

DUOMENŲ BAZIŲ VALDYMO SISTEMŲ TINKAMUMO BIOMEDICININĖMS SISTEMOMS ĮVERTINIMAS

Renata Baronienė, Egidijus Paliulis

Šiaulių universitetas, Technologijos fakultetas

Įvadas

Šiuo metu labai aktuali problema – duomenų kaupimas ir spartus apdorojimas, nes jų kiekiai ir apimtys nuolat didėja. Duomenų bazių valdymo sistema (DBVS) – neatsiejama bet kurios informacinės sistemos dalis. Duomenų bazėse (DB) saugoma esminė organizacijos ar įmonės veiklos informacija. Dauguma rinkoje siūlomų ir įmonėse naudojamų DBVS veikia reliacinio modelio pagrindu [1, 2]. Galima teigti, kad reliacinis modelis vis dar išlieka populiariausias, nors sukurta naujesnių duomenų saugojimo modelių. Analizuojant literatūrą, nepavyko rasti išsamesnių DBVS tinkamumo biomedicininėms sistemoms įvertinimo. DBVS pasiūla didelė, todėl kartais sunku apsispręsti, kurią sistemą reikėtų pasirinkti. Todėl charakteristikų tyrimas ir palyginimai gali padėti pasirinkti reikiamą DBVS konkrečiu jos panaudojimo atveju.

Tyrimo tikslas – atlikti populiariausių reliacinių DBVS kokybinių ir kiekybinių charakteristikų analizę ir lyginimą, siekiant išsiaiškinti, kurios sistemos būtų tinkamesnės biomedicininiam duomenims saugoti ir apdoroti.

Uždaviniai: Išnagrinėti DBVS funkcionavimo bazės galimybes, programavimo kalbų ir SQL standartų palaikymą, DBVS palaikomus duomenų tipus ir jų įvairovę bei DBVS naudojamus duomenų gavimo įrankius.

Situacijos įvertinimas

Analizuojant literatūrą, rasta šiek tiek DBVS tyrimų ir palyginimų, tačiau dauguma jų apsiriboja tik keletu DBVS ir yra bendro pobūdžio [1, 3]. Kai kurių DBVS tyrimų analizė platesnė [2, 4, 5], tačiau problema studijuota gerokai anksčiau. Šiuo metu naujausios DBVS versijos kai kuriais aspektais yra pasikeitusios ir tobulesnės, todėl ankstesnių analizių išvados šiek tiek pasenusios.

Renkantis DBVS biomedicininį duomenų saugojimui ir apdorojimui, labai svarbi duomenų tipų įvairovė ir galimybė naudotis specializuotais duomenų gavybos įrankiais [6, 7, 8]. Nagrinėjant atliktus DBVS tyrimus, pasigesta išsamesnių tyrimų apie programavimo kalbų, duomenų tipų ir SQL standartų palaikymą. Taip pat mažai rašoma apie DBVS naudojamus duomenų gavybos įrankius.

Tyrimo kriterijai

Analizuojant literatūrą, išsiaiškinta, kad DBVS charakteristikos paprastai skiriamos į dvi stambias grupes [1, 2, 3, 4, 5]: kokybines ir kiekybines. *Kokybinės charakteristikos* turinys: apimtis; taikymo sritis; funkcionavimo bazė; darbo patogumas; licencijos rūšis (kaina ir platinimas).

Kiekybinės charakteristikos turinys: palaikomi duomenų ir jų indeksų tipai; šifravimo algoritmai; laikinos ir virtualios lentelės; maksimalus užklausų skaičius; maksimalūs DB, lentelės ir įrašo dydžiai; maksimalus laukelių skaičius.

Nagrinėjant konkretų DBVS panaudojimo atvejį, ne visos charakteristikos vienodai svarbios ir aktualios. Biomedicininiai duomenys pasižymi kompleksiskumu [7, 8], tipų įvairove, todėl labai svarbi DBVS turima duomenų tipų įvairovė. Svarbus duomenų saugumas ir šifravimo algoritmai, DBVS vartojami apdorojimo įrankiai, kurie gali būti *integruoti* (vidiniai) arba *pritaikomi* (išoriniai). Ypač svarbūs duomenų gavybos įrankiai [7, 8]. Skirtingas DBVS vertinti ir lyginti įvestas toks charakteristikų reikšmingumas: duomenų gavybai skirta 40 %, o visoms kiekybinėms ir kokybinėms charakteristikoms – po 30 %. Surinkti kiekvienos DBVS įverčiai perskaičiuojami pagal jų reikšmingumą.

