

Marcin CHABIOR, Magdalena TKACZ  
Uniwersytet Śląski, Wydział Informatyki i Nauki o Materiałach,  
Zakład Systemów Informatycznych

## LABORATORYJNE SYSTEMY ZARZĄDZANIA INFORMACJĄ JAKO UŻYTECZNE ŹRÓDŁO DANYCH W POZYSKIWANIU WIEDZY PRZY UŻYCIU METOD EKSPLOKACJI DANYCH

**Streszczenie.** Artykuł wskazuje na Laboratoryjne Systemy Zarządzania Informacją (ang. LIMS – *Laboratory Information Management System*) jako bazodanowe źródło danych możliwych do wykorzystania w procesie pozyskiwania wiedzy przy użyciu metod eksploracji danych. W artykule opisano zasady funkcjonowania LIMS, kierunki ich rozwoju oraz oczekiwania stawiane tym systemom na dzień dzisiejszy, ze strony laboratoriów badawczych. Na przykładzie systemu *Sample-based Laboratory Information Management System* (SLIMS) pokazano, że przy użyciu aplikacji pośredniczącej można pobrać określone dane zgromadzone w bazie systemu LIMS i wykorzystać je w celu eksploracji danych przy użyciu jednej z metod Data Mining. Opisano ponadto sposób pozyskiwania wybranych danych z bazy danych systemu laboratoryjnego przy użyciu autorskiej aplikacji, odpytującej przez składnię zapytań języka SQL, i użycie tych danych w metodzie drzew decyzyjnych C4.5. Rozwiązanie opisane w artykule pokazuje możliwość rozszerzenia funkcjonalności LIMS o zakres pozyskiwania wiedzy z danych.

**Słowa kluczowe:** lims, zarządzanie treścią, systemy laboratoryjne, eksploracja danych, systemy ekspertowe, algorytm c4.5, drzewa decyzyjne

## LABORATORY INFORMATION MANAGEMENT SYSTEMS AS A USEFUL SOURCE OF DATA TO KNOWLEDGE DISCOVERY USING DATA MINING METHODS

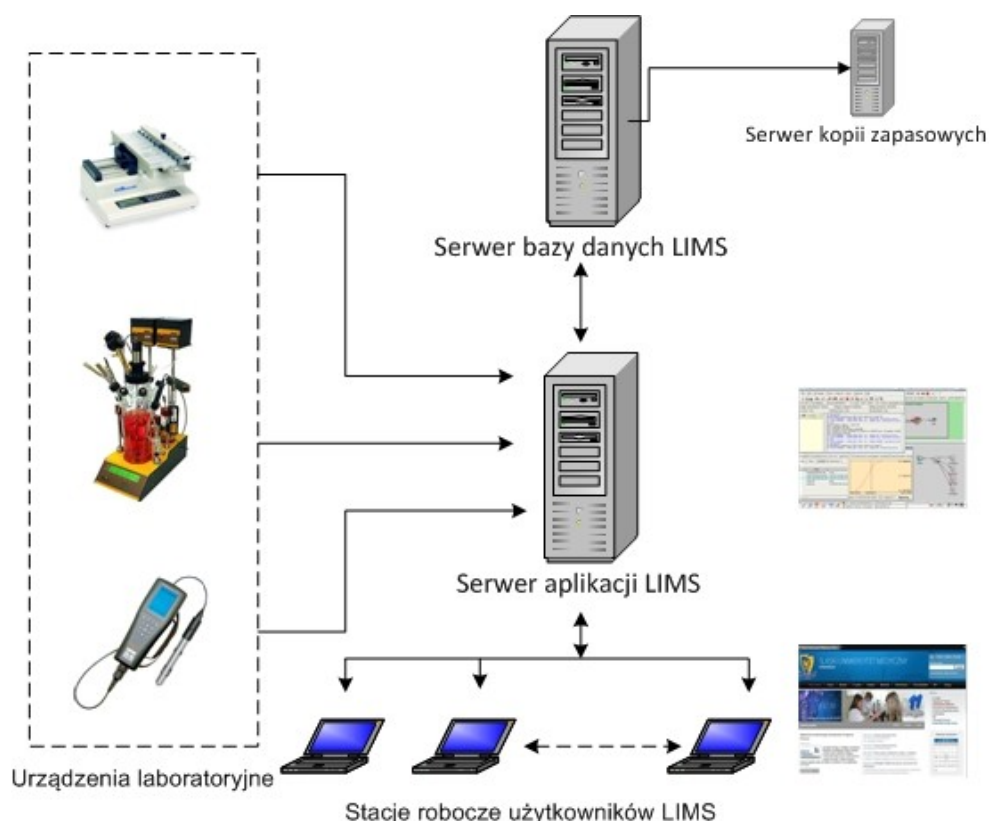
**Summary.** The article describes Laboratory Information Management Systems (LIMS) as a useful source of data to obtain new knowledge by using data mining techniques. Based on the example, article shows that using a simple application user can retrieve data from the LIMS database and use them to explore the data using data mining methods. The article describes also the use of a particular group of data stored in the Genapha SLIMS system to construct a decision tree using the algorithm C4.5.

