

МЕТОДИ ЗА СЕМАНТИЧНО ОПИСАНИЕ НА ЦИФРОВИ ДАННИ В ОБЛАСТТА НА МЕДИЦИНСКИТЕ СИСТЕМИ

METHODS FOR SEMANTIC DESCRIPTION OF DIGITAL DATA IN THE FIELD OF MEDICAL SYSTEMS

Galina Bogdanova

*Institute of Mathematics and Informatics - Bulgarian Academy of Science, Sofia, Bulgaria,
g.bogdanova@gmail.com*

Todor Todorov

*Faculty of Mathematics and Informatics - St. Cyril and St. Methodius University of Veliko
Tarnovo, Veliko Turnovo, Bulgaria,
Institute of Mathematics and Informatics - Bulgarian Academy of Science, Sofia, Bulgaria,
todorvt@abv.bg*

Nikolay Noev

*Institute of Mathematics and Informatics - Bulgarian Academy of Science, Sofia, Bulgaria,
nickey@math.bas.bg*

Abstract

The article discusses modern approaches to the study of patients with cardiac complaints. Methods for semantic description of digital data in the field of medical systems and construction of an ontology for Holter monitoring system are presented. A method of efficient organization of unstructured data used in this system has been studied.

Keywords: Medical Systems, Holter Monitoring System, Ontology, Unstructured Data.

ВЪВЕДЕНИЕ

Холтер мониторинговата система (ХМС) е още един начин за изследване на пациенти с кардиологични оплаквания. Като част от съвременното специализирано медицинско оборудване ХМС борава със специфични данни за пациента и неговия медицински статус.

С това изследване се представят нови методи за обработка, анализ и съхранение на използваните от ХМС дигитални данни. Представят се възможностите за използване на семантични технологии за обработка на тези специализирани медицински данни, взаимовръзките на основните обекти в онтологията и неструктурираните бази данни с които се обработват.

Представената методология е внедрена в проекта „Изследване на приложението на нови математически методи за анализ на кардиологични данни“. Изследвания в областта на това проучване са разгледани в [1], [2], [3] и [4].

УПРАВЛЕНИЕ НА ДАННИТЕ

Неструктурираните данни (или неструктурирана информация) се отнасят за цифрови записи без предварително дефиниран модел или предварително дефинирано значение. Неструктурираната информация обикновено е текст, но може да съдържа и конкретни данни като дата, фигури, факти и т.н.

Така дефинирани данни (без модел) се отличават с много противоречия и затруднения при работа с тях:

- Терминът „неструктурирани“ означава липса на организация в данните, което е очевидно противоречие със значението на термина „данни“ (т.е. подредена и систематизирана информация за бърза работа в големи количества);
- Неструктурираните данни са такава, в които съдържанието не е подредено, но форматът и моделът на данните е стриктно структуриран;
- Основният проблем на неструктурираната информация е формата ѝ на съхранение, което да отговаря адекватно на бъдещо търсене в нея.

Примери за различни подходи за управление и съхранение на неструктурирани данни са изброените тук:

- Релационни бази от данни – данните се организират според релационния модел, като индексират и връзките да позволяват бързо, по-сложно и комплексно търсене;
- Неструктурирани бази от данни – гъвкаво организирани данни, позволяващи мащабируемост, но методите за търсене в базата са нестандартизирани и ниско ефективни;
- Онтологични и таксономични модели – обикновено използвани за допълнително логическо ниво за системите за управление на данни, използващи метаданни за насочване на заявките.