Biomedicinoje diagnostinei vaizdinei medžiagai taikomas DICOM standartas [9]. DICOM apibrėžė informacijos objektus ne tik vaizdiniais, bet ir pacientams, studijoms, ataskaitoms bei kitoms duomenų grupėms.

Tyrimo ribos

Tiriamos tik septynios populiariausios [1, 2] DBVS: *Microsoft SQL Server, Oracle, MySQL, PostgreSQL, DB2, Firebird* ir *Informix*. Minėtų DBVS tyrimuose nagrinėtos palaikomos operacinės sistemos, programavimo kalbos, SQL standartai, šifravimo algoritmai, duomenų tipai ir naudojami duomenų gavybos įrankiai. Duomenys tyrimui vartojami remiantis paskutinėmis išleistomis DBVS versijomis, įvertinant jų gamintojų tinklalapiuose pateiktus duomenis ir oficialiai deklaruojamus dokumentus.

Tyrimo duomenys ir rezultatai

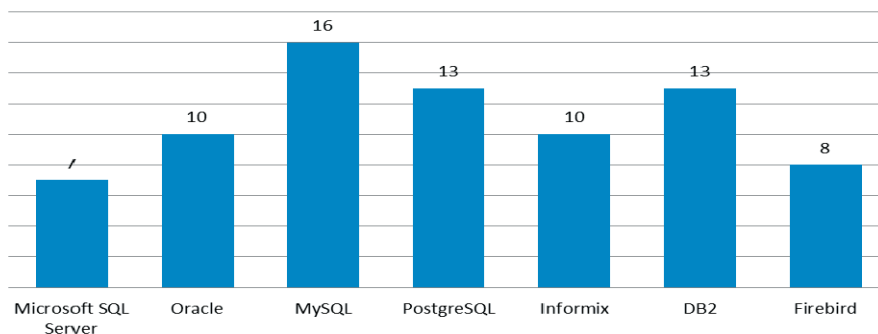
Nagrinėjamos biomedicininį duomenų saugojimui aktualiausios DBVS kokybinės ir kiekybinės

charakteristikos. Svarbiausi kokybinių DBVS charakteristikų elementai (1 lentelė ir 1 pav.): funkcio-

navimo bazė (palaikomos operacinės sistemos, SQL standartai), taikymo sritis (programavimo kalbos).

1 lentelė. *Kokybinės DBVS charakteristikos*

DBVS / Charakteristikos	Palaikomos operacinės sistemos	Programavimo kalbų palaikymas	SQL kalbų standartas
MS SQL Server [10]	Windows Server 2003, Windows Server 2008, Windows Vista, , Windows XP Service	VB, C++, C#, XML	Transact-SQL(T-SQL) ir ANSI SQL
Oracle [11]	Microsoft Windows (32-bit/64-bit);Linux x86;Linux x86-64;Solaris (SPARC) (64-bit);AIX (PPC64); HP-UX Itanium; HP-UX PA-RISC (64-bit)	Java, C, C++ ir PL/SQL	SQL-92
MySQL [12]	Windows 2000/XP/Vista, Windows Server2003/2008, Mac OS, Linux, Unix, Solaris, AIX, HP-UX Itanium, Amiga OS, BSD	C, C++, C#, Java, Perl, Python	SQL-92 , ANSI/ISO SQL
PostgreSQL [13]	Linux, AIX, BSD, HP-UX, SGI-IRIX, Mac OS X, Solaris, Windows	PL/pgSQL, PL/Perl,C, C++ bei Java	PL/pgSQL
Informix [14]	Linux, Windows, Mac OS(10.5.2 ir vėlesnė), Solaris 64-bit, AIX, BSD	Java ir COM+	SQL-92, ANSI-SQL
DB2 [15]	Windows(32bit, 64 bit), AIX, HP-UX, Linux, Linux iSeries, Linux pSeries, Linux xSeries, Linux zSeries, Solaris (x64)	SQL, XQuery, Xpath; C/C++ ; Java (SQLJ); COBOL; PHP; Perl; Python; Ruby on Rails; .NET; ADO.	SQL-92
Firebird [16]	Linux, Windows and Mac OS X	C, C++, Delphi	SQL-99, ANSI SQL.