The solution described in the article, shows the possibility to extend the functionality of the LIMS through knowledge discovery from data

**Keywords:** lims, information management, laboratory systems, data mining, decision tree, algorithm C4.5

## 1. Charakterystyka Laboratoryjnych Systemów Zarządzania Informacją

Laboratoryjne Systemy Zarządzania Informacją są systemami projektowanymi na potrzeby laboratoriów analitycznych i badawczych, celem gromadzenia danych pochodzących z instrumentów laboratoryjnych. Instrumenty te, w postaci chromatografów, spektrometrów i wszelakich analizatorów, generują dane przesyłane do baz danych współpracujących z nimi stacji roboczych bądź bezpośrednio do bazy danych serwera bazodanowego, gromadzącego dane z wielu instrumentów laboratoryjnych. Przykładowa architektura typowego LIMS została pokazana na rys. 1.



Rys. 1. Typowa architektura systemu LIMS  
Fig. 1. Typical architecture of LIMS system

Zasadniczo LIMS pozwala na unifikację obszernych i różnorodnych grup danych biologicznych oraz chemicznych, pochodzących od wielu niezwiązanych ze sobą aplikacji wspierających działanie narzędzi laboratoryjnych na potrzeby prezentacji tychże danych w formie

zbiorów tabel, wykresów oraz raportów. Niezależnie od skomplikowania systemu, LIMS realizują następujące funkcje [1]:

1. Pozyskiwanie danych z instrumentów laboratoryjnych.
2. Analiza danych w powiązaniu z redukcją danych.
3. Generowanie raportów oraz prezentacja danych.
4. Zarządzanie laboratorium przez zarządzanie danymi.

W LIMS dokonuje się konwersja danych zebranych z instrumentów analitycznych do postaci użytecznej dla odbiorcy. W ujęciu tradycyjnym LIMS oparte były na prostej technologii klient-serwer, działającej w sieci lokalnej laboratorium. Rozwiązanie to zapewniało dostęp do danych wyłącznie w fizycznie ograniczonym środowisku, a integracja z systemami bądź aplikacjami wspomagającymi i uzupełniającymi pracę laboratoriów była zazwyczaj skomplikowana, a co za tym idzie kosztowna. Późniejsze oparcie LIMS na technologiach webowych znacznie uelastyczyło architektury tych systemów i dało nowe możliwości w zakresie ich rozwoju. Wykorzystanie technologii XML oraz usług webowych SOAP umożliwia programistom tworzenie skalowalnych systemów LIMS. Statyczne dotąd interfejsy użytkownika zastąpione zostały przez dynamiczne GUI, posiadające wiele funkcjonalności, przekładających się na jakość oraz czas realizacji badań z korzyścią dla obu tych aspektów [2].