Неструктурираните (NoSQL) бази от данни са нов гъвкав модел на управление на данни, различаващ се от релационните бази от данни [5]. По този начин потребителя има прост и хоризонтално мащабируем модел на данните. Нерелационната база данни е добре организирано хранилище с ключово индексирание. Целта и е да оптимизира добавянето и извличането на записи и да подобри производителността на системата при обработка на големи количества съдържание. Нерелационната база данни има нарастваща и значителна роля в обработката мрежови ресурси в реално време и в Интернет. Тази база данни има разпределена и устойчива на грешки архитектура, като информационните обекти може да са на различни мрежови хранилища. По този начин системата е по защитена при проблем с някои от хранилищата и чрез добавяне на още сървъри. Този тип база данни се разширява хоризонтално и се използва за управление на огромна по обем информация, като производителността е преди последователността при обработката на данни.

Схема на база от данни е структурата на системата, описана на официален език, която се поддържа от системата за управление на базата данни. В релационните бази от данни схемата определя таблиците, техните полета и връзките между тях. В неструктурираната база данни схемата е колекция на група документи, всеки от които представлява ред, а колекцията отговаря на таблиците в релационни бази от данни. Колекциите нямат схема, което означава, че различни видове и структурни документи

могат да се съдържат в тях. Мащабирането по хоризонтал означава добавяне на повече възли в схемата. В неструктурираните бази от данни запазването на съдържание може да бъде много по-бързо, защото се възползва от предимството на хоризонталното мащабиране и разпределя данните между възлите.

Описаните характеристики на неструктурираните бази от данни ги прави особено подходящи за представяне на данни, свързани с Холтер мониторинговата система. С тази система е много лесно и ефективно да се опишат различните видове характеристики на различните обекти в архивите, както и връзките между тях. Също така се позволява бърз достъп до данни от множество потребители наведнъж, което е и най-честият случай на използване на такива данни.

Използването семантични средства и онтологии има приложение в различни научни области, например в дигиталното представяне на културното наследство [6]. Онтологията станаха често срещани, особено в глобалната мрежа. Онтологията предоставя на изследователите инструменти за споделяне на данни и знания в дадена област. Тя съдържа термини, понятия и връзки между тях, които биха могли да бъдат обработени автоматично. Използването на общ език за описание на структурата на информацията е ключов момент в изграждането на онтологията [7]. Онтологията по определение е формално описание на понятия, свойства и ограничения в таксонометрична структура [8].

МОДЕЛ НА ДАННИТЕ В ХМС

Целта на това проучване е да се разработи хранилище, съдържащо подробни данни, за анализ и съхранение. Хранилището да има възможности да обработва различни видове записи – текст, числа, графики, мултимедия и др., съдържанието може да не е структурирано, да се използват метаданни като съдържание.

Данните в хранилището са разнообразни с възможни повторения или липси. Разделени са няколко основни типа:

- Данни за пациента – имена, възраст, пол, адрес и други лични данни;
- Здравен статус на пациента;
- Холтерни записи – включително измервателни данни от ЕКГ, техническа информация, анамнеза, обективно състояние, етапна експертиза, история на заболяването, резултат и др.
- Технически данни;
- Информация за изпълнителя на изследването и др.

Параметри на записите получени от Холтер изследването:

