenci i wiele innych na obiekty wirtualne z odpowied-

nimi parametrami i atrybutami. Ich definicje opisane są w DICOM Information Object Definitions (IOD's). IOD's pacjenta zawiera jego płeć, wiek, wagę. Parametrów jest wystarczająco dużo, aby opisać wszystkie istotne medyczne cechy pacjenta. Wszystkie standardowe atrybuty obiektów zebrane są w DICOM Data Dictionary i jest ich około 2000. Atrybuty te są podzielone na około 27 grup w zależności od rodzaju formatowania danych.

2 Struktura i przetwarzanie DICOM

DICOM jest nie tylko sposobem zapisu danych medycznych. Pod tą samą nazwą ukrywa się również protokół komunikacyjny pozwalający na wymianę danych między urządzeniami ten protokół obsługującymi oraz przy zastosowaniu serwisu nasłuchu utworzonego za pomocą języka wyższego rzędu może zapewnić komunikację z serwerem bazy danych (rys. 1). Pozwala to na stworzenie centralnego repozytorium danych medycznych dla całej jednostki, umożliwiając dostęp do danych ze wszystkich urządzeń obrazowych oraz ich analizę poza miejscem wykonywania badań. Takie repozytorium pozwala również na śledzenie zmian w kolejnych wykonywanych badaniach, co daje możliwość pogłębionej analizy procesu leczenia.



***Wymiana między
urządzeniami***

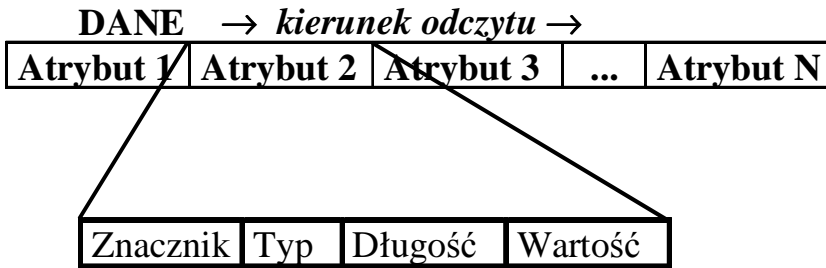
Wymiana sieciowa

Rys. 1. Model komunikacji protokołu DICOM [1], [4], [5]

Standard DICOM ułatwia tworzenie oraz rozbudowę systemów archiwizacji obrazów PACS (ang. *Picture Archiving and Comunication*

Systems) oraz wymianę informacji z innymi systemami informatycznymi obecnymi w medycynie (ang. *Hospital Information System – HIS*).

Format dokumentów zapisanych w tym standardzie jest prosty i składa się z szeregowo zapisanego zestawu atrybutów i ich wartości. Formalnie można go sklasyfikować, jako dokument znacznikowy. Schemat budowy takiego pliku prezentuje rys. 2.



Rys. 2. Budowa ciągu informacji w pliku DICOM [2], [3]

Plik zaczyna się od pustej preambuły (128 bajtów), po której następuje identyfikator w postaci liter DICM, zapisany w kodzie ASCII, który rozpoczyna strumień informacji. Widoczny na rys. 2 ciąg danych składa się z kilku atrybutów [1], [2]. W skład pojedynczego atrybutu wchodzi:

- ♦ Znacznik – 8-cyfrowy wartość określająca zawartości elementu (w notacji szesnastkowej)
- ♦ Typ – reprezentacja wartości wg standardu użytego w pliku DICOM
- ♦ Długość – Liczba znaków, na której zapisano wartość (w systemie szesnastkowym)
- ♦ Wartość– dla typów prostych wartości lub binarna reprezentacja

Tabela. 1. Zawartość pojedynczego przykładowego atrybutu pliku DICOM [2]

Znacznik	Typ	Długość	Wartość
4 Bajty	2 Bajty	2 Bajty	12 Bajtów
0010 0010	PN	0C	Jan Kowalski

W przykładzie na rysunku 3 znacznikiem jest ciąg cyfr 0010 0010. Pierwsze cztery cyfry (czyli 0010) oznaczają, że są to informacje nt. pacjenta. Kolejne cztery dają już pełen numer znacznika, w tym wypadku imię i nazwisko pacjenta. Szczegółowy opis wszystkich

znaczników można znaleźć w dokumentacji standardu DICOM w rozdziale 6 „Data Dictionary”. Następną informacją jest reprezentacja wartości, w przykładowym ciągu jest to PN, czyli Patient Name – zgodnie ze standardem. Kolejną informacją jest długość ciągu wartości, w przykładzie jest to 0C (w systemie szesnastkowym), czyli 12 (w systemie dziesiętnym). Ostatnią częścią jest pole wartości zawierające już konkretną informację. Znacznym utrudnieniem analizy plików DICOM jest fakt, że wiele nawet nowoczesnych urządzeń nie w pełni implementuje ten standard. Nawet dwa skanery tego samego producenta potrafią generować pliki różniące się składnią. Ponieważ jest to jednak typ znacznikowy trudności te daje się przezwyciężyć stosując zaawansowane metody analizy formalnej.