Precyzja oraz wielozadaniowość stosowanych urządzeń laboratoryjnych w sposób bezpośredni przekładają się na ilość gromadzonych w LIMS danych, które muszą zostać odpowiednio przetworzone do formy użytecznej. LIMS muszą posiadać zatem informatyczne rozwiązania pozwalające m.in. na generowanie obrazów graficznych na podstawie danych pochodzących z chromatografu lub spektrometru, konwertowanie surowych danych do formatów umożliwiających działanie w różnych systemach LIMS, walidację danych celem uniknięcia błędów analitycznych, tworzenie raportów i wiele innych [3].

Oczekiwania w stosunku do dzisiejszych LIMS to nie tylko łatwy dostęp do prezentowanych w czytelnej formie danych laboratoryjnych, ale również pozyskanie z tychże danych nowej, nieodkrytej do tej pory wiedzy bądź informacji pozwalających na podjęcie określonych decyzji odnośnie kierunków i sposobów prowadzenia dalszych badań.

## **2. Sample Laboratory Information Management System jako przykład systemu LIMS**

Jednym z przykładowych systemów laboratoryjnych do zarządzania treścią dostępnych na zasadach *OpenSource* jest *Sample Laboratory Information Management System* (SLIMS), udostępniany przez firmę *Genapha*. SLIMS jest zaawansowaną i jednocześnie przyjazną użytkownikowi aplikacją webową, umożliwiającą laborantowi podgląd, edycję i wprowadza-

nie informacji dotyczących prowadzonych, medycznych badań laboratoryjnych. Aplikacja posiada dodatkowo moduły generowania raportów, list oraz wirtualnych mikromacierzy celem wprowadzania danych uzyskiwanych w procesie analizy DNA [4]. Przez interfejs administracyjny SLIMS możliwe jest:

- zarządzanie tematami, próbkami i kontenerami danych,
- przeglądanie danych oraz kompleksowe wyszukiwanie przy użyciu wbudowanego systemu wyszukiwania,
- wprowadzanie walidacji dla określonych pól,
- zarządzania prawami użytkowników,
- tworzenie dynamicznych raportów,
- modyfikowanie widoków danych,
- wykorzystywanie wirtualnych macierzy o liczbie pól 96 lub 384,
- klonowanie danych, w tym danych przechowywanych w wirtualnych mikromacierzach,
- importowanie danych z plików tekstowych oraz eksportowanie danych,
- modyfikowanie kodu źródłowego oraz dodawanie tabel w zależności od potrzeb laboratorium.

Przykładowy widok interfejsu aplikacji pokazano na rys. 2. Dokumentacja oraz pliki źródłowe *Sample Laboratory Information System* dostępne są na stronie <http://genapha.icapulture.ubc.ca/SLIMS/index.jsp>. Działanie SLIMS wymaga zastosowania następujących technologii: Java 1.6.0, JSPs, Hibernate 3.3.1. GA, DB2 i MySQL, Apache Tomcat 6.0.18, NetBeans ISE 6.5, Jasper Reports 3.5.1 oraz JasperSoft's Report 3.5.1 [4].

The screenshot shows the SLIMS application interface. At the top, there is a navigation menu with options: Home, Search, Browse, Lists, Reports, Supporting Data, Upload, Logout, and No active list. Below the menu, there is a search bar and a 'What would you like to do?' dropdown. The main content area displays a table of samples. The table has the following columns: Collection Date, Sample, Sample Type, Material Type, Stock, Volume (ul), Concentration (ng/ul), Barcode, and Container Type. The table contains several rows of data, including samples with IDs like Az1001-1a and various container types such as 96 Well Plate, Box of Tubes, and 384 Well Plate. A magnifying glass is positioned over the 'Create List' button in the top left corner of the table.