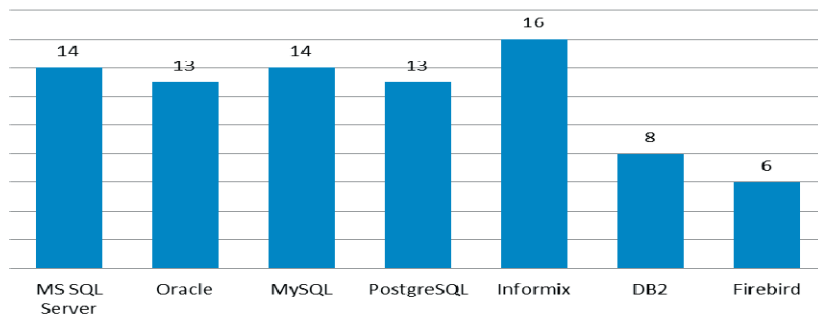
1 pav. Kokybinės charakteristikos apibendrinanti diagrama

DBVS kiekybinių charakteristikų turinys – DBVS turimi duomenų tipai ir jų įvairovė. Duomenų

tipus galima suskirstyti į tris pagrindines grupes: *skaitmeniniai, datos ir simboliniai*.

2 lentelė. *Skaitmeniniai duomenų tipai*

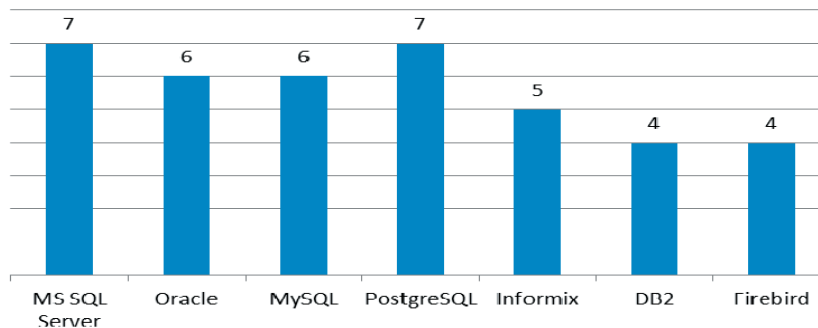
DBVS / Tipai	Sveikasis (Integer)	Pastovaus kablelio (Fix-point)	Slankaus kablelio (Floating-point)	Dvejetainis (Boolean)	Piniginis (Money)	Loginiai (True / False)
MS SQL Server [10]	5	2	3	1	2	1
Oracle [11]	5	3	4	1	-	-
MySQL [12]	7	2	3	2	-	1
PostgreSQL [13]	5	2	2	1	1	1
Informix [14]	7	2	4	1	1	1
DB2 [15]	4	-	2	1	-	1
Firebird [16]	2	2	2	-	-	-



2 pav. Skaitmeninių duomenų tipų apibendrinimas

Taigi, skaitmeninių tipų įvairovė išsiskiria *Informix*, likusioji (didžioji) dalis DBVS turi panašų kiekį skaitmeninių tipų, o mažiausiai – *Firebird*.

3 pav. pateikiamas apibendrintas datos tipų kiekis. Kaip matyti iš diagramos, dauguma DBVS turi panašų datos duomenų tipų kiekį, didelių skirtumų nėra.



3 pav. Datos duomenų tipų apibendrinimas

Nagrinėjamų DBVS simboliniai (eilučių) duomenų tipai, jų įvairovė ir kiekis pateikiami 3 lentelėje, o 4 pav. – suskaičiuotas bendras jų kiekis.

