# Ontology Driven Decision Support System Architecture for Gait Analysis

Ivan Turcin, Vladimir Ergovic, and Marko Lackovic

IBM Croatia, Miramarska 23, 10000 Zagreb, Croatia

*Abstract* — **This paper presents role of domain ontology in construction process and architecture of medical data warehouse and decision support system based on provided ontology. Domain described is in part of musculo-skeletal system of the lower limbs. In this paper we preset a way how to combine generics data mining method with data warehouse model and domain ontology in order to build intelligent system for gait analysis. This paper describes specific domain ontology of lower limbs and a ways how this ontology is mapped to data warehouse models and steps in decision support system creation.**

*Keywords* — **Data warehouse, ontology, expert system, lower limbs, data mining.**

## I. INTRODUCTION

In a number of decision support systems knowledge representation and result interpretation based on the queries over large data warehouse is important problem. Medical domain is characterized by the abundance of existing expert knowledge and practically each of its specializations has a constantly growing and interacting number of relevant guidelines. A long-term goal is knowledge representation in a form that can be used by systems supporting medical decision making. One of these domains is gait analysis or in our case musculo-skeletal system of the lower limbs. We have explored methods and tools available together with generic data mining method as basis method to follow during construction process. So far there were only few researches regarding systematic development of data warehouse models based on ontologies and only one regarding ontology design for musculo-skeletal system of the lower limbs. These researches however didn't give full description of data warehouse and data mining construction process for specific domain of human gait. One of the questions which stayed unanswered is the role of the ontology in method and model. Since the idea behind the ontology is re-usage, we have selected existing ontology as reference for our project. This paper shows integration of the domain ontology with generic data mining method and data warehouse. Our approach includes demographic data (categorical, numerical data) and time sequence data (human gait variables, trial specific variables, etc) of individual classes (joints, muscles [1, 2, 3].

## II. ONTOLOGY BACKGROUND

Ontology is an explicit and structured specification of a concepts and semantic, intelligent relation in a field, where conceptualization is also the abstract model of real phenomena. Today usage of ontologies in medicine is mainly focused on the representation and organization of medical terminologies. Physicians developed their own specialized languages and lexicons to help them store and communicate general medical knowledge and patient-related information. Such terminologies, optimized for human processing, are characterized by a significant amount of implicit knowledge. Medical information systems, on the other hand, need to be able to communicate complex and detailed medical concepts unambiguously [1,4]. This difficult task can be achieved by constructing medical domain ontology for representing medical terminology such as GALEN [4,5]. This large ontology of medical terms, anatomy and drugs is translated into OWL (Ontology Web Language) and it has about 20 MB in size. OWL is most used standard ontology languages today, and it is based on XML format. From an AI (Artificial Intelligence) perspective therefore, ontology is not only a discipline, but also the outcome of the activity of ontological analysis and modeling. This is why we can speak of "ontology of lower limbs". This ontology is example of the so-called "domain ontology", whereas "foundational ontology" represent domain-independent concepts like objects, events, processes. Some benefits of using ontologies are: interoperability, re-usage, data and knowledge integration and sharing. There is skepticism about the impact that ontologies may have on the design and maintenance of real-world systems. Considering the size and complexity of medical ontologies we see issues regarding design and maintaince of the large ontologies. That is why we focus on domain driven ontology. We examined OSMMI (Ontologie du Systeme Musculo-squelettique des Membres Inferieurs) as ontology for our project in Laboratory of Biomechanics in favor of GALEN and other more complex ontologies [1].

Y. Dekhtyar et al. (Eds.): *International Symposium on Biomedical Engineering and Medical Physics*, IFMBE Proceedings 38, pp. 78–81, 2013.
DOI: 10.1007/978-3-642-34197-7_20

## III. PROPOSED ARCHITECTURAL APPROACH

Decision support system construction has multiple approaches. Since our goal is data mining system for gait analysis, we started with generic data mining method which comprises seven steps [6]:

- Defining the issue in a precise statement.
- Defining the data model and the data requirements.
- Sourcing data from all available repositories and preparing the data.
- Evaluating the data quality
- Choosing the mining function and defining the mining run
- Interpreting the results and detecting new information.
- Deploying the results and the new knowledge.