Collection Date	Sample	Sample Type	Material Type	Stock	Volume (ul)	Concentration (ng/ul)	Barcode	Container Type
	Az1001-1a	Blood	Genomic	no	35.79	47.86		96 Well Plate
	Az1001-1a	Blood	Genomic	yes	297.31	137.2		Box of Tubes
	Az1001-1a	Blood	Genomic	yes	1.0	10.0		96 Well Plate
	Az1001-1a	Blood	Genomic	no	0.0	1.0		384 Well Plate
	Az1001-1a	Blood	Genomic	no	5.0	1.0		384 Well Plate
	Az1001-1a	Blood	Genomic	no	5.0	1.0		384 Well Plate

Rys. 2. Wybrany widok interfejsu aplikacji SLIMS  
Fig. 2. Chosen view of SLIMS application

### 3. Wykorzystanie danych systemu SLIMS w procesie tworzenia drzewa decyzyjnego przy użyciu algorytmu C4.5

#### 3.1. Wybór danych z bazy systemu SLIMS

Struktura bazy danych, na podstawie której funkcjonuje SLIMS, składa się z 29 tabel, w których przechowywane są dane opisujące badany materiał biologiczny zarówno pod kątem jego pochodzenia, jak i uzyskanych na jego podstawie wyników laboratoryjnych. Oprócz domyślnie zdefiniowanych tabel, użytkownik systemu może tworzyć własne tabele jako obiekty w SLIMS, co powoduje, że system może być elastycznie rozbudowywany do postaci oczekiwanej, przy jednoczesnym zachowaniu jego podstawowej funkcjonalności. W niniejszym artykule spośród informacji zawartych w systemie SLIMS wybrano następujące, celem wykorzystania ich w procesie eksploracji danych.

Tabela 1

Przykładowe tabele bazy danych systemu SLIMS

Nazwa tabeli w systemie SLIMS	Pobierane informacje
tblcontainercontents	CONCENTRATION – stężenie czynnika w materiale badawczym pobranym od pacjenta
tblcontainers	MATERIALTYPEID – pochodzenie organiczne tkan-ki badanej
tblsimsusers	UAGE – wiek badanego pacjenta

Opisany w niniejszym artykule przykład, na potrzeby przystępnego zobrazowania zagadnienia, opiera się wyłącznie na możliwie najmniejszej grupie danych. Im większa liczba danych wejściowych, tym większa dokładność wyniku uzyskanego ostatecznie w procesie eksploracji danych [5]. Odwołanie się jednak do takich danych, jak: płeć, czynniki środowiskowe, obecność określonych chorób czy też wybrane parametry badań krwi, nie jest niezbędne dla opisanego procesu wykorzystania danych w celu poszukiwania wiedzy, choć zapewne w znaczący sposób przekładają się na jakość i wiarygodność uzyskanych wyników.

#### 3.2. Aplikacja SLIMS-connect, jako autorskie narzędzie do pobierania danych z bazy danych systemu SLIMS

Chcąc uprościć import danych, utworzona została autorska aplikacja umożliwiająca import danych z bazy danych systemu SLIMS. Aplikacja została przygotowana tak, by pobrać wyłącznie te dane, które użytkownik chce zestawić względem siebie celem wyszukania prawidłowości. Dane zaimportowane umieszczane są w nowej bazie danych, w tabeli *dataslims*, z zachowaniem pierwotnych nazw kolumn. Zadaniem aplikacji jest zarówno odpytywanie bazy danych MySQL celem importu, jak i wstępne przygotowanie danych oraz export danych do pliku. Aplikacja przygotowana została z użyciem języka programowania Delphi 8 i funkcjonuje w środowisku MS Windows.