3 Przetwarzanie DICOM po stronie serwera ORACLE

Architektura obiektów zarządzających formatem DICOM składa się z dwóch warstw, pierwsza to warstwa bazy danych, a druga to warstwa klienta, co przedstawią rys. 3

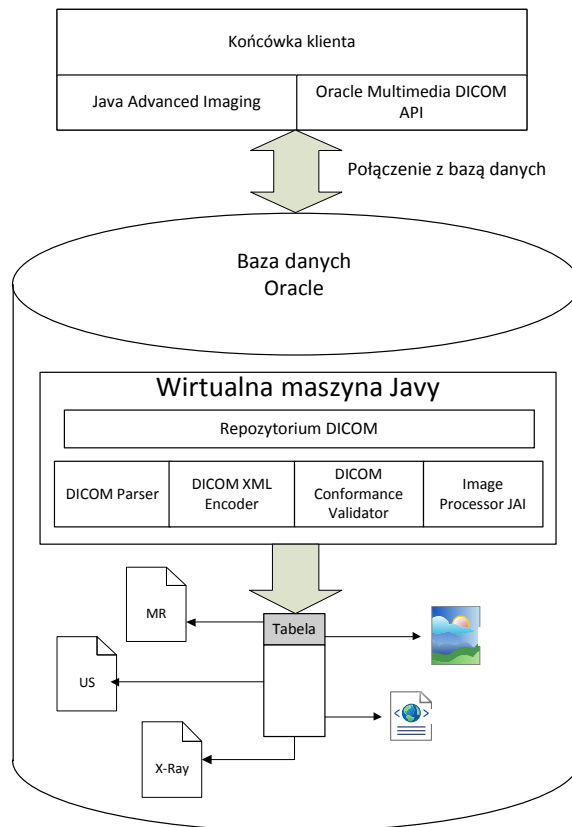
W warstwie bazy danych zawartość DICOM jest przetrzymywana w tabelach. W tabeli jedna kolumna przetrzymuje pełne obrazy, kolejna kolumna przechowuje miniatury tych obrazów, jeszcze inna kolumna przechowuje metadane w formacie XML powiązane z obrazami. W serwer wbudowana jest wirtualna maszyna Javy, którą wykorzystują następujące moduły:

- ◆ DICOM parser
- ◆ DICOM XML encoder
- ◆ DICOM conformance validator
- ◆ Image processor

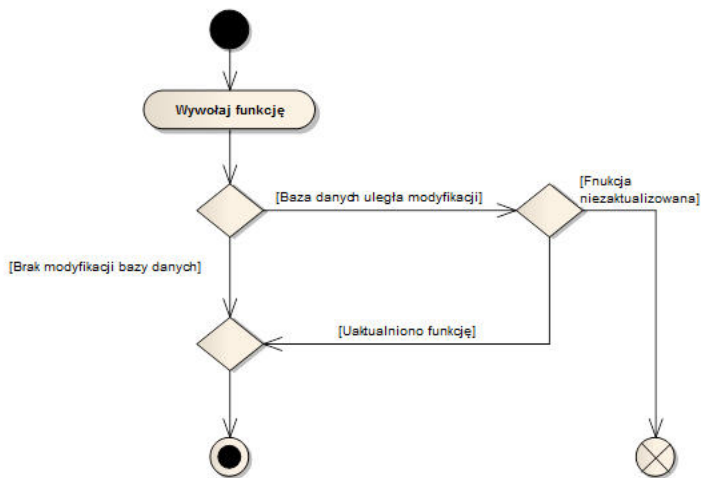
Dicom parser ekstrahuje metadane z oryginalnego pliku. DICOM XML encoder czyta atrybuty z metadanych i zapisuje je do formatu XML. DICOM conformance validator sprawdza poprawność tych danych. Image Processor zawiera Java Advanced Imaging, umożliwiła przeprowadzanie operacji na obrazach np. stworzenie miniatury czy konwersja między różnymi formatami.