Following the method we have defined issues regarding gait disorder based on knee injuries. This step is more oriented to specific issue and in data warehouse we want to cover more broad data which is in domain of musculo-skeletal system of the lower limbs. In second step we have examined existing conceptual data models based on medical domain ontologies. Regarding our domain and interest in gait analysis we have found muscular, skeletal and nervous data facts of the proposed model most interesting. Conceptual data model of skeletal system with fact table and related dimension is shown in Fig 1. In gait analysis this is not optimal solution because additional effort is needed to integrate multiple fact tables and querying becomes more difficult since there exists multiple fact tables for each system. Proposed model from Fig 1 also has denormalized, coarse grained fact table, not suitable for storage of large time series data.

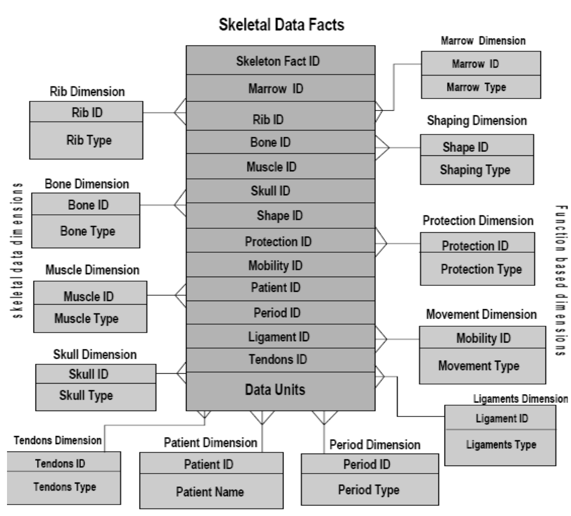


Fig. 1 Skeletal System Multidimensional Ontology (Data warehouse) [2].

In our case, average reading included 10 different parameters sampled at 10 ms for 20 seconds period of movement that produced around 20000 samples for one movement pattern including five steps, without additional temporal and demographic dimensions needed in analysis. Data warehouse is temporal model where measures are organized, collected, and represented as facts which are hold in fact tables. This structural approach enables interpretation of measures in a given context [7]. Facts become more meaningful when enriched using different dimensions, for example time, demographic and anthropometric parameters, vocabulary, etc. Since chorons used in gait analysis are placed on milliseconds scale proposed skeletal fact table becomes candidate for slow-vary dimension table. Ontology is useful input to data modeling and it gives possibility to share the common comprehension of the structure of information and it allows reuse of the implicit knowledge. Presented model although based on one of the ontologies does not satisfy gait analysis requirements since it is too generic. In construction process we used some ideas from it but we based our model on other ontology – OSMMI. OSMMI classes with relationships between them are illustrated with UML (Unified Modeling Language) stereotypes in Fig 2 [1].

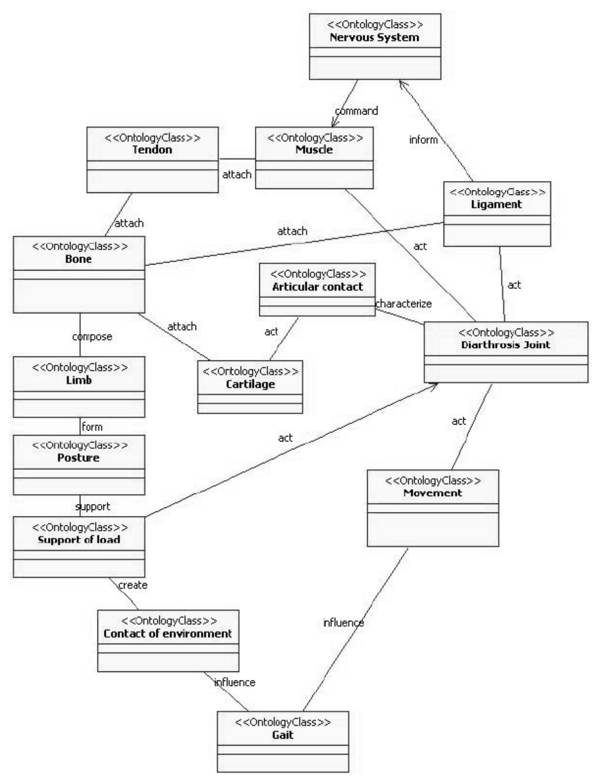


Fig. 2 General overview of the OSMMI.