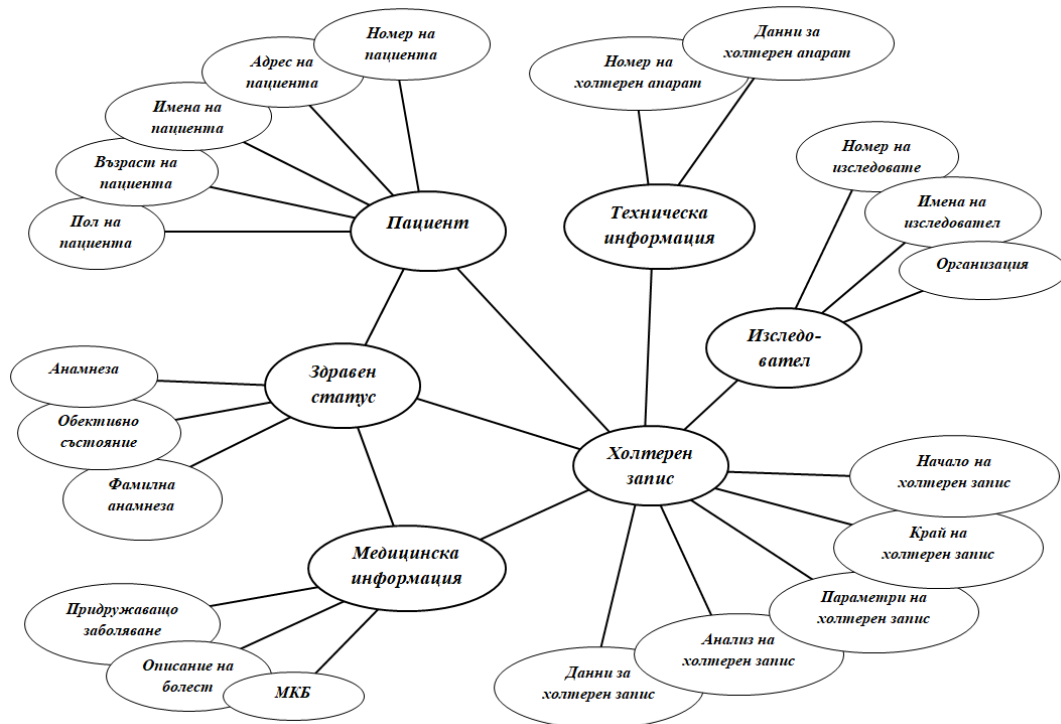
- HR: Средна сърдечна честота;
- Min HR: Минимална сърдечна честота;
- Max HR: Максимална сърдечна честота;
- VE: Систола на камерата;
- VE Pair: Камерна двойка;
- VE Run: Ритъм на камерата;
- VE Big: Биномания на камерата;
- VE Trig: Тригеминал на камерата;

- SVE: Предсърдна систула;
- SVE Pair: Предсърдна двойка;
- SVE Run: Предсърден ритъм;
- SVE Big: Предсърдна бурса;
- SVE Trig: Предсърдна тригеминална анемия;
- L: Пауза.

УПРАВЛЕНИЕ НА ДАННИТЕ В ХМС

Семантичните технологии използвани в ХМС позволяват голяма гъвкавост при обработка на различните данни получени при изследване на пациенти. За целта е необходимо да се изгради онтологична структура, описваща областта на работа на системата в това изследване.

Данните от изследването на пациенти са организирани в онтологичен модел (OWL онтология). Този модел схематично е представен на фигура 1.



Фигура. 1. Схема на ХМС онтология

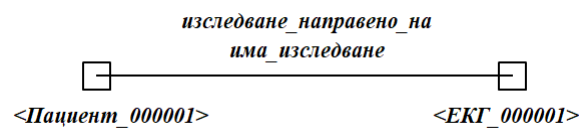
Схемата представя основните ХМС онтологични класове (Пациент, Холтерен запис, Техническа информация, Изследовател, Здравен статус и др.) с техните свойства. Използвана е платформата Protégé за изграждане на OWL елементите на ХМС онтологията [9]. Следните елементи се съдържат в OWL онтология [10], [11] и [12]:

- Елементи – обекти в конкретна област. В този случай всеки запис на изследване, пациент, състояние и др.

- Свойства – представляват връзките между елементите. Например, свойството `<изследване_направено_на>` свързва конкретния елемент `<ЕКГ_000001>` с елемента `<пациент_000001>` или свойството `<има_изследване>` може да свързва елемента `<пациент_000001>` с елемента `<ЕКГ_000001>`. Кое е пример за обратимост на свойствата. Свойствата могат да имат ограничения – забрани, функции, преходи и др.