Przetwarzanie informacji po stronie serwera bazy danych odbywa się przez zastosowanie funkcji i procedur PL/SQL wykorzystujących ORACLE DICOM API, które zostały zgrupowane w postaci przygotowanego pakietu, którego nagłówek ma następującą postać.

```
CREATE OR REPLACE PACKAGE DICOM
IS
FUNCTION addDicomEmptyRow RETURN INTEGER;
PROCEDURE fillDicomData(idNo INTEGER);
PROCEDURE deleteDicomRow(idNo INTEGER);
END;
```



Rys. 3. Widok architektury DICOM Oracle Multimedia

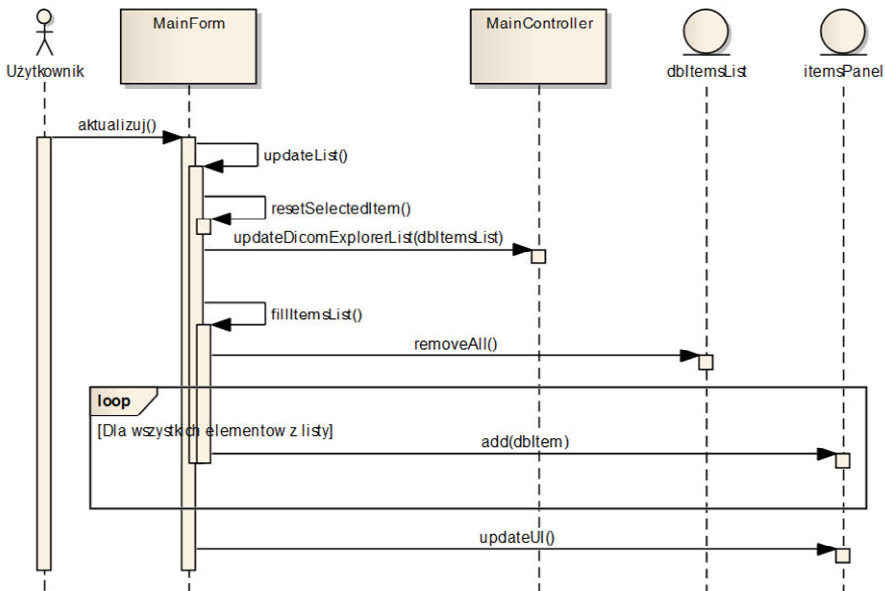


Rys. 4. Wywołanie funkcji składowanej w bazie danych

Jak można zauważyć dopisywanie nowego rekordu jest rozbite na dwa etapy tworzenie pustego wiersza oraz zasilanie go danymi z pliku zawartego w folderze. Każda z funkcji pakietu może zostać wykonana z poziomu końcówki klienta napisanej w języku wyższego rzędu, a jej kod wykona się po stronie serwera bazy danych, według schematu przedstawionego na rys. 4.

4 Budowa aplikacji zarządzającej

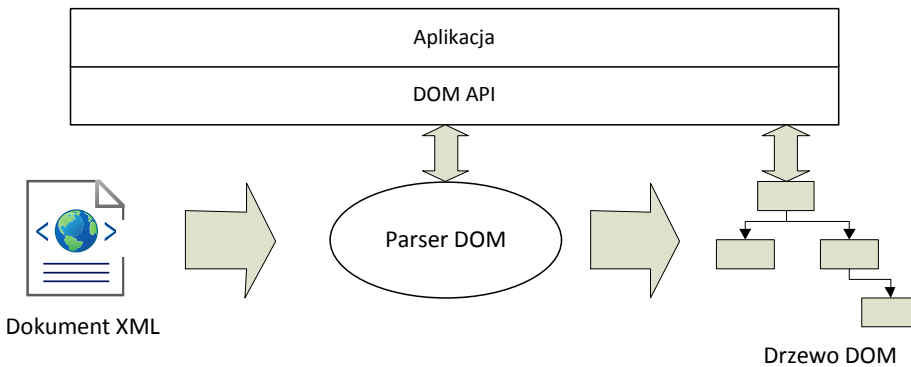
Aby udostępnić operatorowi zasoby dokumentacji medycznej zgromadzone na serwerze ORACLE, została zbudowana końcówka klienta z zastosowaniem języka Java. Diagram tak opracowanej aplikacji przedstawia rys. 5.