OSMMI includes 14 parts (classes). These include: Nervous system, Ligament, Muscle, Tendon, Cartilage, Bone, Limb, Posture, Support of load, Diarthrosis Joint, Movement, Articular contact, Contact of environment and Gait. Semantic of OSMMI is described by relations summarized in table 1 [1].

Table 1 OSMMI relations.

| Name | Description |
|---|---|
| inform | the ligaments inform the nervous system about active signals |
| command | the nervous system command the muscles |
| attach | the muscles are attached to the bones through the tendons; the cartilages and the ligaments are attached to the bones |
| compose | limb is composed by the different bones which correspond to a particular function in gait |
| act | the muscles, the support of load, the movement, and the ligaments act on the diarthrosis joint; cartilage act the articular contact |
| influence | limbs form the correspondent posture |
| support | posture is supported by the support of load; |
| create | support of load is created by the contact of environment |
| characterize | the articular contact characterize the diarthrosis |

Classes are divided into subclasses for finer insights and classification of individual classes for gait analysis. Semantic is based on different relations that make it suitable for implementation in the traditional relational databases, but not for analytical databases. Typical data units held in fact table include measurement of several variables from trial including muscles, bones, ground reaction forces, etc. In our approach we initially include all variables which are part of the typical trial in one large fact table. Justification of this approach is that data in its nature is time series based and additional information is stored in slow vary dimension like trial since it can hold descriptive information (after operation it can be artificial hip as well). Ontology classes are all candidates for dimension tables. Initially this approach creates star like structure, but we need to keep in mind that ontology is formed from the classes and some of these classes have subclasses like posterior ligament (which have two subclasses: arched poplity, oblique poplity), patellar aileron (which again have two subclasses: patellar external aileron, patellar internal aileron) and these classes form snowflake schema. All these data can be derived and conceptualized into ontologically derived data dimensions, based on these categories, functions and activities of human body systems. Simulation models can be deduced based on metadata and data warehousing approaches. An important aspect of human gait is learning the relationship between body structures and their location as interpreted through palpation of relevant landmarks. This examination eases the task of relating knowledge, observation and palpation to make or confirm diagnoses. The conceptualization of data and information has a definite role in the logical and intelligent design of databases and data warehouses in which information is described as an inventory [8] and an asset to the system. Two types of systems are interpreted: machine based (ontology perspective) and natural systems such as human anatomy. Integrated human anatomy [9] is a natural built-in system with well connected entities or dimensions. Dimensions analogous to entities can be described, conceptualizing the relationships. Several associated data dimensions are conceptualized using ontology and stored as metadata. Volume of dimensions from real world situations can be interpreted through ontology stored in meta-data as shown with Fig 3 [6].

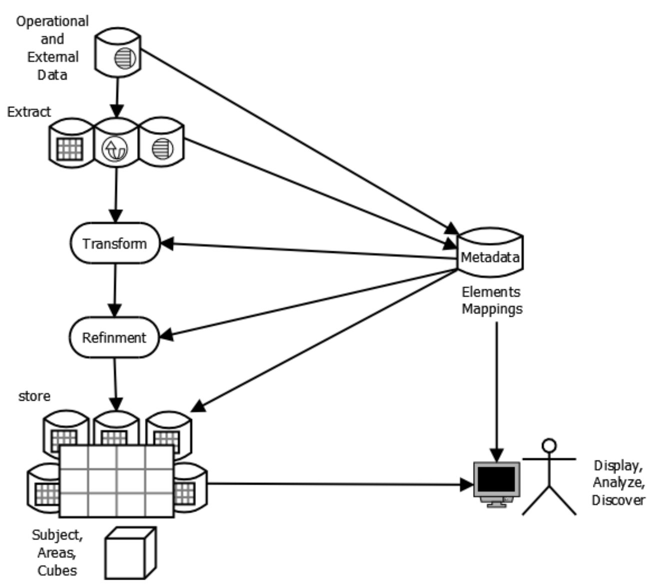


Fig. 3 Metadata in system context.