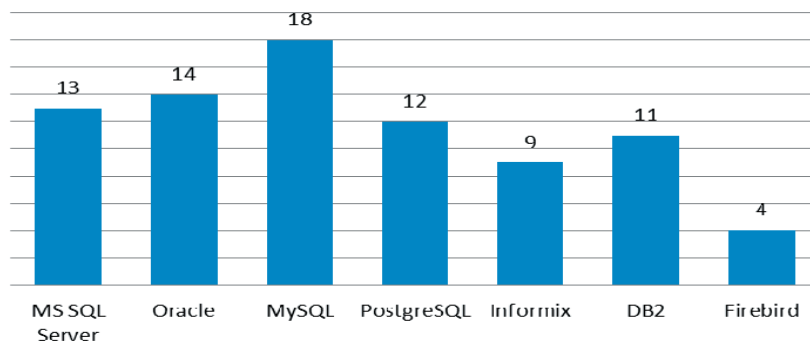
Ketvirtoje lentelėje pateikiami nagrinėjamų DBVS šifravimo algoritmai, jų įvairovė ir kiekis, o 5 pav. suskaičiuotas bendras jų kiekis.

3 lentelė. *Simboliniai duomenų tipai*

DBVS / Tipai	Fiksuoto ilgio (char, text)	Kintamo ilgio (varchar)	Binarinis (blob, binary)	Specifiniai
MS SQL Server [10]	2	4	5	2 (geometriniai, geografiniai)
Oracle [11]	2	3	6	2 (spatial: SDO_geometry, SDO_Raster)
MySQL [12]	4	6	6	2 (SET ir ENUM)
PostgreSQL [13]	1	2	1	2 (serial), 6 (geometriniai), 4 (tinklo adreso)
Informix [14]	3	4	2	-
DB2 [15]	2	1	7	1 (grafiniai)
Firebird [16]	2	1	1	-

4 lentelė. *Palaikomi šifravimo algoritmai*

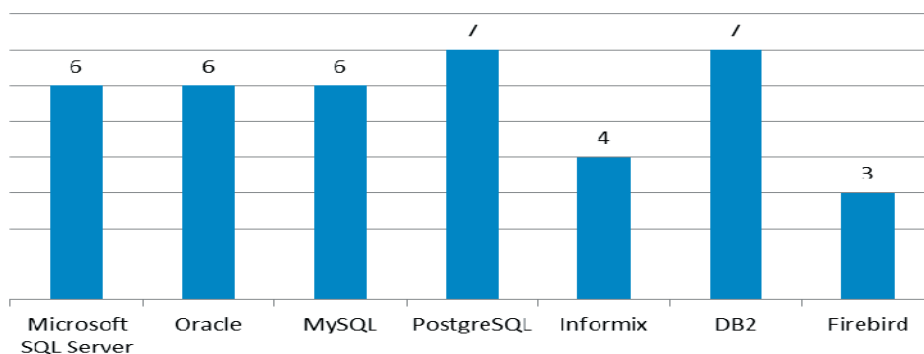
DBVS / Algoritmai	DES	AES	3DES	RC4	FIPS	DESX	MD5	SHA1	Password
MS SQL Server [10]	+	+	+	+	-	+	-	-	+
Oracle [11]	+	+	+	+	+	-	-	-	+
MySQL [12]	+	+	+	-	-	-	+	+	+
PostgreSQL [13]	+	+	+	-	-	+	+	+	+
Informix [14]	+	+	+	-	-	-	-	-	+
DB2 [15]	+	+	+	+	-	-	+	+	+
Firebird [16]	+	+	+	-	-	-	-	-	-



4 pav. Simbolinių duomenų tipų apibendrinimas

Kaip matyti iš 4 pav., simbolinių tipų įvairovė išsiskiria *MySQL*, likusioji (didžioji) dalis DBVS

turi panašų kiekį simbolinių tipų, o mažiausiai – *Firebird*.