Rys. 5. Diagram klas dla końcówki klienta Java

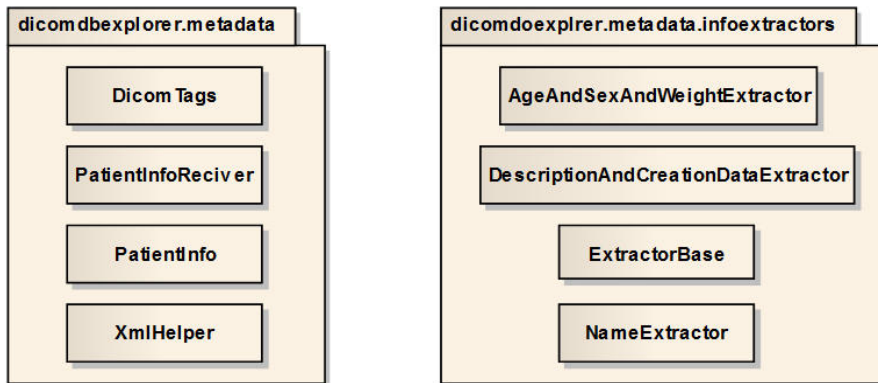
W aplikacji do parsowania metadanych wykorzystana została metoda DOM. Parser dokonuje przekształcenia dokumentu XML w drzewo rys. 6. API parsera umożliwia nam dostęp do funkcji.

Model DOM odtwarza w pamięci strukturę dokumentu, co zapewnia dużą elastyczność oraz udogodnienia przy odczycie, nawigacji i modyfikowaniu dokumentu. Jego wadą są duże wymagania odnośnie zużycia pamięci operacyjnej.



Rys. 6. Przekształcenie danych XML w drzewo DOM

Format DICOM zawiera informacje o pacjentach w postaci meta danych, a klasy je przetwarzające znajdują się w dwóch pakietach, co przedstawia rys.7. Metadane pobierane są z bazy danych w formacie XML, a następnie parsowane w celu wyekstrahowania z nich potrzebnych informacji.



Rys. 7. Pakiety z klasami przetwarzające metadane

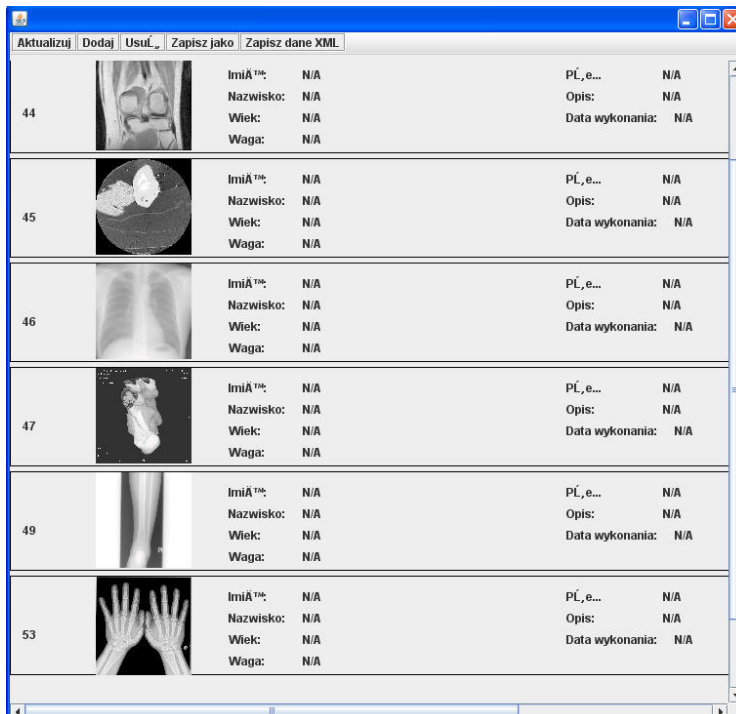
Klasa `DicomTags` zawiera klucze, którymi oznaczone są poszczególne informacje w pliku z metadanymi. Przykładowo klucz z wiekiem pacjenta ma wartość 00101010, a płeć 00100040. Klucze składają się z cyfr, lecz nie są liczbami, posiadają format tekstowy. Każdy z nich składa się dokładnie z ośmiu znaków. Klasa `PatientInfoReciver` służy do tworzenia obiektów oraz wywoływania odpowiednich metod z pakietu `dicombexplorer.metadata.infoextractors`.