Data sourcing and preprocessing comprises the stages of identifying, collecting, filtering and aggregating (raw) data into a format required by the data models. Evaluation of the data model requires validation of chosen variables, handling of outliers and missing values and removal of redundant variables. Ontology based ETL (Extract Transform and Load) is a tool which handles both of these steps. There are several papers that recommend ontology as a tool for selecting appropriate data mining which is part of choosing the mining function step. The results from performing any type of data mining can provide a wealth of information that can sometimes be difficult to interpret and this interpretation is done by business metadata where ontology is stored. In most cases data mining creates mathematical representations of the data that we call models. These models are very important, because they not only provide a deeper insight of patients but can themselves be deployed in, or (re) used by other systems.

## IV. Discussion and conclusion

Although there are numerous obvious benefits of using ontologies in process of decision support system construction such as standardization of terms, knowledge sharing and re-usage, data warehouse model validation and support the indispensable integration of knowledge and data, practitioners should still be focus on specific domain. We were able to build data warehouse model based on OSMMI in a week avoiding numerous questions regarding data requirements and model. Our model has 3 fact tables, 18 dimension tables (based on the ontology classes, 4 which are slowly-vary dimensions) and additional meta-data tables. We re-used ontology in ETL, data refinement and in result interpretation. OWL is interesting XML based language that combined with XSLT (Extensible Stylesheet Language Transformations) enables automatic generation of dimension tables, metadata and skeleton of fact tables. Meta-data and dimension tables can be as well, automatically populated with terms and relationships implicitly stored in OWL for purpose of result interpretation. However we have found that the biggest advantage of ontology is in ETL and in result interpretation.

## Acknowledgment

The results presented are the product of scientific projects “Noninvasive measurements and procedures in bio-medicine”, “Automated motion capture and expert evaluation in the study of locomotion” and “Real-life data measurement and characterization”, realized with the support by The Ministry of Science, Education and Sports, Republic of Croatia.

## References

1. Dao T et al: Ontology of the musculo-skeletal system of the lower limbs, Proceedings of the 29th International Conference of the IEEE EMBS; 2007 Aug 23-26; Lyon, France; p. 386-389.
2. Nimmagadda S et al: Ontology based data warehouse modeling and managing ecology of human body for disease and drug prescription management, Proceedigns of the 2nd International Conference IEEE DEST 2008; 2008 Feb 26-29; Phitsanulok; p. 212-220.
3. Arnold A et al: Computer modeling of gait abnormalities in cerebral palsy: application to treatment planning, Theoretical Issues in Ergonomics Science, Vol 6, No. 3-4, May-August 2005, 305-312.
4. Rector A L et al: Unifying Medical Information using an Architecture Based on Descriptions, Proceedings of the 14th Annual Symposium on Computer Applications in Medical Care , SCAMC 90, 1990, IEEE Computer Society Press, Los Alamitos, California , pp 190-194.
5. Rector A L et al: The GALEN High Level Ontology, In: Fourteenth International Congress of the European Federation for Medical Informatics, MIE-96, Aug. 19-22 2006, Copenhagen, Denmark
6. Baragoin C et al. Mining Your Own Business in Health Care, San Jose, IBM; 2001.
7. Medved V et al: Towards a High Performance Expert System for Gait analysis, Proceedings of the 4th European Congress for Medical and Biological Engineering; 2008 Nov 23-27; Antwerp, Belgium; p. 2105-2108.
8. Tokosumi et al: Medical Ontologies as a Knowledge Repository, Proceedings of the 2007 IEEE/ICME International Conference on Complex Medical Engineering; 2007 May 23-27; Beijing; p. 487-490.
9. Moore K, Dalley A. Clinically Orientated Anatomy, Pennsylvania: Lippincott Williams and Wilkins; 1999.Smith J, Jones M Jr, Houghton L et al. (1999) Future of health insurance. N Engl J Med 965:325–329

Author: Ivan Turcin
Institute: IBM Croatia
Street: Miramarska 23
City: Zagreb
Country: Croatia
Email: ivan.turcin@hr.ibm.com