Ze względu na ewentualne wartości brakujące, punkty oddalone bądź konieczność normalizacji danych, wprowadzono w aplikacji proste rozwiązania, pomocne we właściwym przygotowaniu danych do procesu eksploracji [5]. Wartości brakujące (dla danych liczbowych) mogą być zastępowane średnią arytmetyczną lub wartością średnią modalną. Dla punktów oddalonych zakłada się trzy progi tolerancji, bazujące na wartości średniej – akceptowalne punkty oddalone do +50% wartości średniej arytmetycznej, do +100% wartości średniej arytmetycznej oraz do 200% wartości średniej arytmetycznej. Wszystkie wartości będące poza przedziałem akceptacji są usuwane. W odniesieniu do przekształcania danych wykorzystana została standaryzacja danych, która działa przez obliczenie różnicy pomiędzy daną wartością pola  $X$  i średnią wartością pola  $X$  oraz przeskalowanie tej różnicy przez odchylenie standardowe  $\sigma$  wartości pól, co opisano wzorem (1):

$$X^* = \frac{X - \text{średnie}(X)}{\sigma(X)}, \quad (1)$$

gdzie:  $X^*$  – wartość znormalizowana,  $X$  – wartość pierwotna,  $\text{średnie}(X)$  – średnia wartości pola,  $\sigma(X)$  – odchylenie standardowe wartości pól.

Dane poprawione i oczyszczone mogą zostać wyeksportowane do jednego z następujących formatów plików: \*.txt, \*.csv, \*.xml.

Niniejsze formaty zostały wybrane ze względu na swoją uniwersalność w aspekcie wykorzystania ich przez inne aplikacje realizujące stricte metody eksploracji danych. W niniejszym artykule, dla przeprowadzenia eksploracji danych na opisanym dalej przykładzie, wykorzystano aplikację RapidMiner firmy Rapid-I. Niniejsze oprogramowanie jest narzędziem opartym na licencji open-source, wykorzystywanym do analizy oraz eksploracji danych. Obok tak zaawansowanych narzędzi, jak Statistica firmy StatSoft, SQL Server Analysis Services firmy Microsoft czy GNU R, stanowi ono jedno z pomniejszych, aczkolwiek interesujących oprogramowań realizujących zadania odkrywania wiedzy z danych.

### 3.3. Tworzenie drzewa decyzyjnego przy użyciu algorytmu C4.5 na podstawie danych z systemu SLIMS

Jednym z algorytmów zaimplementowanych w programie RapidMiner jest algorytm drzewa decyzyjnego C4.5. Formalnie drzewem decyzyjnym jest graf-drzewo, którego korzeń jest tworzony przez wybrany atrybut, natomiast poszczególne gałęzie reprezentują wartości tego atrybutu [6]. Zaletą drzew decyzyjnych jest możliwość reprezentowania dowolnie złożonych pojęć pojedynczych oraz wielokrotnych, jeśli tylko ich definicje da się wyrazić w zależności od atrybutów. Ponadto, czas decyzyjny jest ograniczony liniowo przez liczbę atrybutów, a forma prezentacji jest czytelna dla odbiorcy. Nie bez znaczenia jest również łatwość przejścia od prezentacji drzewiastej do reprezentacji regułowej.

Tabela 2

Przykładowe dane pobrane z systemu SLIMS na potrzeby eksploracji

Pacjent	Wiek badanego pacjenta	Stężenie czynnika w materiale badawczym pobranym od pacjenta	Pochodzenie materiału badanego	Ryzyko rozwoju choroby
1	45	małe	węzły limfatyczne	małe ryzyko
2	23	duże	wątroba	duże ryzyko
3	71	średnie	tarczyca	duże ryzyko
4	41	średnie	wątroba	małe ryzyko
5	20	średnie	węzły limfatyczne	małe ryzyko
6	75	małe	tarczyca	małe ryzyko
7	23	duże	tarczyca	duże ryzyko
8	49	średnie	węzły limfatyczne	małe ryzyko