Obiekty klasy PatientInfo służą wyłącznie do przechowywania informacji o zdjęciu i o pacjencie:

- ◆ imię
- ◆ nazwisko
- ◆ wiek
- ◆ waga
- ◆ płeć
- ◆ opis
- ◆ data utworzenia pliku DICOM

Klasa XmlHelper posiada metody pomocnicze przy wydobywaniu danych. Klasy z pakietu dicomdbexplorer.metadata.infoextractors służą do pobierania konkretnych informacji, np. NameExtractor wyszukuje w metadanych znacznik z wartością „00100010”, którą oznaczone jest imię i nazwisko pacjenta.

Wizualnie końcówka posiada pięć zakładek: pierwsza odświeża dane zapisane w bazie, kolejne dwie operują na danych w bazie danych dodając lub usuwając rekord, dwie ostatnie pozwalają na eksport do pliku graficznego lub pliku znacznikowego XML (meta dane). Widok głównego okna aplikacji z anonimowymi danymi pacjentów (co było działaniem celowym) przedstawia rys. 8.



Rys. 8. Widok okna głównego aplikacji z przykładowymi danymi DICOM

Przykładowe dane są publicznie dostępnymi wynikami badań zebranych w repozytoriach takich badań. Są one przeznaczone do prac badawczych poświęconych zastosowaniu formatu DICOM i jego przetwarzaniu. W większości przypadków mają one usunięte dane identyfikujące pacjenta albo dane te są zastępowane wpisami fikcyjnymi.

5 Filtrowanie informacji graficznej

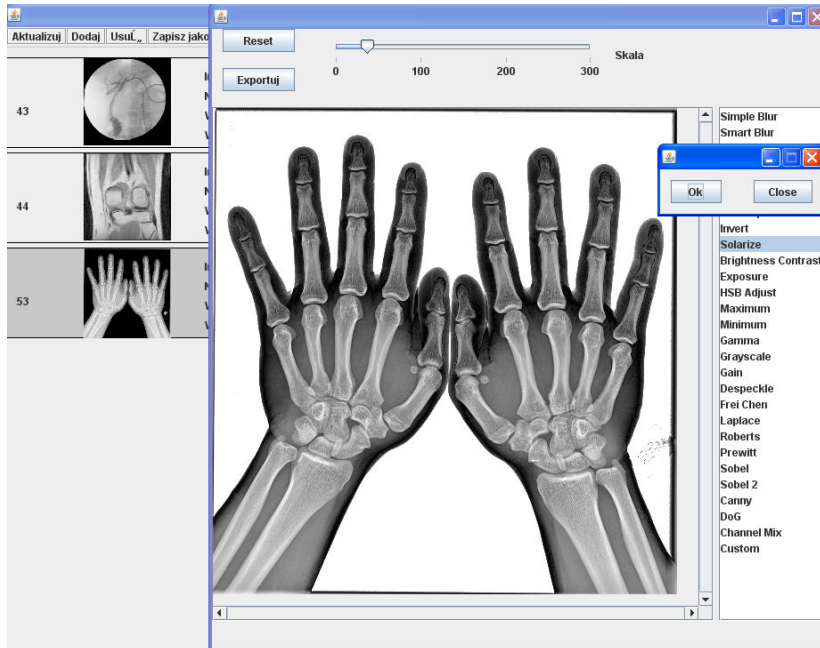
Dane graficzne zawarte w plikach DICOM są bardzo dobrej jakości, co wynika z jakości urządzeń obrazowych, jak również jest podyktowane wymaganiami diagnostycznymi. Powoduje to, że proces przesyłania oraz przetwarzania danych jest długotrwały i stawia wysokie wymagania zarówno w stosunku do silnika bazy danych jak i infrastruktury sieciowej, przez którą końcówka klienta łączy się z bazą.