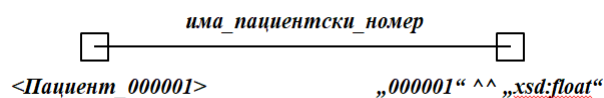
Примери за използвани свойства в ХМС онтологията:

- Обектно свойство на фигура 2, свързващо елемента `<пациент_000001>` с елемента `<ЕКГ_000001>`;



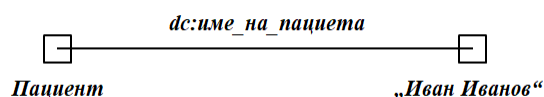
Фиг. 2. Пример за обектно свойство (object property) на ХМС онтология

- Типово свойство на фигура 3, свързващо елемента `<пациент_000001>` със символния низ „000001“, от тип „xsd:float“;



Фиг. 3. Пример за типово свойство (datatype property) на ХМС онтология

- Описателно свойство на фигура 4, свързващо класа `<Пациент>` със символния низ „Иван Иванов“;



Фиг. 4. Пример за описателно свойство (annotation property) на ХМС онтология

- Класове – шаблони за конкретните елементи. Например, класът `<ЕКГ>` представлява елементите на записите на ЕКГ изследванията в тази област на знание. Класовете могат да са подредени йерархично и схематично. Като OWL класовете съдържат набор от правила и условия които елементите трябва да поддържат за да са част от този клас.

Използвани са неструктурирана база от данни и онтологии. Основният проблем в неструктурираните бази от данни (NoSQL DB) е липсата на модел на данните и стандартизиран декларативен език на заявките. Това е причина за необходимостта от индивидуални решения с избора на модела на данните и достъпа до различни части на базата данни. За по-добро управление на търсенето в ХМС хранилището избираме подход, при който данните се съхраняват в неструктурирана база данни. Тази база данни се разширява с онтологичен модел, базиран на анотациите на ХМС онтологията. За

свързване на двата семантични слоя могат да се използват различни подходи. За да се използват стандартизирани езици за семантични заявки като SPARQL, е възможно да използването на програмни библиотеки или специфични апликационни функции за превеждане на заявки от семантично на процедурно ниво. В неструктурираната базата данни има и редица специализирани езици за семантично-процедурен достъп до данни, които в определени ситуации дават редица предимства пред SPARQL или други подобни езици. Този подход, може да бъде приложен успешно в повечето случаи на неструктурирана база от данни. В архива на обекти „камбана“, данните за отделните обекти се съхраняват в база данни на MongoDB (система за обработка на неструктурирани бази от данни) [13]. Вместо да се съхранява информацията в таблици, както при традиционните релационни бази от данни, MongoDB съхранява структурирана информация във формат JSON с динамични схеми. Това прави интегрирането на информация в определени приложения много по-лесно и бързо. Като посредник между разработения модел на онтология и данните се използва Tripod-PHP. Това е RDF система за управление на данни съхранявани в MongoDB. Тази система дава редица предимства в тези случаи, когато броя на запитванията е голям, но относително в капацитета на системата или баланса на четене:запис в базата данни е относително 10:1. Tripod-PHP не поддържа SPARQL заявки, но използва подобен DESCRIBE/SELECT език, което е добро решение с много предимства.

Използването на неструктурирана база от данни (NoSQL DB) и семантичните технологии за търсене се оказва добър избор за представяне, организиране и извличане на данни от Холтер-мониторинговата система.

ЗАКЛЮЧЕНИЕ

Съдържанието на хранилищата на медицинска информация трябва да е добре структурирано и ефективно организирано – да може информацията да се намира лесно и бързо в неструктурирана база данни. Записите от Холтерни изследвания се обработват и анализират автоматично, поради голямото количество информация която съдържат. За да се преодолее сложността на писането на сложни заявки за търсене и да се представят семантични връзки между данните, тази статия се фокусира върху интерактивното генериране на заявки чрез онтологии. Представени са изследвания и методи за семантично описание на цифрови данни в областта на медицинските системи и изграждане на онтология за Холтер мониторингова система. Проучен е метод на ефективна организация на неструктурирани данни, използван при тази система.