Posługując się przykładowymi danymi zawartymi w tabeli 2, można wykazać użyteczność bazy SLIMS w procesie generowania drzewa decyzyjnego przy użyciu algorytmu C4.5. Algorytm C4.5 rekurencyjnie odwiedza każdy węzeł decyzyjny, wybierając możliwy podział, dopóki dalsze podziały są możliwe [7]. Algorytm C4.5 używa pojęcia zysk informacji lub redukcja entropii w celu wybrania optymalnego podziału. Najmniejsza liczba bitów, średnio na symbol, potrzebna do przesłania łańcucha symboli reprezentujących obserwowane wartości  $x$ , jest nazywana entropią  $X$ . Innymi słowy entropia jest średnią ilości informacji określonej na zbiorze prawdopodobieństw wszystkich możliwych realizacji określonego zdarzenia. Jeżeli zatem liczba możliwych komunikatów wynosi  $n$ , ilość informacji zawartej w komunikacie generowanym przez źródło wyraża się w następujący sposób (2):

$$H(X) = - \sum_{j=1}^n p_j \log_2(p_j), \quad (2)$$

gdzie:  $H(X)$  – entropia zbioru wartości  $X$ ,  $p_j$  – prawdopodobieństwo wystąpienia  $i$ -tego komunikatu w zbiorze wynikowym,  $n$  – liczba komunikatów.

Jak zatem widać, wartość tej miary bezpośrednio zależy od wystąpienia różnych komunikatów i osiąga wartość maksymalną, w przypadku gdy prawdopodobieństwa wystąpienia wszystkich komunikatów są równe. Oczywiście, w przypadku gdy prawdopodobieństwo któregośkolwiek z komunikatów przyjmie wartość równą jedności, wynik entropii wyniesie zero. Przyjmując, że dany podział  $S$  dzieli zbiór uczący  $T$  na kilka podzbiorów  $T_1, T_2, \dots, T_k$ , możliwe jest wyliczenie średniego zapotrzebowania  $H_S(T)$  na informację, będącego ważoną sumą entropii  $H_S$  dla pojedynczych podzbiorów  $T_i$ :

$$H_S(T) = \sum_{i=1}^k P_i H_S(T_i), \quad (3)$$

gdzie:  $T_i$  – pojedynczy zbiór danych,  $P_i$  – prawdopodobieństwo wystąpienia komunikatu w  $i$ -tym zbiorze.

Na podstawie podanych wcześniej pojęć można zdefiniować zysk informacji:

$$zysk(S) = H(T) - H_S(T). \quad (4)$$

Innymi słowy zwiększenie informacji wytworzone przez podział zbioru uczącego  $T$ , zgodnie z możliwym podziałem  $S$  w każdym węźle decyzyjnym algorytmu C4.5, wybiera podział optymalny, czyli mający największy zysk informacji.

Wykorzystując algorytm C4.5, możemy dowiedzieć się, które z czynników znacząco wpływają na parametr oznaczony jako „ryzyko rozwoju choroby”. Dla każdego przypadku ujętego w tabeli 2 obliczono entropię lokalną, co, po obliczeniu sumy lokalnych entropii, pozwala na obliczenie zysku informacji zgodnie z regułami tworzenia drzew decyzyjnych metodą C4.5. Uzyskane wyniki dla pierwszej iteracji zestawiono w tabeli 3, a na ich podstawie możliwe jest rozpoczęcie konstrukcji drzewa decyzyjnego.

Tabela 3

Wyniki zysku informacji algorytmu C4.5 potrzebne do zainicjowania drzewa decyzyjnego

Podział	Poddrzewa	Zysk informacji
1	Przedział wiekowy 20 – 45	0,36 bita
	Przedział wiekowy 45 – 70	
	Przedział wiekowy > 70	
2	Stężenie materiału badanego = małe	0,5487 bita
	Stężenie materiału badanego = średnie	
	Stężenie materiału badanego = duże	
3	Pochodzenie materiału badanego = tarczycza	0,1588 bita
4	Pochodzenie materiału badanego = węzły limfatyczne	0,3475 bita
5	Pochodzenie materiału badanego = wątroba	0,0923 bita

Dalsze obliczenia, zgodne z regułami algorytmu C4.5, pozwalają na rozwinięcie drzewa decyzyjnego aż do postaci ostatecznej, w której nie występują już węzły decyzyjne. Wynik działania algorytmu został pokazany na rys. 3. Na podstawie uzyskanych wyników możliwe jest jednocześnie utworzenie czytelnych dla odbiorcy reguł decyzyjnych, opisujących określone zdarzenia dla zadanych danych wejściowych.