Pomimo wysokiej jakości obrazu w wielu przypadkach konieczne jest poddanie go dodatkowej obróbce, która nie musi być dostępna po stronie urządzenia obrazowego. Jedną z podstawowych operacji jest wszelkiego rodzaju filtrowanie, a najpowszechniejszymi filtrami są filtry macierzowe. Definiujemy w nich macierz przekształcenia, najczęściej o wymiarze 3×3 , ale nie istnieją żadne ograniczenia formalne co do tej cechy. Również dopuszczalne jest, aby macierz ta była różna od kwadratowej, co w praktyce prawie nigdy nie ma zastosowania. Wartości wpisane w pola macierzy są wagami, przez które mnożymy cechy punktów (np. nasycenie lub wartości kanałów RGB) z otoczenia punktu centralnego (łącznie z tym punktem), zgodnie ze wzorem (1)

$$V_{k,l} = \frac{\sum_{i=-n}^n \sum_{j=-n}^n (w_{i,j} V_{k+i,l+j})}{\sum_{i=-n}^n \sum_{j=-n}^n w_{i,j}} \quad (1)$$

W wyrażeniu zastosowano skalowanie wyniku przez sumę wartości wszystkich wag, tak, aby zakres wartości barw lub jasności) pozostał bez zmian. Skalowanie takie wymusza, że suma wag nie może być równa 0, natomiast nie ma zakazu, aby były one ujemne. Filtrami tej klasy są na przykład filtry wyostrzające, rozmywające czy też wykrywające krawędzie.

W aplikacji po wybraniu miniatury obrazu możliwe jest uruchomienie oddzielnego okna narzędzia filtrującego dla obrazu źródłowego – rys. 9. Na liście umieszczony został wykaz predefiniowanych filtrów dostępnych z poziomu aplikacji. Wybrany filtr solaryzacji, jest bezparametrowy, dlatego możliwe jest tylko zatwierdzenie jego działania lub powrót do stanu pierwotnego. Na rysunku 9 widać skutek zastosowanego filtra solaryzacji do przykładowego badania dłoni.



Rys. 9. Widok okna filtrującego na tle okna głównego aplikacji z listą filtrów i zastosowanym filtrem Solarize

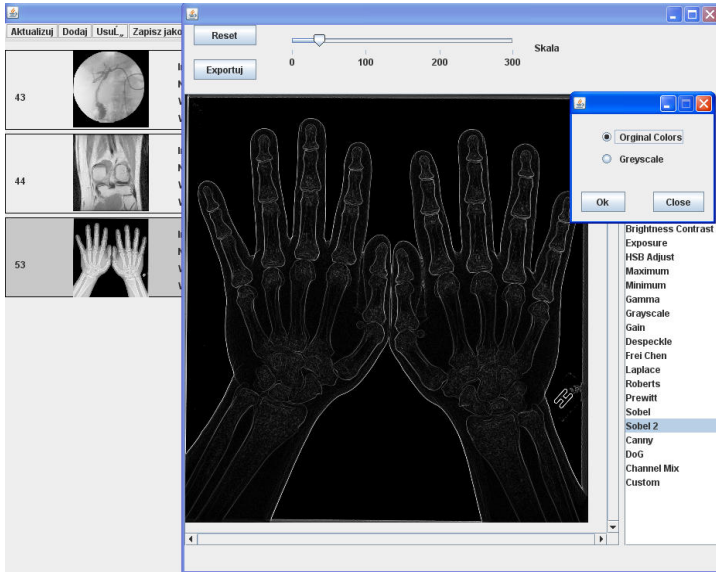
Filtry mogą być składane wielokrotnie, co może prowadzić do efektów niebędących intencją operatora. W takim przypadku istnieje możliwość wycofania filtrowania przyciskiem Reset. Jeśli otrzymujemy pożądany efekt rezultaty pracy mogą zostać zapisane do zewnętrznego pliku graficznego, który może podlegać dalszej obróbce narzędziami zewnętrznymi. Istnieje możliwość powiększania obrazu tak, aby wyeksponowane zostały najbardziej interesujące fragmenty.

Wśród filtrów istnieje spora grupa takich, które posiadają zestaw parametrów. Jednym z przykładów jest filtr wyszukujący krawędzi o nazwie Sobel 2, dla którego możliwy jest wybór pomiędzy filtrowaniem w skali szarości a wykorzystaniem pełnej palety obrazu źródłowego. Skutek działania przedstawia rys. 10.