БЛАГОДАРНОСТИ

Научното изследване е проведено като част от проекта „Изследване на приложението на нови математически методи за анализ на кардиологични данни“ № КП-06-Н22/5 от 07.12.2018 г., финансиран от Фонд „Научни Изследвания“.

ЛИТЕРАТУРА

- [1] Bertaud-Gounot V., Duvauferrier R., Burgun A. 2012. Ontology and medical diagnosis, *Informatics for Health and Social Care*, 37(2), 51-61. DOI.org/10.3109/17538157.2011.590258
- [2] Bogdanova G., Todorov T., Noev N. 2010. Digitalization and security of bulgarian folklore heritage archive, In: *11th International Conference on Computer Systems and Technologies (CompSysTech 10) and Workshop for PhD Students in Computing on International Conference on Computer Systems and Technologies (June 17 - 18, 2010)*, Sofia, Bulgaria.
- [3] Barton A., Rosier A., Burgun A., Ethier J. 2014. The Cardiovascular Disease Ontology. *Frontiers in Artificial Intelligence and Applications*, 267, 409-414. DOI: 10.3233/978-1-61499-438-1-409
- [4] Gedzelman S., Simonet M., Bernhard D., Diallo G., Palmer P., 2005. Building an ontology of cardiovascular diseases for concept-based information retrieval, *Computers in Cardiology*, 255-258, DOI: 10.1109/CIC.2005.1588085
- [5] Georgieva-Tsaneva, G.; Gospodinova, E.; Gospodinov, M.; Cheshmedzhiev, K. 2020. Portable sensor system for registration, processing and mathematical analysis of PPG signals. *Appl. Sci.* 10:1051; <https://doi.org/10.3390/app10031051>
- [6] Georgieva-Tsaneva G, Gospodinova E, Gospodinov M, Cheshmedzhiev K. 2020. Cardio-Diagnostic Assisting Computer System. *Diagnostics (Basel)*. 5:322. doi:10.3390/diagnostics10050322
- [7] Sadalage P. J., Fowler M. 2013. NoSQL distilled: a brief guide to the emerging world of polyglot persistence, Pearson Education.
- [8] Todorov T., Bogdanova G., Shatko E., Noev N. 2018. Digital repository of cultural heritage objects, *International Journal of Mechanical Engineering and Technology (IJMET)*, 9(6), 67-73,
- [9] Gruber T., 1993. Towards principles for the design of ontologies used for knowledge sharing in formal ontology in conceptual analysis and knowledge representation,“ *International Journal of Human Computer Studies*, 43, 907-928.
- [10] Berners-Lee T., Fischetti M., 2004. *Weaving The Web: The Original Design And Ultimate Destiny Of The World Wide Web by Its Inventor*, Demco Media, 246.
- [11] Horridge M., Knublauch H., Rector A., Stevens R., Wroe C. 2004. A Practical Guide To Building OWL Ontologies Using The Protege-OWL Plugin and CO-ODE Tools Edition 1.0, Manchester: The University Of Manchester.
- [12] Antoniou G., Harmelen F. 2009. Web Ontology Language: OWL. Handbook on Ontologies. In *International Handbooks on Information Systems*, Berlin, Heidelberg, Springer, 2009.
- [13] Bartolini C. 2016. Mutating OWLs: semantic mutation testing for ontologies. In *Proceedings of the workshop on domain specific model-based approaches to verification and validation*.
- [14] Bensilamane D., Arara A., Yetongnon K., Gargouri F., BenAbdallah H. 2003. Two approaches for ontologies building: from-scratch and from existing data sources, In: *International Conference on Information Systems and Engineering, ISE*, Montreal, Canada.
- [15] Copeland R., Mongo D.B. 2013. *Applied Design Patterns: Practical Use Cases with the Leading NoSQL Database.*, O'Reilly Media, Inc..