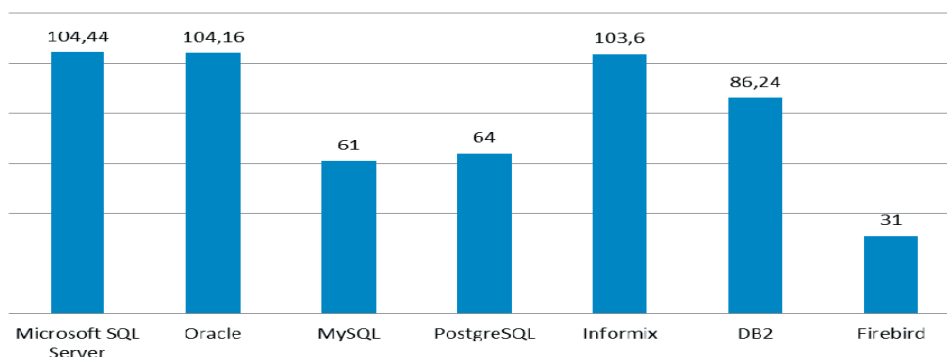


5 pav. Šifravimo algoritmų apibendrinimas

Kaip matyti iš 5 pav., šifravimo algoritmų kiekiu dauguma nagrinėjamų DBVS yra panašios, labai nesiskiria viena nuo kitos. Mažiausiai šifravimo algoritmų turi *Firebird*.

5 lentelė. *Integruoti duomenų gavybos įrankiai*

DBVS	Duomenų gavybos įrankis
MS SQL Server [10]	+
Oracle [11]	+
MySQL [12]	-
PostgreSQL [13]	-
Informix [14]	+
DB2 [15]	+
Firebird [16]	-



6 pav. DBVS charakteristikų apibendrinimas

Išvados

1. Atlikus DBVS analizę, galima teigti, kad nagrinėtos DBVS savo funkcinėmis galimybėmis

yra panašios, tačiau apibendrinus visus kiekvienos DBVS parametrus, kai kurios DBVS surinko daugiau teigiamų įvertinimų (taškų) už kitas.

2. Įvertinus rezultatus, galima pastebėti, kad pirmąja šios DBVS: *Oracle, MS SQL Server, Informix, DB2*.
3. Svarbus kriterijus, renkantis DBVS biomediciniams duomenims apdoroti yra integruotų (vidinių) duomenų gavybos įrankių palaikymas. Juos turi tik *MS SQL Server, Oracle, DB2, Firebird* ir *Informix*.
4. Medicininių duomenų saugojimo standartų palaikymas taip pat labai svarbus kriterijus, renkantis DBVS. Medicininių duomenų saugojimo standarto DICOM palaiką oficialiai deklaruoja tik *Oracle*.
5. Apibendrinus analizės rezultatus, pastebėjimus ir išvadas, galima teigti, kad *Oracle* būtų viena tinkamiausių DBVS biomedicinių duomenų saugojimui ir apdorojimui.
5. Comparison of Database Management Systems. [žiūrėta 2012-09-25]. Prieiga per internetą: <http://www.virtualmv.com/wiki/index.php?title=Comparison_of_Database_Management_Systems>.
6. Thearling. K. Information about data mining and analytic technologies [žiūrėta 2012-09-28]. Prieiga per internetą: <<http://www.thearling.com/>>.
7. How Can Data Mining Help Bio-Data Analysis? [žiūrėta 2012-10-03]. Prieiga per internetą: <<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.2.4890&rep=rep1&type=pdf>>.
8. Biomedical Databases and Data Mining. 2012. <http://www.worldscibooks.com/etextbook/7772/7772_chap01.pdf>.
9. DICOM Medical Image Support. [žiūrėta 2012-09-26]. Prieiga per internetą: <<http://www.oracle.com/technetwork/products/multimedia/overview/dicom11gr2-wp-medimgsupport-133109.pdf>>.
10. MS SQL Server Library. [žiūrėta 2012-09-18]. Prieiga per internetą: <<http://msdn.microsoft.com/en-us/library/bb545450.aspx>>.
11. Oracle Database Documentation Library. [žiūrėta 2012-09-14]. Prieiga per internetą: <<http://www.oracle.com/pls/db111/homepage>>.
12. MySQL Documentation. [žiūrėta 2012-09-17]. Prieiga per internetą: <<http://dev.mysql.com/doc/>>.
13. PostgreSQL 8.3.17 Documentation. [žiūrėta 2012-09-11]. Prieiga per internetą: <<http://www.postgresql.org/docs/8.3/static/index.html>>.
14. Date processing in Informix Dynamic Server. [žiūrėta 2012-10-02]. Prieiga per internetą: <<http://www.ibm.com/developerworks/data/library/techarticle/dm-0510roy/>>.
15. DB2 Express. [žiūrėta 2012-10-04]. Prieiga per internetą: <http://www-05.ibm.com/lt/db2/express_edition.html> [žiūrėta 2012-10-08]. Prieiga per internetą: <http://www.ipc.lt/21z/duomenys/db/integralas/files/DBVS_studija_CD_81228%20final.doc>.
16. FireBird Reference Manual. [žiūrėta 2012-09-18]. <<http://www.firebirdsql.org/en/reference-manuals/>>.