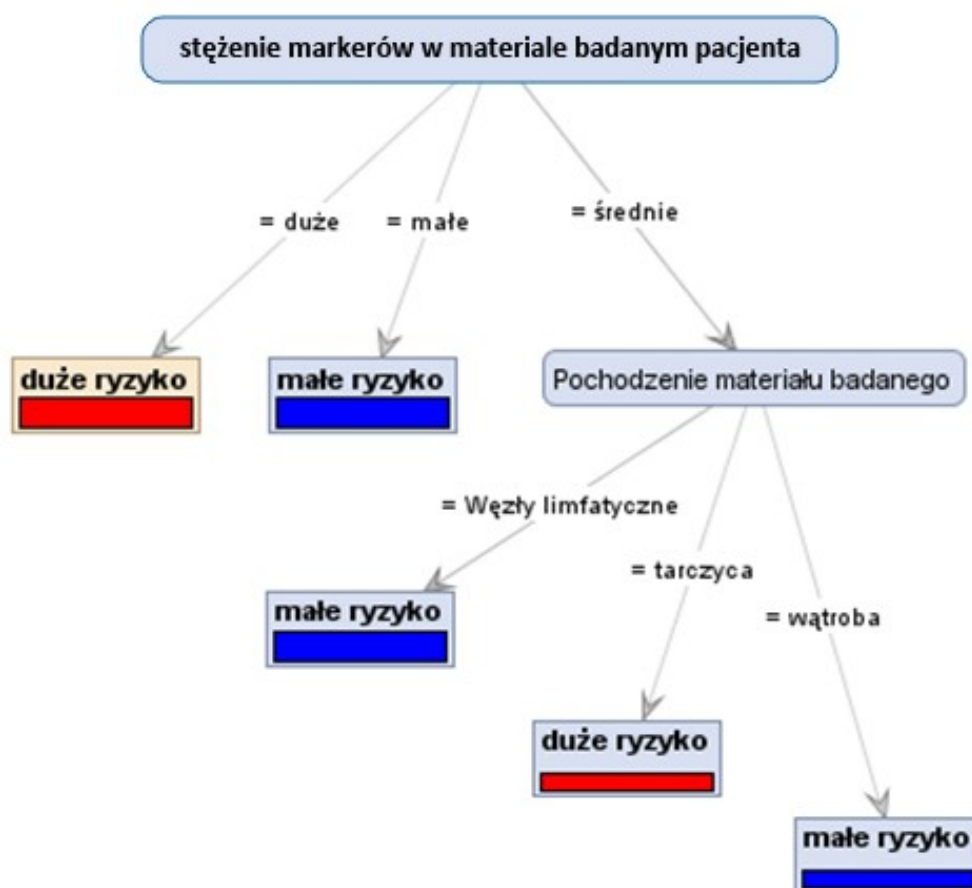
## 4. Wnioski

Laboratoryjne Systemy Zarządzania Informacją posiadają na dzień dzisiejszy rozległe zbiory danych, które, oprócz roli diagnostycznej, mogą być wykorzystane w znacznie szerszym aspekcie badawczym, umożliwiającym pozyskiwanie nowej wiedzy. Opisany w niniejszym artykule przykład bazuje na możliwie najprostszym zbiorze informacji. Jednak wykorzystanie znacznie większej liczby danych i użycie bardziej wyrafinowanych algorytmów eksploracji danych może przynieść w pełni zadowalające efekty. Należy jednak pamiętać, że większość LIMS stanowią rozwiązania komercyjne i niestety tworzone często na indywidualne potrzeby laboratoriów, co powoduje, iż wydobycie danych z tych systemów może być za-



również kosztowne, jak i czasochłonne z uwagi na nieustandaryzowane architektury baz danych.