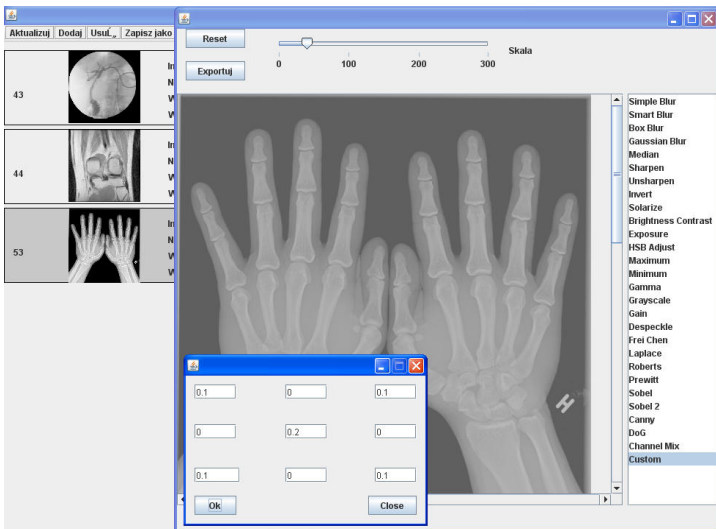
Ponieważ standardowy zestaw filtrów, może okazać się niewystarczający w przypadku konkretnych danych obrazowych, dodany został filtr użytkownika z ręcznie definiowaną macierzą transformacji (filtrowania). Zastosowano standardową jej postać o wymiarze 3x3 oraz elementami będącymi liczbami rzeczywistymi. Przykład definicji filtra użytkownika oraz skutek jego działania został przedstawiony na rys. 11.

Zestaw filtrów jest bogaty, pomimo to istnieje łatwa możliwość jego powiększenia na skutek dodania metody do odpowiedniej klasy obiektowej. Potencjalnie takim rozszerzeniem może być zastosowanie filtrów adaptacyjnych, którym ostatnio poświęcono wiele publikacji naukowych.

Innym podejściem może być wykorzystanie już na poziomie programistycznym składania filtrów, tak, aby dostarczyć odbiorcy, lekarzowi łatwego w użyciu narzędzia. Należy przy tym pamiętać, że operacje filtrowania nie są przemienne, ponieważ praktycznie wszystkie transformacje otrzymywane na skutek ich stosowania są transformacjami nieliniowymi.



Rys. 10. Widok okna filtrującego na tle okna głównego aplikacji z listą filtrów i zastosowanym filtrem Sobel 2



Rys. 11. Widok okna filtrującego na tle okna głównego aplikacji z listą filtrów i zastosowanym filtrem macierzowym definiowanym przez użytkownika

6 Wnioski

W trakcie badań powstał system umożliwiający zarządzanie danymi medycznymi w formacie DICOM spełniający wszystkie zaplanowane założenia. Została odpowiednio skonfigurowana baza danych oraz stworzona aplikacja kliencka.

Zaprojektowany system zapewnia spełnienie głównych założeń takich jak:

- ♦ przetwarzanie plików DICOM – wykorzystanie silnika bazy danych oraz bibliotek dostępnych dla języka Java w celu analizy danych zawartych w plikach DICOM.
- ♦ wykorzystanie bazy danych do składowania informacji o pacjencie – odpowiednie skonfigurowanie bazy danych w celu optymalnego przechowywania informacji o pacjencie.
- ♦ stworzenie użytkownika z dostępem do bazy danych, stworzenie przestrzeni tabel, stworzenie odpowiedniej hierarchii tabel oraz funkcji i procedur składowanych.
- ♦ aplikacja współpracująca z bazą danych – stworzono aplikację łączącą się z bazą danych, która w przejrzysty sposób wyświetla informacje o pacjencie oraz pozwala nimi zarządzać.
- ♦ zaimplementowanie filtrów graficznych do analizy obrazu – w aplikacji zostało zaimplementowanych około trzydziestu filtrów graficznych operujących na obrazach zawartych w plikach DICOM, między innymi są to filtry do wykrywania krawędzi, zmiany nasycenia kolorów czy usuwania szumów.
- ♦ odseparowanie kluczowych zapytań SQL od aplikacji – większość zapytań SQL zostało odseparowanych od aplikacji, dodane zostały, jako funkcje i procedury składowane w bazie danych. Dzięki temu możliwa jest zmiana struktury bazy danych bez modyfikacji aplikacji.

System ten może być dalej rozwijany, przez dodanie nowych filtrów do aplikacji. Ze względu na modułową budowę aplikacji ogranicza się to o utworzenia nowej metody w klasie filtrów. W przypadku większej rozbudowy warto zastanowić się nad wykorzystaniem frameworka Netbeans Platform RCP, z którego korzystają największe firmy i instytucje na świecie takie jak NASA czy The Boeing Company. Ciekawym pomysłem jest także wprowadzenie opcji wyszukiwujących danych DICOM w zależności od wybranych przez użytkownika kryteriów zapisanych w metadanych, a w kolejnym kroku z oznaczonych w procesie przetwarzania grafiki cech obrazu.