Liteatūra

1. Okunis R., Makarevičius M. Populiariausios duomenų bazių valdymo sistemos. [žiūrėta 2012-09-10]. Prieiga per internetą: <<http://www.docstoc.com/docs/20114694/SQL-Server-2000-nors-jau-n%C4%97ra-nauja-DBVS-bet>>.
2. Paliulis E., Baronienė R., 2010, Duomenų bazių valdymo sistemų analizė. *Jaunujų mokslininkų darbai*. Nr. 3 (28). P. 94–99.
3. Compare SQL Server 2008 R2, Oracle 11G R2, PostgreSQL/PostGIS 1.5 Spatial Features. [žiūrėta 2012-09-15]. Prieiga per internetą: <http://www.bostongis.com/PrinterFriendly.aspx?content_name=sqlserver2008r2_oracle11gr2_postgis15_compare>.
4. Jankevičius V. Atvirojo kodo duomenų bazių serverių analizė aptarnavimo paslaugų teikimo valdymo sistemos pagrindu. [žiūrėta 2012-09-18]. Prieiga per internetą: <http://vddb.library.lt/fedora/get/LT-eLABa-0001:E.02~2006~D_20060526_113232-16907/DS.005.0.01.ETD>.

ASSESSMENT OF SUITABILITY OF DATA BASE MANAGEMENT SYSTEMS FOR BIOMEDICAL SYSTEMS

Renata Baronienė, Egidijus Paliulis

Summary

Currently, data collection and rapid processing is a very important problem because the data quantities are always increasing. The database management system (DBMS) is an inherent part of any information system. Analysing literature we failed to find more detailed assessment of suitability of the DBMS for biomedical systems. The offer of the DBMS software is high enough, so sometimes it is difficult to decide which should be chosen. Therefore, the performance analysis and comparison of the DBMS can help to decide the appropriate choice of the DBMS in a specific use case. The aim of this work is to perform the analysis and comparison of qualitative and quantitative characteristics of the most popular relational DBMS and to investigate suitability of the DBMS for biomedical data storage and processing.

Key words: databases management systems, assessment of databases management systems, biomedical systems.

DUOMENŲ BAZIŲ VALDYMO SISTEMŲ TINKAMUMO BIOMEDICININĖMS SISTEMOMS ĮVERTINIMAS

Renata Baronienė, Egidijus Paliulis

Santrauka

Duomenų kaupimas ir spartus jų apdorojimas yra labai aktuali problema, nes duomenų kiekiai ir jų apimtys nuolat didėja. Duomenų bazių valdymo sistema (DBVS) – neatsiejama bet kurios informacinės sistemos dalis. Analizuojant literatūrą, nepavyko rasti išsamesnių DBVS tinkamumo biomedicininėms sistemoms įvertinimo. DBVS pasiūla didelė, todėl kartais sunku apsispręsti, kurią reikėtų pasirinkti. Charakteristikų tyrimas ir palyginimai gali padėti apsispręsti pasirinkant reikiamą DBVS konkrečiu jos panaudojimo atveju. Šio darbo tikslas – atlikti populiariausių reliacinių DBVS kokybinių ir kiekybinių charakteristikų analizę, jų palyginimą, siekiant išsiaiškinti, kurios būtų tinkamesnės biomedicininė duomenų saugojimui ir apdorojimui.

Prasminiai žodžiai: duomenų bazių valdymo sistemos, duomenų bazių valdymo sistemų įvertinimas, biomedicininės sistemos.

Įteikta 2012-10-22