W perspektywie zapewne najbliższych kilku lat, wiodący producenci oprogramowania do zarządzania laboratoriami skłonią się do standardowej implementacji modułów eksploracji danych. Wynika to przede wszystkim z faktu, iż laboratoria badawcze, szczególnie medyczne, przejawiają potrzebę naukowego podejścia do realizowanych zadań, ponieważ daje to możliwość zarówno usprawnienia procesów decyzyjnych, jak również może być przyczynkiem do opracowywania własnych rozwiązań diagnostycznych, co w sposób bezpośredni przekłada się na wyniki ekonomiczne jednostek badawczych.



Rys. 3. Drzewo decyzyjne utworzone na podstawie danych z systemu SLIMS  
Fig. 3. Decision tree created on the basis of the gathered data from SLIMS

## BIBLIOGRAFIA

1. Cowan D.: Informatics for the Clinical Laboratory. Springer, New York 2005.
2. Paszko C., Turner E.: Laboratory Information Management Systems, Second Edition, Revised and Expanded. Marcel Dekker, Inc., 2001.

3. Nakagawa A. S.: LIMS: Implementation and Management. The Royal Chemistry Society, 1994.
4. Rossum T. V., Tripp B., Daley D.: SLIMS – A user-friendly sample operations and inventory management system for genotyping labs. Bioinformatic Advance Access, 30 May 2010.
5. Hand D., Mannila H., Smyth P.: Eksploracja danych. WNT, Warszawa 2005.
6. Larose D. T.: Odkrywanie wiedzy z danych. PWN, Warszawa 2006.
7. Wang J. T. L., Zaki M. J., Toivonen H. T. T., Shasha D.: Data Mining in Bioinformatics. Springer-Verlag, 2005.

Recenzenci: Dr hab. Tadeusz Pankowski, prof. Uniwersytetu im. Adama Mickiewicza  
Dr inż. Aleksandra Werner

Wpłynęło do Redakcji 16 stycznia 2011 r.

## Abstract

Laboratory Information Management Systems (LIMS) have basically been created to collect and process information obtained from various instruments and laboratory equipment. The obtained data are converted into user-friendly, readable and understandable results in the form of lists, tables, reports and charts. A typical LIMS scheme is shown in Figure 1. A very interesting issue concerns the use of data accumulated within a LIMS system on a much broader scale than the one mentioned above. This data can be seen as a useful source containing undiscovered knowledge, in other words, it can be used in the process of data mining.

The data stored in LIMS systems typically have an ordered structure with the simultaneous application of error control mechanisms, which significantly simplifies the implementation of appropriate algorithms for data mining. The article describes the use of a particular group of data stored in the Genapha SLIMS system to construct a decision tree. A sample view of a SLIMS application interface is shown in Figure 2. In order to retrieve data from a MySQL database system, an original application allowing the import of user-specified data, preliminary preparation of the data for further exploration, and finally export of the data to one of three formats was carried out. Sample data taken from the SLIMS system have been summarised in Table 2. A decision tree algorithm C4.5, based on the acquired information, can be used.

Data exported to a \*. csv, 8.txt and 8.xml format are downloaded to the RapidMiner statistical application, where a decision tree using the C4.5 algorithm can be generated on the

basis of the implemented mechanisms, as is shown in Figure 3. The article describes how to create decision trees using algorithm C4.5 and at the same time defines such concepts as entropy (2) or average demand for information (3). Selected results initiating the creation of a decision tree are shown in Table 3.

In conclusion, the article shows LIMS systems as a useful source of data to obtain new knowledge by using data mining techniques. Therefore, LIMS systems can be seen not only as a source of medical information, but also as an area of scientific research.

### **Adresy**

Marcin CHABIOR: Uniwersytet Śląski, Wydział Informatyki i Nauki o Materiałach. Zakład Systemów Informatycznych, ul. Będzińska 39, 41-200 Sosnowiec, Polska.

Magdalena TKACZ: Uniwersytet Śląski, Wydział Informatyki i Nauki o Materiałach. Zakład Systemów Informatycznych, ul. Będzińska 39, 41-200 Sosnowiec, Polska.