7 Literatura

- [1] Revet B., *DICOM Cook Book for implementation in Modalities, chapter 1 and 2*, Jan 1997

- [2] Pelikant A., *Integracja danych medycznych za pomocą plików znacznikowych XML i DICOM*, *Technologie informatyczne w administracji publicznej i służbie zdrowia*, Wydawnictwo SGH 2010
- [3] Pabjańczyk K., Pelikant A., *Implementacja parsera protokołu DICOM, przy użyciu języka T-SQL na platformie SQL Server 2008*, *Studia Informatica*, Volume 32, No 2B (97) Silesian University of Technology Press, Gliwice 2011, pp. 547-559.
- [4] Szalinski M., Pelikant A., *Medical data migration and access management platform*, *System Modeling Control -2005*, Zakopane 2005, *Problemy współczesnej nauki teoria i zastosowania*, Informatyka, Akademicka Oficyna Wydawnicza Exit, Warszawa 2005
- [5] Wojciechowski K., Pelikant A., *Zastosowanie protokołu DICOM oraz technologii J2EE w komunikacji między urządzeniami medycznymi a bazą danych*, *Bazy danych Nowe technologie, Bezpieczeństwo, wybrane technologie i zastosowanie*, Rozdział 22, Wydawnictwa Komunikacji i Łączności, Warszawa 2007, ss. 229-237
- [6] Wang Ch., Huang Y., *DICOM Communication Mechanism and Engineering Project Integration Based on ESB*, 2010 WASE International Conference on Information Engineering, August 2010, pp. 119-122
- [7] Hui Tee S., *Integrating DICOM Information Model with Risk Management Process Area of CMMI for Radiotherapy Applications*, 2010 Second International Conference on Computer Research and Development, May 2010, pp. 73-76
- [8] Selvarani G. A., Annadurai S., *Medical Image Retrieval by Combining Low Level Features and DICOM Features*, 2007 International Conference on Computational Intelligence and Multimedia Applications, December 2007, pp. 587-589
- [9] Rui A., Pereira M., Freire M., *Advanced Querying Architecture for DICOM Systems*, Second International Conference on Systems and Networks Communications (ICSNC 2007), August 2007, pp. 53
- [10] Stanescu L., Ion A., Burdescu D., Brezovan M., *Algorithms and Results in Content-Based Visual Query of the Image Databases Resulting from Dicom Files*, Eighth International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC'06), September 2006, pp. 119-124
- [11] Pelikant A., *Bazy danych w zastosowaniach praktycznych – Monografia WSInf.*, 2007

- [12] Horzelski W., *System rejestracji danych Dyżurowych w Klinice Medycyny Matczyno-Płodowej ICZMP w Łodzi. Bazy danych, struktury, algorytmy, metody*, WKiŁ, Warszawa 2006
- [13] Lance A., Kyte T., *Oracle Database Concepts 11g Release 2*. WWW, 2010
- [14] Eckel, B., *Thinking in Java. Edycja polska. Wydanie IV*. Helion., 2006
- [15] Kuassi M., *Oracle Database Programming Using Java and Web Services*. Elsevier Digital Press, 2006
- [16] Pelikant A., *Programowanie serwera Oracle 11g SQL i PL/SQL*. Helion, 2009
- [17] Pelski S., *Oracle Multimedia DICOM Developers Guide 11g Release 2*. WWW, 2010
- [18] Pelski S., *Oracle Multimedia User's Guide 11g Release 2*. WWW, 2010
- [19] Wrycza, S., *UML 2.1. Ćwiczenia*. Helion, 2007
- [20] Wrycza, S., Marcinkowski, B., Wyrzykowski, K., *Język UML 2.0 w modelowaniu systemów informatycznych*. Helion, 2006
- [21] <http://www.sno.phy.queensu.ca/~phil/exiftool/TagNames/DICOM.html>.
- [22] <http://medical.nema.org>

MEDICAL DATA PROCESSING AND ANALYSIS APPLYING DICOM ON ORACLE SERVER PLATFORM

Summary - Article contains presentations of methods allowing to manage medical devices from a central database server. As the media was used DICOM format, which is the standard for the information exchange in the imaging medical devices – medical scanners. Provides methods for the storage and extraction of metadata on the ORACLE server platform, using the built-in object types and APIs. Beyond the current processing, the graphics filters bank was built for visually information processing. It gives possibility to improve processing image quality and especially exams data by features extraction.